Budapest, Hungary

25-27 September 2012

18th THERMINIC

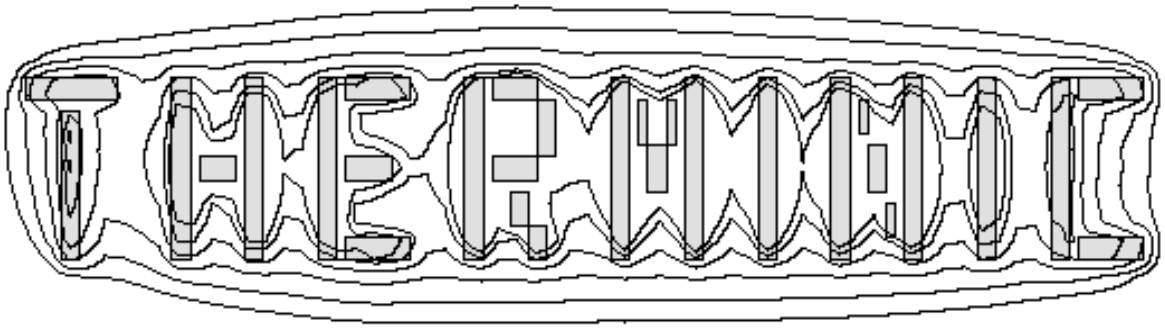
International Workshop on

Thermal investigations of ICs and Systems

Sponsored by

COLLECTION OF PAPERS PRESENTED AT THE

18th International Workshop on THERMal INvestigation of ICs and Systems



Budapest, Hungary

25–27 September 2012

Sponsored by:

IEEE COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY SOCIETY

ISBN: 978-2-35500-022-5
IEEE Catalog Number: CFP12TII-PRT

WORKSHOP COMMITTEE:

General Chair:	Márta Rencz, BME, Budapest, Hungary
Vice General Chair:	Bernard Courtois, CMP, Grenoble, France
Programme Chair:	Peter E. Raad, SMU, Dallas, USA
Programme Vice Chair:	András Poppe, BME, Budapest, Hungary

PROGRAMME COMMITTEE:

J. Altet, U. Politècnica de Catalunya, Spain
T. Baba, Nat. Metrology Institute Tsukuba, Japan
I. Barsony, MFA, Hungary
S. Bouwstra, MEMS Technical Consultancy, The Netherlands
K. Chakrabarty, Duke, USA
O. Chapuis, CIN2-CSIC, Spain
H. Chiueh, National Chiao Tung U., Taiwan
L. Codecasa, Polit. di Milano, Italy
A. Daniel, Intel, USA
R. Egawa, Tohoku U., Japan
V. Eveloy, The Petroleum Inst., UAE
S. Garimella, Purdue U, West Lafayette
Y. C. Gerstenmaier, Siemens, Germany
A. Glezer, The Georgia Institute of Technology, USA
A. Gupta, Freescale Semiconductor Inc., USA
J. Janssen, NXP Semiconductors, The Netherlands
X. Jorda, Centro Nacional de Microelectronica, Spain
W. Luiten, Philips Applied Technologies, The Netherlands
W. C. Maia, THALES-EPM, France
B. Michel, IBM Zurich, Switzerland
A. Napieralski, TU Lodz, Poland
X. Perpina, Centro Nacional de Microelectrónica, Spain
T. Persoons, Purdue University, USA
P. Rodgers, The Petroleum Inst., UAE
A. Rubio, U. Politècnica de Catalunya, Spain
S. Sapatnekar, U. of Minnesota, USA
Y. Scudeller, E.Polytech. U. Nantes, France
A. Shakouri, Purdue University, USA
M. W. Shin, Yonsei University, Korea
E. Suhir, UC Santa Cruz, USA
A. Tay, NUS, Singapore
V. Tsoi, Huawei Technologies, Sweden
B. Vandevelde, IMEC, Belgium
S. Volz, Ecole Centrale Paris, France
G. Wachutka, TU München, Germany
B. Wunderle, TU Chemnitz, Germany
J. Yu, Philips Research, The Netherlands
T. Zahner, OSRAM, Germany

 ISBN:978-2-35500-022-5

 ISBN:978-2-35500-022-5

PREFACE

The **THERMINIC Workshop** is an annual event that makes it possible for researchers from around the world to discuss essential and emerging thermal questions and best practices in the field of microelectronics. These questions are becoming ever more critical with the increasing element density of circuits and with the continual move toward nanotechnology. Dealing effectively with these trends calls for an array of thermal simulation, measurement, and management approaches. Thermal management is expected to become an increasingly dominating cost factor at all levels. The growing power dissipation and mobility of packaged microsystems raise new thermal challenges in the near horizon, making regular discussions among experts in these fields highly desirable. Finally, there is an increasing need for accurate assessment of the boundary conditions used in the analysis of electronic parts, which requires a concurrent solution of the thermal behaviour of a whole system.

Previous THERMINIC Workshops have been held in Grenoble (1995), Budapest (1996), Cannes (1997 and 1998), Rome (1999), Budapest (2000), Paris (2001), Madrid (2002), Aix-en-Provence (2003), Sophia Antipolis (2004), Belgirate (2005), Nice (2006), Budapest (2007), Rome (2008), Leuven (2009), Barcelona (2010), and Paris (2011).

This year, Workshop participants will address the "traditional" thermal management problems as well as thermal-stress-related-reliability issues, emerging novel materials, and advanced modelling and metrology systems, in both the micro- and opto-electronics fields. These issues, including various nanotechnology applications, are of significant importance and of high interest to the engineering community engaged in the field of thermal phenomena in "high-tech" systems.

The programme includes plenary invited talks by Prof. Ephraim Suhir on thermal stress failures in electronics and photonics systems, by Prof. Erik Janzén on silicon carbide as the power device for the future, and by Prof. Ali Shakouri on nanoscale electrothermal energy conversion devices. The programme also includes special sessions on: Dynamic thermometry, Smart power, Thermal stress, and the Therminator project. In total, the schedule comprises 38 oral presentations in 10 sessions, 8 poster presentations, and a vendor session.

We would like to extend our sincere appreciation to the authors for their valuable contributions to the Workshop. We are also grateful to the members of the Programme Committee who have given of their time and expertise to review and help select the papers contained herein. We hope that you will find the contents of these Proceedings beneficial to your own work.

Márta Rencz
General Chair

Bernard Courtois
Vice General Chair

Peter E. Raad
Programme Chair

András Poppe
Programme Vice Chair:

 ISBN:978-2-35500-022-5

 ISBN:978-2-35500-022-5

Table of Contents

Tuesday 25 September 2012

 ISBN:978-2-35500-022-5

Wednesday 26 September 2012

Thursday 27 September 2012

 ISBN:978-2-35500-022-5

25-27 September 2012, Budapest, Hungary

ISBN:978-2-35500-022-5

25-27 September 2012, Budapest, Hungary

Thermal Stress Modeling in Electronics and Photonics: Brief Review

Ephraim Suhir
University of California
1156 High Street, Santa Cruz, CA 95064 USA
Tel. 650-969-1530, fax. 850-968-4611, suhire@aol.com

Abstract-Major findings in the field of modeling of thermal stress in electronic and photonic systems are briefly discussed, with an emphasis on analytical modeling. The addressed areas include the role and significance of thermal stress modeling; bi-metal thermostats and bi-material assemblies models, employed in electronics and photonics engineering, and particularly die-substrate assemblies; attributes of and challenges in analytical and finite-element modeling; thermal stress in solder material and joints; global and local thermal expansion mismatch; assemblies bonded at the ends and assemblies with low modulus adhesive at the ends; design recommendations based on modeling; thermal stress in thin films fabricated on thick substrates; plastic packages of IC devices subjected to thermal loading; thermal stress induced bow and bow-free assemblies; application of the probabilistic approach; thermal stress in coated optical fibers; and thermal stress in some nano-systems. It is concluded that predictive modeling, both analytical and numerical (finite element) should be widely used in the stress-strain analyses and physical design-for-reliability of electronic and photonic systems subjected to thermal loading.

I. THERMAL LOADING AND THERMAL STRESS FAILURES

Thermal loading is the main cause of the finite lifetime of electronic and photonic systems. These systems are comprised of dissimilar materials (Fig.1) and are subjected to the change in temperature and temperature gradients [1-5]. Typical thermal stress failures are ductile rupture, brittle fracture, thermal fatigue, creep, delaminations, excessive deformations and displacements, stress relaxation (that might lead to excessive displacements in optical systems and affect the coupling efficiency), thermal shock, stress corrosion. Elevated thermal stresses and strains can lead also to functional (electrical, optical, heat transfer) failures. If the heat, produced by the chip, cannot readily escape, the thermal stress can result in failure of the p-n junction [6]. Low temperature micro-bending (buckling of the glass fiber within the low modulus primary coating) in dual-coated optical fibers, although might be insignificant enough to lead to appreciable bending stresses and delayed fracture ("static fatigue") in the silica material, can result in significant added transmission losses. Complete failure (loss) in optical coupling efficiency can occur, because of the thermally induced displacement in the lateral (often less than 0.2 micrometers) or angular (often less than a split of one percent of a degree) misalignment in the gap between two light-guides or between a light source and a light-guide. The requirements for the mechanical ("physical") behavior of the materials and structures in optoelectronics and photonics are often based on their functional (optical) performance. The requirements for the structural ("physical") robustness might be less stringent.

II. THERMAL STRESS MODELING

Thermal failures in electronic and photonic materials, components, devices and equipment can be predicted and prevented only if adequate modeling is consistently used in addition, and, desirably, prior, to experimental investigations and reliability testing [7, 8]. Accelerated life testing (ALT) [9-11], the major experimental approach in electronics and photonics engineering, cannot do without simple and meaningful predictive modeling. Pioneering work in modeling of thermal stress in bodies comprised of dissimilar materials was conducted by Timoshenko [12], based on the strength-of-material (structural analysis) approach, and by Aleck [13] who applied theory-of-elasticity technique. Both approaches were later extended and advanced by numerous investigators in application to systems employed in various fields of engineering, including the areas of electronics and photonics. It is noteworthy that Timoshenko published his classical paper in the Journal of the Optical Society of America. The state-of-the-art in the mathematical theory of thermal stresses regardless of a particular application can be found in the classical monograph by Boley and Weiner [14], and in the recently published books by Noda, Hetnarski and Tanigawa [15], and Lanin and Fedik [16]. The current state of knowledge in the field of thermal stress in electronic equipment, including modeling, has been addressed by Lau [1] and by Suhir [17].

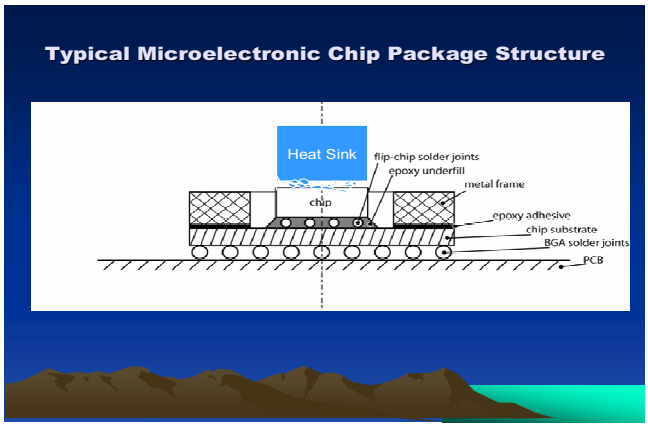


Fig.1. Typical electronic package design

 ISBN:978-2-35500-022-5

III. BI-METAL THERMOSTATS AND OTHER BI-MATERIAL ASSEMBLIES

Timoshenko addressed the thermal stress problem in application to bi-metal thermostats [12]. Many years later Timoshenko's model proved extremely useful for many applications in electronics and photonics and is referenced in numerous publications. There are nonetheless several essential differences between Timoshenko's bi-metal thermostat and assemblies employed in electronic and photonic engineering. First of all, it is imperative that a thermostat structure exhibits high flexural compliance, i.e., is prone to significant bending deformations caused by even small "external" thermal strains. Second of all, the thermostat materials are quite different of the materials employed in typical electronic and photonic structures, and so is their propensity to failure. Metals are among the best structural materials, while silicon and especially compound semiconductors (GaAs, InP, GaN, etc.) are among the worst. The most important difference between a bi-metal thermostat and an electronic or a photonic assembly has to do with the mechanical behavior of these structural elements and their response to thermal loading. Bi-metal thermostat experiences significant bow because of the deliberately introduced and significant CTE mismatch of the strip materials. Therefore the thermal stress related response of primary concern is the normal (bending) stresses in the cross-sections of the thermostat strips. The main concern in electronics and photonics assemblies is usually the interfacial shearing and peeling stresses that could lead to delaminations and cracks in the vulnerable assembly components. These stresses were not addressed by Timoshenko who considered an infinitely long assembly. The problem of thermal stresses in assemblies of finite length, including the interfacial stresses, was addressed only in the late 1970s-early 1980s, primarily in connection with the needs of the electronic industry [14-48]. Zeyfang [18], Grimado [19] and many others applied strength-of-materials approach. Kuo [23], Eischen et al [24], and others used the theory-of-elasticity treatment of the problem. Solutions based on the application of the structural analysis approach enable one to determine, often with sufficient accuracy and always with extraordinary simplicity, the thermal stresses acting in bi-material assemblies. The application of the structural analysis approach resulted in simple closed form solutions and easy-to-use formulas (see, e.g., [20-22]).

Die-substrate assemblies occupy a special place among bi-material assemblies in electronics and photonics, partially because one of the assembly components in this case is a low-expansion and a brittle semiconductor material. One of the first models [20,21] addressed adherend (die or substrate) failure, as well as the adhesive and cohesive failure of the bonding material. The obtained data, based on the concept of interfacial compliance, indicated that the elastic thermal stresses in small-size assemblies (e.g., those with chips, not exceeding, say, 5-7mm in size), increase with an increase in the assembly size and with the decrease in the compliance of the attachment layer. Timoshenko's model can be obtained as a special case of the finite-assembly-size theory, when the assembly is long and/or when the interfacial compliance is zero. The product, kl, of the parameter k of the interfacial shearing stress and half-the-assembly-size l can be used as a suitable criterion of the effect of the assembly size and the interfacial compliance on the thermal stress level (Fig.2). When the product kl is below 3.0-4.0, the assembly can be viewed as a small-size assembly, and the change in this product has a significant effect on the stress. In the region $kl \le 3.0-4.0$ the stresses of all the categories decrease with a decrease in the assembly size and an increase in the interfacial compliance. In the region $kl \ge 3.0-4.0$ the assembly size does not affect the already high stresses. The models [20-22] were applied to many problems in electronics, photonics and beyond by numerous investigators. Luryi and Suhir [27] suggested a new approach to the high quality (dislocation free) epitaxial growth of lattice-mismatched materials in heteroepitaxial semiconductor crystal grown structures. The work [27] triggered a substantial experimental effort and numerous publications. Suhir and Sullivan [28] have developed an axisymmetric version of the [20-22] model for the experimental evaluation of the adhesive strength of molding compounds used in plastic packaging of IC devices. Cifuentes [29] considered the situation when the bonding material at the assembly ends exhibits elasto-plastic behavior. Fan et al [30] used the concept of free-edge energy to predict delaminations in a bi-material system. Wen and Basaran [31] expanded the model [20] to multi-material assemblies. Klein [32] applied "Suhir's analytical method for assessment of interfacial stresses in multilayered structures" and developed "relevant formulas to the mitigation of the impact of thermoelastic stresses in diamond coated ZnS windows". Sujan et al [33] applied "the existing uniform temperature model proposed by Suhir… to account for different temperatures of the layers by incorporating a temperature ratio parameter" and come up with "a correction factor to Suhir's model". Hall et al. [34] were the first to apply the model [20] to tri-material assemblies. An analytical model for a tri-material assembly, in which all the materials are treated as "equal partners" of the assembly was developed in [35].

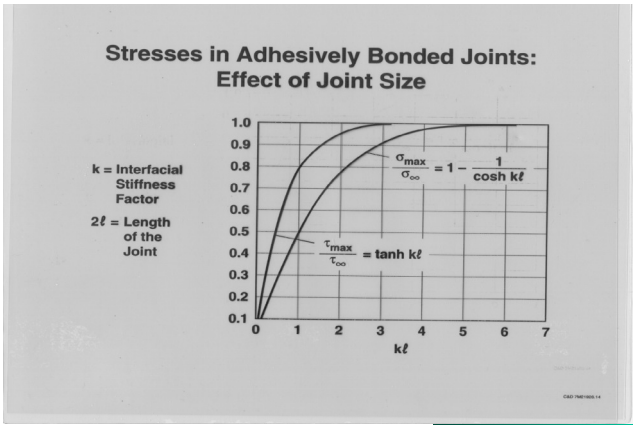


Fig.2. Effect of the assembly size and interfacial stiffness on the induced thermal stresses

IV. ANALYTICAL AND FINITE-ELEMENT MODELING

Finite-element analysis (FEA) has become, since the mid-1950s, the major resource for computational modeling in engineering, including the area of electronics and photonics (see, e.g., the pioneering work by Glaser [36]). Broad application of FEA computer programs has, however, by no means made analytical solutions unnecessary or even less important, whether exact, approximate, or asymptotic. Simple analytical relationships have invaluable advantages, because of the clarity and compactness of the obtained information and clear indication on the role of various factors affecting the behavior of the given structure. It is noteworthy that FEA was originally developed for structures with complicated geometry and/or with complicated boundary conditions, when it might be difficult to apply analytical approaches. Consequently, FEA was originally used in areas of engineering where structures of complex configuration are typical: aerospace, marine and offshore structures, some complicated civil engineering structures, etc. In contrast, a relatively simple geometry and simple configurations usually characterize electronic and photonic assemblies and structures. Such structures can be easily idealized as beams, flexible rods, circular or rectangular plates, or composites of simple geometry lending themselves to effective analytical modeling.

V. SOLDER JOINTS

Numerous experimental and FEA-based studies have been conducted for the evaluation of thermal stresses in, and prediction of the lifetime of, solders joint interconnections in assemblies subjected to temperature excursions (see, e.g., [32-42]). The majority of the today's studies address the predicted thermal fatigue life of ball-grid-array (BGA) interconnections [37-39] experiencing inelastic strains. The emphasis is on lead-free solder materials ("green electronics"). In BGA structures the thermally induced stresses and strains are caused by the mismatch of the chip package and the printed circuit board (PCB). The suggested models predict the solder material fatigue, caused by the accumulated cyclic inelastic strain. The state-of-the-art in the prediction of the lead-free solder interconnect reliability can be found in the recent book by Shangguan [39]. One effective way to reduce the thermal stress in solder joints is by employing a flex circuit [40-42]. Juskey and Carson [42] suggested that flex circuits be used as carriers for the direct chip attachment (DCA) technology.

Solder materials and joints are as important in photonics, as they are in electronics. There are, however, some specific requirements for the photonics solder materials and joints: ability to achieve high alignment, requirement for a low creep, etc. [43-44]. "Hard" (high modulus) solders (such as gold-tin eutectics) have better creep characteristics than "soft" (such as silver-tin) solders. "Hard" solders can result, however, in significantly higher thermal stresses than "soft" solders. For this reason their ability to withstand creep might be not as good as expected, not to mention the short-term reliability of the material, when time dependent responses have not manifested yet.

Solders are often used in electronic and photonics assemblies as continuous attachment layers. In general, as long as all the materials are considered elastic, the models developed for adhesively bonded joints can be applied to soldered assemblies as well. However, elevated stresses at the ends of soldered assemblies often lead to plastic deformations of the material. The problem has become particularly important in connection of using Indium and other low-yield-stress materials as bonding material of choice to bond quantum wells in GaAs lasers to metal substrates (sub-mounts). There is an obvious incentive to assess the size of the plastic zone and, to an extent possible, minimize it. This could be done based on the model [44] that considers and ideally-elastic solder below the yield point and an ideally plastic one above this point.. Clearly, the actual elasto-plastic stress condition is between the two extreme situations addressed by the elastic model [20] and the [44] model. The gap turned out to be quite narrow.

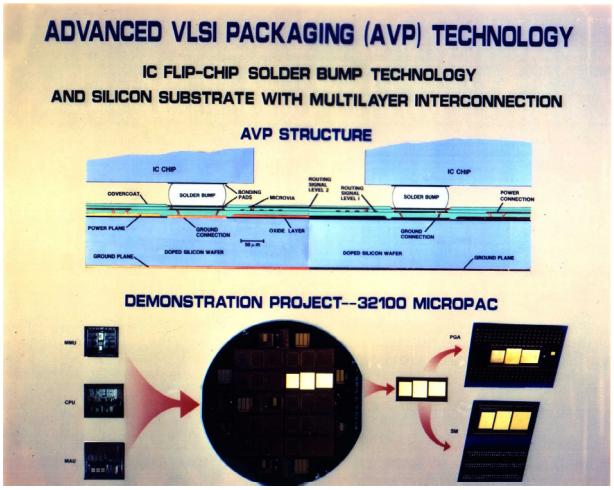


Fig.3. Bell Labs Advanced Si-on-Si VLSI package design

VI. GLOBAL AND LOCAL THERMAL MISMATCH

When the bonding layer is inhomogeneous (as is in solder bump systems, such as, e.g., the one in Fig.3), or when the components are just partially bonded or soldered to each other, both global and local mismatch loading takes place [45-50]. Local thermal expansion (contraction) mismatch loading is due to the mismatch of the dissimilar materials within the bonded or soldered region, while the global mismatch loading is caused by the mismatch of the assembly components (adherends) in the unbonded region. The interaction of the global and local thermal mismatch stresses can be summarized based on the developed models as follows:

- Interfacial shearing stresses caused by the local mismatch are antisymmetric with respect to the mid-cross-section of the bonded area. i.e., are equal in magnitude and opposite in directions (signs).
- Local shearing stresses concentrate at the ends of the bonded area and, for sufficiently long bonded joints and/or joints with stiff interfaces (thin and high-modulus adhesive layer) are next-to-zero in the mid-portion of the bonded area.
- For short-and/or-compliant bonded areas, the local

shearing stresses are more or less linearly distributed over the length of the area, and their maxima at the assembly ends can be significantly lower than the stresses in long-and/or-stiff assemblies.

- Shearing stresses caused by the global mismatch act in the same direction over the entire length of the bonded joint. This direction is such, that in the inner portions of the joints (i.e., in the portions located closer to the mid-cross-section of the unbonded region, which is also the mid-cross-section of the assembly as a whole), the total interfacial stress should be computed as the difference between the local and the global stress. In other words, the interaction of the local and the global stresses at the inner portion of the joint is always favorable, as far as the induced stresses are concerned, i.e., leads to lower total stresses. In the outer portions of the bonded areas, however, the total stress should be computed as the sum of the local and the global stress.
- In the case of small size joints and/or joints with compliant interfaces, the total stress at the outer end of the joint can be considerably larger than each of the stress categories taken separately. Since both the local and the global stresses in short-and/or-compliant joints can be very low compared to the stresses in long-and/or-stiff assemblies, the total stress can be low as well, despite the fact that, for the outer (peripheral) portions of the joints, this stress is obtained as a sum of the local and the global stresses. It is noteworthy, however, that enough "real estate" might be necessary to provide good adhesion, and therefore small size joints with compliant interfaces might not be advisable despite the possible significant stress relief.
- In the case of long-and/or-stiff joints, the global stresses concentrate at the inner edges of the joints and rapidly decrease with an increase in the distance of the given cross-section from these edges. In such a situation, as has been indicated above, the interaction of the local and global stresses is always favorable: at the inner edge of the assembly, this interaction results in the total stress, obtained as a difference between the local and the global stress, while local stress only exist at the outer edge of the assembly.
- For sufficiently long-and/or-stiff bonded joints, the magnitude of the global stress at the inner end is equal to the magnitude of the maximum local stress, so that the total shearing stress is simply zero. In other words, the state of stress is the same for the joint with a continuous bonding layer and for an assembly adhesively bonded at the ends, provided that in the latter case the joint is long enough.

The interaction of the local and the global stresses, with consideration of the effect of the coefficient of thermal expansion (contraction) of the epoxy material itself, was addressed in [51] in application to a glass fiber interconnect whose ends are epoxy bonded into capillaries. The necessity of taking into account the CTE of the adhesive material was due to the fact that the cross-sectional area of the adhesive ring was considerably larger than the cross-sectional area of the glass fiber. For this reason the longitudinal axial compliance of the low modulus adhesive ring was comparable with the compliance of the fiber itself and could not be neglected.

Interfacial stresses that concentrate at the assembly ends can be reduced by using a low modulus bonding material. There is no need to employ a low modulus material throughout the interface: it is sufficient to use it only at the assembly ends [45-47] and/or, if possible, by slanting the edges of the assembly components [46]. The stresses at the ends of polymer coated optical fibers can be reduced by using a low modulus coating at the fiber ends [40]. The mechanical behavior of assemblies with a low modulus adhesive layer at the ends is, in a sense, opposite to the situation that takes place in an assembly adhesively bonded at the ends: in an assembly with a low modulus bond (coating) at the end, it is the midportion of the assembly that is characterized by an elevated Young's modulus of the adhesive (coating), while in the case of an assembly bonded at the ends, its mid-portion is characterized by a "low" (actually, zero) Young's modulus of the "attachment".

VII. DESIGN RECOMMENDATIONS

Based on the modeling of thermal stresses in typical adhesively bonded or soldered assemblies the following general recommendations can be provided:

- equalize the in-plane and bending stiffness of the adherends, and use identical adherends, if possible;
- use as high an adherend in-plane stiffness as possible;
- use low modulus and thick bonds, having in mind, however, that if an adhesive layer is thick enough (say, 2-3mils or so in a typical adhesively bonded or soldered joint), the further increase in its thickness does not lead to an appreciable reduction in the interfacial stresses;
- as an alternative to using a low modulus adhesive throughout the joint, use such an adhesive, for lower interfacial stresses, only at the ends of the joint, i.e. in the region of high interfacial stresses, while a higher modulus adhesive could used, for better adhesion and better heat transfer, in the mid-portion of the assembly;
- vary, if possible, the adherend thickness along the assembly in a proper way and/or slant the adherends edges, if possible, for a thicker adhesive layer at the assembly ends;
- keep the stresses within the elastic range, if possible; with this in mind, lead-free solder joint interconnections, although experience higher stresses and require higher soldering temperatures, might exhibit longer fatigue life that tin-lead solders; "hard" lead-free solders might affect, however, the PCB;
- minimize peeling (in the case of multi-material and thin film structures) and axial (in the case of solder joints) stresses and strains: it is these stresses and strains that are primarily responsible for the short fatigue lifetime of the bond;
- consider using, in soldered assemblies, solders with the adequate yield-stress, for lower peeling stresses in the assembly and satisfactory fatigue life of the solder material itself.

VIII. THERMALLY MATCHED ASSEMBLIES

There is an obvious incentive to employ thermally matched materials in electronic and photonic assemblies [52-56]. Such an assembly was employed particularly in the Bell Labs Si-on-Si flip-chip (FC) design (Fig.3) [51], in a ceramic

Cerdip/Cerquad package [52], and in the advanced low-cost holographic memory storage assembly [53-56], where a highly compliant (thick and low modulus) adhesive of the silicone gel type was used. An actual solder joint in a flip-chip design was substituted in the model [51] by a short ("finite") circular cylinder subjected to the axisymmetric shear loading applied to its plane ends. The analytical predictions were in good agreement with FEA data. It has been demonstrated particularly that although the "external" thermal loading was due to the shearing deformations of the end planes, the maximum stresses and strains were the normal stresses acting in the axial direction.

The physical design of the improved structural design of a ceramic package [52] was based on the requirement that the probability that the normal stress in the cross-sections of the seal glass is always negative (i.e., the glass material is always in compression) is high and that this compression is low enough, so that the interfacial stresses at the glass/ceramic interface do not exceed an allowable level. It was concluded particularly that in order to successfully apply a probabilistic approach (see, e.g., [56]) in the designs of the type in question, customers (like, e.g., Bell Labs) should require that vendors provide information concerning not only the nominal (mean) CTE values, but also the CTE standard deviations. An inhomogeneous adhesive layer was addressed in [54, 55] in application to holographic memory assemblies. The developed models can be used to evaluate and minimize stresses, to determine the width of the peripheral ring, in which the boundaries between the dissimilar segments of the inhomogeneous bond are non-planar (and because of that cannot be used for a satisfactory optical behavior of the holographic bond), to select the most appropriate materials and grating writing techniques, etc.

IX. THIN FILMS

Typical thermal stress failures in thin films fabricated on thick substrates are interfacial delaminations (including delamination buckling), film cracking and blistering. It has been found particularly [57-59] that the thermal stress in the given film layer of a multilayer film structure is due to the thermal expansion mismatch of this layer with the substrate, and not with the adjacent film layers; that the edge stresses in the film are affected by the edge configuration (circular assemblies are somewhat "stiffer" than the rectangular ones, i.e., result in higher stresses that concentrate at a more narrow peripheral portion (ring) of the assembly); and that the stress in a thin film, which experiences bending deformations, is not affected by the assembly bow, while the assembly bow and the stresses in the substrate are strongly affected by the stresses in the film.

The effect of the lattice mismatch of semiconductor materials during crystal growth of thin Germanium (Ge) films on a thick Silicon (Si) substrate was addressed, along with the effect of thermal mismatch, by Luryi and Suhir [27]. It has been shown that by using a "tower-like" surface of the properly engineered substrate (such a surface can be achieved by high-resolution lithography, by employment of porous silicon, etc) one can indeed grow dislocation free semiconductor films of any thickness.

X. PLASTIC PACKAGES

Polymeric materials are widely used in electronic and photonic engineering. They are inexpensive and lend themselves easily to processing and mass production technologies. The reliability of these materials, however, is often insufficient for particular applications. Let us indicate just several models of thermal stress in polymeric materials and plastic packages of IC devices.

Although recent improvements in the mechanical properties of molding compounds, plastic package designs, and manufacturing technologies have resulted in substantial increase in the reliability of plastic packages, there still exists one major industry-wide concern associated with these packages – their moisture-induced failures ("popcorn" cracking). Such failures typically occur during surface mounting the packages onto PCBs using high temperature (220˚C-280˚C) reflow soldering. "Popcorn" cracking is usually attributed to the elevated pressure of the water vapor, generated due to a sudden evaporation of the absorbed moisture, however thermal stresses also play an important role, both directly, due to their interaction with the "mechanical" vapor-pressure-induced stresses in the underchip portion of the molding compound, and indirectly, by triggering the initiation and facilitating the propagation of the interfacial delaminations. It has been suggested [60] that constitutive equations, obtained as a generalization of von-Karman's equations for large deflections of plates, be used as a suitable analytical stress model for the prediction and prevention of structural failures in moisture-sensitive plastic packages. Such a generalization accounts for the combined action of the lateral pressure, caused by the generated water vapor, and the thermally induced loading. The developed model can be used for the selection of the low stress molding compounds, for comparing different package design from the standpoint of their propensity to "popcorn"-cracking, in the development of "figures-of-merit" that would enable one to separate packages that need to be "baked" and "bagged" from those that do not. It has been shown, in particular, that, from the standpoint of structural analysis, the distinction between "thick" and "thin" packages should be attributed primarily to the level of the in-plane ("membrane") normal stresses in the underchip portion of the compound: in "thick" packages this portion exhibits bending only, while in "thin" packages it is subjected to both bending and in-plane loading. The obtained data have indicated that the geometric characteristics of the package (the underchip layer thickness, chip and paddle size, etc.) have a strong effect on the package propensity to failure.

XI. THERMAL STRESS INDUCED BOW AND BOW-FREE ASSEMBLIES

Significant thermal stress induced bow can prevent further processing of the BGA packages or of thin (TSOP) plastic packages of IC devices, can lead to cracking of ceramic substrates in thin over-molded packages, or can have another adverse effect on the design and processing of plastic packages of IC devices. Ceramic substrate cracking has been observed in overmolded packages of IC devices [61, 62]. It has been shown that employment of additional (surrogate) layers can dramatically improve the situation. Two major approaches had been considered: application of a thin

 ISBN:978-2-35500-022-5

surrogate layer of a high-expansion, high-modulus and high-strength layer of a polymeric material fabricated at the opposite side of the ceramic substrate [61] and the application of a thin surrogate layer of a negative-expansion ceramics over the plastic material [62].

XII. PROBABILISTIC APPROACH

Probabilistic models might be useful and even inevitable in situations where the "fluctuations" from the mean values are significant and when the variability, change and uncertainty play a vital role [56]. Probabilistic models enable one to quantitatively assess the degree of uncertainty in various factors, which determine the performance of a product. A good illustration to this statement is the success of the design described in [52], where probabilistic analytical thermal stress modeling was applied to design and manufacture a viable and reliable ceramic package of an IC device. Additional information on the application of the probabilistic approach can be found in Refs. 66 and 67.

XIII. OPTICAL FIBERS

Various problems of thermal stress modeling in bare and coated optical silica fibers were addressed in numerous investigations. Some of them[68-92] are as follows. Low temperature micro-bending can result in substantial added transmission losses in dual-coated optical fibers. This phenomenon is due to the loss of elastic stability of the silica optical fiber within the low modulus primary coating in a dual-coating system. Based on the developed analytical stress models [69,70], it has been shown that the initial curvatures can play an important role in the low temperature behavior of a silica fiber and that, from the standpoint of the possible fiber buckling; certain curvature lengths are less favorable than others. This is because a dual-coated fiber behaves like a narrow-band filter that enhances the initial curvatures, which are close to the post-buckling configuration of the fiber, and suppresses the other curvatures. It has been shown also that the spring constant of the elastic foundation provided by the primary coating has a significant effect on the buckling conditions. Application of the mechanical approach to the evaluation of low temperature added transmission losses [71] enables one, based on the developed analytical stress model, to evaluate the threshold of the low temperature added transmission losses from purely mechanical calculations, without resorting to optical evaluations or measurements. Mechanical behavior and elastic stability of bare, polymer coated and metalized optical fiber interconnects was addressed in [72-92]. The models address bending and axial thermal stresses in an optical fiber interconnect experiencing ends offset, angular misalignment of its end cross-sections, and the effects of thermally induced compressive or tensile loading because of the thermal expansion (contraction) mismatch of the silica material with the material of the enclosure. In those cases when low buckling stresses are a problem, thicker polymer coatings can be used to improve the elastic stability of a polymer coated fiber. It has been shown particularly that, as far as the thermal stresses in the coating (metallization) are concerned, low modulus polymer coatings have significant advantages over high modulus metallizations, and should be preferred despite of their sensitivity to moisture. Shue [82.84,86,87] developed several analytical stress models for the evaluation of the mechanical (physical) behavior of double-coated and jacketed (single-coated) optical fibers at low temperature conditions. Suhir and Vuillamin[73] have demonstrated, based on the developed analytical and FEA models, that the gradient in the distribution of the CTE along one of the diameters of a glass fiber cross-section can be responsible for the undesirable “curling” phenomenon that often occurs during drawing of optical silica fibers. Suhir[74] developed analytical models for evaluating thermal stresses and strains in fused bi-conical

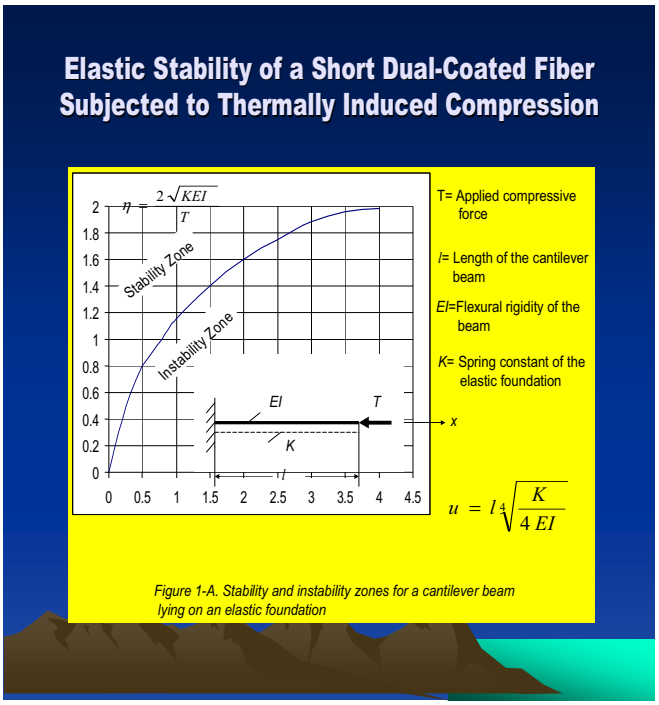


Fig.4. Nano-fiber embedded into a low modulus matrix and subjected to thermally induced compression

taper (FBT) couplers. An effective method for thermostatic compensation of temperature sensitive devices was suggested in [76] application to Bragg gratings. It has been shown that there is no need to use, for particular applications, mechanically vulnerable ceramic materials with a negative CTE: regular and more mechanically reliable materials can be successfully used for the objective in question. Analytical models for the prediction of the critical force in a dual coated optical glass fiber subjected to a combined action of the mechanical and thermal loading enable one to evaluate the elastic stability of dual-coated optical fiber interconnects in optical fiber arrays used in termination structures [88] .Thermal stress can certainly cause material reliability problems in laser chips, but it can result also in spectral broadening and “smile” (near-field non-linearity) in laser arrays. “Smile” (a slight bend of the horizontal line connecting the emitters) is a potentially disturbing property of diode bars. Smile errors can have detrimental effects on the ability to focus beams from diode bars.

 ISBN:978-2-35500-022-5

IX. NANOSYSTEMS

It has been recently demonstrated [82-92] that a newly developed nano-particle material (NPM) can make a substantial difference in the state-of-the-art of coated optical fibers: this material possesses all the merits of the polymer coated and metallized optical fibers without having their drawbacks. Several analytical predictive models for the evaluation of thermally induced and mechanical stresses in carbon-nano-tube (CNT)/carbon-nano-fiber (CNF) arrays were developed by Zhang et al [93, 94] in application to nano-systems intended for heat transfer in high power IC devices. The diagram in Fig.4 can be used to determine if a nano-rod embedded into a low-modulus high-expansion matrix might buckle at low temperature conditions. The same diagram can be used to determine the elastic stability of a dual-coated optical fiber interconnect subjected to the combined action of the thermally induced and external compressive loading. This is, actually, the beauty of an analytical stress model whose formalism can be useful in quite different fields of engineering.

X. CONCLUSION

The role and attributes of, as well as the state-of-the art and major findings in, the analytical thermal stress modeling in electronic and photonic engineering, has been addressed. It has been shown that such modeling is an effective and indispensable tool for the prediction of stresses arising in electronics and photonics materials, structures, packages and systems, subjected to thermal loading. Analytical modeling should be used, whenever possible, in addition to finite-element modeling in any effort aimed at the analysis and rational design of a viable, reliable and cost-effective electronic and photonic system.

REFERENCES

[1] J.H.Lau, ed., "Thermal Stress and Strain in Microelectronics Packaging", Van-Nostrand Reinhold, New York, 1993.

[2] E. Suhir, R.C.Cammarata, D.D.L.Chung, M.Jono, "Mechanical Behavior of Materials and Structures in Microelectronics", Materials Research Society Symposia Proceedings, vol.226, 1991.

[3] E. Suhir, M. Fukuda, C. R. Kurkjian, eds., "Reliability of Photonic Materials and Structures", Materials Research Society Symposia Proceedings, vol. 531, 1998.

[4] E. Suhir, "Thermal Stress Failures in Microelectronic Components - Review and Extension", in A. Bar-Cohen and A. D. Kraus, eds., "Advances in Thermal Modeling of Electronic Components and Systems", Hemisphere, New York, 1988.

[5] E. Suhir, "Thermal Stress Failures in Microelectronics and Photonics: Prediction and Prevention", Future Circuits International, issue 5, 1999

[6] G.A.Lang et al., "Thermal Fatigue in Silicon Power Devices", IEEE Transactions on Electron Devices, vol. 17, 1970

[7] M. Schen, H. Abe, and E. Suhir, eds. "Thermal and Mechanical Behavior and Modeling", ASME, AMD-Vol,1994

[8] E. Suhir, "Thermomechanical Stress Modeling in Microelectronics and Photonics", Electronic Cooling, vol.7, No.4, 2001

[9] E. Suhir, "Accelerated Life Testing (ALT) in Microelectronics and Photonics: Its Role, Attributes, Challenges, Pitfalls, and Interaction with Qualification Tests", Keynote address at the SPIE's 7-th Annual International Symposium on Nondestructive Evaluations for Health Monitoring and Diagnostics, 17-21 March, San Diego, CA, 2002

[10] E.Suhir, "Reliability and Accelerated Life Testing", Semiconductor International, February1, 2005.

[11] E. Suhir, "Modeling of the Mechanical Behavior of Microelectronic and Photonic Systems: Attributes, Merits, Shortcomings, and Interaction with Experiment", Proceedings of the 9-th Int. Congress on Experimental Mechanics, Orlando, Fl., June 5-8, 2000

[12] S.Timoshenko, "Analysis of Bi-Metal Thermostats", Journal of the Optical Society of America, vol.11, 1925.

[13] B.J.Aleck, "Thermal Stresses in a Rectangular Plate Clamped Along an Edge", ASME Journal of Applied Mechanics, vol.16, 1949.

[14] B.A.Boley and J.H.Weiner, "Theory of Thermal Stresses", Quantum Publishers, New York, 1974.

[15] N. Noda, R.B.Hetnarski, Y.Tanigawa, "Thermal Stress", 2-nd ed., Taylor and Francis, Ltd., 2004.

[16] A. Lanin and I.Fedik, "Thermal Stress Resistance of Materials", Springer, 2008

[17] E.Suhir, "Analytical Thermal Stress Modeling in Electronic and Photonic Systems", ASME Applied Mechanics Reviews, invited paper, vol.62, No.4, 2009.

[18] R. Zeyfang, "Stresses and Strains in a Plate Bonded to a Substrate: Semiconductor Devices", Solid State Electronics, vol.14, 1971

[19]P.B.Grimado, "Interlamina Thermo-Elastic Stresses in Layered Beams" , J. Therm. Stresses, vol. 1, 1978.

[20]E. Suhir, "Stresses in Bi-Metal Thermostats", ASME Journal of Applied Mechanics, vol. 53, No. 3, Sept. 1986.

[21] E. Suhir, "Die Attachment Design and Its Influence on the Thermally Induced Stresses in the Die and the Attachment", Proc. of the 37th Elect. Comp. Conf., IEEE, Boston, Mass., May 1987.

[22] E. Suhir, "Interfacial Stresses in Bi-Metal Thermostats", ASME Journal of Applied Mechanics, vol. 56, No. 3, September 1989.

[23] A. Kuo, "Thermal Stresses at the Edge of a Bimetallic Thermostat", ASME Journal of Applied Mechanics, vol.56,1989.

[24] J.W. Eischen, C.Chung and J.H.Kim, "Realistic Modeling of the Edge Effect Stresses in Bimaterial Elements", ASME Journal of Electronic Packaging, vol.112, No.1, 1990.

[25] E. Suhir, "Structural Analysis in Microelectronic and Fiber Optic Systems", Basic Principles of Engineering Elasticity and Fundamentals of Structural Analysis", Van Nostrand Reinhold, New York, 1991.

[26] S. E. Yamada, "A Bonded Joint Analysis for Surface Mount Components", ASME Journal of Electronic Packaging, vol.114, No.1, 1992

[27] S. Luryi and E. Suhir, "A New Approach to the High-Quality Epitaxial Growth of Lattice - Mismatched Materials", Applied Physics Letters, vol. 49, No. 3, July 1986.

[28] E. Suhir and T. M. Sullivan, "Analysis of Interfacial Thermal Stresses and Adhesive Strength of Bi-Annular Cylinders", Int. Journal of Solids and Structures, vol. 26, No. 6, 1990.

[29] A.O.Cifuentes, "Elastoplastic Analysis of Bimaterial Beams Subjected to Thermal Loads", ASME Journal of Electronic Packaging, vol.113, No.4, 1991.

[30] H.B.Fan, M.F.Yuen, E. Suhir, "Prediction of Delamination in a Bi-Material System Based on Free-Edge Energy Evaluation", 53-rd ECTC Proceedings, 2003.

[31] Y.Wen and C.Basaran,"An Analytical Model for Thermal Stress Analysis of Multi-layered Microelectronics Packaging", 54-th ECTC, 2004.

[32] C.A. Klein, "Thermal Stress Modeling for Diamond-Coated Optical Windows", 22nd Annual Boulder Damage Symposium, Boulder, CO, Oct. 24-26, 1990.

[33] D. Sujan et al, "Engineering Model for Interfacial Stresses of a Heated Bimaterial Structure with Bond Material Used in Electronic Packages", IMAPS Journal of Microelectronics and Electronic Packaging, vol. 2, No.2, 2005.

[34] P.M. Hall, et al., "Strains in Aluminum-Adhesive-Ceramic Trilayers", ASME Journal of Electronic Packaging, vol.112, No.4, 1990.

[35] E. Suhir, "Analysis of Interfacial Thermal Stresses in a Tri-Material Assembly", Journal of Applied Physics, vol.89, No.7, 2001

[36] J.C. Glaser, "Thermal Stresses in Compliantly Joined Materials", ASME Journal of Electronic Packaging, vol.112, No.1, 1990.

[37] J.Lau (ed.), "Ball Grid Array Technology", McGraw-Hill, 1995.

[38] J.S.Hwang, "Modern Solder Technology for Competitive Electronics Manufacturing", McGraw-Hill, New York, 1996

[39] D. Shangguan (ed), "Lead-Free Solder Interconnect Reliability", ASM, 2005.

[40] E. Suhir, "Stress Relief in Solder Joints Due to the Application of a Flex Circuit", ASME Journal of Electronic Packaging, vol. 113, No. 3, 1991.

[41] E. Suhir, "Flex Circuit vs Regular" Substrate: Predicted Reduction in the Shearing Stress in Solder Joints", Proc. of the 3-rd Int. Conf. on Flexible Circuits FLEXCON 96, San-Jose, CA, Oct. 1996.

[42] F. Juskey and R. Carson, "DCA on Flex: A Low Cost/Stress Approach", in E.Suhir et al, eds., "Advances in Electronic Packaging", ASME Press, 1997
[43] E. Suhir, "Thermally Induced Stresses in an Optical Glass Fiber Soldered into a Ferrule", IEEE/OSA Journal of Lightwave Technology, vol. 12, No. 10, 1994.
[44] E.Suhir, "Interfacial Thermal Stresses in a Bi-Material Assembly with a Low-Yield-Stress Bonding Layer", Modeling and Simulation in Materials Science and Engineering, vol. 14, 2006
[45] E. Suhir, "Thermal Stress in a Bi-Material Assembly Adhesively Bonded at the Ends", Journal of Applied Physics, vol. 89, No.1, 2001.
[46] E. Suhir, "Bi-Material Assembly Adhesively Bonded at the Ends and Fabrication Method", US Patent #6,460, 753, 2002
[47] E. Suhir, "Predicted Thermal Mismatch Stresses in a Cylindrical Bi-Material Assembly Adhesively Bonded at the Ends", ASME Journal of Applied Mechanics, vol.64, No. 1, 1997.
[48] E. Suhir, "Thermal Stress in an Adhesively Bonded Joint with a Low Modulus Adhesive Layer at the Ends", Journal of Applied Physics Journal, April 2003.
[49]E. Suhir, "Thermal Stress in a Polymer Coated Optical Glass Fiber with a Low Modulus Coating at the Ends", Journal of Materials Research, vol. 16, No. 10, 2001
[50] E. Suhir, "Coated Optical Fiber", US Patent #6,647,195, 2003.
[51] E. Suhir, "Axisymmetric Elastic Deformations of a Finite Circular Cylinder with Application to Low Temperature Strains and Stresses in Solder Joints", ASME Journal of Applied Mechanics, vol. 56, No. 2, 1989.
[52] E. Suhir and B. Poborets, "Solder Glass Attachment in Cerdip/Cerquad Packages: Thermally Induced Stresses and Mechanical Reliability", ASME Journal of Electronic Packaging, vol. 112, No. 2, 1990.
[53] E. Suhir, "Adhesively Bonded Assemblies with Identical Nondeformable Adherends: Predicted Thermal Stresses in the Adhesive Layer", Composite Interfaces, vol.6, No.2, 1999
[54] E. Suhir, "Adhesively Bonded Assemblies with Identical Nondeformable Adherends and Inhomogeneous Adhesive Layer: Predicted Thermal Stresses in the Adhesive", Journal of Reinforced Plastics and Composites, vol.17, No.14, 1998
[55] E. Suhir, "Adhesively Bonded Assemblies with Identical Nondeformable Adherends and "Piecewise Continuous" Adhesive Layer: Predicted Thermal Stresses and Displacements in the Adhesive", Int. Journal of Solids and Structures, vol.37, 2000
[56] E. Suhir, "Applied Probability for Engineers and Scientists", McGraw Hill, New York, 1997.
[57] E. Suhir, "An Approximate Analysis of Stresses in Multilayer Elastic Thin Films", ASME Journal of Applied Mechanics, vol. 55, No. 3, 1988.
[58] E. Suhir, "Approximate Evaluation of the Elastic Thermal Stresses in a Thin Film Fabricated on a Very Thick Circular Substrate", ASME Journal of Electronic Packaging, vol. 116, No. 3, 1994.
[59] E. Suhir, "Predicted Thermally Induced Stresses In, and the Bow of, a Circular Substrate/Thin-Film Structure", Journal of Applied Physics, vol.88, No.5, 2000
[60] E. Suhir, "Failure Criterion for Moisture-Sensitive Plastic Packages of Integrated Circuit (IC) Devices: Application of von-Karman Equations with Consideration of Thermoelastic Strains", Int. Journal of Solids and Structures, vol. 34, No. 12, 1997
[61] E. Suhir and J. Weld, "Electronic Package with Reduced Bending Stress", US Patent #5,627,407, 1997.
[62] E. Suhir, "Arrangement for Reducing Bending Stress in an Electronics Package", U.S. Patent #6,180,241, 2001
[63] E. Suhir, "Device and Method of Controlling the Bowing of a Soldered or Adhesively Bonded Assembly," US Patent #6,239,382, 2001.
[64] E. Suhir, "Bow Free Adhesively Bonded Assemblies: Predicted Stresses", Electrotechnik & Informationtechnik, 120 (6), June 2003.
[65] E. Suhir, "Bow-Free Assemblies: Predicted Stresses", Therminic'2004, Niece, France, Sept. 29-Oct.1, 2004
[66] E. Suhir, "Probabilistic Approach to Evaluate Improvements in the Reliability of Chip-Substrate (Chip-Card) Assembly", IEEE CPMT Transactions, Part A, vol. 20, No. 1, 1997.
[67] E. Suhir, "Thermal Stress Modeling in Microelectronics and Photonics Packaging, and the Application of the Probabilistic Approach: Review and Extension", IMAPS International Journal of Microcircuits and Electronic Packaging, vol.23, No.2, 2000
[68] E. Suhir, "Stresses in Dual - Coated Optical Fibers", ASME Journal of Applied Mechanics, vol. 55, No. 10, 1988.
[69] E. Suhir, "Effect of Initial Curvature on Low Temperature Microbending in Optical Fibers", IEEE/OSA Journal of Lightwave Technology, vol. 6, No. 8, 1988.
[70] E. Suhir, "Spring Constant In the Buckling of Dual-Coated Optical Fibers", IEEE/OSA Journal of Lightwave Technology, vol. 6, No. 7, 1988.
[71] E. Suhir, "Mechanical Approach to the Evaluation of the Low Temperature Threshold of Added Transmission Losses in Single-Coated Optical Fibers", IEEE/OSA Journal of Lightwave Technology, vol. 8, No. 6, 1990.
[72] E. Suhir, "Optical Fiber Interconnect with the Ends Offset and Axial Loading: What Could Be Done to Reduce the Tensile Stress in the Fiber"?, Journal of Applied Physics, vol.88, No.7, 2000
[73] E. Suhir, and J.J. Vuillamin, Jr, "Effects of the CTE and Young's Modulus Lateral Gradients on the Bowing of an Optical Fiber: Analytical and Finite Element Modeling", Optical Engineering, vol. 39, No. 12, 2000
[74] E. Suhir, "Predicted Stresses and Strains in Fused Biconical Taper Couplers Subjected to Tension", Applied Optics, vol. 32, No. 18, 1993.
[75] M. Muraoka, "The Maximum Stress in Optical Glass Fibers Under Two-Point Bending", ASME Journal of Electronic Packaging, vol.123, March 2001
[76] E. Suhir, "Apparatus and Method for Thermostatic Compensation of Temperature Sensitive Devices", US Patent #6,337,932, 2002
[77] E.Suhir, "Optical Fiber Interconnects Having Offset Ends with Reduced Tensile Strength and Fabrication Method", US Patent #6,606,434, 2003
[78] E. Suhir, "Coated Optical Glass Fiber", US Patent #6,647,195, 2003.
[79] Gebizioglu and I.M.Plitz, "Self-Stripping of Optical Fiber Coatings in Hydrocarbon Liquids and Cable Filling Compounds", Optical Engineering, Vol.30, No.6, 1991
[80] E. Devadoss, "Polymers for Optical Fiber Communication Systems", Journal of Scientific and Industrial Research, Vol.51, No.4, 1992
[81] S.T. Shiue and S.B. Lee, "Thermal Stresses in Double-Coated Optical Fibers at Low Temperature", Journal of Applied Physics, Vol.72, No.1, 1992
[82] F. Cocchini, "Double-Coated Optical Fibers Undergoing Temperature Variations-The Influence of the Mechanical Behavior on the Added Transmission Losses", Polymer Engineering and Science, Vol.34, No.5, 1994
[83] S.T. Shiue, "Thermal Stresses in Tightly Jacketed Double-Coated Optical Fibers at Low Temperature", Journal of Applied Physics, Vol.76, No.12, 1994
[84] P. Ostojic, "Stress Enhanced Environmental Corrosion and Lifetime Prediction Modeling in Silica Optical Fibers", Journal of Materials Science, Vol.30, No.12, 1995
[85] S.T. Shiue, "The Spring Constant in the Buckling of Tightly Jacketed Double-Coated Optical Fibers", Journal of Applied Physics, Vol.81, No.8, 1997
[86] S.T. Shiue and W.H. Lee, "Thermal Stresses in Carbon Coated Optical Fibers at Low Temperature", Journal of Materials Research, Vol.12, No.9, 1997
[87] E. Suhir, "Elastic Stability of a Dual-Coated Optical Fiber of Finite Length", Journal of Applied Physics, vol.102, No.5, 2007
[88] E. Suhir, "Elastic Stability of a Dual-Coated Optical Fiber with a Stripped Off Coating at Its End", Journal of Applied Physics, vol. 102, No.4, 2007
[89] D.Ingman, V.Ogenko, E.Suhir, A.Glista, "Moisture Resistant Nano-Particle Material and Its Applications", US Patent #7,321,714B2, 2008
[90] D. Ingman and E.Suhir, "Optical Fiber with Nano-Particle Overclad", US Patent, #7,162,138 B2, 2007
[91] D. Ingman and E.Suhir, "Optical Fiber with Nano-Particle Cladding", US Patent, #7,162,137 B2, 2007
[92] E. Suhir, "Fiber-Optics Structural Mechanics and Nano-Technology Based New Generation of Fiber Coatings: Review and Extension", in E. Suhir, CP Wong, YC Lee, eds. "Micro- and Opto-Electronic Materials and Structures: Physics, Mechanics, Design, Packaging, Reliability", Springer, 2007
[93] Y. Xu, Y. Zhang, E. Suhir, and X. Wang, " Thermal Properties of Carbon Nanotube Array for Integrated Circuits Cooling" Journal of Applied Physics,100, 074302, 2006
[94] Y. Zhang, Y.Xu and E.Suhir, "Effect of Rapid Thermal Annealing (RTA) on Thermal Properties of Carbon Nanofibre (CNF) Arrays", Journal of Physics D: Applied Physics ,39, 2006, pp.4878-4885
[95] E. Suhir, "New Nano-Particle Material (NPM) for Micro- and Opto-Electronic Packaging Applications", IEEE Workshop on Advanced Packaging Materials, Irvine, March 2005.

25-27 September 2012, Budapest, Hungary

Minimizing Thermally Induced Interfacial Shearing Stress in a Thermoelectric Module

Amirkoushyar Ziabari[1,2], Ephraim Suhir[2], and Ali Shakouri[1,2]
[1] Birck Nanotechnology Center, Purdue University, 1205 West State Street, West Lafayette, IN 47907-2057, USA
[2] Baskin School of Engineering, University of California Santa Cruz, CA 95064-1077, USA
Email to: aziabari@purdue.edu

Abstract
The problem of minimizing the level of the thermally induced interfacial shearing stress in a Thermo-Electric Module (TEM) is addressed using analytical and finite-element-analysis (FEA) based modeling. The maximum stress is calculated for different leg sizes. Good agreement between the analytical and FEA predictions has been found. It is concluded that the shearing stress can be effectively minimized by using thinner legs with compliant interfaces.

I. Introduction

Thermal ("internal") loading, and the resulting stresses and deformations can be defined as those that are associated with the change in temperature, and/or as those, which depend on thermomechanical properties of the employed materials. Thermally induced stresses and strains can be due to dissimilar materials that expand/contract at different rates during temperature excursions, and/or to the nonuniform distribution of temperature (e.g., temperature gradient).
Examples of the devices that are prone to be inversely affected by the thermal stresses during their operation are Thermo-Electric Modules (TEMs). TEMs are receiving increasing attention with the development of energy technology and thermoelectric materials [1,2]. Although most of the research focused on improving thermoelectric material properties [3], few researchers pay attention to the mechanical properties of these materials as well as mechanical stability of module under operational condition [4, 5].
In this work, an analytical model is developed to evaluate the level of thermal stresses in a simplified two-leg TEM design. The goal of this analysis is to develop and employ the simplest and most physically meaningful model for the assessment of the role of the finite size (length) of the bonded areas (legs) on the maximum stress in the assemblies of the TEM type. Finite Element Modeling (FEM) software, ANSYS 14 [6], is used to confirm the analytical model presented in this work. The numerical examples are expected to guide the design of a mechanically robust TEM structure.
The rest of this paper is organized as follows. In section II the analytical model is described. In section III case studies are presented and the results for these case studies are discussed. The paper concludes in section IV with a summary and possible future plans.

II. Analytical modeling

A. Interfacial compliance

Analytical modeling has been based on the approach using the interfacial compliance concept [7, 8]. In accordance with this concept, the longitudinal interfacial compliance of a strip subjected to shear loading applied to its long edge (Fig. 1) was found from the following approximate formula for the displacements of this edge:

$$u_0 = -\frac{1-\nu^2}{Ehb}\int_0^x Q(\xi)d\xi + \kappa q(x) \quad (1)$$

Here E and ν are the modulus of elasticity and Poisson's ratio for the strip material, κ is the interfacial compliance, h is the thickness of the strip, b is its width, $q(x)$ is the shear force per unit strip length, and $Q(x)$ is the force at the x cross section. The interfacial compliance κ can be calculated as

$$\kappa = \frac{\sum_k \gamma_k M(u_k) sin\alpha_k x}{Eb\sum_k \alpha_k \gamma_k sin\alpha_k x} \quad (2)$$

where

$$M(u_k) = \left(1+\frac{\nu}{2}\right)\Big[(3-\nu - (1+\nu)u_k cotanh\, u_k)cotanh\, u_k + (1+\nu)u_k - \left(\frac{2(1-\nu)}{u_k}\right)\Big]$$

$$\&\quad G = \frac{E}{2(1+\nu)}, \quad u_k = \alpha_k h = \frac{k\pi}{2}\frac{h}{l},$$

$$\&\quad \gamma_k = \frac{2}{\alpha_k l}\int_0^l \tau_0(x) sin\alpha_k x dx$$

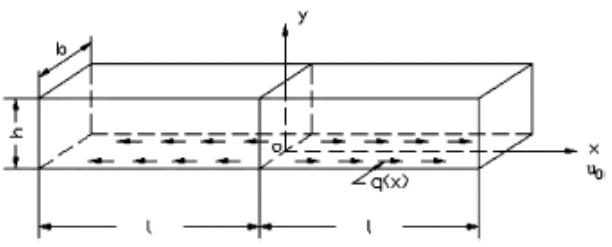


Fig.1. Elongated strip subjected to shear loading.

 ISBN:978-2-35500-022-5

We have shown (Fig.2) that the κ value is dependent on the geometry and the material properties of the structural element in question and is only slightly dependent on the applied shearing load. The general expression (2) can be approximated by simplified relationships [7, 8]

$$\kappa = \begin{cases} \frac{h}{3Gb} & \frac{h}{l} < 0.5 \\ \frac{3-\nu}{2\pi b}\frac{l}{G} & \frac{h}{l} > 2 \end{cases} \quad (3)$$

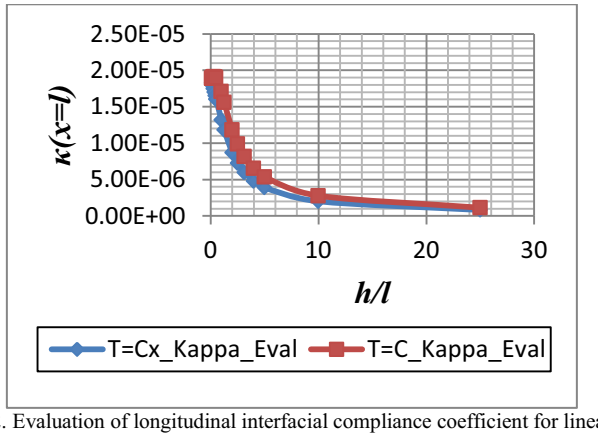


Fig.2. Evaluation of longitudinal interfacial compliance coefficient for linear shear load distribution (blue curve) as well as uniform shear load distribution (red curve).

Accordingly, in our further analysis we use the formulas (3) when the h/l ratio is below 0.5 or above 2, and the general formula (2), when the h/l ratio is between 0.5 and 2.0. The κ values computed for different materials employed in TEM are shown in Fig. 3, assuming that the shear loading is uniformly distributed along the interface.

B. *Shear Stress in a Thermo-Electric Module*

For the TEM schematically shown in Fig.2, the longitudinal interfacial displacements can be predicted using the following approximate equations

$$u_1(x) = -\alpha\Delta t_1 x + \lambda_1 \int_0^x T(\xi)d\xi - \kappa_1\tau(x)$$
$$u_2(x) = -\alpha\Delta t_2 x - \lambda_1 \int_0^x T(\xi)d\xi + \kappa_1\tau(x) \quad (4)$$

which are similar to those used in [7], where, however, dissimilar adherend materials were considered. In the equations (4), α is the coefficient of thermal expansion (CTE) of the components' (adherends') material, Δt_1 and Δt_2 are the change in temperature of the components (from the manufacturing temperature to the operation temperature), λ_1 is the axial compliance of one of the bonded components, E_1 and ν_1 are the elastic (Young's) modulus and Poisson's ratio of the component's material, h_1 is the component thickness, *T(x)* is the axial thermally induced force acting in the cross-sections of the components, τ(x) is the interfacial shearing stress, and κ_1 is the longitudinal interfacial compliance. The latter is due to both the material of the components and the material(s) of the bond.

The condition of the compatibility of the interfacial displacements can be written as

$$u_1(x) = u_2(x) + \kappa_0\tau(x) \quad (5)$$

where κ_0 is the interfacial compliance of the buffering layer, if any. Substituting the displacements (4) into the compatibility condition (5) and solving the obtained equation for the interfacial shearing stress, with appropriate boundary conditions, one can obtain the following expression for the interfacial shearing stress:

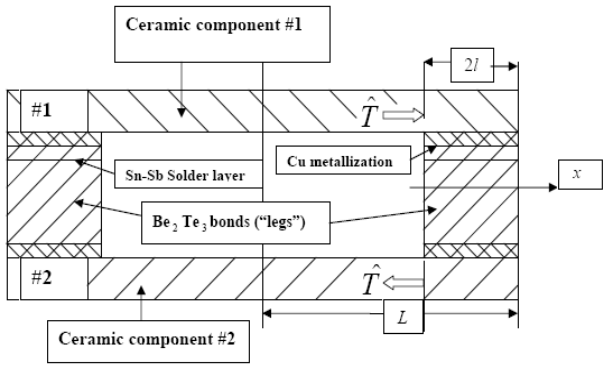


Fig.3. Thermo-Electric Module 2D structure.

$$\tau(x) = k\frac{\alpha\Delta t}{2\lambda_1}\left[\frac{\sinh kx}{\cosh kl} + \frac{\tanh kl}{2kl\left(\frac{L}{2l}-1\right)\sinh 2kl+\cosh 2kl}\cosh k(l-x)\right], \quad \frac{L}{2l} \geq 1 \quad (6)$$

Here

$$k = \sqrt{\frac{2\lambda_1}{\kappa}} \quad (7)$$

is the parameter of the interfacial shearing stress.

The maximum interfacial shearing stress takes place at assembly edge:

$$\tau(l) = k\frac{\alpha\Delta t}{2\lambda_1}\tanh kl\left[1 + \frac{1}{2kl\left(\frac{L}{2l}-1\right)\sinh 2kl+\cosh 2kl}\right] \quad \frac{L}{2l} \geq 1 \quad (8)$$

This relationship indicates that by decreasing the product *kl* one could reduce the maximum interfacial shear stress at the bonding region.

Table I
Mechanical properties of materials employed in TEM

Material	Young Modulus (GPa)	Poisson's ratio
Ceramic Component	380	0.28
Copper Stripe (metallization) on Ceramic	115	0.31
Sn-Sb solder layer	44.5	0.33
Be_2Te_3 leg	47	0.4

ISBN:978-2-35500-022-5

III. Case studies

The analytical solution described in section 2 is applied to the TEM structure Fig. 2. The material properties are given in Table I. Three different assembly sizes of 10, 20 and 40mm (L=5, 10, and 20mm) were chosen. The value of *l*, the half-length of the bonded regions, has been varied to evaluate its effect on the maximum interfacial shear stress. The temperature difference between the top and the bottom components (ΔT) is 130ºC (160-30). Fig. 4 shows how the maximum interfacial shearing stress is changing versus bonded region length for different assembly sizes based on the analytical model. As it could be seen, as the bonded region length decreases the maximum interfacial shear stress reduces.

Finite Element Modeling (FEM) software, ANSYS, has been used to simulate the same TEM assembly. 8 nodes plane 223 element in plane strain mode were used. The structure is meshed with very fine square elements. Each element is $25x25\mu m^2$ and there were around 400000 elements in this structure. The boundary conditions of the simulations were set according to the boundary conditions in the analytical model. The material properties are set according to Table I. The coefficient of thermal expansion (CTE) for the ceramic plates is set to 6.5×10^{-6} (1/ºC).

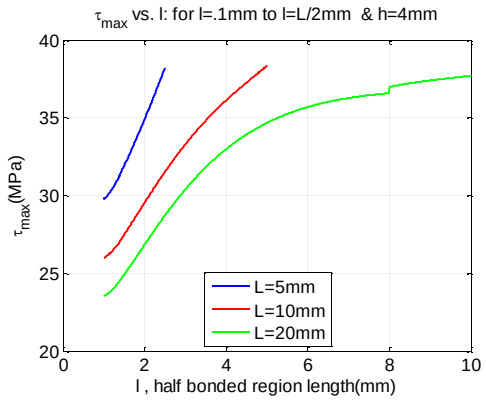


Fig.4. Variation of maximum interfacial shear stress (τ_{max}) vs bonded region's length for different assembly lengths, 2L= 10mm (blue), 20mm (red), and 40mm (green).

The meshed 2D structure and material properties are shown in Fig. 5. The maximum shearing stresses for different assembly and bonded region length sizes, as well as for different leg thicknesses were obtained and the results were compared with analytical solution. The comparison for L=5mm (assembly length 10mm) are shown in Fig. 5, and Fig. 6. As it can be seen in these figures, analytical and numerical solutions are in good agreement.

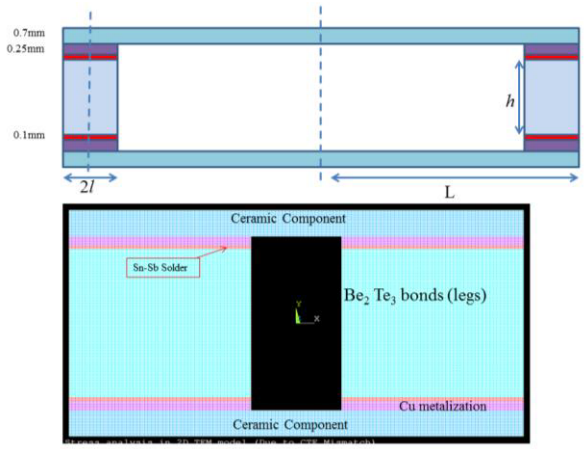


Fig.5. Dimensions, materials, meshing of a 2D TEM with L=3.3*l*.

It is shown that by decreasing the bonded region length the maximum shear stress would drop by a factor of 5. On the other hand, by decreasing the TE leg thickness, the maximum shear stress would increase. Therefore, employment of thinner and longer legs could indeed result in a substantial stress relief, thereby leading to a more mechanically robust TEM. In [5] a similar conclusion was achieved with 3D simulation of 2 leg thermoelectric module.

3D simulation has been done to confirm what is obtained analytically. The meshed structure is shown in Fig. 8. Symmetry is used and a quarter of the model is simulated. In these simulations the assembly length was 10mm and the TE leg thickness was 4mm. By changing the bonded region length from 4mm to 0.2mm (*l* changed from 2mm to 0.1mm), the maximum shear stress dropped by a factor of 13. A comparison between the results is shown in Fig. 9.

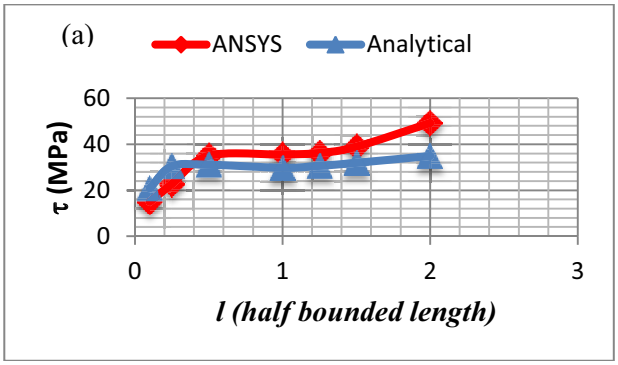


Fig.6 (a). Maximum interfacial shear stress obtained by ANSYS (Red curve) and Analytical model (Blue curve) for h=4mm and L=5mm.

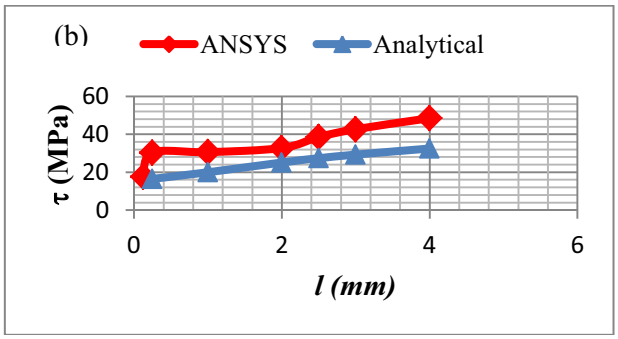


Fig.6 (b). Maximum interfacial shear stress obtained by ANSYS (Red curve) and Analytical model (Blue curve) for h=4mm and L=10mm.

IV. Conclusion

The longitudinal interfacial compliance for the uniform and linearly distributed shear loading along the interface of a long-and-narrow strip has been evaluated in application to assemblies of the TEM type. The evaluated compliances were employed, using analytical modeling, to calculate the maximum shear stress in a TEM design with two legs at the ends. It is shown that the maximum interfacial thermally induced shearing stress occurs at the leg's corner and employment of thinner and longer legs could indeed result in a substantial stress relief. This leads to a more robust design than in the existing (conventional) TEM systems.

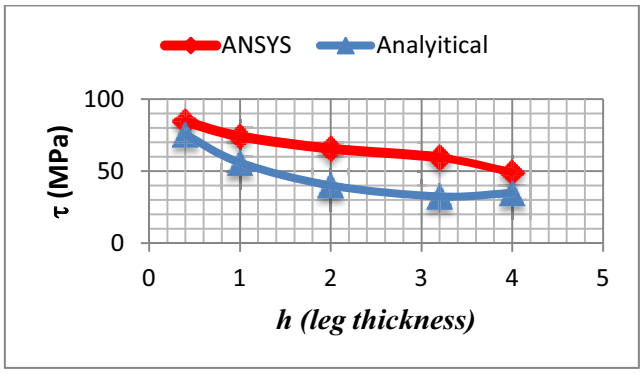


Fig.7. Maximum interfacial shear stress obtained by ANSYS (Red curve) and Analytical model (Blue curve) for L=5mm and *l*=2mm, while h changing from 0.4mm to 4m.

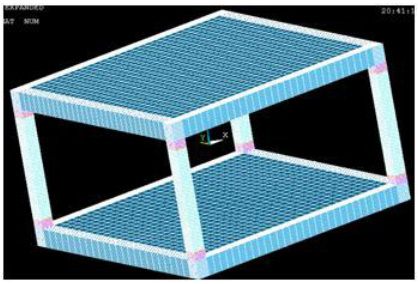


Fig.8. 3D meshed structure.

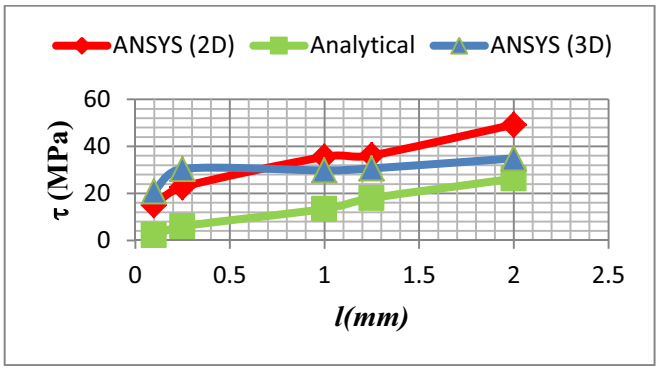


Fig.9. Comparison between Analytical results with 3D and 2D ANSYS results.

REFERENCES

[1] A. Shakouri, "Recent Developments in Semiconductor Thermoelectric Physics and Materials", Annual Review of Materials Research, Vol. 41, pp. 399-431, 2011.

[2] K. Fukutani, and A. Shakouri, "Design of bulk Thermoelectric Modules for Integrated Circuit Thermal Management", IEEE Transaction on Components and Packaging Technologies, vol. 29, no. 4, pp. 750-757, 2006.

[3] J- H. Bahk, R. B. Sadeghian, Z. Bian, and A. Shakouri, "Seeback Enhancement Through Miniband Conduction in III-V Semiconductor Superlattices at low Temperatures", Journal of Electronic Materials, published online Feb.8, 2012.

[4] T. Clin, S. Turenne, D. Vasilevskiy, and R.A. Masut, "Numerical Simulation of the Thermomechanical Behavior of Extruded Bismuth Telluride Alloy Module", Journal of Electronic Materials, Vol. 38, No. 7, 2009.

[5] J.L. Gao, Q.G. Du, X.D. Zhang, and X.Q. Jiang, "Thermal Stress Analysis and Structure Parameter Selection for a Bi_2Te_3 Based Thermoelectric Module", Journal of Electronic Materials, Vol. 40, No. 5, 2011.

[6] ANSYS 14., 2012 ANSYS, Inc.

[7] E. Suhir, "Stresses in adhesively bonded bi material assemblies used in electronic packaging," in *Proc. Elect. Pack. Mat. Science-II, MRS Symp.*, 1986, pp. 133–138.

[8] E. Suhir, "Stresses in bi-metal thermostats," *Trans. ASME J. Appl. Mech.*, vol. 53, no. 3, pp. 657–660, 1986.

25-27 September 2012, Budapest, Hungary

Overview of thermal Studies on photonics devices for reliability, robustness and new design

Yannick Deshayes[1], A. Royon[2,1], R. Baillot[1], L. Béchou[1], L. Canioni[2], Y. Petit[3,2], Th. Cardinal[3], I. Bord[1], B. Levrier[4] and Y. Ousten[1]

[1] *IMS Laboratory, University Bordeaux I-ENSEIRB-CNRS UMR 5218, 351 cours de la Libération, 33405, TALENCE Cedex 5, FRANCE*

[2] *LOMA, University Bordeaux 1, CNRS UMR 5798, 351 cours de la Libération, 33405, TALENCE Cedex 5, France*

[3] *ICMCB, University Bordeaux 1, CNRS UPR 9048, 87, Avenue du Docteur Schweitzer 33608 PESSAC cedex, France*

[4] *Eurelnet, University Bordeaux 1, ADERA, 351 cours de la Libération, 33405, TALENCE Cedex 5, France*

Phone : +33 5 40 00 28 57, Fax : +33 5 56 37 15 45, Email : yannick.deshayes@ims-bordeaux.fr

***Abstract*- This paper presents the main activities on reliability, robustness and failure of photonic devices. The approach is different if the system is manufactured of if it is developed or elaborated. The building reliability is exposed on the new photonic devices using for future numerical data storage for archiving. The robustness is deal with the laser transmitter in marketing position. Finely, the blue light emitting diode is treated with operational reliability. This paper has a main line on the model and measurement of thermal flow or temperature. Since 10 years and actually, the thermal flow is the main cause of degradation of photonic devices. But, additionally, in comparison with electronic devices, the light is the new parameter to take into account in the failure degradation. The blue light interaction with material, especially polymer, is important and the modification of fluorescence change can change function of the photonic devices. This paper underlines the kind of problems.**

I. CONTEXT AND OBJECTIVES

Actual research on reliability is based on study of the material durability and impact on the new design. The building reliability is expected at the elaboration of the devices and specifically for photonic applications. The integration of reliability in the design process is due to the high cost of development and designing of this technology. The knowledge of science, chemical, material and methodologies to understand the failure mechanisms in photonic devices is necessary. This paper proposes some example of association of different laboratory and manufacturer to estimate and build a guideline for conception of new reliable photonic devices.

In parallel, the thermal flow and the temperature distribution is the main cause of degradation like in electronic devices. But the novel problem arrives with the short wavelength transmitter as Blue LED because of the interaction between blue light and plastic material. The plastic material is useful in main technology because of the practical elaboration, good association with semiconductors and the reduction of the packaging cost. But the interaction with blue light is well established and the new solution is engaged by chemical industry to change the chemical composition of plastic material.

II. EXPERIMENTAL SETUP

A. 1.55 µm Butterfly package Laser module

Figure 1 presents a technological description of 1.55 µm Butterfly package Laser module.

Qualification procedures, in particular power drift measurement, must be conducted to validate the system with respect to tolerances through temperature cycling or storage temperature characterizing the limits and the margins of the technology. Actual standards tend to be 500 cycles in the temperature range -40°C/+85°C with a failure criterion of 10% of optical power drift [1]. The methodology of failure diagnostic for optoelectronics components and modules for telecommunication applications imposed to do ageing tests to validate different assumptions coming from the simulation results [2]. In order to calculate levels of stresses and strains after Nd:YAG Laser welding process between the Laser diode platform and the lens holder, simulations are performed using FEM ANSYS software [3]. The different models and boundary conditions are defined in this part. Sub-assembly 1 is composed of the Laser platform and the lens holder essentially in Kovar. The model is based on electrical, thermal and mechanical simulation using a multiphysics approach and will allow to extract isothermal contour plots to evaluate magnitude of thermal gradients in the sub-assembly 1.

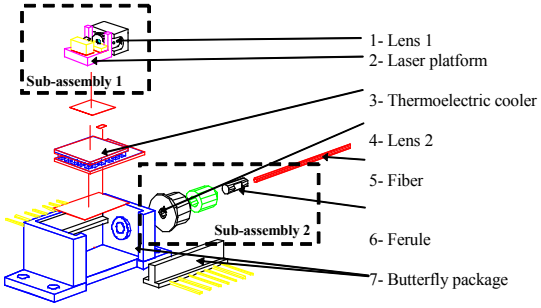


Fig. 1. Description of the Laser module showing the two main sub-assemblies.

B. 472 nm Blue Light Emitting Diode

Figure 2 propose a schematic of 472 nm Blue Light Emitting diode.

Accelerated ageing tests (30mA/85°C/1000h) have been performed on GaN MQW LEDs coated with a silicone oil $(C_7H_8OSi)_n$. Electrical and optical characteristics indicated a 65% loss of optical power and the presence of leakage current on both sides of the chip. First physico-chemical analyses demonstrated that failure mechanisms might be due to a degradation of silicone oil [4, 5, 6]. That last would interact with the chip and let the current to spread over the chip through the oil. Moreover, absorption coefficient of silicone oil has been suspected to increase and contribute to optical losses. The complete NMR 1H and ATR analyses will be investigated before and after ageing tests to precise all failure mechanisms involved in GaN MQW LEDs. Finally all these investigations will be used to extract the physical degradation laws.

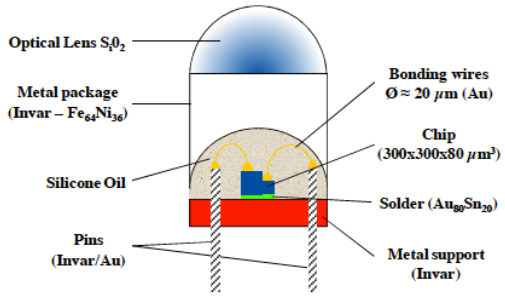


Fig. 2. Schematic diagram of the GaN blue MQWLED packaging.

C. Durability study of structure writing by femtosecond laser for numerical data storage for archival applications

The innovations of recent decades have increased the storage capacity on the media, CD, CD-R and DVD-R up to Blue Ray, but the life span of each new support decrease.

This last point is a real high-stakes topic because nowadays, no support can save files during more than 100 years. Many players in the economic and social life are directly affected by this problem. First we must mention the archivists. Their interest is to find a medium for storing information at vitae eternam. From the PSN group report "It becomes urgent to develop innovative materials for a long-term information storage" (12/2009) [7].

Actual French research is focalized on the possibility to elaborate a durable optical glass disc with femtosecond laser. The first study on this problematic considers thermal stress at 100°C for more than 3100 hours of a fluorescent optical memory composed of laser written silver nanoclusters embedded in glass has been performed [8, 9].

This environmental ageing test is in accordance with the reference of reliability study on CD and DVD supported by LNE (Laboratoire National de Metrologie et d'essais) [10]. Measurements of the luminescence spectra have been carried out at different times, showing a decreasing and an increasing evolution of the red and the blue part of the spectrum, respectively. This evolution has been attributed to the diffusion and the reorganization of different silver species inside the matrix, altering the internal electric field. Tunneling effect based modeling enables the estimation of the lifetime of the memory.

The Archive & Forget French program is based on this first research to elaborate the new glass optical disc. The main key of the success is the compatibility of writing process using femtosecond laser with DVD considering: writing speed and design of data structure.

III. RESULTS AND DISCUSSIONS

A. 1.55 µm Butterfly package Laser module

Actual standards tend to be 500 cycles in the temperature range -40°C/+85°C with a failure criterion of 10% of optical power drift.

Experimental and simulation results lead to give failure modes and assumptions on failure location :

- sudden total optical power drop explained by a break located in the optical fiber core,
- gradual optical power drift outside the failure criteria limit in relation with thermomechanical aspect responsible of columns deformation in sub-assembly 1 and related by stresses relaxation phenomenon,
- gradual optical power drift inside the failure criteria demonstrating the relative instability of optical coupling in Laser module especially on sub-assembly 1.

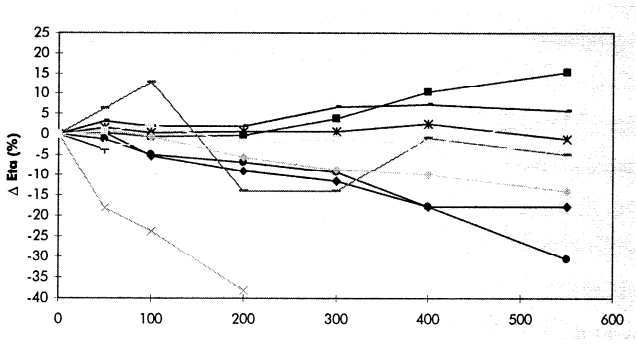


Fig. 3. Ageing test results on 1550 nm InGaAsP/InP Laser module.

Nd:YAG Laser welding process involves a highly focused Laser beam responsible of a non-uniform temperature distribution on the focal print. Simulated energy deposed allows to be close to melting temperature of Kovar material (1723 K). Figure 4 shows the nodal solution contour plot of thermal cartography of Laser platform after first Laser

welding process.

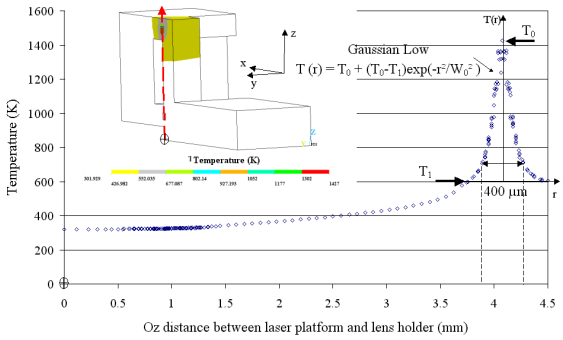


Fig. 4. Variation of the temperature along the column of the Laser submount.

The temperature variation along the column of Laser platform can be fitted by a Gaussian law expressed as :

$$T(r) = T_1 + (T_0 - T_1)\exp\left(-r^2/W^2{}_0\right) \quad (1)$$

with T_0 = 1427 K, the maximal temperature of Laser weld, T_1 = 600K the minimal temperature of Laser weld and W_0 the beam waist defined as the minimum radius of the Laser beam.

Experimental and calculated beam waist values are the same and evaluated around 150 µm. The good agreement between experimental and calculated values validate the simulation approach for Laser Nd:YAG welding process.

Figure 5 clearly shows the residual effective Von Mises stresses close to 55 MPa located in the column base of the Laser Platform after Laser welding.

Thermal gradients (≈ 1100K) along columns of Laser platform induce maximal displacements close to 2 µm located in the column base after Nd:YAG Laser welding process in manufacturing conditions and maximal strains of 0.05%. These displacements are observed by optical axis angular deviation θ (and Δx, Δy and Δz axial deviations.

- An angular deviation θ of 0.02° between sub-assembly 1 and sub-assembly 2.
- Δy, Δz deviations of 10 µm between Laser diode and Lens holder.

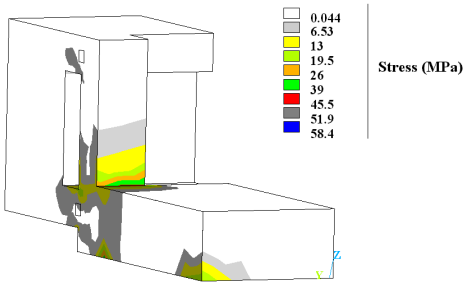


Fig. 5. Residual effective Von Mises stress after Laser welding on the sub-assembly 1.

After Nd:YAG Laser welds, intrinsic and extrinsic stresses can appear. Intrinsic stresses are generally related to only Laser YAG energy deposition on metal surface. The most important accumulated stress is located inside the heat-affected zone (HAZ) caused by plastic deformation and very rapid thermal variation in welded joints [11, 12]. Extrinsic stresses are caused by external loads applied during process and discontinuity of materials on the interface [13]. In our case, the most important external load is represented by pressure strength Fpres used to ensure an adjustment between Laser platform and lens holder.

Relaxation of accumulated stresses in the sub-assembly 1 can occur and could be accelerated by defects induced in the welded zone [14, 15]. Rapid solidification processing in HAZ leads to a metastable phase formation, solid solution or dispersion strengthened alloys and intermetallics and the whole physical phenomenon is at the origin of defects formation located in welded joints [16, 17]. It has been demonstrated that metallic alloys creep fatigue is related to defects rate located in welded joints considered as a metallic alloy zones [18]. In particular, a model based on molecular dynamics calculations, developed by J.D. Vazquez, has discussed on isotropic and anisotropic relaxation phenomenon from simulations of lattice relaxation of metallic alloys considering the sudden appearance of vacancy or an interstitial site in the crystal [19]. This microscopic relaxation model allows to highlight macroscopic effective displacement of system responsible of relaxation phase. Experimental measurements, using in particular an optical method, have been also conducted to observe strains, stresses and fractures of welded joints at the mesoscale level by V.E. Panin [12]. This study has characterized, in bulk material, the accumulated stresses located in HAZ and their evolution after Laser welding process. So our interpretation of gradual optical power drift between the sub-assembly 1 and the pigtail can be explained by relaxation phenomenon and time evolution can be directly related to the number and the location of defects into the welded joints but also in the structure.

B. 472 nm Blue Light Emitting Diode

Active storage (30 mA/85°C/1000 h) has been performed to three GaN MQW LEDs. Three steps of electrical and optical measurements (0 h, 168 h, 500 h and 1000 h) have been achieved. First results highlight the gradual changes of optical power P/P0 about 65% at 1000 h as shown in figure 6.

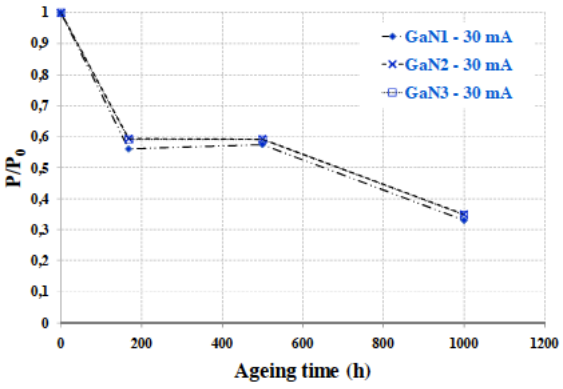


Fig. 6. Variation of the optical power as a function of ageing Time.

Figure 7 plots optical spectra for a constant junction

temperature (300K) before and after ageing tests.

A 65% loss of light emission has been extracted in agreement with total optical power loss. In addition, figure 11 evidences a 3 nm shift of central wavelength (λ_c) towards blue range for a constant T_J. Another mechanism occurs when we look that λ_c keeps constant independently from operating temperature. Such phenomenon cannot be due to chip degradation given the fact that it is well-known in literature that energy gap depends on temperature [20, 21]. That is why we may assume an external mechanism is responsible for both temperature insensitivity of λc and optical power loss. As we have previously demonstrated that silicone oil has a strong effect on GaN-based optical spectra, and considering that polymers are sensitive to blue/near-UV light which is even more the case of λc after ageing (461 nm), we may suggest that silicone coating molecular structure has changed after 1500 h of active storage. This fact could be due to either a different kind of fluorophor molecules or a positive/negative variation of LMWM concentration thereby implying a different fluorescence emission. Fluorescence emission of optical lens has been eliminated from hypotheses as power density of GaN LED chip is too weak (2,88 Wcm^{-2}) to be activated. Therefore, fluorescence analyses have been performed after ageing in addition to 360 nm UV excitation to understand polymer molecular structure change.

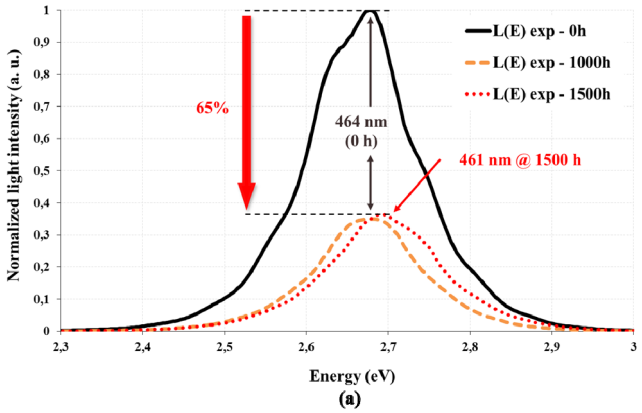


Fig. 7. Optical spectrum before and after ageing tests.

Figure 8 proposes a comparison between fluorescence emission spectra of silicone oil sample at 360nm excitation wavelength before and after ageing.

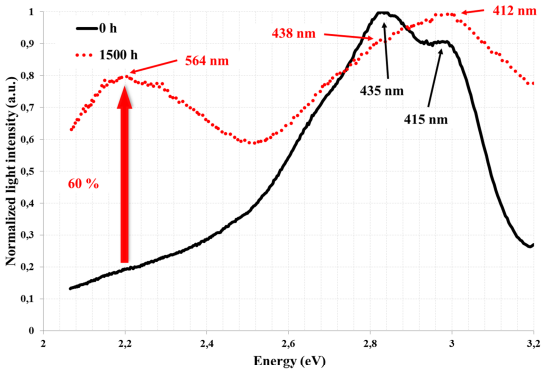


Fig. 8. Fluorescence emission spectra of silicone oil sample at 360 nm excitation wavelength before and after active storage.

Two main mechanisms have been clearly observed after ageing :

- First of all, a strong increase (close to 60%) of fluorescence emission occurred in the red range (564 nm). Here, we may suppose High Molecular Weight Molecules (HMWM) presence could be responsible for such increase

- Another observation has been made on the fluorescence intensity inversion and the spectral shift of the two initial wavelengths (435 nm, 415 nm) at higher energies (from 2,8 eV to 3,1 eV). The most significant shift is that of 3 nm from 415 nm to 412 nm in good agreement with the 3 nm spectral shifts of λc (LED light) observed from 464 nm to 461 nm.

Silicone coating has lost more than 90% of its initial absorption (only coming from fluorophor molecules) when lighted by LED light (464 nm). Under UV excitation, silicone oil fluorescence emission in the red range (564 nm) has increased about 15% that might be slightly linked to the HMWM presence. By looking at 464 nm and below (415 nm, 435 nm), fluorescence efficiency (ηfluorescence) is much better after than before ageing whereas about 70 % of fluorescence emission has been lost after ageing. Moreover, a decrease of fluorophor molecules absorption of 87 % highlights LMWM capability to absorb and reemit fluorescence.

Fluorescence analysis confirms there is a molecular change into molecular structure of silicone oil after ageing. The origin of such change could be linked to polymerization or cross-linking process. Indeed, both light and temperature are often considered as aggravating factors for polymer materials. It has been demonstrated in the literature that either one of those factors or a combination of both, can strongly degrade polymer coating thereby leading to polymerization or cross-linking process [22-25]. At ambient temperature (25°C), the junction temperature of the chip amounts 100°C. Therefore, when the device is operating under ageing conditions (i.e. at 85°C), the junction temperature is suspected to be highest than 160°C. Due to thermal heat diffusion through silicone oil, polymerization or cross-linking mechanism could occur over ageing time (1500 h).

Molecular change has also been confirmed through ^1H NMR and Mass spectrometry analyses.

C. Durability study of structure writing by femtosecond laser for numerical data storage for archival applications

Silver clusters are created by a pulse train from a near infrared (NIR) femtosecond laser focused with a microscope objective inside a FPL glass (Figure 9a). Such a photosensitive glass belongs to the phosphate family and was initially designed for gamma irradiation dosimetry [26,

27]. In the present case, the composition is slightly modified: It is a zinc phosphate glass containing silver ions [28] presenting an ultraviolet (UV) absorption band below 280 nm. Detailed descriptions of its fabrication and properties can be found in the Experimental section.

Following exposure to a high-repetition-rate femtosecond pulse train, the glass presents a broad excitation band (300–450 nm).

When excited by NUV radiation, it emits homogeneous white fluorescence, the properties of which (intensity, spectrum, and lifetime) depend on the irradiation dose (fluence, number of pulses, and repetition rate [29]. This fluorescence is attributed to the presence of Ag_m^{x+} silver clusters with m < 10 the number of atoms and x the ionization degree [26, 30, 31]. These silver clusters are created inside the focusing voxel (Figure 9a) and are arranged into a pipe shape along the propagation axis, with a length corresponding to the Rayleigh range and a wall thickness of about 80 nm [32]. For a given fluence range (from 2 Jcm^{-2} to 6 Jcm^{-2}), the silver clusters are produced without significant linear refractive index change ($< 10^{-4}$), while they exhibit appreciable fluorescence and nonlinear optical properties.

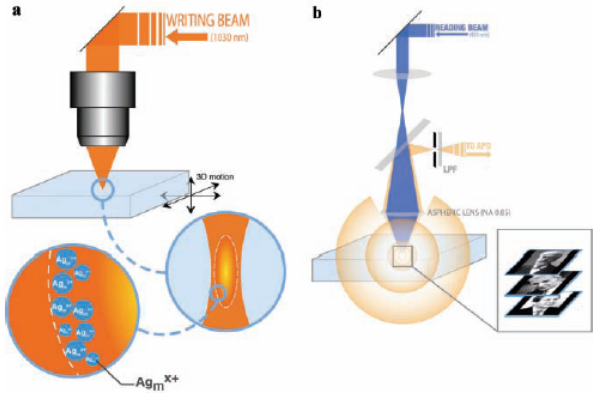


Fig. 9. Fluorescence emission spectra of silicone oil sample at 360 nm excitation wavelength before and after active storage.

This glass has already shown its potential for 3D optical data storage that takes advantage of the third-order susceptibility contrast of the photo-induced clusters with respect to the glass matrix [33].The fluorescence intensities emitted by the silver clusters can be much higher (50 times) than those from fluorescent molecules such as dyes (Rhodamine 6 G at 10^{-4} $mol.L^{-1}$) [34]. The quantum yield of these species is very high, on the order of 17%. [30] In the FPL glass, the emission can be tuned from yellow to red [26, 30]. The fluorescence time decay of the clusters is on the order of a few nanoseconds [26], allowing a reading speed as high as 500 MHz. In the FPL glass, the fluorescence intensity of the silver clusters varies linearly with the irradiance and logarithmically with the deposited number of pulses, within the range of interest. This last behavior is well known in photographic processes [35]. Therefore, when the dose is adjusted properly, the fluorescence intensity of the photo-induced species can be controlled exactly as in silver photographic film.

After 3100 hours of passive storage at 100°C, the fluorescence response not change. The Ag nanoclusters are stable in glass matrix and the diffusion process is stop by the pseudo potential well coming from the different charge concentration organized by femtosecond Laser.

IV. CONCLUSIONS

This paper reports three different results on photonic devices.

The demonstration of robustness is applied on the 1550 nm butterfly transmitters using finite element model. The estimation of residual stresses (55 MPa) coupled to analyses of defects locates in laser welding joints explain the optical losses of transmitters (40%) after 500 thermal cycles. The initial misalignment is around 0,02 ° corresponding to more than 40% of optical losses. The solution find is to reduce the height of columns in sub-assembly 1 to reduce the impact of laser welding process. After this modification, laser transmitters are qualified.

The Blue Light Emitting Diodes are composed by plastic glob top on chip. This part of packaging is sensitive to both temperature and blue light. The modification of molecule in polymer is responsible to both absorption of blue light and fluorescence. The efficiency of fluorescence is weak before the direct lighting. The 65% of optical losses is mainly explained by this last phenomenon. The chemical analyses, based on both 1H NMR and mass spectroscopy, confirm the modification of molecular composition of polymer.

Finely, the new photonic devices using for new numerical data storage support seems to be robust but some ageing tests must be done to estimate the lifetime of it. Considering the good behavior of glass matrix regarding ageing test (Fresnel lens used in lighthouse, Egyptian glass and perovskite structure).

The guideline of this paper is both the consideration of reliability estimation and the thermal analyses of the different photonic devices. Different tools have been used to develop model and understand the failure mechanisms. The link between experimental procedure and models has been performed to build useful tools.

REFERENCES

[1] .L. Goudard, P. Berthier, X. Boddaert, D. Laffitte, J. Perinet. New qualification approach for optoelectronic components. Microelectron. Reliab. 2002; 42 : pp. 1307-1310

[2] R. McLeod, M. Wolkin, V. Morozov, K. Sawyer. Packaging of micro-optics component to meet Telcordia standards. Proc. OFC, Anaheim, CA, USA, March 2002.

[3] Y. Deshayes, L.Bechou, F. Verdier, Y. Ousten, D. Laffitte, JL Goudard, Thermomechanical Stresses and Optical Misalignment in 1550 nm Emissive Optoelectronic Modules Using FEM and Process Dispersions, IEEE Transactions on Components and Packaging Technologies, 31, 4 (2008), pp. 759-766.

[4] R. Baillot, Y. Deshayes, L. Bechou, Th. Buffeteau, I. Pianet, C. Armand, F. Voillot, S. Sorieul, Y. Ousten, Effects of silicon coating degradation on GaN MQW LEDs performances using physical and chemical analyses, Microelectronics reliability, 50 (2010), pp 1568-1573, 2010.

[5] Bochkareva, N.I., Gorbunov, R.I., Klochkov A.V., Lelikov Yu.S.,

Martynov I.A., Rebane Yu.T., Belov A.S. and Shreter Yu.G., "Optical Properties of Blue Light-Emitting Diodes in the InGaN/GaN System at High Current Densities", Semiconductors, 42(11), 1355-1361, (2008).

[6] Meneghini, M., Trevisanello, L.R., Meneghesso, G. and Zanoni, E., "A Review on the Reliability of GaN-Based LEDs", IEEE Trans. Dev. Mat. Reliab. 8(2), 323-331 (2008).

[7] E. Spitz, J.C. Hourcade and F. Laloë, Préservation à long terme de l'information numérique, Rapport du groupe PSN – académies de sciences et des technologies – 51 pages.

[8] Strickler, J. H. & Webb, W. W. Three-dimensional optical data storage in refractive media by two-photon point excitation. Opt. Lett. 16, 1780-1782 (1991).

[9] Bellec, M. Royon, A. Bourhis, K. Choi, J. Bousquet, B. Treguer, M. Cardinal, T. Videau, J.-J., Richardson, M. & Canioni, L. 3D patterning at the nanoscale of fluorescent emitters in glass. Submitted to J. Chem. Phys. C (2010).

[10] J.M Lambert, étude du vieillissement en conditions climatiques sévères de 2 références de CD-R, Fiche R&D N° 202/2005, GIS DON, Avril 2007.

[11] S. Jang. Packaging of photonic devices using Laser welding, Inc : Proc. SPIE, 1996, pp. 138-149.

[12] V.E. Panin, V.S. Pleshanov, S.A. Kobesva, S.P. Burkova. Relaxation mechanism of rotational type in fracture of weld joints for austenic steels. Theoret. and app. Fract. Mechan. 1998 ; 29 : pp. 99-102.

[13] T. Inoue and H. Koguchi. Relaxation of thermal stresses in dissimilar materials (approach based on stress intensity). Intern. Jour. of Sol. and Struct. 1997 ; 34 : pp. 3215-3233.

[14] T. Hayashi and H. Tsunetsugu. Optical module with MU connector interface using self-alignment technique by solder-bump chip bonding, Inc : Proc. 46th Electronic Components and Technology Conference, 1996, pp. 13-19.

[15] W.H. Cheng, Y.D. Yang, T.C. Liang, G.L. Wang, M.T. Sheen and J.H. Kuang. Thermal stresses in box-type Laser packages. Optical and Quant. Electron. 1999; 31 : pp. 293-302.

[16] S. Hariprasad, S. M. L. Sastry and K. L. Jerina. Deformation behavior of a rapidly solidified fine grained Al-8.5%Fe-1.2%V-1.7%Si alloy. Acta Materialia. 1996 ; 44 : pp. 383-389.

[17] W.H. Cheng, W.H. Wang and J.C. Chen. Defect formation mechanisms in Laser welding techniques for semiconductor Laser packaging. IEEE Transact. on Comp., Packag. and Manufact. Techn. 1996 ; 19 : pp. 764-769

[18] T. Asayama. Creep¯fatigue evaluation of stainless steel welded joints in FBR class 1 components. Nucl. Engin. and Design. 2000 ; 198 : pp. 25-40.

[19] J. Dominguez-Vazquez; A.M.C. Perez-Martin; J. Jimenez-Rodriguez; A. Gras-Marti. Relaxation of metals: A model based on MD calculations. Nucl. Instru. & Meth. in Phys. Resea.. 1998 ; 135 : pp. 214-218.

[20] Li, Y., Lu, Y., Shen, H., Wraback, M., Brown, M.G., Schurman, M., Koszi, L. and Stall, R.A., "Temperature dependence of energy gap in GaN thin film studied by thermomodulation", Appl. Phys. Lett. 70, 2458- 2460 (1997).

[21] Teisseyre, H., Perlin, P., Suski, T., Grzegory, I., Porowski, S., Jun, J., Pietraszko, A. and Moustakas, T.D., "Temperature dependence of the energy gap in GaN bulk single crystals and epitaxial layer", J. Appl. Phys. 76(4), 2429- 2435 (1994).

[22] Rabek, J.F., [Photodegradation of polymers : physical characteristics and applications], Springer (1996).

[23] Rabek, J.F., [Polymer photodegradation : mechanisms and experimental methods], Springer (1995).

[24] Hamid, S.H., [Handbook of polymer degradation], Marcel Dekker (2000).

[25] Guillet, J., [Polymer photophysics and photochemistry : an introduction to the study of photoprocesses in macromolecules], CUP Archive (1987).

[26] H. Schneckenburger , D. F. Regulla , E. Unsöld , Appl. Phys. A 1981 ,26 , 23

[27] A. V. Dmitryuk , S. E. Paramzina , A. S. Perminov , N. D. Solov'eva , N. T. Timofeev , J. Non-Cryst. Solids 1996 , 202 , 173.

[28] I. Belharouak , C. Parent , B. Tanguy , G. Le Flem , M. Couzi , J. Non-Cryst. Solids 1999 , 244 , 238.

[29] C. Maurel , T. Cardinal , M. Bellec , L. Canioni , B. Bousquet , M. Treguer , J.-J. Videau , J. Choi , M. Richardson , J. Lumin. 2009 , 129 ,1514.

[30] K. Bourhis , A. Royon , M. Bellec , J. Choi , A. Fargues , M. Treguer , J.-J. Videau , D. Talaga , M. Richardson , T. Cardinal , L. Canioni , J. Non-Cryst. Solids. 2010 , DOI 10.1016

[31] G. De Cremer , E. Coutiño-Gonzalez , M. B. J. Roeffaers , B. Moens , J. Ollevier , M. Van der Auweraer , R. Schoonheydt , P. A. Jacobs , F. C. De Schryver , J. Hofkens , D. E. De Vos , B. F. Sels , T. Vosch , J. Am. Chem. Soc. 2009 , 131 , 3049.

[32] M. Bellec , A. Royon , B. Bousquet , K. Bourhis , M. Treguer , T. Cardinal , M. Richardson , L. Canioni , Opt. Express 2009 , 17 , 10304.

[33] L. Canioni , M. Bellec , A. Royon , B. Bousquet, T. Cardinal , Opt. Lett. 2008 , 33 , 360.

[34] L. A. Peyser , A. E. Vinson , A. P. Bartko, R.M. Dickson , Science 2001 , 291 , 103.

[35] K. Biedermann , Appl. Opt. 1971 , 10 , 584.

25-27 September 2012, Budapest, Hungary

Thermal Stress Analyses of Inductive Proximity Sensors under Thermal Cycling Condition

Sung Yi and Kevin Pickup
Mechanical and Material Engineering Department
Portland State University
Portland, Oregon 97207-075,USA
E-mail:sungyi@cecs.pdx

Abstract- **In the present study, thermally induced stresses in inductive proximity sensors are investigated using the finite element method. High residual stresses are developed in the inductive proximity sensor because of hygro-thermal mismatches among components during the operation. Its operating temperature is from –40°C to 85°C. During thermal cycles, residual stresses inside the sensor are calculated based on global and local finite element analyses. Failure mechanisms and stress distributions in the sensor as well as the components are discussed in detail.**

I. INTRODUCTION

Inductive proximity sensors have been widely used in many industries in order to sense the presence of steel objects without physical contact. The proximity sensor produces an electro-magnetic field within a sensing range and has a semiconductor switching element. Over years, this sensor has faced steadily increasing demand to achieve very high reliability performance on harsh and unforgiving applications. The sensor must operate between –40°C and +75°C temperature conditions with its electronic components kept protected from intrusion of moisture. A cross section of a typical sensor is shown in Fig. 1.

A schematic drawing of a typical sensor is shown in Fig. 1. Electronic components are surrounded by foam resins in order to provide physical protection. Thus the foam resin must possess adequate mechanical strength, good adhesion to various package components, good corrosion and chemical resistance, matched coefficients of thermal expansion (CTE) to the materials, and high moisture resistance in the use temperature range.

In particular, the ability to maintain good adhesion with the various package components under the adverse manufacturing and service conditions is of paramount importance as delamination along interfaces is a major reliability issue for plastic IC packages [1-7]. Residual stresses arise due to hygro-thermal mismatches between foams and transistors or diodes in the inductive proximity sensor and cause delamination and cracks during service. Therefore, the stress analysis of the inductive proximity sensor is essential to achieving accurate prediction of reliability of the product. In the present study, thermally induced stresses in inductive proximity sensors are investigated using the finite element method.

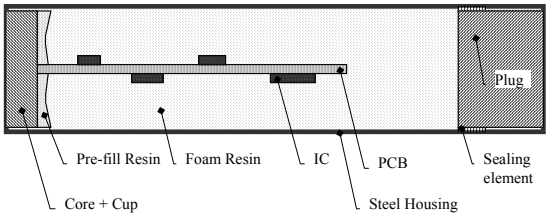


Fig. 1 A typical inductive proximity sensor

II. FINITE ELEMENT MODELS

The sensor consists of many parts including the steel housing, the PCB with the transistors, resistors and diodes, the foam, and so on. The stresses inside the sensor are developed when temperature changes. Moreover, to investigate the stresses inside the transistor or the diode, the outside forces acting on these components should be determined first.

Since the dimensions of the SOT89 transistor and the diode are very small to those of the sensor. Therefore, it is difficult to analyze the sensor including the details of the transistors and diodes using one finite element mesh. The global and local analyses are adapted in order to overcome such a difficulty in this study.

Firstly the global finite element mesh is constructed for the inductive proximity sensor. Then a second-level model is developed for the individual component and areas that are concerned. The displacement field is extracted from the global model, and applied to the local model as boundary conditions at the interfaces between the two models. In the present study, 3-dimension solid elements are used. The special attention has been paid to the failure mechanism of the transistor and the diode. For the local model, the transistor and the diode are included and the very fine meshes are used.

III. Materials

The material properties play an important role in the finite element analysis. Numerical results would be far from the real situation if the inadequate material information were supplied. The materials including foams, PCB, sealing, housing, etc. are characterized. Temperature dependent mechanical properties of those materials are obtained using the dynamic mechanical analyzer and universal material testing machine. The coefficients of thermal expansions are also measured using the thermal mechanical analyzer. The materials properties of these materials are listed in Table 1 and the temperature dependent properties of foams are shown in Fig. 2.

Experimental results show that the properties of foams are dependent on temperature significantly. As shown in Fig. 2, the Young's modulus of the foam decreases rapidly near 65°C that is its glass transition temperature (Tg). It is about 22.5 MPa at 80°C which is only about 2% of 1130 MPa at 55°C.

Table 1. Material Properties (at 25°C)

Material	Young's modulus (MPa)	Poisson's ratio	CTE ($\times 10^{-6}$/°C)
Housing	210000	0.32	8.00
Wire	123000	0.33	9.20
PCB	17700	0.14	21.5
Diode	1160	0.30	3.67
Molding Compound	1280	0.30	16.5
Silicon Die	202240	0.28	3
Core	1131	0.23	2.80
Foam	1190	0.35	41.5
Plug	1260	0.30	118
Sealing	10.9	0.30	157

Boundary Conditions

The sensor is considered as a self-balanced system without any connection to other components. The symmetrical conditions are applied when a quarter of the sensor is analyzed.

The sensor is assembled from several parts and the parts may separate from others during the thermal cycling. In the present study, the contact conditions near two ends are ignored because they play the minor effect on the stresses in the central part of the sensor when the transistor and the diode are main focused on. It is also reasonable if the adhesive property of the foam is considered. A typical experimental curve is shown in Fig. 3. Suhir and Sullivan [1] reported that the typical adhesive strength was about from 10 to 20 MPa. Yi et al. [5] also measured the normal and shear interfacial strengths of the leadframes to the molding compounds and the normal and shear interfacial strengths obtained are about 11.5 and 13.5 MPa, respectively. Therefore, it can be expected that an interface will keep in contacting while the tension stress is less than 10MPa. Otherwise it may be opened. The numerical results show that the maximum tension stress is less than 10MPa, which means the foam would not separate from other components and provide a good adhesion function in working environments.

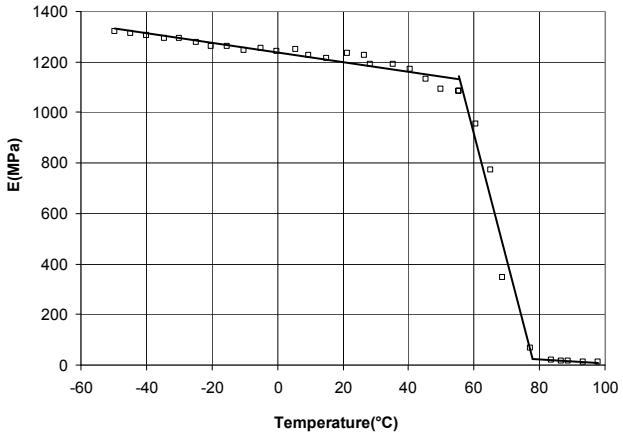


Fig. 2 Young's modulus of the foam

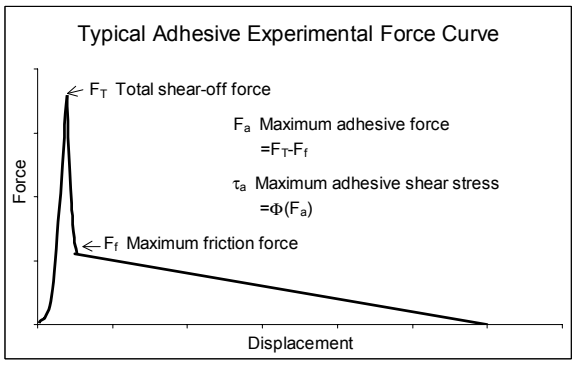


Fig. 3 A typical adhesive strength curve of the foam

Temperature Loading Conditions

The thermal loading cycle is applied to the sensor. As shown in Fig. 4, a cycling period is about 4 hours including 2 hours for the low temperature at -40°C and 2 hours for the high temperature at 85°C. Residual stresses are determined with an assumption that the temperature field is uniform inside the sensor. The loading cycle is divided into two parts: one is from the room temperature (25°C) to the low temperature and the other is from the room temperature to the high temperature. A nonlinear relationship between stresses and temperature is obtained because the glass transition temperature (Tg) of the foam is about 65°C that is in the high temperature loading range.

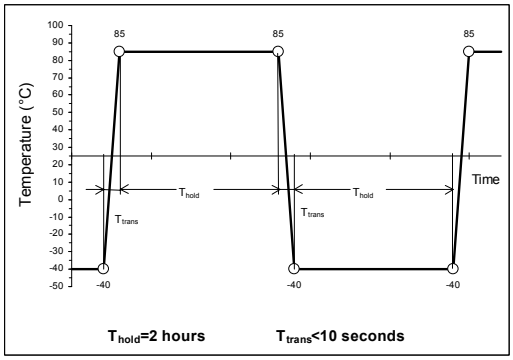


Fig. 4 Thermal cycles

Global Models

Finite element models for the whole sensor are constructed including the housing, the plastic sealing, the plastic cup, the copper coil, the core, the PCB, the foam and so on. The diode and transistor are also meshed to investigate their effect on the global stress fields. The sensor is cut by two symmetrical planes and a quarter is meshed to minimize the model scale without losing any accuracy. The global model is shown in Fig. 5 as assembled with several main parts represented by different colors.

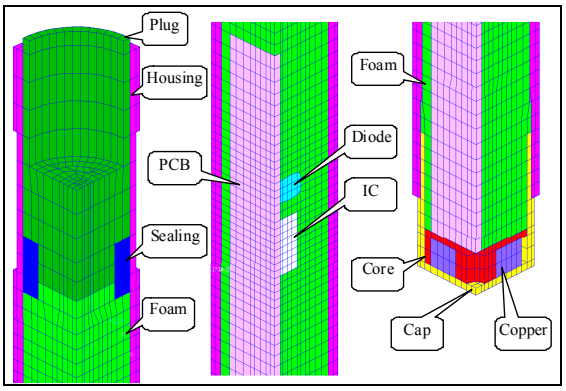


Fig. 5 Global finite element models

Local Models

In the local models, the much-refined meshes are created for the small part of the global models in order to determine the stress fields in the local area. The detail location of the local model is depicted in Fig. 6. The transistor is modeled corresponding to its physical parts including the aluminum leadframes and the molding compound. The diode consists of a glass body and two aluminum terminators. The solder is also modeled as the connection between the components and the PCB. The foam is filled the space in the local models. The local model is shown in Fig. 6.

As the gap between the transistor and the PCB is small, the foam often fails to fill it. In the present study, two filling conditions such as complete filling and cavity are considered. As shown Fig. 7, the foam fills the gap completely or it remains as a cavity.

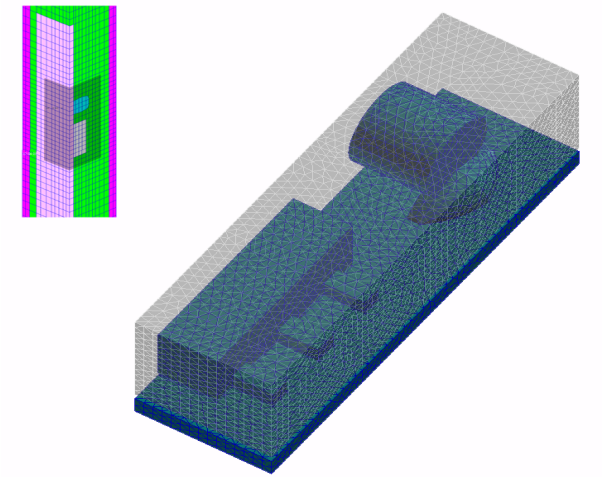


Fig. 6 A local finite element model

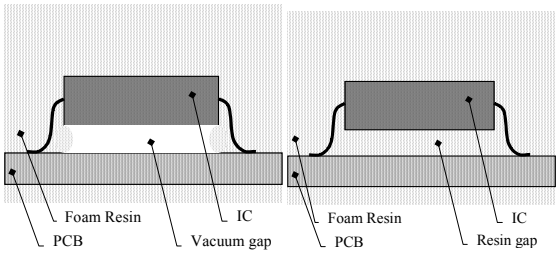


Fig. 7 Two filling models

IV. NUMERICAL RESULTS

Numerical Results of Global Models

Two loading conditions such as case 1 and case 2 are considered in the global models. Case1 is the cooling case in which the temperature is dropping from 25°C to -40°C, while it increases from 25°C to 85°C for case 2.

Loading case 1

Perfect bonding between the foam and the components is assumed for this case. The CTE of the foam is much larger than that of the steel and the steel housing prevents the foam from shrinking. As a result, the compressive stress filed is induced in the foam as shown in Fig. 8.

Von Mises stress is less than 0.1 MPa, which means the shear stresses inside the foam are low. It results from the steel housing prevents the foam from shrinking freely while the temperature dropping because CTE of the later is much greater. The transistor and the diode are under a lifting-off force that may take them from the PCB.

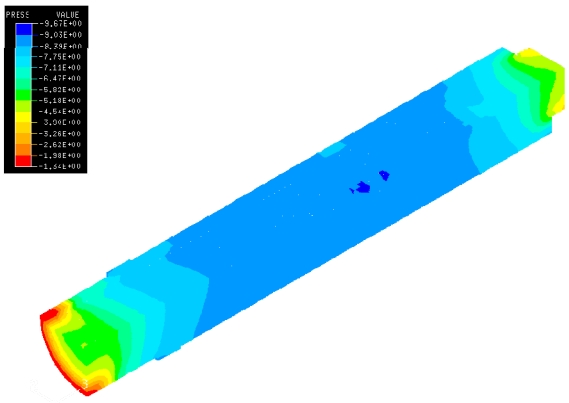


Fig. 8 The global stress distribution for case 1

Loading case 2

The temperature increases from 25°C to 85°C. As the foam has the largest CTE among all components, the foam would push the housing during the temperature increase while expanding. Then the interface between them will keep in contact while the temperature rising. The stresses in the sensor reach their maximum values at about 65°C which is the glass transition temperature (Tg) of the foam. Von Mises stress is quite low. The normal stress field in the foam is positive. The positive stresses push the transistor and the diode towards the PCB and prevent them from being lifted off. Fig. 9 shows the stresses developed in the transistor.

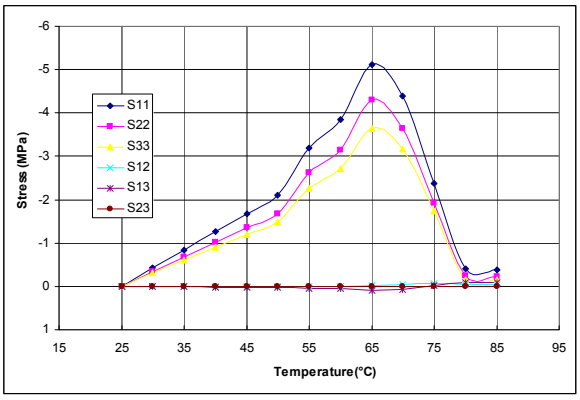


Fig. 9 The stresses in the transistor due to temperature increase for case 2

Numerical Results of Local Analyses

For the local analysis, the displacements obtained from the global analysis are used as boundary conditions. Two filling models for a very thin gap between the transistor and the PCB are considered.

Loading case 1.a

It is assumed that the foam is completely filled into the gap between the chip and the PCB. The stresses at the interface between the aluminum terminators and the encapsulant are calculated. The tension stress distribution is shown in Fig. 10 at the interface and the maximum tension stress is about 13MPa.

The deformation of the diode reflects that a bending force is induced in the current loading case. A clear crack path in which von Mises stress is above 40MPa is indicated in Fig. 11. It is noticed that the crack path starts the end of the diode and extends to the center of the glass body. The symmetrical part can be implied due to the symmetrical conditions. It can be expected that cracks may occur only at the ends of the diode, never start from the central part where the stress is about 11MPa.

Loading case 1.b

In this case, it is assumed that the foam does not fill the gap and leave a cavity. The nromal stress distribution inside the chip is shown in Fig. 12. The maximum tension stress reaches 20MPa in the central area, while the stresses inside diode are similar to case 1.a.

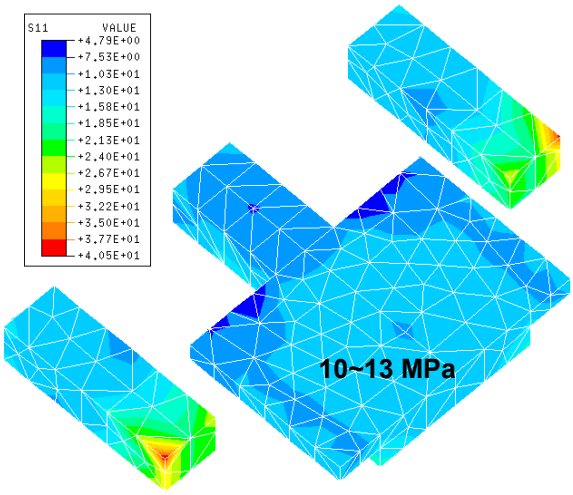


Fig. 10 The tension stress distribution in the chip for case 1.a

Loading case 2.a

In this loading case, the complete filling is assumed. The temperature rises from room temperature to 85 °C. As being pointed out in the global analysis, the stresses reach their maximum values near the glass transition temperature (65°C). All results listed here represent to their values at this temperature. As shown Figs. 13 and 14, the compressive stress is observed in the transistor and the von Mises stress in the diode is less than 8MPa. The magnitudes of stresses are much lower than those in case 1. It can be expected that the sensor is reliable in the high temperature than in the low temperature.

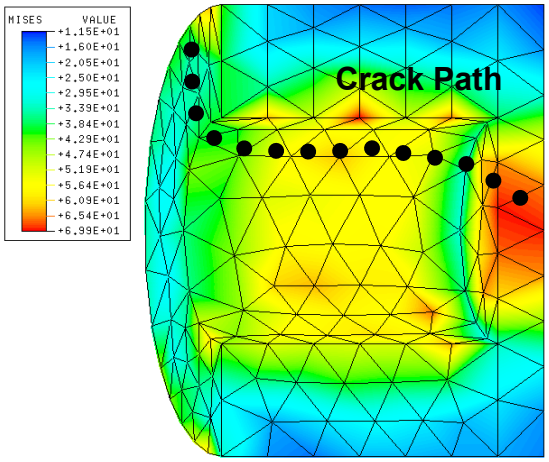


Fig. 11 Von Mises stress distribution in the diode for case 1.a

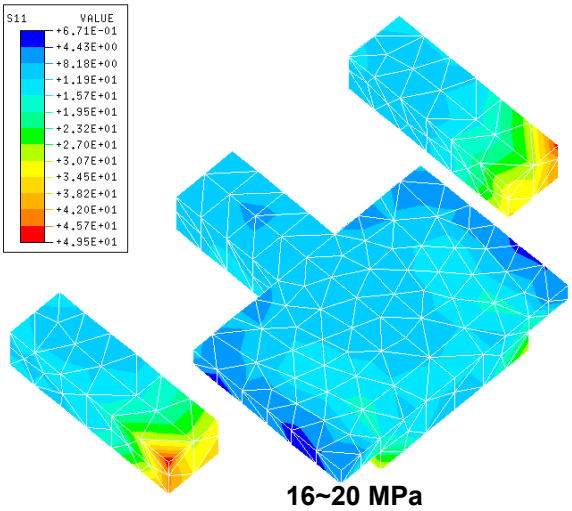


Fig. 12 The stress distribution in the chip and aluminium leadframe for case 1.b

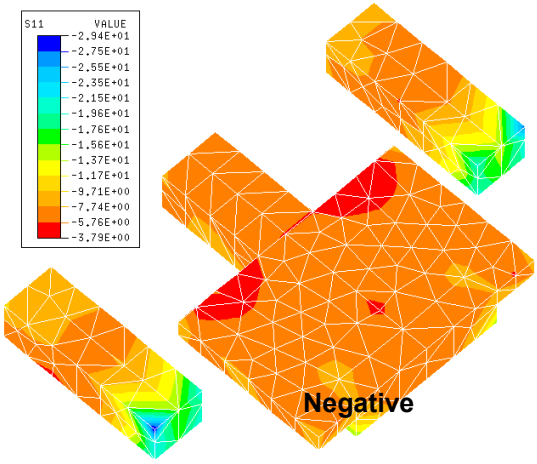


Fig. 13 The stress distribution inside the chip for case 2.a

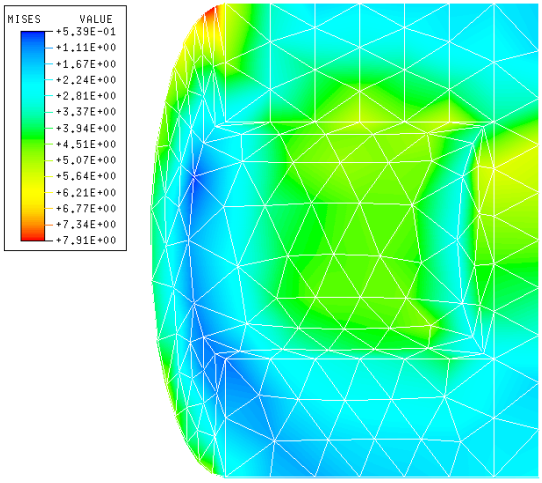


Fig. 14 Von Mises stress distribution in the diode for case 2.a (The maximum value is less than 8 MPa)

Loading case 2.b

Taking the same filling model as loading case 1.b, the stresses in the sensor are investigated for high temperature loading. In this case, both compressive stresses and von Mises stresses are quite lower compared with case 2.a.

V. DISCUSSION AND CONCLUSIONS

As shown in the results from the global models, quite uniform stress fields are observed inside the foam, which means the mesh size is small enough to determine the boundary conditions for the local models. The CTE mismatch among the components of the inductive proximity sensor results in the development of the residual stresses during the thermal cycling. The foam has the large CTE while the steel housing has a much small one. Consequently, the metal housing prevents the foam from expanding or shrinking freely. As a result, the high tensile and compressive normal stresses are developed in the foam. The compressive normal stress is observed in the foam because of the strong adhesive strength of the foam to other components and the metal housing, which can cause the foam can not separate from other components while shrinking. This normal stress may lead to lifting the transistor off the PCB during the cooling. As shown Figs. 10 and 12, the normal stress can be as much as 20MPa at the interface between the aluminum terminators and the encapsulant. The failure may occur along this interface to separate the terminators from the molding compound and then the chip is damaged.

For the high temperature case, the stresses reach their maximum values at the glass transition temperature of the foam 65ºC rather than 85ºC. However, it is also fingered out that the interconnection is much safer for this case since the foam pushes the transistor towards the PCB while expanding. The diode is under bending state since the two aluminum

terminators are connected to the PCB through the solders. A failure mode can be predicted at the temperature of -40ºC. The failure may occur from one end to another by passing the inner central part of the glass body. However, the stresses are very low in the outer central layer. It can be expected that there is no mechanical failure in this area.

REFERENCES

[1] Suhir, E., and Sullivan, T M., ``Novel technique for the evaluation of adhesive strength of molding compounds used in plastic packaging,'' *Proceedings - Electronic Components and Technology Conference*, IEEE Service Center, Piscataway, NJ, USA, 1992, pp. 911-925.

[2] Yi, S. and Sze, K. Y., ``Cooling Rate Effect on Post Cure in Plastic Encapsulated IC Packages,'' *Transaction of ASME: Journal of Electronic Packaging*, Vol. 120, pp. 385-390, 1998.

[3] Yi, Sung, Goh, J. S. and Yang, J. C., "Residual Stresses in Plastic IC Packages During Surface Mounting Process Preceded by Moisture Soaking Test," *IEEE Transactions on Components, Packaging, and Manufacturing Technology-Part B*, USA, Vol. 20, No. 3, March 1997, pp. 247-255.

[4] Yi, Sung and Neo, Guan Hock, Hygrothermo-mechanical Behavior of Mold Compound Materials at Elevated Environment, *Proceedings of the 1998 Workshop on Mechanical Reliability of Polymeric Materials and Plastic Packages of IC Devices***,** Paris, France, ASME, pp. 287-292, 1998.

[5] Yi, Sung, Yue, Chee Yoon, Hsieh, Jang-Hsing, Fong, Liwha, Lahiri, Syamal K., ``Effects of Oxidation and Plasma Cleaning on Adhesion Strength between Copper Leadframes and Molding Compounds'', *Journal of Adhesion Science and Technology*, Vol. 17, No. 7, pp.789-804, 1999.

[6] Kim, H.H., Choi, S.H., Shin, S.H., Lee, Y. K., Choi, S. M. and Yi, Sung, Thermal Transient Characteristics of Die Attach in High Power LED PKG, **Microelectronics Reliability**, Volume 48, Issue 3, pp. 445-454, 2008.

[7] Gao, S., Hong, J.P., Kim, J.S., Yoo, D.J., Jeong, T.S., Choi, S.M. and Yi, Sung, Development of an Ultra-Slim System in Package (SiP), **Journal of the Microelectronics and Packaging Society,** Vol. 15, No. 1, pp. 7-18, 2008.

25-27 September 2012, Budapest, Hungary

MiMAPT: Adaptive Multi-Resolution Thermal Analysis at RT and Gate Level

Mohammadsadegh Sadri, Andrea Barolini, Luca Benini
Micrel Lab, DEIS, University of Bologna
{mohammadsadegh.sadr2,a.bartolini,luca.benini}@unibo.it

***Abstract*—Tight timing/area constraints produce on-chip layouts with non-regular shapes for RTL entities. Thus, grid-like floorplans where RTL entities are abstracted as rectangular blocks for thermal simulation lead to inaccurate results. In addition, spatial and temporal variability of chip workload causes localized temperature variations. Exact localization of hotspots at gate-level necessitates an extremely detailed spatial resolution which is very computationally intensive.**

We propose MiMAPT, a tool capable of performing thermal simulation at RT and gate-level with multiple scales of spatio-temporal resolution. To demonstrate the tool advantages we run various tests for a sample chip. We show that our tool provides high level of flexibility in terms of speed vs. accuracy of results.

I. Introduction

High power densities of today's integrated circuits lead to on-chip thermal hotspots which can compromise chip functionality. As [1] describes, the volumetric power density of a 20nm device is on the order of 10 TW/cm^3. Consequently, researchers and CAD vendors are developing tools to facilitate design-time prediction and identification of thermal hazards.

Spatial and temporal variability of chip workload results in non-uniform on-chip power density and localized temperature variations which significantly affect device parameters. [2] Shows an example of this. Considering the above facts, precise localization of hotspots at gate-level is necessary for many high-performance designs. This however requires an extremely detailed spatial resolution for power/thermal simulation which is very computationally intensive and almost impossible for large designs. This is where multi-scale analysis techniques can help. Conceptually, power and temperature estimation can be done at low resolution when it provides satisfactory level of accuracy at a high level of speed. Resolution (and computational effort) should be increased only for areas of interest, where hot spots are likely. The challenge lies in avoiding false negatives (i.e. missing hot spots) while minimizing false positives to achieve significant speedups w.r.t. fine-grained (gate-level) thermal analysis.

We propose MiMAPT (Micrel Multi-scale Analyzer for Power and Temperature), a tool capable of performing power/thermal simulation at RT and gate-level with multiple scales of resolution and speed. The tool starts coarse-grained transient power/thermal analysis at RTL for a user-defined time interval. It considers non-uniform shapes of on-die units during analysis. MiMAPT switches to accurate gate-level simulation only when a likely hot-spot is suspected. At gate-level it performs iterative power/thermal simulation while refining spatial resolution just for the areas which are suspected to contain hotspots.

II. Related Work

We categorize published results from academic researchers into three areas:

- Thermal simulation platforms which estimate chip/package temperature distribution based on a given power and floorplan.
- Power estimation packages, which calculate dynamic and static power of a target architecture based on activity statistics.
- Concurrent power and temperature computation solutions, which mainly mix the capabilities of the above instruments.

For chip/package thermal simulation [3] introduces Hotspot, a boundary conditions independent compact thermal model. Hotspot package assumes that a per-floorplan-block power trace is given as an input. It performs thermal simulation based on a given floorplan and does not perform automatic mesh refinement and iterative thermal simulation based on the results. [4] and [5] introduce ISAC and NanoHeat. The two packages together create a platform capable of performing thermal simulation at different scales of spatial and temporal resolutions. The tool provides two solvers, one based on Fourier heat conduction, and the other based on Boltzman heat transfer equations (BTE)[1].

The multi-scale analysis capability of the mentioned tools is confined to the provided power trace by the user from outside. This is one important motivation that our solution, MiMAPT, performs joint power and temperature calculation at different scales of resolution.

[6] introduces Logi-Therm. The tool performs concurrent electrical and thermal simulation of standard cell ASIC circuits. It takes standard cells of digital design as basic building blocks and based on cell's power characteristics and switching activity calculates a power/thermal distribution map of the chip. Analysis done by Logi-Therm however are based on a fixed mesh resolution.

[7] introduces ICTherm, a tool developed for thermal analysis of 3D chips. The tool first performs a thermal evaluation of the target with a very high spatial resolution. Based on the results it creates a multi-granularity mesh which is used for thermal simulation. The mesh however is fixed during the rest

 ISBN:978-2-35500-022-5

of the analysis. Instead, MiMAPT changes the mesh dynamically based on the current on-die temperature distribution.

Power estimation is an extensively explored research area: Wattch [8] and SimpleScalar [9] provide basic power models for CPU cores. McPAT [10] extends this capability to wider range of architectures including multi-core clusters with network-on-chips, DRAM-controllers and high-speed interfaces. The above tools however, do not provide power at different levels of granularity. Moreover power estimation is based on pre-defined full-custom architectures thus, in practice these tools do not produce accurate enough results for newly designed hardware and for an ASIC implementation style.

[11] provides a tool for concurrent power, performance and temperature estimation. The tool however works at micro-architecture level and not RTL and so it does not provide enough accuracy for power and temperature estimation especially with new hardware blocks.

As of the author's best knowledge, no academic package, meets the following requirements together:

1- Power and temperature estimation at RT and Gate level with different scales of spatial resolution.
2- Handling non-uniform shapes of design sub-modules for thermal simulation.
3- Seamless integration into major design tool flows and compatibility with widely used library standards.
4- Compatibility and open interfaces with commercial and academic thermal analysis tools (such as Hotspot[3], 3D-ICE[12] and FloTHERM[13]).

Considering commercial software packages, Gradient Design Automation is known to be able to perform concurrent power/thermal analysis with multiple scales of temporal and spatial resolution in transient and steady state [14], [15]. Compared to Gradient's solution, MiMAPT provides wider range of analysis speed and accuracy, since it basically performs thermal analysis at RT level, which is very fast, and switches to gate-level only when higher levels of resolution are demanded.

III. MiMAPT Architecture

Our approach leverages power analysis features provided by state-of-the-art commercial tools. Recent versions of logic synthesis tools (e.g. Cadence RC® and Synopsys DC®) are capable of estimating power at RTL before doing synthesis based on switching activity obtained from functional logic simulation. RTL power estimation can be done very fast, but it is not very accurate[16], [17]. On the other hand, very accurate power at gate level can be obtained after (or during) back-end flow. The power analysis tool should be provided with the finished (placed, routed, clock tree synthesized) design and also switching activity statistics obtained from logic simulation of the finished netlist with timing delays annotated through an SDF file.

In classical thermal simulation flow the power estimation at gate-level is used to obtain per cell power values. This fine-grain power map is then used as input to fine-grain thermal simulation [5], [3]. MiMAPT instead performs power/thermal simulation at RT level with the goal of avoiding

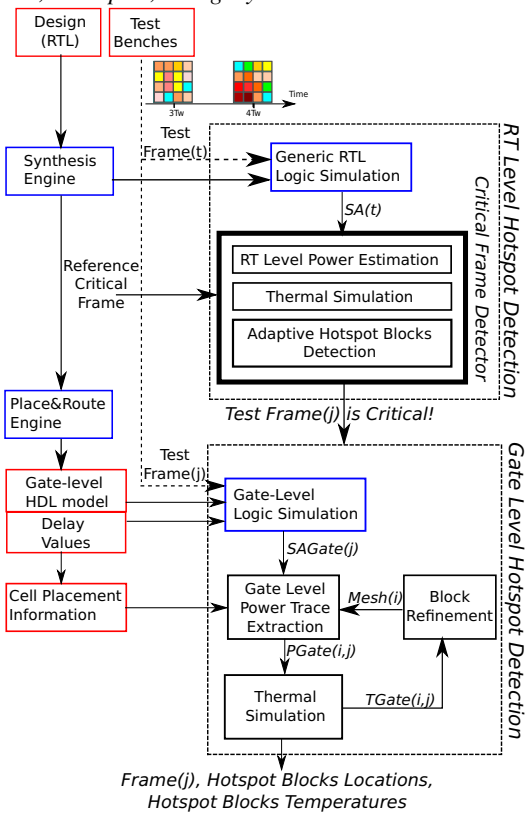


Fig. 1. MiMAPT Block Diagram

time consuming gate-level simulation when hotspots do not exist. This is achieved by using faster but less accurate, RT level power/thermal simulation to filter out non-critical (non-hotspot) portions of a die floorplan. A die area is defined as hotspot when its temperature (T) is higher than a specified threshold (TH). Figure 1 shows the basic building elements of MiMAPT which are two major parts: RTL and gate-level.

We partition the simulation input sequence in subsequences, called *test frames*, typically representing different use cases in a real design. We then dynamically switch to gate-level for any arbitrary *test frame* if needed. We use the RTL state to initialize the gate-level simulation and prepare the input patterns for the *test frame* to simulate at gate-level.

A. RT Level Hotspot Detection

We perform logic simulation for each of the *test frames* at RTL to obtain switching activity ($SA(t)$). This information are then fed to the synthesis tool to obtain power estimation for each of the design sub-modules.

We define a new *thermal floorplan* for the design which divides chip area into equally sized rectangular blocks. For each floorplan block (FB), we calculate what percentage of each sub-module is located inside this block. Based on the percentage we add a fraction of sub-module's power to the total power of the block.

Figure 2 shows this procedure in more detail for a sample design. In this figure, *(a)* shows the defined floorplan for the

 ISBN:978-2-35500-022-5

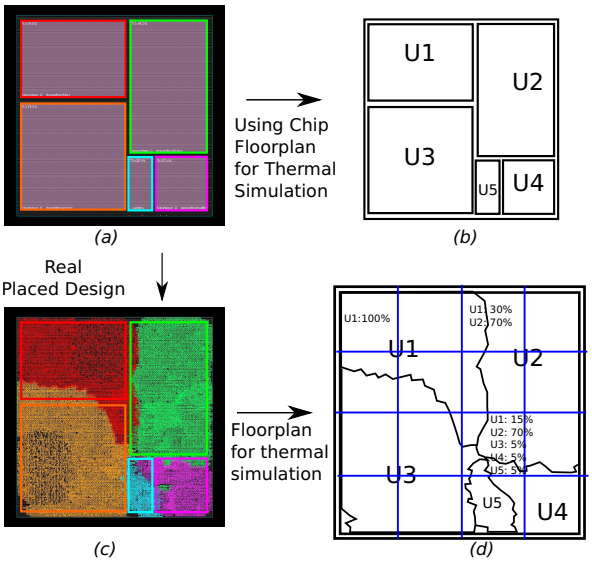


Fig. 2. Generation of power map for thermal simulation at RTL

chip in the layout tool. *(b)* shows the traditional method of creating *thermal floorplan* in which we use the dimensions defined by the default chip floorplan directly. *(c)* shows the chip after placement, as we can see, the real placement of sub-modules is different than the default floorplan thus an accurate thermal simulation is not possible using traditional methods. In *(d)* we show our method of defining a new *thermal floorplan*. Each floorplan block contains one or more design sub-modules. The synthesis tool provides us with the power consumption of each sub-module. Using chip placement information that we have, we obtain the percentage of each sub-module inside each floorplan block by counting the number of cells owned by this sub-module inside the floorplan block. We suppose that the sub-module's power is uniformly divided between its cells thus, we use the obtained percentage to add the fraction of sub-module's power to the total power of the floorplan block.

Final obtained power map is then used by the thermal simulator[3] to estimate per-block temperature map. This thermal map is then compared with a set of adaptive thresholds to identify if the *test frame* contains critical areas. If this situation is detected we trigger the gate level hotspot detection for the same frame. As shown in experimental results, RTL hotspot detection executes significantly faster than the gate-level simulation.

Estimated power at RT level is usually not equal to gate-level power since the design is not fully synthesized yet. Consequently, using a unique temperature threshold value to identify hotspot blocks at RT and gate-level may not lead to accurate detection of hotspots at RTL. Thus we use an adaptive method to detect hotspots at RTL.

We first select the *test frame* in which every design sub-module has highest level of activity and thus power, as reference. We perform thermal simulation for this *test frame* at RTL and gate-level and we mark critical floorplan blocks (FB) by comparing their temperature values at gate-level ($CMapGate$) to a threshold ($THigh$) which contains a safe margin with respect to TH and is slightly lower (e.g. 1°C) than it. We use the computed critical blocks map ($CritMatrix$) for identifying hotspots at RT level for the rest of *test frames*.

For each *test frame* we perform thermal simulation at RTL and we compare each block temperature value with an adaptive threshold. If the block is marked as critical, we act more carefully, thus we create a reduced threshold value by multiplying a coefficient A smaller than 1.0 to the reference block temperature ($CMapRTL$). If the block is not critical we use an increased threshold value by multiplying a coefficient B bigger than 1.0 to the reference block temperature to avoid un-necessary hotspot detection. The smaller values for A make detection of hotspots at RT level safer however, they decrease overall operation speed. Psudocode 1 describes this in detail.

Algorithm 1 Adaptive Hotspot Detection

Require: CMapRTL, CMapGate= Highest-workload frame, temperature map at RTL and gate-level
Require: TMapRTL= current frame temperature map
Require: THigh= threshold value for critical block
Require: A, B: Coefficients ($A < 1.0$) and ($B > 1.0$)
Require: N= Total number of thermal floorplan blocks
for $i = 1 \rightarrow N$ **do**
  $t = CMAPGate(i)$
  if $t > (THigh)$ **then**
    $CritMatrix(i) = 1$
  else
    $CritMatrix(i) = 0$
  end if
end for
for $i = 1 \rightarrow N$ **do**
  $t = TMapRTL(i)$
  if $CritMatrix(i) == 1$ **then**
    if $t > A \times CMapRTL(i)$ **then**
      This Block is hotspot!
    end if
  else
    if $t > B \times CMapRTL(i)$ **then**
      This Block is hotspot!
      $CritMatrix(i) = 1$
    end if
  end if
end for

B. Gate Level Hotspot Detection

If *test frame* (j) is detected as critical in RTL hotspot detector, it will be processed with high accuracy at gate-level to correctly estimate the hotspot position and temperature. This is done by first performing logic simulation at gate-level to obtain circuit switching activity ($SAGate(j)$) used for power estimation. The power map ($PGate(i, j)$) is then converted to temperature ($TGate(i, j)$) using an iterative multi-granularity

 ISBN:978-2-35500-022-5

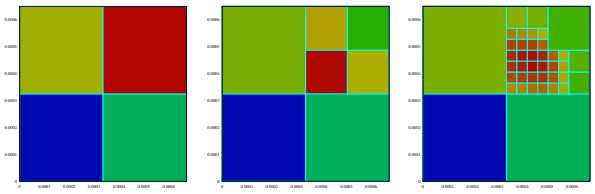


Fig. 3. Increasing spatial resolution of thermal analysis for the area of interest

meshing scheme. Starting from an initial mesh granularity ($initL$) in each iteration (i) we increase spatial resolution for on-die areas which are suspected to contain hotspots. Indeed, in each iteration, we examine the temperature map for the current thermal floorplan. For every floorplan block with a temperature value higher than TH we break the block into $M \times M$ equal sized smaller blocks. The process will continue until the finest spatial granularity ($finL$) is reached. $initL$, M and $finL$ are constants mainly defined by the user. They should be selected according to the chip die area and the desired spatial resolution and the accuracy with which detection of hotspots should be done. Figure 3 shows an example output of our multi-granularity thermal simulation for three continuous iterations.

IV. Experimental Results

We first evaluate our adaptive hotspot detection algorithm at RTL. Based on real power values, we create an ensemble of virtual gate-level and RTL power maps. RTL power maps are obtained by adding Gaussian random variables to per-floorplan block power values at gate-level. We change random variable's characteristics to simulate different situations of RTL power estimation.

Figure 4 shows total power values for 180 different virtual test cases that we have. For each power map at gate-level we create 10 different power maps at RTL by changing the mean of the Gaussian random variable ($\mu = \{-0.2, -0.1, 0.0, 0.1, 0.2\}$). We compare the output of thermal simulation for all of the gate-level and RTL power pairs. For each pair we obtain the number of hotspots and their locations at gate-level and we compare it with the output of hotspot detection methods at RTL.

Figure 5 shows the performance of our adaptive method (*A-Temp*) compared to using a unique user defined threshold value (*TH Only*). In this figure, *(a)* shows the percentage of situations that hotspots exist in the chip and gate-level simulation should be triggered, compared to percentage of situations that each of *A-Temp* and *TH Only* trigger gate-level simulation. *(b)* shows percentage of cases in which the estimated spatial location of hotspot by each of *A-Temp* and *TH Only* is different than its estimated location at gate-level. As we can see, *A-Temp* estimates the spatial location of hotspots correctly in all of the situations. Finally, *(c)* shows the percentage of detected false positives and false negatives for each of *A-Temp* and *TH Only* methods considering all of the 180 test cases. Different than *TH Only*, false-negative for our method is equal to zero which means it captures all of the hotspots completely.

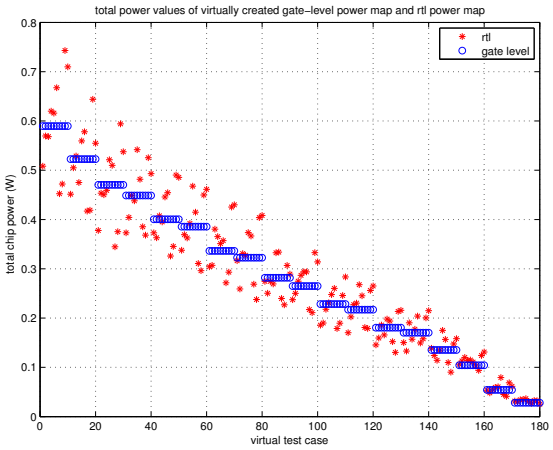


Fig. 4. virtual power values at RTL and gate-level used for evaluation of adaptive method.

TABLE I
Sample Chip : AES, FFT and FPU. (3 separate clock domains).

Block	Area(mm^2)	#Cells	FFs	ClkBuf	F(MHz)
FPU	0.2730	41477	499	13	143
FFT	0.6997	81651	42684	875	525
AES	0.4758	110758	7882	167	1328
Top	1.49	233887	51065	1055	-

To evaluate MiMAPT for a real test case, we create a sample chip containing 3 widely used digital IP blocks (AES, FPU and FFT), and fully perform synthesis, placement, CTS and routing using TSMC 65nm standard-cell library. The results are based on typical corner case of $VDD = 1.2V$ and $T = 25°C$. Table I shows chip's key specifications.

Six different *test frames* are created to evaluate MiMAPT. Figure 6 shows for each *test frame*, total power of the design

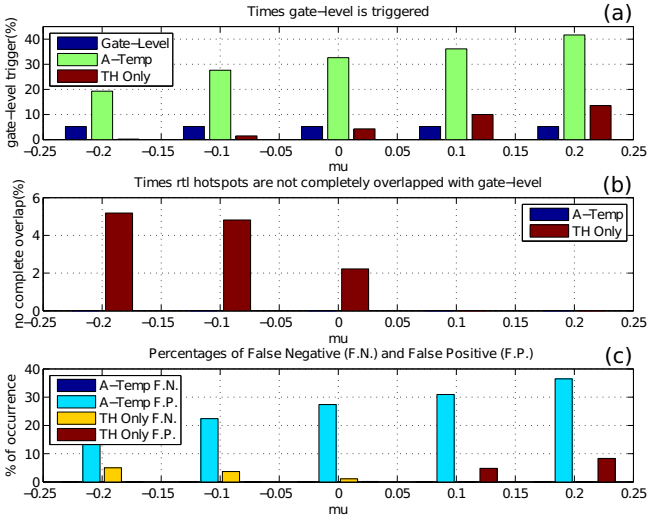


Fig. 5. hotspot detection methods comparison (Adaptive vs. TH only)

ISBN:978-2-35500-022-5

estimated at RTL and gate-level. Duration of each *test frame* is 0.2 seconds.

We execute MiMAPT for all *test frames* and we store spatial and temporal information related to detected hotspots. We then perform power and temperature estimation of the design without MiMAPT for the same set of *test frames* at gate-level and at the finest level of granularity. We perform comparison between MiMAPT and fine-grain simulation for a sample value of $TH = 358K$.

The finest level of granularity ($finL$) is $50um$ and fine-grain floorplan contains 24x24 (total 576) blocks. RTL floorplan and initial floorplan at gate-level are 8x8 blocks ($initL$). For increasing spatial resolution of the multi-granularity mesh at gate-level, we divide each hotspot block into 3×3 equal sized smaller blocks ($M = 3$). A and B coefficients in RTL hotspot detection are 0.9 and 1.1 respectively.

We represent the results in terms of execution time and accuracy of detecting hotspot location and temperature. Among 6 available *test frames*, for 3 of them MiMAPT detects critical blocks at RTL and triggers gate-level, for the other 3, gate-level is not triggered saving time. The difference between estimated temperature by MiMAPT and fine-grain is around 0.02K. For every hotspot block at fine-grain, there exists a corresponding block in MiMAPT which has the same location and size and is announced as hotspot. As a result, the distance between location of hotspots detected by MiMAPT and fine-grain is zero. MiMAPT detects all of hotspots with a very good level of accuracy, thus there is no false negatives. When gate-level is triggered MiMAPT performs two iterations of thermal simulation to achieve required spatial resolution of $50um$. For the first iteration ($i = 1$) the thermal floorplan contains 64 and for the second one ($i = 2$) 168 blocks.

Table II shows the execution time of MiMAPT compared to fine-grain. Considering all the six *test frames*, total execution time at fine-grain is 26520 seconds, it includes required time for gate-level simulation and obtaining switching activities ($Tgate_{sim} = 1610s$) and time for fine-grain thermal simulation ($Tgate_{finthr} = 24910s$). In contrast, total execution time for MiMAPT is 1446 seconds. It contains the time for RTL logic simulation and obtaining switching activities ($Trtl_{sim} = 83s$), time for RTL thermal simulation ($Trtl_{thr} = 72s$) and time for gate-level simulation for each of the 3 detected critical test frames ($Tgate_{sim} = 908s$). Two of the *test frames* are false-positives thus they do not have hotspots at gate level and gate-level thermal simulation will be done for them only at coarse grain ($Tgate_{thrc} = 24s$). One of the test frames is critical and thus contains hotspots at gate-level and triggers MiMAPT multi-granularity engine ($Tgate_{thrm} = 335s$). In total, MiMAPT is 18.3X times faster than fine-grain at the same level of accuracy.

In order to provide a better perspective on the range of possible speedups for MiMAPT, we calculate average MiMAPT time for one *test frame* when it is either a non-critical, false positive or a critical frame. As shown in table III, for our sample chip, for an assumed experiment in which every *test frame* is non-critical, MiMAPT reaches the maximum speedup

TABLE II
EXECUTION TIMES COMPARISON: MIMAPT VS FINE-GRAIN. (TIME IN SECONDS)

Method	RTL logic Sim	RTL Thermal Sim	Gate-level logic Sim	Gate-level thermal Sim	Total
Fine-grain	-	-	1610	24910	26520
MiMAPT	83	72	908	359	1446

TABLE III
AVERAGE MIMAPT SPEED-UP OVER FINE-GRAIN FOR A GENERIC TEST FRAME FOR DIFFERENT FRAME TYPES.

Test frame Type	MiMAPT Time RTL	MiMAPT Time Gate	Example SpeedUp
Non-critical	$Trtl_{sim}$ $+Trtl_{thr}$	-	171X
False positive	$Trtl_{sim}$ $+Trtl_{thr}$	$Tgate_{sim}$ $+Tgate_{thrc}$	13X
Critical	$Trtl_{sim}$ $+Trtl_{thr}$	$Tgate_{sim}$ $+Tgate_{thrc} + Tgate_{thrm}$	7X

of 171X. However when every *test frame* is false-positive, the speedup decreases to 13X. Finally if every *test frame* is critical the speedup reaches a lower bound of 7X.

V. CONCLUSIONS

We proposed MiMAPT, and described its approach to power/thermal simulation at RT and gate level. MiMAPT highly accelerates power/thermal simulation while keeping accuracy at acceptable levels. For the developed sample chip, we observed an speed-up range between 7X to 170X (with a typical value of around 18X) over classical method while providing the same level of hotspot detection accuracy.

ACKNOWLEDGMENT

This work was supported, in parts, by the EU FP7 ERC Project MULTITHERMAN (GA n. 291125) and the EU FP7 Project THERMINATOR (GA n. 248603).

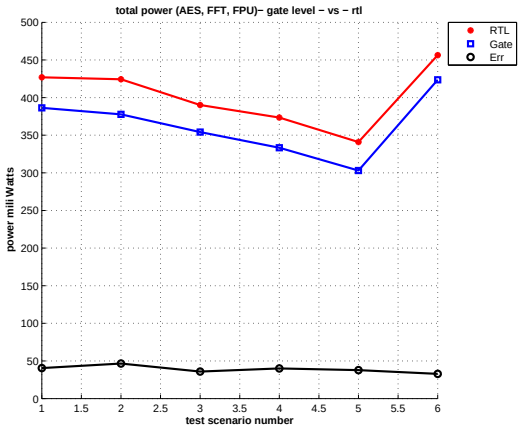


Fig. 6. Total power of each test case at RT and gate-level

ISBN:978-2-35500-022-5

REFERENCES

[1] J. Rowlette *et al.*, "Fully coupled nonequilibrium electronphonon transport in nanometer-scale silicon fets," *IEEE TRANSACTIONS ON ELECTRON DEVICES*, vol. 55, no. 1, pp. 220–232, January 2008.

[2] M. Sadri, A. Bartolini, and L. Benini, "Single-chip cloud computer thermal model," in *Thermal Investigations of ICs and Systems (THERMINIC), 2011 17th International Workshop on*, sept. 2011, pp. 1 –6.

[3] W. Huang *et al.*, "Hotspot: a compact thermal modeling methodology for early-stage vlsi design," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 14, no. 5, may 2006.

[4] Y. Yang *et al.*, "Isac: Integrated space-and-time-adaptive chip-package thermal analysis," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 26, no. 1, jan. 2007.

[5] Z. Hassan *et al.*, "Full-spectrum spatial-temporal dynamic thermal analysis for nanometer-scale integrated circuits," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 19, no. 12, dec. 2011.

[6] A. Timar *et al.*, "Electro-thermal co-simulation of ics with runtime back-annotation capability," in *Thermal Investigations of ICs and Systems (THERMINIC), 2010 16th International Workshop on*, oct. 2010.

[7] A. Fourmigue *et al.*, "Multi-granularity thermal evaluation of 3d mpsoc architectures," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2011*, march 2011.

[8] D. Brooks *et al.*, "Wattch: a framework for architectural-level power analysis and optimizations," in *Computer Architecture, 2000. Proceedings of the 27th International Symposium on*, june 2000.

[9] T. Austin, E. Larson, and D. Ernst, "Simplescalar: an infrastructure for computer system modeling," *Computer*, vol. 35, no. 2, feb 2002.

[10] S. Li *et al.*, "Mcpat: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Microarchitecture, 2009. MICRO-42.*

[11] M.-y. Hsieh *et al.*, "Sst: A scalable parallel framework for architecture-level performance, power, area and thermal simulation," *The Computer Journal*, 2011.

[12] A. Sridhar *et al.*, "3d-ice: Fast compact transient thermal modeling for 3d ics with inter-tier liquid cooling," in *Computer-Aided Design (ICCAD), 2010 IEEE/ACM International Conference on*, nov. 2010, pp. 463 –470.

[13] M. Graphics, *FloTHERM: Optimizing the Thermal Design of Electronics*, May 2011.

[14] G. Design, "Us patent 7823102, thermally aware design modification," Tech. Rep., 2011.

[15] ——, "Us patent 8019580, transient thermal analysis," Tech. Rep., 2011.

[16] Cadence, *Low Power in Encounter RTL Compiler*, Cadence Design Systems, April 2010.

[17] Synopsys, *Design Compiler User Guide*, Synopsys, December 2011.

25-27 September 2012, Budapest, Hungary

Thermal Sensor Distribution Method for 3D Integrated Circuits Using Efficient Thermal Map Modeling

Fatemeh Kashfi and Jeff Draper
Information Sciences Institute, University of Southern California
Marina del Rey, CA 90292
Email: fkashfi@usc.edu, draper@isi.edu

***Abstract*-Three Dimensional Integrated Circuit (3DIC) technology has been introduced to address the interconnect issues in nanometer circuit design that often limit performance improvement and power reduction. However, stacking active layers of silicon will lead to increased power density and overall higher temperatures in a 3D chip implementation for many designs. Thermal sensors are therefore crucial for run-time thermal management of 3DICs. A thermal sensor distribution method customized for 3DICs is introduced in this paper. A new 3D thermal map modeling which facilitates efficient and very fast analyses is embodied in this thermal sensor distribution algorithm. The 3D thermal map modeling is based on scaled hotspot areas, depending on the distance of a stacked layer from the heatsink and also thermal effects of the layers on each other. Our results indicate that for a 4-layer stacked 3DIC, consisting of two layers of quad-core processors and one layer of L2 cache and one layer of main memory, less than 4.4% in maximum sensors reading error can be accomplished with a 53x speedup in the thermal evaluation time and thermal sensor distribution algorithm implementation. Furthermore, the number of needed sensors is 44% lower than that of conventional 2D thermal sensor distribution methods.**

I. INTRODUCTION

Three Dimensional Integrated Circuit (3DIC) technology is emerging to extend Moore's law along a vertical dimension. Device scaling is reaching a limitation because of short-channel effects, increasing variability and power dissipation. 3DIC technology makes it possible to continue Moore's law further not by scaling, but by stacking active layers of transistors on top of each other [1]. 3D stacking can potentially improve interconnect performance by reducing the total wire length on a chip, which decreases both latency and power consumption. Also, 3D stacking has the potential to reduce chip area and makes it possible to integrate dissimilar technologies in one chip. There are also some challenges for this technology to mature and be widely used, e.g., accommodating layers of differing technologies using TSVs (Through Silicon Vias), testing of the stacked layers, and higher junction temperatures due to increased power density. Thermal issues will also worsen because of limitations in using cooling channels between the layers, and because the layers which are further from the heatsink may get overheated [2] [3]. A good understanding of the thermal behavior of 3DICs is necessary for devising design and packaging techniques and power saving methods to mitigate the thermal problems in 3DICs.

Thermal sensors must be allocated inside 3DICs for run-time thermal measurement and management of the chip. The thermal sensor distribution problem has been widely studied for conventional 2D chips. In [4] locations of the sensors are determined after providing an analytical model that describes the maximum temperature difference between a hotspot in the IC and a region of interest. In [5] uniform and non-uniform sensor allocations are compared. The non-uniform allocation identifies an optimal physical location for each sensor such that the sensor's attraction toward steep thermal gradients is maximized. In [6] after locating high energy regions of the chip, two methods of energy-center sensor allocation and energy-cluster sensor allocation are compared, while in the latter, sensor locations are determined using the k-means clustering algorithm, which attempts to strike a balance between hotspot estimation and full thermal characterization. These methods should be further developed and customized for 3DICs, considering the special thermal behavior of this technology.

Any reasonable thermal sensor allocation algorithm for 3DICs requires a 3D thermal map of the IC, and such a thermal map, which is generated by a power profile of the chip, is workload dependent. A thermal sensor allocation algorithm should consider possible thermal maps of the 3DIC to find an optimum number of sensors and their locations for minimum possible reading errors. Using detailed thermal map modeling is therefore very time consuming given all the workload scenarios that must be considered. In this paper we provide a very fast and efficient 3D thermal map modeling approach to be used for thermal sensor allocation. The model, while very fast in execution, yields acceptable error tolerance for the purpose of a thermal sensor distribution algorithm. Furthermore, the approach gives insight for general thermal behaviors in 3DICs, including possible locations of hotspots.

In recent years different methods for thermal map modeling of 3DICs have been proposed. The methods are mainly divided into three categories. The Finite Difference Method (FDM) solves Partial Differential Equations (PDE) of the heat conduction with constraint boundaries. The method is very accurate in modeling both steady-state and transient thermal behavior in the rectangular geometry of a 3DIC. The main disadvantage of this method is its high computational cost [7][8]. Compact modeling divides a 3DIC into thermal grids and based on electrical-thermal analogy equations, solves the thermal relationship between neighboring cells. A transient temperature response is calculated given the physical characteristics and power consumption of units on the die. Compact modeling still very accurately reduces the problem size and expedites thermal simulations [9][10][11]. Fast methods sacrifice some accuracy to reach faster thermal modeling simulation and can be used in thermal management algorithms to get insight into the spatial thermal map of the 3DIC for all possible applications. In [12] a fast thermal simulation for 3DICs is proposed using a Neural Network heuristic. Although the speedup is considerable, using this model still needs a time consuming training phase for random power inputs for each thermal cell.

We approached the problem of 3DIC thermal map modeling from a new angle. Our model is based on common thermal scaling behaviors of 3DICs and can be used for both transient and steady-state thermal analysis of a 3DIC. The method is based on scaled hotspot areas, depending on the distance of a stacked layer from the heatsink and also thermal correlations between the layers. We model each layer's thermal map in a 3DIC as a superposition of its own thermal map, after proper scaling based on its location in the 3DIC, and scaled reflections of other layers' thermal maps on that specific layer. By finding proper thermal

 ISBN:978-2-35500-022-5

scaling factors between layers, a characterization of the chip thermal map for all applications can be obtained.

The thermal sensor distribution method for 3DICs proposed in this paper is developed in two steps. First, a 3D thermal map model is provided to be used in the thermal sensor distribution algorithm. Then a thermal sensor distribution algorithm for 3DICs is proposed which is thermal gradient-aware and employs a k-means clustering algorithm in a 3D manner. Taking advantage of the high thermal correlation between adjacent layers, we claim that any thermal sensor distribution algorithm should be solved as a 3D problem, not for individual layer separately, to avoid excessive assignment of sensors to the same spatial hotspot.

The organization of the paper is as follows: Section II provides the explanation of the proposed 3D thermal map model. In section III an algorithm for thermal sensor distribution for 3DICs is given. Experimental results are presented in section IV, and section V concludes the paper.

II. 3D Thermal Map Modeling for 3DICs

In this section we propose a method for 3D thermal map modeling of 3DICs. In the proposed method we rely on the facts that any hotspot area in a layer inside the 3DIC will enlarge if the layer gets further from the heatsink, and a scaled version of each layer's thermal map will be reflected on other layers. This method helps us to apply a detailed thermal analysis on each 2D layer using the existing IC thermal analysis CAD tools and apply a simplified scaling method to attain the whole 3DIC spatial thermal map.

A. Thermal map scaling based on the layer location in 3DIC

In the first step toward the accomplishment of the 3D thermal map model we study how the thermal map of a single layer will change based on its location in the 3DIC.

Theorem: With the same source of heat, the hotspot area increases as a layer gets further from the heatsink (hotspot area is the area around a heat source that has the possibility of having a temperature greater than a $T_{critical}$).

Proof: Let's assume that a heat source is located in layer i and the hotspot area around it has radius r. The border of the area has a temperature of T_b. The layer's thermal resistance to the heatsink is $R_{i\text{-}HS}$.

By going to layer j which is further from the heatsink we attain $R_{j\text{-}HS} > R_{i\text{-}HS}$. To have an area with the same temperature of T_b on its border, based on Equation (1), R_l (equivalent lateral thermal resistance from heat source to the border) must increase and that means radius r must increase. Therefore with the same source of heat the area of the hotspot in layer j will be greater.

$$\frac{T_s - T_b}{T_b - T_{HS}} = \frac{R_l}{R_{i-HS}} \tag{1}$$

By finding the proper scaling factors, any layer's thermal map can be obtained and then scaled based on its location relative to the heatsink. It is also necessary to model the other layers' thermal effects on a specific layer to obtain the complete solution. Layers' thermal effects on each other will be studied in the next section.

For now we assume that we have only one active layer in the 3DIC (located in layer k), and its thermal map will be solely obtained by a proper scaling of the thermal map of the layer derived from the conventional 2D method that we call *original-thermal-map* of the layer and show by the matrix O. The original-thermal-map of the layer is divided into an $n \times n$ grid, and the average temperatures of the grid cell points form the $n \times n$ matrix O. The scaling model is obtained by data interpolation of the temperature change of each grid cell for each active layer location in the 3DIC. The active layer thermal map matrix in layer k as a function of its original-thermal-map matrix can be formulated as Equation (2). We call this matrix *intermediate-thermal-map* matrix and show it by I.

$$I_{k(n\times n)} = \left(a_k O_{k(n\times n)} + b_k \Delta O_{k(n\times n)} + c_k\right). D_{k(n\times n)} \tag{2}$$

Based on Equation (2) each element of the intermediate-thermal-map matrix in layer k is a function of the corresponding element in the original-thermal-map matrix, $O_{k(n\times n)}$, the horizontal thermal gradient matrix of the original-thermal-map matrix, $\Delta O_{k(n\times n)}$, and the vertical thermal scaling matrix, $D_{k(n\times n)}$, from the heatsink up to layer k. a_k, b_k, and c_k are scalar fitting factors which are functions of lateral thermal conductivity and the location of layer k.

Each element in the thermal gradient matrix $\Delta O_{k(ij)}$ is calculated by taking the difference of the temperature of grid cell *(i,j)* from its neighboring cells' temperatures. This factor arises from the fact that a cell with much hotter neighbor cells suffers more from their temperatures as it gets further from the heatsink, and a larger scaling factor is applied accordingly.

Each element in the vertical thermal scaling matrix, $D_{k(n\times n)}$, is a function of an equivalent vertical thermal conductivity from the corresponding element in the thermal matrix of layer k to the heatsink.

Equation (3) shows how to calculate $D_{k(n\times n)}$.

$$D_{k(n\times n)} = \prod_{i=k}^{i=m-1} D_{i,i+1(n\times n)} \tag{3}$$

In this equation layer m is the closest layer to the heatsink and $D_{i,i+1}$ is an $n \times n$ vertical scaling matrix between layer i and layer $i+1$. Each element in $D_{i,i+1}$ is a function of the vertical thermal conductivity between corresponding grid cells in layer i and $i+1$ and also the location of the layers in the 3DIC relative to the heatsink.

B. Modeling of other layers' thermal effects on a specific layer

The thermal effects of layers in a 3DIC on each other can be studied through primary and secondary paths of heat transfer. The primary path of heat transfer is responsible for conducting the heat toward the heatsink and removing the heat through the heatsink. The secondary path of heat is the one toward the PCB (Printed Circuit Board). In a conventional 2D IC chip, the secondary path of heat transfer is always insignificant, because there are no additional layers of circuitry between the primary heat source and the PCB. Clearly this is not the case in a 3DIC. The secondary path of heat transfer in a 3DIC is more problematic.

If there is a source of heat in the middle of a 3DIC, the layers between the heat source and the PCB will be greatly affected because the path to the heatsink has high thermal resistivity, practically trapping the heat in these layers. Fig. 1 compares the maximum junction temperature of each layer in both the primary and secondary paths of heat transfer in a 3DIC. This sample chip consists of 7 layers (each of them consists of bulk, active Si, and metal sublayers which comprise a total of 21 sublayers under study) with the processor in the middle layer and 3 passive layers both above and below it. Thermal simulation was performed using HotSpot 5.0 [9].

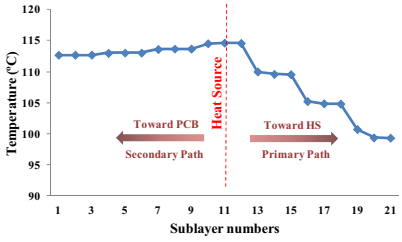


Fig. 1. Primary vs. secondary path of heat transfer in a 7-layer 3DIC using HotSpot 5.0.

 ISBN:978-2-35500-022-5

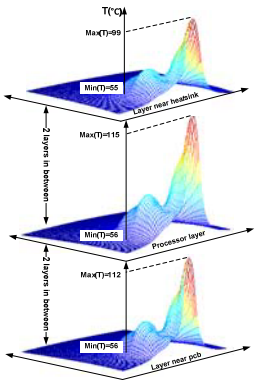


Fig. 2. Thermal map of different layers in a 7-layer 3DIC using HotSpot 5.0.

As can be seen in the figure, the temperature drops dramatically for the layers closer to the heatsink along the primary path of heat transfer, while it remains almost constant for the layers between the source of heat and the PCB along the secondary path of heat transfer.

Through both primary and secondary paths of heat transfer, a scaled replicate of a layer's thermal map will be reflected on other layers. Fig. 2 shows how the thermal map of the processor will be replicated on other layers in both primary and secondary paths for a chip with a processor in the middle layer and 3 passive layers both above and below it. This effect can be formulated as shown in Equation (4).

$$T_{k(n\times n)} = \begin{cases} \left(a_l T_{l(n\times n)} + b_l \Delta T_{l(n\times n)} + c_l\right).P_{kl(n\times n)} & k > l \\ \left(a_l T_{l(n\times n)} + b_l \Delta T_{l(n\times n)} + c_l\right).S_{kl(n\times n)} & k < l \end{cases}$$
$$P_{kl(n\times n)} = \prod_{i=l}^{i=k-1} P_{i,i+1(n\times n)} \qquad (4)$$
$$S_{kl(n\times n)} = \prod_{i=k}^{i=l-1} S_{i,i+1(n\times n)}$$

where $T_{k(n\times n)}$ is a thermal map in layer k generated solely based on the thermal effect of layer l. If T_k is located between T_l and the heatsink, the thermal map generated on it is through the primary path of heat transfer from layer l, and if T_k is located between T_l and the PCB, the thermal map generated on it is through the secondary path of heat transfer from layer l. Each element in the thermal gradient matrix $\Delta T_{l(ij)}$ is calculated by taking the difference of the temperature of grid cell *(i,j)* from its neighboring cells' temperatures. This factor arises from the fact that a cell with much hotter neighboring cells generates hotter grid cells on other layers. a_l, b_l, and c_l are scalar fitting factors which are functions of lateral thermal conductivity and the location of layer l.

The scaling matrix on the primary path is shown by P_{kl} and on the secondary path by S_{kl}. The elements of the scaling matrices are a function of equivalent thermal conductivity between corresponding elements in layer k and l and location of the layers. As the layers between the active layer l and PCB have high thermal resistance to the heatsink, with the same distance from the active layer l, elements in matrix S_{kl} are greater than elements in matrix P_{kl} , and layer l has lower thermal effect on the layers between itself and the heatsink. Each element in the scaling matrix between layer k and l is obtained by the product of corresponding elements in scaling matrices of adjacent layers between layer l and k as shown in Equation (4).

C. Scaling elements calculation

Each element in the matrices of $D_{i,i+1}$, $P_{i,i+1}$, and $S_{i,i+1}$ depends on the equivalent thermal conductivity of that grid cell and the location of the corresponding layer. To compute the equivalent thermal conductivity the fraction of the grid cell being occupied by TSVs must be considered as shown in Fig. 3. This fraction can be calculated using smaller grid cells.

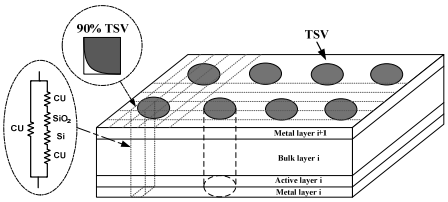


Fig. 3. Vertical thermal scaling matrix elements calculation based on the fraction occupied by TSVs (face-to-back configuration in a flip-chip 3DIC).

1% accuracy yields acceptable error tolerance.

Each scaling element can be calculated as follows:

$$K = (fK_1 + (1-f)K_2)$$
$$K_1 = {}^1/_{R_{TSV}} \quad and \quad K_2 = {}^1/_{R_{cu} + R_{SiO2} + R_{Si} + R_{cu}}$$
$$p = \alpha_{i,i+1}\beta_{(i,i+1),l}K \qquad i \geq l \qquad (5)$$
$$s = \alpha_{i,i+1}\gamma_{(i,i+1),l}K \qquad i \leq l$$
$$d = \delta_{i,i+1}/K$$

In which f is the fraction of the grid occupied by TSVs and R values are thermal resistivity of each layer as a function their height, the size of the grid cells and the material of the layers. As the grid cells are assumed to be equal in size, R values are calculated only once. For each grid cell p, s, and d which are corresponding elements of the scaling matrices of $P_{i,i+1}$, $S_{i,i+1}$, and $D_{i,i+1}$ respectively are calculated as shown in (5). $\alpha_{i,i+1}$ and $\delta_{i,i+1}$ are scaling parameters which are proportional to the distance of the layer from the heatsink. Parameters $\beta_{(i,i+1),l}$ and $\gamma_{(i,i+1),l}$ are inversely proportional to the distance from the source layer l.

D. 3D thermal map modeling

Each layer's thermal map is a combination of its scaled original thermal map and the effects of other layers. By combining Equations (2) and (4), each layer's *final-thermal-map* shown by F can be calculated as follows:

$$F_k = I_k + \sum_{i=k+1}^{m}(a_i I_i + b_i \Delta I_i + c_i).S_{ki} + \sum_{i=1}^{k-1}(a_i I_i + b_i \Delta I_i + c_i).P_{ki}$$
$$I_i = (a_i O_i + b_i \Delta O_i + c_i).D_i \qquad (6)$$
$$I_k = (a_k O_k + b_k \Delta O_k + c_k).D_k$$

P_{ki} models the thermal effect of layer i on layer k, in which layer k is located on its primary path of heat transfer, while S_{ki} models this effect when layer k is located on the secondary path of the heat transfer of layer i. (Layer m is the closest layer to the heatsink.) O_i and O_k are original-thermal-maps of layers i and k, and I_i and I_k are their intermediate-thermal-maps. F_k is the final-thermal-map of layer k.

This model can be used for efficient 3D thermal map estimation of a 3DIC with any number of active layers and any possible application set without requiring tedious, detailed, time-consuming simulations for every application and configuration. The model also gives a very clear insight into the thermal behavior of 3DICs. The model can be used for both steady-state and transient modeling of 3DICs. For the 3DICs employing cooling channels [14] the model can be extended after finding proper scaling factors corresponding to the layer containing cooling channels.

The 3D thermal modeling steps are summarized in Fig. 4. To pinpoint possible hotspots for the thermal sensor distribution algorithm, in step 7 the maximum temperature of each grid cell in each layer for all considered applications is calculated.

Existing 3D thermal analysis tools like HotSpot [9] and 3D-ICE [10] conduct very detailed thermal analyses by finding the thermal resistance and thermal capacitance between thermal nodes to model lateral and

vertical heat conduction paths. As the number of active layers and the number of applications to be run on a 3DIC increase, this kind of detailed thermal simulation is very time consuming, especially for the purpose of thermal control algorithms like thermal sensor allocation which requires a global view of the thermal maps of the 3DIC. Using the proposed method we can attain a significant savings in simulation time especially when the number of active layers and applications increases. The proposed model reduces the complexity of the problem of solving 3DIC thermal modeling by decomposing the 3D problem into interactions among essential components only in 2D layers.

3D Thermal Map Modeling Steps

0 Fill the lookup table of scaling elements
1 For each benchmark
2 For k=1:m
3 Find O_k from 2D thermal methods
4 $I_k = (a_k O_k + b_k \Delta O_k + c_k).D_k$
6 For k=1:m
7 $F_k = \mathrm{Max}\left(I_k + \sum_{i=k+1}^{m}(a_i I_i + b_i \Delta I_i + c_i).S_{ki} + \sum_{i=1}^{k-1}(a_i I_i + b_i \Delta I_i + c_i).P_{ki}\ , F_k\right)$

Fig. 4. 3D Thermal map modeling steps.

III. Thermal Sensor Distribution for 3DICs

The 3D thermal sensor distribution method used in this paper is thermal-gradient-aware using a k-means clustering algorithm in the 3D space and employs the 3D thermal map modeling technique described in section II.

A thermal-gradient-aware k-means clustering algorithm for sensor allocation in conventional 2D ICs is proposed in [5]. However, in this paper we describe the necessity of considering the spatial thermal map of the 3DIC instead of each individual active layer's thermal map for optimum sensor allocation. The k-means clustering algorithm is defined as follows: Given an integer k and a set of n data points $R = \{\vec{a}_i | (x_{i1}, x_{i2}, \dots, x_{im}), i = 1,2,\dots,n\}$ in an m-dimensional space, determine k centers such that the mean-square distance from each data point to its nearest center is minimized [5].

In the sensor allocation problem, n data points are n hotspot locations in the chip, and k centers are optimum sensor positions for monitoring the hotspots. To find the hotspot locations, each layer in the 3DIC is divided into $n \times n$ grid cells, and the spatial thermal map of the 3DIC for the application set is derived. Because of the large set of applications, especially for multi-core processors, thermal map modeling should be fast while accurate enough to give the location of all possible hotspots. A hotspot is any grid cell with possibility of having a temperature higher than a specified $T_{critical}$. A hotspot will be located by its three Euclidian coordinates in the 3DIC. The k-means clustering algorithm will then be solved considering the position of the hotspots in the 3D space. To make the allocation thermally aware, the maximum temperature of each hotspot will be added as the 4th dimension of that hotspot and will be applied to the k-means clustering algorithm. In this way a sensor will be allocated closer to the hotspots with higher temperatures. Although in this way the problem will be solved in a 4 dimensional space, we call the algorithm 3D k-means clustering as it is solved in a 3D Euclidian space in contrast with the problem to be solved for each 2D layer individually that we call 2D k-means clustering algorithm.

Because of the physical adjacency and use of high thermal conductive TSVs between the layers, the thermal maps of adjacent layers are highly correlated to each other. Table 4 shows the thermal correlation between the layers in a typical 4-layer 3DIC. Because of this high thermal correlation any sensor's temperature can also represent the temperature of its immediate neighbor cells in the adjacent layers. The high thermal correlation between the layers also shows the necessity of monitoring spatial hotspots in 3DICs instead of planar hotspots in each layer for further thermal control of the 3DIC. A hotspot in a layer sometimes needs to be controlled by applying thermal control techniques on its adjacent layers.

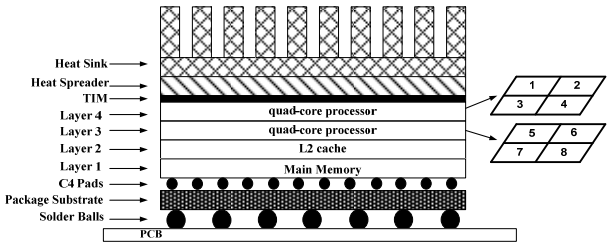


Fig. 5. 3DIC stacked layers (Flip-chip configuration).

In the next section through experiments we show the efficiency of using k-means clustering considering a 3D thermal map of the chip instead of solving thermal sensor allocation for each individual active layer separately.

IV. Experimental Results

This section provides experimental results for the proposed thermal modeling and sensor allocation technique described in previous sections. In all experiments the Alpha 21264 is used as a baseline processor [15]. SPEC2000 benchmarks are used as the benchmarks to be run on the processors [16]. The Wattch infrastructure [17] is used for architectural-level power modeling of the system, and HotSpot 5.0 [9] is used for grid-based thermal simulation of the 3DIC. For the packaging parameters, we used HotSpot 5.0 defaults [9].

The 4-layer 3DIC in our experiments as shown in Fig. 5 consists of 8 cores of Alpha 21264 in 90nm. The top two layers by the heatsink consist of 4 cores each, and the next layer is dedicated to the L2 cache. The last layer which is the closest to the PCB contains main memory. The second layer is rotated 180 degrees to balance the chip power profile [18]. The 3DIC layers are stacked face-to-back. 16MB of L2 cache is shared between all cores, and the size of main memory is 1024MB. The Alpha 21264 configuration is shown in Table 1. The material properties of the 3DIC in our experiments are shown in Table 2.

Each layer is divided into 128×128 grid cells. The elements of the 128×128 scaling matrices *D*, *S*, *P* for each two adjacent layers is obtained based on the percentage of the prism, connecting two corresponding elements in two adjacent active layers vertically, occupied by TSVs. The data is then gathered in a lookup table to be used for any considered

TABLE 1. Alpha 21264 configuration.

Die size	4.65×4.65mm^2
Frequency and Voltage	2GHz, 1.2V
Instruction Queue	64 entries
Functional unit	4IXU, 2FPU, 1BPU
Branch predictor	1K local, 4K global
L1 DCache/core	32KB, 2-way, 64B blocks, 3 cycle lat
L1 ICache/core	64KB, 2-way, 64B blocks, 1 cycle lat
Shared L2 cache	16MB, 8-way LRU, 64B blocks, 25 cycle lat

TABLE 2. 3DIC material properties.

3DIC layers	Specific heat capacity (J/m^{3}K)	Thermal Conductivity (W/m.K)	Thickness (μm)
Bulk (SiO_2)	1.96×10^6	1.2	100
Active (Si)	1.63×10^6	100	0.1
Metal (Cu)	3.45×10^6	400	10
Heat sink	3.55×10^6	400	6900
Heat spreader	3.55×10^6	400	1000
TIM	4×10^6	4	20

TABLE 3. POWER PROFILE OF THE BENCHMARKS.

Benchmarks	Avg power (W)	Power Distribution (%)					
		branch pred	Icache	Dcache	Functional units	Clock	Others
apsi	33.5	2.4	13.4	43.8	7.2	22.5	10.7
equake	25.3	2.6	8.1	47	8.1	20.8	13.4
gcc	28.3	2	6.5	53.2	6.5	19.8	12
bzip	21	4.6	17	14	12.3	31.9	20.2

application. For each two layers this scaling factor is then obtained by a combination of the scaling factors of each two adjacent layers between these two layers.

We evaluated six different applications run on the 3DIC's 8-core processor. These applications are combinations of the benchmarks shown in Table 3. With different power profiles of the selected benchmarks we attained a wide variety of spatial thermal maps for the 3DIC to examine our 3D thermal map modeling and the efficiency of the 3D k-means clustering algorithm for sensor allocation in a 3DIC. The applications are shown in Table 5.

Using the 3D thermal modeling steps shown in Fig. 4, we modeled the spatial thermal map of the 3DIC for our six applications. The maximum of the modeling error observed was in the layer closest to the PCB in application 4. Fig. 6 shows the histogram of the 3D thermal map modeling error for that layer. The layer is divided into 128×128 grid cells, and the error represents the difference between the temperatures of the grid cells calculated with HotSpot 5.0 and the results provided by our proposed 3D thermal map modeling. As shown in the figure the mean of the modeling error is 2.5% while the maximum error is less than 5.5% which is quite acceptable for the purpose of thermal sensor allocation algorithms.

For each individual application, using the same platform, the speedup in using our 3D thermal map modeling is 53x compared to HotSpot 5.0 thermal modeling. HotSpot 5.0 would have to be run again for each application for the entire 3DIC, while our model would just need to be re-evaluated but not recalibrated. This speedup is significant, given the vast number of scenarios that must be evaluated for a complete thermal analysis of typical 3DIC systems and applications.

Because of the high thermal correlation between the layers in a 3DIC we contend that the k-means clustering algorithm should be solved as a 3D problem instead of solving it for each layer individually. Table 4 shows the high thermal correlation of the layers tested for a large set of applications.

To show the efficiency of using the k-means clustering algorithm in the 3D space instead of solving the problem for each individual layer we conducted the following experiment. Each layer in the 4-layer 3DIC is divided into 16×16 macro cells (each macro cell consists of 8×8 grid cells). After deriving the spatial thermal map of the 3DIC and pinpointing possible hotspots, we run both 2D and 3D k-means clustering algorithms.

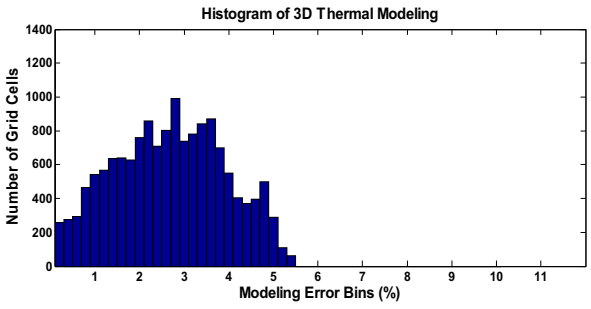


Fig. 6. Histogram of 3D thermal map modeling error.

TABLE 4. THERMAL CORRELATION BETWEEN LAYERS IN THE 4-LAYER 3DIC.

Thermal Correlation	layer 1 Main Memory	layer 2 L2 Cache	layer 3 Quad-core processor	layer 4 Quad-core processor
layer 1	1.00	0.98	0.93	0.87
layer 2	0.98	1.00	0.97	0.92
layer 3	0.93	0.97	1.00	0.97
layer 4	0.87	0.92	0.97	1.00

In both methods a sensor located in a macro cell represents the maximum temperature. For a macro cell that contains no sensor, if its immediate lateral or vertical macro cell neighbors contain sensors, its temperature is the average temperature of its neighbor sensors [4]. This approximation is reasonable because of the high vertical and lateral thermal correlation between the neighbor macro cells but will also be evaluated in the experimental results. If the temperature of any macro cell cannot be determined using this method we say that this macro cell is not in the coverage area. The goal is to cover critical areas, which are any macro cells containing possible hotspots with acceptable sensor reading error tolerance.

Fig. 7 shows percentage of critical area coverage for both 2D and 3D k-means clustering as the number of sensors increases. The sensor reading error tolerance for both methods is kept the same, which is less than 5%. To attain a much smaller error tolerance, the size of macro cells must be set smaller.

We can see that with the same number of sensors and error tolerance, using 3D k-means clustering covers a much higher percentage of the critical macro cells than 2D k-means clustering, due to the high thermal correlation between the layers. For the case using 2D k-means clustering for each layer separately, a sensor allocation results with sensors vertically close to each other because any planar hotspot creates hotspots on adjacent layers too. On the other hand by using 3D k-means clustering a spatial hotspot will be considered one data point, so a smaller number of sensors will be assigned to it. We can see that with 60 sensors we can monitor all critical macro cells in the 3DIC for an error of less than 5% while a much greater 108 sensors are needed if we use a 2D k-means clustering algorithm. Fig. 7 also shows that with increasing the number of sensors and using 2D k-means clustering we may not reach higher coverage, again because of assigning an excessive number of unnecessary sensors to the same spatial hotspots.

The optimum thermal sensor positions using a 3D k-means clustering algorithm for sensor allocation are shown in Fig. 8. We can see that with a minimum number of sensors, for 100% coverage of the critical area and an acceptable reading error of less than 5%, thermal sensors are only located in middle layers and they also monitor their adjacent layers' temperatures. Note that the 3D thermal map shown in the figure is for one application, while the sensor allocation considers all six thermal maps of the 3DIC. Figure shows the macro cells that contain sensors.

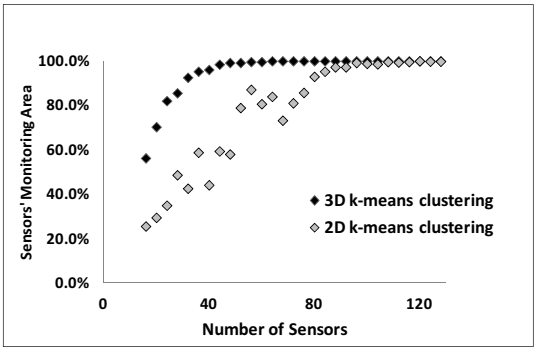


Fig. 7 Sensors' monitoring area vs. number of sensors using 2D and 3D k-means clustering algorithm.

 ISBN:978-2-35500-022-5

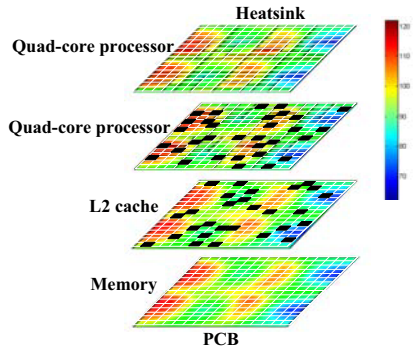


Fig. 8. Thermal sensor allocation using 3D thermal map modeling and 3D k-means clustering algorithm.

Table 5 shows the combination of benchmarks for each application used to create various 3D thermal maps for the 3DIC. The "Max Modeling Error" column shows the maximum grid cell's temperature error of our 3D thermal map modeling as compared to the temperature calculated by HotSpot 5.0 grid-based thermal simulation. The "Max Sensors Reading Error" column shows the maximum sensor reading error for the critical areas that are covered by the proposed thermal sensor distribution. The sensor reading temperature is determined based on the temperature of the grid cell in which it is located and using our 3D thermal map modeling. The maximum temperature of critical macro cells determined by the sensors is then compared with the actual maximum temperature of those macro cells computed with HotSpot 5.0. The column shows the maximum of these errors.

The small magnitude of the sensor reading errors demonstrate the efficiency of our distribution method. The maximum sensor reading error in the critical areas is less than the maximum 3D thermal map modeling error of the whole 3DIC, demonstrating that the model can predict steep thermal gradients quite precisely. This shows another great advantage of using this fast 3D thermal map modeling.

The table shows that the maximum 3D modeling error is 5.46%, and the maximum sensor reading error is 4.40%. This small sacrifice in accuracy is impressive given that the solution was attained with a 53x speedup compared to HotSpot 5.0 thermal modeling and uses 44% fewer sensors than the number needed if using conventional 2D k-means clustering.

V. CONCLUSION

In this paper we proposed a new thermal map modeling and sensor distribution technique for 3DICs. The proposed 3D thermal map modeling relies on two factors: the scaled hotspot area based on the distance of the stacked layer from the heatsink, and the thermal effect of each active layer on other layers. The model is very fast and efficient and gives a clear insight into the thermal behavior in 3DICs. The model is utilized to generate the 3D thermal map of a sample 4-layer stacked 3DIC, consisting of two layers of quad-core processors and one layer of L2 cache and one layer of main memory. For different applications running on the processor the proposed modeling yields maximum error of less than 5.5%, which is quite acceptable for the purpose of a sensor distribution algorithm. The 3D thermal sensor distribution is based on k-means clustering algorithm in 3D Euclidian space. With the proposed method for the 4-layer stacked 3DIC, less than 4.4% error in maximum sensor reading of the temperature of the chip for all applications is accomplished. The algorithm uses the proposed 3D thermal map modeling, which improves evaluation time by 53x for six different applications to be run on the 3DIC, compared with the situation in which detailed 3D map modeling using HotSpot 5.0 is embodied in the algorithm. This speedup will become even more significant as the number of evaluation scenarios increases for more complicated 3DICs and applications. Furthermore, as demonstrated, thermal sensor distribution for 3DICs must be solved as a 3D problem, which results in 44% fewer sensors, as compared with conventional 2D methods, while maintaining the same sensor reading error tolerance.

TABLE 5. MAXIMUM 3D THERMAL MODELING AND SENSOR READING ERROR FOR SIX APPLICATIONS.

Applications	Benchmarks running on core 1 through core 8	Max Modeling Error (%)	Max Sensor Reading Error (%)
1	apsi/equake/gcc/bzip/bzip/gcc/equake/apsi	2.72	2.95
2	apsi/equake/gcc/bzip/apsi/equake/gcc/bzip	2.35	3.27
3	apsi on all cores	2.13	3.28
4	equake on all cores	5.46	4.40
5	gcc on all cores	3.53	4.08
6	bzip on all cores	2.96	3.63

REFERENCES

[1] C. H. Yu, "The 3rd dimension-more life for Moore's Law," *in Proc. IMPACT*, pp. 1–6, Oct. 2006.

[2] Y. Xie, G. Loh, B. Black, and K. Bernstein, "Design space exploration for 3-D architecture," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 2, no. 2, pp. 65–103, Apr. 2006.

[3] G. L. Loi, *et al.* "A thermally-aware performance analysis of vertically integrated (3-D) processor-memory hierarchy," *in Proc. DAC.*, pp. 991–996, June 2006.

[4] K. J. Lee and K. Skadron, "Analytical model for sensor placement on microprocessors," *in Proc. ICCD*, pp. 24-27, Oct. 2005.

[5] S. O. Memik *et al.*, "Optimizing thermal sensor allocation for microprocessors," *IEEE Trans. CAD for ICs*, 27(3), pp. 516–527, Feb. 2008.

[6] A. N. Nowroz, R. Cochran, and S. Reda, "Thermal Monitoring of Real Processors: Techniques for Sensor Allocation and Full Characterization," *in Proc. DAC*, pp. 56-61, June 2010.

[7] P. Li *et al.*, "IC thermal simulation and modeling via efficient multi grid based approaches", *IEEE Trans. on CAD for ICs*, 25(9), pp.1763-1776, Sep. 2006.

[8] C. Xu *et al.*. "Fast 3D Thermal Analysis of Complex Interconnect Structures Using Electrical Modeling and Simulation Methodologies", *in Proc. IEEE/ACM ICCAD*, pp. 658 – 665, Nov. 2009.

[9] W. Huang et al, "Hotspot: A compact thermal modeling methodology for early-stage VLSI design." *IEEE Trans. VLSI Systems*, 14(5), pp. 501–513, July 2006.

[10] A. Sridhar, *et al.*, "3D-ICE: Fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling", *in Proc. IEEE/ACM ICCAD,* pp. 463 - 470 , Nov. 2010.

[11] H. Mizunuma, Y. Lu, and C. Yang, "Thermal modeling and analysis for 3-D ICs with integrated microchannel cooling," *IEEE Trans. CAD for ICs*., 30(9), pp. 1293 – 1306, Sep. 2011.

[12] A. Vincenzi, A. Sridhar, M. Ruggiero, and D. Atienza, "Fast thermal simulation of 2D/3D integrated circuit exploiting neural networks and GPUs," *in Proc. ISLPED*, pp. 151 – 156, Aug. 2011.

[13] S. Sapatnekar, "Addressing thermal and power delivery bottlenecks in 3D circuits," *in Proc. ASP-DAC*, pp. 423-428, Feb. 2009.

[14] A. Coskun *et al.*, "Energy-efficient variable-flow liquid cooling in 3D stacked architectures," *in Proc. DATE*, pp.111–116, 2010.

[15] R. E. Kessler, "The Alpha 21264 Microprocessor," *IEEE Micro*, 19(2):24–36, March-April 1999.

[16] SPEC-CPU2000, Standard performance evaluation council, performance evaluation in the new millennium, version 1.1, *Electronic Resource*, 2000.

[17] D. Brooks, V. Tiwari, and M. Martonosi, "Wattch: A framework for architectural-level power analysis and optimizations," *in Proc. ISCA*, pp. 83–94, May 2000.

[18] C. Zhu, *et al.*, "Three-dimensional chip-multiprocessor run-time thermal management," *IEEE Trans. CAD for ICs,* 27(8), pp. 1479–1492, Aug. 2008.

25-27 September 2012, Budapest, Hungary

Drawing the packaging landscape for CFL applications

F.J. Sluijs, J.H.J. Janssen, E.P.B.J. de Bruin, P. Dijkstra
Gerstweg 2, 6534AE Nijmegen, The Netherlands
+31-24-3536510, j.h.j.janssen@nxp.com

***Abstract*-In CFL lighting applications one of the main topics to address is the thermal situation of the driver electronics. With a relative high local ambient temperature the package for the CFL driver IC needs to be optimal designed for cost, size and thermal behavior. For the CFL application a compact thermal model was made and thermal measurements were done, showing a clear relation between used package technology and maximum CFL output power. The results can be used for optimal co-design of new CFL IC's and their packages.**

Fig 1. CFL application.

I. INTRODUCTION

In line with the worldwide focus on energy saving traditional light bulbs will be banned in the years to come and will be replaced by more intelligent CFL solutions. The widely used standard E14/E27 connectors were designed already in the beginning of the last century with little or no attention for heat transfer. Therefore the current replacement for traditional light bulbs must make use of these connectors. Building reliable electronics in these lamp mounts is by no means a simple task since the lamp itself significantly contributes to a local high ambient temperature. Due to the use of electronic components there are boundary conditions on maximum allowable temperatures for the individual components. In order to understand the thermal management of these applications a study was initiated to investigate these heat transfer mechanisms.

II CFL application and package

In the CFL IC application the output power is usually in the range from 5 up to 25W. We offer several CFL driver IC's in the market with integrated power switches in HV-SOI technology. Fig 1 shows an opened example of a CFL application.

In the CFL application three main factors are limiting the maximum output power of the CFL: 1. The lamp burner heating up the local ambient near the IC, and thus limiting heat transfer between IC and local ambient. 2. The dissipation inside the CFL driver IC, mainly caused by Ron losses in the integrated power switches. 3. The thermal performance of the package of the CFL driver IC.

This study focuses on the thermal behavior of the package within the CFL application. Measurements were done for several existing packages and the new µSIL package. This package was specifically developed for the low to medium power range, since current available package solutions are considered too large and too expensive. For this purpose a miniaturized version of a through-hole SIL package was developed. This novel package is compared to standard SO and currently used DIP outlines for reference. On the other side of the power spectrum a traditional DBS9 power package was evaluated.

The dissipation within the CFL strongly depends on the lamp burner used in the application. The impedance of a CFL burner can vary a lot due to different dimensions used. To limit the variable influence of the CFL itself, all measurements were done with the same type of lamp burner. This implies that with a typical fixed CFL output power the currents through the lamp and driver are also fixed. Then the dissipation within the IC is determined by the current through the integrated switches and the temperature of the IC. This is variable, since the Ron of the HV-SOI switches are temperature dependent. So by measuring the die temperature also the Ron losses and dissipation within the IC can be determined.

III Thermal measurements CFL application

In the CFL application thermal measurements were done using thermocouples and IR imaging techniques. Fig 2 shows an IR image of the CFL PCB, showing the CFL driver IC as local hot spot (indicated with the arrow).

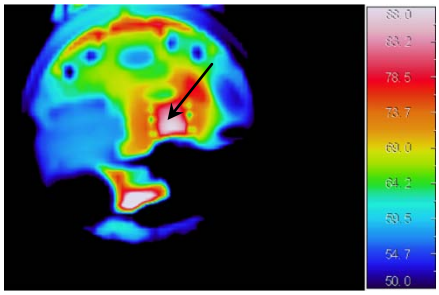


Fig 2. IR image of CFL PCB.

IR measurements are done to identify which components are of interest and heat up during operation and which remain relatively cold. The nature of the IR measurement requires a direct view on the components. To do so the application needs to be open and therefore cannot be used as real application temperatures. For this purpose thermo couples were used.

To get a good understanding of the impact of package technology on the thermal behavior in the CFL application several different packages have been measured. Thermal measurements were done on the µSIL, SO, DIP and DBS packages.

The measurement set up used in the investigation is based on the standard set up as is being used by a major light source manufacturer as shown in Fig 3 [1]. The lamp mount is placed in a tube with a diameter of 100mm and with a length of ~300mm. The tube is placed on a metal grid and elevated from the table leaving the bottom side open to free air.

Fig 3. Measurement setup for CFL.

Thermocouples are glued on the components or positioned in areas of interest as shown in Fig4. The thermo couples are connected to a datalogger on individual channels and the channels are scanned at the desired time interval.

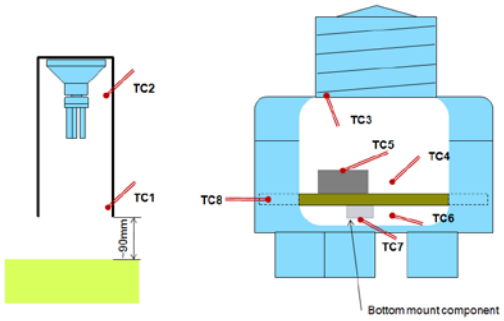


Fig 4. Positions of thermo couples.

Thermal measurements were taken while the system is warming up during operation from room temperature until the temperature in the application stabilizes. Each package configuration was tested under different power conditions. The CFL output power was regulated by means of variable supply voltage.

A typical package used in CFL applications is a DIP package which is reference in this study. Fig 5 shows the warming up curve compared to a low cost SO package solution. In this evaluation the SO package is mounted in two different ways, on the burner side (green line) and opposite the burner (blue line). Both SO configurations are becoming significantly warmer mainly due to the fact that the SO package is much smaller compared to the DIP package. The temperature of the SO package mounted on the burner side becomes even higher due to the warmer environment in which the package is positioned. The difference in air temperature on the burner side and opposite of that is about 5°C.

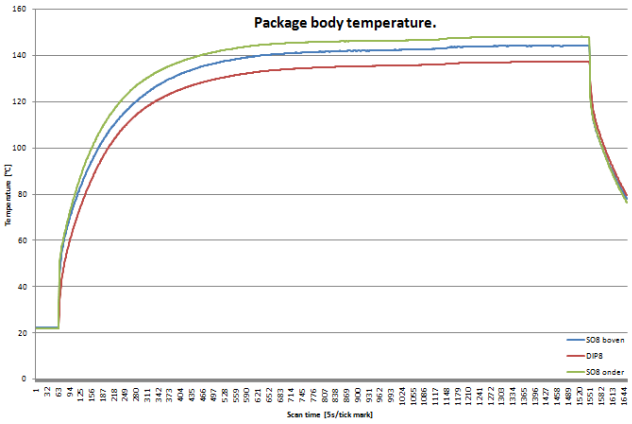


Fig 5. Heating up curves of DIP and SO package.

Several driver IC solution have been evaluated in the reference package. These different IC's clearly impact the package temperature as shown in Fig7.

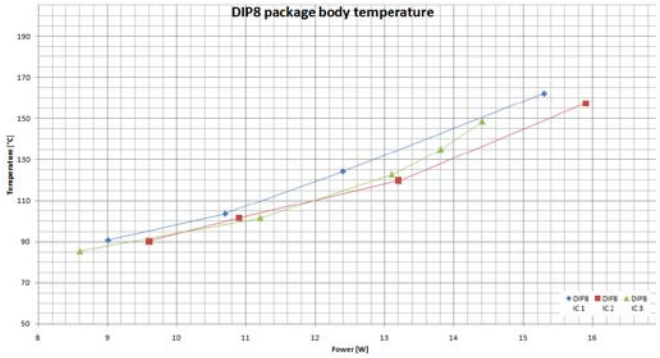


Fig6. Impact of IC type on package temperature.

The DBS package family has some thermal enhancements incorporated in the package which allow for a better thermal performance, an exposed die pad and one fused lead. Measurements are as shown in Fig 8. A deliberately made modification of the application resulted again in a further decrease of the package temperature, showing an interaction between system design and thermal performance.

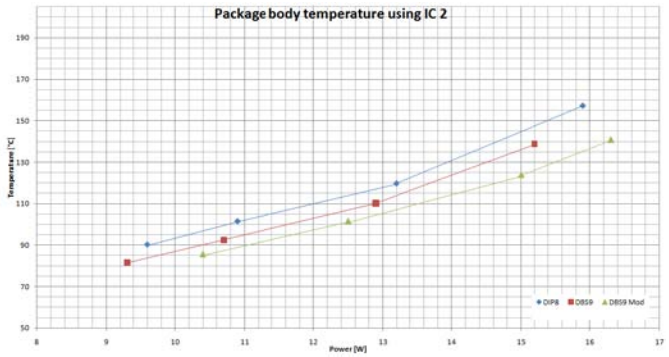


Fig7. Impact of application modifications on temperature.

Further modification of the application by using a different burner type gives the possibility to achieve much higher output power levels then before as can be seen in Fig9 or when power levels are limited the package temperature remains significantly lower.

Fig8. Impact of burner type on temperature.

One of the main drivers in CFL applications is cost. For this purpose a novel package (Fig 9) has been developed with the purpose to include seemingly contradicting requirements such as low cost and high power capability. This has resulted in a new package family which, with modifications, can be used in the low to medium power range.

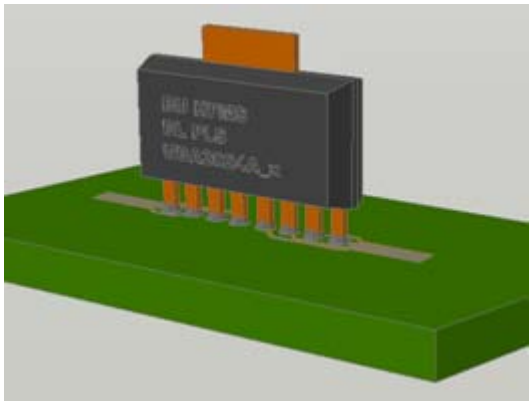


Fig 9. New package development.

The initial temperature measurements are compared with DIP and SO as shown in Fig10.

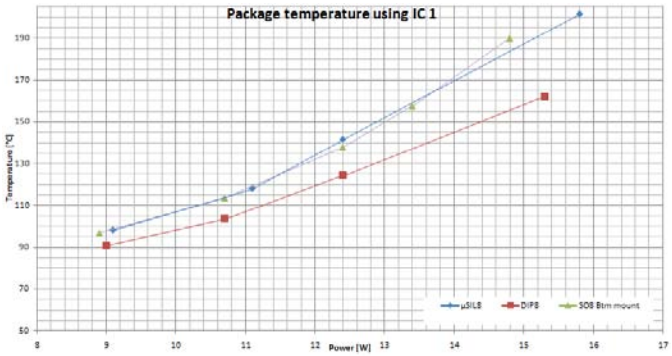


Fig 10. New package development versus DIP and SO.

The measurements show that there is a strong interaction between package technology, IC type and application on the maximum output power and thermal performance of the CFL application.

IV Thermal modeling

One of the objectives of this study was to develop a simulation platform that enables package and IC developers to evaluate new package and/or IC designs in the virtual world. To develop such a platform one needs to make sure that the model predicts the correct temperatures. The above described µSIL package was used to validate this modeling approach. This approach is built on the well known thermal compact model technology [2]. A Finite Element Model (FE-model) for the µSIL package was made. A picture of this model can be seen in fig 11.

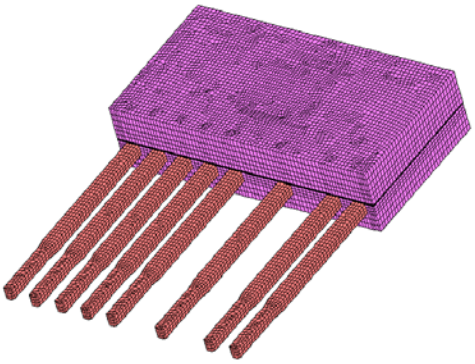


Fig 11. 3D FE-model of the µSIL package.

This model was exposed to the standard set of 38 boundary conditions which is described in [2]. Next, the results of these calculations were fed to the program DOTCOMP. A tool, especially designed to calculate the optimum resistor network [3]. The Compact Thermal Model (CTM) derived is shown in the figure below.

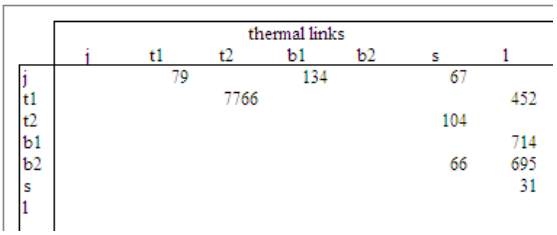


Fig 12. 3D FE-model of the µSIL package.

The maximum differences of the predicted junction temperature and heat fluxes of the CTM are well below 10% of the FE-model, which makes that he derived CTM can be called boundary condition independent [2].

The next step to do was to setup this CTM model as a network assembly in Flotherm which is a powerful 3D Computational Fluid Dynamics (CFD) software tool that predicts airflow and heat transfer in and around electronic equipment, from components and boards up to complete systems. The network assembly and the resistor network as used in Flotherm are given in Fig 13 and 14.

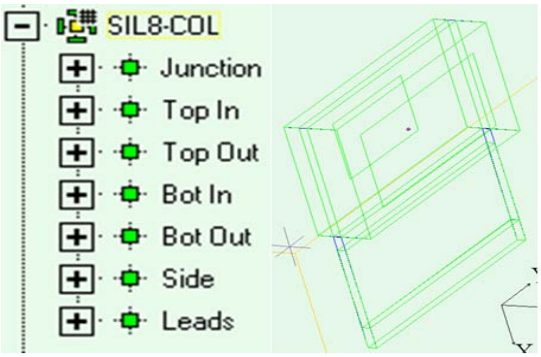


Fig 13. Network assembly in Flotherm.

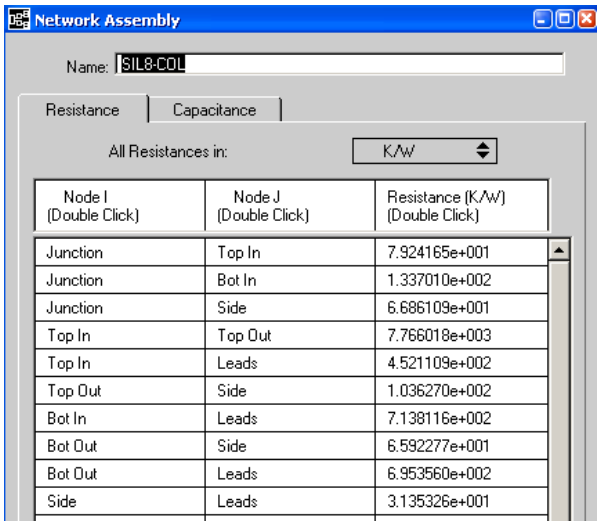


Fig 14. Resistor network in Flotherm.

All the electronic components in the lamp foot and the burners are modeled as cuboids. And for each of these parts estimations were made how much power was dissipated while the burner was in operation. These estimations are given in the table below.

Table 1
Dissipated power for each component

	mW
Fusistor	150
Coil	100
Capacitor	50
Air Coil	500
6x Single tube	11160
IC	450
Total	12410

The total of 12.4W is the measured power drawn from the power grid. The amount of power for each of the CFL tubes was chosen to match the total power.

The Flotherm simulations for the µSIL package are compared with the measurements see table 2. Locations of the thermo-couples can be seen in Fig 4.

Table 2
Comparison between measurement and model

	Measured	Model
T_Tube bottom [°C]	24	26
T_Tube top [°C]	60	52
T_µSIL [°C]	145.5	150
T_lamp mount [°C]	77.9	84
T_board [°C]	98	99
T_air board [°C]	109.1	102

As can be seen there is a reasonable agreement. One other experiment to validate the model has been conducted. The electronics of one CFL light bulb was used to fire the burners in a second light bulb. See figure 15 below. Purpose of this experiment was to determine the individual contributions of the electronic components on one hand and the burner on the other.

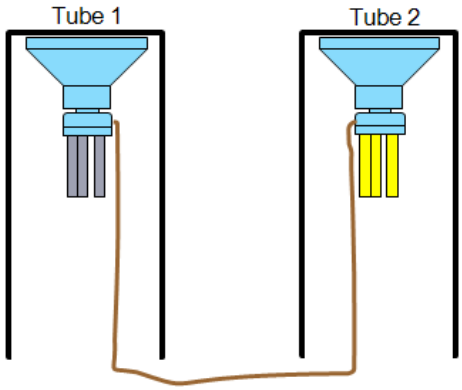


Fig 15. Electronics and burner separated.

Results for electronics only and burners only are given in tables 3 and 4.

Table 3
Electronic only

	Measured	Model
T_Tube bottom [°C]	22	22
T_Tube top [°C]	27	25
T_µSIL [°C]	75.9	74
T_lamp mount [°C]	36.4	47
T_board [°C]	42.8	52
T_air board [°C]	52	54

Table 4
Burner only

	Measured	Model
T_Tube bottom [°C]	24	27
T_Tube top [°C]	60	53
T_µSIL [°C]	124.3	125
T_lamp mount [°C]	78.3	84
T_board [°C]	100.9	102
T_air board [°C]	106.5	97

Again fairly good agreement between model and measurements was achieved.

The same principle has been applied to the DIP and DBS packages showing equal good relation between model and the actual measurements.

V Conclusions

The obtained results clearly indicate that a relation between packaging technology and CFL output power range exists. Furthermore it gives a good direction in which areas thermal bottlenecks can be minimized and thermal enhancements can be incorporated in a new package and application design. Comparing the experiments and modeling results it can be concluded that the developed simulation platform is suitable for future developments of new IC/package design with respect to the thermal performance.
This is a typical illustration that shows that co-design of package and application becomes more important for these kinds of applications.

References

1. H.J.Eggink, “Temperature measurement, an introduction” Philips Applied technologies CP03187, May 2010
2. Lasance, C., Vinke, H., Rosten, H., and Weiner, K.L., 1995, “A Novel Approach for the Thermal Characterization of Electronic Parts”, Proc. of SEMITHERM XI, San Jose, CA, pp. 1-9.
3. Lasance, C.J.M., Hertog, D.D., and Stehouwer, P., “Creation and Evaluation of Compact Models for Thermal Characterization using Dedicated Optimization Software,” Proc. SEMITHERM XV, San Diego, CA, pp. 189–200, 1999.

25-27 September 2012, Budapest, Hungary

Thermoreflectance Measurements for Optically Emitting Devices

Peter E. RAAD[1, 2, 3, *], Pavel L. KOMAROV[1], and Mauro A. BETTIATI[2]
[1] Mechanical Engineering Department, SMU, Dallas, TX 75275, USA
[2] 3S PHOTONICS, F-91625 Nozay Cedex - France
[3] IEEE Senior Member
[*]*Corresponding Author: +1 (214) 768-3043, praad@smu.edu*

Abstract — This work examines the difficulties associated with using optical techniques to measure temperature when the device itself emits a significant level of light over a wide spectrum, making it a challenge to separate the useful measurement signal from the device emission. The specific situation considered here is that of using a thermoreflectance (TR) thermography approach to characterize the thermal behavior of semiconductor laser devices. A lowpass filter was placed in the optical path to minimize the primary laser irradiation on the TR imaging and then the TR response of the region of interest was determined over a wide range of visible light wavelengths to locate the maximum response. TR measurements performed at the optimal light wavelength successfully provided a submicron-resolution map of the active area of sample lasers.

I. INTRODUCTION

Optical methods have been widely used to capture the thermal behavior in areas of interest on microelectronic devices. Optical methods have the major benefit of being non-contact and non-invasive and work by recording the optical energy associated with the material of interest. Methods based on IR physics capture the radiation emanating from a hot surface, while thermoreflectance-based methods measure the relative change of reflected light resulting from a change in the temperature of a surface. These methods have been successfully used to characterize the thermal behavior of devices down to the submicron levels.

This article examines the difficulties associated with using optical measurement techniques when the device itself emits a significant level of light over a wide spectrum, making it a challenge to separate the useful measurement signal from the device emission. The specific situation considered here is that of using a TR thermography approach to characterize the thermal behavior of semiconductor laser devices. The devices used in this work are 980 nm chips with a cavity length of 3.9 mm, whose properties and technology details have been discussed in previous papers [1, and refs. therein] and are wavelength stabilized for improved performance in Erbium Doped Optical Amplifiers (EDFAs). Wavelength stabilization is obtained by locking the emission wavelength to a Fiber Bragg Grating (FBG) inscribed in the fiber pigtail.

II. EXPERIMENTAL APPROACH

The experimental temperature mapping system used in this work is the TMX Scientific T°Imager™ [2], which is based on the thermoreflectance (TR) method [3-6], where the change in the surface temperature is measured by detecting the change in the reflectivity of the sample. A photo and functional schematic of the TTRG system are shown in Fig. 1. The measurement methodology involves a calibration phase and a device activation phase, which have been detailed in previous publications. In the calibration phase, the thermoreflectance coefficient, $C_{TR} = (\Delta R/R)\cdot\Delta T^{-1}$, is determined for each of the surface materials in the region of interest at a given wavelength of light. The activation phase yields the relative change in surface reflectivity in that region of interest, which is then scaled with the calibration data to obtain the temperature map over the DUT.

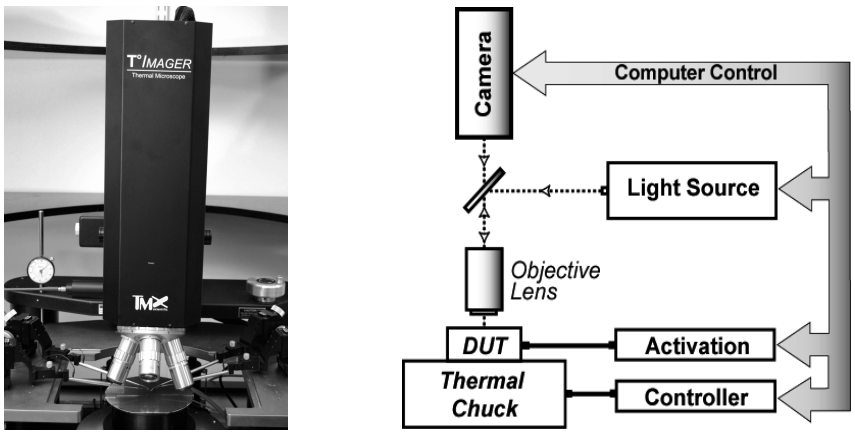


Fig. 1. Photo and Schematic of TRTG system.

Optimizing the response of the TRTG system is achieved by scanning the wavelength range of visible light in order to identify those wavelengths where the TR response is highest. Knowing that the laser devices considered in this work emit light at approximately 980 nm, it was evident that we needed to introduce low pass filters to ensure that the laser energy does not overwhelm the TR signal. Filters, however, also attenuate the useful signal, so it is not practical to use too many of them in the optical path. Optimally, one should perform a spectrum analysis of the light emitted by the laser package under investigation, and then use a single bandpass filter around the wavelength where both the TR response is maximum and the laser emission energy is minimum. To determine the TR response of this laser device, 750 nm lowpass filters were introduced in the optical path to minimize the primary laser irradiation on the TR imaging. Figure 2 depicts the TR response over a wide range of wavelengths in the visible spectrum.

By comparing the TR response to the emission power spectrum of the laser in the visible range (Fig. 3), it is clear that an optimal wavelength for the TR measurements would be 600 nm where the laser's secondary emission is relatively small and away from sharp local peaks which are also temperature dependent. While the TR response is higher at, for example, 665 nm for Laser #976, 600 nm provides the advantage of being the same for the two laser devices considered as well as being in region of higher measurement stability, i.e., where small changes in wavelength do not result in large changes in TR response.

IV. Results and Discussion

The DUT, shown in Fig. 4, was activated at five different current levels from 250 mA to 1000 mA and at two different base temperatures of 25˚C and 45˚C. As an example, Fig. 5 shows a temperature rise map superimposed on the DUT in the thermally active area coinciding with where the laser is emitting at its primary wavelength.

A semi-cylindrical pattern in the temperature field can be observed around the laser emission port. Also, there appears to be an asymmetry toward the left pad. The submicron resolution of the T°Imager system not only makes it possible to detect this asymmetry but also provides a resolved distribution over the middle pad region where the thermal behavior is of interest.

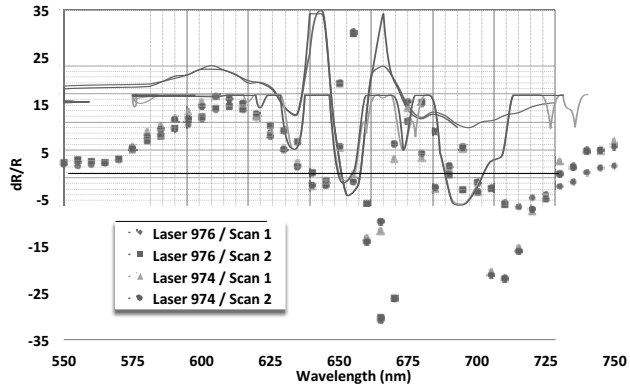


Fig. 2. Thermoreflectance response of region of interest for range of visible light wavelengths

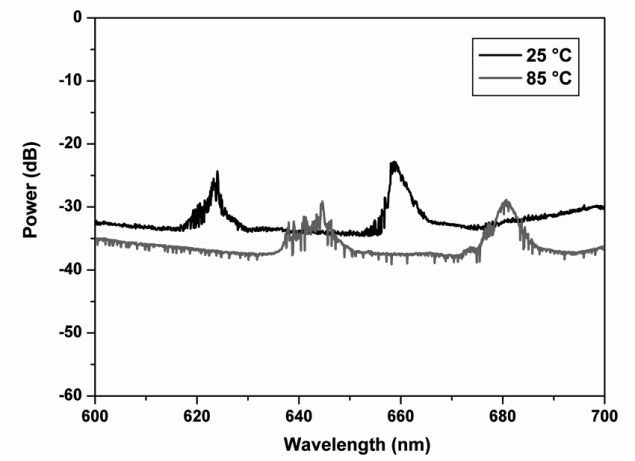


Fig. 3. Power spectrum over a range of visible light wavelengths below the primary laser emission wavelength

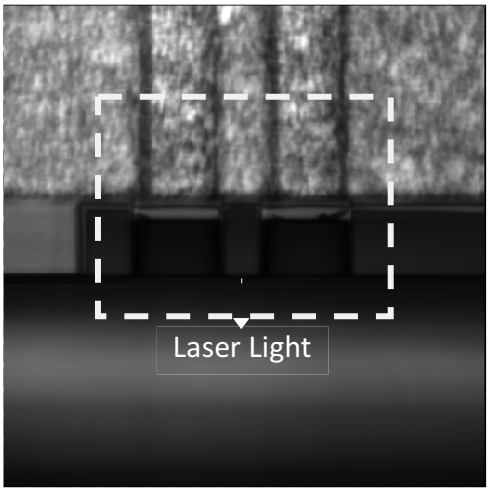


Fig. 4. Image of DUT with active area in the center.

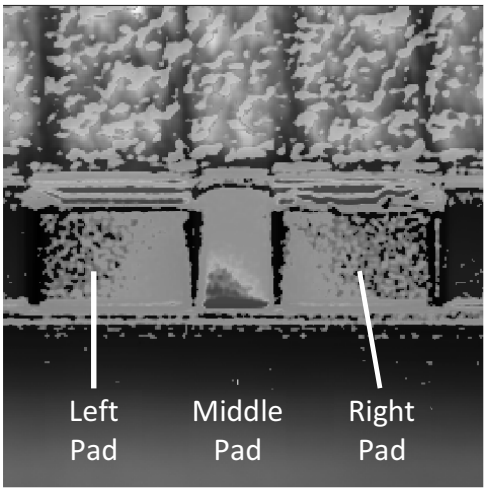


Fig. 5. Thermal map around area from which the laser emits. Field of view corresponds to the dashed area in Fig. 4.

Figure 6 below is a plot of the average temperature values over the three pads for the five current levels at the two different base temperatures. As expected, the middle pad is hotter than the outer ones, and all temperature curves exhibit linear behavior with the applied current. Also, the average temperature values on the right pad are lower then those on the left pad, which is consistent with the asymmetry of the temperature field in the middle pad. Finally, increasing the base temperature results in a consistent increase in the average temperature rise for all levels of applied current.

This work has demonstrated the capability of using an optically-based measurement system to map the temperature field at the deep submicron level of devices that emit strong light, such as lasers. This was accomplished by selecting an optimal TR measurement wavelength from a comparison between the power spectrum of the laser in the visible light range and the TR response over that same range.

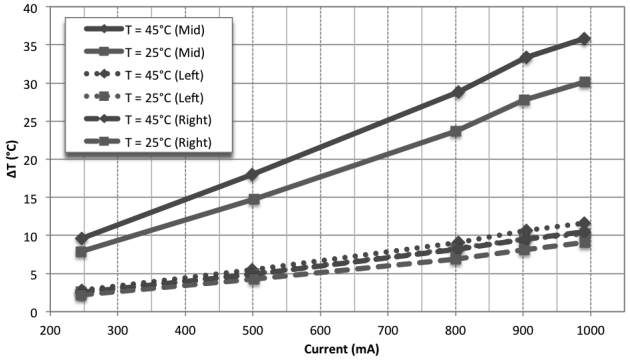


Fig. 6. Plot of average temperatures in thermally active area of interest for Laser #974 at two different base temperatures.

REFERENCES

[1] M.A. Bettiati, C. Starck, F. Laruelle, et al., "Very high power operation of 980-nm single-mode InGaAs/AlGaAs pump lasers," *Proceedings of SPIE*, vol. 6104, 61040F, 2006.

[2] T°Imager™ is a registered trademark and a thermal microscopy system of TMX Scientific, Inc., http://tmxscientific.com.

[3] M. Farzaneh, K. Maize, D. Lüerßen, J. A. Summers, P. M. Mayer, P. E. Raad, K. P. Pipe, A. Shakouri, J. A. Hudgings, and R. Ram, "CCD-Based Thermoreflectance Microscopy: Principles and Applications," *J. Physics D: Applied Physics*, vol. 42, pp. 143001-143021, 2009.

[4] P.L. Komarov, M.G. Burzo, and P.E. Raad, "A Thermoreflectance Thermography System for Measuring the Transient Surface Temperature Field of Activated Electronic Devices," *22nd Annual IEEE Semi-Therm Symposium and Exposition*, Dallas, TX, March 2006.

[5] P.L. Komarov, M.G. Burzo, and P.E. Raad, "CCD Thermoreflectance Thermography System: Methodology and Experimental Validation," *12th International IEEE THERMINIC Workshop*, Nice, France, Sep. 27-29, 2006.

[6] Tessier, G., M.-L. Polignano, S. Pavageau, C. Filloy, D. Fournier, F. Cerutti and I. Mica, "Thermoreflectance Temperature Imaging of Integrated Circuits: Calibration Technique and Quantitative Comparison With Integrated Sensors and Simulations," *J. Physics D: Applied Physics I*, Vol. 39, pp. 4159-4166, 2006.

25-27 September 2012, Budapest, Hungary

Temperature Prediction for Multi-Core Microprocessors with Application to Dynamic Thermal Management

Bartosz Wojciechowski, Janusz Biernat
Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
bartosz.wojciechowski@pwr.wroc.pl

***Abstract*—Temperature is among the most important factors limiting CPU performance. To ensure stable and reliable operation temperature has to be limited. In this paper we present a method for predicting temperature of a multi-core microprocessor based on it's high-level thermal model. Temperature prediction is then utilised in a Dynamic Thermal Management algorithm that can exploit performance optimisation opportunities while ensuring operation below temperature limit. The DTM algorithm uses frequency and voltage scaling of the CPU cores and task migration and was evaluated on a physical computer with a quad-core CPU.**

***Index Terms*—CMP thermal model, power and temperature prediction, Dynamic Thermal Management, DVFS, Linux kernel**

I. Introduction

High power densities in contemporary microprocessors cause hotspots that negatively impact reliability. To ensure stable operation and long lifetime of the devices, power dissipation in hotspots has to be limited. Therefore, temperature is a strong performance-limiting factor. This situation is aggravated by high cost of cooling (e.g. heat sink with forced air convection or water cooling) which can amount to substantial fraction of the total cost of computing device. Moreover, miniaturisation, forced by portability requirements in consumer electronics market and high-density scaling requirements in the High Performance Computing limits the allowable power densities. Also, operation in high temperature leads to extra power consumption, due to exponential dependence of leakage current on temperature.

Traditionally a worst-case analysis is performed during the design to ensure that temperature limits will not be exceeded. However, in most cases both the operating conditions and workloads cause far less thermal stress on the processor than it is assumed in the worst-case scenario. This observation is the foundation of Dynamic Thermal Management (DTM) concept [1] where microprocessor performance and power dissipation is bound more strictly when temperature reaches some predefined limit. Presently most microprocessors employ some form of hardware-controlled DTM. Mechanisms that dynamically limit temperature, such as clock throttling, Dynamic Voltage and Frequency Scaling (DVFS) or thread migration, also reduce performance of the processor. With proactive thermal management it is possible to reduce the negative impact of DTM mechanisms on performance by operating closer to the temperature limit. However, accurate power-performance and thermal models of microprocessor are needed to predict its temperature and act accordingly to the current load.

Maximizing the total processor throughput can stand in contradiction to respecting task priorities. The Operating System holds unique information on task priorities, deadlines and resource utilisation. To be able to exploit optimisation opportunities microprocessor temperature should be managed on the OS-level. To be effective, DTM algorithms must work with low latency. Therefore, any temperature prediction algorithm must work with small computation overhead.

Most of the work on Dynamic Thermal Management for multi-core microprocessors published up to date uses an elaborate simulation setup or analytical modelling to verify the effectiveness of the proposed algorithm or policy. Rao and Vrudhula [2] addressed the problem of efficient online computation of the speeds of different cores of a multi-core processor to maximize throughput subject to an upper bound on the core temperatures. They found an analytical solution to determine the local optimum speeds of cores. Authors of [3] developed a frequency selection algorithm for multi-cores that maximizes operating frequency of the *critical core* (running high-priority task). Optimal processor performance subject to thermal constraints with thread priorities, where each core runs a thread of varied importance was also the subject of [4]. Kadin and Reda [5] developed a frequency planning methodology that maximizes the total performance of multi-core processors. They also analysed the implications of technology scaling on the performance limits of multi-core processors. In [6] central and distributed DVFS management approaches were compared.

With the advent of affordable multi-core processors more works emerged that were focused on aspects of practical implementation of DTM policies on real processors. Authors of [7] show, with an Intel Core 2 Duo processor, that the current Linux scheduler can easily be enhanced with thermal-

ISBN:978-2-35500-022-5

awareness to improve its performance in terms of both temperature conditions and total application throughput. Liu and Quan implemented a number of thermal aware scheduling techniques for a desktop computer with a quad-core processor and studied their performance [8]. Their experimental results demonstrate that the dynamic thermal management implemented in operating system is an effective method to control processor temperature. Similar conclusions were presented in [9] and [10].

Most authors of the papers on DTM of microprocessors employ some form of temperature or hotspot prediction to improve efficiency and throughput of a processor subject to temperature limits. A popular Method of temperature prediction is based on ARMA (Autoregressive Moving Average) model of chip temperature. In [11] ARMA-based approach is validated and compared to other methods based on temperature history. A method of temperature prediction similar to that presented here in terms of thermal model has been proposed in [12]. It's authors use least squares regression to derive state-space thermal model of the microprocessor. Yang et al. predict future power dissipation of CPU cores based on the former measurements which may lead to high prediction errors in some cases [13]. Temperature prediction based on thermal history of each core can be inaccurate under changeable operating conditions, e.g. starting the new tasks on neighboring cores.

Main contributions of this work are: high-level temperature and power-performance models based on measurements of a multi-core microprocessor; a temperature prediction algorithm of low computation overhead and a heuristic DTM algorithm using thread migration and DVFS. Our high-level power-performance model of microprocessor core uses data from performance monitoring counters and current frequency setting to calculate performance and power dissipation. We use an RC thermal model of the CPU to provide temperature predictions based on microprocessor activity history.

II. Power, Temperature and Performance Models

To build high level models of the CPU we collected power, performance and temperature traces from a computer with quad-core CPU Intel Core 2 Q9400. We used a custom program written in C that provided high temporal resolution of 1 ms with minimal overhead on normal operation of the computer. We obtained power-performance-temperature profiles during runs of PARSEC [14] parallel benchmarks. PARSEC is a broad collection of programs with very different behavior in terms of workload size, inter-thread communication and processor resources utilisation. Therefore, the data obtained should be representative of many computing scenarios.

To obtain power consumption trace of the CPU we used a digital power meter. We logged power consumption of the whole computer at the power outlet, and then subtracted the idle power consumption P_{idle} from the trace. Since we minimised the number of peripherals in the computer, the dynamic changes in power consumption could be attributed mainly to CPU and memory activity. Overall power consumption of the

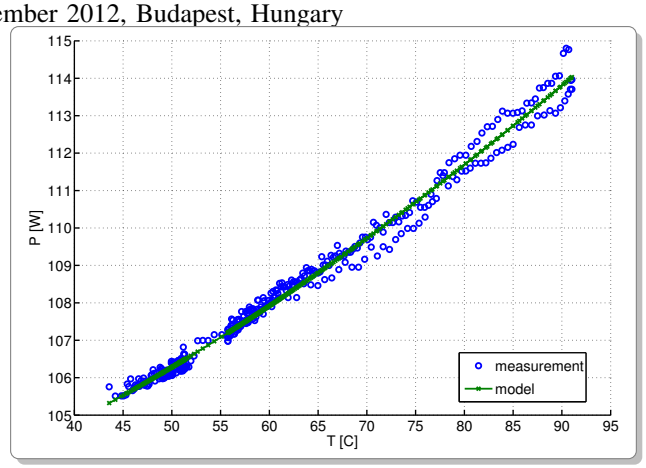


Figure 1. Total power dissipation of a computer as a function of CPU temperature

computer P_c can be described as:

$$P_c = P_{idle} + P_{dyn} + P_{stat} + P_{mem} + \varepsilon, \quad (1)$$

where P_{dyn} is the dynamic power associated with computation in the CPU, P_{stat} is the static, temperature-dependent power dissipated by the processor, P_{mem} is the power dissipated by main memories.

Temperature of each core was read using Digital Thermal Sensors (DTS) embedded in each of the CPU cores [15]. Each sensor has resolution of 1 K and is available through one of Model Specific Registers (MSR) of the core. Changes of the processor temperature can be attributed to intestine power dissipation or changes in ambient temperature and hence it must be controlled. We measured the temperature inside the computer case T_{amb}using the DS18B20 thermometer.

To measure the static power associated with leakage in the CPU we measured power dissipated by the computer while changing the temperature of the CPU from 45 to 90°C under constant workload. The CPU was heated by using a stream of hot air directed at the heat sink. The measurement data along with results of regression against an exponential function $P_c = \alpha + \beta \cdot (T - 45)^{\gamma}$, with α equal to power dissipated at 45°C, are presented in Fig. 1. However, $\gamma \approx 1$ so the static power model can be linearised with minimal error.

We logged two performance metrics, namely Instructions Per Cycle (IPC) and Last-level Cache Misses (LLCM). These are available from MSRs of each CPU core. IPC is a good approximation of the activity of a processor core and LLCM can be related to the amount of off-chip communication (i.e. reads and writes in main memory).

Based on measurements of computer power and numbers of instructions executed we formulated models relating core power and number of instructions committed in a fixed time t_s. Denoting the number of cores as N_c and using k_A to k_E as fitting parameters, we used power, performance and temperature measurements to formulate a high-level power

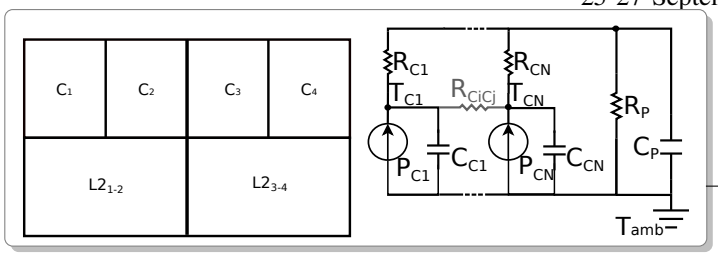


Figure 2. Microprocessor floorplan and RC thermal model of the CPU and package

Table I
ERROR OF TEMPERATURE PREDICTION BASED ON LAST VALUE: MEAN SQUARE ERROR (MSE) AND MAXIMAL ABSOLUTE PREDICTION ERROR ($|\varepsilon|_{\max}$) IN K FOR VARYING PREDICTION PERIODS

	prediction time [ms]					
	1	5	10	20	40	100
SPEC CPU 2000 int						
MSE	0.0018	0.0322	0.0905	0.1658	0.2371	0.3261
$\|\varepsilon\|_{\max}$	0.3	1.4	2.4	3.0	4.0	4.0
SPEC CPU 2000 fp						
MSE	0.0013	0.0234	0.0560	0.1252	0.2078	0.3787
$\|\varepsilon\|_{\max}$	0.4	1.4	2.6	3.8	4.9	5.2
PARSEC -n 4						
MSE	0.0013	0.0238	0.0678	0.1324	0.2086	0.2698
$\|\varepsilon\|_{\max}$	0.7	3.3	4.7	6.0	6.0	8.0

model of a computer:

$$P_k = P_{idle} + k_A + k_B \cdot \sum_{c=1}^{N_c} IPC_c^{k_C} + \\ + k_D \cdot \sum_{c=1}^{N_c} LLCM_c + k_E \cdot \sum_{c=1}^{N_c} (T_c - T_{\min}). \quad (2)$$

From (2) we derived a dynamic power model of one CPU core:

$$P_{core} = P_{core}^{idle} + \frac{1}{N_c} \cdot k_A + \frac{1}{N_c} \cdot k_B \cdot N_c^{k_C} \cdot IPC^{k_C} \quad (3)$$

To get a power-performance model of a processor using DVFS, we run all the benchmarks on all three available frequency settings. The average number of instructions committed in one clock cycle (IPC) changes with the frequency ratio between CPU and the main memory. Based on performed measurements, we formulated the following model to describe IPC as a function of frequency, cache misses and IPC at $f_{\max}$.

$$IPC(f) = IPC(f_{\max}) + k_F \cdot LLCM \cdot \left(\frac{f_{\max}}{f} - 1 \right). \quad (4)$$

We also used performed regression of the following formula on the power-performance traces to find how dynamic power changes with frequency.

$$P_{dyn} = P_{dyn}(f_{max}) \cdot \left(\frac{f}{f_{max}} \right)^{k_G}, \quad (5)$$

Since each frequency has a fixed voltage level, only frequency is used in (5). We obtained $k_G \approx 2.44$.

W use a well known thermal model based on analogy between thermal and electrical phenomena (see Fig 2). One node in a thermal RC circuit was assigned to each core of the CPU as well as portions of level 2 cache. Also Thermal Interface Material (TIM), processor package and heat sink are modeled. Each node is described by it's thermal capacitance and resistance to neighboring nodes. Thermal behavior of such thermal RC circuit can be described with a state equation:

$$\frac{d\mathbf{T}}{dt} = A\mathbf{T} + B\mathbf{P}(s, \mathbf{T}, t), \quad (6)$$

where $\mathbf{T}$ is temperature of the circuit nodes, $\mathbf{P}$ is a column vector of power values dissipated in time t, while $\mathbf{A}$ and $\mathbf{B}$ are matrices describing thermal properties of the system. To obtain $\mathbf{A}$ and $\mathbf{B}$ we used the HotSpot tool [16]. We used default HotSpot configuration except for the chip thickness and heat sink thermal capacitance and resistance which we updated based on the processor datasheet. Also since our processor consists of 2 separate parts connected in one package, each holding 2 cores and half of the total cache memory, we had to manually increase thermal resistance between proper nodes of the model and recalculate the state and input matrices $\mathbf{A}$ and $\mathbf{B}$. Finally, we verified that the model response to a fixed input power reflected closely that of the real processor.

III. PREDICTION OF CPU TEMPERATURE

As can be seen in Tab. I, even short time step $t_s = 5\,\text{ms}$ can lead to temperature change exceeding 3 K. Therefore, predicting temperature equal to previous measurement can lead to high errors and breaking the set temperature limit.

As in [13], [17], to predict the temperature after t_s, we solve (6) for t and substituting:

$$\mathbf{E} = e^{\mathbf{A}t_s} \quad \text{and} \quad \mathbf{F} = \mathbf{A}^{-1}(\mathbf{E} - \mathbf{I}), \quad (7)$$

we get the formula for system temperature:

$$\mathbf{T}(t + t_s) = \mathbf{E}\mathbf{T}(t) + \mathbf{F}\mathbf{P}(t + t_s). \quad (8)$$

Using (8) and the calculated power dissipation (4,5) one can predict temperature of the linear system in time $t+t_s$. Matrices $\mathbf{E}$ and $\mathbf{F}$ describe thermal properties of the system and need only be calculated once. Therefore, each prediction is associated with only two matrix multiplications and a vector addition which add very little overhead to normal OS operation.

To be able to predict temperature for long time steps ($\approx$100 ms), we need a reliable way to predict power dissipation. Since dynamic power is highly correlated to core activity measured in IPC [18], it is possible to use IPC prediction to assess future power dissipation. Many programs exhibit repetitive behavior with recurring phases of low and high activity that result from control flow in loops and in functions [19]. This repetitive behavior can be exploited by using a table-based predictor that is similar to global branch predictors in CPUs. It uses a recent activity (IPC) history to index table with prediction for the next step. With programs

Algorithm 1 Prediction of CPU temperature $\mathbf{T}$

Require: current system temperature $\mathbf{T^{curr}}$, previously predicted temperature $\mathbf{T^{pred}}$, current core frequencies, current IPC and LLCM values of all cores, current speeds of all the cores $\mathbf{S}^{curr} = s_1, .., s_{N_c}$, demanded speeds of all the cores,

Ensure: predicted temperature of CPU cores $\mathbf{T}(t+t_s) = T_1, .., T_{N_c}$

1: **for** each core i **do**
2: update core temperature T_i
3: calculate prediction error $\varepsilon_i = T_i^{pred} - T_i$
4: scale IPC_i on DVFS level change using (4)
5: calculate dynamic power $P_{dyn,i}$ using (3) and scale using (5)
6: calculate static power $P_{stat,i}$ based on linear approximation of the leakage power model
7: calculate dynamic and static power of last-level cache based on it's temperature and cumulated LLCM of all cores
8: update state vector $\mathbf{T}$ using current temperature of the cores
9: adjust temperature of the heat sink using $k_\varepsilon \cdot \sum_{i=1}^{N_c} \varepsilon_i$
10: update input vector $\mathbf{P}$ using previously calculated dynamic power P_{dyn} and static power P_{stat} dissipated in cores and caches
11: calculate $\mathbf{T^{pred}} = \mathbf{T}(t + t_s)$ using (8)

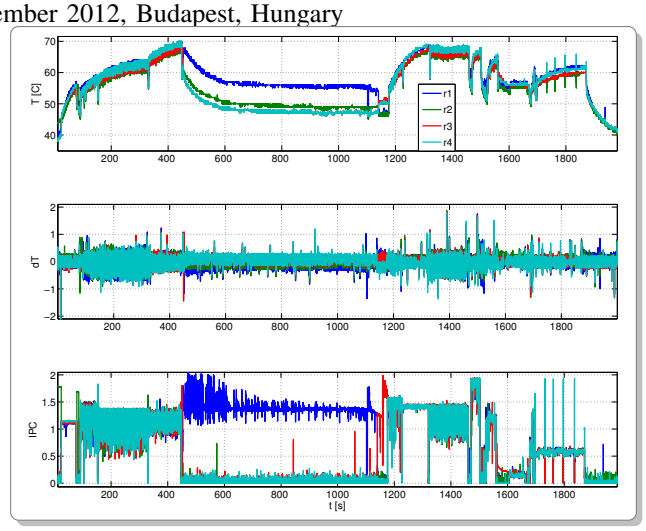


Figure 3. Temperature, temperature prediction error and IPC of 4 cores during execution of full PARSEC benchmark suite

that do not exhibit repetitive behavior our IPC predictor falls back to last-value predictor.

The temperature prediction is presented with Algorithm 1. It can be used to predict temperature of the CPU cores given current temperature (state) of the system, previous prediction, current or predicted performance metrics (IPC and LLCM) of each core, and the speed of each core – DVFS setting. If the DVFS settings are not constant the dynamic power for the next interval is recalculated.

First, Algorithm 1 updates temperature and performance data on all the cores by reading the required MSRs. The IPC for the next period can either be predicted using the table-based approach or assumed constant. In the next step the prediction error is updated and calculates dynamic and static power dissipated by each core using high-level power-performance models. Since temperature of the heat sink is not measured directly, we adjust the relevant part of the state vector $\mathbf{T}$ using previously calculated prediction error. This step also effectively mitigates the problem with changes in the ambient temperature, since $t_s << \tau_{ambient}$. In the last part, the input vector $\mathbf{P}$ is updated and new temperature prediction is calculated.

This temperature prediction algorithm was tested using off-line traces from a quad core computer and also used on-line in a DTM mechanism. We took performance and temperature traces from an Intel Q9400 processor during execution of benchmarks from SPEC CPU 2000 suite and multi-threaded tests from the PARSEC collection. The traces were collected every 1 ms and later sampled every 10, 20, 40 and 100 samples to emulate longer periods.

Fig. 3 holds temperature, temperature prediction error for 20 ms period and IPC plots for all 4 cores during execution of the full PARSEC benchmarks using 4 threads. As can be seen from the middle plot, the prediction error was smaller than 1 K for most of the execution time.

Cumulative results of the prediction algorithm evaluation are presented in Tab. II. It holds the number of prediction errors higher than 2 K and maximal absolute error for different prediction times. The results are for the last-value prediction and our prediction algorithm based on core activity. We evaluated 3 scenarios of IPC prediction: last-value IPC prediction, our table-based IPC prediction and perfect (oracle) IPC prediction. It can be seen that even using last-value prediction of core activity greatly enhances temperature prediction accuracy relative to last-value prediction.

IV. Dynamic Thermal Management

Our DTM algorithm uses temperature prediction to control operation of the CPU under thermal constraints. It uses Thread Migration (TM) to reduce thermal gradients and DVFS to ensure that temperature limits are not exceeded. Based on the current temperature and characteristics of executed tasks the algorithm calculates maximal frequency for each core that will not lead to overheating. We take into account task priority and performance characteristics to ensure sufficient performance of high-priority programs as well as good overall throughput.

When the number of active tasks is much smaller than the number of available cores TM is used to level out temperature between the cores and lower the possible hotspots. Tasks are migrated between cores with the highest differences in temperature, provided that this difference is higher than a predefined threshold T_{th}.

The predicted temperature can be used to find the frequencies of each CPU core that guarantee that maximal temperature will not be exceeded in the next step of the DTM algorithm. Denoting vector of maximal allowed temperatures as $T_{\max}$ one can write:

$$\mathbf{T}_{\max} = \mathbf{ET} + \mathbf{FP}_{\max}. \tag{9}$$

and compute $\mathbf{P}_{\max}$ – the vector of maximal power dissipation allowed in each node of the thermal model during the next step:

$$\mathbf{P}_{\max} = \mathbf{F}^{-1} \left(\mathbf{T}_{\max} - \mathbf{ET} \right) \geq \mathbf{P}'. \tag{10}$$

Table II
NUMBER OF PREDICTIONS WITH ERROR HIGHER THAN 2 K AND MAXIMUM ERROR FOR LAST-VALUE AND OUR PREDICTION METHOD

	$\lvert\mathbb{E}_2\rvert$				$\varepsilon_{\max}$			
pred. method	10 ms	20 ms	40 ms	100 ms	10 ms	20 ms	40 ms	100 ms
last-value T	4	4	310	320	2.88	2.88	9.29	7.7
act. last-value IPC	4	4	36	123	2.84	2.81	3.47	6.50
act. tab. pred IPC	4	4	29	112	2.83	2.85	3.22	6.17
act. oracle pred. IPC	4	4	23	94	2.92	2.82	3.00	5.99

Algorithm 2 Proposed Dynamic Thermal Management

Require: previous temperature prediction $\mathbf{T}^{\mathbf{pred}}$ and it's time t_p
Ensure: new temperature prediction $\mathbf{T}_p$
update performance coefficients (IPC, LLCM) and current frequency of each core
read the list of currently active tasks $\mathcal{Z}$
make initial temperature prediction using the current core frequencies and task-to-core mapping using Alg. 1
sort CPU cores descending by their predicted temperature
for all core pairs $i \in \langle 0..\frac{N_c}{2}-1\rangle$, $j \in \langle N_c-1..\frac{N_c}{2}\rangle$: $T_j - T_i \geq T_{th}^{mig}$ **do**
  if $LLCM_i < LLCM_{th}$ **and** $LLCM_j < LLCM_{th}$ **then**
    switch tasks between cores i and j
update power prediction $\mathbf{P}^{\mathbf{pred}}$ using the new task assignment
determine max power of each core $\mathbf{P}_{\max}$ with inequality (10)
calculate priority weights w_i for each core according to (11)
for cores i sorted by priority weights w_i **do**
  determine max allowable core frequency of i-th core $f_{max,i}$ based on (4) and (5) such that: $P(f_{\max}, i) \leq P_{\max,i}$
  if $i = 0$ or $f_{i-1} == f_{\max}$ **then**
    set $f_i \leftarrow f_{max,i}$
  else if frequency of the previous core $f_{i-1} = f_{\max}$ **then**
    set $f_i \leftarrow min(f_{max,i}, f_i)$
  else
    set the frequency one step smaller than $f_{\max,i}$
update temperature prediction using (8) and save it

Next, using (4) and (5) and iterating over all available frequencies, frequency $f'_{\max}$ can be found for each core that does not lead to power dissipation $\mathbf{P}'$ that would violate the condition: $\mathbf{T}(t+t_s) \leq \mathbf{T}_{\max}$.

Lowering the clock frequency of a core has smaller influence on the execution time of programs with relatively larger number of cache misses (see Eq. 4). Therefore, we assign each core i a weight w_i:

$$w_i = \frac{k_p - p_i}{-k_l \cdot log_{10} LLCM_i}, \tag{11}$$

dependent on current cache misses (LLCM) and priority of the task that it's executing. Then, in case of thermal stress, cores with higher weights have priority to increase their frequency. Coefficients k_p and k_l are manually chosen to equalise the influence of task priority $p_i \in \langle -20, 19\rangle$ and cache misses (we observed LLCM values from 10^{-5} to 10^{-2}).

The proposed DTM mechanism presented with Algorithm 2 was implemented in Python programming language. Standard Linux mechanisms: `taskset` program and `cpufreq` module were used to migrate tasks between cores and change clock frequency of cores. The list of active processes is read using `ps` program and the information on temperature and current CPU activity are read from MSR registers of their respective cores. The algorithm was tested on the `Intel Core 2 Quad Q9400` processor. To ensure that the user-land implementation will activate every 100 ms high priority was assign to it.

To assess the effectiveness of the presented DTM algorithm we prepared scripts running SPEC and PARSEC benchmarks. SPEC integer and floating point programs were started in fours until completion of all programs. Since runtime of individual programs differs, this test reflected the varying workload in real-world computing scenarios. To test the algorithm with multi-threaded programs, we used the PARSEC test suite with default (native) input data sets. Tests were run with 2, 3 and 4 threads to asses the effectiveness of the algorithm is moderate to severe thermal stress scenarios.

All tests runs were separated in time by 600 s to ensure equal thermal conditions. Also room temperature and CPU ambient temperature (inside the PC case) were monitored. To account for cool environment (room temperature of around 23°C) we set temperature threshold to only 55°C. All the tests were run 10 times to minimise the influence of random events on the results. After removing the results with the shortest and longest operating time, mean value of the rest of the results was calculated. We present comparison of our DTM algorithm with simple DTM algorithms using only task migration (TM) and DVFS in Tab. III. Since our DTM implementation has high overhead associated with calling external programs, we run the performance and temperature monitoring part also in the case without DTM.

Our DTM algorithm reduces both the maximal temperature and total time of thermal violation relative to simpler DVFS and TM-based algorithms. Also by migrating tasks from hot to cooler cores it reduces temperature variance between cores. It can be seen that adding temperature prediction to a simple DVFS mechanism, that changes frequency setting by one step when temperature crosses predefined thresholds, results in quite good temperature reduction. However in some situations migrating tasks can improve the total throughput by lowering temperature of the active core and allowing higher clock frequency to be used.

V. DISCUSSION AND SUMMARY

In this paper we have presented a temperature prediction algorithm suitable for dynamic thermal management of multi-core microprocessors. The prediction is based on high level

Table III
EXECUTION TIME, MAXIMAL TEMPERATURE, AND MEAN VARIANCE OF CORE TEMPERATURES OBTAINED WITH NO DTM, REACTIVE AND PREDICTIVE DVFS AND TASK MIGRATION (TM) AND THE PRESENTED PREDICTIVE DTM ALGORITHM.

	SPint	SPfp	PARS2	PARS3	PARS4
	execution time [s]				
no DTM	453	655	3175	2501	2633
DVFS	502	745	3380	2752	2949
pred. DVFS	495	732	3431	2647	2891
TM	456	645	3511	2929	3096
pred. TM	456	650	3317	2700	2637
our DTM	472	732	3360	2695	2811
	maximal temperature [°C]				
no DTM	68	68	69	67	70
DVFS	59	60	60	59	59
pred. DVFS	58	59	59	58	58
TM	67	69	64	67	70
pred. TM	69	68	67	67	70
our DTM	58	57	57	56	58
	no. periods: $max(T) > T_{\max}$				
no DTM	4359	6746	21780	14704	20919
DVFS	52	30	8937	100	3503
pred. DVFS	17	3	15	1	13
TM	4214	6704	23794	14872	23009
pred. TM	4498	6647	19978	14079	21985
our DTM	0	0	5	0	11
	mean variance of core temperatures				
no DTM	2.82	1.68	5.37	4.60	4.74
DVFS	1.49	0.63	3.39	2.85	3.25
pred. DVFS	1.66	0.69	3.92	3.66	3.69
TM	2.15	1.28	2.99	2.59	2.29
pred. TM	4.27	1.68	3.69	3.00	3.04
our DTM	1.23	0.61	1.83	1.98	1.81

power-performance and thermal models of a quad core microprocessor. These models are based on off-line analysis of detailed traces taken from a real computer. The main disadvantage of this approach is the need to repeat the measurements for any other microprocessor. Therefore, the temperature prediction requires an off-line phase in which parameters of the thermal model are determined and would greatly benefit from an automated on-line model learning approach.

The presented dynamic thermal management algorithm utilizes dynamic voltage and frequency scaling combined with task migration. It was implemented as a standalone application for the sake implementation simplicity. However this results in a high overhead associated with calling external programs. This overhead was prohibitively high with short cycle below 100 ms. On the other hand, contemporary operating systems schedule tasks with millisecond granularity and voltage transition takes tens of microseconds. Therefore, the full potential of this approach could not be verified. Implementing the algorithms presented here in the operating system will allow to operate with much lower latencies with negligible overhead.

VI. ACKNOWLEDGEMENTS

This work was supported by Polish National Science Centre (*Narodowe Centrum Nauki, NCN*), under Grant no. N N516 517939.

REFERENCES

[1] D. Brooks and M. Martonosi, "Dynamic thermal management for high-performance microprocessors," in *High-Performance Computer Architecture, 2001. HPCA. The Seventh International Symposium on*, 2001, pp. 171 –182.

[2] R. Rao and S. Vrudhula, "Efficient online computation of core speeds to maximize the throughput of thermally constrained multi-core processors," in *Proc. IEEE/ACM ICCAD*. IEEE Press Piscataway, NJ, USA, 2008, pp. 537–542.

[3] R. Mukherjee and S. Memik, "Physical aware frequency selection for dynamic thermal management in multi-core systems," in *IEEE/ACM ICCAD*. ACM Press New York, NY, USA, 2006, pp. 547–552.

[4] M. Kadin and S. Reda, "Frequency and voltage planning for multi-core processors under thermal constraints," in *Computer Design, 2008. ICCD 2008. IEEE International Conference on*, oct. 2008, pp. 463–470.

[5] ——, "Frequency planning for multi-core processors under thermal constraints," in *Proc. ISLPED '08*. New York, NY, USA: ACM, 2008, pp. 213–216.

[6] M. Kadin, S. Reda, and A. Uht, "Central vs. distributed dynamic thermal management for multi-core processors: which one is better?" in *Proc. 19th ACM Great Lakes Sym. on VLSI*. ACM, 2009, pp. 137–140.

[7] L. Xia, Y. Zhu, J. Yang, J. Ye, and Z. Gu, "Implementing a Thermal-Aware Scheduler in Linux Kernel on a Multi-Core Processor," *The Computer Journal*, vol. 53, no. 7, pp. 895–903, 2010.

[8] G. Liu and G. Quan, "Thermal aware scheduling on an intel desktop computer," in *Southeastcon, 2011 Proceedings of IEEE*, march 2011, pp. 79–84.

[9] X. Yin, Y. Zhu, L. Xia, J. Ye, T. Huang, Y. Fu, and M. Qiu, "Efficient Implementation of Thermal-Aware Scheduler on a Quad-core Processor," in *2011 IEEE 10th Int. Conf. TrustCom*, nov. 2011, pp. 1076–1082.

[10] G. Liu, M. Fan, and G. Quan, "Neighbor-aware dynamic thermal management for multi-core platform," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2012*, march 2012, pp. 187–192.

[11] A. Coskun, T. Rosing, and K. Gross, "Utilizing Predictors for Efficient Thermal Management in Multiprocessor SoCs," *IEEE Trans. on CAD*, vol. 28, no. 10, pp. 1503–1516, oct. 2009.

[12] R. Cochran and S. Reda, "Consistent runtime thermal prediction and control through workload phase detection," in *Design Automation Conference (DAC), 2010 47th ACM/IEEE*, june 2010, pp. 62–67.

[13] J. Yang, X. Zhou, M. Chrobak, Y. Zhang, and L. Jin, "Dynamic Thermal Management through Task Scheduling," in *IEEE ISPASS*, april 2008, pp. 191–201.

[14] C. Bienia, S. Kumar, J. Singh, and K. Li, "The PARSEC benchmark suite: Characterization and architectural implications," in *Proc. 17th Int. Conf. on PACT*. ACM, 2008, pp. 72–81.

[15] M. Berktold and T. Tian, "CPU Monitoring With DTS/PECI," 2009, intel whitepaper.

[16] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. Stan, "HotSpot: A compact thermal modeling methodology for early-stage VLSI design," *IEEE Trans. on VLSI*, vol. 14, no. 5, pp. 501–513, 2006.

[17] V. Hanumaiah, S. Vrudhula, and K. S. Chatha, "Maximizing performance of thermally constrained multi-core processors by dynamic voltage and frequency control," in *Proc. ICCAD '09*. New York, NY, USA: ACM, 2009, pp. 310–313.

[18] B. Wojciechowski, K. Berezowski, P. Patronik, and J. Biernat, "Fast and accurate thermal simulation and modelling of workloads of many-core processors," in *17th Inter. THERMINIC 2011*, Sep. 2011, pp. 1–6.

[19] T. Sherwood, E. Perelman, G. Hamerly, S. Sair, and B. Calder, "Discovering and exploiting program phases," *IEEE Micro*, vol. 23, no. 6, pp. 84–93, 2003.

Thermal design and characterization of solar cell arrays aimed to be used in CubeSat missions

Gyula Horváth[1], David Czifra[2], Gábor Marosy[3], Zsolt Várhegyi[4]
Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics
Magyar tudósok körútja 2, H-1117 Budapest, Hungary
Tel.: +36-1-463-2720[1,4], +36-1-463-3073 [2,3] Fax.: +36-1-463-2973
horvath@cubesat.bme.hu[1], czifra@cubesat.bme.hu[2], marosy@cubesat.bme.hu[3], varhegyi@cubesat.bme.hu[4]

***Abstract* —Besides their important place in space education, the significance of CubeSats as technological and scientific demonstrators is also increasing, due to the miniaturization and their low cost associated with the maturing CubeSat technology. The primary energy sources for these spacecraft are solar cells. It is a key question to maximize the available electrical energy gained from the minimal solar cell area available, while coping with the extreme thermal conditions of space which the panels are exposed to. In this paper we describe the structural design of solar cell arrays aimed at maximal energy yield in space along with the thermal characterization techniques used during the thermal optimization of the structure. The presented procedure resulted in an enhanced solar array design which was successfully used in the Masat-1 CubeSat mission, supported by flight telemetry data regarding solar cell temperature values.**

I. INTRODUCTION

The CubeSat standard is developed by the California Polytechnic State University and the Stanford University together [1]. This standard defines the main characteristics of the satellite, including overall size and the weight. Beyond the mechanical properties it also contains specifications about the operation and the materials ought to be used. However there are no specifications regarding the electronic components and their certifications (screening) and as a consequence a pico-satellite can be sent to space cost-effectively by academic institutes using relatively cheap, COTS components.

The primary energy sources of pico-satellies are solar arrays (Figure 1). In addition, a secondary power source is needed to store the electrical energy and to balance out the temporal differences in the supply (the solar energy supply ceases in the shadow of the Earth) and consumption. The needed extra energy is usually stored in batteries (secondary energy source).

Beyond energy balance, thermal stability is also crucial for any CubeSat mission. Both of these are tightly related to solar arrays. Solar cells usually cover a significant amount of outer surface of the satellite. Because of this, the emissivity and absorptivity of the solar cells largely determine the same properties of the whole outer surface. It is known that the ratio of the emissivity and reflectivity of the outer surface is the most significant factor in the radiative heat exchange balance between the satellite and the outer environment. This ratio, together with the Joule dissipation of the on-board systems, determine the long-time-average temperature of the satellite surface, and via conduction, the temperature of the whole satellite.

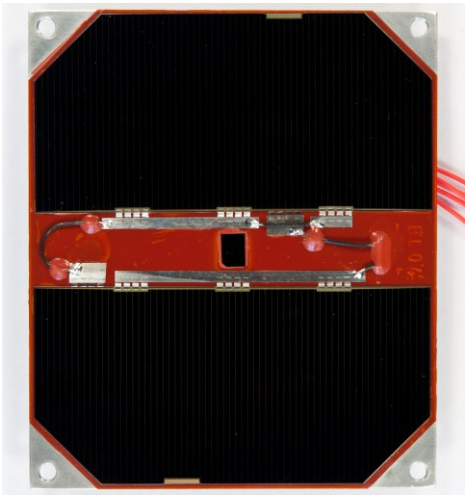


Figure 1: The complete side panel of Masat-1 with solar array

Cell efficiency [2] is highly influenced by the temperature; increasing temperature having a significant adverse effect. This is summarized by Table 1.

Temperature [°C]	80	70	60	50	40	30	20	10	0	-10	-20	-30	-40
Power/panel [W]	2.1	2.1	2.2	2.2	2.3	2.3	2.4	2.4	2.4	2.5	2.5	2.6	2.6

Table 1: The generated power of a solar array function of temperature

If cell efficiency has to be increased, one must aim for a low surface temperature, with minimal spatial and temporal fluctuations to higher values.The average temperature can be influenced, as we have seen, by the thermo-optical properties of the surface and the on-board dissipation.

The thermo-optical properties of the solar cells are largely determined by photovoltaic design criteria, and the surface areas not covered by cells are also occupied by other features such as sensors or fastening elements. There is little room

therefore to modify surface temperature by tuning the thermo-optical properties of the surface. It must be also noted that an average temperature lower than 0°C would cause difficulties in operating the other on-board systems. Decreasing on-board dissipation is also difficult, since one would like to use as much electric power as possible, and increasing solar cell efficiency is done to this end.

The fluctuation of the solar cell temperature, however, is an interesting point to investigate, especially because this phenomenon proved to be a problematic issue during earlier flights. At most pico-satellite missions it can be observed that the temperature of the solar arrays is increased during the daylight phase of orbit. If the solar arrays are significantly overheated they might reach 60-80°C as measured in earlier missions. It can be observed that such a high operational temperature leads to 8-10% power decrease. Conventional satellites might not be affected by this problem because larger safety margin in power balance. At a pico-satellite mission, however, even this decrease can cause serious difficulties in the execution of mission.

More effective energy production could be achieved by decreasing this overheating by optimizing the heat conduction path from solar arrays to solid structure of the satellite. Such a path consists of the several material layers of the side panel on which the solar cells are mounted as well as the layers of adhesive between them and interface area on the structural components directly in contact with the side panel (Fig. 2).

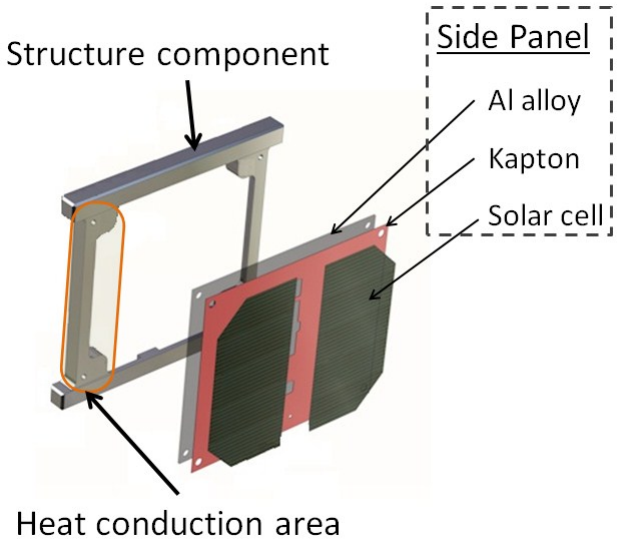


Figure 2: The heat conduction area between side panel and solid structure

Such an optimization would result in a smaller temperature difference between the solar cell surface and the structure and also in smaller temperature differences between the different points of the same panel, when there is a positive or negative heat flux trough the side panels.

The goal is to use materials with small R_{th} for the side panel stack and it is also important to keep good thermal conductivity between the mounting plate and the satellite's primary structure. In this case the heat absorbed by the solar cells can be effectively transferred to the satellite structure. Using this setup lower surface temperature can be reached and additionally another advantage is that it reduces the thermal gradient on side panel. The large temperature gradient in the side panel causes high mechanical stress. Due to cyclic extreme temperature changes solar panels (depending on orbit and orientation) go through aging, combined with the mechanical stress this leads to increased risk of delamination. Similar problem occurs in case of analyzing heatsink fixtures used in power LED applications. If the solar cell delaminates from the surface, then the thermal resistance will be increased significantly between the solar cell and the satellite structure. This leads to increased solar cell temperature and decreased power production. In case of CubeSats you have to take into account these design considerations and comply with mass limitations. Unfortunately, the space environment is significantly limiting the selectable materials, but materials with good heat conductivity properties can be chosen as a heat-spreader for solar cells.

II. STUCTURE OF SOLAR ARRAYS

Space technology sets high expectations concerning applied materials so one is able to choose from a short list of materials for producing structure of solar arrays (for example mounting plate). Furthermore side panels are used for multiple purposes in CubeSat missions, such as radiation shielding, accommodation of sensors, etc. The mass of the side panel assembly has a mass constraint; therefore solar arrays are frequently thinned as much as possible. Typical thickness of mounting plate is 1.6mm or 2.4mm, in few cases it is between 0.4-0.7mm.

Our purpose is to produce a solar array structure for certain space missions with maximal thermal conductance and keep mass of the array low at the same time. First, it is suggested to examine the layer of the substrate structure itself. The most commonly used material for this is FR4 which is mass-effective as its density is 1850kg/m³ but its thermal conductivity is very low (0.8 W/mK).

We suggest the use of Aluminum, which has a very high thermal conductivity (237.7 W/mK). Although, Aluminium density is greater (Aluminium alloy 7075T6 density = 2810 kg/m³) than the density of FR4 we do apply 7075T6 alloy which is a well-used material in space technology. It has good heat conductivity and also has great mechanical parameters. In this case we can use 1mm thick Aluminium alloy instead of the commonly used 1.55mm thick FR4. The mass difference is in our case approximately 2%. Figure 3. presents the difference between the above mentioned two stacks.

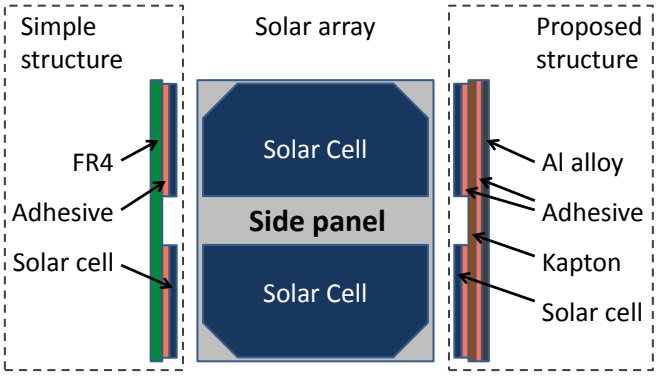


Figure 3: Different solar array structures

The thermal resistance of proposed aluminum alloy-based side panels are significantly reduced compared to the baseline design. Table 2 is clearly indicating the difference in thermal resistance between the proposed and used material of mounting plate. The thermal resistance was calculated using (1).

$$R_{th} = \frac{1}{\lambda}\frac{L}{A} \qquad (1)$$

Material	Heat conductivity λ [W/m·K]	Area A [mm^2]	Thickness L [mm]	Thermal resistance R_{th} [K/W]
FR4	0.8	8000	1.55	0.24219
Aluminium	237.7	8000	1	0.00053
Kapton	0.37	7758	0.13	0.04529
Adhesive	0.2	7758	0.2	0.12890
Solar cell	60.2	6036	0.15	0.00041

Table 2: Thermal resistance of used and proposed materials

The heat capacity of the two materials does not differ significantly. (see Table 3.) Thermal capacitance was calculated using the equation 2.

$$C_{th} = c_v \cdot A \cdot L \qquad (2)$$

Material	Heat capacity c_v [Ws/m^{3}K]	Area A [mm^2]	Thickness L [mm]	Thermal capacitance C_{th} [Ws/K]
FR4	2.000E+06	8000	1.55	24.80
Aluminium	2.391E+06	8000	1	19.13
Kapton	1.668E+06	7758	0.13	1.68
Adhesive	1.430E+06	7758	0.2	2.22
Solar cell	1.703E+06	6036	0.15	1.54

Table 3: Thermal capacitance of used and proposed materials

$$\tau = R_{th} \cdot C_{th} \qquad (3)$$

The thermal capacitance of the two designs are about the same, however, the thermal resistance is significantly decreased, so the calculated (using equ. 3.) time constant of FR4-based mounting plate is 6s, while the time constant of the aluminum-based mounting plate is 0.01s. Of course, beyond the favorable thermal parameters the suggested structure provides further advantages.

First, thanks to Aluminium alloy the satellite becomes less sensitive for radiation in space. Second, we set a Kapton foil into the structure that improves even more the thermal characteristics of structure and brings additional defense for the satellite. Heat capacity and thermal conductivity data of the materials in Tables 1 and 2 were collected from catalogs except for the thermal conductivity of the adhesive, which was measured with a setup for thermal conductivity measurement of thermal interface materials (TIM's), a development of the ASTM D5470-01 standard [3]. In the newly developed setup a power diode is used as a heater, which also serves as the temperature sensing element. The heat-flow dissipated by the diode serves as a probe for the TIM measurements. The thermal interface material is put between the cooling surface of the diode package and an aluminum block kept at known temperature. The distance between the cooling surface of the package and the metal block is adjusted in 10 um steps. By making a power step on the junction of the diode the cooling curve describing the thermal system is captured. The cooling transients at different TIM thickness values are converted to thermal structure functions, which provide the cumulative thermal resistances of the heat flow path. By plotting the resulting thermal resistance values as a function of the TIM thickness, the thermal conductivity of the TIM is calculated from the inverse of the slope of the line fit on the measurement data.

III. SIMULATIONS

First we performed preliminary simulations to examine the chosen materials and we investigated the effect on the solar array temperature.

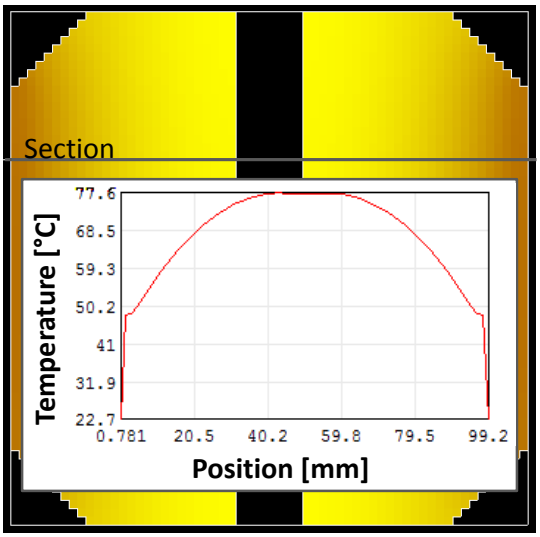


Figure 4: Simulation of simple structure

We made simulations with both of the above described stacks in SunRed thermal simulation tool [4]. We considered the thermal relation between the solar array and the satellite structure. Fig.4. presents spectacularly the structure's effect on solar array. The array is fixed to the structure on its border as the thermal gradient shows. The higher thermal resistance resulted in higher thermal gradient and it means at the same time that the risk of delamination is high. Figure 4. shows to the simple structure in SunRed simulation. In this case the solar cells' temperature in the cross section can reach 77.6°C. Figure 5. presents SunRed simulation of the suggested structure. By comparing the topside temperatures of two structures it can be observed that the Aluminium alloy remarkably reduces solar cell temperature. Due to this it is capable to produce more energy. Thanks to low thermal gradient in this case we don't need to consider delamination.

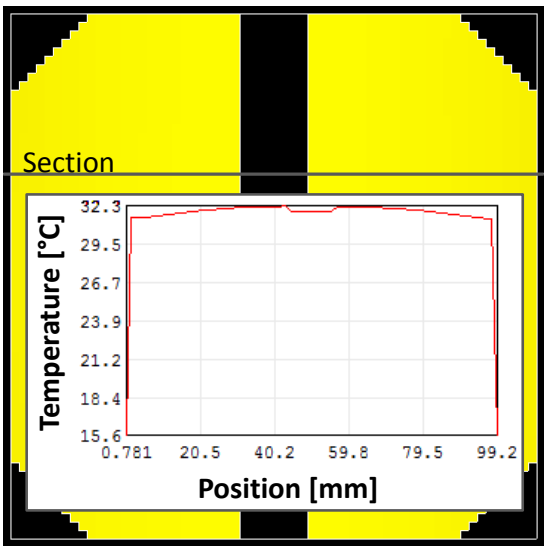


Figure 5: The Simulation of proposed structure

We performed a complete simulation of the mounting plate that was essential for thermal design of Masat-1.

IV. Results

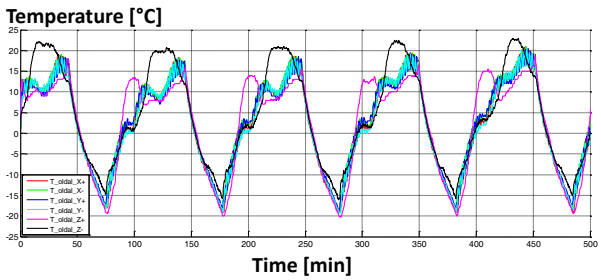


Figure 4: The temperature of side panels

The actual, in-orbit measured temperature of the side panels are shown on Figure 4. The orbital period is clearly visible on the results. The thermal peak of each side panel is lower than in the simulated case because the satellite is rotating along its axes. This led to a periodical temperature swing between a minimal and maximal value.

V. Conclusion

The article describes the procedure for thermal design and characterization of CubeSat's solar arrays used in extreme thermal conditions in orbit. With this method we have defined a mass effective structure of the side panels for Masat-1 mission. The surface temperature of the solar cells and the thermal gradient are significantly reduced, resulting less chance of solar cell delamination. Based on the simulations solar energy production is increased by 10% compared to other FR4 based solutions.

Acknowledgement

The authors would like to express their gratitude to the entire Masat-1 project and team for their outstanding work on the subsystems of the satellite making this project come true, especially András Bükkfejes, Levente Dudás, Sándor Glisics, András Gschwindt, Zoltán Kovács, Tibor Mezei, András Szimler, Tamás Temesvölgyi and Lajos Varga. We would also like to thank the Faculty of Electrical Engineering of Budapest University of Technology and Economics for hosting the project, Space Research Group of the University and European Space Agency (ESA) for granting the launch opportunity.

References

[1] LEE, Simon; MUNAKATA, Riki. CubeSat Design Specification [online]. Rev.12. 2009 [cit.2010-04-22]. http://www.cubesat.org/images/developers/

[2] Azurspace GmbH, „Triple Junction Solar Cell 3G28C Datasheet", http://www.azurspace.com, 2012.

[3] Vass-Várnai András, Sárkány Zoltán, Rencz Márta: Characterization method for thermal interface materials imitating an in-situ environment. MICROELECTRONICS JOURNAL In Press: 8 p. Paper Corrected Proof. (2011)

[4] László Pohl: Multithreading and Strassen's algorithms in SUNRED field solver. In: Proceedings of the 14th International Workshop on THERMal INvestigation of ICs and Systems (THERMINIC'08). Rome, Olaszország, 2008.09.24-2008.09.26. pp. 137-141.(ISBN: 978-2-35500-008-9)

25-27 September 2012, Budapest, Hungary

Heat spreading in a thin longitudinal fin

Wendy Luiten
Philips Research P&PM
High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
Wendy.Luiten@philips.com

***Abstract* -In view of the trend towards higher power densities in ever shrinking geometries, understanding heat spreading fundamentals is gaining in importance. In this paper heat spreading in thin longitudinal geometries is considered. This geometry is of practical interest in one-dimensional Cartesian geometries. A characteristic length is derived and it is shown that this is has physical significance for the distance that heat spreads, and for the total amount of heat cooled away. Furthermore, it is investigated when "thin" is a viable assumption. The use of the characteristic length is illustrated for the case of a line source cooling to a plate. The results are compared to numerical simulations. The work is an extension of the authors' earlier work on heat spreading in infinite longitudinal geometries and heat spreading in infite and finite circular geometries.**

***Keywords*-Heat spreading, thin longitudinal fin, plate heat sink, line source, characteristic length**

Nomenclature

Bi	*Biot number*	*(-)*
h	*heat transfer coefficient*	*W/(m^{2}K)*
k	*thermal conductivity*	*W/(mK)*
L	*length*	*m*
L_c	*characteristic Length*	*m*
q	*transferred heat*	*W*
Q	*scaled transferred heat*	*(-)*
R	*thermal resistance*	*K/W*
T	*temperature*	*C*
t	*thickness*	*m*
w	*width*	*m*
x	*distance along the lenght*	*m*
η	*scaled length distance*	*(-)*
θ	*scaled temperature*	*(-)*
$_0$	*(subscript) at x=0*	
$_e$	*(subscript) at x=L*	

I. INTRODUCTION

Simultaneous advances in miniaturization and performance are a continuing trend in the development of new generations of electronic products. Therefore, source power densities are increasing while at the same time allowable source temperatures stay the same or are lowered. To cool small sources to realistic temperatures, and obtain realistic power densities at the component and system level, heat spreading is gaining in importance.

The thin longitudinal geometry is an important form factor in heat spreading. It covers spreading along the fin height in pin fin and plate fin geometries, spreading along a line in narrow strip geometries, and heat spreading from line sources to a thin plate.

Heat spreading in the thin infinite longitudinal case was addressed previously in [1], and radial heat spreading from a small heat source on a thin plate has been addressed in [2].

The present paper addresses the derivation of the characteristic length for the longitudinal (Cartesian) case in section II, and investigates the physical meaning of this length in section III. In section III also approximations are derived for the temperature distribution along the fin and for the amount of heat transferred to the ambient. In section IV, criteria are given as to under what circumstances a fin can be considered thin. Section V shows an application example for the case of a strip heat source on a thin plate and compares to numerical results.

The results demonstrate the engineering relevance of the derived characteristic length based approximation in enabling quick engineering estimations.

II. DERIVATION OF LENGTH SCALE

In [1] the characteristic length for the infinitely long, thin longitudinal case was derived directly from the differential equation. An alternative approach was taken in [2], inspired by the work of Adrian Bejan [3], [4]. A similar approach is now demonstrated for the longitudinal case.

One of the guiding principles in the constructal approach is that relevant length scales appear when competing physical effects are of the same magnitude. In heat spreading, the competing mechanisms are the conductive heat spreading in the thin plate and the convective heat transfer from the thin plate to the ambient.

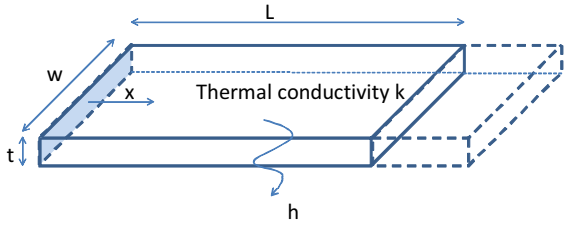


Figure 1: Longitudinal fin geometry

Consider the case depicted in Figure 1. A thin flat plate of thickness t (m), width w (m) is heated at the left side surface at x=0. The far right surface at the end is adiabatic. The

 ISBN:978-2-35500-022-5

thermal conductivity is k W/(mK). Cooling is by means of heat transfer coefficient h W/(m^2K) on one side. The heating and cooling are uniform over the width direction of the plate and the plate is so thin that the temperature is uniform over the thickness[1]. The plate's temperature distribution is a function of the coordinate x along the length direction. The thermal resistances for conductive and for convective heat transfer of a fin section of length L are:

$$R_{conduction} = \frac{L}{twk} \tag{1}$$

$$R_{conduction} = \frac{1}{Lwh} \tag{2}$$

Equating the resistances leads to:

$$L^2 = \frac{tk}{h}$$

$$L = \sqrt{\frac{tk}{h}} \equiv L_c \tag{3}$$

Expression (3) is the same L_c expression as derived from the governing differential equations in [2].

III. WHAT DOES THE LENGTH SCALE MEAN?

We now define a scaled length η and scaled temperature θ, which are used to investigate the physical meaning of L_c.

$$\eta \equiv x / L_c$$

$$\theta \equiv \frac{T - T_{ambient}}{T_0 - T_{ambient}} \tag{4}$$

Furthermore we define L as the total length of the fin, θ_0 =1 as the scaled temperature at $x=0$, the edge of the source, and θ_e as the scaled temperature at the fin tip $x=L$ or in scaled variables $\eta=\eta_e=L/L_c$.

The total heat transferred is given by

$$q = \int_0^L wh(T - T_{ambient})dx = \int_0^{\eta_e} wh\theta(T_0 - T_{ambient})L_c d\eta \tag{5}$$

Which is conveniently scaled as

$$Q \equiv \frac{q}{whL_c(T_0 - T_{ambient})} = \int_0^{\eta_e} \theta d\eta \tag{6}$$

Note that *(6)* demonstrates that Q equals the area below the $\theta(\eta)$ curve.

The temperature field in this geometry is well known [5]:

$$\theta = \frac{cosh(\eta_e - \eta)}{cosh(\eta_e)} = cosh(\eta) - tanh(\eta_e)\,sinh(\eta) \tag{7}$$

A. *Infinite case*

For long fins, η_e>2, tanh( )≈1 and

$$\theta = cosh(\eta) - sinh(\eta) = e^{-\eta} \tag{8}$$

The temperature decays exponentially with rate 1 from θ=1 at η=0, to θ=0. In exponential decay, the influence of the imposed end temperature continues indefinitely but it diminishes fast: At η=1, θ=0.37; at η=2, θ=0.14, and at η=3, θ=0.05. In other words less than 5% of the imposed temperature is found beyond a distance of $3L_c$.

The area below the exponential curve is finite and equal to 1. It was shown earlier that the scaled heat loss to the ambient, Q, is equal to this area. Thus, the heat lost to the ambient is finite as well. From *(6)*

$$Q \equiv \frac{q}{whL_c(T_0 - T_{ambient})} = 1$$

$$q = whL_c(T_0 - T_{ambient}) \tag{9}$$

This shows that the infinitely long rectangular fin with T_0 imposed at the end x=0, experiences heat loss to the ambient *as if a length L_c is heated to T_0*, and the remainder of the fin stays cold.

A linear temperature drop over distance $2L_c$ has an equal area below the curve, and is a good approximation for the exponential decay.

The exponential temperature decay, the step temperature distribution and the linear temperature drop are compared in Figure 2. In all three cases, the temperature loss to the ambient, equal to the area below the curve, is the same. The linear temperature drop with slope $1/(2L_c)$ is a reasonable approximation for the exponential decay.

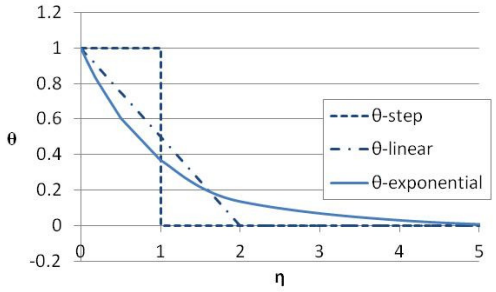


Figure 2: Temperature distribution for the infinite fin

This result has significant engineering applications. The validity of the linear approximation effectively says that in case of a long fin (longer than 2 times L_c), *an approximate distance $2L_c$ from the source is used for cooling and the temperature can be assumed to drop linearly over this distance*. Furthermore, the amount of heat lost is as if a section of length Lc was heated to temperature T_0

Thus, a source of length L_{source} on an infinite thin plate looses heat both from its own area and from the heated area on the plate. A distance of $2L_c$ from the edge of the source is used for cooling. The remainder of the plate stays cold and

[1] Relevant criteria as to "thinness" are found in section IV

does not contribute to heat transfer

$$q = (L_{source} + L_c)wh(T_0 - T_{ambient})$$

This is illustrated in Figure 3.

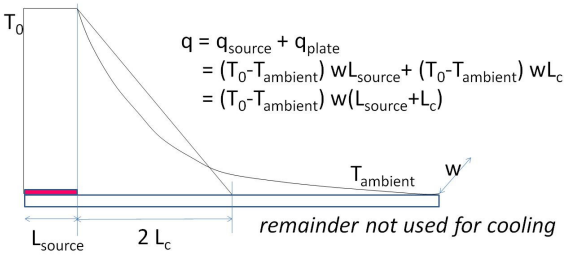


Figure 3: Temperature distribution and heat loss for the infinite fin

B. Finite case

Figure 4 shows the temperature distribution in a finite fin of $1.5L_c$ length. The solution according to *(7)* results in a curved line, and the linear drop with slope $1/(2L_c)$ results in the straight line. At short distances, the straight line overpredicts the temperature, while at the end of the fin, the temperature is underestimated. The area's below both curves, which are related to the amount of heat loss to ambient, are similar in size.

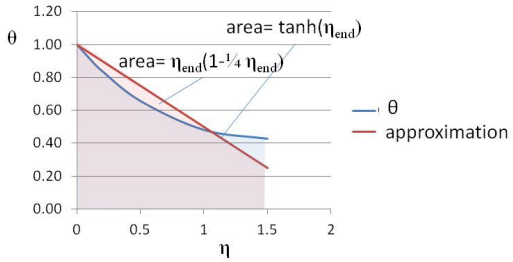


Figure 4: Temperature distribution for a finite fin

It shows that also for the finite case the linear approximation gives a reasonable first quick impression of the temperature drop and heat loss over the fin.

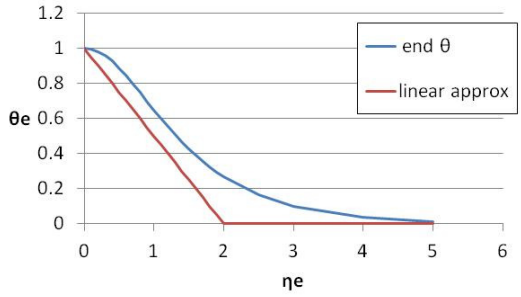


Figure 5: Fin end temperature for finite fin

Figure 5 compares the scaled temperatures at the insulated tip of the fin for a range of fins up to $\eta_e=5$, corresponding to fin lengths up to $5L_c$. This confirms that the linear estimate for the temperature at the tip of the fin, is an underestimation, but can be used as a first estimate.

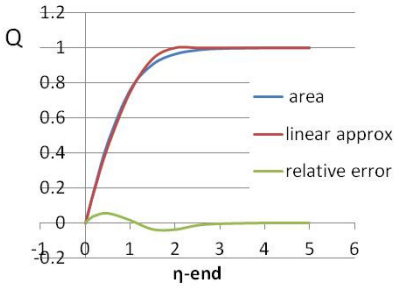


Figure 6: Heat loss for finite fin

Figure 6 compares the transferred heat, so the area below the curves, for the two approaches. It shows that linear approximation is correct within 5%.

This confirms that both for infinite and for finite longitudinal fins, the temperature distribution and heat transfer can be approximated by assuming a linear temperature drop with slope $1/(2L_c)$.

For finite thin fins, $0<\eta_e\leq 2$

$$\theta(\eta) = 1 - \frac{\eta}{2}$$

$$\theta_e = 1 - \frac{\eta_e}{2}$$

$$Q = \eta_e - \eta_e^2 / 4 \qquad (10)$$

For infinite thin fins, $\eta_e > 2$

$$\theta_e = 0$$

$$Q = 1$$

For $0<\eta\leq 2$ $\quad \theta(\eta) = 1 - \frac{\eta}{2}$

For $\eta > 2$ $\quad \theta(\eta) = 0$ $\qquad (11)$

IV. When is the fin thin?

To determine when the fin is sufficiently thin, we turn again to the constructal approach of comparing the competing thermal resistances. *(9)* shows that the fin looses heat as if a distance Lc is at temperature T_0. This corresponds to a the convective resistance of an area of width w and length L_c

$$R_{convection} = \frac{1}{wL_c h} \qquad (12)$$

The conductive resistance over the fin thickness for this area is

$$R_{conduction} = \frac{t}{wL_c k} \qquad (13)$$

The ratio of the two resistances is

$$\frac{R_{conduction}}{R_{convection}} = \frac{th}{k} \equiv Bi \qquad (14)$$

ISBN:978-2-35500-022-5

Comparing the thermal resistances for the two competing effects leads to the appearance of the dimensionless parameter group th/k, which is by definition the Biot numer [6]. If the convective resistance is much larger than the conductive resitance, the temperature drop over the thickness of the fin is negligible, and the fin can be considered thin. This corresponds to a small Biot number.

To illustrate the effect of the Biot number, 2 D conduction numerical simulations are performed on the geometry illustrated in Figure 7

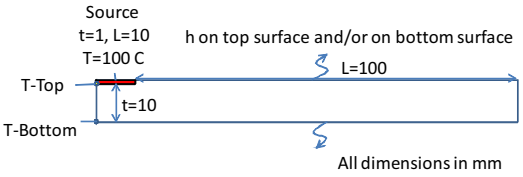


Figure 7: Geometry

A temperature source T=100 C, is embedded in the top surface of a fin. The fin is either cooled on the top surface, on the bottom surface, or on both surfaces. The Biot number is calculated as

$$Bi = \frac{t(h_{top} + t_{bottom})}{k} \qquad (15)$$

The Biot number is varied by applying different combinations of heat transfer coefficients h and thermal conductivity k, and the temperature of the point T-bottom and T-top are compared Because of the presence of the temperature source, T-top is always 100 C. The results are shown in Table 1 and Figure 8. Table 1 shows the calculated temperature opposite the heat source on the bottom of the plate, in case the heat transfer coefficient is applied on the bottom surface ('h on bottom'), on the top and the bottom surface ('top and bottom'), or on the top surface ('top').

	T-top	T-bottom			Tdifference top-bottom		
Biot		h on bottom	top & bottom	top	h on bottom	top & bottom	top
	C	C	C	C	C	C	C
0.00001	100	100	100	100	0	0	0
0.0001	100	100	100	100	0	0	0
0.001	100	100	100	100	0	0	0
0.01	100	97	98	98	3	2	2
0.1	100	86	90	94	14	10	6
1	100	48	61	87	52	39	13
10	100	13	19	80	87	81	20
100	100	6	6	78	94	94	22
1000	100	5	5	78	95	95	22

Table 1

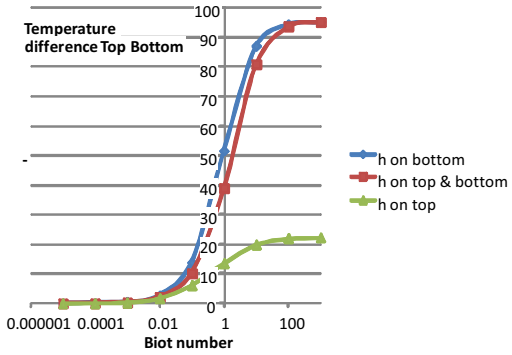


Figure 8: Temperature difference as function of Bi

For the fin to be thin, the three lines should coincide. Figure 8 shows that the fin can be considered thin if Bi≤0.1.

Up to Biot = 0.1, the calculated temperatures at the bottom for the three cooling cases equal to each other within 10%. The temperature difference over the thickness of the plate is below 10% for top cooling only, and below 15% in the bottom cooling and top/bottom cooling case. This means that for Bi≤0.1 there is no significant difference if the cooling is applied on the top or on the bottom, only the total combined heat transfer of the top and the bottom is relevant.

The threshold value Bi=0.1 is confirmed by looking at the temperature distributions: Figure 9 and Figure 10 show the temperature distributions for cooling at the top surface only, cooling at the bottom surface only, and cooling on both top and bottom surface for Bi=0.1 and Bi=1. Figure 9 shows that at Bi=0.1 the temperature distribution for the three cases is quite similar, and isotherms are nearly vertical over the thickness of the fin, which means that the temperature differences between the top surface of the fin and the bottom surface of the fin are small. This corresponds to the assumptions of "thinness."

In contrast, the temperature fields at Bi=1 shown in Figure 10 have non vertical isotherms and significant temperature differences between the top and the bottom surface. Furthermore top cooling, bottom cooling and top/bottom cooling have significantly different temperature distributions. Clearly for Bi=1 the fin is not thin.

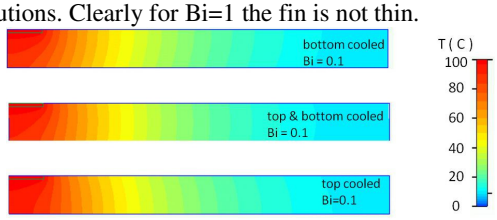


Figure 9: Temperature fields for Bi=0.1

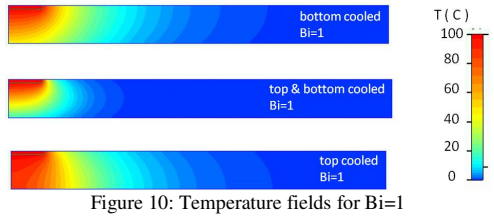


Figure 10: Temperature fields for Bi=1

ISBN:978-2-35500-022-5

V. APPLICATION EXAMPLE

In order to demonstrate the use of the characteristic length approach, cooling of a narrow strip source to a thin plate is considered for various plate materials, and the results are compared to numerical simulations.

A. *Geometry and boundary conditions*

Consider the geometry depicted in Figure 11: an isothermal 5 W strip line source of 1 mm thick, 10 mm wide, located centrally on a plate of 1 mm thickness of dimensions 100 mm x 100mm. The plate is cooled on both sides with h=10 W/m^{2}K, corresponding to free convection cooling with some radiation. What will be the source temperature if the plate is made of plastic (k=0.2 W/mK), thermally conducting plastic (k=1 W/mK), ceramic (k=25W/mK) or aluminium (160 W/mK)? What if natural convection is applied without radiation (h=5 W/m^{2}K), or if forced convection is applied (h=20 W/m^{2}K)? And what if the plate is 10 mm thick plastic with the heated strip embedded in the one side?

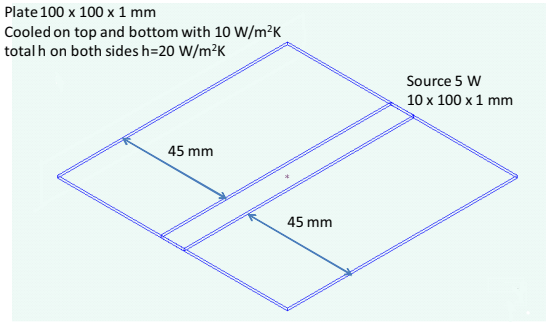


Figure 11: geometry and boundary conditions

B. *Characteristic length based solution*

Table 2 shows the Biot number calculated for the different options, using ***(15)***. Note that, since the plate is cooled on both sides, the heat transfer coefficient in the table is the sum of the heat transfer coefficient on the top and on the bottom surface. In cases a through e, Bi<0.1, so the use of the linear approximation is OK. In case f Bi = 0.2, so some difference can be expected, as the plate is not thin. In case g, Bi=1, so the characteristic length estimate is invalid. The worked case is presented for the purpose of illustration.

		t	h	k	Bi
case		mm	W/m^{2}K	W/mK	
a	plastic	1	20	0.2	0.1
b	thermal plastic	1	20	1	0.02
c	ceramic	1	20	25	0.0008
d	aluminium	1	20	160	0.000125
e	plastic, no radiation	1	10	0.2	0.05
f	plastic, forced conv.	1	40	0.2	0.2
g	thick plastic	10	20	0.2	1

Table 2

The heatspreading on the plate is approximated by application of *(10)* and *(11)*.Table 3 shows the calculated L_c and corresponding scaled fin length η_e, scaled power Q and scaled temperature at the edge of the plate, θ_e,

	L_c	η_e	Q	θ_e
case	mm			
a	3.2	14.23	1	0
b	7.1	6.36	1	0
c	35.4	1.27	0.87	0.36
d	89.4	0.50	0.44	0.75
e	4.5	10.06	1	0
f	2.2	20.12	1	0
g	10.0	4.50	1	0

Table 3

The results in Table 3 show that in case a, b, e, f and g the plate acts as an infinite fin. The 45 mm distance between the edge of the source and the edge of the plate is much larger than $2L_c$, the distance that the heat will spread over the plastic. The edge of the plastic plate is not heated by the source and will be at ambient temperature. In contrast, if the plate is made from ceramic or aluminum, as in case c and d, the characteristic length L_c is much larger, heat will spread further, and the edge of the plate will be heated by the source.

The source will transfer heat to the ambient in three ways: by its own surface, by heatspreading to its left side on the plate, and by heatspreading to its right side on the plate.

The R_{th} for heatspreading on the plate is found by rewriting *(9)*:

$$R_{th-plateL/Rside} = \frac{T_0 - T_{ambient}}{q} = \frac{1}{QwL_c h} \tag{16}$$

The R_{th} for the source surface itself is given by

$$R_{th-source\_selfcooling} = \frac{1}{wL_{source} h} \tag{17}$$

				$\Delta T_{over\ ambient}$	
	$R_{th\text{-}plate}$	$R_{th\text{-}source}$	$R_{th\text{-}total}$	source	edge
case	K/W	K/W	K/W	C	C
a	158.1	50.0	30.6	153	0
b	70.7	50.0	20.7	104	0
c	16.3	50.0	7.0	35	12.7
d	12.7	50.0	5.6	28	21.1
e	223.6	100.0	52.8	264	0
f	111.8	25.0	17.3	86	0
g	50.0	50.0	16.7	83	0

Table 4

Table 4 shows the thermal resistances for spreading on the plate *(16)* and for self cooling of the source *(17)*, as well as the total thermal resistance from application of the three thermal resistances: a plate resistance on the left of the source, the source itself, and a plate resistance to the right, in parallel. Furthermore, the resulting calculated temperatures for the source and for the edge of the plate are shown, based on the source power dissipation of 5 W and relative to the ambient temperature.

C. *Comparison to numerical solution*

2D conduction numerical simulations were performed, with a heat transfer coefficient applied to the top and to the bottom surface.

ISBN:978-2-35500-022-5

	ΔT over ambient					
	simulation		estimated		difference	
	source	edge	source	edge	source	edge
case	C	C	C	C	C	C
a	154	0	153	0	0.9	0.0
b	104	0	104	0	0.2	0.4
c	36	18	35	13	0.5	5.7
d	27	24	28	21	-1.3	2.7
e	264	0	264	0	0.0	0.0
f	87	0	86	0	0.9	0.0
g	120	2	83	0	36.4	2.1

Table 5

Table 5 shows the approximated source and edge temperatures are in good agreement with the simulation results except for the last case. This was as expected for case a through e, which had Bi <0.1. Some difference was expected for case f, with Bi=0.2. In case g, with Bi=1, no agreement was expected between the estimation and the numerical result as the estimation was invalid; the plate is not thin.

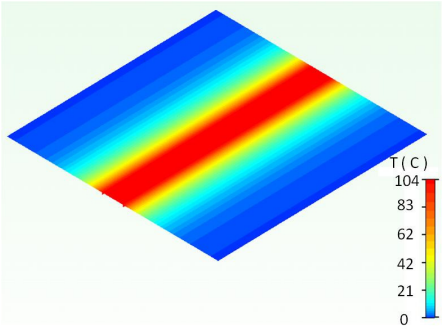


Figure 12: Temperature field for case a

Figure 12 shows the simulated temperature field for case a. This confirms that in case a, the plate acts as an infinite fin and the heat from the source does not reach the edge of the plate.

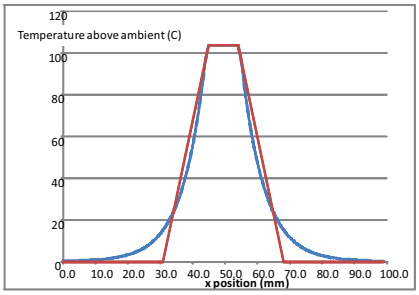


Figure 13: Temperature comparison

Figure 13 shows the simulated temperature distributions versus the linear approximation for case a. The temperature distributions are in good agreement.

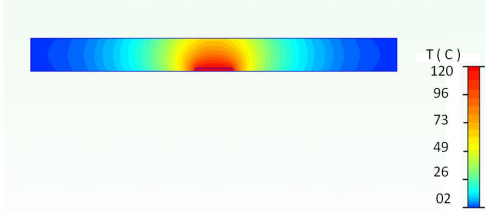


Figure 14: cross-sectional temperature distribution for case g

Figure 14 shows the cross-sectional temperature distribution for case g. It confirms that indeed the 10 mm plastic plate cannot be considered thin, as was already apparent from Bi=1. Therefore, the characteristic length approximation is invalid, and the estimated value from the approximation does not agree with the numerical results.

VI. SUMMARY AND CONCLUSION

The characteristic length approach from was extended to longitudinal fins of finite length. It was shown that the temperature field extending from the heated edge of the fin can be approximated by a linear drop towards ambient temperature over a distance of $2L_c$, with

$$L_c \equiv \sqrt{\frac{tk}{h}}$$

For finite thin fins, $0<L\leq 2L_c$

$$\theta(\eta) = 1 - \frac{\eta}{2}$$

$$\theta_e = 1 - \frac{\eta_e}{2}$$

$$Q = \eta_e - \eta_e^2 / 4$$

For infinite thin fins, $\eta_e > 2$

$$\theta_e = 0$$

$$Q = 1$$

For $0<\eta\leq 2$ $\quad \theta(\eta) = 1 - \frac{\eta}{2}$

For $\eta > 2$ $\quad \theta(\eta) = 0$

Furthermore it was shown that the plate is sufficiently thin if the corresponding Biot number is <0.1 , which boils down to $t<0.1\ k/h$.

The characteristic length approach has significant engineering value as it enables quick engineering estimations and a means of translating thermal quantities to dimensions and material choices relevant to the mechanical design engineer.

The Characteristic length approach was demonstrated for cooling of a line source on a thin plate of different thermal conductivities, and results agreed well with expectations.

ACKNOWLEDGMENTS

The support of Philips Electronics and of the TVP innovation sites Eindhoven and the Bruges, is gratefully acknowledged.

REFERENCES

[1] Luiten, W, "characteristic length and cooling circle", Semitherm 26, San Jose 2010

[2] Luiten, W, "heat spreading from a small source on a thin plate", Semitherm 28, San Jose 2012

[3] Bejan, A. "Shape and structure, from engineering to nature", Cambridge University press, 2000

[4] Bejan, A. and Lorente, S."Design with constructal theory" John Wiley & sons, 2008

[5] H. Y. Wong, "heat transfer for engineers", Longman London, 1977

[6] http://en.wikipedia.org/wiki/Biot_number

 ISBN:978-2-35500-022-5

25-27 September 2012, Budapest, Hungary

Electro-thermal characterization of a differential temperature sensor and the thermal coupling in a 65nm CMOS IC.

Josep Altet[1], José Luis González[3], Dídac Gomez[1,*] , Xavier Perpiñà[2], Stephane Grauby[4], Cedirc Dufis[1], Miquel Vellvehi[2], Diego Mateo[1], Stefan Dilhaire[4], Xavier Jordà[2].

1. Electronic Engineering Dep. Universitat Politècnica de Catalunya. Spain

2. IMB-CNM (CSIC). Bellaterra – Spain

3. CEA-Leti, Minatec Campus, Grenoble, France

4. Laboratoire Ondes et Matière d'Aquitaine, Université Bordeaux I. France.

***Abstract*- This paper explains the design decisions and the different measurements we have done in order to characterize the thermal coupling and the characteristics of temperature sensors embedded in a integrated circuit implemented in a CMOS 65nm technology. The circuit contains a 2GHz linear power amplifier, MOS transistors behaving as heat sources and two differential temperature sensors. Temperature measurements performed with the embedded sensor are corroborated with an Infra-Red camera and a laser interferometer used as thermometer.**

I. INTRODUCTION

One of the applications of embedding temperature sensors in a silicon die along with other circuits (so called circuits under test) is to characterize either their structural integrity [1, 2] (i.e. the presence of structural defects) or their performances [3].

Examples can be found in the literature where the circuit under test is a high frequency analog circuit: the work of [4] presents the characterization of the central frequency and 1dB compression point of a 1GHz low noise amplifier from low frequency temperature measurements. In [5], the efficiencies of two Power Amplifiers (PAs), operating at 2GHz and 60GHz, respectively, are monitored through DC temperature measurements. The basic principle behind this proposal is that as temperature depends on the power dissipated by the circuit, and power depends on the voltage and current signals within the circuit, then temperature is an indirect way to observe the magnitude of voltages and currents.

The understanding of the correlation between the temperature measurements and the performances of the high frequency analog circuit requires of an accurate thermal coupling characterization between the circuit to monitor and the temperature sensor placed close to it, as well as a good characterization of the temperature sensor that is used to perform the measurements.

The aim of this paper is to show the design decisions and the thermal coupling characterization that exists between differential temperature sensors and devices/circuits embedded in the same silicon die and to show the complete procedure we follow in order to understand the behavior of the sensor. The integrated circuit (IC) we use in this paper to perform the measurements is the same that has been used in [5] to correlate DC temperature measurements with high frequency performances of a 2GHz CMOS power amplifier fabricated in a commercial 65nm CMOS technology.

The paper is structured as follows: in the following section II we describe the integrated circuit. Section III presents a detailed characterization of the temperature sensor. Temperature measurements performed with the embedded sensor are corroborated with other off-chip temperature sensing strategies: an IR camera and a laser interferometer used as thermometer. These comparisons are performed in section IV. Finally, section V concludes the paper.

[*] D. Gómez is currently with Broadcom.

II. Integrated Circuit Description

Fig. 1 shows the schematic of the PA designed for a 2GHz transmitter for coax-cable communications. It has been implemented using a CMOS 65nm process. The inductors L_1 and L_2, as well as the C_{DC} capacitors, are off-chip components, and are used to center the PA in one of four possible sub-bands (2-2.5GHz). Details about this design can be found in [6]. The main characteristics measured from the PA are: Gain @2.3GHz = 17.8 dB, $P_{DC}@V_{DD}=1.1$ V = 96mW, OCP1 = 10.5 dBm.

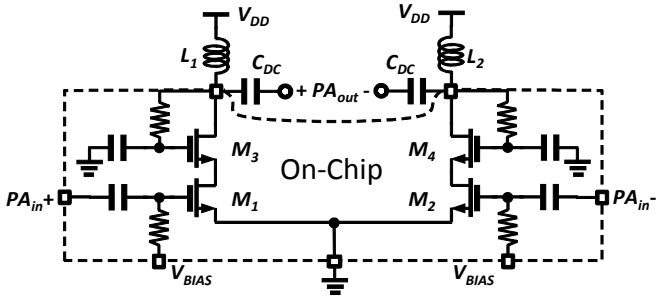


Fig. 1. Schematic of the 2GHz Power Amplifier.

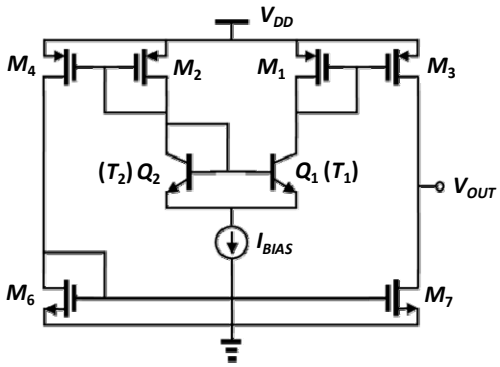


Fig. 2. Schematic of the differential temperature sensor embedded in the same IC than the Power Amplifier.

Fig. 2 shows the schematic of the differential temperature sensor embedded with the power amplifier. This sensor is based on the structure published in [7]. The name differential comes from the fact that the sensor output voltage variations are proportional to the difference of temperature of two devices of the sensor circuit, acting as temperature transducers. The two temperature transducers are the bipolar transistors Q_1 and Q_2, whose temperature is T_1 and T_2 respectively. Then, the output voltage of this sensor can be written as:

$$V_{OUT} = S_{TD}\left(T_1 - T_2\right) + V_{Offset} \quad (1)$$

where S_{TD} is the differential sensitivity of the sensor in V/°C, and V_{offset} is the sensor output voltage when there is no thermal imbalance between both temperature transducers. In this design, the temperature transducers are vertical NPN bipolar transistors built using the deep-nwell/pwell /n+diffussion structure available in this CMOS process. The working principle is as follows: when the temperature transducers operate with different temperature, the differential pair is unbalanced. The current mirrors and the high impedance of the output stage convert the change appearing in the transistors collector current provoked by changes of the temperature difference ($T_1 - T_2$) into changes of the output voltage V_{OUT}. To avoid self-heating, the power consumption of the sensor is $P_{DC}(@V_{DD}=1.2V)$ = 800 µW.

Fig. 3 shows the placement of the temperature sensor and the PA devices in the IC layout. The temperature transducer Q_2 is placed close to the power amplifier (at 25 microns from the PA Cascode MOS transistor of one of the branches of the PA, M_3 in Fig. 1). The temperature transducer Q_1 is placed at 240 microns from Q_2, together with the other devices that form the temperature sensor. The goal of this placement is to make the temperature imbalance between both transducers proportional to the power dissipated by the PA. As thermal coupling will affect more Q_2 than Q_1, the former is called hot transistor, whereas we named the latter cold transistor. The hot transistor is placed closer to the cascode MOS transistor (M_3 in Fig. 1) than the Input MOS (M_1 in Fig. 1) as the first will experience higher RF amplitude in the drain to source voltage, thus having a power dissipation with stronger correlation to the RF PA behavior [4].

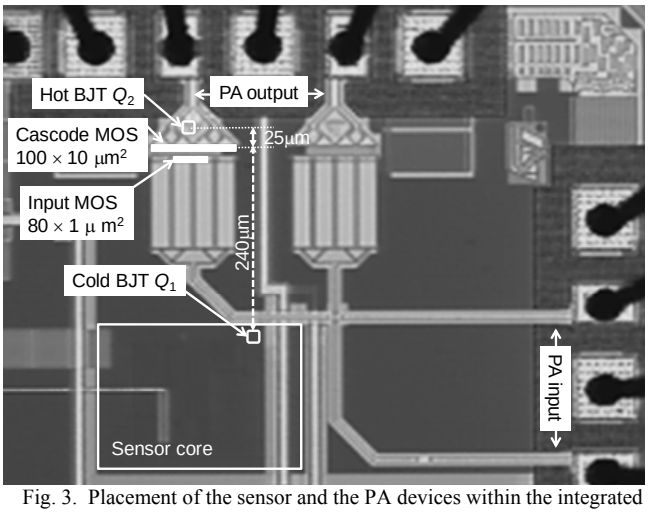


Fig. 3. Placement of the sensor and the PA devices within the integrated circuit layout.

To perform a detailed thermal coupling characterization, in the same silicon die we have placed another identical differential temperature sensor (same dimensions, same device placement, same distance between hot and cold transducers) jointly with three heat sources. The heat sources are MOS transistors in diode configuration, i.e. with their gate tied to the drain and connected to an external pin, whereas their source is tied to ground. Two of the heat sources are sized 20/0.5. The first, named Heat Source A in Fig. 4, is placed at 11 microns from the hot transducer, whereas the second, called Heat Source C in Fig. 4, at 11 microns from the cold transducer. The third transistor, called Heat Source B in Fig. 4, has been placed at 25 microns from the hot temperature transducer. This latter heat source has the same dimensions than the PA cascode transistor (100/10) and the distance between the hot transducer and this transistor is the same than the existing between the PA

cascode transistor and the hot transducer of its temperature sensor. This last heat source is intended to provide an environment for thermal characterization as close as possible to the one existing between the cascode transistor in the power amplifier and the hot temperature transducer.

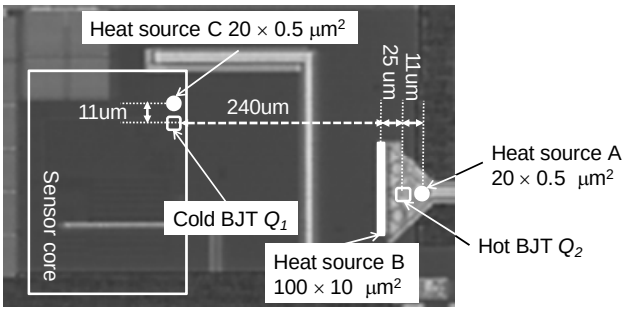


Fig. 4: Detail of the placement of the temperature sensor used for the thermal coupling characterization and the three devices acting as heat sources..

III. Temperature Sensor Characterization

A. Characterization of the differential sensitivity. DC measurements

The differential operation of the sensor can be characterized if the temperature between both transducers is unbalanced. To achieve this, we activate only one of the heat sources.

First, we characterize the heat sources: Fig. 5 shows the power dissipated by each heat source as a function of the gate voltage applied to its gate. As we will not be able to measure the real temperature at the transducer location, we will characterize the temperature sensors as a function of the power dissipated by each heat sources. As a matter of a fact, this is more interesting to our final application: to derive the power dissipated by the Power Amplifier by observing the sensor output voltage.

Fig 6 shows the sensor output voltage evolution as a function of the power dissipated by each heat source when the other two are off (Sensor power supply V_{DD} = 1.2V). The sensor output voltage variations that appear when the heat source A and C are activated show the symmetrical behavior of the sensor expressed by equation (1): when the heat source A is active, the hot transducer increases its temperature and the sensor output voltage decreases. On the other hand, when the heat source C is active, the cold transducer increases its temperature and the sensor output voltage increases. As these two heat sources are identical in size and placed at the same distance from each temperature transducer, they are supposed to induce similar temperature increase to the closest temperature transducer. However, heat sources B and A are different in size and distance. Therefore, it is interesting to see that due to the different power density and to the different distance between the heat source and the transducer, the plot obtained when Heat source B is activated is different from the one obtained when the Heat source A is activated: clearly, for the same amount of dissipated power, the heat source A induces a higher temperature increase to the hot temperature transducer than the heat source B.

Since we cannot measure the temperature of the transducers, we can only express the sensitivity of the sensor as a function of the power dissipated by each heat source. We define the sensitivity as the slope of the static transfer function (i.e. the derivative of the functions extracted in Fig. 6). To get this figure of merit, we have fit the measurements of Fig. 6 with a second order polynomial function. Fig. 7 shows the three sensitivities (expressed in V/W) as a function of the power dissipated for the three heat sources obtained in this way. There are some interesting points to analyze from the results shown on these plots: first, as expected from equation (1), the sensitivity associated to the heat source A is negative, whereas the one associated to the heat source C is positive. However, from equation (1), we would expect identical sensitivity in absolute value, which is not the case. This can be attributed to several factors. The first one is the sensor topology itself: the differential sensitivity depends linearly on the sensor output resistance [7]. Due to the topology of this sensor, the output resistance has strong dependence on the operating point of the two output MOS transistors: M_3 and M_7 in Fig. 2. So, if the output voltage value changes, so does the sensor output resistance. In addition to this, heat source C also modifies the temperature of the sensor core devices. Moreover, the base current of both bipolar devices depends on temperature, which introduces an additional source of unbalance. Finally, the low V_{DD} value and the fact that the information is given through voltage and not current: when the output voltage is close to any of the supply rails, the sensor enters in saturation and sensitivity nulls.

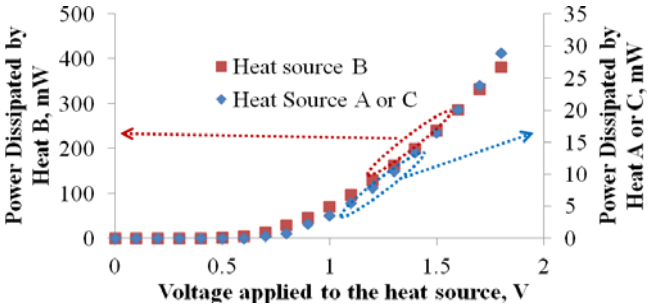


Fig. 5: Power dissipated by each heat source as a function of the voltage applied to its gate. Left Axis: Heat source B. Right Axis: Heat sources A or C. Sample 1.

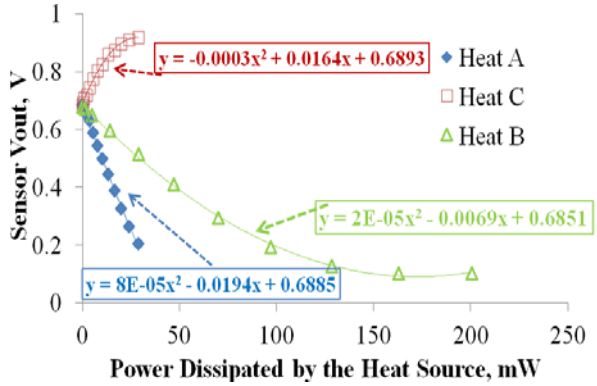


Fig. 6: Sensor output voltage as a function of the power dissipated by each heat source (the other two are off). VDD = 1.2V. Sample 1.

The dependence of the sensitivity on the sensor output voltage and on the proximity to the rail voltage can be easily demonstrated by measuring the evolution of the sensor output voltage as a function of the power dissipated by heat source A in two different conditions. First, when both heat source B and C are off, second when the heat source C is dissipating 12.48mW (i.e. 1.2V applied to its gate). This is plotted in Fig. 8. As it can be seen, the fact of activating the heat source C introduces an offset to the sensor output voltage (as expected from equation 1). This plot shows also the polynomial approximation that we can get from the measurements. The discrepancy with the fit got in Fig. 6 comes from the fact that these measurements are performed on another sample, thus we observe as well process variations. The main difference between the sensitivity measured with the sample 2 from the one measured with the sample 1 is that in the sample 2 the sensor starts to saturate when its output voltage reaches 250 mV approximately. The sensitivity extracted from the previous measurements is plotted in Fig. 9. As it can be seen, the sensitivity that we obtain when the heat source C is activated is much more constant, due to the fact that the sensor output voltage is always far away from saturation.

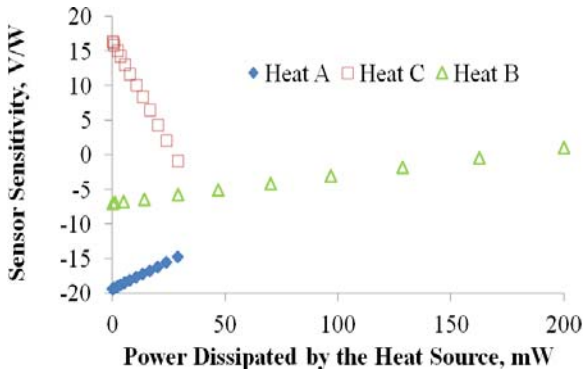


Fig. 7: Sensor differential sensitivity (V/W) as a function of the DC power dissipated by each heat source. Sample 1.

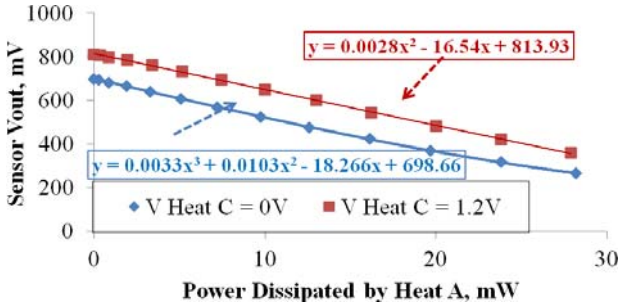


Fig. 8: Sensor output voltage as a function of the power dissipated by heat source A. Two values of power dissipation by heat source C: 0W and 12.48mW. V_{DD} = 1.2V. Sample 2.

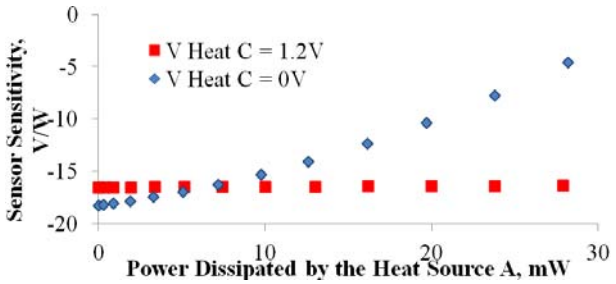


Fig. 9: Sensor differential sensitivity (V/W) as a function of the DC power dissipated by the heat source A. Sample 2.

B. Characterization of the common sensitivity. DC measurements

Ideally the output voltage of the sensor should be only sensitive to the difference of temperature between both temperature transducers. There are two factors that make the output voltage change due to variations in the common temperature of both transducers (i.e. $(T_1+T_2)/2$). One factor is the systematic mismatches that are present between the two branches of the differential pair, either introduced by process variations, or introduced by the design. An example of such systematic mismatch is the base current that biases both temperature transducers, which is taken from one of the branches. Another factor is the finite value of the output resistance I_{BIAS} in Fig. 2.

Common sensitivity should be low in order to avoid output voltage changes due to ambient temperature changes or due to the power dissipated by devices that have been intentionally placed equidistant from both temperature transducers, in order to precisely cancel the sensitivity of the output voltage to the power dissipated by these devices.

In order to evaluate the common sensitivity of the sensor, we activate simultaneously heat source A and heat source C with the same voltage source. In this configuration they dissipate the same amount of power and generate similar temperature increases at the transducer location. Results are shown in Fig. 10 for both samples. As it can be seen, the output voltage varies from sample to sample. Whereas in sample 1 it changes about 60mV, it only changes about 15mV in Sample 2. In both cases, variations are almost zero when the power dissipated by the heat sources has a low value.

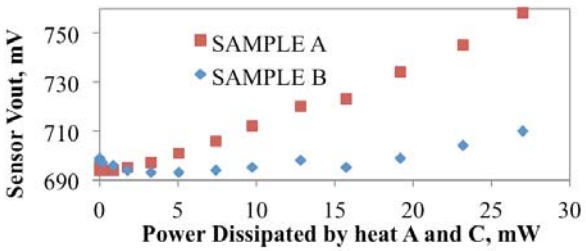


Fig. 10: Sensor output voltage as a function of the power dissipated simultaneously by heat sources A and C. Sample 1 = Sample A. Sample 2 = Sample B.

C. AC Differential Measurements.

The goal of this section is to extract the bode diagram that relates the amplitude and phase of the spectral component of the sensor output voltage at the frequency *f* when the heat source B dissipates power following a sinusoidal function at the same frequency. The interest of such measurement is to characterize the frequency response of the sensor and the frequency response of the thermal coupling, which is of interest when heterodyne temperature measurements are performed [3, 4]. To achieve a controllable sinusoidal power dissipation, the heat source B is biased with a gate voltage of V_g=1+0.05·cos(2π*ft*). The MOS transistor is a non-linear device. Therefore, this sinusoidal voltage will generate a dissipated power that can be written as the sum of different harmonic functions of frequency multiple of the input frequency *f*. According to the results shown in Fig. 5, these harmonic functions are superimposed to a continuous term of 70mW. From Fig. 7, the DC sensitivity of the sensor is -4.1 V/W when the dissipated power is 70mW. Focusing on the fundamental harmonic of the temperature variation, the measured bode diagram is shown in Fig. 11. As it can be seen, at 100Hz (the lowest frequency measured) the sensor has a response of 4.08V/W of magnitude and a phase shift of 173˚, very close to the expected -4.1V/W. As the frequency increases, the sensitivity of the sensor decreases, due to the limited bandwidth of both the temperature sensor and the thermal coupling.

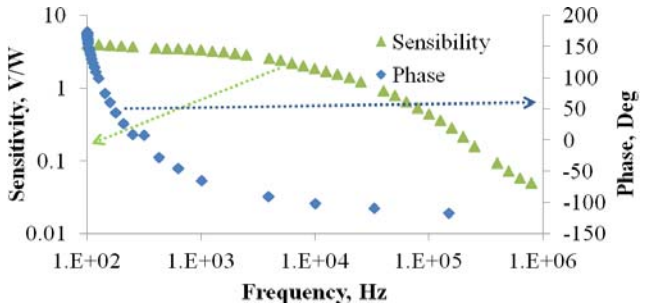


Fig. 11: Frequency response of the sensor. Sensor biased with Vgate= 1V + 0.05V·cos(wt).Amplitude and phase of the sensor output voltage as a function of the input frequency. Sample 1.

IV. Corroboration with Other Measuring Strategies

In addition to the sensor characterization performed using the devices that dissipate power in a controllable way that are placed within the IC, we corroborate the information extracted from the embedded sensor with other sensing strategies. On one hand, this can give additional information about the sensor. On the other hand, this corroboration ensures that the information given by the sensor depends on temperature variations and not due to electrical couplings that may appear in the IC between the dissipating devices and the sensor.

A. IR measurements.

We have used IR measurements in order to validate DC temperature measurements. The IR measurement set-up consists of a camera that senses the IR radiation in the short-wavelength range (3-5 μm). The lens used allows for a spatial resolution of 6 μm. The sample has been mounted on a xyz micropositioning stage to focus and select the inspected areas. In this paper we are going to show measurements of the average IR emissions coming from an area of 6 μm x 6 μm (i.e. the reading provided by just on pixel of the 2D image provided by the camera). In the measurements, no emissivity correction has been performed. However, when relative measurements are performed IR measurements do not depend on the sample surface emissivity and this compensation is not required [8].

To perform the corroboration, we have biased the power amplifier with 1.2V and we have swept the V_{BIAS} (see Fig. 1) from 0V to 0.9V. This bias voltage controls the DC current that flows through the amplifier and, therefore, its DC power dissipation. The lower the V_{BIAS} voltage, the lower the DC current flowing through the amplifier and lower its power dissipation. Fig. 12 compares the power dissipated by the power amplifier, the information given by the embedded sensor and the IR camera. We have focused the camera over the cascade transistor M_3 (see fig. 1). Values are normalized to its maximum value. As it can be seen, both the readings from the camera and from the embedded sensor are directly proportional to the DC power dissipated by the amplifier, meaning that the embedded temperature sensor is correctly reacting to the temperature increases generated by the power amplifier.

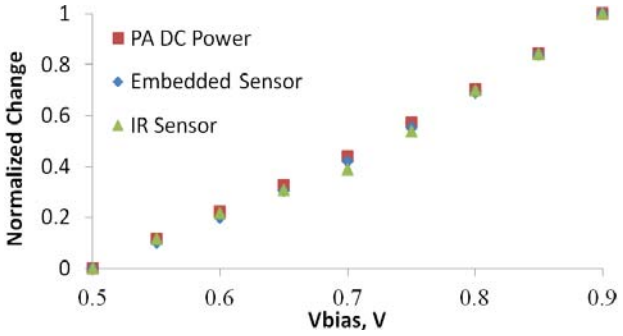


Fig. 12: Comparison of the power dissipated by the power amplifier, the embedded sensor output voltage and the IR camera readings as a function of the power amplifier bias voltage. Values are normalized to its maximum value.

B. Interferometry measurements

With the help of a laser interferometer we have validated the AC sensor characterization. Details about the laser interferometer used as thermometer can be found in [9]. Fig 13 indicates where the laser has been focused next to heat source B (Diode in Fig. 13) in order to perform the measurements. Fig. 14 compares the measurements performed with the interferometer and the differential sensor, normalized to its maximum value. In this figure we have also included the amplitude of the bode diagram that exists between the heat source B and the hot temperature transducer obtained by solving the heat transfer equation in the Laplace domain. The agreement between the laser and

the differential sensor measurements indicate that it is the thermal coupling which is imposing the lowest cut off frequency, not the differential temperature sensor.

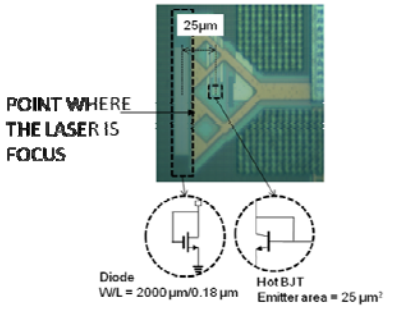


Fig. 13: Point where the laser is focus in order to perform interferometric measurements. Diode is the heat source B: MOS transistor in diode configuration.

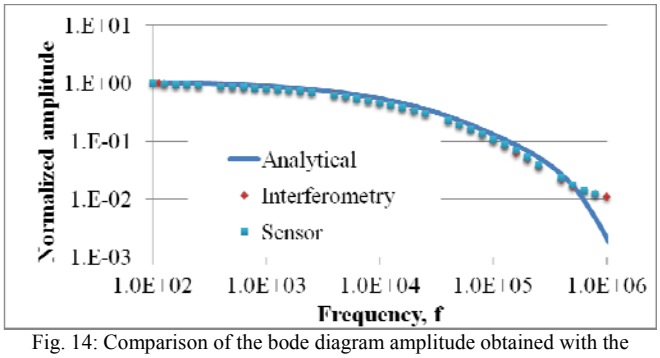


Fig. 14: Comparison of the bode diagram amplitude obtained with the interferometer, the embedded temperature sensor and the analytical solution of the thermal coupling between the heat source B and the Hot transducer. Values are normalized to its maximum value.

IV. CONCLUSIONS

In this paper we have presented design decisions and the strategy we have followed in order to characterize the thermal coupling and the performances of a differential temperature sensor implemented in a standard 65nm CMOS technology.

We have characterized the differential and common modes operation of the embedded temperature sensor by using MOS transistors connected in diode configuration that are used as controllable heat sources. Such results indicate that the on-chip sensors are suitable for tracking the temperature changes produced by different operating condition of the RF circuits.

The measurements performed with the embedded temperature sensor have been corroborated with external sensing techniques: an infra-red camera and a laser interferometer used as thermometer.

ACKNOWLEDGMENT

This work has been partially supported by the project TEC2008-01856, RyC2010-07434 (Ramón y Cajal contract), TEC2008-05577 (THERMOS) from the Spanish Government.

REFERENCES

[1] J.M. Soden, R.E. Anderson, "*IC Failure Analysis: Techniques and Tools for Quality and Reliability Improvement,"* Proceedings of the IEEE. Vol. 81, no. 5, pp. 703-715, May 1993.

[2] L. Abdallah, H. G. Stratigopoulos, S. Mir, J. Altet, "Testing RF circuits with true non-intrusive built-in sensors", Proc. DATE 2012, 1090-1095

[3] D. Mateo, J. Altet, E. Aldrete-Vidrio "An approach to the electrical characterization of analog blocks through thermal measurements" Proc. Therminic 2005, 59-64.

[4] Aldrete-Vidrio E, Mateo D, Altet J, Salhi MA, Grauby S, Dilhaire S, Onabajo M, Silva-Martinez J (2010) Strategies for built-in characterization testing and performance monitoring of analog RF circuits with temperature measurements. Measurement Science and Technology 21(7):075104(10pp).

[5] J.L. González, B. Martineau, D. Mateo, and J. Altet "Non-invasive monitoring of CMOS power amplifiers operating at RF and mmW Frequencies using an on-chip thermal sensor," *IEEE Radio Frequency Integrated Circuits Symposium Digest of Papers*, 2011.

[6] C. Dufis, J.L. Gonzalez, "Design of a Wideband Class-A Power Amplifier for Wireline Communication, Proc. DCIS 2010, pp. 339-343.

[7] Altet J, Rubio A, Schaub E, Dilahire S, Claeys W (2001) Thermal coupling in integrated circuits: application to thermal testing. IEEE J Solid-State Circuits 36(1):81-91.

[8] M. Vellvehi, X. Perpiñà, G. L. Lauro, F. Perillo, and X. Jordà. "Irradiance-based emissivity correction in infrared thermography for electronic application", Review of Scientific Instruments, 82, 114901 (2011).

[9] J. Altet, E. Aldrete-Vidrio, D. Mateo, X. Perpina, X. Jorda, M. Vellvehi, J. Millan, A. Salhi, S. Grauby, W. Claeys and S. Dilhaire, "A heterodyne method for the thermal observation of the electrical behavior of high-frequency integrated circuits", Measurement Science and Technology, 19, 114704 (2008).

25-27 September 2012, Budapest, Hungary

Electro-thermal modeling and measurements of SiGe HBTs

Mountakha Dieng*, Abdel-Kader El Rafei*, Raphael Sommet*, Raymond Quéré*
* XLIM-UMR CNRS n°6172 – 7, rue Jules Vallès – 19100 Brive La Gaillarde
Tel: 0033 555 867 301, Fax: 0033 555 861 426, E-mail:mountakha.dieng@xlim.fr

***Abstract*- This paper deals with a comparison between electro-thermal nonlinear simulation and measurements performed on a Silicon Germanium (SiGe) Heterojunction Bipolar Transistors (HBTs). The first part describes the simulation approach developed with ANSYS 3D Finite Element (FE) model. The second part describes the measurement process. This one is based on Low Frequency electrical impedance measurements generally achieved with S-parameters Vector Network Analyzer setup. Finally the FE model is compared to a Cauer circuit extracted to represent the measured data.**

I. INTRODUCTION

HBTs have become of great interest for high power and high frequency applications. The needs to increase electrical performances, cut-frequency as well as the reduction of geometrical dimensions induce many changes which sometimes cause an increase of the device temperature. Self-heating in bipolar transistors is a crucial point for many circuits that used silicon (Si) and silicon dioxide (SiO2). The increase of self-heating for advanced SiGe technologies is mainly due to the use of the silicon dioxide in the Deep Trench Isolation (DT) used to reduce the parasitic capacitances of the circuit and their poor thermal dissipation properties [1].

The first section presents the simulation aspects and the FE model of the SiGe HBTs performed with ANSYS. The second one is dedicated to the measurement method and some results concerning the theory to determine the thermal impedance using h-parameters approach. Measurements will be fitted with a Cauer network. Finally in the last section the transient simulation of the model is performed with Advanced Design System (ADS) and compared to the FE approach.

II. ANSYS THERMAL MODEL

The transistor used in this study is a two finger Infineon's high-frequency SiGe bipolar technology B7HF200 [3]. A cross section of this device is given in figure 1.

This transistor is a two finger transistor with four metallization levels. Only the first one is represented on figure 1, but all are fundamental for the thermal aspects and have been taken into account for the thermal study.

The major parts of the materials that make up this transistor are silicon (Si) and silicon dioxide (SiO2). The thermal conductivity of SiO2 increases the thermal impedance.

The Deep Trench protects the circuit for the parasitic capacitances and increases also the thermal effects which are harmful for the circuit reliability. The material properties used for the simulation are non-linear. The thermal conductivity K_T (1) [4] and the specific heat C_T (2) [4] of some materials depend on the temperature according the following laws.

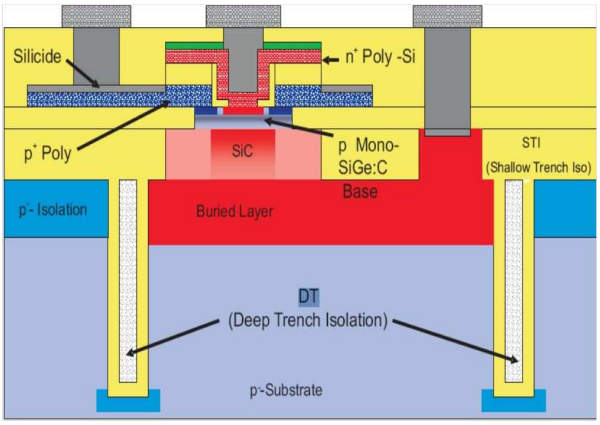


Fig. 1. Schematic cross section NPN transistor [9].

$$K_T = K_{T_0}\left(\frac{T_0}{T}\right)^b \quad (1)$$

$$C_T = C_{T_0} + C_1\left(\frac{\left(\frac{T_0}{T}\right)^\beta - 1}{\left(\frac{T_0}{T}\right)^\beta + \frac{C_1}{C_{T_0}}}\right) \quad (2)$$

K_{T0} and C_{T0} are respectively the thermal conductivity and specific heat at 300K. C_1 is a constant depending on the material. The parameter values for the simulation are summarized in Table 1 and 2.

Material	K_{T0}(W/K.m)	b
Si	148	1.3
Ge	60	1.25
SiO_2	1.38	-0.33
SiC	340	1.5

Table I. PARAMETER VALUES FOR THE THERMAL CONDUCTIVITY

SiC parameters obtained from [5] are relative to a doping level near 10^{17} cm–3.
The thermal conductivity of SiGe depends on the two materials percentage with formula (3).

$$K_{SiGe}(T_0) = \frac{1}{\frac{0.7}{K_{Si}} + \frac{0.3}{K_{Ge}} + \frac{0.7 \times 0.3}{C_k}} \quad (3)$$

The exponent b of formula (1) is linearly interpolated because of the lack of experimental data at temperatures other than 300K [4]. C_k = 2.8 W/K.m is introduced to reduce the thermal conductivity with the increase of material composition [4].

$$b_{SiGe} = 0.7 \times b_{Si} + 0.3 \times b_{Ge}$$

Material	C_{T0} (J/K.Kg)	C_1(J/K.Kg)	β
Si	711	255	1.85
Ge	360	130	1.3
SiO_2	1000	820	1.5

Table II. PARAMETER VALUES FOR THE SPECIFIC HEAT

For SiGe the expression of the specific heat is calculated by expression (4) with the percentage of germanium of 30%.

$$C_T(SiGe) = 0.3 \times C_T(Ge) + 0.7 \times C_T(Si) \quad (4)$$

The ANSYS element used for the mesh of the SiGe transistor (2x0.3x5.8µm2) is SOLID 70 (Fig.2).

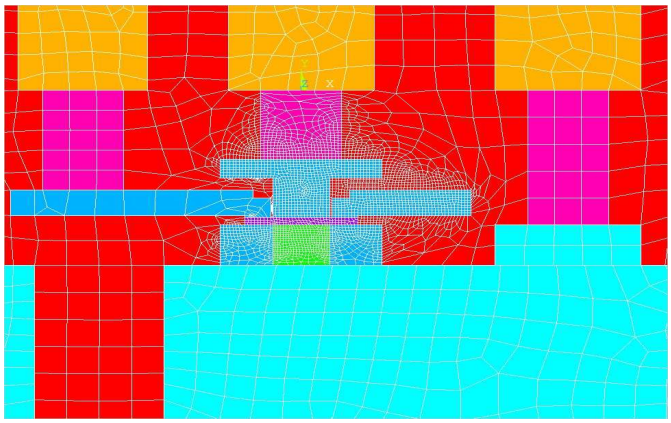


Fig. 2. Finite Element model of the B7HF200

Conditions for the simulation are the following. The dissipated power is 20 mW for both transistors so 10mW for each transistor and the baseplate temperature is 300K.
A transient simulation is performed between 10ns and 3ms. Simulation results are shown in the figure 3.

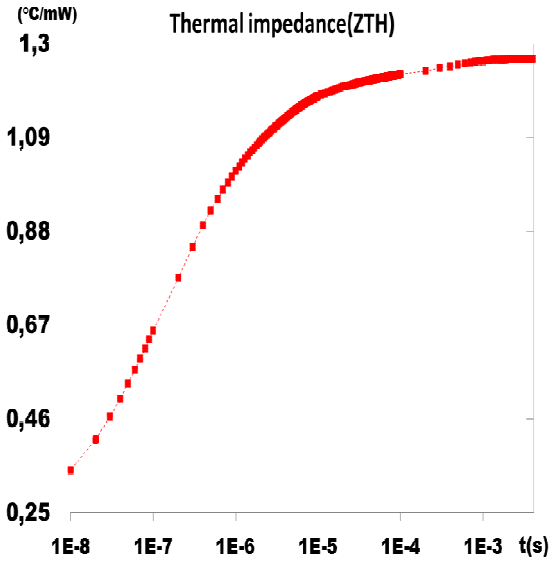


Fig. 3.ANSYS Thermal impedance model of B7HF200.

The thermal resistance obtained by ANSYS simulation is 1280 °C/W as shown in figure 3.

III. THERMAL IMPEDANCE MEASUREMENT

The method used to determine the device temperature is an electrical based method. This method has been developed in our labs since several years and has proved to be efficient on several technologies.

Generally speaking, a transistor can be considered as a two port network as shown in figure 4.

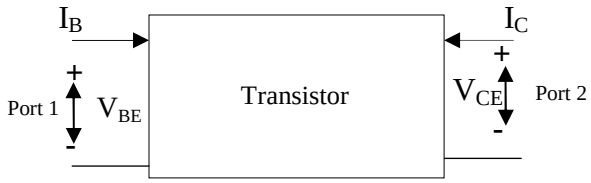


Fig. 4. Two port network

When the device is biased, non-linear relationships between currents, voltages, and what we call the junction temperature that can be expressed by equation (5).

$$V_{BE} = f(I_B; V_{CE}; T_J)$$
$$I_C = g(I_B; V_{CE}; T_J) \quad (5)$$

T_j results from a heat dissipation in the thermal impedance of the device.

$$T_J = T_0 + Zth\ P \quad (6)$$

Small signal parameters can be measured around this DC

bias point. They correspond to the linearization of the device around this fixed point. The terminal controlling any part is specified by voltage V_{BE} and V_{CE} at port 1 and 2 respectively and current I_B and I_C entering at port 1 and 2 respectively.

Several basic relationships between electrical AC currents and voltages can be derived (Z-parameters, Y-parameters, S-parameters…).

Among these various relationships, the h-parameters depend respectively on currents and voltage parameters with an hybrid form as shown in equation (7).

If the input current I_B and output Voltage V_{CE} are taken as independent variables, the input voltage V_{BE} and output current I_C can be written as:

$$\begin{aligned} v_{BE} &= h_{11}\, i_B + h_{12}\, v_{CE} \\ i_C &= h_{21}\, i_B + h_{22}\, v_{CE} \end{aligned} \quad (7)$$

Measurements are based on the S-parameters by using a low frequency Vector Network Analyzer (VNA). The S-parameters are then transformed into h-parameters using classical transformation [8]. The current collector I_C and the base-emitter voltage V_{BE} [11] depend on the collector-emitter voltage V_{CE}, the base current I_B and the temperature junction T_J.

It has been shown in [6] [12] that the thermal impedance can be derived from the bias parameters (I_{B0}, V_{BE0}, I_{CO}, V_{CE0}), the thermal behavior of V_{BE} vs T (where ϕ is the slope) and I_C vs T (where γ is the slope) and finally the h-parameters. The relationship between h-parameters, the current, the voltage and the thermal impedance of the transistor is given in [2]. Computation of the thermal impedance detailed in [2] is remembered here:

$$\tilde{Z}_{TH} = \frac{h_{12} - h_{12_{ISO}}}{\Phi\left(Ico + h_{12} I_{B0} + h_{22_{ISO}} V_{CE0}\right) + \gamma V_{CE}\left(h_{12} - h_{12_{ISO}}\right)} \quad (8)$$

The index "ISO" denotes the values of the h-parameters when the temperature is maintained at a constant value. ϕ, γ are respectively the derivatives of Vbe and Ic versus temperature. Many simplifications can be performed regarding the numerical value of parameters. They are presented in [2].We obtain finally:

$$\tilde{Z}_{TH} \approx \frac{h_{12}}{\Phi Ico} \quad (9)$$

So the thermal impedance for HBTs is the image of h12.The thermal impedance magnitude and phase [2] obtained by using expression (9) is shown in figure 5.

The thermal resistance R_{TH} is 1270°C/W (value obtained at low frequency). Phase of Zth is 90° at 200MHz and magnitude equal to zero. It means that thermal effects do not affect the device if the frequency of the AC signal is higher than this frequency limit.

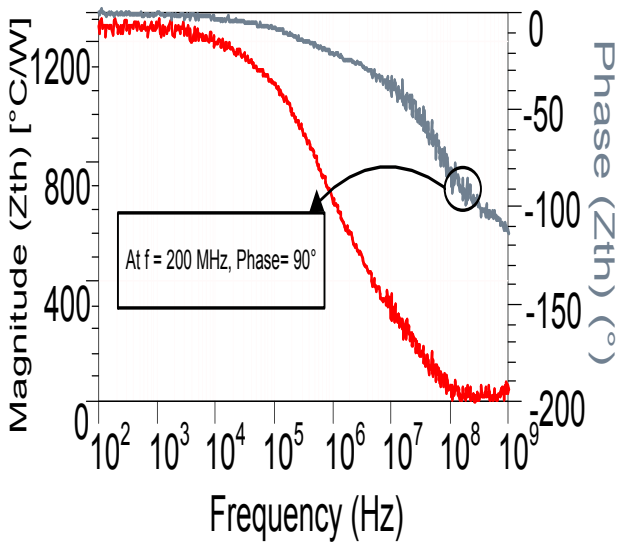


Fig. 5: Thermal impedance measurements

In the ADS circuit simulator a simplified Cauer network model has been developed to fit with frequency measurements. The Cauer network time constants [7] depend on all the resistances Ri and capacitors Ci, R_{TH} is the thermal resistance and Z_{TH} the thermal impedance. The Cauer network model is composed of 9 resistors and 8 capacitors. All the resistors and capacitors are proportional respectively to R_0 and C_0. This relationship is given in (9).

$$R_i = R_0 \times K^i \text{ and } C_i = C_0 \times K^i \quad (9)$$

R_1=0.1 Ohm R_0=310 Ohm;
C_0= 1.9e^{-11} F K=2.5.
R_0, C_0 and K are obtained by optimization

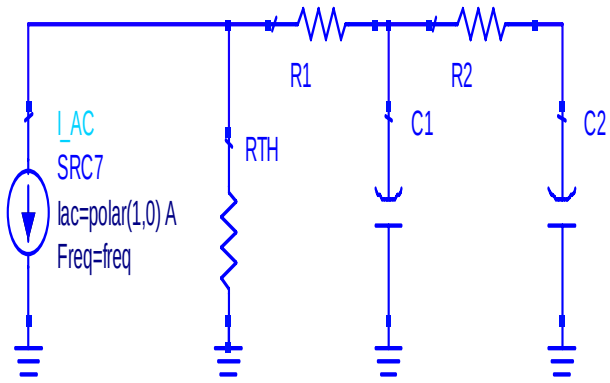


Fig. 6. Cauer network for transient thermal impedance [2].

The Cauer model is compared to measurements in the frequency domain of measurements: 10Hz-100MHz.

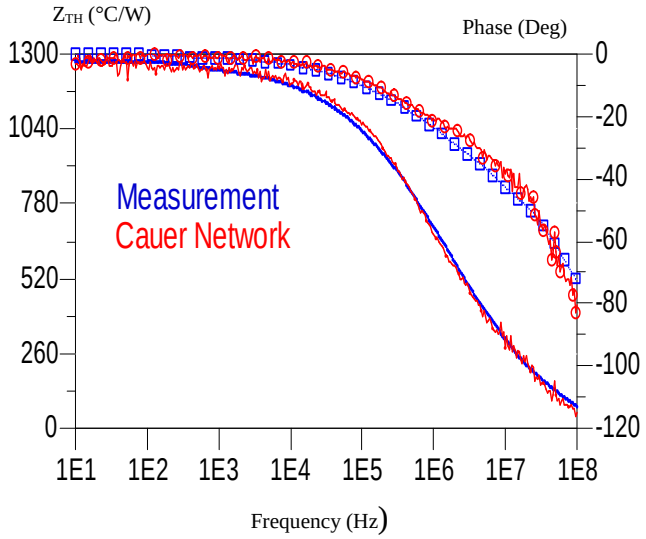


Fig. 7: Comparison between measurements and Cauer network

There is a good agreement between measurements and the Cauer Network model [10]. This step validates our lumped distributed model.

IV. COMPARAISON ANSYS MODEL AND CAUER NETWORK MODEL

In this section we compare the simulation results obtained in section II with the Cauer model but this time in the time domain.

Transient response (fig.6) between 0.01µs-1ms is compared with the finite element solution of the ANSYS model depicted in figure 3, only the magnitude is compared and presented in figure 8.

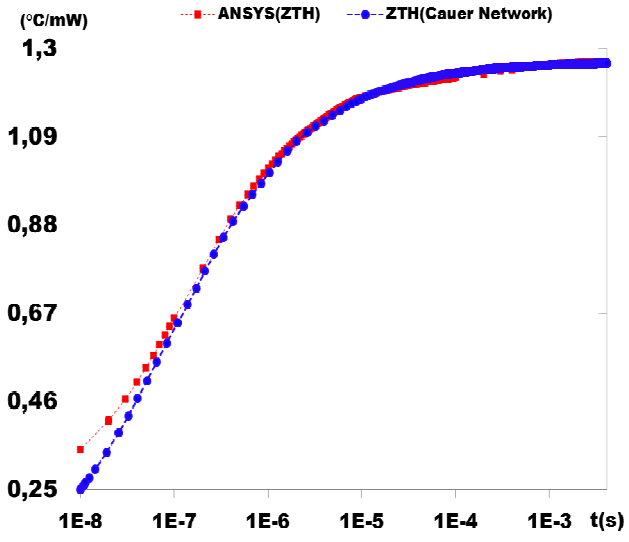


Fig. 8. Comparison transient Cauer network and ANSYS simulation

Figure 8 exhibits also a good agreement between Cauer network transient model and ANSYS simulations.

The difference between the two thermal resistances is 10°C/W or 0.8 percent. This can explained by the fact that some materials used are linear they are independent of temperature namely copper, aluminum and poly-silicon. Moreover the transistor model described in ANSYS is not the real model. Some assumptions have been made to limit the number of nodes for decreasing the computation time.

V. CONCLUSION

In this paper a comparison between 3D thermal model of SiGe bipolar transistors performed in ANSYS and original measurements obtained with h-parameters approach has been presented. Results cross validated both approaches on the band [10Hz-100MHz]. A Cauer Network model is proposed for electro-thermal coupled simulation.

REFERENCES

[1] J.-S. Rieh, D. Greenberg, Q. Liu, A. J. Joseph, G. Freeman, and D. C. Ahlgren, “Structure optimization of trench-isolated sige hbts for simultaneous improvements in thermal and electrical performances, Electron Devices, IEEE Transactions on, vol. 52, no. 12, pp. 2744–2752 2005.

[2] Abdelkader El Rafei, Alah Saleh, Raphael Sommet, Jean Michel Nebus and Raymond Quere, Fellow, IEEE “Experimental characterization and modeling of the thermal behavior of SiGe HBTs”

[3] T. Meister, H. Knapp, H. Schafer, K. Aufinger, R. Stengl, S. Boguth, R. Schreiter, M. Rest, W. Perndl, M. Wurzer, T. Bottner, and J. Bock, «High-speed sige hbt technology and applications to mm-wave circuits,” in Silicon Monolithic Integrated Circuits in RF Systems”, 2004. Digest of Papers. 2004 Topical Meeting on, Sept. 2004, pp. 61 – 64.

[4] http://www.iue.tuwien.ac.at/phd/palankovski/node34.html

[5] Jon C. Freeman Glenn Research Center, Cleveland, Ohio “Channel Temperature Model for Microwave AlGaN/GaN HEMTs on SiC and Sapphire MMICs in High Power, High Efficiency SSPAs”

[6] A. El Rafei, R. Sommet, and R. Quere, “Electrical measurement of the thermal impedance of bipolar transistors,” Electron Device Letters, IEEE, vol. 31, no. 9, pp. 939 –941, sept. 2010.

[7] F. N. Masana: A straightforward analytical method for extraction of semiconductor device transient thermal parameters. Microelectronics Reliability 47(12): 2122-2128 (2007)

[8] Vadim Issakov “Microwave circuit for 24GHz Automotive Radar in silicon-Based Technologies”

[9] http://www.elektroniktutor.de/grundlg/hparam.html

[10] F. N. Masana: “A new approach to the dynamic thermal modeling of semiconductor packages”. Microelectronics Reliability 41(2001): 901-912.

[11] F. X. Sinnesbichler , G. R. Olbrich: “Electro-Thermal Large-Signal Modeling of SiGe HBTs”. Microwave Conference, 1999.

29th Euro0pean.

[12] A. de Souza, J.-C. Nallatamby, M. Prigent, and R.Quere, "Dynamic impact of self-heating on input impedance of bipolar transistors," Electronics Letters, vol. 42, no. 13, pp. 777-778, june 2006.

25-27 September 2012, Budapest, Hungary

Electro-thermal High-Level Modeling of Integrated Circuits

Jean-Christophe KRENCKER, Jean-Baptiste KAMMERER, Yannick HERVÉ, and Luc HÉBRARD, *Member, IEEE,*
Institut d'Electronique du Solide et des Systèmes (InESS)
23, rue du Lœss – BP 2067037 Strasbourg Cedex 02
[krencker, kammerer, herve, hebrard]@iness.c-strasbourg.fr

***Abstract* —** Operating temperature and temperature gradients are of critical concern in the design of planar integrated circuits (ICs) and are bound to be exacerbated in the upcoming 3D technologies. However, a thermal aware design of ICs allows thermal issues to be kept to the minimum. Previously, a simulator integrated in the Cadence® environment that allows electro-thermal simulations to be carried out at transistor levels has been presented. In this paper, a methodology used to build high-level electro-thermal model compliant with this simulator is detailed.

Index Terms — Electro-thermal, Circuit Simulation, High-Level Model, Verilog-A, VHDL-AMS.

I. INTRODUCTION

The operating temperature as well as the temperature distribution over integrated circuits (ICs) is a major issue. Indeed, increase of delays in lines [1-3], power leakage, mismatches in analog circuits, thermal noise or mechanical stress are some of critical hindrances due to high operating temperatures [4-8]. In the upcoming 3D technologies such issues are bound to be ever more critical than in planar technologies [5, 7, 8]. If these issues cannot be totally avoided, it is at least possible to minimize them by optimizing the floor plan of ICs [6, 7, 9-11]. Therefore, it becomes necessary to take accurately the electro-thermal behavior of the IC into account since the early stages of the design flow.

In our previous works [12-13], a direct electro-thermal simulator developed for this purpose has been presented. Its operation principle consists in generating a thermal network representing the silicon die and to couple it to the electrical schematic of the IC. For this purpose, the standard electrical models of the devices are replaced by their electro-thermal counterparts. Those electro-thermal models have an additional terminal which is used to keep track of the temperature and to inject the heat flux generated by the device in the thermal network.

Standard electrical simulations cannot be performed for LSI or VLSI systems at transistor level. For such systems, simulations rely on electrical high level behavioral modeling [14-15]. This consists in describing the electrical behavior between the different analog or digital ports of functional sub-circuits with a reduced set of equations. For instance, the electrical behavior of an operational amplifier (op-amp) can be described by an arctangent like function that links the output voltage to the differential input. Concerning electro-thermal simulations, high level modeling is even more critical since both the electrical and the thermal aspects are modeled. However, accurate high-level electro-thermal models need to be sensitive to the temperature profile that appears on the sub-circuit during the simulation. For example, a temperature difference between the transistors of the differential pair of an op-amp will be responsible for an offset variation.

The proposed high-level modeling principle is described in the next section. Then, three different modeling approaches which depend on the surface occupied by the considered electronic sub-circuit are presented. In the third section, the modeling of a proportional to absolute temperature (PTAT) sensor formerly presented in [13] is detailed. Then, the results obtained from the simulations at transistor level of a test chip including 16 PTAT sensors are compared to the results obtained with two different high-level models of the sensor.

II. HIGH-LEVEL ELECTRO-THERMAL BEHAVIORAL MODELING

A. Principle

In a conventional electrical high-level modeling process, one assumes a flat and constant temperature profile over the modeled sub-circuit. Such models are built through the analysis of the electrical behavior between the electrical ports of the sub-circuit. The more, since conventional electrical simulations operate at a preset temperature, the temperature sensitivity is extracted through parametric simulations. The high-level model of a n-terminals sub-circuit can thus be described by a set of n equations:

$$V_i = f_i(V_1,\ldots, V_n, I_1, \ldots, I_n, T) \quad (1)$$

with $i \in [1, n]$ and where V_i and I_i are the voltage and the

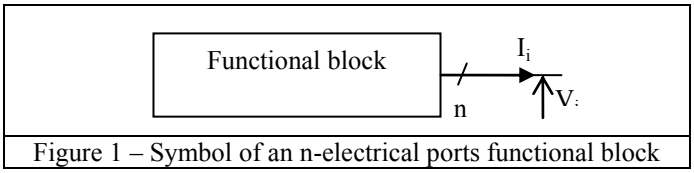


Figure 1 – Symbol of an n-electrical ports functional block

 ISBN:978-2-35500-022-5

current at the i^{th} port and T is the constant operating temperature of the functional sub-circuit during the simulation. However, such an approach is not suited for accurate high-level electro-thermal modeling.

In electro-thermal analysis the information of temperature profile is available in the thermal network and this profile may be heterogeneous during the simulation. Also, the heat generated by each component has to be reported in the thermal network. This information can be included in the model by adding as many thermal terminals as components included in the modeled sub-circuit. In such a situation, for an electrical sub-circuit (figure 1) which has *n* electrical ports and made of *m* components, *m* thermal terminals are needed. Thereby, an electro-thermal model can be extracted through parametric analysis where the electrical behavior between each electrical port is studied for every setup of components temperature. The extracted model can thus be written as follow:

$$\begin{cases} V_i = f_i(V_1, \dots, V_n, I_1, \dots, I_n, T_1, \dots, T_m) \\ P_j = g_j(V_1, \dots, V_n, I_1, \dots, I_n, T_1, \dots, T_m) \end{cases} \quad (2)$$

with $i \in [1, n]$ and $j \in [1, m]$ and where P_j are the powers dissipated (heat flux) by the *m* components and passing through the j^{th} thermal terminal. Finally, the obtained model is very accurate but may be too complex and may be too much time consuming to extract. Building such high-level model is thus meaningless. However, the complexity of this model can be strongly reduced by using only a few thermal terminals. The temperature map can be sampled from the thermal network and thus an approximate heat flux map can be reported into the thermal network. The number of thermal terminals sets the tradeoff between accuracy and simplicity. This point is detailed in the following subsections.

B. Single thermal terminal modeling

Previously, it was stated that generally the conventional high-level modeling approach is not suited to build an electro-thermal model. However, if the thermal profile of a functional sub-circuit remains essentially flat when used in its targeted application then such an electrical model can be converted into an electro-thermal one by simply adding a single thermal terminal to sample in the thermal network the average temperature over the surface covered by the sub-circuit. This assumption is generally correct in small sub-circuits made of only few components. Nevertheless, in this case the heat flux generated from the sub-circuit can only be reported as homogeneous in the thermal network. Consequently, the temperature dependence of the sub-circuit can be extracted through conventional electrical simulations for a given range of temperatures. Finally, the high-level electro-thermal model is written so that its electrical behavior depends on the temperature at its single thermal terminal (instead of the conventional fixed temperature) and the heat flowing through this terminal is defined as equal to the electrical power dissipated by the whole sub-circuit. The extracted model can be written as:

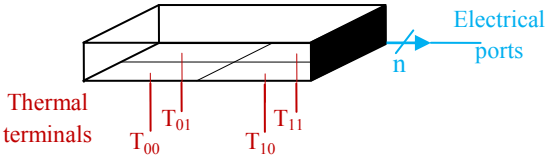


Figure 2 – Symbol of a four thermal terminals model

$$\begin{cases} V_i = f_i(V_1, \dots, V_n, I_1, \dots, I_n, T) \\ P = g(V_1, \dots, V_n, I_1, \dots, I_n, T) \end{cases} \quad (3)$$

with $i \in [1, n]$.

C. Multiple thermal terminals modeling

When the requirements for one thermal terminal modeling are not met, several thermal terminals must be used. Usually, such modeling is needed for sub-circuits of large dimension where a heterogeneous temperature map may appear on it. In this case, the surface of the sub-circuit is split in *p* columns and in *q* lines. This leads to a $p \times q$ thermal terminals model. Here, the thermal terminals are used to sample the temperature map of the sub-circuit and to report the heat flux map generated by the sub-circuit in the thermal network. The resolution of this map is one p^{th} of the sub-circuit width and one q^{th} of its length. Consequently, the power consumption map and its electrical behavior when the sub-circuit is submitted to different temperature profile need to be analyzed. For didactical reasons, the modeling principle of a 2×2 thermal terminals high-level model is described in the next paragraph.

a) Temperature dependence extraction:

In this case, the governing equation set (2) is reduced to the following system:

$$V_i = f_i(V_1, \dots, V_n, I_1, \dots, I_n, T_{00}, T_{01}, T_{11}, T_{10}) \quad (4)$$

where T_{xy} is the temperature of the thermal terminal corresponding to the zone located in x_{th} column and the y_{th} line (figure 2).

Running a parametric analysis by setting fixed temperatures on the areas covered by the T_{xy} thermal terminals creates thermal discontinuities in the temperature profile of the block and convergence problems arise. As a result, another temperature setup has to be devised. The temperatures at each T_{xy} terminals represent four averages of values sampled in a two-dimension temperature distribution (scalar field). From these temperatures, an interpolated approximation of the real distribution is used in the model. Thus, these temperatures represent projections of the effective thermal distribution on a

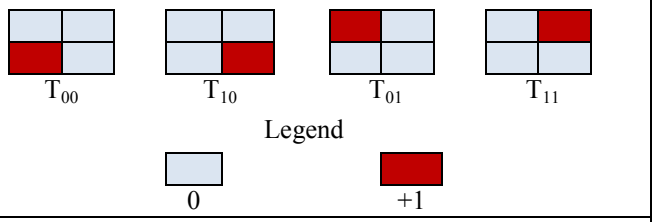


Figure 3 – Four zones base vectors

specific four vectors base presented in figure 3. However, many different bases may be used to describe the temperature scalar field. Especially, the temperature profile can be expressed as a linear combination of the average temperature over the whole sub-circuit and its spatial derivatives:

$$V_i = f_i\left(V_1,\ldots,V_n, I_1\ldots,I_n, T_{avg}, \frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial^2 T}{\partial x\,\partial x},\right) \quad (5)$$

This formalism is more convenient since these four parameters can be controlled by setting specific thermal boundary conditions during the extraction process of the model. Also, during electro-thermal simulations these parameters are evaluated in the high-level model from the individual temperatures obtained on the thermal terminals. In the case of a block which area is divided in four identical regions, the correspondence for each of the thermal terms is given by the following equations set:

$$\begin{aligned} T_{avg} &= \frac{T_{00}+T_{10}+T_{01}+T_{11}}{4} \\ \frac{\partial T}{\partial x} &= \frac{T_{00}+T_{10}-T_{01}-T_{11}}{2.W} \\ \frac{\partial T}{\partial y} &= \frac{T_{00}+T_{01}-T_{10}-T_{11}}{2.L} \\ \frac{\partial^2 T}{\partial x\,\partial y} &= \frac{T_{00}+T_{11}-T_{10}-T_{01}}{4.W.L} \end{aligned} \quad (6)$$

where W and L are the width and length of the modeled sub-circuit.

To extract the model's parameters, it is necessary to generate the electro-thermal network representing the portion of the die holding the sub-circuit and to apply the thermal boundary condition depicted in figure 4. However, the self-heating of the components must be disabled during the extraction process otherwise artifacts on the required imposed temperature profile will appear. To extract the dependence on the first order derivatives, the opposite sides of the thermal network are connected to ideal temperature sources while the other ones are left unconnected (adiabatic conditions). Therefore, data are obtained through parametric simulations in which the temperature set on one side is increased by dT while the temperature on the other side is decreased by dT. Finally, the influence of the gradients can be modeled as a perturbation of the electrical behavior obtained for a flat

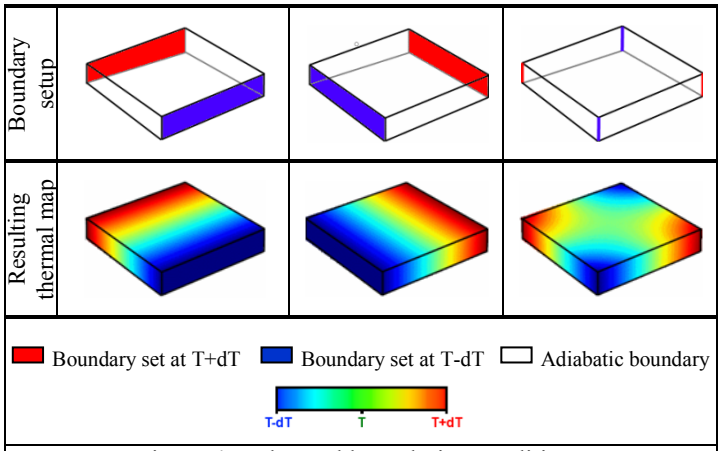


Figure 4 – Thermal boundaries conditions

temperature profile. Similarly, the second order derivative dependent terms are extracted by controlling the temperatures of two opposites' edges by connecting them to the same ideal temperature source while the remaining edges orthogonal to the top surface of the die are connected to a second temperature source (figure 4). Again, by increasing the temperature difference between the two temperature sources while maintain the average temperature constant, the dependence over $\frac{\partial^2 T}{\partial x \partial y}$ is extracted and modeled as a second order perturbation.

b) Heat distribution extraction:

The power distribution of the block has to be distributed between the four thermal terminals. Unfortunately, it cannot be done by enabling the self-heating in the models of the block components. Indeed, the heat produced by the devices would flow through the thermal network and would modify the thermal map imposed by the boundary temperature sources. However, since the heats produced by the components correspond to their electrical power consumptions, the needed information is get from the electrical parameters obtained in the previous simulations set: the electrical power is calculated for each component and the total power dissipated from the components located in each area is affected to the corresponding thermal terminal.

III. RESULTS

In order to compare the results given by an electro-thermal simulation to experimental data, we designed a 2.5×2.5mm chip made of a 4 × 4 temperature sensors matrix [13]. These temperature sensors are PTAT (Proportional To Absolute Temperature) functional sub-circuits made of 39 elementary components (MOSFET, bipolar transistors, resistors and capacitor). This sub-circuit possesses one electrical terminal, its output, which provides a voltage proportional to its temperature. The more, eight resistors (two by side) are

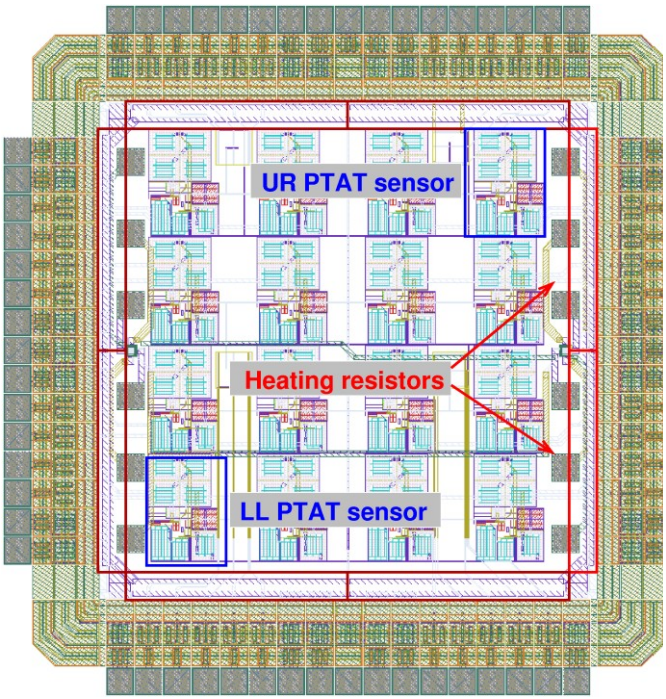


Figure 5 – Layout of the experimental chip

placed around the sensor matrix and are used to warm up the chip and create thermal gradients (figure 5).

We first simulated the behavior of the full chip at transistor level for several voltage steps (0, 2.95, 3.5 and 4.5 volts) applied on the two right resistors. At low level, the number of nodes of the full electro-thermal network was about 72,500 and the simulation took 115 minutes to perform a 1,600 seconds transient simulation in which a 615 seconds voltage step is applied.

In order to increase the simulation speed, we developed a four thermal terminals high-level model of our elementary PTAT sensor. Since many devices in this temperature sensing circuit must be matched, we expect to find a strong dependence of the output voltage on T_{avg}, $\frac{\partial T}{\partial x}$, $\frac{\partial T}{\partial y}$ and $\frac{\partial^2 T}{\partial x \partial y}$.

The output voltage of the PTAT sensors has thus been modeled by the following equation:

$$V_{out} = A(T_{avg}) + B\left(\frac{\partial T}{\partial x}\right) + C\left(\frac{\partial T}{\partial y}\right) + D\left(\frac{\partial^2 T}{\partial x \partial y}\right) \quad (8)$$

From the low-level simulation of a single PTAT sensor, the functions A, B, C and D have been extracted accordingly to the procedure described previously. Since the thermal sensor power density is very low and almost homogeneous, the heat flow passing through each thermal terminal has been evenly split. The more, in the particular case of our test chip, the thermal distribution essentially depends of the heat produced by the heating resistors. The electro-thermal network obtained in this configuration was made of roughly 12,000 nodes and it took 5 minutes and 45 seconds to make the same transient

Voltage [V]	Relative error [%] 1T	Relative error [%] 4T
0	0	0
2.95	-2.46	0.6
3.5	-2.35	0.8
4.5	-2.56	0.7

a – Upper right sensor

Voltage [V]	Relative error [%] 1T	Relative error [%] 4T
0	0	0
2.95	-0.7	-0.1
3.5	-0.9	-0.1
4.5	-1.4	-0.1

b – Lower left sensor

Figure 6 – Relative error between low-level and high-level simulations

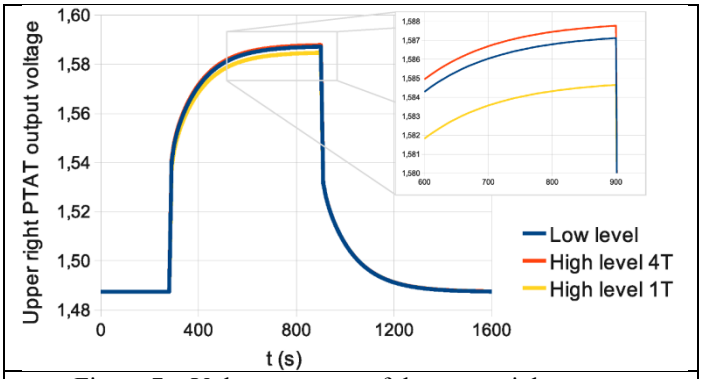


Figure 7 – Voltages output of the upper right sensor

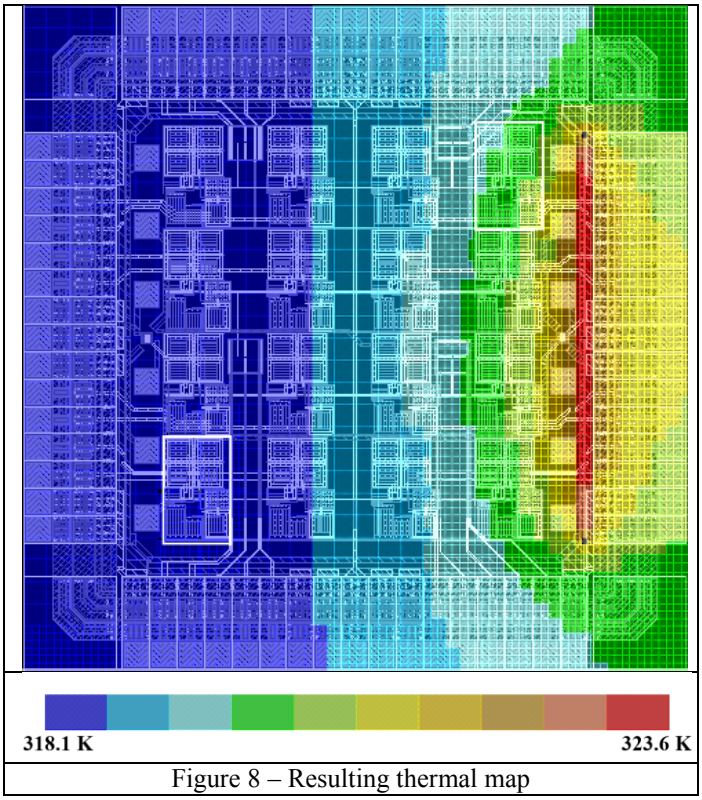


Figure 8 – Resulting thermal map

ISBN:978-2-35500-022-5

simulation. The simulation speed has thus been increased by a factor 20 while the accuracy loss is always lower than 1% (figure 6.a).

To demonstrate the benefits provided by using a multi-thermal terminals model, we also ran the same simulations with our high-level model in which the functions B, C and D have been removed. The output voltage equation in this last case is the following and the high-level model behaves as if it has only one thermal terminal:

$$V_{out} = A(T_{avg}) \quad (9)$$

Here, the number of nodes of the electro-thermal network remained the same and the simulation time was roughly identical. However, as shown in figure 7, the accuracy of these last simulations was strongly reduced, especially for the upper-right sensor which is placed in a region where strong thermal gradients take place (figure 8). These results are summarized in tables 6.a and 6.b.

IV. CONCLUSION

A high-level electro-thermal modeling methodology is presented in this paper. This methodology consists in extracting the behavior of the sub-circuit as a function of its temperature distribution from low-level simulations. Depending on the size of the sub-circuit, its area is split into a reduced number of sub-areas. The dependence on the spatial derivatives of the temperature distribution is extracted from low-level simulations. Our approach consists in evaluating the thermal distribution from the knowledge of the temperatures sampled in each sub-area. A convenient way to do this is to express the electrical quantities as functions of the average temperature and its spatial derivatives over the covered area. Indeed, by using our electro-thermal simulation tool, it is easy to create linear thermal distributions at low level thanks to perfect temperature sources connected to the underlying thermal network. The efficiency of this methodology has been demonstrated. It allows reducing dramatically the number of thermal terminals which, in turns, allows strongly reducing the underlying thermal network leading to faster while accurate simulation at chip level.

ACKNOWLEDGEMENT

This work is supported by French National Agency (ANR) through Nanoscience and Nanotechnology Program (Project 3D-IDEAS).

REFERENCES

[1] Sassone, A.; Wei Liu; Calimera, A.; Macii, A.; Macii, E.; Poncino, M.; , "Modeling of thermally induced skew variations in clock distribution network," Thermal Investigations of ICs and Systems (THERMINIC), 2011 17th International Workshop on , vol., no., pp.1-6, 27-29 Sept. 2011

[2] A. Ajami, K. Banerjee, and M. Pedram, "Modeling and analysis of nonuniform substrate temperature effects on global ULSI interconnects,"IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 24, no. 6, pp. 849-861, June 2005.

[3] S. Bota, J. Rossello, C. de Benito, A. Keshavarzi, and J. Segura, "Impact of thermal gradients on clock skew and testing," IEEE Design Test of Computers, vol. 23, no. 5, pp. 414-424, may 2006.

[4] Pedram, M.; Nazarian, S.; , "Thermal Modeling, Analysis, and Management in VLSI Circuits: Principles and Methods," *Proceedings of the IEEE* , vol.94, no.8, pp.1487-1501, Aug. 2006 doi: 10.1109/JPROC.2006.879797

[5] Lee, H.-H.S.; Chakrabarty, K.; , "Test Challenges for 3D Integrated Circuits," *Design & Test of Computers, IEEE* , vol.26, no.5, pp.26-35, Sept.-Oct. 2009 doi: 10.1109/MDT.2009.125

[6] Jeng-Liang Tsai; Chen, C.C.-P.; Guoqiang Chen; Goplen, B.; Haifeng Qian; Yong Zhan; Sung-Mo Kang; Wong, M.D.F.; Sapatnekar, S.S.; , "Temperature-Aware Placement for SOCs," *Proceedings of the IEEE* , vol.94, no.8, pp.1502-1518, Aug. 2006 doi: 10.1109/JPROC.2006.879804

[7] Van der Plas, G.; Limaye, P.; Loi, I.; Mercha, A.; Oprins, H.; Torregiani, C.; Thijs, S.; Linten, D.; Stucchi, M.; Katti, G.; Velenis, D.; Cherman, V.; Vandevelde, B.; Simons, V.; De Wolf, I.; Labie, R.; Perry, D.; Bronckers, S.; Minas, N.; Cupac, M.; Ruythooren, W.; Van Olmen, J.; Phommahaxay, A.; de Potter de ten Broeck, M.; Opdebeeck, A.; Rakowski, M.; De Wachter, B.; Dehan, M.; Nelis, M.; Agarwal, R.; Pullini, A.; Angiolini, F.; Benini, L.; Dehaene, W.; Travaly, Y.; Beyne, E.; Marchal, P.; , "Design Issues and Considerations for Low-Cost 3-D TSV IC Technology," *Solid-State Circuits, IEEE Journal of* , vol.46, no.1, pp.293-307, Jan. 2011 doi: 10.1109/JSSC.2010.2074070

[8] Franzon, P.D.; Davis, W.R.; Steer, M.B.; Lipa, S.; Eun Chu Oh; Thorolfsson, T.; Doxsee, T.; Berkeley, S.; Shani, B.; Obermiller, K.; , "Design and CAD for 3D integrated circuits," *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE* , vol., no., pp.668-673, 8-13 June 2008

[9] Jain, A.; Alam, S.M.; Pozder, S.; Jones, R.E.; , "Thermal-electrical co-optimisation of floorplanning of three-dimensional integrated circuits under manufacturing and physical design constraints," *Computers & Digital Techniques, IET* , vol.5, no.3, pp.169-178, May 2011

[10] J. Liu, S. Dong, Y. Ma, D. Long, and X. Hong, *"Thermal-driven symmetry constraint for analog layout with CBL representation,"* in Proc. IEEE/ACM Asia South Pacific Design Autom. Conf., Jan. 2007, pp. 191–196.

[11] Lin, M.P.-H.; Hongbo Zhang; Wong, M.D.F.; Yao-Wen Chang; , *"Thermal-Driven Analog Placement Considering Device Matching,"* Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on , vol.30, no.3, pp.325-336, March 2011 doi: 10.1109/TCAD.2010.2097308

[12] Krencker, J.-C.; Kammerer, J.-B.; Hervé, Y.; Hébrard, L.; , "Direct electro-thermal simulation of integrated circuits using standard CAD tools," *Thermal Investigations of ICs and Systems (THERMINIC), 2010 16th International Workshop on* , vol., no., pp.1-4, 6-8 Oct. 2010

[13] Krencker, J.-C.; Kammerer, J.-B.; Herve, Y.; Hebrard, L.; , "3D electro-thermal simulations of analog ICs carried out with standard CAD tools and Verilog-A," Thermal Investigations of ICs and Systems (THERMINIC), 2011 17th International Workshop on , vol., no., pp.1-4, 27-29 Sept. 2011

[14] B. Antao, "AHD Languages – A Must for Time-Critical Designs," IEEE Circuits and Devices Magazine, pp.12-17, 1996

[15] G. Gielen, and R. Rutenbar, "Computer-Aided Design of Analog and Mixed-Signal Integrated Circuits," Proc. IEEE, vol.88, pp.1824-1854, Dec. 2000

25-27 September 2012, Budapest, Hungary

Development and Performance of LED Metal Package with an Integrated Reflector/Heat Slug Structure

Jin Hwan Kim[1], Jong Hwa Choi[2], So Hyeon Mun[2], Moo Whan Shin[1*]
[1] *School of Integrated Technology and YICT College of Engineering, Yonsei University*
162-1 Songdo-dong, Yeonsu-gu, Incheon, 406-840, Korea.
[2] *Department of Material Science and Engineering, Myong Ji University,*
38-2 Nam-Dong, Yongin, Kyunggi, 449-728, Korea.
[*] *mwshin@yonsei.ac.kr*

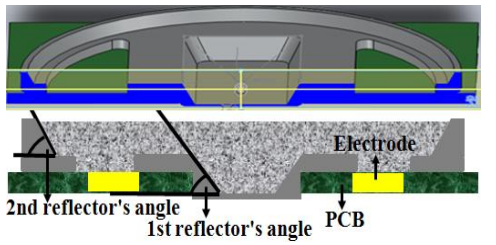
***Abstract*- This paper reports on the development of new design of Light Emitting Diode (LED) metal package and its thermal and optical performance. The new package exhibits a reflector integrated with a heat slug as one body. Thermal and optical performance of the package was evaluated and compared with that of commercial plastic LED package for reference. The new metal package showed excellent thermal performance compared with the plastic package ($R_{th,j-a}$ of 27 K/W vs. 49 K/W and T_j of 57 ℃ vs. 86 ℃ at a driving current of 350 mA). Unexpected from the much better thermal performance, optical power from the metal package was not noticeably higher than that from the plastic package when measured by an integration sphere. Comparison of optical simulation results from 50 cm radius far field and near field receiver indicated a significant optical loss at the reflector surface of the metal package. It was demonstrated that the external optical power of the metal package be significantly improved by the change of reflector design. The external optical power was improved from 145.29 mW to 170.15mW by optimizing the two reflectors' angles (from 80 °and 80 ° to 55 ° and 50 °, for the 1st and 2nd reflector, respectively).**

I. INTRODUCTION

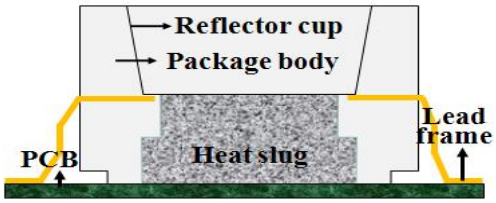
High power Light Emitting Diodes(LEDs) have been widely used in many applications due to their advantages over conventional lighting sources such as incandescent lamp or fluorescent lamp. The advantages include low power consumption, long life time, and environmental friendliness. They opened up many new applications such as backlighting for displays, automotive lighting, and general indoor/outdoor lighting [1]. As the demand for light output and the driving power increase, thermal management of LED packages became one of critical issues for successful application of LEDs for high power operation. High temperature at the junction of LED leads to degradation of optical performance and is directly related with a life time of devices. Therefore, effective thermal removal from the junction is the key to meet the high flux requirements in many applications of LED Package [2]. Various thermal designs with different materials and structures of LED packages have been developed and commercialized [3-5]. In general, thermal resistance is reported to be very high for LED packages with plastic bodies such as Polyphthalamide (PPA) or FR4 mainly due to the low thermal conductivities of the materials (typically less than 1 Wm-1K-1) [6]. Of particular interests are the ceramic materials as LED packages because of their various advantages over plastic materials including high thermal conductivity. Thermal analysis on the ceramic packages has been reported and better thermal characteristics were demonstrated compared with plastic package [7]. However, higher thermo mechanical stress was calculated from the ceramic package due to its lower elastic modulus. Discontinuity of packaging materials between the heat slug and package body (reflector) results in a barrier to efficient thermal dissipation, which is a common case for plastic and ceramic packages. The mismatching of coefficient of thermal expansion (CTE) between two attached packaging materials leads to delamination in many cases when the LED package experiences thermal stress from self heating or fabrication processing. The delamination is known as one of major factors which degrade the performance and life time of LEDs [8]. In this paper, we propose a new LED metal package which has an all-in-one structure of heat slug and reflector. Thermal and optical performance of the package has been evaluated and compared with other commercial plastic package. Optical reflection from the new package was carefully investigated using an optical simulator for far filed and near field receiver. The structure of metal package was optimized as a function of reflector angle for the improved optical performance.

II. EXPERIMENTAL METHOD

The new metal package exhibits one body frame of Cu reflector cup and heat slug as is shown in Fig. 1(a). Nano-sized Cu powder was mixed with organic binder and squeezed into a mold for shaping as a metal package followed by a sintering and polishing process. The surface of the package was electroplated with 2㎛ thickness of Ni and 4㎛ thickness of Ag. The package has inside (1st) and outside (2nd) reflector and the reflector angle is defined as an angle between the surface of heat slug and the reflector's surface as is depicted in Fig. 1(a). Two holes were formed by a punching process for the connection of electrodes on the upper side of PCB to the wires from LED chip. One GaN-based LED chip



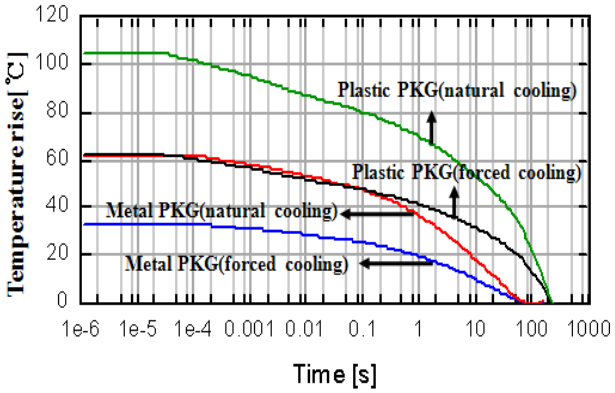
(a)



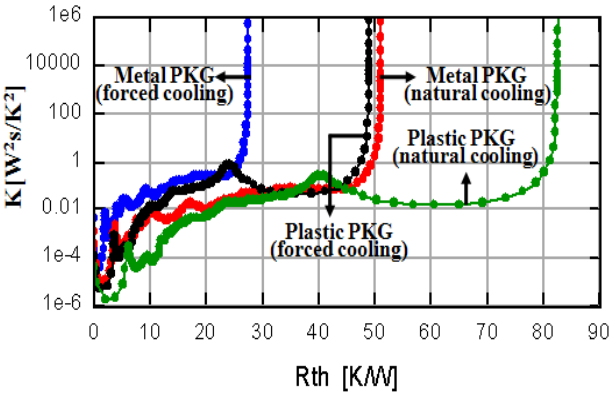
(b)

Fig. 1 Schematics of high power LED packages; (a) metal package and (b) plastic package.

(0.8mm x 0.8mm x 0.09mm, C460XB900 by CREE) was placed in the center of the cup. For a comparative thermal evaluation of the new metal package, the same chip was also loaded into a typical plastic package where the reflector is made of PPA (Fig. 1(b)).



(a)



(b)

Fig. 2 Measured (a) cooling curves and (b) structure functions from the metal and plastic packages.

Thermal analysis was performed by utilizing a Thermal Transient Tester (T3Ster, MicReD Ltd.). Based on the thermal R-C network and structure function theory, the heat path can be determined quantitatively and detailed process of thermal analysis is described elsewhere [9, 10]. The junction temperature and thermal resistance of the LED packages were determined under the driving current of 350 mA at an ambient temperature of 25℃ for both natural cooling and forced cooling conditions. Optical characteristics of the samples were measured by using 50 cm radius integrating sphere (OPI-1000, WITHLIGHT Co., Ltd) with the temperature of the package bottom being fixed at 25℃. To evaluate the detailed optical performance of high power LED packages, a numerical optical simulator, LightTools (Optical Research Associates, America) was utilized.

The simulator is based on Monte Carlo method and makes use of the generation of photons with random direction according to a distribution function describing the nature of the light emission. In this analysis, the light source was considered as being isotropic from a single point situated in the center of the device [11]. The light path from a light source was evaluated using optical ray tracing to calculate the final light distribution. Optical rays started from random locations and directions of the source, traced through the optical system, and collected on receivers specified by distance and area.

III. Results and Discussions

The metal package showed better thermal performance than the conventional plastic package under the both natural and forced cooling conditions as is shown in Fig. 2(a). The junction temperatures were around 57℃ and 86℃ for metal and plastic package, respectively under the forced cooling condition and 85.5℃, 127℃ under the natural cooling condition. The lower junction temperatures from the metal packages led to lower thermal resistances as is expected from the relationship;

$$R_{th} = \frac{T_j - T_a}{P} = \frac{\Delta T}{P} \quad (1)$$

where R_{th} is the thermal resistance, T_j the junction temperature, T_a the ambient temperature, P power, and $\triangle T$ the junction temperature rise. Under the forced cooling condition, the measured thermal resistances are 27 K/W and 49 K/W for metal and plastic package, respectively (Fig. 2(b)). The values are in agreement with calculated ones using the Eqn. (1).

Structure functions in Fig. 2(b) clearly represent the excellent thermal dissipation of metal package in the thermal paths from the chip to die attach, die attach to heat slug, and heat slug to the air. Optical characteristics of the metal LED packages have been measured in an integrating sphere with one meter of diameter. The optical power per unit area in the detector of integrating sphere is given by the Eqn. (2) [12];

$$\hat{P} = \Phi_{in} \frac{\rho}{1-\rho(1-f)} \frac{(1-\rho_D)}{4\pi R^2} \quad (2)$$

where Φ_{in} is the luminous flux, R the radius of the integrating sphere, ρ the reflexibility of inside wall of integrating sphere, f the area ratio for total area integrating sphere, ρ_D the surface reflexibility of detector. The measured optical power of metal package was 161 mW. Unexpectedly, the optical performance from the metal package is only slightly higher than that from the plastic package which exhibited optical power of 152.4 mW. Only considering the thermal performance, metal LED package is better than plastic LED package. However, higher optical efficiency and uniform color distributions are also or even more important key issues for real application of the metal LED package compared with the thermal performance. Therefore, we investigated the extraction of light and its interaction with the surface of packages by using optical simulator, LightTools. First, the optical power was simulated for both metal and plastic packages setting a distance of far field receiver as 50 cm which is consistent with the size of the integrating sphere used in the measurement. The simulated optical powers were 145.29 mW and 138.72 mW for the metal and plastic package, respectively. They are in good agreement with the measured ones, but the simulated values are slightly lower than the measured ones. The reason for the difference is still under investigation. However, it was confirmed that the significant advantage of thermal performance from the metal package over the plastic package did not lead to excellence in optical performance, which is consistent with the measured data. Thus, optical simulation was performed for the near field receiver with a distance of 8.5 mm from the LED chip so that the light flux can be determined before any possible light loss by interaction with the package surface. Fig. 3 compares the light flux at the near field receiver for the metal package and plastic package. It is clear that more light flux is generated from the chip loaded into the metal package rather than the plastic package. The near field optical powers from the metal and plastic packages are 199.42 mw and 144.72 mw, respectively.

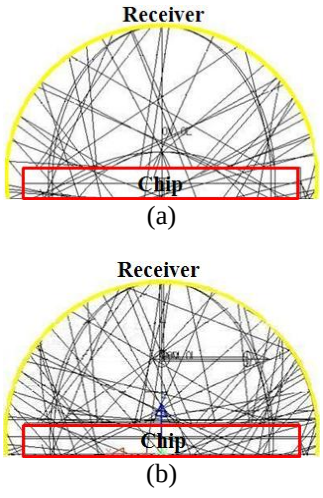


Fig. 3. Ray tracing for near field receiver; (a) plastic package, (b) metal package.

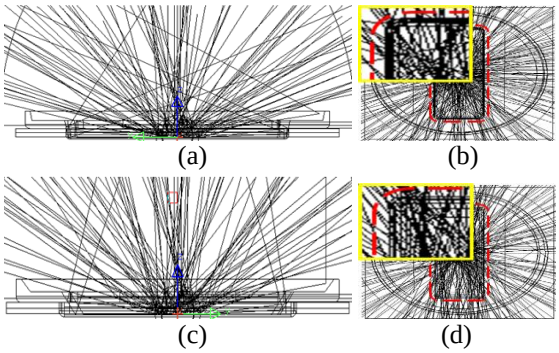


Fig. 4. Simulated ray tracing in the reflector cup of different 1st and 2nd reflector angle; (a) side view for 80°-80°, (b) top view for 80°-80° (original structure), (c) side view for 55°-50°, (d) top view for 55°-50° (optimized structure).

The result is a clear demonstration that the new metal package has an advantage in light emission over the plastic package due to the higher reflectivity of metal. Comparison between the data of the near field and far field receivers leads to a conclusion that the metal package has higher light extraction in near the chip, but the emitted light experiences more interaction with the package surface before it reaches the far field receiver (surface of integrating sphere as well in measurement). It means that the reflecting structure is not efficient so an optimization is required.

In order to investigate the optical performance variation with reflector angles, we simulated the optical performance of sixty two metal packages with different pair of 1st and the 2nd reflector angle in a range 40° to 85°.

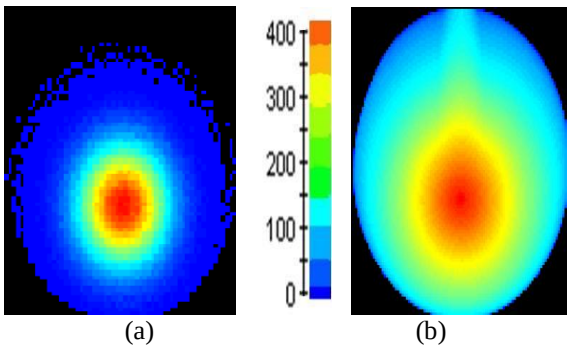


Fig. 5. Distribution of luminous intensity of the metal LED package with (a) original structure and (b) optimized structure.

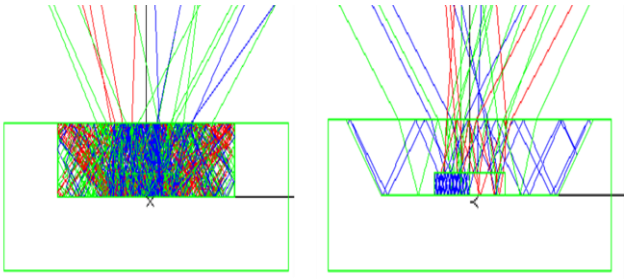


Fig. 6. Light propagation in the package of different reflector angles (a) 90 degree (b) 50 degree.

Actually, it was found that the reflector angles in metal package (Fig. 1(a)) make significant effects on the light interaction after emission from the chip. Both the 1st and the 2nd reflector angle in the fabricated metal packages is 80°.

Fig. 4 presents the simulated ray tracings for the fabricated metal LED package structure and for optimized structure. Compared with the original structure (145.29mW), the optimized structure exhibits much higher optical power of 170.15mW. The inlets in Fig. 4(b) and (d) clearly demonstrate more interaction of rays with package for the original structure. The distribution of luminous intensity of the metal LED package with original structure and optimized structure was shown in Fig. 5. Comparison the original structure (Fig. 5(a)) with the optimized structure (Fig. 5(b)) shows a great improvement of the luminous intensity and the optical extraction efficiency increased from 77.1% to 89.7%. Optical extraction efficiency variation with various reflector angles was shown above. In order to make the explanation straightforward, a simplified ray trace results with limited rays were plotted in Fig. 6. At a high reflector angle, the emitted photons impinge the side surface of the reflector several times before they escape from the reflector cup (Fig. 6(a)). These photons have high propagation angle relative to the vertical axis of the package. This means that the photons are difficult to be extracted from reflector cup. As the reflector angle decreases to a certain lever, photon tracing time becomes shorter (Fig. 6(b)) and photons escape from the reflector cup at a smaller incident angle. Therefore, the photon absorption by the package materials such as reflector and chip is reduced, and more photons are extracted out of the package. It is worth noting that there is a turning point for the extraction efficiency if we continue to decrease the reflector angle. In this paper, we did not give the results after the turning point. The variation of reflector angle is a good choice to improve the optical efficiency.

It was confirmed that the optical power from the developed metal package was significantly diminished by the non-optimized reflector angles, otherwise it could excel the plastic package in optical performance as well as thermal performance.

IV. Conclusion

We proposed a new LED metal package with one body structure of a heat slug and a reflector and investigated its thermal and optical characteristics. The metal package showed a good thermal performance inherently stemmed from the one body structure with a good thermal conduction. By comparative analysis in far field and near field simulation, the optical performance of the metal package was found to be significantly suppressed by the non-optimized reflector angles inside the package. Improved optical performance was demonstrated by the optimization of the angles of the 1st and the 2nd reflectors.

ACKNOWLEDGMENT

This work was supported by the Components & Materials Technology Development program of Ministry of Knowledge Economy. [High Color Rendering LED-BLU Panel For Digital Display]

Support for M. W. Shin is from the MKE (The Ministry of Knowledge Economy), Korea, under the “IT Consilience Creative Program” supervised by the NIPA (National IT Industry Promotion Agency)"
(NIPA-2010-C1515-1001-0001)

References

[1] W.K Jeung, W.K Jeung, S.H Shin, S.Y Hong, S.M Choi, S.Yi, Y.B Yoon, H.J Kim, S.J Lee, K.Y Park, "Silicon-Based, Multi-Chip LED Package", Electronic Components and Technology Conference, 2007. ECTC '07. Proceedings. 57th, pp. 722-727, 2007

[2] H.H Kim, S.H Choi, S.H Shin, Y.K Lee, S.M Choi, S.Yi, "Thermal transient characteristics of die attach in high power LED PKG", Microelectronics Reliability, Volume 48, Issue 3, pp. 445-454, Mar. 2008.

[3] G. Rong-feng, T. Da-lei, W. Xing, Z. Wen-qing, “Thermal, Mechanical and Optical Analysis of SiC-based LED”, Electronics Packaging Technology Conference, 2008. EPTC 2008. 10th, pages 939-944, 2008

[4] Wei-Hao Chi, Tsung-Lin Chou, Cheng-Nan Han, Kuo-Ning Chiang, “Analysis of Thermal Performance of High Power Light Emitting Diodes Package”, Electronics Packaging Technology Conference, 2008. EPTC 2008. 10th, pp. 533-538, 2008.

[5] Mingsheng Ma, Zhifu Liu, Yongxiang Li, "Thermal performance of high power LED package based on LTCC", Electronic Packaging Technology and High Density Packaging(ICEPT-HDP), 2011 12th International Conference, pp. 1-4, 2011

[6] L. Kim, M.W. Shin, “Thermal Resistance Measurement of LED Package with Multichips”, Components and Packaging Technologies, Volume 30 , Issue 4, pp. 632-636, 2007

[7] L. Kim, M.W. Shin, “Thermal Resistance Measurement of LED Package with Multichips”, Components and Packaging Technologies, Volume 30 , Issue 4, pp. 632-636, 2007

[8] J. Hu, L. Yang, M.W. Shin, "Mechanism and thermal effect of delamination in light-emitting diode packages", Microelectronics Journal, Volume 38, Issue 2, pp. 157-163, Feb. 2007

[9] W.J. Hwang, P. Szabo’, G. Farkas, M.W. Shin, Proceedings of the 10th International Workshops on Thermal Investigations of ICs and Systems(THERMINIC), pp. 91–95, Sep. 2004

[10] V.Sze´kely, T.V. Bien, “Fine structure of heat flow path in semiconductor devices: a measurement and identification method”, Solid-State Electron. 31, pp. 1363–1368, 1988

[11] Aldo Badano and Jerzy Kanicki, “Monte Carlo analysis of the spectral photon emission and extraction efficiency of organic light-emitting devices”, Journal of Applied Physics, Volume 90, Issue 4, pp. 1827-1830, 2001

[12] M. L. Rastello, E. Miraldi, P. Pisonid, “Luminous-flux measurements by an absolute integrating sphere”, Applied Optics, Vol. 35, Issue 22, pp. 4385-4391, 1996

25-27 September 2012, Budapest, Hungary

A disruptive technology for thermal to electrical energy conversion

O. Puscasu[1,2,3], S. Monfray[1], C. Maître[1], JP. Pemeant[1], F. Boeuf[1], G. Savelli[3], F. Gaillard[3], D. Guyomar[2], T. Skotnicki[1]

[1]*STMicroelectronics (Crolles 2) SAS, 850 rue Jean Monnet, 38926, Crolles Cedex, France*

[2] *LGEF, INSA Lyon, 8, Rue de la Physique 69621 Villeurbanne Cedex, France*

[3]*CEA Liten,17 rue des martyrs 38054 Grenoble Cedex 9, France*

Abstract

A disruptive approach to thermal energy harvesting is presented. The new technique can be used for powering ultra-low power electronics and autonomous systems. We propose a two step conversion of heat into electricity: thermal to mechanical and mechanical to electrical. Devices can work in a wide range of temperatures: from -40°C to 300°C, and the available mechanical power density is in the order of 1 mW/cm^2. We evidenced that one of the keys to improve the generated power density is downscaling of individual devices. To demonstrate this point, laws modeling downscaling have been established and show that the miniaturization of the devices by a factor k increases the generated power density by the same factor, thanks to the higher speed of heat transfer.

I.Introduction

Heat is one of the most abundant energy sources that can be converted into electricity in order to power circuits. Significant efforts have been made in order to harvest heat through the development of thermoelectric generators (TEGs), based on the Seebeck effect [1]. They require rare materials based on bismuth telluride (Tellurium has the same abundance as Platinum, and is one of the most rare solid elements on Earth: 1µg/kg), or complex nanostructures in order to work at ambient temperatures. These materials show low thermal insulating properties, which makes keeping temperature gradients difficult and the use of a heat sink necessary (Fig. 1).

An innovative way of harvesting, that allows to avoid the above mentioned difficulties is presented in this work. It enables the fabrication on thin modules that work without a heat sink at temperatures close to ambient. The key point of the integration of this technology is the ability to keep an important gradient on a device body by intelligent control over the thermal flow (Fig. 2).

II. Working principle and first results

An elementary device of our technology is based on a thermal bimetal that snaps when it is being heated and then cooled down (Fig. 3). Thus if such a bimetal is put between a hot and a cold surface, it oscillates and generates mechanical power. By actuating a piezoelectric during snap action, electric pulses are generated (Fig. 4). When the piezoelectric is connected to a harvesting circuit, electrical energy is stored.

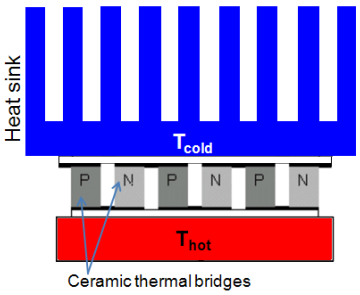


Figure 1. Seebeck module made of ceramic materials. A heat sink is needed.

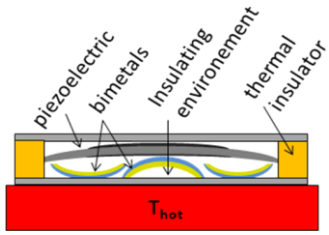


Figure 2. Module with thermal conductivity controlled by bimetals: no heat sink necessary

A thermal bimetal is a double layer consisting of a material of high coefficient of thermal expansion (>$10\cdot10^{-6}$ K^{-1}) and a material of low CTE (≈$1\cdot10^{-6}$ K^{-1}). It is necessary to give the layer a concave shape in order to make it snap, that is to move quickly from one stable position to another under the effect of rising or decreasing temperature [2].

By using a bimetal that snaps up at 122 °C and down at 117 °C, with an area of 7 cm^2, and a 0.3 mm thickness, an electrical power of 12 µW has been obtained on a hot plate at 130 °C. The bimetal oscillates with a frequency of 0.5 Hz and develops a kinetic energy of 16 mJ per snap, that is 1,1 mW/cm^2 of mechanical power available.

By using a bimetal that snaps up at 50 °C and down at 40 °C, a force as high as 3.6 N is generated (Fig. 5).

To retrieve a maximum of mechanical power, the bimetal must have a snap frequency as high as possible. To achieve this: 1) the dimensions of the bimetal should be as low as possible to retrieve the temperature of the heat source (or the

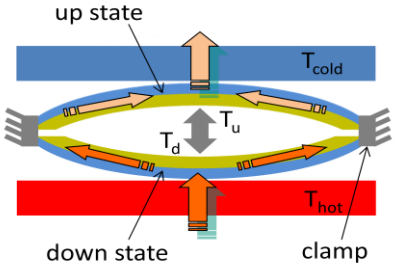


Figure 3. Up and down state of a bimetal. T_u and T_d are the state change temperatures.

ISBN:978-2-35500-022-5

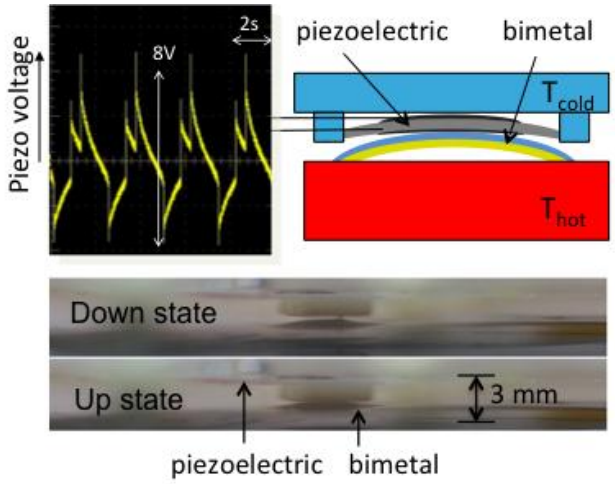


Figure 4. Example of measured signal for a piezoelectric actuated by a bimetal, T_u= 122°C. Signals are generated during snap action, which is the transition between the up and down state of the bimetal. 8V peak to peak amplitude can be easily reached.

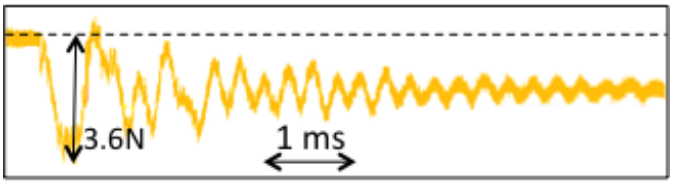


Figure 5. Example of force signal developed by a bimetal working between 50 and 40 °C. An amplitude of 3.6 N is reached for a bimetal thickness as low as 0.2 mm during the first instants of the impact on the force gauge.

cold sink) as quickly as possible, while maintaining sufficient deflection in order to hit the mechanical-to-electric transducer, 2) the two temperatures of the bimetallic transition (up and down) have to be the closest, 3) the number of bimetals must be high (matrix) and their miniaturization must be technically possible, as shown in the next section.

III. Thermo mechanical conversion by a bimetal. Scaling laws.

The heat flow equations governing our devices are established in this section. They lead to scaling laws that can enable power gain.

The heat flow from a hot source to a heated body can be modeled using the heat equation:

$$\frac{\partial^2 T}{\partial x^2} = \frac{1}{\alpha_T}\frac{\partial T}{\partial t} \quad (1)$$

with α_T – thermal diffusivity, and x – the considered thermal propagation direction.

By supposing a bimetal at an initial temperature T_{cold} in contact with a heat source at constant temperature T_{hot}, as shown in Fig. 3, the time dependence of the average temperature during the heating process can be expressed by:

$$\bar{T}(t) = \frac{8}{\pi^2}\exp\left(-\pi^2\alpha_T\frac{t}{L^2}\right)(T_{cold} - T_{hot}) + T_{hot} \quad (2)$$

with α_T – average thermal diffusivity of the bimetal materials. The constant α_T can be expressed as the ratio between the

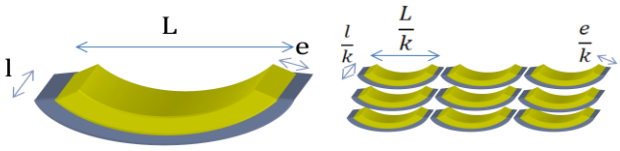


Figure 6. Replacement of a large bimetal by a matrix of small bimetals occupying the same surface

thermal conductivity and the product of the density to the thermal capacitance: $\alpha_T = k/(\rho c_p)$. L is the length of the bimetal.

It can be observed from (1) that the time needed to heat up a bimetal to a target temperature is proportional to its squared length:

$$t \sim L^2 \quad (3)$$

By supposing that the target to reach is the snap temperature, it can be understood that the smaller the bimetal is, the faster the snap action will take place. The time needed to have it will be k^2 times smaller when the size is divided by a factor k. The same law will govern the cooling process. The snap frequency of a bimetal oscillating between a hot and a cold source is inversely proportional to the sum of the heating and cooling time. This implies that by diminishing the size of a bimetal by a factor k, the oscillation frequency will be increased by k^2:

$$f \sim \frac{1}{L^2} \sim k^2 \quad (4)$$

Analytical expressions show that snap temperatures should not change with scale [2].

The thermal stresses created during the heating step will lead to the accumulation of elastic energy inside the bimetal. The snap action will allow its liberation by conversion into kinetic energy. Calculations using the stress distribution in [3] show that the evolution of the elastic energy with temperature can be written as:

$$E_{elast} = \frac{Y}{32}(\alpha_2 - \alpha_1)^2 \Delta T^2 V_b \quad (5)$$

with Y – the average Young modulus of the materials, α_2-coefficient of thermal expansion of the upper layer, α_1 – coefficient of thermal expansion of the down layer, ΔT – difference between snap temperature and the ambient, V_b – the bimetal volume.

This energy will be partially converted into kinetic energy during snap action:

$$E_k = \frac{\rho V_b \cdot v^2}{2} \quad (6)$$

with ρ - density of the materials, and v – maximal bimetal speed.

The kinetic energy will be transmitted to the piezoelectric and make it vibrate. The bimetal speed resulting from the former expression is independent of scale:

$$v \sim \sqrt{\frac{Y}{\rho}} \cdot (\alpha_2 - \alpha_1)\Delta T \quad (7)$$

By replacing one large bimetal with several small ones occupying the same surface, with sizes divided by k (Fig. 6), the total mass will be decreased by k, since the thickness will

ISBN:978-2-35500-022-5

be smaller. Meanwhile the frequency of each element will be multiplied by k^2. The result will be a k times higher transmitted mechanical power:

$$\frac{P_{tot\,s}}{P_l} = \frac{N_b E_{ks}}{E_{kl}} \frac{f_s}{f_l} = N_b \frac{V_s}{V_l} \frac{L_l^2}{L_s^2} = k = \sqrt{N_b} \quad (8)$$

with $P_{tot\,s}$ – the total mechanical power transmitted by small bimetals, P_l – the mechanical power transmitted by a large bimetal. N_b is the number of bimetals and V is the volume of a bimetal, the index s refers to small bimetals, and the index l to the large one. The main parameter of the calculation is k – the scaling factor.

IV. Electromechanical conversion by the piezoelectric

The force developed by the bimetal for a given ΔT scales down as k^2 [4], which leads to a deflection in the piezoelectric divided by k and an average stress that does not vary with scale, as for a piezoelectric cantilever:

$$< \sigma_L >= \frac{3}{4} \frac{sA}{L^2} \quad (9)$$

with s – thickness, A – deflection, L_p – length of the piezoelectric.

The resulting piezoelectric voltage will depend on the stress and the piezoelectric thickness [5].

$$U = \frac{d_{31}}{\varepsilon} \cdot \sigma_L \cdot s \sim \frac{1}{k} \quad (10)$$

with all the constants and variables referring to the piezoelectric: U – voltage, d_{31} – bending charge constant, ε – dielectric constant, σ_L – longitudinal stress.

The electric energy stored on the piezoelectric due to bending can be expressed using the plain capacitor formula:

$$E_{el} = \frac{C_p U^2}{2} (11);\ C_p = \varepsilon \frac{S}{s} \quad (12)$$

with E_{el} – electric energy on the piezoelectric, C_p – capacitance, S – surface of the piezoelectric.

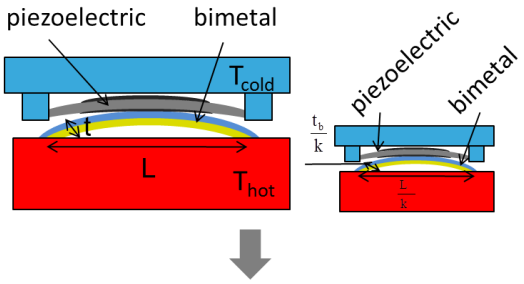
The previous equations allow predicting the change in generated electric power when a large bimetal associated to a piezoelectric is replaced by several small bimetals, each acting on a miniaturized piezoelectric.

By combining (3) and (9) to (12) one can deduce that the generated electric power for several miniaturized piezoelectric-bimetal components will be k times larger compared to a single component occupying the same surface:

$$P_{eltot} = N_s \cdot E_{el} \cdot f \sim k \quad (13)$$

Thus miniaturization of the devices can enable power gain (Fig. 7). Small scale harvesters can be built using LSI processes. The predicted gain can help in obtaining superior electric power compared to Seebeck modules.

One of the encountered challenges as known from literature and shown by (10) is the voltage decrease with the thickness of the piezoelectric. The voltage should be kept high enough in order to overcome the threshold of the harvesting circuit. Also, decreasing the scale will accelerate the heat transfer through the bimetals, so the degree of miniaturization is to be chosen depending on the available external temperature difference.



frequency (f)	↗	k^2
snap temperatures (T_u, T_d)	↗	same
bimetal force (F)	↘	k^2
piezoelectric and bimetal deflection (A)	↘	k
stress and strain inside piezoelectric		same
piezoelectric capacitance (C_p)	↘	k
piezoelectric voltage (U)	↘	k
transmitted thermal power/surface (P_t)	↗	k
mechanical power/surface (P_m)	↗	k
electrical power/surface (P_{eltot})	↗	k

Fig. 7 Evolution of the working parameters and the generated power with scale for our disruptive technology

V. Conclusion

A disruptive thermal energy harvesting technology has been presented. It is based on the conversion of heat into movement by thermal bimetals that act on piezoelectrics and thus generate electrical signals.

The devices we developed can work in a wide temperature range, including ambient conditions (-40 °C to 300 °C). The mechanical power developed by a bimetal that snaps up at 122 °C is of 1,1 mW/cm^2. By converting this amount into electric energy, high-end autonomous sensor applications could be addressed.

The generated electric power density in the above-mentioned case is of 1,7 µW/cm^2 and can be enhanced at macro scale by increasing the electromechanical efficiency through mechanical optimization of the structure.

We evidenced that power density can also be increased by miniaturizing the harvesters. The predicted gain is equal to the scaling factor for each device. The optimal working scale will depend on the needed electrical power and the amplitude of the generated signals, which should be higher than the threshold of the harvesting circuit. This is an approach that can enable a superior performance for our technology compared to Seebeck devices, without need for a heat sink.

REFERENCES:

[1] Hudak et al. Journal of Applied Physics 103, 101301 (2008)
[2] Wittrick. W. H., Stability of a bimetallic disk, The Quarterly Journal of Mechanics and Applied Mathematics (1953).
[3] S. Timoshenko, Analysis of bi-metal thermostats, J. Opt. Soc. Am. 11 (1925) 233-256
[4] Kanthal Thermostatic Bimetal Handbook, Kanthal AB.
[5] Tadigadapa et al. Meas. Sci. Technol. 20 (2009) 092001

25-27 September 2012, Budapest, Hungary

The thermal design of the thermal cutter of an antenna opening mechanism employed on a pico- satellite

Zoltán Gy. Kovács[1], Gábor E. Marosy[2], Gyula Horváth[3]
Department of Electron Devices, Budapest University of Technology and Economics
Magyar tudósok körútja 2, H-1117 Budapest, Hungary
Tel.: +36-1-463-2798[1], +36-1-463-2720 [2,3] Fax.: +36-1-463-2973
kovacsz@eet.bme.hu[1], marosy@eet.bme.hu[2], horvath@eet.bme.hu[3]

***Abstract* - In CubeSat-sized satellites the effective placement of the communication system's antenna is crucial both from weight and volume-wise point of view. Due to the size of the antenna, the only way it could be located inside the satellite was in a folded state. Thus a procedure had to be devised to open the antenna in a safe and reliable way using a thermal cutter in a wide temperature range. This paper summarizes the design considerations of the thermal cutter. It touches on the measurements and tests that were conducted in order to verify the design on the ground. The thermal cutter described in this article worked flawlessly on the Masat-1 CubeSat that was put into orbit in 2012. The results of the measurements conducted during the opening procedure in space are also described in the paper.**

I. INTRODUCTION

The communication unit with its antenna is a vital unit in every spacecraft. Radios on CubeSats usually operate in the 70 cm range, thus the antenna is 17 cm long [1][2]. This means that the antenna is larger than the satellite (which is 100x100x100mm), so a solution had to be found to fold and safely open the antenna in the extreme space environment. The price, reproducibility, consumption and reliability were important factors. The antenna was fastened by using a plastic fiber in its furled state. The fiber was cut with a thermal cutter when the antenna was unwound. In order to be able to calculate the parameters of the system, a thermal model of the thermal cutter was created [3] [4] [5]. Calculations were made difficult by the fact that the battery's terminal voltage depends on its current state and that the nominal operating range of the thermal cutter was between -40°C and 80 °C. The question to be answered was that how long it takes for the cutter to get into a state of severe degradation in which it is no longer operable. We have created a novel procedure to calculate such a system and verified using a climate chamber. The procedure was successfully applied in space in 2012.

II. BOUNDARY CONDITIONS OF THE DESIGN

The melting point of the applied plastic fiber was determined by measurements (T_{cut} = 130 °C), but in order to ensure a clean cut, the temperature gradient around the melting point had to be at least 20 °C/s.

$$\left.\frac{dT}{dt}\right|_{T_{cut},P_{min},T_{0min}} \geq 20\left[\frac{^{\circ}C}{s}\right] \qquad (1)$$

The solder materials used to fasten the component had a melting point of 260 °C (T_{melt}). These two temperature values acted as boundary condition in the design. The battery's nominal voltage was 4.2 V, but it was limited to 3.5-4.1 V to ensure a safe operation and long lifetime. Its current was limited to 2.1 A (I_{max}).

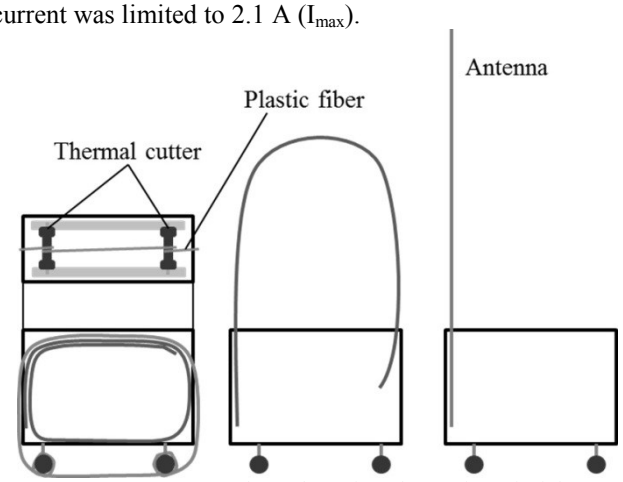


Fig. 0. – Antenna box schematic and operating principle

In order to increase the reliability of the system, two thermal cutters were placed on top of the antenna box. The two cutters operate in parallel thus providing redundancy for the opening mechanism. Due to mechanical considerations 0.6 W 1% THC resistors were chosen as cutters despite the fact that more power is needed to overheat them than in the case of lower power resistors. The operating range of the chosen resistor was -55 °C – 155 °C, thus the resistor had to be used outside its nominal operating range.

III. ELECTRONIC DESIGN PROCESS

According to the cutting temperature (T_{cut}) and the gradient criterion (1), the needed temperature can only be reached by overloading the resistor. The parameters of the thermal cutter were examined in a controlled overloaded state. The temperature of the resistor (T_{cur}) was measured using a thermal camera at given overload factors (OL), while the voltage (V_{cur}) and current (I_{cur}) of the resistor were recorded.

$$OL = \frac{P_{cur}}{P_{nom}} \qquad (2)$$

The resistance and current power of the resistor were determined using the values above (see Fig.1.).

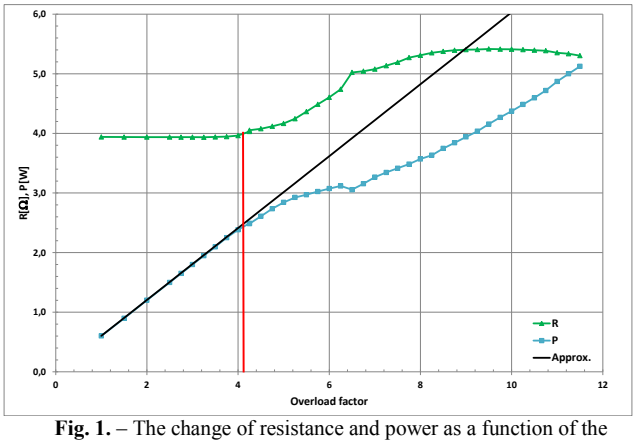


Fig. 1. – The change of resistance and power as a function of the overload

The measurement showed that the resistance is highly nonlinear at an overload factor of 4 and it starts to increase thus the resulting power dissipation is smaller than expected. Another diagram created during the aforementioned measurements shows that the resistor sustains its electrical parameters until 280 °C, above which the resistance increases due to thermal degradation.

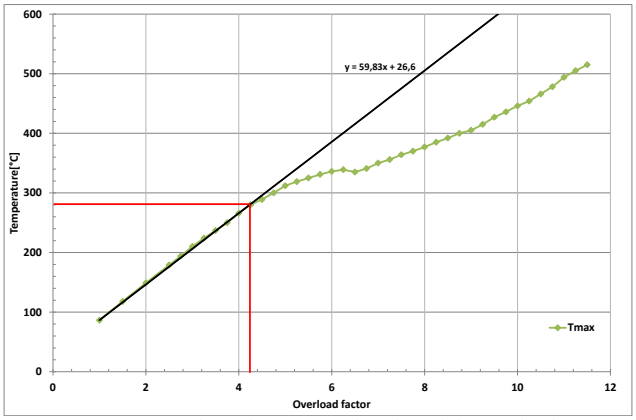


Fig. 2. – The resistance as a function of the overload

In summary it can be stated that the temperature of the resistor should never exceed 280 °C, otherwise irreversible degradation occurs. The electronic model of the system created with respect to the previously described boundary conditions can be seen in Fig. 3.

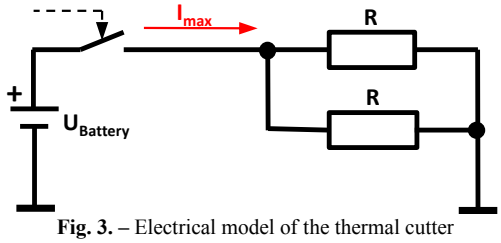


Fig. 3. – Electrical model of the thermal cutter

Elements of the E24 series meeting the electrical conditions were investigated. The thermal resistance of the resistors (R_{th_esti} = 50 °C / 0,6 W ≅ 83.3 °C/W) had to be determined in order to acquire the maximal overload rate. [6]

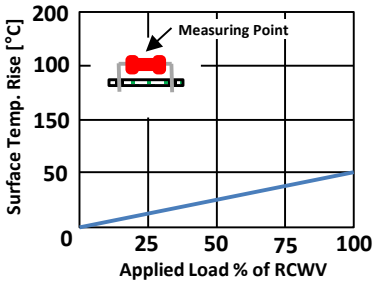


Fig. 4. – The temperature rise of the applied resistor as a function of the dissipated power

Using the estimated thermal resistance (R_{th_esti}) minimum and maximum criteria can be expressed for the nominal value of the resistor (R_{nom}).

$$\frac{R_{th\_esti}\cdot(U_{Battmax})^2}{T_{err}-T_{max}} \geq R_{nom} \geq \frac{R_{th\_esti}\cdot(U_{Battmin})^2}{T_{err}-T_{min}} \qquad (3)$$

According to the maximum current (I_{max}) criterion above, another limitation can be expressed for the minimum value of R_{nom} which is stronger than (3).

$$\frac{R_{th\_esti}\cdot(U_{Battmax})^2}{T_{err}-T_{max}} \geq R_{nom} \geq \frac{2\cdot U_{Battmax}}{I_{max}} \qquad (4)$$

The gradient criterion can be defined in parametric form (5) by using the time derivative of the function of time (8).

$$P_{min}\cdot\sum_{i=1}^{n}\frac{R_{thi}}{Tau_i}\cdot e^{-\frac{t}{Tau_i}} \geq 20\left[\frac{^{\circ}C}{s}\right] \qquad (5)$$

The following diagram shows the chosen resistance value and the corresponding estimated temperature minimum and maximum. The maximum and the minimum values of the diagram were counted by using the following two equations (6) (7).

$$T_{min}(R_{nom}) = T_{0min} + R_{th\_esti}\cdot\frac{(U_{Battmin})^2}{1.01\cdot R_{nom}} \qquad (6)$$

$$T_{max}(R_{nom}) = T_{0max} + R_{th\_esti}\cdot\frac{(U_{Battmax})^2}{0.99\cdot R_{nom}} \qquad (7)$$

The limiting factors, cutting temperature and the resistance corresponding to the maximal current are also shown. It can be clearly seen in the diagram (Fig.4.) above that the values between 3.9 Ω and 5.6 Ω satisfy the conditions.

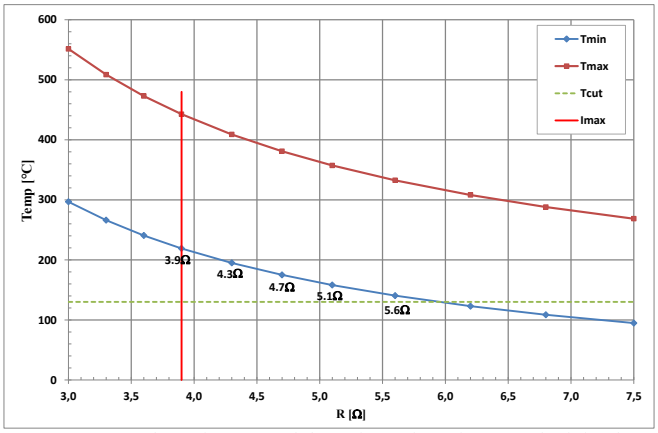


Fig. 4: - The resistance and the estimated maximum and minimal temperature showing the limiting factors

IV. THERMAL DESIGN

In the previous section the thermal parameters of the cutter were estimated using the parameters of data sheets. This method made it possible to give an estimation for the resistance value but further investigations were required to get a full picture of thermal behavior of the system. The measurements had to be performed in a still air chamber using a thermal camera in order not to alter the relatively small time constant of the system. The thermal transient was recorded with a resolution of 1 s and a curve was fitted on the measurement data in the time domain using (8).

$$T(t) = T_0 + P \cdot \sum_{i=1}^{n} R_{thi} \cdot (1 - e^{-\frac{t}{Tau_i}}) \qquad (8)$$

The fitting was performed for the values n = 2..4. Table 1 shows the calculated coefficients and their corresponding mean squared errors. It can be seen that the error doesn't decrease significantly for an order higher than 2. The first R_{th} element of the second order model is very close to the value given in the data sheet. This verifies the model and shows that the R_{th_esti} value used for the electrical calculations is in a good accordance with the real R_{th} value.

Table I.
The coefficients and mean square errors of the fitting

n	R_{th1}	Tau_1	R_{th2}	Tau_2	R_{th3}	Tau_3	R_{th4}	Tau_4	Mean Squared Error
2	*85,886*	*7,408*	*105,577*	*35,154*	--	--	--	--	**127,04**
3	*42,696*	*6,207*	*46,454*	*9,292*	*102,434*	*35,861*	--	--	**126,55**
4	*42,621*	*6,204*	*46,528*	*9,289*	*51,138*	*35,873*	*51,297*	*35,848*	**126,55**

In conclusion it can be stated that the system can be approximated to the needed extent using two time constant, so there is no need for a higher order model for the further calculations.

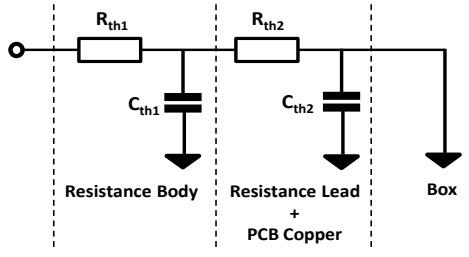


Fig. 5. - Thermal model of the system

Using the results above the following formula (9) can be deduced:

$$T(t, U_{Batt}, T_0) = T_0 + \frac{U_{Batt}^2}{R} \cdot \sum_{i=1}^{n} R_{thi} \cdot (1 - e^{-\frac{t}{Tau_i}}) \qquad (9)$$

In the expression above the U_{Batt} voltage was stepped by 0.1 V steps while the T_0 temperature was stepped by 0.1 °C steps in order to determine the moment (t_{err}) when the system reaches the T_{err} temperature.

$$T(t_{err}, U_{Batt}, T_0) = T_{err} \qquad (10)$$

The aforementioned experiments were conducted in a still air chamber. With respect to the fact that the application is to operate in vacuum, the previously calculated temperature value (280 °C) was reduced by 30 °C for safety reasons and the t_{err} moments were determined using this figure. The time step was small enough to allow an accurate calculation of T_{err} with a maximum error of 0.1 °C. The system's temperature reaches the highest allowed value at the resulting t_{err} moments. This limitation in time ensures that the thermal cutter's temperature is kept below the degradation limit at all times. The value of the gradient was defined at a t_{cut} time by the gradient criterium (1) (5) supposing the worst case (Fig.6.). It can be seen in the diagram that two resistance values (3.9 Ω and 4.3 Ω) meet the gradient criterion.

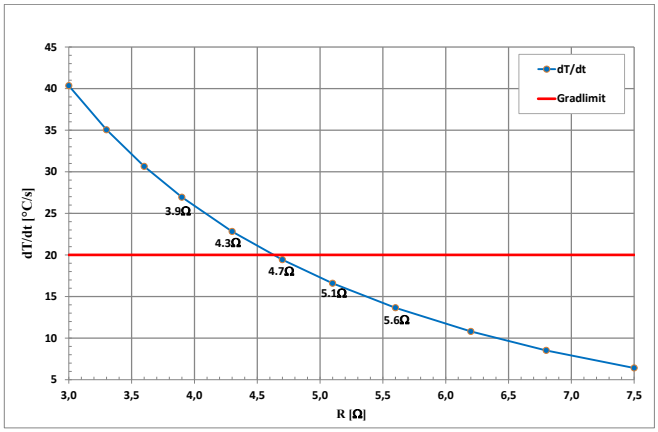


Fig. 6. – The gradient values at a t_{cut} time depending on the resistance value

Both of them satisfied the conditions but the 3.9 Ω was chosen for security reasons and for practical reasons based on the analysis of the video recording of the opening procedure. This choice resulted in a higher current consumption but also a steeper thermal gradient in the proximity of the cutting temperature. The task at this point is solved but there is another problem to solve.

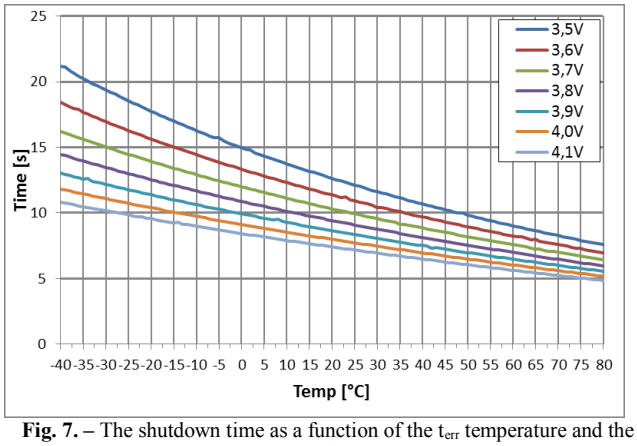


Fig. 7. – The shutdown time as a function of the t_{err} temperature and the U_{Batt} voltage

The resulting (Fig 7.) data array of 847 element is a difficult task to store in embedded systems. An efficient storing method was needed. A polynomial function was fitted on the curves. The order of the approximation that yields the desired accuracy had to be determined. The strong non-linearity at the lower section of the function excluded the low order fitting, thus a third order polynomial was chosen (11).

$$t_{err}(U_{Batt}, T_0) = A[U_{Batt}] \cdot T^3 + B[U_{Batt}] \cdot T^2 + C[U_{Batt}] \cdot T + D \qquad (11)$$

The coefficients of the polynomial (A[n], B[n], C[n], D[n]) were determined using numerical methods. The approximation reduced the number of need elements from 847 to 28. The error introduced by the approximation also had to be investigated. The following equation (12) was used to determine the difference of the originally calculated and approximated T_{err} values at the t_{err} moments.

$$\left| T(t_{err}(U_{Batt}, T_0), U_{Batt}, T_0) - T_{max} \right| < 1\ °C \qquad (12)$$

The error was below 1 °C in the entire operating range, which was considered acceptable. The formula above was implemented in a table form in C language in the satellite. A function taking the current temperature at a precision of 1 °C and the current battery voltage at a precision of 0.1 V as parameters returned the t_{err} value in milliseconds.

V. RESULTS

The system described above was tested in a climate chamber at several temperature values and at different stages of the battery. The opening procedure was recorded using a video camera. The difference between the real and predicted opening times was below 0.5 s. The thermal cutters always remained operable afterwards. Having completed the thorough and exhaustive tests, the system was used on board Masat-1. The antenna opening mechanism worked impeccably on board the pico-satellite which was put into orbit on 13 February 2012. The length of opening procedure can be determined from the logged data (Fig. 8.): it took 6.4 seconds at 16 °C external temperature and 3.8 V battery voltages.

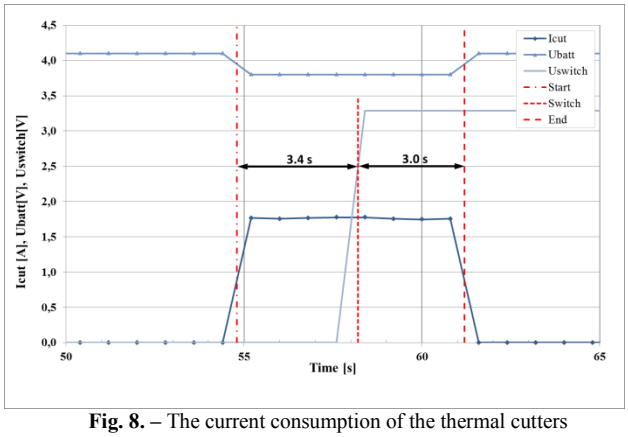


Fig. 8. – The current consumption of the thermal cutters

The figure above contains a 3 s software delay, so the real length of the procedure was 3.4 s. The estimated time for the given temperature and voltage data was 3.5 s according to the algorithm, thus the error of the prediction was 0.1 s.

VI. SUMMARY

We have developed a design method for highly reliable thermal cutters applied in a harsh environment. The method was put into practice in the design of a pico-satellite. The component was thoroughly tested on ground and in space.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the entire Masat-1 project and team for their outstanding work on the subsystems of the satellite making this project come true, especially Tamás Temesvölgyi, Levente Dudás, Lajos Varga, Sándor Glisics, Dávid Czifra, Zsolt Várhegyi. We would also like to thank the Faculty of Electrical Engineering of Budapest University of Technology and Economics for hosting the project and ESA (European Space Agency) for putting the satellite into orbit.

REFERENCES

[1] CubeSat Design Specification Rev. 12. http://www.cubesat.org/images/developers/cds_rev12.pdf (2012.03.31.)

[2] H. Heidt, J. Puig-Sauri, A. Moore, S. Nakasuka, and R. Twiggs, "CubeSat: A new generation of Picosatellite for education and industry low-cost space experimentation," in Proc. 14th Annu. AIAA/USU Conf. Small Sat., Logan, UT, Aug. 2000, SSC00-V-5.

[3] Dudás, Levente; Varga, Lajos; Seller, Rudolf, "The communication subsystem of Masat-1, the first Hungarian satellite", Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2009. Edited by Romaniuk, Ryszard S.; Kulpa, Krzysztof S. In Proc. of the SPIE, Volume 7502 (2009)., pp. 75020L-75020L-10 (2009).

[4] Joseph Costantine, Youssef Tawk, Christos G. Christodoulou, Jeremy Banik, and Steven Lane, "CubeSat Deployable Antenna Using Bistable Composite Tape-Springs", in Antennas and Wireless Propagation Letters, Vol.11, pp.: 285 – 288, 2012

[5] Joseph Costantine, Youssef Tawk, Christos G. Christodoulou, Jeremy Banik, and Steven Lane, "CubeSat Deployable Antenna Using Bistable Composite Tape-Springs", in Antennas and Wireless Propagation Letters, Vol.11, pp.: 285 – 288, 2012

[6] Metal Film Resistors datasheet http://www.yageo.com/documents/recent/Leaded-R_MF0_2011.pdf (2012.03.31.)

CFD Transient Model of the Buoyancy Heat Transfer for a Heat Sink: Effects of Geometry Rotation

Carmine Sapia, Giorgio Sozio
Applied Electronics Department – University of Rome "ROMA TRE"
Via della Vasca Navale 84 – 00146 – Rome (Italy)
Tel. +39.06.57337328 – Fax +39.06.57337026 – E-mail. sapia@uniroma3.it

Abstract- **The natural convection heat transfer from extended surfaces is, so far, the primary method of electronics cooling. With finned surfaces in air and natural convection, it is possible to manage a thermal power of 0.1 W/cm^2 with a temperature difference of ΔT = 80 ° C [1, 2].**

Cooling techniques based only on natural convection are of particular interest in all those situations where demands for miniaturization and low noise are dominant, and it is also a simple, reliable and low cost.

The study has been started analyzing the buoyancy induced heat transfer for an heat sink, commercially available, made by two aluminum horizontal rectangular thick fins per side positioned on a wide horizontal base plate. The time evolution of natural convection flow in the heating phase of the structure starting from t=0 (cold) to the development of the steady state has been investigated. The visualization of the buoyancy flow for the structure was implemented numerically by developing a computational fluid dynamics (CFD) model. The numerical solution of the problem was carried out by the discretization of the space in analysis. The mesh used were optimized according to an hexa unstructured structure, with a maximum size along the three axes fixed in 0.0075 m, 0.015 m, 0.01 m and a maximum spacing of 149.8 10^{-6} m. The post-processing tolerance has been set in 10^{-4}.

The time discretization for the transient analysis was 0.001 s (time step) and 20 iterations for each time step, in order to perform a transient analysis. The convergence criteria imposed (residues) for transient analysis were 0.001 for the flow and 1x10^{-7} for energy. Each solution reached convergence with an average time of about 13 days.

With a thermal load of 1 W localized in the middle of the base plate the temperature observed for the hot spot at the steady state in the commercial configuration was 60.8 ° C.

To enhance the readability of the extension of the stagnation areas of overheated flow in convective motion some appropriate numerical probes were positioned. The CFD model was also developed to monitor the evolution of the boundary layer, highlighting, in the volume surrounding the heat sink, the time evolution of the low-speed flow recirculation areas.

It was observed that the commercial heat sink presents a lower speed level of flow recirculation in the region between the vertical profiles in the side channels, as well as an extensive central area of stagnation.

It was also studied the effect of the heat sink geometry rotation on the natural convection heat transfer.

Keeping constant the structure of the heatsink in this work were investigated the effects connected with the simple rotation of the geometry and the expected increase of the heat exchange due to this choice

We suppose a vertical rotation of 90 deg to match the natural buoyancy flow direction. The proposed configuration leads to a significant improved performances. The geometry rotation allows a better flow recirculation speed. Stagnation areas are quite disappeared. In this way it is possible to reach a steady state conditions in a shorter time and lower (51.2 °C) operating temperatures for the hot spot of the system.

I. INTRODUCTION

Natural or forced convection heat sinks is still the most used electronics cooling method. They are preferred due to their inherent simplicity, reliability and low long term costs in cooling simple or multi-chip circuit boards. The consistent trends in microelectronic design goes to higher packaging densities and higher power dissipation rates.

In order to reduce the time to bring products to market the prediction of thermal performance of electronic equipment is becoming a necessity. The method of choice for making these predictions is mostly computational fluid dynamics (CFD) [3-5].

This paper presents a study oriented to analyze the convective heat transfer of a fin heat sink for power transistor. The analysis of the Thermo Fluid Dynamics has been performed through the development of a numerical model with Fluent inc. CFD software.

A question that arises very frequently in the choice of a heat sink is whether a given area of the base should choose one with very thick fins or one with fins very sparse.

An heat sink fins will have a very dense larger surface area for heat exchange but a lower coefficient for heat exchange because of the resistance of the fins to the fluid motion. In addition, an heat sink fins will have a very sparse coefficient of heat more but less heat exchange surface.

You can determine optimal spacing that maximizes the heat transfer coefficient of the heat sink for a given area of the base WL, and W and L respectively, width and height of the base of the heatsink vertically oriented. If the fins can be

considered isothermal and the thickness of the single wing t is small compared to the distance between fins s, the optimal spacing for the vertical fins of a heatsink is determined by the formula proposed by Bar-Cohen and Rohsenow [7]:

$$S_{ott} = 2.714 \frac{L}{Ra^{1/4}} \tag{1}$$

where the length L in the vertical fin is regarded as characteristic length to calculate the number of Rayleigh.

The coefficient of heat exchange in the case of optimal spacing is:

$$h = 1.31 \frac{\lambda}{S_{ott}} \tag{2}$$

The thermal power exchanged by natural convection fins can be calculated with the equation:

$$\dot{Q} = h(2nLH)(T_s - T_\infty) \tag{3}$$

where n is the number of fins on the heatsink and T_s is the superficial temperature of the fin heat sink.

The heat sinks with fins close together are not useful in the case of natural convection [7]. The presence of fins close together, forming narrow channels, it tends to block or suppress the fluid, especially when the heat sink is long, an effect that degrades the heat exchange characteristics of the heat sink.

The heat sinks with many vertical fins close together are not useful in the case of natural convection [6, 8]. In this case the narrow channels between vertical profiles tends to be an obstacle for the free convective motion of the heat transfer fluid.

II. Numerical Model

The analysis of the natural convection flow for the heat sink was implemented numerically by developing an appropriate model of computational fluid dynamics (CFD).

Anticipating, already at the design stage, how the natural convection fluid is distributed in the structure is fundamental. It allows thus to optimize the heat exchange process through the design of appropriate working volumes with effective reduction in the inefficient areas of heat stagnation. The visualization, even numerical, of the convective heat fluxes [9] plays a crucial role in the design of electronic components and systems.

Fluent uses a finite volume formulation of the Navier-Stokes and energy equations to discretize the problem.

The convection phenomenon is described [3-5] in these methods from a system of Continuity, Momentum and Energy equations. These equations describe the behavior and the evolution of the speed components and of the thermodynamic properties: pressure *p*, density ρ and temperature T in function of the time and the spatial coordinates. This system of equations includes the Navier-Stokes equations and the energy equation.

The continuity equation, or equation for the mass conservation is:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\rho \vec{v}\right) = 0 \tag{4}$$

The equation for the momentum transport, or momentum equation, is described in a inertial reference system, from:

$$\frac{\partial \left(\rho \vec{v}\right)}{\partial t} + \nabla \cdot \left(\rho \vec{v}\vec{v}\right) = -\nabla p + \nabla \cdot \left(\overline{\overline{\tau}}\right) + \rho \vec{g} + \vec{F} \tag{5}$$

Where p is the static pressure, ρ is the density, $\vec{v}$ is the velocity vector, *t* is the time, $\overline{\overline{\tau}}$ is the stress tensor, $\vec{g}$ is the gravity vector and *F* contains other source terms that run-off from the heating elements, the sources, etc.

The stress tensor $\overline{\overline{\tau}}$ is:

$$\overline{\overline{\tau}} = \mu \left[\left(\nabla \vec{v} + \nabla \vec{v}^T \right) - \frac{2}{3} \nabla \cdot \vec{v} I \right] \tag{6}$$

where μ is the molecular viscosity, *I* is the unit tensor, and the second term in the second member is the effect of the volume expansion.

The energy equation in a region of fluid can be written in terms of sensible enthalpy h as

$$\frac{\partial (\rho h)}{\partial t} + \nabla \cdot \left(\rho h \vec{v}\right) = \nabla \cdot \left[(k + k_t) \nabla T \right] + S_h \tag{7}$$

Where *k* is the molecular conductivity, k_t is the conductivity due to turbulent transport ($k_t = c_p \mu_t / \mathrm{Pr}_t$) and the source term S_h includes any volumetric heat defined.

It was analyzed the thermal exchange by natural convection of a fin heat sink, commercially available, two vertical profiles each side (ST338K) made of extruded aluminum. The object (figure 1) has been modeled in a Fluent ambient. This extruded aluminum material has been considered with the following specifics: density 2700 kg/m^3, specific heat 0.929 $kJ / Kg \cdot K$, conductivity 209 $W / m \cdot K$.

Referring to the real structure the dimension of the sink are: width 4.3×10^{-2} m; length is 2.02×10^{-2} m. The vertical profiles are 1.5×10^{-3} m thick, and 1.55×10^{-2} m tall, and the base is 2×10^{-3} m thick. The fin heat sink has been placed on a vertical support electrically controlled in temperature with a power load of 1 W. The cabinet has been defined as a rectangular prism 0.15 m wide, 0.3 m tall and 0.2 m long. The heat transfer fluid is air, with a temperature of 20° C, with pressure of 1 atm and a laminar flow regime set.

The numerical solution of the problem leads to a discretization of the analysis space. The selected meshes structure are optimized with a hexahedral unstructured shape with maximum size along the three Cartesian axes of 0.003 m and a maximum space of 199.8×10^{-6} m. The post-processing tolerance has been set in 10^{-4}.

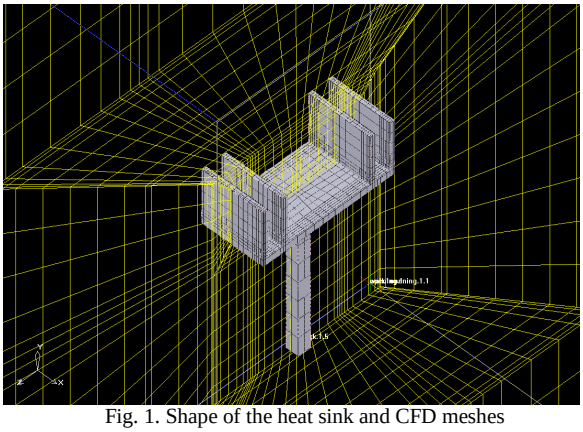


Fig. 1. Shape of the heat sink and CFD meshes

A non-stationary analysis has been performed, with a time step for the temporal discretization of 0.001 s and 20 iterations for each time step. The residuals imposed for the convergence of the solution for the transitory analysis are $1x10^{-3}$ for the flux and $1x10^{-7}$ for the energy. The solution reached convergence with an average computing time of about two weeks.

III. Results

A. *Original geometry*

The convective flow generated during the heating of the heat sink described has been analyzed. Particular attention was devoted to the transient phase of the phenomenon during the heating of the structure: the distribution of the recirculating flow from t = 0 (cold fin) to the steady-state has been analyzed.

The images are related to the speed profiles obtained by cutting planes, parallel to the x axis, at the middle of the extension of the heat sink (cut plane 0.506).

To better analyze the temporal evolution of velocity profiles of the convective recirculation flow a series of reference points has been placed (Figure 3) starting at a height from the base of $1.5x10^{-3}$ m and equally spaced of 0.003 m. A same distribution of the numerical probes has been positioned for the side channel identified by the vertical profiles.

The results are presented as velocity profiles (m/s) of the convective fluid circulation across the entire field of view (Figure 2), with particular attention to time evolution monitored through the measurement points (Figure 3).

In Figure 2 we can observe the convective flow emerging from the structure at t = 0.5 s of the transient heating. Natural convection flow are already organized in the lateral channels between the two vertical profiles; in other areas is only weak, as confirmed by the time trends recorded on the measuring points (figure 3). The region covered by the flow is mainly related to the vertical profiles of the structure.

In Figure 3 is shown the time window (3 sec) in which the main changes occurs. The steady state is reached after 5 s.

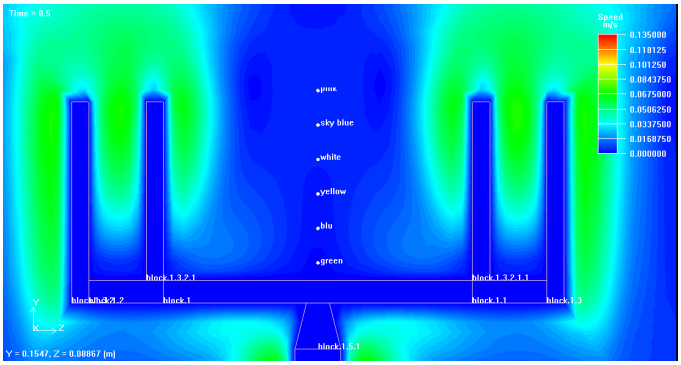


Fig. 2. Speed map at t=0,5 s.

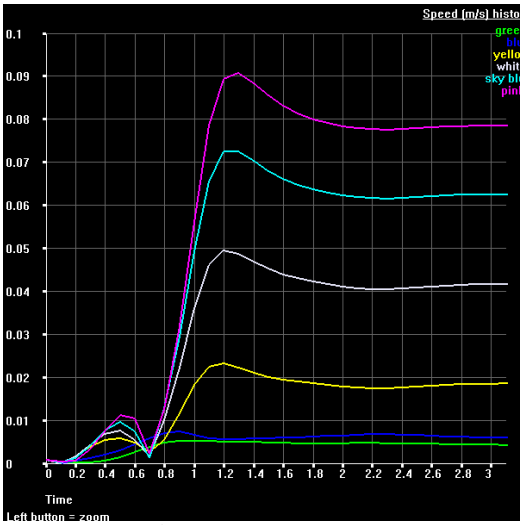


Fig. 3. Speed vs time for the numerical probes in the center channel.

Alter 1 sec heating (figure 4) the recirculation zones appear more structured: high speed flow (red) becomes visible at each side channel. At the top of the structure begins to organize a single central stagnation zone, nearly triangular. The thermal "bubble" is positioned in the middle monitored by numerical probes point identified as "white".

At the baseplate, in the ground of the heatsink, occurs others discrete areas of stagnation. The overheated air below the structure can emerge only for convective floating laterally beyond the extent of the fin base plate (Figure 4-5).

In figure 3 can be observed that the points with higher

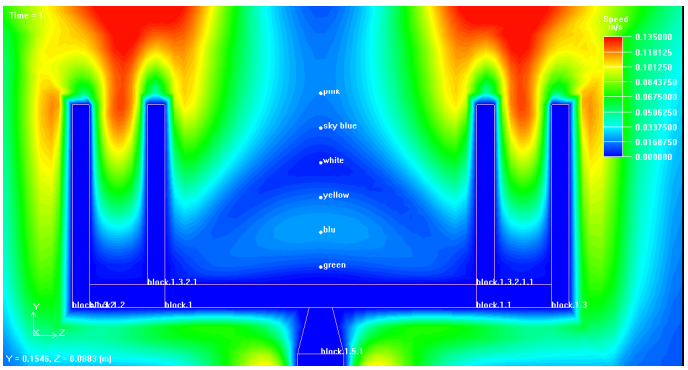


Fig. 4. Speed map at t=1 s.

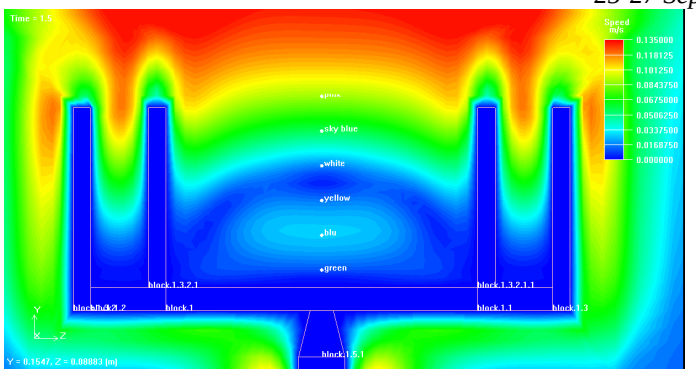


Fig. 5. Speed map at t=1,5 s

quote show a speed evolution that corresponds to their distance from the basis of the heat sink. The speed history show a minimum in 0.7 s. The lower points blue and green, that are always in the low recirculating zone, show a more constant trend, even if with low circulation speed.

The convective heat transfer on lateral channels identified by vertical profiles, shows a growing tendency to expand laterally extending above the central area. The convective recirculation in the central area is still of low intensity, as confirmed by the trends in figure 3.

In figure 5 is presented the velocity profile configuration, after 1.5 sec, of the convective flow for the examined structure. The central portion is now fully compressed at the top, according to a domed structure. At this moment the convective circulation is also well developed in the top center and it is observed a clearly defined border in the area of stagnation. The low recirculation bubble reaches its maximum compression ($1.15x10^{-2}$ m quote from the heat sink baseplate); the border region descends until it reaches the point identified with white. The numerical probes reach their maximum speed of recirculation after 1.3 s.

At the time of t=2 s the central area of stagnation of the fluid overheated is than organized into three discrete lobes. It can be seen that the dome reaches its maximum height at the top ($12.8\ 10^{-3}$ m quote). The border region in the area of stagnation is in the point of maximum expansion, above the area of placement of the numerical probes just mentioned and next to the sky blue point. The stagnation zone returns to expand until it reaches a height level, from the heat sink

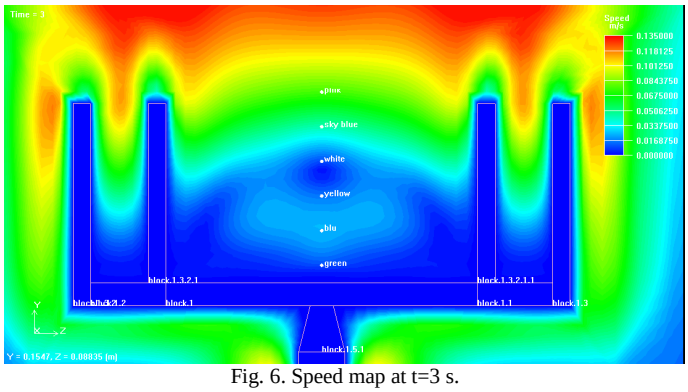


Fig. 6. Speed map at t=3 s.

base, of $12.3x10^{-3}$ m. The height described is almost stable and represents the final extension of the zone of stagnation. The steady state is reached after 5 s.

The boundary of the stagnation zone extends from a minimum of $4.5x10^{-3}$ m. In the time evolution the border region tends to gradually increase their height level up to a distance of $6.1x10^{-3}$ m. When the steady state has been reached the area with low recirculation, of stagnation for the cooling air, is confined within a distance of $6.1\ 10^{-3}$ m from the base plate of the heat sink.

The peak temperature observed measured in the middle of the heat sink at a quote of $5x10^{-3}$ m from the base fin (hot spots) is 60.8 ° C and is reached after a time of 1 sec.

In the lateral channel the maximum speed registered after 0.7 s is 1.05 m/s, the stabilizes after 2 s at 1 m/s.

B. Effects of 90 deg vertical geometry rotation

We have investigated the effects of a simple rotation of the original fin heat sink geometry. The new orientation was obtained after a 90° vertical rotation of the whole structure.

The buoyancy flows (in a gravitational field) are always in vertical direction, against the direction of the gravity vector. After the 90° vertical rotation the aerodynamical obstacle for the buoyancy flow, represented by the heat sink baseplate, is now reduced only to the cutting section of the structure.

In the following images the results of the CFD analysis are showed, according to the numerical model developed.

The natural convection flows are now better organized and developed. Figure 7-11 shows the transient heating phase of the rotated geometry. It's possible to observe that the stagnation areas with low air recirculation speed are quite completely disappeared.

The same numerical probes of the situation A (original geometry) were disposed in the central and lateral channel in order to better understand the buoyancy flow distribution in the areas surrounding the heat sink

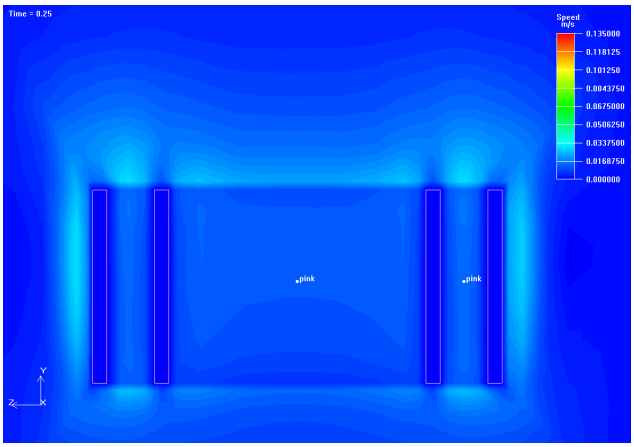


Fig. 7. Speed map at t=0.25s. CFD probes in central and side channels.

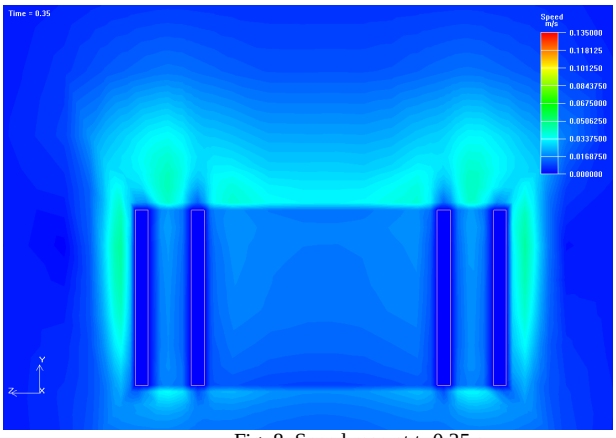


Fig. 8. Speed map at t=0.35 s

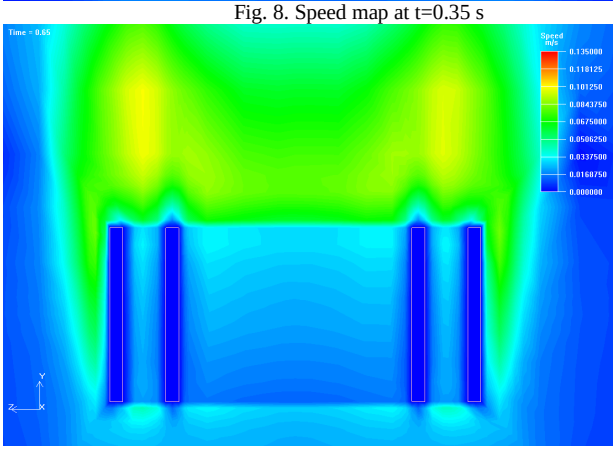


Fig. 9. Speed map at t=0.65 s.

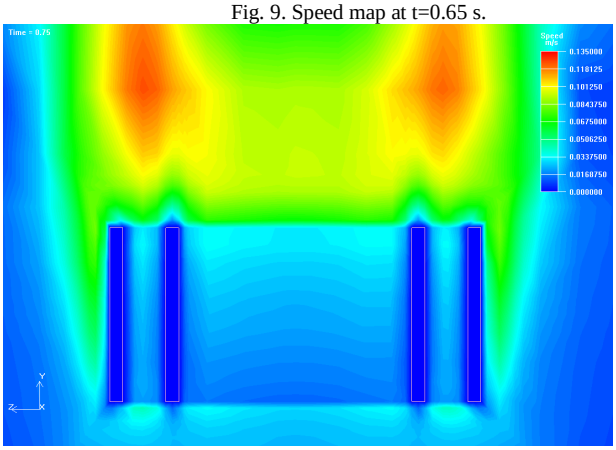


Fig. 10. Speed map at t=0.75 s.

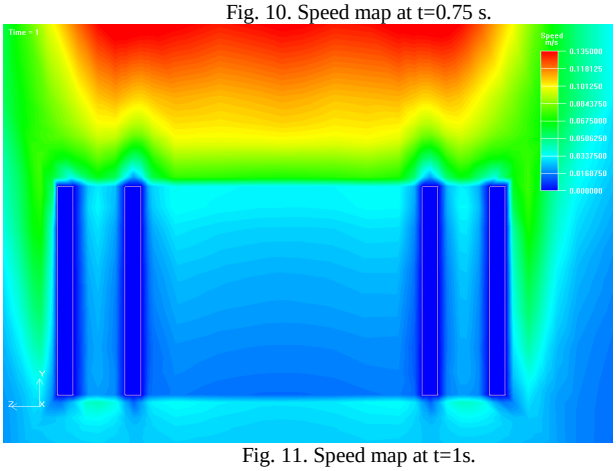


Fig. 11. Speed map at t=1s.

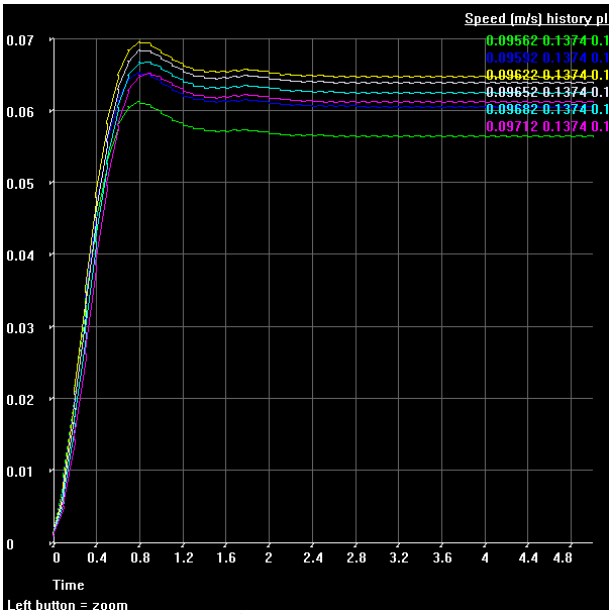


Fig. 12. Speed vs time for the numerical probes in the center channel.

It can be observed that quite the same natural convection flow speed are measured inside the central channel in all the observation points (fig 12). In the lateral channel (fig 13) the green and blue patterns, related to points closer to the heat sink baseplate, are both partially slowed down. The most intense convective flows are located outside the finned heat sink.

The absence of the main aerodynamical obstacle (heat sink baseplate) to the free emersion of the buoyancy connective flows, leads to a shorter transient heating time. Temperature reaches the steady state after about 1 second. The maximum temperature in the hot spot is reduced to 51.2° C (fig. 14).

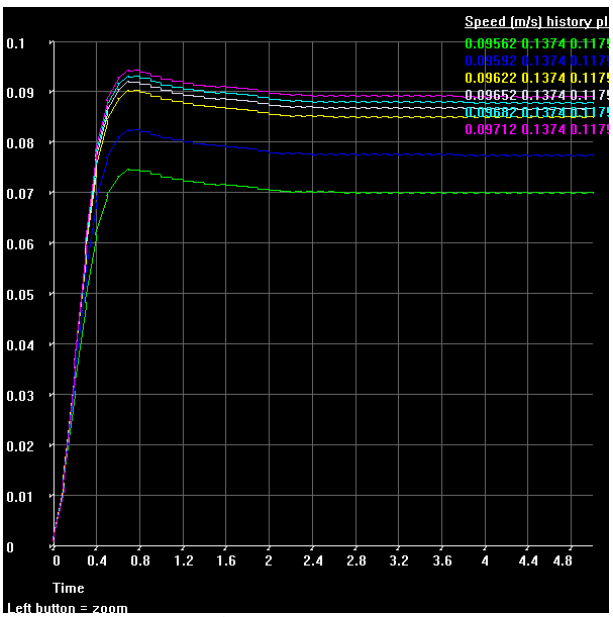


Fig. 13. Speed vs time for the numerical probes in the lateral channel..

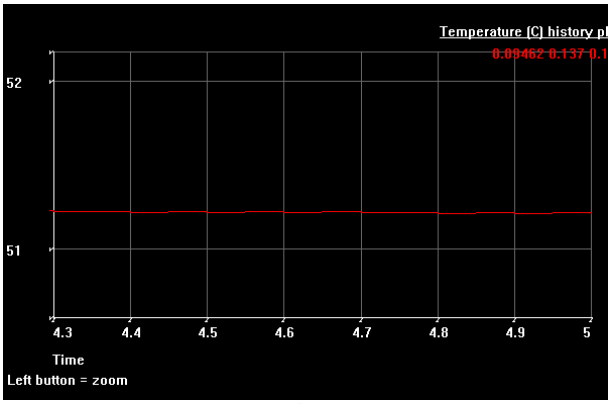


Fig. 14. Temperature vs Time for the hot spot at t=5 s: steady state

The 90 deg vertical geometry rotation clearly shows that the buoyancy flow is now better organized, with stagnation areas considerably reduced. This makes possible to reach the steady state in a shorter time with system's hot spot lower temperature.

Changing the orientation to the fin heat sink seems to be a cheap solution to increase the natural convection heat transfer. But a cooling geometry rotation needs often more working volume in the electronic system and different electrical interconnections. It's not always possible or simple to apply.

In the Figure 15 it is showed the buoyancy flow organization for the rotated finned heat sink with a larger field of view for the CFD cutting plane selected.

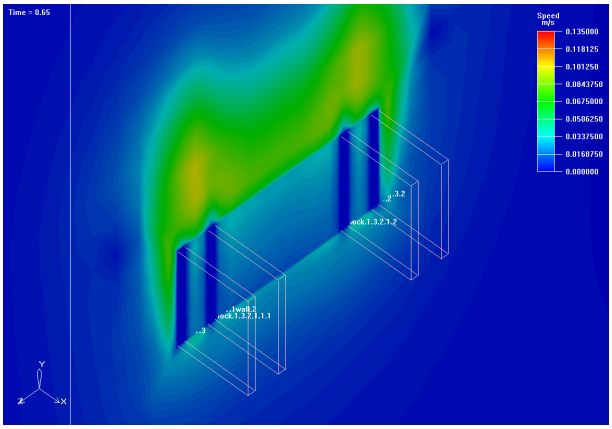


Fig. 15. Speed map at t=0.65 s.

IV. Conclusions

The results of the CFD analysis of the natural convection heat transfer for an heat sink has been presented. The computational model developed was tested on a particularly recurrent geometry in the cooling of electronic components (heat sink ST338K). With the CFD model developed was possible to analyze the development of natural convection flow and the hot spot peak temperatures.

Particular attention has been paid, in this work, to the study of the transient phase of the convective phenomenon. The work was focused to monitor the time history of the convective flow in order to show the distribution and the time evolution of the natural convection flow highlighting low-speed recirculating areas (stagnation zones). The heating phase of the structure was analyzed, starting from t = 0 (cold) to the steady-state. The stationary configuration, is reached after a time of 5 sec.

To enhance the readability of the boundary of stagnation zones were placed some numerical probes both in the large central body and in the narrow lateral channels.

The time history shows how the convective heat transfer appears to be initially more developed on the side vertical profiles. Later (t=1.5 s) the area of high speed convective recirculation is closed on top with a clear identification of lower "bubbles" of stagnation for the overheated fluid. The peak temperature observed is 60.8 ° C reached after a time t = 1 s.

The recirculating speed map history shows how the boundary zone proceeds compressing the heat sink base. The lower distance from the base is reached after 1.5 seconds; the region of low-speed recirculation extends to an height of 11.5×10^{-3} m. The stagnation zone expands to a maximum height of 12.8×10^{-3} m (t = 2s) and stabilizes at a distance of 12.3×10^{-3} m (t = 3s) almost stable up to the steady state (t = 5 s).

A simple change was also investigated in order to reduce the stagnation zones and the hot spot working temperature: a 90° vertical rotation along the z axe..

The proposed configuration has improved performance. The stagnation areas was clearly reduced leading to a significant reduction in the system hot spot temperature. The operating temperature of the hot spot of the system is stabilized at 51.2 ° C for the 90 deg rotated geometry. In the original "horizontal" configuration the temperature stabilizes at 60.8 ° C. A cooling geometry rotation needs often more working volume in the electronic system and different electrical interconnections; this operation not always simple to apply.

REFERENCES

[1] Kim, S.J., Lee, S.W., *Air cooling technology for electronic equipment*, CRC Press, Boca Raton, LA, 1996.

[2] Incropera, F. P., *Convection heat transfer in electronic equipment cooling*, J. Heat Transfer, vol. 110 No. 4, pp. 1097-1111, 1988.

[3] D. Anderson, J. Tannehill, and R. Fletcher. *Computational Fluid Mechanics and Heat Transfer*. McGraw-Hill Book Company, New York, 1984.

[4] S.V. Patankar, *Numerical Heat Transfer and Fluid Flow*, McGraw-Hill Book Company, New York, 1980.

[5] C.A.J. Fletcher, *Computational Techniques for Fluid Dynamics*, Springer – Verlag Heidelberg New York, 2nd edition 2006.

[6] R. Bahadur and A. Bar-Cohen, *Thermal design and optimization of Natural Convection Polymer Pin Fin heat Sinks*, IEEE Transactions on Components, Packaging and Manufacturing Technologies, vol. 238, No. 2, pp 238-246, 2005.

[7] A. Bar-Cohen and W. M. Rohsenau, *Thermally Optimum Spacing of Vertical, Natural Convection Cooled, Parallel Plates*, Journal of Heat Transfer, February 1984, vol. 106, Issue 1, pp 116-124.

[8] S. Narasimhan and J. Majdalani, *Characterization of Compact Heat Sink Models in Natural Convection*, IEEE Transactions on Components, Packaging and Manufacturing Technologies, vol. 25, No. 1, pp 78-86, 2002.

[9] R. L. Linton and D. Agonafer, *Coarse and Detailed CFD Modelling of a Finned Heat Sink*, IEEE Transactions on Components, Packaging and Manufacturing Technologies, vol. 18, part A, No. 3, pp 517-520, 1995.

25-27 September 2012, Budapest, Hungary

Integrated microcooler structure realized by wet chemical etching

Gy. Bognár, B. Plesz, P. G. Szabó, G. Hantos
Budapest University of Technology and Economics, Department of Electron Devices
H-1117 Magyar tudósok krt. 2, Budapest, Hungary
E-mail: <bognar|plesz|szabop|hantos>@eet.bme.hu

Abstract – In this paper we summarize the first steps of a new idea to create cost effective microcooler structures with high performance. Our basic idea is to use a cheap etching process instead of reactive ion etching or LIGA technologies. However before we go further into varying the layout in order to increase the cooling efficiency, the performances of the new structures have to be compared against others which were formed either by RIE or LIGA. In order to validate our structures' cooling capability thermal transient tests were performed. These tests proved are expectation as the cooling efficiencies were almost identical.

I. INTRODUCTION

The number of integrated components on the surface of integrated circuits and the operating frequency of IC-s are continuously increasing. In modern processors (e.g.: CPU, GPU, DSP) the dissipated power per unit area is reaching 100W/cm^2. Furthermore the stacked dies and 3D packaging possibilities result in even more elevated temperatures of the packaged devices. Also in high power applications (e.g.: power transistors, power LEDs) the temperature of the applied devices must be kept within an acceptable range avoiding the decreasing efficiency and failures of operation. The generated heat must be extracted on the shortest and lowest thermal resistance ways, thus the active or passive cooling of these devices is prominently important. The major cause of electronic failures is the elevated temperature; more than 55 % of failures are caused by overheating giving an urgent demand for new approaches in cooling solutions.

Another important and emerging field with specific cooling requirements is concentrator solar cells. As the efficiency of solar cells is decreasing with the increasing temperature the appropriate cooling is not only important by the means of failure prevention, but energy optimization. In this case not only the capacity, but the effectiveness (i.e. the power consumed for cooling) of the applied cooling system is a crucial aspect.

Conventional air-cooling systems are reaching their limits, thus new designs are investigated in order to be able to cope with the increased heat flux and to keep junction temperatures in an acceptable range [1][2][3]. An urgent demand for more efficient cooling has risen, which resulted in micro-scale thermal management solutions. Other approaches include application of integrated micro-refrigerators attached close to the hot-spot locations. In these applications high heat-flux (up to 300 W/cm^2) can be achieved with localized cooling [4].

Other often mentioned cooling solutions are applying microscale cooler structures. The basic idea of these devices is to use small, submillimeter size channels or patterns to bring the coolant as close as possible to the regions where the power is dissipated. Several approaches of microcoolers have been published, reaching from active (externally driven coolant transport) to passive (coolant transport driven by dissipation heat) devices, or from discrete devices to integrated implementation [1][2][3].

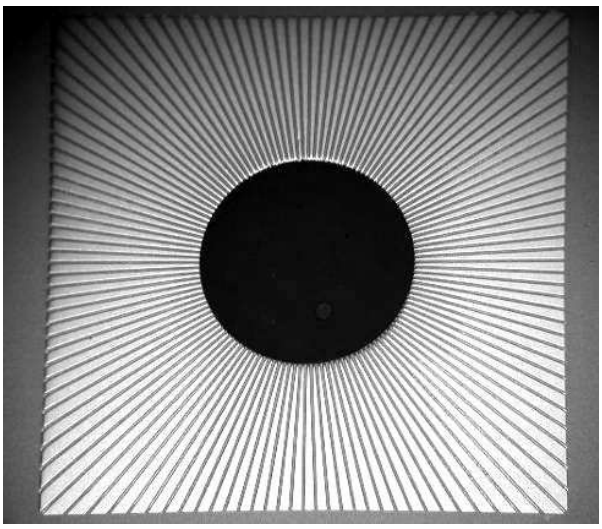


Fig. 1 The previous microcooler plate

Liquid cooling structures realized by Through Silicon Via (TSV) and Reactive Ion Etching (RIE) techniques were introduced in [5]. However these cooling structures were not realized in the silicon die itself. By applying TSV and RIE techniques the integrated cooling system of a 3D packaged Multi-Processor System-on-Chip (MPSoC) stacked die structures were presented [9]. However applying RIE and DRIE in the fabrication process flow has some major disadvantage: only one wafer can be etched at once (single wafer processing), and the RIE equipment and the process itself are expensive. A promising method for integrated microcoolers could be the realization of micro channel structures in the semiconductor device by wet etching that has the advantages of allowing batch processing by using relatively cheap etchants. By realizing the potential of wet chemical etching we aim to investigate and develop etching

processes for the cheap high throughput processing for microchannels in integrated microcoolers which will be presented in the second chapter.

The measurement of the exact thermal properties of different cooler structures is also a great challenge. Commercially available measuring systems and well known methods give only the thermal resistance from junction to ambient of the overall system. To be able to characterize the thermal impedance of only the components under investigation new approaches need to be developed.

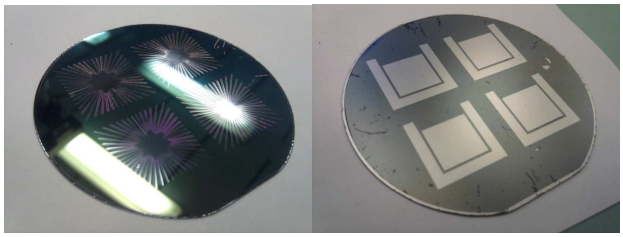
The thermal transient measurement is a well established method for the thermal characterization of electric components [10][11]. The measurement results of this technique do not only cover the junction to ambient resistance, but the thermal impedance (i.e. the thermal resistance and the heat capacitance) of each single component in the main heat path can be determined as well. Thermal transient testing has already been successfully applied to the characterization of non-integrated microcoolers, and proved to be a powerful and reliable tool for their thermal characterization [6][7][8]. In the third chapter the new ideas to improve the existing measurement techniques and setups for the application with integrated microcoolers in semiconductor devices will be presented, resulting in a reliable and stable combined electro-thermal characterization method for electronic devices containing integrated microcoolers.

In parallel with the characterization methods thermal models were developed that can also be applied later for the design and the simulation of integrated microcoolers. Cross verified with the data received from the measurements, simulations based on the created models will help to predict further possibilities of different integrated microcooler structures.

II. SAMPLE PROCESSING

Integrated microcoolers were processed by wet anisotrpic etching of <100> n-type silicon wafers in tetramethyl-ammonium-hydroxide (TMAH). The etching was performed in a 25 % solution of TMAH at 85°C resulting in an etch speed of ca. 0,3 microns/minute. The etching solution contained no additives, and was performed in an ultrasonic bath. Due to the ultrasonic bath the bubbles formed during etching are quickly removed from the silicon surface. This ensures homogenous etching even in case of batch processing. In addition the combination of ultrasonic excitation and slow etch speed results in a very low surface roughness.

The samples were formed on 2" single side polished wafers, 4 devices with the size of 15 x 15 mm on each wafer. The channels for the microcooler were etched on the polished surface in radial directions (Fig. 2a). At the intersection in the middle of the chip a cavity of ca. 4 mm in diameter was formed for the gas inlet. On the other side of the wafer a p-n junction of the size 9 x 9 mm was formed by boron diffusion for each of the 4 microcooler structures. An aluminium metal contact is evaporated on the p layer and on the n substrate around the p-n junction (Fig 2b) to ensure good electrical contact to the sample. In an alternate version the metal layer is not deposited on the p-n junction. In this case the p-n junction can act as a solar cell and can be excited with light.



a) b)

Fig. 2 Processed wafers with samples: a) polished side with etched channels, b) backside with diodes and metal contacts

Finally the wafers were diced and the channels on the polished side of the wafer were closed with borofloat 33 glass attached by anodic wafer bonding. For the gas inlets a hole with a diameter of 1 mm was drilled into the closing glass substrate by laser cutting. One of the finished devices can be seen in Fig. 3.

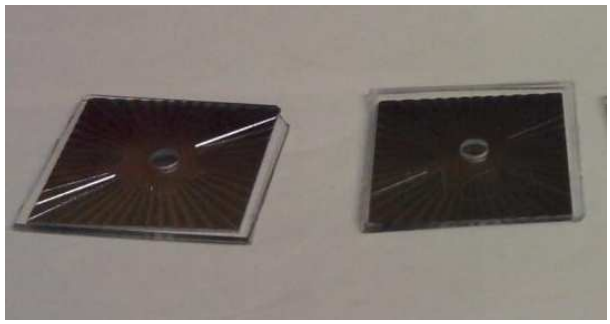


Fig. 3 Completed samples

III. DESIGNING THE NEW MEASUREMENT SETUP

The measurements have been carried out by thermal transient testing. However a new thermal characterization method are developed to determine the exact heat transfer coefficient and partial thermal resistance of the microcooler structure itself with different channel patterns and fluids flow rates. Since the cooler structures are formed in the backside of the semiconductor substrate that incorporates different electronic elements (simple diodes, solar cells, bipolar or MOS integrated circuits) on the topside a new measurement setup and support structure was also developed. In this case not only the fluid inlet(s) outlet(s) should be formed but the electrical connectivity should be provided as well. During the characterization steps the constant fluid flow rate must be ensured.

By measuring the temperature change caused by a heat dissipation step the thermal resistivity of the whole structure can be easily determined. But in our investigations it was necessary to determine only the thermal resistance and capacitance (impedance) of the microcooler structure itself in order to give the exact thermal properties and compare the measured results with the simulated ones. Hence from the results of the thermal transient testing the integral structure function was derived. From this the thermal properties of the different parts of the heat flow path(s) can be determined.

During initial measurements the dissipating element is a power transistor on which the microcooler structure (semiconductor chip with integrated microcooler device) is placed. In this case the integrated devices are not used for power dissipation but due to this the measurement results can be directly compared to the previous measurements results of nickel microchannel plate formed by LIGA technology and presented in [6][7][8].

In future the integrated electronic devices will act as both heat source and temperature sensor during the measurements. In case of the integrated cooler structure expectedly the thermal resistance from junction to ambient will decrease due to the decreasing thermal resistivity between the active area and the cooling channels.

In order to ensure the validity of the results the parallel heat flow paths should be eliminated besides providing proper electrical connection to the electronic devices on the top side and coolant supply to the microcooler structure in the back side of the semiconductor. Forasmuch the etched channels of the microcooler are covered by glass plate and the structure is placed onto a thermal insulator material. As a result the heat flow path from the microcooler structure (from the backside of the semiconductor devices) to the ambient by thermal conduction is nearly eliminated. On the top side of the semiconductor substrate the electrical conductivity is formed by pressing a proper printed wiring board (PWB) to the surface of the semiconductor device. During the measurements we must deal and calculate with the parallel heat flow path (heat flow from the active area of the semiconductor through the PWB to the ambient). As we expected we managed to determine the parallel heat flows by measuring an empty measurement setup (means without gas flow and without microcooler structure).

The arrangements and pattern of the channels influence the property and characteristics of fluid flow and the thermal properties as well. By increasing the cross section of the heat flow path (semiconductor surface contacting the flowing fluids) the partial thermal resistivity of the microcooler decreases hence the heat transfer coefficient increases.

IV. MEASUREMENT RESULTS

We used the measurement support structure presented previously in [6][7][8]. An NPN power transistor was applied as a dissipator element and the silicon chip with integrated microcooler structure was placed on top of it. The thermal properties of the cooling structure (junction-to-ambient thermal resistance and the partial thermal resistance of the microcooler structure) can be derived from the thermal transient response of the system. So the aim of the measurement is to obtain the integrated structure functions against different rate of gas flow.

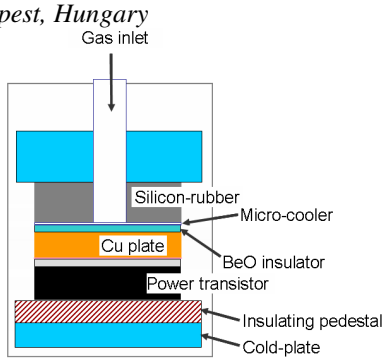


Fig. 4 Measurement setup

The measurement setup consists of the followings from bottom to top:

- Aluminium heatsink
- heat insulator sheet
- NPN transistor
- Copper heatsink
- Silicon substrare with integrated microcooler structure faces against the copper heatsink
- Glass sheet (which is part of the microcooler in this case) providing electrical and thermal insulation as well
- Silicon tube and sheet for ensuring gas inlet
- Aluminium clamping faces

During the measurement a 3.5 W step function dissipation was applied. The measurements have a 900 second runtime after a 900 second delay because this is the approximated time after the system reaches steady state.

In Fig. 5. the measurement results of the new integrated microcooler can be seen.

As expected the R_{thja} junction-to-ambient thermal resistance decreases with the increasing rate of gas flow (compressed air). In Fig. 5 the red and black curves indicate the integrated structure functions at coolant flows of 60 l/h and 120 l/h.

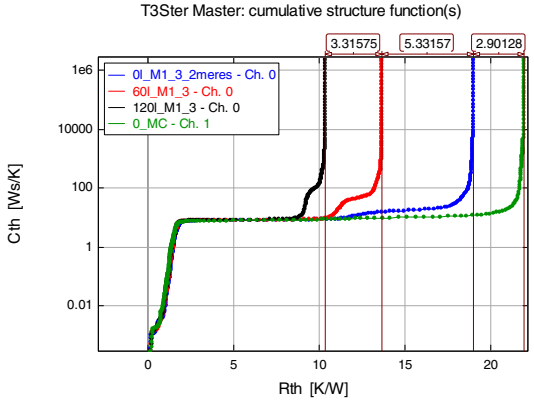


Fig. 5 Integrated structure functions against different rate of gas flow

It is worth comparing the results of the current measurements performed on the integrated microcooler with previous results gained from measurements on LIGA-formed microcooler plates. The comparison can be seen in Fig. 6 The black (0 l/h), brown (60 l/h) and green (120 l/h) curves

are the results of the measurements of the new microcooler design while the 1st grey (0 l/h), blue (60 l/h) and 2nd grey (120 l/h) curves indicate the results of the previous measurements of a microcooler sample formed by LIGA technology. It is interesting to see that despite the R_{thja} (junction-to-ambient thermal resistance) values at 0 l/h are differs approximately by 2 K/W between the new and the old design the values at 60 l/h and 120 l/h are identical. It is also noticeable that the curves show no differences until ~7 K/W which indicates the unchanged measurement environment. From this it can be seen that the new integrated microcooler design shows identical thermal properties as the old design (R_{thja} and R_{thMC}) if the coolent flows are equal.

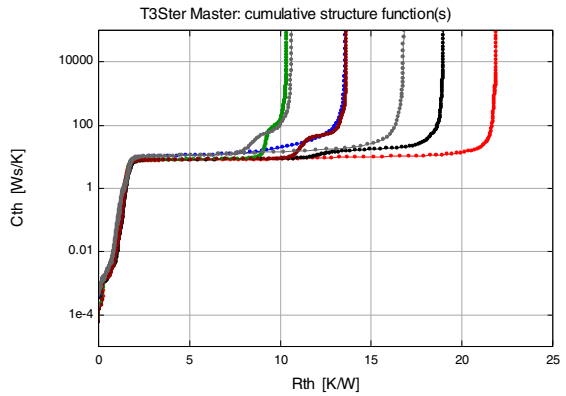


Fig. 6 Integrated structure functions against different flow rates compared the measurement results in 2005

IV. SIMULATION

Besides the measurements numerical analysis or analytical approximations are also crucial in the characterization process. The thermal model which is the result of the thermal transient measurement can also be calculated by using the material properties and geometric data. This approximated model can be matched with the one of the thermal transient testing and can be simulated by using industry standard SPICE simulators to verify the behaviour of the original structure. If the results are adequate then model can be used later to predict the behaviour of other structures with different layout.

V. CONCLUSION

So far we investigated the possibilities of manufacturing a microcooler structure by cheap wet chemical etching instead of the high cost RIE and LIGA techniques. However the performance of these microcooler structures has to be tested against the more expensive ones. The thermal transient tests showed that with the same layout near identical cooling efficiency can be achieved with the structures made by wet chemical etching and those made by LIGA in 2005. Our further aim is to investigate the possibilities of increasing the efficiency by layout modifications for which we will use electrothermal modelling as well to obtain more sophisticated and replicable results.

ACKNOWLEDGMENT

The authors wish to express their gratitude towards laboratory assistant Zsuzsa Harmath and master student Ádám Luhály who participated in the processing of the integrated microcooler samples. The authors also thank Tamás Kárpáthy from the HAS RCNS for performing the wafer bonding of the samples. The work was partially supported by the JEMSiP_3D ENIAC JU project (No 120016) of NKTH (No OMFB-00166/2010).

REFERENCES

[1] Sauciuc et al.: Air-cooling extension – performance limits for processor cooling applications, Proceedings of the XXth SEMI-THERM Symposium, pp. 74-80, (2003)

[2] Chen Li and R.A. Wirtz: Development of a high performance heat sink based on screen-fin technology, Proceedings of the XXth SEMI-THERM Symposium, pp. 55-60, (2003)

[3] S.V. Garimella: Advances in mesoscale thermal management technologies for microelectronics, Proceedings of the 10th THERMINIC Workshop, pp. 185-206, (2004)

[4] G. Jeffrey Snyder et al: Hot Spot Cooling using Embedded Thermoelectric Coolers, Proceedings of the XXIIth SEMI-THERM Symposium, (2006)

[5] Gong Yue Tang et al: Integrated Liquid Cooling Systems for 3-D Stacked TSV Modules, IEEE TRANSACTIONS ON COMPONENTS AND PACKAGING TECHNOLOGIES, VOL. 33, NO. 1, (2010)

[6] Kohári Zs, Bognár Gy, Horváth Gy, Poppe A, Rencz M, Székely V Cross-verification of Thermal Characterization of a Microcooler. JOURNAL OF ELECTRONIC PACKAGING 129:(2) pp. 167-171. (2007)

[7] Desmulliez, M. P. Y., Pang, A. J., Leonard, M., Dhariwal, R. S., Yu, W., Abraham, E., Bognar, Gy., Poppe, A., Horvath, Gy., Kohari, Zs., Rencz, M. Fabrication and Characterization of a Low-Cost, Wafer-Scale Radial Microchannel Cooling Plate. IEEE TRANSACTIONS ON COMPONENTS AND PACKAGING TECHNOLOGIES 32:(1) pp. 20-29. (2009)

[8] W Yu, M P Y Desmulliez, A Drufke, M Leonard, R S Dhariwal, D Flynn, G Bognar, A Poppe, G Horváth, Z Kohári, M Rencz High-aspect-ratio metal microchannel plates for microelectronic cooling applications. JOURNAL OF MICROMECHANICS AND MICROENGINEERING 20:(2) 5 p. Paper 025004. (2010)

[9] Mohamed M. Sabry: Towards Thermally-Aware Design of 3D MPSoCs with Inter-Tier Cooling, Design, Automation & Test in Europe Conference & Exhibition (DATE), (2011)

[10] Székely Vladimír, Tran Van Bien. Fine structure of heat flow path in semiconductor devices: A measurement and identification method. Solid-State Electronics, 31(9):1363–1368, 1988.

[11] Székely Vladimír, Rencz Márta, Poppe András, Courtois Bernard. New way for thermal transient testing. 1998.

25-27 September 2012, Budapest, Hungary

Simulations on Vanadium Dioxide Thin Film as Thermographic Material

Márton C. Bein, János Mizsei
Budapest University of Technology and Economics
Department of Electron Devices
Magyar tudósok körútja 2., H-1117 Budapest, Hungary
Email: {bein, mizsei}@eet.bme.hu

***Abstract*—Vanadium dioxide (VO_2) has been extensively investigated due to its thermal-induced semiconductor-metal transition (SMT) at about 67 °C; transition temperature can be lowered by doping. Besides the rise in electrical conductivity of up to 3-4 orders of magnitude the phase transition is followed by other changes in different physical properties, including an optical reflectance drop near to its half. Possible applications include thermally controlled electrical and optical switches.**

In this paper thermal simulations of a proposed thermographic arrangement are presented. The structure consists of a 100 nm thick VO_2 thin film deposited on the passivated surface of a silicon chip. The goal of these simulations is to predict the acquirable resolution of the proposed method. Abstract heat sources are defined on the silicon surface and the temperature of the thin film is calculated. Thermal mapping could be performed by optical observation of the isotherm at transient temperature, while the dissipation and the cooling of the chip is swept.

Possible benefits of the proposed method include the simple and reliable use of the once-deposited solid phase film, maintaining a few micron resolution, similar to liquid crystals, as predicted in earlier electro-optical measurements.

***Index Terms*—vanadium dioxide, thermographic film, simulation**

I. Introduction

Thermal-induced semiconductor-metal transition (SMT; traditionally mentioned as metal-insulator transition, MIT) of vanadium dioxide (VO_2) has been investigated for more than 50 years [1], [2]. In pure VO_2 the transient temperature is about 67 °C, however, it can be lowered near to room temperature by doping with metallic ions, e.g. tungsten or molybdenum [3], [4]. The transition temperature shows hysteresis with a width of 1 °C for macroscopic bulk crystals, but it has been shown, that for thin films and nanoparticles this value can be 10-15 °C or even larger [5]. Above the transition temperature VO_2 bulk shows metallic conductivity while for lower temperatures it acts like an intrinsic semiconductor. During the transition conductivity can increase by up to 3-4 orders of magnitude [6].

MIT is accompanied with a quite abrupt change of many physical parameters, first of all that of the crystal lattice structure. However, it is still not clear, whether the electrical changes are consequences of the structural changes or these have the same cause [7]. Semiconducting and metallic phases have different reflectance in the visible region, thus the transition is observable using an optical microscope [6], [8].

A dominant part of the integrated circuit failures attributable to thermal causes, largely cracks due to different thermal expansion and overheating. Consequently, thermal analysis, simulation and management of ICs have grown to a separate profession in the past 20 years. One of the analytical techniques is thermography, where the thermal map of the investigated (IC) surface is obtained. There are several methods developed, e.g. contactless techniques including the use of thermographic cameras and laser interferometers as well as use of thermochromic coatings like fluorescent compounds with temperature-dependent decay time and thermochromic liquid crystals [9], [10].

Optical properties presented above allow vanadium dioxide to be used as thermochromic coating for heat-protecting intelligent windows [5], and as thermographic layer for IC thermal mapping. A 100 nm thick VO_2 film is partially transparent in both phases [11], thus the surface of the analyzed integrated circuit will be visible during thermographical analysis and results could be matched precisely with topology.

Expected benefits of thermographic application include the simple and reliable use of the once-deposited solid phase films and a possibly higher resolution compared to liquid crystal thermography [12]

II. Thermographic arrangement

A proposed thermographic arrangement is presented on Fig. 1. A VO_2 layer of about 100 nm is to be deposited over the chip. As the thermographic layer is (semi)conductive it could short the wiring; an insulating layer must be applied underneath. Typical ICs are covered with passivating phosphosilicate glass (PSG) except for the pad ring; therefore if the VO_2 layer is laid on the glass it will not disturb the circuit.

The pad ring of the chip is not covered by glass, this means that the conductance of the thin film is to be considered and eliminated here as well. Evident, but not developed solution is the masking of the VO_2 layer. Another way could be the use of a discontinuous coating i.e. consisting of non-percolating grains with a size of at most a few μm.

Different optical reflectance and transmittance of the semiconducting (lighter) and metallic (darker) phase is observable by optical microscopy from above. The border of these regions gives the isotherm of the transient temperature. The hysteresis of the phase transition can not be neglected, however, if the

 ISBN:978-2-35500-022-5

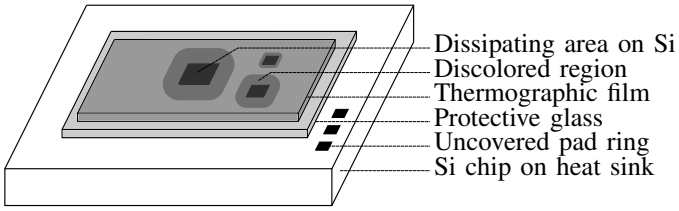


Figure 1. Proposed arrangement of VO_2 as thermographic thin film over an IC. The continuous film must be isolated from the IC surface as it could shorten the wiring; protective PSG could be used. The isotherm of the transient temperature is at the border of the lighter (semiconducting) and darker (metallic) region.

temperature change during the formation of the regions is monotonic accuracy is not limited by it.

The back side of the chip must be attached to a heat sink; otherwise the temperatures in the chip will equalize due to the poor cooling relative to the high thermal conductance of the silicon. One steady-state dissipation arrangement with a fixed heat sink temperature will give one particular isotherm. An isotherm map can be obtained if the temperature of the cold plate is swept [13]. Dissipation map can be calculated from the thermal map [14].

The main advantage of the proposed technique is that the layer has to be deposited once and as a solid material can be used for several times. In both phases the 100 nm thick layer is still partially transparent and the structure below it can be seen.

Although this structure looks quite simple, there is no elaborated process of the VO_2 layer deposition compatible with a CMOS IC, especially because the high temperature and/or oxidative atmosphere used in common techniques [15] damages wiring.

III. Model Setup

A quasi-realistic model used in the simulations is presented on Fig. 2. The structure consists of dissipating stripes on the surface of a silicon chip covered with insulating glass region and a 100 nm thick VO_2 thermographic layer. The thermal properties of the glass are supposed to be similar to SiO_2.

The model represents a 3D periodic structure which can be simulated in 2D; this can be explained in two steps. First, the temperatures do not change with y dimension as this dimension of the device is supposed to be infinite relative to the typical sizes of the structure and there is no change in this direction. Second, the structure repeats along x dimension with a period length of a_x (considered infinite times as well) and this means that there is no heat flux horizontally between the symmetric "unit cells", one of which is shown as a dashed rectangle on the figure. A unit cell with its properties is presented on Fig. 3.

The parameters used in the simulations are shown on Table I; some of them are fixed. The thickness of VO_2 is negligible compared to the other layers, thus not varied. A typical wafer thickness and average dissipation is chosen. Dissipation per

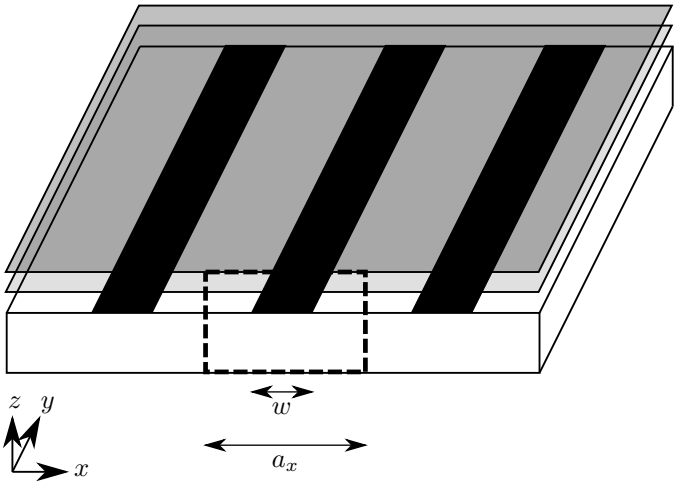


Figure 2. Periodic model used in the presented simulations. Black stripes represent the dissipating areas on the Si surface. The periodicity of the structure and the width of the stripes are a_x and w, respectively. The insulating glass and VO_2 layers are continuous over the structure. The dashed area showing a part of the cross section is a kind of "unit cell" as the whole structure can be built from it with translation operations.

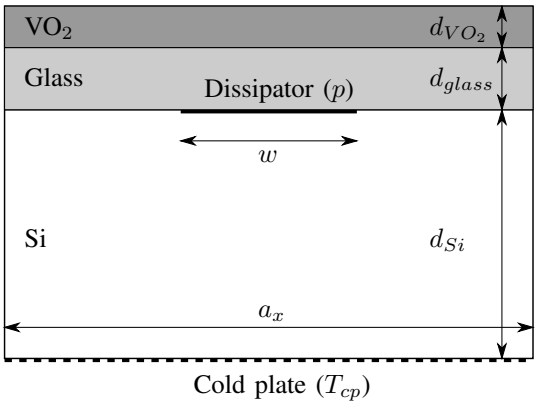


Figure 3. 2D unit cell of the periodic model.

unit area of the stripe is calculated from the average dissipation and the fill factor of it: $p = P\frac{a_x}{w}$

Output of the simulation was the temperature along the upper surface of the VO_2 layer. The dynamics of these data is analysed in order to estimate the achievable resolution.

IV. Results and Discussion

First of all, the linearity of the structure is examined. The thermal model used for silicon and glass (SiO_2) are mostly

d_{Si}	$500\,\mu m$
d_{VO_2}	$100\,nm$
d_{SiO_2}	$0.1\,..\,3\,\mu m$
$P_{average}$	$100\,W/cm^2$
a_x	$10\,..\,100\,\mu m$
w	$0.05\,..\,0.5\,a_x$
T_{cp}	$330\,..\,340\,K$

Table I
Parameters used in the model presented on Fig. 2

 ISBN:978-2-35500-022-5

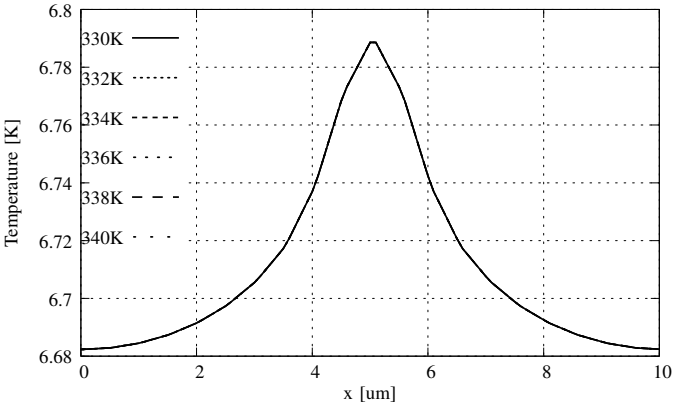


Figure 4. Temperature profile of a structure with 100 nm glass layer, 10 μm period length 1 μm stripe width and various cold plate temperature. Plots are normalized with the latter. The result is practically identical which means that nonlinearity plays no role in this setup.

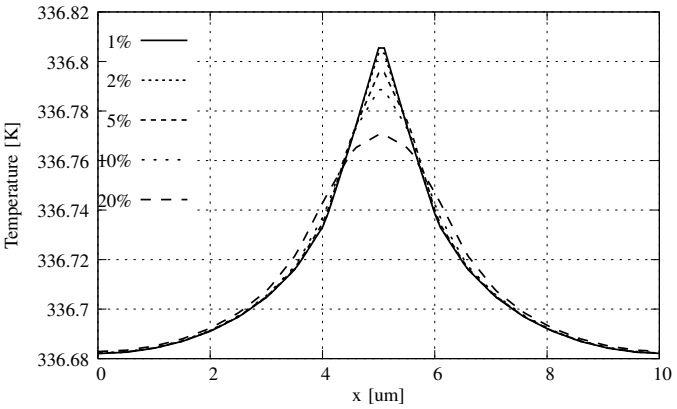


Figure 5. Temperature profile of structures with 10 μm period length and 330 K cold plate temperature and different dissipator widths (fill factor shown in %). One can see that maximum slope does not differ significantly.

linear, while VO_2 shows some change in thermal conductivity at the phase transition.

On Fig. 4 the temperature profile of the same structure for different cold plate temperatures is to be compared. Plots are normalized with this "base" temperature and are practically identical which means that the role of the nonlinearity of the 100 nm thick VO_2 layer in the behaviour of the structure can be considered marginal, therefore we will not analyze the temperature profile as a function of cold plate temperature.

On Fig. 5. structures with 10 μm period length and 330 K cold plate temperature are compared. One can see that maximum slope does not differ significantly compared to the variation in the width of the stripes.

The maximum slope for the sample with $100\,nm$ wide dissipator stripe is $78\,mK/\mu m$, while the one with $2\,\mu m$ wide stripe has $44\,mK/\mu m$ slope. This means that the slope is dominated by the layer structure rather than the fill factor of the dissipator. These slopes necessitate fine thermal resolution of cold plate.

In support of this conclusion we compared the maximum slope as a function of glass thickness (Fig. 6). Thin insulating

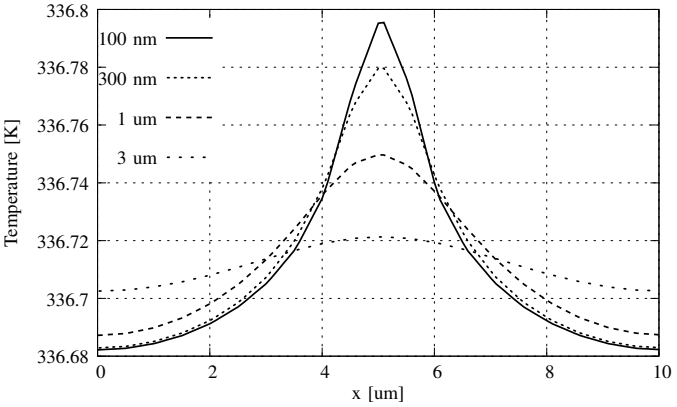


Figure 6. Temperature profile of structures with 10 μm period length, 500 nm wide dissipator elements, 330 K cold plate temperature and different dissipator glass thickness. The slope is now significantly affected, however, other layers have an important role as well, which can be observed on thin glass layers.

layers do not change the slope dominantly (100 nm and 300 nm thickness), showing that the good thermal conductor silicon spreads heat along the system. On the other hand, thicker glass layers greatly slur the temperature profile.

PSG layers are typically a few μm in thickness. Considering this Fig. 6 suggests that the surface temperature map of a submicron chip is limited by the glass.

On all the figures presenting thermal profiles of the chip surface we can observe moderate temperature differences which can be surprising if we mention that the average dissipation is 100 W/cm^2. In steady-state the excellent thermal conduction of silicon balances the temperature of the fine structures making the thermography harder, while maintaining a more homogeneous heat distribution helps the circuit work evenly.

V. Summary

In this paper a possible thermographic arrangement using VO_2 as the thermochromic layer is presented. A 2D model is designed and simulations were carried out with different geometries. the typical temperature differences on the surface of a structure with 10 μm period length is at about 0.1 K. As a conclusion we must draw that the resolution in space is limited by the protecting glass layer of the chip and in temperature by the punctuality of the cold plate. The inaccuracies caused by the hysteresis of the VO_2 layer is possibly eliminated by monotonic temperature sweep, however, it must be investigated in the future.

VI. Acknowledgement

This work is connected to the scientific program of the "Development of quality-oriented and harmonized R+D+I strategy and functional model at BME" project. This project is supported by the New Széchenyi Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002).

References

[1] F. J. Morin, "Oxides which show a metal-to-insulator transition at the neel temperature," *Phys. Rev. Lett.*, vol. 3, no. 1, pp. 34–36, Jul 1959.

 ISBN:978-2-35500-022-5

[2] A. Zylbersztejn and N. F. Mott, "Metal-insulator transition in vanadium dioxide," *Phys. Rev. B*, vol. 11, no. 11, pp. 4383–4395, Jun 1975.
[3] F. Kong, M. Li, S. Pan, Y. Zhang, and G. Li, "Synthesis and thermal stability of w-doped vo2 nanocrystals," *Materials Research Bulletin*, vol. 46, no. 11, pp. 2100 – 2104, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0025540811002881
[4] S. Xu, H. Ma, S. Dai, and Z. Jiang, "Study on optical and electrical switching properties and phase transition mechanism of mo^{6+}-doped vanadium dioxide thin films," *Journal of Materials Science*, vol. 39, pp. 489–493, 2004.
[5] J. Nag and R. F. Haglund, "Synthesis of vanadium dioxide thin films and nanoparticles," *Journal of Physics: Condensed Matter*, vol. 20, no. 26, p. 264016, 2008.
[6] G. Gopalakrishnan, D. Ruzmetov, and S. Ramanathan, "On the triggering mechanism for the metal–insulator transition in thin film vo_2 devices: electric field versus thermal effects," *Journal of Materials Science*, vol. 44, pp. 5345–5353, 2009.
[7] G. Stefanovich, A. Pergament, and D. Stefanovich, "Electrical switching and mott transition in vo 2," *Journal of Physics: Condensed Matter*, vol. 12, no. 41, p. 8837, 2000.
[8] M. C. Bein and J. Mizsei, "Contracting current paths in vanadium dioxide thin films," in *Thermal Investigations of ICs and Systems (THERMINIC), 2011 17th International Workshop on*, sept. 2011, pp. 1 –4.
[9] J. Altet, S. Dilhaire, S. Volz, J.-M. Rampnoux, A. Rubio, S. Grauby, L. D. P. Lopez, W. Claeys, and J.-B. Saulnier, "Four different approaches for the measurement of ic surface temperature: application to thermal testing," *Microelectronics Journal*, vol. 33, no. 9, pp. 689 – 696, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0026269202000514
[10] J. Kölzer, E. Oesterschulze, and G. Deboy, "Thermal imaging and measurement techniques for electronic materials and devices," *Microelectronic Engineering*, vol. 31, no. 1–4, pp. 251 – 270, 1996, <ce:title>Electron and Optical Beam Testing of Electronic Devices</ce:title>. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167931795003487
[11] S. Heinilehto, J. Lappalainen, H. Jantunen, and V. Lantto, "Ir-wavelength optical shutter based on ito/vo2/ito thin film stack," *Journal of Electroceramics*, vol. 27, pp. 7–12, 2011, 10.1007/s10832-010-9604-9. [Online]. Available: http://dx.doi.org/10.1007/s10832-010-9604-9
[12] W. Liu and B. Yang, "Thermography techniques for integrated circuits and semiconductor devices," *SENSOR REVIEW*, vol. 27, no. 4, pp. 298–309, 2007.
[13] A. Csendes, V. Székely, and M. Rencz, "Thermal mapping with liquid crystal method," *Microelectronic Engineering*, vol. 31, no. 1–4, pp. 281 – 290, 1996, <ce:title>Electron and Optical Beam Testing of Electronic Devices</ce:title>. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167931795003509
[14] X. Wang, A. Shakouri, S. Farsiu, and P. Milanfar, "Extraction of power dissipation profile in an ic chip from temperature map," in *Semiconductor Thermal Measurement and Management Symposium, 2007. SEMI-THERM 2007. Twenty Third Annual IEEE*, march 2007, pp. 51 –56.
[15] S. Heinilehto, J. Lappalainen, V. Lantto, and H. Jantunen, "Pulsed laser deposition and e-beam evaporation of vanadium dioxide thin films for ir-photonics applications," *Proc. SPIE 7022*, vol. 7022, no. 1, p. 70221G, 2007.

 ISBN:978-2-35500-022-5

25-27 September 2012, Budapest, Hungary

Practical Dynamic Thermal Management of Multi-core Microprocessors

Bartosz Wojciechowski, Marek A. Bawiec
Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
bartosz.wojciechowski@pwr.wroc.pl

***Abstract*—Here we present an implementation of thermally-aware DVFS governor in a form of Linux 3.2 module. Our thermal governor operates by reading digital thermal sensors placed in CPU cores and pro-actively adjusting operating frequency of individual cores to maintain temperatures below given threshold. We evaluate our method using state-of-the-art parallel benchmarks from PARSEC suite. Apart from evaluation of our thermally-aware DVFS governor, we present insights into operation of a modern high-performance CPU with 6 cores and 2 hardware threads per core.**

***Index Terms*—Dynamic Thermal Management, temperature-aware DVFS, Linux kernel, SMT/CMP**

I. Introduction

Temperature is one of the main factors that limit performance and greatly impact microprocessor reliability. Technology scaling leads to very high power densities resulting in hot-spots in processor cores under heavy load. The other important factor is the cost of high-performance cooling. Apart from price of the cooling solution itself, there is the matter of energy usage by the heat sink fan and the cost of extra air conditioning related to computer operation. Other important factors are noise pollution caused by fans and weight and size of cooling package, that limit portability and scalability.

Temperature is a complex function of processor load and cannot be easily predicted [1]. Also, performance is a non-linear function of frequency, especially with Simultaneous Multithreading (SMT) where multiple threads share resources of a single core [2]. Another performance-limiting factor in multi-core processors is memory contention [3]. Eventually, maximizing performance in contemporary chip multiprocessors (CMPs) with SMT under temperature limits is a multidimensional problem [4].

Due to quadratic dependence of dynamic power on voltage, Dynamic Voltage and Frequency Scaling (DVFS) is an effective mechanism to restrict temperature [5]. In contemporary processors frequency transition times are of the order of microseconds allowing for fast response with negligible overhead. DVFS is widely used as energy conservation mechanism in present-day computing systems.

DTM can be made on hardware or Operating System level. Hardware mechanisms are robust but do not deal with the information on executed tasks that the OS collects. Therefore, using solely hardware DTM excludes performance optimization. Dynamic Thermal Management of microprocessors is the subject of many papers and many works on thermally-aware scheduling for Linux systems exist (e.g. [6]–[8]).

Here, we present a thermally-aware DVFS manager in a Linux 3.2 module that maintains temperatures of each CPU core below the given threshold. Our DVFS governor works by measuring temperature of all the cores as well as performance status of all logical CPUs of an SMT microprocessor. Then for each core a DVFS level is selected that will not lead to excessive temperature.

Most of the work on Dynamic Thermal Management for multi-core microprocessors published up to date uses an elaborate simulation setup or analytical modelling to verify the effectiveness of the proposed algorithm or policy, e.g. [9]–[13]. However analytical and simulation-based works rarely deal tend to take arbitrary assumptions about thermal parameters of microprocessor packages, CPU power dissipation or properties of thermal sensors such as accuracy, that influence applicability of developed methods.

With the advent of affordable multi-core processors more works emerged that were focused on aspects of practical implementation of DTM policies on real processors. Authors of [6], [8], [14], [15] show that the current `Linux` scheduler can easily be enhanced with thermal-awareness to improve its performance in terms of both temperature conditions and total application throughput. Their experimental results demonstrate that the dynamic thermal management implemented in operating system is an effective method to control processor temperature.

The main contributions of this work comprise temperature-aware DVFS governor for Linux called `cpufreq_temp`, that limits temperature and seeks optimisation possibilities; insight in operation of a contemporary CMP/SMT processor; method for Digital Thermal Sensor (DTS) error compensation; we also provide DTS error evaluation for the processor used for testing.

II. Measurement of CPU temperature, performance and power consumption

Our experiments were done on a high-performance 6-core double-threaded processor `Intel Core i7 970`. It's schematic layout is shown in Fig. 1. There are 6 cores placed in-line with 2 large areas of L3 cache at the bottom. The total die size is 239 mm^2. With TDP of 130 W the CPU is equipped

ISBN:978-2-35500-022-5

with heavy, high-performance heat-sink. The CPU supports 13 frequency levels from 1.6 GHz ($f_{\min}$) to 3.2 GHz ($f_{\max}$). There are 2 hardware threads per core, i.e. two sets of architectural registers and the core uses higher of the 2 values set in the appropriate control registers.

We examined performance and thermal behavior of our test computer using `cpuburn` programs, which provide constant and uniform load in all threads, and `PARSEC` [16] benchmark set. The total running time of the test set was 52, 32, 21, 19 and 18 s for 1, 2, 4, 8 and 16 threads, respectively. As can be seen, using more than four threads does not lead to big performance improvements. These are however accumulated results and do not reflect the behavior of individual programs. In `Linux 3.2` kernel the cores **c0** through **c5** were mapped to pairs of logical CPUs respectively: (2,8), (0,6), (4,10), (5,11), (1,7), (3,9).

A. CPU power measurements

We measured power consumption of the whole computer during idle and full load. Then we subtracted the idle power in room temperature (103 W in our case) to get an estimate value; obviously, such method of CPU power consumption measurement is prone to some error caused by a non-linear efficacy of the computer's power supply, limited efficiency of the CPU voltage regulators [17] and instantaneous change in power consumption by peripheral devices (such as hard drives, or a graphic card).

To check power consumption of each core, we run the computer idle for 10 s and then every 20 s we added one thread of a `burnP6` program, putting 2 threads on the first core and then moving to the next. With every two threads, P_k rose about 15 W. The differences of up to 1.6 W can be attributed to rising temperature and static power. We discovered that the relationship between DVFS level and power consumption is surprisingly nearly linear by running the processor under full load while decreasing voltage settings one step at a time.

B. Temperature measurements

Digital Thermal Sensors (DTS) in Intel's processors [18] are visible to software through a set of Model Specific Registers (MSR) as a difference from maximal available temperature T_{jmax}. They are individually calibrated but are subject to normal factory variation and their accuracy is best at temperatures close to T_{jmax}. CPU temperature depends on the

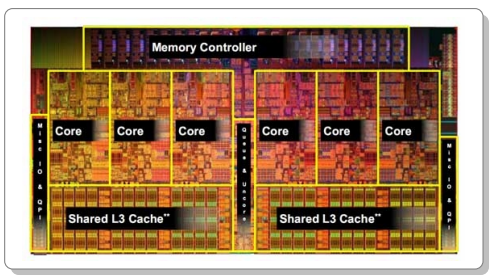


Figure 1. Die photo of `Core i7` processor (Source: `Intel`'s marketing materials)

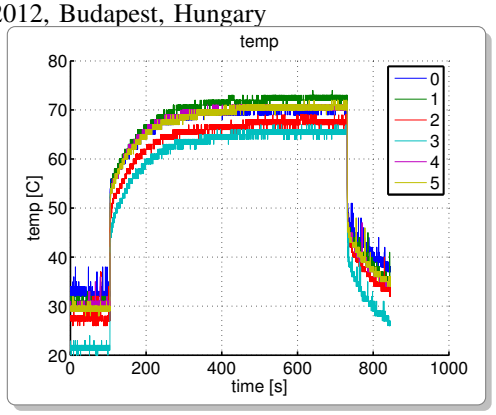


Figure 2. Thermal response of the processor to a constant-IPC load. Temperature errors not corrected.

ambient temperature inside the PC case T_{amb}. We measured it using a 1-wire `Dallas 18B20` digital thermometer placed in the air streamed into the heat sink. We observed changes in T_{amb} of up to 10 K.

During normal operation the changes in temperature can be substantial and reach 3.3, 4.7, 6.0 and 8.0 K in 5, 10, 20, 100 s respectively. Therefore a simple reactive DTM mechanism should provide a trade-off between temperature limit and latency. With longer operating period, the processor should work with higher thermal guard-band, and hence, with lower performance.

In this work we use a fixed CPU fan speed and hence we have constant die-to-ambient thermal resistance R_{amb}. Total thermal resistance of CPU package and cooling solution that we measured is approximately 0.37 K/W (based on power consumption swings and temperature differences between idle and load steady-state, see Fig 2. Based on the time of increase of the temperature, we estimated thermal capacitance of the heat-sink to be around 240 J/K. The power consumption of the whole computer was 103 W in idle state, 197 W during full load and slowly increased to 204 W with the increase in CPU temperature. This last increase of 7 W can be associated with increased leakage power.

There are three potential sources of differences in temperature readings under uniform load on all cores. The first is the difference in power consumption of different cores caused by factory variation causing non-uniform power dissipation. However, we saw no such differences. In our experiments, with identical workload, all cores dissipated the same power. Other source of temperature difference is the layout of the processor. Due to thermal coupling to nearby cores and *uncore* elements (e.g. memory controllers, IO, etc.) some cores may run hotter than others. The last source of temperature difference is the error of the thermal sensors.

By default the DTSs in our CPU show very different temperatures in each core (see Fig 2) both in stress and during long idle periods. When heated with 12 instances of `cpuburn` program the measured steady-state core temperatures oscillated around: 72, 71, 67, 65, 70, 71, for each of the cores respectively. During idle periods, the steady state temperatures

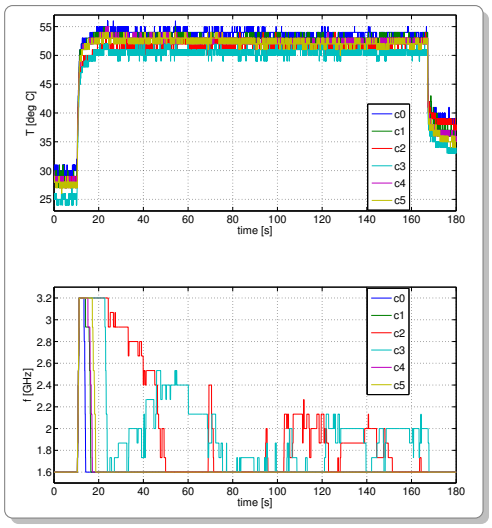


Figure 3. Frequency balancing under fixed temperature limit: temperature and core frequency traces

were: 33, 31, 27, 22, 30 and 31. Temperature inside the PC was ranging between 26 during idle periods and 29 in high stress. The room temperature oscillated slightly around 23°C.

To find the correcting factors for all the temperature sensors, we wrote a simple program that balanced frequencies of the cores subject to uniform workload and temperature limit. It adjusted frequency and voltage settings of each core not to cross the temperature limit and reduced differences in frequency between all the cores by correcting DTS readouts. Temperature and DVFS traces. Its operation can be seen on Fig. 3. The final modifiers that we achieved that provided good operation of our DVFS governor were: -2, -1, +5, +9, +0 and +0 K for cores **c0**–**c5**.

III. Thermally-aware DVFS governor

We implemented a reactive DTM mechanism in the form of a temperature-aware DVFS governor. It was designed and tested as a Linux kernel module for the 6 core microprocessor with SMT. By default Linux kernel supports five frequency scaling governors, namely: `conservative`, `ondemand`, `userspace`, `powersave` and `performance`. Each logical processor (as seen by the OS) can have a governor assigned to it separately. `Conservative` and `ondemand` governors adjust core frequency based on the current load. Once the load exceeds a predefined up-threshold, clock level is increased by a unit (`conservative`) or set to maximum frequency (`ondemand`). When CPU load is below down-threshold, the frequency is decreased one step at a time. Powersave and performance governors set the lowest and the highest available level. The `userspace` governor provides an interface through the `sysfs` for users and applications to set the desired frequency level[1].

Our proposed governor makes decisions based on global information about all the logical processors (as seen by the OS) and the temperatures of physical cores. Since only one frequency can be assigned to one core, we assume that the same governor would be assigned to all logical CPUs.

[1]The interface is available in the form of a file in /sys/devices/system/cpu/cpuN/cpufreq/scaling_setspeed, where N is the number of logic processor

Table I
Execution time and energy consumption of a subset of PARSEC benchmarks using different DVFS governors. Parameters: $(t_s, T_{\max}, T_{th})$

DVFS governor	time [m]	energy [kJ]
cpufreq_temp (80,55,50)	14:38	123
cpufreq_temp (40,55,50)	14:30	123
cpufreq_temp (10,55,50)	14:17	124
ondemand	11:10	104
conservative	11:10	105
powersave	19:33	153
performance	10:58	103

A thermally aware DVFS governor to be used in a general-usage operating system, must have the following properties: low latency, low overhead, assigns high frequencies to high-priority tasks, assign lower frequencies for tasks characterized by very frequent references to main memory, assigns higher frequencies to cores with 2 threads and/or high IPC.

The governor operates in three phases. In the first one, information about temperature and activity (instructions committed, cache misses) of all the logical CPUs is updated. Then load[2] of each logical CPU (in range 0–11) is updated. When the temperature is below the predefined threshold, the operating principle is the same as in the case of the `conservative` governor. When T_{th} is exceeded, the core frequency is reduced so the temperature never exceeds $T_{\max}$.

The value of response in case of surpassing T_{th} is calculated as a sum of Proportional, Integral and Derivative modules forming a PID controller. This controller is reset every time the temperature drops below T_{th}. The temperature is tracked so that every time the previous value is known. Our governor operates on each core periodically, with default period of $t_s = 80$ ms. This step length ms provides negligible overhead of the governor with reasonably high guard-band $T_{\max} - T_{th}$.

In Figure 4 the thermal trace of the processor is presented during execution of the PARSEC set with `native` input sets and 8 threads. After short pre-heating for 60 seconds, the benchmark set is run three times with our `cpufreq_temp` governor with varying latency (see Tab. I). Then `ondemand` and `conservative` governors are used. They result in almost the same performance and processor thermal behavior. The `powersave` governor sets the $f_{\min}$ causing effectively doubling the computation time but reducing the temperature to below 50°C. The `performance` governor sets the $f_{\max}$ throughout the benchmark execution which results in the best performance. The exact execution times and energy used by the computer are shown in Tab. I. Our frequency governor achieves robust temperature reduction and provides a reasonable trade-off between performance, energy consumption and temperature under heavy load.

[2]The fraction of time a processor is busy executing an active process.

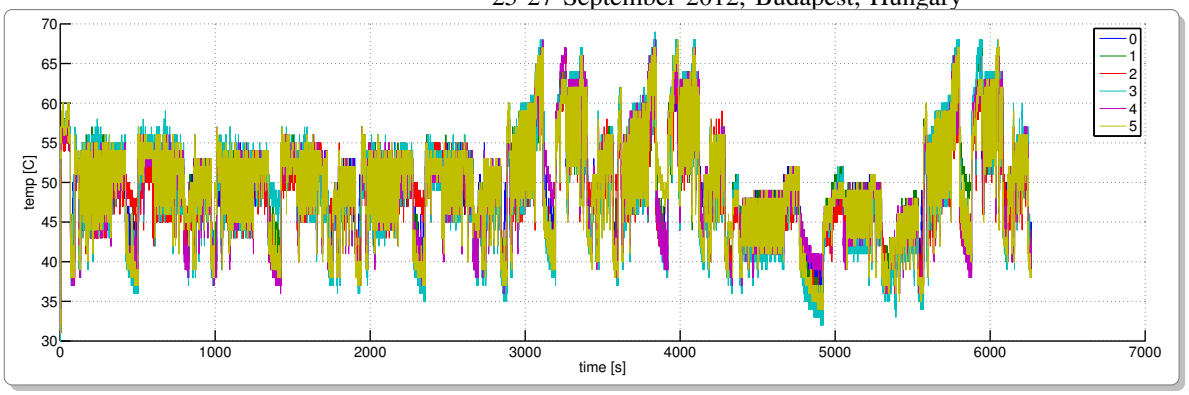


Figure 4. Temperature trace of all six cores when running `PARSEC` with eight threads and `native` input set and different DVFS governors. Left to right: three times with **cpufreq_temp** and latency t_s of 80, 40 and 10 ms , then using `ondemand`, `conservative`, `powersave` and `performance` governors

IV. Discussion and summary

Low temperature limits used in experiments reflect the really mild operating conditions that the experiments were performed in. In real-life situations ambient temperatures can easily be 10–20 K higher. This however does not affect the results, since temperature of the cores is simply a sum of ambient temperature and dissipated power divided by thermal resistance to ambient.

Although this paper shows the efficacy in reducing temperature, a comprehensive comparison against hardware throttling and more extensive testing is still needed. This can either be done by lowering the threshold at which the CPU starts to limit it's performance or by putting the computer in more severe thermal conditions. Also, a broad testing of optimisation possibilities is still needed. It remains to be verified that our proposed DVFS governor operates equally efficiently with heterogeneous workloads (in terms of program types and their priorities).

During the course of this research one thing become apparent: that a global power-performance-temperature-energy global optimisation of CPU operation is needed on the OS level. Therefore, our next goal will be to add temperature awareness to the scheduler of the OS and ensure proper use of all the computing resources under thermal limits. Contemporary operating systems are not temperature-aware when it comes to scheduling. Proper utilisation of the ever-growing computing resources will be of vital importance in the years to come.

In this work we have presented capabilities of a contemporary multi-core microprocessor with simultaneous multi-threading regarding on-line temperature and performance measurement. We have also presented a temperature aware DVFS governor for Linux OS and CPUs with symmetric multiprocessing that allows for an effective trade-off between performance and operating temperature.

References

[1] W. Huang, M. Allen-Ware, J. Carter, E. Cheng, K. Skadron, and M. Stan, "Temperature-aware architecture: Lessons and opportunities," *Micro, IEEE*, vol. 31, no. 3, pp. 82–86, may-june 2011.

[2] D. T. Marr, F. Binns, D. L. Hill, and e. a. Glenn Hinton, "Hyper-Threading Technology," *Intel Technology Journal*, vol. 06, 2002.

[3] M. Jahre, M. Grannaes, and L. Natvig, "A Quantitative Study of Memory System Interference in Chip Multiprocessor Architectures," in *11th IEEE Int. Conf. HPCC '09.*, june 2009, pp. 622–629.

[4] J. Kong, S. W. Chung, and K. Skadron, "Recent thermal management techniques for microprocessors," *ACM Comput. Surv.*, vol. 44, no. 3, pp. 13:1–13:42, Jun. 2012.

[5] H. Hanson, S. Keckler, S. Ghiasi, K. Rajamani, F. Rawson, and J. Rubio, "Thermal response to DVFS: analysis with an Intel Pentium M," in *ACM/IEEE ISLPED*, aug. 2007, pp. 219–224.

[6] L. Xia, Y. Zhu, J. Yang, J. Ye, and Z. Gu, "Implementing a Thermal-Aware Scheduler in Linux Kernel on a Multi-Core Processor," *The Computer Journal*, vol. 53, no. 7, pp. 895–903, 2010.

[7] J. Yang, X. Zhou, M. Chrobak, Y. Zhang, and L. Jin, "Dynamic Thermal Management through Task Scheduling," in *IEEE ISPASS*, april 2008, pp. 191–201.

[8] G. Liu and G. Quan, "Thermal aware scheduling on an intel desktop computer," in *Southeastcon, 2011 Proceedings of IEEE*, march 2011, pp. 79–84.

[9] R. Mukherjee and S. Memik, "Physical aware frequency selection for dynamic thermal management in multi-core systems," in *IEEE/ACM ICCAD*. ACM Press New York, NY, USA, 2006, pp. 547–552.

[10] R. Rao and S. Vrudhula, "Efficient online computation of core speeds to maximize the throughput of thermally constrained multi-core processors," in *Proc. IEEE/ACM ICCAD*. IEEE Press Piscataway, NJ, USA, 2008, pp. 537–542.

[11] M. Kadin and S. Reda, "Frequency and voltage planning for multi-core processors under thermal constraints," in *Computer Design, 2008. ICCD 2008. IEEE International Conference on*, oct. 2008, pp. 463–470.

[12] ——, "Frequency planning for multi-core processors under thermal constraints," in *Proc. ISLPED '08*. New York, NY, USA: ACM, 2008, pp. 213–216.

[13] M. Kadin, S. Reda, and A. Uht, "Central vs. distributed dynamic thermal management for multi-core processors: which one is better?" in *Proc. 19th ACM Great Lakes Sym. on VLSI*. ACM, 2009, pp. 137–140.

[14] X. Yin, Y. Zhu, L. Xia, J. Ye, T. Huang, Y. Fu, and M. Qiu, "Efficient Implementation of Thermal-Aware Scheduler on a Quad-core Processor," in *2011 IEEE 10th Int. Conf. TrustCom*, nov. 2011, pp. 1076–1082.

[15] G. Liu, M. Fan, and G. Quan, "Neighbor-aware dynamic thermal management for multi-core platform," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2012*, march 2012, pp. 187–192.

[16] C. Bienia, S. Kumar, J. Singh, and K. Li, "The PARSEC benchmark suite: Characterization and architectural implications," in *Proc. 17th Int. Conf. on PACT*. ACM, 2008, pp. 72–81.

[17] F. Mesa-Martinez, M. Brown, J. Nayfach-Battilana, and J. Renau, "Measuring performance, power, and temperature from real processors," in *Proc. Work.on Exp. CS*. ACM New York, NY, USA, 2007.

[18] M. Berktold and T. Tian, "CPU Monitoring With DTS/PECI," 2009, intel whitepaper.

25-27 September 2012, Budapest, Hungary

A Thermal Model for The PCB Structure Including Thermal Contribution of Traces and Vias

Yabin Zhang, Alessio Pennatini, Paolo Emilio Bagnoli
Department of Information Engineering, Universit à di Pisa
Via G. Caruso 16, 56122 Pisa Italy
Tel: +39 050 2217503, Fax: +39 050 2217522, e-mail: ch.zhyb@gmail.com, a.pennatini@iet.unipi.it, p.bagnoli@iet.unipi.it

***Abstract*- The present model is mainly targeted at thermal simulation of multilayer PCB structure, realizing 2-D layout analysis and taking into account temperature-dependent Joule heating in traces as well as thermal conduction in traces and vias. The structure is analyzed by loading 2-D layout bitmaps of each metal layer in the program, and heat sources as well as traces are meshed in square elements due to the requirement of rectangular heat-source elements in the derivation of analytical solution of the temperature in dielectric layers. The main calculation process is in matrix operation and the model is programmed in the Matlab. The model runs in an iterative way by exchanging heat density and temperature distribution between two main programs, electric solver and thermal solver. The electric solver is able to analyze electrical distributions in each trace as long as there are sufficient known conditions of rms electric potentials or rms branch currents for some pads connected to the trace. The thermal solver integrates both the analytical solution of the rectangular homogeneous structure with square heat sources and the numerical analysis of non-uniform heat transfer coefficient (HTC) distribution and thermal conduction in traces and vias. In order to examine the algorithm in the model, the comparison between the model and a CFD software is given for a two-layer structure. Particularly, the model is more efficient in analyzing the structure with LDR (large dimension ratio) between the whole structure and heat sources.**

I. BACKGROUND

Thermal analysis of the PCB structure is important in the reliability analysis of electronic systems due to non-negligible contribution of PCB structure in the heat transfer, especially the thermal conduction contribution by traces and vias. On the other side, the Joule heating effect in traces with high current could also contribute to the overheat failure in traces or devices.

Presently, nearly most commercial CFD software based on FEM have already integrated the function of calculating Joule heat in conductors [1-3]. However when there is a large dimension ratio (LDR) between heat sources and whole object, long time of meshing in triangle-mesh FEM is unavoidable in a normal PC. While, our model has immunity on such problem, because the thermal solver is developed based on an analytical solver, Djoser [4]. Such merit is due to the immunity on the size of heat sources in the analytical solution [4-6]. Another merit of the analytical solution is the capability of solving solution solo in expected points or regions but not necessarily the whole structure. Further explanation can be found in the Section II.

Nevertheless, since heat sources used in the analytical solution should be divided into a number of square elements, a square-mesh method was used in calculating the electrical distribution in traces, so that some intrinsic defects of square mesh are unavoidable. First, irregular shapes cannot be completely reconstructed. The other is the 'average effect' in the quantities solved for each square element, especially there is a little error in the fluid densities.

On the other side, effective orthotropic thermal conductivity of each layer or of each discretized small volume is commonly used in some models [7] or in some software [1,3]. But to the knowledge of the authors, there is no thermal model for the PCB that tries to analyze the thermal conduction in element level in traces, of which the real layout is kept invariant through the whole simulation, and no effective conductivity is generated.

Actually, the triangle-mesh FEM is possible to realize such 'accurate' level analysis as exemplified in the Section III, but probably the huge number of small elements for discretizing traces reduce the running efficiency in thermal analysis, thus limits the analysis in such level. So our motivation is to propose a different way to analyze the thermal performance of a PCB structure at element level even though the model is still not so fast. The proposed model integrates both the immunity of analytical solution on LDR problem and powerful capability of finite-element numerical calculation in analysis of irregular and in-homogeneous structure.

Moreover, when an average HTC couldn't accurately represent the thermal boundary of any surface, non-uniform HTC distribution can be considered in a numerical way, further explanation is given in the next section.

II. BRIEF INTRODUCTION OF THE ALGORITHM

The model was programmed in the Matlab 7. There are three hypotheses used in the algorithm. First, the thermal contact between traces and dielectric layers is assumed perfect and vertical temperature distribution inside slim metal traces is neglected. Second, due to thin thickness of the PCB structure, only heat transfer through the top and bottom surfaces to ambient is taken into account. Third, the influence of temperature on the thermal conductivity of materials is not taken into account.

In the model, square mesh density is equal to the resolution of layout bitmaps. The program recognizes traces, pads and dielectric surface area depending on the different digital data of

different color and the data 'cluster' of each trace or pad is further localized. Then, constant heat sources (ICs and Power devices), and pad potential or rms current flowing in some branch route in each trace can be defined. Also thermal boundary conditions for the board, such as ambient temperature, average HTC or HTC distribution on surfaces can be defined.

The flowchart in Fig.1 shows the brief calculation procedure in the program. This iterative calculation strategy is valid due to uniqueness characteristic of steady state, if existing, under certain thermal boundary conditions. Thermal resistance matrix in analytical solution and numerical heat transfer matrix in traces and vias or based on the HTC distribution are prepared in the Pre (preprocess) Program.

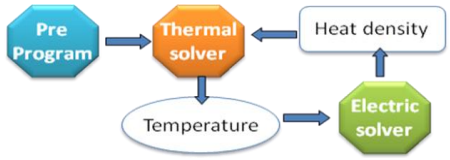


Fig. 1. Brief flowchart of the model.

A. *Electric Solver*

Since traces are numerically divided into a number of square elements, Kirchhoff's current law is applied to calculate the relation between electric potential (EP) and electric conductance (EC) for any trace element and its four adjacent elements as shown in Fig.2. Furthermore, a relation for deriving EP of unknown elements in the trace can be expressed in the form of matrix operation when there are sufficient known conditions:

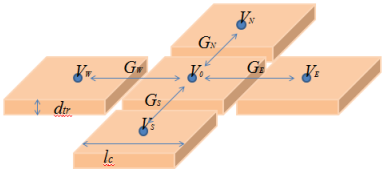


Fig. 2. Schematic of electric conduction between metal elements.

$$M_{EC} * V_{tr} = K_p \Rightarrow V_{tr} = M_{EC}^{-1} * K_p \tag{1}$$

In which, M_{EC} is the EC relation matrix, V_{tr} is the EP array of all unknown elements in a trace, K_p is the array including all terms related to elements with known EP. If the EP of all pads connected with the trace are known, the EP distribution in the trace can be directly derived based on (1).

When the EP of pads are not usually totally known, but some currents flowing in some trace branches are given, those unknown pads EP should be derived first to apply (1). Since the current in one pad (as a branch current in one trace) is determined by EP of both pad elements and trace elements in contact-boundary between each pad and the trace branch as well as EC between boundary elements, such branch current is certainly linearly determined by EP of all the pads connected within one trace:

$$\begin{cases} I_1 = C_{v(1,1)}V_1 + C_{v(2,1)}V_2 + \cdots + C_{v(n,1)}V_n \\ I_2 = C_{v(1,2)}V_1 + C_{v(2,2)}V_2 + \cdots + C_{v(n,2)}V_n \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ I_{n-1} = C_{v(1,n-1)}V_1 + C_{v(2,n-1)}V_2 + \cdots + C_{v(n,n-1)}V_n \\ I_n = I_1 + I_2 + \cdots + I_{n-1} \end{cases} \tag{2}$$

In which, n is the number of pads connected in one trace, I_1~I_n are branch currents in the trace, V_1~V_n are EP of each pad, $C_{v(1,1)}$~$C_{v(n,n-1)}$ are constant coefficients depending on trace structure. In (2), there are $n-1$ effective equations and $2n-1$ variables, so it's necessary and sufficient to derive any group of $n-1$ variables when other n variable are known. In another way, at least one pad EP and other $n-1$ combined conditions of the branch currents or pad EP should be given (there could also be two quantities from one pad as two known conditions). The next mission is to derive all the coefficients in (2).

For example, when the EP of m ($m<n$) pads are unknown but there are m branch currents, a group of equations can be derived:

$$\begin{cases} V_1 = C_{v1} + C_{i(r+1,1)}I_{r+1} + \cdots + C_{i(r+m,1)}I_{r+m} \\ V_2 = C_{v2} + C_{i(r+1,2)}I_{r+1} + \cdots + C_{i(r+m,2)}I_{r+m} \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ V_m = C_{vm} + C_{i(r+1,m)}I_{r+1} + \cdots + C_{i(r+m,m)}I_{r+m} \end{cases} \tag{3}$$

In which, I_{r+1}~I_{r+m} are branch current variables of those pads with known branch currents; V_1~V_m are EP variables of those pads with unknown EP; C_{v1}~C_{vm} are constant coefficients combined by terms with known EP; $C_{i(r+1,1)}$~$C_{i(r+m,m)}$ are coefficients independent of any current or EP.

To solve $m+1$ coefficients in each equation of (3), $m+1$ independent sub-equations should be given, where $m+1$ coefficients are unknown variables but m branch currents and one pad EP are known coefficients (electrical coefficients). That means $m+1$ groups of electrical coefficients should be known. According to (1), $m+1$ combinations of unknown pads EP could be assumed to derive each group of electrical coefficients. After each round of calculation in (1) under each combination of assumed EP, I_{r+1}~I_{r+m} should be calculated according to the relation of the branch currents and EP of elements in contact boundary between the pad and the trace. Then $m+1$ independent sub-equations to solve any group of coefficients in one equation of (3) can be obtained.

Finally, by substituting those known branch currents, all unknown pad EP can be derived from (3). Then one more step of calculation of (1) will give the actual EP distribution in the trace, which is in turn used for deriving the flowing currents in each trace element according to EC between elements. Finally Joule heat in each element can be solved according to the following equation:

$$p_0 = \frac{1}{2G_0} * (I_N^{\,2} + I_S^{\,2} + I_W^{\,2} + I_E^{\,2}) \tag{4}$$

Where I_N, I_S, I_W and I_E represent the currents flowing in four

directions of an element, and G_0 represents the self-EC of the element.

B. Thermal Solver

The thermal solver integrates the analytical solution for rectangular dielectric layers and numerical analysis of thermal conduction in traces and vias to take into account the in-homogeneity in the PCB. In multilayer structure, since the heat exchanged between layers is not known thermal conditions, the solution cannot be derived one layer by one layer independently. Instead, the solution should be solved from a set of equations:

$$\begin{cases} T_{t,1} = R_{t,1,Tin} * Q_{1,Tin} - R_{t,1,Tout} * Q_{1,Tout} + R_{t,1,Bin} * Q_{1,Bin} \\ T_{b,1} = R_{b,1,Tin} * Q_{1,Tin} - R_{b,1,Tout} * Q_{1,Tout} + R_{b,1,Bin} * Q_{1,Bin} \\ T_{b,1} = T_{t,2} \\ Q_{1,Bin} = Q_{2,Tout} \\ \vdots \\ Q_{N'-1,Bin} = Q_{N',Tout} \\ T_{t,N'} = T_{b,N'-1} \\ T_{t,N'} = R_{t,N',Tin} * Q_{N',Tin} - R_{t,N',Tout} * Q_{N',Tout} + R_{t,N',Bin} * Q_{N',Bin} \\ T_{b,N'} = R_{b,N',Tin} * Q_{N',Tin} - R_{b,N',Tout} * Q_{N',Tout} + R_{b,N',Bin} * Q_{N',Bin} \\ T_{b,N'} = T_{t,N'+1} \\ Q_{N',Bin} = Q_{N'+1,Tout} \\ \vdots \\ Q_{N-1,Bin} = Q_{N,Tout} \\ T_{t,N} = T_{b,N-1} \\ T_{t,N} = R_{t,N,Tin} * Q_{N,Tin} - R_{t,N,Tout} * Q_{N,Tout} - R_{t,N,Bout} * Q_{N,Bout} + R_{t,N,Bin} * Q_{N,Bin} \\ T_{b,N} = R_{b,N,Tin} * Q_{N,Tin} - R_{b,N,Tout} * Q_{N,Tout} - R_{b,N,Bout} * Q_{N,Bout} + R_{b,N,Bin} * Q_{N,Bin} \end{cases} \quad (5)$$

Which shows the general analytical solutions of each surface of the dielectric layer in one *N*-layer structure. Subscripts *t* and *b* refer to top and bottom surface in one dielectric layer respectively. Quantity *R* is the thermal resistance matrix for the elements in one surface under certain heat density *Q*, and is related to Fourier series in the analytical solution. Subscripts *Tin* and *Bin* as well as *Tout* and *Bout* refer to quantities respectively related to heat transferred into or out from the top surface and bottom surface in one dielectric layer.

The heat generated on bottom surface $Q_{N',Bin}$ of N'^{th} dielectric layer except the last N^{th} layer is assumed as only contributing to the $N'+1^{th}$ dielectric layer below in the expressions, while the actual thermal contribution of $Q_{N',Bin}$ to N'^{th} dielectric layer is taken into account by the heat $Q_{N'+1,Tout}$ transferred from the $N'+1^{th}$ dielectric layer, which is equal to $Q_{N',Bin}$. Also contact surfaces between two dielectric layers are assumed to be at the same temperature distribution. The Derivation and explanation of the analytical solution can be found in other papers [4,5].

On the other hand, for the surface elements in traces or vias, Q_{Tin} or Q_{Bin} is not in the same distribution of the Joule heat generated. Therefore, thermal conduction in traces and vias was analyzed by applying Kirchhoff's law in the analysis of thermal conduction. There are two main possible thermal conduction in metal layers as shown in Fig.3 and in (6). First is thermal conduction between one central trace element and other adjacent trace elements and via elements. Second thermal conduction is among one central via element and other adjacent trace elements and via elements, also including adjacent via elements in vertical direction. Particularly, the hollow part in a via is approximately considered as thermal insulation, and thermal conduction calculated in the analytical analysis is compensated in the numerical calculation. Each via in one dielectric layer is simply considered as comprised of two main parts in vertical direction, and each part is meshed in a number of square elements. In any two adjacent metal layers, when the via element in the lower metal layer is inside its upper dielectric layer, the distance between two vertical via elements is $d_i/2$ due to the symmetrical placement. When the via element in the lower metal layer is also inside its lower dielectric layer, the distance between two vertical via elements is d_i.

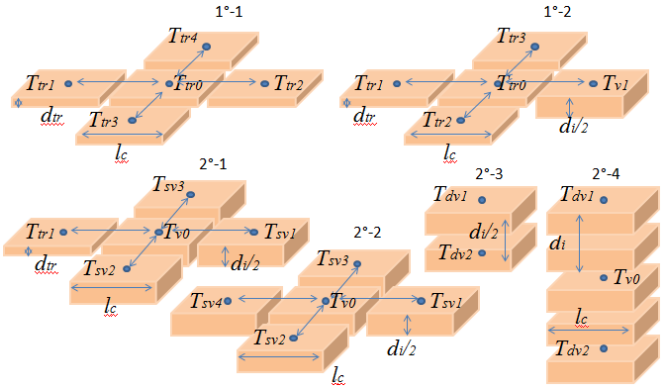


Fig. 3. Schematic of possible forms of thermal conduction between metal elements.

$$\begin{cases} k_m d_{tr}/l_c^2 (n_{tr}T_{tr0} - T_{tr1}...) + 2k_m d_{tr}/l_c^2 (n_v T_{tr0} - T_{v1}...) = p_{tr0} - q_{tr0} \\ 2k_m d_{tr}/l_c^2 (n_{tr}T_{v0} - T_{tr1}...) + (k_m - k_i)d_i/2l_c^2 (n_{sv}T_{v0} - T_{sv1}...) \\ + (k_m - k_i)/d_i (n_{dv}T_{v0} - T_{dv1}...) + 2(k_m - k_i)/d_i (n_{iv}T_{v0} - T_{iv1}...)] = p_{v0} - q_{v0} \\ \text{or for cosidering thermal insulation in the hollow via:} \\ -k_i d_i/2l_c^2 (n_{sv}T_{v0} - T_{sv1}...) + -k_i/d_i (n_{dv}T_{v0} - T_{dv1}...) \\ + -2k_i/d_i (n_{iv}T_{v0} - T_{iv1}...) = p_{v0} - q_{v0} \end{cases} \quad (6)$$

In (6), *k* refers to thermal conductivity, *d* refers to thickness, l_c refers to side length of one element, *n* refers to the number of adjacent metal elements, *p* refers to joule heat density generated in the central element and *q* is the density of heat transferred from the central element to the dielectric layer. So the heat transferred from the center element to other adjacent metal elements is expressed as *p*-*q*. Subscript *m* relates the quantity to general metal element, *tr* for trace element, *i* for dielectric layer, *v* for via element, *sv* for adjacent via elements in same metal layer, *dv* for adjacent via elements in different dielectric layers, *iv* for adjacent via elements in the same dielectric layer, number *0* for the central metal element.

Certainly, the ICs and power devices are major heat sources on the PCB. So far, the model is able to consider the ICs or devices as surface heat elements, while the heat transferred to ambient through package plastic layer on the top is neglected.

When one average HTC couldn't represent the heat transfer from the PCB surface to the ambient, but an HTC distribution is given, the HTC thermal boundary conditions seem impossible to be taken into account in the analytical solution. Instead, numerical analysis of such condition can be realized according to the linear relation between the heat transferred and temperature difference. For example, if there are two HTC distributions on both surfaces in a two-layer structure, the first equation in (6) expressing the thermal conduction among traces elements change to the following forms:

$$\begin{cases} k_m d_{tr} / l_c^2 \,(n_{tr}T_0 - T_{tr1}...) + \cdots + h_{t,m}T_0 = p_m - q_m \\ k_m d_{tr} / l_c^2 \,(n_{tr}T_0 - T_{tr1}...) + \cdots + h_{b,m}T_0 = p_m - q_m \\ or \quad k_m d_{tr} / l_c^2 \,(n_{tr}T_0 - T_{tr1}...) + \cdots = p_m - q_m \end{cases} \tag{7}$$

Where $h_{t,m}$ or $h_{b,m}$ refers to corresponding HTC of one metal element in the top or bottom surface. The third sub-equation refers to any metal element covered by ICs, thus there is no direct heat exchange between the element and the ambient. In the similar way, for other dielectric surface elements, the heat density may be expressed in one of the following equations:

$$\{q_i = -h_i T_0 \quad \text{or} \quad q_i = q_c \tag{8}$$

Where q_i is the surface heat density transferred to the dielectric layer. When the dielectric surface element is covered by some IC, q_i is equal to the density of heat transferred from the top IC through that square region, q_c. Or else, q_i is equal to the heat exchange to the ambient under specified HTC, h_i. So the matrix equations representing both the analytical solution and numerical heat transfer for a two-layer structure with HTC distributions on external surfaces can be given as:

$$\begin{cases} T_{m1} = R_{\theta,m1,TinM}Q_{TinM} + R_{\eta,m1,BinM}Q_{BinM} + R_{\varphi,m1,TIC}Q_{TIC} + R_{\zeta,m1,BIC}Q_{BIC} \\ T_{m2} = R_{\theta,m2,TinM}Q_{TinM} + R_{\eta,m2,BinM}Q_{BinM} + R_{\varphi,m2,TIC}Q_{TIC} + R_{\zeta,m2,BIC}Q_{BIC} \\ T_{i1} = R_{\theta,i1,TinM}Q_{TinM} + R_{\eta,i1,BinM}Q_{BinM} + R_{\varphi,i1,TIC}Q_{TIC} + R_{\zeta,i1,BIC}Q_{BIC} \\ T_{i2} = R_{\theta,i2,TinM}Q_{TinM} + R_{\eta,i2,BinM}Q_{BinM} + R_{\varphi,i2,TIC}Q_{TIC} + R_{\zeta,i2,BIC}Q_{BIC} \\ A_{mv1}T_{m1} + A_{v2}T_{m2} = P_{m1} - Q_{TinM} \\ A_{v1}T_{m1} + A_{mv2}T_{m2} = P_{m2} - Q_{BinM} \\ Q_{TIC} = h_{t,i}T_{i1} + Q_{t,c} \\ Q_{BIC} = h_{b,i}T_{i2} + Q_{b,c} \end{cases} \tag{9}$$

$$\begin{pmatrix} h_{t,i,1} & 0 & 0 & \cdots & 0 \\ 0 & h_{t,i,2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & h_{t,i,N_i} \end{pmatrix} \text{ and } \begin{pmatrix} h_{b,i,1} & 0 & 0 & \cdots & 0 \\ 0 & h_{b,i,2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & h_{b,i,N_i} \end{pmatrix} \tag{10}$$

In which, quantities with subscripts θ, η, φ or ζ refer to thermal resistance under surface heat Q_{TinM}, Q_{BinM}, Q_{TIC} or Q_{BIC} respectively; subscript M relates the quantities to surface metal elements, subscript *TIC* and *BIC* relate the quantities to surface dielectric elements covered by the ICs or devices on top surface and bottom surface respectively. $Q_{t,c}$ or $Q_{b,c}$ is the known density of constant heat transferred from ICs or devices to the covered dielectric surface; P_{m1} and P_{m2} are densities of heat generated in metal elements including Joule heat and heat transferred from upper ICs if covered; T_{i1} and T_{i2} refer to temperature of surface dielectric elements; A_{mv1}, A_{mv2}, A_{v1} *and* A_{v2} are the matrix including those numerical coefficients in (6) and (7). Particularly, A_{v1} *and* A_{v2} are the complementary matrix of coefficients to relate those via elements in the bottom layer with those via elements in the top layer. After several steps of derivation, all unknown heat densities and temperature can be solved. Furthermore, chip temperature can be calculated based on the junction-case thermal resistance given in the datasheet of a device.

III. Examination of the Algorithm in a Two-Layer Structure

The model was tested by comparing the results of a simple two-layer board as shown in Fig.5 with the CAD model built in the COMSOL Multiphysics. The dimension of the board is 50mm×37.5mm×1.66mm with trace thickness 35μm. The resolution of the layout is in 360×270. The ambient temperature is assumed at 20 ℃, and a uniform HTC is defined in order to examine the core algorithm with consideration of thermal contribution in traces and vias.

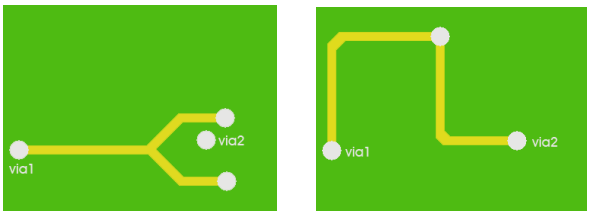


Fig. 4. Layout of the two-layer structure.

Without consideration of ICs and devices, the solution equations for this kind of layer structure can be simply expressed as:

$$\begin{cases} T_{m1} = R_{\theta,m1,Tin}Q_{Tin} + R_{\eta,m1,Bin}Q_{Bin} \\ T_{m2} = R_{\theta,m2,Tin}Q_{Tin} + R_{\eta,m2,Bin}Q_{Bin} \\ A_{mv1} * T_{m1} + A_{v2} * T_{m2} = P_{m1} - Q_{Tin} \\ A_{v1} * T_{m1} + A_{mv2} * T_{m2} = P_{m2} - Q_{Bin} \\ T_{S1} = R_{\theta,S1,Tin}Q_{Tin} + R_{\eta,S1,Bin}Q_{Bin} \\ T_{S2} = R_{\theta,S2,Tin}Q_{Tin} + R_{\eta,S2,Bin}Q_{Bin} \end{cases} \tag{11}$$

In which, after deriving Q_{Tin} and Q_{Bin} from the first four equations, temperature distribution T_{S1} and T_{S2} of two surfaces under any dedicated resolution can be further given. Therefore, during the iteration between electric solver and thermal solver, only the temperature and heat transferred from the metal elements have to be solved.

There are 32761 (181*181) sets of eigenvalues β, μ and γ

in the thermal analytical solution. The resolution of temperature of traces is in 360×270 during iteration. But when steady-state is obtained, to obtain a more refined result and more accurate comparison with the COMSOL, the resolution of the temperature distribution was increased to 480×360. One arbitrary set of electrical parameters as given in Table I were defined to the pads and traces.

TABLE I
ELECTRICAL PARAMETERS DEFINED IN THE TWO-LAYER MODEL

Layer1	Via 1	Pad 1.2	Pad 1.3	Via 2
Value	IP: 6A	EP:0.04V	EP: 0V	EP: 0.01
Layer2	Via 1	Pad 2.2		Via 2
Value	IP: -6A	EP: 0.1555V		EP: 0.01

Where IP represents current flowing in the trace-branch connected with corresponding pad.

In the COMSOL, we loaded 'shell, conductive media DC' model and 'general heat transfer' model to implement the electro-thermal collaborative simulation. A CAD model was built according to the layout. The traces were recognized as thin 'shell' on the surface of the PCB, which is similar to the assumption of neglecting the vertical thermal diffusion in the traces in our model. In the COMSOL, all traces in the board were meshed in 4660 triangles, and the PCB was meshed in 48520 tetrahedral blocks and the total number of degrees of freedom is 88251. The simulation was running on a computer with Core-2 Quad 9400 CPU and 8G memory.

The output results are shown in following figures with the left from our model and the right from the COMSOL:

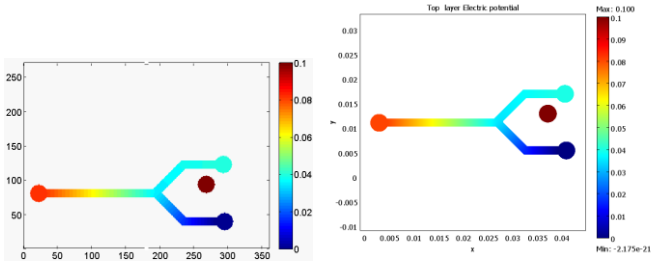


Fig. 5. EP simulation results of first layer

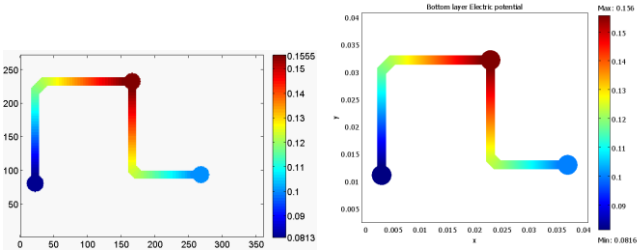


Fig. 6. EP simulation results of second layer

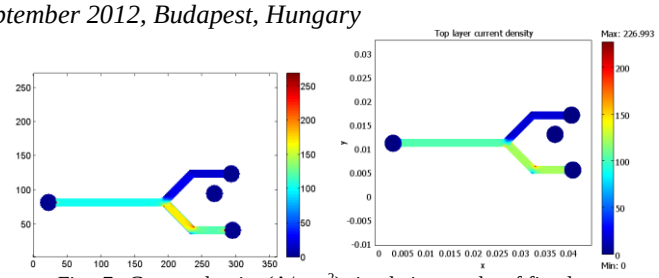


Fig. 7. Current density (A/mm^2) simulation results of first layer

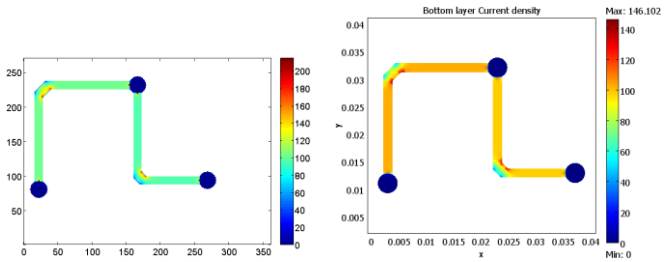


Fig. 8. Current density (A/mm^2) simulation results of second layer

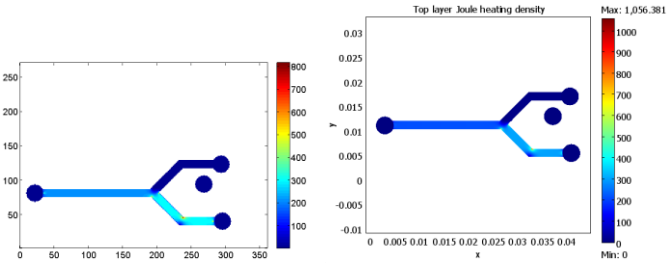


Fig. 9. Power density (W/cm^3) simulation results of first layer

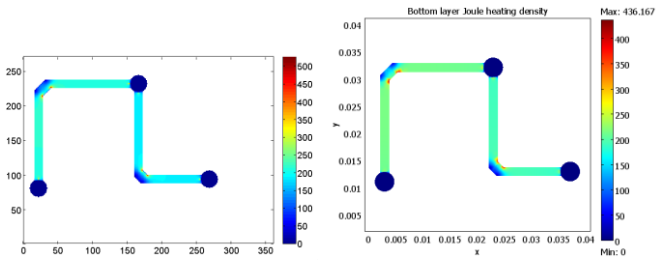


Fig. 10. Power density (W/cm^3) simulation results of second layer

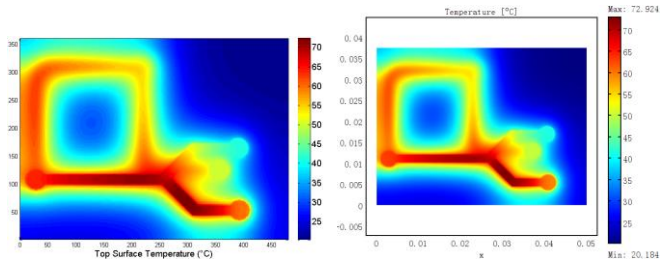


Fig. 11. Temperature simulation results of first layer

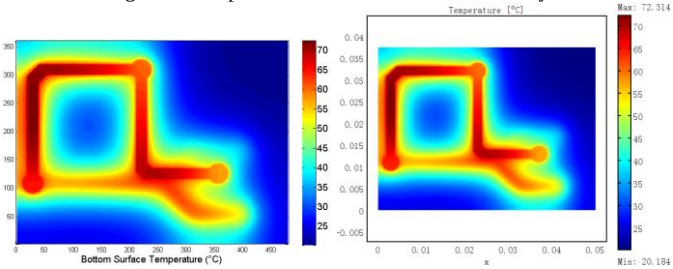


Fig. 12. Temperature simulation results of second layer

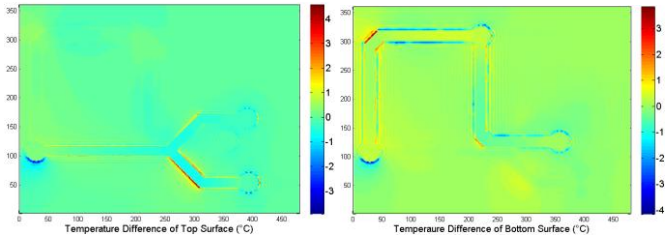


Fig. 13. Difference of Temperature results between two programs

So the EP results from two programs seem quite consistent to each other. But there are a few differences in the maximum values of density results despite the color scales are similar. However, these differences seem not influencing the consistence of results of temperature. Because with large number of square mesh elements, the heating power totally generated in a relatively small area in traces is close to the real case.

According to the results of difference in Fig.13, the average accuracy of the temperature results is more than 99%. It took about 1 minute to reach the thermal steady state in our model. But about 20 minutes were required in calculating the temperature of the whole surface, which is much longer than 1 minute, the simulation time in the COMSOL. However, in the model only 2-D layout bitmaps should be prepared, which can be fast obtained from the PCB EDA software, but not building a CAD model as required in the COMSOL.

IV. LDR MODEL

A simple LDR model was built to show the LDR immunity merit of the model. Suppose there are 25 square spot heat sources of 0.01W each uniformly distributed on the top surface of a rectangular FR-4 epoxy board, of which the dimension is in 5cm×5cm×1.66mm and thermal conductivity is set to 0.3W/mK. The surface is meshed in 600×600 resolution, so that the LDR is 600:1. If the board is attached to a heat sink at ambient temperature on the bottom and is thermal insulated on the top surface, the steady-state temperature rise in the heating spot region (24×24 pixel elements) can be given:

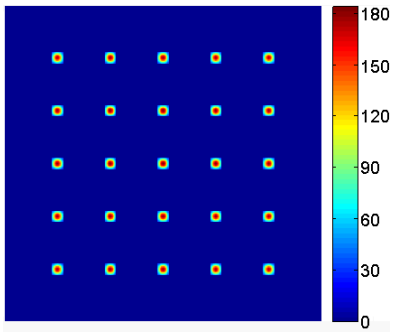


Fig. 14. Surface Temperature rise of a LDR structure.

Calculation time for total 14400 spots is less than 1 minute. While, the same structure was not able to be successfully meshed in the COMSOL due to out-of-memory problem.

V. CONCLUSIONS

The model is able to realize the thermal simulation of the multilayer PCB structure with consideration of the thermal contribution of traces and vias. Fast analysis of electrical distribution in traces with multiple pads were realized in the electric solver. Integration of numerical calculation with the analytical solution helps overcome the difficulty in the modeling due to the in-homogeneity of the PCB structure. The algorithm has been validated through the comparison with a commercial CFD software, despite of a few errors occurred in fluid densities due to the intrinsic defects of the square mesh.

The main potential of this model is from the capability of modeling the LDR structure. When there are small ICs or power devices mounted on a large PCB, LDR could be a serious problem for the commercial CFD software, because structure meshing could be failed due to the huge number of elements to be meshed. Certainly, simulation time of the model should be far more improved from the programming point of view.

ACKNOWLEDGMENT

Thanks should be given to our colleagues in the Laboratory of Antennas, Microwave and EMC for their friendly and patient help during the development of the model.

REFERENCES

[1] Dogruoz M.B., Nagulapally M.K., "Effects of Trace Layers and Joule Heating on the Temperature Distribution of Printed Circuit Boards: A Computational Study", *Journal of Thermal Science and Engineering Applications*, vol.1, pp. 022003-1 - 022003-10, 2009.

[2] Heat Transfer Module Model library, COMSOL AB.2008.

[3] B. Blackmore, "Validation and Sensitivity Analysis of an Image Processing Technique to Derive Thermal Conductivity Variation within a Printed Circuit Board", 25th IEEE SEMI-THERM Symposium2009, pp. 76-86.

[4] F. Stefani, P.E.Bagnoli, M. Beccai, "The Djoser Thermal Simulator For Electronic Devices And Systems: Didactical Features For Improving Reliability Culture In Academic And Industrial Environments", ICERI Int. Conf. of Education, Research and Innovation, Madrid, 2008, pp. 10.

[5] P.E.Bagnoli, Yabin Zhang, "Electro-thermal simulation of metal interconnections under high current flow", *Microelectronics Reliability*, vol.50, pp. 1672-1677, Sep./Nov. 2010.

[6] Yuri S. Muzychka, "Influence coefficient method for calculating discrete heat source temperature on finite convectively cooled substrates", *IEEE Trans. Comp. Packag. Technol.vol*. 29. pp. 636-643. September 2006.

[7] Shankaran G.V., Singh R.K., "Election of Appropriate Thermal Model for Printed Circuit Boards in CFD Analysis", 10th Intersociety Conf. Thermal and Thermomechanical Phenomena in Electronics Systems2006, pp. 234-242.

25-27 September 2012, Budapest, Hungary

Miniaturized Frictionless Fan Concept for Thermal Management of Electronic

Schacht, Ralph[1,2,4], Hausdorf, Alexander[2], Wunderle, Bernhard [1,3,4], Michel, Bernd [3,4]
[1]Fraunhofer ENAS, Chemnitz, [2]Lausitz University of Applied Sciences, Senftenberg,
[3]Technical University Chemnitz,, [4]Joint Lab Berlin
Technologie Campus 3, Chemnitz, D-01968
E-mail: ralph.schacht@enas.fraunhofer.de

***Abstract* - This paper introduces a miniaturized frictionless fan concept, which is similar to a piezo-electric driven fan principle and that can fits into a size of a matchbox. This type of fan has been employed for the enhancement of heat transfer by increasing the fluid circulation in regions which are otherwise stagnant.**

The introduced fan is based on a flexible blade whose vibration is driven by means of a magnetic field.

As opposed to the piezo-electric fan, the blade material can be based on a polymer or steel blade. A static mechanical FE-analysis has shown good potential for a reliable, frictionless (i.e. highly reliable) fan compared to the piezo-electric one or to small axial or radial fans.

A flow rate of ~11 l/min could be reached. This is $^1/_3$ below to the compared piezo-electric fan. But the advantage of the introduced fan compared to the piezo-electric fan is its low driving voltage of 2.5 V instead of 120 V.

The paper will describe the principle and first experimental results.

I. INTRODUCTION

Reliable miniaturized fans are of interest as they can enhance heat transfer by increasing the fluid circulation in regions which are otherwise stagnant.

Available market-going miniaturized axial and radial fans using ball bearings have a low mean time to failure, because their movement entails friction. Frictionless fans like piezo fans or synthetic jet drivers use a blade or a membrane. In case of the piezo blade structure [1, 2] the mean time to failure is low and 6 to 7 times more expensive. The synthetic jet fan has an acceptable mean time to failure, but its weight is heavy, because of the used magnetic coil [3][4-7].

For that, a reliable fan with a high mean time to failure, a low weight, low power consumption, small outline, no noise, low cost and an acceptable flow rate is necessary. The paper describes the results of a first concept study of a magnetic field driven, frictionless blade fan and its working principle. The flow rate, the heat transfer coefficient as well as the influence of surrounding walls around the blade fan will be examined. At least, three prototypes and its properties as well as a blade shape study to enhance the performance will finish this paper.

II. FRICTIONLESS BLADE FAN PRINCIPLE

The idea is to have a reliable clamped blade. Two magnetic coils stimulate in 180° phase shift the PVC-U blade with a steel clip in position of the coils. The steel clip is necessary for the magnetic excitation. As blade material a PVC-U is used.

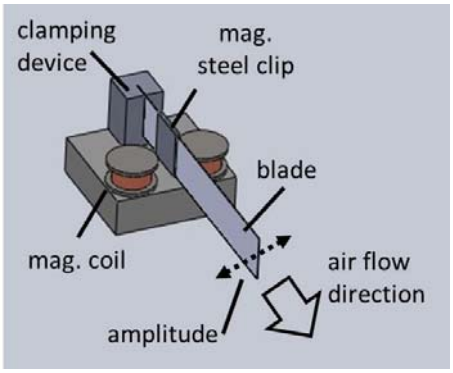


Fig.1 Blade fan assembly.

Figure 1 shows a schematic of the blade fan principle. An air flow will be generated in direction of the blade tip.

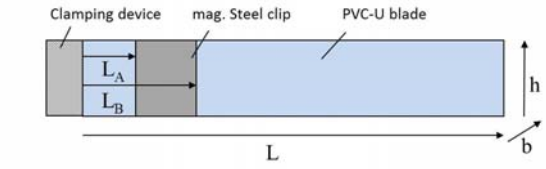


Fig.2 Blade dimensions and steel clip position (side view).

Table 2 states the blade dimensions and the position of the steel clip.

Table1
BLADE DIMENSIONS (SEE FIGURE 2)

Parameter	value[mm]
L_A	9
L_B	20
L	35…90
b	0.2
h	12

The flow rate is a function of the amplitude y and the operated natural frequency. The blade fan creates mainly air flow near its tip. The fan can be operating in its 1th and 2nd natural frequency, which is dependent of the blade length. No blade movement was observed at the 3rd natural frequency. At a blade length over ~ 45 mm, the second natural frequency shows a much better air flow rate then the 1st one.

III. DETERMINATION OF AIR FLOW RATE

A schematic of the used set-up to determine the fan performance is shown in Fig. 3a. The used set-up is referring to the AMSE/ANSI Standard 210-07 [9], because a recommended set-up was not available yet. Therefore the pressure

determination has to be done later. In a first step, only the flow rate could be measured with a hot wire-anemometer from KIMO CTV210 [8] (working ranges: 0 to 1 m/s and 0 to 30 m/s).

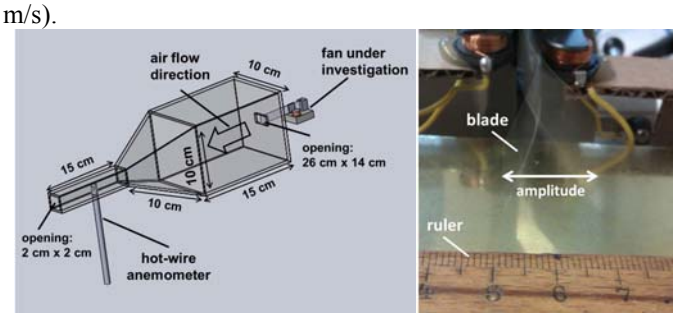


Fig.3 a) Measurement set-up b) Amplitude measurement [11]

The amplitude measurement was done as shown in figure 3b [10].

A maximum flow rate of $\mathring{V}$ ~ 10.5 l/min could observed for a blade length of L = 70 mm. A maximum amplitude y could be noticed for blade length of L = 80 mm (figure 4).

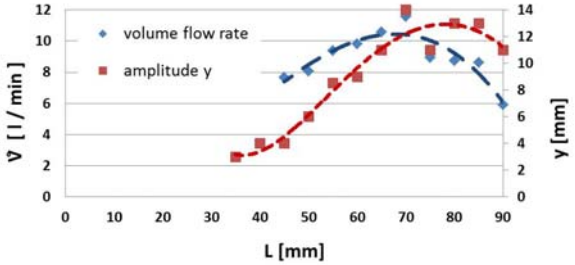


Fig. 4 Flow rate $\mathring{V}$ and amplitude y as function of blade length L.

A. *Validation with a piezo fan*

A comparison between the introduced blade fan and an in [2] investigated piezo fan is given in figure 5.

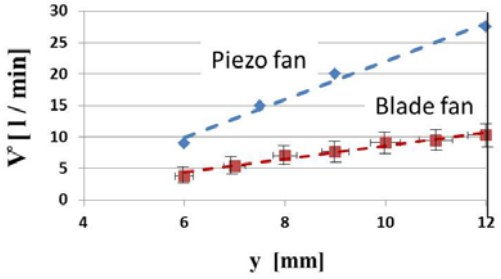


Fig.5 Comparison of the flow rate between piezo [2] and new blade fan as function of the amplitude y.

Because the piezo fan was characterized in the AMSE/ANSI Standard 210.07 recommend flow box, the given validation in figure 5 gives only a tendency. Considering this, it can ascertain that the piezo fan has a six times higher flow rate slope with increasing amplitude. One reason could be that the piezo fan was operate in 1st natural frequency mode instead of the blade fan, which was operate in 2nd natural frequency mode.

IV. INFLUENCE OF CABINET

The main idea is to protect the blade fan against damage with a cabinet. Therefore further on the influences of surrounding walls will be examined.

Supposing the amplitude y limits the air inlet into the cabinet, three different cabinet sizes, given in table 2, were taken under investigation.

Figure 6 presents the measurements results. To verify the flow rate $\mathring{V}$, the characteristic for no encapsulation is also given. It can be seen, that the flow rate $\mathring{V}$ for cabinet 1 follows the behavior with no encapsulation. For cabinet 2 the flow rate $\mathring{V}$ reaches a maximum plateau at amplitude y = 9 mm. For cabinet 3, only two measurement data could be determined, because below y = 8 mm and above y = 9 mm the flow rate was too low for the anemometer.

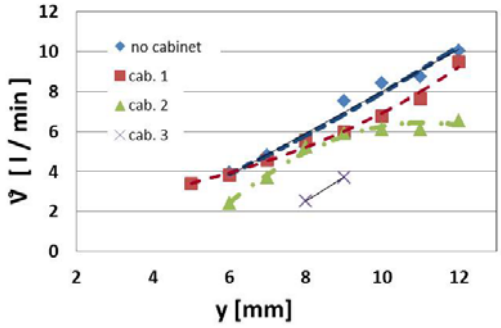
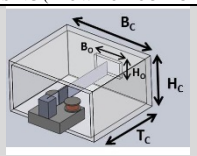


Fig. 6 Flow rate as function of the amplitude for different cabinet sizes (see table 2). P_V = 0.5 W, blade length L = 70 mm.

For cabinet 3, only two measurement data could be determined, because below y = 8 mm and above y = 9 mm the flow rate was too low for the anemometer.

Table 2
CABINET SIZES TO DETERMINING THE INFLUENCE OF THE SURROUNDING BLADE FAN HOUSING (FLOW BOX OUTPUT OPENING DIMENSIONS: B_O = 26 mm and H_O = 14 mm)

	**Cab. 1**	**Cab. 2**	**Cab. 3**
Height (H_C)	40 mm	25 mm	20 mm
Width (B_C)	80 mm	50 mm	40 mm
Depth (T_C)	70 mm	70 mm	70 mm

As a design rule for a closed side wall cabinet it could found that a minimum of a factor of 6 of the maximum amplitude y should be used to get a maximum flow rate.

A. *Cabinet openings*

In case to force a miniaturized fan, the influence of air inlets at the side or top walls was also taken under investigation.

Therefore, the vertical and horizontal side walls were opened and the flow rate $\mathring{V}$ was observed during increasing the opening length O (figure 7).

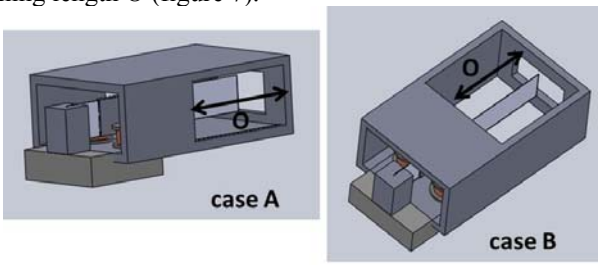


Fig. 7 Openings at vertical side walls (case A) and horizontal side walls (case B) using cabinet size 3.

Figure 8 shows the influence on the flow rate as function of the opening length O at vertical side walls (case A) and horizontal side walls (case B).

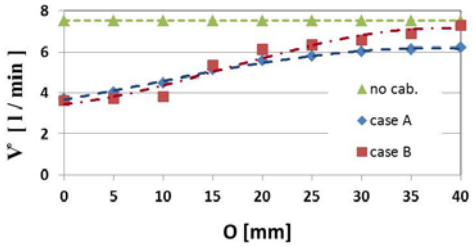


Fig. 8 Flow rate as function of the opening length O at vertical side walls (case A) and horizontal side walls (case B) using cabinet size 3 (amplitude y = 9 mm).

While increasing the openings length O, the flow rate shows an exponential increase. It can also be seen that the openings at the horizontal side walls (case B) lead more air into the cabinet as the vertical side wall openings (case A).

V. BLADE FAN CHARACTERISCS

A. *Determination of heat transfer coefficient*

The heat transfer coefficient α can be determined by equation 1,

$$\alpha = \frac{P_{V,heat}}{A \cdot (T_S - T_{amb})} \qquad (1)$$

with $P_{V,heat}$ [W] as power loss in a heating resistor and T_S [K] and T_{amb} [K] as surface and ambient temperature.

To have a small outlined heating source surface and an easy to control heat und temperature measurement, a thermal test chip (TTC), existing of a 3 x 3 thermal test die (TTD) assembly, was used. The TTC is mounted in flip-chip technology on a FR-4 board (figure 9a).

Figure 9b shows a schematic of such a TTD with its implanted heat resistance and diode.

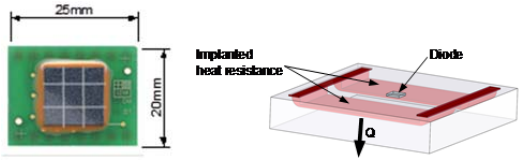


Fig. 9 Thermal test chip (TTC) assembly (a) and Structure (b).

With the implanted diode the chip temperature T_J can be measured because of the temperature-depending physical effect of the diode forward voltage U_F, during a constant diode current I_F, is used (equation 2).

$$T_J \sim U_{F,Diode} = \frac{K\,T}{e} \ln\left(\frac{I_{F,Diode}}{I_S}\right) @\ I_{F,Diode} = const \qquad (2)$$

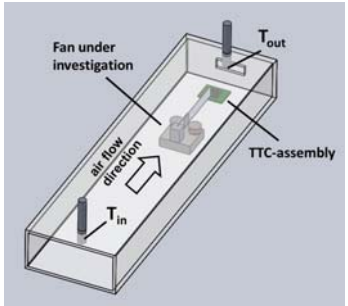


Fig. 10 Test set-up to determine the transfer coefficient and the self-heating of the blade fan [12].

To determine the heat transfer coefficient it is in first approximation assumed that T_S is close to T_J , because of the low thermal resistance between junction and surface of the silicon chip.

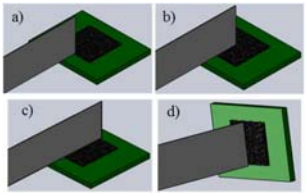


Fig.11 Blade positions to determine the heat transfer coefficient using a TTC.

To find the optimum position for the blade fan achieving the best heat transfer coefficient, the blade was positioned with no blade overlap (a), half overlap (b), full overlap (c) and a vertical position to the TTC surface (d), see figure 11.

Table 3
BLADE DIMESIONS AND MEASURMENT PROPERTIES AS WELL AS TTC PROPERTIES; WHILE DETERMINING THE HEAT TRANSHER COEFFICIENT

Blade length	70 mm
Blade height	12 mm
Amplitude y	9 mm
Frequency	16.92 Hz
R_{TTC}	11 □
A_{TTC}	139.24 mm²

The measurements were done using the set-up given in figure 10. The blade and TTC properties are given in table 3.

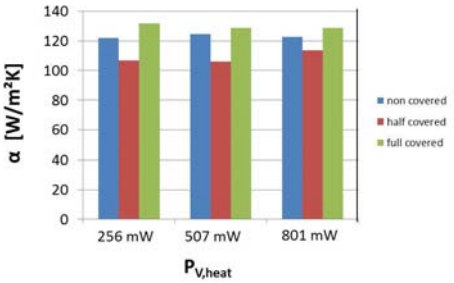


Fig. 12 Measured heat transfer coefficient α for different blade position (see figure 14) and different TTC power dissipation.

The results of this examination are given in figure 12. It was found that there is no dependence on the heat dissipation $P_{V,heat}$ but on the blade position with respect to the heating surface of the TTC. Best results are reached in the vertical position. Measurement results in [11] show for the half covered position a ~ 25 % lower heat transfer coefficient. For the non and fully covered positions the results are diverge.

B. *Static mechanical stress*

To have a numerical estimation about the reliability of the PVC-U blade (Polyvinylchlorid-strong in no stretched form), a static mechanical stress FE-Simulation was done (figure 13).

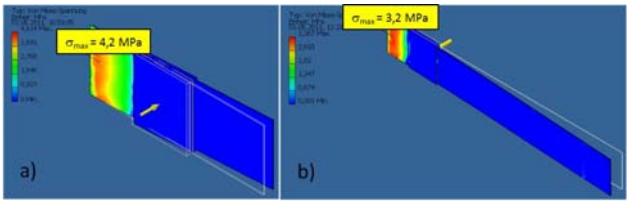


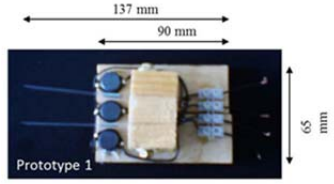
Fig.13 Static mechanical FE-simulation (Mises stress) for blade length L a) 35 mm and b) 90 mm.
The amplitude was in case a) y = 3 mm and b) y = 11 mm.

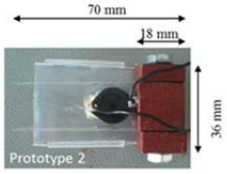
The FE-simulation were done for a short (35 mm) and a long (90 mm) blade length. The results allow a first statement, that the maximum stress in the clamping region is about 5-6 times smaller than the maximum fracture stress of PVC-U (σ_{break} = 20 MPa) [12].

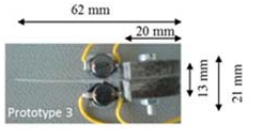
VI. MINIATURIZATION

A. *Prototypes*

In case of having a small outline of the blade fan, a study for three prototypes was done. The investigated outline dimensions and a property list for the blade fan prototypes. are given in figure 14. It can be seen, that there is a linear behavior between the power dissipation and flow rate, while reducing the size





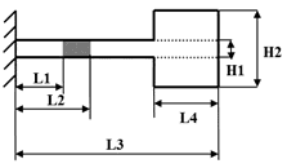


	Proto-type 1	Proto-type 2	Proto-type 3
Nr. coils	3	1	2
Nr. blades	2	2	1
P_V [W]	1	0.5	0.192
Freq. f [Hz]	17.8	33.38	48.20
$\dot{V}$ [l / min]	9.675	6.3	2.92

Fig.14 Realized prototype blade fans

B. *Optimization - Rectangle front shape blade*

To optimize the flow rate $\dot{V}$ a design study of the blade shape was done.



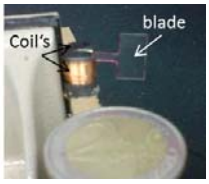


Fig. 15 a) Design of rectangle blade shape (steel clip position: L_1 = 4 mm, L_2 = 8 mm). b) Set-up for rectangle blade shape investigation.

Figure 15a shows a schematic of the changed shape and its parameter, which were taken under investigation. Figure 15b shows a picture of the test set-up, with which it is possible, to have a blade fan of a size around 26 mm in diameter (2 € coin).

Under investigation were the height and length of the rectangle blade form at the front of the blade (L_4, H_2) with respect to the blade height at the clamping (H_1) and the overall length (L_3).

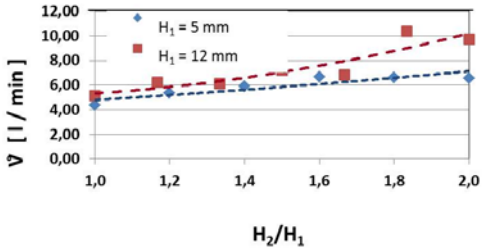


Fig. 16 Flow rate $\dot{V}$ as function of the aspect ratio of the blade shape H_2/H_1, (L_3 = 45 mm, L_4 = 15 mm).

Figure 16 shows the dependence between the height H_1 and the aspect ratio of the blade shape ratio H_2/H_1.

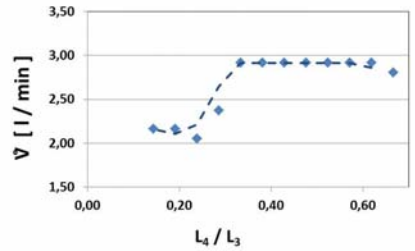


Fig. 17 Flow rate as function of the aspect ratio of the front blade shape L_4/L_3 (H_1 = 2 mm, H_2 = 12 mm, L_3 = 31 mm).

It can be seen that the flow rate $\dot{V}$ increases non linearly, which become stronger with an increasing height H_1.

Also determined was the aspect ratio L_4/L_3 of the blade shape (figure 17). It could observed, that there is a plateau between L_4/L_3 = 0.35 to 0.6.

VII. CONCLUSIONS

A new type of frictionless working fan was introduced and its principle function as well as the first experimental results was highlighted. The fan shows a good potential for reliability and for miniaturization and has the capability to enhance heat transfer by increasing the fluid circulation in regions which are otherwise stagnant. In a first static mechanical FE-Simulation it could found, that the maximum stress is 6-7 times smaller to fracture stress. The flow rate can be nearly doubled, while using air wall inlets in the surrounding cabinet walls. The flow rate can also be enhanced using a rectangle shape at the front of the blade. Further work has to be done in characterizing its pressure behavior and the influence of the self-heating and the magnetic field onto the electronic components as well as long time studies with respect to reliability.

REFRENCES

[1] Mark Kimber, Kazuhiko Suzuki, Nobutaka Kitsunai, Kenichi Seki, and Suresh V. Garimella, “Quantification of Piezoelectric Fan Flow Rate Performance and Experimental Identification of Installation Effects” 978-1-4244-1701-8/08/$25.00 ©2008 IEEE.

[2] Mark Kimber, Kazuhiko Suzuki, Nobutaka Kitsunai, Kenichi Seki, and Suresh V. Garimella, “Pressure and Flow Rate Performance of Piezoelectric Fans” IEEE TRANSACTIONS ON COMPONENTS AND PACKAGING TECHNOLOGIES, VOL. 32, NO. 4, DECEMBER 2009.

[3] Kercher, Lee, Brand, Allen, and Glezer, "Microjet Cooling Devices for thermal Management of Electronics," IEEE Transactions on Components and Packaging Technologies, vol. Vol. 26, pp. 359-366, Juni 2003

[4] Axial fan, www.sunon.com

[5] Radial fan, www. Titan-cd.com

[6] Piezo fan, www.piezo.com (Part Number: RFN1-LV-02)

[7] Membran fan, SynJet® Low Profile Cooler. Co.Nuventix (www.nuventix.com)

[8] Data Sheet Anemometer http://www.sentronic-instruments.com

[9] Ju Hyun Yoo, Jae Il Hong, Wenwn Cao,“Piezoelectric ceramic bimorph coupled to thin metal plate as cooling fan for electronic devices” 0924-4247/00/$ © 2000 Elsevier Science S.A.

[10] Laboratory Methods of Testing Fans for Certified Aerodynamic Performance Rating, ANSI/AMCA Standard 210-07.

[10] Tolga Acikalin, Sydney M. Wait, S V. Garimella, Arvind Raman, “Experimental Investigation of the Thermal Performance of Piezoelectric Fans” Heat Transfer Engineering, 25(1): 4-14, 2004

[12] Carlowitz, *Polyvinylchlorid-hart-Folien (Hart-PVC-Folien)*, München: Carl Hanser, 1995, p. 483.

25-27 September 2012, Budapest, Hungary

Dual Cold Plate System with Active Temperature and Heat Flux Control

Marcin Janicki, Zbigniew Kulesza, Tomasz Torzewicz, Andrzej Napieralski
Department of Electronics and Computer Science, Technical University of Lodz
Wolczanska 221/223, 90-924 Lodz, Poland
Tel: +48 42 6312654, fax: +48 42 6360327; e-mail: janicki@dmcs.pl

***Abstract*-This paper presents in detail a practical realization of the measurement system allowing active control of cooling conditions during thermal characterization of electronic devices and their packages in the dual cold plate cooling assembly. The cooling conditions are adjusted based on temperature and heat flux measurements owing to the use of Peltier thermo-electric modules controlled by the PID algorithm.**

I. INTRODUCTION

The proper thermal characterization of electronic systems often requires numerous transient temperature measurements which should be performed in precisely controlled cooling conditions. Such measurements can be taken using the Dual Cold Plate (DCP) cooling assembly which stabilizes package temperature during measurements. However, the cold plates might not be sufficient to keep constant package temperature.

Therefore, this paper presents an enhanced version of the DCP cooling system realizing in practice the concept of the variable thermal resistance. This concept is not new and was already presented in literature [1]-[3], but the current version possesses some novel features, which were not discussed before. Namely, this new realization allows the registration of thermal transients not only in precisely controlled cooling conditions but also with known contact resistance values. This was possible owing to the application of a tensometer bridge and a special squeezing force adjusting weights. Such enhancements might be especially useful for the transient dual interface measurements described in the recently published JEDEC standard [4].

The following section of this paper presents briefly the design and the calibration of the cooling condition control circuit. Particular attention is paid to the presentation of the heat flux control unit. The operation of the system is verified in practice based on a simple example where the heat flux from a diode package is stabilized during measurements taken with variable power dissipation in the device.

II. SYSTEM DESIGN

The measurement system was developed as an extension to the standard DCP measurement stand. This system allows the registration of thermal transients with very high temporal resolution in various cooling conditions and with different values of contact thermal resistances without the necessity of reassembling the entire experimental setup.

The particular system components responsible for the heat source triggering and the recording of thermal responses as well as the detailed design of the temperature control circuit were already presented in [5] and [6]-[7] respectively, thus here the authors will focus mainly on the unit responsible for the heat flux measurement, which is a new feature compared to the above mentioned publications.

A. Hardware Description

The main component of the entire cooling condition control system is the processor with the AVR core. This processor issues appropriate control signals based on the information obtained from electrical, thermal and mechanical sensors. The communication with the system is enabled via a set of serial and USB I/O ports as well as the user interface containing LCDs, keyboard and LEDs.

The control of cooling conditions in the system is exercised by two relatively large (40 mm x 40 mm) Peltier Thermo-Electric Modules (TEMs) which are inserted on both sides of a tested device between the cold plates and spacers directly touching its package. The TEMs, having the nominal thermal power of 90 W, are operated at 17 V DC voltage. They are powered by two separate power modules containing their own power supply units with a transformer and a cooling fan.

The power units are based on the TDA7294 analog full bridge amplifiers, which provide current to the TEMs. These amplifiers, equipped with internal short-circuit overheat and protection circuits, are controlled by the symmetrical voltage and they can provide 10 A of current forcing the current flow through the TEMs in both directions, hence producing either heating or cooling action. According to the measurements, the variation of the control voltage by one mV changes the TEM current by approximately 4.5 mA.

Digital control signals for both TEMs are generated in the main processor. First, these signals are converted in the 12-bit MCP4922 DACs to the analog form and then they are further processed in the voltage level shifting circuit to produce the symmetrical control voltage for the bridge amplifiers.

The measurement of the instantaneous electrical power provided to the TEMs is possible owing to the application of the dedicated ACS712 current sensors and measurement amplifiers monitoring the voltage supplied to the TEMs. The analog signals coming from these sensors are converted into the digital domain in the internal 10-bit converters contained in the main processor.

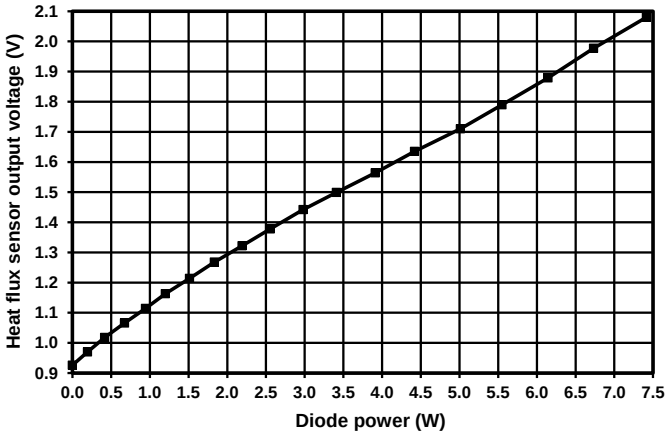


Fig. 1. Heat flux sensor calibration curve.

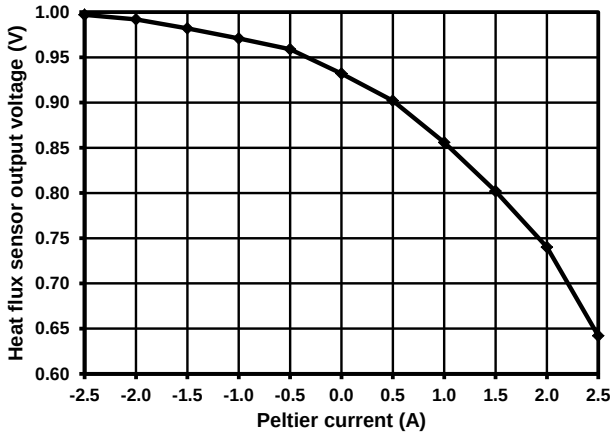


Fig. 2. PID regulator characteristic.

The temperature measurements at different locations in the system are taken with thermocouples connected to dedicated AD595 amplifiers, which render possible the compensation of the cold junction. The heat flux from the package is sensed with commercially available 27036-3 thin sensor integrated with an T-type thermocouple manufactured by Rhopoint. The readings from this sensor are highly temperature dependent thus a special calibration curve provided by the manufacturer has to be used when reading signal out of this sensor. Since the nominal sensitivity of this sensor is just 0.317 $\mu V/(W/m^2)$ its output signal is initially amplified 100 times in the INA128 amplifier. The output signals from all these thermal sensors are supplied to the main processor after their conversion into the digital domain in the 12-bit MCP320 ADCs.

B. System Calibration

The cooling conditions of a tested device or package could be controlled by the adjustment of TEM current based on the information coming from temperature and heat flow sensors. Then, implementing an appropriate current control algorithm it would be possible to force virtually any kind of boundary conditions, e.g. prescribed temperature, heat flux or even heat transfer coefficient. In such a case, the software code running in the processor monitors all the system thermal parameters and controls the operation of the TEMs thus permitting the real time adjustment of cooling conditions. However, for the proper operation of the algorithm all the system components must be carefully calibrated.

The particular algorithm chosen by the authors to control the TEM current is the well-known PID algorithm [8], which has already been used in [7] to maintain constant temperature on the TEM sides facing the package of a tested device. This was realized based only on the information coming from the temperature and electrical sensors. Here, the objective of the regulation is reformulated to maintain during measurements constant heat flux. Thus, now both temperature and heat flux data is required for the control algorithm as it was originally proposed in [9].

According to the theory [8], the PID algorithm parameters required to adjust the heat flux can be determined from the open loop gain of the control system. This means that in the considered case the dependence of the heat flux sensor output voltage on both the TEM current and the power dissipation in a tested device have to be known.

Thus, first the actual sensitivity of the heat flux sensor has to be determined. For this purpose, a power diode was placed in the DCP stand. Its package was thermally insulated from the top and firmly attached to the bottom cold plate with the application of thermal grease, thus forcing unidirectional heat flow. The experimentally measured dependence of the output sensor voltage on the dissipated power is presented in Fig. 1. As can be seen, the obtained characteristic is fairly linear in a relatively wide range of dissipated power values with the slope of 160 mV/W. Now, knowing the precise contact area this value could be converted into the actual heat flux units.

Another unknown characteristic which is still required for the control algorithm to determine the open loop gain is the sensitivity of the regulated parameter, i.e. the heat flux, on the TEM current. This experimentally measured curve is shown in Fig. 2. As can be seen, this dependence is highly nonlinear what means it has to be somehow linearized near the nominal operating point so as to find the optimal algorithm parameter values. An additional difficulty in controlling the heat flux consists in the nonlinearity of the curve presented in Fig. 2 because the control voltages required to adjust the heat flux value differ significantly for positive and negative Peltier TEM current values.

III. EXPERIMENTAL RESULTS

The feasibility of controlling the heat flux has been verified in practice based on the example presented in this section. The experiment consisted in maintaining the heat flux flowing through the heat slug of a power diode constant for different values of dissipated power, what in fact corresponds to the significant degradation of the device cooling and the increase of its thermal resistance.

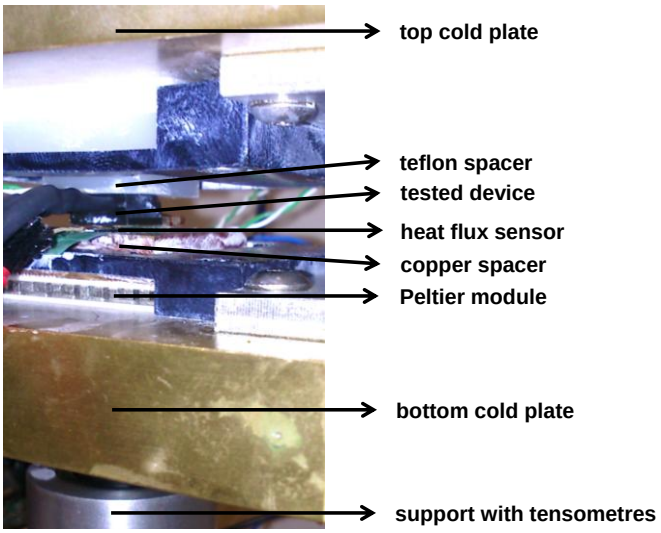


Fig. 3. Device mounting details.

A. *Measurement Setup*

The measurements were performed in the earlier described DCP stand. The tested device was a third generation silicon carbide power diode rated for 6 A and 600 V packaged in the TO-220 package. The diode, as pictured in Fig. 3, was firmly squeezed between the cold plates. During the measurements the package was thermally insulated from the top cold plate by the thick Teflon spacer and generated heat was removed from the device towards the bottom cold plate through the copper spacer and the Peltier TEM; all visible in the photo. The heat flux sensor was placed between the heat slug of the package and the bottom spacer.

First, the power of 1.0 W was dissipated in the diode. Then, it was increased to 1.5 W and 2.0 W, each time after at least 300 seconds when the thermal steady state was approximately reached. Except for the heat flux, the temperature values were measured by thermocouples located on both sides of the TEM and the heat slug. The water flow through the cold plates was forced by a thermostat and its temperature was set to 30 ˚C.

B. *Heat Flux Control*

Initially, the measurements were taken without any control of the heat flux. The recorded variation of the voltage at the heat flux sensor output is presented in Fig. 4 by the lighter noisy curve. Additionally, the solid black line represents the changes in the power dissipation. The numbers on the left side of the chart indicate the output voltage whereas the ones on the right correspond to dissipated power. As can be seen, the obtained results prove that the heat flux sensor operates correctly and its response remains in quite good agreement with the earlier measured sensor calibration curve pictured in Fig. 1.

Even more information can be inferred from Fig. 5, where the changes in the power dissipation are plotted together with the recorded temperature values. The lighter line corresponds to the heat slug temperature whereas the darker double line represents the copper spacer temperature. Now, the numbers on the left side of the chart indicate the measured temperature values.

Looking at the figure one can easily notice that during the experiment the temperature of the copper spacer, located between the heat flux sensor and the TEM, remained fairly constant, whereas the temperature of the heat slug increased significantly. As a result, when power increased from 1.0 W to 2.0 W, the temperature difference between the heat slug and the spacer almost doubled from 22 K to 42 K. Except for the non-perfectly unidirectional heat flow within the structure, this slight nonlinearity can be explained by improved cooling through the outer package surfaces due to enhanced radiation and convection cooling at elevated temperatures.

Another comment which has to be made is that both the temperature and heat flux settling times are relatively long and they exceed three minutes. This fact could be a serious problem for the real time stabilization of the heat flux.

Next, this experiment was repeated, but this time the PID control algorithm was activated and it was set to maintain the same heat flux value as if the power of 1.0 W was dissipated in the diode.

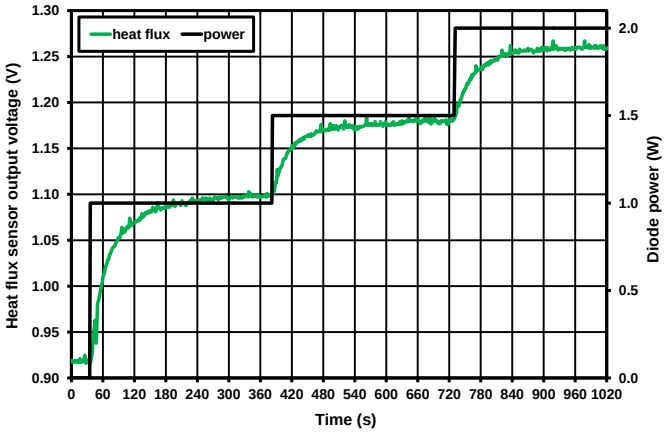


Fig. 4. Uncontrolled variations of surface heat flux.

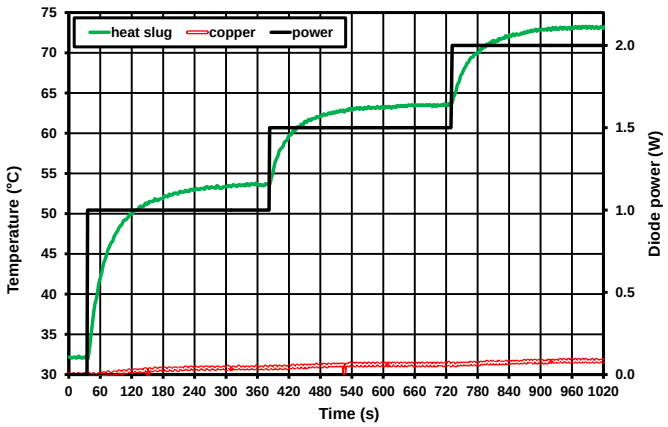


Fig. 5. Variation of temperature without heat flux control.

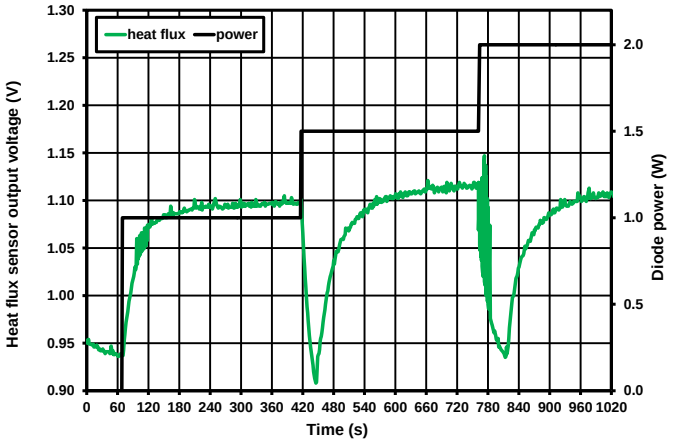


Fig. 5. Controlled variations of surface heat flux.

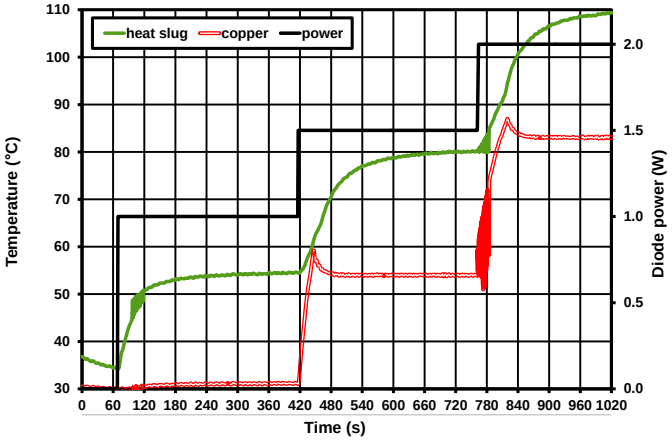


Fig. 6. Variation of temperature with heat flux control.

The results of the second experiment are shown in Figs. 5-6 representing respectively the measured variations of the heat flux and the temperature values at the previously considered locations. As can be seen in Fig. 5, the PID algorithm indeed manages to attain the desired value of the heat flux in the steady state maintaining at the same time the temperature difference between the heat slug and the copper spacer in the range of 23 K÷26 K. However, this time again long settling times are observed. Moreover, the algorithm first opposes the changes in the heat flux starting from the zero heat flux condition and then it tends towards the desired value.

Concluding, this first attempt to demonstrate in practice the possibility to maintain constant surface heat flux during the entire measurement was successful. However, still more fine-tuning of the PID control algorithm parameters is required. Moreover, taking into account that the heat flux sensor itself introduces important thermal resistance and capacitance into the main heat flow path, in order to decrease the settling time future versions should rely on the stabilization of temperature differences rather than on keeping constant heat flux sensor indications.

IV. CONCLUSIONS

This paper discussed in detail the practical problem of real time control of cooling conditions during the transient thermal measurements of an electronic device. The presented results demonstrated the feasibility of realizing a PID control system which, when provided with adequate heat flux or temperature data, is capable of maintaining during measurement not only constant temperature but also heat flux or thermal resistance, independently from the power dissipation in a tested device.

Such a versatile and precise control of cooling conditions offers virtually unlimited freedom of shaping the surface heat exchange coefficient value without the necessity to repeatedly disassemble the entire experimental setup. This feature might be especially useful for the generation of boundary condition independent compact thermal models or for the determination of the junction-to-case thermal resistance according to the recent JEDEC Standard JESD51-14 [10].

ACKNOWLEDGMENT

The research presented in this paper was supported partly by the grant of the Polish Ministry of Scientific Research and Higher Education No. N515 2423 37.

REFERENCES

[1] V. Székely, A. Nagy, S. Török, G. Hajas, and M. Rencz, "Realization of electronically controlled thermal resistance," *Microel. J.*, vol. 31, pp. 811-814, 2000.

[2] V. Székely, G. Mezősi, "Design issues of a variable thermal resistance," *Proc. 12th Therminic*, Nice, France, 2006, pp.186-190.

[3] V. Székely, S. Török, and E. Kollár, "Improvements of the variable thermal resistance," *Proc. 13th Therminic*, Budapest, Hungary, 2007, pp. 180-183.

[4] D. Schweitzer, H. Pape, L. Chen, R. Kutscherauer, and M. Walder, "Transient dual interface measurement - a new JEDEC standard for the measurement of the junction-to-case thermal resistance," *Proc. 27th Semi-Therm*, San Jose, USA, 2011, pp. 222 – 230.

[5] M. Janicki, Z. Kulesza, P. Pietrzak, and A. Napieralski, "Multichannel system for real time registration of electronic circuit temperature response," *Proc. 17th MIXDES*, Wroclaw, Poland, 2010, pp. 357-360.

[6] M. Janicki, Z. Kulesza, T. Torzewicz, and A. Napieralski, "Automated stand for thermal characterization of electronic packages," *Proc. 27th Semi-Therm*, San Jose, USA, 2011, pp. 199-202.

[7] M. Janicki, Z. Kulesza, T. Torzewicz, and A. Napieralski, "Active control of circuit cooling conditions for transient thermal measurement purposes", *Proc. 17th THERMINIC*, Paris, France, pp. 215-218.

[8] K.J. Åström, and T. Hägglund, *PID Controllers: Theory, Design, and Tuning*. ISA, 1995.

[9] M. Rencz, E. Kollár, and V. Székely, "Heat flux sensor to support transient thermal characterization of IC packages," *Sensor Actuat. A-Phys.*, vol. 116/2, pp. 284-292, 2004.

[10] JEDEC Standard JESD51-14, "Transient Dual Interface Test Method for the Measurement of the Thermal Resistance Junction to Case of Semiconductor Devices with Heat Flow Trough a Single Path," 2010, available at www.jedec.org.

25-27 September 2012, Budapest, Hungary

Silicon Carbide – The Power Device for the Future

Erik Janzén and Olof Kordina
Linköping University
Department of Physics, Chemistry, and Biology, IFM, 581 83 Linköping, Sweden
+46 13 28 10 00, +46 13 13 75 68, erija@ifm.liu.se

***Abstract* – Silicon Carbide (SiC) is a wide bandgap semiconductor with highly appreciated properties making it ideal for high power, high frequency, and high temperature applications. One of these properties is the high thermal conductivity (λ) which is mentioned in almost every single publication, yet very few measurements of λ on SiC have been made and essentially no knowledge has been accumulated on the limiting factors on λ or on what role λ plays in a finished device. Improvement of λ may be achieved if the material is isotope enriched i.e. the SiC is dominatingly $^{28}Si^{12}C$. In this paper we will highlight the properties of SiC as a power device material with specific emphasis on λ. The results from $^{28}Si^{12}C$ show that the layers are very low doped and have an isotope purity of 99% or better.**

I. INTRODUCTION

What makes SiC so attractive for power applications? In the perfect world a power device blocks a very high voltage in the reverse direction and at the same time conducts perfectly in the forward direction. However no power device material is perfect, so let us look at what limits the forward direction using the simplest model. In a unipolar device, such as a Schottky diode or MOSFET, all current transport is achieved by the few free electrons that are manifest in the active layer (fig. 1). The resistance (R_{on}) is hence obviously determined by the doping and thickness of the active layer. It is thus an advantage if the active layer can be made as thin and high doped as possible.

Blocking a voltage, however, requires a rather thick layer with a low doping in order to keep the field at the junction below the critical field E_{max}. Here SiC has a major advantage as E_{max} is roughly ten times higher than that of Si which in essence means that an active layer of SiC for a specific voltage can be made ten times thinner and ten times higher doped leading to an R_{on} which is 100 times lower than that of Si [1].

The thickness and doping in the reverse direction for a unipolar 2kV SiC diode is hence roughly 20 µm thick and the doping in the mid 10^{15} cm^{-3}. Obviously the equivalent device in Si the active layer would be too thick and too low doped to produce a satisfactory R_{on} and therefore the high voltage devices made from Si will be bipolar where R_{on} is not determined by the doping of the active layer but rather the total amount of carriers present in the layer at a specific instance.

In a bipolar device holes and electrons can be injected into the active layer to form an electron-hole plasma. The amount of carriers that are present in the layer is determined by the carrier lifetime (τ) and hence R_{on} can be quite low even for high voltage devices. The drawback is however that when the device switches from the on-state to the reverse-state all the carriers present in the layer will rush to their respective contact and give rise to a short current pulse which we call the reverse recovery current. As a switching device, the unipolar device, where switching losses are limited to the RC losses, is far superior to the bipolar device.

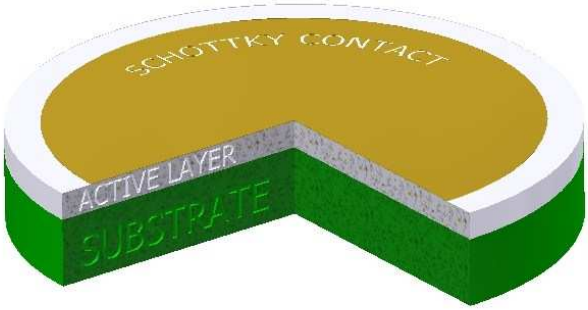


Fig. 1. Illustration of a Schottky diode

A lower carrier lifetime will of course give a better switching device but at the same time R_{on} will become higher so an appropriate balance must be found. Fig. 2 illustrates simply how the carrier distribution looks depending on whether τ is high or low. Most carriers are located close to the anode and cathode. When τ is low, more carriers must be injected i.e. the forward voltage drop will be higher in order to reach the desired current.

Controlling the carrier lifetime is hence of very high importance as the characteristics of the finished bipolar device greatly depends on it.

Most power devices are used in applications where fast switching is required. Motor drives for instance take a DC current and make an AC with variable frequency. The faster the switching the more sinusoidal will the AC current be and the losses will be reduced. Analyzing the losses one will see

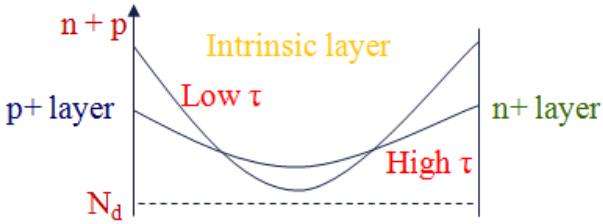


Fig. 2. Illustration of how the carrier lifetime impacts the carrier distribution in a PiN diode.

that the switching losses account for the largest share of the losses in circuits using bipolar devices. However, when unpolar devices are used, the switching losses are substantially reduced and conduction losses and reverse leakage losses are of greater concern. In most high voltage applications such as in motor drives (hybrid cars), bipolar Si diodes are used as freewheeling diodes together with a Si IGBT or MOSFET (CoolMOS or similar). The reverse recovery from the freewheeling diode is huge and this gives rise to the large switching losses. If the bipolar Si diode is substituted with a SiC Schottky diode, these switching losses become very small fig. 3 [2].

As the junction temperature increases the mobility of the carriers decreases which is reflected in a larger forward voltage drop. An excellent example of this may be seen in fig. 4 which shows the thermal characteristics of a SiC Schottky diode made by Infineon [3]. Additionally, the leakage in the reverse direction increases with increasing temperature. Keeping the junction temperature low is hence of high importance and here packaging can help a lot.

Thankfully SiC has a very high thermal conductivity. Low doped SiC has a λ which is 390 W/mK and 490 W/mK parallel and perpendicular to the c-direction, respectively [4]. This is substantially higher than that of Si but much less than λ of diamond. It is surprising to note that in any publication on SiC you read, λ is mentioned yet very few studies have been made. Very much knowledge on the material can be accumulated if λ is studied on a number of different samples grown under different circumstances. In which way do extended intrinsic defects impact the thermal conductivity? Can a large concentration of vacancies impact the thermal conductivity? These are questions to which we yet need to seek an answer.

A known way to improve λ is to reduce the isotope disorder in the material. This has been done in Si, Ge, and diamond but never in SiC with the purpose to increase λ. Simple estimates show that λ of isotope enriched SiC should be 36% higher than that of natural SiC [5].

From a commercial standpoint, λ is the foremost parameter determining the current density i.e. the number of chips on a wafer is determined solely by λ..

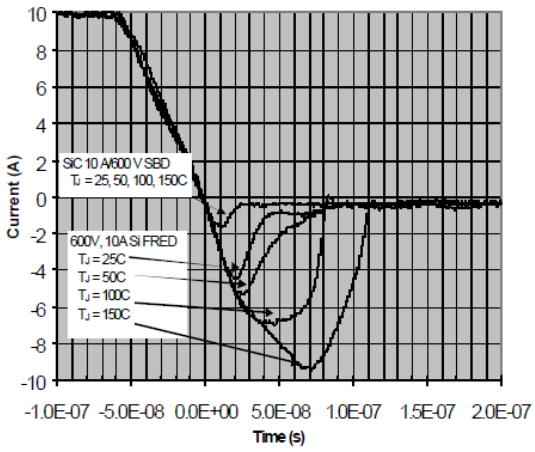


Fig. 3. Comparison of the reverse recovery losses between a SiC Schottky freewheeling diode and a fast recovery Si diode together with a Si IGBT in an inverter configuration for solar panels: Even at room temperature the SiC Schottky diode greatly outperforms the FRED diode. At higher junction temperatures the FRED diode performs even worse.

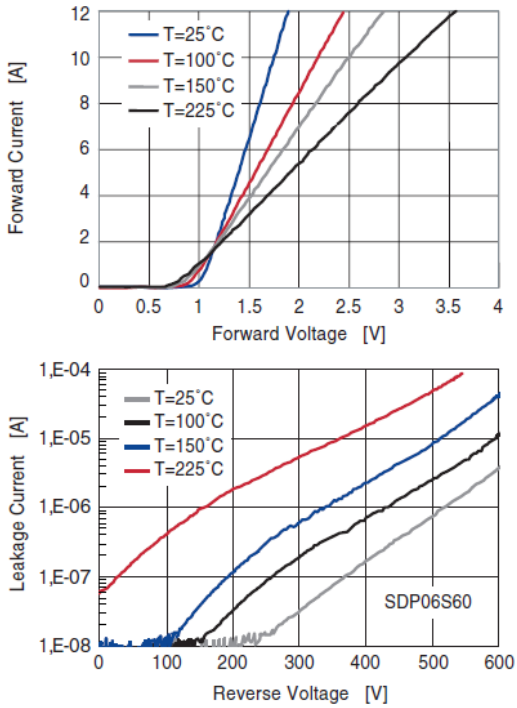


Fig. 4. The thermal characteristics of a SiC Schottky diode.

II. Properties of SiC

Besides the high breakdown electric field strength and the high thermal conductivity, SiC has a high saturated electron drift velocity v_{sat} which is advantageous in high frequency devices. Most high frequency devices are however now moving towards group III/N type devices since one can obtain a higher power density from these. However, the highest power densities are obtained on devices grown on semi insulating (SI) SiC substrates since these have a higher thermal conductivity.

SiC is a wide bandgap material but the bandgap depends on which polytype that is referred to. All polytypes have the same composition and in the basal plane they all look the same, however they differ in the stacking sequence of the Si-C bilayers in the c-direction.

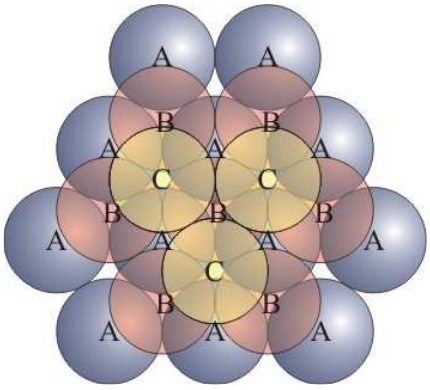


Fig. 5. The closed packed hexagonal structure of SiC as seen along the c-direction. Note that on the 'A' layer the next layer can either be on a 'B' position or a 'C' position. Depending on how the layers are stacked on top of each other, different polytypes will be obtained.

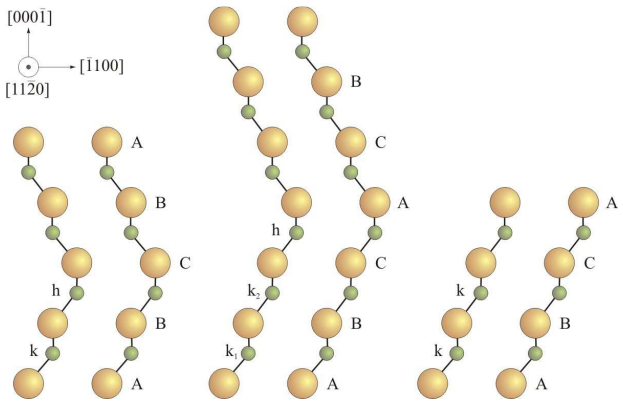


Fig. 5. The closed packed hexagonal structure of SiC as seen along the c-direction

The three most common polytypes are 3C, which is the only cubic structure in SiC, 4H, which is the most commonly used polytype, and 6H. The number denotes the number of bilayers of the stack sequence before it repeats itself and the letter denotes the resulting crystal structure. The polytypes differ in all optical and electronic properties; the bandgap is for instance 3.2 eV for 4H and 2.4 eV for 3C.

III. Growth of SiC

A. *Crystal Growth*

SiC does not have a molten phase, instead it sublimes at high temperatures. The crystal growth normally is done using a process called seeded sublimation growth also known as Physical Vapor Transport (PVT) and modified Lely technique which was first put into practice by Tairov and Tsvetkov [6]. The technique is very simple; a container is filled with a powder of SiC and closed with a lid that has a seed wafer attached to it. When the container is heated small axial and radial thermal gradients are placed on the system. The axial thermal gradient will ensure that material is transported in a positive way from the source to the seed. The radial gradient is made such that the growth rate is slightly lower on the crystal edge which is vital in order to obtain a good crystal structure. When the temperature reaches more than about 1800 °C, the SiC will begin to sublime appreciably and, on account of the thermal gradient, material will start to transport from the source to the seed.

The growth is usually done at seed temperatures in excess of 2000 °C. Axial thermal gradients are several tens of degrees/cm and the growth pressure is around 10 mbar to promote a fast transport. The growth rates are between 0.1 – 1 mm/h. The best quality material is of course obtained at the lower growth rates where the thermal gradients are low.

A related technique was developed in the early ‘90s which is called High Temperature Chemical Vapor Deposition (HTCVD) [7]. In this technique, the source material is replaced with gases, normally silane and ethylene and a carrier gas of He or a mixture of He and H_2 (fig. 6). The silane and ethylene will decompose already in the injector to form micro-clusters of SiC which are transported to the hot zone where they sublime. The sublimed species will move upwards to condense on the slightly colder seed. Growth rates and temperatures are in the same range as with the PVT technique but the pressure is around 400 mbar. The transport of the species and the micro-crystals is achieved by the carrier gas. The choice of carrier gas is a fine point as it is important that the gas is inert and has a good thermal conductivity. He has a thermal conductivity comparable to H_2, but Ar is a poor choice as the conductivity is only one tenth of that of H_2. Using only H_2 will lead to aggressive etching of the crystal, however, a small amount of H_2 is advantageous for the growth both on terms of growth rate and crystal quality.

If the micro-crystals do not fully sublime in the hot zone it can be seen as smoke coming out of the reactor. The temperature where one can observe smoke or not (the smoke limit) is very well defined. A change in only a few degrees can move the process from showing smoke or not. The crystal quality will be severely compromised if the sublimation of the micro-crystals is not complete. The choice of carrier, gas, pressure, thermal gradient, and precursor chemistry all impact on the smoke limit. The choice of carbon precursor is of particular importance; propane is for instance very bad whereas methane or ethylene is substantially better. The intriguing reactions occur in the injector region where the micro-crystals form. A higher pressure will for instance promote homogeneous nucleation and the particle size will increase. Larger micro-crystals will be more difficult to sublime and hence the smoke limit will increase.

The purity of the crystals is outstanding using the HTCVD technique as the gases can easily be obtained in a very pure form. The crystal quality is comparable to the PVT technique. Unfortunately, the crystal quality with both techniques is

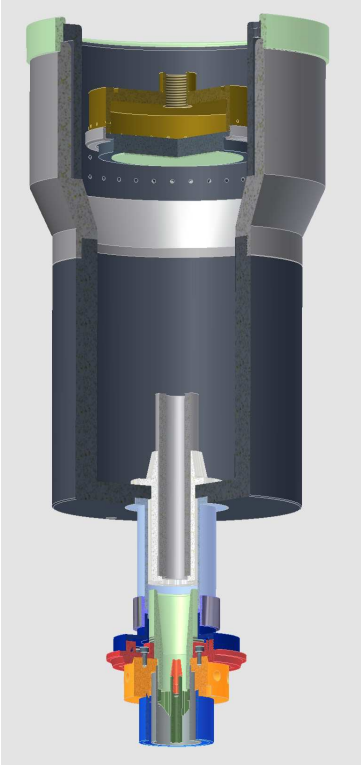


Fig. 6. CAD model of the HTCVD system. The gases enter at the bottom in a co-axial injector. The precursor gases enter in the center and most of the carrier gas is entered in the outer part of the injector. This is to avoid parasitic depositions in the injector. The micro-crystals formed in the gas phase through homogeneous nucleation are transported to the hot-zone where they sublime and the species condense on the seed.

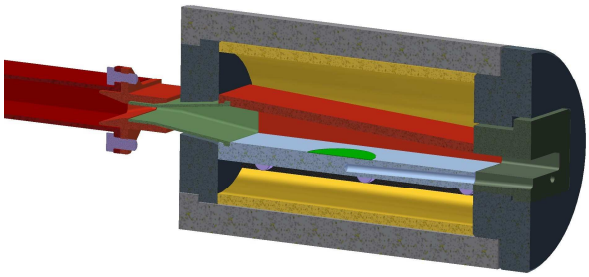


Fig. 7. CAD model of a horizontal hot wall CVD susceptor. The wafer is placed on the bottom of the susceptor. A tilted ceiling is placed over the bottom. The gases enter via a liner and a transition piece and inlet. The susceptor is heated by a cylinder which in turn is heated via an rf coil. Rigid thermal insulation helps load everything inside the quartz tube.

hampered by the large thermal gradients which tend to build in stresses in the material. Additionally the hard attachment to the lid is also a concern. The attachment is needed to prevent Si from the seed crystal leeching out, however, the difference in CTE and temperature dependence of the CTE between the graphite and SiC has to be well matched otherwise the crystal can crack during cool down.

The HTCVD has not made it commercially to any great extent. The high purity is of course nice for making SI substrates for high frequency devices but conducting substrates for power devices is today exclusively produced using the PVT technique.

B. Epitaxial Growth

All epitaxial SiC layers are today grown using Chemical Vapor Deposition (CVD). There are several different styles of reactors but today most reactors are either hot-wall reactors or warm-wall reactors. The hot-wall reactor was first introduced in the early ’90 and was an immediate success on account of the very high purity and quality material that could be produced in these. The concept has been developed since the ’90 and today there are horizontal reactors with a 3 x 3” capacity giving a 1.6% and 4.8% thickness and doping uniformity, respectively [8]. There are also larger planetary reactors with a capacity of up to 10 x 100 mm which has a thickness and doping uniformity of 2.4% and 3.7%, respectively [9].

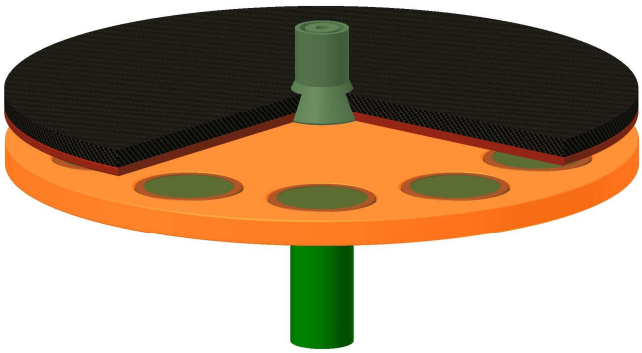


Fig. 8. Principle of a warm wall planetary reactor. The gases enter through the injector in the center. The main plate rotates with mechanical rotation via the shaft seen below. The individual satellites rotate by gas foil rotation. The main plate is heated by a pancake rf coil, however the ceiling is radiatively heated by the main plate. There is usually some thermal insulation on top of the ceiling to keep the temperature as high as possible.

The standard chemistry is silane, propane (or ethylene) and a massive flow of hydrogen as carrier gas. Typical growth conditions are 1500 - 1600 °C, 100 – 200 mbar, and a C/Si ratio between 0.8 – 2. N-type and p-type doping is achieved by additions of N_2 and trimthyl-aluminum (TMA), respectively. This chemistry is proven and gives growth rates in the order of 10 µm/h. Increasing the growth rate proved to be difficult with the standard growth conditions due to excess homogeneous nucleation. The Si clusters could be so large that they would fall down on the wafer surface and leave an imprint on it as seen in figure 9. Lowering the pressure and increasing the carrier gas helps reduce homogeneous nucleation but it proved difficult as it led to increased etching counteracting the intended and expected growth rate increase. Tsushida et al. went to the extreme in this respect and grew epitaxial layers at 40 mbar with a carrier flow of 50 – 70 SLM resulting in growth rates of 250 µm/h [10]. Undeniably, this is an impressive result however it proved much more successful to use a chlorinated chemistry. The Cl – Si bond is much stronger than the Si – Si bond which makes this chemistry attractive. Using a chlorinated chemistry, growth rates in excess of 100 µm/h could be obtained [11]. Leone et al. attempted to grow bulk using the $SiCl_4$ resulted in growth rates of 350 µm/h [12]. The efficiency of using chlorinated silanes directly is better than using silane and adding HCl. This can simply be understood considering that the active growth specie in the process is $SiCl_2$ which quite readily is formed from e.g. trichlorosilane directly whereas it is more difficult to form it from silane: No less than six bonds must be broken and there has to be a “meet and react” process between the Si atom and the two chlorine atoms.

Typical background doping is in the low 10^{14} cm^{-3}. This is sufficient for power devices up to roughly 20 kV. Additionally layers up to about 200 µm thick can be grown with good morphology.

IV. DEFECTS

Controlled amounts of point defects are needed in order to make the devices work as intended but other than those we do not want any defects. N-type doping is achieved by adding nitrogen. Due to the crystal structure, nitrogen in 4H can occupy two sites that differ in the number of second closest neighbors. This gives nitrogen two different binding energies, one for nitrogen on the cubic site and one on the hexagonal site. 6H consequently has three inequivalent sites and 3C, of course, only one. The nitrogen donors in 4H has thus a binding energy of roughly 61 meV and 125 meV. Diffusion is almost nonexistent in SiC and all dopants must be either grown during epitaxy or implanted. Boron is known to diffuse slightly at 1600 °C but it is essentially fixed. Acceptors are

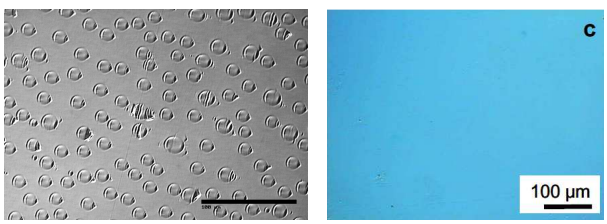


Fig. 9. Left, imprints from Si droplets that have fallen down and evaporated. Bar length is 100 µm. Right, optimized morphology of a thick layer using chlorinated chemistry.

deeper in SiC but the more interesting defects to discuss are those that are not so nice to have at all times. It is believed that the carbon vacancy C_v, which is a mid gap defect, is the one that limits the carrier lifetime. Typically, carrier lifetimes are around 1 µs on as-grown epimaterial and it depends greatly on the growth parameters [13].

There are several disturbing structural defects in SiC. Micropipes, which are small hollow cores penetrating the wafers, used to be a major concern, however, the density of micropipes has gradually been reduced to acceptable levels. It is even possible to obtain micropipe free material today. Screw dislocations and threading edge dislocations at high concentrations will impact the performance of the power devices in a negative manner, however, the basal plane dislocations (BPD) are here by far the biggest problem. When a bipolar device is run in the forward direction, a BPD can transform into a Shockley type stacking fault leading to a higher forward voltage drop. It is believed that the stacking fault acts as a quantum well structure which can trap carriers and hence limit the current flow [14]. Stacking faults are undesirable for another reason as we believe that these could limit the thermal conductivity since they are extended and covers an appreciable area on the surface and they form a double interface that could limit the thermal conductivity. This is one reason for studying the thermal conductivity in greater detail in SiC.

V. Isotope Enriched SiC

Though λ may be high for SiC, it will definitely be commercially interesting if λ can be increased granted that the cost of producing the epi and wafers will not be too high. This has been tried in Si previously and not been a commercial success. There was no consensus as to how much the isotope enrichment actually improved λ at RT. The cost of the enriched silane is also prohibitive which would make the epitaxy shockingly pricey to produce. The situation is somewhat different for SiC. The cost to produce SiC wafers and epi does not lie so much on the precursor materials but rather on the hot zone consumables and the post processing. All gases, precursor gases and other gases needed in the process, make up for less than 10% of the total cost of making an epiwafer. Furthermore, it is believed that the effect on λ due to isotope enrichment will be substantial in SiC. As mentioned previously, one estimate shows that λ will be 36% higher for $^{28}Si^{12}C$ as compared to natural SiC. Obviously, the enriched precursor gases must come down in price if this is going to be a commercial success but a price around $2 – $5/g would make it interesting commercially. The increase in price of the epiwafer can be offset by a higher packing density of the chips. For some applications enriched epi would be sufficient, but high frequency devices would greatly benefit in using an enriched SI wafer with a higher λ.

There are other interesting applications for isotope enriched SiC such as fuel coating for nuclear fuels, quantum computers, and fundamental spectroscopic studies. These applications can help drive the market towards an industrial scale where the price of the precursors will be more attractive.

At Linköping University an isotope enriched project recently started with the aim to determine the thermal conductivity of enriched SiC and to make fundamental studies on quantum computing applications and detailed spectroscopy. Enriched $^{28}SiH_4$ and $^{12}CH_4$ was used as precursor gases. The target is to produce as thick and pure epitaxial layers as possible suitable for measuring thermal conductivity. The meaning of 'pure' is here that the layers should be pure in terms of point defects, structural defects, and isotope composition.

Centrifuged SiF_4 was converted into silane. The SiF_4 contained significant amounts of boron and phosphorous which both are disturbing dopants in SiC. Particularly B is difficult which produces a deep acceptor in SiC.

Methane is also not ideal to use for epitaxial growth of SiC as the morphology usually is very rough however this is the only enriched hydrocarbon that is available at any significant amounts to our knowledge. Normally growth is conducted in a silicon limited growth regime (higher C/Si ratios) where the growth rate is limited by the supply of Si to the surface. Additionally, methane works well at substantially higher temperatures so tests were performed using a lower C/Si ratio where the growth rate is in a carbon limited regime. This usually renders a substantially better morphology though at the price of a higher nitrogen doping. The morphology was indeed better and using a growth temperature of 1700 °C the morphology was quite acceptable. Some new type of small triangular features could be seen on the surface which probably is related to stacking faults but the density of these was very low.

The growth conditions for the enriched runs were thus a growth temperature of 1700 °C, growth pressure 100 mbar, C/Si ratio 0.8, Si/H_2 ratio 0.25%, and Cl/Si ratio of 4. In total seven runs of enriched material were made and the thickest layers were 80 µm thick. Several substrates were put inside the reactor to maximize the yield. The different types of substrates were: 4H Si-face and C-face 8 deg off axis, 6H Si-face 3.5 deg off axis, 4H Si-face 4 deg off axis, and one 3C sample.

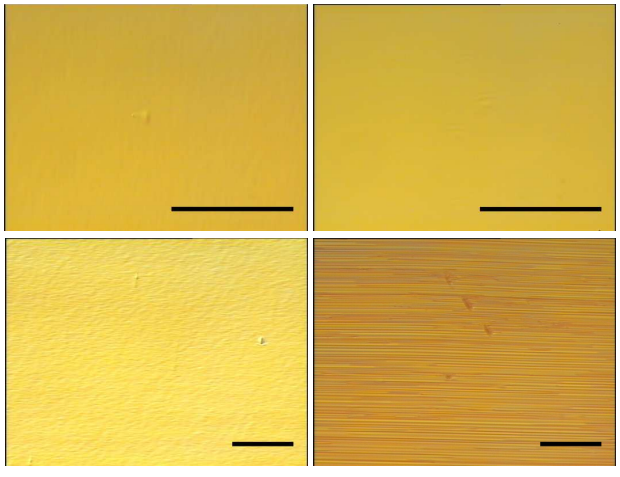


Fig 10. The surfaces of a thick 4H SiC epitaxial layers grown on 8 deg off axis substrates. Top Left. Si-face isotope enriched SiC using methane as carbon source. Note the small triangular defect on the surface. Top Right. Si-face natural SiC using ethylene as carbon source. Bottom Left. C-face isotope enriched using methane and Bottom Right. Natural SiC using ethylene as carbon source. Note the large difference in morphology on the C-face which is caused by the carbon precursor. Bar length is 100 µm

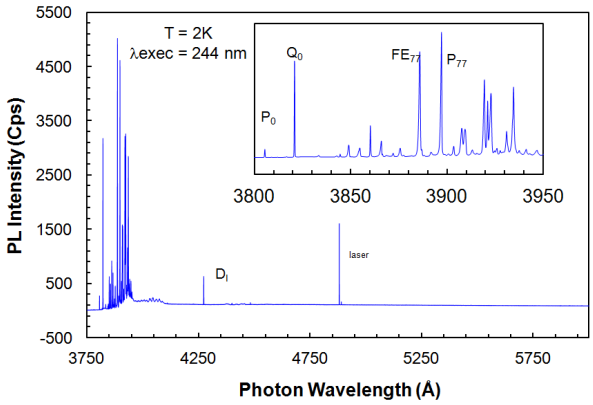


Fig. 11. PL spectrum of an isotope enriched sample. The spectrum is dominated by near bandgap luminescence consisting of nitrogen BE and FE phonon replicas. No luminescence from deep defects can be seen.

SiC coated graphite is usually used for the growth of epitaxial SiC, however, in this case we could not use SiC coated graphite due to the risk of contaminating the layer with ^{29}Si, ^{30}Si, or ^{13}C. For this application, TaC coated graphite was used instead. TaC is a high temperature refractory and the Ta - C bond is very strong so there is no risk of any ^{13}C coming from the TaC coating. However, the TaC coating can form micro-cracks which can expose the underlying graphite which then would become a source of ^{13}C contamination. The decision was nevertheless made in favor of the TaC coating.

PL measurements were made at 2K and using a 244 nm frequency doubled Ar ion laser as excitation source. All samples were measured and we could see that there were no traces seen of either boron or phosphorous. The background doping was in the low 10^{14} cm^{-3} range as determined by comparing the free exciton (FE) FE_{77} line with the Q_0 nitrogen bound exciton (BE) line. At longer wavelengths one can often see stacking fault related luminescence and luminescence from 3C inclusions but no such luminescence was observed as can be seen in fig. 11. The only defect related line that was observed is luminescence from the D_1 center which is very often seen in epitaxially grown SiC.

The line widths are known to be very sharp for isotope enriched Si and in this case we could not resolve the line width of the nitrogen BE as they were limited by the spectral response of the monochromator. More detailed PL measurements are ongoing using a monochromator with a higher resolution.

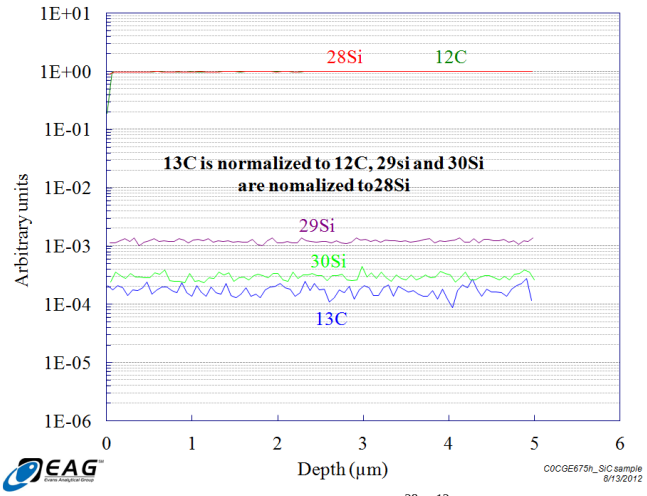


Fig. 12. SIMS data from an isotope enriched ^{28}Si^{12}C sample. The purity is 99.85% and 99.98% on ^{28}Si and ^{12}C, respectively.

Figure 12 shows the SIMS measurements from the first enriched layer that was grown. The isotope purity is indeed very good; 99.85% on ^{28}Si and 99.98% on ^{12}C.

Thermal conductivity measurements have not been performed at the time of this paper's writing

VI. SUMMARY

Silicon carbide has matured into a useful material suitable for high power devices. Several troubling issues have been solved and the market for SiC devices is growing. There are still several issues remaining to be solved with the material. In particular basal plane dislocations which limit the forward current in bipolar devices is a concern. The thermal conductivity is high but is limited by the isotope disorder. Layers of isotope enriched SiC have been produced for demonstrating an improved thermal conductivity.

ACKNOWLEDGEMENTS

Support for this work has been provided by the Knut and Alice Wallenberg foundation and the Swedish Research council.

[1] J. A. Cooper, Jr. and A. Agarwal, Proceedings of the IEEE, Vol. 90, issue 6, p 956, (2002)
[2] Michael O'Neill, Cree Research application document, "Silicon Carbide Diodes Make Solar Power Systems More Efficient", 2008, www.cree.com
[3] Christian Meisner, Roland Rupp, Holger Kapels, Michael Krach, and Ilia Zverev, Infineon application document, "ThinQ Silicon Carbide Schottky Diodes: An SMTP Circuit Designer's Dream Come True!", www.infineon.com
[4] A. Powell, J. Jenny, St. Müller, H.McD.Hobgood, V. Tsvetkov, R. Leonard, and C. Carter, Jr., International Journal of High Speed Electronics and systems, Vol. 16, No. 3, 751, (2006)
[5] D. T. Morelli, J. P. Heremans, and G. A. Slack, Phys Rev. B, Vol. 66, 195304, (2002)
[6] Tairov, Y. M., and V. F. Tsvetkov, J. Cryst. Growth, Vol. 43, 1978, p. 209.
[7] A. Ellison, B. Magnusson, C. Hemmingsson, W. Magnusson, T. Iakimov, L. Storasta, A. Henry, N. Henelius, and E. Janzén, MRS Proceedings, Volume 640, (2000).
[8] J. J. Sumakeris, M. K. Das, S. Ha, E. Hurt, K. Irvine, M. J. Paisley, M. J. O'Loughlin, J. W. Palmour, M. Skowronski, H. McD. Hobgood, C. H. Carter, Jr., Mater. Sci. Forum, 483-485, 155, (2005).
[9] C. Hecht, R. Stein, B. Thomas, L. Wehrhahn-Kilian , J. Rosberg, H. Kitahata, F. Wischmeyer, Mater. Sci. Forum, 645- 648, 89, (2010).
[10] H. Tsuchida, M. Ito, I. Kamata, M. Nagano, T. Miyazawa, and N. Hoshino, Mat. Sci Forum, 645 – 648, pp 77 – 82, (2010).
[11] For a review see: Henrik Pedersen, Stefano Leone, Olof Kordina, Anne Henry, Shin-ichi Nishizawa, Yaroslav Koshka, and Erik Janzén; Chemical Review , Publication Date (Web): December 2, 2011 **DOI:** 10.1021/cr200257z
[12] S. Leone, F.C. Beyer, A. Henry, C. Hemmingsson, O. Kordina, and E. Janzén, Crystal Growth and Design, Vol. 10, 3743 – 3751, (2010)
[13] L. Lilja, J. Hassan, I. Booker, J.P. Bergman, and E. Janzén, Mat. Sci. Forum, Vol. 715 – 720, 161, (2012)
[14] Seoyong Ha and J. P. Bergman, MRS Bulletin, Vol. 30, issue 4, 305, (2005)

25-27 September 2012, Budapest, Hungary

The Thermoreflectance Approach in Dynamic Thermometry and Thermophysical Property Measurement for Thermal Design of Electronics

Peter E. Raad[1, 2, *] and Pavel L. Komarov[1]
[1] Mechanical Engineering Department, SMU, Dallas, TX 75275 USA
[2] IEEE Senior Member
[*] *Corresponding Author: +1 (214) 768-3043, praad@smu.edu*

Abstract — This presentation is a part of the Special Session on Dynamic Thermometry and Thermophysical Property Measurement for Thermal Design of Electronics, and aims at providing a perspective on the physics of thermoreflectance (TR) and its application to measuring the thermal properties of thin-film electronic materials and the surface temperature fields of active devices. The existence of a dependence between the temperature of a surface and the reflectivity of that surface has long been known. But, since that dependence is generally weak, on the order of 10^{-3} or 10^{-4} per Kelvin, it was not immediately obvious how to leverage thermoreflectance physics as a thermal metrological tool. More recently, however, several research teams around the world have found success in the use of TR as a means of measuring thermal properties and temperature, and there have been early successful efforts in productizing thermal measurement instruments based on the TR approach.

I. INTRODUCTION

As an optical method, the TR approach has the advantage of being non-contact and non-invasive. Since materials reflect light differently at different wavelengths, the TR approach has the added benefits of making it possible to (a) achieve smaller spot sizes by measuring at low wavelength in the visible spectrum, and (b) vary the probing light wavelength so as to maximize the TR response from a given material surface. The facts that microelectronic devices have been shrinking in size and are typically covered with reflective materials have contributed to the ascendency of the TR approach. In effect, thermoreflectance extends the resolution range of existing well-established optical approaches, such as IR, and makes it possible to measure the thermal performance of a broad range of materials that may not be made easily compatible with a method that is inherently limited to a single wavelength of light.

Thermoreflectance-based methods work by measuring the relative change of reflected light resulting from a change in the temperature of a surface. When trying to obtain the thermal property of a thin-film material, one needs to force a change of temperature in that surface and record the change in reflected light. This is achieved by a so-called "pump and probe" approach. Since heat dissipation in thin films is a fast phenomenon, it has been traditional to use a pulsed laser to heat the sample under test (the pump part) and a second CW laser of a different color to continuously monitor the change in reflectivity of the sample's surface (the probe part). In this arrangement, an appropriately fast and sensitive photodetector is used to record the transient signature of the probing laser's light that is reflecting from the surface of interest (while sensibly ignoring the output of the pumping laser). Once a thermal response has been obtained, it can be used to drive an optimization procedure in which the heat equation is solved repeatedly until the unknown thermal property is determined. The pumping and the probing can be done on the same side of a sample or alternatively one can pump on one side and probe the other side. Also, depending on the thickness (rather thinness) and diffusivity of a material of interest, the pumping can be done with nano-, pico-, and even femto-second pulsed lasers. In the pico- and femto-second regimes, additional difficulties would have to be addressed, including data acquisition, since digital converters are not fast enough to capture at those speeds, and the introduction of ballistic or wave equations to describe the thermal process for extracting the unknown thermal property.

When trying to measure the temperature field in an area of interest on a microelectronic device, a pumping laser is no longer required since the device is essentially heating itself. The probing part remains the same as before with the amplitude of the reflected laser light changing depending on whether the device is on or off. The relative difference in reflectivity recorded with the help of a photodetector yields the change in temperature after appropriate calibration. The advantage of using a laser and a fast photodetector is in the ability to capture fast transients. The disadvantages are (a) having to measure a single spatial point at a time and (b) not having a wide range of light colors available. To address these disadvantages, it is possible to use an array of photodetectors or more conveniently a highly sensitive CCD camera. This way it is possible to illuminate and measure an entire area at once as opposed to a single point at a time, and it is also possible to use a monochromater to select the light wavelength that provides the best possible TR response for a given surface material of interest.

Given the continued miniaturization of microelectronic devices coupled with the increase in their complexity and that of their thermal management solutions, a real need exists for measurement tools particularly at the deep submicron scales. Thermoreflectance methods have been advancing over the last decade and it is expected that these advances will continue and make real contributions to the evolutionary successes of microelectronic designs. The presentation will include details and examples for each of these approaches.

REFERENCES

[1] M. Farzaneh, K. Maize, D. Lüerßen, J. A. Summers, P. M. Mayer, P. E. Raad, K. P. Pipe, A. Shakouri, J. A. Hudgings, and R. Ram, "CCD-Based Thermoreflectance Microscopy: Principles and Applications," *J. Physics D: Applied Physics*, vol. 42, pp. 143001-143021, 2009.

[2] G. Tessier, M.-L. Polignano, S. Pavageau, C. Filloy, D. Fournier, F. Cerutti and I. Mica, "Thermoreflectance Temperature Imaging of Integrated Circuits: Calibration Technique and Quantitative Comparison With Integrated Sensors and Simulations," *J. Physics D: Applied Physics I*, vol. 39, pp. 4159-4166, 2006.

[3] P.L. Komarov, M.G. Burzo, and P.E. Raad, "CCD Thermoreflectance Thermography System: Methodology and Experimental Validation," *12th International IEEE THERMINIC Workshop*, Nice, France, Sep. 27-29, 2006.

25-27 September 2012, Budapest, Hungary

Nanoscale Thermal Metrology for Electronic Devices

Stefan Dilhaire, Wilfrid Claeys
LOMA - Université Bordeaux 1
351, cours de la Libération, Talence, 33405, France
stefan.dilhaire@u-bordeaux1.fr

Text unavailable at the time of printing.

25-27 September 2012, Budapest, Hungary

Challenges and Opportunities for Transient Thermal Imaging of Microelectronic Devices

Kazuaki Yazawa[1,3], Dustin Kendig[1], Daniel Hernandez[2], Kerry Maze[2], and Ali Shakouri[2,3]
[1]Microsanj, LLC. (Santa Clara, California, USA),
[2]University of California Santa Cruz (Santa Cruz, California, USA),
[3]Birck Nanotechnology Center, Purdue University (West Lafayette, Indiana, USA)
3287 Kifer Road, Santa Clara, CA 95051 USA
Phone: (408) 256-1255, Fax: (801) 405-3850, e-mail: kaz@microsanj.com

***Abstract*- In the development of electronic devices, transient thermal information is necessary to validate whether the entire device or a specific part of the device is operating properly. This becomes even more important for higher frequency operation of such devices. Sometimes, this is more important for the circuit designers, process engineers, or chip architects, rather than the thermal or package engineers who mostly need just a few points or locations of time-averaged temperature data during the design phase. For the time domain, recent developments in thermoreflectance imaging allow us to achieve a wide range of time scales from 50 nanoseconds up to milliseconds. This study provides an example of transient thermal imaging on a test chip along with the thermoreflectance imaging technique. The discussion covers the relationship of spatial resolution and time resolution considering the 'time constant' component of the chip. Spatial resolution for thermoreflectance is limited by the diffraction of the illuminating light, time resolution is limited by the high speed electrical signal management, and temperature resolution is limited by the signal-to-noise ratio. Time averaging, therefore, plays an important role in determining temperature resolution. The scientific inter-relationship between these factors adds to the complexity. The modeling work described above provides a guideline for the 'scope' and 'probe' depending on the dimensional scale of interest.**

I. INTRODUCTION

The thermal characterization and evaluation of microelectronic devices has been an ongoing challenge for designers of electronic products. Since the devices themselves are very small, the time response is very fast [1]. This fundamental characteristic has been convenient for shrinking the transistors to exponentially improve the functional density per cost, e.g. Flops per dollar for CPU [2]. In spite of the evolution of circuit and process technology, thermal characterization technology has fallen behind. For recent ICs, embedded thermal sensors provide timely temperature information at locations of interest [3]. But this is not suitable for devices other than logic ICs.

Additionally, emerging packaging technologies, such as multiple die-stacking, have further complicated thermal analysis of packaged devices. As a practical matter, the number of temperature sensors that can be embedded is approaching the limitation of the pins/area because of high density signal transactions between the chips in a stack [4]. Hotspot temperatures have been a limiting factor especially in multi core processors and still remain the main challenge to homogenizing the temperature in a chip. The hotspot temperature impacts device peak performance and reliability. The characterization of hotspots in short lengths of time is therefore essential to validate device design and reliability and to ensure quality control in production. The smaller scale and distributed heat sources prevalent in today's semiconductor devices require thermal mapping techniques with sub-micron spatial resolution and time resolution in sub microseconds.

II. THERMAL IMAGING

A. Characterization setup

A typical thermoreflectance imaging setup is illustrated in Figure 1. A microscope with a CCD camera in the optical assembly is located above the device-under-test (DUT), which is on the small X-Y table. An external LED light source is attached aside of the optics to illuminate the DUT. The entire setup is placed on a vibration isolation bed.

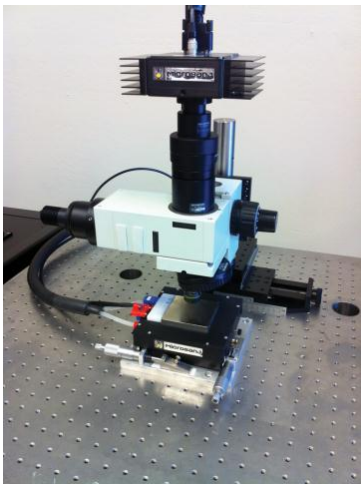


Figure 1: The basic setup of the thermoreflectance imaging.

With this technique, as outlined in Yazawa et al [5] and Burzo et al [6], thermoreflectance imaging is based on detecting the intensity of light reflectance and measuring the reflectance change in temperature change in theory. To commercialize the two-dimensional thermal mapping, our systems typically use CCD camera [7] except through silicon imaging since the illumination light can be visible in wave length. The external illumination by LED is actively controlled for shuttering the illumination synchronized timing with driving the device (DUT).

One of the advantages of thermoreflectance in comparison to IR emission imaging is that the light reflectivity is a function of the wave length. The reflectivity is a material dependent property and independent of the surface structure. Therefore, unlike emissivity for IR imaging, this reflectivity is constant for same material and fabrication process. This relationship is represented by the thermoreflectance coefficient (C_{th}) described as

$$C_{th}(\lambda) = \Delta T/(\Delta R(\lambda)/R(\lambda)) \text{ [K]} \quad (1)$$

Typical top surfaces of electronics devices have a limited variation of passivation layers. Once the C_{th} is determined by known temperatures, it can be used for a number of characterizations. Note that some metal materials show a steep change(s) of C_{th} in spectra, e.g. the thermoreflectance coefficient of copper changes its sign positive to negative at around 577 nm investigated by Rosi et al [8]. Also a similar characteristic on the C_{th} of semiconductors has been reported by Matatagui et al [9].

B. Transient thermal imaging

Microelectronics devices typically have complex workloads in time. The electrical capacitances of each component in such devices are designed to work correctly in time. However, the thermal mass or the thermal capacitances are sometimes not considered during the design of the circuitry. The temperature distribution caused by such complex operational modes remains unknown during the design phase. The temporal thermal performances therefore cannot be characterized until the devices are manufactured. Local and temporal temperatures of the components in the devices are critical for predicting the performance and reliability of the circuit.

Following Figures 2 to 4 show an example of a simple circuit designed as a thermal test vehicle (TTV) for thermal packaging testing. At steady state (see Figure 2), the temperature of the heater blocks (4 arrays x 2 heaters/array) and the substrate are quite similar with moderate external heat transfer. In a short time after applying the pulse current (100 µsec), the thermal map shows a different image than the steady state. The heaters show peak temperature (red) while the substrate is close to room temperature (blue). This is due to the thermal mass of the substrate as its thermal diffusion speed is not fast enough to change temperature up to that of the heaters. Figure 4 shows the local temperature response of the heater and the substrate.

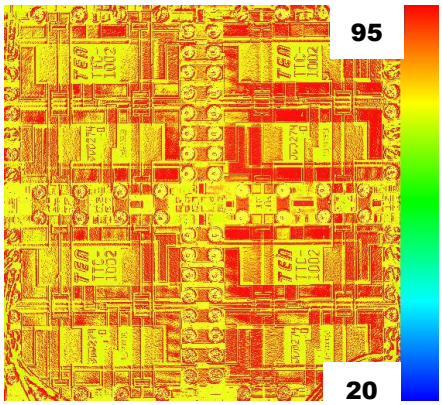


Figure 2: Steady state TR image for 29 V quasi DC bias (0.5 Hz). Image is taken in 2 minutes average.

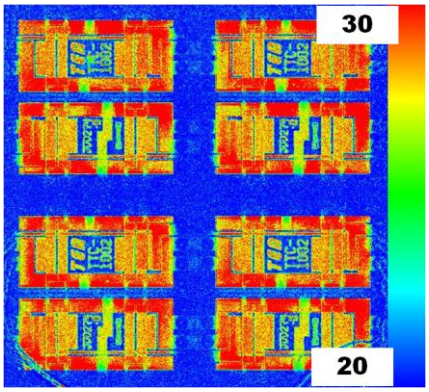


Figure 3: Transient TR image @ 100 µs for 1 ms and 60 V for the same device as Figure 2.

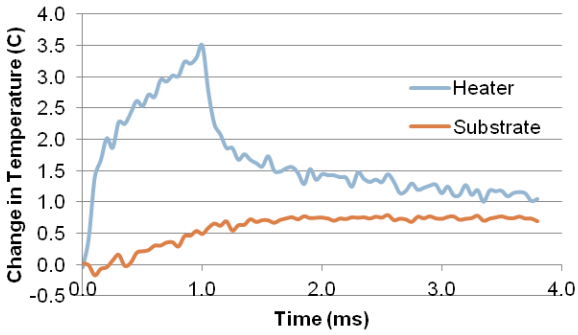


Figure 4: Thermal response chart of the same data as Figure 3. 10 µs/data point for time axis. The applied bias is 30V.

C. High speed thermal imaging and spatial resolution

Transient imaging gives richer information of the temperature in transition. When investigating small time scale events, the spatial resolution must also be high enough to observe the thermal diffusion in the device (DUT) itself, (See Eq. 2). Otherwise, the pixels will display an irregular temperature step, which actually induces unwanted noise in the thermal map.

The one dimensional diffusion equation is described as

$$\frac{\partial T}{\partial t} = \frac{1}{\alpha}\frac{\partial^2 T}{\partial x^2} \tag{2}$$

where: T is temperature, t is time, and α is the thermal diffusivity. Time resolution for the location of interest must have a relation to the spatial resolution x and temperature resolution ΔT. Here we consider a time step, Δt, from a certain point of time, t_0, and the resulting temperature change ΔT from T_0. Spatial resolution is defined as the number of pixels, n, in the length, L, of the region of interest. The cross sectional area of the region is A.

$$T_0\left(1-e^{\left(-\frac{t_0+\Delta t}{rc}\right)}\right)-T_0\left(1-e^{\left(-\frac{t_0}{rc}\right)}\right)\le \Delta T \tag{3}$$

$$x = L/n$$

Since the temperature change in time at a certain location is not constant, the largest change must be considered. It may be found at or near the time constant. Note that this is only an approximation and may not always be accurate since the system is discrete not one-dimensional. The time constant of the unit pixel, rc, is found as,

$$rc = \frac{L}{n\beta A}\frac{\rho C_p A L}{n} = \frac{L^2}{n^2\alpha} \tag{4}$$

Solving Eq. (3)

$$\frac{\Delta t}{rc} \le -1-\ln\left(e^{-1}-\frac{\Delta T}{T_0}\right) \tag{5}$$

Rewriting by using known parameters,

$$\alpha\frac{\Delta t}{x^2} \le -1-\ln\left(e^{-1}-\frac{\Delta T}{T_0}\right) \tag{6}$$

If we assume a temperature error of 1% or less, the right term of the above equation becomes 0.0276 and is independent of the diffusivity or length resolution. Thus as a design guide, it is easier to remember that the time resolution must be set to

$$\Delta t = \frac{0.02}{\alpha}x^2 \tag{7}$$

As an example, for silicon, α=8.4x10^{-5} m^2/s, therefore the time resolution is approximately 24 nsec for a 10 μm spatial resolution and a 1% temperature reading error.

High speed imaging in a range of more than a few MHz (sub microseconds of time resolution) also requires a special consideration for electrical circuitry. As is well known, high speed signals require impedance matching, especially for driving the device and interlocking the illumination timing. Unmatched impedance between the current driver and the device will cause signal reflections in the circuit and typically lead to a large overshoot. This could cause a shift in the lock-in timing or a completely different bias condition than intended.

D. Signal-to-noise ratio and averaging

It has been demonstrated, among experts in this technical area, that the deviation of the captured light intensity converges to a specific value if we average the intensity over a significant amount of time. Since the C_{th} is usually very small, e.g. in the range of 1e-4 for metals and 1e-5 for semiconductors, the deviation of the light intensity is usually widespread for a uniform temperature with an acceptable time average, such as a few minutes. We have done an analysis of the averaging on the semiconductor device. The tested device was an LED device with a 4 x 4 array of junctions on the chip. Figure 5 shows the device structure and Figure 6 shows the histogram of the reflectance intensity for a 15 minute averaging time. The reason for choosing this device was to separate the Electro Luminescence (EL) signal, and the reflectance of the illuminated external light. The details of this separation technique are presented in ref. 10. The illumination wave length was determined to be 405 nm for the separation with a 457 nm blue LED device. The averaging analysis was carried out with 5x optics. Figure 7 shows the standard deviation of the reflectance intensity ($\Delta R/R$) as a function of averaging time. The data points fall on the square root trend line. If we project a longer averaging time, the data points will ultimately converge. For this particular LED sample, the standard deviation is reduced by more than 50% for an averaging time of 15 min (900 sec) compared to a 3 min (180 sec) averaging time.

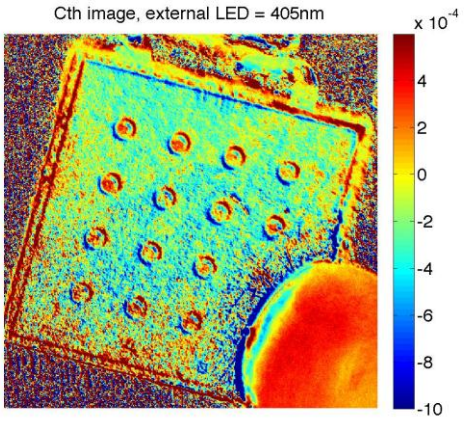


Figure 5: Example of a thermal image for a multi-junction LED device

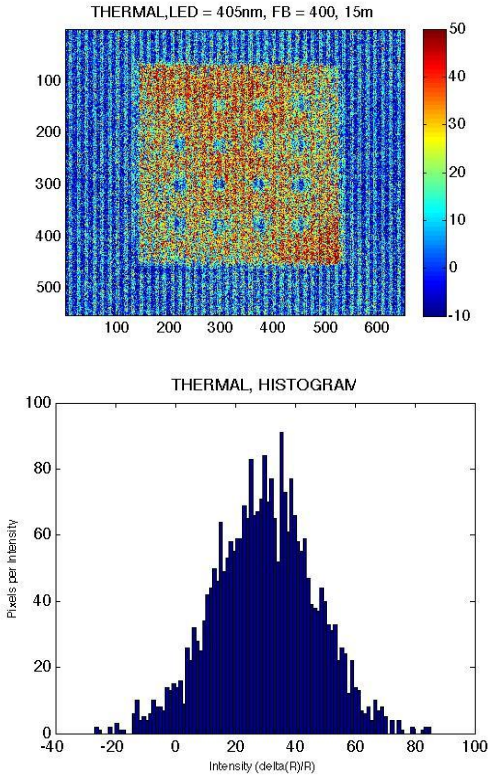


Figure 6: Example of histogram of intensity for an averaging time of 15 minutes.

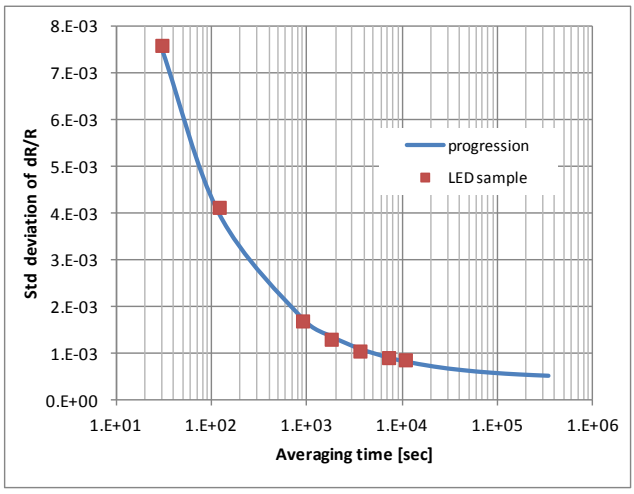


Figure 7: Standard deviation of ΔR/R (σ) as a function of the averaging time. The σ converges to 4.43×10^{-4} for an extremely long averaging time and is an order of magnitude smaller than the mean value, 3.3×10^{-3} for ΔR/R.

III. CONCLUSIONS

We described an example of the transient thermal characterization using the thermoreflectance imaging technique. With a 50 ohm impedance match for high speed signals, the technology enables a time resolution ranging from nanoseconds to milliseconds. With consideration of the thermal diffusion scale, higher time resolution requires higher spatial resolution. We also showed that this relationship is device material dependent. The spatial resolution for thermoreflectance is also limited by the lighting diffraction while temperature resolution is limited by the signal-to-noise ratio. We showed the effectiveness of time averaging and the trend to converge to a reasonable value for the standard deviation. In this case, the standard deviation was one order of magnitude smaller than the mean value even though there was only weak reflection from the semiconductor device. The results of the study provide a guideline for considering thermoreflectance transient thermal imaging for high speed devices. We described how the time resolution and the spatial resolution relate to the temperature reading accuracy and also provided a relationship for determining the time resolution for an available spatial resolution given the thermal diffusivity for the device material and a 1% temperature reading error.

ACKNOWLEDGMENT

Authors give special thanks to Mr. Siegel for providing the thermal test vehicle, and to Dr. Christofferson for providing images of the Thermal test sample.

REFERENCES

[1] K. Yazawa, D. Kendig, J. Christofferson, A. Marconnet, and A. Shakouri, "Ultrafast Transient and Steady State Thermal Imaging of CMOS Integrated Circuit Chips with and without Package Thermal Boundaries", Proceedings of ITherm 2012 - Inter Society Conference on Thermal Phenomena, (2012).

[2] M.L. Barton and G.R. Withers, "Computing performance as a function of the speed, quantity, and cost of the processors", Proceedings of the 1989 ACM/IEEE Conference on Supercomputing, (1989).

[3] 10. D. Pham, T. Aipperspach, D. Boerstler, M. Bolliger, R. Chaudhry, D. Cox, P. Harvey, P. M. Harvey, H. P. Hofstee, C. Johns, J. Kahle, A. Kameyama, J. Keaty, Y. Masubuchi, J. Pille, S. Posluszny, M. Riley, D. Stasiak, M. Suzuoki, O. Takahashi, J. Warnock, S. Weitzel, D. Wendel, K. Yazawa, Overview of the Architecture, Circuit Design, and Physical Implementation of a First-Generation CELL Processor, IEEE Journal of Solid-State Circuits, Vol. 41, No. 1, pp.179-196, (2006).

[4] W.R. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A.M. Sule, M. Steer, P.D. Franzon, "Demystifying 3D ICs: the pros and cons of going vertical", IEEE Design & Test of Computers, Vol. 22, Issue 6, pp. 498-510, (2005).

[5] K. Yazawa and A. Shakouri, Ultrafast Submicron Thermoreflectance Imaging, Electronics Cooling Magazine Spring Issue 2011, ITEM Publications, pp 10, 12-15, (2011).

[6] M. G. Burzo, P. L. Komarov, P. E. Raad, "Thermo-Reflectance Thermography for Submicron Temperature Measurements", Electronics Cooling Magazine Vol. 14, February 1st, ITEM Publications, (2008).

[7] M. Farzaneh, K. Maize, D. Lüerßen, J.A. Summers, P.M. Mayer, P. E. Raad, K.P. Pipe, A. Shakouri, R. J. Ram and J. A. Hudgings, "CCD-based thermoreflectance microscopy: principles and applications", Journal of Physics D: Applied Physics 42, 143001, 20, (2009).

[8] R. Rosei and D. W. Lynch, "Thermomodulation Spectra of Al, Au, and Cu", Physical Review, Vol. 5,No. 10, (1972).

[9] E. Matatagui, A. G. Thompson, and M. Cardona,"Thermorefelectance in Semiconductors", Physical Review, Vol. 176, No. 3, (1968).

[10] D. Kendig, K. Yazawa, and A. Shakouri, "Thermal Imaging of Encapsulated LEDs", Proceedings of the 27th IEEE SEMI-THERM Symposium, pp 310-313, (2011).

25-27 September 2012, Budapest, Hungary

Development of pulsed light heating thermoreflectance methods under configurations of rear heating / front detection and front heating / front detection

Tetsuya Baba, Naoyuki Taketoshi, Takashi Yagi
National Metrology Institute of Japan (NMIJ)
National Institute of Advanced Industrial Science and Technology (AIST),
Umezono 1-1-1, Tsukuba, Ibaraki 305-8563, Japan
Telephone: +81-29-861-4394, fax: +81-298-4323, e-mail: t.baba@aist.go.jp

***Abstract*- Thermoreflectance technique holds promise where radiation thermometry faces potential limitations: in temperature measurements around room temperature and of objects requiring sub-micron resolution. Time response of 500 kHz has been experimentally demonstrated.**

In order to measure thermal diffusivity of thin films, National Metrology Institute of Japan (NMIJ) has developed rear face heating / front face detection picosecond pulsed light heating thermoreflectance methods. This configuration is essentially equivalent to the laser flash method which is the standard method to measure thermal diffusivity of bulk materials. Thermal diffusivity values of thin films can be determined reliably from the thickness of a thin film and the heat diffusion time across a thin film.

For application to variety of thin films synthesized on non-transparent substrates, systems have been developed for measuring thermal diffusivity and / or thermal effusivity of thin films under front face heating / front face detection configuration. Area of a diameter of several 10 μm on thin film front face is heated by pulsed light and the same position is irradiated by the probe beam. Then, the history of front face temperature is observed by the thermoreflectance method. In this method, the thermal diffusivity can be calculated from the cooling rate after the pulse heating.

In this paper, pulsed light heating thermoreflectance methods under configurations of "rear heating / front detection" and "front heating/front detection" are compared in order to discuss their complementary role, advantages, limitations, and calibration method.

I. INTRODUCTION

Non-contact optical methods are essential in dynamic temperature measurements. Applications of radiation thermometry as well as reflectance thermometry have been applied to surface temperature measurement of semiconductor device production processes and device evaluations [1-6].

Reflectance thermometry has been applied to measure thermophysical property of thin films because it can measure temperature change much faster than other thermometry method such as radiation thermometry. For example, NMIJ developed a pulsed light heating thermoreflectance method where a thin film on transparent substrate is heated by a picosecond pulsed laser [7-8] or nanosecond [9] pulsed laser from the substrate side, and temperature change of the specimen front face is measured by change of the reflectivity. Since geometrical configuration of this method is same as the conventional laser flash method measuring thermal diffusivity of bulk materials, thermal diffusivity of the thin film can be determined reliably based on analytical solution of one dimensional heat diffusion [10-11].

However, many thin films of practical use are synthesized on non-transparent substrate. In these cases, the thin film must be heated from open surface and temperature must be detected by the reflectance from the open surface of the thin film since the boundary between the thin film and the substrate cannot be irradiated by the laser beam

Temperature response after picosecond pulse heating by mode locked dye laser was observed by Paddock and Eesley for the first time [12]. However, it is not easy to calculate thermal diffusivity values from the temperature response curves observed by this method as discussed later.

On the other hand, measurements under front face heating / front face detection heating configuration have been successfully applied to dielectric and semiconductor thin films from 500 nm to 10μm thick by nanosecond pulsed light heating with pulse duration of the order of 10 ns [13, 14].

In this article, the problems and the approach to calculate the thermophysical property of the thin film quantitatively from the temperature response curve observed by front heating / front detection pulsed light heating thermoreflectance method especially heated by picosecond picosecond light pulse [15, 16].

II. ANALYTICAL MODEL

A. Response function method

The temperature response of the semi-infinite solid material after an impulse heating of unit intensity to the solid surface at time, t', is represented by the following Green's

function [10].

$$\mathrm{G}(x,t\mid 0,t') = \frac{1}{b\sqrt{\pi(t-t')}}\exp\left(-\frac{x^2}{4\alpha(t-t')}\right), \quad (1)$$

where x is the distance of the observing position from the solid surface, t is the time of observation. The thermophysical properties of the solid material are represented as α is the thermal diffusivity, b the thermal effusivity defined as $b=\sqrt{\lambda c\rho}=c\rho\sqrt{\alpha}$, λ is the thermal conductivity, c is the specific heat capacity, and ρ is the density.

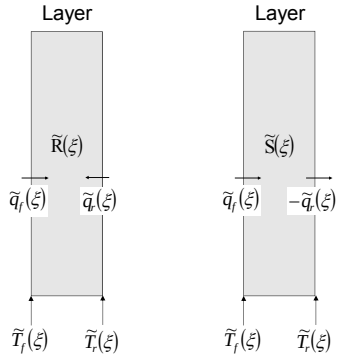


Fig. 1. Transfer matrix (left) and quadrupole matrix (right)

When the front face of a flat layer is pulse-wise heated, transient temperature distribution across the layer is given by the following Green's function.

$$\mathrm{G}(x,t\mid 0,t') = \frac{1}{c\rho d}\left[1+2\sum_{n=1}^{\infty}\cos\frac{n\pi x}{d}\cdot\exp\left[-n^2\pi^2\frac{t-t'}{\tau}\right]\right], \quad (2)$$

where the definition of $x, t, t', c, \rho, \alpha$ are same as equation (1), d is the thickness of the layer, τ is the characteristic time of thermal diffusion across the layer defined as $\tau = d^2/\alpha$.

The temperature response of the point x at the time t by the arbitrary function $f(t)$ is expressed by the following convolution integral.

$$T(x,t) = \int_{-\infty}^{t}\mathrm{G}(x,t\mid 0,t')\,f(t')dt' . \quad (3)$$

When the steady state temperature of the layer is constant, T_0, the temperature response at the both faces at time t are represented by the following equations under incoming heat flux density $q_f(t)$ to the front face and $q_r(t)$ to the rear face [10].

$$\boldsymbol{T}(t) = \boldsymbol{T}_0 + \int_0^t \mathrm{R}(t-t')\boldsymbol{q}(t')dt', \quad (4)$$

$$\boldsymbol{T}_0 = \begin{bmatrix} T_0 \\ T_0 \end{bmatrix}, \boldsymbol{T}(t) = \begin{bmatrix} T_f(t) \\ T_r(t) \end{bmatrix}, \mathrm{R}(t) = \begin{bmatrix} \mathrm{R}_{ff}(t) & \mathrm{R}_{fr}(t) \\ \mathrm{R}_{fr}(t) & \mathrm{R}_{rr}(t) \end{bmatrix}, \boldsymbol{q}(t) = \begin{bmatrix} q_f(t) \\ q_r(t) \end{bmatrix},$$

where $\mathrm{R}(t)$ is the impulse response function matrix of a parallel layer adiabatic to the environment and the element $\mathrm{R}_{ij}(t)$ is the temperature rise at the face "i" at time "t" after the unit impulse heating to the face "j" at time "0". The subscript i or j represents either front face "f" or rear face "r".

$$T_f(t) = T_0 + \int_0^t\left(\mathrm{R}_{ff}(t-t')q_f(t') + \mathrm{R}_{fr}(t-t')q_f(t')\right)dt', \quad (5)$$

$$T_r(t) = T_0 + \int_0^t\left(\mathrm{R}_{rf}(t-t')q_f(t') + \mathrm{R}_{rr}(t-t')q_r(t')\right)dt'. \quad (6)$$

The initial temperature T_0 can be assumed to be zero for simplicity without losing generality.

The Laplace transform of $f(t)$ is expressed as $\tilde{f}(\xi)$ defined by the following equation where ξ is the Laplace parameter.

$$\tilde{f}(\xi) = \int_0^{\infty}\exp(-\xi t)f(t)dt . \quad (7)$$

Laplace transform of the equation(4) is,

$$\tilde{\boldsymbol{T}}(\xi) = \tilde{\mathrm{R}}(\xi)\cdot\tilde{\boldsymbol{q}}(\xi), \quad (8)$$

$$\tilde{\mathrm{R}}(\xi) = \begin{bmatrix} \tilde{\mathrm{R}}_{ff}(\xi) & \tilde{\mathrm{R}}_{fr}(\xi) \\ \tilde{\mathrm{R}}_{rf}(\xi) & \tilde{\mathrm{R}}_{rr}(\xi) \end{bmatrix}. \quad (9)$$

where $\tilde{\mathrm{R}}(\xi)$ is called as the transfer matrix as shown in left side of Fig. 1.

B. Uniform single layer

Since Green's function of a uniform single layer is expressed by equation (2), the elements of the impulse response matrix are expressed as follows,

$$\mathrm{R}_{ff}(t) = \mathrm{R}_{rr}(t) = \frac{1}{c\rho d}\left[1+2\sum_{n=1}^{\infty}\exp\left(-n^2\pi^2\frac{t}{\tau}\right)\right], \quad (10)$$

$$\mathrm{R}_{fr}(t) = \mathrm{R}_{rf}(t) = \frac{1}{c\rho d}\left[1+2\sum_{n=1}^{\infty}(-1)^n\exp\left(-n^2\pi^2\frac{t}{\tau}\right)\right]. \quad (11)$$

Laplace transformation of equation (10) and (11) are

$$\tilde{\mathrm{R}}_{ff}(\xi) = \tilde{\mathrm{R}}_{rr}(\xi) = \frac{1}{c\rho d}\left[\frac{1}{\xi}+2\sum_{n=1}^{\infty}\frac{1}{\xi+n^2\pi^2/\tau}\right] = \frac{1}{b\sqrt{\xi}}\coth\left(\sqrt{\xi\tau}\right) \quad (12)$$

$$\tilde{\mathrm{R}}_{fr}(\xi) = \tilde{\mathrm{R}}_{rf}(\xi) = \frac{1}{c\rho d}\left[\frac{1}{\xi}+2\sum_{n=1}^{\infty}\frac{(-1)^n}{\xi+n^2\pi^2/\tau}\right] = \frac{1}{b\sqrt{\xi}}\operatorname{cosech}\left(\sqrt{\xi\tau}\right) . \quad (13)$$

C. Quadrupole matrix

Explicit expression of equation (8) is

$$\begin{bmatrix} \tilde{T}_f(\xi) \\ \tilde{T}_r(\xi) \end{bmatrix} = \begin{bmatrix} \tilde{\mathrm{R}}_{ff}(\xi) & \tilde{\mathrm{R}}_{fr}(\xi) \\ \tilde{\mathrm{R}}_{rf}(\xi) & \tilde{\mathrm{R}}_{rr}(\xi) \end{bmatrix}\begin{bmatrix} \tilde{q}_f(\xi) \\ \tilde{q}_r(\xi) \end{bmatrix}. \quad (14)$$

As shown in Fig. 1, this equation is transformed to a quadrupole matrix expression which relates the pair of the heat flux density and temperature at the front face to the same pair at the rear face as follows.

$$\begin{bmatrix} -\tilde{q}_r(\xi) \\ \tilde{T}_r(\xi) \end{bmatrix} = \tilde{\mathrm{S}}(\xi)\begin{bmatrix} \tilde{q}_f(\xi) \\ \tilde{T}_f(\xi) \end{bmatrix}, \quad (15)$$

$$\tilde{\mathrm{S}}(\xi)=\begin{bmatrix}\cosh\left(\sqrt{\tau\xi}\right) & -b\sqrt{\xi}\cdot\sinh\left(\sqrt{\tau\xi}\right)\\ -\frac{1}{b\sqrt{\xi}}\cdot\sinh\left(\sqrt{\tau\xi}\right) & \cosh\left(\sqrt{\tau\xi}\right)\end{bmatrix}, \quad (16)$$

where quadrupole matrix $\tilde{\mathrm{S}}(\xi)$ is a function of thermal effusivity b and heat diffusion time $\tau = d^2/\alpha$ [10]. It should be noted that $\tilde{q}_f(\xi)$ and $-\tilde{q}_r(\xi)$ are the vectors to the same direction as shown in right side of Fig. 1.

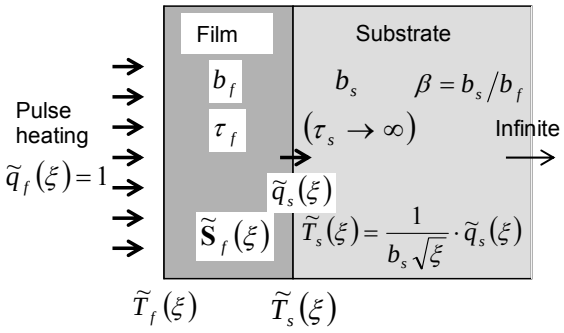


Fig. 2. Temperature response, heat flux density, and thermophysical properties in a film/substrate

D. *Film/substrate model*

A thin film synthesized on a semi-infinite substrate is illustrated in Fig. 2. The parameters to specify this film/substrate model are thickness of the film d_f, thermal diffusivity of the film α_f, thermal effusivity of the film b_f, thermal diffusivity of the substrate α_s, and he thermal effusivity of the substrate b_s [10]. Temperature response at the film substrate boundary, $T_s(t)$, is calculated as convolution integral of the heat flux density effusing into the substrate, $q_s(t)$, and the Green's function of the semi-infinite substrate as shown in equation (1) .

$$T_s(t,0)=\int_0^t \frac{1}{b_s\sqrt{\pi(t-t')}}q_s(t')dt' . \quad (17)$$

Laplace transform of this equation is

$$\tilde{T}_s(\xi)=\frac{1}{b_s\sqrt{\xi}}\cdot\tilde{q}_s(\xi) . \quad (18)$$

The pair of $q_s(t)$ and $T_s(t)$ is related to the pair of the heat flux density onto the film surface, $q_f(t)$, and the temperature at the film surface, $T_f(t)$, via the quadrupole matrix of the film, $\tilde{\mathrm{S}}_f(\xi)$, as the equation (15).

$$\begin{bmatrix}\tilde{q}_s(\xi)\\ \tilde{T}_s(\xi)\end{bmatrix}=\tilde{\mathrm{S}}_f(\xi)\cdot\begin{bmatrix}\tilde{q}_f(\xi)\\ \tilde{T}_f(\xi)\end{bmatrix}. \quad (19)$$

Substituting $q_f(t)=\delta(t)$, the delta function at time "0", the simutaneous equations (18) and (19) are solved and $T_f(t)$ is expressed as follows.

$$\tilde{T}_f(\xi)=\frac{1}{b_s\sqrt{\xi}}\frac{\coth(\sqrt{\xi\tau_f})+\beta}{\coth(\sqrt{\xi\tau_f})+\beta^{-1}}, \quad (20)$$

where the thermal diffusion time across the film is defined as $\tau_f = {d_f}^2/\alpha_f$ and the thermal effusivity ratio of the substrate to the film is defined as $\beta = b_s/b_f$ [10].

The equation (20) is expressed as the function of $\exp\left(-2\sqrt{\xi\tau_f}\right)$ by substituting $\coth(x)=[(1+\exp(-2x))]/[1-\exp(-2x)]$ as,

$$\tilde{T}_f(\xi)=\frac{1}{b_f\sqrt{\xi}}\frac{1+\gamma\exp\left(-2\sqrt{\xi\tau_f}\right)}{1-\gamma\exp\left(-2\sqrt{\xi\tau_f}\right)}$$
$$=\frac{1}{b_f\sqrt{\xi}}\left(1+2\sum_{n=1}^{\infty}\gamma^n\exp\left(-2n\sqrt{\xi\tau_f}\right)\right), \quad (21)$$

where $\gamma=\left(b_f-b_s\right)/\left(b_f+b_s\right)=(1-\beta)/(1+\beta)$.

The surface temperature is obtained by Inverse Laplace transformation of equation (21).

$$T_f(t)=\frac{1}{b_f\sqrt{\pi t}}\left(1+2\sum_{n=1}^{\infty}\gamma^n\exp\left(-n^2\frac{\tau_f}{t}\right)\right). \quad (22)$$

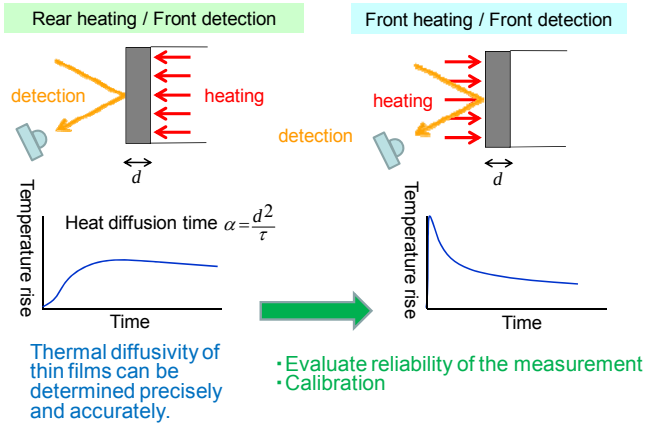


Fig. 3. Pulsed light heating thermoreflectance methods

II. Analysis of Observed Temperature Response

A. *Pulsed light heating thermoreflectance method*

As shown in left side of Fig. 3, temperature response of the surface opposite to the heated face of the thin film is observed after heat diffuses across the known thickness by the pulsed light heating thermoreflectance method of the rear face heating / front heating configuration, where thermal diffusivity of the thin film is calculated by the analysis principally same as the laser flash method which is the standard method to measure the thermal diffusivity of bulk material reliably [11].

On the other hand, as shown in right side of Fig. 3, thermal diffusivity of the first layer and thermal effusivity of the second layer are calculated from difference of the cooling rate of the specimen front face by the pulsed light heating

thermoreflectance method of the front heating / front detection configuration [12-16].

The heating pulsed light must be absorbed and sensitivity of thermoreflectance temperature measurement must be high enough to observe temperature response with good signal to noise ratio. Since most materials do not satisfy these conditions, it is common solution to synthesize a metallic thin film which satisfy these conditions (e.g., molybdenum) of thickness around 100nm in front face of the specimen [10].

As shown in Fig. 4, heat absorbed at the surface of the metallic thin film diffuses across the metallic thin film first. Then, it effuses into the second layer thin film to be measured and diffuses across the second thin film. Finally, heat effuses into the substrate and the temperature returns to original level before the pulse heating when time passes long enough. The shape of the cooling curve of the surface changes dependent on thermal diffusivity of the first layer, thermal diffusivity and thermal effusivity of the second layer, and thermal effusivity of the substrate.

Here, it is assumed that thickness of the substrate is thick enough and the boundary thermal resistance between the metallic thin film and the second layer thin film, the boundary thermal resistance between the second layer thin film and substrate are negligibly small.

If all thermal effusivities of the metallic thin film, the second layer thin film, the substrate are equal each other, the surface temperature changes following the middle solid curve in Fig. 4 which is identical to the curve to be observed for semi-infinite uniform thermal effusivity. Only information obtained from this observation is that the thermal effusivities of the second layer thin film and the substrate are close to that of the metallic thin film.

When the thermal effusivity of the second layer thin film is smaller than a thermal effusivity of the metallic thin film, and the thermal effusivity of the substrate is smaller than the thermal effusivity of the second layer thin film, cooling is suppressed by heat effusion into the second layer and the substrate as the top curve in Fig. 4.

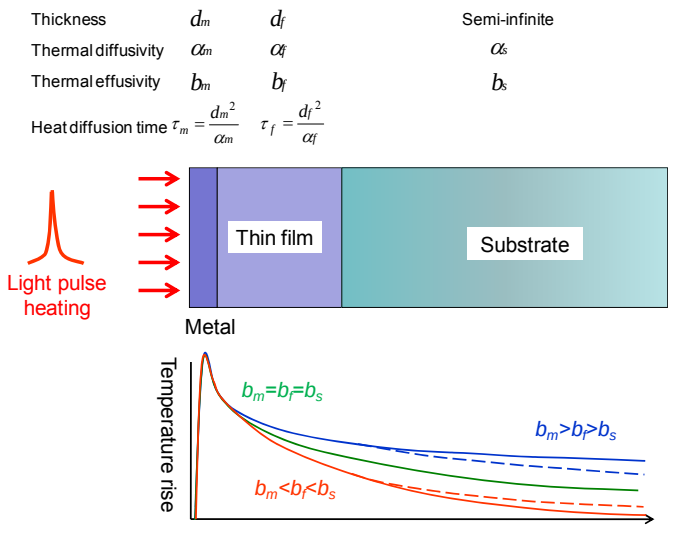


Fig. 4 Temperature responses at the surface of metal thin film after heated by a light pulse

On the contrary, if the thermal effusivity of the second layer thin film is larger than a thermal effusivity of the metallic thin film, and the thermal effusivity of the substrate is larger than the thermal effusivity of the second layer thin film, cooling is enhanced by heat effusion into the second layer and the substrate as the bottom curve in Fig. 4.

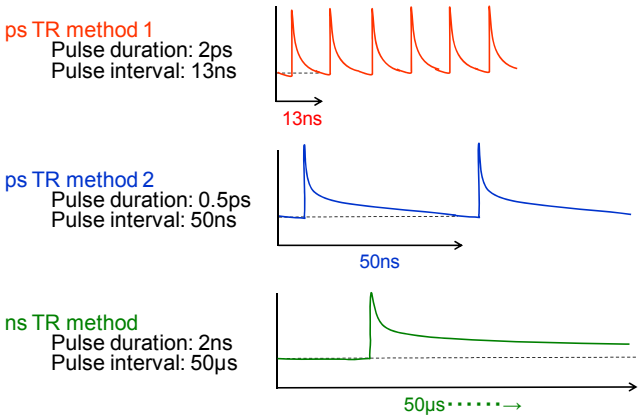


Fig. 5 Temperature responses dependent on repetition interval of heating light pulses

B. *Effect of the repetition pulsed light heating*

Discussion of the previous section assumes that the specimen is heated by a "single" pulsed light and the temperature response is observed long time enough for the temperature response returns to the original level before the pulsed light heating. The assumption of a single pulsed light is valid for thermal diffusivity measurements of bulk materials by the laser flash method.

In the case of picosecond pulsed light heating thermoreflectance method, a specimen heated by pulse train of 2 ps duration with repetition of 76 MHz emitted from a mode-locking titanium sapphire laser. Then, temperature response follows the same curve repeated with the interval of 13 ns as shown by top trace of Fig. 5 [17].

A new picosecond thermoreflectance system has been developed using the fiber laser. The pulse duration is 0.5 ps and repetition of the pulsed light train is 20MHz. Then, temperature response follows the same curve repeated with the interval of 50 ns as shown by middle trace of Fig. 5 [11].

In the case of nanosecond pulsed light heating thermoreflectance method, the pulse duration is 2 ns and repetition of the pulsed light train is 20 kHz. Then, temperature response repeats with the interval of 50 μs as shown by bottom trace of Fig. 5. Since this interval is considered as infinitely long for the thin film specimen to thickness several 100nm, the specimen temperature returns to the initial value and converges flat. This means that the observed temperature response curve agrees with the temperature response after "single" pulsed light irradiation.

However, the temperature response curve just after pulse heating observed by the nanosecond pulsed light heating instrument might be distorted by time resolution of the order of 10 ns which is not fast enough to observe heat diffusion across thin films thinner than 1 μm [18].

Fig. 5 demonstrates that apparent cooling rate changes depending on pulse irradiation interval of three types of pulsed light heating thermoreflectance instruments.

In the case of picosecond pulsed light thermoreflectance method, it must be considered about the distortion of the signal by the phase detection in addition to the effect of the lock-in detection [17, 19].

IV. SUMMARY

It is not easy to calculate thermophysical property values quantitatively from apparent temperature response curve observed by the picosecond pulsed light heating thermoreflectance methods. Reliable correction method should be established in order to calculate a thermophysical property quantitatively from the temperature response curve of the pulsed light heating thermoreflectance method of the front face heating / front detection.

On the contrary, thermal diffusivity can be calculated quantitatively by the rising part of the temperature response curve observed by the picosecond pulsed light thermoreflectance method of the rear face heating / front detection instead of the cooling part of the temperature response curve.

Comparison with the thermophysical properties obtained from the front face heating / front detection measurement with those obtained from the rear face heating / front detection method for the same specimen is the key information in order to establish quantitative measurement the front face heating / front detection picosecond pulsed light heating thermoreflectance method.

ACKNOWLEDGMENT

The authors would like to express their sincere thanks to Dr. Juntaro Ishii, Dr. Yoshiro Yamada and Dr. Yukiko Shimizu about information of advanced radiation thermometry and thermoreflectance thermometry method developed in National Metrology Institute of Japan.

REFERENCES

[1] W. Claeys, S. Dilhaire, et al., "Laser probing of thermal behavior of electronic components and its application in quality and reliability testing." Microelectronic Engineering 24 (1-4), 1994, pp. 411-420.

[2] S. Dilhaire, S. Grauby, et al., "Calibration procedure for temperature measurements by thermoreflectance under high magnification conditions." Applied Physics Letters 84 (5), 2004, pp. 822-824.

[3] Y. Shimizu, J. Ishii, et al., "Reflectance thermometry for microscale metal thin films." Japanese Journal of Applied Physics 46, 2007, p. 3117.

[4] J. Christofferson, K. Maize, et al., "Microscale and nanoscale thermal characterization techniques," Journal of Electronic Packaging 130, 2008, 041101.

[5] J. Christofferson, K. Yazawa, et al. "Picosecond transient thermal imaging using a CCD based thermoreflectance system," Proceedings of the 14th International Heat Transfer Conference IHTC14, Washington, DC, USA, August 8-13, 2010.

[6] Y. Yamada, and J. Ishii, "Emissivity compensation utilizing radiance distribution in thermal images for temperature measurement of electronic devices," Japanese Journal of Applied Physics 50 (11), 2011, 11RE04.

[7] N. Taketoshi, T. Baba and A. Ono, "Observation of heat diffusion across submicrometer thin films using a picosecond thermoreflectance technique," Japanese Journal of Applied Physics., 38, 1999, pp. L1268-1271.

[8] N. Taketoshi, T. Baba and A. Ono, "Development of a thermal diffusivity measurement system for metal thin films using a picosecond thermoreflectance technique," Meas. Sci. Technol., IOP, Bristol, 12, 2001, pp.2064-2073.

[9] T. Baba, " General needs on nanoscale thermal metrology and the japanese program on this subject," Proceedings of Workshop on Thermal Investigations of ICs and Systems (Therminic 2004), Sophia Antipolis, France, 29 September – 1 October 2004, pp. 241-249.

[10] T. Baba, "Analysis of one-dimensional heat diffusion after light pulse heating by the response function method," Japanese Journal of Applied Physics. 48 (2009) 05EB04.

[11] T. Baba, N. Taketoshi, and T. Yagi, "Development of ultrafast laser flash methods for measuring thermophysical properties of thin films and boundary thermal resistances," Japanese Journal of Applied Physics 50 (11), 2011, 11RA01.

[12] C. A. Paddock and G. L. Eesley, "Transient thermoreflectance from thin metal films," J. Appl. Phys., AIP, Melville, 60, 1986, pp.285-290.

[13] M. G. Burzo, P. L. Komarov, and P. E. Raad, "Influence of the metallic absorption layer on the quality of thermal conductivity measurements by the transient thermo-reflectance method," Microelectronics journal 33 (9), 2002, pp. 697-703

[14] P. L. Komarov, and P. E. Raad, "Performance analysis of the transient thermo-reflectance method for measuring the thermal conductivity of single layer materials," International journal of heat and mass transfer 47 (14), 2004, pp. 3233-3244.

[15] N. Taketoshi, T. Baba, A. Ono, "Picosecond thermoreflectance measurements of thermal diffusion in film/substrate two layer systems," Thermal Conductivity24 / Thermal Expansion 12, vol.24, 1999, pp.2 89-302.

[16] T. Baba, N. Taketoshi, "Analysis of heat diffusion in multi-layer thin films by a response function method," Proc. Eurotherm 57, Poitiers, 1999, pp. 285-292.

[17] N. Taketoshi, T. Baba and A. Ono, "Electric delay technique in a picosecond thermoreflectance method for thermophysical property measurements of thin films," Rev. Sci. Instrum., 76, 2006, pp. 1-8.N.

[18] T. Yagi, K. Ohta, K. Kobayashi, et al, "Thermal diffusivity measurement in a diamond anvil cell using a light pulse thermoreflectance technique, " Measurement Science & Technology , 22, 2011, 024011

[19] Taketoshi, T. Baba, E. Schaub, A. Ono, "Homodyne detection technique using spontaneously generated reference signal in picosecond thermoreflectance measurements," Rev. Sci. Instrum., 74, 2003, pp. 5226-5230

25-27 September 2012, Budapest, Hungary

Measurement of the thermal impedance of GaN HEMTs using "the 3ω method"

Mustafa Avcu
7 rue Jules VALLES, 19100 Brive La Gaillarde, France
mustafa.avcu@brive.unilim.fr

***Abstract*- This paper deals with an accurate characterization method dedicated to the determination of the thermal impedance of Gallium Nitride based High Electron Mobility Transistors (GaN HEMTs). The method is inspired by the "3ω method" initially proposed by D.G. Cahill in order to measure the thermal conductivity of bulk materials or layers and our previous works and experiments on the thermal resistance measurements. It is demonstrated that the voltage oscillation at the third harmonic is a real image of the thermal impedance of the device in the frequency domain. Both theoretical approach and test bench are discussed.**

I. INTRODUCTION

GaN HEMTs are extremely promising for RF power electronics applications but they are affected both by self-heating and traps phenomena [1] [2]. These low frequency memory effects limit the output power and reduce the device linearity. It remains a difficult task to characterize separately these phenomena with electrical measurements. Numerous methods have been proposed in the literature for the temperature determination (Raman, optical, electrical…) [3] [4] which often require complex equipment and well-prepared samples. On the contrary, electrical methods can be applied on any samples and have the main advantage of allowing "in situ" measurements. However, the drawback of performing a global approach supposes a "unique temperature description" which is certainly an approximation for large multi-finger devices such as high-power HBTs or HEMTs [5]. The challenge is often to separate thermal effects from trapping effects. A method using low frequency [S]-parameters has been proposed in [6] to characterize the thermal impedance of HBTs independently of their size and technology. Recently, an original method to estimate the thermal resistance through the measurements of I_{dmax} and R_{ON} has been proposed for GaN HEMTs [7]. In this method the thermometer is the "on state resistor" of the device, usually called R_{ON}. Our idea here consists to apply to the HEMTs the method initially proposed by D.G. Cahill [8] to characterize the thermal conductivity of bulk materials or layers. This method relies on the measurement of the third harmonic response of the device image of a "thermal excitation". The novelty consists in using this method with a transistor with a fixed gate voltage and a small signal applied to the channel. Driven by this small drain signal, the channel resistance observed on the IV curve (R_{ON}) is linear and traps are not excited. Measurements [1] reveal also a linear behaviour of R_{ON} with the baseplate temperature what is very interesting regarding the "3ω method".

II. THERMAL IMPEDANCE METHODOLOGY CHARACTERIZATION

In order to study the thermal dispersive effects of GaN HEMTs, we have developed a new "3ω method" which allows extracting the thermal resistance. In order to achieve accurate measurements of the thermal resistance (R_{TH}) and to extract the time constants of the thermal impedance Z_{TH} without trapping effects, the DC bias of the device is set to $V_{GS0} = V_{DS0} = 0V$. As mention before, in this linear area the resistance in the ohmic zone (R_{ON}) can be used both as heat source and thermal probe.

A. *Theoretical approach*

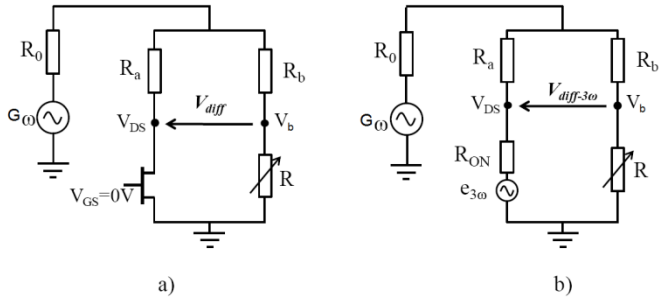


Fig. 1. The measurement principle (Wheatstone bridge): the transistor is replaced by its Thévenin equivalent model.

The main difficulty of this method is the measurement accuracy due to the very low voltage magnitude at the third harmonic V_{DS} (3ω) which is approximately 80 dB below the fundamental signal. To solve this problem, we use a Wheatstone bridge as described in "Fig. 1" and a lock in amplifier. At fundamental frequency, we just have the contribution of the generator G. At third harmonic, the "thermal excitation" generated by the transistor is noted $e_{3\omega}$. The transistor is placed in series with resistor R_a. The balancing branch is composed of resistor R_b in series with the variable resistor R. The idea is to adjust the variable resistor R in order to balance the bridge at the fundamental frequency. Then it is possible to measure with a great accuracy V_{DS} at the third harmonic frequency. We decompose the study into two parts: the fundamental frequency and the third harmonic frequency. At fundamental

frequency, when the bridge is balanced V_{DS} -V_b = 0V.

$$R_{ON}R_b = R_aR \quad (1)$$

We suppose that all the resistors of the bridge but R_{ON} are stable with the temperature (by component choice). So the transistor is the only resistor of the bridge generating harmonic signals. The voltage generated by the transistor at the 3rd harmonic frequency is $e_{3\omega}$ and V_{DS} as well as V_b are proportional to $e_{3\omega}$:

$$V_{DS} = \frac{R_a(R_0+R_b+R)+R_0(R_b+R)}{(R+R_b)(R_{ON}+R_a+R_0)+R_0(R_{ON}+R_a)} e_{3\omega} \quad (2)$$

$$V_b = \frac{RR_0}{(R+R_b)(R_{ON}+R_a+R_0)+R_0(R_{ON}+R_a)} e_{3\omega} \quad (3)$$

At third harmonic, the difference between V_{DS} and V_b is:

$$V_{diff\text{-}3\omega} = \beta\, e_{3\omega} \quad (4)$$

Where β, is a coefficient depending only of resistors included in the Wheatstone bridge. The resistor R_{ON} (T) as shown in [1] can be expressed by a linear equation depending on temperature:

$$R_{ON}(t) = R_{ON0}\left(1 + \alpha(T - T_{REF})\right) \quad (5)$$

Where R_{ON0} is the value of the resistor at the reference temperature T_{REF} and α is the temperature coefficient. ΔT = T – T_{REF} represents the temperature variation depending on the instantaneous dissipated power P_{DISS} (t):

$$T = Z_{TH}(t) * P_{DISS}(t) \quad (6)$$

$$P_{DISS}(t) = V_{DS1}\, I_{DS1} cos^2(\omega t) \quad (7)$$

The parameter P_{DISS} has a DC component and an AC component at second harmonic (2ω). V_{DS1} and I_{DS1} are respectively the magnitude of the voltage and the current at the fundamental frequency. By substituting (6) and (7) in (5), we obtain the expression of R_{ON} versus time:

$$R_{ON}(t) = R_{ON0}\left[1 + \alpha\left(Z_{TH}V_{DS1}\frac{1+\cos(2\omega t)}{2}\right)\right] \quad (8)$$

The I_{DS1}.cos (ω.t) current flows through the R_{ON} resistor, thus in the frequency domain it is possible to express the 3rd harmonic of V_{DS} (t) due to the thermal signal.

$$e_{3\omega} = \frac{\alpha\, R_{ON0}Z_{TH}V_{DS1}I_{DS1}^2}{4} \quad (9)$$

From equations (9) and (5), thermal impedance Z_{th} can be expressed by:

$$Z_{TH} = \frac{4V_{diff-3\omega}}{\alpha\beta R_{ON0}^2 I_{DS1}^3} \quad (10)$$

The important thing to note here is that the ratio $\frac{V_{diff-3\omega}}{I_{DS1}^3}$ must be constant at the first order; β depends on the elements of the Wheatstone bridge, can be expressed by:

$$\beta = \frac{R_a(R_0+R_b+R)+R_0R_b}{(R+R_b)(R_{ON0}+R_{cable}+R_0+R_a)+R_0(R_{ON0}+R_{cable}+R_a)} \quad (11)$$

R_0 is the output resistor of generator and R_{cable} is the resistor of cable which connects the transistor at resistor R_a. (R_{cable} =0.32 Ω).

B. Experimental technique

The setup developed to perform these measurements is presented in "Fig. 2". The magnitude of measured signals at the third harmonic is in the microvolts range. The calibration of the setup is made possible thanks to resistors which exhibit low temperature coefficient (lower than 50ppm / °C). The component choice is critical for accurate measurements. The voltage source must be chosen with very good Total Harmonic Distortion (THD). However, theoretically even if the generator is not perfect at ω and presents a small nonlinearity, the simple balancing of the Wheatstone bridge normally ensures the cancellation of the generator contribution to $V_{diff\text{-}3\omega}$. Our source can generate a sinusoidal signal with maximum amplitude of 10V (50 Ohms). This generator drives an amplifier with an output resistor R_0 of 2.1 Ohms thus improving the available current in the bridge. The main constraint of the Wheatstone bridge relies on a resistive bridge with elements thermally insensitive. For these reasons Birge and Nagel recommended the choice the wire-wound resistors [9]. Moreover, in order to increase the current in the "R_a branch", the choice of R_a and R_b must present a ratio of 10 (even 100). A capacitance may be added to compensate the inductive effects due to cables at high frequency (HF), in order to measure Z_{TH} on a wide frequency range.

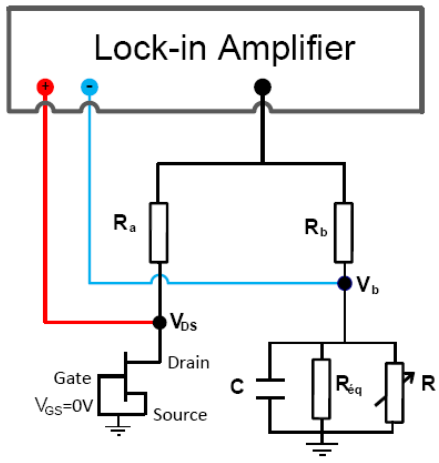


Fig. 2. Diagram of the set up to measure the thermal impedance Z_{TH} with the 3ω method.

III. GAN HEMT THERMAL RESULTS

The extraction of the thermal impedance has been performed on a coplanar (600µm thickness) AlInGaN / GaN HEMTs with 8 gate fingers of 75µm width and 0.25µm length, from III-V Lab. First, we present the measurement of the voltage $V_{diff\text{-}3\omega}$ for different magnitude values V_{DS1} at a single frequency (121 Hz).

Figure 3 shows the extraction of R_{ON} resistor for different V_{DS} values for a given frequency. One can notice that R_{ON} is almost constant ($\approx 3.7\Omega$) which confirms the approximation of linear behavior of the transistor in the ohmic zone. Moreover, R_{ON} is decomposed a sheet resistance (R_{sh}) of the semiconductor between the strips, the contact resistance (ρ_c) and the specific contact resistivity (R_c) of metal ohmic contacts to semiconductor [*10*].

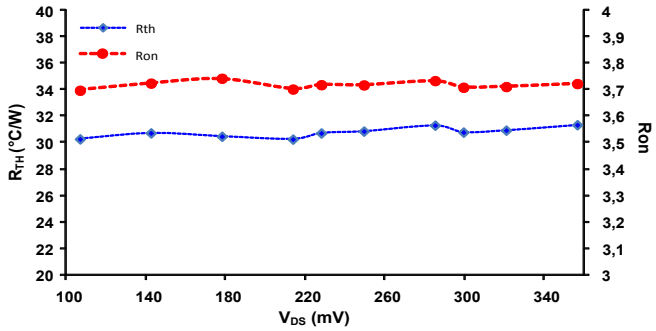


Fig. 3. Measured R_{ON} and R_{TH} versus V_{DS1}.

As we saw in formula (10), the ratio $V_{diff\text{-}3\omega} / I_{ds1}{}^3$ must be constant. The term α of $R_{ON}(T)$ has been determined as shown in [7]. β is computed with the knowledge of the bridge resistors (α = 0.0036, β=0.343).

Figure 4 shows that the ratio $V_{DIFF3\omega} / I_{ds1}{}^3$ is almost constant versus V_{DS}. The resulting R_{TH} value is plotted in Figure 3. Moreover, the magnitude of $V_{DS}(\omega)$ is limited in the range of ohmic transistor that avoids adding a term to the third harmonic from transistor, the study is very thin.

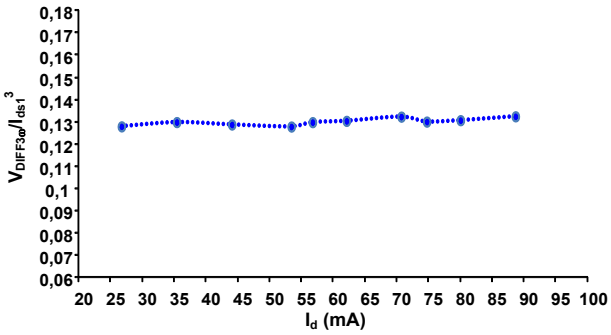


Fig. 4. Report $V_{DIFF3\omega} / I_{ds1}{}^3$ measured for different V_{DS1}.

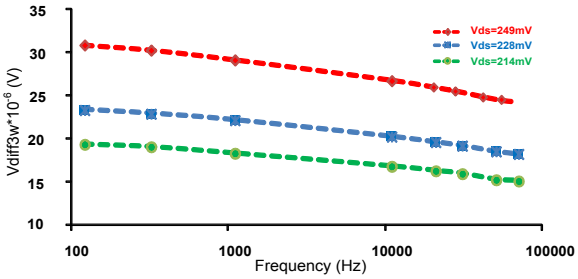


Fig. 5. The amplitude of temperature oscillations for different V_{DS1}.

Fig. 5 highlights the changes in oscillation temperature, which decreases with frequency for different magnitudes of V_{DS} (here 214mV, 228mV, 249mV). Fig. 6 shows that the real part of the thermal impedance obtained is the same for different magnitudes of V_{DS}. This result is very encouraging, regarding the reliability of the measurement and the extraction procedure.

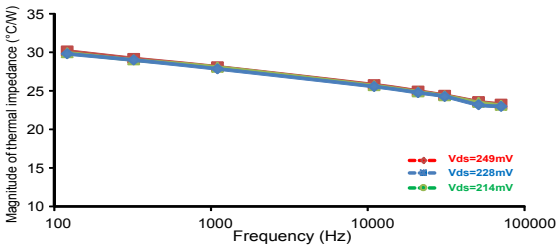


Fig. 6. The real part of the thermal impedance of transistor measured versus frequency for different V_{DS1}.

IV. CONCLUSION

We have presented a new accurate and repeatable method for measuring the thermal impedance of GaN HEMTs without trapping effects because the DC bias of the device is set to $V_{GS0} = V_{DS0} = 0V$. This method is based on the electrical measurements of the third harmonic frequency response as a real image of the thermal impedance. It is to our knowledge the first time that this method is used for transistor devices. We still need to set a Wheatstone bridge accurate enough to be balanced up to 10MHz in to order to extract all the time constants.

ACKNOWLEDGMENT

This work was performed in the framework of the common lab MITIC (XLIM / III-V lab) supported by the FP7 program ALINWON.

REFERENCES

[1] Mouginot, G.; Sommet, R.; Quéré, R.; Quarch, Z. ; Heckmann, S. ; Camiade, M.;, "Thermal and trapping phenomena assessment on AlGaN/GaN microwave power transistor," *Eumic (2010)*, pp. 110-113.

[2] M. Faquir, "Analysis of the physical mechanisms limiting performance and reliability of gan based hemts," Ph.D. dissertation, Université de Modène & Univerdité de Bordeaux 1, 2009.

[3] M Kuball, "Measurment of temperature in active high-power

AlGaN/GaN HFET's usin Raman spectroscopy," *IEEE Electron Device Letter*, vol. 23, no. 7-9, 2002.

[4] Peyretaillade, T.; Perez, M.; Mons, S.; Sommet, r. Auxemery, Ph.; Lalaurie, J.C.; Quere, R.;, "A pulsed-measurement based electrothermal model of HBT with thermal stablity prediction capabilities," *IEEE MTT-S International*, vol. 3, pp. 1515-1518, Jun 1997.

[5] Xiong, A.; Sommet, R.; De Souza, A.A.L.; Quere, R.;, "Dual approach for HBT thermal impedance," *Thermal Inveatigation of ICs ans Systems, 2008. Thermic 2008. 14 th International Workshop on*, pp. 190-194, 24-26 Sept 2008.

[6] El Rafei, A.; Sommet, R.; Quere, R.;, "Electrical measurement of the thermal impedance of bipolar transistors," *IEE Electron Device*, vol. 31, pp. 939-941, 2010.

[7] Sommet, R.;Mouginot, G.;Quéré, R.;Quarch, Z.;Camiade, M.;, "Thermal modeling and measurements of AlGan/GaN HEMTs including thermal boundary resistance," *Microelectronics Journal*, july 2011.

[8] D. G. Cahill, "Thermal conductivity measurement from 30 to 750k : the 3ω; method," *Review of Scientific Instrumens*, vol. 61, no. 2, pp. 802-808, 1990.

[9] N. O. Birge and S.R. Nagel, "Wide-frequency specific heat spectrometer," *Review of Scientific Instruments* , vol. 58, no. 8, pp. 1464-1470, 1987.

[10] Fontseré, A.; Pérez-Tomas, A.; Fernadez-Martinez, P.; Baron, N.; Chenot, S.; Cordier, Y.; Moreno, J.C.; Jennings; M.R.; Gammon; P.M.; Walker, D.;, "GaN Ohmic contact resistance vs temperature," Proc. Spanich Conf. Electron Devices (CDE), vol. 2011.

25-27 September 2012, Budapest, Hungary

A Simple and Approximate Analytical Model for the Estimation of the Thermal Resistances in 3D Stacks of Integrated Circuits.

François de Crécy,
CEA, LETI, Minatec Campus,17 rue des Martyrs, F38054 Grenoble Cedex 9, France.
(33)438 783 218 francois.decrecy@cea.fr

Abstract

The purpose of this paper is to describe a simple analytical model, useful for system architects, which describes the thermal resistances between a hot spot and the cold heat sink in steady state conditions with constant material properties, while taking into account the impact of the lateral spreading.

The total thermal resistance Rth_total is modeled as dependent on Rth_1D, the thermal resistance in the Z direction, perpendicular to the layers, and Rth_3D, the effects of the lateral spreading of the heat (the X and Y directions) from the hot spot.

In a 1st step, the different layers are homogenized to obtain an equivalent anisotropic thermal conductivity per layer. It must be noticed that the presence of copper TSV increases the thermal conductivities in the Z direction but decreases this conductivity in XY directions, due to the very negative effect of the SiO2 layer surrounding the Cu.

These conductivities are used in a 2nd step to compute Rth_1D, combining the individual thermal resistance of each layer. The 3rd step estimates Rth_3D by a simple analytical model. The few adjustable coefficients are fitted on several thousand FEM simulations.

I. INTRODUCTION

The 3D stacking of integrated circuits has multiple advantages, but may result in higher temperature of the most active surface and the thermal management of such a 3D stacking may be a challenge [1], [2], [3]. The purpose of this paper is to describe a very simple and approximate analytical model of thermal resistances between a hot spot and cold heat sinks in steady state conditions. As our objective is not to obtain a precise model, but an easy-to-use simple model for preliminary estimations of temperatures and sensitivity studies, valid within a very large range of factors, we assume that the variations of conductivities with temperature are second order phenomena and we use only constant material properties and the same temperature for the different heat sinks. These assumptions lead to linear computations and results and allow the use of the concept of thermal resistance.

II. LAYER-BY-LAYER MODELING

The 3D stack is modeled as a stacking of different plane layers of same external dimensions: a main assumption of our model is to consider each layer as homogeneous but anisotropic, with a conductivity kZ in the Z direction (normal to the layer) and another conductivity kXY in X or Y direction (within the plane of the layer). For example, for a three-die stack, we typically model 15 different layers, as depicted in figure 1:

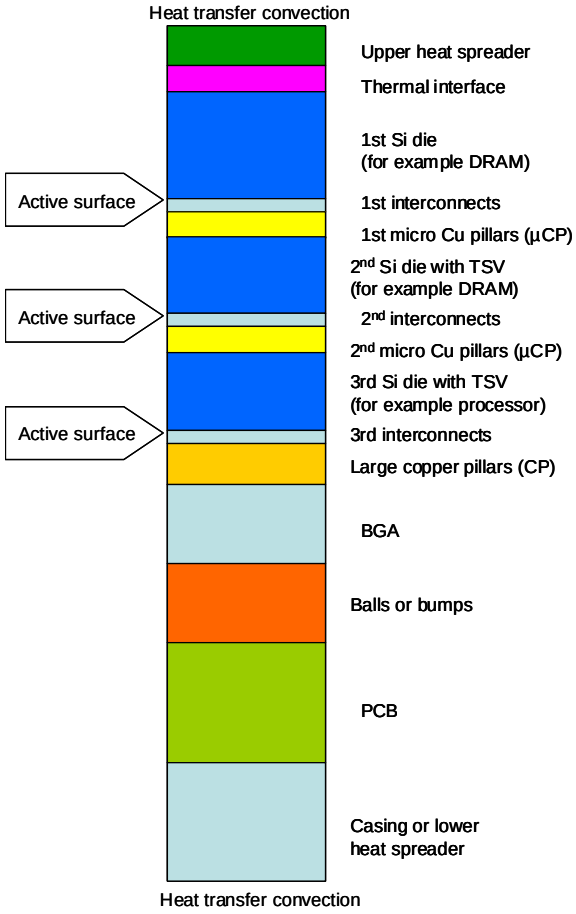


Figure 1: schematic of a stacking of different layers

There are also convective heat transfer boundary conditions on both upper and lower surfaces of the stack.

III. HOMOGENIZATION

This means that we have to homogenize some complex structures such as TSV, Interconnects, micro copper pillars (µCP), large copper pillars (CP) or bumps in the corresponding layer. For each layer, this is done in two steps: we firstly use a very simple, rough and approximate deterministic model, and then we add (or multiply by) a corrective polynomial based on several hundreds of Finite Element Method (FEM) simulations, by varying each input factor (geometric parameters and conductivities) over a large range and using the techniques of Design of Simulations (DOS) [4]. To do so, we model a characteristic structure, if possible periodic, within a parallelepiped and we impose two different temperatures on the opposite faces with adiabatic lateral faces. The equivalent conductivity in this direction is deduced from the total power transferred and the dimensions of the box.

Example: the micro Copper Pillars (µCP) layer

As an example, we present the homogenization of the µCP layer. Each µCP is done of two Cu studs and a solder and is surrounded by a polymer underfill.

To model the lateral (X or Y) equivalent thermal conductivity, the rough model is obtained by two parallel heat flows, the core circular region is replaced by a square of the same area, as explained in fig. 2:

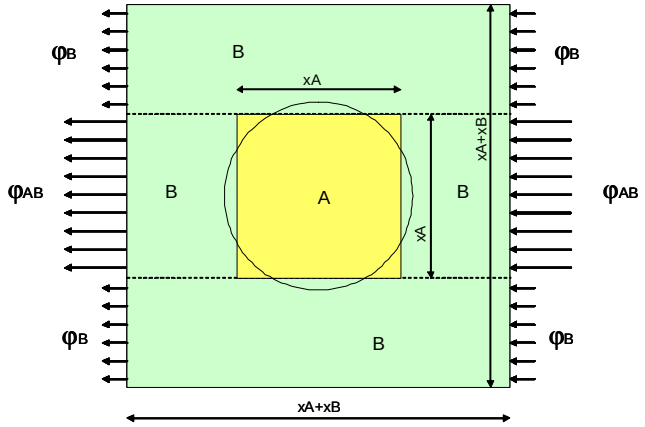


Figure2: rough simplistic model
for lateral equivalent conductivity in a µCP layer

The rough equivalent lateral thermal conductivity is:

$$k\_rough\_X = \frac{xB}{xA+xB}kB + \frac{xA \cdot kA \cdot kB}{xA \cdot kB + xB \cdot kA}$$

where kA and kB are the conductivity of material A (for example copper or solder) and material B (for example underfill).

To model the axial (Z) equivalent thermal conductivity, the rough model is obtained by two parallel heat flows, one in the µCP, the other in the underfill, as explained in fig. 3:

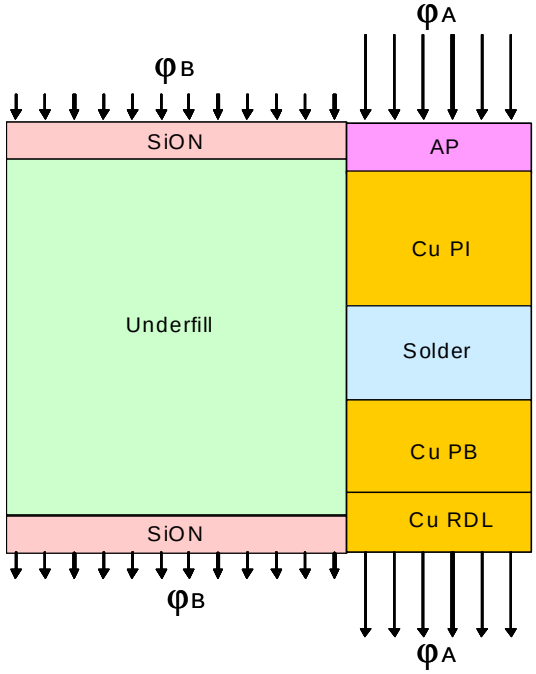


Figure3: rough simplistic model
for axial equivalent conductivity in a µCP layer

The rough equivalent axial thermal conductivity is:

$$k\_rough\_Z = \frac{kA \cdot SA + kB \cdot SB}{SA + SB}$$

where kA is the equivalent axial conductivity of the µCP and kB is the equivalent axial conductivity of the underfill:

$$kA = \frac{\Delta Z}{\frac{th\_AP}{k\_Al} + \frac{th\_CuPI + th\_CuPB + th\_CuRDL}{k\_Cu} + \frac{th\_solder}{k\_solder}}$$

and

$$kB = \frac{\Delta Z}{\frac{2 \cdot th\_SiON}{k\_SiON} + \frac{th\_underfill}{k\_underfill}}$$

where th stands for thickness and k for conductivities.

We then run a set of 702 Finite Element Modeling (FEM) simulations using ANSYS software spread over a large parameter range according to a DOS [4] for the geometric dimensions and conductivities.

These 702 FEM simulations are used to fit polynomial correction factors defined by:

*K_equiv_X = k_rough_X * correction_factor_X*
*K_equiv_Z = k_rough_Z * correction_factor_Z*

The polynomial correction factors depend on the geometric dimensions and conductivities, and an analysis of variance shows that the most influential parameters are the radius and pitch of the µCP. The presence of aligned

TSV in the adjacent Si layers is taken into account by a coefficient and has significant influence.

These equivalent conductivities in lateral (*K_equiv_X*) or axial (*K_equiv_Z*) directions will be used in global 1D and 3D simulations.

IV. THE OTHER LAYERS

We will not detail the modeling of each layer, but the methodology remains the same: a rough analytic model improved by a polynomial correction factor.

Silicon with TSV

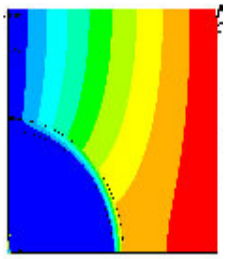


Fig 4: Temperature field around a TSV

The presence of copper TSV with its SiO2 barrier within the silicon slightly increases the equivalent conductivity of the Si layer in Z direction, but slightly decreases the equivalent conductivity in X or Y direction. This is due to the low conductivity of the SiO2 barrier. The simulations have been run with thickness of SiO2 ranging from 0.1µm to 0.5µm.

Interconnects

The Cu interconnects have been homogenized by modeling (FEM simulations with ANSYS) a great variety of typical structures. The results are statistically analyzed and surface responses are built for the equivalent thermal conductivities, depending only on the conductivity of the dielectric and the metal layer rate. Due to the huge variety of the individual tested geometries, the coefficient of determination are not really good but these interconnects have moderate influence on the overall thermal behavior of the stack.

Copper pillars or bumps

The rough model for large Copper Pillar (CP) or bumps is quite similar to the µCP model.

BGA

The BGA is modeled as a stack of successive sub-layers mainly composed of copper or dielectric and the equivalent conductivities depends of the thickness and conductivities of the sub-layers.

Balls

The radius, height and pitch of the balls are the predominant factors of the conductivities of the ball layer.

V. THERMAL RESISTANCE

As the geometries and the conductivities are independent of the temperature, and as the boundary conditions are linear (convection) toward the same heat sink temperature, the resulting temperatures are linear with the power and we can use the thermal resistance defined by:

RTh = max_difference_temperature / total_power

max_difference_temperature is the difference between the maximum temperature and the temperature of the heat sink.

When there is only one hot spot or a limited number of independent hot spots, we model RTh as the sum of two terms:

$$RTh = RTh\_1D + RTh\_3D \cdot \frac{W\_hotspot}{W\_total}$$

where *W_hotspot* is the power of the hot spot and *W_total* is the total power, the sum of the powers uniformly distributed over each active surfaces and the power of all hot spots. If there are several independent hot spots, we must consider the highest value of the quantity *RTh_3D*W_hotspot* and not the sum of these quantities.

Rth_3D represents the effect of higher temperature due to localized hot spots and the lateral spreading of the heat. Thereby *RTh_1D* determines the temperatures that would be established if the total power of each active surface was uniformly distributed over this active surface. The mean temperature of each active surface is the solution of the classical linear system:

$$\frac{T_k - T_{k-1}}{R_k} = W_k + \frac{T_{k+1} - T_k}{R_{k+1}} \quad k=1,\ldots,d$$

k is the index of the active surface and varies from 1 to d, the number of active surfaces. T_k is temperature of the k^{th} active surface. $T_0 = T_{d+1}$ is the temperature of the external heat sink. W_k is the total power of the k^{th} active surface (uniform + hot spots). R_k is the thermal resistance in Z direction between the active surfaces k and k-1:

$$R_k = \frac{1}{S} \cdot \sum_i \frac{th_i}{\lambda_i}$$

The summation is done over the layers between active surfaces k and k-1 or between the first and last active surface and the external heat sink. A term 1/(S*h) must be added to R_1 and R_{d+1} to simulate the convective heat transfer toward the heat sink, where h is the convective heat transfer coefficient.

RTh_1D is deduced from the temperatures $T_1, \ldots, T_d$ and the total power by

RTh_1D = (max(T_1, …, T_d)-T_0) / W_total

RTh_3D: localized hot spots

In case of localized hot spots, the maximum temperature is higher than in the case of the same total power with uniform heat flux. *RTh_3D*W_hotspot* represents this increase of temperature. The heat must spread in both Z and XY directions. The heat spreading in XY directions

depends on the product kXY_i* th_i (transverse_conductivity* thickness) of the layers: the highest this product is, the most efficient is this spreading and the lowest is *RTh_3D*. A surface of influence, depending on the kXY_i* th_i, is assigned to the hot spot and *RTh_3D* is modeled by

RTh_3D = A(1/S_influence - 1/S_die)*

This surface of influence is the intersection of the surface of the die and an area extending around the hot spot on a characteristic distance L_carac, naturally taking into account the effects of eccentricity and form factor of the hot spot (Fig 5)

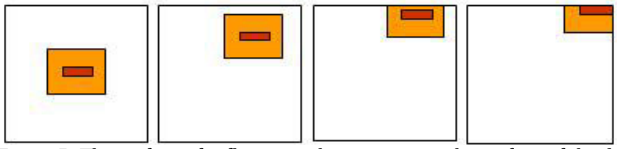


Figure 5: The surface of influence is the intersection the surface of the die (white rectangle) and an extension (orange) of the hot spot (red) on a distance Lx and Ly depending on a characteristic distance L_carac

The die is supposed to be rectangular, x varying from x0 to xmax, y varying from y0 to ymax. The hot spot is also rectangular, x varying from x1 to x2 and y varying from y1 to y2. The surface of influence is *S_influence = Lx*Ly* with:

$$Lx = L\_carac * \left(1 - Exp\left(\frac{-(x1 - x0)}{L\_carac}\right)\right) + (x2 - x1)$$
$$+ L\_carac * \left(1 - Exp\left(\frac{-(x\max - x2)}{L\_carac}\right)\right)$$
$$Ly = L\_carac * \left(1 - Exp\left(\frac{-(y1 - y0)}{L\_carac}\right)\right) + (y2 - y1)$$
$$+ L\_carac * \left(1 - Exp\left(\frac{-(y\max - y2)}{L\_carac}\right)\right)$$

The characteristic distance *L_carac* represents the lateral distance on which the hot spot is very influent. It is defined for each active surface and it depends on the products kXY_i*th_i of the different layers and on the thermal resistance in Z direction between this active surface and the current layer:

$$kXYT_k = \sum_{i=1}^{d} kXY_i \cdot th_i \cdot \left(C0 + \sum_{j=i}^{k} \frac{th_j}{kZ_j} + C2 \cdot \left(\sum_{j=i}^{k} \frac{th_j}{kZ_j} \right)^2 \right)^{-1}$$

(the index k represents the active surface and d is the total number of layers). C0 and C2 are adjustable coefficients. $kXYT_k$, which have the dimension of the square of a thermal conductivity, is characteristic of the stacking with respect to the active surface number k, independent of the power, position, size and form of the hot spots.

L_carac is modeled by:

$$L\_carac = D_{05} \cdot \sqrt{kXYT} + D_1 \cdot kXYT + \left(E_{05} \cdot \sqrt{kXYT} + E_1 \cdot kXYT\right) * EXP\left[\frac{-Dh}{F_1 \cdot kXYT}\right]$$

Dh is the equivalent diameter of the hot spot defined by Dh=2*dx*dy/(dx+dy)

with dx=x2-x1 and dy=y2-y1. Actually, the corrective term *E*exp(-Dh/F)* is influent only for small equivalent diameters Dh.

Optimization on a large set of simulations

The coefficients A, C0, C2, D_{05}, D_1, E_{05}, E_1 and F_1 have been fitted over a set of 7572 FEM simulations according to a Design of Simulations, simulating 1, 2 or 3 stacked die, between 9 to 15 layers. The dimensions of the die and of the hot spot, the position and form factor of the hot spot, the thicknesses and equivalent anisotropic conductivities of each layer, the convective heat transfer coefficients on top and bottom surface, the resulting *RTh_1D* and *Rth_3D* vary over large factor ranges given in table 2.

As the heat transfer coefficient at top and bottom of the die is typical of forced air convection, the convection term is usually predominant in *RTh_1D* and *RTh_1D* is usually predominant over *RTh_3D*.

We obtain the coefficients of the Table 1:

TABLE 1: ADJUSTABLE COEFFICIENTS FOR RTH_3D ANALYTICAL MODEL, FITTED OVER 7572 FEM SIMULATIONS

A	K*mm²/W	235.63
C0	mm²*K/W	7.921
C2	W/(K*mm²)	0.0438
D_{05}	mm²*K/mW	0.724
D_1	mm^3*K^2/mW^2	0.3246
E_{05}	mm²*K/mW	-0.724
E1	mm^3*K^2/mW^2	-0.2053
F1	mm^3*K^2/mW^2	0.07787
R^2		0.937
RSD	K/W	5.7

Where R^2 is the coefficient of determination for the *RTh_3D* model (the R^2 for *RTh* itself is much better: R^2=0.9978) and RSD is the Residual Standard Deviation.

The quality of fit of this methodology may be viewed by the figure 6:

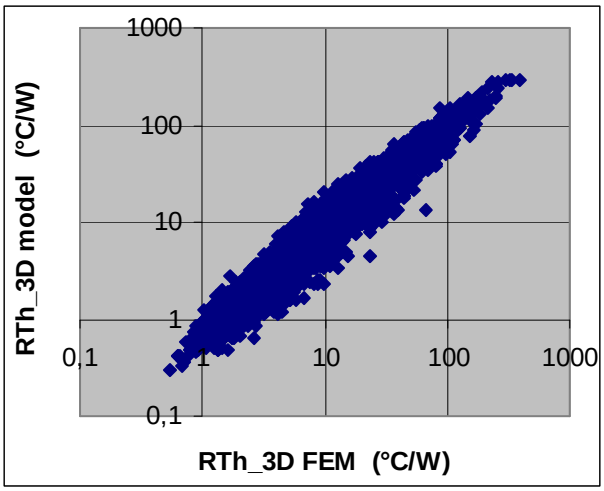


Figure 6a: dispersion of the model for RTh_3D

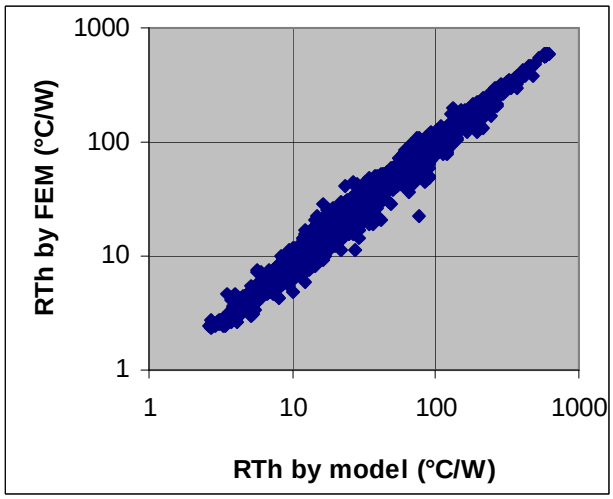


Figure 6b: dispersion of the model for RTh_total

TABLE 2: FACTOR RANGES (MINIMUM, 3 PERCENTILES AND MAXIMUM) OF THE 7572 FEM SIMULATIONS USED TO FIT THE COEFFICIENTS OF TABLE 1

		min	20%	50%	80%	max
Number of Silicon dies and active surface		1	1	2	3	3
Surface of the die	mm²	50	108	525	907	1024
Form Factor of the die (xmax/ymax)		1	1	1,49	1,91	2
Equivalent Diameter Dh of the hot spot	mm	0,20	0,34	1.36	2.45	3.5
Form Factor of the hot spot (dx/dy)		1	1	4.08	6.63	8
Eccentricity of the hot spot		0	0,31	0,51	0,73	0,99
Thickness of the Si dies (w or woTSV)	µm	40	81	327	654	750
Lateral conductivity of Silicon (w or woTSV)	W/(m*K)	105	110	133	152	160
Axial conductivity of Silicon (w or woTSV)	W/(m*K)	140	146	160	177	180
Total thickness of the µCP layers	µm	20	21	35	48	50
Lateral conductivity of the µCP layers	W/(m*K)	0,8	1,07	3,5	5,9	6,2
Axial conductivity of the µCP layers	W/(m*K)	1	1,73	10.5	19	20
Total thickness of the large Cu pillar layer	µm	35	41	70	99	105
Lateral conductivity of the large Cu pillar layer	W/(m*K)	1	1,66	4,85	8.01	8,7
Axial conductivity of the large Cu pillar layer	W/(m*K)	2	3,06	8	12,9	14
Thickness of the BGA layer	µm	150	177	300	424	450
Lateral conductivity of the BGA layer	W/(m*K)	10	23	80	138	150
Axial conductivity of the BGA layer	W/(m*K)	1	1,34	3,00	4,64	5
Thickness of the balls or bumps layer	µm	80	108	240	373	400
Lateral conductivity of the balls or bump layer	W/(m*K)	0,2	0,25	0,50	0,75	0,8
Axial conductivity of the balls or bump layer	W/(m*K)	0,5	0,85	2.73	4,58	5
Thickness of the PCB	µm	500	587	1000	1411	1500
Lateral conductivity of the PCB	W/(m*K)	10	22	80	137	150
Axial conductivity of PCB	W/(m*K)	1	1,33	3,00	4,65	5
Thickness of casing or heat spreader	µm	500	585	1000	1409	1500
Axial & Lateral conductivity of casing or heat spreader	W/(m*K)	10	43	130	229	250
Heat transfer coefficient on upper side	W/(m²*K)	5	10,5	31,6	54,5	60
Equivalent heat transfer coefficient on lower side	W/(m²*K)	13	100	250	557	1200
RTh_1D	K/W	1,3	5.66	10.7	43.6	1171
RTh_3D by FEM simulations	K/W	0,54	2.54	6.08	17.4	367

VI. How to use this method?

Once you have determined the axial (kZ) and lateral (kXY) thermal conductivity of each layer, the next step is to determine the total power per active surface. This total power per active surface is then used to compute the mean temperature of each active surface by solving the classical linear system. The maximum temperature and the total power will give *RTh_1D*.

For the effect of localized hot spots, compute $kXYT_k$ and *L_carac* for each active surface. If there are several hot spots, the method is useful if the hot spots are independent, that is to say that if the lateral distance between hot spots is large compared to *L_carac*. Compute the surface of influence and *RTh_3D* for each hot spot. Look for the maximum value of the products

*RTh_3D*W_hotspot*

and use this increase of temperature to determine the actual *RTh*.

VII. Conclusion

We propose a model and methodology to quickly obtain approximations of the maximum temperature in a 3D stack of several active dies. It depends on only 8 coefficients fitted over 7572 simulations. This model may be used to easily do approximate sensitivity studies or for a first optimization by system architects.

Acknowledgement

Part of this work has been funded by the European project 3DIM3

References

[1] S.Im and K. Banerjee, *Full chip thermal analysis of planar (2D) and vertically integrated (3D) high performance ICs*, Tech. Dig. IEDM, 2000, pp 727-730

[2] K. Banerjee, M. Pedram and A. Ajami, *Analysis and optimization of thermal issues in high-performance VLSI*, ACM/SIGDA International Symposium on Physical Design (IPSD), pp 230-237, April 2001

[3] K. Matsumoto, S. Ibaraki, K. Sakuma, K. Sueoka, H. Kikuchi, Y. Orii and F. Yamada, *Thermal Resistance Evaluation of a Three-dimensional (3D) Chip Stack*, Electronics Packaging Technology Conference (EPTC) pp614-619, 2010

[4] T.J. Santner, B.J. Williams, W.I. Notz, 2003, *The Design and Analysis of Computer Experiments*, 2003, Springer

25-27 September 2012, Budapest, Hungary

New thermal method for the characterization of solid materials at different temperature levels

Ali Assy[1, 2, 4] Séverine Gomès[1, 2, 4, a)] Patrice Chantrenne[1, 3, 4] Nicolas Pavy[5]
Jayalakshmi Parasuraman[5,] Xavier Kleber[1, 3, 4] and Philippe Basset[5]
1) Université de Lyon
2) CETHIL, UMR-5008, CNRS, Université Lyon 1
3) MATEIS, UMR-5510, CNRS, Université Lyon 1
4) INSA de Lyon, Lyon F-69621, France
5) Université Paris-Est, ESYCOM, ESIEE Paris
Bât. Sadi Carnot 9 rue de la physique 69621 Villeurbanne Cedex FRANCE
0033 472 436 428, 0033 472 438 811, *severine.gomes@insa-lyon.fr*

Abstract **- A new method for evaluating the thermal properties of solid materials at different temperatures is proposed. An array of parallel and long resistive lines is deposited on the surface of the sample that is shaped in a suspended film several hundred µm in thickness. One of the resistive lines of this array is heated with an alternative current. The surface temperature changes in DC and AC regimes are detected by measuring the electrical resistance change of each of the other lines of the array. The length of wires was chosen so that they may be assumed isothermal and heat diffusion operates perpendicularly to them. By measuring the dependence of the surface alternative temperature rise on the modulation frequency *f* and on the separation between the heating wire and the probe wires, the thermal diffusivity of sample can be determined. The adjustment of the amplitude of the alternative current in the source allows controlling the sample temperature at which the thermal diffusivity is evaluated. For the validation of the proposed method, pure silicon samples were first studied. The experimental bench was set up and resistive source and probes were experimentally characterized. Results obtained at different temperatures until 500 K for pure silicon are in good accordance with the ones obtained elsewhere in the scientific literature by modeling as well as other experimental techniques.**

I. INTRODUCTION

Bulk materials and films thermal characterizations are very important for different application domains (microelectronic, nuclear, materials for aeronautics…). To achieve such determinations, thermal waves are very convenient tools which can be used in order to explore the matter and to measure thermal parameters. Usually the thermal source is photothermally created: the sample is then illuminated with amplitude modulated light which may be absorbed at the surface or in the bulk. A spot, a straight line or a domain of the surface can be illuminated. They correspond respectively to cylindrical and plane and 1D propagation regimes. The surface temperature variation can be measured either with a mirage setup [1, 2] or a photoreflectance microscope [3-5]. Thermal parameters at ambient temperature are then generally extracted from the fitting of curves calculated with a model of heat spreading in the sample adapted to the propagation regime experimentally used with the temperature variations measured for different modulation frequencies or distances to the thermal source. Measurements at different levels of temperature by these methods need a particularly heavy conditioning of the setup [6]. Also based on the use of thermal waves, a new method is here proposed for evaluating the thermal diffusivity α of solid materials at different temperatures. As the 3ω method [7], this method combines the principle of thermal metrology techniques with contact and in modulated regime.

II. PRINCIPLE OF THE METHOD

A. Principle

For the experiments, the material to be studied is shaped in a flat suspended film. As shown in Fig. 1, an array of parallel and long resistive lines is deposited on its surface. One of the resistive lines of this array is heated with an alternative current I of magnitude I_{ac} and frequency $f=\omega/(2\pi)$ with ω the angular pulsation. The power P dissipated in this line may be written as:

$$P = R.I^2 = R.I_{ac}^2/2 + R.I_{ac}^2/2 .\cos(2\omega t) \quad (1)$$

where R is the electrical resistance of the heated wire.

Then, the conductive diffusion of this heating source energy in the material produces a continuous and a 2ω alternative sample surface temperature rises called $\Delta T_{DC}(x)$ and $\Delta T_{2\omega}(x)$ respectively, where $|x|$ is the distance between the heating source and the other resistive wires used as thermal resistive probes (see in Fig. 1). The length of all the wires was indeed chosen so that they may be assumed isothermal along their length (in the Y direction in Fig. 1) and heat diffusion operates perpendicularly to them (in the X direction).

The surface temperature $\Delta T_{2\omega}(x)$ change is detected by measuring the electrical resistance change of each of the probe wires $\Delta R_{2\omega}(x)$:

$$\Delta R_{2\omega}(x) = R_0(x)\beta\Delta T_{2\omega}(x) \quad (2)$$

where $R_0(x)$ is the electrical resistance of the wire at ambient temperature T_a and β is the temperature coefficient of the electrical resistivity of the wire material.

 ISBN:978-2-35500-022-5

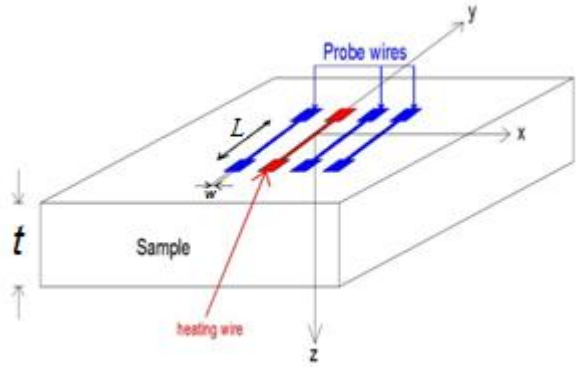


Fig.1. Schematic of the implemented sample.

For this measurement of electrical resistance change, a direct current i_0, of magnitude sufficiently small to avoid a significant heating, is injected in the probe wires. The second harmonic voltage resulting at the wire ends is then analyzed with a lock-in amplifier technique. It is given by:

$$\Delta V_{2\omega}(x) = \Delta R_{2\omega}(x) i_0 = R_0(x)\beta\Delta T_{2\omega}(x) i_0 \quad (3)$$

By measuring the dependence of the surface alternative temperature rise on the heating wire modulation frequency $F=2f$ and on the separation $|x|$ between the heating wire and the probe wires, the thermal diffusivity of the sample can be determined.

B. Modeling

The determination of the thermal diffusivity α will be all the easier as the thermal diffusion length μ in the sample material, given by:

$$\mu = \sqrt{\alpha/2\pi f} \quad (4)$$

where

$$\alpha = k/(d.C) \quad (5)$$

with k the material thermal conductivity and d and C its density and heat capacity respectively , is large compared with the thickness t of the suspended film. In this first approach of the measurement, these conditions will be assumed verified while working at small frequencies (see later). In this case, the thermal gradient in the sample thickness may be assumed null (in the Z direction) and the thermal behavior of the sample may be described as one-dimensional in the X direction. From the resolution of the heat transfer equation and in the assumption of a punctual section of the heating wire (with respect to the fact that $\mu >> 2\pi w$, where w is the width of the wire), the second harmonic temperature decay measured along a distance x from the heater may then be written as:

$$\Delta T_{2\omega}(x) = \Delta T_{2\omega}(x=0)\exp(-x/\mu)\exp(-ix/\mu) \quad (6)$$

where $\Delta T_{2\omega}(0)$ is the second harmonic temperature at the heater (x=0).

The thermal diffusivity α can be deduced from the attenuation of the magnitude $\exp(-x/\mu)$ and the variation of the phase $(-x/\mu)$, both as a function of x. Let α_{am} *and* α_{Φ} the thermal diffusivity extracted from the magnitude attenuation and the phase respectively. From the experimental point of view, these data are estimated from the signal measured by the probe wires and from (3), (4), (5) and (6). This involves a good knowledge of the electrical properties of the material composing the wires.

The adjustment of the amplitude of the alternative current I_{ac} in the source allows controlling the sample temperature at which the thermal diffusivity is evaluated.

Let us remark that in this first approach the thermal contact resistance between the wires and the sample surface is considered null and the section of the probe wires is also assumed punctual.

II. Sample of Test

A. The Reference Material

Because of its well-known thermal properties versus the temperature, we used a sample of pure silicon as reference material. Table I gives the value of α of silicon calculated and experimentally determined in the literature for different levels of temperature [8-10]. The density was determined by *Smakula et al.* [11], d=2329 kg/m^3, for high pure silicon at a temperature of 298 K. We considered it constant as a function of temperature in our work, as the thermal expansion coefficient does not exceed 1 % at temperature below 600 K [12].

B. The Wire Array

As described in Fig. 1, the heating wire, is located in the middle of the sample surface. The probe wire are located on both sides of the heater. The probe wires are distant 75 µm from each other; the closest one being away from the heater of 225 µm and the furthest one of 675 µm. The main process for fabricating these wires array is photolithography and lift-off. The schematic diagram of the fabrication process flow is shown in Fig. 2. After a substrate cleaning, a 2 µm–thick nLOF negative photoresist is spun on the silicon wafer. The resist is baked on a hotplate at 110°C for 1 min. A UV exposure is then carried out (Fig. 2-a). A post bake is performed at 110°C on a hotplate for 1 min, immediately followed by the resist development in a AZ-351B solution for 60 s, leaving an array of line openings into the resist layer on top of the sample. After the photolithography step, a bilayer chromium / gold, respectively of 50 nm-thick (Cr) and 450 nm thick (Au), is deposited on top of the sample by means of sputtering (Fig. 2-b).

TABLE I
THERMAL PROPERTIES OF PURE SILICON VALUES

Temperature (K)	C (J/Kg.K)[8]	k (W/m.K) [9]	$\alpha_{calculated}$ [a] ($.10^{-5}$ m²/s)	α_{exp} [10] ($.10^{-5}$ m²/s)
300	703	148	9.03	8.8
400	760	105	5.93	5.33
500	800	80	4.29	3.8

[a] calculated with (5) and the values of k and C from the table I.

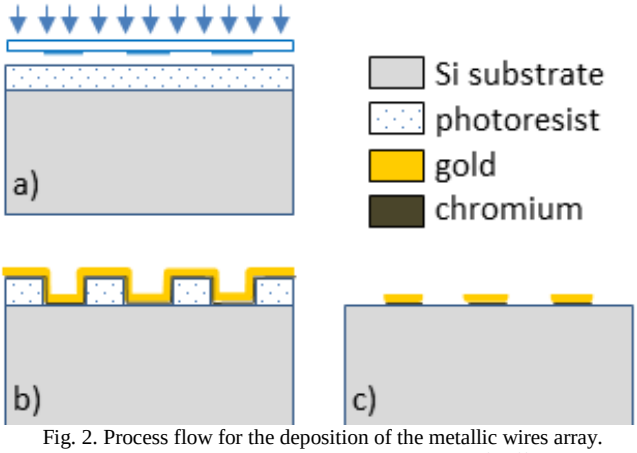


Fig. 2. Process flow for the deposition of the metallic wires array. (a) Photolithography, (b) Metal deposition, (c) Lift-off process.

The chromium is used as an adhesion layer to allow the gold to stick on silicon. Finally the lift-off step is performed in an acetone bath under ultrasonic excitation (Fig. 2-c).

A specific sample was prepared for the test. Its thickness, width and length were 200 µm, 7 mm and 10 mm respectively.

C. Characterization of the Wires

Table II gives the value of the section S, the electrical resistivity ρ and its temperature coefficient β experimentally determined for the deposited wires. S was measured by Atomic Force Microscopy in tapping mode. $\rho=R.S/L$, with L the wire length, was determined from the measurement of the wire electrical resistance R by the 4-wire method. The wires are of several millimeters of length. Their length was measured with an optical microscope. β was also estimated from wire electrical resistance measurements while the chip (sample + wires) was placed in a specified oven, and was heated at different temperatures T: $R=R_0(1+\beta(T-T_a))$.

III. Numerical Model and Simulation Results

Numerical simulations of the thermal behavior of the Si sample were done with COMSOL Multiphysics to estimate the frequency domain of validity for the simple 1D approach of the measurement proposed in the section II of this paper. Based that the temperature is uniform along the length of the wires, a 2D model was used for that.

A. Description of the System

As the width of the silicon sample is millimetric in size and the heating wire is located in its middle, the sample may be assumed symmetric. A volumetric term ($P = RI_{ac}^2/V = \rho I_{ac}^2/S^2$ where V is the source wire volume), calculated with the experimental data given by Table II, was inserted in the section of the source wire. Thermal losses by natural convection (h= 5 w/m^{2}K) were considered on the entire sample surface.

The thermal contact resistance between the wire and the sample was supposed null. The thermal properties of silicon

TABLE II
PROPERTIES OF THE WIRES

Parameter	Value
Section S (m²)	$1{,}66\pm0{,}15.10^{-12}$
Electrical resistivity ρ (Ω.m)	$1{,}12\pm0{,}1.10^{-7}$
Coefficient of temperature β (K^{-1})	$8{,}8\pm0{,}05.10^{-4}$

(k, d and C) and their dependence on temperature, deducted from Table I, were used for the calculations.

B. Simulation Results

Simulations were made for different frequencies and alternative currents in the heating wire. For these simulations, the number of the elements in the mesh was 1300 elements and the time between two steps was at least twenty times smaller than the heat power frequency. The temperature components T_{DC} and $T_{2\omega}$ were extracted from the simulation of the total temperature variation versus time at x=0 (location of the heating wire) and at the different distances x of the probe wires from the heater.

For $f \leq$ 80 Hz, the validity of the 1D model allows determining the thermal diffusivity of Si with a maximum uncertainty of 3 % by fitting the simulated $\Delta T_{2\omega}(x)$ profile with the 1D model equation (6). As an example, we discuss here the results obtained for a frequency of 80 Hz and a current I_{ac} of 60 mA.

Fig. 3 gives the simulated temperature variation versus time at x = 0 for a frequency of 80 Hz and a current I_{ac} of 60 mA. For this value of I_{ac}, the input of the thermal diffusivity of the sample was α= 4.6 10^{-5} m^2/s and T_{DC} =486 K and $T_{2\omega}$ = 1.1 K. The simulation showed that the sample is isothermal along x, as the difference in T_{DC} between the heating wire and the furthest probe wire does not exceed 1K. Besides these results show that the values of $T_{2\omega}$ can be larger than three orders of magnitude smaller than those of T_{DC}: the very low cut-off frequency of the system justifies this difference.

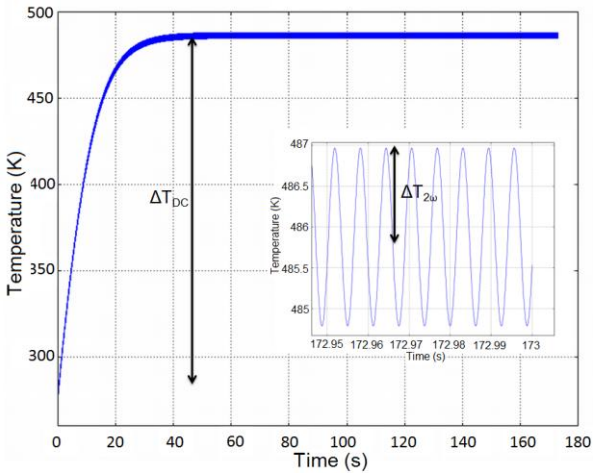


Fig. 3. Variation of the heater temperature as a function of time. f= 80 Hz.

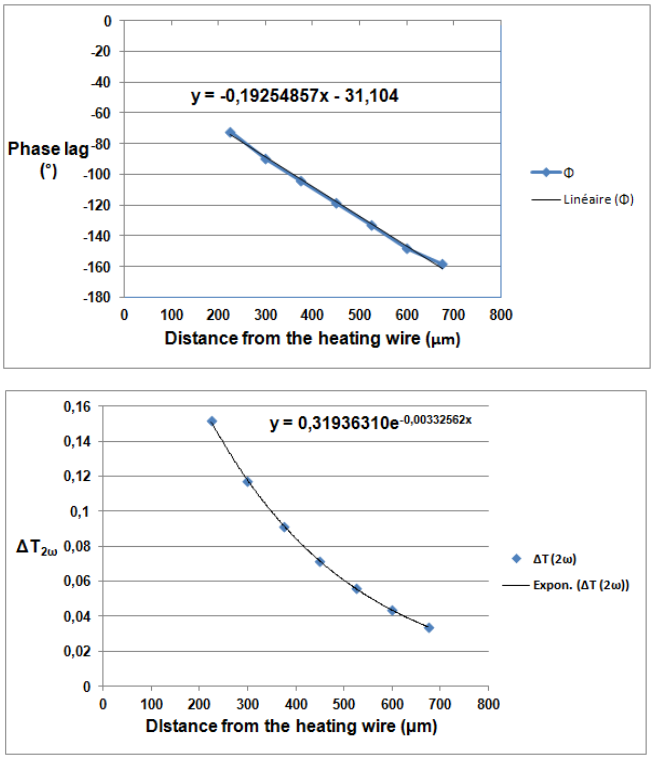


Fig. 4. Simulated variation of (a) the phase lag (b) $\Delta T_{2\omega}$ as a function of the distance from the heating wire for I_{ac}=60 mA and f=80 Hz. α= 4.6*10^{-5} m^2/s.

Fig. 4 (a) and (b) show the magnitude $\Delta T_{2\omega}$ and the phase lag Φ of the calculated second harmonic temperature respectively as a function of x.

Using (6), the thermal diffusivity α was extracted. The value obtained is the same than the input value (4.6 10^{-5} m^2/s) with an uncertainty of 3%. This results show that 1D model can be tested in a first approach to characterize the sample for $f \leq 80$ Hz: the sample thickness t is smaller than μ when f decreases.

IV. Experimental Results

Experiments were performed for frequencies smaller than 80 Hz. Two different magnitudes of 40 and 60 mA were generated in the heating wire for this domain of frequencies. An example of the measurement at a current of 60 mA is given by the Fig. 5 (a) and (b) for different frequencies. The exponential tendency of $\Delta T_{2\omega}$ via and the linear tendency of Φ are found for each current.

Using the equations 3 and 4, α_{am} and α_{Φ} were extracted. The table III gives the values of the identified diffusivity average and its uncertainty for both used currents. Comparing between the experimental values in the table III and the literature values in the table I, we found that the values of the thermal diffusivity experimentally determined with the 1D model lie in the range of the values previously determined.

This is relied to the fact that till now there is no thermal contact resistance taken into account in these simulations.

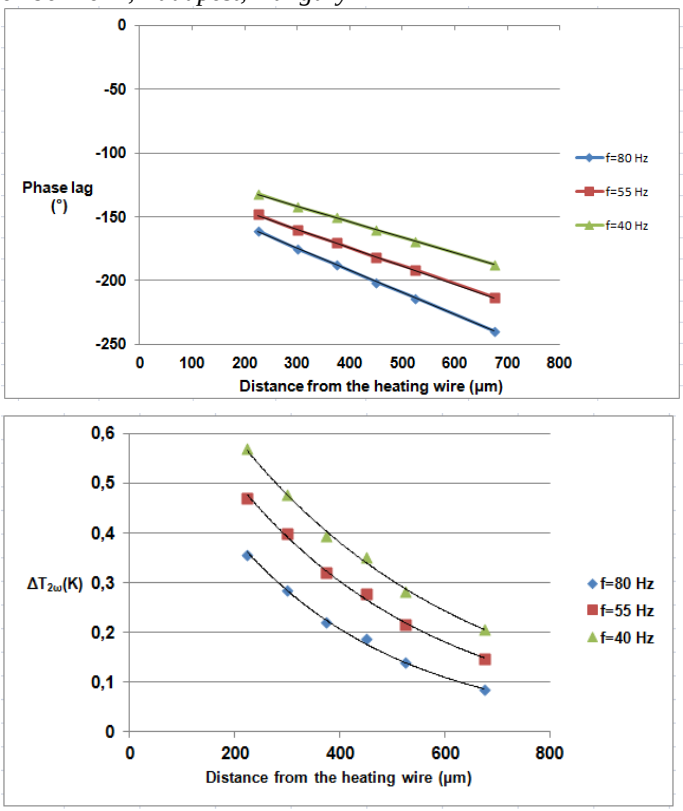


Fig. 5. Experimental variations of (a) the phase lag (b) $\Delta T_{2\omega}$ as a function of the distance from the heating wire for I_{ac}=60 mA and for different frequencies f.

TABLE III
SI THERMAL DIFFUSIVITY EXPERIMENTALLY DETERMINED

I_{ac} (mA)	α_{am} (.10^{-5} m^2/s)	α_{Φ} (.10^{-5} m^2/s)	T_{DC} (K)[a]
40	6.5±0.5	6.7±0.5	395
60	5±0.1	5.5±0.1	485
[a]Estimation from COMSOL Multiphysics simulations for the corresponding current I_{ac}.			

V. Conclusion and Perspectives

This work reports the first step of validation of a new technique based on thermal metrology techniques with contact and in modulated regime for estimating the thermal properties of solid materials at different temperatures. For this first step of validation, the technique was applied to the estimation of the thermal diffusivity of a pure silicon sample. In a first approach of the measurement, the thermal diffusivity of the material was extracted from the fitting of measurements with curves calculated with a simple 1D model of heat spreading in the sample. The silicon thermal diffusivity values obtained at different temperatures are in agreement with the ones of the scientific literature. The simple 1D model used in this approach is however shown to allow an estimation of the thermal diffusivity of the material with an uncertainty around 3% and this for a restricted frequency domain. Besides this uncertainty was estimated from the fitting with it of the results of numerical simulations performed for a system description already

simplified compared with the real system (no thermal resistance between the heater and the probes in contact with the sample, heating and probe wires of punctual section). A more realistic model has to be developed to reduce the uncertainty of the measurement, to extend the frequency domain of use of the technique and to envisage the application of the technique to the characterization of complex materials such as multilayered materials.

At this stage of the development of the method proposed, the perspectives of this work will naturally focus the improvement of the model of heat spreading in the heater, probes and sample.

ACKNOWLEDGMENT

The authors acknowledge financial support from the ANR-07-PNANO-047. We are also pleased to acknowledge the technical contribution of our colleague E. Pané Farré and the SOITEC Company (Grenoble-France) for providing us the silicon samples.

REFERENCES

[1] A.C. Boccara, D. Fournier, J. Badoz, "Thermo-optical spectroscopy: Detection by the mirage effect'', *Appl. Phys. Lett*, vol. 36, pp. 130-132, 1979.

[2] F. Charbonnier, D. Fournier, "Compact design for photothermal deflection (mirage): Spectroscopy and imaging'', *Rev. Sci. Instrum*, vol. 57, pp. 1126-1128, 1986.

[3] A. Rosencwaig, J. Opsal, W.L. Smith, D.L. Willenborg, "Detection of thermal waves through optical reflectance '', *Appl. Phys. Lett*, vol. 46, pp. 1013-1015, 1985.

[4] L.J. Inglehart, A. Broniatowski, D. Fournier, A.C. Boccara, F. Lepoutre, "Photothermal imaging of copper-decorated grain boundary in silicon'', *Appl. Phys. Lett*, vol. 56, pp. 1749-1751, 1990.

[5] L. Pottier, "Micrometer scale visualization of thermal waves by photoreflectance microscopy'', *Appl. Phys. Lett*, vol. 64, pp. 1618-1619, 1994.

[6] D. Rochais, H. Le Houëdec, F. Enguehard, J. Jumel and F. Lepoutre, "Microscale thermal characterization at temperatures up to 1000°C by photoreflectance microscopy. Application to the characterization of carbon fibres", *J. of Phys. D: Appl. Phys.*, vol. 38, pp. 1498-1503, 2005.

[7] D.G. Cahill, "Thermal conductivity measurement from 30 to 750 K: the 3ω method'', *Rev. Sci. lnstrum*, vol. 61, pp. 802-808, 1990.

[8] G. Chen, "Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices'', *Phys. Rev. B*, vol. 57, pp. 14958–14973, 1998.

[9] C.J. Glassbrenner and G.A. Slack, "Thermal Conductivity of Silicon and Germanium from 3 K to the Melting Point'', *Phys. Rev.*, vol. 134, pp. 1058–1069, 1964.

[10] H. R. Shanks, P. D. Maycock, P. H. Sidles, and G. C. Danielson, "Thermal Conductivity of Silicon from 300 to 1400 K'', *Phys. Rev*, vol. 130, pp. 1743–1748, 1963.

[11] A. Smakula and V. Sils, "Precision density determination of large single crystals by hydrostatic weighing'', *Phys. Rev.*, vol. 99, pp. 1744–1746, 1955**.**

[12] H. Watanabe, N. Yamada, and M. Okaji, "Linear thermal expansion coefficient of silicon from 293 to 1000 K'', *Int. J. Thermophys.*, vol. 25, pp. 221-236, January 2004.

25-27 September 2012, Budapest, Hungary

Modelling of Heat Transfer in Microdroplets as Microreactors

Ferenc Ender[1], Gusztáv Hantos[2]
Budapest University of Technology and Economics
Department of Electron Devices
Budapest, Hungary
Email: [1]ender | [2]hantos@eet.bme.hu

***Abstract*—A novel approach of modeling heat transfer in Taylor flow is presented. The model addresses especially the applications of Lab-On-a-Chip systems and focuses on the thermal behavior of individual droplets. Heat transfer is investigated at different boundary conditions as heat generation inside the droplet and heat generation at the droplet surfaces. A novel approximation of convective mass and heat transfer is presented as well.**

I. Introduction

Segmented flows are attracting high attention due favorable heat transfer behavior in the microscale. It is widely recognised that such flows show a higher heat transfer coefficient practically exploited for cooling intergrated circuits. Heat transfer issues were investigated by simulation by many authors [1]–[5]. Numerical models were presented as well by Muzychka et al. (2011), Walsh et al. (2010), Leung et al. (2010). [2], [3], [6].

Slug or Taylor flow as a special type of segmented flows has further advantages in the field of bioanalitics. This type of flow appears as consequent slugs of two phases which could be gas and liquid or two immiscible liquids and was first investigated by Taylor (1961) [7]. The aforementioned adventages are based on the following key features according to Theberge et al. (2010), [8]). The separated slugs form individual microreactors, which

- provide compartments in which species or reactions can be isolated
- are monodisperse and therefore potentially suitable for carrying out quantitative studies
- provide the possibility to work with extremely small volumes and single cells or molecules
- offer the ability to perform very large numbers of experiments

This paper addresses the bioanalitical use, where numerous investigation of possible applications and fluid mechanics including mass transfer were reported by many researchers, but the design aspects of thermal relations were not analyzed in depth. A review of the literature was presented by Theberge et al. [8]. As a general approach the thermal behavior of Taylor flow is usually investigated as a whole system comprising a few dozen of droplets in terms of cooling electronic devices. Though the numerous works are presented in this field numerical models are limited to average Nusselt number appriximations which ignore the thermal relations of the individual droplets, and interprets all droplets together. In contrast, the thermal relations should be analyzed in individual droplets in case of monolithical microreactors as each droplet performs individual reaction. These reactions can occur on the droplet surface and/or inside the liquid slug and can vary in time as well, while the droplets, slugs and the channel wall itself are thermally coupled to each other. Therefore the

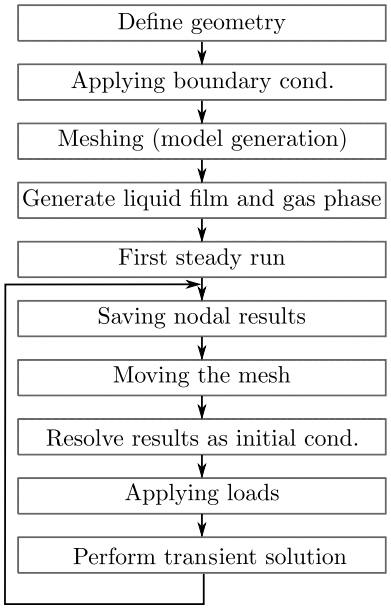


Fig. 1. Flowchart of the ANSYS implementation

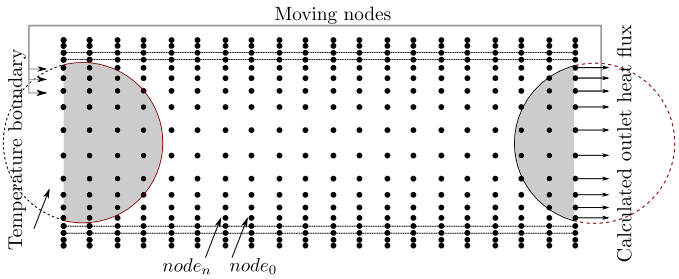


Fig. 2. Moving mesh technique and applied boundary conditions

ISBN:978-2-35500-022-5

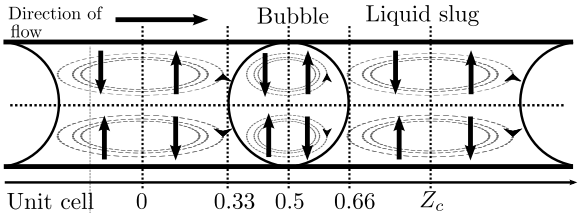


Fig. 3. Schematic representation of inner circulations. Fluid volume is divided into sections marked by arrows for modelling the circulation effect by mass transfer

thermal modelling requires to count the droplets individually and the thermal couplings of the whole system as well.

Our final goal is to develop a thermal compact model which can be integrated in the design flow of monolith microreactor based Lab-on-a-Chip devices. Thermal compact models are widely used in design practice of electrical systems, where these reduced order models yield fast results in early design phases contrary to finite element simulation which requires much more time. Typical simulation run-times of the Taylor-flow problem vary in wide scale from 38 to 1460 hours [4], [5]. A thermal compact model for integrated circuits with cooling microchannels was presented by Zhang et al. (2002), [9]. The model using thermal resistances and controlled voltage sources is able to predict the temperature distribution of the channel wall, but not for individual droplets. Numerical models based on *Nu* prediction were presented by Muzychka et al. (2011), Walsh et al. (2010), Leung et al. (2010). [2], [3], [6]. A compact model for modelling fluid dynamics in Taylor flow was presented by Jousse et al. (2005) [10].
This paper introduces a simplified model for modelling heat transfer in monolith microreactors while keeping the running times between reasonable limits. The cases of surface and volumetric heat generation is detailed in depth while the classical constant heat flux boundary condition acts as a reference.

II. Model generation

In order to generate the model the following requirements may be set. The model:

- is a good approximation of the heat transfer in Taylor flow at low Reynolds numbers typical for Lab-on-a-Chip systems,
- handles transient effects applied to the individual droplets,
- is linear, nonlinear parameters may also involved,
- involves any effects of internal forces (e.g. internal circulation) by simplified models and correlations,
- works with any types of boundary conditions (constant temperature and flux, heat generation inside the droplets etc).
- yields the local temperature variation and/or local heat flux coefficient as outcomes,
- describes the transient behavior of the system,
- is scalable. As a borderline case the model should be implemented by a special R-C network.

TABLE I
Summary of material properties

Fluid	**Density** kgm^{-3}	**Din. viscosity** $Pas \times 10^{-5}$	**Thermal cond.** $Wm^{-1}K^{-1}$	**Specific heat cap.** $Jkg^{-1}K^{-1}$	**Surface tension** Nm^{-1}
Water at 25^oC	997.0	89	0.6	4182	0.072
Nitrogen at 25^oC	1.145	2.1	0.0242	1040	
PDMS	965	-	0.15	1460	-
Si	2329	-	4980	710	-

A. Setting up the model solver

As a first implementation of the model a finite element simulation in *ANSYS Multiphysics 14* environment was performed. The model satisfying the aforementioned requirements was described in APDL (Ansys Parametric Design Language). The step-by-step process of the APDL code can be seen as a flowchart on Fig 1.
The two-dimensional geometry is defined first. Channel geometry, droplet size and length act as input parameters. Material and fluid parameters such as conductivity or heat capacitance, number of droplets, fluid velocity are also defined here. Any type of boundary conditions can be applied e.g. constant wall heat flux, constant wall temperature, volumetric heat generation inside the droplets etc.
A *fixed frame computational domain* is used which includes a part of the channel and a given number of droplets. In order to preserve the energy balance the number of droplets is kept constant during the simulation (the volumetric sum of the droplets is always constant).
The above steps are repeated until the desired simulation time is reached. Local values of the channel wall temperature and fluid mean temperature are available at all iteration steps.
The mesh is realized by a quadratic non-uniform meshing. The radial distribution of the nodes corresponds to the reverse Gaussian distribution, as the density of the nodes is higher at the wall-liquid interface region than in the middle of the channel. This non-uniformity of the mesh reduces the sum node number and therefore decreases simulation time. The distance of two node d_n at the normalized axial position r is

$$f(r,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

where $\mu = 3$ and $\sigma = 3$, the channel-diameter to channel-wall ratio is 10.2.
Material parameters are assigned to each node. Gas and liquid phase, channel wall and liquid film layers are treated as different materials.
Conduction and convection effects are considered by additional equations at affected nodes and will be discussed later.
A first, *steady-state* run is performed with all body force loads deactivated and the nodal temperature solution is saved. Mass

transfer is modelled by a movable mesh. The mesh of the fluid area is shifted by one node in each iteration step in the direction of fluid flow (Fig 2.). The nodes have a consecutive numbering from the left to the right and from the bottom to the top. Though the value of the nodes being shifted is formulated through a *modulo function*

$$\text{shift}_{eff} = \text{mod}(\text{iteration}_n, \text{cellnum}) \tag{2}$$

where $iteration_n$ is the number of the actual iteration while $cellnum$ is the number of nodes in axial direction. Therefore the nodes behave as a closed loop, the droplet exiting at the outlet immediately enters at the inlet.
On the other hand the loop is opened from thermally point of view, as a *Dirichlet* boundary condition is applied at the inlet (2). The sum heat exiting at the outlet is applied as a *Neumann* boundary condition and calculated as follows:

$$\dot{q}_{out} = \frac{A \cdot \sum_{n}^{i=1} c_{node} \cdot T_{node}}{\Delta t} \tag{3}$$

The nodal temperature result of the last run is loaded again as an initial condition in the following step. Body force loads are applied and a transient simulation is performed. Simulation time depends on fluid velocity as

$$\Delta T = \frac{d_{n,ax}}{v} \tag{4}$$

where $d_{n,ax}$ and v are the axial distance of the nodes and the fluid mean velocity respectively.
The liquid film layer thickness is calculated on a basis of the correlation proposed by Aussillous and Quere states as follows:

$$\frac{\delta_F}{R_0} = \frac{1.34Ca^{2/3}}{1 + 2.5(1.34Ca^{2/3})} \tag{5}$$

where δ_F and Ca is the film thickness and the capillary number, respectively. A mesh resolution of at least $\delta_F/2$ is set within the film area. The film layer is therefore modelled by a conductive liquid layer, as it was also suggested by He et al. (2009) [11].

B. Modelling convection

As a basic objective of the model it was stated that the solving time mainly caused by the iterative calculating of inertia forces must be reduced. In order to keep the model as simple as possible a constant velocity profile is used. This simplification is an essential condition to reduce the model complexity and therefore reduce running times. Obviously the approximation of the single phase parabolic velocity profile by a unit-step like function may cause an intolerable error. It will be shown that in Taylor flow the aforementioned errors may be satisfyingly handled.

- In laminar flow a thermal boundary layer develops very close to the wall and the characteristic effects of heat transfer occurs within this layer (assuming $Pr \approx 1$)
- In spherical bubbles the centerline part of the axial velocity profile more homogenous than in one phase flow thus the constant profile approximation does not result intolerable error.
- In thermal-entry length problems where the centerline part of the velocity profile is constant a higher Nusselt number is observed [12] in correlation with the experimental data observed by Talimi et al. [13]. In Taylor-flow a periodic thermal-entry length problem occurs. The thermal profile develops in every slug furthermore much faster than in single phase flow [5]. Therefore the sum error caused by the constant velocity profile approximation vanishes at the phase boundaries.

TABLE II
SUMMARY OF TEST CASES

Case No.	Re	Sensor	Slug	Reaction	R value
1	1	$650\mu m$	$560\mu m$	Surface	0.012
2	1	$400\mu m$	$560\mu m$	Surface	0.014
3	1	$800\mu m$	$1120\mu m$	Surface	0.062
4	1	$800\mu m$	$1120\mu m$	Volumetric	0.044
5	1	$650\mu m$	$560\mu m$	Volumetric	0.012
6	0.1	$650\mu m$	$560\mu m$	Surface	0.014
7	0.1	$400\mu m$	$560\mu m$	Surface	0.011
8	0.1	$800\mu m$	$1120\mu m$	Surface	0.109
9	0.1	$800\mu m$	$1120\mu m$	Volumetric	0.036

The aforementioned assumptions are reflected in Fig 5. where homogeneously mixed two phase flow under constant heat flux was simulated in Ansys *FLUENT*. The temperature rise from the channel inlet-to-outlet is plotted at different Reynolds numbers. It can be seen that the model serves reasonable results at wide spectra especially at very low Reynolds numbers acting the scope of our interest.

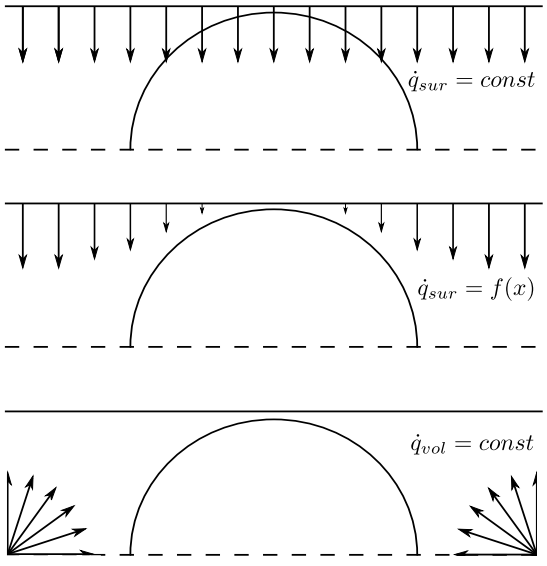


Fig. 4. Boundary conditions applied: a) constant wall heat flux, b) reaction at the liquid slug surface, c) reaction in the liquid slugs

C. Modelling internal circulation

The enhanced heat transfer experienced in Taylor flow is caused by the internal circulation of mass within the slugs. As the velocity field is not calculated by the model, the heat transfer enhancement effect of mass circulation should be considered by extending the conductive heat transfer equation by convective terms.

Let us assume a slug with internal circulation and a constant heat flux applied at its upper tangent. As the upper part of the slug is heated a thermal gradient occurs and heat is transferred in radial direction due to conductive heat transfer. Within the slug the heat is distributed by the constant mass circulation, therefore an improved heat transfer is experienced in radial direction, as it was experimentally shown by Muzychka et al. (2011) [6]. In axial direction thermal diffusion effects are dominant. The axial temperature distribution in the bubble changes rapidly due to about 20 times higher thermal diffusivity than of the fluid.

The above effects are estimated by a constant convective mass transfer applied at the different regions of the flow pattern regarding the dominant mass flow direction as depicted in Fig 3. The mass velocity equals the fluid mean velocity thus the complete changing of the mass occurs within one slug-length as it was pointed out by Muzychka et al. [6]. The following equation is solved for each element (four nodes form an element)

$$\int_V \left(\rho c \delta T(\tfrac{\partial T}{\partial t} + \underline{v}^T \nabla T) + \nabla^T(\delta T)([D]\nabla T)\right) = \int_{S_1} \delta T q^* dS_1 + \int_{S_2} \delta T h_f (T_w - T_m) dS_2 + \int_V \delta T \ddot{q} dV \quad (6)$$

where ρ, c, q, $[D]$ are fluid density, heat capacitance, heat flux and conductivity matrix respectively. Heat transfer occurs through internal fluid surfaces S_1 and fluid-wall interface S_2 at local wall and mean fluid temperatures T_w and T_m. Heat is stored in the fluid volume V, and finally the fluid may moving by meanstream velocity of v. Note, that in axial direction the mass transfer is modelled by the moving mesh, and the second term of the first integral is only evaluated while modelling internal mass fluxes.

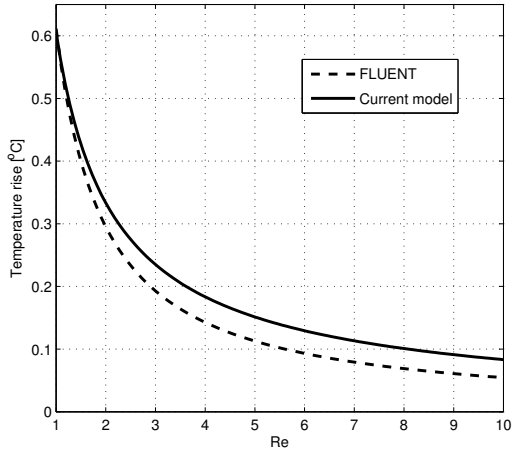


Fig. 5. Comparison of channel inlet-to-outlet temperature rise under constant heat flux of homogeneously mixed two phase flow and Taylor flow modelled by the current model

III. Test cases

Taylor flow of water and nitrogen gas with a capillary number of 0.0045 under different boundaty condition has been solved. See material properties in Table I. A sum of 8 droplets were introduced into the channel of diameter $320\mu m$ and a mainstream velocity of $0.0028m/s$ to $0.00028m/s$ resulting a Reynolds number of 0.1 to 1. Channel wall is built of $30\mu m$ PDMS and a part of the channel wall is substituted by Silicon in order to modelling temperature sensor device.

A constant wall heat flux boundary condition was used to validate the model as the analytical result of this case is known. Under constant heat flux boundary condition the surface heat flux is constant regardless which phase is present locally (Fig. 4a).

An approximation is used to model the heat generated due the surface reaction of the wall and the liquid slugs. This is the case where the wall is coated by enzymes and the liquid slugs carry its substrate. Regarding the Michelis-Menten kinetics the enzyme reaction could constantly generate heat if the substrate concentration is much higher than the enzyme concentration. It could be proven by the mass circulation inside the liquid slugs, therefore the heat generation effect should be more intensive where the circulation effect is dominant. This function is approximated by the Fourier row of the unit step function as follows:

$$\dot{q}_{sur}(x) = h \cdot \left(1 - \left(0.5 + 0.5 \cdot (\cos(x) + 0.5\cos(x)^2)\right)\right) \quad (7)$$

where h is the heat flux constant and x is the axial position (Fig. 4b).

Another relevant case is when the biological reaction occurs inside the liquid slugs (i.e. both the substrate and the catalyst are present inside the slug). This case is approximated a constant volumetric heat production for each node of the liquid volume (Fig. 4c).

The possibility of sensing the temperature changes develops due to biological reactions was also considered. A *sensor chip* made of silicon was inserted the channel wall just before the outlet. All heat flux generation were ignored at the sensor area. The thickness of the sensor chip is $30\mu m$. The time function of the average temperature of the sensor chip was investigated. The problem was solved for different sensor sizes and liquid slug lengths. The liquid slug is $560\mu m$ (short slug) or $1120\mu m$ (long slug). According to the slug length in case of short slugs the sensor length is $400\mu m$ and $650\mu m$ for short and long sensors, respectively. In case of long slugs the sensor length is $800\mu m$ (see Table II).

ISBN:978-2-35500-022-5

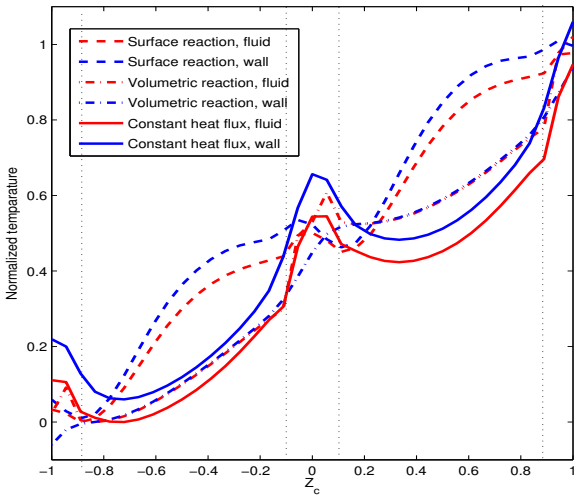


Fig. 6. Temperature profiles along the wall and fluid regions from $Z_c = -1..1$ at different boundary conditions

IV. Results

A. Performance analysis

The model implemented in ANSYS *Multiphysics* 14 was ran on an Intel Core i5 4 core CPU with 4 GB RAM under 64 bit Debian GNU/Linux (kernel 3.2.0-2-amd64). In order to achieve reasonable results an axial resolution of $80\mu m$ was applied for a channel diameter of $320\mu m$ which yields a sum of 2736 nodes. For the channel length of $6mm$ a sum of 75 iterations are needed to achieve the thermal equilibrium resulting a total runtime of 20 minutes.
The build up of the model requests $4.96s$ while the model solver runs $13.29s$ in each iteration. The *solve-time-to-build-time-ratio* is therefore 2.68 which grows rapidly for more complex problems, for a problem size of 21888 nodes the ratio becomes 3.94. This issue raises the need of parallel processing to reduce solution times. Figure 7. shows the build and solve times for different problem sizes and for serial and parallel processing. It can be seen that for smaller problem sizes the serial processing (1 CPU) gains shorter run times both for model building and solve processes. The *1 CPU-to-4-CPU-runtime-ratio* is 0.77 for 2736 nodes, however even for a bigger problem of 21888 nodes it remains below one, exactly 0.91. The iterations could not be threaded as the result of the last run is always requested, parallel processing makes sense only at huge node numbers. It is expected that for more bigger problems the parallel processing gains shorter runtimes than the serial one.

B. Thermal profiles

Thermal profiles along the channel axis was investigated under various boundary conditions detailed before. The temperature profiles are plotted at the axial position of the 2nd droplet at $Z_c = -1..1$ (Fig 6.) It can be seen that the thermal profile can vary characteristically under different boundary conditions applied.
For *constant wall heat flux* boundary condition the usual solution for thin wall was experienced. Due to the very moderate heat capacitance of the wall and very low Reynolds numbers of $Re = 1$ the wall can reach the thermally stable state and the wall and fluid temperatures run parallel. As heat generation is present at both phases the wall-fluid temperature difference does not differ significantly along the channel.
In case of surface reaction modelling, heat generation only occurs at the liquid slug. The heat generation rate depends also on axial position based on Equation 7. The temperature rises rapidly from $Z_c = -1$ to -0.5 where the slope of the temperature rise decreases due to the temperature gradient between the slug and gas bubble. Regarding the significantly smaller heat capacitance of the gas bubble its temperature reaches one of the liquid slug. Due to conductive heat transfer from the wall its temperature could be slightly higher than the slug's. Due to the high heat capacitance of the liquid slugs, far enough heat can be stored to drive the temperature sensor. Temperature gradients develop another way in case of modelling volumetric heat generation. Obviously a much bigger amount of heat is generated under this boundary condition. The heat generated at the liquid-gas surface heats up the gas bubble while its mean temperature rises above the liquid slug's temperature. Due to the internal heat generation the temperature is higher at the channel centerline then those of the wall, unlike all other cases. That is why the temperature of the gas droplet is higher than the wall temperature as well.

C. Temperature measurement analysis

The temperature of the liquid slugs is measured by a silicon sensor modelled by a silicon-plate inserted in the wall. Due the 1 : 33000 ratio of heat conductance of PDMS and silicon the heat stored in the liquid slugs and gas bubbles is conducted rapidly into the sensor substrate and rises its temperature. The temperature rise becomes proportional to the amount of heat stored inside the phases. Note, that temperature sensing should

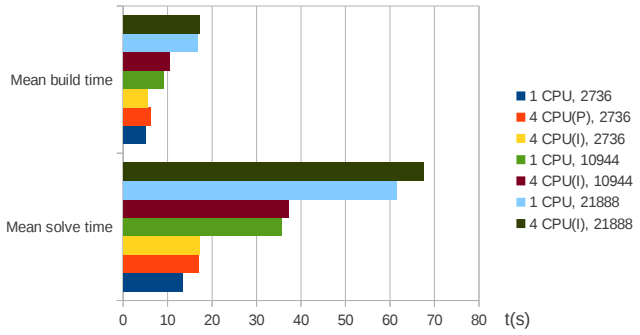


Fig. 7. Running time of each iteration for different problem sizes. Message passing interface (MPI) (P): PCMPI (I): Intel-MPI

be applied to the slugs individually, so each slug should be separated by the measurements. That is why the phases should be measured separately otherwise the heat differences of the phases would be averaged decreasing the sensor signal.

The main properties of the test cases can be seen in Table II. The results of the test cases are compared to each other through the amplitude of the output signal T_{osc} (the time dependent mean temperature of the sensor) and a normalized value called R and defined as follows:

$$R = \frac{T_{osc}}{T_{max}} \tag{8}$$

where T_{max} is the temperature of the droplet exiting at the outlet.

In the cases 1, 5 and 6 the sensor area is longer than the liquid slugs resulting the aforementioned averaging effects therefore gains very low R values. Worse values do not affected by the way of heat generation.

Low R value may be also experienced if the ratio of slug and sensor heat capacitance is too small. In case 2 and 7 short slugs are flowing under surface heat generation which results very low oscillation and R value as well.

Better results can be achieved by keeping the length of slugs higher than the sensor length. Case 3 and 4 show good R values regardless the way of heat generated. Obviously higher oscillation amplitude can be achieved by volumetric heat generation in case 4.

Droplet velocity can also affect both T_{osc} and R values. Lower velocities enable longer contact time at the sensor yielding higher heat accumulation. Lower velocity means also less convective heat transfer ratio increasing the slope of the temperature profile along the channel, which results also higher temperature. Case 8 and 9 are reflecting the aforementioned facts yielding very reasonable R values.

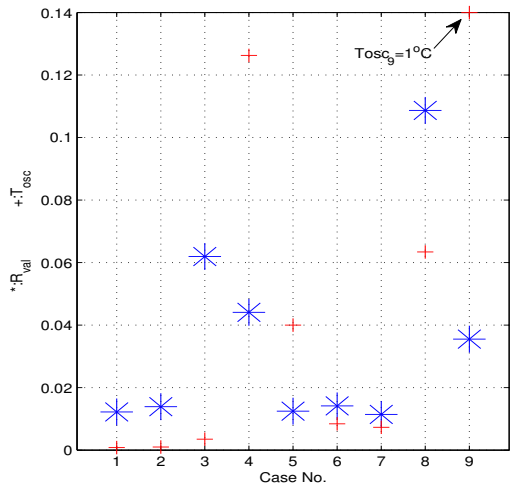


Fig. 8. Temperature measurement analysis at different combinations of boundary conditions, droplet and sensor sizes. Normalized goodness factor R is marked by *blue stars*, temperature oscillaton T_osc is marked by *red crosses*

V. Conclusion

Based on the general properties of Taylor flow a simplified model was introduced to describe heat transfer in microchannels. In order to validate the model it was implemented in a finite element solver, ANSYS *Multiphysics*.

Note the results of temperature measurement analysis where the same heat generation rate was applied in all cases. The effectiveness of temperature sensing can vary by a factor of 10 for the same heat generation rate only by changing the droplet size, velocity and the sensor size.

An other limiting factor of heat loss, due to heat convection or radiation from the channel outer surface was ignored in this analysis. This factor could worse the R values at low velocities which forms the problem into an optimization task.

As all of the equations building up each finite element used in this model could be substituted linear electric circuit models, the model is suitable to be implemented as a linear network. In the future work the possibilities of such implementation will be investigated. Model parameters regarding the kinetics of biological reactions should be also refined by using more complex models.

References

[1] D. Lakehal, G. Larrignon, and C. Narayanan, "Computational heat transfer and two-phase flow topology in miniature tubes," *Microfluidics and Nanofluidics*, vol. 4, pp. 261–271, May 2007.

[2] S. S. Leung, Y. Liu, D. F. Fletcher, and B. S. Haynes, "Heat transfer in well-characterised Taylor flow," *Chemical Engineering Science*, vol. 65, pp. 6379–6388, Dec. 2010.

[3] P. a. Walsh, E. J. Walsh, and Y. S. Muzychka, "Heat transfer model for gas–liquid slug flows under constant flux," *International Journal of Heat and Mass Transfer*, vol. 53, pp. 3193–3201, July 2010.

[4] R. Gupta, D. F. Fletcher, and B. S. Haynes, "CFD modelling of flow and heat transfer in the Taylor flow regime," *Chemical Engineering Science*, vol. 65, pp. 2094–2107, Mar. 2010.

[5] A. N. Asadolahi, R. Gupta, D. F. Fletcher, and B. S. Haynes, "CFD approaches for the simulation of hydrodynamics and heat transfer in Taylor flow," *Chemical Engineering Science*, vol. 66, pp. 5575–5584, Nov. 2011.

[6] Y. S. Muzychka, E. J. Walsh, and P. Walsh, "Heat Transfer Enhancement Using Laminar Gas-Liquid Segmented Plug Flows," *Journal of Heat Transfer*, vol. 133, no. 4, p. 041902, 2011.

[7] G. Taylor, "Deposition of a viscous fluid on the wall of a tube," *J. of Fluid Mechanics*, vol. 10, no. 10, pp. 161–165, 1961.

[8] A. B. Theberge, F. Courtois, Y. Schaerli, M. Fischlechner, C. Abell, F. Hollfelder, and W. T. S. Huck, "Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology.," *Angewandte Chemie (International ed. in English)*, vol. 49, pp. 5846–68, Aug. 2010.

[9] L. Zhang, J.-m. Koo, L. Jiang, M. Asheghi, K. E. Goodson, A. Member, J. G. Santiago, and T. W. Kenny, "Measurements and Modeling of Two-Phase Flow in Microchannels With Nearly Constant Heat Flux Boundary Conditions," *Measurement*, vol. 11, no. 1, pp. 12–19, 2002.

[10] F. Jousse, G. Lian, R. Janes, and J. Melrose, "Compact model for multi-phase liquid-liquid flows in micro-fluidic devices.," *Lab on a chip*, vol. 5, pp. 646–56, June 2005.

[11] Q. He, Y. Hasegawa, and N. Kasagi, *Heat Transfer Modelling of Gas-Liquid Slug Flow without Phase Change in a Micro Tube*. No. November 2009, 2009.

[12] W. M. Kays and M. E. Crawford, *Convective heat and mass transfer*. McGraw-Hill, Inc., 1993.

[13] V. Talimi, Y. Muzychka, and S. Kocabiyik, "A review on numerical studies of slug flow hydrodynamics and heat transfer in microtubes and microchannels," *International Journal of Multiphase Flow*, vol. 39, pp. 88–104, Oct. 2011.

25-27 September 2012, Budapest, Hungary

Quasi – 3D approach for BGA Package Thermal Modeling.

K.O. Petrosyants, N.I. Rjabov
Moscow State Institute of Electronics and Mathematics (Technical University)
3, Bolshoy Trekhsvyatitelsky per., 109028 Moscow, Russia
eande@miem.edu.ru

***Abstract*-The quasi-3D BGA package thermal model is proposed. The general 3D heat transfer problem is correctly transformed to the set of 2D equations for temperature distribution in different layers of BGA package construction. It is shown that quasi-3D model provides the reasonable accuracy for standard and thermally enhanced BGAs. The software tool OVERHEAT-BGA is developed to obtain an numerical solution with considerably (6 – 10 times) reduced CPU time in comparison with universal 3D simulators.**

The good agreement between designed and measured temperatures for standard and XP BGA packages is achieved.

I. INTRODUCTION

The ball grid array (BGA) is a packing solution for integrated circuits which provides the high I/O density, low assembly cost, low package profile and high electrical and thermal performance [1-3]. In response to the increasing demand for higher speed and power of semiconductor devices together with die shrinkage functional integration, various thermally enhanced BGAs were inverted to increase package power dissipation capability, such as Heat Slug Plastic BGA (HS-PBGA) [4]. Bottom Heat Slug BGA (BHS-BGA) [5], Enhanced BGA (EBGA) [6], Super BGA (SBGA) [7], Ultra BGA [8], Extra Performance BGA (XP-BGA) [9], etc.

So a thermal design is necessary to provide optimal temperature distribution within the package and attachment components changing of geometry, materials, process and application environment for different BGA package constructions. For this purpose the complex heat transfer problem in BGA packages is numerically solved using universal 3D simulators ANSYS [10], Flotherm [11], MSC/PATRAN [12], COSMOS [9] and others. However the estimation of sensitivity to various geometry and material parameters for large CFD and/or FEM models requires computationally demanding simulations, particulary for non-linear problems. Model reduction is an efficient means to obtain an accurate solution with acceptable CPU time [13,14].

In this paper the new quasi-3D approach for BGA package thermal modeling is proposed. The general 3D heat transfer problem is correctly transformed to the set of 2D equations for temperature distributions in different layers of BGA package construction. As a result the numerical solution and data interpretation are considerably simplified. In spite of this, the properties of material and temperature distribution in each constructive layer of BGA package are described and presented in more details without ensemble averaging.

II. QUASI-3D NUMERICAL MODEL OF BGA PACKAGE

The key points of quasi – 3D approach and the most important features of the software tool are considered by an example of the XP-BGA package developed as a cost competitive package for thermal margin [9]. The multilayer structure of XP-BGA is presented in Fig. 1.

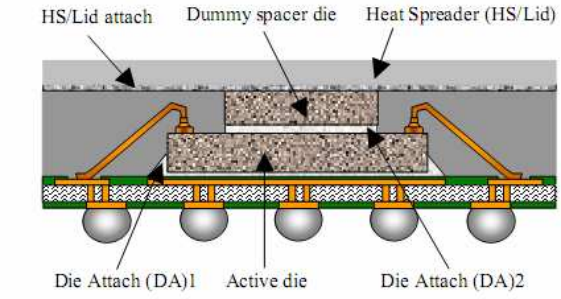


Fig. 1. Schematic of XP-BGA.

The vias placed under the active die (in central region) carry out the function of heat transport and the vias not under the die (peripheral region) – the functions of electrical contacts and heat transport. With respect to the standard BGA package XP mode has two pronounced features: a dummy spacer die on top of the active die and a flat heat spreader or lid on top of the package. The former offers a direct thermal path form the active surface to the heat spreader and the latter provides a good thermal coupling to the peripheral solder balls by reducing spreading thermal resistance.

For thermal analysis of BGA the quasi – 3D approach is used [15]. The ball diameter ZB, die and substrate thicknesses ZD and ZS are many smaller, then package and die layout sizes in plane xy. As a result, the temperature distribution along z axe can be assumed piece linear.

In this way, the classic 3D heat transfer equation transforms into the set of 2D heat transfer equations for the temperatures on package top $T_1(x,y)$, heat spreader bottom $T_2(x,y)$, spacer die top $T_3(x,y)$, spacer die attachment top $T_4(x,y)$, active die top $T_5(x,y)$, active die attachment top $T_6(x,y)$, substrate top $T_7(x,y)$, balls top $T_8(x,y)$. This approach allows for reduce computational cost essentially [15]. The temperature of heat sink top is assumed to be equal to

ambient temperature, on package top is convective heat exchange. These equations are:

$$\frac{\partial}{\partial x}\left[\lambda_1(x,y)\frac{\partial T_1}{\partial x}\right]+\frac{\partial}{\partial y}\left[\lambda_1(x,y)\frac{\partial T_1}{\partial y}\right]+ +\alpha\left(T_{AMB}-T_1\right)+\lambda_1(x,y)\frac{T_2-T_1}{Z_1}=0 \quad (1)$$

$$\frac{\partial}{\partial x}\left[\lambda_\xi(x,y)\frac{\partial T_\xi}{\partial x}\right]+\frac{\partial}{\partial y}\left[\lambda_\xi(x,y)\frac{\partial T_\xi}{\partial y}\right]+ +\lambda_{\xi-1}(x,y)\frac{T_{\xi-1}-T_\xi}{Z_{\xi-1}}+\lambda_\xi(x,y)\frac{T_{\xi+1}-T_\xi}{Z_\xi}=0 \quad (2)$$

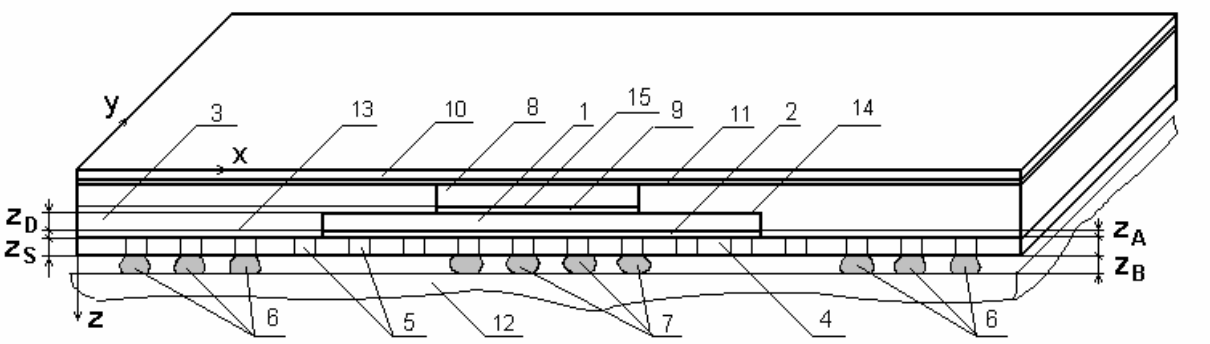
$$\xi=2,3,4,6,7$$

$$\frac{\partial}{\partial x}\left[\lambda_5(x,y)\frac{\partial T_5}{\partial x}\right]+\frac{\partial}{\partial y}\left[\lambda_5(x,y)\frac{\partial T_5}{\partial y}\right]+ +\lambda_4(x,y)\frac{T_4-T_5}{Z_4}+\lambda_5(x,y)\frac{T_A-T_5}{Z_5}= =-P(x,y) \quad (3)$$

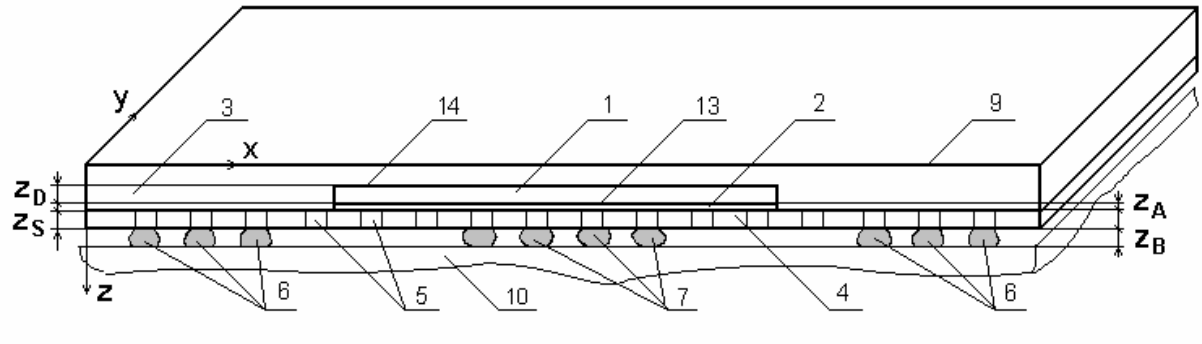
$$\frac{\partial}{\partial x}\left[\lambda_8(x,y)\frac{\partial T_8}{\partial x}\right]+\frac{\partial}{\partial y}\left[\lambda_8(x,y)\frac{\partial T_8}{\partial y}\right]+ +\lambda_7(x,y)\frac{T_7-T_8}{Z_7}+\lambda_8(x,y)\frac{T_{PCB}-T_8}{Z_8}=0 \quad (4)$$

Boundary conditions for (1) - (4) are established on side surfaces of package:

$$\left.\frac{\partial T_\xi}{\partial x}\right|_{x=0}=\left.\frac{\partial T_\xi}{\partial x}\right|_{x=X_S}=\left.\frac{\partial T_\xi}{\partial y}\right|_{y=0}=\left.\frac{\partial T_\xi}{\partial y}\right|_{y=Y_S}= =\alpha(T_\xi-T_{AMB}),\quad \xi=1,\ldots,8, \quad (5)$$



a)



b)

Fig. 2 3D geometrical BGA package models for Extra Performance (a) and standard (b) versions.

1 – active die, 2 – die attachment (DA1), 3 – molding, 4 – substrate (temperature distribution on its top is T_7, on its bottom is T_8), 5 – via, 6 – perimeter balls, 7 – thermal balls, 8 – dummy spacer die, 9 – die attachment (DA2), 10 – heat spreader, 11 – heat spreader attachment, 12 – heat sink, its temperature is T_{AMB}, 13 – plane of attachment top (temperature distribution is T_6), 14 – plane of active die top (temperature distribution is T_5), 15 – plane of dummy spacer die bottom (temperature distribution is T_4).

where: T_ξ – temperature distributions (see Fig. 2); T_{AMB} – ambient temperature; T_{PCB} –temperature of PCB surface; P – power density on die top; α – coefficient of heat exchange; λ_ξ, z_ξ – thermal conductivity and thickness of BGA package constructive ξ-layers, ξ=1,…,8; X_S, Y_S – package sizes on x and y.

Whole system of equations (1) - (4) describes thermal regime of XP BGA package (Fig. 2,a).

The model (1) - (4) is universal and can be generated for another type of BGA package with horizontal multilayer construction, for example, HS-PBGA [4], TBGA [10], HS-BGA [4], etc.

For standard BGA package geometrical model is automatically simplified: 8 – dummy spacer die, 9 – die attachment, 10 – heat spreader, 11 – heat spreader attachment are extracted (see Fig. 2b). As a result the equation system (1) - (4) reduced into the set of equations (1), (3) - (4).

The system of partial differential equations (1) - (4) is solved by finite difference method described in [15]. The software tool **Overheat-BGA** was developed. Its input data consist of three types of parameters:

1. parameters, describing BGA package: number and type of each construct layer, its sizes, physical and thermal properties of the material;
2. powers of active dies;
3. parameters directing the simulation process: $M_X \times M_Y$ grid sizes, accuracy of solutions.

Output data:

1. $M_X \times M_Y$ matrix of temperatures T_1, …, T_8 in (i,j) grid points;
2. multi-color temperature maps in plane xy for temperatures T_ξ;
3. average and maximal values of temperatures T_ξ.

III. Modeling Results

Thermal modeling of standard BGA and XP BGA packages was carried out using the software tool Overheat – BGA. External sizes of both packages are 17×17 mm, number of perimeter balls – 112, number of central thermal balls – 16 (see Fig. 3). Total power is P=4 W and uniformly distributed on the die top. Package sizes are presented in Table I [8]. Table II lists the material properties used for this study.

Localized 340×340 grid was used to capture the temperature distributions T_1, …, T_8. It was assumed that a converged result has been achieved if the junction temperature error is less than 0.01, ^{0}C. I.e. this structure has two perpendicular symmetry axes, we can calculate one quarter of it.

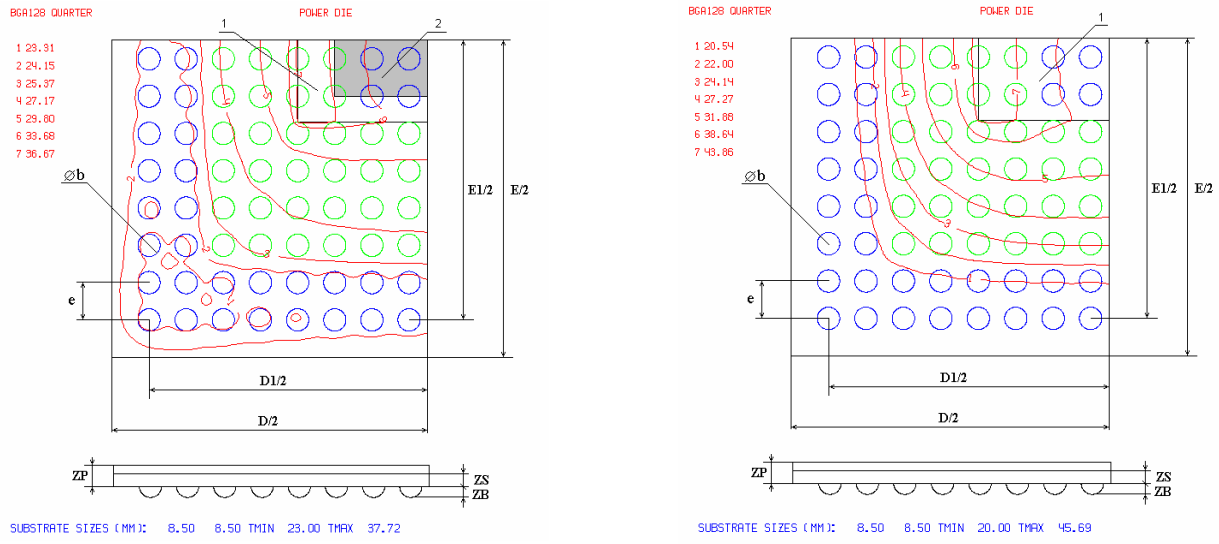
TABLE I
Parameters of BGA package

D, mm	17
D1, mm	15
E, mm	17
E1, mm	15
b, mm	0.6
e, mm	1
ZB, mm	0.5
ZP, mm	1.15
ZS, mm	0.36
Number of perimeter balls	192 or 112
Number of thermal balls	16
Die size, mm	7×4.4

Thermal analysis results of BGA and XP BGA packages with 112 perimeter vias and 16 thermal vias are shown in Fig. 3 where die top temperature distribution T5(x,y) (see eq. (3)) is presented.

Maximal and average temperatures of the die top are presented in Table III.

Analyzing the temperature distribution in the plane of active die top $T_5(x,y)$ for BGA and XP BGA, we can see, that using of XP BGA structure allows to reduce T_{MAX}, T_{AV} on 8 and 5 ^{0}C, respectively (17.5% reduction). This results are in good agreement with [9] where a 17% reduction in the free Θ_{JA} was achieved.



a) b)

Fig. 3 Temperature distribution on active die top for:
a) XP BGA; b) standard BGA. (1 – active die; 2 – dummy spacer die)

The importance of thermal balls as efficient heat parts between the package and the PCB was also investigated. It was shown by computer modeling, if the number of thermal balls in the standard BGA package was reduced from 16 to 8 the maximum temperature on the die top T_5 was increased by 6.7 ^{0}C. The removal of all thermal balls increased the die temperature by 17.1 ^{0}C.

TABLE II
THERMAL CONDUCTIVITY OF CONSTRUCTIVE MATERIALS

Material	Thermal Conductivity [W/(m K)]
Solder	50
Molding compound	0.8
Die attachment	2
Silicon Die	140
PCB	0.36
Copper	390

TABLE III
MAXIMAL AND AVERAGE TEMPERATURES ON THE DIE TOP, AMBIENT TEMPERATURE IS 20^{0}C.

Number of thermal balls	Temperature of die top, ^{0}C			
	maximal		average	
	XP BGA	stand. BGA	XP BGA	stand. BGA
16	37.7	45.7	35.8	40.8

IV. VALIDATION OF QUASY – 3D MODEL

The standard 17×17 mm^2 BGA package with 8.2×8.2 mm^2 die, a 3-row peripheral ball array was selected to compare the modeling and measurement results which are presented in Table IV. Total power is P=4 W, number of perimeter balls – 112, number of central thermal balls – 16 (see Fig. 3).

TABLE IV
COMPARISON OF THE MODELED AND MEASURED RESULTS OF 17×17 MM^2 BGA

	Θ_{JA}, K/W	Average temperature of die, ^{0}C	Die temperature increase after balls removal
Measurement [12]	23.3	41.6	57%
Quasi-3D model	23.9	40.8	60%

It is seen that the proposed quasi-3D BGA package model provides the results with reasonable accuracy.

V. CONCLUSIONS

The mathematical model based on the quasi-3D approach was proposed for BGA package thermal modeling. The software tool Overheat-BGA was developed to provide the numerical solution and data interpretation considerably simplified in comparison with universal 3D simulators.

The effectiveness of the new model and software tool was demonstrated for standard and thermally enhanced BGAs.

Pentium 4 CPU time for XP BGA and standard BGA temperature distributions (Fig. 3) simulations using quasi-3D model was 20 min. and 15 min. For comparison, the same simulations using 3D COSMOS model took 3 hours and 2 hours 30 min. accordingly.

By our experience, Overheat-BGA program may be utilized by the design engineers with a minimum of training and can be effectively used to study the different types of BGA packages to provide optimal thermal performance, changing the geometry, properties of materials, process and application environment.

ACKNOWLEDGMENT

This work was supported by Russian Foundation for Basic Research, grants 10-07-00689 and 12-07-00506.

REFERENCES

[1] Lau, J., Nakayama, W., Prince, J., and Wong, C. P., *Electronic Packaging Design, Materials, Processes, and Reliability*, McGraw-Hill, 1998.
[2] Harper, C. A., *Electronic Packaging & Interconnection Handbook*, McGraw-Hill, Second Edition, 1997.
[3] Lau, J., *Ball Grid Array Technology*, McGraw-Hill, 1995.
[4] Zahn B.A. Development of Improved Thermal Performance Embedded Heat Slug Plastic Ball Grid Array Packages//Proc 53-th Electronic Components and Technology Conf. 2003, p.p. 1490-1497.
[5] Bei, et al. Thermally Enhanced IC Chip Package//US patent N 6614660.
[6] Karnezos M. An EPBGA Alternative//Advansed Packaging, v. 7, Issue5, 1998,p.p. 90-96.
[7] Guenin B.M. et al. Analysis of Thermally Enhansed Ball Grid Array Package//IEEE Trans.-CPMT-A, v.18, N 4, 1995, p.p. 749-757:
[8] Miller H. Innox and Substrate Technologies Goal: Make from Fit Function//ChipScale Review, Jan. 2001.
[9] Y.Y. Ma, Desmond Y.R. Chong, C.K. Wang, and Anthony Y.S. Development of Ball Grid Array Packages with Improved Thermal Performance Sun. http://www.utacgroup.com/library/ EPTC2005_F2.5_P0149_Thermal_XP-BGA.pdf.
[10] H.J. Bo. Thermal Characterization of Tape BGA Package by Modeling//Proc. SPIE, 2003, v. 4229 (2000), p.p. 202-208.
[11] J. W. Jones, J. Li, Masaki Shiratori, Yuichi Yamazaki, Jidong Yang, Qiang Yu. Reliability Analysis of BGA Packages – a Tool for Design Engineers. http://www.mscsoftware.com/support/library /conf/auc99/p03599.pdf.
[12] Elkady Y. A., Thermal Performance of Ball Grid Arrays and Thin Interface Materials. A Dissertation Submitted for the Degree of Doctor of Philosophy. Auburn, Alabama August 8, 2005, 176,.
[13] E. Zukowski, J. Wilde, E.B. Rudnyi, J. Korvink. Model Reduction for Thermo-Mechanical Simulation of Packages//Collection of Papers. 11-th Intern. Workshop on THERMal INvestigation of ICs and Systems (THERMINIC-2005), Belgirate, Italy, Sept. 2005, p.p. 134 – 138:
[14] H. Oprins, M. Cupak, G. Van der Plas, P. Marchal, B. Vandevelde, A. Srinivasan, E. Cheng. Fine Grain Thermal Modeling of 3D Stacked Structures//Collection of Papers. 15-th Intern. Workshop on THERMal INvestigation of ICs and Systems (THERMINIC-2009), Leuven, Belgium October 2009 p.p. 45 – 49.
[15] Petrosjanc K.O., Rjabov N.I., Kharitonov I.A., Maltcev P.P. Software System for Semiconductor Devices, Monolith and Hybrid ICs Thermal Analysis. Proceedings of European Design Automation Conference EURO-VHDL'95 Sept. 1995, Brighton, UK. p.p. 360 -365.

25-27 September 2012, Budapest, Hungary

Context-Aware Energy Consumption of Mobile Wireless Communication

Marius Marcu and Sebastian Fuicu
"Politehnica" University of Timisoara
2 V. Parvan Blv., Timisoara, 300223, Timis, Romania
+40256403263, mmarcu@cs.upt.ro, sfuicu@cs.upt.ro

***Abstract*-Smart mobile devices like smartphones and tablets have fast penetration rates into almost all domains of our life. Frequent usage of mobile applications on such devices increases their resources usage and energy consumption, reducing thus their mobility and autonomy. Modern power management mechanisms are thus overwhelmed by today power hungry mobile applications. Therefore new power management and energy efficiency mechanisms are evaluated. One such mechanism is the integration of higher system abstraction layers into the power management process with the aim to build a cross layer energy management solution. In our work we aim to integrate high level user context information into the energy management solution in order to increase the energy consumption efficiency. This paper investigates the influence of user and communication infrastructure contexts on the energy consumption of wireless communication.**

I. INTRODUCTION

Nowadays, mobile wireless devices take active part to our daily activities as they increase in processing power, memory and storage capacities and number of built-in sensors. The cost users have to pay for using these enhancements is the reduction of the battery lifetime and consequently the decrease in autonomy and mobility [1]. Present battery technologies are currently able to sustain one day autonomy of frequently applications usage by smart mobile devices, compared to almost one week autonomy of common voice communication mobile devices. Therefore, further research efforts are needed in order to surpass current limitation of smart mobile devices autonomy and to increase the energy efficiency of mobile applications execution.

Most power management mechanisms are implemented and located at the lower layers of the mobile device, physical and operating systems layers [2]. The integration of the upper layers of mobile systems into power management mechanisms can decrease their energy consumption. Application-level power management mechanisms [3], take advantage of usage patterns associated with mobile applications and user activities such as email retrieval and web browsing.

In addition, the user intervention into the system's power management mechanism is going to be taken into account. The user is the best able to decide how to use the remaining energy within the battery. However, most of the users have no capacity to understand all aspects of power management mechanisms as they are implemented today. Therefore power management techniques should propagate to higher levels of the system, more specifically to the user applications level, in a more user friendly mode. There are two possible ways to implement user friendly power management interfaces: user-control interfaces and user-awareness interfaces. The first one allows users to actively interact with power management mechanisms in a much more informed way in order to select the most appropriate parameters based on their habits. The second one permits the higher-level applications and operating system to make the users aware of application specific energy consumption in order the user to change their behavior. The goal of these user interfaces is to allow the users to get information about the impact of the running applications on the energy consumption and in case of poor energy resources to have the ability to decide which are the priority applications which are allowed to utilize the remaining energy.

However, promoting power management mechanisms to user level is not an easy task due to the high user resistance to changes or due to the minimum effort expected from the user. Therefore we try to find a solution for this problem, because we consider that that final user has the most important information needed by power management mechanisms: what are the most important tasks of the user in order to allow them to finish using the available energy in the battery. In our work we aim to integrate high level user context information into the energy management solution without significant user intervention. This paper discusses the how user context influences the energy consumption of wireless communication interfaces.

The rest of the paper is organized as follows. Section II presents the state of the art of context based energy management and optimization. Section III describes the theoretical aspects of the work and the proposed solution design. Section IV discuss the measurement results we obtained based on the proposed test cases. Finally we conclude the paper in the last section.

II. RELATED WORK

The main advantage of mobile devices is that they allow users to get access to services anywhere, anytime through wireless communication medium. There are two widespread families of mobile networking technologies: the data communication services offered by mobile telephony networks and the 802.11 wireless LAN standards known as

well as WiFi technology. The first family contains the older GSM technology together with the newer 3.5G and 4G technology. The second family comprises the 802.11 a/b/g/n group of standards.

The [4] shows the pervasive computing improvements from 1990, where CPU clock frequency has attained 800 times higher in recent 12 years while battery density has just doubled in the same period. As we see, the battery life is crucial in portable wireless devices, especially in cellular systems, and thus we have significantly to save the power consumption at the terminal side

Generally speaking, context-awareness means that one is able to use context information. A system is context-aware if it can extract, interpret and use context information and adapt its functionality to the current context of use [5]. Our intention is to use the context information in such a way that will help the mobile device to choose among the three mobile technologies with the aim of preserving the energy as much is possible, based on the particularities of each wireless network technology. The [6] showed that Wi-Fi and cellular network interfaces (802.11b and GSM/EDGE) have energy profiles with complementary strength. Through a theoretical analysis the article showed that judiciously choosing between network interfaces can considerably improve battery lifetime under a broad range of application requirements.

In order to optimize the power consumption of a wireless transfer there are two kinds of approach. On one hand are the solutions which try to reduce the consumption by switching between multiple radio interfaces, trying to select the one which lid to a lower consumption and the other approach, regardless of wireless connection type, tries to conserve the energy using different strategies for data transfer.

As we can see in [7] the authors did an extensive measurement study of data transfer over the GSM, 3G and WiFi, showing that energy consumption is intimately related to workload and not just the total transfer size. The authors observe that besides the energy necessary for data transfer there are two overhead components named ramp energy and tail energy. Ramp energy overhead occurs in transitioning from the low power state to the high power state and the tail energy component is wasted in high-power states after the completion of a typical transfer. Summarizing the observations made by the authors, in 3G, a large amount of energy (nearly 60%) is spent by tail energy overhead. In GSM the tail time, is much smaller compared to 3G (6 vs. 12 secs) and the lower data rate of GSM implies that more energy is spent in the actual transfer of data. In WiFi, the association overhead is comparable to the tail energy of 3G, but the data transfer itself is significantly more efficient than 3G for all transfer sizes.

Using the same type of strategy, the authors of [8] has explored the idea of switching among multiple radio interfaces in order to reduce overall power consumption. In this paper the authors propose a system which gives the possibility to a wireless mobile device to automatically switch between multiple radio interfaces, in this case WiFi and Bluetooth, with the aim of increase battery lifetime. The main contribution of their work is an exploration of the policies that enable a system to switch among multiple interfaces, each interface having different characteristics.

The other approach is reflected in [9] which takes into account only the 3G technology for data transfer and propose a very promising strategy using timer alignment methods for reducing energy consumption. The authors observe that timing requirements for always-on applications are usually not strict and therefore application timers can be grouped together by using a timer alignment method. Using this method the authors succeed to reduce the power consumption of always-on applications by 50-60%. The main reason for the reduction is the alignment of network timers which implies a reduced number of radio access activations.

In [10] the authors consider only the WiFi connectivity and present a measurement study of the energy consumption of VoIP applications over WiFi-based mobile phones. The authors find that intelligent scanning strategies in WiFi can reduce power consumption for VoIP applications. The authors of [11] measured the energy consumption for Youtube-like video streaming applications in mobile phones using both WiFi and 3G. Their focus is on the energy utilization of various storage strategies and application-level strategies such as delayed-playback and playback after download. The author of [12] discovered that TCP downloads during 3G voice calls reduce energy consumption 75–90% compared to consecutive execution. Parallel downloading during VoIP calls lead to 30–40% savings and parallel TCP session reduce energy consumption by 0–20%.

A proxy-based energy adaptation framework is proposed in [13] where the authors utilize lossless data compression with the aim of saving energy on mobile devices when receiving data from wireless networks. The paper has three major contributions: a) proposing a practical solution adopting a decision-making mechanism based on context-aware policies, b) taking into account the compression ratio of different data types and the client contexts such as the energy utility of decompression, c) taking network conditions (throughput and SNR) into account when estimating the communication cost.

The solution proposed by the authors of [14] cannot be includes neither in the first category of papers nor in the second, because the article evaluate the energy saving that can be achieved with the energy-aware cooperative management of the cellular access networks of two operators offering service over the same area. The switching in not made between different radio interfaces but among different operators. This approach may become feasible only if operators are willing to cooperate, by accepting the competitor's subscribers as roaming customers.

We propose an approach which allows the utilization of the solutions developed to optimize the consumption of a mobile devices communicating by using a single type of wireless network interface, but in the same time our approach offers the possibility to choose, based on the contextual information, the best variant between available wireless network interfaces. A similar approach is proposed by [15] where context is used by battery management component.

III. User Context and Energy Consumption

The current research covering context-awareness expands from defining and understanding context to proposing different context-aware frameworks and applications. The term of context and context-aware in the computing domain was first used in 1994 by Schilit, Adams, and Want [16] but it's fully development was possible due to the spread of smart mobiles and tablets. The idea proposed in [16] was to make use of context information and adapt applications accordingly in order to provide the users with specific services based on their current context. The aspects of context that can be taken into account are location, time, people, devices and services nearby and changes in the previous aspects.

Both context and energy consumption aspects are very complex to model and implement by real-life applications. Therefore in order to reduce the complexity of the solution, the first step of our work was to address energy efficiency of wireless communication and to identify the main context aspects that impact the energy efficiency of the device while using wireless communication. The main goal of the paper is thus to discuss context aspects that influence energy efficiency of wireless communication. Every mobile user usually has several contexts while he benefits of network services using his mobile device. Such contexts may include home, company, school, friends, clubs, etc.

Wireless interface of the mobile device is usually shared by several user applications that needs network services. These applications are grouped function of their context execution, in two categories: context dependent and context independent. Context dependent wireless applications are used within a specific user context, like company, and wireless communication is bounded to this context. Context independent wireless applications on the other hand are not bounded with a specific user context and can consume network services from various contexts. Context independent wireless applications in their turn can be further divided into three classes based on the type of urgency of remote data access: instant, deferred and predicted access. Instant remote data access of wireless applications is needed when user requests urgent online data access, that cannot be delayed and that was not predicted and available locally. Deferred remote data access is initiated by the user, but data are not needed at the request time. Deferred data can be accessed at a later time when data is available within the best energy efficient user context. Predicted remote access is the access to local available data that was correctly predicted and downloaded in advance.

There is a kind of applications which are not requiring instant access to network services or remote data. For these types of applications is very suitable to have the possibility of choosing the best user context for wireless communication. The user context will be chose based on the history of energy efficiency of that user context. These kinds of applications are those which are used intensive on mobile device and generate the highest power consumption. Trying to optimize the communication generated by these applications will have significant impact on overall power consumption of the system.

Energy efficiency of wireless communication in different user contexts is estimated for every application which makes use of wireless network interfaces. In order to estimate the energy efficiency, power consumption and transfer rates of every wireless transfer must be known. Our solution is based on system monitoring, meaning that both power consumption and transfer performance are estimated online from measurable system's parameters. Wireless communication power consumption is estimated from power consumption of the whole system obtained using battery API.

Battery energy measurements specify the current status of the remaining energy available in the system. Battery measurements are valuable metrics for improving mobile systems and applications power management. Power management strategies applied for lower levels in the system usually do consider neither battery parameters nor other system parameter; they only implement certain policies, in most of the cases based on timeout or usage predictions. Mobile systems are subject to different and complex user and usage models, each with its own battery life profiles. Monitoring these profiles and detecting the changes in the network status we can estimate the power consumption of mobile wireless interface.

The proposed monitoring solution in Fig. 1 is based on a layered architecture containing the user interface layer (GUI layer), energy-efficiency logic specific layer (Profiling layer), component parameters monitoring (Monitoring layer) and the data acquisition layer (Data layer). The bottom layer is running on top of the host operating system (OS) and it provides a platform independent method to access and collect specific parameters from the underlying hardware and software components. Data acquisition layer uses the specific APIs provided by OS to read component related measurements which are further transferred to the upper layer in a uniform platform independent way. But the implementation of data access layer is platform dependent; therefore specific implementation should be developed for every new targeted platform or device. This layer provides access to read battery, telephony, radio, CPU and memory parameters.

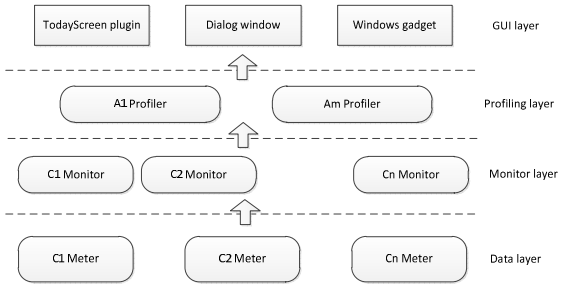


Fig. 1. Energy efficiency monitoring layers and modules.

The middle layer is platform independent and it implements the energy estimation logic. These layers contain intelligent component monitors and the energy-efficiency profilers that match the power consumption measurements and workload related phases to estimate the energy

consumed by every specific workload. The monitor components store historical context related data which are used to calculate energy costs of each workload in different contexts.

On top of the logic layer is placed the user interface layer which provides mobile users with simple access to energy costs of the last workloads and an estimation of the available time considering the current energy status of the battery and the current running workloads. The GUI layer is platform dependent and is designed to allow development of application specific GUI components: today screen or normal dialogs.

IV. Context Aware Energy Monitoring Test Results

In order to test the proposed solution we implemented a prototype for a HTC HD2 smart phone with Windows Mobile 6.5 operating system. We used this system because it offers OS API and unmanaged programming access at the system's components. The application was developed using Visual Studio 2008 development tools. We developed a client/server test application. The client is running on the client side and can transfer (upload and download) various data with the server application. Two communication protocols we used between client and server: TCP and HTTP. We selected one data block size to be used and randomly generated content. The wireless communication tests were run from three user contexts: home, school (shown as B413, B413 or HOL in the figures) and a partner software company (shown as LST in the figures). For each context 4 tests were executed as combinations between used protocols (TCP and HTTP) and wireless technology (WiFi and HSDPA). Every test were repeated at least three times and recorded for analysis.

The estimation of WiFi power consumption values in the three different user contexts are shown in Fig. 2. The user context located at the school laboratory we have several test executions at different distances from the main base station (WiFi access point). The parameter in the contexts that describes the distance from the base station is the RSSI.

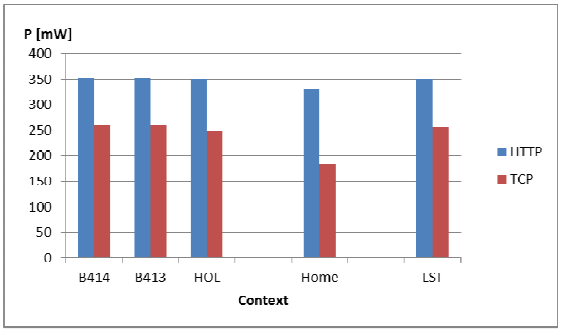


Fig. 2. WiFi power consumption function of user context.

Power consumption of the mobile device depends on the type of protocol used for communication: HTTP consumes with ~0.4 more power than TCP communication (Fig. 2). On the other hand the wireless communication transfer rates we measured vary in very large limits with the context of communication. For example the device used at home is a wireless access point that uses 3G connection for the uplink connection (Fig. 3). Therefore the energy efficiency of wireless communication depends also on user context (Fig. 4).

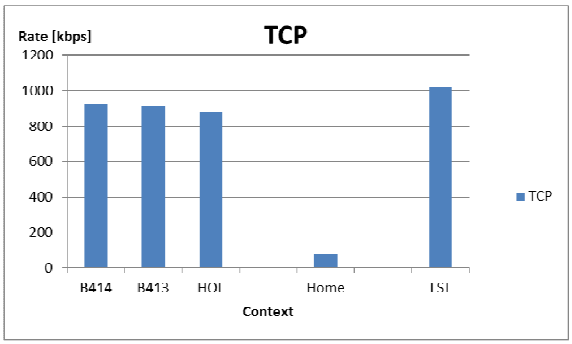


Fig. 3. WiFi transfer rates depending on user context.

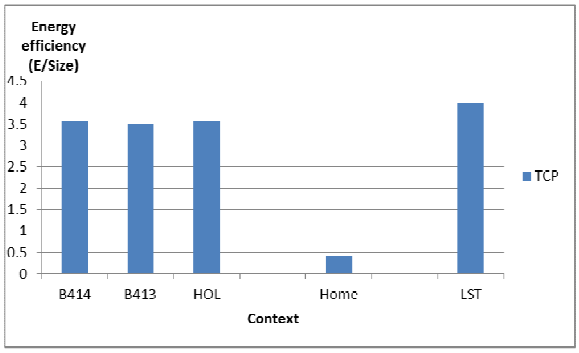


Fig. 4. Energy efficiency of different context.

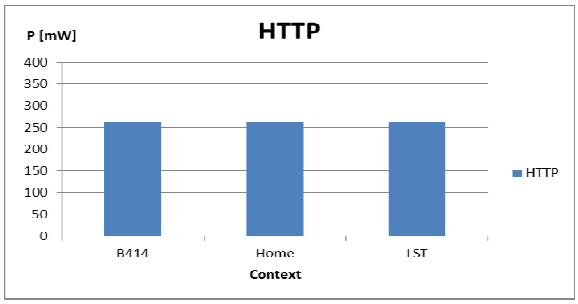


Fig. 5. HSDPA power consumption function of context.

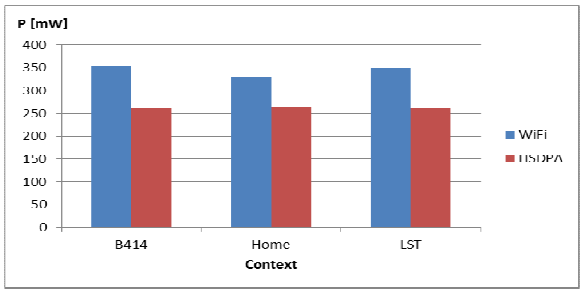


Fig. 6. HSDPA vs. WiFi power consumption.

We tested also various wireless medium access modes and technologies. Besides WiFi we used HSDPA connection. The tests results for HSDPA show similar power consumption levels for every context (Fig 5. and Fig. 6).

In our tests we obtained similar energy efficiency results for the same context and there are significant differences

between test executions in different contexts. This makes possible implementation of context based selection for wireless transfers in order to increase the efficiency of the running mobile applications.

V. CONCLUSIONS

The main contribution of our study is that it tries to offer o solution which allows the coexistence of two different types of approaches designed to reduce the power consumption of a mobile device. In fact, our solution may act "over" every kind of solutions mentioned in section II, which propose optimization for data transfer over a single type wireless network interface, without influencing their performance.

Based on contextual information gathered in time, it will decide which wireless network interface leads to a lower power consumption. In the same time the choice guaranties that communicating parameters will satisfied the requirements of running applications. In the future we intend to add the possibility for a user to define different kind of scenarios which will allow a more fine grained decision to be taken.

REFERENCES

[1] S. Cordova, "Harvesting the Future of Mobile", Green Innovations, ST-Ericsson Press, Feb. 2011.

[2] H. Zeng, C. S. Ellis, A. R. Lebeck, and A. Vahdat, "ECOSystem: managing energy as a first class operating system resource," SIGOPS Oper. Syst. Rev., vol. 36, no. 5, pp. 123–132, Oct. 2002.

[3] L. M. Feeney and M. Nilsson, "Investigating the Energy Consumption of a Wireless Network Interface in an Ad Hoc Networking Environment", Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2001, pp. 1548-1557, Anchorage, USA, Apr. 2001.

[4] J. Paradiso and T. Starner," Energy scavenging for mobile & wireless electronics", in IEEE Pervasive Computing, 4(1):18–27, 2005.

[5] J. Hong, E. Suh, S.-J. Kim, "Context-aware systems: A literature review and classification" in Expert Systems with Applications, vol. 36, no. 4, pp. 8509–8522, May 2009.

[6] A. Rahmati and L. Zhong, "Context-for-wireless: context-sensitive energy-efficient wireless data transfer," in Proceedings of the 5th international conference on Mobile systems, applications and services, New York, NY, USA, 2007, pp. 165–178.

[7] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications", in IMC'09, November 4–6, 2009, Chicago, Illinois, USA.

[8] T. Pering, Y. Agarwal, R. Gupta, and R. Want, "CoolSpots: Reducing the Power Consumption of Wireless Mobile Devices with Multiple Radio Interfaces," in Proceedings of the 4th international conference on Mobile systems, applications and services, Uppsala, Sweden, 2006, pp. 220–232.

[9] V. Könönen and P. Paakkonen, "Optimizing power consumption of always-on applications based on timer alignment," in 2011 Third International Conference on Communication Systems and Networks (COMSNETS), 2011, pp. 1–8.

[10] A. Gupta and P. Mohapatra, "Energy Consumption and Conservation in WiFi Based Phones: A Measurement-Based Study," in 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07, 2007, pp. 122-131.

[11] Y. Xiao, R. S. Kalyanaraman, and A. Yla-Jaaski, "Energy consumption of mobile youtube: Quantitative measurement and analysis", in NGMAST '08: Proceedings of the 2008 The Second International Conference on Next Generation Mobile Applications, Services, and Technologies, pages 61–69, Washington, DC, USA, 2008.

[12] J. Nurminen, "Parallel connections and their effect to battery consumption of a mobile phone", in 7th IEEE Consumer Communications and Networking Conference (CCNC 2010), Las Vegas, Nevada, USA, January 2010.

[13] Y. Xiao, M. Siekkinen, and A. Yla-Jaaski, "Framework for Energy-Aware Lossless Compression in Mobile Services: The Case of E-Mail," in 2010 IEEE International Conference on Communications (ICC), 2010, pp. 1–6.

[14] M. A. Marsan and M. Meo, "Energy efficient management of two cellular access networks," in SIGMETRICS Perform. Eval. Rev., vol. 37, no. 4, pp. 69–73, Mar. 2010.

[15] N. Ravi, J. Scott, L. Han and L. Iftode, "Context-aware Battery Management for Mobile Phones: A Feasibility Study", Sixth Annual IEEE International Conference on Pervasive Computing and Communications, 2008. PerCom 2008.

[16] B. N. Schilit, N. Adams, and R. Want, "Context-Aware Computing Applications", IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'94), Santa Cruz, CA, US 1994.

25-27 September 2012, Budapest, Hungary

Thermal aware design methodology for small signal discrete products

S. Holland, M. Röver, H.-J. Kühl, H.-J. Witt, R. Quade
NXP Semiconductors
Stresemannallee 101, 22529 Hamburg, Germany
E-mail: steffen.holland@nxp.com

***Abstract* - A methodology for electro-thermal simulations of discrete semiconductor devices on PCB level is presented. By combining the finite element simulation tools FloTherm and Sentaurus the computation time can be reduced so that it is possible to include the surrounding ambient into the electro-thermal model. This is achieved by extraction of thermal resistances using FloTherm which are inserted into Sentaurus as thermal boundary conditions. This approach has been verified with thermal measurements for a set of discrete devices in different packages and PCB layouts. The simulation results agree well with measurement.**

I. INTRODUCTION

The ongoing miniaturization of electronic appliances asks for increased electrical performance of semiconductor products in smaller and smaller packages. Frequently this leads to the situation that the thermal characteristics are limiting the electrical performance of the product. In addition the thermal management of the final application, for instance a smart phone, becomes ever more challenging with the ongoing densification on the circuits boards. Thermal characteristics can only be measured at the final product resulting in a very long design loop and the risk of cost intensive re-spins in the development of discrete products. Therefore a methodology for accurately modeling electro-thermal behavior on board level is needed.

Electro-thermal modeling of semiconductor devices is very challenging with respect to computing resources. A full finite element electro-thermal model of an application has many nodes which lead to very long solving times. This limits electro-thermal tools essentially to the simulation of the semiconductor die neglecting its surroundings. One way to reduce the amount of computation time is to use distributed lumped element models [1]. Another approach is to generate compact models to describe the electro thermal behavior [2]. Temperature distributions of detailed PCB boards can be simulated quickly with such an approach [3]. Even though electro-thermal modeling with finite elements is very computation time consuming, tools which only solves the thermal behavior can handle larger structures with much more nodes due to the fact that fewer equations have to be solved. Thus simulations of PCB-level structures are possible. Within the Therminator project a simulation flow for the electro-thermal behavior of discrete semiconductor devices on PCB-level has been set up and calibrated. The tools FloTherm and Sentaurus are used, but the flow is not limited to these. In general it can be used with every combination of an thermal-only solver and an electro-thermal simulation tool.

The paper is organized in three sections. In the first section we describe the simulation tools and flow for thermo-electrical simulations for a discrete device on PCB level as developed in the Therminator project. The second section covers the calibration setup and measurement. The third section shows the results. A discussion and an outlook is found at the end of the paper.

II. SIMULATION FLOW SETUP

FloTherm is a finite element simulation tool which takes only thermal effects into account. It includes thermal diffusion, radiation and convection into transient or steady state computations [4]. Due to the restriction to Cartesian coordinates big models with up to 5 x 10^6 cells can be handled in reasonable time scales. This makes FloTherm capable of simulating a discrete semiconductor device with its ambient surrounding like a PCB, housings or even further devices or heat sources in one model. Furthermore it is possible to use thermal compact models in FloTherm to further reduce the number of nodes.

Sentaurus is a finite element semiconductor device simulation tool. All relevant semiconductor physics models are implemented, thus allowing to reproduce the behavior of silicon based devices. Sentaurus is able to do 3D simulations and apart from semiconductor materials various other insulators and metals are available which enables full PCB level simulations. However, as the computing time increases with the number of nodes there is a practical resource limit since complex simulations would take days of computing time to finish. Thus there is the need to reduce the number of nodes as much as possible without sacrificing the accuracy of the results.

In this work the complete device and a small part of the PCB is modeled in Sentaurus. The model of BC847, a bipolar npn transistor with a die size of 300×300 µm² in a SOT23 package is shown in Fig. 1.The PCB up to a length of 2 mm away from the package is incorporated into the structure. Such a structure has about 50000 nodes. A transient simulation with four CPU cores takes about 3 hours. In order to model the remaining of the PCB thermodes (thermal contacts) are placed at the end of the modelled structure which set boundary conditions for the temperature distribution. They are indicated by purple lines in Fig. 1. These are the interfaces between Sentaurus and FloTherm and model the surrounding ambient. There are five

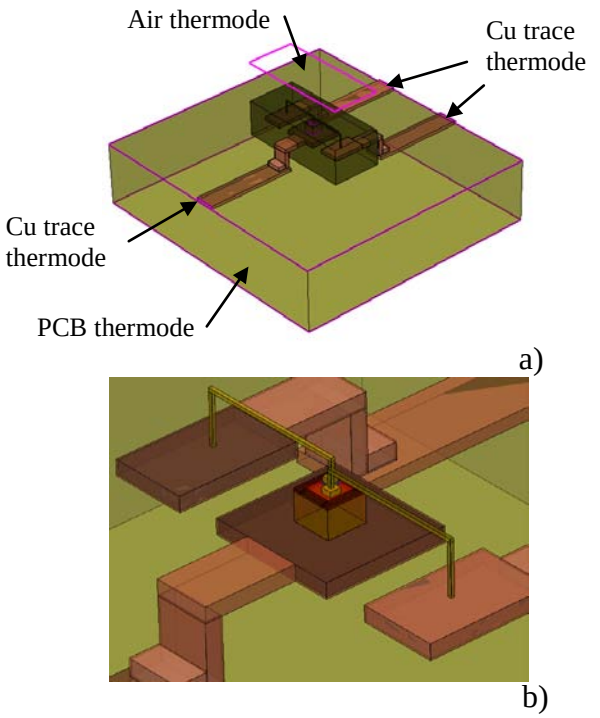


Fig. 1: a) Simulation structure used in Sentaurus of BC847, a bipolar transistor in SOT23. The purple lines indicate the boundaries of the thermodes used to model the surrounding ambient. b) Magnified region inside the package. To reduce the number of nodes the cylindrical bond wires are approximated by rectangles. For better visibility the mould compound is not shown.

different main heat dissipation paths of a device: three are attached to the end of the copper traces; one is set around the PCB edge which models the heat dissipation within the PCB material; the last one is placed directly over the package modeling the air convection.

Sentaurus allows defining lumped thermal surface resistances directly in the thermode statement [5]. These resistance values are determined by performing simulations with FloTherm. As these thermal resistances are connected in parallel it is not possible to determine each thermal resistance from one simulation only due to the fact that, although the temperatures at different positions can be determined by FloTherm, the ratios of power dissipation through the different paths are unknown. The thermal resistance of each path has to be determined by an independent simulation. With this approach it is ensured that the power can dissipate only along this path. The thermal resistance can be calculated by

$$R_{th} = \frac{T_{bound}[°C] - T_{amb}}{P[W]} \qquad \text{eq. (1)}$$

where T_{amb} is the ambient temperature (typically 25°C), P is the applied power and T_{bound} is the temperature at a position which coincides with the edge of the modelled structure. Sentaurus needs surface resistances (units are $K \cdot cm^2/W$) not thermal resistances. Thus the obtained R_{th} described above has to be multiplied with the area of the thermode. The setup of FloTherm to obtain the thermal resistances is described in the following. For the copper traces only the metal is present, radiation and convection is turned off. Each trace is simulated separately. The temperature T_{bound} is read at a 2 mm distance to the heat source (placed on a solder pad for the package) which coincides with the end of the simulated structure in Sentaurus. These simulations define the first three thermodes describing the copper traces in Sentaurus. For the determination of the R_{th} of the PCB convection and radiation are turned on and the package material property in FloTherm is defined as non-conducting. This eliminates or at least reduces the convection directly above the device. The copper traces are not present in this simulation. A fixed heat power is applied to the collector terminal of the package and the temperature T_{bound} is read in a distance of 2 mm.

Sentaurus is not able to simulate convection. The convection is modelled as solid state heat diffusion with an effective thermal conductivity. Therefore the device is placed in FloTherm on a PCB which is defined as non-conducting, radiation is turned off and the copper traces are not present. A power is applied to the device and the only dissipation path is now convection. At a distance d above the package the temperature is read as well as the temperature at the surface of the package. The two temperatures together with the power allow calculating the effective thermal conductivity of the ambient:

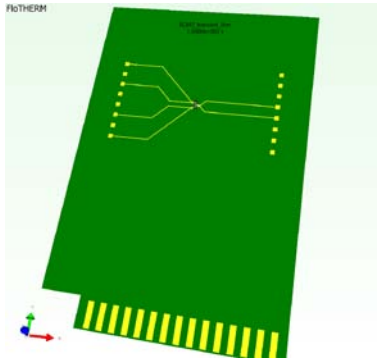
$$\lambda \left[\frac{W}{cm \cdot K}\right] = \frac{P\,[W]}{(T_{package} - T_{1mm})[°C] \cdot d} \qquad \text{eq. (2)}$$

The thermal resistance is calculated as usual by taking the temperature T_{bound} of the ambient in the distance d. Please note that if the PCB and the packages remain unchanged these values can be reused for different devices.

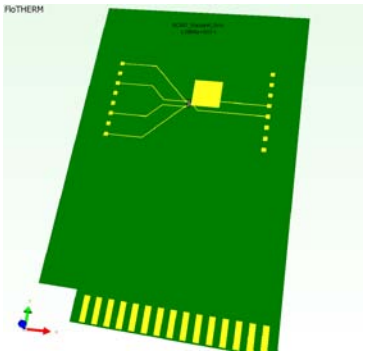
III. CALIBRATION

The devices are soldered on different PCB versions and Z_{th} curves are measured. The layout of the PCBs used is shown in Fig. 2. Two versions are used. The first has the standard reflow soldering footprint for SOT23 connected with 250 µm wide and 70 µm thick copper traces forming force and sense contacts. The second version has an additional copper heat sink with 1cm² directly connected to the collector pad (Fig. 2 b)). Both versions are JEDEC compliant. Multilayer PCBs are used to change the effective thermal conductivity of the board. All multilayer PCBs are made of alternating copper and FR4 layers as shown in Fig. 3. The layout of the top layer is the same as for single layer PCBs while the other three remain unstructured. The pn-junction of the emitter base junction is used as temperature sensor, which has to be calibrated at a fixed external temperature. This is done by placing the board into a chamber with given temperature between 25°C and 150°C while measuring the forward voltage drop V_f for different currents. A measurement current I_m is chosen in the range where V_f is linear with temperature. In order to avoid self heating of the device during measurement this current should be small. The slope of the obtained curve gives the temperature coefficient t_c which is used for calculating *Zth*. Furthermore this measurement generates the data for the DC calibration of the devices in Sentaurus.

Zth is now obtained with the following procedure. First a heating current I_{power} is chosen such that at steady state a



a)



b)

Fig. 2: PCB layout used in this work. A) std. footprint and b) footprint with 1cm² heat sink. The device is placed in the middle of the copper traces. The layout is JEDEC compliant.

junction temperature of 150°C is obtained. The ambient temperature is room temperature which is 21°C. To obtain Z_{th} for a given time a pulse of I_{power} of this length is applied to the device. The voltage drop at the end of the pulse V_{fpower} is recorded. Afterwards the pulse current is reduced to I_m. Due to intrinsic delays the tester is not able to measure immediately after the current has been reduced. Therefore it measures V_f at different times after the pulse and the value at the switching point is approximated with a parabolic fit. This extrapolated voltage drop V_{m0} is stored together with the pulse length. The procedure is repeated for pulse lengths between 10 µs and 1000 seconds.

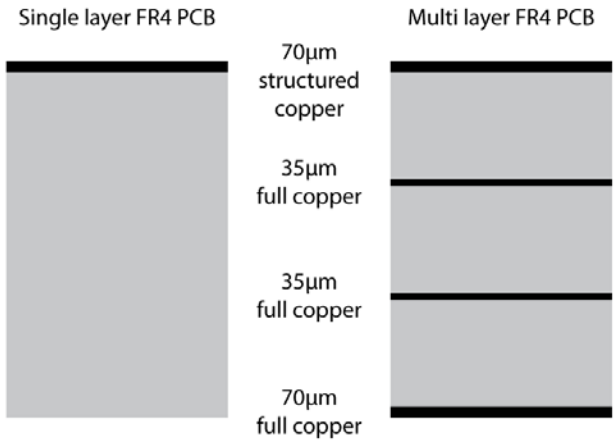


Fig. 3: left: Single layer PCB structure, right: Multi layer PCB structure

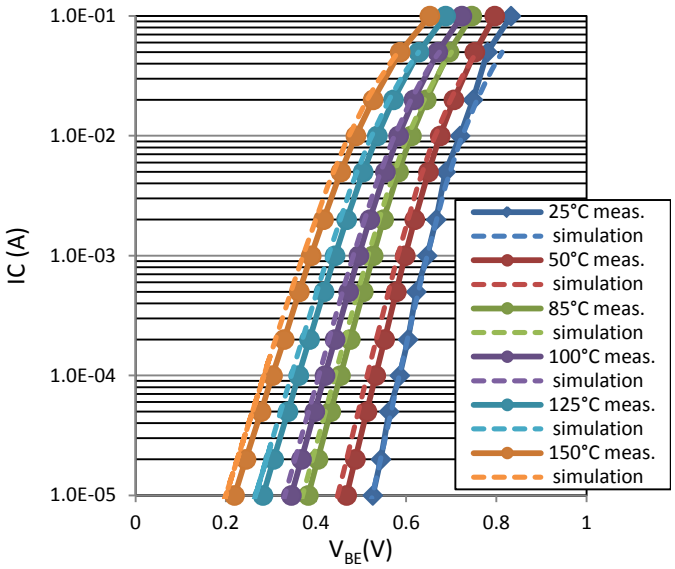


Fig. 4: DC IV curve of BC847 for different temperatures. The current is given and the voltage is read out. At lower currents and elevated temperatures the difference between simulation and measurement is about 20mV which corresponds to ~7 K at 150°C. The origin of this systematic offset is not known.

The time dependent *Zth* is calculated by the equation

$$Z_{th}(t) = \frac{V_{m0}(t) - V_{mcold}}{(t_c * V_{fpower} \cdot (I_{power} + I_m))} \qquad \text{eq. (3)}$$

where V_{mcold} is the voltage drop at room temperature when measured with I_m.

The ambient has a strong influence on the thermal behavior of the device. Thus not only the material parameters have to be characterized but also the geometrical dimensions. A cross section of the PCB at the position of the soldered device delivers the solder and copper trace thicknesses.

For an accurate electro-thermal simulation the simulated DC IV-curve has to agree with measurement for a wide range of external applied temperatures. In Fig. 4 a comparison of simulation and measurement is depicted in the range of 25°C and 150°C. They show a good agreement for all temperatures. At lower currents and elevated temperatures the difference between simulation and measurement is about 20mV which corresponds to ~7 K at 150°C. The origin of this systematic offset is not known.

IV. RESULTS

In order to verify that the methodology gives accurate results different devices in different packages were measured and simulated. Table 1 gives an overview of the used packages which cover a broad range of sizes, from being SOT883 the smallest and SOT89 the largest one. In the following detailed results are shown for BC847 (a bipolar npn transistor in the SOT23 package) from NXP Semiconductors. If not explicitly stated all simulation results are obtained from Sentaurus.

Table 1: Overview of investigated packages, covering a broad spectrum of different sizes as well as different technologies.

	SOT23	SOT883	SOT89	SOT1061
Package				
Dimension (mm)	2.9 × 1.3 × 1.0	1.0 × 0.6 × 0.5	4.5 × 2.5 × 1.5	2.0 × 2.0 × 0.65
Type	Gullwing	Exposed heat sink	Exposed heat sink	Exposed heat sink

The full transient thermal behavior of the full model (PCB with BC847) was simulated with FloTherm. Fig. 5 shows the simulated and measured Z_{th} curves for a PCB with standard footprint and with a 1cm² copper heat sink. Measurement and simulation agree well within a +/-10% error margin. Such a full model has about 10^6 cells. Simulation time for a full transient solution is about 3-5 hours with eight parallel solvers.

The results of the transient simulation show that the model represents the system well. However, for the extraction of thermal resistances a steady state solution is sufficient, which takes less than 30 min simulation time. The transient solution for 1000s and the steady state solution of the model deliver the same result. FloTherm describes the DC and transient behavior very well, thus it can be expected that the extraction of the thermal surface resistance gives accurate values. The measurements together with the simulated transient voltage of the emitter base junction of BC847 on a standard footprint obtained by Sentaurus are shown in Fig. 6 a). The voltage does not change much until 50 ms, i.e. the heating is not yet strong enough to change the temperature strongly as seen in Fig. 6 b). At larger times heating becomes increasingly important, the temperature of the pn-junction of emitter and base rises and the its voltage drops. Around 100 seconds the temperature does not increase any further and steady state is reached. The voltage shows a sudden change in slope in the simulations at around 30 seconds. This is most likely due to the fact that the remaining surrounding PCB with its copper traces is only modeled by thermal resistance boundary conditions in Sentaurus but the heat capacitances of the copper traces and the PCB were neglected. This leads to two effects: the voltage curve is not smooth (no RC-circuit) and the simulated temperature becomes higher than in measurement.

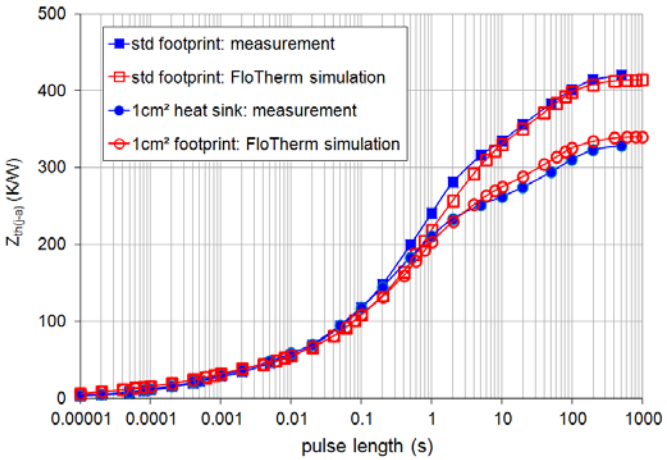


Fig. 5 Comparison of measured Zth for different pulse lengths (symbols). The blue squares represent measurement with a standard footprint, the blue circles measurements where a 1cm² heat sink has been added on the PCB in the direct vicinity of the device. The red open symbols show corresponding FloTherm simulations. For the described simulation flow a transient FloTherm computation is not needed but only a steady state analysis.

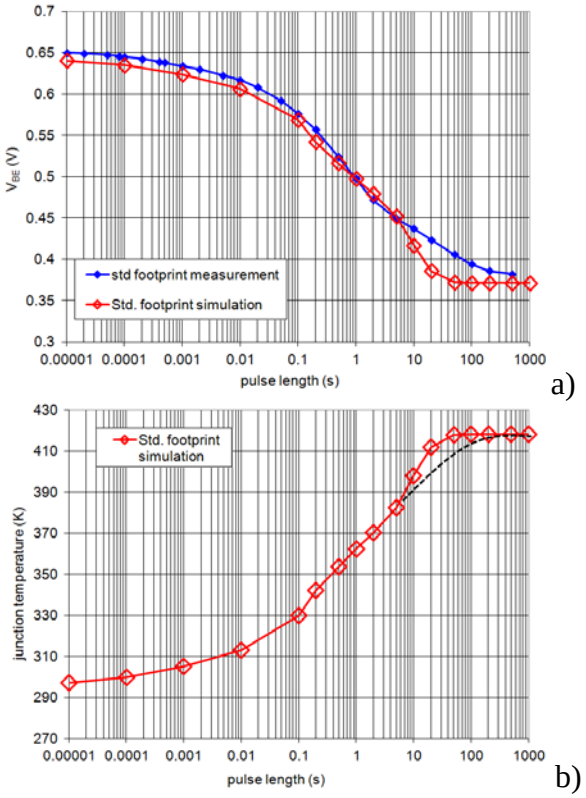


Fig. 6: a) Transient voltage of BC847 in a SOT23 package on a standard footprint PCB. The blue symbols indicate the measurement, the red symbols are simulation results. An abrupt change in slope is observed around 30s. This is most likely due to the fact that the remaining surrounding PCB with its copper traces is only modeled by thermal resistance boundary conditions in Sentaurus but the heat capacitances of the copper traces and the PCB were neglected. b) Simulated junction temperature. The dashed curve is an illustration of how the temperature could change if a thermal capacitance was included in the thermal boundary conditions.

Table 2: Simulated junction temperature obtained by Sentaurus at a pulse length of 1000 s (steady state). During the measurement the heating current was chosen such that in steady state the junction temperature reaches 150°C.

Type	Type	Package	Std. footprint (°C) (single layer PCB)	1cm² heat sink (°C) (single layer PCB)	Std. footprint (°C) (ML layer PCB)
BC847	BJT	SOT23	145	144	not measured
BZX84C7V5	PN Diode	SOT23	142	not measured	not measured
PBSS4160T	BJT	SOT23	150	not measured	not measured
BC847BM	BJT	SOT883	149	147	not measured
PBSS4480X	BJT	SOT89	155	not measured	not measured
BF620	BJT	SOT89	153	143	not measured
PMEG2010EPA	Schottky	SOT1061	152	152	not measured
PMEG4020EPA	Schottky	SOT1061	141	146	148

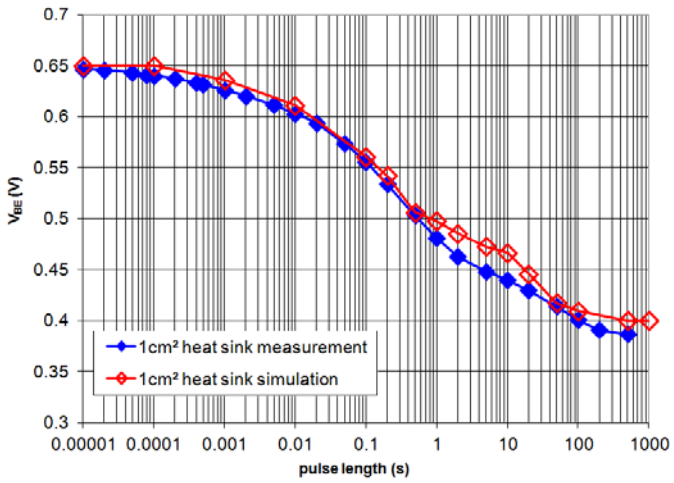


Fig. 7: Transient voltage of BC847 in a SOT23 package PCB with 1cm² heat sink. The blue symbols indicates measurements, the red symbols are simulation results. In contrast to Fig. 6 the simulated curve follows the measurement much better The reason is that in the simulation structure is larger because the full heat sink is modeled and therefore additional thermal capacitances are taken into account.

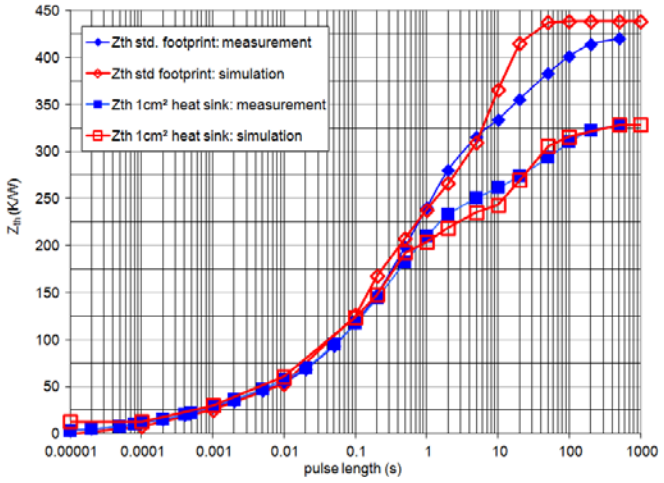


Fig. 8: calculated and simulated Z_{th} by using eq. (3). The effect of the 1cm² heat sink is now clearly visible: Z_{th} is only about 75% of the value of the Z_{th} measured with the standard footprint as more power is needed to reach a junction temperature of 150°C at 1000 s (steady state)

At smaller times the thermal capacitances are still incorporated in the modeled structure and the agreement between simulation and measurement is thus good. At large times the system has reached its steady state and therefore thermal capacitances no longer play a role and agreement is good again. The same measurement was done with the same device on a PCB with a 1cm² heat sink. The results are shown in Fig. 7. One can see that, in contrast to the simulation of the standard footprint (shown in Fig. 6) the simulated curve follows the measurement much better. The reason is that the simulation structure is larger (10 mm instead of 2 mm PCB in every direction) because now the full heat sink is modeled and therefore additional thermal capacitances are included in the Sentaurus structure. At a first sight the V_{BE} curves of Fig. 7 and Fig. 8 are very similar. This is due to the fact that each measurement is done in a way so that in steady state the junction temperature achieves 150°C. Transient Z_{th} curves can be compared much better since they also take the input power into account. Z_{th} is extracted by using (eq. 3) for both PCBs and shown in Fig. 8. A good agreement is achieved between simulation and measurement for different PCB layout except for the standard heat sink for a pulse length between 50 s and 200 s due to the missing thermal capacitances.

Table 2 shows simulated junction temperatures obtained by Sentaurus at a pulse length of 1000 s (i.e. steady state). The package types are very different in size, with SOT883 the smallest and SOT89 the largest one. They are packages with an exposed heat sink and thus have a different design than SOT23. In addition different PCB layouts have been tested. Each temperature is very close to 150°C and agrees well with the measured temperatures since the heating current was chosen in a way that in steady state the junction temperature reaches 150°C. This means that the described methodology works very well for a broad range of devices, packages and different PCB layouts.

In the following the main influences to the simulations are described. In general it is not important to model detailed shapes of the discrete device, but the reduction to Cartesian shapes is sufficient for an accurate representation of the system. This is true for PCB and sub-package sized structures. The exact thickness of the copper traces on the PCB has to be known. Junction temperatures of the devices can vary by more than 20°C by changing the copper thickness of the traces from 70 µm to 90 µm. The copper thicknesses of each employed PCBs in this work have been measured and adjusted appropriately in the simulation tools. For the FloTherm simulation the direction of the gravity has to be set according to the measurement. Changing the gravity from parallel to the PCB plane to orthogonal can change the junction temperature by up to 10°C. Since the PCB covers large areas which can effectively dissipate heat by radiation, radiation in FloTherm has to be switched on and surfaces of PCB and package are defined as radiating. In the package itself the thermal parameters of lead frame and die attach material are important factors for an accurate simulation. These two parameters become in general more important the smaller the package becomes. Especially for the die attach material it has to be noted that it is not important to model the exact metal layer composition of the crystal back side metallization. An eutectic attach it is thus modeled by a single gold layer of 2 µm, for glued crystals the back metallization can be omitted completely and only the glue itself has to be taken into account.

V. CONCLUSION

A methodology for electro-thermal simulations of discrete semiconductor devices on PCB level is presented. It combines the thermal-only simulation tool FloTherm with the device simulator Sentaurus. FloTherm is used to extract the thermal boundary conditions for Sentaurus. It has been verified that FloTherm models the thermal behavior of the system correctly. The results obtained with Sentaurus agree with measurement within a 10% error margin over a broad range of devices, packages and different PCB layouts which serves as a proof of the validity of the methodology.
In this model up to now thermal capacitances of the system are not included in the boundary conditions. Only thermal resistances have been considered. The extension to a full compact thermal model of the system surrounding the modeled structure in Sentaurus would be the next step to model more accurately the transient electro-thermal behavior of discrete devices.

ACKNOWLEDGMENT

We gratefully acknowledge funding by the Therminator project via the FP7 programme.

REFERENCES

[1] G. Bazzano, D.G. Cavallaro, G. Greco, A. Grimaldi, S. Rinaudo, “2D Thermal propagation analysis of discrete power devices based on an innovative distributed model technique and CAD framework”, 16th international workshop on Thermal investigations of ICs and Systems (Therminic), 2010

[2] E.G.T. Bosch, “Thermal compact models: an alternative approach”, IEEE Transactions on components and packaging technologies, 2003

[3] M. Rencz, V. Szekely, A. Poppe, B. Courtois, “Co-simulation of dynamic compact models of packages with the detailed models of printed circuit boards“, IEEE/SEMI International Electronics Manufacturing Technology Syposium, 2002, IEMT 2002, 27th

[4] FloTherm User Guide, Version 9.3, Mentor Graphics

[5] Sentaurus device user guide, Version D-2010.03, March 2010, Synopsys

25-27 September 2012, Budapest, Hungary

Coupled electro-thermal model for simulation of GaN power switching HEMTs in circuit simulators

Steve Stoffels, Herman Oprins, Denis Marcon, Karen Geens, Xuanwu Kang, Marleen Van Hove and Stefaan Decoutere
Affiliation(s): Imec Leuven, Kapeldreef 75, 3001 Leuven
Mailing address: Kapeldreef 75, 3001 Leuven, Belgium
+32-16-288506, stoffels@imec.be

Abstract- **In this paper we will present an electrical equivalent electro-thermal model for simulating GaN-on-Si high electron mobility transistors (HEMTs) in circuit simulators. The extraction procedure for the thermal sub-circuit will be shown, which is based on finite element modeling. The model will be verified by electrical measurements and micro-Raman measurements.**

I. INTRODUCTION

The promise of GaN is to deliver a paradigm shift in the performance of power switching semiconductor components. GaN-based transistors presented in the literature are already exhibiting better on-state resistances, switching losses and gate charge compared to state of the art silicon components [1]. Because of these properties, GaN components can exhibit much higher current densities, for smaller volumes than equivalent silicon power components [2]. When operated in a DC-DC convertor, these transistors switch from the off-state at high drain bias, to the on-state with a large output current but low drain bias. During this transition, they traverse through a highly resistive region (the saturation regime) where they dissipate substantial amounts of heat, thus increasing the channel temperature [3]. An increased channel temperature might degrade several key parameters of a transistor, such as leakage current, on-resistance, maximum current and might also have an influence on the reliability of the components. It is therefore important to characterize the heating of the component under these transient conditions and predict how a component will behave in a circuit when considering heating. Moreover, the choice of the substrate on which the HEMTs are grown has a profound impact on the self heating. Silicon has a worse thermal conductivity than e.g. SiC, but it is an order of magnitude less expensive than SiC and available in large wafer sizes, while still offering good performance. For this reason our work is focused on GaN-on-Si HEMTs. For this work a fully coupled thermal-electrical compact model will be presented which can predict the transient heating of GaN power components under switched mode operation, for a GaN-on-Si power technology. The model will be validated using electrical DC and micro-raman measurements.

In previous work we have investigated the validity of the Angelov model, when applied to GaN-On-Si power transistors [4]. In that work it was found that with some small modifications to the Angelov model [5], it was possible to accurately describe the output behavior of the transistors. Furthermore, we went on to calibrate the model at different chuck temperatures, which allowed us to find the relation between the model temperatures and the temperature. For this work we will couple the electrical model of [4] to a thermal sub-circuit, which can describe the temperature variation in the channel. Silicon has a lower thermal conductivity compared to e.g. SiC, meaning that it is more difficult to dissipate the heat generated in GaN-on-Si transistors. An accurate description of the thermal behavior of the GaN-On-Si component in steady and transient state is therefore essential to design and simulate the behavior of the transistors. It is important that the impact on the electrical parameters of processing steps, such as e.g. wafer thinning or carrier bonding, can be evaluated during the design phase. Therefore, a methodology based on the finite element method (FEM) was developed. A thermal model, representing the device and substrate was constructed in a FEM-tool and reduced through model order reduction to be usable in a circuit simulator.

The next section will give an overview of the electro-thermal model. In the following section the extraction procedure will be shown. In the fourth section the model will be verified by comparing with Raman and electrical measurements.

II. DESCRIPTION OF THE COMPACT MODEL

The electro-thermal model was constructed by coupling two electrical equivalent models, one describing the electrical part and a second sub-circuit for the thermal part (Figure 1). The model parameters for the electrical part were extracted from measurements and calibrated at different temperatures. The high temperature calibration and electrical model were described in previous work [4]. By performing the calibration, the model is able to predict the electrical response at all temperatures. Simulation of the junction temperature, used to calculate the exact value of the electrical model parameters, is performed by the thermal sub-circuit. For the thermal sub-

circuit a foster network of series connected RC circuits was chosen (Figure 1). This provides a sufficient level of abstraction for the thermal system, such that it can be used in a circuit simulator. The values for the resistors and capacitors in the thermal sub-circuit were extracted from finite element simulations by performing a transient thermal simulation and model order reduction. A fully parametric thermal model of the GaN power transistor was constructed, which allowed us to easily study different device topologies. The model and a simulated thermal profile (at t=1s) are shown in Figure 2. In a next step, an equivalent temperature was chosen, which represents the channel temperature in the lumped system. The value of the equivalent temperature was chosen to be equal to the average temperature under the gate finger (explained in detail in section III). The electrical and thermal parts of the model were coupled by exploiting the equivalences between thermal lumped systems and electrical systems. According to this equivalence, the temperature increase in the component is equal to the voltage over the thermal sub-circuit when the value of the current through the thermal sub-circuit is equal to the power dissipated in the component.

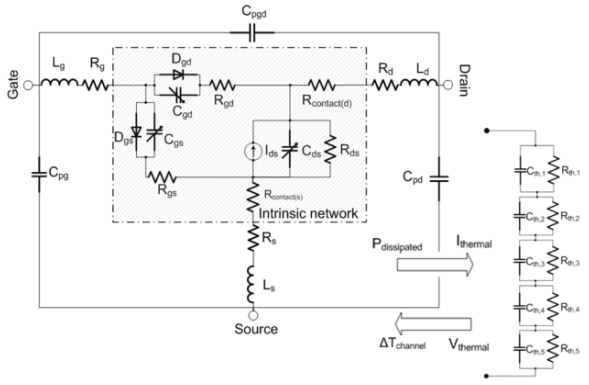


Figure 1 - Equivalent thermal sub-circuit, which is a Foster network, i.e. a series connection of parallel RC branches. The voltage over this circuit gives the temperature rise in a component, when the value of the current is equal to the power dissipation in the component.

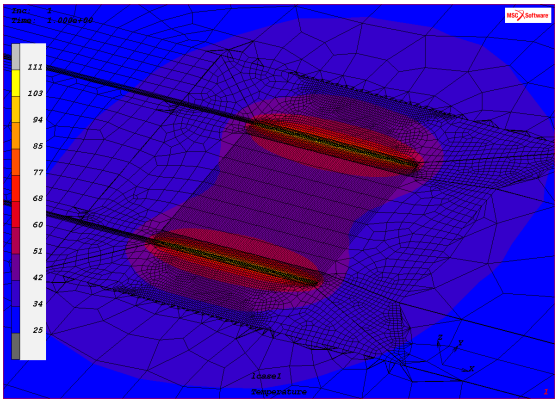


Figure 2 – Simulation of the thermal profile (at t=1s), performed with the finite element model constructed for a 2 finger GaN transistor. Each gate finger is 100µm long.

III. EXTRACTION OF THE THERMAL SUB-CIRCUIT

The extraction of the thermal sub-circuit was performed by means of finite element simulations. The equivalent model should be able to accurately simulate both the electrical DC and switching operation of the transistor. The finite element model should thus also capture the thermal transient behavior of the component. A complex 3D finite element model (FEM) representing a unit cell of the device on the Si substrate was constructed using the software code Msc.Marc[TM.]. The first step in defining the model is setting up the geometry, meshing and defining the material properties of all the layers. Figure 3 shows the geometry and the mesh of the constructed model for the unit cell of the GaN device. On the right hand side of Figure 3, the mesh of the model is shown in detail for the region close to the gates. The material properties for the different layers are summarized in Table .

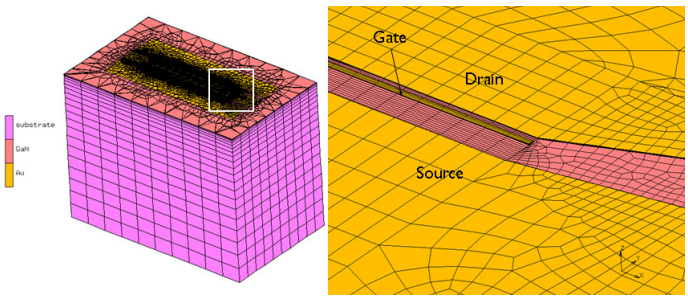


Figure 3 - Mesh and geometry of the GaN hemt.

Table I:

Conductivity and thickness of the materials used in the simulations

Material	Thermal conductivity [W/m-K]	Thickness [µm]
GaN	$130\left(\frac{T}{300K}\right)^{1.4}$	2
Au	310	0.15
SiO_2	$130\left(\frac{T}{300K}\right)^{1.4}$	0.15

In a next step the boundary conditions and initial conditions of the model need to be set. The boundary conditions depend in large part on the ambient in which the device is placed and on the way the device is packaged and mounted. The experiments for this work were performed on a bare die test structure, mounted on a probe station with a thermal chuck to perform measurements. The measurement conditions could in this way easily be controlled. For this measurement topology it can be assumed that the thermal chuck acts as an isothermal bath at constant temperature. The silicon substrate, at the bottom of die, is connected through a thermal resistance to the chuck. This condition can be mimicked by using a convective boundary condition, which allows specifying both the ambient temperature of the chuck as well as the thermal impedance between wafer and chuck. Using such a boundary condition has the added advantage that the thermal impedance can be modified to simulate e.g. the impedance of a package or PCB without needing large changes in the model. The value of the thermal interface resistance or thermal boundary resistance has been extracted by the fitting of finite element models of

multifinger GaN devices to the results of thermal measurements using micro Raman spectroscopy [6]. All other surfaces of the model are considered to be adiabatic.

Heat is generated in the channel region of the transistor due to joule heating. In saturation most of the power will be dissipated in the channel region under the gate, owing to the high impedance of this region under these conditions. For the model, the assumption is made that heat is generated uniformly along the length and width of the channel region under the gate. The channel length where the heat was generated was used as a model parameter, and was modulated with respect to the gate length. The calibration of the channel length parameter was performed by comparing the simulated steady state temperature to micro-Raman measurements of the channel temperature (Figure 4). It was found that the value for the modulated channel length needed to be set 1.5 times larger than the gate length to match experimental curves to simulated ones. This value of the channel length is in accordance with calibrated values which were obtained in previous work [6]. For the transistor considered in this analysis, the gate length was 1.5 µm, leading to a modulated channel length of 2.25 µm; the width for one finger was 100µm. To extract the thermal impedance of the structure, a unit power of 1 W was dissipated and the thermal response was monitored. For the considered two finger device, the total power dissipation of 1W corresponds to a heat dissipation of 5 W/mm in the channel region. This is translated to a heat flux of 2222 W/mm^2 used as boundary condition in the model.

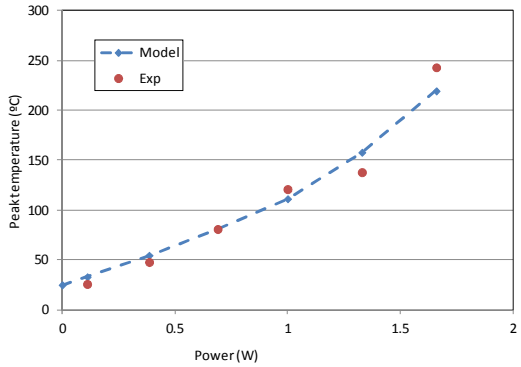


Figure 4 - Peak temperatures in the channel measured for different power dissipations. The measured temperatures were used to calibrate the model parameters. Simulations results for the calibrated model are also shown on the figure.

The goal of the model is to extract an equivalent thermal network which is able to capture the behavior of the component under switching conditions. For this we need to be able to simulate the transient behavior, thus the equivalent network should consist of thermal resistances and thermal capacitances. A steady state simulation is sufficient to extract the thermal resistances, however in order to extract the values of the thermal capacitances, a transient simulation is required. For this type of time dependent simulation, the specific heat and density of the materials are required material properties. In the transient simulation, a unit step for the power dissipation is applied as an initial condition. During the simulation, the temperature distribution in the model is calculated at each time step. Given the small dimensions of the heat generating regions, the initial heating will occur very fast. An initial time step of 10^{-12}s was used, so that the transient behavior could be captured, even for the fast initial heating. The result of the transient simulation is a 3D thermal map of the temperature at each point in the component for all the time steps. The thermal map after one second of heating is shown in Figure 2.

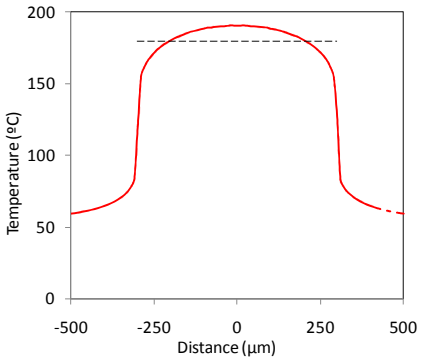


Figure 5 - Temperature profile in the GaN layer parallel to the gate. The dashed line indicates the average gate temperature

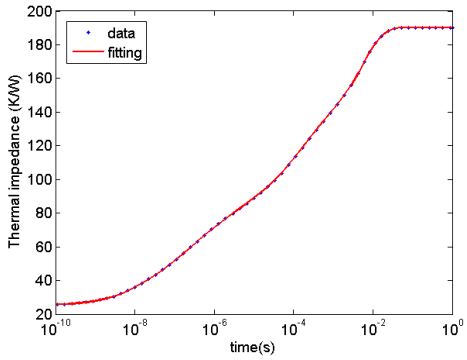


Figure 6 - Transient response of the average channel temperature for a power dissipation of 5W.

The main motivation for building the thermal model is to know the impact of temperature on the electrical behavior of the component. Therefore we need to investigate the thermal response in the channel region. In the linear regime the source/gate/drain channel regions all exhibit a linear behavior, dissipating equal amounts of power. However, once saturation is reached the carrier concentration in the channel under the gate will quickly start to decrease (pinch off), increasing the gate-channel resistance and the power dissipation of this region. Due to its high resistance, the region under the gate will be the main factor in determining the electrical output characteristics in saturation. The temperature for the compact model should therefore be represented by the maximum channel temperature along the length of the gate region. This maximum channel temperature can be determined for each point along the width of the gate. However when inspecting the value versus position along width of the gate (Figure 5), it is obvious that the maximum channel temperature does not remain constant. It is essential for the equivalent circuit that the channel temperature can be represented by a single value. To choose the proper value, one has to realize that the total current along the gate is the summation of all the contribution of unit

length gate segments. Due to the summation the effect of the temperature variation will be averaged along the width of the gate. Therefore, the temperature used in the model was the maximum channel temperature averaged along the width of the gate. The value for this temperature was used as a metric for determining the transient temperature response, shown in Figure 6.

The next step is to reduce the order of the finite element model such that it can be used in a circuit simulator. It is possible to describe the thermal response with a network of thermal resistors (R_i) and capacitors (C_i). This follows the thermal response of a one dimensional heat flow, which can be described as a series of following exponential functions

$$T(t) = \sum_{i=1}^{N} R_i \left(1 - \exp\left(\frac{-t}{\tau_i}\right)\right) \quad (1)$$

where τ_i is the thermal time constant

$$\tau_i = R_i C_i \quad (2)$$

with R_i and C_i the thermal resistance and thermal capacitance respectively. Each term of equation (1) can be represented by an equivalent thermal network consisting of a parallel connection of a thermal resistance R_i and a thermal capacitance C_i.

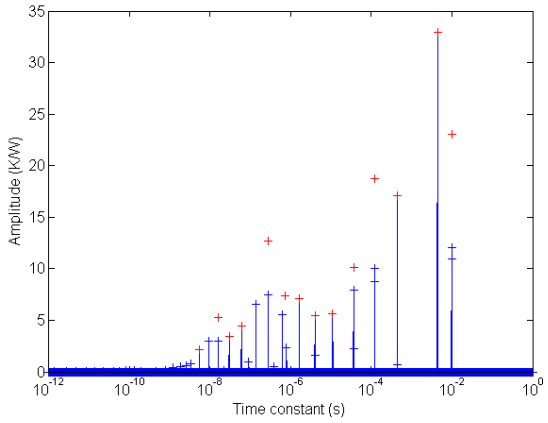
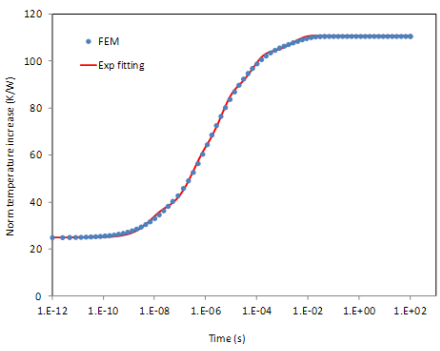


Figure 7 - Thermal time constant spectrum. Blue dots initial distribution, red crosses: after clustering.

Following procedure is used to convert the transient response extracted from the finite element simulations to an expression in terms of R_i and C_i. First the time constant spectrum of the transient thermal response is analyzed. This spectrum shows the contribution (amplitude) of each time constant. Figure 7 shows the time constant spectrum for the transient response of the average channel temperature of the 2 finger GaN transistor. The different time constants correspond to the different components in the thermal path. The very small time constants (below the µs range) correspond to the local heating in the GaN. Larger time constants (ms range) correspond to the heating of the Si substrate. In case of a packaged sample, larger time constants (tens of ms to seconds) correspond to the heating of the package and the even larger time constants (typically tens of seconds and above) to the convection cooling of the PCB on which the package is mounted.

In order to generate the time constant spectrum, first a discrete array with time constant values is generated. In the case of the heating of the GaN device, the values of the time constants are chosen between 10^{-12}s and 1s. Since the components of the transient response are best presented on a logarithmic scale, the time constant values are chosen using a logarithmic distribution between the lower and upper limit. In the second step, for each time constant and the DC component which represents the constant temperature off set, the corresponding contributions are found using a least square fitting of the transient response. The results of this step are shown as blue markers in Figure 7. In the next step, the detailed information of the fine grid of time constants is reduced by clustering the results of neighboring time constants. Due to the discretization of the time constant, small difference can occur based on the initial choice of the time constants. By clustering the results after the calculation of the amplitudes, the impact of the initial choice of the time constant array is limited. For the clustering, a tolerance is specified which is used as a criterion to decide which time constants to eliminate. The results of the clustering are shown by the red markers in Figure 7. Using the time constants and the corresponding amplitudes in equation(1), the transient response can be reconstructed. In Figure 6 the reconstructed response is shown as a solid red line. It can be seen that this representation gives a very good fitting of the response over the whole range of time constants.



Parameter	Value
R0	25
R1	7.6
R2	18
R2	28
R4	22
R5	11
C1	3.71E-04
C2	3.81E-06
C3	1.35E-07
C4	1.24E-08
C5	6.36E-10

Figure 8 - Fitting of an 5th order exponential series to the transient response of the average channel temperature of the two finger GaN device.

In the final step, the number of terms in the RC network can be reduced. After the clustering, the number of RC components is typically in the order of 20 to 50. For the case presented in Figure 7, 28 components are used after the clustering step. The objective of the extraction of the RC network is coupling it to the electrical simulator. For this application, a network with more than 20 components, attached to each node in the electrical simulation would take too much time to simulate. Therefore it is required to reduce the number of components in the RC network, without deviating significantly from the original transient response. In Figure 8 the reconstructed transient response is shown for 5 components in the thermal network. In case of a package device, additional larger time should be included to represent the thermal behavior of the package.

IV. VALIDATION

In a first experiment the DC characteristics of the unit cell were measured. Under these conditions the transistor is continuously dissipating power. The amount of power which is dissipated depends on the operational point. In the linear regime a limited amount of power is dissipated because of the low voltage drop over the component. However, once the component enters the saturation regime, the dissipated power will start to increase rapidly, because of the increasing resistance of the component. The power dissipation is given by the following relation:

$$P_{diss} = V_{DS} \cdot I_{DS} \qquad (3)$$

Because of the increasing power dissipation, and thus heat generation, the materials in the power transistor will start to heat up and therefore the value of the material properties will change. The change in material properties will also affect the output characteristic of the transistor. Therefore in the DC measurements at high V_{DS} and high current levels, one will start to see temperature dispersion [3]. This temperature dispersion can be simulated with the coupled electro-thermo model and gives a way to validate the modeling approach.

For the measurements V_{DS} was swept from 0V to 20V in steps of 0.5V and V_{GS} was swept from -2V to 2V in steps of 0.2V. The simulations were performed with the electro-thermal model presented in this work. The simulation was performed for the same range of biasing conditions used for the measurements.

The graphs produced by measurements and simulations were overlaid, shown in Figure 9. A good correlation can be seen between simulated and measured results. This already gives a first indication that the compact model gives a good abstraction for the actual device performance.

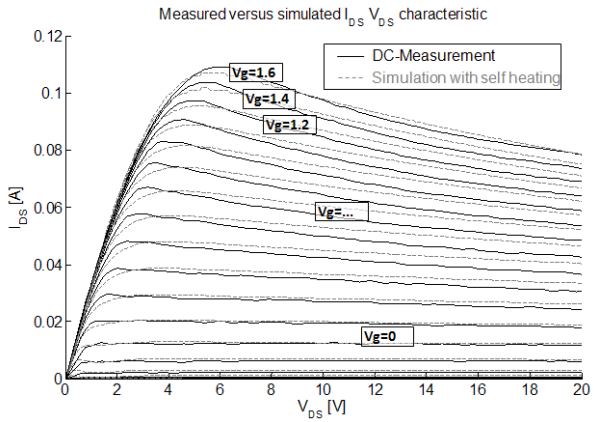


Figure 9 - Comparison between DC measurements (full black line) and the equivalent circuit compact model with the thermal sub-circuit (dotted gray line).

In a second experiment micro-Raman thermal spectroscopy was performed on the sample. For this type of measurements a laser beam is focused through a microscope on a sample. Due to interaction between the laser beam and the phonons in the material, an inelastic (or Raman) scattering of the laser beam will occur causing a shift in the frequency of the laser beam. The phonon distribution is temperature dependent and will thus cause a varying frequency shift with changes in the temperature [7].

Also strain in the material induces a change in the phonon distribution and related Raman frequency shift. GaN is a piezoelectric material and electrical fields will induce strain in the material. To exclude the effects of piezoelectric strain on the measurements, a calibration was done in the off-state of the transistor. Sweeping the V_{DS} in this state does not significantly heat the transistor as the dissipated power in this regime is extremely low due to the low leakage current. However, the large voltage will induce a piezoelectric polarization and thus strain in the material. In the off-state, the strain and temperature effects can thus be decoupled and used to renormalize the measurement results in the on-state.

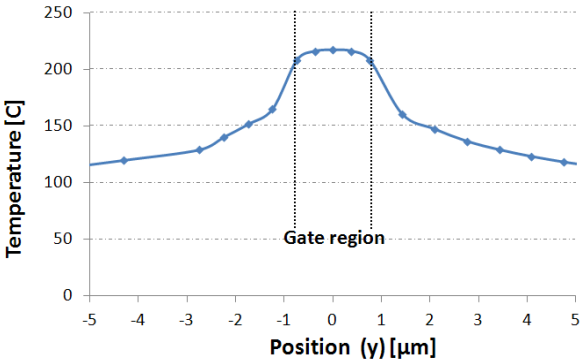


Figure 10 - Simulated thermal profile. Cross section of the thermal profile along the length of the gate, through one of the gate etch region. The center of the gate was taken as the reference position for the y-coordinate.

The micro-Raman measurements can give accurate temperature information of the material with submicron spatial resolution. The temperature point was probed on a single point of the transistor, near the edge of the gate field plate. The field plate is located 1µm away from the gate edge. The temperature quickly drops of after the edge of the gate. Therefore, to extract the exact temperature under the gate, a thermal model was constructed to simulate the temperature profile under the gate and extrapolate the exact temperature; the thermal model for the transistor was the same as was used to extract the thermal sub-circuit presented in section III. A steady state simulation was performed to determine the temperature profile. The result of the thermal simulation (Figure 10) indicates that the temperature drops quickly near the edge of the gate and the temperature in the channel is 30% higher, compared to the edge of the field plate. This value can be used to calculate the exact channel temperature from the micro-Raman measurements at the field plate edge.

The actual micro-Raman measurements were performed at 6 different bias points and at a chuck temperature of 25°C. The bias points were taken at a V_{GS} of 1.6 and V_{DS}

varying from 0V to 15V. For the simulation an extra node was added in the circuit, which can output the channel temperature at each simulated operational point. We then proceeded to simulate the I_DV_D characteristic for V_{GS} from -2V to 1.6V (0.2V steps) and V_{DS} from 0 to 20V (0.5V steps) and plotted for each point the temperature given by the temperature node in the compact model. The simulated and measured data were then overlaid showing a good correspondence over the whole measured range (Figure 11). This gives confidence that the thermal sub-circuit is simulating the correct temperature behavior.

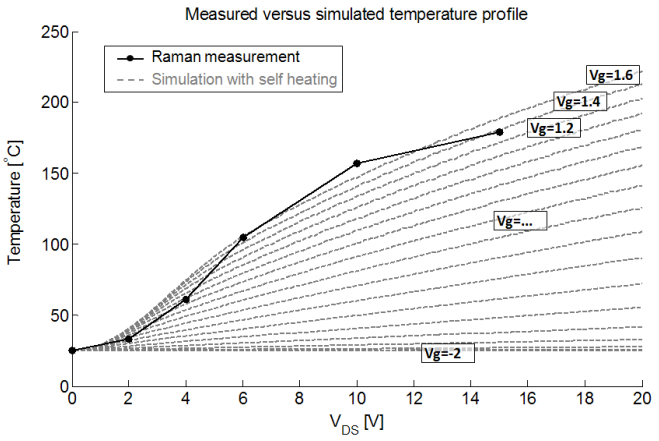


Figure 11 - Comparison between the temperature profile from Raman measurements at Vg=1.6V (full black line) and the electro-thermal compact model with Vg varied from -2V to 1.6V (dotted gray line).

V. CONCLUSIONS

In this work we have developed an electro-thermal model for predicting the electrical behavior of GaN-on-Si transistors in a circuit simulator. The methodology has been set up in a general way such that impact of substrate, geometry, wafer thinning, etc... can be investigated during the design phase.

It was shown that, after calibrating the thermal contact resistance and the heat generating channel length in the finite element model, we could get a good correlation between simulated and measured DC electrical characteristics. Furthermore, the temperatures were also simulated a various operational points and compared to temperatures measured with micro-Raman measurements. Also for this case there was a good agreement between the model and temperatures. This indicates that, once the model has been properly calibrated, it is able to accurately predict electrical and thermal characteristics of GaN-on-Si HEMTs.

VI. AKCNOWLEDGEMENTS

This work was performed under the framework of the European project Therminator (ICT-2009.3.2-248603-IP). The authors acknowledge the group of Martin Kuball at Bristol University for performing the micro-Raman measurements.

REFERENCES

[1] N. Ikeda, S. Kaya, J. Li, Y. Sato, S. Kato, and S. Yoshida, "High power AlGaN/GaN HFET with a high breakdown voltage of over 1.8 kV on 4 inch si substrates and the suppression of current collapse," in *Power Semiconductor Devices and IC's, 2008. ISPSD '08. 20th International Symposium on*. IEEE, May 2008, pp. 287-290.

[2] N.Ikeda, Y. Niiyame, H. Kambayashi, Y. Sato, T. Nomura, S. Kato, S. Yoshida, "GaN Power Transistors on Si Substrates for Switching Applications". *Proceedings of the IEEE*, Vol. 98, No. 7, July 2010 pp. 1151-1161.

[3] Duncan A. Grant, J. Gowar, "POWER MOSFETS: Theory and applications", *John Wiley & Sons*, 1989, ISBN-0-471-82867-X

[4] S. Stoffels, D. Marcon, K. Geens, X Kang, G. Van Der Plas, M. Van Hove and S. Decoutere, "High temperature calibration of a Compact Model for GaN-on-Si Power Switches" 2011. *Therminic 2011* pp. 159-163.

[5] I. Angelov et al., "Large-signal modelling and comparison of AlGaN/GaN HEMTs and SiC MESFETs," in *2006 Asia-Pacific Microwave Conference*, 2006, pp. 279-282.

[6] Das, J.; Oprins, H.; Ji H.; Sarua, A.; Ruythooren, W.; Derluyn, J.; Kubal, M.; Germain, M.; Borghs, G., 2006, Improved Thermal "Performance of AlGaN/GaN HEMTs by an Optimized Flip-Chip Design", *IEEE Trans. Electron Devices*, vol. 53 (11), Nov. 2006, pp. 2696-2702.

[7] M. Kuball, J. M. Hayes, M. J. Uren, I. Martin, J. C. H. Birbeck, R. S. Balmer, and B. T. Hughes, "Measurement of temperature in active high-power AlGaN/GaN HFETs using raman spectroscopy," *IEEE Electron Device Letters*, vol. 23, no. 1, pp. 7-9, Jan. 2002.

25-27 September 2012, Budapest, Hungary

Temperature dependent timing in standard cell designs

András Timár, Márta Rencz
Budapest University of Technology and Economics
Department of Electron Devices
Budapest, Hungary 1111
Email: timar|rencz@eet.bme.hu

***Abstract*—This paper proposes a methodology to simulate temperature dependent timing in standard cell designs. Temperature dependent timing characteristics are derived from standard delay format (SDF) files that are created by synthesis tools automatically. A case study is also presented in this paper where the temperature dependent frequency variation of a ring oscillator is simulated demonstrating the necessity of temperature dependent timing simulations. An adaptively refineable partitioning method for simulating standard cell designs logi-thermally is proposed as well. This paper also introduces recent enhancements in the CellTherm logi-thermal simulator developed in the Department of Electron Devices, BME, Hungary.**

***Index Terms*—temperature dependent delays, grid, logi-thermal, electro-thermal, simulation**

I. Introduction

In this paper the most recent improvements of the CellTherm [1], [2] simulator engine is introduced. The CellTherm logi-thermal simulator is capable of simulating standard cell integrated circuit designs given with Verilog structural description. The simulator couples a standard compliant logic simulator (e.g. QuestaSim©, Incisive©) and thermal solver engine and calculates device temperatures in function of simulation time. CellTherm also reads the full layout, power and timing data of the design and calculates temperature-dependent delays of the consisting standard cells. This way CellTherm not only can detect hot-spots in the design but can annotate device timing and propagation delays during the simulation. CellTherm also can create heating animations of the design and watch out for thermally induced timing violations with the help of the logic simulator.

II. Related work

In [3] Pable et al. deals with ultra-low-power signaling challenges caused by process, voltage and temperature (PVT) variations. Exponential dependency of subthreshold drive current on V_{th} and temperature in subthreshold operating region makes process and temperature variations of great interest while designing robust ULP systems. Small variation in the device V_{th} will translate into exponential variation in bias current and hence the device delay and power dissipation.

Rebaud et al. in [4] describe a new monitoring system, allowing failure anticipation in real-time, looking at the timing slack of a pre-defined set of observable flip–flops. They propose adaptive voltage scaling (AVS) and adaptive body biasing (ABB) to compensate PVT variations.

Lin et al. in [5] introduce a novel 9 transistor SRAM cell where PVT variations are taken into account. The proposed CMOS SRAM cell is PVT tolerant.

Kumar and Kursun in [6] attract attention to the fact that temperature-dependent propagation delay characteristics of CMOS integrated circuits will experience a complete reversal in the near future. They demonstrate that the speed of circuits in a 45-nm CMOS technology is enhanced when the temperature is increased at the nominal supply voltage. This is a quite interesting phenomenon contrary to the older technology generations.

In [7] Sánchez-Azqueta et al. introduce a CMOS ring VCO design where PVT variations were taken into account. To overcome the limitation of the PVT variations, a tuning range of about 20% is sufficient for their ring VCO.

In [8] the leakage current, active power and delay characterizations of the dynamic dual V_t CMOS circuits in the presence of process, voltage, and temperature (PVT) fluctuations are analyzed based on multiple parameter Monte Carlo method.

In [9] Winther et al. show that using wirelength as the evaluation metric for floorplanning does not always produce a floorplan with the shortest delay. They propose a temperature dependent wire delay estimation method for thermal aware floorplanning algorithms, which takes into account the thermal effect on wire delay.

[10] presents the temperature influence on energy consumption and propagation time delay in CMOS ASIC circuits with several measurements.

III. Design for demonstration

Our case study was a $10\mathrm{mm} \times 10\mathrm{mm}$ standard cell digital circuit with a 4-bit D-flip-flop chain and a ring oscillator circuit. The design is partitioned into 10×10 tiles where the temperatures are calculated. This design is a fictional design and cell sizes are intentionally enlarged to be able to demonstrate the effect of temperature variations and evolving hot-spots on cell propagation delay. Power dissipations for logic transitions in the cells are also fictional values large enough to spectacularly demonstrate the mentioned effects.

Standard Delay Format (SDF) files were bred to define inverter cell delays with which the temperature dependent

ISBN:978-2-35500-022-5

frequency of the ring oscillator can be demonstrated.

In Fig.1 the schematic layout of the design is shown. In the upper part of the chip the four D-flip-flops form the exciting circuit. The dissipated powers in the DFFR cells are intentionally chosen to be 1000-times larger (1mW) than the inverters' dissipated power per logic transition (1μW). In the lower part of the layout is the ring oscillator formed by 10 inverters and a kick-in NAND gate.

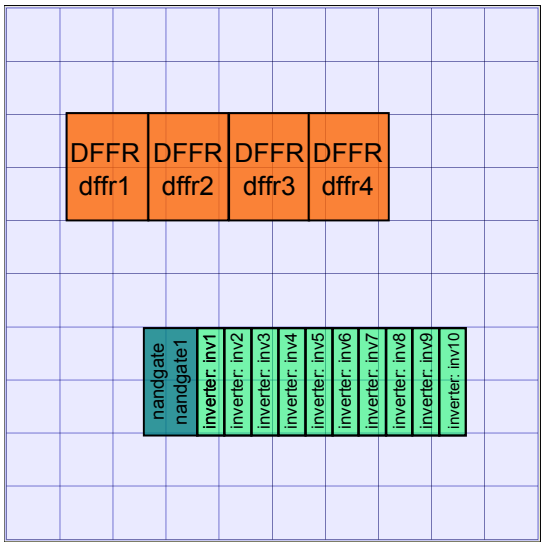


Fig. 1. Design layout with 10×10 partitioning mesh

IV. Grid partitioning

The surface of the standard cell IC is divided into subregions called *partitions* where dissipated powers are accumulated and temperatures are calculated. The partitioning approach speeds up simulation times and initial database parsing in large designs containing more than 1000 standard cells. This approach makes the initial thermal model generation practically insensitive to the number of standard cells. Of course logic simulation can take longer for designs containing high number of cells, the time taken by the thermal engine to solve equations remains the same because thermal equations are generated for the partitioning grid not the standard cells.

Powers dissipated in partitions and temperatures of cell instances are calculated using the partition area and overlapping cell area ratio. This means that if a standard cell falls partly into a partition, then the cell's power dissipation is taken into account proportionally to the overlap ratio.

Fig.2 depicts the power distribution of the design. In our test case, the DFFR flops are driven in a counter-like pattern, that is, switching activity of the flip-flop chain can be specified with (1),

$$A(\text{dffr4}) = 2 \cdot A(\text{dffr3}) = 4 \cdot A(\text{dffr2}) = 8 \cdot A(\text{dffr1}) \quad (1)$$

where $A()$ means the switching activity of the cell. The dissipated dynamic power per logic transition is proportional to the switching activity.

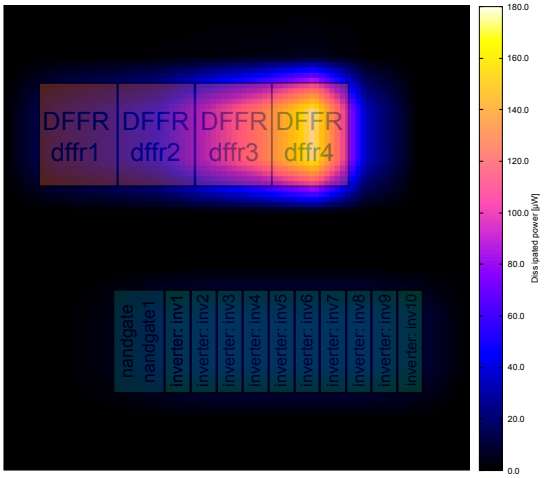


Fig. 2. Power map of test design

In the power map it can be clearly seen that the power dissipation of *dffr4* cell is the largest. The ring oscillator's power dissipation can also be seen in Fig.2 but it is much less than the flip-flop chain's dissipation even though its oscillating frequency is larger than the switching frequency of the DFFR chain. The resulting temperature map is shown in Fig.3.

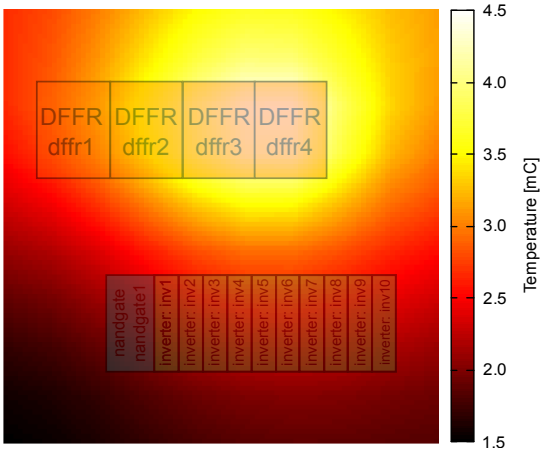


Fig. 3. Temperature map of test design

During the partitioning process partitions can be subdivided into subpartitions down to an unlimited depth. This way temperature and delay resolution in critical areas of the circuit can be refined. In a basic scenario the partitions can be sliced into 4 equal parts that can further be sliced into anouther 4 parts as shown in Fig.4

V. Data serialization

The initial database for the logi-thermal simulation is generated from the *LEF/DEF* layout files, Liberty *.lib* files, and *SDF* files. In a design with thousands of standard cells this initial database creation can take hours to complete. For example, in a design containing 1490 cells manufactured on a 65nm STMicroelectronics process the initial database generation and

 ISBN:978-2-35500-022-5

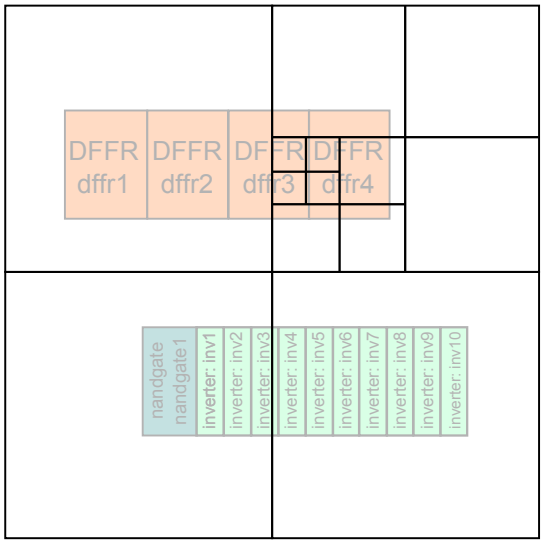


Fig. 4. Subpartitioning area of interest

thermal model generation took 56 minutes. This initial process of internal database creation and thermal model generation has to be done only once in the beginning of the simulation, however, when running simulations with different stimuli the database creation has to be done every time a new simulation is started.

In this paper a serialization method is described that speeds up initial loading times of simulations. After the initial database creation in the beginning of the simulation, CellTherm can be asked to serialize the created internal database to disk.

In computer science, in the context of data storage and transmission, serialization is the process of converting a data structure or object state into a format that can be stored (for example, in a file or memory buffer, or transmitted across a network connection link) and "resurrected" later in the same or another computer environment [11].

The serialized data can be read back from disk in fractions of seconds when starting a new simulation run thus the time consuming process of thermal model generation and database creation can be skipped. The deserialized data in memory will result in the same data as if the database creation phase had been run.

A comparison is presented here between the serialized and the pure initial loading time of the simulator. When loading without serialization the load time was 99.654 seconds. The serialized load time resulted to 0.262 second measured by the TCL command *time*. The thermal model generation time depends only on the structure and size of the chip die and packaging which does not change from simulation to simulation.

VI. Temperature-delay functions from SDF

Temperature-delay functions for the standard cells are calculated from synthesizer-generated SDF files. SDF files contain timing and delay data for cells in the placed and routed design. Synthesis tools usually generate these SDF files within voltage, process and temperature corners. Timing data are present in the SDF file for the worst case, nominal and best case corners. From these corner cases the corresponding temperatures and thus the related delays can be extracted. By interpolating the extracted temperature-delay corners delays can be calculated for arbitrary temperature values.

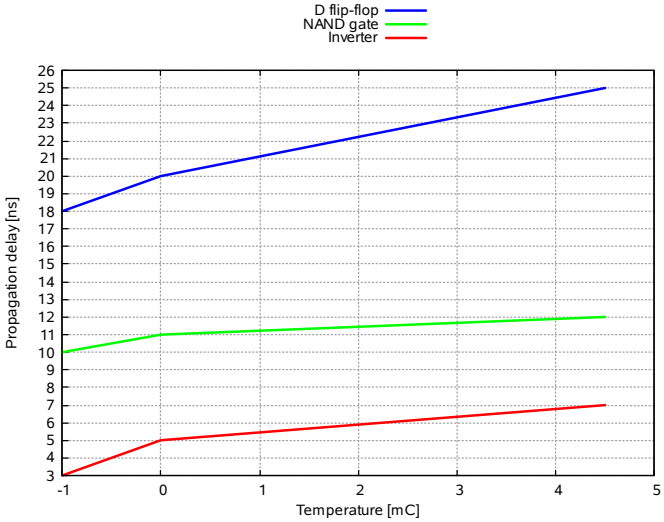


Fig. 5. Delay versus temperature functions for standard cells

VII. Temperature dependent delay simulation

In the test vehicle showed in Fig.1 the temperature-dependent frequency variation of the ring oscillator is demonstrated. By the dissipation of the D-flip-flop chain the resulting temperature map on the IC surface will not be uniform thus each inverter cell in the ring oscillator inverter-chain will have a different propagation delay. The delays are calculated from the standard cell's current temperature and updated in every simulation timestep. The varying delays of the cells will mistune the frequency of the ring oscillator. The changing frequency over simulation time is monitored constantly by the logi-thermal simulator and displayed to the user in every simulation timestep. The output frequencies of the ring oscillator after different simulation times can be seen in Table I.

VIII. Results

Temperature curves versus simulation time for the DFFR flip-flop chain and the *inv5* cell is shown in Fig.6. It can be observed that the circuit reaches steady-state temperature near the 4th second. The temperatures throughout this paper are differential not absolute temperature values.

In Fig.7 the period and frequency of the ring oscillator is depicted versus simulation time. It is clearly visible that with the increasing simulation time the temperature also rises thus the period of the ring oscillator starts to rise also. As the temperature of the functioning circuit rises the period of the ring oscillator also rises according to the temperature-delay functions in Fig.5. This increase in the period means a decrease

 ISBN:978-2-35500-022-5

TABLE I
OUTPUT FREQUENCY OF RING OSCILLATOR DEPENDING ON DEVICE TEMPERATURE

Simulation time [ms]	Period [ns]	Frequency [MHz]
10	122.0000	8.196721
250	131.6340	7.596821
500	136.3500	7.334067
750	139.0860	7.189796
1000	140.7680	7.103887
2000	143.2520	6.980705
3000	143.7240	6.957780
3250	143.7700	6.955554
3500	143.7980	6.954200
3750	143.8240	6.952942
3800	143.8240	6.952942
3900	143.8360	6.952362
4000	143.8420	6.952072

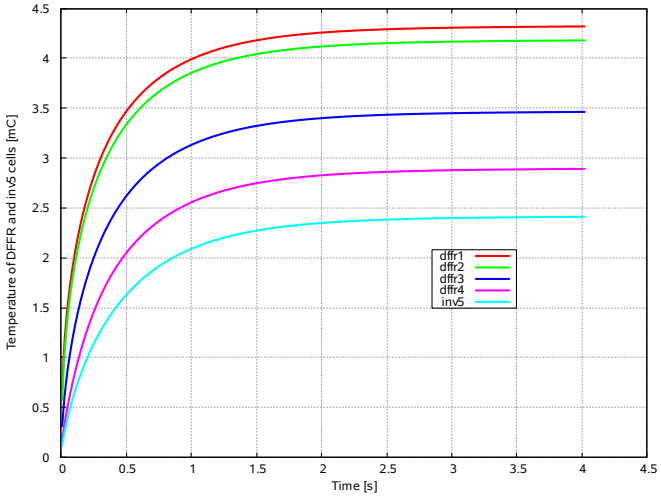


Fig. 6. Temperature of *DFFR* and *inv5* cells

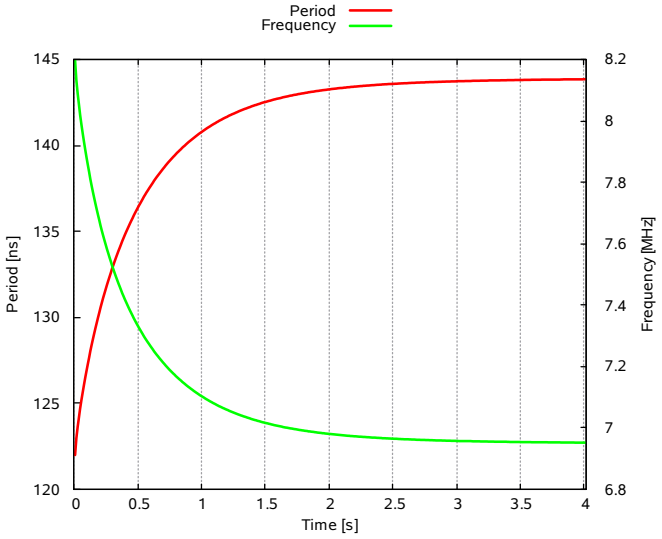


Fig. 7. Period and frequency of ring oscillator over time

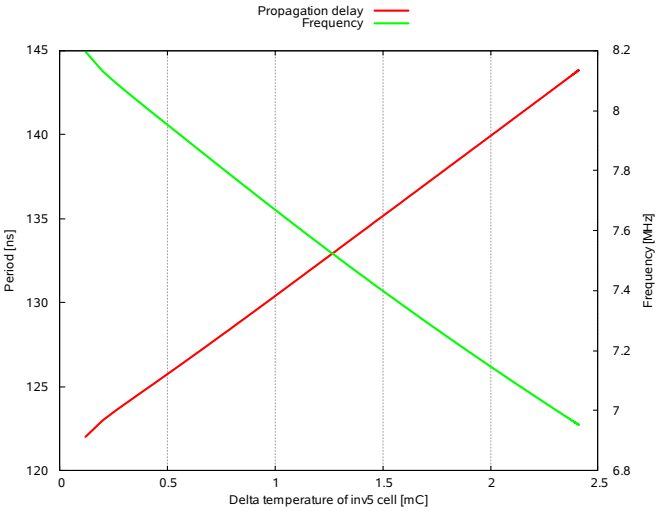


Fig. 8. Period and frequency of ring oscillator versus temperature of *inv5* cell

in the oscillation frequency as shown by Fig.7. The frequency of the oscillator drops from the initial 8.1967 MHz to 6.952 MHz in the 4th second.

As the thermal steady state is reached near the 4th second, the frequency also reaches a steady state at the 4th second. This phenomenon is clearly visible in Fig.7.

In Fig.8 the period and frequency function is depicted versus the temperature of the *inv5* cell.

IX. SUMMARY

In this paper a methodology for simulating temperature dependent propagation delays in a ring oscillator circuit is demonstrated. A special demonstration circuit has been created to spectacularly demonstrate the effect of circuit self-heating on propagation delays and operating frequency.

Temperature dependent delays are calculated from synthesizer-generated SDF files making thermal-aware logic simulations possible.

A grid partitioning method has been introduced where the created thermal model is independent of the number of standard cells in the circuit. This approach speeds up initial thermal model generation and does not scale with increasing standard cell count.

This paper also proposed a method of serialization that can speed up initial database loading for the logi-thermal simulation with CellTherm.

Finally, simulation results are shown to prove the concept introduced in this paper.

X. ACKNOWLEDGEMENT

This work was partly supported by the IP 248603 THERMINATOR FW7 project of the European Union and by the Hungarian Government through TÁMOP-4.2.1/B-09/1/KMR-2010-0002.

 ISBN:978-2-35500-022-5

REFERENCES

[1] A. Timar, G. Bognar, A. Poppe, and M. Rencz, "Electro-thermal co-simulation of ICs with runtime back-annotation capability," in *Thermal Investigations of ICs and Systems (THERMINIC), 2010 16th International Workshop on*, Barcelona, Spain, oct. 2010, pp. 1 –6.

[2] A. Timar, G. Bognar, and M. Rencz, "Improved power modeling in logi-thermal simulation," in *Thermal Investigations of ICs and Systems (THERMINIC), 2011 17th International Workshop on*, Paris, France, sept. 2011, pp. 1 –6.

[3] S. Pable and M. Hasan, "Ultra-low-power signaling challenges for subthreshold global interconnects," *Integration, the VLSI Journal*, vol. 45, no. 2, pp. 186 – 196, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167926011000770

[4] B. Rebaud, M. Belleville, E. Beigné, C. Bernard, M. Robert, P. Maurine, and N. Azemard, "Timing slack monitoring under process and environmental variations: Application to a DSP performance optimization," *Microelectronics Journal*, vol. 42, no. 5, pp. 718 – 732, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0026269211000292

[5] S. Lin, Y.-B. Kim, and F. Lombardi, "Design and analysis of a 32nm PVT tolerant CMOS SRAM cell for low leakage and high stability," *Integration, the VLSI Journal*, vol. 43, no. 2, pp. 176 – 187, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167926010000040

[6] R. Kumar and V. Kursun, "Reversed temperature-dependent propagation delay characteristics in nanometer CMOS circuits," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 53, no. 10, pp. 1078 – 1082, oct. 2006.

[7] C. Sánchez-Azqueta, S. Celma, and F. Aznar, "A 0.18 μm CMOS ring VCO for clock and data recovery applications," *Microelectronics Reliability*, vol. 51, no. 12, pp. 2351 – 2356, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0026271411001697

[8] J. Wang, N. Gong, L. Hou, X. Peng, R. Sridhar, and W. Wu, "Leakage current, active power, and delay analysis of dynamic dual vt cmos circuits under P–V–T fluctuations," *Microelectronics Reliability*, vol. 51, no. 9–11, pp. 1498 – 1502, 2011, proceedings of the 22th European Symposium on the RELIABILITY OF ELECTRON DEVICES, FAILURE PHYSICS AND ANALYSIS. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0026271411002010

[9] A. Winther, W. Liu, A. Nannarelli, and S. Vrudhula, "Temperature dependent wire delay estimation in floorplanning," in *NORCHIP, 2011*, nov. 2011, pp. 1 –4.

[10] A. Golda and A. Kos, "Temperature influence on power consumption and time delay," in *Digital System Design, 2003. Proceedings. Euromicro Symposium on*, sept. 2003, pp. 378 –382.

[11] (2012, 16th April, 10:38) Serialization. [Online]. Available: http://en.wikipedia.org/wiki/Serialization

 ISBN:978-2-35500-022-5

25-27 September 2012, Budapest, Hungary

Layout Constrained Body-Biasing for Thermal Clock-Skew Compensation

Valerio Tenace Sandeep Miryala Andrea Calimera
Alberto Macii Enrico Macii Massimo Poncino

Politecnico di Torino, 10129 Torino, ITALY

Abstract— **In this work we investigate on the actual capability of adaptive body bias to be applied on Clock Distribution Networks (CDNs) for dynamic compensation of thermally-induced skews. Selectively applying different bulk polarizations to the clock buffers, i.e., Forward Body Bias (FBB) to speed-up, or Reverse Body Bias (RBB) to slow down, it is possible to recover the timing phase shifts which accumulate along the clock tree paths due to on-chip thermal gradients. However, the physical implementation of such technique is not trivial: While a fine tuning of the clock buffers would require to apply a cell-by-cell biasing, design constraints due to semi-custom layout rules impose a higher granularity, that is, a physical cluster of cells, i.e., a layout row. We therefore propose a row-based ILP formulation that considers the aforementioned physical constraints and that provides optimal body bias assignment for thermal clock skew minimization. As will shown for a set of realistic benchmarks mapped into an industrial 40nm technology, our solution allows real skew compensation under different thermal profiles while avoiding unfeasible conditions where buffers placed in the same row require different bulk polarizations.**

I. Introduction

One of the main concerns during the implementation of modern digital ICs is the design of the Clock Distribution Network (CDN).

A reliable CDN guarantees the clock edges to reach flip-flops approximately at the same time such to minimize the *clock-skew*, namely, reduce race conditions between paths avoiding circuit de-synchronization and incorrect data latching. The special characteristics and attributes of the clock signals, however, do not facilitate this design phase [1]. First, as many thousands of flip-flops are placed on the same circuit, the clock signal is typically loaded with very high fanout. Second, since sequential elements are irregularly scattered across the layout, the clock wires can span the whole die traveling over the longest distances. Third, the clock waveforms, which propagate at the highest speed of any signals, need to be clean and sharp in order to guarantee reasonable noise margins.

While all these issues have been properly addressed with the introduction of automatic tools for the synthesis of regular CDN structures, like buffered *trees* and *H-trees* [1]–[3], the advent of nanometric technologies made the problem resurfacing again. On the one hand, the global interconnects, typically adopted for CDNs, become more highly resistive at each technology node; as side effect the number of buffers required to drive the clock signal increases substantially exacerbating the problem of power consumption. On the other hand, CDNs mapped with scaled technologies show unpredictable behaviors as both MOS transistors and metal wires are very sensitive to parametric variations [4]. Temperature, in particular, has shown to be one of the most critical source of variability when considering global interconnects and CDNs [5], [6], not just for the higher and higher peak temperatures devices must tolerate, but mostly because of the large spatial and temporal gradients which appear across the layout (more than 50°C in high performance ICs as reported in [4]).

It is well known that devices working at different temperatures may show significant performance mismatches. This induces different branches of the CDN to have unbalanced delays, that is, branches crossing hot regions get slower, while those crossing cold regions get faster. The resulting difference generates clock-skews [7], [8] which may vary, dynamically, depending on the workloads.

To compensate dynamic clock-skews, the authors of [9]–[11] pioneered a new design paradigm based on the use of *adaptive* devices, i.e., adjustable buffers whose delay can be tuned, at run-time, depending on the temperature distribution. Although different embodiments of tunable buffers have been proposed in the literature, from those which adopt supply voltage scaling, to those that make use of programmable capacitive loads to increase the fan-out [9], the idea of using the body bias as performance control knob turned out as one of the best solution. It does not require any modification of the standard gates, and most technology libraries offers support for it.

Body biasing works on the principle of applying a forward bias voltage (FBB) or a reverse bias (RBB) to decrease or increase gate delays. To compensate the skew variations, FBB can be applied to those buffers which belong to slow branches of the CDN, i.e., those which cross cold regions, while RBB to those buffers which belong to fast branches of the CDN, i.e., those which cross cold regions. The right sequence of RBB/FBB configurations brings back the CDN to have zero skew. Obviously, this requires the capability of independently applying different bulk polarizations to each individual buffer in the CDN. While this was not an issue for old technologies (above 90nm) where dedicated bulk contacts were available in each gate, for CMOS technologies below the 65nm single bulk contacts are no longer available as they were replaced with ad-hoc contact cells that must be uniformly placed through the layout, typically every 30 to 50 μm [12].Unfortunately

this prevents body-biasing to be applied cell-by-cell, thereby limiting the efficacy of most of the proposed strategies (e.g., [11]), which, instead, do not consider those rules.

In this work we propose a new ILP formulation that operates at the row level in order to find a proper clustering of buffers that can enable adaptive body bias for thermal clock skew compensation. The optimal selection returned through the ILP can be used as reference to fill the LUT of a thermal management unit that will drive the tuning during in-field operations [13]. Experimental results conducted on a selection of large-scale benchmarks mapped into an industrial 40nm technology and obtained from industry-strength physical design tools, show that the clock skew induced by realistic thermal profiles can be effectively reduced (6% on average) still maintaining compliance with physical rules.

The remainder of paper is organized as follows. Section II gives a brief summary of the state-of-the-art in thermal-aware clock synthesis. Section III provides the readers with basic thermal models, while Section IV and V describe the body-bias architecture and the proposed ILP. Finally, simulation and analysis results are discussed in Section VI. Section VII closes the paper with a brief summary of the work and final remarks.

II. Previous Works

Standard approaches for the design of the CDN historically focused on generating a clock network (typically a tree) with minimum wire-length, zero [14], [15] or bounded skew [16], [17], and possibly combined with wire sizing and/or with buffer insertion [18]. All these approaches assume a constant temperature along the clock network.

Clock tree optimization under spatial thermal gradients was first studied in [19], where the authors modified the traditional Deferred-Merge Embedding (DME) method proposed in [14] to search for nodes in the vicinity of merging points that can minimize clock skew for both uniform and non-uniform thermal profiles. Such kind of methods, however, rely on design-time optimizations for non-uniform, but static thermal profiles.

To manage time-varying gradients, a different design paradigm was presented by the authors of [10], [11], [13], who proposed a new post-silicon solution. The latter, which falls under the class of *adaptive strategies*, is based on the use of dynamically adjustable delay clock buffers that can be tuned to compensate, at the run-time, the clock-skew. The main differentiation factor among the proposed adaptive techniques is the type of knob used to regulate the delay of the buffers. Among all the possible solutions [20], *Adaptive Body Bias* (ABB) has been proven to be a key alternative. It does not require any modification of the standard gates, and most technology libraries offers support for it. ABB exploits the body effect of MOS transistors allowing step-wise modulation of the buffer delays.

A final class of solutions exploit the capability of some cells to be temperature insensible when powered under a specific supply voltage, the zero-temperature coefficient voltage [21]. In [22], for instance, the authors make use of programmable temperature compensation devices to shift the ZTC of the buffers around the nominal voltage, therefore obtaining temperature insensitive devices. The resulting buffers, however, show area penalties and high sensitivity to process variations.

III. Thermal Effects on Clock Distribution Networks

The clock skew is formally defined as the maximum difference between the (source to sink) arrival times $D_i \quad D_j$ of any pair of nodes (i, j) belonging to the set of sinks S of the tree, Equation1.

$$Skew = |max\{D_i \quad D_j\}|, \ \forall \ (i,j) \ \in S, \ with \ i \neq j \quad (1)$$

The arrival time at each sink is the result of the additive delay contributions of wire segments and buffers that break up the clock-path from the root. Both wires and buffers may vary their behavior significantly depending on the operating temperature. Concerning metal wires, a larger temperature gets a linear increase of the metal resistivity, as described by Equation 2:

$$R_0(x) = R_0(1 + \beta \cdot T(x)) \quad (2)$$

where, R_0 is the resistance at reference temperature (room temperature), β is the temperature coefficient that depends on the type of material ($3.9e^{-3}$ for Copper), while T represents the local temperature of the wire resulting from the diffusion of heat from the substrate and the self-heating effect [23]). Notice that, for long wires, as the case of clock trees, T may change depending on the actual position along the wire x.

Concerning MOS transistors, one of the main parameters affected by temperature is the carrier mobility μ, which, due to larger temperature induced lattice vibrations, decreases with increasing temperature[1]. The relation governing such dependence is shown in Equation 3, where T is the actual junction temperature and T_0 is the nominal temperature (about 300k); m is the temperature coefficient which is about 1.5 but may vary depending on the process.

$$\mu(T) = \mu(T_0) \cdot \left(\frac{T_0}{T}\right)^m \quad (3)$$

As described by the alpha power law model [24], Equation 4, lower mobility reflects in a smaller current capability of the active transistors, which, in turn, affects the speed of the CMOS buffers that get slower as temperature increases.

$$I_d \ \alpha \ \mu(T)(V_{dd} \quad V_{th})^\alpha \quad (4)$$

Even if the clock signal propagates on a symmetric network with both wires and buffers showing a monotonic temperature relationship, the presence of thermal gradients may cause substantial timing skew. Branches of the tree crossing hot regions get slower, while branches that run over cold regions get faster. FigureIII depicts a scenario where two paths, *Route-to-Sink1* and *Route-to-Sink2*, having same length, thus zero

[1] In this work we assume CMOS devices do not show Inverted Temperature Dependence [21]

clock skew, can result in different delays due to the thermal effects.

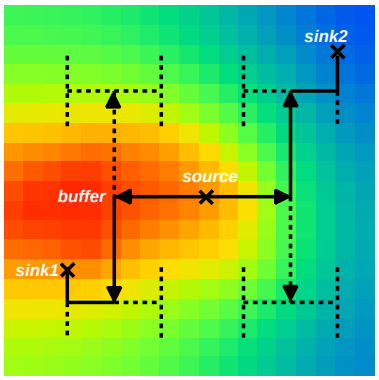


Fig. 1. Clock branches reaching different thermal regions develop unequal delays.

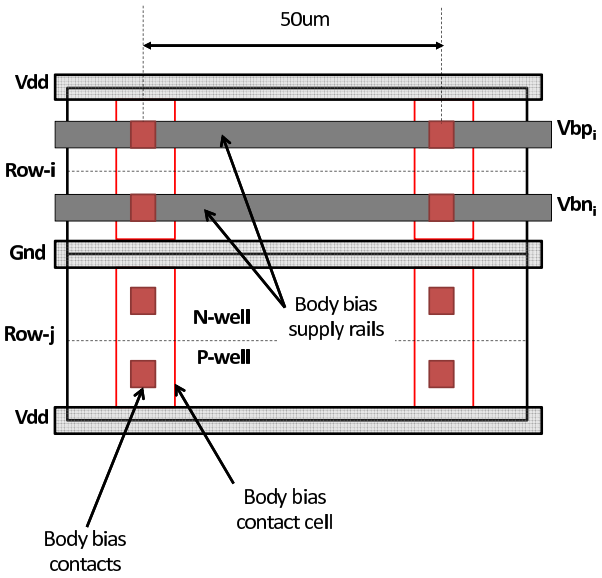


Fig. 2. Abstract view of row-based ABB layout [12]

IV. Adaptive CDNs

Recent works [10], [11], [13] have proposed the use of adaptive strategies to compensate, dynamically, the effects of temperature on the clock skew. All of them are based on the same idea, that is, the use of tunable clock buffers through which compensate delay variation along the timing arcs. However they can be classified in terms of i) control knob used to adjust the buffers delay, ii) granularity at which the control knob operates on the buffers, iii) implementation of the control structure.

In this work we use body-bias as main control knob and we introduce a layout constrained row-based scheme where the body bias is applied at the row-level, namely, each individual row of the layout has its own independent polarization voltage that is controlled dynamically depending on the thermal profile of the die.

a) **Control Knob:** body-bias can alter the threshold voltage of a device exploiting the body effect of MOS transistors. The body effect can be modeled by Equation 5, which describes the dependence of the threshold voltage V_{th} from the bulk voltage V_b:

$$V_{th} = V_{T0} + \gamma(\sqrt{|V_{sb} + 2\phi f|} \quad \sqrt{2\phi f}) \qquad (5)$$

with V_{T0} the threshold voltage under zero substrate bias, γ is the body effect parameter, V_{sb} is the voltage between the source and the bulk terminals, and $2\phi f$ is the surface potential. A Forward Body Bias (FBB), i.e., $V_{b_n} = +V_{FBB}$ for nMOS and $V_{b_p} = Vdd \quad V_{FBB}$ for pMOS, with $V_{FBB} > 0$, reduces the V_{th} of the transistors that can drain more current making the buffer faster. As side effect, FBB exponentially increases the sub-threshold leakage power [25]. On the contrary, a Reverse Body Bias RBB, i.e., $V_{b_n} = \quad V_{RBB}$ for nMOS and $V_{b_p} = Vdd + V_{RBB}$ for pMOS, with $V_{RBB} > 0$, increases the V_{th} getting slower buffers. A positive effect of RBB is a lower leakage current.

The selection of V_{FBB} and V_{RBB} is a technology dependent process. Depending on the internal characteristics of the transistors, like doping concentration of the source/drain-bulk pn junctions, RBB and FBB polarizations can be defined as to guarantee the best trade off between delay variation, power and reliability. For our technology, a low-power 40nm technology with nominal V_{dd} of 1.1V, we used symmetric values, i.e., $V_{FBB} = V_{RBB} = 0.\ V$.

b) **Granularity:** From a conceptual point of view, the body bias should be applied, separately and independently to each individual buffer of the clock tree. Unfortunately, physical design rules of modern CMOS technologies do not support cell-level body biasing, as the bulk contacts are no longer available in the cells. On the contrary, silicon vendors provide designers with special cells, the body bias contact cells, which are placed through the layout rows and driven by the body bias generator. The design rules require the body bias contact cells to be placed every 30 to 50μm (for the technology we have used). Therefore, the minimum grain at which ABB can be applied is a portion of the row, i.e., a sub-row. However, since multiple biases per row seriously complicates the routing of the body bias supply rails, in this work we consider the row as atomic unit, i.e., each row can be polarized with a dedicated voltage (V_{b_p} and V_{b_n}, as shown in Figure 2).

c) **Architecture:** Concerning the control structure, we borrowed the idea proposed in [13], where an embedded hardware mechanism, called the thermal management unit (TMU), is used to translate on chip thermal profiles (provided by dedicated temperature sensors) into the proper bias configuration that minimizes the skew. Figure 3 illustrates the proposed architecture. Data collected from the sensors are used to point a specific row in the TMU table; the one that contains the body-bias configurations for compensating the skew generated by the current thermal profile. Conceptually, rows containing

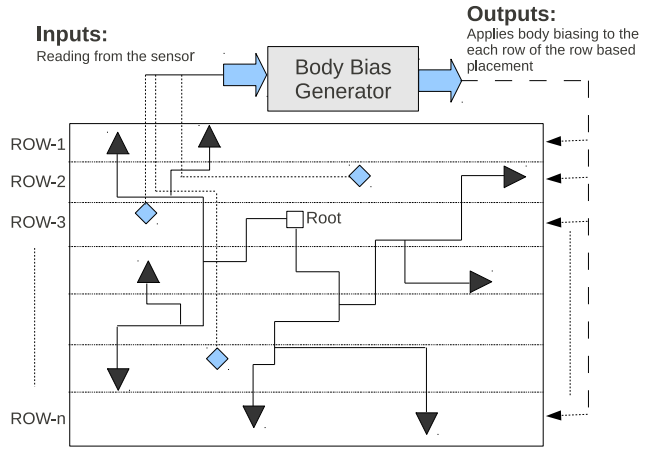


Fig. 3. Revisited TMU architecture [13] for skew compensation using row-based ABB

buffers that belong to slow paths are Forward Body Biased (FBB), while, rows that contain buffers of fast paths are Reversed Body Biased (RBB). To notice that rows may contain buffers belonging to both slow and fast paths concurrently. This impose constraints frustration during optimization.
The optimum body bias configuration, which is stored in the TMU table, is carried out by means of an off-line characterization step [13]. For each applied thermal profile, the optimal body bias selection (RBB or FBB) of each row is obtained by solving an ad-hoc ILP model (described in the next section).

V. The ILP Model

In this section we introduce the mathematical formulation of the problem we are trying to solve thorough an Integer Linear Programming (ILP). We want to find an optimal bias configuration of the layout rows such to compensate the clock skew under a given thermal profile.
Let assume that the gates in the design have been placed among N different layout rows. We assign two 0-1 integer variables to each row i, x_i and y_i, whose meaning is described in 6:

$$x_i = \begin{cases} 0 & \text{zero bias} \\ 1 & \text{if FBB} \end{cases} \qquad y_i = \begin{cases} 0 & \text{zero bias} \\ 1 & \text{if RBB} \end{cases} \tag{6}$$

The first main constraint concerns the clock skew, which has to be maintained below a user defined bound B_{skew}. This is described in Equation 7, where the difference between the propagation delay D of any pair of paths p_n and p_k belonging to the set of root-to-sink paths Π is forced to be lesser than B_{skew}.

$$0 \leq D_{\pi_k} \quad D_{\pi_n} \leq B_{skew} \qquad \forall \pi_k, \pi_n \in \Pi \mid k \neq n \tag{7}$$

The delay D_{π_i} of the i-th path π_i is given by the sum of the propagation delays of the clock buffers and wire segments in the path itself:

$$D_{\pi_i} = \sum_{j=0}^{J} d_j + d_{net,j} \qquad \forall j \in \pi \tag{8}$$

Where $d_{net,j}$ is the delay introduced by wire load in the fanout of the buffer, while d_j is the propagation delay of each buffer. As described in the following Equation, each d_j can be further expanded as a function of the body bias voltage applied to the row i that hosts buffer j:

$$d_j = d_0 \quad x_i \delta_{j,FBB} + y_i \delta_{j,RBB} \tag{9}$$

where d_0 is the nominal propagation delay of the cell under zero biasing, $\delta_{j,FBB}$ represents the delta delay of the buffer in case of FBB, and $\delta_{j,RBB}$ represents the delta delay in case of RBB. To calculate d_j, we make use of LUTs containing pre-characterized delay under different loads, transition time and operating temperatures.
Depending on the bias configuration associated to each row i, namely, the pair (x_i, y_i), the clock skew can be compensated. Obviously, each row needs an *unique* bulk polarization, FBB or RBB; This condition is forced by the following constraints.

$$x_i + y_i \leq 1 \qquad \forall i \in N \tag{10a}$$

The resulting ILP model is then reported in Equation 11, where the cost function to minimize, the Equation 11a, represents the total number of rows polarized at a voltage other than zero. Minimizing such metric we reduce the impact of body-biasing to the other gates in the circuit.

$$\text{min: } \sum_{i}^{N} (x_i + y_i) \tag{11a}$$

$$0 \leq D_{\pi_k} \quad D_{\pi_n} \leq B_{skew} \qquad \forall \pi_k, \pi_n \in \Pi \mid k \neq n \tag{11b}$$

$$x_i + y_i \leq 1 \qquad \forall i \in N \tag{11c}$$

$$x_i \in \{0, 1\} \qquad \forall i \in N \tag{11d}$$

$$y_i \in \{0, 1\} \qquad \forall i \in N \tag{11e}$$

VI. Experimental Results

A. Toolchain internals

The design framework we implemented to validate the proposed methodology relies on different commercial tools like Synopsys Design Compiler (DC), used for logic synthesis, Synopsys IC Compiler (ICC) used for the physical implementation (placement and clock tree routing), Synopsys HSPICE used for transistor level simulations. ILP problems, instead, have been solved using the open source tool Mixed Integer Programming Solver $p\_ o ve$ [26].
The entire flow consists of three main stages.

1) **Synthesis**: Starting from a RTL description, the design goes through standard logic/physical synthesis, which yields a gate-level netlist containing both the logic gates and the clock tree. Once those steps are accomplished an in-house tool [8] extracts a transistor-level netlist of the clock network which includes extracted view of

the buffers (taken from an industrial 40nm technology library) and parasitics back-annotated from the SPEF file[2].

2) **Body-Bias Assignment**: The SPICE netlist is simulated along with thermal map (assumed to be given) via HSPICE, in order to obtain path delays and skew information used for the construction of the ILP, as described in section V. The solver is then fed with the ILP model and run. The returned solution describes the bias configuration to be applied to each row.
3) **Validation**: The bias configurations returned from the ILP solver are finally applied to the original netlist and statically analyzed within ICC. For this stage we make use of timing libraries characterized under different bulk polarizations: FBB ($V_{FBB} = 0.\ V$, RBB ($V_{RBB} = 0.\ V$). Synopsys NCX has been used to collect the timing information.

B. Benchmarks and simulation results

Table I reports the main characteristics of the three circuits used as benchmarks.

TABLE I SYNTHESIZED BENCHMARKS

Design	# Rows	# Buffers	Clock Skew (ps)
switch_6x6	66	35	72.87
cf_fir_33	177	138	197.52
FPU4	305	182	224.74

To emulate different thermal behaviors, the chip die is divided in eight equally sized layout regions, clockwise labeled from one to eight, starting from the top left. We therefore assume the availability of a dedicated temperature sensor for each region. Table II shows the four thermal profiles we considered during simulations.

Table III reports, for each benchmark, the amount of clock skew compensation ($\Delta Skew$) we can achieve under different thermal maps. The comparison is made between the physical constrained ILP model presented in this paper (column *Row-Based*) and a standard ILP model (column *Cell-Based*) that does not take into account any physical constraints, that is, each buffer can be assigned, separately, to any body bias. Such cell-based formulation is very similar to the row-based one, but boolean variables representing ABB configuration, i.e., x_i and y_i (please refer to SectionV), are referred to each single buffer in the clock tree rather than rows.

As expected, the cell-based ILP, which has more degrees of freedom, shows a better skew compensation, 7.55% on average, w.r.t. the row-based ILP, 5.18% on average. Nevertheless, the cell-based approach returns a non feasible solution for most of the cases under test. The column *Feasibility* reports the number of rows that produce frustrated constraints, namely, rows containing buffers that should be assigned to a certain bulk polarization but cannot because of the presence of other buffers polarized at different body bias. Only for one case, the cell-based ILP returns a feasible solution, the *switch_6x6* under map # , for which 0 rows show constraint violations. On the contrary, the row-based approach is always feasible, as each row is assigned to a single body bias.

One additional comment concerns the potential overheads introduced by the row-based body biasing. Changing the bulk polarization of a row affects not just the clock buffers, but also the other cells placed in that row. This may alter other metrics of the circuit, as the leakage power (FBB reduces the V_{th} increasing the sub-threshold current) and the maximum operating frequency (RBB increases the V_{th} making logic gates slower). However the proposed ILP has been constructed in order to minimize such effects. This has been achieved by minimizing the number rows assigned to any FBB or RBB (maximum number of rows left to zero bias). Experimental results have shown that leakage overhead is always below 5%, for any benchmarks and for any thermal profiles; larger delay penalties, instead, have been registered, 10% on average. The reason is that the ILP does not consider that RBB on cells belonging to the critical paths directly affect the overall speed of the circuit. We are working on new formulations and heuristics that consider timing penalties during optimization.

[2]SPEF: Standard Parasitic Exchange Format.

VII. CONCLUSIONS

In this work we illustrated the basic idea behind the construction of an ILP model and the implementation of a in-house framework for dynamically compensating the clock skew variation due to thermal gradients. It turned out that our ILP-based approach is an effective option to recover clock skew degradation. Results show that physical constrained compensation of the clock skew is possible, however it will come at expense of worse slack timing among the circuit's critical paths. A possible improvement of this work could be to consider slack constraints in the ILP model, thus limiting the degradation of the slack. Nevertheless, including timing constraints may lead to less effective skew compensations or, in the worst cases, it could lead to an unfeasible solution to the problem.

REFERENCES

[1] E. Friedman, "Clock distribution networks in synchronous digital integrated circuits," *Proceedings of the IEEE*, vol. 89, no. 5, pp. 665–692, may 2001.
[2] P. Ramanathan, A. Dupont, and K. G. Shin, "Clock distribution in general vlsi circuits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 5, pp. 395–404, may 1994.
[3] N. H. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*. Pearson Education, Inc, 2005.
[4] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. of Design Automation Conference*, June 2003, pp. 338–342.
[5] K. Banerjee and A. Mehrotra, "Global (interconnect) warming," *Circuits and Devices Magazine, IEEE*, vol. 17, no. 5, pp. 16–32, sep 2001.
[6] A. H. Ajami, K. Banerjee, and M. Pedram, "Modeling and analysis of nonuniform substrate temperature effects on global ulsi interconnects," *IEEE Transactions on CAD*, vol. 24, no. 6, pp. 849 – 861, jun 2005.
[7] S. Bota, J. Rossello, C. de Benito, A. Keshavarzi, and J. Segura, "Impact of thermal gradients on clock skew and testing," *Design Test of Computers, IEEE*, vol. 23, no. 5, pp. 414 –424, may 2006.

TABLE II CHIP DIE THERMAL REGION TEMPERATURES (IN °C)

	Reg #1	Reg #2	Reg #3	Reg #4	Reg #5	Reg #6	Reg #7	Reg #8
Map #1	100	75	25	75	100	25	100	50
Map #2	50	100	75	25	25	50	25	100
Map #3	75	50	75	50	100	25	50	25
Map #4	25	100	50	25	75	50	25	100

TABLE III CLOCK SKEW COMPENSATION

			Cell-Based		Row-Based	
	# Rows	Thermal Map	Δ Skew	Feasibility	Δ Skew	(FBB, RBB)
cf_fir_33	66	Map #1	8.84%	14	6.39%	(17, 6)
		Map #2	7.68%	6	4.36%	(13, 4)
		Map #3	7.72%	6	5.33%	(12, 4)
		Map #4	8.67%	12	6.82%	(16, 7)
switch_6x6	177	Map #1	11.74%	2	7.98%	(2, 2)
		Map #2	11.54%	2	4.28%	(4, 1)
		Map #3	8.63%	2	7.41%	(5, 1)
		Map #4	5.54%	0	4.34%	(4, 1)
FPU4	305	Map #1	5.70%	26	4.28%	(41, 6)
		Map #2	3.59%	28	2.27%	(32, 7)
		Map #3	4.69%	28	3.75%	(28, 4)
		Map #4	6.25%	31	4.97%	(39, 7)
		avg	*7.55%*		*5.18%*	

[8] A. Sassone, W. Liu, A. Calimera, A. Macii, E. Macii, and M. Poncino, "Modeling of thermally incuduced skew variations in clock distribution network," in *Therminic'11:Thermal Investigations of ICs and Systems*, Sep. 2011, pp. 1 – 6.

[9] A. Chakraborthy, K. Duraisami, A. Sathanur, P. Sithambaram, E. M. L. Benini, and M. Poncino, "Modeling of thermally incuduced skew variations in clock distribution network," in *ISLPED'06:Proc. of International Synposium on Low Power Electronics and Design*, oct 2006, pp. 162–167.

[10] J. Long, J. C. Ku, S. Memik, and Y. Ismail, "Sacta: A self-adjusting clock tree architecture for adapting to thermal-induced delay variation," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 18, no. 9, pp. 1323 –1336, sept. 2010.

[11] T. Ragheb, A. Ricketts, M. Mondal, S. Kirolos, G. Links, V. Narayanan, and Y. Massoud, "Design of thermally robust clock trees using dynamically adaptive clock buffers," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 56, no. 2, pp. 374 –383, feb 2009.

[12] A. Sathanur, A. Pullini, L. Benini, G. D. Micheli, and E. Macii, "Modeling of thermally incuduced skew variations in clock distribution network," in *DATE'09: Proceedings of the conf. on Design, Automation and Test in europe*, Mar. 2009, pp. 162 – 167.

[13] A. Chakraborty, K. Duraisami, A. Sathanur, P. Sithambaram, L. Benini, A. Macii, E. Macii, and M. Poncino, "Dynamic thermal clock skew compensation using tunable delay buffers," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 16, no. 6, pp. 639 –649, jun 2008.

[14] T.-H. Chao, Y.-C. Hsu, J.-M. Ho, and A. Kahng, "Zero skew clock routing with minimum wirelength," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 39, no. 11, pp. 799 –814, Nov. 1992.

[15] R.-S. Tsay, "An exact zero-skew clock routing algorithm," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 242 –249, feb 1993.

[16] J. Cong and C.-K. Koh, "Minimum-cost bounded-skew clock routing," in *Circuits and Systems, 1995. ISCAS '95., 1995 IEEE International Symposium on*, vol. 1, apr-3 may 1995, pp. 215 –218 vol.1.

[17] D. J. H. Huang, A. B. Kahng, and T. C.-H.A., "On the bounded-skew clock and steiner routing problems," in *Design Automation, 1995. DAC '95. 32nd Conference on*, 1995, pp. 508 –513.

[18] Y. Chen and D. Wong, "An algorithm for zero-skew clock tree routing with buffer insertion," in *European Design and Test Conference, 1996. ED TC 96. Proceedings*, mar 1996, pp. 230 –236.

[19] M. Cho, S. Ahmedtt, and D. Pan, "Taco: temperature aware clock-tree optimization," in *Proc. of International Conference on Computer-Aided Design*, Nov. 2005, pp. 582–587.

[20] S. Tawfik and V. Kursun, "Dual supply voltages and dual clock frequencies for lower clock power and suppressed temperature-gradient-induced clock skew," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 18, no. 3, pp. 347 –355, march 2010.

[21] A. Calimera, R. Bahar, E. Macii, and M. Poncino, "Temperature-insensitive dual- synthesis for nanometer cmos technologies under inverse temperature dependence," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 18, no. 11, pp. 1608 –1620, nov. 2010.

[22] D. Wolpert and P. Ampadu, "Exploiting programmable temperature compensation devices to manage temperature-induced delay uncertainty," *IEEE Transactions on Circuits and Systems*, vol. 59, no. 4, pp. 735 – 748, Apr. 2012.

[23] A. Ajami, K. Banerjee, and M. Pedram, "Modeling and analysis of nonuniform substrate temperature effects on global ulsi interconnects," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 24, no. 6, pp. 849 – 861, june 2005.

[24] T. Sakurai and A. Newton, "Alpha-power law modfet model and its applications to cmos inverter delay and other formulas," *Journal of Solid State Circuits*, vol. 25, no. 2, pp. 584 – 594, Apr. 1990.

[25] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 305 – 327, feb 2003.

[26] *Introduction to lpsolve*, 2012. [Online]. Available: http://lpsolve.sourceforge.net/5.5/

25-27 September 2012, Budapest, Hungary

A quasi-SPICE electro-thermal simulator

Albin Szalai†, Zoltán Czirkos†, Vladimír Székely*†
†Budapest University of Technology and Economics
*Mentor Graphics Hungary Ltd. Mechanical Analysis Division
szalai|czirkos@eet.bme.hu|vladimir_szekely@mentor.com

***Abstract*—Integrated circuits exhibit coupled electro-thermal phenomena. The circuit elements dissipate heat, which propagates in the chip, rising the temperature of all circuit elements. Since the operation of the latter is temperature dependent, their dissipation will also change. This change results in the variation of the element voltages and currents, causing further changes in the dissipation, etc. Therefore we are facing a pair of couplings: an electronic → thermal one and a thermal → electronic one, meaning that all integrated circuits constitute essentially a mutually coupled electronic and thermal system.**

In this paper a detailed discussion of the electro-thermal phenomena is presented regarding the practical implementation. The final implementation is a SPICE compatible electro-thermal solver (qSPICE), which implements some of the major industry standard models with extended coupled thermal capabilities. The accuracy of qSPICE is demonstrated on multiple examples.

I. Introduction

The severe thermal problems of today's VLSI integrated circuits highlighted the necessity of thermal and electro-thermal simulations. Although solutions for electro-thermal simulation of monolithic integrated circuits appeared in the early 70's from various independent research teams [1]–[4], the same problem was rediscovered in the 90's and other solutions have also been published [5]–[8].

The importance of electro-thermal simulation can be easily demonstrated with the example of a simple op-amp (Fig. 1) – which is an ideal benchmark circuit for electro-thermal circuit simulators because it is highly sensitive to thermal offset. The relatively high dissipation of the output stage (T14, T20) warms up the input stage (T1, T2) which moves the operating points and thus changes the transfer characteristics. This effect strongly depends on the layout of the circuit. Despite of the fact that this topic is at the focal point of research today, there is still no wide-spread industry standard solution. The direction of the investigations is two-fold: modelling of the thermal system and developing a solver for the problem. The general, well-known procedure to handle a thermal system is to apply some kind of finite element method (FEM). Another solution was introduced in the 90's which aims to create an electric equivalent subcircuit that represents the thermal behaviour [9]–[11]. Besides the need to create an accurate model of the thermal system, the solver engine also needs to be prepared to be able to handle the two different types of models together.

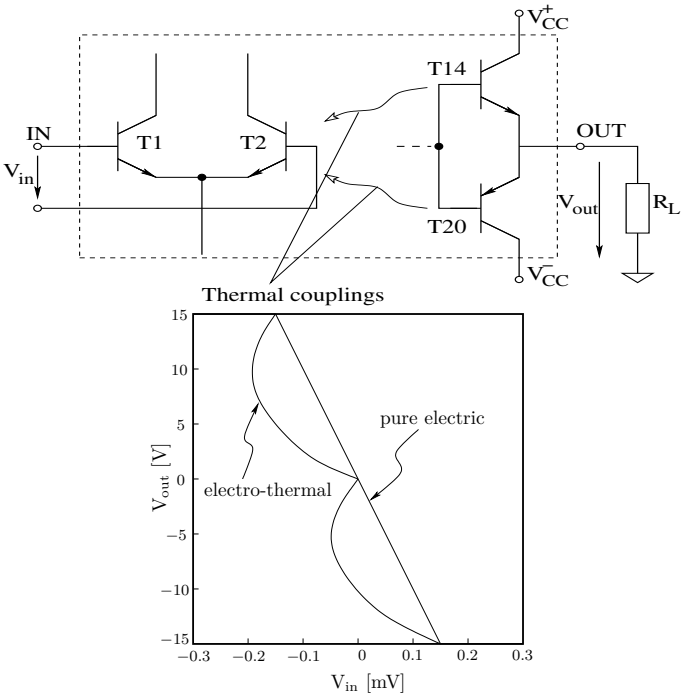


Fig. 1. Thermal feedback in an op-amp – a typical situation, where there is a significant difference in the results between "electrical-only" and electro-thermal simulations

There are two basic approaches of developing an electro-thermal solver: either the thermal problem is mapped into an electrical equivalent and an electric solver performs a co-simulation of both the electric and thermal subsystems co-existing inherently in an IC chip [2], [12]–[14], or two independent simulators, a thermal simulator and a circuit simulator iterate in order to deliver the solution of an electro-thermal problem [15]–[18]. Both methods have advantages and disadvantages. However, we are of the opinion that the first method (co-simulation by the simultaneous solution of the electrical and thermal problem) is superior, since it is applicable to a wider range of electro-thermal problems.

None of the earlier realizations of electro-thermal simulators became wide-spread, for various reasons. One aspect is the question of integration in EDA-flows (either in an IC-flow or in a PCB-flow – if discrete components on

PCB substrate have strong temperature dependence like individual LEDs in an MCPCB assembled RGB module [19]) – most of the published realizations lack this. The second aspect concerns the set of model equations used to describe the semiconductor devices: on one hand these should be "standard" models like the BSIM models for MOS transistors, on the other hand these models should be capable of describing the (junction) temperature dependence of the device operation as well as the device's self-heating in consistence with the electrical operation. The third aspect is the accurate and efficient modelling of the "chip+package+ambient" as a thermal system. This paper provides details of these three aspects and the problems we faced while we tried to address the above issues.

The work in our paper is dedicated to transistor level thermal simulation studies, where the electrical netlist of the circuit is combined with a SPICE-like RC model of the thermal environment. The importance of this approach can be demonstrated by Fig. 1, which shows the open loop transfer characteristics of an op-amp simulated with a pure electric simulator and an electro-thermal one. The latter is able to explore the thermally induced effects on the electric properties. In this particular case the large dissipation of the transistors in the output stage heats up the input stage and causes a shift in their characteristics. This effect strongly depends on the physical layout. The thermal environment is defined by the chip (or stack of chips if 3D SiP packaging solutions are treated) and its close environment (e.g. bonding wires/bumps, die attach and to some extent the lead frame). This simulation requires equation-level implementation of the thermal extension of semiconductor device models in the circuit simulator being used. The device models extended in this way are also complemented with an auxiliary thermal node and a thermal branch – the current of which is the power dissipated by the active semiconductor component. The thermal node is terminated by the thermal RC model extracted from the detailed physical model of the chip environment by a tool developed at BME. In this model the electrical resistors and capacitors represent thermal resistivity and thermal capacitance. The thermal RC model and the electrical netlist are solved simultaneously by the circuit simulator, which is equipped with device models having electro-thermal extension at equation level. This special methodology enables SPICE-like electro-thermal simulation. Our paper concentrates on chip level analog problems, but the thermal modelling approach (the Laytherm program) used in our electro-thermal simulation system is also used in one of our logi-thermal solutions (CellTherm [20]). From our perspective a PCB as a substrate can be modelled in the same way as an IC chip as a substrate, therefore our electro-thermal simulator is applicable to PCB substrates as well, populated with discrete analog semiconductor devices. The difference is that the compact dynamic thermal models of the semiconductor devices are to be included into the thermal model. These package models have to be inserted between the thermal node of the semiconductor device model (junction node) and between the thermal node of the thermal model of the substrate corresponding to the footprint of the discrete component on the PCB [21]. With this difference in mind, the paper in the following does not distinguish the IC and PCB design flows; it will concentrate on the IC design flow only.

This paper describes the following:

- principles of the analog simulation flow,
- state of our prototype of electro-thermal simulator (qSPICE),
- test cases,
- outlook on further development.

II. Principles of the analog simulation flow

There are two methods for electro-thermal simulation. One of the methods is to perform the transistor level simulation by SPICE, ELDO, SABER, or similar circuit simulators and use a thermal simulator that is usually a FEM program (e.g. ANSYS) connected in an iteration loop (Fig. 2). One simulator uses the updated results of the other one in the iterative process. The method is called the relaxation method (Fig. 2). The advantage in this

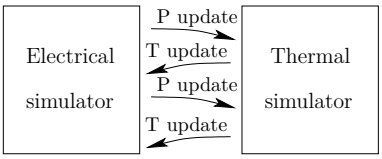


Fig. 2. Simplified scheme of the relaxation method

case is the relative simplicity of the implementation. The drawback is that very fast changes cannot be considered, and in case of strongly coupled thermal problems the simulator coupling frequently cannot find convergence. Frequency domain calculation is usually not possible.

The other possibility for electro-thermal simulation is the direct method or simultaneous iteration (Fig. 3). In these programs the thermal system is represented by an electrical model network, that has common nodes with the electrical-only network, the so-called thermal nodes. The iterative solution takes place simultaneously for the thermal and electrical sub-networks. The advantage in this case is the capability of considering very fast changes as well. The drawback is that it requires a more complex implementation than the relaxation method. A further benefit of the direct method is that it makes AC simulation possible.

The direct method needs the compact thermal RC model of the chip, the package and the ambiance as seen from the dissipating/temperature sensitive layout features. These compact models are calculated in our case by BME's Laytherm software using a layout description as the input. Our goal is to realize electro-thermal simulation using the

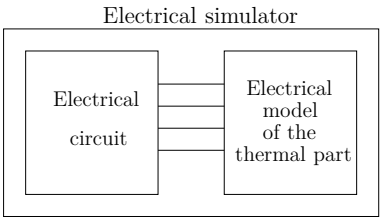


Fig. 3. Simplified scheme of the electro-thermal simulation using the direct method. The electrical netlist is completed with thermal nodes at the thermally relevant devices, which nodes are terminated by the model of the thermal system.

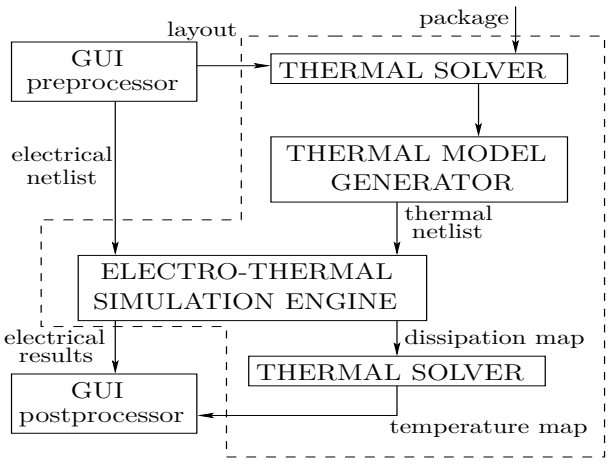


Fig. 4. The simulation flow of BME's solution to the electro-thermal simulation problem: components within the dashed line are part of the system.

direct method as it provides more accurate results. The overall structure of such a simulator is shown in Fig. 4. The electro-thermal simulation engine together with the thermal solver and the thermal model generator are at the center of our attention, we are not concerned with graphical user interface (GUI).

For compact modeling of the chip, the package and the ambiance the chip surface layout is taken (e.g. LEF/DEF or custom format) and loaded by Laytherm. The chip, die attach and leadframe layer structure is defined in Laytherm. Laytherm performs the tasks of the THERMAL SOLVER and the THERMAL MODEL GENERATOR blocks of the diagram in Fig. 4: its input is a layout description, based on which it creates RC equivalent circuits. Layout shapes can reside on any layer interface, which allows modeling stacked die structures. Package model beyond the leadframe (or die attach) is added to the chip's detailed model as a dynamic compact thermal model (Fig. 5). If there is heat-flow at the top, a similar model can be added. Such models can be based on the so-called structure functions provided either by physical measurements using Mentor Graphics' T3Ster (thermal transient tester) system [22], or on structure functions calculated from simulated thermal transient curves provided by a

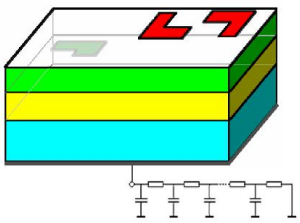


Fig. 5. Simplified view of the thermal model of the physical structure of the realized circuit.

numerical simulator (e.g. FloTHERM, ANSYS, etc.).

The ultimate model treated by an electro-thermal simulator using the direct method is shown in Fig. 6.

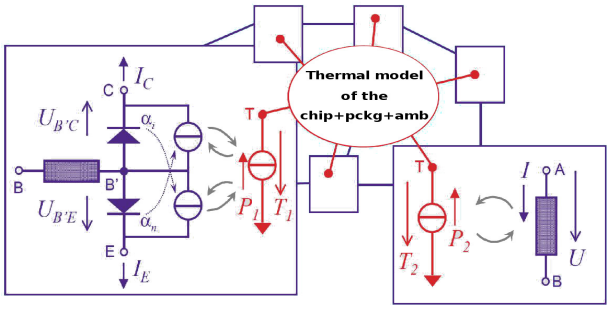


Fig. 6. The simplified model of an electro-thermal circuit simulation problem: gray arrows represent the electro-thermal transconductances, red dots represent the thermal nodes of the electro-thermal device models.

III. Requirements towards an electro-thermal solver

These requirements can be grouped as follows.

A. "Must have" requirements

The junction temperature of the semiconductor devices in a netlist must be maintained individually for every instance of the same model contained in the circuit description. Therefore:

- The device models need to be complemented with a thermal node which is the interface towards the compact thermal model.
- The model needs to drive its thermal node with a current source that corresponds to its dissipation.
- The temperature dependent quantities in the electrical model equations have to be taken into account as well, they need to be updated (recalculated) during the simulation whenever the device temperature changes.

B. "Nice to have" requirements

- If the simulator conforms to industry standard interfaces, it is much easier to integrate into existing design flows. For this purpose it is necessary to:
  - accept SPICE netlist format as an input,

ISBN:978-2-35500-022-5

– use the SPICE model card format,
– implement the SPICE model equations with their parameter names,
– accept standard layout formats (such as GDSII, OASIS, LEF/DEF).

IV. Current status of qSPICE

The development has started on two tracks.

One of the approaches was to extend an industry standard simulator's abilities to enable it to perform electro-thermal simulation. This way the "nice to have" requirements are satisfied by default and only the "must have"-s need to be implemented.

The advantage of an industry standard solver is that it implements all the major widespread device models. Hence our primary approach was to use these already implemented models for electro-thermal simulation. Despite the fact that all the industry standard SPICE models describe the temperature dependence correctly, today's SPICE simulators implement these models as "constant temperature at simulation-time". In order to achieve the goals of the task, one needs to modify the simulator at a very deep level at which only its developers can operate. We had close consultations with EDA vendors, but it turned out that the only way to deal with this problem would have been to reimplement these models correctly, and in the case of some complex models, even the solver would have been to be modified as well.

The other approach was to use BME's proprietary electro-thermal solver. This engine is equipped with the toolset that is needed for electro-thermal simulation so it fulfils the "must have" requirements. The task here is to realize the standard interfaces and to implement the standard models.

At this stage our solver can deal with all the major problems introduced above. We have verified it with measurements. Our target circuit was the well-known μA741 operational amplifier (Fig. 7). With this the classical benchmark problem illustrated by Fig. 1 can be well studied: what is the effect of the thermal feedback from the output stage to the two transistors of the differential pair of the input stage, and how does this effect depend on the actual physical realization (layout, packaging style). We studied two layout variants (from two different vendors) as illustrated in Fig. 8. The transistors marked with yellow circles are simulated using their electro-thermal model. We were able to acquire this integrated circuit from different vendors with two different layouts, which gave us the ability to verify our electro-thermal solver at different thermal scenarios. The simulated open-loop transfer characteristics can be seen with different loads (Fig. 9). These results demonstrate that the structure of the physical layout (see Fig. 8) has a major effect on the electrical behaviour through thermal coupling. Fig. 10 shows the AC simulation of the same circuit in a feedback

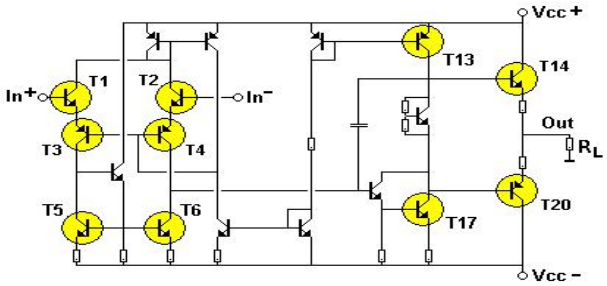


Fig. 7. Schematic of the μA741 operational amplifier.

configuration together with measurement results. We have multiple information on these figures.

- The simulation and measurement results have an excellent matching. This verifies our electro-thermal simulator.
- It is an interesting fact, that in certain feedback configurations, at the lower frequency range we found frequency values where one of the samples behaved as ideal inductor while the other as ideal capacitor. The only difference between the two was their physical layout.

Another subtask was to implement the nice to have features mentioned above. We created an input/output frontend/backend for the SPICE netlist compatibility purpose. With this converter our simulator can act as a SPICE compatible simulator, which is now easy to integrate with any industry standard design environment.

A. List of the implemented models

Based on this method we implemented the following models:

- SPICE MOS models (with Meyer capacitance models)
  – Shichman-Hodges
  – analytical
  – semi-empirical
- EKV (Enz-Krummenacher-Vittoz) MOS model
- Gummel-Poon BJT model
- Si/Al contact model (Seebeck effect)

For the SPICE compatibility layer and the thermal compact model generation we created a framework called qSPICE (quasi-SPICE) that is transparent to the user. We can carry out pure electrical simulations using the SPICE netlist format, and we do not have to alter the netlist for electro-thermal simulations either. If the layout is available the framework extracts the thermal compact model using Laytherm.

B. Verification

The engine and the implemented models has been verified both in pure electrical and electro-thermal domain. In the pure electrical domain we carried out the same simulations with our qSPICE and Berkeley's SPICE3f5.

ISBN:978-2-35500-022-5

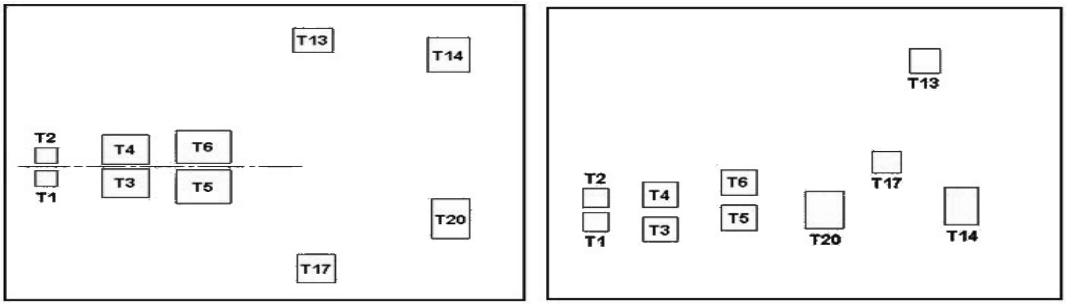


Fig. 8. The two different layouts for the μA741

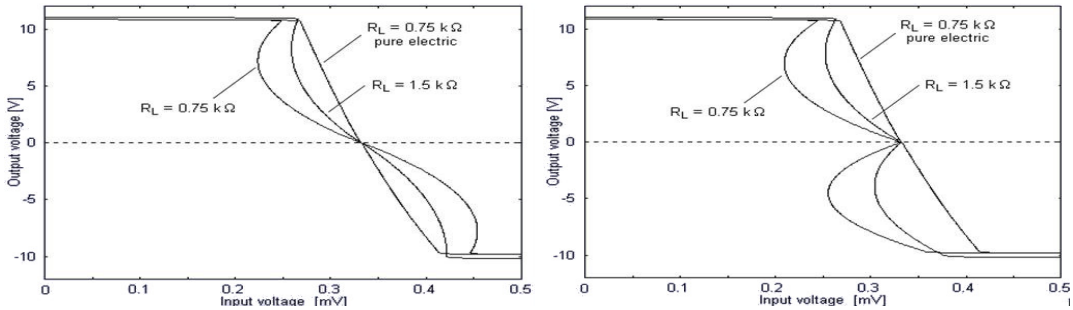


Fig. 9. Transfer characteristics with different loads

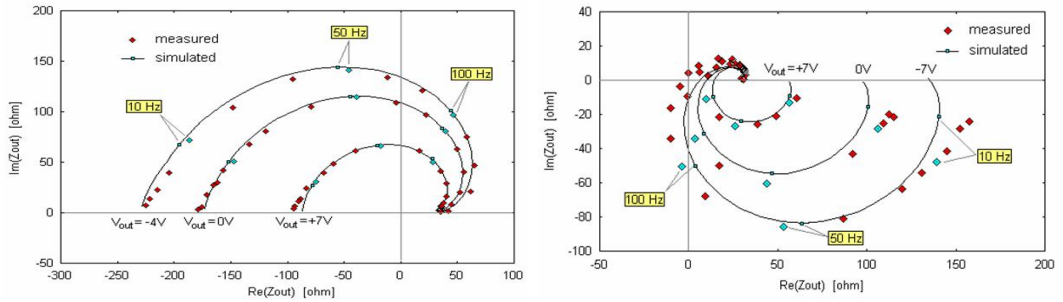


Fig. 10. Behaviour in the frequency domain with different loads

We performed the simulations with multiple benchmark circuits with various device models.

The matching error between the results was below the level of the numerical noise. In the electro-thermal domain the performed simulations were verified with published benchmark results. [23]

C. Example

In our basic example (diff. pair with output stage) we created a pure electrical simulation and an electro-thermal simulation where our input transistors and our output transistor have a compact thermal model.

On Fig. 12 the different simulation results for pure electrical and electro-thermal simulations can be seen. Fig. 12 (B) shows the relative temperatures of the important transistors for the nominal temperature (300K).

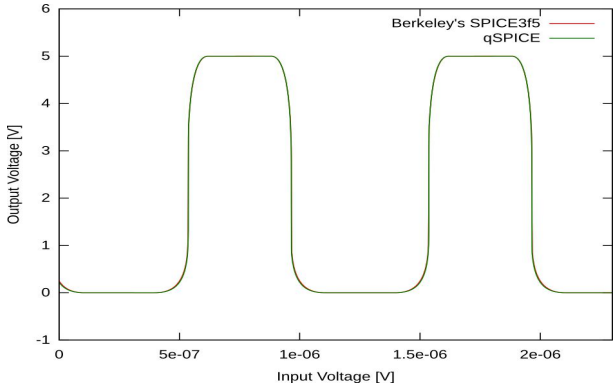


Fig. 11. Transient analysis of a CMOS inverter stage.

V. Conclusion

The objective of this work was to create an electro-thermal simulator that complies with industry standards

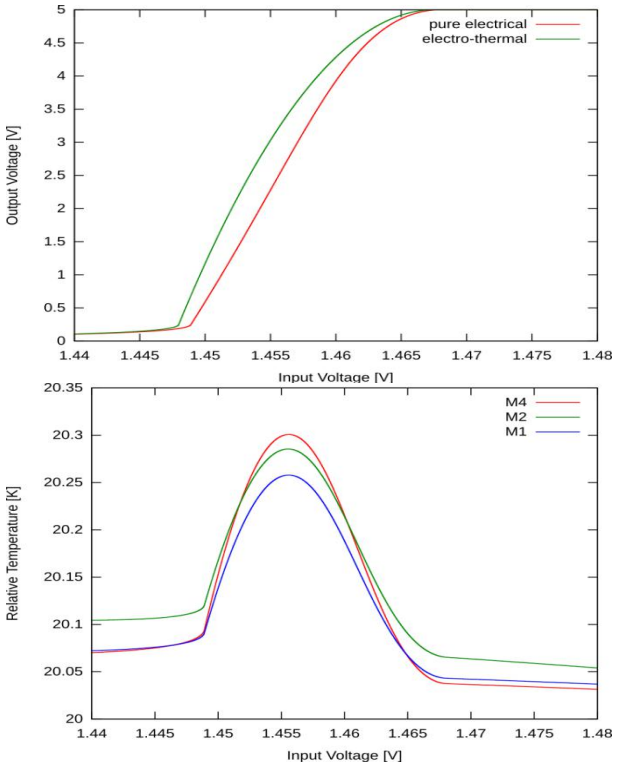


Fig. 12. (A) transfer function of a simple differential pair, (B) relative temperatures of transistors.

and is able to deal with 3D SiP packaging solutions as well.

Having found that the extension of an industrial solver is not a viable option, we decided to concentrate our efforts on developing a new electro-thermal solver and extending it with industry standard models and interfaces (the "nice to have" features mentioned above).

Our simulator, qSPICE satisfies the original objectives of the task as it is capable of determining the thermal coupled electric behaviour of complex systems. The input formats and device models are standard implementations.

References

[1] V. Székely and K. Tarnay, "Accurate algorithm for temperature calculation of devices in nonlinear-circuit-analysis programs," *Electronics Letters*, vol. 8, no. 19, pp. 470–472, 1972.

[2] V. Székely, "Accurate calculation of device heat dynamics: a special feature of the TRANS-TRAN circuit-analysis program," *Electronics Letters*, vol. 9, no. 6, pp. 132–134, 1973.

[3] K. Németh, "On the Analysis of Nonlinear Resistive Networks Considering the Effect of Temperature," *IEEE Journal of Solid-State Circuits*, vol. 11, no. 4, pp. 550–552, 1976.

[4] K. Fukahori and P. R. Gray, "Computer Simulation of Integrated Circuits in the Presence of Electrothermal Interaction," *IEEE Journal of Solid-State Circuits*, vol. SC-11, no. 6, pp. 834–846, 1976.

[5] S. Lee and D. J. Allstot, "Electrothermal Simulation of Integrated Circuits," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 12, pp. 1283–1293, 1993.

[6] W. Petegem, B. Geeraerts, W. Sansen, and B. Graindourze, "Electro-thermal simulation and design of integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 2, no. 29, p. 143, 1994.

[7] W. Kao and W. Chu, "ATLAS: an integrated thermal layout and simulation system for ICs," in *Proc. European Design & Test Conference '94*, 1994, pp. 43–48.

[8] T. Veijola and M. Andersson, "Combined Electrical and Thermal Parameter Extraction for Transistor Model," in *Proceedings of ECCTD'97*, Budapest, 1997, pp. 754–759.

[9] V. Székely and A. Poppe, "Novel tools for thermal and electrical analysis of circuits," *Electrosoft*, vol. 1, no. 4, pp. 234–252, 1990.

[10] C. J. Lasance, H. Rosten, and J. Parry, "The world of thermal characterization according to DELPHI-Part II: Experimental and numerical methods," *IEEE Transactions on Components and Packaging Technologies*, vol. 20, no. 4, pp. 392–398, 1997.

[11] C. J. Lasance, D. Den Hertog, and P. Stehouwer, "Creation and evaluation of compact models for thermal characterisation using dedicated optimisation software," in *Fifteenth Annual IEEE Semiconductor Thermal Measurement and Management Symposium*. Ieee, 1999, pp. 189–200.

[12] V. Székely, A. Poppe, M. Rencz, and A. Csendes, "Electro-thermal simulation: a realization by simultaneous iteration," *Microelectronics journal*, vol. 28, pp. 247–262, 1997.

[13] M.-N. Sabry, A. Bontemps, V. Aubert, and R. Vahrmann, "Realistic and efficient simulation of electro-thermal effects in VLSI circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 5, no. 3, pp. 283–289, 1997.

[14] V. Székely, A. Poppe, M. Rencz, A. Csendes, and A. Páhi, "Self-consistent electro-thermal simulation: Fundamentals and practice," in *Proceedings of 1st THERMINIC Workshop*, 1995, pp. 188–194.

[15] K. Petrosjanc, N. Ryabov, I. Kharitonov, and P. Kozynko, "Electro-thermal simulation: a new Subsystem in Mentor Graphics IC Design Flow," in *Thermal Inveatigation of ICs and Systems*, no. October, Leuven, Belgium, 2009, pp. 70–74.

[16] Y.-K. Cheng, E. Rosenbaum, and S.-M. Kang, "ETS-A: a new electrothermal simulator for CMOS VLSI circuits," in *Proceedings ED&TC European Design and Test Conference*. Paris, France: IEEE Comput. Soc. Press, 1996, pp. 566–570.

[17] T. Li, C.-H. Tsai, and S.-M. Kang, "Efficient transient electrothermal simulation of CMOS VLSI circuits under electrical overstress," in *1998 IEEE/ACM International Conference on Computer-Aided Design*. San Jose: Acm, 1998, pp. 6–11.

[18] S. Wünsche, C. Clauß, P. Schwarz, and F. Winkler, "Electro-thermal circuit simulation using simulator coupling," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 5, no. 3, pp. 277–282, 1997.

[19] A. Poppe, G. Farkas, V. Székely, G. Horváth, and M. Rencz, "Multi-domain Simulation and Measurement of Power LED-s and Power LED Assemblies," in *IEEE SEMI-THERM Symposium*, Dallas, USA, 2006, pp. 191–198.

[20] A. Timar and M. Rencz, "Real-time heating and power characterization of cells in standard cell designs," *Microelectronics Journal*, pp. 1–9, Apr. 2012.

[21] M. Rencz, V. Székely, and A. Poppe, "A Methodology for the Co-Simulation of Dynamic Compact Models of Packages With the Detailed Models of Boards," *IEEE Transactions on Components and Packaging Technologies*, vol. 30, no. 3, pp. 367–374, Sep. 2007.

[22] A. Vass-Várnai, R. Bornoff, S. Ress, Z. Sárkány, S. Hódossy, and M. Rencz, "Accurate Thermal Characterization of Power Semiconductor Packages by Thermal Simulation and Measurements," in *Proceedings of 12th Symposium on Design, Test, Integration & Packaging of MEMS/MOEMS*, no. May, Aix-en-Provence, France, 2011, pp. 324–329.

[23] V. Székely, A. Páhi, A. Poppe, and M. Rencz, "Electro-thermal simulation with the SISSI package," *Analog Integrated Circuits and Signal Processing*, vol. 21, pp. 21–31, 1999.

25-27 September 2012, Budapest, Hungary

Yield enhancement by logi-thermal simulation based testing

Gergely Nagy, László Pohl, András Timár, András Poppe
Budapest University of Technology and Economics
Department of Electron Devices
Budapest, Hungary 1111
Email: nagyg | pohl | timar | poppe@eet.bme.hu

***Abstract*—This paper proposes a method for yield enhancement in digital integrated circuit manufacture using a temperature dependent logic simulation tool.**

In an industrial environment the time slot dedicated to the logic testing of a single integrated circuit needs to be as short as possible in order to boost production. During this short period thermally induced errors might remain hidden due to long thermal time constants.

This paper introduces a methodology to determine the steady-state die temperature where a short logic test is able to reveal logic faults. The evolved die temperature is simulated with a logi-thermal simulator engine that performs logic simulation by taking self-heating into account.

We propose that the testing should take place at an elevated temperature where the temperature dependent failures arise. This approach makes it possible to detect otherwise hidden defects while keeping testing times short.

***Index Terms*—logic test, test automation, thermal aware**

I. Introduction

In today's digital integrated circuit production logic testing after manufacture plays an important role in achieving a high yield. To accomplish this the logic testing should be able to determine if a chip is faulty with the highest possible precision.

Due to the fact that an exhaustive test is time consuming only short tests are feasible in production. These tests are usually carried out at room temperature where thermally induced defects might remain hidden and the length of the testvectors prevents the chips from heating up to their typical operating temperature.

Using a logi-thermal simulator [1], [2] the evolved operating temperature of the circuit can be determined by simulations already at design time, long before manufacture. If a heating step is introduced in the process line directly before the testing, a short testvector can detect thermal problems, thus maintaining the production speed and enhancing the yield.

This virtual temperature map can then also be used to assist the preparation of the DUT before testing. The operating modes and components that are the most sensitive to temperature change can be detected. This makes it possible to define the shortest test vector that can reveal an error caused by the elevated temperature.

II. Simulation framework

The simulator consists of two simulator engines: a logic engine, that is more complex than a conventional logic simulator in the sense that it handles gate delays and logs gate activities and a thermal engine, that calculates the chip's temperature distribution using the activity (and dissipation) data provided by the logic engine (Fig. 1).

The logic engine is fed with the schematic description of a digital circuit and the properties of the gates: the logic function that describes their operation, the function that describes the gate delay's dependence on temperature and the dissipation of the gate when it is active.

The thermal engine's input is the physical layout of the circuit and (if the desired precision necessitates and the information is available) the details of the chip's package and further thermal ambience.

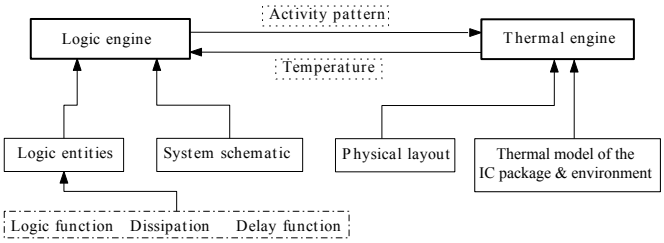


Figure 1. The architecture of the logi-thermal simulator

The architecture of the framework was designed in a very flexible way. The cooperation of the two engines is realized through a very thin interface which decreases the cross dependencies to a minimum. As a consequence, the engines in the framework are easily replaceable. Users can be given the opportunity for example to choose between a fast but less accurate and a very accurate but slower thermal engine. Simulation times and precision can thus be tailored to the needs of engineers using the simulator.

The simulation is operated by an instance of the *logithermal_simulator* class. It organizes the activities of the logic and the thermal engines. It is also an *observable* class [3] and it notifies its observers every time a time step is over. This mechanism is used by loggers that record important values calculated during the simulation and save them in several output files. For example, the VCD file that stores the waveforms of a circuit is created by such a logger.

The same design pattern is used by the logic modules' ports to keep track of the simulation time.

ISBN:978-2-35500-022-5

A. The logic engine

The logic engine was designed to ease the extension of the gate library and to give the maximum flexibility in the definition of the behaviour of elements.

The logic engine is coordinated by a *logic_simulator* instance that contains the modules and instructs them to scan their inputs and to check whether there is a change scheduled at any of their outputs at every simulation cycle.

The logic modules can have an arbitrary number of ports that can be configured to act as inputs, outputs or inouts and can be of any size. Delays can be defined in a port-to-port manner and are handled by the port objects. This means that when a new module type is defined, delays need not be given or taken care of, it is done automatically based on description files to be discussed later.

Every logic module is loaded dynamically at the beginning of the simulation. The gates are stored in dynamically linked libraries (DLLs or SOs) and only the ones instantiated in the simulated design are loaded. If a library is present at a certain location in the simulator engine's directory structure, it can be used as a source of logic modules. Hence the gate library can be extended by simply creating a new dynamic library and placing it at the appropriate location.

A new module can be created by inheritence. The base class of all modules is a very intelligent entity. It knows everything that a module needs to know for logithermal simulation: it can handle ports, knows how to respond to the change of input signals, has a delay (realized by its ports), has a dissipation and records its own activity (whether it changed its state in the latest time step or not).

There is only a very short list of things to define in order to create an entirely new module. The module's ports need to be created in the constructor by giving their name, their size and their direction. The method that describes the behaviour of the module is pure virtual in the base class so it needs to be defined in every subclass. The method gets the value of the module's inputs and yields the resulting output values. Whether the new output values are different from the previous ones (and thus action needs to be taken) and the delays that are introduced by the module are determined and handled by the ports, so it doesn't have to be given by the user code.

When extending the gate library, one can focus on the actual logic function that the modelled gate realizes. The ease of realizing functions of any complexity is ensured by the fact that the complete toolset of the C++ language is available. This makes it possible to modell entire CPUs in a single logic module. In fact, the level of abstraction is not bound or limited in any way – the same design might contain modules as simple as a logic gate and as complex as multicore processors next to eachother.

B. The thermal engine – SUNRED

The physical layout of a circuit is fed to the thermal engine as a list of bounding boxes with names exactly matching the module instance names in the logic circuit description. The engine calculates with rectangular dissipators. The solution method used in the thermal engine is SUNRED.

SUccessive Network REDuction is a solution method for the Finite Differences Method (FDM) field simulation models. The applicable fields are defined by partial differential equations (PDE-s): Laplace-equation and Poisson's equation [4]. Different implementations of the SUNRED algorithm can solve thermal, electrostatic or electro-thermal fields [5], [6]. In case of the logi-thermal simulation we start from the following form of the heat equation:

$$\lambda \cdot \operatorname{div} \operatorname{grad} T = \lambda \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) = \\ = p(x,y,z) + c_{th} \frac{\partial T}{\partial t} \quad (1)$$

Finite differences methods approximate the solution of PDE-s by replacing the differential equations with difference equations. In one dimension this means

$$f'(a) \simeq \frac{f(a+x) - f(a)}{x} \quad (2)$$

By decreasing x the approximation of the differentials becomes more accurate. The FDM method divides the examined geometrical domain into a finite number of simulation cells and calculates the (2) difference in the center of each section. In 2D or 3D this division results in a finite differences grid as shown in Fig. 2 (a). The approximating difference equation of (1) is the following:

$$\lambda \frac{T_{i-1,j,k} - T_{i,j,k}}{\Delta x^2} + \lambda \frac{T_{i+1,j,k} - T_{i,j,k}}{\Delta x^2} + \\ + \lambda \frac{T_{i,j-1,k} - T_{i,j,k}}{\Delta y^2} + \lambda \frac{T_{i,j+1,k} - T_{i,j,k}}{\Delta y^2} + \\ + \lambda \frac{T_{i,j,k-1} - T_{i,j,k}}{\Delta z^2} + \lambda \frac{T_{i,j,k+1} - T_{i,j,k}}{\Delta z^2} = \\ = p_{i,j,k} + c_{th\ i,j,k} \frac{\mathrm{d}T_{i,j,k}}{\mathrm{d}t} \quad (3)$$

Finite differences can be modeled as resistors between two neighboring nodes. The SUNRED method divides each resistor into two parts as illustrated by Fig. 2 (b). E', W' etc. denote *East*, *West* etc. neighboring nodes, C indicates the center node. According to the thermal-electrical analogy, in the thermal case the resistors are thermal resistances. Fig. 2 (c) shows the SUNRED cell model in 3D.

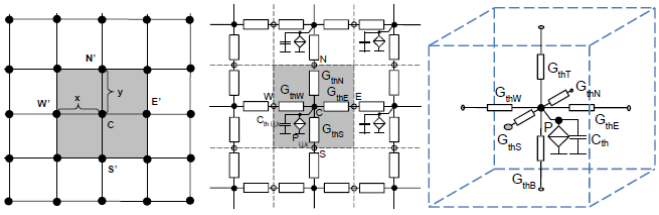


Figure 2. (a) Finite differences grid in 2D, (b) SUNRED model in 2D, (c) SUNRED cell in 3D

The dissipation is realized as current sources (Fig. 2 (c)). The Successive Network Reduction algorithm solves this network for the given boundary conditions, excitations and thermal loads, and determines all the temperatures and heat fluxes [6]. Fig. 3. illustrates the steps of the method. The base algorithm is a direct solution method by its nature; there is no need for iteration but for dynamic simulation iteration is inevitable. Time domain solution of (3) as an ordinary differential equation (ODE) uses the backward Euler method [7] where the thermal capacitances of Fig. 2. are replaced by the Fig. 4. circuit equivalent. This iteration, after the first step however, requires much less computation than a DC solution or the first iteration step because the network reduction is not required as the structure of the network does not change. Network reduction uses matrix to matrix multiplications and matrix inversions, the solution of time steps requires only matrix to vector multiplications. This results in an approx. 6 times shorter solution time.

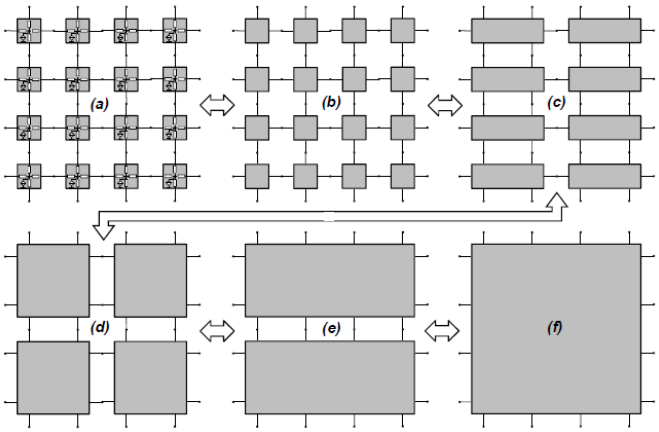


Figure 3. The successive network reduction method

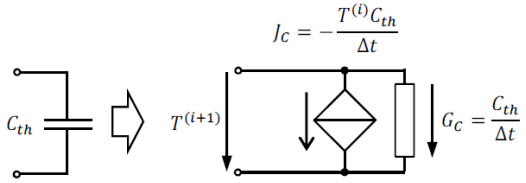


Figure 4. Backward Euler equivalent circuit of a thermal capacitance

III. Industrial input and output formats

Throughout the development of the simulator framework a design goal was to provide industrial standard means to feed information to it and to also yield results in standard formats. Not every input format is standard though. This is due to the uniqueness of the task, but whenever a standard was available, it was applied.

The delays in a circuit can be defined using the *Standard Delay Format* (SDF), an IEEE (originally OVI) standard for the representation and interpretation of timing data for use at any stage of an electronic design process [8].

SDF is a symbolic expression based language that can describe cell and interconnect delays in a temperature and supply voltage dependent way. The delays of a cell are given in a port-to-port manner. Delays can be assigned to individual module instances or to every instance of the same type.

The waveforms resulting from the logic simulation are saved in the Value Change Dump format (VCD), which is also an IEEE standard (actually it was defined as part of the Verilog standard – IEEE Standard 1364-2001) [9]. It is an ASCII-based format as well and is recognized by most EDA tools.

IV. Simulation results

The sample circuit used to demonstrate the proposed method is a simple checksum generator circuit, a little subcircuit that can become a bottleneck in a system. The pseudo code of the BSD checksum algorithm [10] is given below.

```
1 checksum = 0
2 while there_is_input() {
3   word = get_next_word()
4   checksum = (checksum >> 1) + ((checksum & 1) << 15)
5   checksum = checksum + word
6 }
```

Listing 1. The algorithm calculating the BSD checksum

The simple circuit realizing the BSD checksum algorithm can be seen in Fig. 5. The input dataflow is fed to the input buffer in 16 bit chunks. The adder calculates the sum of the previous checksum and the current data word. The shift register loads the result of the adder and then performs a rotation to the right. The register is switched between these two operating modes through the $l/\overline{r}$ (load/rotate) input.

The data arrives at a given rate and the clock and control signals of the register are set accordingly. The adder on the other hand is a combinational circuit that has to perform the operation during the time slot between the arrival of two data words.

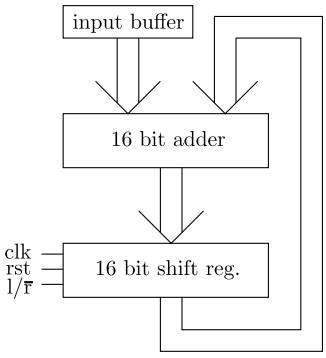


Figure 5. The schematic structure of the checksum generator

The figures 6-8. show the waveforms resulting from the simulations performed at 25 °C, 45 °C and 50 °C respectively. At 25 °C and 45 °C the circuit operates flawlessly although the increase in the delay of the adder's output can be observed

ISBN:978-2-35500-022-5

in the latter waveform. At 50 °C the delay exceeds the time slot available for the calculation of the sum, and thus the operation fails. According to our logithermal simulation, if such a circuit is heated up to 50 °C before running a few μs long test then thermally induced errors remaining unrevealed at room temperature can be detected.

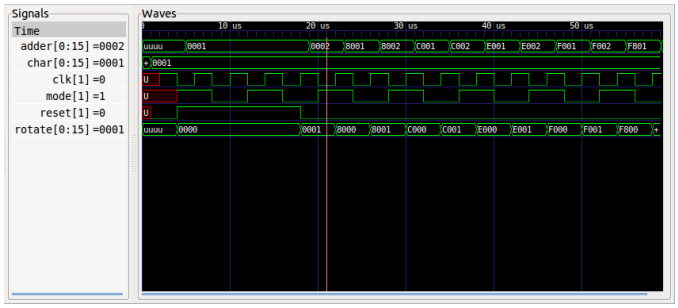


Figure 6. The waveforms of the checksum generator at 25 °C

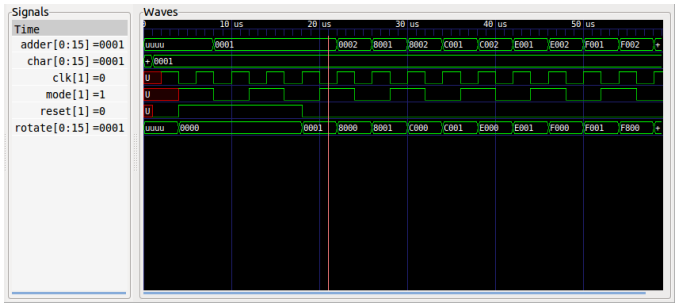


Figure 7. The waveforms of the checksum generator at 45 °C

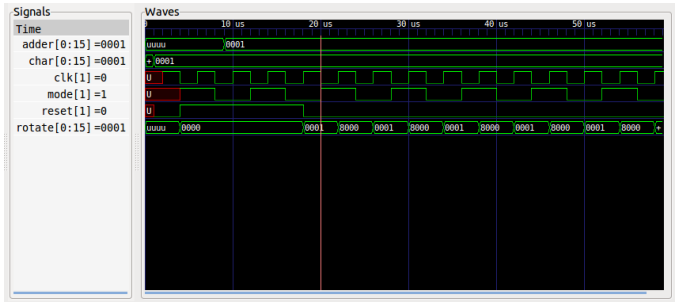


Figure 8. The waveforms of the checksum generator at 50 °C

The calculated thermal map of the circuit is shown in Fig. 9. It clearly shows the hot spots and how the active subcircuits' dissipation effects the temperature of the neighbouring modules.

V. Summary

We have proposed a methodology to determine the steady-state die temperature where a short logic test is able to reveal logic faults that remain hidden when a short testvector is run at room temperature. The operation of a digital system was simulated using our logi-thermal simulation framework

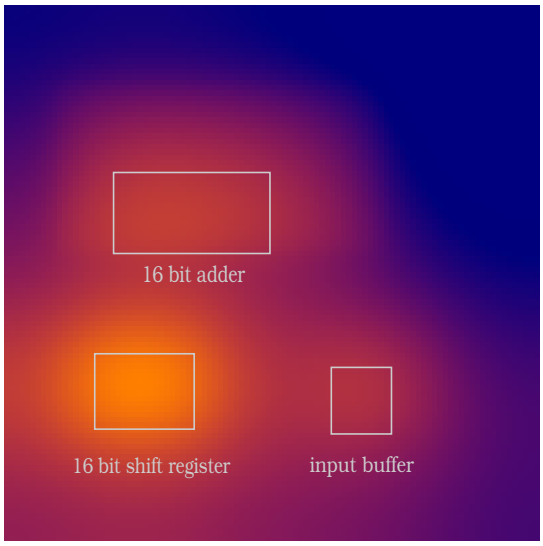


Figure 9. Thermal map resulting from the logi-thermal simulation of the checksum generator

that takes ammbient temperature and the dissipation of active circuits into account.

VI. Acknowledgement

This work was partly supported by the IP 248603 THERMINATOR FW7 project of the European Union and by the Hungarian Government through TÁMOP-4.2.1/B-09/1/KMR-2010-0002.

References

[1] G. Nagy and A. Poppe, "Simulation framework for multilevel power estimation and timing analysis of digital systems allowing the consideration of thermal effects," in *Test Workshop (LATW), 2012 13th Latin American*, april 2012, pp. 126–130.

[2] A. Timar, G. Bognar, and M. Rencz, "Improved power modeling in logi-thermal simulation," in *Thermal Investigations of ICs and Systems (THERMINIC), 2011 17th International Workshop on*, Paris, France, sept. 2011, pp. 1–6.

[3] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1994.

[4] A. D. Polyanin, *Handbook of Linear Partial Differential Equations for Engineers and Scientists*. Chapman & Hall/CRC Press, Boca Raton, 2002.

[5] V. Szekely, "SUNRED a new thermal simulator and typical applications," in *Thermal Investigations of ICs and Systems (THERMINIC), 1997 3rd International Workshop on*, sept. 1997, pp. 84–90.

[6] L. Pohl, Z. Kohari, and V. Szekely, "Fast field solver for the simulation of large-area OLEDs," *Microelectronics Journal*, vol. 41, no. 9, pp. 566–573, 2010.

[7] B. P. Demidovich and I. A. Maron, *Computational Mathematics*. Editorial Mir, Moscow, 1973.

[8] *Standard Delay Format Specification*. Open Verilog International, 1995.

[9] *IEEE Standard Verilog Hardware Description Language, IEEE Std 1364-2001*. IEEE, 2001.

[10] FreeBSD. (2011) BSD checksum. [Online]. Available: http://svnweb.freebsd.org/base/stable/9/usr.bin/cksum/sum1.c?revision=225736&view=markup

ISBN:978-2-35500-022-5

25-27 September 2012, Budapest, Hungary

Nanoscale Electrothermal Energy Conversion Devices

Je-Hyeong Bahk[1,2], Younes Ezzahri[2,3], Kazuaki Yazawa[1,2], Bjorn Vermeersch[1,2], Gilles Pernot[2,4], Ali Shakouri[1,2,*]

[1] Birck Nanotechnology Center, Purdue University, 1205 West State Street, West Lafayette, IN 47907-2057, USA
[2] Baskin School of Engineering, University of California Santa Cruz, CA 95064-1077, USA
[3] Département Fluides, Thermique, Combustion, Institut Pprime, CNRS-Université de Poitiers-ENSMA, ENSIP Bâtiment de mécanique, 2 rue Pierre Brousse, F86022 Poitiers Cedex, France.
[4] LOMA, Université de Bordeaux, 33400 Talence, France
[*] Shakouri@purdue.edu

***Abstract*- Energy consumption in our society is increasing rapidly. A significant fraction of the energy is lost in the form of heat. In this talk we introduce thermoelectric devices that allow direct conversion of heat into electricity. Some new physical concepts and nanostructures make it possible to modify the trade-offs between the bulk electrothermal material properties through the changes in the density of states, scattering rates, and interface effects on the electron and phonon transport. The potential to increase the energy conversion efficiency and bring the cost down to $0.1-0.2/W will be discussed. We also describe how similar principles can be used to make micro refrigerators with cooling power densities exceeding 500 Watts per square centimeter. Hybrid liquid/solid-state cooling will be shown to have the potential to reduce the total cooling power requirement significantly by selective removal of hot spots. Finally, experimental results will be presented for thin film thermal conductivity of nanostructured materials using a femtosecond laser pump-probe technique. We describe how the ballistic and diffusive components of heat transport can be identified. The transition between energy and entropy transport in nanoscale devices will be discussed.**

I. INTRODUCTION

Thermoelectric is a solid-state energy conversion technique that can directly convert heat to electricity and vice versa. Since the working fluid is electrons/holes, thermoelectric devices have many unique advantages over conventional energy conversion systems including: no moving parts or vibration, high reliability and durability, compactness, and easy control.[1] It is thus very useful for various applications such as hot-spot cooling on micro-chips, and electrical power generation from waste heat. The energy conversion efficiency of thermoelectric devices is directly related to the thermoelectric figure of merit, $ZT=S^2\sigma T/(\kappa_e+\kappa_l)$, of the materials used, where S is the Seebeck coefficient, σ is the electrical conductivity, T is the absolute temperature, and κ_e and κ_l are the electronic and lattice thermal conductivities, respectively. If ZT is larger than 3-5, the thermoelectric device can be competitive with traditional mechanical energy conversion systems. In the 1950s, Bi_2Te_3 alloys were found to have ZT~1 at room temperature, but since then, the progress has been slow because all the properties comprising ZT are mutually coupled, and it is extremely difficult to enhance one property without affecting another.

Recent advances in the material synthesis and growth techniques and pioneering theoretical studies have enabled breakthroughs for the enhancement of thermoelectric materials beyond ZT~1. In the early 1990s, Hicks and Dresselhaus [2] theoretically predicted a significant enhancement of ZT in low-dimensional materials such as quantum wells and wires. Since their work, along with timely availability of advanced materials, a great deal of research on nanostructured materials has been conducted for thermoelectric applications.[3,4] Indeed significant improvements of ZT values have been reported during the past few decades. However, most of the improved ZT values were achieved mainly by the reduction of lattice thermal conductivity, and the significant enhancement of thermoelectric power factor, $S^2\sigma$, has not been realized to date. We will discuss several difficulties as well as future directions for the power factor enhancement and, in Section II, summarize the recent results of ZT enhancement achieved for various nanostructured materials.

Applications of TE devices have been limited due to their low energy conversion efficiency. An important factor for any energy conversion device is cost per watt analysis. Yazawa and Shakouri [5] recently analyzed the cost-efficiency trade-off of TE devices and found that TE devices can be competitive with other sustainable energy technologies in terms of cost per watt even with the current efficiency values when the design of the device is optimized. We revisit the cost analysis of TE devices in Section III.

Another advantage of TE devices is their ability to be combined with other conventional liquid cooling techniques [6], which offer an additional degree of freedom for removing both background heating and hot spots in integrated circuit chips. Hot spots and non-uniform temperature distribution in the chip can degrade the performance and reduce the reliability. Unfortunately, most of the existing cooling techniques cannot remove the hot spots selectively and they have to operate in a suboptimal fashion and overcool the entire chip [7]. To overcome these difficulties, one solution is to use hybrid solid-state and liquid cooling, which is described in Section IV.

Time domain thermoreflectance (TDTR) is a femtosecond pulsed laser technique capable of characterizing thermal properties of thin films [8,9] and therefore highly valuable to

assess the performance of thermoelectric materials [10-12]. Thermoreflectance imaging is another thermal characterization technique utilizing optical measurement of the variation of light reflectivity with temperature on the illuminated surface. Using a high resolution CCD camera, temperature mapping is possible without a scan. Joule heating and Peltier cooling can be separated since they have different dependencies on the magnitude and direction of current. Thermal conductivity measurements of thin film nanostructured materials using TDTR and the details of the two characterization techniques will be discussed in Sections V and VI.

Based on Shastry's work [13], we analyze both the ballistic (non-thermal) and diffusive (thermal) energy transport by electrons in metals occurring for a very short period of time. It is found that a damped oscillatory behavior of the energy density, Green's function, due to the ballistic transport is dominant at very short time less than the scattering time and at very short length scale, and then disappears as the diffusive transport becomes significant. Both spatial and temporal impulse responses are obtained and analyzed to separate the ballistic and diffusive components of the transport in Section VII. Finally, a summary is followed to conclude the paper.

II. Nanostructured Thermoelectric Materials

The theoretical prediction of a large ZT enhancement in quantum wells and wires by Hicks and Dresselhaus [2] is based on the enhancement of the thermoelectric power factor due to the modification of the electronic density of states in the in-plane direction of those low-dimensional materials. The quantum confinement in the low-dimensional materials creates sharp features in the density of states, which increases the asymmetry in the differential conductivity with respect to the Fermi energy, thereby enhancing the Seebeck coefficient.[14] However, most of the recently reported ZT enhancements did not come from the power factor enhancement. Lattice thermal conductivity was reduced due to the increased interface and boundary scattering of phonons in the nanostructured materials. Some of the reasons that an enhancement of the in-plane power factor in quantum wells or along nanowires is practically difficult could be: 1) increased carrier scattering at the interfaces, which reduces the electrical conductivity, 2) quick disappearing of the sharp features in the density of states due to size non-uniformity in the confinement directions, 3) the reduction of band degeneracy due to the quantum confinement, and 4) the wide distribution of carrier energy compared to the width of the sharp features in the density of states. Furthermore, Kim et al.[15] recently pointed out in their theoretical calculations based on the Landauer formalism that the enhancement in Seebeck coefficient is modest and the major improvement, at most 50% over bulk, comes from the enhancement in effective 3D electrical conductivity in a lower dimensional system with a high packing density.

In the cross-plane direction of superlattices, two scenarios for the Seebeck coefficient enhancement may be possible. Firstly, low-energy carriers can be filtered out of the transport due to the presence of potential barriers in the cross-plane direction of heterostructure superlattices. Hot electrons having energies higher than the barrier height can go over the barriers and contribute to the transport. This energy filtering scheme can enhance the Seebeck coefficient [16, 17] Theoretical calculations showed that a large enhancement of Seebeck coefficient and, thus, ZT is possible when the lateral momentum of carriers is not conserved during the transport over barriers.[18] If the lateral momentum is conserved in analogy with the total internal reflection of light, the emission current over the barrier could be so low that the power factor may not be largely enhanced even with the large Seebeck enhancement. Also, the number of conducting channels in both the well and barrier layers can be a limiting factor in enhancing the power factor.[19] Secondly, when the superlattice period is much smaller than the mean free path of carriers, coherent multiple interferences by the periodic superlattice potential can form mini-bands, which modify the density of states, and thus, can enhance the Seebeck coefficient.[20, 21] A large enhancement of Seebeck coefficient was predicted at low temperatures in the miniband transport regime, but the power factor was not enhanced due to the suppression of group velocity of carriers.[21] However, further optimization of design and selection of materials may be possible to lead to power factor enhancement, even at room temperature or higher.

In the materials with embedded nanoparticles, modification of carrier scattering time by nanoparticle scattering can enhance the power factor.[22-24] Ionized nanoparticles create spatially slowly-varying potentials around them, which are different from the potential by conventional ionized impurities. The actual enhancement of power factor comes from the increased electrical conductivity at a given Seebeck coefficient, rather than the Seebeck enhancement.[24] Core-shell nanoparticles can create energetically sharp scattering characteristics via resonances, which can significantly enhance the Seebeck coefficient at low temperatures.[25] Recently a modulation doping concept was proposed to enhance the electrical conductivity and thus, the power factor.[26] Heremans et al. demonstrated a large power factor enhancement in Tl-doped PbTe at high temperatures up to 770 K, and attributed this enhancement to the distorted density of states by the Tl resonant level inside the valence band of PbTe.[27] The ErAs:InGaAlAs semimetal/semiconductor nanocomposites showed enhanced ZT~1.3 at 800 K over bulk InGaAlAs due mainly to the thermal conductivity reduction by increased phonon scattering by a few nm-sized ErAs nanoparticles.[28]

Fig. 1 shows several important ZT values, both n- and p-type, that were recently reported for nanostructured materials over a wide temperature range. The ZT values of the conventional thermoelectric materials such as Bi_2Te_3, PbTe, and SiGe are also shown in the figure for comparison. ZTs of about 1.5-1.7 have been reported in several different nanostructured materials in the last ten years. ZTs higher than 2 were also reported in a few cases, but further experimental verification is necessary as some of the data are not independently verified [3]. It is noted that most of the ZT improvements in Fig. 1 were obtained via reduction of the lattice thermal conductivity, except for the two cases [26,27]

in which enhanced power factor was the main factor leading to high ZTs.

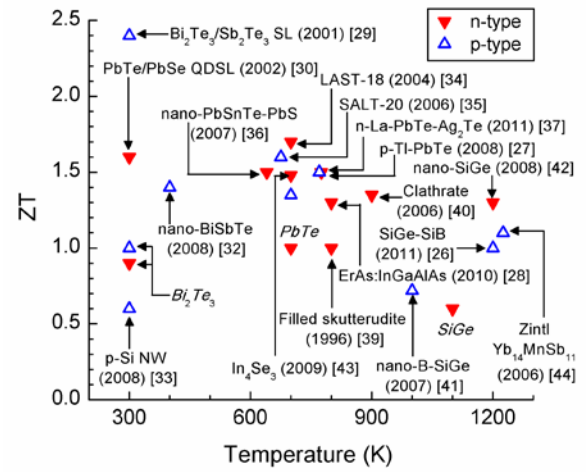


Fig. 1. Advances of thermoelectric figure of merit, ZT, reported over a temperature range from 300 K to 1200 K.

At near room temperature or 300 K, Bi_2Te_3 or PbTe-based nanostructured materials showed significant enhancement in ZT over their bulk counterparts by a reduction of thermal conductivity. The Bi_2Te_3/Sb_2Te_3 superlattice [29] was reported to have a cross-plane thermal conductivity value of 0.22 W/mK, which is more than two times lower than that of $Bi_{0.5}Sb_{1.5}Te_3$ alloy, and attributed to the increased interface phonon scattering in the superlattice. A large ZT of 2.4 was reported for this superlattice, but to our best knowledge, this value has not been reproduced. PbTe/PbSe nano-dot superlattices by Harman et al. [30] showed ZT~1.6 at this temperature due to the very low thermal conductivity of 0.33 W/mK that was inferred indirectly from thermoelectric cooling measurements. Recent direct measurements using the time-domain thermoreflectance technique, however, did not reproduce this value.[31] Recently, nanostructured p-type BiSbTe bulk alloys prepared by hot pressing ball-milled nanopowders showed ZT~1.2 at room temperature and ~1.4 near 400 K.[32] Recently, Hochbaum et al. fabricated and measured rough silicon nanowires of ~50 nm in diameter, which showed a ZT ~ 0.6 at room temperature.[33] This ZT value is still lower than 1, but is significantly higher than the ZT ~ 0.01 of bulk silicon.

At elevated temperatures, near 600 K or higher, several PbTe-based nanostructured materials showed high ZTs of 1.5~1.7 over the best bulk PbTe of ZT~1. The material system, $AgPb_mSbTe_{2+m}$, also known as LAST-m, becomes spontaneously nanostructured when cooled from the melt, which helps reduce the thermal conductivity. With selected m values, e.g. LAST-18, the thermal conductivity was found to be only about 30% of that of PbTe, and thus a ZT~1.7 at 700 K was achieved.[34] A p-type variation of LAST-m, $NaPb_mSbTe_{2+m}$ (SALT-m), also showed ZT~1.6 at 675 K for m~20 with a very low thermal conductivity of 0.85 W/mK.[35] The strain field created by nano-inclusions in these material system is believed to scatter phonons effectively to reduce the thermal conductivity. The $(PbSnTe)_x(PbS)_{1-x}$ system is phase-separated into PbTe-rich and PbS-rich regions, which become coherent nanostructures suppressing the lattice thermal conductivity. A ZT ~ 1.5 at 640 K was reported for this material system of n-type with x~0.08.[36] The n-type La-doped Ag_2Te-PbTe system has nano-scale (50-200 nm) Ag2Te precipitates formed in PbTe, and showed ZT~1.5 at 775 K.[37] Skutterudites and clathrates have complex cage-like crystal structures with voids in which 'rattler' atoms are inserted to effectively scatter acoustic phonons.[38] $CoSb_3$-based filled skutterudites such as p-type $LaFe_3CoSb_{12}$, and n-type $CeFeCoSb_{12}$ showed ZT > 1 at 800 K and higher.[39] Recently, Czochralski-grown clathrate $Ba_8Ga_{16}Ge_{30}$ showed ZT~1.6 at 1100 K. [40]

Beyond 1000 K, SiGe has long been known to be a good thermoelectric material, and its ZT is typically 0.5 ~ 0.6 at 1100 K. After nanostructuring by hot pressing and ball milling, ZT of p-type B-doped SiGe was improved to 0.7 at 1000 K.[41] ZT~1.3 at 1200 K was also reported for nanostructured n-type SiGe. [42] Two-phase SiGe-SiP nanocomposites had ZT > 1 at 1200 K, and this ZT enhancement was attributed to the modulation doping effect that enhanced mobility and thus electrical conductivity.[26]

III. Cost Analysis of Thermoelectric Devices

Energy conversion from a waste heat is important since we are dissipating more than half of the energy from natural resources. Thermoelectric is suitable for this application due to relatively lower operating temperature compared to mechanical systems. Fig. 2 shows the one-dimensional model that we used for theoretical analysis [5].

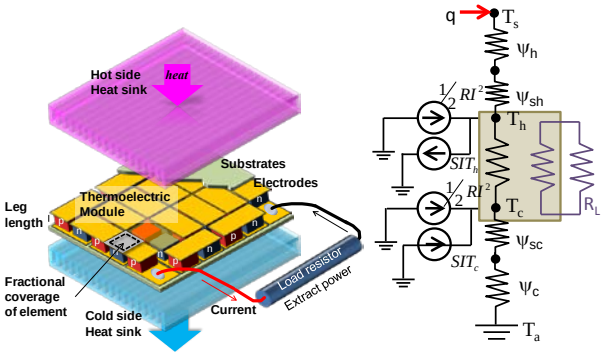


Fig. 2. Thermoelectric waste heat power generation system and electro-thermal network model.

Since the heat source is free of cost, the minimum mass of the material used to produce the maximum power output is the criteria to finding the lowest cost per useful energy produced. The maximum power output, W_{max}, is found in Eq. (1) to be proportional to the square of overall temperature difference from the heat source to the ambient and also inversely proportional to the sum of the external thermal resistances including module substrates. While, the optimum TE element (leg) length d_{opt} is found as Eq. (2),

$$w_{\max} = \frac{Z}{4(1+m)^2}\frac{1}{A\sum\Psi}(T_s - T_a)^2 \quad (1)$$

$$d_{opt} = m\kappa FA\sum\Psi \quad (2)$$

$$m_{opt} = \sqrt{1+ZT} \quad (3)$$

where ΣΨ is the sum of external resistances, A is area of substrate, κ is the thermal conductivity of TE material, and F is fractional area ratio of the TE leg (fill factor). These relations are derived for symmetric thermal contacts with hot and cold reservoirs. There are slight changes for asymmetric systems [45]. Based on this analysis, one can see that the fill factor does not affect power output until reaching less than 1% if the leg length is scaled accordingly for thermal impedance matching. This constant thermal resistance ideally yields an immediate volume reduction of the TE element itself. Without considering the parasitic losses, such as heat leakage and contact resistances, the cost per power output of TE could be reduced by 1/10,000 for 1% of fill factor. Realistic modules and power generation systems, however, still need the substrates and heat sinks with constant volume. As well as, the thermal spreading resistance increasing with reduced fill factor. Considering these impacts, the cost of the system is found for different fill factors as shown in Fig. 3. It can be seen that the reduction of thermal conductivity has the greatest impact on the $/W metric for waste recovery systems [46].

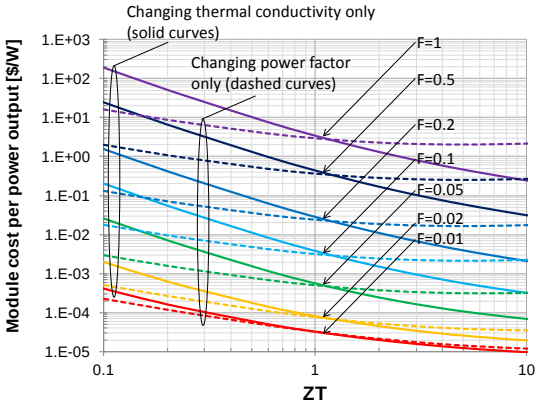


Fig. 3. Module cost performance [$/W] at the maximum power design. Thermal conductivity κ is 1.5 W/mK and power factor is $3.35\text{x}10^{-3}$ W/mK^2 for ZT=1. The number of elements per unit area N is 100 cm^{-2}.

IV. Hybrid Solid-State Micro-Refrigeration / Liquid Cooling

Conventional cooling by microchannel reaches the limit in the reduction of the thermal resistance due to the insensitive heat transfer coefficient to the coolant flow rate, resulting in a massively large amount of power required for a small improvement in cooling. We studied the hybrid scheme using a thermoelectric in conjunction with microchannels as shown in Fig. 4. The fundamental performance model is independent to the kinds of thermoelectric materials, whereas this particular figure shows a Si/SiGe supperlattice cooler. The total cooling power should increase by injecting an electrical current to the thermoelectric for pumping the local heat. In turn, pumping the excess heat flux at the hot spot helps to relax the requirement of the overall heat sink performance.

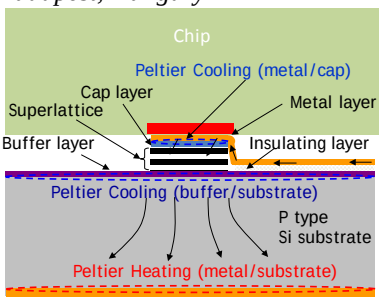


Fig. 4. Thermoelectric hybrid scheme structure diagram. vMicro-channel heat sink is not shown.

Analytic modeling allowed us to analyze the performance when we change the design of the TE element and the drive current to get the minimum cooling power to meet the junction temperature, T_j, requirement not exceeding the maximum allowable temperature, T_{j_max}. Fig. 5 shows an example of the T_j dependance on the above two design parameters.

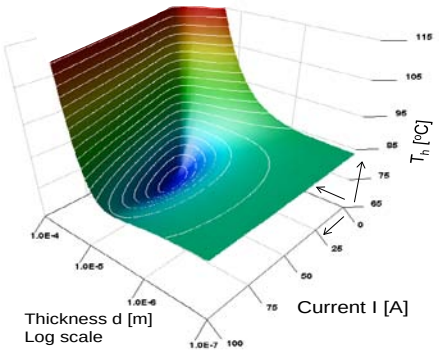


Fig. 5. The junction temperature, T_j, as a function of TE cooler thickness, d, and drive current, I. ZT=0.5 and ambient temperature T_a=35 °C are assumed.

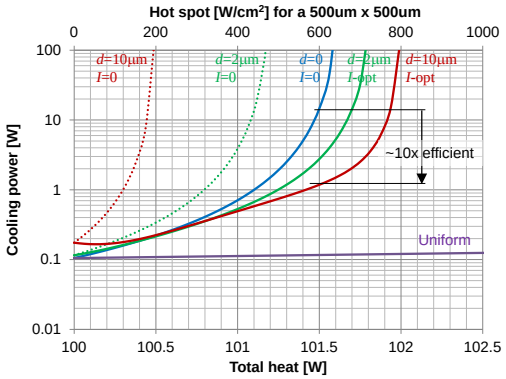


Fig. 6. Total cooling power as a function of the chip power and the hot spot heat flux. Tj_max=85 °C, ZT=0.5 are assumed.

The result shown in Fig. 6 shows that this hybrid scheme could allow us to reduce the total cooling power by nearly 1/10 of what we may need for conventional microchannel cooling at 600 W/cm^2 for a 500 μm x 500 μm single hotspot on a uniformly heated 1 cm x 1 cm chip with 100 W/cm^2.

V. Thermoreflectance Imaging Technique

Device level thermal characterization requires identifying the two-dimensional temperature gradients and the localized

temperatures, thus, thermal mapping is important. Due to the feature size (1-10s microns) of the device with adjacent heaters at a wide range of operating temperatures, conventional Infrared (IR) imaging is not suitable. We developed a thermoreflectance imaging technique with electrical interlocking. Thermoreflectance stands for the reflectivity change versus the temperature changes of the surface. This relationship is nearly the constant. It is a material surface property for a particular wave length of the externally illuminating light [47-49]. By using an inter-lock technique, the temperature change with applied bias can be measured at the accurate timing of ON and OFF. Fig. 7 shows an example of a Si/SiGe superlattice cooler fabricated on a silicon substrate [48]. The heating load imitating the cooling target device, e.g. LSI chip with a hotspot, was fabricated on top of the TE device, so that we can identify the junction temperature.

The power demands of either energy systems or cooling are typically intermittent. The transient response characteristics cause different overall efficiencies. Therefore, transient temperature imaging is critical to determine the performance of the energy conversion devices. The electrical signal interlocking and the image interleave in software enable us to compose the frame-by-frame animation as the time line image in addition to the time response data set. Fig. 8 shows an example of time line images demonstrating the capability of sub-nano second time resolution [50].

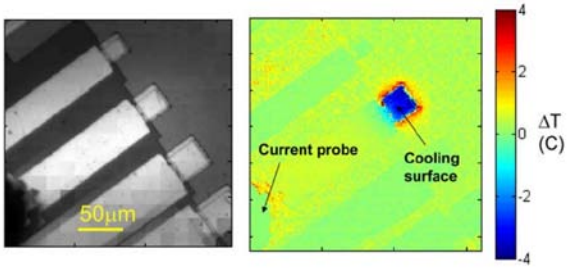


Fig. 7. Thermal image of a 50 x 50 μm^2 microrefrigerator at an applied current of 500 mA. Stage temperature is 30 °C and the device is cycled at a frequency of ~1 kHz [48].

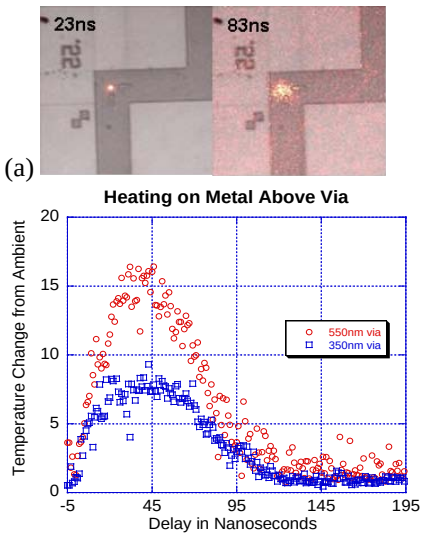


Fig. 8. Picosecond thermal imaging of sub-micron vias [50]. (a) shows the overlay of thermal image on optical image. (b) shows the temperature response to a pulsed power input.

VI. Thermal Identification of Thin Films With Time-Domain Thermoreflectance

The measurement principle of time-domain thermoreflectance is schematically illustrated in Fig. 9. Optical pulses from a femtosecond Ti:Sapphire laser (typical parameters: ~300 fs pulses, repetition rate 76-80 MHz, wavelength ~780 nm) are split into two beams undergoing separate treatment which are then recombined onto the sample under study. The 'pump' beam induces localized heating of the sample while the 'probe' beam monitors the thermally induced changes in surface reflectivity. The delay between pump and probe can be accurately varied by means of a mechanical delay line. Fitting the observed temperature decay as a function of pump-probe delay to a diffusive heat spreading model enables us to identify the unknown thermal parameters.

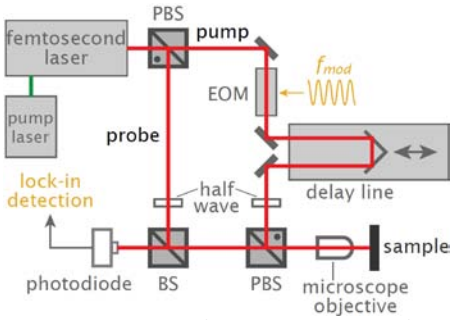


Fig. 9. Schematic principle of a typical TDTR setup for thermal characterization. EOM = electro-optic modulator, (P)BS = (polarizing) beam splitter.

A thin metallic film, typically aluminum, is deposited on top of the sample to act as transducer for the measurement (Fig. 10). The film efficiently converts the absorbed pump pulse into heat and provides a good temperature dependence of surface reflectivity under IR illumination. To achieve sufficient signal-to-noise ratios, the pump beam is modulated at a frequency f_{mod} (typically in the 1-20 MHz range) such that lock-in detection of the probe signal at the same frequency can be utilized. We note here that higher modulation frequencies correspond to shorter effective thermal penetration lengths into the sample. This is very useful for characterization of thin films as it enables one to minimize the influence of the substrate parameters on the measured thermal response. Knowledge of both in-phase, V_{in}, and out-of-phase, V_{out}, signal components on the lock-in can be exploited by using the ratio function, V_{in}/V_{out}, for further processing and analysis. This reduces experimental artifacts associated to the variation of the pump beam size with pump-probe delay and acts as a normalization process [8]. Fig. 11 gives an example of measurement data obtained for a GaAs substrate covered by a 90 nm thick aluminum film.

The thermal parameters of the sample under study can be determined by comparing the experimental data to a three dimensional heat spreading model [51]. Usually only two free parameters remain in the fitting procedure, namely the Kapitza resistance, r_K, of the aluminum/film interface and

thermal conductivity, *k*, of the film. The thermal parameters of the transducer and substrate are fixed to their literature value while the transducer thickness can be determined from picosecond acoustics (see the acoustic echo in the inset of Fig. 11(a)). It is important to point out that analysis of the theoretical signal is, unfortunately, not straightforward. The single pulse response calculated by the model, i.e. the temperature response of the sample surface to a single laser pulse, must first be processed to account for the cumulative temperature effects of the multiple laser pulses and the amplitude modulation of the pump beams before comparison with experiments can be made. Derivations of the final signal form in either time or frequency domain are available elsewhere [51-53].

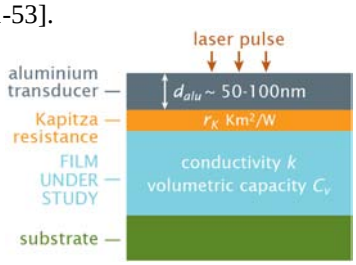


Fig. 10. Typical sample layout for TDTR thermal characterization.

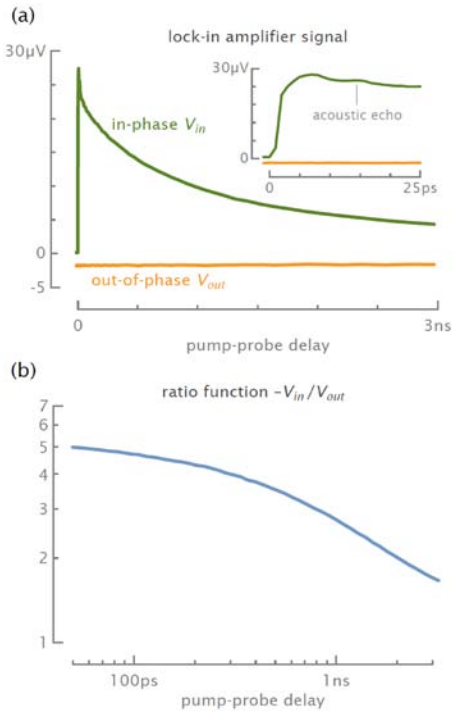


Fig. 11. Measured lock-in signals on a GaAs substrate: (a) in-phase and out-of-phase signal components with inset showing acoustic echo; (b) corresponding ratio function in double logarithmic scale.

Detailed sensitivity analyses show it is preferable to determine the Kapitza resistance from measurements over the entire available pump-probe delay at high modulation frequency. This value can then be used in the fitting procedure to determine the thermal film conductivity at other modulation frequencies. Fig. 12 illustrates identified thermal conductivities for a GaAs substrate. As we can see, an accurate estimation of the transducer thickness is crucial to obtain reliable thermal conductivity data. The picosecond acoustics provide an estimated uncertainty of ±5% on the thickness, which corresponds to a variation of roughly ±10% in the identified thermal conductivity. Errors for the laser spot size of the pump and probe beams on the sample (typically ±7%) also introduce uncertainties in the thermal conductivity, but to a much lesser extent.

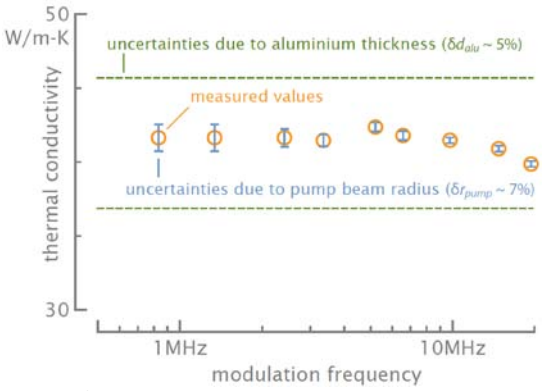


Fig. 12. Identified thermal conductivity and uncertainty contributions for GaAs substrate at different modulation frequencies.

VII. Nonthermal and Thermal Regimes of Energy Transport in Metals

Energy deposits into and propagates through a material in different ways depending on: the excitation, the structure of the material, and the nature of the energy carriers. At short time and length scales, Fourier's law becomes invalid and many non-Fourier heat conduction models, both local and nonlocal, have been developed to overcome problems associated with Fourier's model (e.g., infinite speed of propagation of heat) [54-56]. At very short time and length scales, the distinction between diffusive and ballistic regimes of energy transport becomes very relevant [57-59]. In addition to this distinction, the nonlocal effects may also become very crucial at very short length scales [56].

Using the recently developed formalism of Shastry [13], we have analyzed the transition between the nonthermal (ballistic) and thermal (ballistic-diffusive) electron energy transport in metals assuming a 1D problem in which the top free surface of the metal is excited by a delta power source [60,61]. The starting point of the analysis is the response function, N_2, introduced by Shastry [13]. This function gives a measure of the change in the energy density and hence the temperature at various points in the metal in response to the applied excitation at the top free surface of the metal and, as such, represents the energy (heat) Green's function of the metal. The Shastry-Green function, N_2, is a very simple expression in the case of a metal:

$$N_2(\omega,q)=\frac{-i+\tau_q\omega}{\omega+i\tau_q\omega^2-iD_eq^2} \qquad (4)$$

This expression is obtained by turning off the coupling factor, ξ, between the charge and energy modes. As a matter of fact, and as shown by Shastry, ξ can be expressed using the high frequency limit of the thermoelectric figure-of-merit Z^*T ($\xi=Z^*T/(1+Z^*T)$)[13], and since metals are very poor

thermoelectric materials with a very low ZT, the decoupled limit is therefore justified. In Eq. (4), ω is the angular frequency, q is the electron wave vector, D_e is the electronic thermal diffusivity, and τ_q is the total electron scattering time, which, in general, is a function of q. The detail of the calculations can be found in references [60,61]. Here we shall focus on giving and summarizing the key results of the abovementioned analysis.

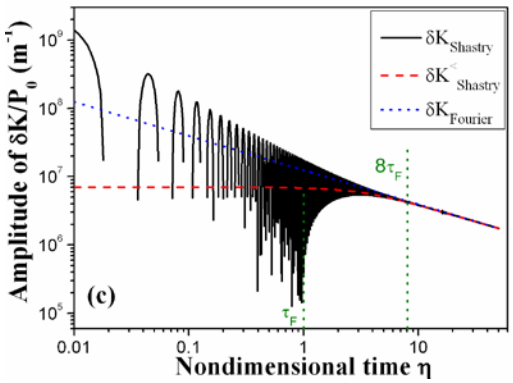


Fig. 13. Comparison between the behaviors of the thermal contributions (dashed line), the sum of the nonthermal and thermal contributions (solid line) to the total electron energy density Green's function at the top free surface of gold at room temperature, with Fourier's model (dotted line) as a function of the non-dimensional time, $\eta=t/\tau_F$.

In our treatment, we assume the case of a constant scattering time $\tau_q = \tau_F$, which we consider to be the average scattering time of electrons over the Fermi surface of the metal. We show that the total energy density Green's function $\delta K(t,x)$, which is proportional to the inverse Fourier transform of $N_2(\omega, q)$, can be expressed as a sum of two contributions; a *"nonthermal"* and *"thermal"* [60,61]. As can be seen in Fig. 13, at the top free surface of the metal, the thermal contribution $\delta K^C(\eta,0)$, undergoes two different temporal regimes. For $\eta<1$, before any scattering event happens, $\delta K^C(\eta,0)$ is almost constant, then starts decaying exponentially with time up to about $\eta\sim 8$, from which it changes the trend and starts following a Fourier type energy diffusion law. On the other hand, the nonthermal contribution $\delta K^>(\eta,0)$ (not shown in Fig. 13, instead we show the sum of the nonthermal and thermal contributions) shows a dampening oscillating behavior as a function of time. While it dominates at very short times, $\delta K^>(\eta,0)$ becomes insignificant after about $\eta\sim 8\text{-}10$. $\delta K^>(\eta,0)$ features two types of oscillations; a fast and a slow one. The fast oscillation is characterized by a period given by $\theta_F^{\text{time}}=2\sqrt{3}a/v_F$, where a refers to the metal lattice constant and v_F is the Fermi velocity. This fast oscillating behavior in the energy transport ($\theta_F^{\text{time}} \approx 1fs$ at room temperature for a typical noble metal) is a consequence of the band cut-off due to the discrete character of the crystalline lattice. This leads to Bragg reflection of electrons in a metal. This fast oscillating behavior can be viewed as an energetic analogous to the conventional Bloch oscillation in the charge density of the conduction band electrons [62,63]. On the other hand, the slow oscillation behavior has a period given by $\theta_F^{\text{time}}=4\pi\tau_F$ and it describes the nature of the thermalization process, which means the transition from the nonthermal regime to the thermal regime of energy transport. This transition occurs in a damped oscillatory manner with a pseudo-period proportional to the total scattering time of electrons.

The analysis made at the top free surface of the metal can easily be extended to any location, y. We report the spatial behaviors of the total energy density Green's function $\delta K(\eta,y)$ as calculated based on Shastry's formalism, in comparison with the Cattaneo and Fourier models [60,61]. The amplitude of $\delta K(\eta,y)$ decreases as the observation point is moved through the metal far from the excitation location (top free surface). Moreover, because of the causality requirement, the energy density vanishes for locations $y>\eta$. The locations beyond the energy pulse front remain unaltered while at locations before the front, the energy density tends to a Fourier-type behavior at long time scales.

Similarly to the temporal behavior, the spatial behavior of $\delta K(\eta,y)$ features fast and slow oscillations. The fast spatial period is given by by $\theta_F^{\text{space}}=2a$ and it is nothing other than the size of the Wigner–Seitz unit cell, which is the reciprocal of the FBZ, while the slow spatial period is proportional to the MFP ($\Gamma_F=v_F\tau_F$) of the electron, $\theta_F^{\text{space}}=(4\pi/\sqrt{3})\Gamma_F$. Similar to the oscillations in the time domain, the oscillations in the space domain describe the same sequence of physical phenomena. Namely, the nonthermal (ballistic) transport of the energy density at short time scales, when the distribution of the electronic system in the conduction band of the metal is still in a nonequilibrium state, followed by an oscillating transition to the thermal (ballistic-diffusive) regime, when the electronic distribution tends toward an equilibrium thermal distribution and a temperature can be defined. The ratio of the spatial to the temporal periods of the fast and slow oscillations is the same for both oscillation types and it is given by the speed of the energy pulse, $U=\sqrt{D_e/\tau_F}$.

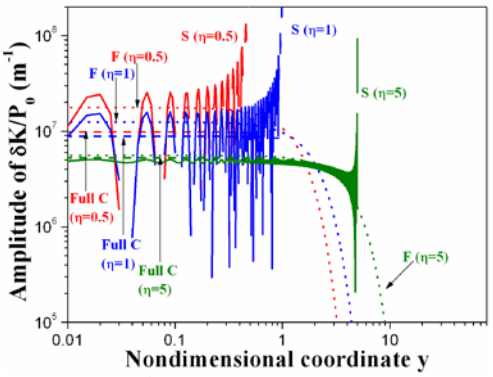


Fig. 14. Spatial behavior of the total electron energy density Green's function of gold at room temperature at different times, η, as calculated based on Shastry's model (S, solid), in comparison with Cattaneo's model (C, dashed) and Fourier's model (F, dotted). The non-dimensional coordinate in the x-axis is defined as $y=x/\sqrt{D_e\tau_F}$.

It may appear shocking or anomalous to have negative values of the electron energy density at short time scales. One should not forget, however, that at these short time scales, the electronic distribution in the conduction band of the metal is still in a non-equilibrium state and the energy density is mostly transported ballistically before it thermalizes. As the electronic distribution tends to an equilibrium thermal one,

there is an exchange in energy density between different locations of the metal both ballistically and through diffusion (represented by D_e). This nonthermal (ballistic) transport of the energy density dampens out with time very quickly as the nonthermal regime transitions to a full diffusive thermal regime. Electron-electron interactions, not considered here, can redistribute the electrons in the momentum space, and affect the ballistic distance traveled by electrons with different wavelengths.

VIII. SUMMARY

Nanoscale electrothermal energy conversion devices have a wide range of applications. Thermoelectric devices are a viable candidate for cost-effective sustainable energy applications. Although the low energy conversion efficiency of the current TE devices limits its applications, recent advances in thermoelectric material research based on nanostructured materials have shown promising enhancement of the efficiency. The cost analysis of TE devices shows that TE devices can be as competitive as conventional mechanical energy conversion systems even with current efficiency values when it comes to cost per watt performance. A hybrid cooling scheme, combining TE micro-coolers with micro-channel liquid cooling, is very effective for removing hot spots that plague today's micro-chips. Thermal transport in thin film nanostructured materials can be precisely studied using the time-domain thermoreflectance technique. The thermoreflectance imaging technique is also useful for fast temperature mapping of sample surfaces without a scan providing sub-microsecond temporal and submicron spatial resolutions. Finally, we have applied Shastry's model to electron energy transport in metals, and interesting separation of ballistic and diffusive transports in temporal and spatial analysis is revealed due to the unique oscillatory behavior of the ballistic component in very short time and length scales.

ACKNOWLEDGMENT

A.S. acknowledges the funding from the Center for Energy Efficient Materials, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Basic Energy Sciences under Award Number DE-SC0001009 which supported the work of B.V. and J.-H.B; DARPA/Army Research Office, Contract No. W911NF0810347 supporting the work of G.P.; and the Interconnect Focus Center, one of the five research centers funded under the Focus Center Research Program, a DARPA and Semiconductor Research Corporation Program which supported the work of Y.E. and K.Y.

REFERENCES

[1] L. E. Bell, Science 321, 1457 (2008).
[2] L. D. Hicks, M. S. Dresselhaus, Phs. Rev. B 47, 12727 (1993).
[3] C.J. Vineis, A. Shakouri, A. Majumdar, M.G. Kanatzidis, Advanced Materials 22, 3970 (2010).
[4] A. Shakouri, Annual Review of Materials Research 41, 399 (2011).
[5] K. Yazawa, A. Shakouri, Environ. Sci. Technol 45, 7548 (2011).
[6] A. Bar-Cohen, M. Arik, M. Ohadi, Proc. IEEE 94, 1549 (2006).
[7] V. Sahu, Y. Joshi, A. Federov, Proceedings of ITherm (2008).
[8] R. M. Costescu, M. A. Wall, D. G. Cahill, Phys. Rev. B 67, 054302 (2003).
[9] S. Huxtable, D. G. Cahill, V. Fauconnier, J. O. White, J.-C. Zhao, Nature Mater. 3, 298 (2004).
[10] G. Pernot, et al. Nature Mater. 9, 491 (2010).
[11] L. E. Cassels, et al., J. Vac. Sci. Technol. B 29, 03C114 (2011).
[12] P. G. Burke et al., J. Electron. Mater. 41, 948 (2012).
[13] B. S. Shastry, Rep. Prog. Phys., 72, 016501 (2009).
[14] A. Shakouri, M. Zeberjadi, Chapt. 9, in Thermal Nanosystems and Nanomaterials, ed. by S. Volz, (Springer Berlin / Heidelberg, 2009).
[15] R. Kim, S. Datta, M. S. Lundstrom, J. Appl. Phys. 105, 034506 (2009).
[16] A. Shakouri, J. E. Bowers, Appl. Phys. Lett. 71, 1234 (1997).
[17] G. D. Mahan, L. M. Woods, Phys. Rev. Lett. 80, 4016 (1998).
[18] D. Vashaee, A. Shakouri, Phys. Rev. Lett. 92, 106103 (2004).
[19] R. Kim, C. Jeong, M. S. Lundstrom, J. Appl. Phys. 107, 054502 (2010).
[20] Z. Bian et al., Phys. Rev. B 76 , 205311 (2007).
[21] J.-H. Bahk, R. B. Sadeghian, Z. Bian, A. Shakouri, J. Electron. Mater. 41, 1498 (2012).
[22] S. V. Faleev, F. Leonard, Phys. Rev. B 77, 214304 (2008).
[23] M. Zebarjadi et al., Appl. Phys. Lett. 94, 202105 (2009).
[24] J.-H. Bahk, Z. Bian, M. Zebarjadi, P. Santhanam, R. Ram, A. Shakouri, Appl. Phys. Lett. 99, 072118 (2011).
[25] J.-H. Bahk, P. Santhanam, Z. Bian, R. Ram, A. Shakouri, Appl. Phys. Lett. 100, 012102 (2012).
[26] M. Zebarjadi et al., Nano Lett. 11, 2225 (2011).
[27] J. P. Heremans et al., Science 321, 554 (2008).
[28] J. M. O. Zide et al., J. Appl. Phys. 108, 123702 (2010).
[29] R. Venkatasubramanian, E. Siivola, T. Colpitts, B. O'Quinn, Nature 413, 597 (2001).
[30] T. C. Harman, P. J. Taylor, M. P. Walsh, B. E. LaForge, Science 297, 2229 (2002).
[31] Y. K. Koh, C. J. Vineis, S. D. Calawa, M. P. Walsh, D. G. Cahill, Appl. Phys. Lett. 94 , 153101 (2009).
[32] B. Poudel et al., Science 320 , 634 (2008).
[33] A. I. Hochbaum et al., Nature 451 , 163 (2008).
[34] K. F. Hsu et al., Science 303, 818 (2004).
[35] P. F. P. Poudeu, J. D'Angelo, A. D. Downey, J. L. Short, T. P. Hogan, M. G. Kanatzidis, Angew. Chem. Int. Ed. 45 , 3835 (2006).
[36] J. Androulakis et al., J. Am. Chem. Soc. 129, 9780 (2007).
[37] Y. Pei, J. Lensch-Falk, E. S. Toberer, D. L. Medlin, G. J. Snyder, Adv. Funct. Mater. 21, 241 (2011).
[38] G. S. Nolas, J. Poon, M. Kanatzidis, Mater. Res. Soc. Bull. Vol. 31, p. 199, Mar. 2006.
[39] B. C. Sales, D. Mandrus, R.K. Williams, Science 272, 1325 (1996).
[40] A. Saramat et al., J. Appl. Phys. 99, 023708 (2006).
[41] M. S. Dresselhaus et al., Adv. Mater. 19, 1043 (2007).
[42] W. Wang et al., Appl. Phys. Lett. 93 , 193121 (2008).
[43] J.-S. Rhyee et al., Nature 459, 965 (2009).
[44] S. R. Brown, S. M. Kauzlarich, F. Gascoin, G. J. Snyder, Chem. Mater. 18, 1873 (2006).
[45] K. Yazawa and A. Shakouri, J. Appl. Phys. 111, 024509 (2012)
[46] K. Yazawa and A. Shakouri, J. Mater. Res. 27, 1211 (2012).
[47] J. Christofferson and A. Shakouri, Rev. Sci. Instrum. 76, 024903-1-6 (2005).
[48] A. Shakouri, Proc. IEEE 94 (8), 1613 (2006).
[49] K. Yazawa and A. Shakouri, Electronics Cooling Magazine Spring Issue 2011, ITEM Publications, pp. 10 (2011).
[50] J. Christofferson, K. Yazawa, A. Shakouri, Proc. 14th International Heat Transfer Conference (IHTC14) (2010).
[51] D. G. Cahill, Rev. Sci. Instrum. 75, 5119 (2004).
[52] A. J. Schmidt, X. Chen, G. Chen, Rev. Sci. Instrum. 79, 114902 (2008).
[53] S. Dilhaire, G. Pernot, G. Calbris, J.-M. Rampnoux, S. Grauby, J. Appl. Phys. 110, 114314 (2011).
[54] A. A. Joshi, A. Majumdar, J. Appl. Phys. 74, 31 (1993).
[55] F. X. Alvarez, D. Jou, Appl. Phys. Lett. 90, 083109 (2007).
[56] D. Y. Tzou, ASME J. Heat Transfer 117, 8 (1995).
[57] G. Chen, Phys. Rev. Lett. 86, 2297 (2001).
[58] G. Chen, ASME J. Heat Transfer 124, 320 (2002).
[59] M. E. Siemens, Q. Li, R. Yang, K. A. Nelson, E. H. Anderson, M. M. Murnane, H. C. Kapteyn, Nature Mater. 9, 26 (2010).
[60] Y. Ezzahri, A. Shakouri, Phys. Rev. B 79, 184303 (2009).
[61] Y. Ezzahri, A. Shakouri, ASME J. Heat Transfer 133, 072401 (2011).
[62] C. Waschke, H. G. Roskos, R. Schwedler, K. Leo, H. Kurz, K. Köhler, Phys. Rev. Lett. 70, 3319 (1993).
[63] T. Dekorsy, R. Ott, H. Kurz, K. Köhler, Phys. Rev. B 51, 17275 (1995).

25-27 September 2012, Budapest, Hungary

The SMARTPOWER project: goals and tasks

Afshin Ziaei
Thales Research and Technology
1, Avenue Augustin Fresnel, 91767 Palaiseau, France
afshin.ziaei@thalesgroup.com

Text unavailable at the time of printing.

25-27 September 2012, Budapest, Hungary

Technical Requirements for the Development of a High Performance RF module for Airborne Radar Application

Franck Vouzelaud, Philippe Kertesz, Julie Mazeau and Yves Mancuso – Thales Systèmes Aéroportés
2 avenue Gay-Lussac – 78851 Elancourt Cedex – France
Tel +33 (0)1 34 81 60 00 – www.thalesgroup.com

***Abstract* - In the framework of the Smartpower project, the technical requirements are key input parameters to develop the technologies and materials that would be necessary for the packaging and thermal management of high power RF modules used in Active Electronically Scanned Array based on GaN transistors. These essential requirements are linked to the very demanding environment constraints for airborne applications and the necessities linked to an optimal electrical RF performance for the packaging. The major mechanical and thermal architecture requirements of a liquid cooled antenna are provided to detail the main characteristics expected from a highly conductive Thermal Interface Material which has to be developed and also the geometrical requirements expected from a low-cost compact packaging which should include industrial constraints for manufacturing. Thermal management requirements linked to the integration of a thermal sensor into the packaging are explained in parallel to the management of power dissipation coming from a GaN high power amplifier. The outcome of the technologies developed by the partners of the consortium will be integrated into a demonstrator that is presented.**

I. INTRODUCTION

Smartpower project aims to develop new technologies and materials for packaging and thermal management to achieve an efficient and cost-effective implementation of GaN-based power modules for RF transmitter systems. Among the different purposes of packaging for electrical or electronic equipment, a major one is to provide protection from physical damage, mechanical forces, and atmospheric or chemical contamination that may exist in a typical operating environment of airborne applications.

Packaging has been defined in a broad sense to, not only include the materials and technologies required to provide electronic components with physical protection and electronic connections, but also include system architecture and partitioning approaches, power management, thermal management, data flow & timing, and input/output interfaces.

Smartpower explores the potential of extending the performance of miniaturized electronic applications through the use of Wafer-Level Packaging techniques.

This technology has already been evaluated for digital applications, but the Smartpower project will enable it to host the more fragile chips required for RF and MM wave applications (i.e. GaAs, SiGe or GaN semiconductors with high dissipated power densities).

This paper is expected to give a detailed description of some of general requirements necessary for material and technology development linked to RF module packaging and thermal management of the chips with high power densities.

II. GENERAL DESCRIPTION

For a better clarity, it is necessary to first introduce a general description of Front End part of an Active Electronically Scanned Array where high power RF modules are located. In Fig. 1, 4 main elements can be seen. The RF transmitters are duplicated to reach the necessary size for the antenna, and they are brazed on a Power Control Board (PCB). The sub-assembly is then attached to a cold plate (attachment not shown). This cold plate has 2 main functions, provide a mechanical rigidity to the PCB with its modules, and cool down the electrical components which are inside the RF modules. The thermal link between the modules and the cold plate is obtained through a thermal Interface Material, usually referenced with the number 2, since it is outside the electronic packaging.

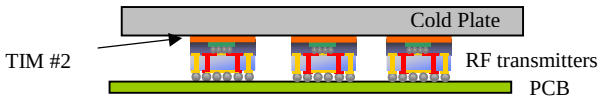


Fig. 1. General description of Front End of AESA

Typically for airborne applications, the cold plate is made of an aluminum alloy. The cooling is generally obtained with liquid circulation such as Polyalphaolefin (PAO), inside channels using corrugated structures or fins to create a turbulence and therefore obtain high heat transfer coefficients which are usually between 6000 to 8000 $W/m^2.K$ depending of the geometry of the corrugation, the fluid used and the other general parameters such as fluid velocity, fluid temperature etc.

The RF module is mainly composed of 2 RF GaN components: a driver and a High Power Amplifier (HPA) which are used to amplify the input signal. A thermal sensor is also part of the module to measure a reference temperature.

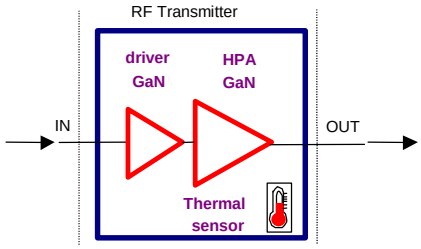


Fig. 2. Schematic of the RF Transmitter with packaging

It is well known that GaN component performance is temperature dependant. Therefore amplitude and phase of the RF signal have to be corrected based on the temperature of the components. Usually, an accurate correction would use the junction temperature i.e. the temperature of the transistor. However, this temperature can not be directly measured therefore, the thermal sensor measures an average temperature of the packaging/module, and this temperature is used indirectly to apply corrections.

III. ENVIRONMENT

Regarding the environment, the electronic packaging is submitted to sever constraints, it is expected that developed packaging should withstand a given level of environmental solicitations without failure (detailed specifications of the environmental conditions can be found in the reference [7]). A representation of the external interfaces is provided through the Fig. 3. Even if experience is a major advantage, an analytical approach through Design Failure Mode and Effect Analysis can be required. In this representation the packaging has been considered as a system, whereas PCB and cold plate have been considered as external elements to the system. In this chapter the constrains are therefore based on experience, standards and analyses.

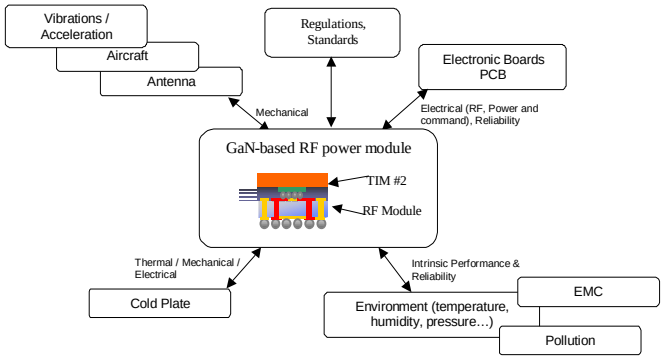


Fig. 3. External Interfaces

For size, volume and weight even no target is provided, it is expected that the solution will be as compact as possible. The only constrain is that the packaging must fit into the mesh size of the antenna which is typically 14 mm for X band application.

Regarding material selection, European regulations [1] and [2] are applicable and therefore have to be considered.

For simplification, the constrains are often described as test procedures or experimental validations that the packaging should pass. Tests are often conducted on passive daisy chain test vehicles, in a first approach to reduce the complexity of the test and verify the electrical continuity of the signal before and after the test to detect reliability failure. Typically the packaging must sustain several tests that are both representative of the external environment seen by the packaging during a normal life cycle, and the industrial process or have been proven as relevant to reach required reliability for considered applications. Main tests are provided below:

- 500 thermal shocks between ambient temperatures from –55°C to +125°C
- High temperature storage during 1000 hours at an ambient temperature of +125°C,
- Moisture & temperature during 1000 hours at an ambient temperature of 85°C with 85 % of Relative Humidity (RH) or Highly Accelerated Stress Test (HAST) at an ambient temperature of 110°C with 85 % RH according to [3].
- Package Integrity Check after three solder reflow simulation (double side bonding + a repair operation simulation).

IV. THERMAL MANAGEMENT

Thermal management is one of the key challenge for the development of a new innovative packaging. The dissipated power by the HPA and the driver have to be removed by minimizing the temperature gradient or temperature difference that will be created between the junction temperature of the die and the temperature of the cold plate $\Delta T_{j\text{-}cp}$. The challenge is usually difficult to master since many parameters are involved in the calculation of the temperature gradient. It includes the die itself ΔT_{die}, the packaging ΔT_p, the TIM #2 ΔT_{TIM}, and the cold plate ΔT_{cp}. The overall equation (1) is given below.

To manage this equation, and provide only essential information to Smartpower partners that will develop technologies and materials. The problem is further simplified to the calculation of the terms ΔT_{TIM} and ΔT_p.

$$\Delta T_{j\text{-}cp} = \Delta T_{cp} + \Delta T_{TIM} + \Delta T_p + \Delta T_{die} \qquad (1)$$

ΔT_{cp}: With liquid cooling, based on the experience on existing system, it is assumed that the performance of the cold plate can maintain its external skin temperature, in the worst case scenario, at 85°C.

ΔT_{die}: This is probably the most difficult number to calculate for GaN components. Recent publications from the Thirteenth InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems “ITherm 2012” can give a good overview of the latest techniques used to simulate and measure the junction temperature of high electron mobility transistors (HEMTs)

transistors [4],[5] and [6]. GaN on SiC substrate HEMTs for X band application are still under-development. Maximum junction temperature T_{jmax} is not fully known today since reliability studies are on-going. Usually die manufacturers guaranty from 10^5 to 10^6 hours for this temperature. For the moment it is commonly assumed that T_{jmax} will be in the range from 175°C to 200°C. The value of ΔT_{die} is currently calculated with existing thermal model of the GaN HPA around 55°C for a peak power density given in Table 1 and for long pulses. This result also assumes that the packaging thermal performance has limited influence on the value ΔT_{die}. This approximation is correct if the thermal conductivities of different material layers from the die are considered at the most critical temperature.

From this analysis it can be concluded that the temperature gradient from the TIM and the packaging should not exceed the value from equation (2).

$$\Delta T_{TIM} + \Delta T_p \leq 35°C \text{ to } 60°C \qquad (2)$$

For the correct estimation, through simulations, of the 2 values ΔT_{TIM} and ΔT_p it is necessary to provide detailed information that can create the correct power flux at the bottom of the die. The problem is identical for both the HPA and the driver. In this paper the method is explained with the example of the HPA only.

A illustration of the power distribution on the top surface of a GaN HPA is provided in the Fig. 4. The power is dissipated mainly on the 2 stages of the transistors. As explained to re-create, the power flux at the bottom of the die, a simplified version of the thermal description of the die is necessary.

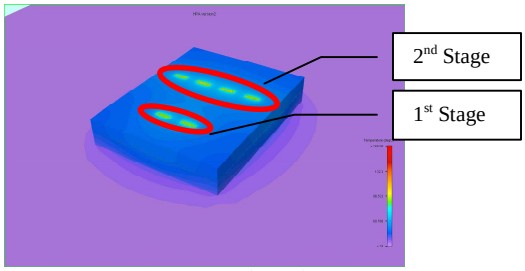


Fig. 4. Power distribution on the top surface of a GaN HPA (example)

The elements necessary for this description are illustrated in Fig. 5 and provided in the Table 1:

- the description of the transistor through its simplified geometry,
- the description of the different layers of the die and
- the power densities in the transistor.

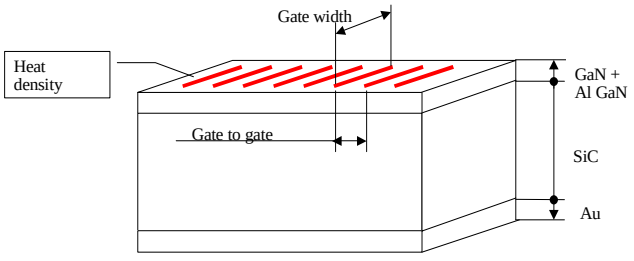


Fig. 5. 3D Cross Section of GaN die / Transistor

	HPA GaN - Thermal
Peak Power density	3 W/mm
Gate width	1st Stage = 8 x 75 µm 2nd Stage = 8 x 100 µm
Size of the dissipated surface area	1st Stage = 8 x 75 µm x 2.8 µm 2nd Stage = 8 x 100 µm x 2.8 µm Gate to gate distance = 40 µm
Peak Power dissipated by 1 transistor	1st Stage = 4 x 1.8 W 2nd Stage = 8 x 2.4 W
Peak Power dissipated by GaN Die	30 W 3.6 W dissipated out of the transistors
Material Thickness	GaN with Al = 2 µm SiC = 93 µm Au = 5 µm

Table 1. Thermal parameters for GaN HPA

Another important parameter is related to type of thermal calculation that should be performed: transient vs. permanent. Radar are working in a pulsed mode mixing emission and reception with a duty cycle and a pulse length. Usually the maximum duty cycle considered is below 25%. Therefore the average power dissipated by the HPA is reduced from the peak power of 30W to the value of 7.5W (Peak Power * DC).

From the calculation of thermal time constants (thermal resistance * heat capacity * mass) of the different materials, pulse length is short enough to avoid a calculation in transient outside the die. Therefore a good approach is to superpose the calculation in the die with transient simulation (using peak power) with a study state calculation in the packaging (using average power). However this superposition method has to be verified based on pulse length and thermal time constants of the different material used in the packaging.

From the methodology, it can be easily deduced that the thermal sensor integrated into the packaging can not measure the junction temperature of the die, which is located in the transistor close to the gate. The sensor measures, what is considered as a study state reference temperature (of the packaging) in comparison to the junction temperature. Therefore the performance of the HPA, which is temperature dependant, can only be corrected from an indirect measure. However thermal time constant should be fast enough (<2s) to capture thermal variation of the packaging and precise enough (<2°C) in the overall temperature range from –55°C to +150°C to apply the performance corrections.

V. THERMAL INTERFACE MATERIAL #2

The specifications of the TIM must be compatible with the general requirement from reference [7] and the specification defined for the packaging in this paper. However it is also

important to highlight the requirements that as often considered as secondary but are vital from an industrial point of view and for system maintenance.

The TIM #2 from Fig. 1 is used as a thermal interface between mechanical structure (cold plate) and the packaging (RF module). The purpose of the TIM #2 is to provide an efficient thermal interface between the module and the cold plate plus compensate for mechanical dispersions. TIM #2 is generally referred as a gap-filler.

From an equipment point of view, we need to have the RF modules soldered onto a unique PCB. The different modules have a height which varies according to assembly processes. The interval of tolerance (IT) is usually ± 10% of the nominal height.

Therefore TIM #2, in addition to a good thermal performance, must compensate for the mechanical dispersion induced by the fabrication process and must completely fill in the air gap during the mounting operation (refer to Table 2).

	TIM #2
Thickness (BLT)	0.2 to 1 mm
Thermal Conductivity	mini 15 W/m.K
Disassembly	Possible
Volume Resistivity	> 10^2 MΩ.m

Table 2: TIM #2 Main characteristics

When a reliability study is performed on the front end part of current antenna, the results show that Mean Time Before Failure (BTFM) of a simple RF module is not sufficient to cover the total system life time. Therefore it is assumed that it will be necessary to replace part of them during maintenance operations. The design of the front end has to take into account this requirement: TIM #2 should allow the disassembly of the cold plate from the RF modules without destroying the modules or the cold plate.

Operation of disassembly is expected to be performed at room temperature but specific tooling are possible.

The list of standard requirements for industrialization are provided below:

- The TIM implementation should be compatible with standard industrial deposition methods (syringe, stamping …).
- The cleaning agent, for the surface preparation, must respect the environment. There must be in accordance with the norm ISO14001 (for example isopropanol).
- The preferable storage conditions should not exceed a temperature lower than -25°C. Products having ambient temperature (20°C) or cold temperature (5°C) storage conditions are preferred.
- Each product will have a shelf life clearly identified. This shelf life must be greater than 6 months and products with a shelf life greater than 12 months are preferred.

VI. DEMONSTRATOR

A functional demonstrator will be designed and manufactured by SMARTPOWER partners to evaluate the performance of the different developed technologies (refer to Fig. 6).

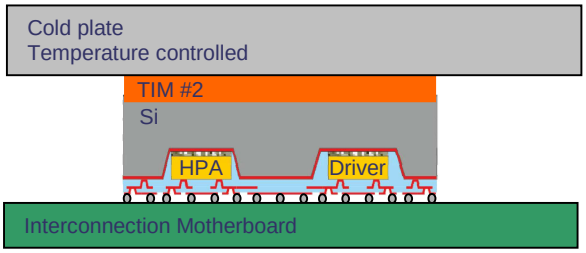


Fig. 6. Schematics of the Demonstrator

The demonstrator will include the RF module packaged by IZM using a Wafer-Level Packaging Technique according to the concept described in Fig.2. Similar technology has already been considered by Thales in the past using organic packaging [8]. The module will be eventually balled as a micro-BGA and soldered to an interconnection motherboard (PCB) designed in order to allow the test of the different functions of the amplification X band module and measure the imbedded temperature sensor.

The demonstrator will be assembled to a cold plate with the TIM #2 from Chalmers. The temperature of the cold plate will be controlled externally to simulate the different temperatures seen by the system.

ACKNOWLEDGMENT

The authors wish to thank UMS which is in charge of developing and manufacturing the X-Band GaN HPA, for providing all necessary data.

REFERENCES

[1] Regulation No 1907/2006 – "Registration, Evaluation, Authorization and Restriction of Chemicals (REACH)"

[2] ROHS 2002/95/EC – "Restriction of the use of certain Hazardous Substances in electrical and electronic equipment"

[3] JESD22-A110-B – "Highly Accelerated Temperature and Humidity Stress Test"

[4] Joyce H. L. Ling, Andrew A. O. Tay, Kok Fah Choo and Weiguo Chen "Thermal Characterization and Modelling of a Gallium Arsenide Power Amplifier MMIC" *ITherm 2012*

[5] AmirHossein Aminfar, Elah Bozorg Grayeli, Mehdi Asheghi, Kenneth E. Goodson "Analysis of HEMT Multilayered Structures using a 2D Finite Volume Model" *ITherm 2012*

[6] Tapan G. Desai, Daniel Piedra, Richard Bonner, Tomas Palacios "Novel Junction Level Cooling in Pulsed GaN Devices" *ITherm 2012*

[7] RTCA DO-160G – "Environmental conditions and test procedures for airborne equipment" – Dec. 2010

[8] J. Galière, J.L. Valard, E. Estèbe "Millimetre-wave MMIC packaging compatible with surface-mount technology (SMT)" *12th GAAS® Sysposium Amsterdam 2004*

25-27 September 2012, Budapest, Hungary

Room Temperature Transfer of Carbon Nanotubes on Flexible Substrate

Di Jiang[a], Yifeng Fu[b], Yan Zhang[c] and Johan Liu[a,c]

[a]Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-412 96 Göteborg, Sweden
[b]SHT Smart High-Tech AB, Fysikgränd 3, SE-412 96, Gothenburg, Sweden
[c]Key Laboratory of Advanced Display and System Applications & SMIT Center, School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200072, China

Mailing address: Johan Liu, Kemivägen 9, Göteborg, 41296, Sweden
Telephone: +46 31 772 30 67
e-mail address: johan.liu@chalmers.se

Abstract- **In this paper we report a novel method of transferring thermally grown vertically aligned carbon nanotubes (VA-CNTs) onto flexible substrates at room temperature with a single-step process. The transfer process is carried out by placing the CNT forests upside down on a double sided thermal release adhesive tape and peeling off the silicon substrate. Scanning electron microscope (SEM) is used to observe the transfer results. Also a second transfer using the same method but a thermal tape with higher release temperature is repeated on the as-transferred CNTs forests. The results show that this method is able to provide a novel process for transferring CNT forests at room temperature. This process will help to bring close the low cost fabrication of vertically aligned CNT structures for electronics.**

I. INTRODUCTION

Carbon nanotubes (CNTs) have been envisioned as one of the most promising engineering material due to their excellence in thermal[1], mechanical[2] and electrical[3] properties. Recently, interconnect bumps made of CNTs have been studied as an alternative of traditional metal interconnect bumps, due to the excellent properties of CNTs. Among all the CNT production methods, thermal chemical vapor deposition (CVD) has the advantage of flexible patterning by photolithographically defining the growth catalyst metals, and therefore is most widely adopted for producing functional CNT structures.

However, during the thermal CVD process used to produce CNTs, high temperature above 500°C is commonly required. This particular feature is not compatible with most of the existing semiconductor processes and materials, thus hinders the integration of CNTs with existing technology. Solders[13] and polymers[14]-[15] have been proposed to transfer the grown CNTs from their original substrate onto the target one. Yet the reflow of solder or cure of polymers still requires heating and cooling of the whole transfer system. In this paper, we propose a single-step room temperature transfer process of the VA-CNTs, which takes the advantage of commercially available thermal release adhesive tapes and thus reduces the processing temperature. We also demonstrated a second transfer of VA-CNTs using two thermal release tapes with different release temperature.

II. EXPERIMENTAL

In the experiment, VA-CNT forests are grown on a silicon substrate with 10nm-Al2O3/1nm-Fe barrier-catalyst layer patterned by photolithography and e-beam evaporation. Next, the substrate is placed in an oven at 700°C with flows of 150 sccm C2H2 and 650 sccm H2 at the pressure of 4.6 mbar. The growth time varies from 1 min to 5 min, depending on the desired CNT height. The whole growth process is carried out in a commercially available CNT growth machine (Black Magic, Aixtron).

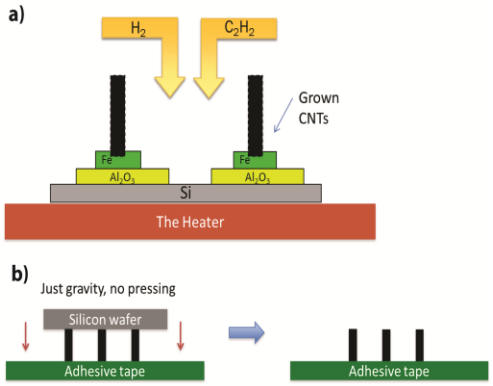


Figure 1. Illustration of the transfer process.

After the growth of patterned VA-CNT forests, the substrate with grown CNTs are placed up-side down on a double sided thermal release adhesive tape, as illustrated in Figure 1. The CNTs substrate is kept on the tape for 30 seconds before peeling it off. The CNTs are thereby transferred onto the adhesive tape with minimum damage of the CNT forest structures. It should be noted that before putting the transferred VA-CNT forests into SEM chamber

for examination, a thin layer of 5nm Au was sputtered onto the whole sample, in order to prevent the cure of polymers on the adhesive tape. This cure process is introduced by the bombardment of electron beams in the SEM machine and can destroy the adhesive tape surface flatness.

III. RESULTS AND DISCUSSION

As we can see from Figure 2(a), an array of 100um diameter cylindrical VA-CNT forests are transferred by this simple yet effective method. A 100% transfer rate is achieved with most of the VA-CNT forests kept in good shape. No significant deformation or damage is observed after the transfer process. A close examination of as-transferred individual VA-CNT forest in Figure 2(b) clearly shows that the VA-CNTs remain in a vertically straight shape, comparing with the just-grown VA-CNTs before the transfer. Smaller sized VA-CNTs (diameter 40μm) are also tested in our experiment and the results are shown in Figure 2(d).

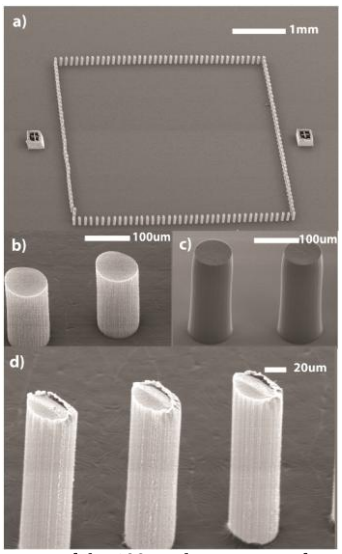


Figure 2. a) SEM picture of the 100um diameter transferred VA-CNT forests array. b) SEM of the individual VA-CNT forests within the array. c) SEM picture of the original VA-CNT forests. d) SEM picture of the 40um diameter transferred VA-CNT forests.

Further examinations were performed on the interface between transferred VA-CNTs and adhesive tape. In Figure 3(a), it is clearly shown that the VA-CNTs are well wetted by the polymer layer on the adhesive tape. When peeling the VA-CNT forests off the silicon substrate, this well-wetted strong bond is believed to produce much stranger force than the Van de Waals force between the silicon substrate and the CVD-grown VA-CNTs, which leads to a pressure-free transfer process as described in the experimental part.

We also observed that the peeling of VA-CNTs produced irregular leftovers on the silicon substrate. As shown in Figure 3(b), a small portion of the VA-CNT tips are left on the silicon after the transfer process. And the distribution of such leftovers seems to be uneven across the CNT-silicon interface. An explanation of this phenomenon is that the manual peeling operation introduced non-uniform peeling force during the transfer. This can be proved by the slightly deformed VA-CNT tips shown in Figure 2(b) and 2(d).

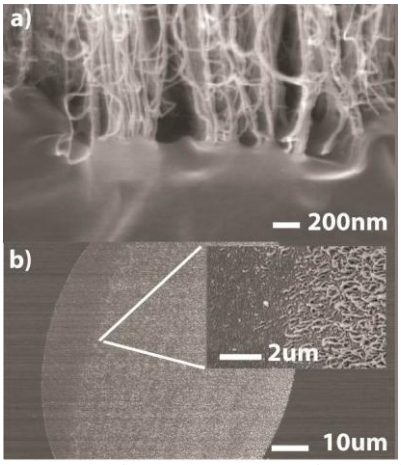


Figure 3. a) High magnification SEM picture of the interface between transferred VA-CNT forest and the adhesive tape. b) The leftovers on silicon substrate after transfer process.

Next, we performed another transfer of the as-transferred VA-CNT forests taking advantage of the difference in thermal release temperature of two adhesive tapes. As illustrated in Figure 4(a), the CNTs were transferred from a 90°C-release adhesive tape to a 120°C-release tape. The transfer tapes were heated to 90°C so that the low temperature release tape lost its adhesion while the high temperature one remained adhesive.

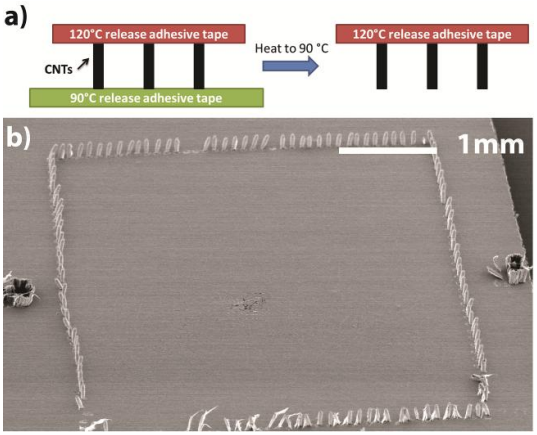


Figure 4. a) Illustration for the second transfer process. b) SEM picture of the 100um diameter transferred VA-CNT forests array after the second transfer process.

It can be found in Figure 4(b) that the quality/result of the second transfer is much worse than that of the first one. Although most of the VA-CNT forests are transferred onto the second adhesive tape, the structures of the forests are severely deformed. However, an interesting and unique finding was observed, as shown in Figure 5, a certain level of densification can be found on both the roots and tips of the transferred VA-CNT forests. This densification phenomenon is similar to those in the previously researches [16–18]. We can estimate from the SEM picture that the VA-CNT forest is densified to about 50% of its original diameter. The cause of this densification can be attributed to the volume shrinkage of the adhesive polymer during thermal release or curing

process, which is the nature of most polymer-based materials. In this case, the VA-CNTs followed obviously the shrinkage of adhesive polymers during the heating process, which happened both on the tips and roots of the VA-CNT forests, therefore forming the spindle-like appearance as shown in Figure 5.

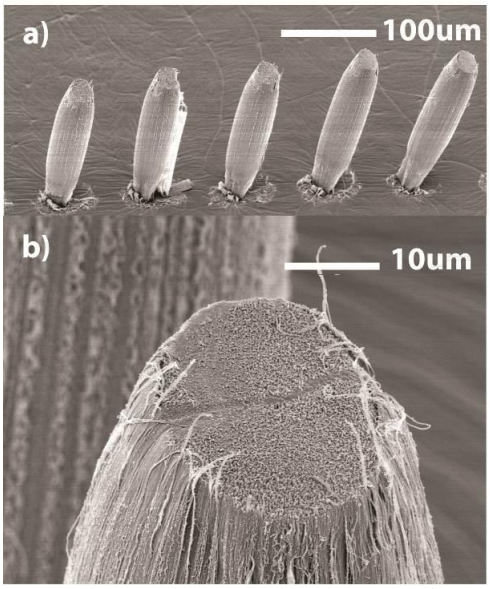


Figure 5. a) SEM picture of the result of the second transfer process. VA-CNT forests diameter is 40um. b) SEM picture of the densified VA-CNT forest tip after the second transfer.

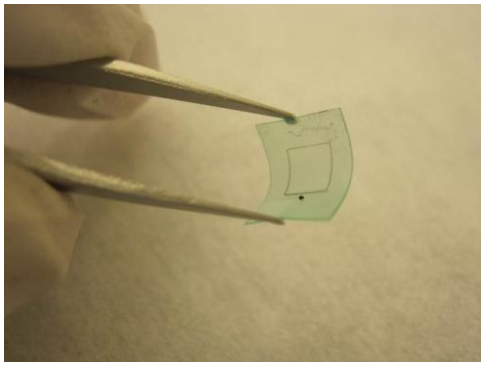


Figure 6. Transferred VA-CNT forests on a flexible substrate.

In principle, our room-temperature transfer process can be used on any material surface that the adhesive polymer applies. It is especially useful when transferring VA-CNTs on flexible materials (Figure 6). It has been tested in our experiment that the bending of the flexible substrate will not affect the integrity of the VA-CNT forests structures. This feature, together with the excellent mechanical and electrical properties of CNTs, reveals the possibility of integration VA-CNT into flexible electronics[19–21] at a very lost cost. It should also be noted that our transfer process does not require particular cleanroom environment and wafer processing technology. Therefore, it is highly possible to implement this process in a roll-to-roll fashion on a large scale, which would further reduce the overall cost of this technology.

IV. CONCLUSIONS

In this paper, we have demonstrated a novel process of transferring vertically-aligned carbon nanotubes onto a flexible substrate with a simple step performed at room temperature. The simplicity of this method enables the low cost fabrication of vertically aligned CNTs, which is desired in many electronics applications such as CNT based interconnect bumps. Due to the flexible nature of CNTs, the whole system is invulnerable to bending and twisting, and the whole transfer system can be designed in a roll-to-roll manner.

ACKNOWLEDGMENT

This work was supported by EU programs “Thema-CNT”, “Smartpower”, “Nanotec”, “Nanocom”, “Mercure”, and the Swedish National Science Foundation (VR) under the project “on-chip cooling using thermo-electrical device with the contract no: 2009-5042. This work was also carried out within the Sustainable Production Initiative and the Production Area of Advance at Chalmers. We also acknowledge the support from the Chinese Ministry of Science and Technology for the International Science and Technology Cooperation program of China (No.2010DFA14450). Thanks are also given to the Chinese governmental 02 program (2011ZX02602).

REFERENCES

[1] P. Kim, L. Shi, A. Majumdar, and P. L. McEuen, “Thermal Transport Measurements of Individual Multiwalled Nanotubes,” *Physical Review Letters*, vol. 87, no. 21, p. 215502, Oct. 2001.

[2] M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, “Exceptionally high Young’s modulus observed for individual carbon nanotubes,” *Nature*, vol. 381, no. 6584, pp. 678-680, Jun. 1996.

[3] S. Frank, P. Poncharal, Z. L. Wang, and W. A. de Heer, “Carbon Nanotube Quantum Resistors,” *Science*, vol. 280, no. 5370, pp. 1744 -1746, Jun. 1998.

[4] G. Che, B. B. Lakshmi, E. R. Fisher, and C. R. Martin, “Carbon nanotubule membranes for electrochemical energy storage and production,” *Nature*, vol. 393, no. 6683, pp. 346-349, May 1998.

[5] L. Hu et al., “Highly conductive paper for energy-storage devices,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21490 -21494, Dec. 2009.

[6] S. N. Cha et al., “Fabrication of a nanoelectromechanical switch using a suspended carbon nanotube,” *Applied Physics Letters*, vol. 86, no. 8, pp. 083105-3, Feb. 2005.

[7] J. E. Jang et al., “Nanoelectromechanical switch with low voltage drive,” *Applied Physics Letters*, vol. 93, no. 11, pp. 113105-3, 2008.

[8] M. Dragoman et al., “Nanoelectromechanical switches based on carbon nanotubes for microwave and millimeter waves,” *Applied Physics Letters*, vol. 90, no. 11, pp. 113102-3, Mar. 2007.

[9] J. E. Jang et al., “Nanoelectromechanical switches with vertically aligned carbon nanotubes,” *Applied Physics Letters*, vol. 87, no. 16, pp. 163114-3, Oct. 2005.

[10] E. Minoux et al., “Achieving high-current carbon nanotube emitters,” *Nano Letters*, vol. 5, no. 11, pp. 2135–2138, 2005.

[11] D.-ming Sun et al., “Flexible high-performance carbon nanotube integrated circuits,” *Nat Nano*, vol. 6, no. 3, pp. 156-161, Mar. 2011.

[12] K. Lee, Z. Wu, Z. Chen, F. Ren, S. J. Pearton, and A. G. Rinzler, “Single Wall Carbon Nanotubes for p-Type Ohmic Contacts to GaN Light-Emitting Diodes,” *Nano Letters*, vol. 4, no. 5, pp. 911-914, May 2004.

[13] A. Kumar, V. L. Pushparaj, S. Kar, O. Nalamasu, P. M. Ajayan, and R. Baskaran, “Contact transfer of aligned carbon nanotube arrays onto conducting substrates,” *Applied Physics Letters*, vol. 89, no. 16, pp. 163120-163120-3, Oct. 2006.

[14] H. Jiang, L. Zhu, K.-sik Moon, and C. P. Wong, "Low temperature carbon nanotube film transfer via conductive polymer composites," *Nanotechnology*, vol. 18, no. 12, p. 125203, Mar. 2007.
[15] T. Y. Tsai, C. Y. Lee, N. H. Tai, and W. H. Tuan, "Transfer of patterned vertically aligned carbon nanotubes onto plastic substrates for flexible electronics and field emission devices," *Applied Physics Letters*, vol. 95, no. 1, p. 013107, 2009.
[16] M. De Volder et al., "Diverse 3D Microarchitectures Made by Capillary Forming of Carbon Nanotubes," *Advanced Materials*, vol. 22, no. 39, pp. 4384-4389, Oct. 2010.
[17] T. Wang, K. Jeppson, and J. Liu, "Dry densification of carbon nanotube bundles," *Carbon*, vol. 48, no. 13, pp. 3795-3801, Nov. 2010.
[18] D. N. Futaba et al., "Shape-engineerable and highly densely packed single-walled carbon nanotubes and their application as super-capacitor electrodes," *Nat Mater*, vol. 5, no. 12, pp. 987-994, Dec. 2006.
[19] Q. Cao et al., "Medium-scale carbon nanotube thin-film integrated circuits on flexible plastic substrates," *Nature*, vol. 454, no. 7203, pp. 495-500, Jul. 2008.
[20] G. Gruner, "Carbon nanotube films for transparent and plastic electronics," *Journal of Materials Chemistry*, vol. 16, p. 3533, 2006.
[21] T. Y. Tsai, C. Y. Lee, N. H. Tai, and W. H. Tuan, "Transfer of patterned vertically aligned carbon nanotubes onto plastic substrates for flexible electronics and field emission devices," *Applied Physics Letters*, vol. 95, p. 013107, 2009.

25-27 September 2012, Budapest, Hungary

Graphene Heat Spreader for Thermal Management of Hot Spots in Electronic Packaging

Zhaoli Gao[1,2], Yong Zhang[1,3], Yifeng Fu[4], Matthew Yuen[2], Johan Liu[1,3]
[1] Department of Microtechnology and Nanoscience, Chalmers University of Technology
Kemivägen 9, SE-412 96, Gothenburg, Sweden
[2] Department of Mechanical Engineering, Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong SAR, China
[3] Key State Laboratory of New Displays and System Applications and SMIT Center, School of Automation and Mechanical Engineering, Shanghai University, 149 Yanchang Rd., Shanghai, 200072, China
[4] SHT Smart High Tech AB, Fysikgränd 3, Se 412 96, Sweden
Corresponding author: jliu@chalmers.se

***Abstract*-Monolayer graphene was fabricated using thermal CVD for the application of heat spreader in electronic packaging. Platinum (Pt) micro-heater embedded thermal testing chips were utilized to evaluate the thermal performance of the graphene heat spreader. The hot spot temperature was decreased by about 5 °C at a heat flux of up to 800W/cm2. It is possible to further improve the thermal performance of graphene heat spreader by optimizing the synthesis parameters and transfer process.**

I. INTRODUCTION

Thermal management of hot spots with high heat flux is critical in electronic packaging industry. Active solutions include thermoelectric [1, 2], enable site-specific and on demand cooling in electronic devices. For passive solutions without power consumption, a heat spreader is promising. Metal materials with high thermal conductivity, such as Cu and Al, are widely used to dissipate the heat from the hot spots, and graphene with ultrahigh thermal conductivity of 5300 W·m-1·K-1 [3, 4], has shown to be a promising heat spreader material. Recently, Yan et al. [5] reported the application of exfoliated graphene quilts (few-layer graphene) in the thermal management of the high-power AlGaN/GaN transistor. The temperature of the hot spot was reduced by ~20 °C, extending the transistor lifetime to one order higher. To further improve the thermal performance of the graphene heat spreaders, layer number control of the fabricated graphene is necessary, as the intrinsic thermal conductivity of few-layer graphene increases with reducing layer numbers [3]. Moreover, compared with the exfoliation methodology, thermal CVD is more feasible to fabricate large area monolayer graphene [6-8], allowing an enhancement of the thermal performance of the graphene heat spreader.

In this paper, application of the thermal CVD prepared monolayer graphene as the heat spreader is demonstrated. Pt was selected as the heating and temperature sensing material due to its excellent linearity of the electric resistance-temperature relationship up to 800 °C [9, 10]. The fabricated graphene was then transferred onto the thermal testing chips embedded with Pt micro heaters, acting as the hot spots. This was used to evaluate the thermal performance of the graphene heat spreaders. A micro-Raman spectrometer (RM3000, Renishaw) using laser excitation wavelength at 514.5 nm was employed to identify the quality and layer numbers of the graphene.

II. EXPERIMENT

A. Graphene Synthesis

A thermal CVD (Black Magic, AIXTRON Nanoinstruments Ltd.), as shown in Figure 1 (a), was used for graphene synthesis. The E-beam evaporated 1μm Cu thin films on SiO_2 substrates were cleaned with acetone, IPA and DI water. They were then transferred onto the heating stage with a thermal couple attached on the stage surface. C_2H_2 and Ar gas were chosen respectively as the carbon precursor and the gas carrier. Figure 1 (b) shows the synthesis profile. The reactor chamber was initially pumped down to a base pressure of 0.5 mbar. 1000 sccm Ar and 20 sccm H_2 were then introduced. The temperature was raised at a rate of 300 °C/min to about 900 °C and maintained for 5 min. 5sccm C_2H_2 was introduced at a flow rate of 5 sccm for 5 min. Finally, all the gas supplies were turned off, and the residual gas was pumped to a pressure of ~0.1 mbar. The reactor chamber was subsequently naturally cooled to room temperature with 1000 sccm Ar and 20 sccm H_2.

B. Fabrication of Thermal Testing Chip

The micro heaters and temperature sensors were made of titanium/platinum/ gold (Ti/Pt/Au) with a thickness of 20/150/30 nm. A lift-off process was used, and the thin films were deposited using an e-beam evaporator (PVD 255, Kurt J. Lesker Company). Pt was chosen as the temperature sensing material due to its remarkable temperature-resistance linear relationship. Meanwhile, it has a high melting point of around 1770 °C, thus it can be used as a micro heater when driven by electric current. Due to its designed dimension of 390 μm × 400 μm, it functions as a hot spot of the thermal testing chip with a dimension of 1 cm × 1 cm. The Ti acts as

the adhesion layer between the Pt and the SiO_2 substrate, and the Au layer on top can ease the soldering of the test chip to the power supply and sensing circuits. A 30 nm SiO_2 insulation layer was deposited on the testing chip using a sputter machine (MS150, FHR). The testing chips were calibrated with a standard resistance temperature detector (LABFACILITY PT100 XF-316-FAR) at various temperatures in the furnace. In order to eliminate the wiring and contact resistance contribution, the four point probe methodology was used here to characterize the electric resistance of the fabricated thermal testing chip. The Pt resistance can be described as

$$\mathbf{R(T) = \alpha T + R_0} \quad (1)$$

where α is the coefficient of electric resistance to temperature, and R_0 is the electric resistance at 0 °C. Graphene would then be transferred onto the testing chips utilizing the wet chemical process.

C. Graphene Transfer

A wet chemical process was adopted to transfer the graphene onto the thermal testing chip. Firstly, graphene synthesized on the Cu thin film was spin coated with a PMMA layer at 3500 rpm and baked at 110°C for 3 minutes. 30% FeCl3 solution was prepared for the Cu thin film etching. The PMMA layer with graphene was then floated in the FeCl3 solution, which was diluted with DI water. The PMMA supported graphene was transferred onto the calibrated thermal testing chip. Finally, the PMMA was dissolved in hot acetone, and the graphene was laid on top of the hot spot, acting as a heat spreader.

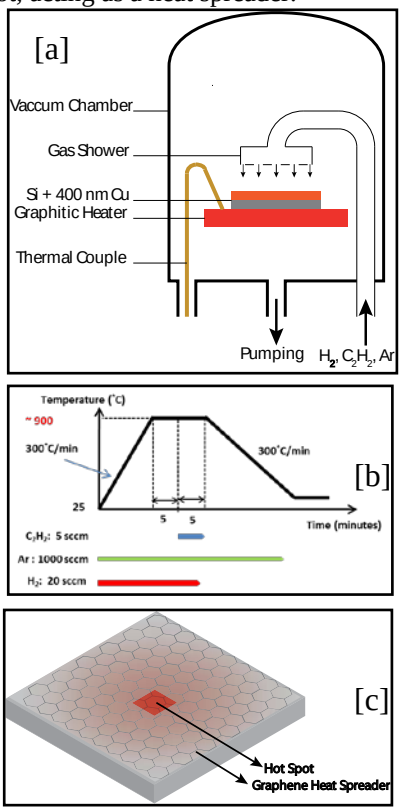


Fig. 1. (a) Cold-wall thermal CVD furnace for graphene synthesis, (b) thermal CVD process profile for graphene synthesis, (c) schematic of graphene heat spreader on thermal testing chip.

III. RESULTS

A. Graphene Synthesis

Raman spectroscopy was utilized for the graphene characterization. Due to the unique electron bands of graphene, Raman spectroscopy can clearly distinguish a monolayer graphene from few layers ones (n = 1~5) [11, 12]. As illustrated in Figure 2, the G band (~1588 cm^{-1}) and 2D band (~2697 cm^{-1}) are clearly identified. The low intensity of D band, located at ~1350 cm^{-1}, indicates high quality of the grapheme heat spreader with little defect. The inset shows the expanded 2D band, which indicates a monolayer graphene.

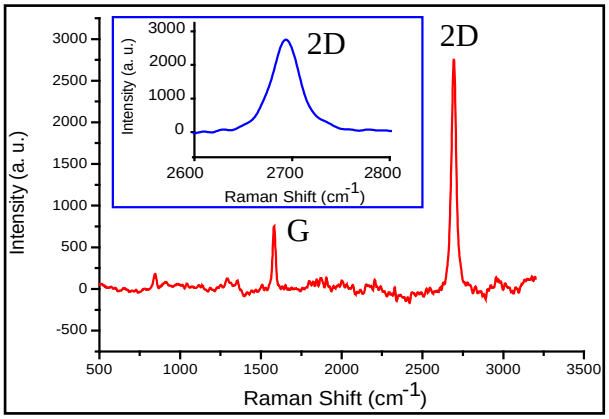


Fig. 2. Raman spectrum of synthesized graphene. Inset is the expanded 2D band indicating single layer grapheme.

B. Testing Chip Performance

Figure 3 (d) shows the calibration results of the Pt micro-heater with a dimension of 390μm × 400μm. Satisfaction linearity of the resistance-temperature relationship was obtained. Equation 2 shows the temperature-resistance relationship.

$$\mathbf{R = 0.113 \times T + 74.2} \quad (2)$$

To evaluate the thermal performance of the graphene heat spreaders, electric current was supplied to the Pt micro-heaters, generating heat flux. The electric resistance was measured in-situ, and then converted to temperature by Equation 2. A temperature drop of the Pt micro-heater can be observed at certain heat flux using the graphene as the heat spreader.

C. Thermal Performance of Graphene Heat Spreader

Figure 3(a), (b) and (c) show optical images of the transferred graphene on the thermal testing chip before and after the PMMA removal. The Pt micro-heater was fully covered by the graphene heat spreader. As shown in Figure 3 (e), the temperature of the hot spot was decreased by about 5°C, when the Pt micro-heater was driven by a heat flux of from 100 W/cm^2 to 800W/cm^2.

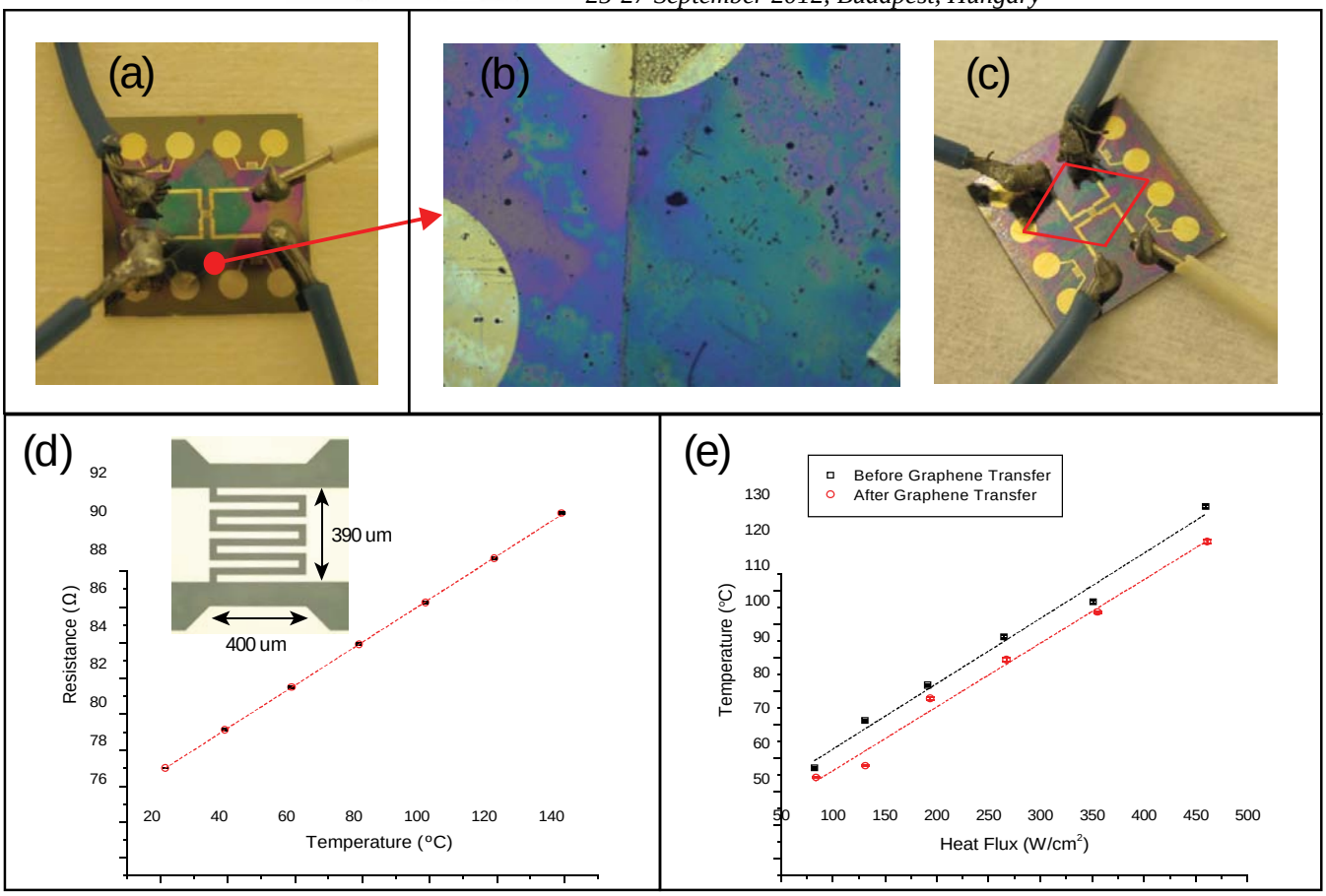


Fig. 3 Thermal performance of graphene heat spreader. (a) Thermal testing chip with PMMA supported graphene heat spreader attached onto the hot spot. (b) Optical image of the graphene heat spreader onto the testing chip. (c) Graphene on the heater testing chip after PMMA removing. (d) Electric resistance – temperature relationship of the Pt heater, inset is the optical image of the Pt heater. (e) Temperature – heat flux relationship of the Pt heater before and after the graphene heat spreader transfer.

IV. Conclusions

Monolayer graphene was demonstrated to be a promising heat spreader material in electronic packaging. The thermal performance of the graphene heat spreader was evaluated by Pt micro-heater embedded thermal testing chips. The hot spot temperature was decreased by about 5 °C with a heat flux of up to 800W/cm^2.

Acknowledgment

This work was supported by the EU programs “Smartpower”, “Thema-CNT", "Mercure", "Nanocom", "Nanoteg" and "Nanotec" and the Sustainable Production Initiative and the Production Area of Advance at Chalmers. The authors are grateful for the support of the Swedish National Science Foundation (VR) under the project "Thermoelectric Nanostructures for On-Chip Cooling" (2009-5042), Swedish Governmental Agency for Innovation Systems under the Eranet program "Nano TIM" as well as the financial support from the Chinese Ministry of Science and Technology international collaboration program (2010DFA14450). Thanks are also given to the Chinese governmental 02 program (2011ZX02602).

References

[1] I. Chowdhury, et al., "On-chip cooling by superlattice-based thin-film thermoelectrics," Nat Nano, vol. 4, pp. 235-238, 2009

[2] A. Bar-Cohen and P. Wang, "Thermal Management of On-Chip Hot Spot," Journal of Heat Transfer, vol. 134, pp. 051017-11, 2012

[3] A. A. Balandin, et al., "Superior Thermal Conductivity of Single-Layer Graphene," Nano Letters, vol. 8, pp. 902-907, 2008

[4] A. A. Balandin, "Thermal properties of graphene and nanostructured carbon materials," Nat Mater, vol. 10, pp. 569-581, 2011

[5] Z. Yan, et al., "Graphene quilts for thermal management of high-power GaN transistors," Nat Commun, vol. 3, p. 827, 2012

[6] X. Li, et al., "Large-Area Synthesis of High-Quality and Uniform Graphene Films on Copper Foils," Science, vol. 324, pp. 1312-1314, 2009

[7] Y. Fu, et al., "Templated Growth of Covalently Bonded Three-Dimensional Carbon Nanotube Networks

Originated from Graphene," Advanced Materials, vol. 24, pp. 1576-1581, 2012

[8] S. Jie, et al., "Low Partial Pressure Chemical Vapor Deposition of Graphene on Copper," Nanotechnology, IEEE Transactions on, vol. 11, pp. 255-260, 2012

[9] Y. Fu, et al., "A complete carbon-nanotube-based on-chip cooling solution with very high heat dissipation capacity," Nanotechnology, vol. 23, p. 045304, 2012

[10] Q. Wenmin, et al., "Development of multi-functional sensors in thick-film and thin-film technology," Measurement Science and Technology, vol. 11, p. 1111, 2000

[11] A. C. Ferrari, et al., "Raman Spectrum of Graphene and Graphene Layers," Physical Review Letters, vol. 97, p. 187401, 2006

[12] A. Gupta, et al., "Raman Scattering from High-Frequency Phonons in Supported n-Graphene Layer Films," Nano Letters, vol. 6, pp. 2667-2673, 2006

[13] J. H. Seol, et al., "Two-Dimensional Phonon Transport in Supported Graphene," Science, vol. 328, pp. 213-216, 2010

25-27 September 2012, Budapest, Hungary

Enhancement of the thermal properties of a vertically aligned carbon nanotube thermal interface material using a tailored polymer

H. Le Khanh[1,3], L. Divay[1], Y. Ni[2], P. Le Barny[1], E. Leveugle[1], E. Chastaing[1], F. Wyczisk[1], A. Ziaei[1], S. Volz[2], J. Bai[3].

[1] Thales Research and Technology, Campus Polytechnique, 1, Avenue Augustin Fresnel 91767 Palaiseau cedex France
[2] Laboratoire d'Énergétique Moléculaire et Macroscopique, Combustion, Ecole Centrale Paris, CNRS UPR288, PRES UniverSud Paris Grande Voie des Vignes, 92295 Châtenay-Malabry Cedex (France)
[3] Laboratoire de Mécanique des Sols, Structures et Matériaux, Ecole Centrale Paris, CNRS UMR8579, PRES UniverSud Paris Grande Voie des Vignes, 92295 Châtenay-Malabry Cedex (France)

Abstract- This paper presents our results on thermal properties measurements of thermal interface materials based on vertically aligned carbon nanotubes. In this system, the total interface resistance is the sum of thermal resistances created by the contact between the growth surface and the CNTs, the intrinsic resistance of the CNTs array, and the contact between the loose end of the CNTs and the opposite substrate. The latter is reported to be the limiting factor in the optimization of the global interface resistance, and in this work, polymers are used to enhance the thermal contact. Two polymers are studied. It is shown that the use of a polymer as interface material, despite its low thermal conductivity, leads to an improvement of the total thermal resistance. Experimental data and equilibrium molecular dynamics (EMD) show that the thermal contact resistance when using a reactive polymer – able to establish covalent bonds with the CNTs - is at least three times lower than that with the Van der Waals interaction polymer. The reactive polymer also enhances adhesion between CNTs and the polymer, which allows the mechanical complete transfer of the aligned CNTs array from growth substrate to a copper substrate. This transfer process allows to expand applications for VACNTs.

I. INTRODUCTION

The traditional TIMs are composite materials with thermal conductivities to the order of 1 to 5W/mK, and are used as film with thicknesses up to hundreds of micrometers leading to interfaces with thermal resistances larger than ~10^{-5} m²K/W.The development of new thermal interface materials (TIMs) is today a key element for the engineering of advanced thermal dissipation architectures.

Carbon nanotubes (CNTs) used as fillers seem to be good candidates to enhance the thermal conductivity of composite TIMs. CNTs are flexible, and the reported thermal conductivity of an isolated CNT is up to ~3000 W/mK [1], higher than that of diamond.. Indeed, thermal conductivities up to hundreds of W/mK were measured on aligned CNTs samples [2]. As the heat flow is perpendicular to the film plane in a thermal interface, vertically aligned carbon nanotubes (VACNTs) are required for a good heat transfer.

For a VACNT based TIM, one issue is to ensure a low interface resistance at the CNT-superstrate contact. This resistance is indeed identified as a major contribution to the total resistance of such an interface [3]. Several methods have been proposed to deal with this issue, such as the application of pressure on the interface [4]; or the thermocompression bonding, which uses a metal as an interface material between the CNT and the superstrate [5,6]. Even more sophisticated methods employing the covalent grafting of the nanotubes on the superstrate have been reported and give low resistance values [7]. The two last methods have the advantage to allow the development of a mechanical adhesion between the CNT and the superstrate.

To improve the contact between the CNTs and the superstrate, here we have used a heated polymer thin film as an adhesive layer at the CNT/superstrate interface. Two different polymers were studied. Polyethylmethacrylate (PEMA), which interacts with the CNT exclusively through Van der Waals forces, and HLK5, an azide functionalized polymer. Azide moieties are used to enhance thermal contact at the CNT/polymer contact owing to their ability to form covalent bonds with CNTs.

The samples are characterized by their thermal resistance using flash methods to measure their thermal resistances. The effective thermal conductivities of the VACNT and the contact resistances are deduced by varying the thicknesses of the CNTs array. The total resistance values are compared to the properties of samples without any interface material.

II. SAMPLE PREPARATION

The VACNTs are provided by SHT Smart High Tech. They are grown by thermal CVD process onto wafers with a 10nm Al_20_3 diffusion barrier layer and iron catalyst. Three series of samples with 10, 50 and 100µm heights and 1cm² surfaces are studied.

A. Azide functionalized HLK5 polymer preparation

The polymer HLK5 (fig. 1) is synthesized by esterification of the poly(vinylphenol) and 6-azido hexanoic acid. The IR and NMR spectroscopies confirm the structure of HLK5. NMR also allows to estimate that about 93% of side chains bear the azide function, so the grafting rate is about 93%. The glass transition temperature of HLK5 is measured to be ~4°C by differential scanning calorimetry.

ISBN:978-2-35500-022-5

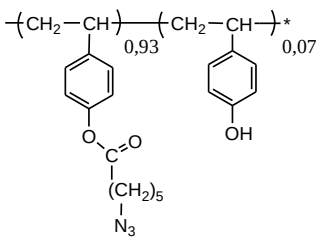


Fig. 1. The structure of HLK5 polymer

B. *Dry contact assemblies*

A first series of samples are obtained directly bringing into contact the VACNT with different CNT lengths on silicon with a 1cm² optical quality polished copper superstrate (fig. 2a).

C. *PEMA bonding*

The second series of samples involve the use of a thin polyethylmethacrylate layer at the CNT-copper contact (fig. 2b).

First, the PEMA thin films are spin-coated at 3000 rpm onto the surface of the Cu square from filtered 1,1,2-trichloroethane solutions. The PEMA concentration in the solutions is 80g/L. The polymer thicknesses are then measured using a profilometer and a value of 0,6µm is obtained.

The two surfaces : Si-CNTs and Cu-PEMA are then brought into contact. The sample is submitted to 66 kPa and heated at 120°C during 90 minutes. This temperature is higher than the glass transition temperature of the PEMA (66°C), allowing the polymer to flow and form a bonding layer between the CNT tips and the copper upon cooling.

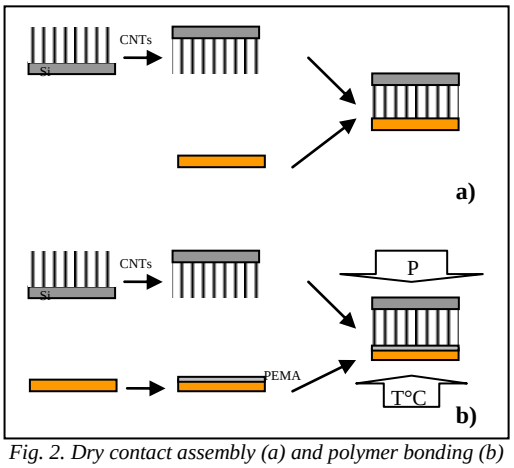


Fig. 2. Dry contact assembly (a) and polymer bonding (b)

D. *HLK5 bonding*

The third series of sample is realized with HLK5. The process is similar with PEMA bonding, but the HLK5 thicknesses are only 0,4µm. The glass transition temperature of the HLK5 is lower than that of the PEMA (4°C), so the sample is heated at only 100°C during 90 minutes. These conditions allow covalent bonds to be formed (fig. 3).

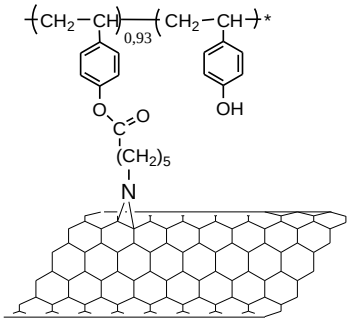


Fig. 3. The covalent bond between HLK5 and CNT

III. EXPERIMENTAL METHODS

The interfaces thermal properties are characterized by measuring the total thermal resistance of the multilayered samples with a thermal impedancementer (QuickLine-50 Thermal Interface analyser Anter Corporation) at room temperature.

The thermal impedancemeter allows the measurement of the complete thermal resistance of the sample excluding the resistances of the substrate (Si) and superstrate (Cu). The resistance value includes the intrinsic resistance of the TIM and the contact resistances at the substrate/material interface. By varying the thickness of the material under study, it is possible to extract the intrinsic conductivity of the material, providing the boundary resistances are not modified by the thickness variation. The dependence of the resistance with the thickness of the sample is then linear. The slope of the line is the inverse of the intrinsic thermal conductivity. The offset allows the determination of the resistances that are not affected by the thickness modification i.e. the contact resistances.

IV. RESULTS AND DISCUSSION

A. *Azide functionalized polymer-CNT interaction*

In order to verify the formation of covalent bonding during the interface closing with HLK5, an Auger Electron Spectrocopy (AES) study was conducted.

We synthesised another polymer : HLK9 (fig. 4a). HLK9 is very similar to HLK5, and in particular bears the same azide based functional groups as side chains. It is however based on a siloxane backbone. Main chain Si atoms can be used as AES probes on the surface of the CNTs. After mixing CNTs and HLK9 at 100°C for 90 min, the CNTs were taken out and washed extensively with dichloromethane (CH_2Cl_2). It is assumed that the repeated washing of the CNT allows the solubilization of the physisorbed HLK9. Therefore, remaining silicon can be attributed to HLK9 chain covalently bonded to the carbon nanotubes. Samples were washed in dichloromethane extensively until no HLK9 molecules were detected in dichloromethane by infrared spectroscopy. The AES surface survey on CNT and on the substrate (GaAs wafer) are presented in Fig. 4b. The presence of Si and S from the AES survey of CNTs in Fig. 4b indicated that HLK9 chains were connected to CNT through covalent bonding, thus HLK5

chains should have covalent bonding with CNT via the same mechanism.

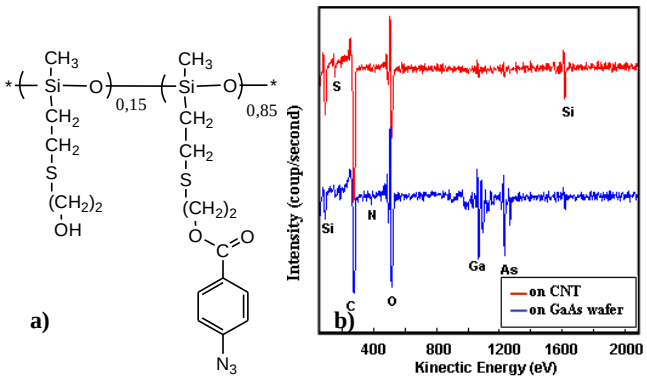


Fig. 4: The structure of HLK9 (a) and Auger electron spectroscopy surface survey on washed CNTs and GaAs wafer (b).

B. Thermal resistance comparison

The PEMA layer (0,6µm) is here thicker than the HLK5 layer (0,4 µm). From our recent work on the dependence of the resistance with the PEMA film thickness [8], we have been able to deduce the resistance values of 0,4 µm thick PEMA samples. This adjustment allows to compare the thermal resistance between PEMA bonding samples and HLK5 bonding samples.

The figure Fig. 5 shows the dependence of the measured thermal resistance as function of the length of the CNTs. For the "dry contact" series, values of 4.10^{-5} to 7.10^{-5} m²K/W are obtained. For the PEMA bonding series, values ranging from $2,1.10^{-6}$ to $1,15.10^{-5}$ m²K/W are obtained. And for the HLK5 bonding series, values of $1,4.10^{-6}$ to $9,13.10^{-6}$ m²K/W are obtained.

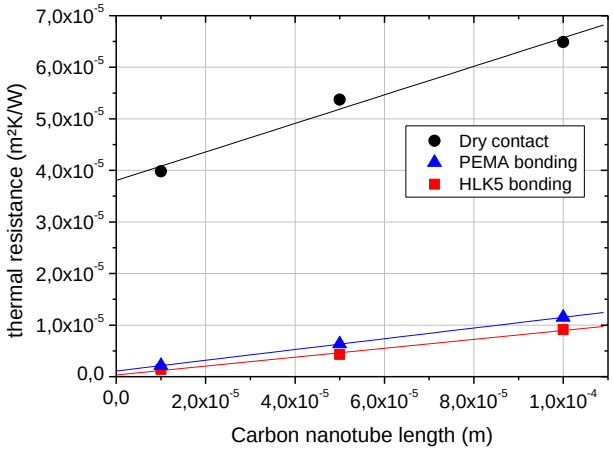


Fig. 5. Thermal resistance of the samples for dry contact and polymer bonding as function of the CNT length. The polymer bonding series contains a 400nm polymer layer as interface material.

The thermal resistance as a function of the CNT length is as expected found to be linear in all cases. The intrinsic thermal conductivity, as well as the interface resistances, can be extracted from the slope of the line and its offset respectively. The results give a ~4W/mK intrinsic thermal conductivity for the CNT turf, and ~$3,9.10^{-5}$ m²K/W for the interface resistances. This last value contains the resistance found at the CNT-Si contact and the CNT-Cu contact. The contact resistances values represent 38 to 95% of the total interface resistance depending on the CNT length.

For the PEMA bonding sample series, a value of ~10W/mK is found for the intrinsic conductivity of the VACNT forest, and $1,1.10^{-6}$ m²K/W for the contact resistances. It has to be pointed out that the contact resistance in that case contains contributions from the Si-CNT, CNT-PEMA and PEMA-Cu contacts, as well as the resistance originating from the PEMA layer itself. In this second series, the contact resistances represent 10 to 52% of the total resistance.

For the HLK5 bonding sample series, a value of ~12W/mK is obtained for the intrinsic conductivity of VACNT array, and the contact resistance is only $3,4.10^{-7}$ m²K/W. The values represent 4 to 24% of the total resistance for this series.

Considering these results, and as pointed out in a previous publication [9] it is confirmed that the addition of a polymer layer allows a decrease of the total resistance of the multilayered sample. Both an increase of the intrinsic thermal conductivity and a decrease of the contact resistance are observed. These results can be interpreted as follow :

For the dry contact samples, the effective contact surface is limited because of the length dispersion of the CNTs in the VACNT assemblies. Only the taller CNTs are in contact with the copper superstrate, the shorter CNTs being isolated with an air layer. As reported by Cola *et al.* [4], the application of pressure on the assembly allows a reduction of the thermal resistance of the sample [4]. Under pressure, the taller CNT bend and the shorter CNTs are able to contact the superstrate. This phenomenon has the effect to increase the effective contact surface at the CNT-superstrate interface thereby increasing both the thermal conductivity and decreasing the thermal contact resistance.

The addition of a polymer layer can as well increase that surface. The application of heat and pressure can allow the polymer to flow, and the CNTs to enter the polymer layer. In this manner, the polymer replaces the air film for the shorter CNTs, allowing a reduction of the total thermal resistance by the increase of the number of CNTs connected to the copper. The effective contact surface increase seems more beneficiary to the thermal properties than the addition of a low conductivity layer.

When PEMA layer is substituted by HLK5, the total resistance of sample is decreased. More than 25% decrease of total resistance were found in all the considered CNT lengths, indicating that optimized thermal paths are generated with CNT-polymer covalent bonding. The contact resistance values at zero CNT length with PEMA bonding and HLK5 bonding were found to be 1.1 mm²K/W and 0.34 mm²K/W, respectively. This contact resistance contains contributions of resistances from the Si-CNT, CNT-polymer, polymer-Cu contact, as well as the intrinsic resistance from polymer. As the polymer layer is quite thin and it fully contacts the copper surface, the resistance originating from polymer and that between polymer and Cu can be considered as negligible. As a result, the thermal interface resistance

between CNT and polymer should be the dominating factor to the thermal contact resistance. The thermal contact resistance between CNT-PEMA is 3 times as large as that between CNT-HLK5. This result can be attributed to the beneficial effect of covalent bonding to the low CNT/polymer thermal resistance.

This effect is confirmed by molecular simulation. Equilibrium molecular dynamic (EMD) simulations were conducted to predict the thermal contact resistances of CNT/PEMA and CNT/HLK5 [9]. The thermal resistance between CNT and one monomer of PEMA and that between CNT and one monomer of HLK5 are calculated as 15.0 and 4.92 GK/W, respectively. Which confirms that the factor of three decrease in the thermal contact resistance can be attributed to the chemical bonding of the polymer to the CNTs tips.

These covalent bonds have a simpler mechanical effect on this contact : the enhancement of the adhesion. Indeed, when submitting the assembly to mechanical traction, CNTs can be separated from the growth substrate and fully transferred to the copper substrate : The CNT/polymer adhesion is stronger than the CNT/silicon adhesion. As such transfer cannot be performed on the samples bonded with PEMA, it can be concluded that the covalent bonds play a role in this strong interaction.

The possibility to transfer the CNTs on new substrates has been studied previously by other teams using different bonding materials : metal layers [10] or sophisticated molecular self assembled monolayers [7,11]. This new polymer process (fig. 6) possesses the advantage over these techniques to offer the possibility to transfer the CNTs on metal surfaces without the need for vacuum techniques [10], microwaves irradiations [7], or prior chemical functionalization of the CNTs surface [11], while keeping a good thermal contact between the CNTs and the new substrate.

This process opens the way to a wider use of VACNTs as thermal interface materials as they can be transferred on surfaces that could not withstand the growth conditions of high quality carbon nanotubes. For example, VACNT could be transferred directly on a microelectronic device.

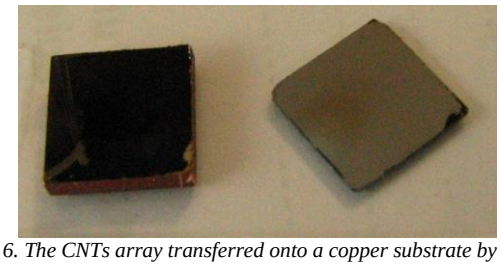


Fig. 6. The CNTs array transferred onto a copper substrate by HLK5 bonding process.

IV. CONCLUSION

We have demonstrated that the use of a polymer thin film as a CNT/superstrate interface material is shown to reduce the thermal resistances of the samples by as much as one order of magnitude compared to the direct CNT/superstrate contact. The lowest thermal resistance samples show a resistance as low as $1{,}4.10^{-6}$ m^2K/W with 10µm length VACNTs using HLK5. The CNTs length dependant thermal resistance measurements suggest that the polymer thin film allows both a reduction in the CNT/superstrate contact resistance and an enhancement of the VACNT intrinsic thermal conductivity. An interpretation would be to consider the polymer thin film as a material able to increase the CNT/superstrate contact surface thereby reducing the total resistance of the samples, despite the low intrinsic conductivity of the polymer. The thermal contact resistance when using azide-functionalized polymer with C-N bond is at least three times lower than that with the Van der Waals interaction polymer.

In addition, the azide functionalized polymer allows to transfer the CNTs array. This low-temperature transfer process allows to expand applications for VACNTs.

ACKNOWLEDGMENT

This work is supported in part by the European Commission (EC) 7th Framework Programme (FP7), under the Grant Agreement n° 288801 (SMARTPOWER).

REFERENCES

[1] Kim, P.; Shi, L.; Majumdar, A.; McEuen, P. L.; "Thermal transport measurements of individual multiwalled nanotubes" *Phys. Rev. Lett.* ; **2001**, *87*, 215502

[2] Hone, J.; Llaguno, M. C. N.; Nemes, M.; Johnson, A. T.; Fischer, J. E.; Walters, D. A.; Casavant, M. J.; Schmidt, J.; Smalley, R. E.; "Electrical and thermal transport properties of magnetically aligned single wall carbon nanotube films" *J. Appl. Phys. Lett.*, **2000**, *77(5)*, 666

[3] Cola, B.A.; Xu, J.; Cheng, C.; Xu, X.; Hu, H.; Fisher, T.S.; **"**Photoacoustic Characterization of Carbon Nanotube Array Thermal Interfaces"*J. Appl. Phys*., **2007**, *101*, 054313

[4] Cola, B.A.; Xu, J.; Fisher, T.S.; **"**Contact mechanics and thermal conductance of carbon nanotube array interfaces" *International Journal of heat and Mass Transfer*, **2009**. *52*, 3490-3503

[5] Johnson, R. D.; Bahr, D. F.; Richards, C. D. ; Richards, R. F.; McClain, D.; Green, J.; Jiao, J.; "Thermocompression bonding of vertically aligned carbon nanotube turfs to metalized substrates" *Nanotechnology*, **2009**, *20*, 065703

[6] Hamdan, A.; Cho, J.; Johnson, R.; Jiao, J.; Bahr, D.; Richards, R.; Richards C.; "Evaluation of a thermal interface material fabricated using thermocompression bonding of carbon nanotube turf" *Nanotechnology*, **2010**, *21*, 015702

[7] Lin, W. ; Zhang, R. ; Moon, K-S. ; Wong, C. P.; " Molecular phonon couplers at carbon naotubes/substrate interface to enhance interfacial thermal transport" *Carbon*, **2010**, *48*, 107-113

[8] Le Khanh, H.; Divay, L.; Le Barny, P.; Leveugle, E.; Chastaing, E.; Demoustier, S.; Ziaei, A.; Volz, S.; Bai, J.; "Thermal Interface based on Vertically aligned Carbon nanotubes : An Analysis of the Different Contributions on the Overall Thermal Resistance" *in 16th International Workshop on the Thermal Investigations of ICs and Systems*, Barcelona, Spain, 6-8 October **2010**.

[9] Ni, Y.; Le Khanh, H.; Chalopin, Y.; Bai, J.; Le Barny, P.; Divay, L.; Volz, S.; "Highly efficient thermal glue for carbon nanotubes based on azide polymers" *Appl. Phys. Lett.*, **2012**, *100*, 193118.

[10] Chen, M. X.; Song, X. H.; Gan, Z. Y.; Liu, S.; "Low temperature thermocompression bonding between aligned carbon nanotubes and metallized substrate", *Nanotechnology*, **2011**, *22*, 345704.

[11] Lin, W.; Xiu, Y.; Jiang, H.; Zhang, R.; Hildreth, O.; Moon, K-S.; Wong, C. P.; "Self-Assembled Monolayer-Assisted chemical Transfer of In situ Functionalized Carbon Nanotubes", *J. Am. Chem. Soc.*, **2008**, *130*, 9636.

25-27 September 2012, Budapest, Hungary

Thermal transient characterization of pHEMT devices

Zoltán Sárkány, Gábor Farkas, Márta Rencz
Budapest University of Technology and Economics
Department of Electron Devices
1117 Budapest, Magyar tudósok körútja 2.
sarkany@eet.bme.hu; farkas@eet.bme.hu; rencz@eet.bme.hu

***Abstract*- In this paper the issues of thermal transient measurement of heterostructure field-effect transistors are discussed. It is shown that the measurement setups used for MOSFET devices may produce artifacts. Even though the measured transients may seem to be thermal, at the calibration of the measured temperature sensitive parameter no temperature dependency is observed. The influence of the gate-lag effect is discussed, which is a plausible reason of the electric transients with a time constant in a few millisecond range.**

Finally a practical measurement setup is presented utilizing the forward voltage of the Schottky barrier gate as a temperature sensitive parameter.

I. INTRODUCTION

In microwave power electronics the compound semiconductor FETs such as GaAs and GaN based MESFETs and HEMTs play a very important role. High frequency amplifiers built from these devices are the heart of many wireless communication applications, e.g. in mobile phone base stations, space and military appliances where reliability is a key issue. For prolonged lifetime beside other parameters the device temperature has to be kept as low as possible. This can be realized by careful thermal design where the simulation results have to be supported by measurements.

Until recent years the measurement of thermal parameters has been realized by surface scanning methods like liquid crystal thermography [1] or infrared microscopy [2]. These methods are able to provide temperature map over the surface but the thermal resistance of the device under test (DUT) can only be calculated with limited resolution and the dynamic behavior cannot be investigated.

In case of common semiconductor components as diodes, bipolar junction transistors, MOSFETs and IGBTs the thermal transient measurement is a widely accepted test method with well-developed measurement setups for each type of these devices [3].

At a thermal transient measurement the heating (cooling) is induced by applying a power step on the device. This nearly always occurs by sudden change of electric quantities, voltage or current at certain device poles. Accordingly, we speak about voltage or current jump measurements.

As we know all semiconductor characteristics is temperature sensitive. That makes all characteristic curves depicting voltage versus current behavior shifting when the temperature of the device changes.

However, this voltages and currents cause a self-heating of the device. Using a curve tracer we shall experience different curve set when measuring in pulsed mode with (nearly) constant device temperature and when measuring at each voltage/current point until a thermal equilibrium (steady state) is reached.

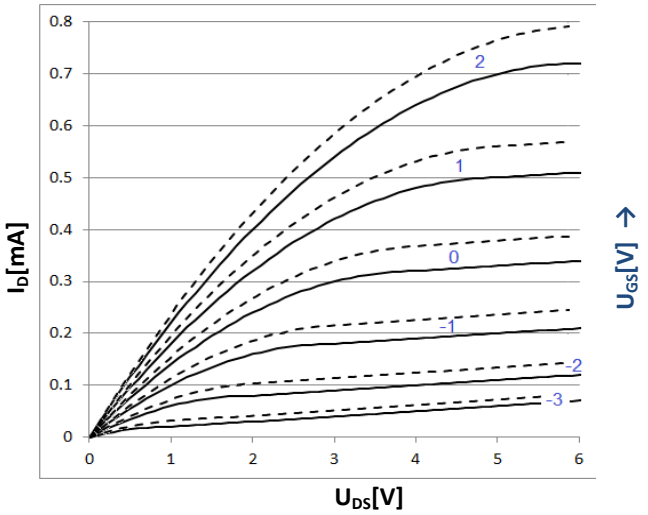


Fig.1 Output characteristics of a MOSFET, static (solid) and pulsed (dashed)

Fig.1 illustrates the effect of self-heating by presenting a set of drain current versus drain-source voltage curves of a MOSFET, each curve belonging to different gate-source voltage. The dashed curves represent results from pulsed measurements where the device is nearly at ambient temperature. The solid lines show the operating point when a long waiting time lets the device reach high temperature. (Fig.1 is an Excel sheet calculation of a rough model just illustrating the concept, not an actual measurement.) These effects are broadly treated in the literature such as [5][6].

Obviously, the static measurement represented by the solid lines depends not only on the applied voltage and current but also on the actual whole heat conducting path between the device and the ambient.

During a cooling transient measurement we move from a static hot characteristics set to a static cold characteristics set through a bunch of in-between cases, where power and cooling mounts dictate the pace of the change. Still, in most cases we can translate these effects to a “thermally sensitive parameter”, especially because we carry out a voltage-to-temperature calibration process at low power levels, where the device temperature and the ambient temperature are very near to each other disregarding the cooling path.

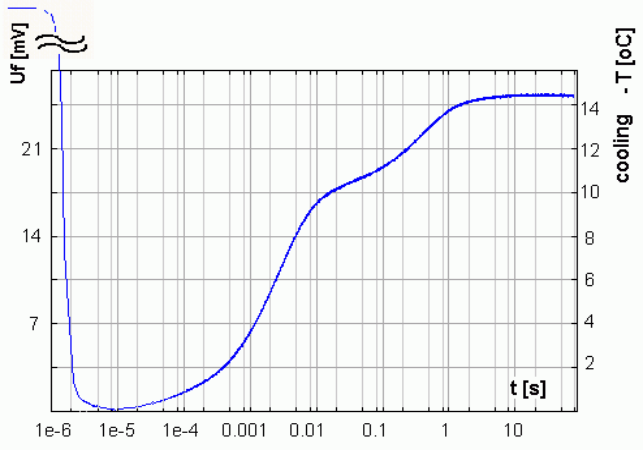


Fig.2 Recorded transient scaled in voltage change and temperature change

Fig.2 shows the result of the process discussed so far. We get a compound change of the selected parameter, as Fig.1 suggests seriously influenced by both electric and thermal causes. What makes life easier is that we can typically use the implicit assumption that the *electric* changes occur very fast while *thermal* changes take place in longer times.

In case of MOSFET devices with the characteristics shown in Fig.1 we can directly see one of the reasons of the early electric change, this is the channel length modulation. Moving along constant I_D current the change of the U_{DS} voltage causes a change in U_{GS}. This is a fast effect and so it can be incorporated into the correction of the early transient. MOS devices also show slow effects like the migration of charge particles on the oxide interface. However, these appear in the transients at very small amplitude, as such they are treated rather as low frequency noise than as signal.

If the temperature dependence of the parameter is nearly linear then we can scale our transient e.g. in voltage and temperature. In Fig.2 the electric and thermal portions of the transient can be well separated at 10 µs.

A popular measurement arrangement for transient measurement of MOSFETS is shown in Fig.3. The power step is induced by sudden change in U_{DG} or I_S, then U_{GS} is recorded.

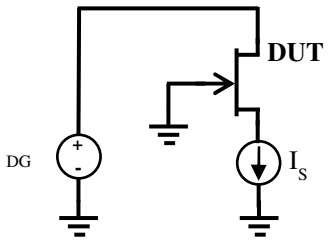


Fig.3 Commonly used n-MOSFET measurement setup

II. ARTIFACTS IN TRANSIENT MEASUREMENTS

All the measurement results presented in this paper had been measured on a GaAs based power pHEMT device fabricated on a semi-insulating GaAs substrate with 0.35 µm channel length and 8mm width.

The electrical behavior of HEMT and MESFET devices is similar to MOSFETs. Since the electrical characteristic of the device had not been published in the datasheet; supposing enhancement mode characteristics the common measurement setup of Fig.3 had been selected as a first test.

The power step was induced by applying a voltage step on U_{DG} and the cooling curve was recorded as the transient of U_{GS}. Two measured transients captured at different power levels can be seen in Fig.4.

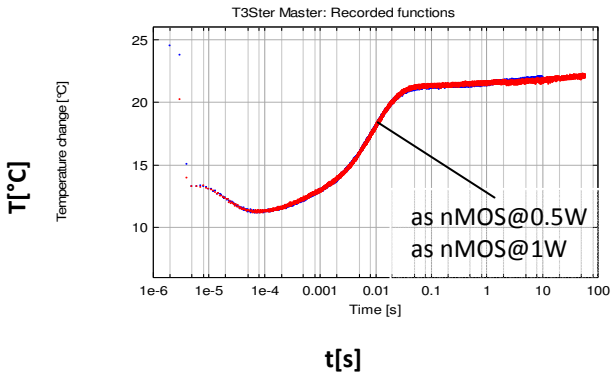


Fig.4 Transient measurement results in NMOS setup switching between two different power levels.

As it can be seen on the graph the thermal transient is quite short with moderate level of noise while the electric transient seems to end at around 80 µs. The quasi steady state is reached at around 0.1 s.

It has to be realized that the transient plots fit perfectly even though they were measured at different power levels which indicates that the power calculation or the interpretation of the measured transient given above is wrong. By measuring the parameters of the powering applied we found that there was no source current flowing through the device and with further measurements it has been shown that the DUT has depletion type characteristic. This proves that the nice curve recorded in this setup has non-thermal roots.

There is no general measurement method for depletion type devices though it is possible to set a bias point by applying a constant gate-source voltage (V_{GS}) and using the channel resistance (R_{DSON}) for measurement. For applying sufficient power on the device we found a suitable bias point at V_{GS}= -0.55 V and used the current jump method changing from 500 mA to 25 mA. This resulted in 1.3W power change. Repeating the current jump at V_{GS}= -0.6 we measured a power step of 1.7W in the same arrangement.

The measured transients seem to reach steady state before 0.1 s as shown in Fig.5. Unlike the previous measurement now the transients measured on R_{DSON} seem to be proportional to the applied power step.

The y axis in Fig.5 is scaled now in quasi-temperature, based on typical temperature sensitivity values of semiconductors. After the measurements seemed to yield valid thermal transients we attempted to calibrate the temperature-voltage correlation, which is normally monotonous; and (in most cases) can be approximated with a linear relation.

The calibration results can be seen in Fig.6. The points represented by crosses are the measured voltages

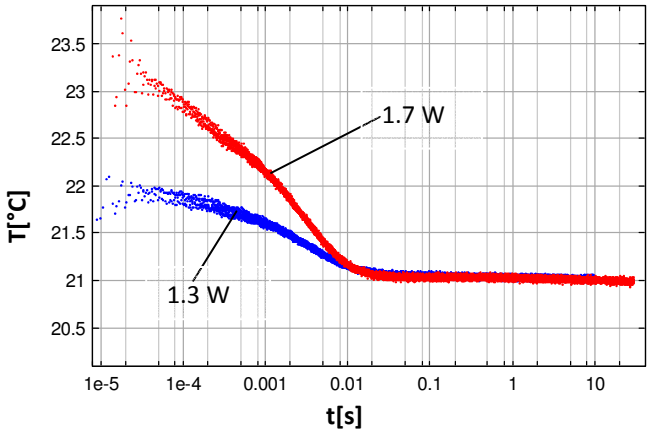


Fig.5 Transient curves measured on the channel resistance

corresponding to the temperature of the cold-plate the DUT is mounted on, in between we see the fitted regression line. The diagram shows that the change of the drain-source voltage is negligible, it is not temperature dependent.

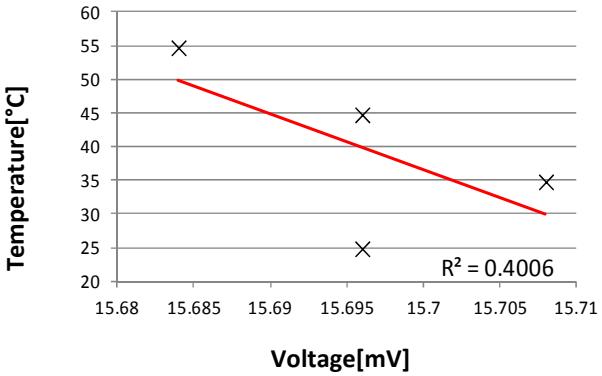


Fig.6 "Temperature sensitivity calibration" curve of the D-S voltage at constant current, k = 4.8e-7 mV/K

There are similarities between the transients in MOSFET setup and the ones measured on the channel resistance. On one hand the highest time constants are in the range of a few milliseconds. The time constant of the die of a MOSFET device falls into the same range but the time constant of the package must be much larger. On other hand it is proved that these are not thermal transients.

Generalizing our observations above; we can state that the length of the electric transients of discrete semiconductor devices is determined mostly by the parasitic capacitances, inductances and resistances. In most cases the whole switching time falls in the few microsecond range but it is rarely longer than 100 µs. Moreover HEMT and MESFET devices are designed to be used at high frequencies.

However they were designed to be used in small-signal application. In their articles [7][8] Kazushige Horio et. al. discussed the large signal behavior of GaAs based MESFET devices. They investigated the slow current transients which are often observed if drain or gate voltage changed abruptly which is called drain-lag (gate lag). In their article it can be seen that the time constants of this drain-lag are in the same range as our transients. Based on comparison of measurements and simulation results they showed that the drain-lag is the transient manifestation of the charge trapping.

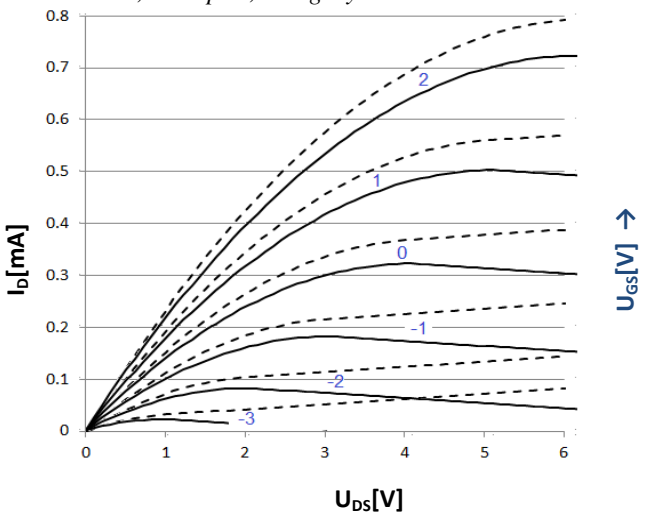


Fig.7 Output characteristics of a HEMT, static (solid) and pulsed (dashed)

Since we apply a sudden power step during the measurement which induces an abrupt voltage change the drain-lag effect is a probable explanation of the measured long electric transients.

At the measurement time we had no detailed data sheet of the HEMT which we measured. As these devices are used as high frequency amplifiers at one optimal operating point; typically only the gain and bandwidth parameters at this point are specified. A modeled characteristics in Fig.7 hints that besides transients caused by the drain-lag effect we have also problems with setting the proper operation point. Selecting a starting point on the negative slopes of the static characteristics can cause oscillations. Other well discussed experiments are presented in [4].

III. SOLUTION

After the unsuccessful measurement attempts, we decided to use the gate-source two-pole as Schottky diode for sensor element. The textbook scheme of the measurement setup is illustrated in Fig.8 . The behavior of the circuit can be decomposed into two steps.

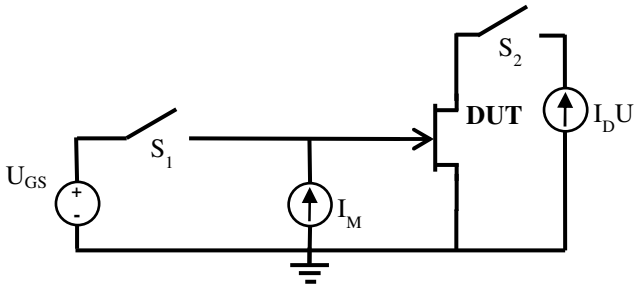


Fig.8 Ideal schematic for HEMT measurement

During heating switch S1 and S2 are closed. The transistor is switched on by proper gate-source voltage and the drain current flows through the device. The drain-source voltage is as it comes from the device characteristics. We selected an operating point at a stable section not causing oscillation.

During cooling both S1 and S2 are open. There is no drain current, and the forward biased gate Schottky diode works as sensor using a constant sensor current I_M.

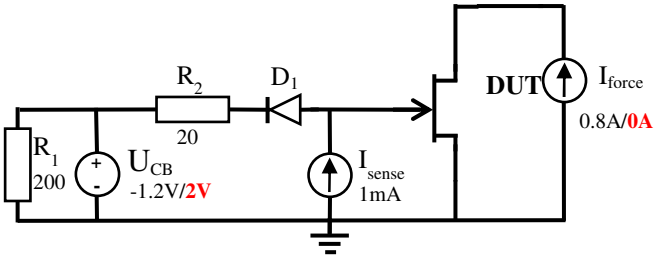


Fig.9 Schematic of the real measurement setup, for generators with two assigned values the first value applies for heating, the **second for cooling**

In the thermal transient measurement equipment S2 is realized internally as a programmed current source switching on and off. The voltage source can also switch between two voltage levels, as a consequence for realizing S1 only a diode had to be applied which behaves as shorted switch at negative gate voltage and as open circuit at positive, such not influencing the gate voltage in the latter case.

The real measurement setup can be seen in Fig.9 .

During heating D1 is forward biased by R2 (20Ω) and negative U_{CB} (-1.2 V), with e.g. 0.6 V forward voltage drop on the diode. The heating current I_{force} (800 mA) flows through the device. The HEMT gate is reverse biased, hence there is no gate current.

During the cooling D1 is reverse biased thus U_{CB} (which is positive) is separated. The HEMT gate is forward biased by I_{sense} (1mA) sensor current. I_{sense} must be much lower than allowed maximum Schottky current.

In Fig.10 the measured transient responses can be seen at two different power levels. The transient curves normalized by the applied power step fit perfectly (Fig.11). The temperature elevation is proportional to the applied power step. The calibration had also been carried out and the temperature sensitivity showed good linearity.

IV. Conclusion

In all fields of power electronics in order to increase the lifetime and reliability the thermal management is a key issue. Beside the simulations the thermal properties also have to be measured after the manufacturing. Thermal transient testing is a commonly used method for the characterization of discrete electronic components but the measurement results have to be treated carefully in case of special devices.

In this paper we presented two measurement setups for HEMT devices which yield transient measurement data that are deceptive. These results can be misinterpreted due to their similarities to thermal transients, long and strange transients have to be verified before we could believe them.

We also presented a measurement setup utilizing the gate Schottky diode as temperature sensor element. Using this setup the measured transient was not influenced by long electric transient, we were able to produce the real thermal transient which is characteristic to the heat flow path. This setup can be used for the thermal transient measurement of theoretically any type of HEMT and MESFET device.

Acknowledgement

This work was partially supported by the 288801 SMARTPOWER integrated project of the Framework 7 Program of the EU and JEMSiP_3D ENIAC JU project (No 120016) of NKTH (No OMFB-00166/2010).

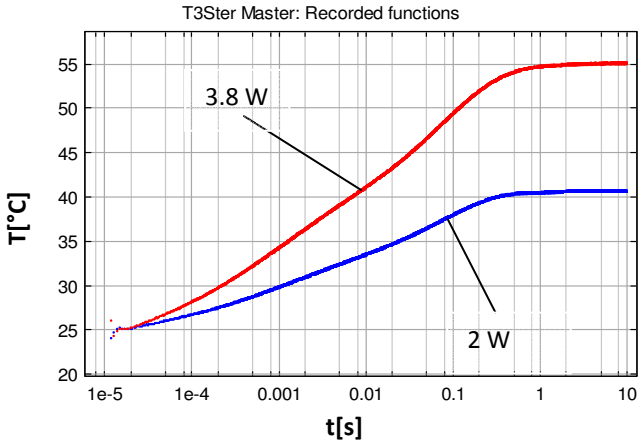


Fig.10 Measured transient using the gate Schottky diode as sensor

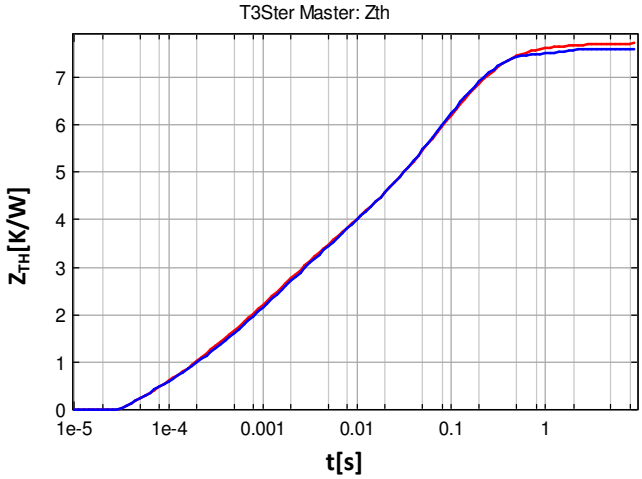


Fig.11 Transient curves normalized by power-step with initial transient correction

References

[1] C. E. Stephens, F. N. Sinnadurai, "A surface temperature limit detector using nematic liquid crystals with an application to microcircuits", Journal of Physics E: Scientific Instruments, Vol. 7, No. 8, 1974, pp. 641 - 643.

[2] L. G. Walshak, W. E. Poole, "Thermal resistance measurement by IR scaning", Microwave Journal, Vol. 16, 1977, pp. 62-65.

[3] V. Székely et al., "New approaches in the transient thermal measurements," Microelectronics J., vol. 31, pp. 727–733, 2000.

[4] J. Desmaris, et al., "Fabrication and characterization of submicron AlGaAn/GaAn HEMTS", Chalmers University , http://www.ep.liu.se/ecp/008/posters/008/ecp00808p.pdf

[5] K.A. Jenkins, K. Rim, "Measurement of the effect of self-heating in strained-silicon MOSFETs", Electron Device Letters, Vol. 23, Issue 6, pp 360-362. June 2002

[6] Zheng Chen et al., "Characterization and Modeling of 1.2 kV, 20 A SiC MOSFETs", Proc. of ECCE 2009, pp. 1480 – 1487, Sept. 2009

[7] K. Horio and Y. Fuseya, "Two-dimensional simulations of drain-current transients in GaAs MESFETs with semi-insulating substrate compensated by deep levels," IEEE Trans. Electron Devices, vol. 41, pp. 1340–1346, Aug. 1994.

[8] K. Kunihiro and Y. Ohno, "A large-signal equivalent circuit model for substrate-induced drain-lag phenomena in HJFETs," IEEE Trans. Electron Devices, vol. 43, pp. 1336–1342, Sept. 1996..

25-27 September 2012, Budapest, Hungary

Modelling and Characterisation of Smart Power Devices

B. Wunderle[1], M. Abo Ras[2], M. Springborn[1], D. May[1], J. Kleff[3], H. Oppermann[3], M. Töpper[3], T. Caroff[4], R. Schacht[5], R. Mitova[6]

[1]Technische Universität Chemnitz, Reichenhainer Str. 70, 09107 Chemnitz, Germany
Email: bernhard.wunderle@etit.tu-chemnitz.de
[2]Berliner Nanotest und Design GmbH, 12489 Berlin, Germany
[3]Fraunhofer Institute Reliability and Microintegration IZM, 13355 Berlin, Germany
[4]Commissariat à l'énergie atomique et aux énergies alternatives - LITEN, 38054 Grenoble, France
[5]University of Applied Sciences Lausitz, 01968 Senftenberg, Germany
[6]Schneider Electric, 38050 Grenoble, France

This paper deals with the system design, technology and test of a novel concept of integrating Si and SiC power dies along with thermo-electric coolers in order to thermally manage transients occurring during operation, thus turning it into a smart power device. The concept features double-sided cooling as well as new materials and joining technologies to integrate the dies such as transient liquid phase bonding/soldering. Coupled-field simulations are used to predict thermal performance and are verified by especially designed test stands to very good agreement. This paper is the first in a series of planned publications on the ongoing work in the project.

I. INTRODUCTION

The thermal performance and thermo-mechanical reliability of power assemblies is largely materials- and technology-driven. Therefore, progress in this respect depends on the integration of novel semiconductor as well as joining materials along with specially developed processes within a viable system concept (for a review see e.g. [1, 2]). For power conversion in the automotive, energy or aerospace sector, new emerging power semiconductor technologies, respectively SiC and GaN, have demonstrated their higher performances than market-going technologies (Si and GaAs) in terms of reduction of losses, increase of efficiency and power handling. Consequently, to increase the efficiency of entire power modules, achieve smaller form factor and make them cost effective, lightweight and reliable, these SiC and GaN components need to be packaged with respective novel high-temperature joining technologies, assuring low thermal resistance and mechanical stresses, assuring high reliability, within the constraints of the (liquid) cooling concept.

In the presented case of interest, twelve Si (later also SiC) power dies (typical *six-pack* converter) need to be cooled with a limited thermal budget and integrated into a commercial standard casing for industrial applications (motor drive). Further, thermal transients need to be managed by implementation of a thermo-electric cooler (TEC). This calls for a completely new thermal management concept and system architecture as it is depicted in figure 1.

To meet the given challenges, the time-honoured design of a chip soldered onto direct copper bonded (DCB) substrates and contacted with heavy wire bonds will not work. An integrated TEC requires a second heat path from the top side, resulting in a double sided cooling approach for the envisaged design.

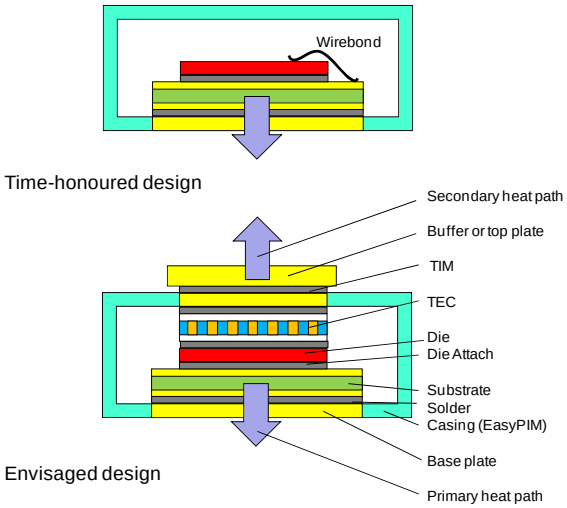


Fig. 1. Schematic of time-honoured concept for power chip cooling (top) and envisaged new concept of double-sided cooling concept with integrated thermo-electric cooler (btm).

The use of TEC for thermal management of power systems seems a very inefficient solution, so the pros and cons have to be carefully weighted. In general, they would only be applied if no other solution is available. So TECs have in the past been proposed for hot-spot cooling in microprocessors [3-5], where they do not substantially increase thermal dissipation due to very small area coverage only. For power modules this would not apply – save in a transient case as is suggested here – and so this needs to be proved, and if yes, under what conditions.

Using TECs as needed here, entails double sided cooling

with all its challenges. It has been shown [2,6], that double sided cooling represents an ingenious way to reduce size, weight and costly chip area in a power module. Here, up to 40 % performance increase are predicted. Some solutions already exist [2,7], all of which require development of new joining technologies due to issues of processability, planarity, high-temperature compatibility and reliability [8]. As there are no wire bond interconnections, reliability problems [9] associated with them are eliminated.

Recent developments constitute silver sintering [10] which already outdoes Sn-based solder in lifetime [1], diffusion soldering (transient liquid phase bonding/soldering) with e.g. the CuSn system [8,11] or nanosponge technologies [12]. Often, processes have to be specifically adapted and reliability questions still to be answered, especially for the case of double sided die attach to become a mature technology.

In the following, we will discuss first some general design considerations, then the selection of the different components and materials, the development of required technologies, performance prediction and characterisation methods. Finally, thermal performance will be demonstrated by an assembly of reduced complexity, using an already packaged power chip. First, tasks will be carried out on Si dies. In a second phase, SiC dies will be used.

II. SYSTEM REQUIREMENTS & FEASIBILITY

In this paper, we contemplate a 4 KW converter for a 3-phase motor drive application with a bias of 390-540 Volts and medium currents of around I = 10 Amps. The total power loss during normal operation amounts to ca. N_{tot} = 70 W with an additional t = 60 seconds transient overload case (at low frequencies) of P_{tot} = 42 W (see also table 1). This is P_{th} = 6 W for each IGBT and 1 W for each diode. Maximum operating junction temperature is T_j = 130 °C, being reached by a liquid cooling heat transfer coefficient of ca. h = 3000 W/m^{2}K for normal operation (N_{tot}), when putting the heat sink at a temperature of 95°C. After one overload peak, the system has enough time to equilibrate again to steady state conditions.

Table 1: Specifications

Quantity	Symbol	Value
Total normal power module	N_{tot}	70 W
Total overload power module	P_{tot}	42 W
Normal power per IGBT	N_{th}	11.6 W
Overload power per IGBT	P_{th}	6 W
Total Overload Power IGBT	$P_{OL} = N_{th} + P_{th}$	17.6 W
Overload power per Diode		1 W
Overload time	t_{th}	60 s
Max. Heatsink temperature	T_{HS1}	95 °C
Max. junction temperature	T_J	130 °C
Max. T_j increase during overload	ΔT	5 K

Specification requires that during overload the junction temperature will not change by more than ΔT = 5 K. This cannot be accommodated by the time-honoured design. However, in order to stick with the same commercial package (EasyPIM), i.e. not to increase the form factor, the use of a TEC is envisaged.

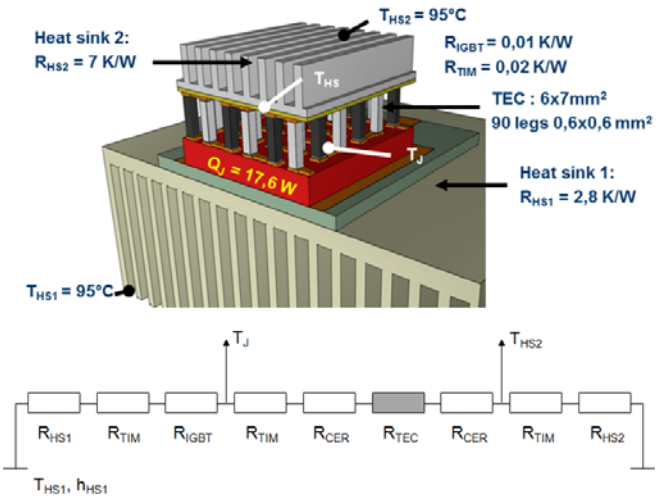


Fig. 2. Schematic for first calculation and circuit diagram. Values refer to table 2.

A first simulation shows the feasibility of such a concept, using a one-dimensional heat path simulation, taking into account the individual thermal resistance contributions as given in figure 2 and table 2.

Table 2: R_{th} data and boundary conditions used.

Quantity	Symbol	Value
Temperature of the heatsink 1	T_{HS1}	95 K/W
Thermal resistance of the Heatsink 1	R_{HS1}	2,7 K/W
Temperature of the heatsink 2 (assuming infinite capacity)	T_{HS2}	95°C
Thermal resistance of the Heatsink 2	R_{HS2}	7 K/W
Thermal resistance of the TIM	R_{TIM}	0,02 K/W
Thermal resistance of the IGBT	R_{IGBT}	0,01 K/W
Thermal resistance of AlN substrate for TEC	R_{cer}	0,05 K/W

The study was carried out on a P_{OL} = 17.6 W IGBT. In this steady state case, the IGBT is located between two heat sinks and mounted with a thermoelectric cooler, as can be seen in figure 2 for the double side cooling configuration. The thermal resistance of the ceramic is assumed to be R_{cer} = 0.05 K/W on both sides and the thermal resistance of the thermal interface material (TIM) was taken to be R_{TIM} = 0.02 K/W. The thermoelectric cooler was especially designed to cool down the IGBT to T_j = 125 °C while minimizing its power consumption. The TEC (c.f. figure 5) is made up of 90 thermoelectric legs (0.6 mm x 0.6 mm x 0.45 mm height) and its thermal resistance is around R_{TEC} = 11.5 K/W. With

the TEC switched off, the junction temperature reaches T_j = 136°C at P_{OL} = 17.6 W as it adds a non-negligible thermal resistance between the IGBT and the second heat sink and the heat path is not laid out for that power under normal operation conditions.

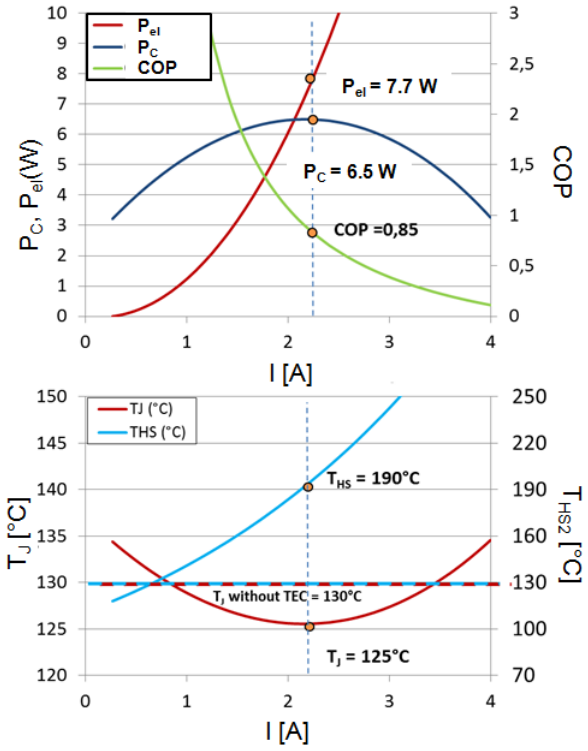


Fig. 3. First feasibility study for one IGBT with an assumed 50 % overload power of P_{OL} = 17.6 W. Boundary conditions (top) and results (btm).

As is seen from figure 3, during an overload of P_{th} = 6 W, the chip can be even cooled down to 125 °C (instead of T_{max} = 130°C, as $P_C > P_{th}$) using a TEC at I = 2.2 A (P_{el} = 7.7 W). In this configuration, the cooling will induce an increase of around ΔT = 60 K in the top heat sink at a low coefficient of performance COP = 0.85, i.e. P_H = 14.2 W, adding up to an excess thermal energy Q_H = 852 J per IGBT to be handled. The COP (c.f. also section VI-B), which is calculated as

$$COP = \frac{P_C}{P_{el}} = \frac{P_H}{P_{el}} - 1$$

is low, but this efficiency is of no concern as P_{el} is absolutely negligible compared to the overall electrical power. So this first calculation shows that a TEC would be indeed a feasible solution, given that the heat path for the second heat sink can be optimised to significantly lower T_{HS2}.

III. CONCEPTION & DESIGN

From a systems perspective, there are several critical points to address when selecting materials and technologies: Thermal performance, local hot spots, electrical insulation (due to high voltage), mechanical and reliability-related aspects, processability, compatibility and availability considerations, form factor as well as cost:

A: Thermal performance aspects

Using a TEC to manage the overload transients requires a double sided die attach as well as a thermally conductive connection to the TEC itself. Both solutions, attaching TECs individually to each die (as in figure 1) or attach them on to the top DCB for system cooling (figure 4) would be thermally viable. However, considerations of planarity (processability) and electrical contacting of the individual TECs rule out the first variant.

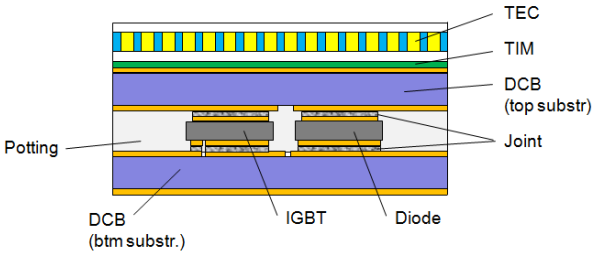


Fig. 4. Design variant showing the system-mounted TEC and flip-chip-mounted IGBT and diode.

B: Processability / compatibility / availability aspects

As TEC technology, two variants, depicted in figure 5, are available for that purpose: Si-wafer-level assembled thin-film TECs or AlN-ceramic mounted TECs, both of which can be designed to fit the system size of the top DCB substrate, either as one or in several parts. TIMs for attaching the TECs to the DCB can be very thin bond line thickness (BLT) adhesives due to the small thermal mismatch of silicon to AlN. Both are available with dedicated metal finishes.

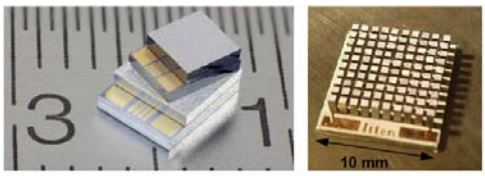


Fig. 5. Different enabling TEC technologies. Micropelt *mpc-d701 (left) and a model from* LITEN.

During assembly, the DCB is each time used as a platform for flip-chip die bonding, including the IGBT which is contacted gate and emitter pad pointing downwards as shown in figure 4. As joining material CuSn-transient liquid phase bonding (TLPB), transient liquid phase soldering (TLPS) or silver sintering are foreseen, allowing a low BLT joint for low thermal resistance. The necessary metallisations on the die (Cu or Ag) will be provided accordingly and serve as buffer layers to accommodate thermo-mechanical stresses

as the CuSn-intermetallics forming during processing are brittle. All of those technologies are stable at high temperatures. TLPS has furthermore the advantage to compensate for DCB warpage and can therefore be used as solution for the last joint. A more detailed description of these technologies follows below.

C: Mechanical / reliability aspects

There is considerable thermal mismatch between the dies and the DCBs. Double sided attach will potentially increase thermo-mechanical stress. Physics-of-failure-based lifetime modelling including damage and fracture-mechanical modelling [13] is to accompany the process development phase. Regarding potential reliability problems due to the TLPB bond, Ag-Sintering is envisaged as an alternative. The Ag-Sintered bond is more ductile and may accommodate stresses rather than shift failure to the interfaces or even the die. Similar considerations for global thermal mismatch will be subject of evaluation when it comes to mechanically attach the DCBs to the TEC as well as to a casing.

D: Electrical insulation aspects

Dies with a voltage class of 1200 V will be used. To assure a large enough distance between collector and emitter (i.e. distance between top and bottom DCB) for high-voltage compatibility, which is usually accounted for by a rim around the perimeter of the chip of several hundred microns, existing gaps will be filled with polymer insulator material (potting, etc.) instead. In addition, epoxy underfill will be used to prevent flashovers between the collector potential to the bottom DCB.

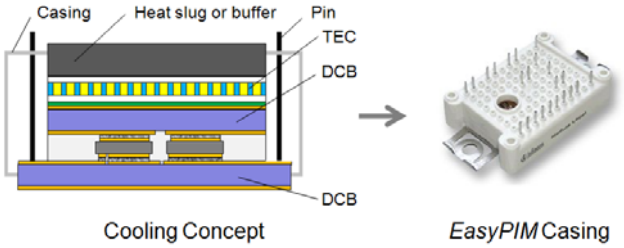


Fig. 6. The design has to fit into a commercial EasyPIM casing.

E: Dimensionality / form factor

Further, form factor considerations require the design to comply with a standard power-electronics casing, such as e.g. EasyPIM (figure 6). Pins which pass through the casing can be connected by an adopted PCB-board. Again, two options exist for the reverse side cooling. Either one may use a customer interface which allows to attach a cold plate to a top heat slug, or a thermal buffer. As the thermal energy to be buffered is not very large, one may consider the latent heat of a material at a phase transition temperature of around 150 °C. To give space for the heat slug or buffer the EasyPIM will be adopted by removing the top casing partially.

F: Local hot spots

A further issue to be clarified is the potentially harmful occurrence of hot spots due to non-uniform heat transfer in double sided cooling as depicted in figure 7. Especially during transient loading this may cause temperature peaks.

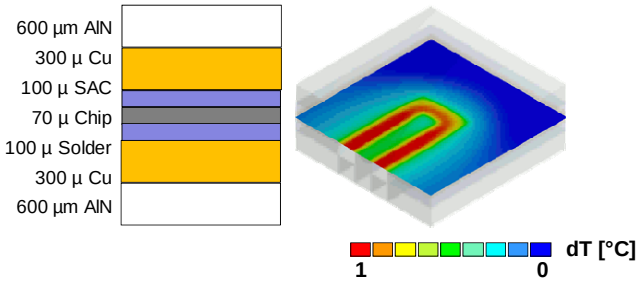


Fig. 7. Transient thermal FE-simulation of center gate architecture power die to estimate hot spots intensity.

Finite Element (FE) simulations were performed to that end, showing a maximum hot spot of $\Delta T = 1$ K step load compared to passive areas on chip, which is negligible. So the concept is set.

IV. JOINING TECHNOLOGY

Three different bonding technologies are foreseen to realize the packages described earlier. All are high temperature stable and have high thermal conductivity.

A: Transient Liquid Phase Bonding (TLPB)

Characteristic for the TLPB is the use of a layer stack consisting of a low temperature melting and a high temperature melting metal between two substrates. By heating this layer stack, it melts and a reaction between the liquid and the high melting temperature layer occurs. The resulting intermetallic phases have higher melting temperatures than the low temperature melting material. For the envisaged package rolled or electroplated copper will be used as high temperature melting layer and electroplated tin will act as the low temperature melting layer. The resulting intermetallic phases Cu_3Sn and Cu_6Sn_5 have decomposition temperatures of 640 °C, respectively 415 °C.

To ensure that the tin layer has totally reacted and the bond is high temperature stable, thin tin layers of 3 µm to 5 µm are commonly used. The reflow is mainly done with pressures in the range of 0.1 MPa to 10 MPa [14]. This guarantees a sufficient contact area between substrates and interlayer. Due to the small amount of solder the surface roughness of the substrates is of high importance. In this case the copper of the DCB has to be polished.

The dies are delivered with a TiNiAg metallisation for the bottom side (Collector). To minimize the process steps, one of the envisaged TLPB bonds will be done with a copper substrate, a tin interlayer, but a TiNiAg metallisation for the

other substrate. First bonding trials with interlayer thicknesses of 3 µm and 5 µm tin show a bonding interface with very low void ratio, see figure 8. The formed intermetallic phases have to be checked by SEM.

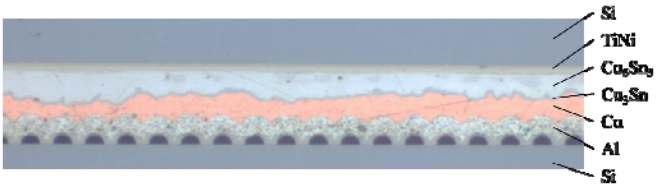


Fig. 8. TLPB-bond.

B: Transient Liquid Phase Soldering (TLPS)

The idea of TLPS is similar to that of TLPB. For TLPS no pressure is needed during reflow and the solder thickness is higher. For the TLPS, we will also use the copper–tin system. The process flow has already been described by Ehrhardt et al. [15]: SnAgCu solder paste with added copper powder and flux is printed on a copper substrate. It has been demonstrated that a one-step process is not targeting. Loading the paste with 40 wt.-% copper balls leads to a highly porous bond [8].

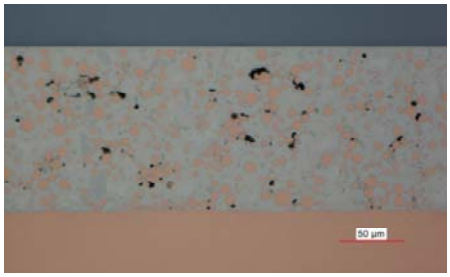


Fig. 9. TLPS-bond.

Good results have been achieved with a two-step process using a solder paste with 40 wt.-% copper balls and a proprietary binder. The cross-section of such a TLPS bond can be seen in figure 9. In a first step, solder is melted under pressure so the bond is compressed and solder is partially squeezed out. In a second step the copper oxides are removed under reducing atmosphere. The process temperature is below the melting temperature of the solder. Afterwards the temperature is increased and the solder, which has been squeezed out, is flowing back. Further investigations to simplify the process flow are on-going.

C: Ag-Sintering

Ag-Sintering can be done with or without pressure. The porosity of the bond can be tuned by the pressure. Figure 10 shows cross-sections of bonds done with 10 MPa and 30 MPa. The sinter paste is printed on a substrate with adequate metallisation. For Ag-Sintering this could be silver or gold. The counterpart is placed and the whole assembly is heated with or without the application of pressure.

D: Copper-Tin Intermetallic compounds

Copper tin phases have a thermal conductivity of 34 W/mK for Cu_6Sn_5, respectively 70 W/mK for Cu_3Sn [16]. Bulk silver has a thermal conductivity of 418 W/mK. Precision measurements have recently shown a reachable 370 W/mK for a sintered Ag die attach [17]. Therefore all three bonding technologies are excellent choices for challenging thermal management. For all of them pressure has to be applied during the process, therefore tooling is relatively expensive. Ag-Sintering without pressure could be an alternative but is still under development.

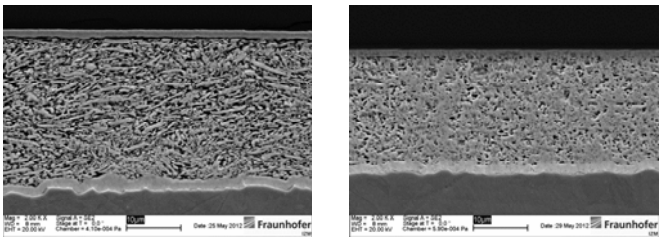


Fig. 10. Sintered silver bond.

We will use at least one TLPS bond in the process flow of the packages to compensate the warpage of the DCBs. TLPB is not suitable for warpage compensation as the bond thickness is very small. For Ag-Sintering different heights will result in different compression ratios which should be avoided.

V. ELECTRO-THERMAL SIMULATION

In order to analyse thermal concepts at field level and to support and verify network level estimations, finite element simulations are performed. Besides accurate material data models, the implementation of TE coolers has to be done carefully.

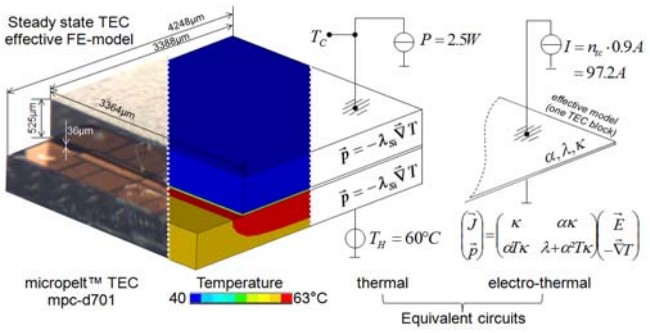


Fig. 11. Field-coupled TE-simulation, principle of effective FE-model (TEC mpc-d701).

In this work, coupled electro-thermal (multi-physics) TEC models are going to be applied to consider the (absolute) temperature dependent heat pumping performance of the Peltier effect. Also, instead of modelling each TE couple, the material data of an effective TE layer is adjusted to match

documented and measured cooling performance.

Figure 11 drafts the steady-state coupled-field physics as explicitly presented in [18,19] and the applied steady-state TEC modelling principle using the example of the TEC mpc-d701 at one specific load condition (0.9A/2.5W/60°C).

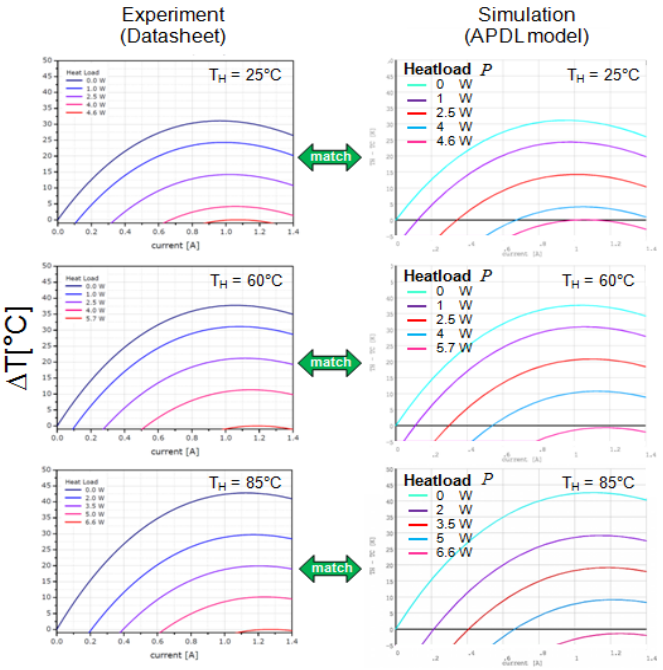


Fig. 12. TEC performance adjustment TEC mpc-d701.

The symbols α, κ and λ refer to the effective material values of the effective TE layer: mean Seebeck coefficient, electrical and thermal conductivity. T, T_C and T_H are absolute, cold and hot side temperature, P, $\vec{p}$, I, J, E and n_{tc} are cold side heat load, heat flux density, operating current, current density, electric field strength and number of thermoelectric couples within the TEC. Now, to adjust the TEC's thermal cooling performance, which (in case of one TE block) can be given as

$$\Delta T = \frac{\frac{h_{TE}}{2\kappa} I^2 + P - \alpha T_H I}{-\alpha I - \frac{\lambda}{h_{TE}}} \quad , \quad h_{TE} = \frac{height_{TE}}{Area_{TE}}$$

α, κ and λ were investigated using the following Ohm's law relations and applying correction coefficients for κ and λ:

$$R_{el,\Delta T=0} = \frac{n_{tc}^2 h_{TE}}{\kappa f} , \alpha = \frac{\alpha_{Bi_2Te_3}}{cc_\alpha} , f = \frac{Area_{TE}}{Area_{TEC}}$$

$$R_{Th,TE} \approx R_{Th,TEC} - \frac{h_{Hot}}{\lambda_{Hot}} - \frac{h_{Cold}}{\lambda_{Cold}} \approx cc_\lambda \frac{h_{TE}}{\lambda}$$

$R_{el,\Delta T=0}$ refers to the TEC's electrical resistance at a ΔT (and so Seebeck voltage) of zero. $R_{Th,TE}$ and $R_{Th,TEC}$ are the thermal resistances of the TE layer and the TEC (including the bordering layers), that can be derived from the ΔT curve shift at different heat loads.

For the mpc-d701 TEC, the effective model is in good agreement with its datasheet performance, using the following material data:

$$\alpha = 292.5 \frac{\mu V}{K} \quad , \quad \kappa = 4493 \frac{A}{Vm} \quad , \quad \lambda = 0.40435 \frac{W}{mK}$$

Since this is an effective model, the picked material data do vary from typical values in the literature. Especially thin film TEC require considerable correction coefficients of around 0.7-0.8 because of the increasing impact of the metallization layer. However, figure 12 compares model and datasheet performance data at different hot side temperatures. It can be seen, that the model correlates well within small errors regarding the ΔT_{max} and I_{max} shift while heat load increases and of course over all heat pumping performance with the absolute hot side temperature.

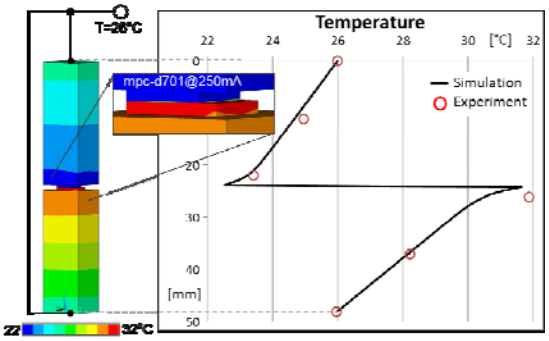


Fig. 13. TEC mpc-d701 under test, measurement vs. simulation.

For further verification, simulations of the TEC testing discussed in chapter VI have been made. Figure 13 shows the TEC tested at 250 mA and the TEC tester temperature distribution within the thermal reference blocks. Both measurement and simulation show the typical asymmetric heat flow condition (due to the TEC's power dissipation) and also field narrowing effects at the TEC ports.

So, already in good agreement, the ongoing work is going to pursue the goal of easy and precise FE assisted TEC characterization and the extraction of easy-to-implement TEC-models by measurement, being capable of sample related simulations. Also transient behaviour is going to be taken in consideration.

VI. ELECTRO-THERMAL CHARACTERISATION

Thermo electrical cooler TECs are heat pumps. To describe its pump performance two important characteristic values have to be taken into account.

A: Thermo-electrical figure of merit (ZT):

In general the figure of merit (Z) is a quantity used to characterize the performance of a device. The figure of merit for thermoelectric devices (ZT) is defined as is the relation between the electrical voltage and the voltage caused by the Seebeck effect. [20]

$$ZT = \frac{V_{th}}{V_{el}}$$

The effective figure of merit of a TEC module can be measured by the so called modified Herman method [21], see figure 14. The TEC will be sourced with low constant DC current (few mA). Once the TEC module reaches the thermal steady-state, the total measured voltage on the TEC is the sum of the electrical voltage and the thermal voltage generated by the Seebeck effect. When the constant current is switched off, only the voltage V_{th} caused by the Seebeck effect is measured.

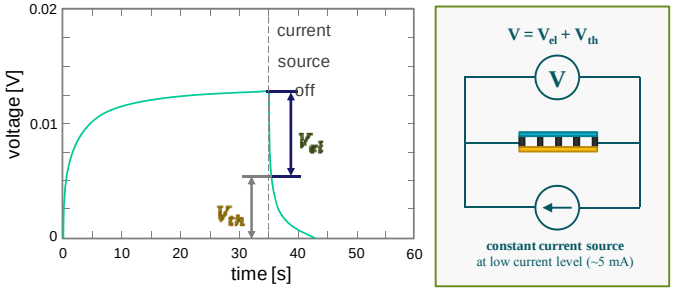


Fig. 14: Principle of ZT measurement.

B: Coefficient of performance (COP):

The coefficient of performance (COP) is the relation between the heat absorbed at the cold side of TEC (P_C) and the input electrical power (P_{el}) as shown in figure 15.

$$COP = \frac{P_C}{P_{el}}$$

Where the pumped heat of the hot side of TEC (P_H) is the sum of the absorbed heat at the cold side (P_C) and the input electrical power (P_{el})

$$P_H = P_C + P_{el}$$

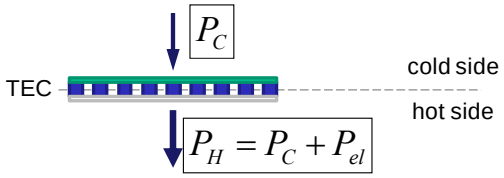


Fig. 15: principle of COP measurement.

To measure the COP of TEC on module level a test stand shown in figure 17 was designed, which is based on the well-tested *TIMA* design [22]: The test stand allows measurement of the heat flow through the both side of TEC by metal heat flow sensors. The surface temperature of both TEC side can also be measured by the metal heat flow sensors schematically depicted in figure 16. The geometry and the thermal conductivity of the metal block must be known. The heat flow j_{th} through the metal block can be calculated by following equation:

$$j_{th} = \frac{P}{A} = \frac{\Delta T}{L} \cdot \lambda,$$

where ΔT signifies the temperature drop across the metal block, L its length, λ its thermal conductivity A its cross section area.

The temperatures on the cold side and hot side surfaces can be calculated by extrapolation of the linear function.

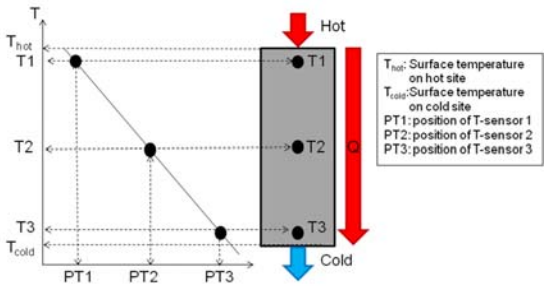


Fig 16: Schematic of metal heat flow sensor.

The TEC is measured between two metal heat flow sensors. To simulate the real operation condition of the TEC, the heat flow sensors on the cold side of the TEC is connected to a heater and the sensor on the hot side of the TEC is connected to heat sink.

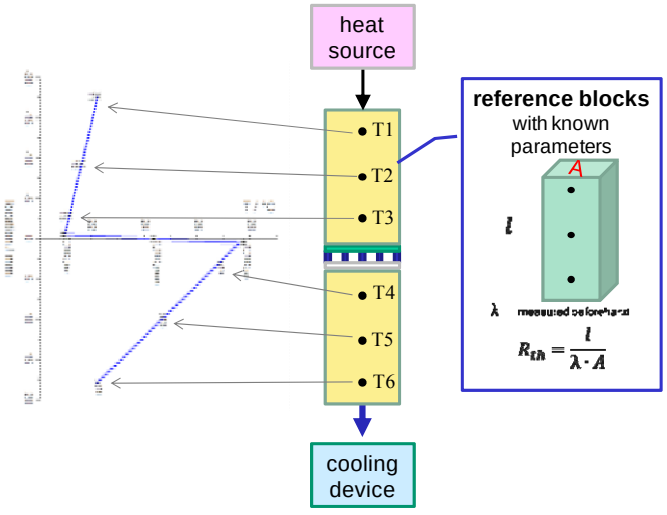


Fig. 17. Concept of test stand for COP characterisation of TEC: Test setup (middle); topical temperature curve (left); reference metal block as heat flow sensor (right)

By automated measurement and controlling software

following operations can be performed: Variation and measurement of electrical power of TEC, regulation of the heat source power, calculation of the heat flow on cold and hot side of TEC, calculation of coefficient of performance and calculation of error bars of COP with respect to all measured values.

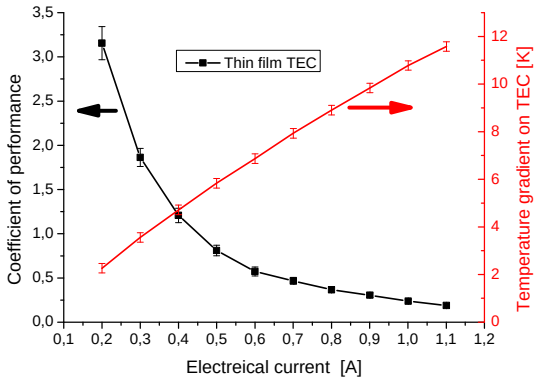


Fig. 18. ΔT_{TEC} and COP as function of electrical current of WL thin film TEC (figure 5) at RT.

The COP of commercial ceramic-based TEC and WL thin film TEC was characterized by the in above described test method. Figure 18 shows the COP of thin film TEC as function of the electrical current. The dimensions of the TEC are A = 5 x 5 x 0.5 mm^3. For COP measurement the input electrical power was varied between 0.2 W and 7 W while the TEC was kept at room temperature. The result shows a typical trend of COP and ΔT_{TEC} as function of electrical current. The COP decreased with increasing the electrical current.

VII. FIRST ELECTRO-THERMAL DEMONSTRATION

To show the potential of a TEC base smart cooling solutions for power electronic devices, a special demonstrator concept (shown in figure 19) was designed and is being realized.

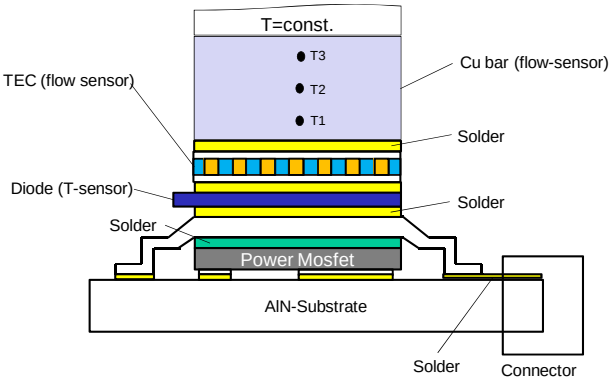


Fig. 19. Schematic of first demonstrator set-up for characterisation.

This is necessary to fit the boundary conditions given by the TEC characterization platform (described in the previous section) and to separate thermo electrical characterization issues from those of process development of TEC integration.

Fig. 20. Packaged power-die (CanPAK) *for the first demo.*

The demonstrator (figure 19) consists of a readily available and solderable power MOSFET transistor (see figure 20) soldered onto a ceramic board. The thin film based (figure 5) TEC (provides by project partner) is mounted on top of the *CanPAK* device. To measure the case temperature of the MOSFET a very thin, thermally negligible, chip diode (thickness 70 µm) is placed between transistor and TEC. On the top of TEC a metal based thermal flow sensor measures the pumped heat in parallel to the TEC which is then also used as heat flow sensor. As jointing technologies soldering or sintering are envisaged. Gluing would reduce drastically the measuring accuracy due to its very process dependent and low thermal conductivity and is therefore ruled out.

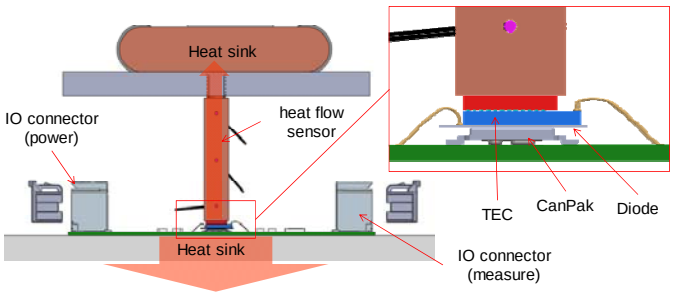


Fig. 21. Realization of characterisation set-up of first demonstrator.

To characterize the core pieces of the future smart system the entire test setup provides (figure 21) the following items: 1. Heat generation by MOSFET transistor, 2. transient load profiles and temporarily overload scenarios are possible, 3. full control the heat pumped by TEC, measurement of all relevant temperatures of TEC and MOSFET transistor, 4. measurement of TEC heat pump by metal heat flow sensor and 5. measurement of the AlN substrate temperature by NTC (as monitor and reference).

Using the completed demonstrators and the designed test stand can answer the question of whether smart TEC integration can be used to relieve transient thermal loads. At the time of writing of this paper no assemblies are finished yet.

CONCLUSIONS & OUTLOOK

In this paper we have studied the design, technology and characterisation of a double-sided cooled power-die with top-mounted thermo-electric cooler under transient load. We have found out the following:

- Under given boundary conditions the proposed concept is thermally viable. This could be proved by combined simulative and experimental approach.
- Joining technologies and corresponding process flows have been defined in view of the challenges of high voltage, warpage, conductance, metallurgy and reliability. They also maintain high-temperature headroom. Critical points in the design are identified to best knowledge for now and are worked on in the form of building blocks for the design of the module.
- The concepts demands substantial characterisation and modelling accompanying its optimisation. Therefore, a TEC characterisation test stand has been realised as well as first demonstrator of lower complexity to allow measurement of the main parameters. First thermo-electric simulations agree well with the first measurements.

ACKNOWLEDGMENTS

The authors would like to thank W. Maurer, Infineon Technologies AG, Germany, for valuable discussions. We appreciate the support of the EU FP 7 Integrated Project "Smartpower", Grant #288801.

REFERENCES

[1] T. Stockmeier. *From Packaging to 'Un'-Packaging – Trends in Power Semiconductor Packaging.* Proc. Power Semiconductor Devices and IC's (ISPSD) conf., 2008.

[2] J. Schulz-Harder. *Review on Highly Integrated Solutions for Power Electronic Devices*. Proc. Integrated Power Systems (CIPS) conf., 2008.

[3] A. Bar-Cohen and P. Wang. *On-chip Hot Spot Remediation with Miniaturized Thermoelectric Coolers*. Microgravity Sci. Technol. 21 (Suppl 1). pp. 351–359, 2009

[4] G.J. Snyder, M. Soto, R. Alley, D. Koester and B. Conner. *Hot Spot Cooling using Embedded Thermoelectric Coolers*. Proc. 22nd Semitherm Conf. 2006.

[5] F. Schindler-Saefkow, M. Jägle, B. Wunderle, H. Böttner and J. Nurnus. *Thermal-Electric Simulation of Micropeltier Elements for Hot Spot Cooling.* Micromaterials and nanomaterials 6 (2007), S.218, ISSN: 1619-2486, 2007.

[6] S. Dieckerhoff, S. Guttowski and H. Reichl. *Performance Comparison of Advanced Power Electronic Packages for Automotive Applications.* Proc. Automotive Power Electronics Conf., Paris, June 21-22, 2006.

[7] M. Schneider-Ramelow, T. Baumann and E. Hoene. *Design and assembly of power semiconductors with double-sided water cooling.* Proc. CIPS, 2008.

[8] M. Schneider-Ramelow, M. Hutter, H. Oppermann, J.-M. Göhre, S. Schmitz, E. Hoene and K.-D. Lang. *Technologies and trends to improve power electronic packaging*. Proc. 44th Int. Symp. On Microelectronics, 2011.

[9] J. Göhre, M. Schneider-Ramelow, U. Geissler, K.-D. Lang. *Interface Degradation of Al Heavy-Wirebonds on Power-Semiconductors during Active Power Cycling measured by the Shear Test*. Proc. PCIM 2010, Nuernberg, Germany.

[10] U. Scheuermann. *Reliability challenges of automotive power electronics.* J. Microelectronics Reliability 49, pp. 1319–1325, 2009.

[11] K. Guth, D. Siepe, J. Göhrlich, H. Torwesten, R. Roth, F. Hille and F. Umbach. *New assembly and interconnects beyond sintering methods*. Proc. PCIM 2010, Nuernberg, Germany.

[12] H. Oppermann, L. Dietrich, M. Klein and B. Wunderle. *Nanoporous Interconnects*. Proc. 4th ESTC, Berlin, Germany, Sept. 13-16, 2010.

[13] B. Wunderle and B. Michel. *Lifetime Modeling for Microsystems Integration – from Nano to Systems*. J. of Microsystem Technologies, Vol 15, No 6, pp 799 - 813, June 2009

[14] G.O. Cook, C.D. Sorensen. *Overview of transient liquid phase and partial transient liquid phase bonding.* Journal of Material Science, pp. 5305-5323, 2011.

[15] C. Ehrhardt, M. Hutter, H. Oppermann, K.-D. Lang. *Transient Liquid Phase Soldering for lead-free joining of power electronic modules in high temperature applications.* HiTEC 2012.

[16] H.-T. Lee, M.-H. Chen, H.-M. Jao, T.-L. Liao. *Influence of interfacial intermetallic compound on fracture behaviour of solder joints.* Materials and Engineering A358, pp. 134-141, 2003.

[17] B. Wunderle, M. Klein, L. Dietrich, M. Abo Ras, R. Mrossko, D. May, R. Schacht, H. Oppermann, B. Michel, H. Reichl. *Advances in Thermal Interface Technology: Mono-Metal Interconnect Formation, Processing And Characterisation.* Proc. 12th Itherm Conf, Las Vegas, USA, June 2-5, 2010

[18] E.E. Antonova, D.C. Looman. *Finite elements for thermoelectric device analysis in ANSYS.* ICT 2005, pp. 215-218.

[19] M. Jaegle. *Simulating Thermoelectric Effects with Finite Element Analysis using COMSOL.*, Poster Contribution at the ECT 2007.

[20] Harman, T. et al. *Measurement of Thermal Conductivity by Utilization of the Peltier Effect.* Journal of Applied Physics vol. 30, no. 9, pp. 1351-359, 1959.

[21] Park, W. et al. *Effect of thermal cycling on commercial thermoelectric modules*. IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM) 2012, May 30 - June 1, San Diego, CA

[22] M. Abo Ras, R. Haug, R. Schacht, C. Monory-Plantier, D. May, B. Wunderle and B. Michel. *Automated test system for in-situ testing of reliability and aging behaviour of thermal interface materials*. Proc. 14th Nanotech Conf., Boston, Ma, USA, June 13-16, 2011

Author Index