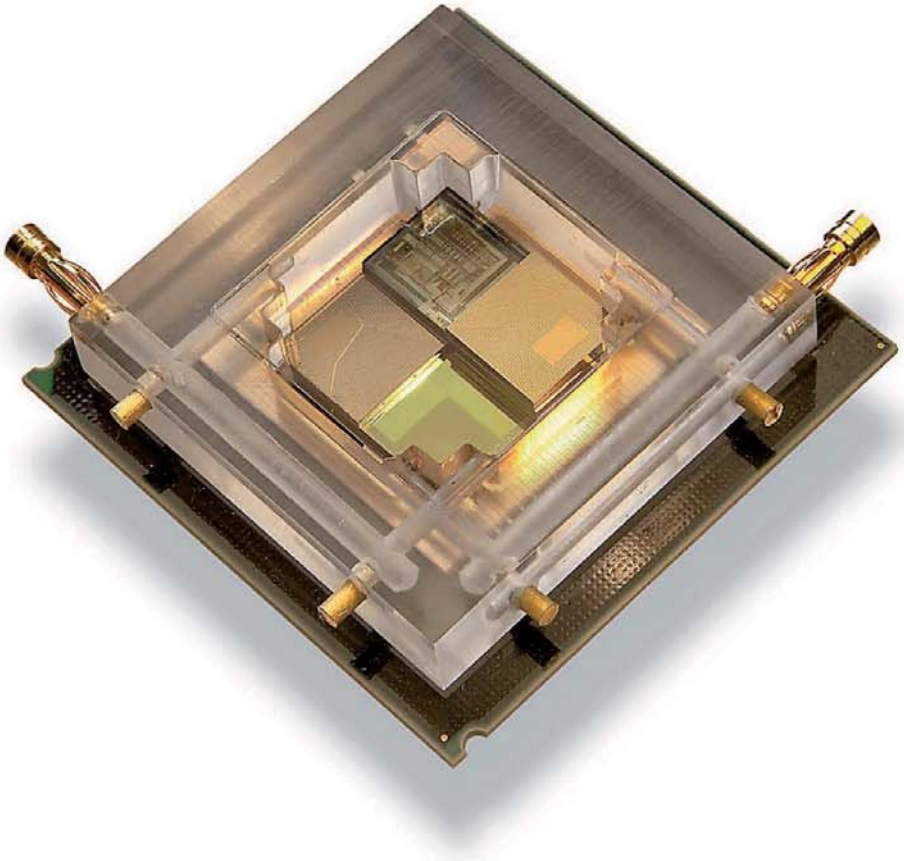
Interlayer Thermal Management of High-Performance Microprocessor Chip Stacks

Thomas Brunschwiler



Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

Interlayer Thermal Management of High-Performance Microprocessor Chip Stacks

Interlayer Thermal Management of High-Performance Microprocessor Chip Stacks

von Diplom-Ingenieur
Thomas Brunschwiler
aus Thalwil

von der Fakultät *IV* - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades eines

Doktor der Ingenieurwissenschaften
– Dr.-Ing. –

genehmigte Dissertation.

Promotionsausschuss:

Vorsitzender:	Prof. Dr.-Ing. Roland Thewes	TU Berlin
1. Gutachter:	Prof. Dr.-Ing. Dr.-Ing. E.h. Herbert Reichl	TU Berlin
2. Gutachter:	Prof. Dr. Dimos Poulikakos	ETH Zürich
3. Gutachter:	Prof. Dr.-Ing. Bernhard Wunderle	TU Chemnitz

Tag der wissenschaftlichen Aussprache: 16. Januar 2012

Berlin 2012

D 83

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
1. Aufl. - Göttingen : Cuvillier, 2012
Zugl.: (TU) Berlin, Univ., Diss., 2012

978-3-95404-034-6

D 83

Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

1. Auflage, 2012
Gedruckt auf säurefreiem Papier

978-3-95404-034-6

Summary

Vertical integration of integrated circuit dies offers tremendous opportunities from an architectural as well as from an economical standpoint. Memory proximity supports performance scaling, and might enable significant energy savings. Partitioning of the corresponding functionalities and technologies into individual tiers can improve yield and modularity substantially.

The paradigm change of stacking active components has a direct impact on heat-removal concepts and is therefore the motivation of this thesis. A stack comprised of a single logic layer in combination with multiple memory dies was identified as the limit for traditional back-side heat removal. To minimize junction temperatures, a stacking sequence with the high heat-flux component in close proximity to the cold plate is proposed.

Interlayer cooling is the only volumetric heat-removal solution that scales with the number of dies in the stack. Hence, the focus of this thesis has been to identify the potential of interlayer cooling and to provide a modeling framework. Fundamental heat-transfer building blocks, such as unit-cell geometries, fluid structure modulation, fluid focusing, as well as four-port fluid delivery supporting power-map-aware heat removal, are discussed. Moreover, the theoretical foundation was experimentally validated on resistively heated convective test cavities. Therefore, specific bonding and insulation schemes were developed. Finally, the interlayer cooling performance was demonstrated on a pyramid chip stack.

A multi-scale modeling approach for the efficient design of non-uniform heat-removal cavities was proposed. Periodic arrangements of heat-removal unit-cells in the cavities are described by the porous-media approximation. Their characteristics are represented by the directional and velocity-dependent modified permeability and convective thermal resistance. An extended tensor description was developed to map the pressure gradient to the DARCY velocity. These parameters were derived from detailed numerical heat and mass transport modeling for arbitrary angle-of-attack of the fluid, using a set of novel routines that support periodic hydrodynamic and thermal boundary conditions. For pin-fin arrays, a biased fluid flow towards directions with maximal permeability could be observed. Field-coupling between the two-dimensional porous and adjacent three-dimensional solid domains was performed to derive the temperature field in the chip stack, including heat spreading in the silicon die. The modeling results are conservative and deviate less than 20 % from the measured junction temperatures, when considering the temperature dependency of the coolant viscosity. This is a very good value considering the immense complexity reduction, resulting in a low computational time of less than 20 min on a desktop computer, to derive the mass transport and junction temperatures within a chip stack.

Sputtered AuSn 80/20 was investigated as eutectic thin-film bond to form leak-tight interfaces with mechanical, electrical, and thermal functionality, as part of the technology development, to enable the use of water as coolant. The resulting bond quality was characterized for various underbump metallizations, atmospheres, and reflow/force profiles. The implementation of a differential pumped chamber allowed the use of formic acid in the flip chip bonder to reduce the tin oxide on the solder surface. The transient liquid-solid nature of the thin-film solder process explains the sensitivity on the underbump metallization and the heat ramp. Finally, processing guidelines supporting the design of leak-tight bond interfaces were summarized. Acceptable intermetallic compound formation was achieved at heat ramps of 100 $\frac{\mathrm{K}}{\mathrm{min}}$ and with chromium as wetting layer. A bondline thickness of 4 µm and a Teflon support provided sufficient compliance to form successful bonds considering the wedge errors of the flip chip bonder.

Waterproof, two-level metallizations to mimic processor-like, non-uniform power maps with background and hot-spot heaters were developed for the implementation of single- and multi-cavity test sections. Pin-hole-free dielectric layers (1 µm PECVD Si_3N_4 / 100 nm ALD Al_2O_3) were achieved by conformal thin-film deposition.

Numerous heat transfer assessments yielded the following insights:

The limited heat capacity and flow rate of the coolant were identified as the major contributor to the thermal gradient in convective interlayer heat removal, even when water using as coolant. This is due to the small hydraulic diameter defined by the interconnect density (pitches < 200 µm) and the length of the cross-flow heat exchange cavity (≥ 10 mm).

The circular pin-fin in-line unit-cell was identified as the optimal heat transfer geometry for heat capacity limited cross-flow heat transfer. It results in the highest porosity, beneficial for efficient mass transport, compared with microchannels and other pin shapes at a given minimal radius constraint. Improved convective heat transfer towards the outlet of the cavities caused by transient vortex shedding was observed at increased REYNOLDS numbers (> 100) in the pin-fin in-line case.

Fluid cavities with four-port fluid delivery and heat removal geometry modulation need to be considered for chip stacks larger than 2 cm^2 and a interconnect pitch of ≤ 50 µm. Their effectiveness was demonstrated with cavities that were either partially fully or half populated with pin-fin arrays. These arrangements result in a significant increase in local fluid flow compared with uniform heat transfer cavities.

Microchannels have proved to dissipate heat efficiently to multiple fluid cavities in the chip stack because of the improved die-to-die coupling, caused by the 50 % fin fill factor. This is advantageous for disparate tier stacking. The high-power die can benefit from heat dissipation into cavities adjacent to low-power tiers.

Additional recommendations, critical for electro-thermal co-design, are also discussed: i) Heat spreading in the silicon helps to mitigate hot-spots below a critical spatial dimension of 1 mm. ii) High heat flux macros should be placed towards the fluid inlet and die corners if the two- or four-port configuration is implemented, respectively. iii) A manifold width of 1 mm should be considered to achieve a fluid maldistribution below 1 % between the fluid cavities. iv) A 1.6 ms thermal time constant was derived for an interlayer cooled chip stack. Hence, predictive cooling-loop control schemes need to be implemented to account for the comparable high pump time constant.

Finally, for the first time, the superiority of interlayer cooling as a volumetric heat-removal method could be experimentally demonstrated on the pyramid chip stack test vehicle with four fluid cavities and three power dissipating tiers. Aligned hot-spots were included with 250 $\frac{W}{cm^2}$ heat flux each. A total power of 390 W, corresponding to a 3.9 $\frac{kW}{cm^3}$ volumetric heat flow, could be dissipated on the 1 cm^2 device at a 54.7 K junction temperature increase. In comparison, back-side cooling would result in a junction temperature increase of 223 K with respect to the fluid inlet temperature of the microchannel cold plate.

Using the results of the present work, it is now possible to design and predict mass and heat transport in an interlayer cooled chip stack, with the support of the proposed best-practice design rules in combination with the validated multi-scale modeling framework. The scalable nature of interlayer cooling will enable “Extreme-3D-Integration” with computation in sugar cube form factor chip stacks, extending integration density and efficiency scaling beyond the “End-of-2D-Scaling”.

Zusammenfassung

Die vertikale Integration von integrierten Schaltungen weist nicht nur bezüglich neuer Mikroarchitekturen sondern auch in ökonomischer Hinsicht ein enormes Potential auf. Die unmittelbare Nähe von Speicherkomponenten und integrierten Schaltungen ermöglicht eine beträchtliche Leistungssteigerung sowie eine signifikante Verbesserung der Energieeffizienz. Die Partitionierung von unterschiedlichen Funktionalitäten und Technologien auf individuelle Ebenen verbessert Fabrikationsausbeute und Systemmodularität substantiell.

Der Paradigmenwechsel hin zum „Stapeln" von aktiven Komponenten (der Chips) hat eine direkte Auswirkung auf die Wahl der Kühlkonzepte und war deshalb die Motivation dieser Dissertation. Als Obergrenze hinsichtlich des traditionellen Kühlens via Chip-Rückseite, wurde ein Chip-Stapel bestehend aus einem Logik- und mehreren Speicherbauteilen identifiziert. Um die Transistortemperatur zu minimieren, wird eine Stapelsequenz vorgeschlagen, in der die Komponente mit hohem Wärmefluss in unmittelbarer Nähe des Kühlkörpers platziert wird.

Die einzige volumetrische Methode zur Wärmeabführung, die mit der Anzahl der Chips im Stapel skaliert, ist die so genannte Zwischenlagenkühlung. Deshalb liegt der Fokus dieser Dissertation darauf, ihr Potential zu definieren und eine Modellierungsmethodik zu entwickeln, mit der die Zwischenlagenkühlung ausgelegt und optimiert werden kann. Des weiteren werden fundamentale Aspekte der Zwischenlagenkühlung diskutiert, wie die Geometrie der Wärmeabfuhreinheitszellen, deren Modulation, das Fokussieren der Kühlflüssigkeit, wie auch die vierseitige Flüssigkeitseinspeisung, welche der lokalen Wärmeflussverteilung im Chip-Stapel Rechnung tragen. Ausserdem werden die theoretischen Grundlagen mittels widerstandsbeheizter Testbauteile experimentell validiert, für deren Realisierung spezifische Verbindungs- und Isolationsmethoden entwickelt wurden. Schließlich wird die Leistungsfähigkeit der Zwischenlagenkühlung anhand eines Pyramiden-Chip-Stapels gezeigt.

Ein Multiskalen-Modellierungsansatz wurde für die effiziente Auslegen von ungleichmäßigen Wärmeentzugskavitäten vorgeschlagen. Wärmeabfuhreinheitszellen in periodischer Anordnungen werden mittels der Porösen-Medien Theorie angenähert. Dabei werden ihre Eigenschaften durch die Richtungs- und Geschwindigkeitsabhängige modifizierte Permeabilität und den konvektiven thermischen Widerstand repräsentiert. Um den Druckgradienten auf die DARCY Geschwindigkeit abzubilden, wird eine erweiterte Tensor-Beschreibung entwickelt. Die entsprechenden Parameter werden aus detaillierten Wärme- und Massentransportmodelrechnungen für beliebige Strömungsrichtungen hergeleitet unter Verwendung neu entwickelter Rechenroutinen, welche die Anwendung von periodischen hydrodynamischen und thermischen Randbedingungen ermöglichen. Im Falle der Pin-Fin Arrays stellt sich heraus, dass die Strömungsrichtung generell hin zur maximalen Permeabilität der Struktur tendiert. Um das Temperaturfeld im Chip-Stapel unter Berücksichtigung der Wärmespreizung im Siliziumbauteil zu berechnen, wird eine Feldkopplung zwischen der zweidimensionalen porösen und der angrenzenden dreidimensionalen festen Domäne eingeführt.

Die Ergebnisse der Modellrechnungen sind konservativ und weichen – bei Berücksichtigung der Temperaturabhängigkeit der Viskosität des Kühlmediums – weniger als 20 % von der gemessenen Transistortemperatur ab. Dies ist ein sehr gutes Ergebnis, speziell man dabei auch die immense Reduzierung der Rechenkomplexität bedenkt, dank der die Bestimmung des Massentransports und der Transistortemperaturen innerhalb des Chipstapels in einer Rechenzeit von weniger als 20 min auf einem Arbeitsplatzrechner möglich wird.

Im Rahmen der Technologieentwicklung wurde gesputtertes AuSn 80/20 als eutektische Dünnfilmverbindung untersucht, um leckfreie Verbindungen mit mechanischer, elektrischer und thermischer Funktionalität herzustellen, welche den Einsatz von Wasser als Kühlflüssigkeit ermöglichen. Die Verbindungsqualität wurde für unterschiedliche Unterlotmetallisierungen, Atmosphären, sowie Temperatur-/Kraftprofile untersucht. Der Einsatz einer differentiell gepumpten Kammer ermöglicht die Verwen-

dung von Ameisensäure im Flip-Chip-Bonder zur Reduzierung des Zinn-Oxids auf der Lotoberfläche. Das transiente Flüssig-Fest-Verhalten des Dünnfilmlötprozesses erklärt dessen Empfindlichkeit auf die Unterlotmetallisierung und die Heizrate. Die entsprechenden Erkenntnisse sind für die Erstellung von Prozessrichtlinien zur Herstellung von leckagefreien Verbindungen verwendet worden. Akzeptable intermetallische Phasenbildung konnte mit Heizraten von 100 $\frac{K}{min}$ und unter Verwendung von Chrom als Benetzschicht erzielt werden. Eine Lotschichtdicke von 4 µm, in Kombination mit einem Teflon-Untersatz, resultiert in einer ausreichenden Nachgiebigkeit, um selbst unter Berücksichtigung des Keilfehlers des Flip-Chip-Bonders zuverlässige Verbindungen zu erzielen.

Zur Realisierung von Testbauteilen mit Einzel- und Multikavitäten wurden wasserkompatible Schichtabfolgen mit integrierter Metallisierung auf zwei Ebenen entwickelt. Diese erlauben es, die ungleichmäßige Leistungsdichteverteilung auf einem Prozessor mittels Hintergrund- und Hot-Spot-Heizern nachzuahmen. Konforme defektfreie dielektrische Schichten (1 µm PECVD Si_3N_4 / 100 nm ALD Al_2O_3) wurden mittels Dünnschichtabscheidung hergestellt.

Aus den im Rahmen dieser Dissertation gemachten Wärmeübertragungs-Untersuchungen konnten folgende Erkenntnisse gewonnen werden.

Der resultierende Wärmegradient der konvektiven Zwischenlagenkühlung wird massgeblich von der begrenzten Wärmekapazität und der Durchflussmenge des Kühlmittels beeinflusst, selbst wenn Wasser als Kühlmedium eingesetzt wird. Dies erklärt sich aus dem geringen hydraulischen Durchmesser, der durch die elektrische Verbindungsdichte (Abstand < 200 µm) und die Länge der Kreuzstrom-Wärmetauscher-Kavität ($\geq$ 10 mm) definiert wird.

Die Pin-Fin in-line Einheitszelle wurde als optimale Wärmeübertragungsgeometrie im Falle von Wärmekapazitäts-limitiertem Kreuzstrom-Wärmetauscher Betrieb identifiziert. Maximale Porosität resultiert von runden Pin Fins, vorteilhaft für effizienten Stofftransport bei einer Vorgabe eines minimalen Radius.

Im Falle von Pin-Fin in-line Strukturen verstärkt sich bei erhöhten REYNOLDS-Zahlen (> 100) aufgrund der transienten Wirbelablösung die konvektive Wärmeübertragung in Richtung des Kavitätsauslasses. Chip-Stapel ab einer Grösse von mehr als 2 cm^2 und mit elektrischen Verbindungsabständen von weniger als 50 µm erfordern Kavitäten mit vierseitiger Flüssigkeitseinspeisung, sowie entsprechender Modulation der Wärmeabfuhrgeometrie. Deren Wirksamkeit wurde anhand von Kavitäten, die teils mit voll und teils mit halb bestückten Pin-Fin-Arrays versehen wurden, getestet. Solche Anordungen führten – verglichen mit uniform bestückten Kavitäten – zu einem signifikanten Anstieg der lokalen Flüssigkeitsgeschwindigkeit.

Mikrokanäle dagegen zeichnen sich dadurch aus, dass bei ihnen der 50 % Kühlrippen-Füllfaktor zu einer erhöhten Chip-zu-Chip-Koppelung führt, dank der sie Wärme effizient in mehrere Kavitäten des Chip-Stapels abführen können. Dies ist im Falle von Chip-Stapeln von unterschiedlichen Komponenten von Vorteil. So kann ein Hochleistungschip auch von der Wärmedissipation in Kavitäten von Chips mit geringer Leistungsdissipation profitieren.

Weitere Faktoren von entscheidender Bedeutung für ein elektro-thermisches Co-Design werden ebenfalls diskutiert: i) Hot-Spots können von der Wärmespreizung im Silizium profitieren, falls diese auf weniger als eine kritische Ausdehnung von 1 mm beschränkt werden. ii) Makros mit hohem Wärmefluss sollten entweder in die Nähe des Flüssigkeitseinlasses oder in den Chip-Ecken platziert werden, falls die zwei-, beziehungsweise die vierseitige Flüssigkeitsspeisung gewählt wird. iii) Die Breite des Flüssigkeitseinlasses sollte mindestens 1 mm betragen, um die Abweichung der Flüssigkeitszufuhr zu den einzelnen Kavitäten auf weniger als 1 % zu beschränken. iv) Für Chip-Stapel mit Zwischenlagenkühlung wurde eine thermische Zeitkonstante von 1.6 ms errechnet. Dies lässt auf die Notwendigkeit von vorausschauenden Regelkreisen für den Kühlkreislauf schließen, da die Zeitkonstanten der Pumpen um Größenordnungen höher sind.

Schließlich wurde zum ersten Mal experimentell mittels Pyramiden-Chip-Stapel mit vier integrierten Flüssigkeitskavitäten und drei dissipierenden Schichten die Überlegenheit der Zwischenlagenkühlung als volumetrische Kühlmethode demonstriert. In diesem Experiment konnten aufeinander ausgerichtete Hot-Spots mit jeweils einem Wärmefluss von 250 $\frac{W}{cm^2}$ gekühlt werden, was einer Leistungsableitung von 390 W bzw. einem volumetrischen Wärmefluss von 3.9 $\frac{kW}{cm^3}$ entspricht, welcher im Stapel mit 1 cm^2 Fläche zu einem Temperaturanstieg von 54.7 K führt. Im Vergleich, normale Rückseitenkühlung würde

hier zu einem Anstieg der Transistorentemperatur von 223 K, relativ zur Vorlauftemperatur der Kühlflüssigkeit der Mikrokanalkühlers, führen.

Mit diesen Erkenntnissen ist es nun möglich, mit Hilfe der vorgeschlagenen Designrichtlinien und in Kombination mit der validierten Multiskalen-Modelliermethodik, den Massen- und Wärmetransport in einem zwischenlagengekühlten Chip-Stapel auszulegen und vorherzusagen. Die Skalierbarkeit der Zwischenlagenkühlung ermöglicht die Realisierung von „extremer 3D-Integration“, also von Chip-Stapeln mit „Zuckerwürfel-Formfaktor“, um eine Steigerung der Integrationsdichte und Effizienz über die der traditionellen 2D-Skalierung inhärenten Limiten hinaus zu ermöglichen.

Acknowledgements

First and foremost I offer my sincerest gratitude to my promoters, Prof. Herbert Reichl, Prof. Dimos Poulikakos and Prof. Bernhard Wunderle, for their confidence and support throughout my thesis. I want to mention especially the extraordinary mentoring by Prof. Bernhard Wunderle. His profound comprehension of a wide variety of engineering and scientific disciplines was always an indispensable resource. My scientific understanding matured through numerous constructive discussions with him at inspiring locations. The effort of the members of the graduation committee, Dr. Hermannn Oppermann and Dr. Bruno Michel is deeply appreciated.

The work described herein was mainly performed at the IBM Zurich Research Laboratory in Rüschlikon (ZRL). I am very grateful to Dr. Bruno Michel, the manager of the Advanced Thermal Packaging team and to Dr. Walter Riess, the head of the Science and Technology department, for their continuous credit during my doctoral thesis. Many of the scientific findings would not have been possible without the support and exchange with my colleagues at the ZRL. I am indebted to the members of the Advanced Thermal Packaging team. The excellent and extensive work of this group in the field of electronics cooling and packaging science is not only the groundwork for this thesis, but also leading-edge in corporate research. Sincere thanks to Stephan Paredes for his careful experimental support, Reto Wälchli for the knowledge exchange in computational fluid-dynamic modeling, Werner Escher for discussions about heat removal test stands, and Javier Goicochea for theoretical considerations concerning porous media modeling. Furthermore, I would like to thank Patrick Ruch for the numerous daily discussions in our office, Yassir Madhour for his motivation to push the frontiers of interlayer cooling even further, and Hugo Rothuizen for his support on questions related to ANSYS CFX. The mental, scientific and typographic support from Tobias Lamprecht, Nikolaj Moll, Charlotte Bolliger and Tobias Kraus is also highly appreciated. I enjoyed all the fruitful and creative conversations with them. A special thanks goes to all the technicians, in particular Ute Drechsler, Martin Witzig, Daniele Caimi, Urs Kloter, and Richard Stutz, from the Science and Technology department, for their highly valued technical support in micromachining. In addition, I would like to thank all the colleagues from IBM Reasearch in Yorktown Heights and Tokyo, with whom I had the chance to collaborate with during the thesis.

I also enjoyed the inspiring scientific exchange during my stays at the Fraunhofer Institute for Reliability and Microintegration IZM in Berlin. The technological discussions and support with respect to solder technology by Dr. Hermann Oppermann, Jessica Kleff, and Dr. Martin Hutter were essential for this work. Numerous conversations with Prof. Bernd Michel, Dr. Olaf Wittler, Dr. Hans Walter, Dr. Astrid Gollhardt, Gerd Schlottig, Dr. Hossein Shirangi, Emmanouella Dermitzaki, Daniel May, Mohamad Abo Ras, Mohammed Bouazza, Raul Mrossko, and Dr. Jürgen Keller broadened my scientific background. The high level of technological knowledge found in the institute was remarkable.

CMOSAIC, a Nano-Tera sponsored project allowed me to collaborate with Gözde Toral and William Cesar from Prof. Yusuf Leblebici's Microelectronic System Laboratory at EPFL, during their master thesis. Further exchange of ideas with respect to different aspects of interlayer cooling was also possible with the many other Professors, pre and post doctoral students from the LTCM and ESL institute at EPFL in Lausanne, as well as the LTNT and FML at ETH in Zurich. Especially to name are Prof. John Thome, Prof. David Atienza, Prof. Wendelin Stark, Dr. Navid Borhani, Dr. Igor Zinovik, Dr. Manish Tiwari, Dr. Fengda Sun, Arvind Sridhar, Mohamed Sabry, Adrian Renfer, and Fabio Alfieri.

The fast and competent response from Steffen Peters CADFEM and Dr. Roland Bender ANSYS CFX were very important during the development of the multi-scale modeling approach.

Contentedly I am looking back on many exciting experiences during my PhD thesis and many unique places I visited during this intense time. I had the chance to meet exceptional persons among my coworkers. I would like to thank them for creating such a inspiring atmosphere, and also for the fun we had during the various recreational activities outside of the laboratory.

Last but not least, I would like to express my gratitude to my girlfriend Miriam, my family, and my friends for their support during this endeavor and beyond.

Contents

1 Vertical Integration of High-Performance Processor-Memory Stacks: Motivation & Conception

Since more than four decades MOORE's Law dictates the pace of economic transistor integration in the integrated circuit (IC) industry. It postulates the doubling of switching elements every 18 month on a single microprocessor die (chip) [1]. In the same time period several key technologists predicted the end of this trend within one or two generations - this never happened. Down to 130 nm transistor node dimensions integration was manly achieved through "classic" DENNARD scaling. Gate-oxide thickness, transistor length and width are scaled with a constant factor improving integration density and clock-frequency at constant power density. Shrinking the switch dimension into the deep-sub-micron range resulted in degraded device performance in all subsequent generations [2]. Hence, the introduction of enhancers such as strained Silicon, high-k metal-gates, low-k wire dielectrics and improved thermal management were responsible to sustain MOOR's Law to 32 nm node dimensions in 2010. For future nodes, the need of innovations is becoming more accentuated as indicated by the INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS[1] (ITRS). They identified vertical integration as a key technology to keep delivering "more for less" also beyond the complementary metal-oxide-semiconductor (CMOS) era.

1.1 Driving Forces and Enabling Technologies

Die-on-die and package-on-package stacking is used in mobile applications to deliver substantial computing performance in a minimal form factor for years already [3]. Individual stratas are fabricated in standard CMOS technology with subsequent wafer thinning before stacking. Unfortunately, the silicon substrate acts as a natural barrier for direct communication between transistors on different levels along the shortest possible distance. Consequently, signals and power have to be routed along the die periphery. This results in severe bandwidth constraints which are not acceptable for high-performance microprocessor applications. They demand for thousands of electrical in- and outputs (IO) as described by RENT'S rule for complex systems [4]. Only a method accepting true area-array-electrical interconnects between dies would fulfill the communication needs within a 3D processor memory stack.

The advent of economic fabrication methods to form anisotropic patterns into silicon with subsequent conformal metal filling were the prerequisites to enable the implementation of die stacks using vertical through-silicon via (TSV) communication. Such processes were introduced in the early 90ties and became state-of-the-art in the industry. Deep reactive-ion etching (DRIE), the so called "BOSCH process", was developed to fabricate high-aspect-ratio microelectromechanical systems (MEMS) such as sensors and fluidic devices [5, 6]. The IC-industry was challenged to reduce wiring resistance and capacitance by replacing sputtered low-aspect-ratio aluminum wiring with high-aspect-ratio electroplated copper utilizing the dual-damascene process [7] in the same time-period. The avenue for vertical integration was foreseeable with these technologies maturing in a high-volume production environment .

1.1.1 Performance Benefits of Vertical Area Array Electrical Interconnects

Vertical integration opens several new possibilities for the design of microprocessors. The benefits are manifold, such as (i) reduced interconnect wiring length, (ii) improved memory to core bandwidth, as

[1] http://www.itrs.net/

well as (iii) heterogeneous integration with its implications on improved computing performance at lower power and cost [8, 9].

In "classical" transistor scaling, switching delay in transistors is improved as the device dimensions are reduced. In contrast, signal propagation between transistors in the wiring layers is altered [10]. The RC_{delay} scales inverse proportional to the square of wiring pitch P and thickness T, due to increased parasitic capacity and resistance as illustrated in Equation 1.1:

$$RC_{delay} = 2\rho\epsilon(\frac{4L^2}{P^2} + \frac{L^2}{T^2}), \tag{1.1}$$

with metal resistivity ρ, permittivity of the dielectric ϵ, at a wiring length L. The result is an expected 2.5 ps transistor delay compared to a 250 ps RC_{delay} per mm in 32 nm technology [11]. At the same time, the die foot print of high-end server microprocessors is increasing, accentuating the so called "wiring crisis" even further. Additional signal repeaters are needed to assist global wire signal transport affecting total power consumption. So far, evolutionary technological improvements such as low-k dielectrics and an increasing number of metal levels to provide global wires with larger cross-section were introduced. 3D integrated architectures would change the perspective completely by minimizing global wiring length proportional to $1/\sqrt{n}$, with n representing the number of stacked dies [12]. This enhances power efficiency, as well as latency.

A further advantage of 3D integration reduced core-to-cache bandwidth bottleneck especially present for multi-core microprocessors. Their efficiency relays on the low latency to fetch data. Therefore, memory hierarchies with fast but small capacity level zero (L0) cache at closest proximity to the core, followed by more distant and larger capacity L1 and L2 cache are implemented on-chip. An imbalance of cache-to-core ratio leads to core stalling due to "cache-miss" events. This hurts especially if cores have to wait for off-chip data to be fetched, causing hundreds of unused cycles. The unbalanced growth of processor-performance of 60 % per annum, compared to a 10 % per annum memory access-time improvement will continue in case of 2D processor architecture and is the main factor comprising the "memory wall". A heuristic observation postulates a reciprocal dependency of the miss-rate relative to the square-root of the cache size. Considering this, doubling the number of cores at a constant off-chip memory bandwidth results in an 8-fold cache size demand. This leads to ever increasing die footprints which soon will hit the limit defined by the maximal lithographic projection area of about 6 cm^2 [13]. Stacking of cache on cores with TSV vertical interconnects results in massive bandwidth and minimal latency to an enormous amount of memory. Separation of functionality to different layers (heterogeneous packages) reduces also the fabrication complexity of individual dies and therefore allows the implementation of e.g. dynamic random-access memory (DRAM) for L2 cache with 8-fold denser packing density, than static random-access memory (SRAM) cells. Production cost will drop and enables additional functionalities in a chip stack as proposed by the "MORE THAN MOORE" initiative.

Several studies tried to quantify the discussed advantages for high-performance processors. BLACK et al. demonstrated the circuit-level potential of the INTEL CORE DUO™processor with shared L2 cache on top of the cores. The number of cycles per memory access was reduced up to 55 %. Further, a logic-on-logic stack was considered with floating point unit on top of a x86 core. It results in a simultaneous 15 % performance increase and a 15 % power reduction for a SPEC2000 benchmark [14]. KGIL et al. proposed a memory hierarchy change for high-throughput multi-thread applications such as tier 1 web servers [15]. Large capacity DRAM layers are attached at the backside of the logic die with high-bandwidth connectivity replacing on-chip L2 cache. The gained space on the logic die can be used for additional processing cores improving computation throughput while each core can run at modest frequency and therefore improving energy-efficiency. The study concludes with a 14 % performance improvement at a 55 % power reduction over a baseline 2D approach.

Architectural and physical design tools have to be developed to take full advantage of the additional degree-of-freedom from vertical integration. Especially algorithms to optimize TSV arrangement with minimal impact on wiring congestion have to be considered, to take full advantage of global wire reduction. Early studies investigating physical-via-impact conclude with an optimal partitioning granularity on unit level (a unit being a large logical entity such as a floating-point unit) and beyond (Figure 1.1) [16].

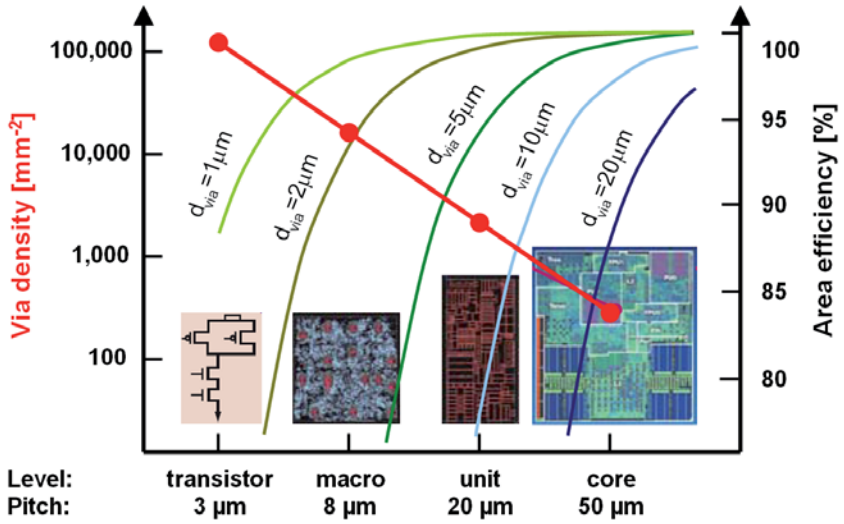


Figure 1.1: 3D partitioning optimum when considering TSV-impact: The available silicon area for transistors (area efficiency) is plotted for different TSV diameters (blue to green lines). Corresponding via densities needed for the communication at different levels is depicted with a logarithmic scale (red line).

This results in TSV pitches ranging from 4 µm to 200 µm[2].

3D integration enables also new business strategies. Intellectual property (IP) sharing is simplified compared to system-on-a-chip (SoC) integration. Functionalities are partitioned to individual stratas and make the designed and fabrication by individual legal entities possible. This approach also enables the use of generic functionality across a product portfolio. Products can be improved and upgraded with minimal cost and time-to-market,simply by replacing one of the layers in the stack, while keeping all the other strata the same. Both aspects increase the number of dies processed for a certain functionality. Therefore, costs are expected to drop according to the principle of economy-of-scale.

1.1.2 Enabling Technologies: Vertical Interconnects - Substrate Thinning - Alignment & Bonding

Two basic approaches were proposed to fabricate a multitude of electric active layers on-top of each other including vertical electrical interconnects. In the *bottom-up* approach a monolithic integration on a single substrate is intended. The bottom-most layer is fabricated using standard CMOS fabrication technology. Subsequent silicon layers are deposited, transistors are defined and are connected by metal wiring layers. It is challenging to achieve high quality subsequent semiconductor layers due to temperature limitations of around 400 °C imposed by the metal wires in place. Epitaxial film quality has to be achieved for high-performance applications to derive charge-carrier-mobilities comparable to bulk silicon. Therefore, localized or low-temperature processes such as laser assisted solid-phase recrystallization [17], metal-induced-lateral overgrowth [18], as well as the implementation of seed agents such as germanium or nickel [19] were proposed to transform the amorphous silicon layers into monocrystalline layers. Another concern is the increasing number of process steps which scales with the number of active layers and comprises device yield.

Hence, the *top-down* approach utilizing 2D die stacking seams to be economically favorable and was proposed in the 80ies due to the aforementioned implications [20]. Each active layer is built on high-quality silicon substrates resulting in optimal transistor performance. Individual layers can be verified prior to stacking, rendering Known-Good-Die philosophy applicable in case of die-to-die or die-to-wafer bonding. Wafer-to-wafer bonding is economical favorable for high-yield devices, due to increased through-put by utilizing the full advantage of batch processing [21]. The main process steps to form 3D chip stacks are: (i) implementation of vertical interconnects, (ii) wafer thinning, and (iii) alignment with subsequent bonding.

[2]This finding will strongly influence the performance of interlayer cooling, as will be discussed in section 1.2.

Vertical Interconnects

Vertical-electrical-interconnects, so called through-silicon vias (TSVs), are the main ingredients in vertical integration and were proposed by ANTHONY in 1981 [22]. TSV fabrication includes the following process steps [23, 24]:

- VIA-ETCHING: DRIE to form high-aspect-ratio holes into the silicon substrate.
- VIA-LINING: Deposition of interconnect supporting layer system - electrical passivation (SiO_2, Si_3N_4), diffusion barrier (TaN / Ta), adhesion (Ti) and seed (Cu) functionality.
- VIA-FILLING: Chemical vapor deposition (CVD) of tungsten (W) or copper (Cu) or electroplating of Cu to form the electrical conductor body.

Four main TSV types are differentiated depending on their implementation during IC fabrication. (i) *Via-first* refers to TSV implementation prior to front-end-of-line[3] (FEOL) processing. Only poly-silicon as a conductor material is compatible with subsequent FEOL processing at temperatures up to 1000 °C. These vias suffer from poor electrical conductivity and cause therefore high resistive losses. TSVs fabricated after transistors implementation, between FEOL and back-end-of-line[4] (BEOL) processing are named (ii) *via-middle*. Due to the relaxed process temperatures during BEOL deposition of ≤ 450 °C conductor materials with improved electrical conductivity such as tungsten (W) and copper (Cu) are applicable. (iii) *Via-last* and (iv) *via-after-bonding* are fabricated typically from copper after completion of the main IC-fabrication.

Feasible TSV heights, diameters, and pitches are strongly dependent on the processes and material involved. They also influence the detailed geometry of the TSV. Economic and void free cylindrical Cu vias can be electroplated with aspect-ratios up to 6:1, using levelers to optimize conformal filling. High mechanical stress is built-up at the TSV-Si interface induced at temperature peaks during fabrication due to the large miss-match in coefficient of thermal expansion (CTE) between Cu (17 ppm/K) and Si (3 ppm/K). The stress increases with the TSV dimension and causes fracture above a critical via dimension. This defines an upper bound for the diameter of cylindrical Cu vias of about 20 µm. For larger diameters an annular via design with a trench width smaller than the critical dimension is proposed. An impedance of 4.4 mΩ was measured at a 50 µm diameter and 150 µm via height with a 4 µm annular ring width [25, 26]. For high-frequency signal transmission the effective impedance of annular and cylindrical TSVs are equivalent considering skin-effects. CVD-tungsten with a low CTE of 5 ppm/K further mitigates mechanical stress, but suffers from reduced electrical conductivity of 9 µΩ cm compared to 2 µΩ cm for copper. Only layer thicknesses below 4 µm are feasible due to the slow CVD deposition-rate resulting in cylindrical vias with diameters ≤ 8 µm. Trenches with 4 µm width and a depth of up to 150 µm can be filled void free thanks to the high conformality of the process. Large vias will be represented by an assembly of several bar-shape or concentric trenches. A seven bar via with 80 µm diameter and 90 µm height results in an impedance of 25 mΩ [27]. Figure 1.2 concludes the TSV dimensional space for the different technologies discussed above.

Via-first and middle have to be processed by integrated device manufacturers (IDMs) or foundries and cause minimal wiring congestion. Via-last and after-bonding can also be performed by outsourced assembly and test (OSATs) organizations. They are fabricated at lower aspect ratios and cause additional wire-blockage in the BEOL wiring layers. Via-last and via-after-bonding are already available on the market according to a market study performed by YOLE DÉVELOPPEMENT [28]. Via-middle processes are still in the development phase. Via-first are in production for sensor applications, but are not considered as a candidate in high-performance applications due to the low electrical conductivity.

Wafer-Thinning

Wafer-thinning in the range of 10 to 150 µm is necessary to minimize vertical communication distance, as well as to maximize TSV densities. The initial thickness of a 300 mm substrate is 780 µm. First, a grinding step is used to thin the wafer down to about 150 µm. It is followed by chemical-mechanical polishing (CMP) and a plasma etch step, to minimize the risk for defects in the junctions. A desired

[3] The term FEOL refers to all process steps preceding the first wiring metal level deposition, mainly being the formation of the transistors.

[4] The term BEOL refers to all processes concerned with metal wiring fabrication. They follow FEOL processing.

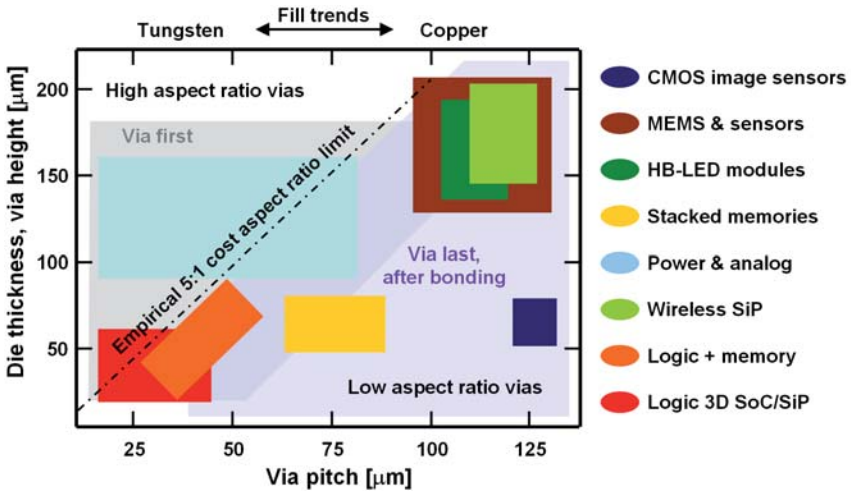


Figure 1.2: TSV pitch and height window for different interconnect density demand, defined by individual applicaitons. Prefered material types and aspect ratios of via first and last are indicated as well [28].

thicknesses of 30 to 150 µm is targeted for bulk devices. Selective wet-etching is used to remove the silicon at the backside of the buried-oxide entirely for silicon-on-insulator (SOI) stratas [29]. The wafer is temporarily bonded onto a handling substrate for mechanical support prior to the thinning process. A spin-on thermoplastic or thermoset compatible with grinding, exposure to chemicals, as well as the stack bonding temperatures performs the temporary bond. Debonding of the handling substrate is done at elevated temperature, where the adhesive becomes viscous and the parts can be separated by a sliding procedure at 200 to 300 °C or by thermal decomposition of the adhesive at 350 to 400 °C [30, 31]. It was also proposed to use electrostatic chucks instead of adhesive bonding, to minimize the risk of thin wafer damage during the debonding step [32].

Alignment & Bonding

Stack-formation can be pursued in serial mode, such as die-to-die (D2D) and die-to-wafer (D2W) bonding or in batch-mode, namely wafer-to-wafer (W2W) bonding. The later method results in highest throughput, but implies a uniform wafer and die size across the chip stack. It is only economical for high yield devices, since stack yield is equal to the die yield to the power of strata numbers. Optimal die size and maximal stack yield on the other hand can be achieved by selecting known-good-dies (KGDs) followed with individual die alignment and bonding as in case of D2D or D2W processing. KGDs can also be reassembled on a carrier substrate, to form a KGD-wafer allowing for high throughput W2W bonding with a resulting high yield.

Face-to-face (F2F) or face-to-back (F2B) strata arrangements are possible[5]. In the F2F arrangement no handling wafers are needed since wafer thinning can be performed after bonding. Furthermore, a high die-to-die interconnect-density can be realized, independent on TSV pitch. At least one F2B bond results for stacks comprised of more than two dies, with the need of TSV and handling wafer.

The main limiting factor for the vertical interconnect density are alignment limitations in the order of ± 0.5 and 2 µm for D2W and W2W bonding respectively. Distortions and wafer bowing caused by non-uniform intrinsic stress and non-isothermal temperature distribution during bonding as the root-cause. Tight alignment tolerance for W2W bonding is more demanding, since alignment is performed across a 200 or 300 mm substrate. This tolerance issues result in a practical minimal TSV pitch of 5 µm [33].

The bond itself serves multiple-purpose, such as mechanical integrity, electrical and thermal connectivity at low impedance. Four main bonding systems and combinations are proposed:

Metal-metal thermocompression bonds can be performed between extremely flat Cu-Cu surfaces. Microscopic contact points start to deform at elevated temperatures of 350 to 400°C under the applied pressure. The resulting inter-diffusion of copper across the interface in the subsequent 60 min annealing step under inert nitrogen atmosphere forms a low electrical impedance bond. Mechanical

[5]The active side of the strata carrying the transistors is called face.

integrity is achieved with a copper coverage > 13 %. Additional dummy copper area have to be added at a low number of interconnects, to improve mechanical strength, as well as thermal transfer between stratas [34].

Fusion bonding between dielectric layers such as SiO_2 are typically performed at high temperatures of 1000 °C. Initial VAN-DER-WAALS bonds from pre-bonding are replaced by covalent bonds desired for high bond-strength. Pre-surface treatments using plasma activation or hydroxyl surface termination by wet-chemical processes reduces the annealing temperature substantially to 400 °C. A proprietary process described by ZIPTRONIX using ammonia termination does not rely on heating at all. Covalent bonds are formed at room temperature without chemical byproducts such as water [35]. High surface quality is a prerequisite, achievable through careful CMP. The electrical contact is formed later by via-after-bonding processes or by hybrid bonding as in case of dielectric-dielectric bonds.

Thin film eutectic soldering takes advantage of the liquid metal phase existent during reflow, adjusting for roughness and non-planarity between the bond pads, therefore relaxing the surface quality requirements. Typical pressures of around 5 bar are applied. The most widely reported system is CuSn, discretely deposited using electroforming of tin on top of copper. Copper dissolves into the liquid tin and forms the eutectic phase at reflow temperatures of 260 °C. An increasing reflow dwell time results in solidification of the initially liquid phase. It is initiated by the increasing copper concentration resulting in a high-melting temperature intermetallic compound. This irreversible process allows the formation of bonds at low reflow temperatures, but results in bondlines with higher thermal stability. This enables assembly hierarchies required for sequential die stacking [36, 37, 38].

Adhesive bonding relays on spin-on thermoplastics such as polyimide or benzocyclobutenes (BCB). They become viscous at elevated temperature and form compliant bonds at around 350 °C and applied pressures of 5 bar. A major concern is the poor thermal conductivity of such bondlines of about 0.2 $\frac{\mathrm{W}}{\mathrm{m\,K}}$ rendering a high thermal barrier for heat conduction [29].

Hybrid bonding combines the advantages of metal-metal thermocompression and adhesive bonding. Openings on the metallic receptacles are formed by patterned intermediate adhesive and act as a self-alignment structures. Metallic protrusions from the counter-part are locked into these structures. The polymer film thickness is reduced and the metal bond is performed during thermocompression bonding [29]. “Direct bond interface” is an extension of the pure room temperature oxide fusion bond proposed by ZIPTRONIX. The interface consists of SiO_2 and copper pads in nearly coplanar arrangement. The pads are some nanometers recessed. Therefore a room temperature oxide bond can be achieved as described before. A compressive stress can be achieved on the metal pads due to an increased metal thickness expansion compared to the dielectric layer at elevated temperatures. This results in equivalent metal-to-metal bond quality as derived from thermocompression bonding [35].

To fulfill the requirements of the targeted product chip stack a stringent process sequence has to be defined from the broad variety of technologies available. Some processes will become an industry standard for given applications as technologies mature.

1.1.3 Vertical Integration Product Roadmap

Vertical integration will also have a significant impact on the supply-chain of semiconductor packages. Several players are capable to ramp-up high-volume fabrication of 3D specific technology. Foundries have the highest flexibility in TSV fabrication. Substrate manufacturers and OSATs are only capable to implement via-first or via-last approaches, respectively. Extensive alignment and bonding knowledge is already available at MEMS manufacturers and OSATs. For foundries, these processes are completely new. IDMs have the unique advantage to optimize all process steps during stack formation and do not rely on standardization of components. Whereas the major vertical integration break-through relys on

Application	Wafer starts	Wafer size
CIS	550k	8″, 12″
MEMS / Sensors	158k	6″, 8″
Power / RF	22k	6″
Memory	10k	12″
HB-LED	2.5k	8″

Table 1.1: 3D TSV market share of different applications by wafer starts in 2009 according to Yole Développement [39].

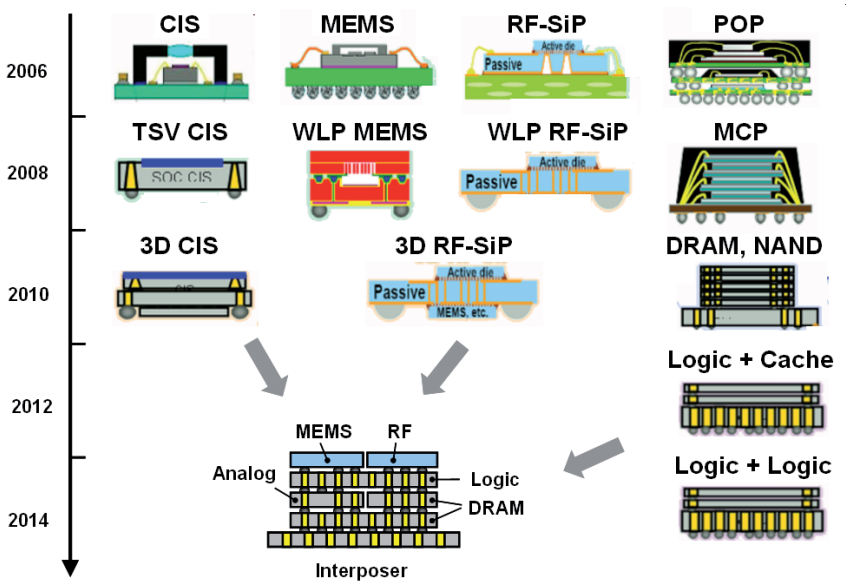


Figure 1.3: Vertical integration roadmap of diverging market segments such as CMOS image sensors, microelectromechanical systems, transceiver modules and memory / logic applications[39].

standardization of IC interfaces and 3D technology to allow stacking of dies designed and fabricated at different suppliers.

CMOS image sensors (CIS) were dominating the 3D TSV market in 2009, taking a share of nearly 3/4 by the number of wafer shipped according to a market study performed by Yole Développement [39]. MEMS sensor applications accounted for the other quarter. Power, radio frequency (RF), and memory applications, as well as high-brightness light-emitting diodes (HB-LEDs) with integrated TSV technology were brought to the market recently (Table 1.1).

Elpida announced to ship first vertical integrated 8-layer DRAM stacks in 2010. The next evolutionary step will be to implement such memory stacks side-by-side with a high-performance microprocessor on a silicon carrier, allowing for high density signal lines between the components [40]. Such system-in-package (SiP) assemblies will dominate the 3D TSV market from 2014 on according to market forecasts. A moderate 10 % share for memory-on-logic and logic-on-logic is predicted until 2015 (Figure 1.3).

About eighteen 300mm foundries are transforming to become TSV compatible world wide. This sites are equally distributed between North America, Europe, South East Asia, Korea and Japan. Taiwanese manufacturers take the lead in TSV fabrication by CIS production for mobile-phones up to now.

1.2 Thermal Management Concepts and Limitations

1.2.1 Power Dissipation Characteristics of IC-Dies

Power delivery and heat removal to and from ICs became the limiting factors for a second time in history of electronic components [41]. It caused a paradigm change in technology and system architecture. Heat fluxes of of bipolar mainframe dies exceeded heat removal capabilities of cold plates in the late

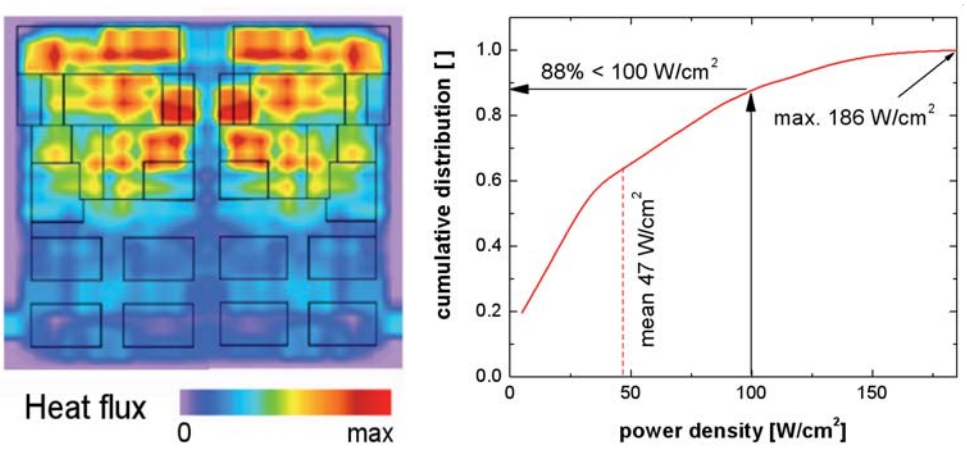


Figure 1.4: Measured power map of a duo-core processor at peak load using spatially-resolved infrared imaging [45] (left) (reprint with permission). cumulative distribution of the duo-core processor heat flux at peak load (right).

80ies. The industry was forced to change to energy efficient CMOS transistor technology and had to cope with a substantial cock-frequency drop. Fortunately, CMOS switches scale favorable and allowed ever increasing clock-frequencies at shrinking node dimensions in the following decade. The power dissipation at that times was caused primarily by the active power (P_{active}) of the transistors, which is a second-order function of the supply voltage (V_{dd}) and proportional to the gate capacitance (C) and clock frequency (f) [42]:

$$P_{active} \propto \frac{1}{2} C f V_{dd}^2. \tag{1.2}$$

The active power density turned out to be invariant down to 130 nm gate length even at increasing clock frequency. This, thanks to linear gate capacitance and supply voltage reduction with respect to the transistor physical dimension at constant-field scaling. Proportional V_{dd} to gate length scaling became impractical at smaller nodes due to parasitic effects. Passive power dissipation due to gate and subthreshold leakage became significant and account for nearly half the power dissipated in today's 45 nm devices [43, 44]. Due to the lack of alternative transistor technology the second paradigm change affected the microprocessor architecture. Clock frequency roadmaps to more than 10 GHz were abandoned and the multi-core era was entered. Total power consumption and clock-frequency are kept more or less constant. The microprocessor performance is now scaled by adding cores to the die. The next and third paradigm change will be a consequence of the multi-core architecture with it's increasing demand for cache size and bandwidth as discussed already in 1.1.1, leading to vertical integrated microprocessor - cache chip stacks in the near future.

The power map, displaying the spatial heat flux of a duo-core microprocessor was measured at peak load using spatially-resolved infrared imaging by HAMANN et al. [45] (Figure 1.4 left). High heat flux zones are caused by logic macros, compared to low power cache regions. The cumulative heat flux distribution is shown in Figure 1.4 (right). The peak heat flux of 186 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ is close to four times the average heat flux (47 $\frac{\mathrm{W}}{\mathrm{cm}^2}$). It is defined as the total power dissipated divided by the chip area. In this example, only 12 % of the chip area dissipates more than 100 $\frac{\mathrm{W}}{\mathrm{cm}^2}$. This strong heat flux contrast is responsible for local hot-spots influencing signal delay and differential aging of transistors.

1.2.2 Thermal Response of IC-Dies

Temperature control of ICs is key to guarantee their reliability and to operate them at a defined freqeuency [41]. Charge carrier mobility and electrical conductivity of metal wires is reduced at elevated temperature. This alters the switching time of transistors and signal transmission delay. Furthermore, device efficiency is degraded by an exponential leakage current increase with temperature [46]. Hence, sub-ambient cooling utilizing refrigeration loops was proposed. The additional power dissipated by these compression loops with a coefficient-of-performance (COP) of about 3 to 5 exceeds the power sav-

APPLICATION	T_{jmax} [°C]
High-performance logic	95
Low-performance logic	125
Memory devices	125
Handheld devices	125
Automotive electronics	175

Table 1.2: Maximal junction temperature T_{jmax} for various applications according to ITRS.

ings from reduced junction temperature by far [47]. Therefore, the opposing trend to improve system efficiency by hot water cooling in conjunction with low temperature gradient microchannel cold plates can be observed. It allows to eliminate all refrigeration loops in the datacenter and is called free-cooling mode. The high quality output heat at a level of 65 °C can be sold to neighborhood-heating-networks where it's re-used for space heating purpose. It offsets the carbon emission and reduces total-cost-of-ownership of the datacenter [48]. Efficient computing becomes significant as world-wide power consumption of datacenters increases due to the consolidation trend of information and communication technology (ICT) [49].

Component life time (L) depends on aging effects, such as electro migration, diffusion, relaxation, delamination, and voiding. Most of these mechanisms depend exponentially on temperature (T) as described by the ARRHENIUS law:

$$L(T) = A\left(e^{\frac{E_a}{kT}} - 1\right) \tag{1.3}$$

with a system-specific constant A, the activation energy E_a and BOLTZMANN constant k [50]. The maximum junction temperature is then defined by the accepted mean-time-between-failure (MTBF) of the application. Typical industry standards defined by ITRS are listed in Table 1.2.

1.2.3 Established Heat Removal Concepts

Today's microprocessors are mounted on a wiring board using flip-chip technology. Individual solder balls with a spacing $\geq 150\,\mu m$ serve as electrical interconnects for signaling or power delivery purpose. This leaves the entire backside of the die for heat removal. A copper lid adhesively attached to the laminate spans the silicon die and acts as thermal conductive and mechanical protective interface from die to heat sink. The ubiquitous availability and compatibility of air with electronic components renders air cooling as the heat removal technique of choice. Air fins increase the total surface area and reduce the convective thermal resistance from solid to air by a factor of 100. A chip heat flux of up to 70 $\frac{W}{cm^2}$ can be dissipated. Higher heat fluxes can be removed with improved heat spreading in the heat sink base. Water- or methanol-filled heat pipes or vapor chambers are used accordingly. Fluid evaporates on top of the microprocessor die and condenses in the cold plate periphery with temperatures below the fluids boiling point. The fluid transport to the heat source is performed by capillary forces in wicking structures. A critical heat flux[6] of around 140 $\frac{W}{cm^2}$ constitutes the upper performance limit of such devices. Most efficient heat removal in terms of exergy destruction (low temperature gradient) and minimal pumping power is performed by microchannel cold plates utilizing water as a coolant. Heat flux levels up to 400 $\frac{W}{cm^2}$ could be demonstrated thanks to the high volumetric heat capacity of water which is 4183 $\frac{kJ}{m^3\,K}$, compared to air with 1.27 $\frac{kJ}{m^3\,K}$. The fluid flow in the microchannel cold plate was divided into six parallel flow sections populated with staggered fins. A cold plate thermal resistance of 7 $\frac{K\,mm^2}{W}$ was measured at a pressure of 0.35 bar and a volumetric flow rate of 1.7 $\frac{l}{min}$ [51]

Thermal interface materials account for efficient heat conduction between adjacent solid parts, such as the silicon die and the copper cold plate. Particle filled oils or adhesives at concentrations above the percolation threshold result in effective thermal conductivities of up to 6 $\frac{W}{m\,K}$. These materials form a

[6] All liquid is evaporated in the wicking structure at the critical heat flux causing a sudden drop in heat transfer. A thermal runaway of the system with catastrophic failure is the result.

compliant interface between CTE mismatched components, reducing thermo-mechanical stress during thermal cycling. Effective thermal resistances of 6 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ can be achieved [52].

1.2.4 Back-side Heat Removal Limits of Vertical Integrated Chip Stacks

Contemporary 2D thermal packaging technology will be used for 3D chip stacks initially, to mitigate development risk and to leverage on existing manufacturing infrastructure. Hence, the back-side heat removal potential needs to be evaluated (Figure 1.5 left). Heat fluxes of the active stratas are accumulating in the chip stack. Additional thermal barriers such as BEOL layers and bonding interfaces have to be considered. The arrangement, as well as the stacking sequence of high heat flux macros and dies for non-uniform in-plane and die-to-die power dissipation becomes an important design parameter [53, 41].

A first statement about scalability of back-side heat removal is possible with a compact thermal model assuming one-dimensional heat flux [54]. The chip stack is discretized into sub-components and is represented as equivalent resistor network with attached current sources (Figure 1.5 right). The linear equationsystem is solved using the GAUSS JORDAN elimination algorithm to derive individual junction temperatures T_{jn}.

The wiring layer thermal resistance R_b is calculated for a microprocessor and memory die considering BEOL dimensions for 45 nm node technology as reported by ITRS[7]. Effective-medium-theory is applied to define the effective thermal conductivity k_{eff_i} of individual wiring layers i. They are composed of copper k_{Cu} wires and vias with an area fill factor f and the inter-metal and inter-layer dielectrics k_{IMD} and k_{ILD}, respectively:

$$k_{eff} = f k_{Cu} + (1 - f) k_{ILD}. \tag{1.4}$$

The effective thermal resistance of individual BEOL layers is equal to the thickness t to effective thermal resistances ratio. Their sum

$$R_b = \sum_{i=1}^{n} \frac{t_i}{k_{eff_i}} \tag{1.5}$$

results in a thermal impedance of 1.43 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ and 3.50 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ for the memory and microprocessor wiring BEOL layers (Table 1.3).

A state-of-the-art thermal interface resistance of 6 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ and a cold-plate thermal resistance of 7 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ were used as reported in section 1.2.3. The thickness of the thinned dies was assumed to be 100 µm, compared to the 500 µm thick die interfacing the cold plate. Typical heat flux values (section 1.2.1) for memory stratas and microprocessor cache and logic (hot-spots) areas of 10, 60, and 240 $\frac{\mathrm{W}}{\mathrm{cm^2}}$ are assumed, respectively. The maximum acceptable junction temperature T_{jmax} of the microprocessor unit (MPU) is set to 80 °C at a fluid inlet temperature T_{fin} of 20 °C. The memory temperature limit is set to 95 °C. Critical temperatures for memory-logic and logic-logic stack configurations are calculated and depicted in Table 1.4 and visualized in Figure 1.6, respectively.

The temperature budget is not exceeded for a single MPU and multiple memory layers. MPU stacks of two dies can only be tolerated with misaligned high heat flux macros. In general, it is advantageous to place the high heat flux strata (MPU) close to the cold plate. Additionally, stacking of congruent high power dies should be prevented. It is questionable if these thermal design rules match with electrical needs. A large amount of TSVs on a regular grid have to be placed in the bottom dies to deliver the current to the top MPU[8]. This TSV grid causes a substantial silicon real-estate loss and constrains macro placement in the lower dies [55]. It demonstrates clearly - electrical and thermal design-rules are contradictory and trade-offs altering system performance are inevitable. Only marginal return-on-investments can be expected from 3D technology if thermal constraints are limiting the electrical design and therefore the device performance.

Hence, new, scalable heat-removal concepts have to be developed. In this context the proposal to include thermal vias on the die to improve heat conduction across large thermal barriers in the stack sounds like a drop in the ocean (besides adding additional routing congestion) [56].

[7] http://public.itrs.net

[8] Up to 3/4 of the electrical interconnects to a 2D MPU are today used to deliver power. This number will increase with further supply voltage reduction, unless on-chip step-down voltage conversion becomes feasible.

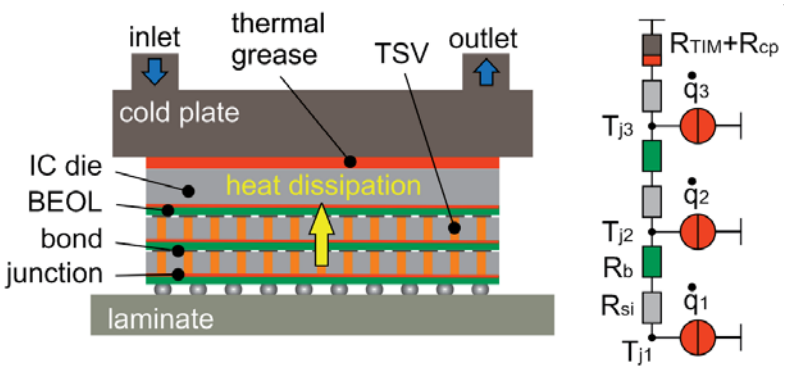


Figure 1.5: Back-side heat removal schematic illustrating chip stack with attached cold plate (left). Analogous resistor network representing a three-die stack with individual thermal resistances for the silicon substrate R_{Si}, the BEOL wiring layers R_b, the thermal interface R_{TIM}, and the cold plate R_{cp}. The heat flux is imposed by current sources $\dot{q}_n$ with resulting junction temperatures T_{jn} (right).

BEOL Layers	Microprocessor	Memory
Global / intermediate / metal 1 layers [count]	3x / 8x / 1x	1x / 2x / 1x
Total thickness [nm]	4180	2037
Effective thermal resistance [$\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$]	3.499	1.434

Table 1.3: Number of wiring levels in the BEOL layers and its effective thermal resistance according to ITRS dimensions for a 45 nm technology node microprocessor and memory die.

MPU Budget: $\Delta T_{jmax-fin}$ = 60 K	$\Delta T_{jmax-fin}$ [K]
11x memory / MPU logic / cold plate	59.0
MPU logic / 2x memory / cold plate	54.4
MPU logic / MPU cache / cold plate	60.8
2x MPU logic / cold plate	111.1

Table 1.4: Temperature gradient from fluid inlet to maximal junction temperature of the MPU ($\Delta T_{jmax-fin}$) for different chip stack configurations.

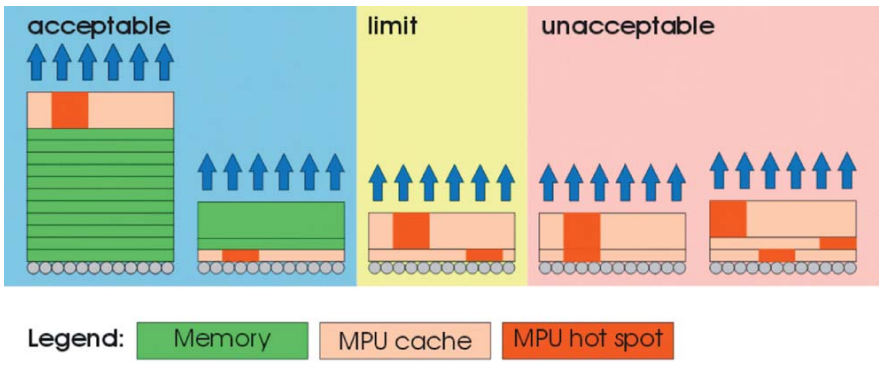


Figure 1.6: Chip stack configuration with acceptable, close to the limit and unacceptable junction temperatures for a 60 K thermal budget.

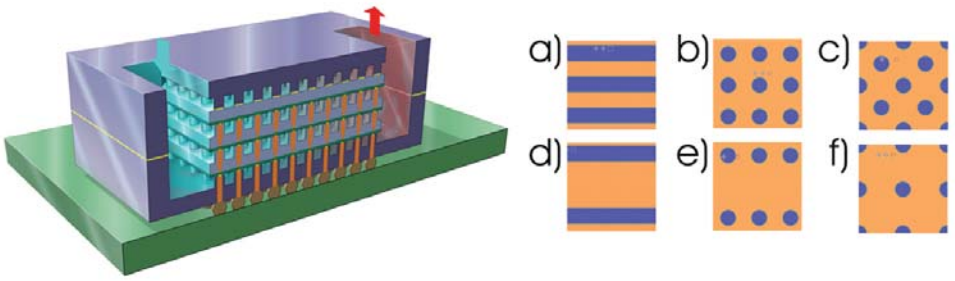


Figure 1.7: Artistic view of a interlayer cooled chip stack using forced convection (left). TSV-compatible heat transfer structures (blue: pin or fin), a) microchannel, b) pin-fin in-line, c) pin-fin staggered, d)-f) analogous structures, but half-populated (hp) instead of full-populated (fp) (right).

1.2.5 Interlayer Thermal Management

Volumetric heat removal concepts scalable with the number of stratas have to be investigated to keep 3D integration a successor of 2D MPUs with a enough room for improvement for several generations. This approach can potentially be performed by lateral heat removal between active dies relying on conduction or convection. Conductive layers between dies smooth out heat flux peaks and can dissipate the heat laterally out of the package. Layer thicknesses of more than 1 mm are needed to be thermally effective [57]. This limits the technology to peripherally-routed or low vertical interconnect density 3D packages and small die sizes (below $1\,cm^2$). Alternatively, thin form-factor vapor chambers may be implemented between stacked dies [58]. So far, the implementation of vertical-interconnects has not been demonstrated and the thermal performance was marginal. Both methods are not compatible with high interconnect densities and heat fluxes in large footprint chip size stacks. Only forced convection seams to be a feasible candidate for volumetric heat removal in high-performance vertically integrated chip stacks.

1.3 Convective Interlayer Heat Removal - a Scalable Concept

A coolant is forced by a pressure gradient through embedded fluid cavities between individual dies in the chip stack in case of convective interlayer cooling (Figure 1.7 left). The fluid cavities are defined by adjacent dies and solder ball arrays or fluid geometries etched into the backside of one of the silicon dies. The heat transfer structures need to be compatible with the area array arrangement of the electrical interconnects. Individual TSVs are integrated into micochannel walls or pins of pin-fin arrays (Figure 1.7 right). Dielectric coolants can be used without the need of electrical insulation, whereas hermetically sealing concepts are important to prevent hydrolysis or electrochemical corrosion if water is used.

Basic convective interlayer cooling was investigated by several researchers [59, 60, 61]. TAKAHASHI demonstrated heat removal of $25\,\frac{W}{cm^2}$ for die stacks with peripheral interconnects and with water as coolant. Dies with a size of 10 mm formed a parallel-plate heat-transfer configuration at a spacing of 10 µm. CHEN investigated dielectric coolants (FC-77) at a die spacing of 50 µm resulting in $50\,\frac{W}{cm^2}$ heat dissipation. Parallel-plate heat transfer geometries are not feasible for high-performance applications with the need for high density area array TSVs and the associated high heat-fluxes. KOO concluded with a critical heat flux of $140\,\frac{W}{cm^2}$ for a $2\,cm^2$ chip stack using two-phase heat transfer in microchannels and water as refrigerant. 300 µm high channels with a pitch of 800 µm were considered, resulting in a moderate TSV density. Integrated coolant delivery through the printed circuit board (PCB) was demonstrated by DANG et al. [62]. The fluid manifold consists of holes with a 50 µm diameter etched through the dies, resulting in a 2.5% silicon real-estate loss and a significant additional pressure drop. Buried microchannels were proposed. They were etched into the silicon die backside and where filled temporarily with a sacrificial spin-on polymer. Porous SiO_2 was deposited on top of the channels by

plasma-enhanced chemical vapor deposition (PECVD), to form the microchannel cap after a planarization step . The sacrificial polymer decomposed during a subsequent annealing step, clearing the channels. Finally, the nano-pores of the SiO_2 layer are sealed by an additional dielectric layer. Water at a relative pressure of 1 bar could be pumped through the capillaries.

The listed studies assumed all uniform power dissipation and basic heat transfer structures with opposite in- and outlets, resulting in poor performance with respect to heat removal and/or interconnect density. Additional innovations are necessary to make interlayer cooling a serious candidate for the heat removal of future vertically integrated high performance microprocessor chip stacks.

1.3.1 Implementation Concepts: Via Sealing and Pressure Balanced Package

A pressure balanced package design is proposed to prevent tensile or shear stress in the chip stack imposed by the pressure gradient in the fluid. A manifold and a silicon interposer (with underfilled solder ball array support) form the fluid containment supporting the pressure difference between fluid and ambient. The symmetric arrangement of the fluid cavities results in equal local pressures on both sides of the die (Figure 1.8 left). Hence, only compressive stresses are induced into the silicon dies and bond interfaces. Additionally, no fluid delivery structure needs to be implemented into the chip stack, allowing non-uniform die sizes and a maximal use of silicon die area.

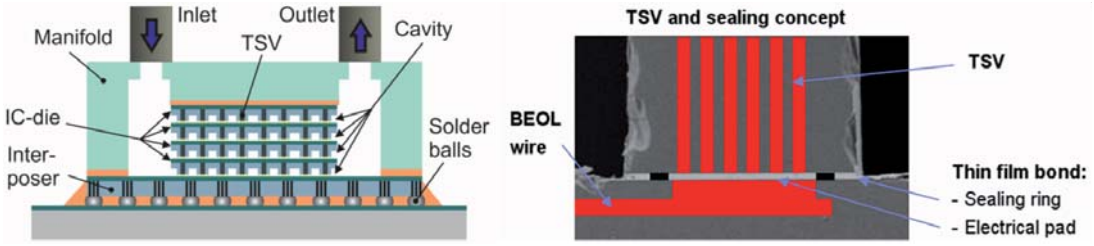


Figure 1.8: Pressure balanced interlayer cooling package at face-to-back stacking sequence (left). Cross-section of the proposed ring-pad TSV sealing architecture (right).

A face-to-back stacking sequence of the n stratas results in $n+1$ cavities. Whereas only $n/2+1$ fluid cavities at double the height exist for back-to-back / face-to-face arrangements. Additionally, the transistors and BEOL wiring levels are spatially separated from the fluid by the silicon slab (approximately 50 µm thickness) and not only by a 1 to 10 µm thick dielectric layer. Hence, reduced signal damping due to the coolant proximity (especially important in case of water with a high permittivity and loss tangent at high frequencies) without the need for shielding can be expected.

Potential microchannel or pin-fin aspect ratios are defined by TSV fabrication limits, as discussed in section 1.1.2. CVD tungsten vias with heights of 150 µm were demonstrated. With advances in technology 300 µm might be possible in the near future. Electroplated copper vias are today processed with a 6:1 aspect ratio, resulting in diameters of 25 µm for via heights of 150 µm. Again, with advancing technology 30 µm diameter for 300 µm height seam to be feasible. In this study a minimal silicon thickness of 50 to 100 µm was considered to assure mechanical integrity, leaving room for cavity heights of 100 to 200 µm.

A ring-pad sealing concept is proposed to enable the use of water as a coolant (Figure 1.8 right): a bar shaped TSV is embedded in a silicon post or rib. A thin film solder pad connects the BEOL layer of the lower strata electrically, mechanically, and thermally to the TSV. It is surrounded by a thin film solder sealing ring, separating the fluid from the electrical interconnect. The silicon structure finally acts as a pin fin or fin in case of microchannels, transferring heat from solid to the fluid and from die-to-die throughout the chip stack. The sealing ring adds 20 µm to the 30 µm diameter via, resulting in a final 50 µm post diameter or wall width. This leaves a height to diameter/width aspect ratio range for the heat transfer structure of 2:1 to 4:1.

The arrangement of TSVs is assumed to be periodic and equidistant in x and y direction (fully populated (fp)) (Figure 1.7 right a-c). At locations with lower bandwidth need, interconnect densities can be reduced, and half-populated (hp) fluid structures are considered (Figure 1.7 right d-f).

As chip stack footprint, the range from 1 to 4 cm^2 is relevant for high performance microprocessor stacks and therefore was explored in this study.

1.3.2 Internal Flow Heat Transfer Characteristics: a Sensitivity Analysis at constant Geometrical Ratio

A basic scaling analysis assuming uniform heat flux and heat transfer cavity for each strata is performed to understand the characteristics and to identify significant parameters in interlayer cooling. The junction temperature response from fluid inlet to outlet in internal cross-flow heat exchange mode with the simplification of one-dimensional heat conduction is shown qualitatively in Figure 1.9 (left). Heat conduction through the silicon and the BEOL layers causes a constant temperature drop $\triangle T_{cond}$. The convective heat transfer coefficient is maximal at the fluid inlet and is reduced, approaching an asymptotic value after the entrance-length, where fully-developed boundary layers exist. Consequently, the convective thermal gradient $\triangle T_{conv}$ increases and converges to a constant value. The linear increase of the coolant temperature $\triangle T_{heat}$ is caused by the accumulation of sensible heat in the fluid. Hence, the maximum junction temperature T_{jmax} is located at the fluid exit.

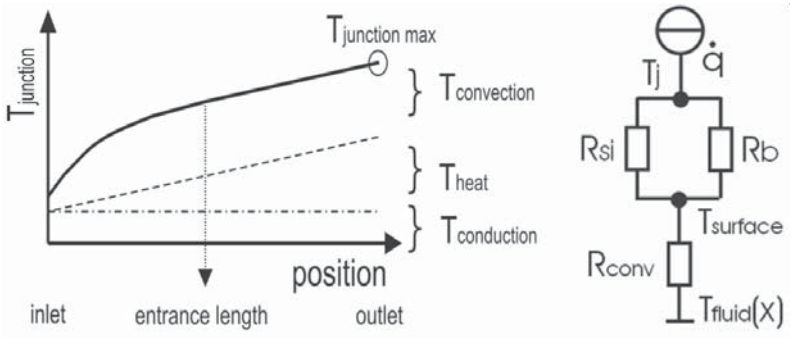


Figure 1.9: Junction temperature response from inlet to outlet with uniform heat flux and cross-flow heat-exchange mode (left). Analogous resistor network of the heat transfer configuration considering symmetric, double-side heat removal (right).

A quantitative parametric study using correlation-based prediction for microchannels and numerical modeling for staggered pin fin structures was performed to identify feasible TSV interconnect density ranges. In this gedankenexperiment we assume a infinite number of stratas resulting in periodic thermal boundary conditions with double-side heat flux conditions for an individual cavity. 100% fin efficiency is assumed and can be justified for the low aspect ratio structures proposed (Table 1.5). The critical heat flux $\dot{q}_{critical}$ was computed using an equivalent resistor network (Figure 1.9 right) considering the maximal junction temperature $T_{jmax-in}$ budget of 60 K. The total thermal resistance R_{tot} can be defined as the sum of conductive R_{cond}, convective R_{conv} and fluid thermal resistance R_{heat} as proposed by TUCKERMAN and PEACE in [63]:

$$R_{tot} = R_{cond} + R_{conv} + R_{heat}, \tag{1.6}$$

and in more details:

$$R_{tot} = \frac{1}{\frac{k_{Si}}{t} + \frac{1}{R_b}} + \frac{A_{heater}}{hA_t} + \frac{A_{heater}}{\dot{V}\rho c_p}, \tag{1.7}$$

with the silicon thermal conductivity k_{Si}, the silicon slab thickness t, the effective thermal resistance of the BEOL layers R_b, the chip stack squared footprint $A_{heater} = l^2$, the total wetted surface area $A_t = 2(w_c + h_c)\frac{l^2}{w_c + w_w}$ and average heat transfer coefficient h, the volumetric flow rate $\dot{V}$, the water density ρ and heat capacity c_p.

INCROPERA [64] derived the following correlation for the NUSSELT number $Nu = \frac{hd_h}{k_f}$ at fully-developed flows with heat transfer across all four microchannel sidewalls with channel height h_c, channel width w_c, and axial heat flux boundary condition:

$$Nu = 3 + \frac{1}{\frac{1}{AR_c} + 0.2} \, with \, AR_c = \frac{h_c}{w_c} > 1. \tag{1.8}$$

The DARCY friction factor fr REYNOLDS number $Re = \frac{vd_h}{v}$ product for rectangular channels with fully-developed hydrodynamic boundary layer at an average bulk velocity v and hydraulic diameter $d_h = \frac{4h_c w_c}{2(h_c+w_c)}$ is defined with a fourth-order polynomial:

$$frRe = a_4 AR_c^4 + a_3 AR_c^3 + a_2 AR_c^2 + a_1 AR_c + a_0, \tag{1.9}$$

with coefficients

$$a_0 = 23.992, \, a_1 = -32.169, \, a_2 = 43.147, \, a_3 = -29.464, \, a_4 = 8.722. \tag{1.10}$$

The pressure drop results from:

$$\triangle P = 4 \frac{frRe}{Re} \frac{l}{d_h} \frac{\rho v^2}{2} \tag{1.11}$$

with ρ as the fluid density.

The pressure drop $\triangle P$ and convective thermal resistance R_{conv} of low-aspect-ratio staggered pin-fins with equal x- and y-spacing and double-side-heat flux was derived through numerical conjugate heat and mass transfer modeling (details are revealed in subsection 2.3.1) due to the lack of reliable correlations.

A chip stack with a 2 cm^2 footprint, heat transfer structure aspect ratio h_c/w_w of 3:1 and TSV-pitch to channel wall ratio of 4:1 are used in the parametric study. The fin efficiency of these heat transfer structures at the given boundary conditions is in all cases > 95% and renders the assumption of isothermal fin temperature as valid. Typical server applicaiton cold plate operation with 30 kPa pressure head and a thermal budget of 60 K is considered (Table 1.5).

PARAMETERS OF HEAT-TRANSFER STRUCTURE	VALUES
Stack footprint (l^2)	2 cm^2
Silicon slab thickness (t)	100 µm
Pitch / diameter ($(w_w+w_c)/w_w$)	4:1
Height / diameter (h_c/w_w)	3:1
Pressure drop from inlet to outlet $\triangle P$	30 kPa
Thermal budget $\triangle T_{jmax-in}$	60 K
Volumetric flow rate per cavity ($\dot{V}$)	< 1 $\frac{l}{min}$

Benchmark: 200 µm pitch, 50 µm wall width / pin diameter, 150 µm cavity height

Table 1.5: Attributes considered in the parametric study.

Two heat transfer modes could be identified. They are separated by a global heat-flux maximum (Figure 1.10). For microchannels (MC) with pitches > 200 µm, $\triangle T_{conv}$ dominates $\triangle T_{tot}$, whereas for relevant pitches below 200 µm $\triangle T_{heat}$ is dominating the thermal budget. This makes sense by considering the first-order dependence of the thermal gradients on the hydraulic diameter, which is

$$\triangle T_{conv} \propto \frac{\dot{q} d_h}{k_f Nu}, \, but \tag{1.12}$$

$$\triangle T_{heat} = \frac{\dot{Q}}{\dot{V} \rho c_p} \propto \frac{\dot{q} l^2 \mu}{d_h^3 \triangle P \rho c_p}, \tag{1.13}$$

for a given pressure drop $\triangle P$. The convective thermal resistance scales linear with the hydraulic diameter and becomes therefore significant at large TSV pitches. The fluid temperature increase is inverse proportional to the third power of d_h, therefore dominant at small TSV-pitches. The superposition

of both effects result in the global heat flux maximum, which is shifted to larger pitches for pin-fin staggered (PFS) structures with improved heat-transfer coefficient, but increased friction factor. Hence, the heat-transfer structure should be optimized for maximal mass flow rate at interlayer cooling relevant pitches smaller than 200 µm (defined by interconnect density needs, section 1.1.1). This stands in contrast to conventional back-side cold plates with the convective thermal resistance as the main optimization criteria. These heat exchangers are operated at a 10-fold increased volumetric flow rate thanks to reduced fin height constraints. A non-significant temperature increase of 2 to 4 °C results.

The study also demonstrates clearly the limits of interlayer cooling if basic heat removal cavities and moderate pressure heads are considered. The heat removal capability drops from 150 to 50 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ per strata by reducing the pitch from 200 to 100 µm. Hence, more sophisticated heat removal concepts are introduced in the following section 1.3.3 and an electro-thermal co-design is proposed in section 1.3.4 extending interlayer cooling to 50 µm interconnect pitches. For larger die footprints this building-blocks are even more important: the fluid temperature increases proportional to the power of two with respect to the die length at a given pressure boundary condition.

A summary of the first-order sensitivity analysis is presented in Table 1.6. The influence of the coolants thermo-physical properties is also listed. The utilization of dielectric coolants such a fluorinerts (e.g. FC-77) would render electrical insulation obsolete, reducing package complexity. Unfortunately, their viscosity is larger and the volumetric heat capacity is less compared with water. Hence, FC-77 as a representative dielectric fluid performs about three times worse than water in the fluid temperature dominant heat transfer regime (Table 1.7). Therefore, dielectric coolants are not an alternative to water in single-phase cooling. They might be an option in two-phase heat transfer mode using their latent heat of vaporization to minimize the needed volumetric flow rate. Engineered fluids with suspended nanoparticles in water (so called nanofluids) were studied intensively in recent years with the hope to surpass thermal properties of water. Heat conduction benefits as large as 12 % for silica nanoparticles suspended in water at a volumetric fill of 16 % were reported, improving convective heat transfer at constant volumetric flow rate. This benefit comes with dramatic increase in dynamic viscosity of a factor of three at drop in coolant heat capacity of 26 % degrading the coolant mass flow rate at pressure drop boundary conditions accordingly. Hence, reduced heat transfer capabilities compared to water were measured in backside cold plates [65, 66, 67]. The performance drop would be even worse for interlayer cooling with smaller hydraulic diameters.

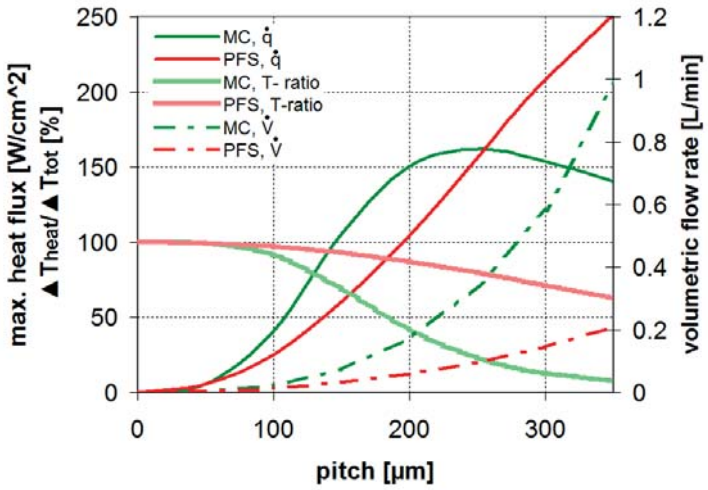


Figure 1.10: Heat-removal capability versus interconnect pitch at constant aspect ratios and uniform heat flux $\dot{q}$ for microchannel (MC) and pin-fin staggered (PFS) geometries are depicted at REYNOLDS numbers < 600. Additionally, the corresponding temperature-gradient ratio $\triangle T_{heat} / \triangle T_{tot}$ and the volumetric flow rate $\dot{V}$ are shown.

PARAMETER	CRITICAL HEAT FLUX SCALING
Die size (l)	a^{-2}
Hydraulic diameter (d_h)	a^3
Fluid thermal conductivity (k_f)	1
Fluid volumetric heat capacity (ρc_p)	a
Fluid dynamic viscosity (μ)	a^{-1}
Pressure drop from inlet to outlet ($\triangle P$)	a
Thermal budget ($\triangle T_{jmax-in}$)	a

Table 1.6: Critical heat flux dependency on individual parameters in case of fluid temperature dominant heat transfer in microchannel cavities.

THERMO-PHYSICAL PROPERTIES AT 25 °C	WATER	FC-77	NANOFLUID
Thermal conductivity [$\frac{\mathrm{W}}{\mathrm{m\,K}}$]	0.607	0.061	0.728
Volumetric heat capacity [10^6 $\frac{\mathrm{kJ}}{\mathrm{m^3\,K}}$]	4.170	1.958	3.086
Dynamic viscosity [10^{-6} Pa s]	890.1	1282.0	3000

Table 1.7: Thermo-physical properties of potential coolants at 25 °C. Nanofluid: water with 16 vol% suspended silica nanoparticles [67].

1.3.3 Modular Heat and Mass Transfer Building Blocks

Power map aware heat removal was proposed for backside cold plates to mitigate hot-spots on 2D microprocessors, improve pump efficiency, and to minimize exergy destruction. The benefit compared to uniform heat removal is marginal due to the high thermal resistance of the thermal interface between the die and cold plate. It results in a close to uniform heat flux distribution at this interface as a consequence of heat spreading in the 780 µm thick silicon die. Heat spreading is minimal for a theoretical package without thermal interface and a silicon die thickness of 1 µm resulting in a volumetric flow rate reduction of 78 % at modulated heat transfer in the cold plate [68]. Interlayer cooling is an implementation of this near-junction heat removal at minimal silicon slab thickness and the fluid in direct contact with its surface. Hence, hot-spot aware heat removal should be considered as discussed in [69]. Its foundation being the following heat and mass transfer building blocks.

Unit-Cell Geometry Optimization

The local fluid friction and heat transfer coefficient depends on the geometry of the elementary heat transfer unit-cell. It is defined by the parallel endwall pair separated by the cavity height h_c in z-direction. High-aspect-ratio microfabrication technologies, such as deep reactive ion etching (DRIE), allow the incorporation of fin geometries with arbitrary shapes in x- and y-direction. The minimal geometrical requirement is to completely embed the cylindrical TSV in the silicon fin. Continuous and discontinuous topologies are the basic fin types resulting in non-communicating and communicating cavities (Figure 1.11). Microchannels and pin-fin arrays are two representations of each type[9]. The optimal local fin shape has to be defined according to the cavity length, the interconnect density, and the power map. At chip stack footprints larger than 1 cm^2, uniform heat transfer and TSV pitches ≤ 200 µm unit-cells with minimal fluid friction result in best system performance (section 1.3.2). Hence, parallel plates would be the optimal unit-cell structure. As a consequence, fins should only be placed at locations where TSV are needed. Therefore, fully and half-populated pin and microchannel arrays are considered in this study (as described in section 1.3.1). They reflect die locations with high and low communication and power demand.

The previous scaling exercise (Figure 1.10) indicated the relevant REYNOLDS number Re regime for interlayer cooling being between 50 and 600. It allows the assumption of laminar flows in the cavity.

[9] Pressure balancing transverse to the flow direction occurs in communicating cavities. The flow field is non-trivial compared to non-communicating fin structures. The fluid will propagate predominantly along the path of least flow resistance.

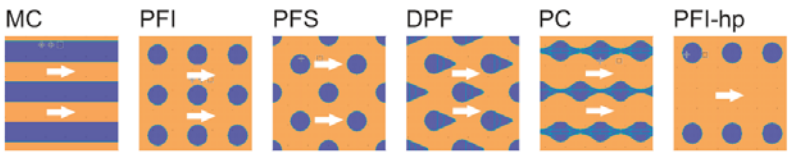


Figure 1.11: Example of unit-cell geometries, including non-communicating (MC, PC) and communicating structures (PFI, PFS, DPF).

The pressure drop in internal flows with fins in place is composed of viscous and form drag. At low REYNOLDS numbers the viscous drag is dominant. It is a consequence of developing boundary layers caused by no-slip conditions at the cavity wall. It scales with Re and the surface area exposed to the fluid. Non-linear inertia terms resulting in form drag become important with increasing Re and obstacles redirecting the fluid. Hence, different geometrical parameters are of concern depending on the considered flow regime. The cavity surface area needs to be minimized to optimize viscous drag. The fin shape and its dimension transversal to the flow direction have to be considered at larger Re. Drop-like shapes with minimal projected area might mitigate the formation of dissipative wakes behind the fin. Whereas, square pin shapes are most prone to wake formation, followed by circular pin geometries [70, 71, 72]. Inertia terms in microchannel flows become only important in the turbulent regime at $Re > 2300$.

Heat Transfer Structure Modulation

Unit-cell geometry variation is one possibility to account for local heat flux changes. It could be the variation of channel width (Figure 1.13 left), pin diameter or the pitch. Switching between different uni-cell types is also feasible. The pumping power of a cavity with uniform and a modulated channel width is compared in the following example. Both cases have to comply with the thermal budget of 65 K for the given one dimensional power map with varying heat flux in longitudinal flow direction (Figure 1.14 left). Compact thermal modeling with one dimensional heat flux is used to derive junction temperatures (as discussed in subsection 1.3.2). The convective thermal resistance and the pressure gradient versus channel width ratio can be computed considering Equation 1.7, 1.8, 1.9 and is presented in Figure 1.12 for a 1 cm^2 cavity footprint, 150 µm channel height and 0.1 $\frac{l}{min}$ volumetric flow rate. The implementation of TSVs in the channel wall constrains the width-to-pitch ratio $w_c/pitch$ to the range of 0 to 0.5. The microchannel convective thermal resistance is independent of the fluid velocity if fully developed boundary layers and laminar flow are considered. It increases close to proportional with respect to the channel width-to-pitch ratio up to a value of 0.5 due to the hyperbolic dependency of the heat transfer coefficient with respect to w_c. The convective resistance drops again for the 200 µm pitch example at channel widths beyond the 0.5 width-to-pitch limit due to the increase in wetted area (Figure 1.12 left). The pressure gradient depicted in Figure 1.12 (right) drops dramatically with increasing channel width for the given volumetric flow rate. The cost to improve the local heat transfer is enormous and should therefore only be applied at specific locations to cap the junction temperature if needed.

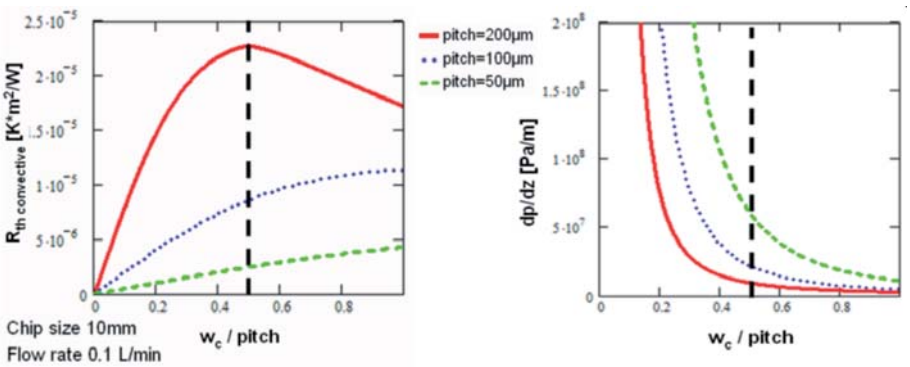


Figure 1.12: Effective microchannel convective thermal impedance vs. microchannel width-to-pitch ratio at a channel height of 150 µm (left). Pressure gradient at a constant volumetric flow rate for different pitches and varying width-to-pitch ratios (right).

A TSV pitch of 200 µm and a fluid cavity height of 150 µm is chosen for the uniform versus modulated microchannel test case considering a 1 cm^2 chip stack footprint. The maximal allowed channel width is 100 µm and turned out to be the optimal width with respect to pumping power for the uniform channel case. The flow rate was adjusted to utilize the thermal budget available completely.

The optimal local channel width is defined numerically by the "Channel-Walk Method" in case of the modulated microchannel cavity (Figure 1.13 right). The channel is discretized into equidistant segments in the flow direction resulting in n unit-cells with a size of A_{cell}. The local fluid temperature along the channel can be derived with

$$T_{f(n)} = T_{f(n-1)} + \frac{\dot{q}_{n-1} A_{cell}}{\dot{V} \rho c_p}, \tag{1.14}$$

assuming a given flow rate $\dot{V}$ and one-dimensional heat flux.

The maximal convective thermal resistance for each discretized cell is then computed by subtracting the conductive temperature gradient and fluid temperature increase from the total thermal budget, multiplied by the local heat flux as shown in the following equation:

$$R_{convmax(n)} = \dot{q}_n (\triangle T_{jmax-in} - (T_{f(n)} - T_{f(1)}) - R_{cond} \tag{1.15}$$

The local channel width ($w_{c(n)}$), is modulated according to the following cases:

$$w_{c(n)} = \begin{cases} w_{cmax} = \frac{pitch}{2} & \text{for } R_{conv}(w_{cmax}) \leq R_{convmax(n)} \\ f(R_{convmax(n)}) & \text{for } R_{conv}(w_{cmax}) > R_{convmax(n)} \end{cases} \tag{1.16}$$

considering the maximal channel width-to-pitch ratio of 0.5. The total pressure drop is then defined for the resulting geometry. This sequence is repeated for a range of flow rates. The geometry of the run with minimal pumping power requirement is then selected to be optimal.

The most efficient channel geometry compared to the uniform channel case with their associated pressure drops are presented in Figure 1.14 (right). Channel widths below the 100 µm constraint are implemented only on high heat flux locations with a monotonic decrease towards the outlet due to the increasing fluid temperature. The junction temperature responses are depicted in Figure 1.14 (left). Overall, the pressure and the pumping power requirements are about 2 and 5 times lower than for the optimal uniform channel-width geometry (Table 1.8).

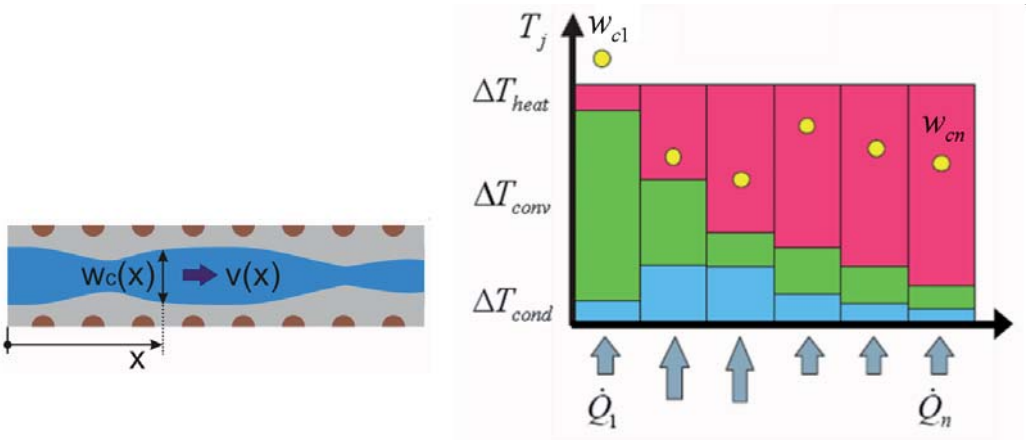


Figure 1.13: Top view of a modulated microchannel with the constraining electrical interconnects indicated in orange (left). Graph indicating the optimization routine: The channel width $w_{c(n)}$ influencing the convective temperature gradient is modulated to meet the junction temperature limit on all chip locations (right).

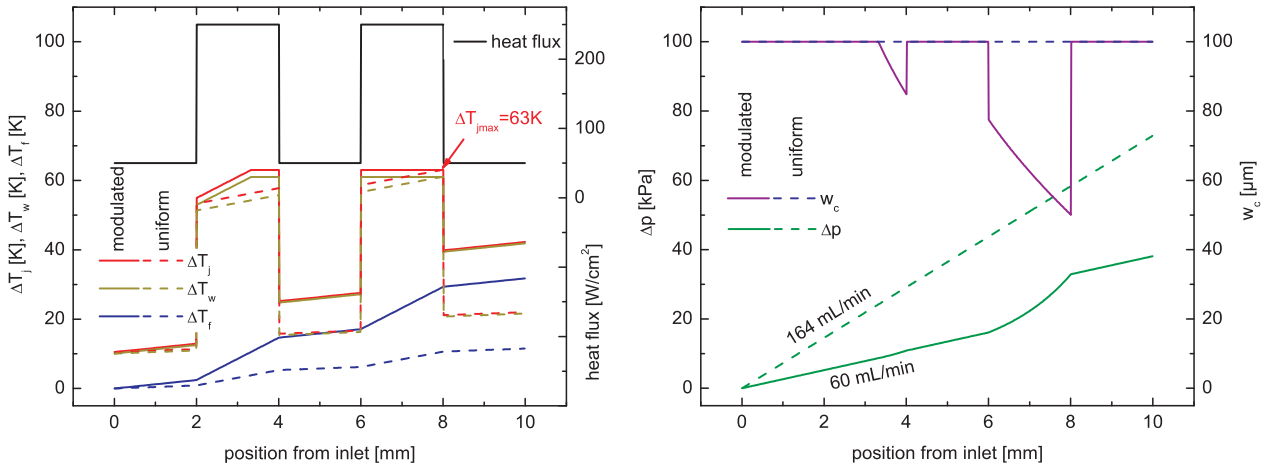


Figure 1.14: One dimensional power map with varying heat flux in flow direction. 40 % hot-spot area and a heat flux ratio (hot-spot to background) of 5 is considered in the test case. Resulting temperatures for the uniform and modulated channel width case are plotted as well (left). The optimal channel width when considering developed boundary conditions only and at a channel-width constraint of half the interconnect pitch is depicted. The resulting pressure drop is compared for the uniform and modulated channel width case (right).

Parameters	Uniform	Modulated
Maximum width [µm]	100	100
Minimum width [µm]	100	50.6
Pressure drop [kPa]	72	38
Volumetric flow rate [$\frac{ml}{min}$]	164	59.5
Pumping power [mW]	199.8	37.8

Table 1.8: Comparison of optimal microchannel structure at uniform and modulated channel width.

Fluid Focusing / Defocusing towards and from Hot-Spots

Fluid "focusing" and "defocusing" is another option to mitigate hot-spots by increasing local fluid velocity. This can be performed by branching and combining junctions in microchannel networks (Figure 1.15). Unit-cell modulation results in flow redistribution in case of communicating heat transfer geometries. Additional guiding structures (such as walls) can also be implemented. Finally, a low pressure drop path to deliver and drain coolant to and from the hot-spot location exits. An increased local flow rate results mitigating the local fluid temperature increase and reducing the convective thermal resistance in case of short microchannels and pin-fin arrays. In general, elements redirecting the fluid to specific locations are associated with increased fluid friction and result in a drop in total volumetric flow rate. Hence, fluid focusing only pays-off in case of high heat flux contrast power dissipation. Furthermore, a balanced cavity design is important to prevent coolant starvation at low heat flux locations.

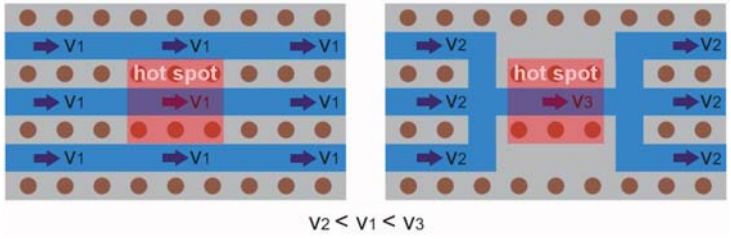


Figure 1.15: Uniform (left) and focusing / defocusing fluid delivery towards and from a hot-spot location (right).

Peripheral Fluid Delivery: Four-Port Architecture

All four sides of the chip stack periphery should be used as in- and outlets to maximize the total coolant delivery to the fluid cavity. This configuration is named as four-port (4-port) architecture (Figure 1.16 right). In the previous examples only two opposing sides of the stack were used for fluid delivery and is called two-port (2-port) architecture, accordingly (Figure 1.16 left). The channel length in a four-port with opposing in- and outlets increases linear from cavity corner to its center line. Its maximal channel length is equal to the length of all channels in the two-port at a square chip stack footprint. The four-port channel flow rate increases hyperbolically towards the chip corner at pressure-boundary conditions. This renders the four-port fluid delivery most beneficial in case of hot spots in corner zones. A quad-core test case (four core microprocessor) using microchannels as heat-transfer structures at a fixed channel width is compared with a two-port configuration (Figure 1.17 left). A chip size of 20 by 20 mm^2, a channel height and width of 100 µm, with a interconnect pitch of 200 µm are chosen. The maximum allowed junction to fluid inlet temperature increase is set to 60 K. The peak heat flux is 238 $\frac{W}{cm^2}$ at a total power dissipation of 311 W. The power map is discretized and the conductive and convective temperature gradients and the fluid temperature are computed for a given pressure boundary condition, as described in the "Channel-Walk Method" (subsection 1.3.3). The resulting maximal junction temperature increase is compared with the 60 K temperature budget. The pressure drop is then adjusted using the NEWTON optimization algorithm for the next iteration. The stop criteria is set to a temperature difference of 0.1 K. Compared to the two-port case a fourfold reduction in pumping power is observed due to a drastic reduction in pressure drop (Table 1.9). The flow rate of the four-port design is slightly higher owing to the short channels in the corners with high fluid velocity. Note that the low pumping power solution is the one with the highest area ratio of zones having a temperature close to the maximum junction temperature (compare Figure 1.17 middle with right).

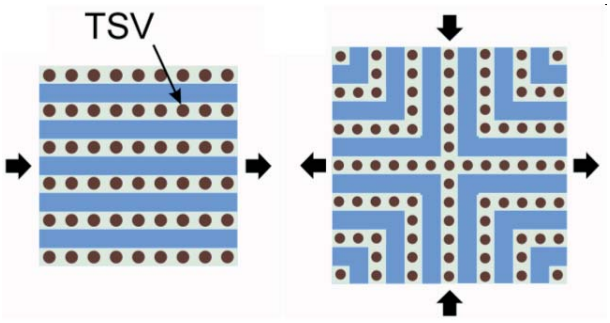


Figure 1.16: Top view of the fluid cavity in two-port (left) and four-port microchannel fluid delivery configuration (right).

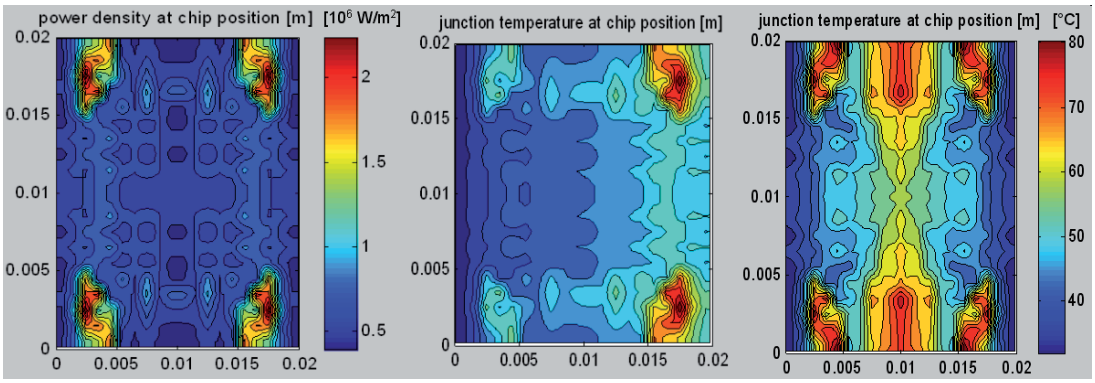


Figure 1.17: Power map of the considered quad-core processor [$\frac{W}{m^2}$] (left). Resulting junction temperature [°C] complying with the thermal budget of 60 K for the two-port (middle) and the four-port fluid delivery architecture (right).

PARAMETERS	TWO-PORT	FOUR-PORT
Pressure drop [kPa]	313	70
Volumetric flow rate [$\frac{ml}{min}$]	309	396
Pumping power [mW]	1612	462

Table 1.9: Comparison of the two- and the four-port architecture for the given quad-core power map.

1.3.4 Electro-Thermal Co-design / Optimization Framework

Thermal aware floorplanning considering the characteristics of interlayer cooling should be utilized to improve the system performance and efficiency even further. The acceptable heat flux towards fluid inlets is increasing for instance. Highly active macros should therefore be placed close to these locations, allowing to increase the clock-frequency. This benefit might be more than compensated by the communication penalty associated by the rearrangement of the floorplan. Therefore a electro-thermal co-design is inevitable.

The power map and the interconnect density distribution defined in the previous floorplan iteration is used as the input parameter for the fluid cavity optimization routine. This algorithm takes advantage of the mentioned heat transfer building-blocks and defines the optimal non-uniform heat transfer cavity for the given pumping power. The maximal clock-frequency and system performance can then be defined with respect to the available thermal budget. The resulting temperature field is fed back to guide the rearrangement of the floorplan. Macros in high temperature zones are moved towards low temperature inlet zones and aligned hot-spots on different tiers are offseted for the next iteration (Figure 1.18). In this analysis the computational performance in floating-point operations per second (FLOPS) or the system efficiency defined as the energy dissipated for one floating-point operation (FLOPS/watt) is the optimization parameter.

1.4 Scope and Organization of the Thesis

Interlayer cooling performance seams not to be compatible with high-performance microprocessor heat fluxes providing only low interconnect densities of less than 10 TSV per mm^2 according to current studies. From the IC-design perspective, densities of 100 to 10'000 TSV per mm^2 are required to use the full potential of 3D integration. So far, only basic heat transfer structures such as straight microchannels or periodic pin arrays were assumed in conjunction with uniform power dissipation. This assumption is not adequate for real microprocessors with a high heat flux contrast of up to 5:1 for logic compared to cache macros. The corresponding spatial interconnect density is also non-uniform. The TSV density in high power zones is double compared to low power zones. Therefore, efficient thermal management solutions have to be spatially modulated. The fundamental heat removal building blocks to design hot spot aware fluid cavities were introduced in section 1.3.3. The main challenge besides the reliable integration of interlayer cooling in an IC-package is to provide efficient tools to design multiple non-uniform heat removal cavities in the stack and to predict resulting temperature fields. Up to now, either correlation based compact thermal modeling or detailed numerical conjugate heat and mass transfer approaches were proposed and used. The fluid flow in the cavity is represented by a resistor network in the first method and has to be known initially. Therefore, only heat transfer in non-communicating unit-cells or uniform pin-fin arrays with trivial stream lines can be treated [73]. The second method allows generic designs of non-uniform cavities with communicating unit-cells, but is computationally much more demanding than the compact model. Therefore, only cavities with moderate complexity up to several hundred pins can be incorporated for acceptable response time needed for the electro-thermal co-design. The main focus of this present investigation is the development of numerical design strategies which are able to capture complex, non-uniform fluid cavities with up to 100'000 pins at minimal computational cost. The proposed methodology is validated by extensive experimental heat and mass transfer investigations. The individual test sites are representing subcomponents and the collective of a

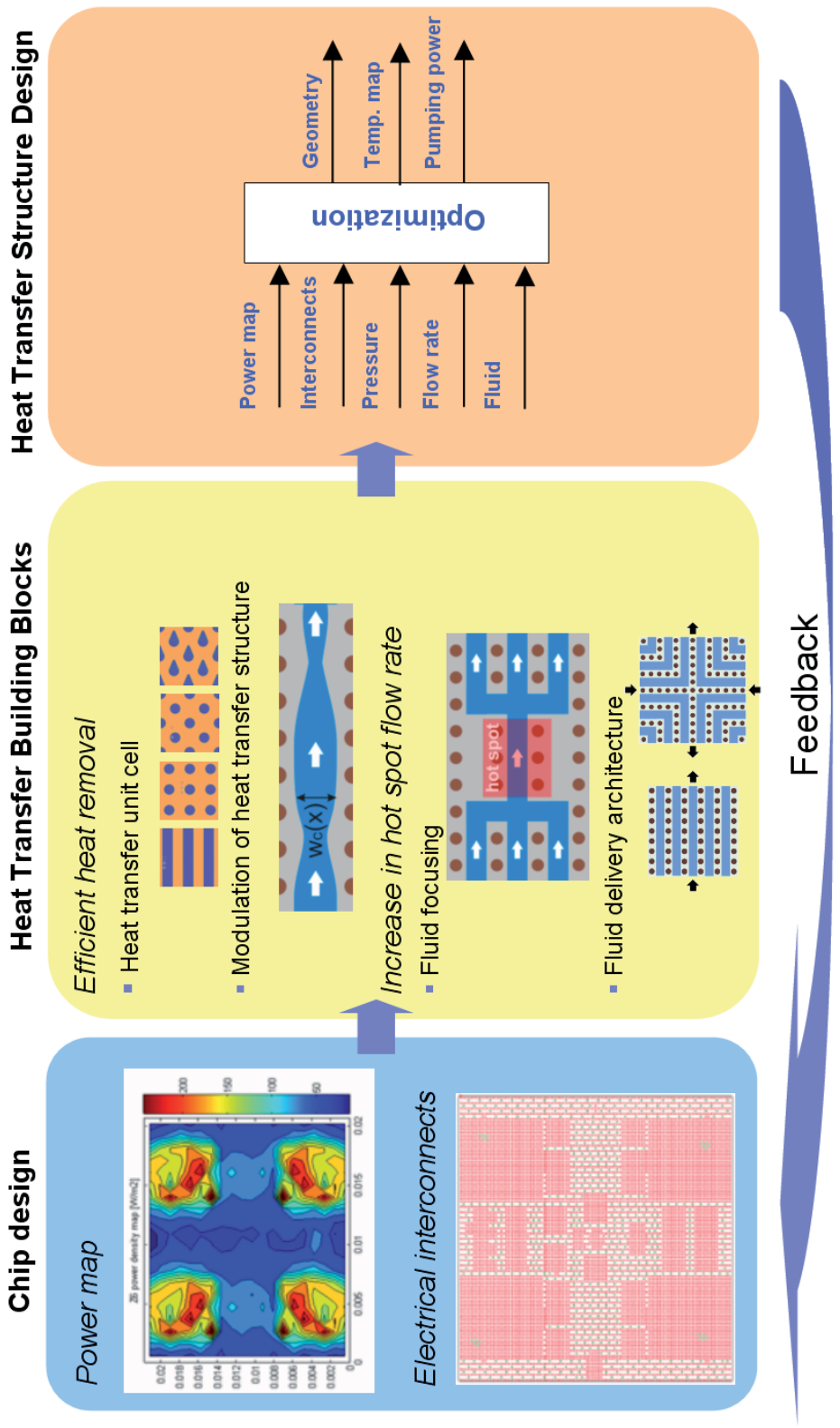


Figure 1.18: Proposed electro-thermal co-design feedback loop with power map and electrical interconnects as input parameter for the fluid cavity optimization algorithm. The resulting non-uniform heat transfer cavity takes advantage of the identified heat transfer building-blocks.

vertical integrated chip stack including interlayer fluid cavities. Interlayer cooling specific bonding and sealing technologies are developed to demonstrate a reliable implementation of the concept. A list of specific design-rules for the safe implementation is presented as well. The window of opportunity for interlayer versus backside cooling is defined finally in a benchmark study case.

The organization of this thesis is as follows

Chapter 2 covers the theoretical foundation of various potential heat and mass transfer modeling approaches to design complex interlayer cooling fluid cavities. The multi-scale approach using detailed sub-domain modeling to derive attributes of individual unit-cells and its use in the high-level porous media model will be the main focus. In this context, the characteristics of periodically arranged pin fins with respect to arbitrary angle-of-attacks is discussed. Hence, periodic hydrodynamic and thermal boundary conditions are proposed for the sub-domain models to derive the permeability and convective parameters efficiently. Field-coupling between solid and porous domains is introduced to complete the modeling protocol, making it viable to design complete chip stacks with multiple cavities and tiers.

Chapter 3 describes the process development of interlayer cooling specific technologies, as well as test vehicle design and fabrication. Low thermal resistance bonding with eutectic gold-tin (AuSn 80/20) thin film solder in reducing atmosphere is discussed in particular. A ring-pad sealing architecture is discussed to prevent electrical shorting between individual bondlines through the fluid. Additionally, a water compatible BEOL layer sequence is proposed. Single cavity test sites are designed for detailed uniform and non-uniform heat removal investigations at two or four-port fluid delivery using spatial thermal imaging by infrared thermography. A flexible electro-fluidic interface was designed and connected to a fluid loop to easily connect to individual test dies. A pyramid chip stack design with three active stratas and four fluid cavities is further described, representing a product style chip stack to complete the validation of the modeling approach. The chapter concludes with an error estimation defining the expected accuracy of the collected data points in the various experiments.

Chapter 4 reports on various experimental and numerical findings. Unit-cell efficiency of different geometries is discussed and compared. For pin-fin arrays a set of NUSSELT number and friction coefficient correlations is derived. Additionally, the effect of vortex shedding was observed. The validity and limits of the multi-scale modeling approach is investigated by detailed numerical modeling, as well as experimental results from non-uniform heat transfer test sites. The most appropriate mathematical description of the periodic porous media is defined for different heat transfer unit-cell types. Most important parameters to incorporate into the model are discussed, to minimize the modeling time and complexity. Finally, realistic interlayer cooling performance including thermal cross-talk between tiers is demonstrated in the pyramid chip stack.

Chapter 5 benchmarks the interlayer and the backside cooling approach against each other. The characteristics of each method with their window of opportunity is defined. Additional interlayer cooling aspects, such as manifold design, hot-spot mitigation by heat spreading in thinned dies, as well as thermal time constants important for temporal workload allocation are discussed. Design-rules for unit-cell shapes, power map aware heat removal and electro-thermal co-design are given. Finally, an interlayer cooling performance evolution roadmap is sketched out. The chapter concludes with a recommendation for future research topics related to the reliable implementation, advanced heat transfer and the extension of the modeling framework.

2 Multi-Scale Modeling with Porous-Media Approach for Heat and Mass Transfer Design

To address individual hot-spots on a single microprocessor die by backside cold plates is not very efficient as long as thermal interfaces are implemented (see subsection 1.3.3). Therefore, uniform heat removal is targeted with reduced design complexity. Branched fluid networks feeding individual normal flow heat removal unit-cells are proposed [74, 75]. To derive the cold plate thermal resistance a detailed conjugate heat and mass transfer model needs to be performed for one of the unit-cells only thanks to their periodic arrangement. A heat conduction model is finally utilized to define the junction temperature with the power map and the uniform cold plate thermal resistance as boundary conditions and the fluid inlet temperature as reference. This sequential design methodology with a detailed heat and mass transfer followed by the global heat transfer model is valid as long as the coolant temperature entering any unit-cell is close to the reference temperature. This is true for normal flow cold plates, but only true for cross-flow cold plates at high flow rate resulting in a minimal fluid inlet to outlet temperature difference of 2 to 4 K. This does not apply for interlayer cooling with low volumetric flow rate and the associated high fluid temperature increase (see subsection 1.3.2). Hence, heat and mass transfer needs to be solved also in the global model to capture the fluid temperature increase throughout the heat transfer cavity. The complexity is further increased by non-uniform heat removal geometries to mitigate local hot spots. Therefore, novel modeling concepts will be discussed in this chapter pointed towards an efficient interlayer cavity design.

2.1 Detailed Conjugate Heat and Mass Transfer

The most general approach to derive the temperature field in the chip stack with multiple active stratas and heat removal cavities is to solve the conjugate heat and mass transfer problem in three-dimensions. All components are represented in their very detail. This results in the most accurate prediction quality at the expense of a high computational cost.

2.1.1 Fluid Flow Basics of Newtonian Viscous Media

Continuum Mechanics in the Microscale

The study of fluid mechanics in conventional, macroscopic length scale treats the fluid as a continuum. All quantities such as density, velocity, temperature, and pressure are defined everywhere in space and change continuously from point to point. These are not points in the geometrical sense, they are rather small sampling volumes. The continuum approximation is valid as long as all fluid properties are continuous. The number of molecules in the sample volume needs to exceed 10^4 to result in less than 1 % statistical variation for kinematic and thermodynamic properties. Transport properties describe the interaction between fluid molecules and fluid molecules with the flow boundaries. Again, a continuous behavior can be observed as a result from a large quantity of interactions. Therefore, a sampling volume with a cube size 10 times the molecules characteristic interaction length scale should be considered. This conditions are satisfied at dimensions of 1 µm and 10 nm for gasses and fluids at normal conditions, respectively [76]. Therefore the continuum approximation for water in channels with hydraulic diameters larger than 10 µm as considered in this study is valid. This also implies the cogency of similarity theory with the use of dimensionless quantities in fluid flow and heat transfer in such problems.

Mass and Momentum Conservation: Navier-Stokes Equation

The governing equations of fluid flow represent mathematical statements of the conservation laws in physics. Mass conservation applied to a fluid element means: "Rate of increase of mass in fluid element = Net rate of flow of mass into fluid element". The unsteady, three-dimensional mass conservation or continuity equation in a compressible fluid can be written as:

$$\frac{\partial \rho}{\partial t} + div(\rho \vec{v}) = 0, \tag{2.1}$$

with the fluid density ρ, the time derivative $\frac{\partial}{\partial t}$, and the velocity $\vec{v}$. The assumption of incompressibility with constant fluid density is valid for fluids and simplifies Equation 2.15 to

$$div(\vec{v}) = 0. \tag{2.2}$$

The momentum equation is derived from NEWTON's second law and states: "Rate of increase of momentum of fluid particle = Sum of forces on fluid particle". Surface and body forces are differentiated. Surface forces result from pressure gradients and the flow of a viscous media. Body forces $\vec{F}$ such as gravity are included as a source term in the following equation. Considering NEWTONIAN fluids such as water with proportional viscous stress to deformation rate yields the NAVIER-STOKES equation for incompressible fluids with constant dynamic viscosity μ:

$$\rho(\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v}) = -\nabla p + \mu \nabla^2 \vec{v} + \vec{F}. \tag{2.3}$$

The left-hand-side of the equation representing the inertia per volume. In more detail: $\partial \vec{v}/\partial t$ being the unsteady and $(\vec{v} \cdot \nabla)\vec{v}$ the convective acceleration term. The later is time independent and the only non-linear term in the equation. The right-hand-side represents the divergence of stress, with ∇p as pressure gradient and $\mu \nabla^2 \vec{v}$ as viscous term and the additional body force $\vec{F}$. Equation 2.3 can be simplified considering only steady state fluid flow in microchannels with insignificant hydrostatic pressure changes due to the small dimensions and reads

$$(\vec{v} \cdot \nabla)\vec{v} = -\frac{1}{\rho}\nabla p + \nu \nabla^2 \vec{v}, \tag{2.4}$$

with ν being the kinematic viscosity.

The velocity and pressure field of the fluid flow can be solved with the governing Equations 2.2 and 2.19 in conjunction with the appropriate set of boundary conditions. Observation of viscous fluid flow reveal that both normal and tangential components of the fluid velocity at a solid boundary must be equal to the boundary velocity itself. For non-moving boundaries $\vec{v} = 0$. The tangential component of velocity is known as the no-slip boundary condition [77].

2.1.2 Heat Conduction and Energy Conservation

Heat conduction in the solid and fluid phase is governed by Fourier's law

$$\vec{q} = -k \nabla T, \tag{2.5}$$

with the local heat flux $\vec{q}$, the material thermal conductivity k, and the temperature gradient ∇T.

The energy equation is derived from the first law of thermodynamics which states: "Rate of increase of energy of fluid particle = Net rate of heat added to fluid particle + net rate of work done on fluid particle". Radiative effects, viscous dissipation, and the work done by pressure change are not considered in this study. This is justified by the convection dominated heat transfer regime at solid-fluid and solid-gas interfaces resulting from the low temperatures (smaller than 100 °C) expected in the chip stack and the low BRINKMAN number of about $1/1000$ indicating heat conduction dominated temperature gradients in the coolant[1]. The only relevant heat sources in IC-devices are power dissipation in the transistors

[1] The BRINKMAN number is a measure of the importance of the viscous heating relative the conductive heat transfer in the coolant and is defined as $Br = \frac{\mu v^2}{k(T_s - T_f)}$. Br equals $1/6000$ considering 20 °C water at 1 $\frac{\mathrm{m}}{\mathrm{s}}$ velocity and a temperature gradient of 10 K.

and wiring layers. Therefore, source terms in the fluid can be neglected. Averaging over an elemental volume sample in the fluid phase yields

$$\rho c_p \vec{v} \cdot \nabla T_f = \nabla \cdot (k \nabla T)_f \tag{2.6}$$

and in the solid phase reads

$$0 = \nabla \cdot (k \nabla T)_s + \dot{q}'''. \tag{2.7}$$

Here the superscript s and f refer to the solid of fluid phase, respectively and c_p is the specific heat at constant pressure of the fluid. The rate of change of thermal energy per unit volume of fluid due to convection is equal to $\rho c_p \vec{v} \cdot \nabla T_f$ and due to heat conduction equal to $\nabla \cdot (k \nabla T)$. The heat generation per unit volume is $\dot{q}'''$.

The boundary conditions on the interface between solid and fluid phase are: (i) conservation of heat flux

$$k_s \nabla T_s = k_f \nabla T_f \tag{2.8}$$

and (ii) equal temperatures

$$T_s = T_f. \tag{2.9}$$

2.1.3 Limitations of Detailed Chip Stack Modeling

The mass flow rate, the pressure drop, and the temperature field in the chip stack could be numerically derived using e.g. the finite volume method (FVM) considering the governing equations and the listed boundary conditions discussed in the previous subsection. The first step in the FVM is to divide the solid and fluid volume into a finite number of control volumes (nodes) with the variables of interest located at their centroid. This procedure is called meshing. The differential form of the governing equations is apply on each control volume in the subsequent step. The resulting non-linear algebraic system of equations is called the discretized equation and is solved in a direct or iterative manner. The number of equations to be solved equals the number of unknowns. It is the product of the number of control volumes with the unknowns per control volume. For laminar, steady state heat and mass transfer considering constant material properties and incompressible NEWTONIAN fluids the three principle velocities u, v, w, the pressure p, and the temperature T have to be considered.

A mesh with 40'000 control volumes is needed (see subsection 2.3.2) to accurately model the fluid dynamics and thermal behavior of one unit-cell of the chip stack. It represents only a single element of the IC-die and the fluid cavity including its fin geometry with the footprint of the through silicon via (TSV) pitch squared. The unit-cell number equals 160'000 assuming a TSV pitch of 100 µm, a chip stack with four fluid cavities, and an footprint size of 4 cm^2. Hence, the physical domain is represented with 6.4 Billion control volumes. Today, problems with up to 10 Million nodes are solved on single workstations within a reasonable response time. A high-performance-computing infrastructure is necessary for larger models with up to 1 Billion nodes to make a parametric study feasible. Consequently, more efficient modeling approaches are important for an economic study and design of interlayer cooled chip stacks at today's computational performance.

2.2 Multi-Scale Modeling: Heat and Mass Transfer in Porous Media

Multi-scale modeling concepts are widely used in the IC-design phase. It is motivated by the large difference in dimensions, ranging from few nanometers to centimeters for switches and complete packages, respectively, the vast amount of components integrated, as well as the large difference in time scales ranging from femto seconds to seconds of the observed phenomena of interest. Hierarchical concepts are key to partition the problem into tasks with reduced complexity. A detailed investigation of subcomponents (sub-domain modeling) with the extraction of relevant parameters (effective properties)

describing the macroscopic behavior of the ensemble of these elements is performed initially. These results are used as input parameters in the next higher hierarchy level. The approach is also known as the effective medium or volume averaging theory and is successfully used to efficiently analyze thermal, electrical, and structural problems in IC components and packages.

2.2.1 Volume Averaging of Porous Media

Fluid cavities populated with microchannel or pin-fin arrays can be interpreted as a porous media. The concept of volume averaging is valid for problems at length scales L much larger than the characteristic dimension l of the representative elementary volume, also known as REV (Figure 2.1):

$$L \gg l. \tag{2.10}$$

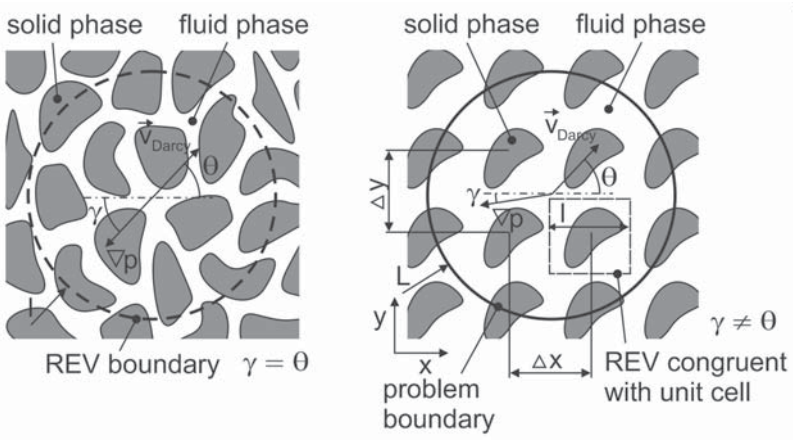


Figure 2.1: Visualization of the representative elementary volume (REV) in an isotropic, but non-deterministic (left) and an engineered periodic porous media. The minimal size of a modeled flow domain L necessary if effective media theory is used is indicated for the periodic porous media represented by the pin-fin array (right).

The REV includes an ensemble of pores to average statistical variations between individual elements in case of a non-deterministic porous medias found in geology or such as metallic foams (Figure 2.1 left). The expansion of the REV above a critical value results in a minimal change of the derived effective parameters. The critical REV in case of engineered porous media with periodically arranged pins at invariant shapes is congruent with one unit-cell of this heat transfer array (Figure 2.1 right).

The mass transfer in porous media is governed by the characteristic permeability and porosity. The porosity φ is defined as the ratio between the fluid volume V_f and the total volume V_{tot} representing the REV:

$$\varphi = \frac{V_f}{V_{tot}}. \tag{2.11}$$

The local volume averaged velocity of the fluid transported in the porous media is denoted as DARCY velocity $\vec{v}_{Darcy}$. It is also called the superficial velocity and is defined as:

$$\vec{v}_{Darcy} = \frac{1}{V_{tot}} \int_{V_f} \vec{v}\, dV = \varphi \frac{1}{V_f} \int_{V_f} \vec{v}\, dV, \tag{2.12}$$

compared to the true velocity which is defined as:

$$\vec{v}_{true} = \frac{1}{V_f} \int_{V_f} \vec{v}\, dV = \frac{\vec{v}_{Darcy}}{\varphi}. \tag{2.13}$$

The volumetric flow rate $\dot{V}$ in integral form through a plane A_p in the porous media reads:

$$\dot{V} = \int_{A_p} \vec{v}_{Darcy} \vec{n} \, dA, \tag{2.14}$$

with the normal vector $\vec{n}$ of the plane.
The mass conservation equation is altered as follows:

$$\frac{\partial}{\partial t} \varphi \rho + div(\rho \vec{v}_{Darcy}) = 0. \tag{2.15}$$

2.2.2 From Darcy Flow to the Extended Navier-Stokes Equation

The flow of NEWTONIAN fluids in a porous media in the creeping-flow regime at Re_K (REYNOLDS number based on the pore diameter) below unity was extensively applied in seepage flows in geology for non-deterministic agglomeration of particles. It is described by DARCY's law, a reduced form of the general momentum equation. Inertia forces are neglected since viscous forces are dominating the momentum transport at such low fluid velocities. The superficial fluid velocity $\vec{v}_{Darcy}$ (also called the seepage velocity) depends linear on the pressure gradient ∇p and is anti-parallel to it (Figure 2.1 left):

$$\nabla p = -\frac{\mu}{K} \vec{v}_{Darcy}. \tag{2.16}$$

The proportionality constant is the dynamic viscosity μ to the permeability K ratio. The permeability has the dimension of $(length)^2$ and is defined by the geometry of the porous media and is independent of the considered fluid. Typical porosity and permeabilities are listed in Table 2.1. Microchannels considered in this study exhibit similar porous properties as sand. Orthotropic permeability can be described with a second-order tensor $\mathcal{K}$. The more general DARCY equation is written as:

$$\vec{v}_{Darcy} = -\mu^{-1} \mathcal{K} \nabla p. \tag{2.17}$$

The resulting velocity vector is in general not anti-parallel to the pressure gradient (Figure 2.1 right). It is preferentially biased towards the axis of highest permeability, with minimal resistance towards fluid flow. The use of the DARCY's equation with the missing viscous and inertia term will result in physically wrong results for some cases, especially at increasing REYNOLDS numbers. Free-slip conditions at porous-solid boundaries, missing viscous drag exchange at porous-fluid interfaces, as well as missing recirculations occurring at flow separation zones are such examples.

A quadratic extension of DARCY's equation was proposed to account for non-linear drag at higher fluid velocities and is called the FORCHHEIMER equation:

$$\nabla p = -\frac{\mu}{K} \vec{v}_{Darcy} - c_F K^{-1/2} \rho_f \left| \vec{v}_{Darcy} \right| \vec{v}_{Darcy}, \tag{2.18}$$

where c_F is a dimensionless form-drag constant and ρ_f the density of the fluid. Form drag compared to surface drag becomes significant at Re_K larger than unity. The transition from the DARCY regime to the FORCHHEIMER regime for general porous media should be considered at Re_K values from 1 to 10 [78]. Laminar flow can be expected up to Re_K of 150 at least. The non-linear drag behavior is more pronounced for pore networks which cause significant flow direction changes and is called tortuosity. Pin-fins in staggered arrangements with resulting small boundary layers are such an example. The steady laminar flow is prone to periodic oscillations (vortex shedding), resulting in an unsteady, laminar flow at increasing flow velocities. The transition to the chaotic turbulent flow regime occurs at Re_K of about 300 [79].

The momentum loss in the porous media as described in Equation 2.16 or 2.18 can be included as a source term $\vec{F}$ in the enhanced NAVIER-STOKES equation

$$(\vec{v}_{Darcy} \cdot \nabla) \vec{v}_{Darcy} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \vec{v}_{Darcy} + \vec{F}, \tag{2.19}$$

for steady-state flows.

MATERIAL	POROSITY φ	PERMEABILITY K [m^2]
Brick	0.12–0.34	$4.8 \times 10^{-15} - 2.2 \times 10^{-13}$
Sand	0.37–0.50	$2 \times 10^{-11} - 1.8 \times 10^{-10}$
Cigarette		1.1×10^{-9}
Microchannel 100 µm *pitch*, $pitch/w_c = 2$, $pitch/h_c = 1$	0.50	7.17×10^{-11}

Table 2.1: Properties of common porous materials [80] compared with a benchmark microchannel case.

2.2.3 Heat Transfer through a Porous Medium

The local-thermal-equilibrium model is one basic concept to describe heat transfer in a porous media. It assumes local thermal equilibrium so that $T_{pf} = T_{ps} = T$, with T_{pf} and T_{ps} as the temperatures of the fluid and solid phase, respectively. The missing temperature gradient results in adiabatic conditions between the two phases. Heat is conducted in parallel in the fluid and solid phase. Hence the classic heat transfer Equation for the fluid phase 2.6 is adjusted to take into account the solid phase in the porous media and is written as [80]:

$$\rho c_p \vec{v}_{Darcy} \cdot \nabla T_{pf} = \nabla \cdot (k_m \nabla T), \tag{2.20}$$

with the effective thermal conductivity k_m described as the arithmetic mean of the solid and fluid thermal conductivity in case of parallel heat conduction in the two phases such as:

$$k_{m1} = (1 - \varphi) k_{ps} + \varphi k_{pf}, \tag{2.21}$$

or the weighted harmonic mean in case of serial heat conduction in the porous media:

$$\frac{1}{k_{m2}} = \frac{(1 - \varphi)}{k_{ps}} + \frac{\varphi}{k_{pf}}. \tag{2.22}$$

The described simplification is only valid for small temperature differences between the phases compared to spatial temperature changes. This is not valid for fluid cavities considered in this study, were the temperature difference between the solid and fluid phase is large compared to the thermal non-uniformity of the pins normal to the flow direction (see subsection 2.3.2). Hence, the two-temperature model [80] accounting for heat transfer between the two phases has to be considered. Equation 2.20 is replaced by a pair of coupled governing equations for steady state heat transport in the two phases of the porous media. The heat transport in the solid phase reads

$$0 = (1 - \varphi) \nabla \cdot (k_{ps} \nabla T)_{ps} + h(T_{pf} - T_{ps}) A_0, \tag{2.23}$$

compared to the heat transport in the fluid phase which is

$$(\rho c_p) \vec{v}_{Darcy} \cdot \nabla T_{pf} = \varphi \nabla \cdot (k_{pf} \nabla T_{pf}) + h(T_{ps} - T_{pf}) A_0. \tag{2.24}$$

Heat transfer between the phases depends on the fluid to solid temperature difference, the proportionality constant h representing the convective heat transfer coefficient, as well as the contact area density $A_0 = A_{fs}/V_{fs}$ with A_{fs}as wetted area and V_{fs} as the volume of the porous domain.

2.2.4 Porous - Solid Domain Interaction

The thermal boundary conditions at the porous-solid interface in case of the two-temperature model are:

$$T_s = T_{ps} = T_{pf}, \tag{2.25}$$

$$k_{pf} \nabla T_{pf} = \varphi k_s \nabla T_s, \tag{2.26}$$

$$k_{ps}\nabla T_{ps} = (1-\varphi)k_s\nabla T_s. \tag{2.27}$$

They indicate equal temperatures at and continuity of heat flux across the interface. Therefore, only no-slip hydrodynamic boundary conditions are meaningful at these interface boundary conditions. The no-slip condition at the porous-solid interface results in additional fluid friction superimposed to the momentum losses caused by the porous media itself. Accordingly, the permeability of microchannels and pin-fin arrays has to be computed at infinite height, to account for the endwall-effect only once. Hence, the fluid interaction at the endwalls would not be captured in all its details. To use the permeability computed from the cavity unit-cell including the endwalls would result in a pessimistic predict on of the mass transfer by considering endwall damping twice. The same accounts for the heat transfer, were h is representing the averaged convective heat transfer from the solid to the fluid phase in the porous media. The heat transfer across the solid-porous interface is also calculated without considering detailed flows close to the unit-cell endwall [81]. These errors are non-relevant in case of large cavity height to pin array pitch ratio and for porous media with multiple pores in direction normal to the endwalls. They can be large in case of interlayer cooling cavities at low-aspect-ratio structures with one single pore in direction normal to the endwalls.

Hence, the porous media is modeled as a quasi two-dimensional domain which is linked to the three-dimensional solid domain using field-coupling to account for the energy transfer (Figure 2.2). The quasi two-dimensional domain results in free-slip boundary conditions at the solid-fluid interface and an invariant fluid velocity normal to the cavity endwall. Hence, the total momentum loss is described with the permeability only including both, friction from endwalls as well as from the pin or channel geometry. The heat transfer $\dot{q}_{ss}$ between the upper and lower endwall at a temperature difference of ΔT_{ss} is defined as:

$$\dot{q}_{ss} = \frac{\Delta T_{ss}}{R_{cond}}, \tag{2.28}$$

with the effective conductive thermal resistance between the endwalls of

$$R_{cond} = \frac{t}{k_{pin}(1-\varphi)}, \tag{2.29}$$

at a spacing of t and the porosity φ.

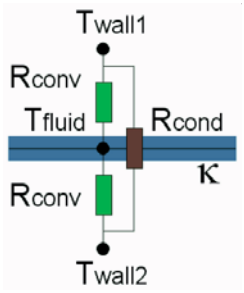


Figure 2.2: Resistor network representing thermal field-coupling between the quasi two-dimensional porous domain and the adjacent three-dimensional solid domain indicated by the endwall nodes.

The effective thermal resistance approximation through the solid phase of the porous media is valid for low thermal conductive fluids with respect to the thermal conductivity of the pin material. This is true in case of the more than two order of magnitude higher thermal conductivity of silicon and copper compared to water. Therefore, only the effective thermal conductivity of the pin k_{pin} is considered . Heat conduction in streamwise direction in case of microchannels walls is neglected this approach, results in a conservative temperature estimate. For pin-fins the assumption is precise. The heat transfer from solid to the fluid phase in the porous media is defined as

$$\dot{q}_{sf} = \frac{T_w - \overline{T}_{pf}}{R_{conv}}, \tag{2.30}$$

with the local solid wall T_w to average local fluid $\bar{T}_{pf}$ temperature and the convective thermal resistance R_{conv} defined from sub-domain modeling (Equation 2.32). The quasi two-dimensional character of the fluid domain demands for orthotropic thermal conductivity of water with infinite heat conduction normal to the cavity surface and can be implemented with two finite volumes only normal to the cavity endwall.

2.2.5 3D Solid to 2D Porous Media Field-Coupling

The character and the boundary condition dependency of the individual coupling elements used in the postulated field-coupling concept have to be defined and validated.

Convective Resistance at Double-Side Heat Flux

A parallel plate configuration at constant but disparate heat fluxes $\dot{q}_1$ and $\dot{q}_2$ imposed at opposing endwalls is the most simplistic cross-flow heat transfer case (Figure 2.3 left). The conductive thermal resistance in the field-coupling model becomes infinite due to the absence of solid structures such as pins or channel walls coupling between the bottom and top wall. Hence, the equivalent resistor network can be simplified by eliminating R_{cond} (Figure 2.3 left).

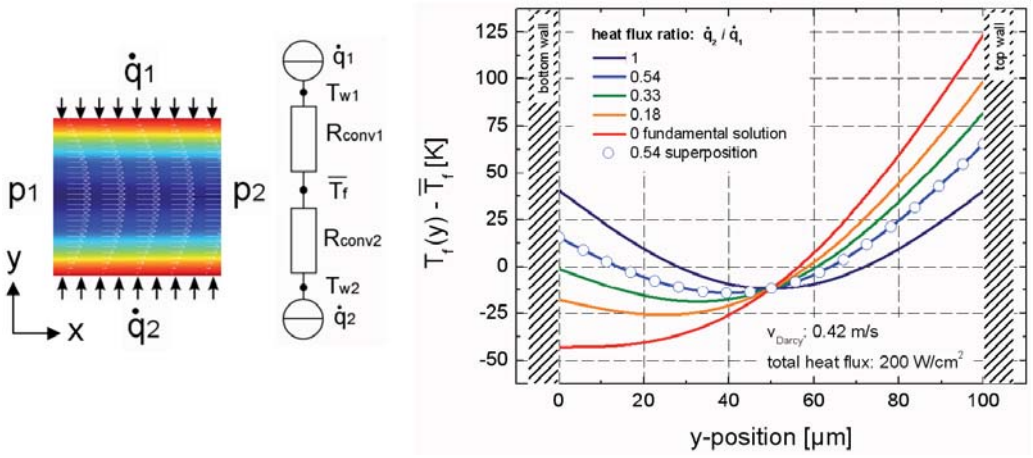


Figure 2.3: Cross-section through a parallel plate cavity with constant but disparate heat flux boundary conditions. The vectors represent the fluid velocity and the color indicate the fluid temperature in case of fully-developed hydrodynamic and thermal boundary layers. The heat transfer can be described with the simplified equivalent resistor network (left). Fluid temperatures $T_f(y)$ with respect to the average fluid temperature $\overline{T}_f$ are plotted for different heat flux ratios $\frac{\dot{q}_2}{\dot{q}_1}$. The CFD temperature profile at a heat flux ratio of 0.54 is compared with the linear combination (Equation 2.31) of the fundamental solution (right).

The fluid temperature $T_f(y)$ normal to the wall surface with respect to the average fluid temperature $\overline{T}_f$ across the cavity is derived from a two dimensional computational fluid dynamic (CFD) model. Fully-developed fluid dynamic and thermal boundary layers at different heat flux ratios ($\frac{\dot{q}_2}{\dot{q}_1}$) are considered and presented in Figure 2.3 (right) for a gap of 100 µm, $v_{Darcy} = 0.42\,\frac{\mathrm{m}}{\mathrm{s}}$, and a constant total heat flux $\dot{q}_1 + \dot{q}_2$ of 200 $\frac{\mathrm{W}}{\mathrm{cm}^2}$. The fluid profile is symmetric and the wall temperatures are higher than the average fluid temperature $\overline{T}_f$ at a heat flux ratio of one. The top wall temperature increases due to the increased power dissipation at that location with reducing heat flux ratio. The inverse happens at the bottom wall. The bottom wall temperature becomes smaller than the average fluid temperature at heat flux ratios smaller than 0.33. All fluid temperature profiles can be derived through a superposition of the fundamental solution at a heat flux ratio of 0 ($\dot{q}_1 = 200\,\frac{\mathrm{W}}{\mathrm{cm}^2}$.and $\dot{q}_2 = 0\,\frac{\mathrm{W}}{\mathrm{cm}^2}$), as indicated by [82]

$$T_f(\dot{q}_1, \dot{q}_2, y, y_{max}) = \overline{T}_f + \frac{\dot{q}_1}{\dot{q}_0} T_0(y) + \frac{\dot{q}_2}{\dot{q}_0} T_0(y_{max} - y), \tag{2.31}$$

with the mean fluid temperature $\overline{T}_f$, the position y normal to the cavity wall, and the cavity gap y_{max}. The fluid temperature at a heat flux ratio of 0.54 is plotted as a result from the CFD model (blue line) and is composed (blue dots) from the fundamental solution (red line) as an example (Figure 2.3 right).

The convective thermal resistance R_{conv_j} for wall j and the NUSSELT number Nu derived from the detailed conjugate heat and mass transfer model can be defined as

$$R_{conv_j} = \frac{T_{wall} - \overline{T}_f}{\dot{q}_j} = \frac{d_h}{Nu \cdot k_f}, \tag{2.32}$$

with hydraulic diameter d_h and the thermal conductivity of the fluid k_f. The results are presented in Figure 2.4. The NUSSELT number as well as the convective thermal resistance depend strongly on the heat flux ratio. R_{conv_2} becomes even negative and the NUSSELT number function exhibits a pole for $T_{wall} - \overline{T}_f = 0\,\mathrm{K}$. A negative convective thermal resistance results in parallel heat flow with respect to the temperature gradient which violates FOURIE's law, but makes sense if the average fluid temperature and not the minimal fluid temperature across the cavity is used to define R_{conv_2}. Of course, the minimal fluid temperature is always smaller or equal to the wall temperature. Hence, the heat flow is anti-parallel to the temperature gradient. SHAH and LONDON obtained the NUSSELT number for parallel plate cavities at a given heat flux ratio and reads:

$$Nu_1 = \frac{140}{26 - 9(\frac{\dot{q}_2}{\dot{q}_1})}, \; Nu_2 = \frac{140}{26 - 9(\frac{\dot{q}_1}{\dot{q}_2})}. \tag{2.33}$$

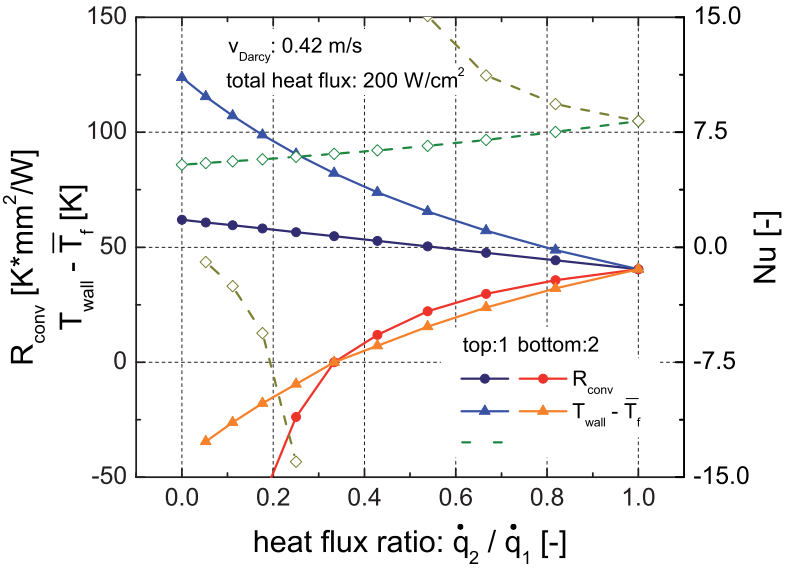


Figure 2.4: The heat flux ratio dependent convective thermal resistance, wall to average fluid temperature difference, and NUSSELT number derived from the detailed CFD model are shown.

To consider source heat flux ratio dependent convective thermal resistances would increase the complexity of the field-coupled modeling concept. The heat transfer present in interlayer cooled microchannel cavities can not be represented by pure heat flux or thermal boundary conditions at the solid-to-fluid interface, compared to the discussed parallel plate case. It could be less sensitive to the imposed source heat flux ratio due to thermal coupling of adjacent dies through the channel walls. Hence, a test case with conductively coupled top and bottom wall and a finite silicon die thickness was performed to answer this question. Microchannels with a typical wall fill factor of 50 % and a two order of magnitude higher thermal conductivity of the walls compared to coolant are considered (Figure 2.5 left). The convective thermal resistances were set as parameters in the equivalent resistor network and were derived from a least square fit based on resulting temperature and heat fluxes calculated from a CFD model. Fully developed hydrodynamic and thermal boundary layers were assumed. The results are plotted in

Figure 2.5 (right). The fitted value for the bottom and top convective thermal resistance are equal and are close to the convective resistance of 17.95 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ reported in literature for constant axial heat flux, but isothermal cross-section at a heat flux ration of one [82]. Any wall temperature in the microchannel even for the most extreme case with the heat flux ratio of zero is larger than the average fluid temperature. Therefore, all thermal resistances stay positive. The coupling between the cavity walls seams to distributes the heat nearly equally. The heat flow from the bottom wall to the fluid is only 9 % reduced compared to the heat flow from the top wall to the fluid at the most extreme case with single side heat flux only. Hence, a variation from 1 to 0.91 of the channel wall to fluid heat flux ratio is observed in microchannels. The convective thermal resistance can therefore be assumed to be invariant to the source heat flux ratio. This is validated with temperatures derived from the equivalent resistor network model at constant NUSSELT number provided by correlation of SHAH and LONDON [82] for microchannels with respect to temperatures derived from CFD modeling. A maximal error of 1.29 % is observed. To conclude: source heat flux ratio dependent thermal resistances have to be considered for thermal decoupled or weakly coupled fluid cavities such as parallel plates. A strong thermal coupling results in close to isothermal condition with a minimal heat flux difference between the bottom and top solid-to-fluid interface and allows the use of source heat flux invariant convective thermal resistances.

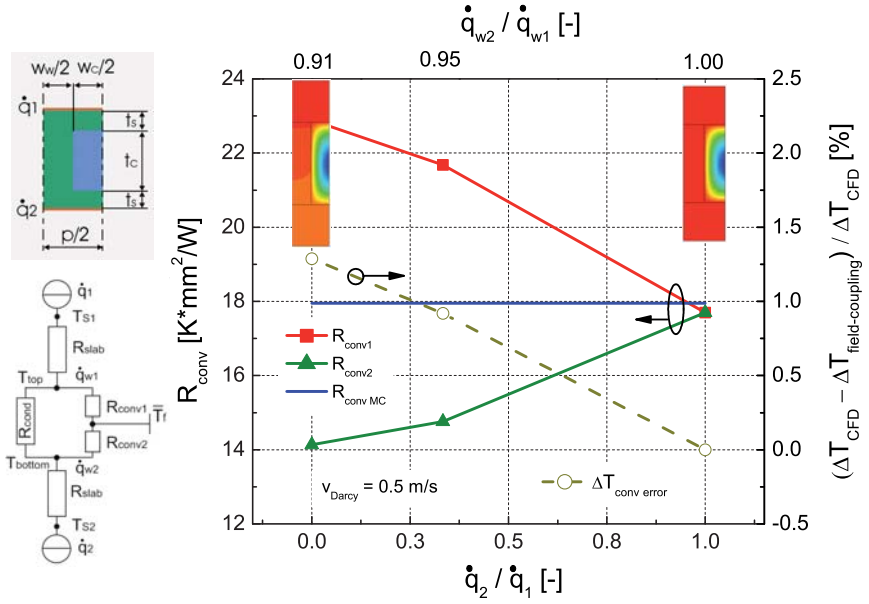


Figure 2.5: The equivalent resistor network representing a single microchannel with double side heat sources is shown (left). The fitted convective thermal resistances representing the heat transfer in a microchannel cavity with additional silicon slabs are presented as a function of source and wall heat flux ratio. The temperature field for the source heat flux ratio equal to one and zero are depicted as a contour plot. R_{convMC} (blue line) derived from correlations [82] is shown and used as the heat flux independent convective thermal resistance in the equivalent resistor network. The resulting junction temperature error with respect to the CFD model is shown as well (right).

Field-Coupling Validation on a Multi-Stack Example

A multiple-cavity CFD model is performed to validate the multi-scale modeling concept with field-coupling at heat flux ratio invariant convective thermal resistances (details of the implementation can be found in the appendix B). Asymmetric solid domains with BEOL layers on one and a Si-slab on the other side of the junction are considered. Four cavities populated with microchannels at a channel width of 50 µm, channel pitch and height of 100 µm, and at a length of 8.4 mm are implemented. The domain width is kept at half a microchannel pitch thanks to symmetry reason as shown in Figure 2.6 (middle). The DARCY velocity is set to 0.25 $\frac{\mathrm{m}}{\mathrm{s}}$ in each cavity and a uniform heat flux of $\dot{q}_n$ of 100 $\frac{\mathrm{W}}{\mathrm{cm^2}}$ is imposed at the junction of all dies. The material properties of the fluid and solid representing water and silicon are set constant at a reference temperature of 25 °C. A schematic of the porous-solid domain model with the two-dimensional cavity and the three-dimensional solid domains, coupled through thermal resistors

are depicted in Figure 2.6 (right). The thermal conductivity of the 12 µm BEOL layer considered was set to 2.25 $\frac{\mathrm{W}}{\mathrm{m\,K}}$. The silicon slab thickness t_s was assumed to be 50 µm, the convective thermal resistance $R_{conv_1} = R_{conv_2} = 17.95\,\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$, and the conductive thermal resistance $R_{cond} = 1.54\,\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ [82].

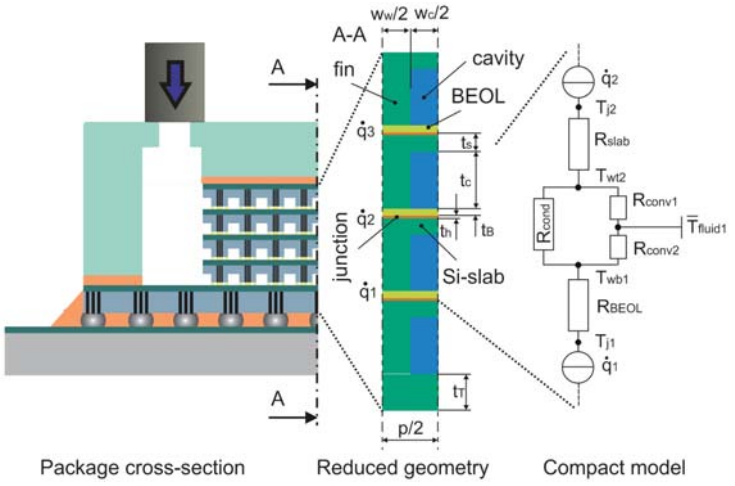


Figure 2.6: Different abstraction levels of a interlayer cooled package starting with a cross-section of the detailed physical implementation, followed by the defeatured geometry used for the detailed CFD model, as well as the schematic of the field-coupled porous-solid representation are presented from left to right.

The resulting temperature field and profile of the detailed CFD and the field-coupled porous-solid domain model is compared in Figure 2.7 (the detailed implementation of the field-coupled porous-solid domain model is explained in appendix B). The temperature normal to the chip stack surface (in direction of the z-axes) in the channel wall and fluid center and along the three junctions in flow direction are depicted in Figure 2.7 (left) and (right), respectively. A maximal junction temperature deviation of 0.5 K is observed between the models. This results in an error of 2 % with respect to the minimal to maximal junction temperature difference of 25 K.

2.3 Parameter Extraction from Efficient Sub-Domain Modeling with Periodic Boundary Conditions

A conjugate heat and mass transfer model of the detailed fluid cavity has to be performed to derive the effective fluid dynamic and thermal parameters describing the porous domain. The permeability and convective thermal resistance at developed hydrodynamic and thermal boundary layers are of interest. ALFIERI et al. [81] studied circular pin fin arrays in in-line arrangement and were using 20 pin-fin unit cells as computational domain. They extracted the effective parameters from cell 16 assuming hydrodynamic and thermal fully developed conditions. Several researchers [83, 84, 85] proposed to consider the periodic nature on pin arrays in cross-flow heat exchange mode to minimize computation cost. Thanks to the periodicity only one unit-cell of the heat transfer geometry has to be depicted as the computational domain. ROYCHOWDHURY et al. [83] implement periodic hydrodynamic and thermal boundary conditions for two-dimensional fluid domains, but did not consider additional solid domains. WÄLCHLI et al. [84] included solid domains and solved the problem in an iterative procedure using a PERL script. The fluid inlet temperature was defined as the fluid outlet temperature of the previous CFD run. This results in reduced memory demand due to a 20x reduction in mesh nodes compared to [81]. But still, 16 completely converged runs are necessary to derive the parameters. Heat fluxes across the solid periodic interface were also neglected. Hence, the implementation of a computational efficient procedure to model unit-cells with solid and fluid domains is proposed in the following section as described in [86]. Effective parameters for heat transfer geometries at varying REYNOLDS number and pressure gradients parallel to the geometry symmetry line are presented. Furthermore, the influence of the angle-of-attack in pin-fin arrays is investigated [87].

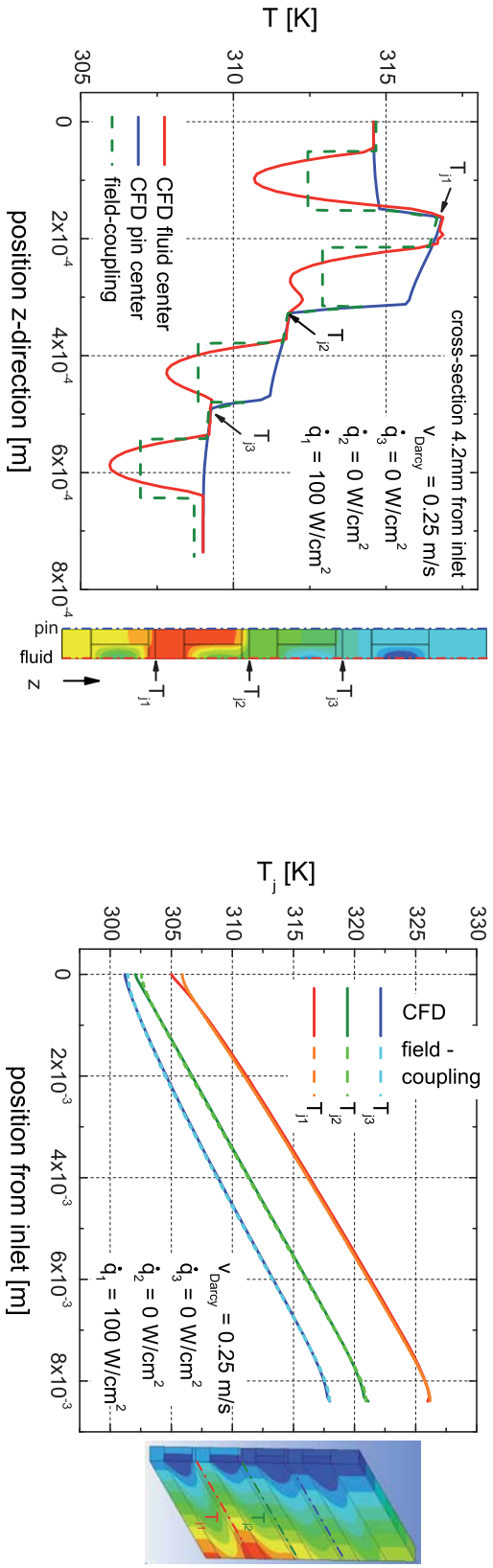


Figure 2.7: Field-coupled porous-solid domain model validation with respect to the detailed CFD model. The resulting temperature field of the detailed CFD model is presented by the contour plots. The comparison of resulting temperatures derived from the two modeling methods are compared in both plots. In z-direction through the stack (left) and at the junction in streamwise direction (right).

2.3.1 Sub-Domain Model using Periodic Boundary Conditions

The permeability and convective thermal resistance at developed hydrodynamic and thermal boundary layers are considered as a first-order approximation in the porous domains or in compact thermal models. The periodic arrangement of individual pin-fins, the uniform heat flux as thermal boundary condition, and the assumption of constant material properties render the problem streamwise-periodic in both the fluid and the solid domain (Figure 2.8). Thus, the flow pattern repeats with the unit cell length L_{cell}. The pressure Δp_{cell} and the fluid temperature change ΔT_{cell} between specific locations on the periodic boundaries become constant:

$$\Delta p_{cell} = p(\vec{r} + \vec{L}_{cell}) - p(\vec{r}) = p(\vec{r} + 2\vec{L}_{cell}) - p(\vec{r} + \vec{L}_{cell}), \tag{2.34}$$

$$\Delta T_{cell} = T(\vec{r} + \vec{L}_{cell}) - T(\vec{r}) = T(\vec{r} + 2\vec{L}_{cell}) - T(\vec{r} + \vec{L}_{cell}), \tag{2.35}$$

with the position vector $\vec{r}$ and the periodic length vector $\vec{L}_{cell}$. The temperature change can be related to the heat flow $\dot{Q}$ dissipated at the heat source in one unit-cell and is

$$\Delta T_{cell} = \frac{\dot{Q}}{\dot{m} c_p} = \frac{2 \dot{q} A_{cell}}{\dot{m} c_p}, \tag{2.36}$$

with the junction heat flux $\dot{q}$, the projected area of the unit-cell A_{cell} (surface S_j in Figure 2.9 left), the coolant mass flow rate $\dot{m}$, and the specific heat capacity c_p. Such problems can be analyzed by restricting the numerical model to a single unit-cell representing a quarter of the total pin-fin cavity cell as shown in Figure 2.9 (left).

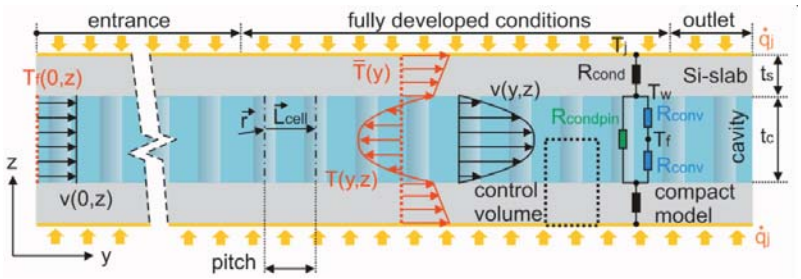


Figure 2.8: Cross section of a fluid cavity with a periodic arrangement of heat-transfer structures and imposed double-side heat flux. The velocity and temperature profile (within the fluid and solid) with respect to the mean temperature in streamwise direction become periodic in the section of fully developed conditions. The control volume used in the sub-domain model to extract the permeability and convective thermal resistance is indicated by the dashed box. The various components used in the compact thermal model to assess the pin shapes are shown as thermal resistance network.

The permeability K is defined according to DARCY's law:

$$K = \mu \frac{v_{Darcy}}{\bar{p}_{in} - \bar{p}_{out}}, \tag{2.37}$$

where μ is the dynamic viscosity of the fluid and $\bar{p}_{in}$ and $\bar{p}_{out}$ denote the area average pressure at the inlet and outlet of the unit cell, respectively. The DARCY velocity in this case is defined as

$$v_{Darcy} = \frac{\int_{S_{f11}} v dA}{w \cdot h}, \tag{2.38}$$

where v is the local inlet velocity across the fluid inlet area in streamwise direction, and w and h are the width and the height of the unit cell cavity, respectively.

The representative convective thermal resistance R_{conv} is calculated according to

$$R_{conv} = \frac{\overline{T}_w - \overline{T}_f}{\dot{q}_j}, \tag{2.39}$$

with $\dot{q}_j$ the source heat flux, $\bar{T}_w$ the area average wall and $\bar{T}_f$ the mass flow average fluid temperature.

Computational fluid dynamic (CFD) solvers are used to numerically compute the conjugate heat and mass transfer problem. Hydrodynamic, but no thermal periodic boundary conditions (BC) are available for periodic problems. Therefore, a one-way coupling modeling sequence was defined (neglecting the temperature dependent viscosity, which would alter the flow field) to also include the energy transport in a single unit-cell with periodic BCs imposed. The implementation in this study was done on ANSYS CFX, a commercially available finite volume method (FVM) solver [88]. It uses a pressure-based formulation proposed by RHIE and CHOW [89]. The resulting linear equation system is computed with an algebraic multi-grid solver [90]. The proposed sequence can be applied for pressure gradients imposed along the symmetry lines of the unit-cell geometry. This results in two opposing faces ($S_{f11} + S_{s11}$, $S_{f12} + S_{s12}$) with mass and heat transfer, and two opposing faces with symmetry boundary conditions ($S_{f21} + S_{s21}$, $S_{f22} + S_{s22}$), as shown in Figure 2.9 (left).

1) Hydrodynamic periodic BC: First, the velocity and pressure field in the fluid domain are computed. The steady-state Navier–Stokes equation in conjunction with mass conservation for incompressible fluids at constant thermal properties is used as governing equation at a given Δp_{cell} at periodic hydrodynamic BC (blue arrows). No-slip and symmetry (symmetry sign) conditions are set to the fluid-solid interface and the cavity central plane, respectively (Figure 2.9 right). The iteration loop is left if the residuum becomes $< 1 \times 10^{-6}$.

2) Thermal periodic BC: a) The energy transport equation is solved for a first iteration with a uniform inlet temperature $T_{in_0} = T_{ref}$ and the velocity field computed in step 1). b) The fluid inlet temperature for the subsequent iteration $T_{in_{n+1}}$ is then defined from the outlet temperature field T_{out_n} corrected by the mass flow averaged fluid outlet temperature $\bar{T}_{out_n}$ minus the reference temperature T_{ref} as

$$T_{in}(x,z)_{n+1} = T_{out}(x,z)_n - \left(\frac{\int_{S_{f12}} v T_{out_n} dA}{\int_{S_{f12}} v dA} - T_{ref} \right). \tag{2.40}$$

c) The energy transport equation is now solved with the temperature field $T_{in_{n+1}}$ as the new thermal inlet boundary condition, using again the velocity field from run 1). This sequence is repeated from step 2b) onward until the energy imbalance and the average residuum become $< 0.1\,\%$ and $< 5 \times 10^{-5}$, respectively.

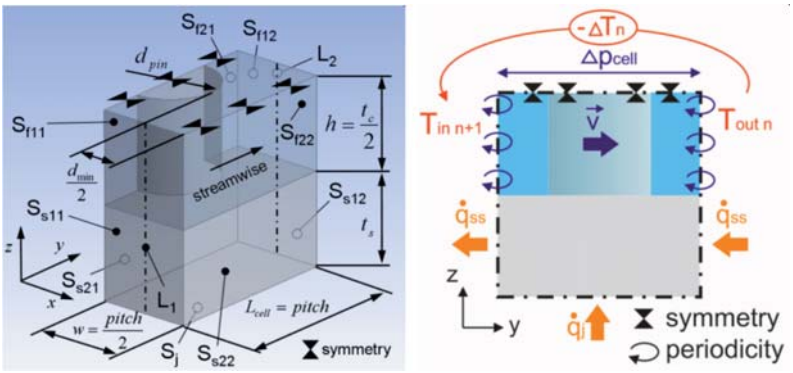


Figure 2.9: Definition of surface, line, and dimension names in the control volume used for sub-domain modeling. It represents a quarter pin-fin cell with silicon slab (left). Hydrodynamic and thermal boundary conditions used in the periodic modeling approach (right).

Routines enabling the mapping of temperature fields from one surface to another have to be developed to implement thermal periodic BCs. ANSYS CFX is a coupled solver for fluid dynamics and heat transfer that uses a node-centered discretization scheme with finite numerical control volumes around the nodes of the mesh. This scheme strongly affects the implementation of the field mapping. In a node-centered scheme, the results (e.g. the temperature values) are calculated and stored for the nodes. This is also true at boundaries. Accordingly, the temperature values at the outlet are available to the nodes.

On the other hand, fluxes are calculated on control volume surfaces. The variable values required for the flux calculation are taken at so-called integration points, which are representative of the control volume surface. The same applies to fluxes on boundaries. At the inlet, an enthalpy flux boundary condition, which is a function of the fluid temperature, has to be defined to solve the energy transport equation. Therefore the inlet boundary temperature has to be available at the integration points. As these points do not coincide with the nodes, the node outlet temperatures have to be interpolated on the inlet integration points.

The implementation is performed with three main USER FORTRAN modules (Figure 2.10): (i) The first module maps boundary vertices at the fluid outlet (S_{f12} in Figure 2.9 left). It generates a list of boundary vertices and is called after solver startup. (ii) The second module extracts and stores the variable values for the temperature T_{out_n} after every solver iteration. The vortex list generated by the mapping module is used for this extraction. (iii) The third module interpolates the temperature values on the inlet boundary integration points (S_{f11} in Figure 2.9 left). This interpolation allows the use of nonconformal meshing of the inlet face and the corresponding outlet face. Additionally, a subroutine is available in this module that performs the subtraction of the offset temperature ΔT_n.

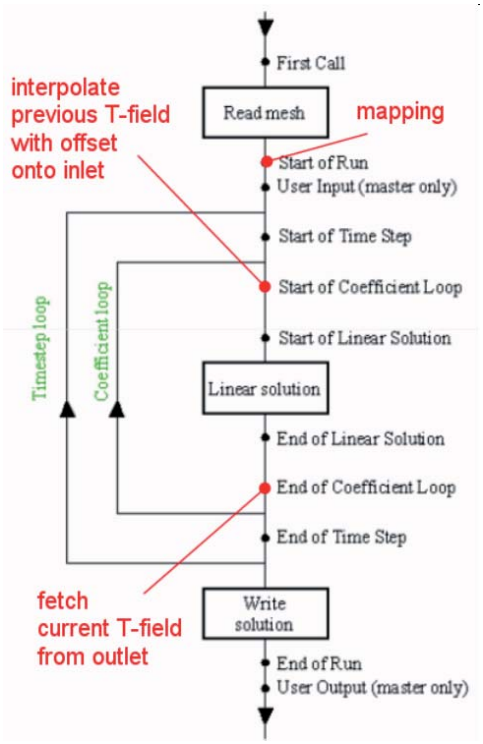


Figure 2.10: Flow diagram of the CFX solving sequence. Also indicated are the calls of additional USER FORTRAN routines (in red) to map the fluid temperature outlet field of run n to the fluid inlet temperature field of run $n+1$.

Note, that the sequence described here only takes into account the advection term, but not the heat conduction from one unit-cell to the next in streamwise direction. This is a good approximation for fluids with low thermal conductivity ($k_{water} = 0.6 \frac{\mathrm{W}}{\mathrm{mK}}$). For highly conductive solids, such as silicon ($k_{si} = 130 \frac{\mathrm{W}}{\mathrm{mK}}$), the adiabatic BC on surfaces S_{s11} and S_{s12} results in almost isothermal conditions, as can be seen in the left-hand-side (gray) panel of Figure 2.11 (left). The graph shows the temperature difference between the central line at the inlet (L_1) and at the outlet (L_2). The constant ΔT_{cell} can be observed in the fluid, but not in the solid.

A streamwise heat flux $\dot{q}_{ss}$ has to be imposed on the solid-to-solid faces (S_{s11}, S_{s12}) to fulfill the thermal periodic condition with a constant ΔT_{cell} also in the solid. A 40-unit-cell model (such as shown in Figure 2.8), representing a fluid cavity with 40 subsequent pin-fin unit-cells, was used to validate the periodic BC concept and to study the streamwise heat flux $\dot{q}_{ss}(x,z)$ uniformity in the solid.

Uniform velocity and thermal BCs for the fluid inlet were considered. The pressure drop and convective thermal resistance (defined in Equation 2.32) of each unit-cell were computed. The cells with

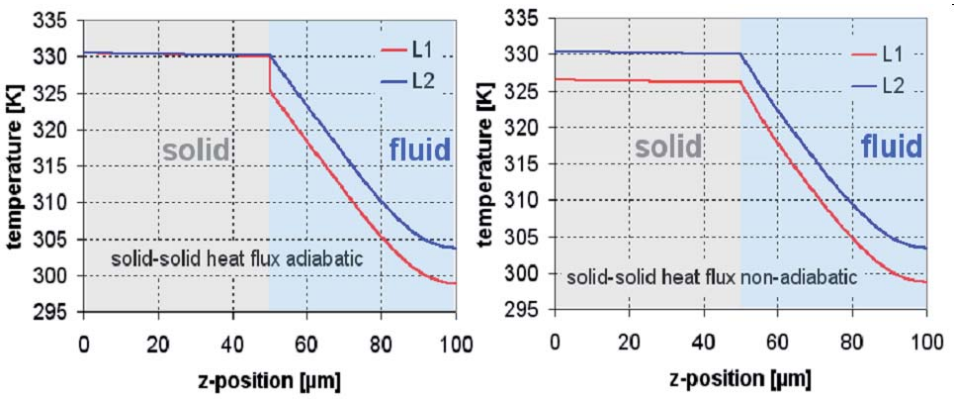


Figure 2.11: Central inlet and outlet temperature along lines L_1 and L_2 (Figure2.9 left) of the converged temperature field with adiabatic (left) and imposed uniform heat flux $\dot{q}_{ss}$ (Figure 2.9 right) at the solid-to-solid interface S_{s11}, S_{s12} (Figure 2.9 left).

asymptotic values were identified as being representative for the comparison with the periodic model. The heat-flux non-uniformity was less than 4.6 % in the worst case at the minimum velocities ($v_{Darcy} = 0.1\,\frac{\mathrm{m}}{\mathrm{s}}$) of interest for the given material set and geometries. This allows the approximation of an iso-heat flux anti-parallel to the streamwise direction at the solid inlet face (S_{s11}) and the solid outlet face (S_{s12}) (Figure 2.9 right). A proportional controller (Figure 2.12) was used to set the unknown heat flux $\dot{q}_{ss}$ correctly. It adjusts the heat flux after each solver iteration (n) according to

$$\dot{q}_{ss_{n+1}} = K_p(\Delta T_{f_n} - \Delta T_{s_n}) + \dot{q}_{ss_n}, \tag{2.41}$$

with the mass flow averaged ΔT_f and the area averaged ΔT_s temperature difference from the inlet to the outlet faces of the fluid (S_{f11} to S_{f12}) and solid (S_{s11} to S_{s12}), respectively. The proportionality constant K_p was derived with the 40-unit-cell model according to the ZIEGLER–NICHOLS method and is set to be $0.5 \times 10^6\,\frac{\mathrm{W}}{\mathrm{m^2\,K}}$ [91, 92]. This value was used in all models and proved to be stable without overshoot, with minimum error, and optimum sensitivity.

The central temperature at the cell inlet and outlet including the adjusted solid-to-solid heat flux is presented in Figure 2.11 (right). Finally, a constant temperature difference ΔT_{cell} can be observed in the solid and fluid domain.

The quality of the periodic BC modeling approach discussed here is validated with respect to fully developed thermal conditions in the 40-unit-cell model as described before (Table 2.2). The permeability and convective thermal resistance values deviate less than 1.1 % from the reference. The computational time (CPU time) could be reduced by a factor of 26 and justifies the use of the proposed periodic BC approach to efficiently derive the permeability and convective thermal resistance values needed in the system-level porous-media and compact thermal models.

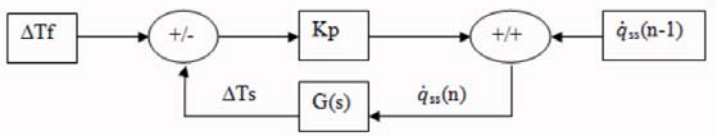


Figure 2.12: Proportional controller to adjust the streamwise heat flux $\dot{q}_{ss}$ (Figure 2.9 right) at the solid-to-solid interface.

PARAMETER	40-UNIT-CELL ASYMPTOTIC	PERIODIC BC ADVECTION ONLY	PERIODIC BC: ADVECTION AND S-S HEAT FLUX
In-out temp. diff., ΔT_f, ΔT_s [K]	5.079 / 5.082	5.169 (1.77) / 0 (inf)	5.202 (2.42) / 5.201 (2.34)
Solid-solid heat flux, $\dot{q}_{ss}$ $[\frac{\mathrm{W}}{\mathrm{cm}^2}]$	8.202×10^6	0 (inf)	8.090×10^6(-1.36)
DARCY velocity, $\vec{v}_{Darcy}$ $[\frac{\mathrm{m}}{\mathrm{s}}]$	0.118	0.119 (1.32)	0.119 (1.32)
Permeability, K [m^2]	1.191×10^{-10}	1.197×10^{-10}(0.5)	1.197×10^{-10}(0.5)
Conv. therm. resistance, $R_{conv}[\frac{\mathrm{K\,mm}^2}{\mathrm{W}}]$	2.205	2.209 (0.181)	2.203 (-0.091)
CPU time [min]	202	7	7

Table 2.2: Periodic BC modeling validation in absolute and in percental (in parentheses) deviation with respect to developed conditions (asymptotic) derived from the 40-unit-cell model. Pin-fins with circular geometry and an imposed pressure gradient of 1×10^6 $\frac{\mathrm{Pa}}{\mathrm{m}}$ is considered. (s-s is short form for solid-solid)

2.3.2 Parameter Set of Individual Heat Transfer Geometries in Symmetry Direction

The shape and arrangement of the pin-fins is an important parameter defining the efficiency of the heat-transfer system by minimizing the pumping power needed for a given thermal budget. Unfortunately, nearly all studies for low-aspect-ratio structures investigate only single-side heat-flux boundary conditions [70, 72, 93, 94]. For interlayer cooling, double-side heat-flux is the correct boundary condition to consider. Hence, the effective parameters for basic and advanced heat transfer structures and its characteristics is defined in the following section using conjugate heat and mass transfer modeling with periodic boundary conditions, as discussed before.

Basic Heat Transfer Structure Characteristics

In this section correlations describing the permeability κ and convective thermal resistance R_{conv} for parallel plate (PP), microchannel (MC), as well as for pin fin (PF) structures (geometry definition Figure 1.11) with cavity height and heat transfer structure pitch of 100 µm at microchannel wall width and pin diameter of 50 µm will be presented. These parameters are relevant to derive the mass transport using the extended NAVIER-STOKES equation with the source term $\vec{F}$ as discussed in Equation 2.19 and for the energy transport using field-coupling introduced in section 2.2.4. The corresponding porosity, conductive thermal resistance, and the ratio of wetted to projected area of the individual structures are listed in Table 2.3.

PIN SHAPE		POROSITY [-]	CONDUCTIVE THERMAL RESISTANCE $[\frac{\mathrm{K\,mm}^2}{\mathrm{W}}]$	RATIO OF WETTED TO PROJECTED AREA $[\frac{\mathrm{m}^2}{\mathrm{m}^2}]$
Parallel plate		1.000	inf	2.00
Microchannel	fp	0.500	1.54	3.00
	hp	0.750	3.08	2.50
Pin fin	fp	0.804	3.92	3.18
	hp	0.902	7.84	2.59

Table 2.3: Comparison of porosity, conductive thermal resistance, and ratio of wetted to projected area of the heat transfer cavities discussed.

The REYNOLDS number range of interest is defined by feasible boundary conditions, as well as the dimensions of the fluid cavity. A pressure gradient of 1×10^7 $\frac{\mathrm{Pa}}{\mathrm{m}}$ is considered as a benchmark, which is a result from a pressure difference of 1×10^5 Pa and a chip length of 10 mm. This results in REYNOLDS numbers between 50 and 500 depending on the cavity geometry considered. The upper limit of the pressure gradient to be considered is 1×10^8 $\frac{\mathrm{Pa}}{\mathrm{m}}$ and results from a fluid path length of 1 mm which exists in the corner of a four-port configuration (Figure 1.16). Whereas, the lower limit of the pressure gradient to be considered is 1×10^5 $\frac{\mathrm{Pa}}{\mathrm{m}}$ and results from a pressure drop as low as 1×10^4 Pa across the

cavity (Figure 2.15). The solid domain considered in the numerical models consists of a 50 µm thick silicon slab and the pin-fin itself.

The discretization of the computational domains was done with unstructured hexaedral meshes using the meshing software ICEM. O-grids were implemented to improve the mesh angle quality for the squared unit-cell and the circular pin shape. Inflation layers using a growth factor of 1.2 were used to resolve the fluid dynamic, as well as the thermal boundary layers in the fluid domain (Figure 2.13). A mesh refinement study to define the critical node number to result in a mesh independent solution was performed prior to each investigation. The partial differential equations were solved numerically with a finite volume second order high resolution discretization scheme using the solver CFX 12. The coupled algebraic multi-grid solver is used in conjunction with a convergence criterion for the root mean square mass, momentum and energy residuals of 10^{-6}. The DARCY velocity $v_{Darcy}(n)$ for a given node number n relative to $v_{Darcy}(n_{max})$ at the reference node number n_{max} of 94'000 is plotted in Figure 2.13. A minimal node number of 30'000 has to be considered for deviations of less than 1 % from the reference mesh result.

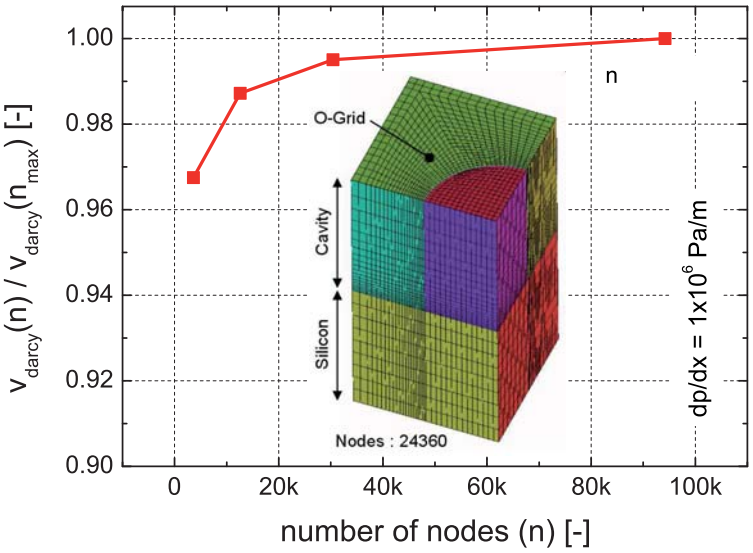


Figure 2.13: Mesh refinement study, comparing the resulting v_{Darcy} for a given node number with v_{Darcy} at 94'000 nodes and a pressure gradient of $1 \times 10^6 \frac{\text{Pa}}{\text{m}}$.

Laminar flow can be expected for all test cases with pressure gradients up to $1 \times 10^7 \frac{\text{Pa}}{\text{m}}$ as stated before. The momentum transport in the creeping flow regime is dominated by viscous damping. The stream lines adapt to the solid surfaces and the pressure gradient varies linear with changing fluid velocity. This can be observed for REYNOLDS numbers below 10 (Figure 2.14 left, 2.15). Above this value, two stationary, symmetric recirculations start to appear down stream of the pins (Figure 2.14 center). They grow in size up to a $Re \sim 150$ where they completely occupy the space between the pins (Figure 2.14 right). The main mass transport occurs now besides the pins and appears to be similar to the microchannel flow. Inertia terms start to dominate the momentum transport at this REYNOLDS numbers with the consequence of a non-linear pressure gradient to velocity relationship for PFI, but not for MC structures. The convergence criterion only converges marginal and starts to oscillate at pressure gradients larger than $1 \times 10^8 \frac{\text{Pa}}{\text{m}}$ with REYNOLDS numbers of more than 900. This behavior indicates the transition from steady state to transient laminar or even turbulent flow and would need to be solved as a transient problem.

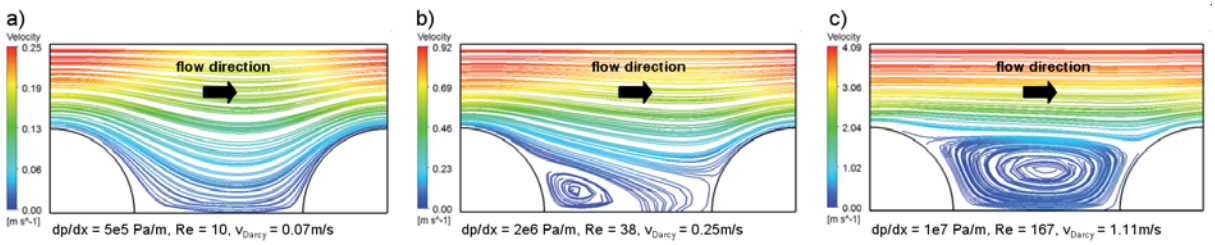


Figure 2.14: Fluid flow indicated by streamlines in the cavity center-plane for pin-fin arrays in in-line (PFI) arrangement with periodic boundary conditions in the range of interlayer cooling relevant pressure gradients.

The permeability in the original DARCY law (Equation 2.16) is modified to be dependent on v_{Darcy} and the fluid temperature T_f to describe the mass transport in the fluid cavity for REYNOLDS numbers > 1, such as

$$\kappa(v_{Darcy}, T_f) = \frac{-\mu(T_f) \cdot v_{Darcy}}{\frac{dp}{ds}(v_{Darcy}, T_f)}. \tag{2.42}$$

More complex correlations or look-up tables compared to the second order polynomial FORCHHEIMER equation (Equation 2.18) with the permeability and form-drag constant as coefficients can reflect the detailed fluid dynamic behavior of the flow and justifies the definition of the modified permeability. It allows to account for generic types of pressure gradient to velocity responses improving modeling accuracy.

Resulting modified permeabilities of pin-fin arrays from numerical modeling at fluid temperatures of 25 °C according to the definition of Equation 2.42 are compared to parallel plate and microchannels derived from correlations (Figure 2.15).

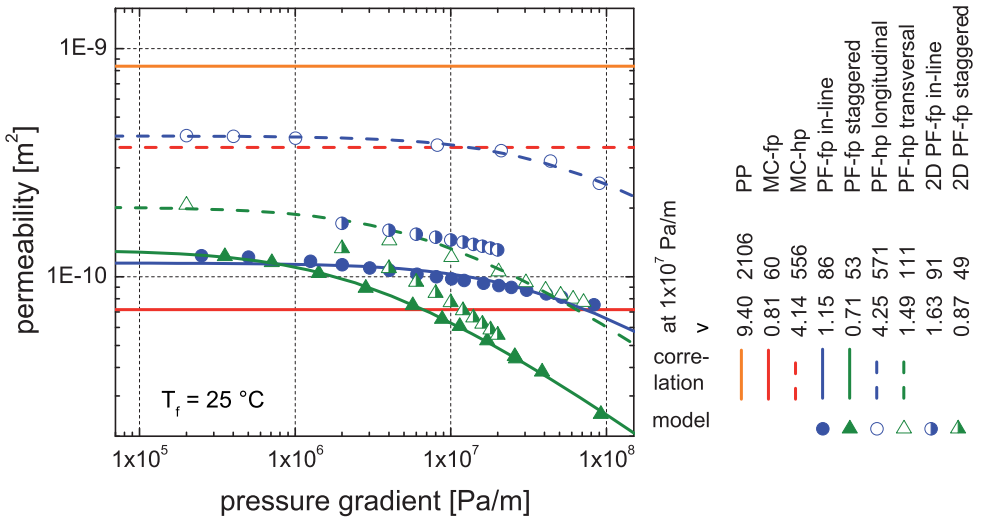


Figure 2.15: Modified permeability as defined in Equation 2.42 of parallel plate (PP), microchannel (MC), and pin-fin (PF) cavities are presented from numerical modeling (dots). All values are computed at a water temperature of 25 °C. The orientation sensitivity of PF is demonstrated by the in-line and staggered case at full population, as well as for transversal and longitudinal flows in case of half populated PF. The permeability of infinitely long PF is plotted (2D case) as a comparison. The table also lists the DARCY velocity dependent REYNOLDS number for a pressure gradient of $1 \times 10^7 \frac{\text{Pa}}{\text{m}}$. Lines demonstrate the quality of the correlations presented in Table 2.4.

The parallel plate with a porosity value of one and low wetted surface per area exhibits more than one order of magnitude increased permeability than the fully populated microchannel. The permeability is a constant throughout the range of pressure gradients considered in both cases. Fully populated pin-fin arrays in in-line or staggered orientation exhibit equal permeabilities at low REYNOLDS numbers, due to the invariant porosity and the diminishing influence of the inertia term in the momentum transport.

At increased pressure gradients the permeability of both geometries become non-linear. The flow in the staggered arrangement is exposed to a tortuosity $\tau > 1$. It is defined as $\tau = L_{streamline}/L_{direct}$ with $L_{streamline}$ being the length of the stream line and L_{direct} being the shortest distance from point A_n to its periodic counterpart A_{n+1} from fluid cell n to $n+1$, respectively. Inertia terms are especially dominant at high tortuosity, which explains the difference between staggered and in-line flows at high pressure gradients. The dominance of the momentum change caused by the pins at large pressure gradients in staggered arrangement becomes evident by the convergence of the modified permeability of pin-fins at infinite length (2D case) to the one of pin-fins at low aspect ratio includes endwalls. A sparse TSV population is implemented at chip locations with low communication needs. Therefore, channel walls or pin fins can be removed locally. Half-populated microchannels and pin-fin arrays are the result, with substantially increased permeability. The longitudinal arrangement of pin-fins at half population with 250 µm minimal fluid constriction compared to the 50 µm in case of transversal arrangement should be preferred. The lines in Figure 2.15 represent correlations describing the modified permeability and are of the form of

$$K(v_{darcy}) = -\frac{\mu_{25°C}}{a \cdot v_{darcy} + b}, \tag{2.43}$$

with coefficients a and b derived from a least square fit, the dynamic viscosity $\mu_{25°C}$ of water at 25 °C, and the DARCY velocity v_{darcy}. The coefficients are listed in Table 2.4.

MODIFIED PERMEABILITY [m^2] dimension: 100 µm height and pitch, 50 µm wall width		
PP	8.367×10^{-10}	
MC fp	7.171×10^{-11}	
MC hp	3.684×10^{-10}	
Pin fin	$K(v_{darcy}) = -\frac{\mu_{25°C}}{a \cdot v_{darcy} + b}$ with μ =0.89 mPa s at 25 °C	
	a [$\frac{\text{Pa s}^2}{\text{m}^3}$]	b [$\frac{\text{Pa s}}{\text{m}^2}$]
PF in-line fp	7.840×10^5	7.755×10^6
PF staggered fp	1.041×10^7	6.801×10^6
PF longitudinal hp	4.756×10^4	2.151×10^6
PF transversal hp	1.531×10^6	4.410×10^6
	$K(v_{darcy}, T_f) = -\frac{\mu(T_f)}{a \cdot v_{darcy} + b(T_f)}$	
PF in-line fp	1.457×10^6	$-9.499 \times 10^4 \frac{1}{°C} \cdot T_f + 9.985 \times 10^6$
PF staggered fp	9.552×10^6	$-1.225 \times 10^5 \frac{1}{°C} \cdot T_f + 1.065 \times 10^7$

Table 2.4: Correlations and their coefficients describing the modified permeability (Equation 2.42) of the cavity structures at variable or constant temperature at fully developed hydrodynamic boundary conditions.

The temperature field in the solid and fluid domain at a base heat flux of 100 $\frac{\text{W}}{\text{cm}^2}$ and a pressure gradient of 1×10^7 $\frac{\text{Pa}}{\text{m}}$ is presented in Figure 2.16 (left). The fluid and solid temperature increase from unit-cell inlet to outlet (in x-direction) is not significant at this flow rates. The location of lowest fluid temperature is congruent with the location of highest fluid velocity. It is found between the pins in transversal direction and the fluid cavity center. The fluid temperature in the recirculation zone (Figure 2.14 right) is quite high due to the minimal mass exchange with the main fluid flow. The solid domain with the low aspect ratio pin-fins (height to diameter equals two) is close to isothermal especially at low pressure gradients (Figure 2.16 right). The resulting fin efficiency is >97 % for both pressure cases.

The convective thermal resistance R_{conv} values according to Equation 2.39 were derived from the thermal models for the same heat transfer structures discussed in Figure 2.15 with constant material properties defined at 25 °C (Figure 2.17).

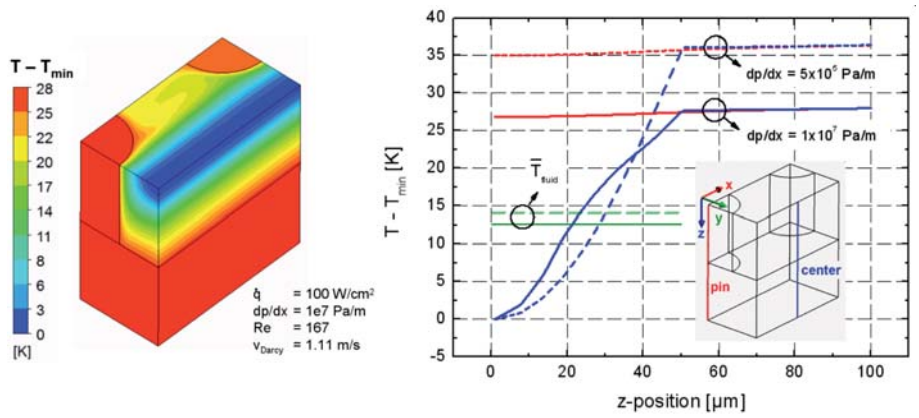


Figure 2.16: Temperature field of a PFI unit-cell at a pressure gradient of $1 \times 10^7 \frac{\text{Pa}}{\text{m}}$ with respect to the minimal temperature in the fluid at a heat flux of $100 \frac{\text{W}}{\text{cm}^2}$ (left). Detailed temperature comparison of the solid and liquid domain along the indicated lines oriented parallel to the z-direction for two pressure gradients.

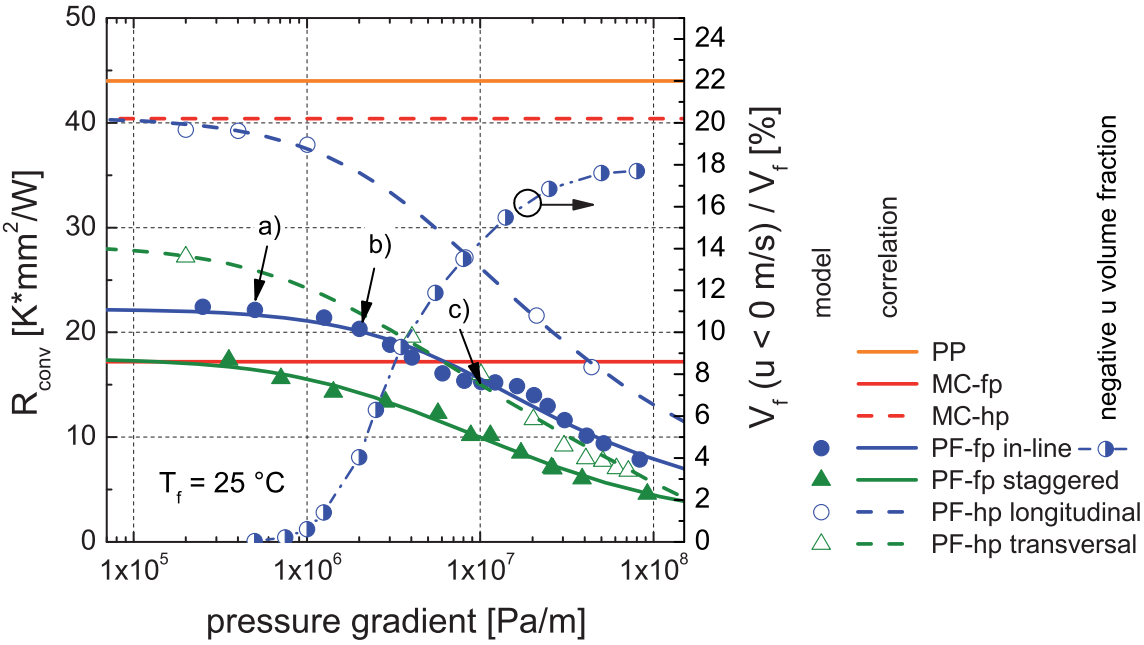


Figure 2.17: Convective thermal resistance R_{conv} as defined in Equation2.39 for parallel plate (PP), microchannel (MC), and pin fin (PF) cavities are presented from numerical modeling (dots). All values are computed at a water temperature of 25 °C. The orientation sensitivity of PF is demonstrated by the in-line and staggered case at full population, as well as for transversal and longitudinal flows in case of half populated PF. Lines demonstrate the quality of the correlations presented in Table 2.5. The arrows indicate the operation point of the stream line images presented in Figure 2.14 (a: left, b: center, c: right) for the PFI-fp case. The volume of the negative velocity component u (velocity in x-direction) in the unit-cell is plotted relative to the total fluid domain volume for the PFI-fp case. It is a measure for the recirculation size between the pins.

The fully populated microchannel exhibits a 2.6 times lower R_{conv} compared to the parallel plate. The wetted area and the hydraulic diameter in case of the PP (Table 2.3) is smaller and larger, respectively, resulting in increased heat diffusion length. Heat of the PF fp in-line structure is less efficiently dissipates to the fluid compared to the MC fp even at a larger wetted area and the same hydraulic diameter. This can be explained by the inefficient use of the wetted area for the heat transfer as depicted in Figure 2.14 (left). Only low fluid velocities exist between the pins in down stream direction and contribute only moderate to the total heat dissipation. A PF orientation change from in-line to staggered results in equal or improved heat transfer compared to MC even at low flow rates. This due to the tortuosity which causes high flow rates even behind the pins and therefore uses the wetted area more efficiently.

The convective thermal resistance is also reduced for PF cases analog to the permeability at increased pressure gradients. A local saddle point can be observed at $1 \times 10^7 \frac{\text{Pa}}{\text{m}}$ for the PF fp in-line structure. The onset of reducing R_{conv} seams to coincide with the appearance of recirculations behind the pins. The volume of the fluid domain with negative velocity component u (velocity in x-direction) with respect to the total fluid volume is plotted to support this statement. It represents the relative volume of the backflow in the recirculation. It increases monotonically and reaches an asymptotic value at a pressure gradient of $5 \times 10^7 \frac{\text{Pa}}{\text{m}}$. Additionally, the operation point of the stream line images presented in Figure 2.14 (a: left, b: center, c: right) are indicated. It seams, the saddle point coincides with the recirculation starting to span from one pin to the other. The effect of half populated MC and PF structures results in an approximate twofold increased R_{conv}. This is not true for transversal oriented pin-fins at hp. They result in an equal or slightly smaller R_{conv} compared to PF-fp in-line at moderate to high pressure gradients, due to the higher fluid velocity at equal pressure gradient.

The structure of the correlation used to define R_{conv} as a function of the DARCY velocity is

$$R_{conv} = c / \left(\frac{v_{darcy} + d}{1\,m/s} \right)^n + e \tag{2.44}$$

with coefficients c, d, e and n derived in the non-linear least square fit. The coefficients are listed in Table 2.5. Additional permeability and convective thermal resistance values are computed for different pin shapes and are assessed in the appendix A.

CONVECTIVE THERMAL RESISTANCE $[\frac{\text{K mm}^2}{\text{W}}]$ dimension: 100 µm height and pitch, 50 µm wall width				
PP	44.0			
MC fp	17.2			
MC hp	40.4			
Pin fin	$R_{conv} = c / \left(\frac{v_{darcy}+d}{1\,m/s} \right)^n + e$ at 25 °C			
	c $[\frac{\text{K mm}^2}{\text{W}}]$	d $[\frac{\text{m}}{\text{s}}]$	e $[\frac{\text{K mm}^2}{\text{W}}]$	n [-]
PF in-line fp	25.27	1.3662	1.533	0.6365
PF staggered fp	25.27	1.3500	1.533	1.5204
PF longitudinal hp	59.36	2.0374	−6.650	0.3220
PF transversal hp	132.80	0.2967	−113.600	0.0550

Table 2.5: Correlations and their coefficients describing the convective thermal resistance of the cavity structures at fully developed hydrodynamic and thermal boundary conditions.

The energy transport in the chip stack between the tiers is defined by the conductive thermal resistance of the pins and microchannel walls in the cavity. The R_{cond}, including the porosity and the wetted area per projected area for the discussed geometries is listed in Table 2.3. The MC fp results in strongest thermal tier-to-tier coupling thanks to the high channel wall fill ratio of 50 %.

A sensitivity analysis at a reference pressure gradient of $1 \times 10^7 \frac{\text{Pa}}{\text{m}}$ was performed, to account for geometrical imperfections, as well as for fluid temperature variations in the experiment. The sensitivity of v_{Darcy} with respect to pin diameter changes is plotted in blue in Figure 2.18 (left). It is getting more significant for small nominal pin diameters and pitches. It reaches a value of $-10 \frac{\%}{\text{µm}}$ for a 25 µm pin diameter and 50 µm pitch. This is expected, since a one micrometer pin diameter change causes a larger relative hydraulic diameter change at small, compared to large pitches. In general, the results demonstrate the importance of precise fabrication methods.

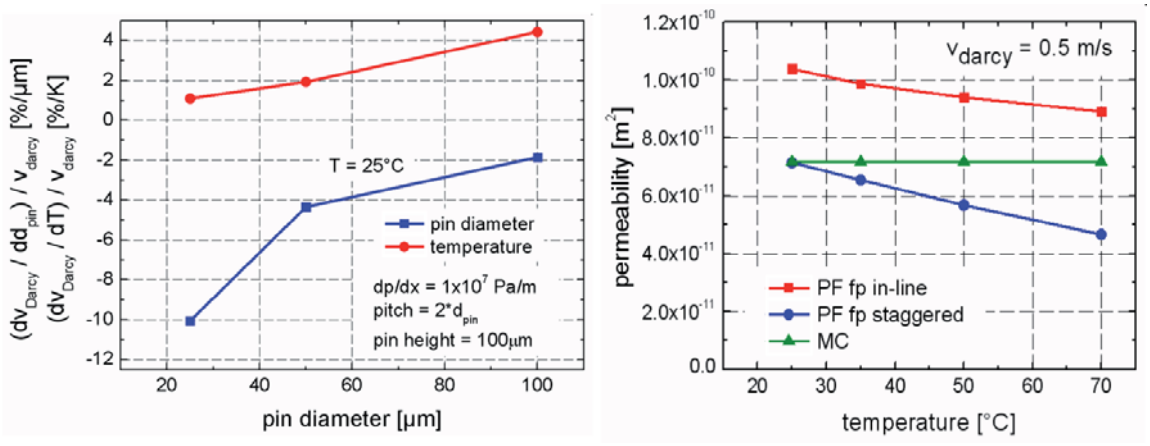


Figure 2.18: Sensitivity analysis indicating the dependency of v_{Darcy} on pin diameter and fluid temperature changes versus pin diameter. The values are computed for pin-fin in-line structures at a pin height of 100 µm and a pin pitch double the pin diameter (left). Temperature dependency of the modified permeability (as defined in Equation 2.42) for the microchannel (MC) and pin fin (PF) fully populated (fp) in in-line and staggered arrangement (right).

Additionally, the temperature influence on the flow rate is significant too. It results in a 2 % change per K in case of the reference geometry with 50 µm pin diameter (Figure 2.13 left red line). This effect has to be included in the models, since fluid temperature increases from inlet to outlet of up to 40 K can be expected. This results in a maldistribution of coolant between equidistant channels at spatial non-uniform power dissipation. The positive velocity to fluid temperature coupling causes increased flow rates in channels with high power dissipation and helps therefore to mitigate hot spots. The root cause of this effect is the water viscosity drop as shown in Figure 2.19. This property drops by 53 % for inter-layer cooling relevant fluid temperatures of 20 to 60 °C. The fluid density is the other parameter relevant in mass transfer. It is reduced by only 1.5 % in this temperature range and can therefore be treated as invariant. The volumetric heat capacity and the thermal conductivity of water are the relevant parameters defining the convective heat transfer and change by −1.5 % and 9 %. The thermal conductivity in the solid silicon is reduced by 13 % in the same temperature range.

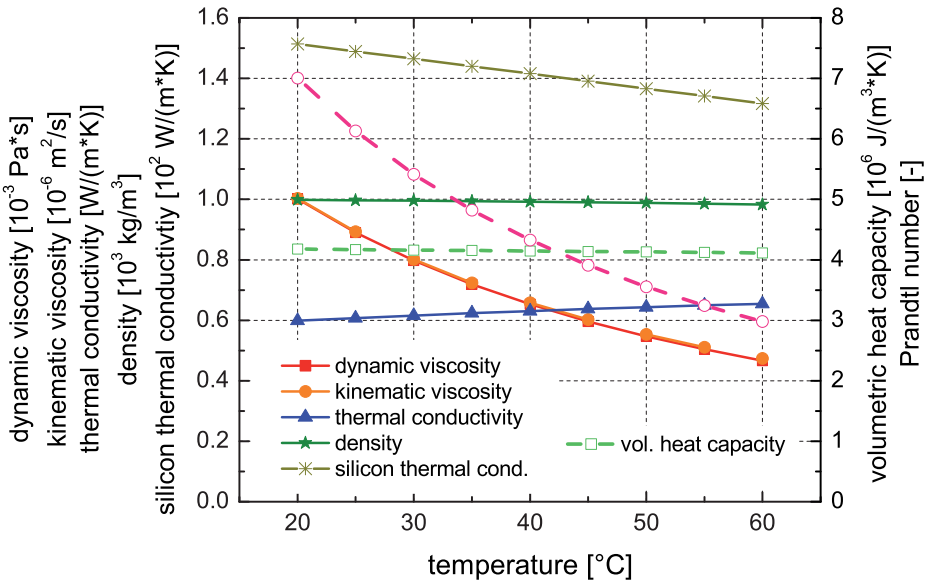


Figure 2.19: Temperature dependency of water and silicon material properties important in heat and mass transfer for the interlayer cooling relevant temperature range (water properties [95], silicon thermal conductivity [96]).

The modified permeability with respect to the fluid temperature for MC and PF fully populated (fp) in in-line and staggered configuration is summarized in Figure 2.18 (right) for DARCY velocities of 0.5 $\frac{\mathrm{m}}{\mathrm{s}}$. The value is invariant for the MC, but drops for both PF geometries if defined according to the modified DARCY's law (Equation 2.42). This behavior can be attributed to the high sensitivity of viscous, but close to invariant dependency of inertia damping with respect to fluid temperature changes. The viscous and the inertia damping scale proportional with the coolant viscosity and density, respectively. Only the viscosity changes significant in the relevant temperature range. Hence, the modified permeability has to drop for geometries with a significant contribution of inertia damping at a given flow rate. This is strongly the case for the pin-fin staggered design.

Accordingly, the modified permeability Equation 2.43 is extended with the temperature dependent dynamic viscosity $\mu(T_f)$ and coefficient $b(T_f) = f \cdot T_f + g$ as:

$$K(v_{darcy}, T_f) = -\frac{\mu(T_f)}{a \cdot v_{darcy} + b(T_f)}, \tag{2.45}$$

to account for temperature variations in the fluid cavity. The values of coefficient a, f, and g are listed in Table 2.4 for pin-fin structures. The accuracy of the data fit for PF-fp in-line and staggered is within $\pm 10\,\%$ for the temperature range from 25 to 70 °C as can be seen in Figure 2.20.

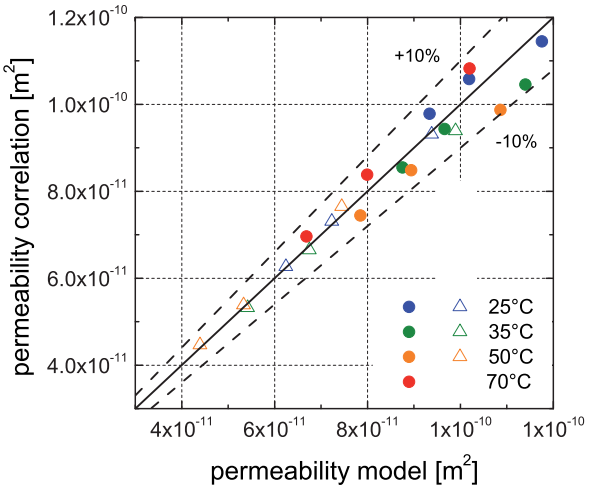


Figure 2.20: Quality of temperature dependent modified permeability correlation for PF-fp in-line and staggered in the temperature range of 25 to 70 °C.

2.3.3 Parameter Set for Pin-Fin Arrays at Arbitrary Angle-of-Attack

Pressure gradients parallel to the symmetry lines of the pin-fin arrays were considered in the previous sections. This resulted in an anti-parallel flow direction, which is predominant at uniformly populated heat transfer cavities in two-port operation. The more general case with an arbitrary flow direction θ, here referred to as angle-of-attack (AoA), is present in non-uniform heat transfer cavities or in cavities with four-port fluid delivery as described in [87]. The pressure gradient being the root-cause to drive the flow is not anti-parallel with respect to the velocity vector (Figure 2.21). Up to know, only one study could be identified considering discrete AoA for pin arrays at low aspect ratio. METZGER et al. [97] derived the friction factor and NUSSELT number from an experimental investigation for flow directions of 5°, 19°, 32° and 45° in the turbulent flow regime. They conclude with a monotonic friction factor increase between in-line and staggered pin arrangement. This results stay in contrast to investigations performed on tube bundles (pin arrays with infinite aspect ratio), which observed a friction factor maximum at flow directions of 32° [98]. But, the maximal NUSSELT number was identified for a flow direction of 19° in both studies.

The modified permeability dependency on the AoA will be investigated in the following section for laminar flows and pin-fin arrays with circular pins in full or half population. The full unit-cell (dashed

box in Figure 2.21) of the pin-fin array geometry including the endwalls and two fluid dynamic periodic boundary pairs (S_{11} with S_{12} and S_{21} with S_{22}) are considered to derive the DARCY velocity vector $\vec{v}_{Darcy}$ at a given pressure gradient ∇p. It results from pressure boundary conditions imposed across the unit cell in x and y-direction (Δp_x, Δp_y) :

$$\nabla p = \begin{pmatrix} \frac{\Delta p_x}{l_x} \\ \frac{\Delta p_y}{l_y} \end{pmatrix}, \tag{2.46}$$

with the unit cell length l_x and l_y. Dimensions of 100 µm pin pitch and cavity height and 50 µm pin diameter are used as a reference. The DARCY velocity vector $\vec{v}_{Darcy}$ is defined from the resulting velocity field as:

$$\vec{v}_{Darcy} = \begin{pmatrix} \frac{\dot{V}_{S12}}{l_y \cdot h} \\ \frac{\dot{V}_{S22}}{l_x \cdot h} \end{pmatrix}, \tag{2.47}$$

with the volumetric flow rate $\dot{V}_{s12}$ and $\dot{V}_{s22}$ at face $S_{12}$and S_{22}, the unit cell length l_x and l_y, as well as the cavity height h.

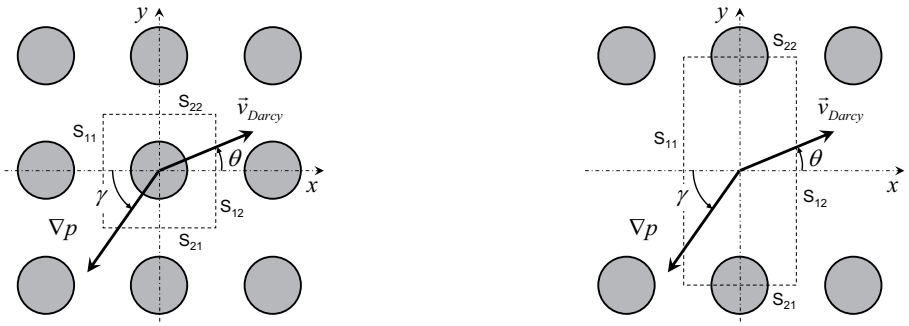


Figure 2.21: Fluid flow through a periodic pin-fin cavity at full (left) and half population (right), with an arbitrary angle-of-attack (AoA) θ and the non-anti-parallel pressure gradient ∇p.

The resulting velocity vector field in the cavity center plane of the fully populated pin-fin array at a pressure gradient direction γ of 22.5° and a pressure gradient of $1 \times 10^5 \frac{\text{Pa}}{\text{m}}$ and $1 \times 10^8 \frac{\text{Pa}}{\text{m}}$ is presented in Figure 2.22. The pressure gradient direction γ to AoA θ offset diminishes at creeping flow (viscous damping dominant flow regime) due to the convergence of the in-line and staggered modified permeability as discussed in Figure 2.15 (Figure 2.22 left). Whereas the mass flow occurs preferentially along the in-line direction at high pressure gradients due to the significant inertia damping resulting in a higher modified permeability in in-line, compared to the staggered direction (Figure 2.22 right).

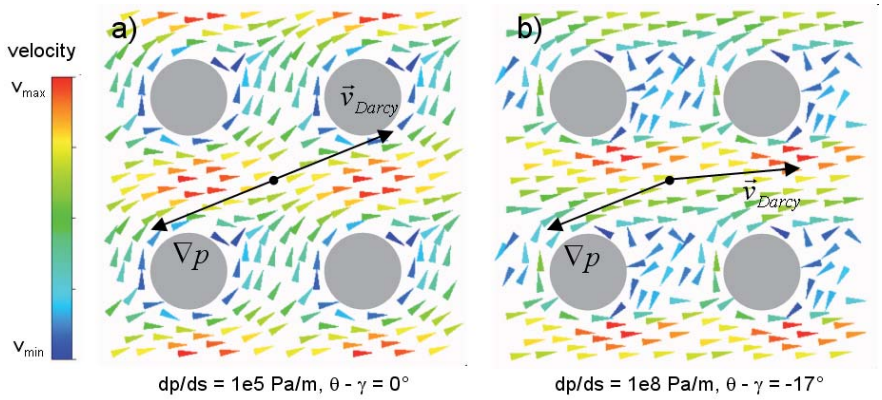


Figure 2.22: Mass flow direction θ at a given pressure gradient direction γ of 22.5° for a pressure gradient magnitude of 10^5 and $1 \times 10^8 \frac{\text{Pa}}{\text{m}}$.

A more detailed analysis is depicted in Figure 2.23 with resulting flow directions θ for a varying pressure gradient magnitude and direction γ for a cavity with fully populated pin-fin array and the 2D case, considering infinitely long pin-fins. At pressure gradients $<1 \times 10^6 \frac{\text{Pa}}{\text{m}}$ the flow direction in the fluid cavity is anti-parallel to the pressure gradient. Above this value, the flow direction for γ $<22.5°$ converges towards the in-line direction. For $\gamma \geqq 22.5°$ a moderate convergence towards the staggered direction up to $3.5 \times 10^6 \frac{\text{Pa}}{\text{m}}$ can be observed, with a subsequent convergence change towards in-line direction up to $1 \times 10^8 \frac{\text{Pa}}{\text{m}}$. At such high pressure gradients only the datapoints representing flow direction with γ =37.5° tends towards a staggered flow. The minimal modified permeability of the pin-fin structure seams to exists for a AoA of around 30°. This behavior of flow separation is more pronounced for infinitely long pin fins (2D case) without endwall damping.

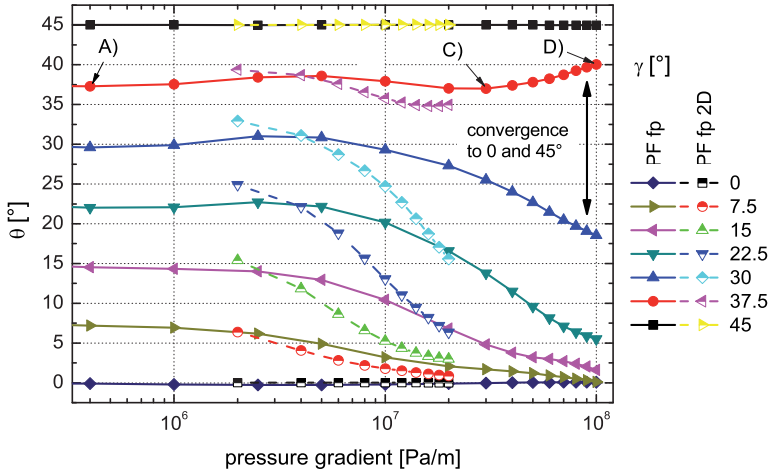


Figure 2.23: Detailed analysis of the flow convergence effect, indicating a local permeability minimum at a θ of around 30° at high pressure gradients. A fully populated pin-fin cavity and infinitely long pin-fins (2D case) are considered. The flow direction θ is plotted for different pressure gradients directions γ.

The corresponding vector fields for the half populated pin-fin case is shown in Figure 2.24. Recirculations spanning the entire space between the pins in down stream direction exist for longitudinal or transversal flow directions (0° and 90°). The recirculations are distorted and reduced in size due to bypassing fluid for all other cases. The flow is biased in longitudinal direction up to an AoA of 60°. Additional details are depicted in Figure 2.25. At high pressure gradients two local permeability minima at around 20° and 75° seam to exist. This also corresponds with the findings of to the fully populated case with also two, but symmetric local minima in the flow directions range of 0° to 90°. Even in the creeping flow regime a deviation from an anti-parallel flow to pressure gradient direction can be observed. This can be explained by reduced viscous damping at the larger hydraulic diameter of 120.0 µm in case of longitudinal flow compared to the 66.6 µm for the transversal flow. Accordingly, the mass flow shows an offset towards the longitudinal direction compared to the pressure gradient.

The modified permeability defined according to Equation 2.42 using the norm of $\vec{v}_{Darcy}$ and ∇p for a DARCY velocity of $1 \frac{\text{m}}{\text{s}}$ is plotted for pin-fin fully populated with endwalls and for infinitely long pins (2D), as well as for pin-fin half population in the polar plot at logarithmic scale (Figure 2.26 left). The corresponding absolute value of the flow direction to pressure gradient direction deviation is presented in Figure 2.26 (right). The angular characteristics of PF fp with low aspect ratio and infinitely long pins is equivalent. The endwalls of the low aspect ratio case result only in a direction indipendent offset attributed to additional viscous damping. The two-fold symmetry of the PF hp geometry is clearly reflected, as well as maximal permeability in longitudinal direction. The 60° local maxima (and its symmetry pairs) is also clearly visible.

Different methods to describe the angle dependent permeability in a porous domain are compared next. A basic sinusoidal interpolation could be applied if only the permeability values of in-line and staggered or longitudinal and transversal arrangement are known. This reads:

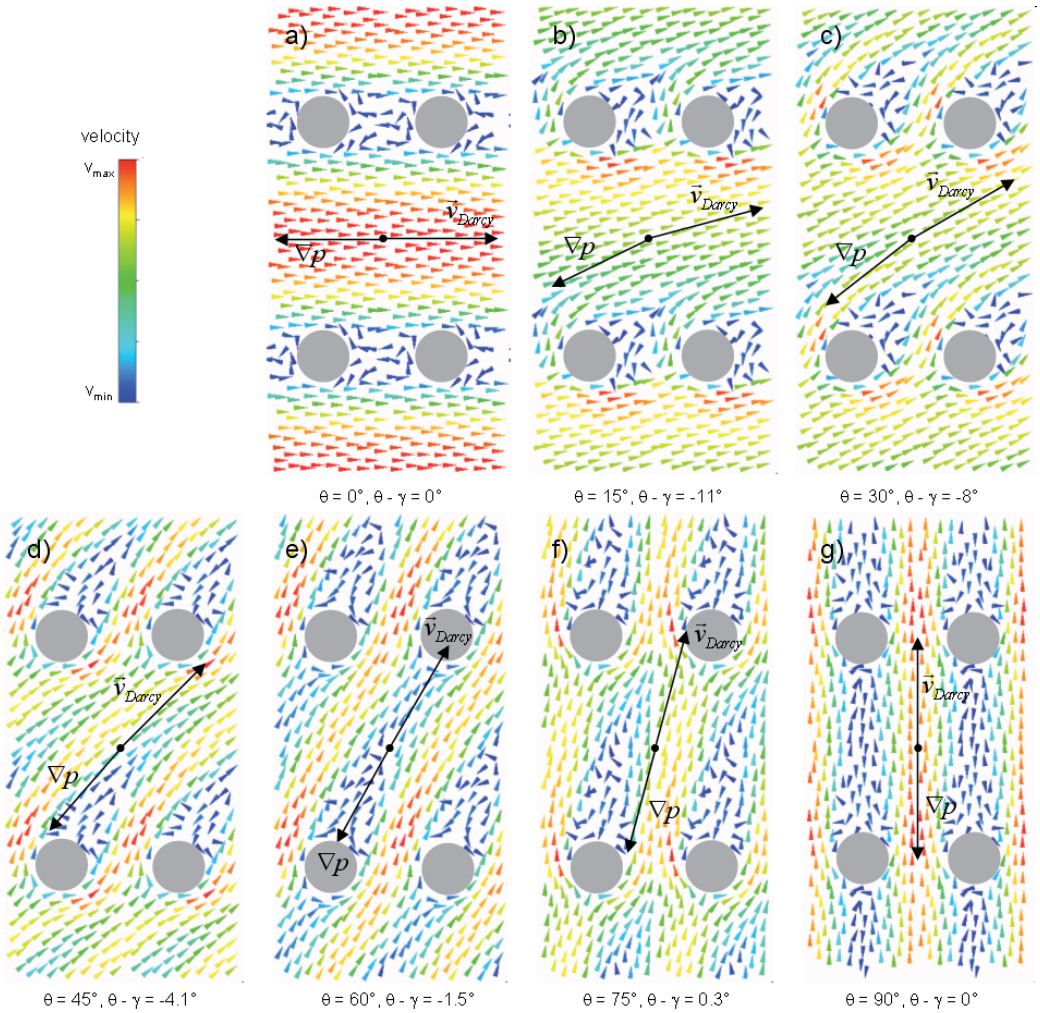


Figure 2.24: Velocity vector field in the center plane of the pin-fin half populated cavity at a pressure gradient of $1 \times 10^7 \frac{\text{Pa}}{\text{m}}$ with indicated γ to θ offset at different DARCY flow directions.

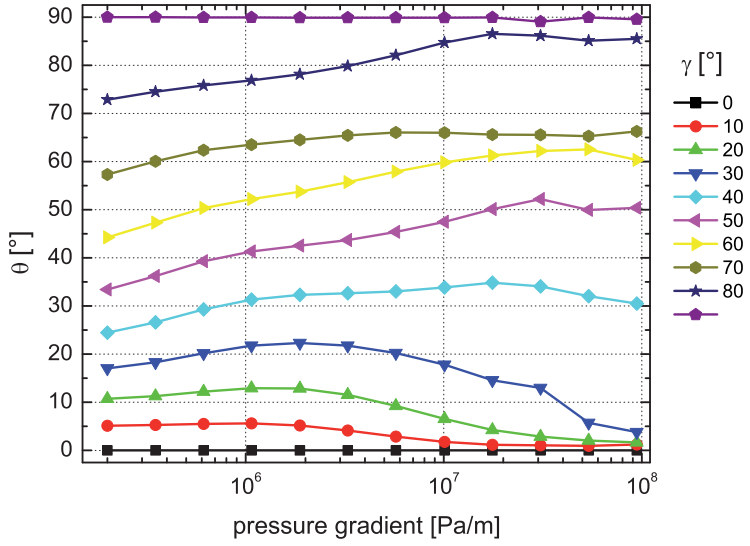


Figure 2.25: Detailed analysis of the flow convergence effect in a cavity with half populated pin-fins, indicating two local permeability minima at a θ of around 20° and 75° at high pressure gradients. The flow direction θ is plotted for different pressure gradient directions γ.

$$K(\theta) = \frac{\kappa_1 + \kappa_2}{2} + \frac{\kappa_1 - \kappa_2}{2} cos(n\theta), \tag{2.48}$$

with indices 1 for in-line or longitudinal and 2 for staggered or transversal and the coefficient $n = 4$ and $n = 2$ in case of full and half population, respectively. An other attempt is to use a tensor notation with the principle permeabilities κ_{xx} and κ_{yy} :

$$K = \begin{pmatrix} \kappa_{xx} & 0 \\ 0 & \kappa_{yy} \end{pmatrix}. \tag{2.49}$$

This method is used to describe orthotropic materials and yields an elliptic angular permeability with a two-fold symmetry. Therefore, only unit-cell geometries with a two-fold symmetry such as PF hp might be approximated with this method. κ_{xx} would represent the longitudinal and κ_{yy} the transversal permeability. A more general method would be to map the pressure gradient vector exactly to the DARCY velocity vector by the use of a tensor notation with pressure gradient dependent (norm $\frac{dp}{ds}$ and direction γ) principle coefficients such as depicted in:

$$K(\frac{dp}{ds}, \gamma) = \begin{pmatrix} \kappa_{xx}(\frac{dp}{ds}, \gamma) & 0 \\ 0 & \kappa_{yy}(\frac{dp}{ds}, \gamma) \end{pmatrix}. \tag{2.50}$$

The corresponding velocity vector can be defined for each pressure gradient. The reverse assignement is not definitely due to the convergence of the flow towards the minimal pressure gradient directions. The values of the principle coefficients are derived from the detailed numerical study for the pressure gradient of interest with angular resolution of 7.5° and 15° in the range of 0° to 45° and 0° to 90° in case of full and half populated pin-fins, respectively. The results are stored in a look-up table . An interpolation scheme based on distance weighted averaging of the closest three points from the look-up table is used later to perform the mapping for generic pressure gradient directions.

The accuracy of the proposed descriptions for PF fp is demonstrated in Figure 2.27 with respect to results derived from detailed modeling (blue line). The sinusoidal interpolation (green dashed line) considering only one frequency (Equation 2.51) is not able to capture the permeability minima between the in-line and the staggered direction, resulting in a poor coefficient of determination (R^2-value) of 0.64 and a large mean absolute error of $13.5 \times 10^{-12}\,\mathrm{m}^2$ (Figure 2.27 left). Furthermore, only anti-parallel flows to pressure gradients are predicted (Figure 2.27 right). Compared, the variable tensor notation (Equation 2.50) depicts the permeability and the offset from anti-parallelism correctly with a R^2-value quality of 0.97 and 0.93 and mean absolute error of $2.1 \times 10^{-12}\,\mathrm{m}^2$ and 0.5°, respectively (red triangles).

Values resulting from the orthotropic description (Equation 2.49) for the pin-fin half populated geometry is additionally included in Figure 2.28 (brown dashed line). The description quality is improved from a mean absolute error of $86.8 \times 10^{-12}\,\mathrm{m}^2$ to $53.6 \times 10^{-12}\,\mathrm{m}^2$ compared to the sinusoidal description (green dashed line). Furthermore, the tensor model results in an angle offset, but at a too large magnitude, resulting in a mean absolute error of 18.8°. Again, the extended tensor description performs best and results in a mean absolute error of $14.6 \times 10^{-12}\,\mathrm{m}^2$ and 2.19°.

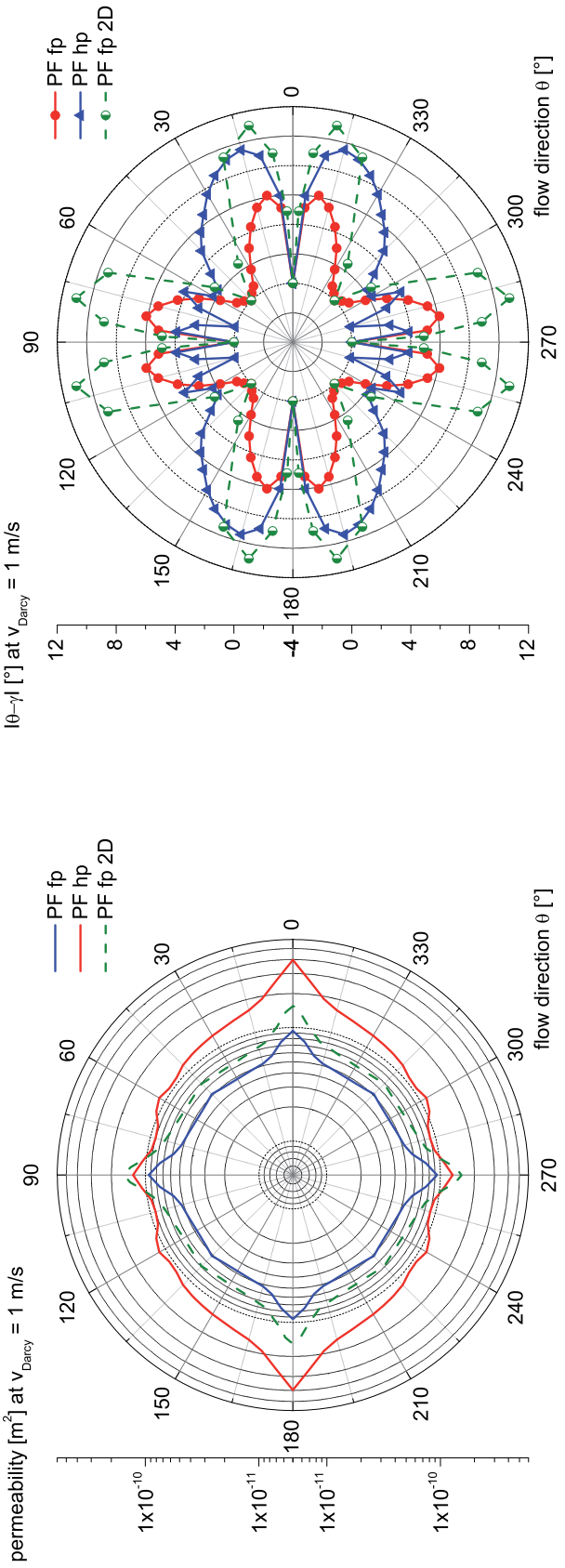


Figure 2.26: Permeability (left) and absolute direction difference between the DARCY velocity and the pressure gradient vector (right) as defined in Figure 2.21 for fully (PF fp)and half populated (PF hp) low aspect ratio pin-fins and fully populated, infinitely long pins (PF fp 2D).

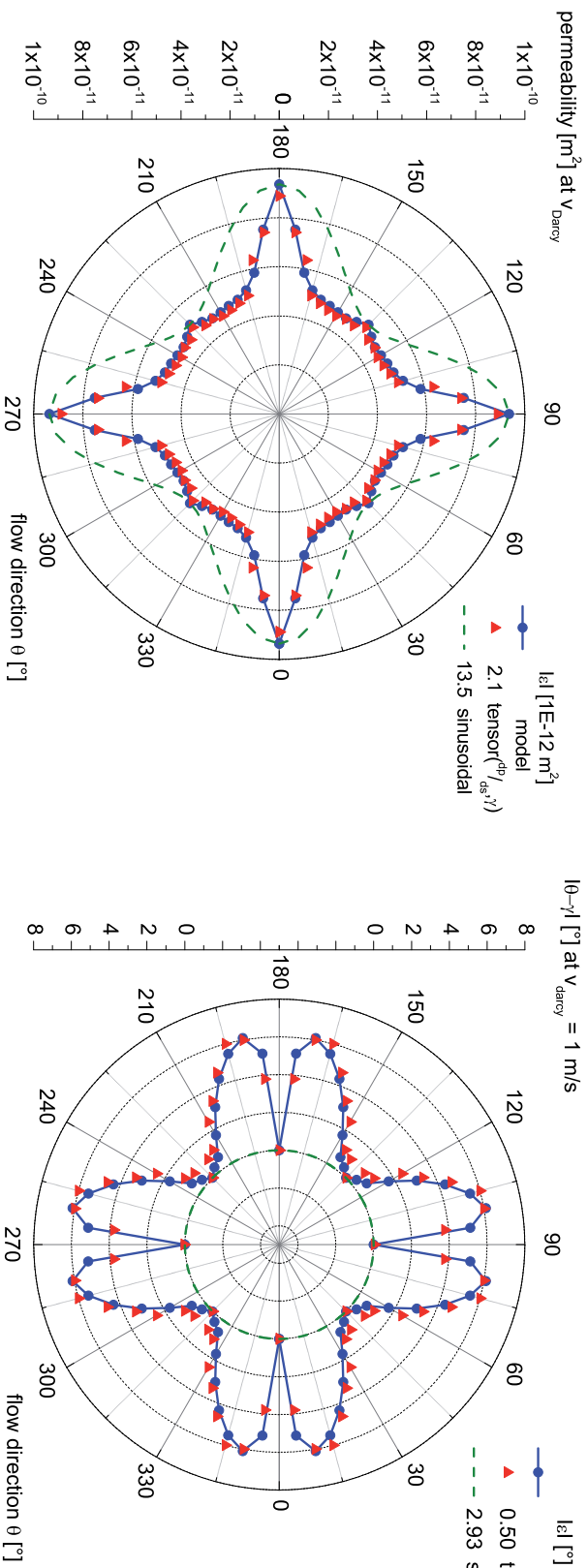


Figure 2.27: Comparison of methods to describe the angular permeability (left) and angle offset (right) to results derived by detailed modeling (blue lines) for PF fp at a DARCY velocity of 1 $\frac{\mathrm{m}}{\mathrm{s}}$. The results of the extended tensor description and the basic sinusoidal correlation are plotted in red and green, respectively. The mean absolute error $|\bar{\epsilon}|$ is listed as well.

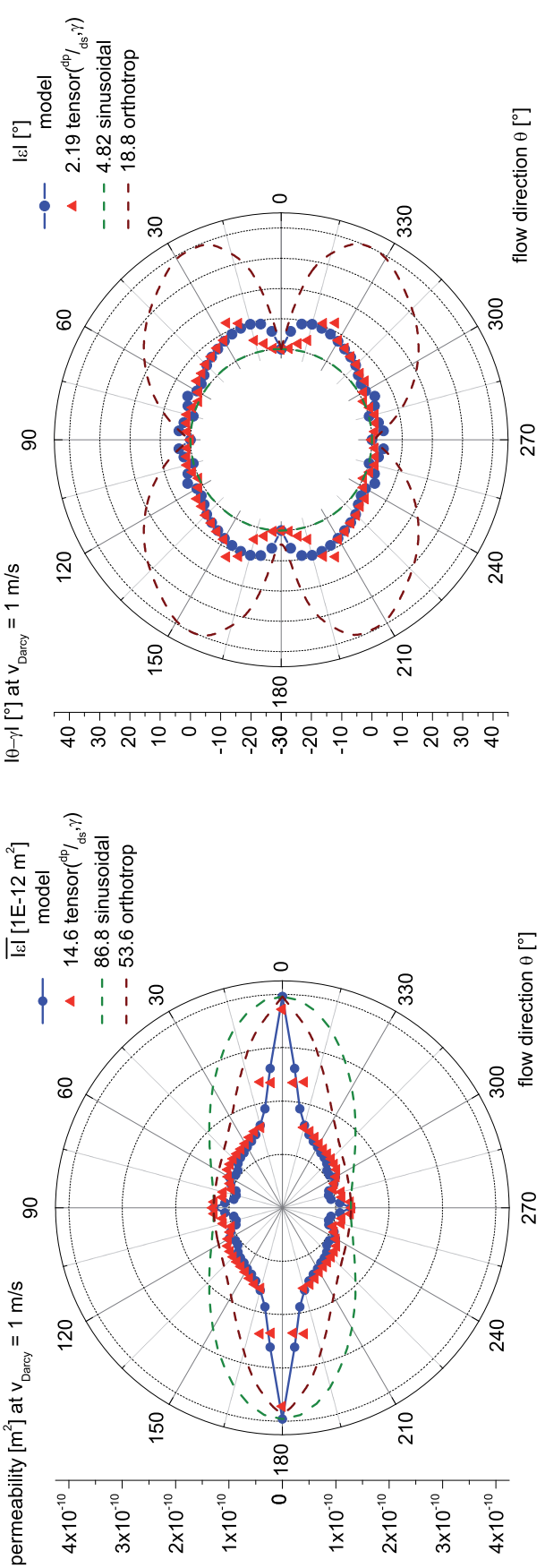


Figure 2.28: Comparison of methods to describe the angular permeability (left) and angle offset (right) to results derived by detailed modeling (blue lines) for PF hp at a DARCY velocity of 1 $\frac{\mathrm{m}}{\mathrm{s}}$. The results of the extended tensor description (red), of the basic sinusoidal correlation (green dashed line), and of the orthotropic description (brown dashed line) are plotted. The mean absolute error $|\bar{\epsilon}|$ is listed as well.

The complete characteristics (permeability and angle offset) of the pin-fin full and half populated cavity at 25 °C is summarized for the pressure gradient range of interest ($1 \times 10^5 \frac{\text{Pa}}{\text{m}}$ to $1 \times 10^8 \frac{\text{Pa}}{\text{m}}$), with respect to the AoA in Figure 2.29 and 2.30, respectively. An additional look-up table with the modified permeability at a fluid temperature of 60 °C is computed and is used in the porous media model to also account for the temperature changes in the coolant.

The description complexity of the scalar convective thermal resistance is low, compared to the vector-to-vector mapping in case of the DARCY velocity. Two periodic thermal boundary conditions are needed in addition to the hydrodynamic periodicity on the identical surface-pairs (S_{11} with S_{12} and S_{21} with S_{22}, Figure 2.21) to compute the AoA dependent heat transfer coefficient. The mapping method described in section 2.3.1 allows to couple only one boundary-pair and can therefore only be used for flows parallel to the geometrical symmetry line. A computational more expensive solution would be to model a pin array in all the details and to probe the needed thermal parameters for the cell with developed flow conditions. The number of pins needed to reach developed boundary layers for PF fp in-line was in the range of 40, as discussed in section 2.3.1. This is already the largest possible case to solve on a single 32-bit server with the physical memory limit of 4 GB. The number of unit-cells to be included in a model with arbitrary flow angles would be 40x40. Such problems would need a high-performanc computer to be solved. Therefore, a basic sinusoidal interpolation between the derived convective thermal resistance values for flows parallel to the geometrical symmetry lines (Table 2.5) is used

$$R_{conv}(\theta) = \frac{R_{conv_1} + R_{conv_2}}{2} + \frac{R_{conv_1} - R_{conv_2}}{2} cos(n\theta), \tag{2.51}$$

with indices 1 for in-line or longitudinal and 2 for staggered or transversal flow direction and the coefficient $n = 4$ and $n = 2$ in case of full and half population, respectively. The resulting interpolation for pin-fin fully and half population in the range of DARCY velocities of interest are depicted in the polar contour plots in Figure 2.31.

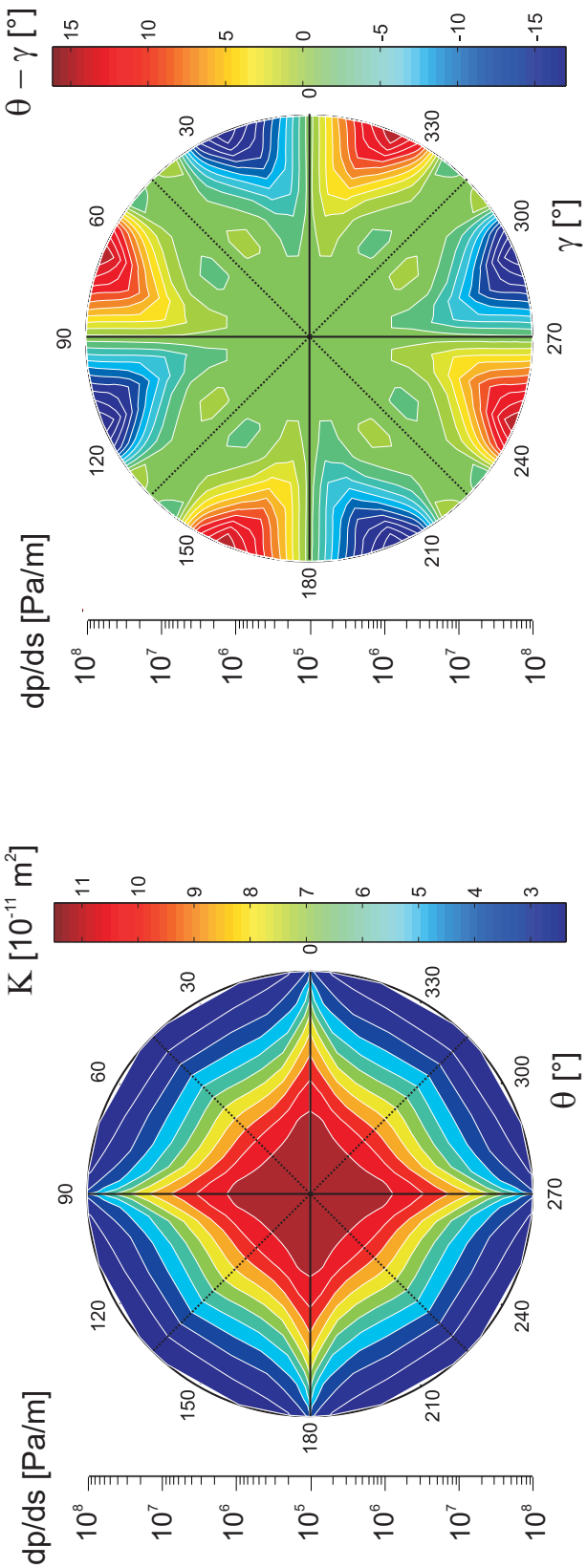


Figure 2.29: The permeability (left) and the angle offset (right) with respect to the angle-of-attack and the pressure gradient direction, respectively is plotted for fully populated pin-fin arrays at 25 °C and the pressure gradient range of interest.

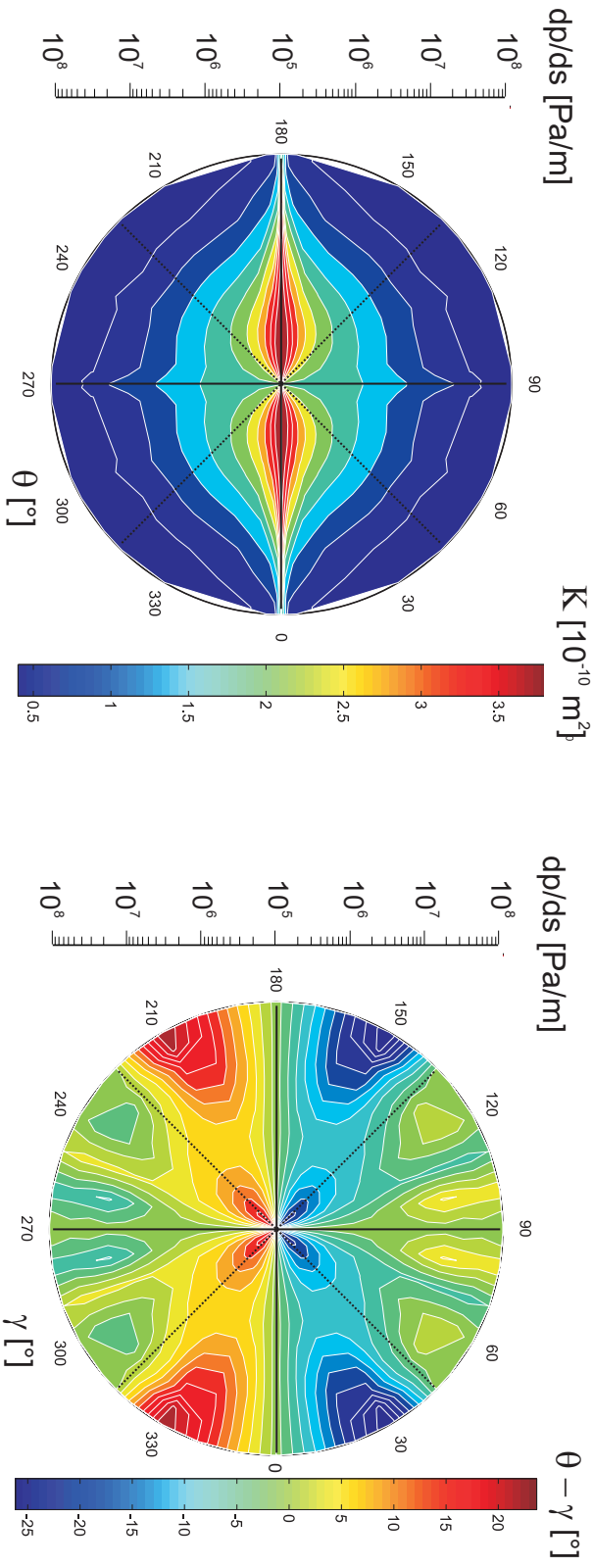


Figure 2.30: The permeability (left) and the angle offset (right) with respect to the angle-of-attack and the pressure gradient direction, respectively is plotted for half populated pin-fin arrays at 25 °C and the pressure gradient range of interest.

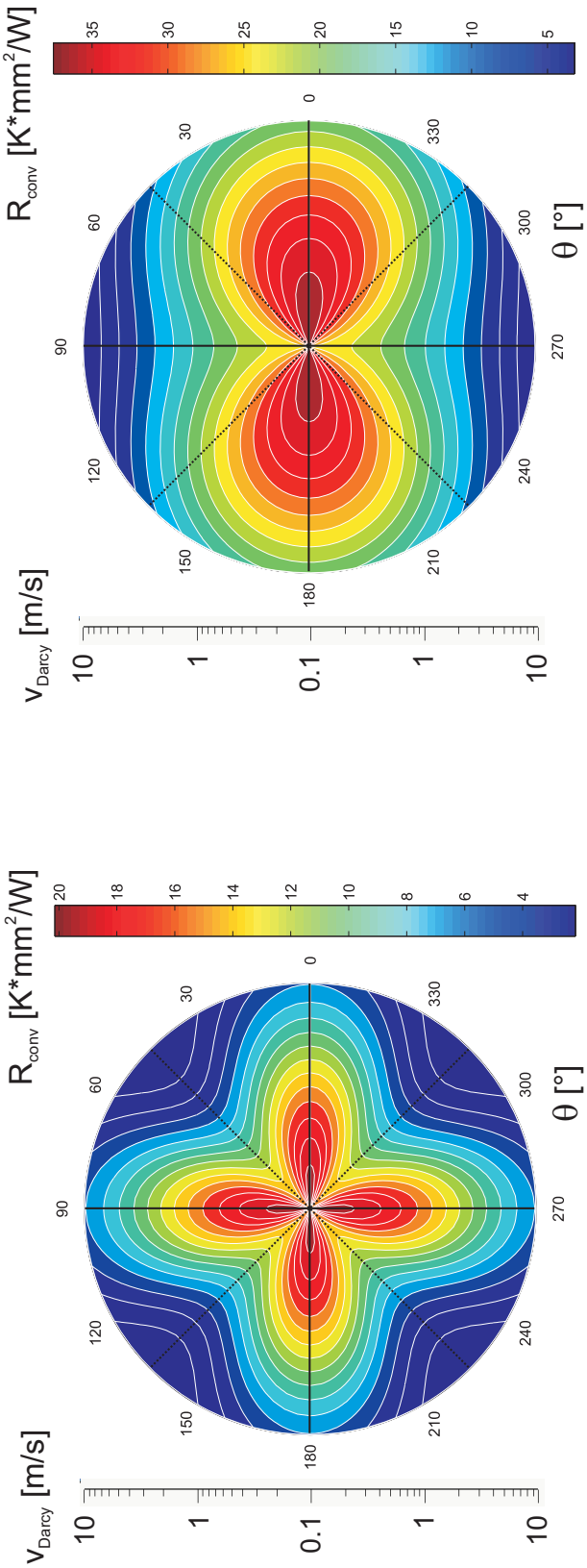


Figure 2.31: Convective thermal resistance of fully (left) and half (right) populated pin-fin cavities derived from a sinusoidal interpolation of numerically computed values for flows parallel to the geometrical symmetry lines. The contour plot depicts values in the DARCY velocity range of interest and with respect to the angle-of-attack.

3 Eutectic Bonding, Test Cavities, and Experimental Apparatus for Characterization

The incorporation of fluid cavities into a chip stack triggers the development of novel bonding, sealing, and shielding technologies, to maintain power and signal integrity. This is espcecially important if water is used as the coolant. Hence, fluid cavities etched into the silicon die backside in combination with eutectic thin film bonding including ring-pad sealing concepts are proposed, to prevent shortage be tween TSVs. Additionally, water-proof metal wiring layers are reported on in this chapter. Multiple heat transfer test sights with single or multiple fluid cavities were finally implemented with the developed technology. They allow for detailed heat and mass transfer investigations in combination with the described test facility, using spatially resolved infrared imaging. They are the experimental foundation of the results presented in chapter 4 and allow the validation of the modeling concepts introduced in chapter 2.

3.1 Eutectic AuSn 80/20 Thin Film Bonding: Low Thermal Resistance, Leak Tight Interface

In this section the rational behind the sealing concept and the eutectic Au-Sn thin film solder selection is depicted. A subsequent discussion on the significant process parameters, including the system modifications for reductive atmosphere thermo-compression bonding, is given. Finally, the bond quality was characterized, to identify the thermodynamic stability of underbump metallizations (UBM), to identify bond-line uniformity and void content.

3.1.1 Solder Technologies for Structural and Thermal Interfaces

The bonding interface between individual IC-dies in the stack has to fulfill electrical, thermal, and mechanical requirements. As shown in Table 3.1, bonds with satisfactory characteristics may be formed by applying either a soldering process, a thermally and electrically conductive adhesive, or by performing copper-to-copper thermocompression bonding.

Bonding Technology	Thermal Conductivity $[\frac{W}{mK}]$	Bond-Line Thickness [µm]	Thermal Resistance $[\frac{K mm^2}{W}]$
Acceptable values		≤ 10	≤ 1.00
Solder	40 to 90	4 to 100	0.04 to 0.25
Polyimide	0.2	0.1 to 1	0.50 to 5.00
Thermal adhesive	7	1 to 5	0.14 to 0.71
Cu-Cu bond	360	0.1 to 1	0.0003 to 0.0028
Si-Si fusion bonding	140	0	0

Table 3.1: Bond process overview, including their thermal properties [99].

All these bonds result in a thermal resistance smaller than 1 $\frac{K mm^2}{W}$. Additional criteria like processing robustness with respect to surface topography and a high reliability, resulted in the selection of a solder technology for bonds. The use of adhesives was abandoned as these are prone to swelling in contact with fluids, with a subsequent degradation of adhesion strength. Cu-Cu thermocompression bonding was

abandoned due to the need of chemical-mechanical polishing (CMP) to guarantee an intimate surface-to-surface contact. This process step increases the manufacturing complexity substantially.

To seal the electrical active TSV from the water (coolant), a sealing pad or ring needs to surround the TSV pad in case of microchannels or pin fin arrays, respectively (Figure 3.1). The TSV pad serves electrical, thermal, and mechanical purposes. The sealing structure, on the other hand, needs to be leak tight, but does not have to support any electrical aspects. It is important to maximize the porosity of the cavity for fluid dyamic reasons as this will result in a high coolant mass flow. Hence, the sealing structure dimension should be minimal. For a TSV diameter of 20 µm a TSV pad diameter of 25 µm, with a sealing pad / ring spacing and width of 5 µm is desirable, resulting in a channel width or pin fin diameter of 50 µm. Accordingly, only thin film solder with a thickness $\leqq 10$ µm is applicable to prevent solder bridging between sealing structure and TSV pad.

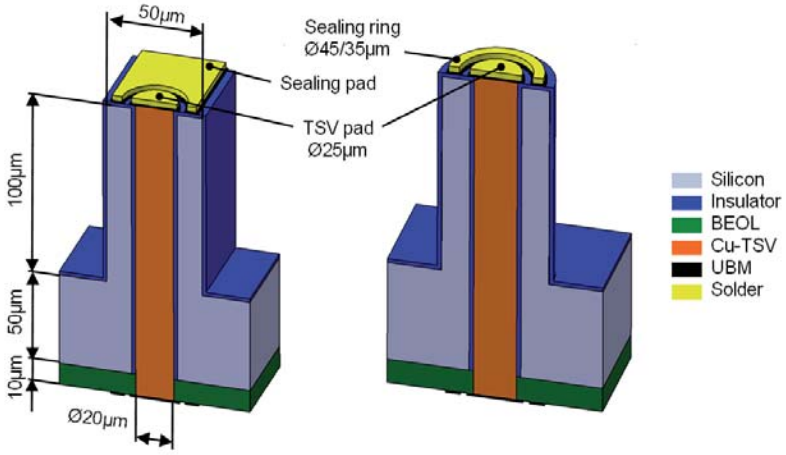


Figure 3.1: Cross-section and possible dimensions of a microchannel (left) and pin-fin (right) unit-cell, depicting the sealing concept with a sealing pad or ring surrounding the electrically active TSV pad.

The selection of a thin film solder was constrained by the BEOL temperature limit of 400 °C, as well as subsequent solder reflow temperatures in higher-level packaging. The bonding temperature should thus not exceed 400 °C whereas the final bonded interface material should be able to withstand solder reflow temperatures. The typical reflow temperature for a solder is 30 K higher than its liquidus temperature. It is ∼250 °C in case of SnAg 96.5/3.5, one of the lead free solders used as controlled-collapse-chip-connection (C4), so this was the minimum solder temperature the material must withstand. Eutectic AuSn 80/20 (numbers indicate weight percent) is a solder that fulfills both given temperature criteria as indicated in Table 3.2. The eutectic solder results in a high thermal conductivity and a high tensile strength. Additionally, eutectic AuSn 80/20 can be expected to have a high chemical stability against corrosion due to its high gold content. Eutectic AuSi was not selected due to its high reflow temperature close to the BEOL temperature limit.

Material [weight %]	Liquidus Temperature [°C]	Thermal Conductivity $[\frac{W}{mK}]$	Tensile Strength [MPa]
Acceptable values	250 to 400	high	> 10
AuSn 80/20	278	57	198
SnAg 96.5/3.5	221	26	26.7
SnAgCu 95.6/3.5/0.9	217	-	48
PbSn 63/37	183	41	54
In	157	84	4.5
AuSi 96.8/3.2	363	∼318	280

Table 3.2: Selection of solder alloys and their properties [99].

The binary phase-diagram of the Au/Sn system is depicted in Figure 3.2. Its complexity results from the existence of four different intermetallic compounds and two eutectic phases. The gold rich eutectic (AuSn 80/20) is most prominently used as a solder alloy, due to its low melting temperature of 278 °C as wells as for its superior mechanical properties, compared to the tin rich eutectic AuSn 10/90, which consists of brittle intermetallic. The gold rich eutectic is composed of the stable ζ' (Au_5Sn) and the δ (AuSn) phase (as indicated in Figure 3.2). The $\zeta'$and δ - phase can be differentiated in a secondary electron scanning electron microscope analysis by their signal intensity. Areas with increased electron density (gold rich) appear as bright gray due to the increased scattering probability with primary electrons (Figure 3.2 inlet).

With regard to process stability, it is a challenge that the solder liquidus temperature depends strongly on the gold tin ratio in the gold rich eutectic of the phase diagram. Around the eutectic point a 1 wt. % shift in gold or tin concentration results in a 46 K or a 9 K liquidus temperature increase, respectively. A slightly tin richer system should be targeted to minimize process variations.

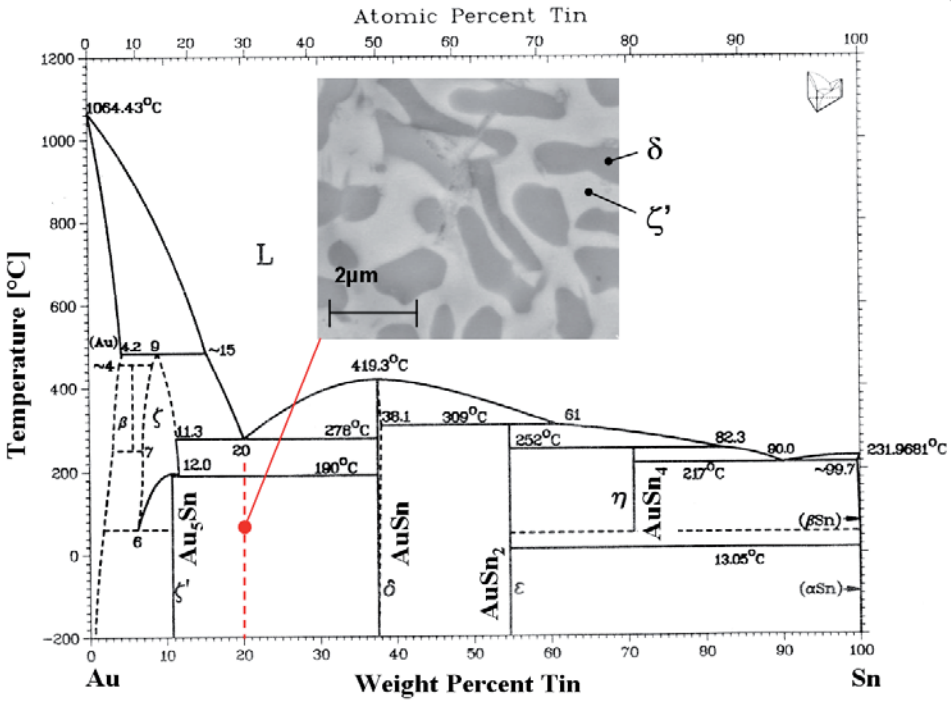


Figure 3.2: Binary Au-Sn equilibrium phase-diagram [100]. The morphology ($\zeta'$and δ phase) of the AuSn 80/20 eutectic is depicted on the secondary electron SEM image (inset).

Electroplating, evaporation, and sputtering are the three major Au/Sn deposition techniques reported. Electroplating is the most cost effective technology for thick solder layers $\geqq 10\,\mu m$. Due to the lack of a eutectic plating process, the gold and tin layers are deposited in subsequent steps with a thickness ratio corresponding to the targeted final composition (80/20) [101, 102, 103, 104]. A reflow in a liquid medium has been proposed to form the eutectic and thereby prevent oxidation before bond-line formation [105, 106]. A multi-layer electron-beam evaporation process of four pairs of alternating layers of Sn (0.35 µm) and Au (0.2 µm) was performed by KATZ et al. to achieve a bond-line with an overall eutectic composition [106]. Sputtering however is suitable to perform this directly from a eutectic target. Thin layers with a thickness uniformity that was satisfactory for this study can be obtained. A close to eutectic target (AuSn 79/21) was selected due to the aforementioned process stability issues. The layer thickness was varied between 1.5 and 4 µm. It is possible to pattern such films by lift-off techniques. This process step was not included for the fabrication of the thermal test vehicles for simplicity reasons, since they do not include TSVs. Hence, continuous, un-patterned films could be deposited.

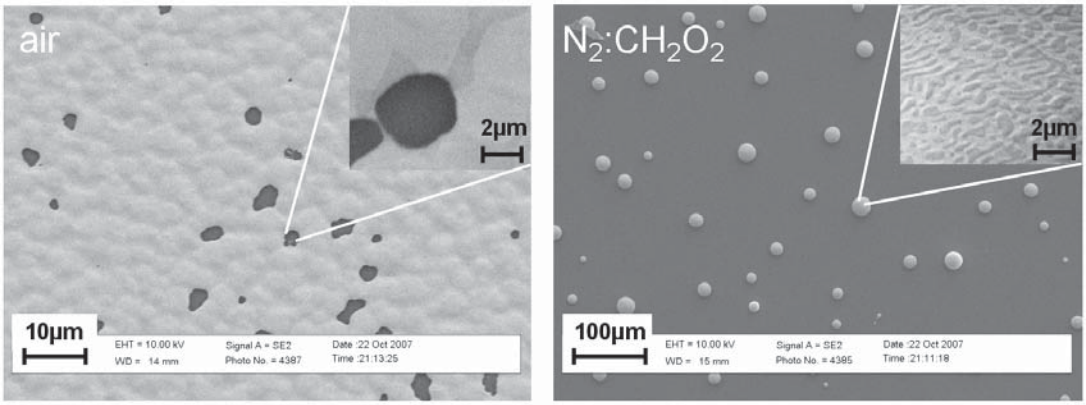


Figure 3.3: AuSn 79/21 film on wet-oxide after reflow on a hot plate at 400 °C for 1 min, exposed to the ambient (left) and under formic acid atmosphere (right).

3.1.2 Formic Acid Assisted Thermo-Compression Bonding

Due to the high content of gold in the eutectic AuSn 80/20 composition, only a small amount of tin oxide would be expected on its surface. However, X-ray photoelectron spectroscopy (XPS) analysis revealed the opposite: namely 57.4 at. % of tin on the surface, compared to the 29.5 at. % of tin in the bulk of the eutectic [107]. A correspondingly high concentration of oxygen was observed. Segregation has been identified to be responsible for the enrichment of tin on a eutectic AuSn 80/20 surface [107]. The driving mechanisms responsible for segregation of one constituent of a composition are a difference in surface free energy between the elements, the heat of mixing of the alloy (as this reflects the bond strength between the elements), and the lattice strain energy. The surface free energy of Sn is 0.6 $\frac{\mathrm{J}}{\mathrm{m}}$ right below the melting point. Since this is smaller than for Au with 1.4 $\frac{\mathrm{J}}{\mathrm{m}}$, a tendency of Sn segregation can be expected. However, the expectation is limited to ζ' - phase , with a long-range ordering at a reduced symmetry, resulting in a super-structure with a large unit cell. For the well ordered δ - phase, no segregation could be observed [108]. Several studies have been presented where it has been attempted to identify the tin oxide composition on the surface and concluded with the presence of SnO and SnO_2 [107].

A simple reflow experiment was performed as a part of this study which revealed that a tin oxide layer on an AuSn 80/20 eutectic film will be detectable due to its influence on dewetting behavior. A layer of 1 µm eutectic AuSn 80/20 was sputter deposited directly onto a wet grown thermal oxide layer present on a silicon substrate. A UBM layer was omitted. After a reflow process on a hot plate at 400 °C for 1 min at atmospheric pressure, the surface roughness was observed to have increased, probably due to recrystallization, and only partial dewetting was observed (Figure 3.3 left). This was taken as an indication of the presence of a tin oxide layer.

In an initial attempt to eliminate the tin oxide formation, a 50 nm gold or platinum layer was sputter deposited as a capping layer for the eutectic solder. After one week of storage at room temperature, no Sn peak was visible in an XPS analysis of the capped samples with Au or Pt. For an uncapped Au-Sn sample, tin oxide peaks were observed right after deposition. Both capped samples were subsequently exposed for 1 min to 200 °C on a hot plate at atmospheric pressure. A tin oxide peak could be identified in case of the Au cap, but still not for the Pt sample. A further exposure of the Pt capped sample to 355 °C for 1 min resulted in a tin oxide peak. This could be expected since the temperature was larger than the liquidus temperature of the eutectic solder. In conclusion, the capping method using Pt could be used to mitigate the solder oxidation process until the reflow temperature is reached. However, the complexity of the now ternary solder alloy is increased [109] due to the additional Pt cap layer. The liquidus temperature right at the bonding interface is correspondingly more difficult to predict, and for this reason the method of Pt capping was abandoned.

The strategy to use reducing agents was investigated as a second attempt to eliminate tin oxide formation. KUHMANN et al. [110] identified tin oxide growth on eutectic AuSn 80/20 to be self-limiting, as on eutectic PbSn, by means of Auger electron spectroscopy (AES). They concluded that a maximal oxide thickness of 3.5 nm was formed after storage at room temperature independent of time within the

range or three days to three month. Furthermore, they listed the standard Gibbs free energy $\triangle G^0$ of the reaction of forming SnO or SnO_2 from pure tin. The values for both reactions are negative and identical within 1 % ($-4.9 \times 10^5 \frac{\mathrm{J}}{\mathrm{mol}}$ to $-4.5 \times 10^5 \frac{\mathrm{J}}{\mathrm{mol}}$ in the temperature range from 150 °C to 350 °C). The values indicate spontaneous oxidation of the tin under standard conditions. Tin oxide decomposition would be expected only at temperatures which are higher than acceptable for reflow. Tin oxide reduction was prominent at standard reflow temperatures [110] at reduced atmospheric pressures (1×10^{-4} Pa) and at the introduction of a partial pressure of hydrogen. Metal based wafer-to-wafer bonding is typically performed in a vacuum chamber where several process gases may easily be introduced. To perform die-to-die bonding in vacuum is normally not considered as economically feasible. The use of forming gas (N_2:H_2) at atmospheric pressure mitigates the growth of additional tin oxide at elevated temperature, but does not reduce the metal oxide below temperatures of 350 °C [111, 112]. Therefore, alternative reducing agents were investigated. At reflow temperatures $\geqq$ 200 °C carboxylic acids have proven to be efficient, even at atmospheric pressure. Only a small amount of energy is needed to release carboxyl pairs from formic acid, which therefore outperforms acetic and acrylic acid with respect to reduction kinetics and uniformity [113]. The gaseous formic acid to metal oxide reaction can be described as follows [114]:

At temperatures between 150 °C and 200 °C the formic acid reacts with the solder oxide to form a compound,

$$MeO + 2HCOOH = Me(COOH)_2 + H_2O. \tag{3.1}$$

At temperatures above 200 °C the compound decomposes into carbon dioxide and hydrogen,

$$Me(COOH)_2 = Me + 2CO_2 + H_2, \tag{3.2}$$

where Me represents the metal.

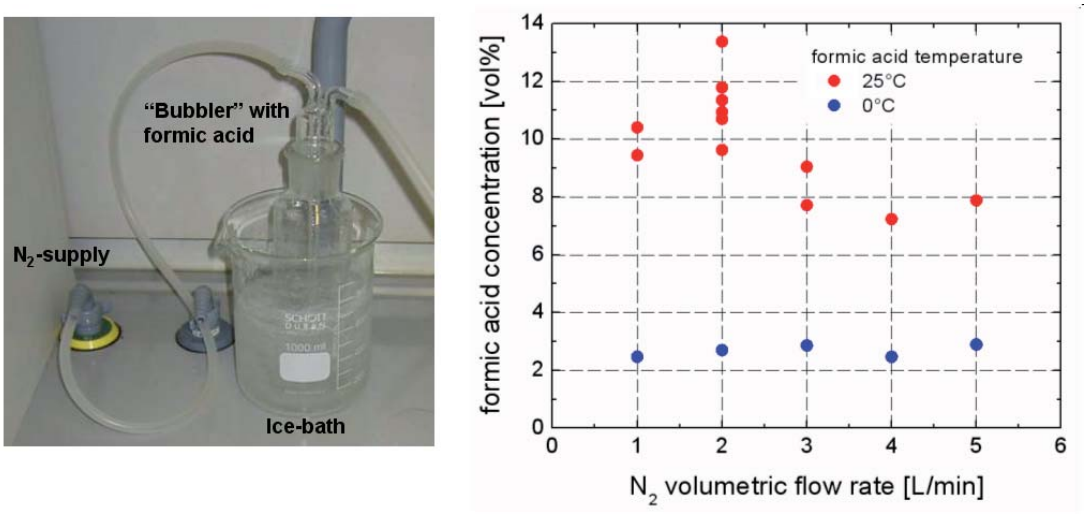


Figure 3.4: Ice-water bath tempered "bubbler", filled with formic acid and connected to the nitrogen source (left). Formic acid concentration vs. nitrogen flow rate at 25 °C and 0 °C.

The wetting behavior of eutectic PbSn solder was investigated on different UBMs in presence of varying formic acid to nitrogen (acting as carrier gas) concentration. The lower limit to sufficiently reduce the surface oxides was determined to be 1.6 vol.% in case of a Ni/Au UBM and 4 vol.% in case of copper as UBM. No carboxylic residue could be observed on the surfaces after reflow up to a concentration of 7 vol.%. The later value was thus set as an upper limit to prevent later corrosion of the sample or bonding chamber [112, 114]. A "bubbler" system was installed to supply a concentration of 2 to 3 vol.% of formic acid to a reflow chamber on a hot plate. Nitrogen was used as an inert gas to transport the formic acid having a boiling point of 100.7 °C (Figure 3.4 left). The concentration depends strongly on the vapor pressure of the formic acid, which can be controlled with temperature, whereas it is almost independent

of the nitrogen flow rate as depicted in Figure 3.4 (right). An ice-water bath was used to achieve the targeted concentration, which was controlled by combining measurements with a nitrogen flow meter and measurements of the formic acid weight change over time. The previous reflow experiment with 1 µm of eutectic AuSn 80/20 sputter deposited directly onto wet grown thermal oxide was repeated in this reductive atmosphere. The initially smooth film dewetted completely at 300 °C and formed Au-Sn spheres (Figure 3.3 right), indicating a close to eutectic metallurgy and a successful surface oxide reduction. The fast dewetting and resulting spherical shapes of the balls is caused by the surface tension and the low viscosity of the eutectic above the melting point, which was measured to be 0.6 N m and 0.9 mPa s, respectively [107].

A flip chip bonding tool (FC6) from KARL SUESS with micron-sized alignment accuracy was used for die-to-die bonding. A differential pumped reduction-chamber was designed to expose the samples during reflow to formic acid (Figure 3.5 left). The main chamber is partially enclosed by the upper chuck. A gap between both parts, complies chip thickness variations, but results in leakage of formic acid, which is removed with a vacuum pump in a concentric groove, surrounding the main chamber. A second concentric groove is introduced and pressurized with nitrogen to a higher level than inside the main chamber, to prevent the release of reducing agent into the flip chip bonding system which is equipped with fragile optics, or further into the operator's area. The differential pumped reduction chamber, and how it works during the reflow process is shown in Figure 3.5 (right).

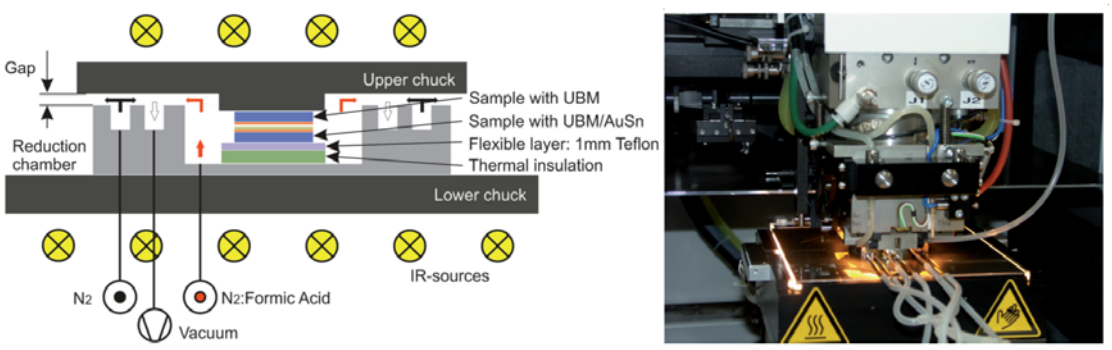


Figure 3.5: Cross-section of the differential pumped reduction chamber. The preferred die-to-die bonding configuration with compliant Teflon layer and bottom thermal insulation is indicated (left). FC6, including the differential pumped reductive chamber, during reflow (right).

3.1.3 Thin Film Solder Bond-Line Formation

A set of samples with varying parameters were prepared to study the bond-line formation of eutectic AuSn 80/20 thin films with relevance for the fluidic thermal test vehicles. Three types of silicon die pairs were fabricated: a mesa structure for pull-testing (Figure 3.10 left), a pin-fin cavity for pressure tests (Figure 3.10 right), and the thermal test vehicle cavity (Figure 3.11). One of the 525 µm thick silicon dies in each pair was patterned by deep reactive ion etching (DRIE) to a depth of 100 µm. A mesa diameter of 4 mm was structured for pull-tests and a cavity size of 9.5 x 9.5 mm^2 with a post area fill ratio of 50 % was implemented for pressure tests. The fluid thermal test die size was 16 x 16 mm^2 and comprised fluid structures, which are surrounded by a rim, which acted as a seal. The pin-fin array with a 21 % area fill, spread across the central area of 10 x 10 mm^2. All dies had a 300 nm thick layer of wet grown thermal oxide below unstructured UBM layers. Three different UBM systems were deposited and named as follows:

- "Platinum": 100 nm Ti / 50 nm Pt
- "Chromium": 100 nm Ti / 200 nm Cr / 50 nm Au
- "Nickel": 100 nm Ti / 200 nm Ni / 50 nm Au

A layer of solder was sputter deposited from a AuSn 79/21 target on top of the UBM layers on either the structured or the unstructured die surface. Thicknesses ranging from 1.5 µm to 4 µm were deposited and the solder layers were left unpatterned.

The flip chip bonding tool was operated in the thermo-compression mode. The die pairs were temporary mounted by applying vacuum on the chucks of the FC6, followed by a wedge error compensation procedure. The dies were brought into contact at a minimal force during a subsequent touch-down sequence. This initialization procedure was run prior to each reflow process. A N_2 : formic acid (2 to 3 vol.%) flow rate of 0.3 $\frac{l}{min}$ was then supplied to the reduction chamber. The heater (infrared lamps) control loop of the lower and upper chuck ramped the chuck temperatures linearly from the initial temperature $T_{initial}$ to the reflow temperature T_{reflow} (Figure 3.6 left). A force F was applied when the first critical temperature T_{c1} was reached. To leave time for proper oxide removal T_{c1} was set 30 K above the metal oxide reduction temperature of the formic acid which is 200 °C. After a short dwell-time at T_{reflow}, which was chosen to be at least 30 K higher than the liquidus temperature of the alloy, the heating system was switched off and the chucks with samples cooled down by natural convection. The force was released, once the second critical temperature was reached, defined 80 K below the liquidus temperature of the eutectic. The formic acid supply was stopped and the upper chuck lifted after vacuum release. In order to get an initial idea about the bonding strength, a simple test was performed by introducing a crack on the back-side of an upper die of the bonded couple with a scriber. The bonded stack was exposed to a bending force which resulted in crack-propagation. The resulting fracture surfaces were inspected by optical microscopy, to identify wetting and failure modes. Further, more sophisticated testing is described in subsection 3.1.5.

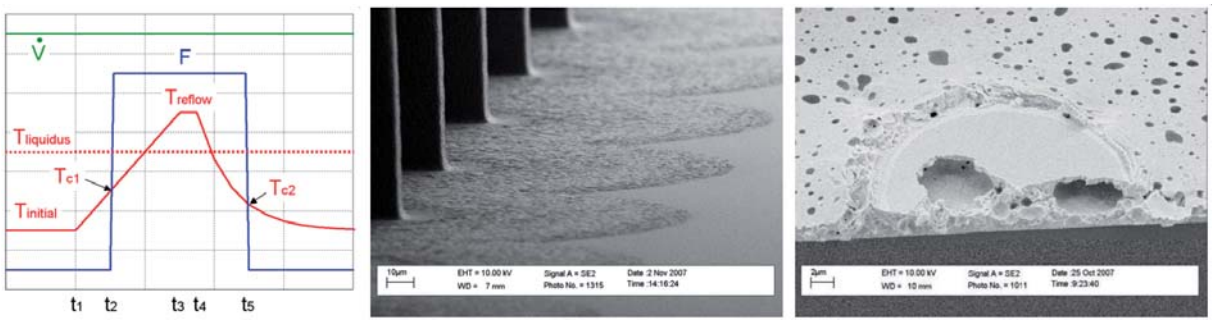


Figure 3.6: Exemplary FC6 thermo-compression process flow (left). Solder wettability on the unstructured UBM (flat surface), with solder deposited on the fin structures only (center). Fractured sample (pin detached) indicates large voids within the bondline. The solder was deposited on the flat die (right).

Initial tests, with platinum wetting layer and a 4 µm thick solder layer on the structured die resulted in poor wetting of the solder and a correspondingly low bond quality. A maximal temperature ramping of 100 $\frac{K}{min}$ up to T_{reflow} of 350 °C supported by both the upper and lower chuck of the FC6 was applied in the trials. Extensive solid-state interdiffusion and tin segregation towards the platinum wetting layer during temperature ramping - resulting in an alloy with a higher liquidus temperature than the targeted reflow temperature - was believed to have prevented the solder from melting at the set reflow temperature (subsection 3.1.4). The lower die was thermally decoupled by a ceramic thermal insulation (4 mm thick MACOR (CORNING) die) from the lower chuck, to improve the heating rate of the dies to be bonded beyond what was the normal limit of the tool. The lower chuck constitutes the largest thermal mass in the system and is therefore the limiting element with respect to the heat ramp (Figure 3.5). In addition, the initial temperature was increased to 150 °C. The upper and lower chucks were operated at their individual maximal temperature ramp, resulting in an asymmetric chuck heating. The time for the dies to reach 350 °C was reduced from 3 min to 30 s, thanks to the high dynamics of the upper chuck. The reduced ramping time resulted in proper wetting and a good bond quality.

The basic fracture test was performed on bonded thermal test cavities. A non-uniform bond quality was observed with a gradual drop in solder wetting from one edge of the sample to the opposite edge. Even after wedge error compensation in the FC6, a residual non-parallelism of up to 1 µm across the 16 mm dies was measured. Only a limited amount of liquid-phase, possibly in the range of 1 to 2 µm thickness is expected for a 4 µm thick solder, which is prone to intermetallic compound (IMC) formation with the UBM. It therefore can not compensate for the wedge error. To get both surfaces of the die pair into intimate contact, a compliant layer (1 mm thick Teflon die) was introduced between the lower

die and the ceramic (Figure 3.5). Accordingly, the bond uniformity was improved and a leak tight seal with high bonding strength was achieved across the whole surface. When solder was deposited on the patterned die, and thus also on the fins in case of a microchannel test structure, a uniform solder wetting radius of 115 µm on the unstructured UBM on the opposing flat die could be observed after reflow (Figure 3.6 center). When the basic fracture test was performed for these samples the crack had propagated through both dies rather than following the bond interface, indicating a quite high bond strength.

Significant voiding could be identified from inspections of fractured surfaces of cracked samples (Figure 3.6 right). Outgassing could also be observed with optical microscopy during reflow experiments on a hot plate. Gas coalescence manifested in bubble growth within the bulk solder and culminated in the collapse of large bubbles at the surface. Four samples with 4 µm of Au-Sn on the chromium-type UBM were exposed to a reflow temperature of 375 °C without formic acid on a hot plate for 2, 7, 15, and 100 s. The morphology of the surface was later inspected by scanning electron microscopy (SEM) (Figure 3.7). Already after 2 s, the surface topology indicated bulk bubble formation. Black spots were attributed to small surface bubbles, covered with a tin oxide membrane. A maximal bubble size of up to 20 µm in diameter was observed after 7 s of reflow exposure time. The tin oxide membrane was now loaded to its yield strength and collapsed, leaving crater like structures and residues of sub 100 nm thick tin oxide membrane on the solder surface (Figure 3.7 15 s). After 15 s at the reflow temperature, surface roughness recovery is observed. Release of argon incorporated during the sputter deposition process (base pressure 1×10^{-7} mbar, process pressure 3×10^{-3} mbar, process gas Ar) was considered as a possible source for the bubble formation. Unfortunately, no spectrometric analysis could be performed to proof this hypothesis. RF bias sputtering could be used to remove the Ar impurities during the film condensation process, with the risk of an altered Au-Sn composition due to the unequal sputter yield of gold and tin atoms.

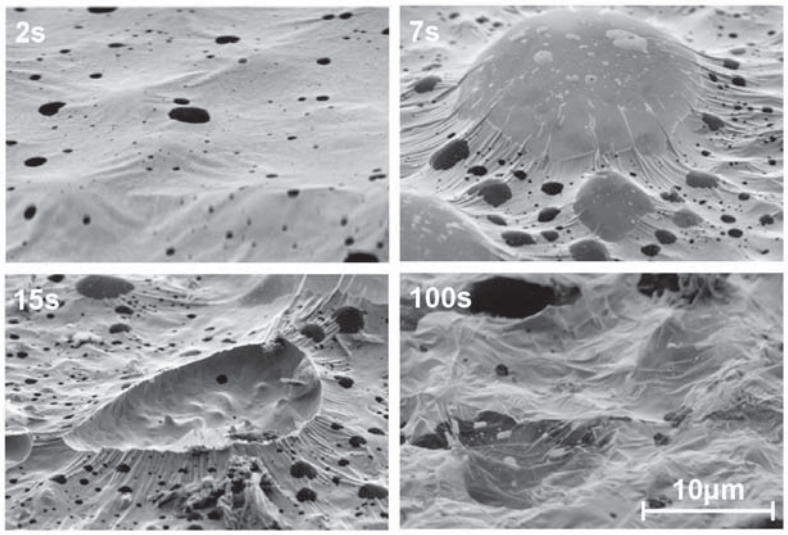


Figure 3.7: SEM inspection of the outgassing kinetics of a 4 µm thick Au-Sn film, sputter deposited on a chromium wetting layer (100 nm Ti / 200 nm Cr / 50 nm Pt). The samples were reflown on a hot plate at 375 °C with an exposure time of 2, 7, 15, and 100 s without reducing agent.

A sensitivity analysis was performed to minimize the gas release with the consequence of void formation in the bond-line. This, to minimize the bond strength and heat conduction degradation of the interface. No "bubble" formation was observed at reflow temperatures $\leqslant$ 340 °C. The surface roughness increased at this temperature and was attributed to recrystallization in the solder itself. Unfortunately, the solder wettability was poor at these temperatures. Further investigations showed enhanced "bubble" formation on the chromium, compared to the nickel or platinum UBM system. This might be explained by the inertness of the chromium. Cr does not react as easily as Ni or Pt with the solder, potentially resulting in a larger liquid solder volume during the reflow sequence (as described in subsection 3.1.4). At a smaller Au-Sn layer thickness of 1.5 µm, the voiding effect was also reduced, probably due

to the correspondingly lower amount of incorporated gas. Finally, a trade-off between bubble formation and solder wettability was performed, by the selection of a reflow temperature of 350 °C.

The influence of the applied force (500 or 150 N) on the bond-line thickness and a potential compression of voids was investigated. The bond-line thickness was independent of the applied force and yields a thickness below 1 µm between the pins, whereas it still was close to the initial Au-Sn thickness in the large sealing area. Solder squeezing seams to be more hindered at the 1 to 3 mm wide bond frames than on the 50 µm wide pins.

The most significant parameters investigated in the sensitivity analysis and their impact on solder quality are listed in Table 3.3.

PARAMETER	MAIN IMPACT	EXPLORED
Au-Sn composition	change in liquidus temperature	79/21
Au-Sn thickness	bond compliance	1.5 to 4 µm
Formic acid	tin oxide reduction	0 to 0.3 $\frac{\mathrm{l}}{\mathrm{min}}$
Heat ramp	wettability of solder	RT to reflow 3 min (100 $\frac{\mathrm{K}}{\mathrm{min}}$)
	(IMC formation)	150 °C to reflow 30 s (400 $\frac{\mathrm{K}}{\mathrm{min}}$)
Reflow temperature	wettability of solder, void formation	300 °C to 380 °C
Force	bond-line thickness	100 N to 500 N
Compliance	uniform area contact	without / 1 mm Teflon part

Table 3.3: Most significant parameters influencing bond-line quality.

The standard reflow process used to produce the thermal test vehicles is:

- Chromium UBM on a 300 nm thick wet grown thermal oxide: SiO_2 / 100 nm Ti / 200 nm Cr / 50 nm Au
- Die sequence: unstructured bottom die, with UBM and 4 µm sputtered Au-Sn from 79/21 target / structured top die with UBM
- Lower bond chuck configuration: reduction-chamber, thermal insulation (4 mm MACOR), compliant material (1 mm Teflon)
- Reducing agent: 0.3 $\frac{\mathrm{l}}{\mathrm{min}}$ of 2 to 3 vol.% formic acid with nitrogen carrier (ice-bad tempering)
- Asymmetric heating: $T_{initial} = 150\,°\mathrm{C}$, $T_{reflow} = 350\,°\mathrm{C}$, $t_3 - t_1 = 30\,\mathrm{s}$, $t_4 - t_3 = 5\,\mathrm{s}$,
- Applied force: $F = 500\,\mathrm{N}$, starting from $T_{c1} = 230\,°\mathrm{C}$, release at $T_{c2} = 200\,°\mathrm{C}$

3.1.4 Thermodynamic Stability of Under Bump Metallizations

Reactions, such as diffusion and dissolution between and of the UBM and the solder can result in local changes of the solder composition, KIRKENDALL voids, and intermetallic compound (IMC) formation. The formation of brittle and large grained IMC reduces the reliability of solder joints and should be prevented accordingly. The influence of the UBM is especially important in case of thin film solders ($\leqq 10$ µm), where the formation of IMC consumes up to 100 % of the solder with a subsequent change in liquidus and solidification temperatures. The typical reaction kinetic stages during a bonding process are:

1. Ramp-up solid-state interdiffusion: Already during heat-up, diffusion between the wetting / barrier layer and the solder layer occurs and can form regions of IMC.
2. Melting: Above the liquidus temperature, the solder starts to melt between phase boundaries ($AuSn$ and Au_5Sn) and becomes viscous. Additionally, some IMC formed during phase 1 will dissolve in the liquid solder.

3. Substantial formation of IMC: Mass transport kinetics in the liquid metal is improved by convection and results in a substantial ternary intermetallic phase with the associated barrier / wetting layer.

4. Solid-state interdiffusion: Additional IMC growth is supported by grain boundary diffusion of the barrier / wetting layer through the solid IMC, resulting in a non-uniform IMC thickness.

5. Solidification of solder: Once, one component of the solder (typically Sn) reacted with the barrier / wetting layer, a local depletion of this component results in a deviation from the eutectic stoichiometry and may initiate a local partial freezing of the solder. The local freezing can also be induced by dissolution of the barrier / wetting layer in the liquid solder and the subsequent increase of the eutectic temperature of the changed composition.

The local freezing phenomena may also be exploited for sequential formation of bond-lines. A solder-hierarchy can be formed even with identical solders. After reflow of a first bond-line, the liquidus temperature of this interface material is increased substantially and does not remelt during reflow of a second bond-line [115]. Au-Sn on a nickel wetting layer was studied by TSAI and HUTTER et al. [116, 105]. They observed the formation of a $(Au, Ni)Sn$ layer at the nickel interface and a residual $(Au, Ni)_5Sn$ phase on top. The first phase has a structure similar to the δ, the second, like the ζ' - phase, but contains small amounts of dissolved nickel of 8.6 and 1.1 at.%, respectively (Ni substitutes Au in the AuSn lattice). Additional metallization schemes considering platinum and chromium were studied by KATZ et al. [106]. In general, it was observed that an increased UBM chemical stability (inertness) mitigates premature freezing during reflow, but results in reduced adhesion due to the lack of intermetallic formation. Platinum was identified as a very reactive UBM with first local freezing sights after 5 s and complete freezing after 30 s at 320 °C. PtSn IMC formation was observed already below the liquidus temperature of Au-Sn. The complete 200 nm thick Pt layer is consumed after 15 s with increased Sn kinetics at reflow temperature and forms additional complex ternary IMC with the Au-Sn [109]. Titanium, which was present under the Pt layer and reacts with gold and might degrade the bond quality further. The high reaction rate of the Pt UBM could explain the poor wetting quality observed in the first reflow experiments with the slow heating rate of 100 $\frac{\mathrm{K}}{\mathrm{min}}$ up to T_{reflow}, documented in subsection 3.1.3. For a nickel UBM the solidification time is five times longer, compared to platinum, which might be a result of the very limited solubility of Ni in Au below temperatures of 350 °C. Highest thermodynamic stability was observed with chromium UBM, but resulted in partial dewetting [106].

For bond-lines with the considered UBM systems (subsection 3.1.3 beginning) a spatial energy dispersive X-ray spectroscopy (EDX) analysis (line-scan along red dashed line across the bond-line, as indicated on the SEM images, Figure 3.8) was performed to identify the IMC formation and solder composition after the standard reflow (subsection 3.1.3 end) and was compared with the SEM cross-sections. The interface with solder deposited onto the UBM is located towards position zero, and the bonded (soldered) interface is located towards the right side of the graph (Figure 3.8 spectras). Charging effects resulted in a distortion of the beam position, therefore the bond-line appear to be larger in the line-scan than as measured in the SEM cross-section. Accordingly, the SEM dimensions should be considered and the Si peaks used as references in the spectra. It should also be noted, that the spatial resolution of the EDX analysis is in the range of 1 to 2 µm at a kinetic energy of 30 keV of the primary electrons. It is equal to the lateral dimension of the electron interaction volume, from were photons are generated .

In case of the platinum UBM, a symmetric phase-separation with a 1 µm thick tin rich δ - phase (AuSn) at the interface and 2 µm of ζ' - phase (Au_5Sn) in the center results from the solidification process (Figure 3.8 left). The accurate identification of the Pt distribution is difficult due to the close proximity and resulting convolution of the Pt with the Au peak. Still, one small Pt peak close to the bond interface is observed, where also a clear Ti peak is visible. Additionally, a 156 nm thick layer interfacing the silicon oxide is visible in the SEM image, supporting this observation (the initial UBM is composed of 100 nm Ti / 50 nm Pt). It seams, that most of the Pt layer on the AuSn 80/20 covered die is dissolved in the δ - phase, resulting in a ternary IMC and an initial Au interaction with the exposed Ti, as described by KATZ et al. [106]. This strong interaction with Pt indicates a high susceptibility to the reflow process dynamics in case of Pt UBM. With a nickel UBM, the accumulation of Sn towards the UBM interface is reduced to a layer of about 300 nm (Figure 3.8 middle). A slightly increased phase-separation towards

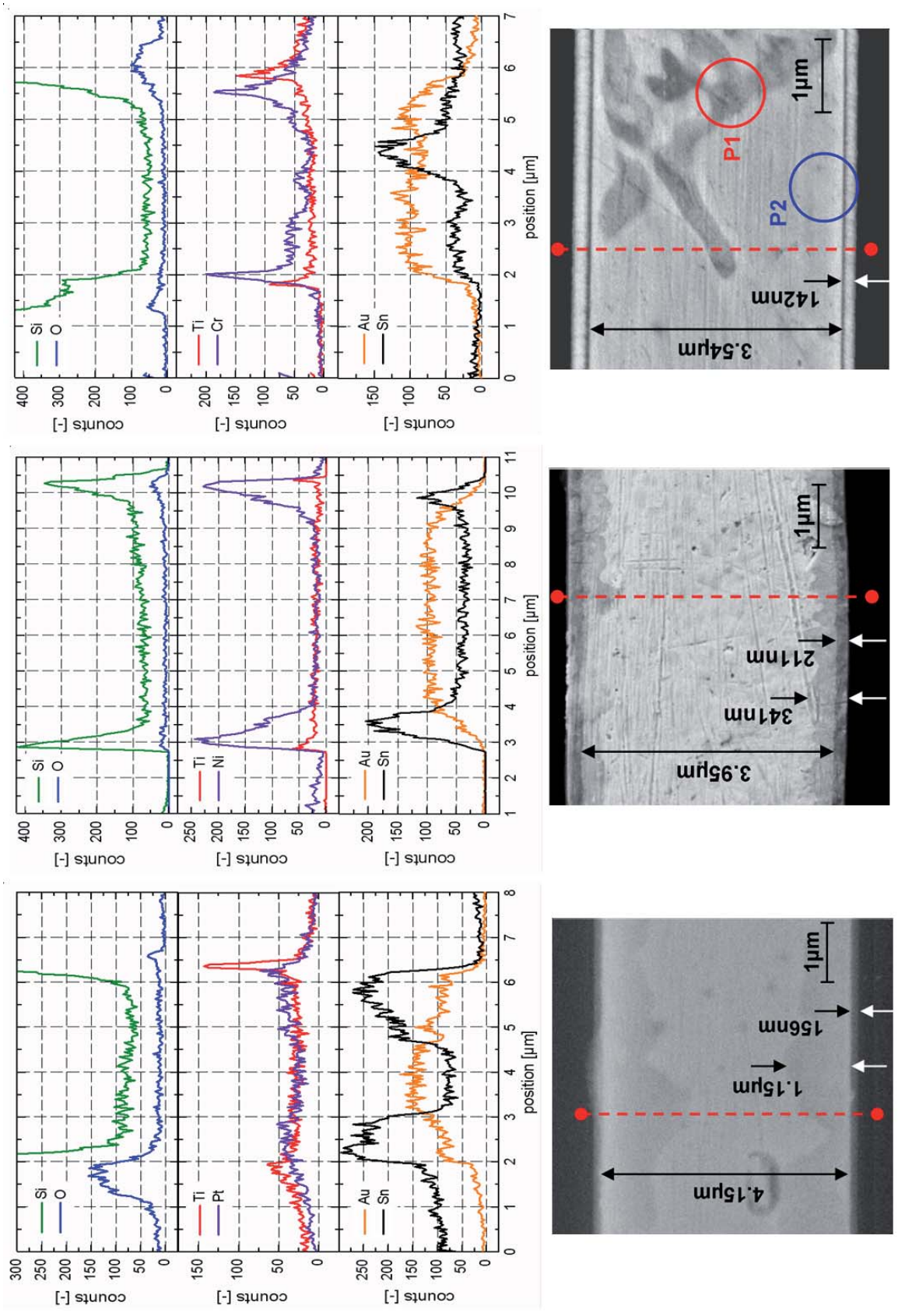


Figure 3.8: EDX line-scans across the Au-Sn bond-lines with Pt-UBM (left), Ni-UBM (middle), and Cr-UBM (right) after reflow (process parameters: subsection 3.1.3 end). The corresponding SEM images indicate the location of the line-scans (red dashed line, with position zero being the lower end), the layer thicknesses and topology. The Au-Sn layer was sputter deposited onto the UBM of the bottom silicon die.

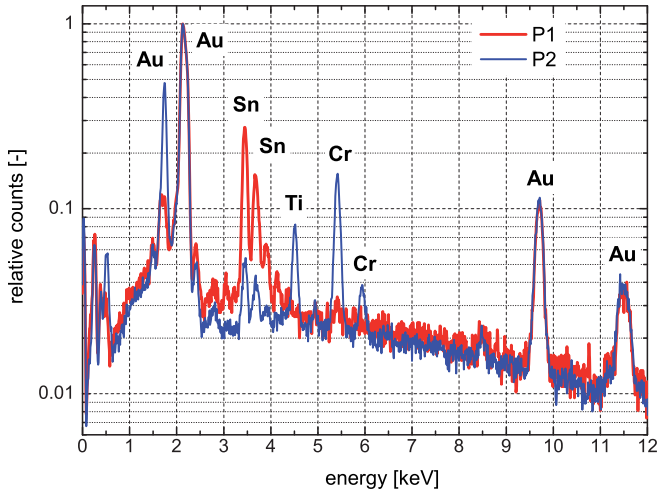


Figure 3.9: EDX spectra at location P1 (δ - phase) and P2 (ζ' - phase close to the UBM) of the Cr-UBM bond-line cross-section, as indicated in Figure 3.8 (right).

the initially Au-Sn covered UBM interface is visible, but still both titanium peaks in the spectra and both UBM layers at a thickness of 211 nm can be identified. The nickel layer seams still to perform as a continuous barrier layer.

The sample with Cr UBM indicates the superior thermodynamic stability and inertness of the chromium wetting / barrier layer (Figure 3.8 right). A uniform δ and ζ' - phase distribution as observed in the bulk eutectic could be identified in the cross-section. It appears that due to the relative dimension of these phases with respect to the bond-line thickness and a possible tendency of local δ - phase accumulation, a spatial non-uniform phase distribution would exist. Accordingly, the local liquidus temperature would vary substantially. The individual EDX spectras of the δ - phase and the ζ' - phase close to the Cr-UBM (indicated by the circles in Figure 3.8 right) are presented in Figure 3.9.

3.1.5 Solder Process Qualification

The structural quality of the bond-interface was analyzed in a self-made pull-tester, using a pneumatic cylinder connected with a spring to the test specimen. The applied load was monitored with a tension load cell KD40S (TRANSMETRA). Patterned silicon samples (10 x 10 mm^2) with a silicon mesa diameter of 4 mm, were used to amplify the tensile stress in the solder interface, compared to the glue interface. Cohesive fracture occurred through the silicon mesa base, at a nominal tensile stress of 19$\pm$6 MPa (Figure 3.10 left and middle). A pressure test, to characterize the use of Au-Sn as bonding and sealing material in microfluidic applications was performed on a 10 x 10 mm^2 die pair. One die is patterned with a sealing rim of 120 µm width, surrounding an area of 9.5 x 9.5 mm^2, filled with an array of 150 µm wide pins, arranged at a 200 µm pitch, resulting in an area fill of 56 % (Figure 3.10 right). The sealing functionality was maintained by the bond-line up to the maximal test pressures of 1×10^6 Pa, without any sign of degradation. The results are satisfactory for the use of the cavities at a nominal applied pressure of 1×10^5 Pa. The sealing rim in case of the fluid thermal test vehicle is much larger and ranges from 2 to 3 mm. The bond uniformity was inspected by scanning acoustic microscopy (SONOSCAN, D-9000) with a 100 MHz transducer (Figure 3.11).

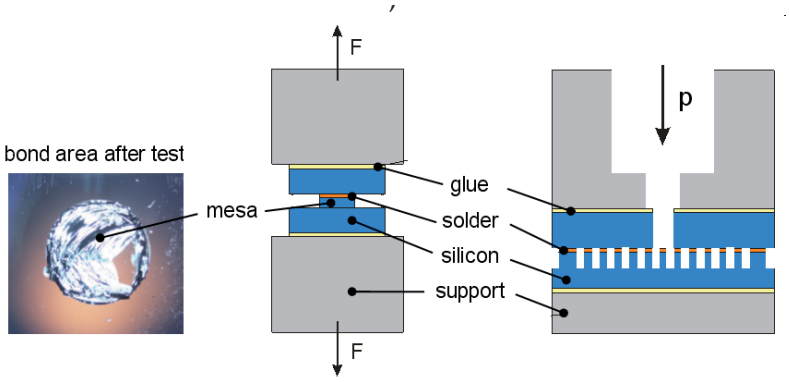


Figure 3.10: Pull test schematic using a stress amplification mesa, to increase the nominal tensile stress in the solder (center). Photograph of a specimen after pull-test, showing fracture through the silicon mesa base (left). Schematic view of the pressure test site (right).

The silicon to the water filled cavity interface 2, with an impedance drop from 20×10^6 to $1.5 \times 10^6 \frac{\mathrm{Ns}}{\mathrm{m}^3}$ results in a strong negative reflection, as indicated with the blue color (Figure 3.11 left, the corresponding C-SAM scan is shown on the upper right). A void free bond-line causes a positive and a negative reflection at the interface 2 and 3 respectively (Figure 3.11 upper right sketch), due to the increased acoustic impedance of the metal compared to silicon. At a bond-line thickness of 4 µm both reflections overlap due to their short time difference and the finite peak width of the initial acoustic signal, resulting in a low amplitude, axisymmetric reflection (Figure 3.11 lower right C-SAM scan). The color scale of the C-SAM map is adjusted to increase the visual sensitivity in this low amplitude regime (Figure 3.11 left). No large voids could be identified in the sealing area (they would appear in blue color). The minor non-uniformity, indicated by a color change from red to yellow, could be caused by solder thickness variations or a low content of small voids below the C-SAM spatial resolution, which is in the range of 100 µm. This also limits the inspection of individual pin-fin bonds with a 100 µm diameter. They are only visible as black dots and can not be investigated in detail. But the uniform signal resulting from the pins across the heat-transfer area could suggest a uniform bond quality (Figure 3.11 inlet, black dots).

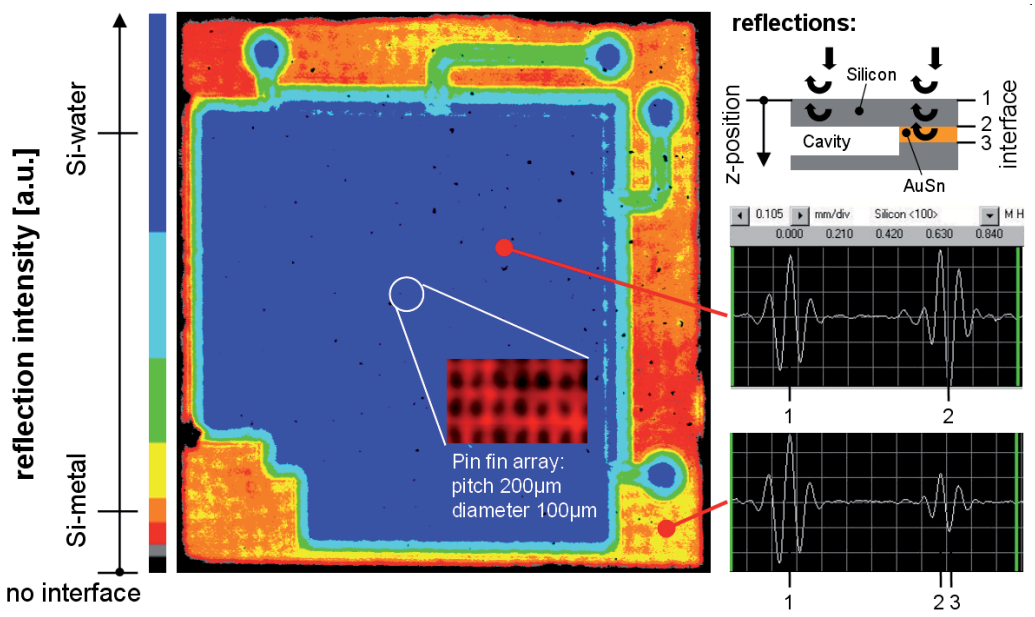


Figure 3.11: C-SAM amplitude map of the second reflection, caused by interface 2 and 3 of a fluid thermal test cavity with the outer dimension of 16 x 16 millimetre. The green lines and blue circles are fluid structures (left). Two local response spectras (with silicon distance as x-scale), one in the cavity area (top) and one at the sealing rim (bottom), show the change in amplitude and shape of the reflection from interface 2 and 3 (reflection 1 acts as reference peak). The individual interfaces are defined in the upper right sketch, showing the cross-section of the bonded die pair.

A more local investigation, to identify voids present in the bond-line, was performed by SEM inspection of polished cross-sections (Figure 3.12 left). An areal void fraction of about 2 % between a pin and the flat surface was identified. Furthermore, minimal "bubble" formation was observed on the solder film surface, with the final reflow sequence (Figure 3.12 right). The wetting quality from the continuous solder layer on the flat die to the microchannel lamellas is also acceptable. The scalloping on the lamella sidewalls, caused by the cyclic nature of the DRIE process, causes a non-uniform solder wetting. Local UBM thickness modulation on the sidewalls might cause this discontinuity. Large surface roughness could have a significant impact on the convective heat transport during the later experiments. Fluid trapped in surface pores, acts as a thermal insulation layer. At the peak-to-peak roughness of 0.6 µm (inspected by the SEM) a thermal resistance penalty of smaller than 1 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ has to be considered accordingly.

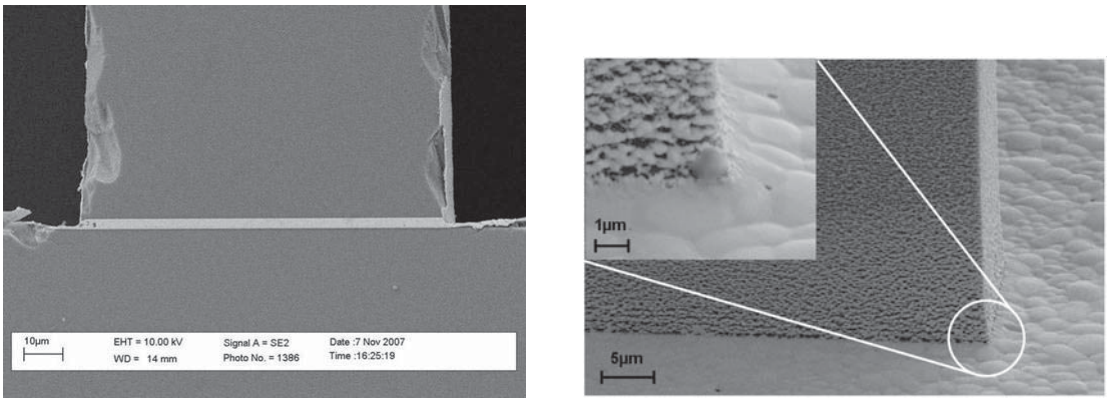


Figure 3.12: SEM inspection of a polished bond-line cross-section, indicating low voiding in the standard process (left). Side-view to a microchannel lamella bonded to the solder covered flat die, revealing low "bubble" formation and adequate meniscus formation (right).

3.2 Water Coolant Compatible Multi-Metal Layer System

The deposition of water coolant compatible multi-metal layers on the test sites is needed to implement the different electrical functionality. Joule heating in resistive metal heaters is considered to mimic power dissipation of transistors. Local resistance thermal detectors (RTDs) are used to derive surface temperatures by Kelvin sensing. Electrical wiring traces are used to connect the active elements on the chip. Finally, electrical interconnection from printed circuit board (PCB) to on-chip pad metallizations is performed with wire-bonds or spring-loaded probes. Patterned dielectric layers are incorporated to electrically insulate between multiple metal layers. This enables the placement of RTDs on top of heaters, compared to the side-by-side design in case of only one metal layer. The later arrangement causes heat flux non-uniformities, resulting in a local temperature drop at the RTD location, by discontinuous heater designs. More complex power maps can be implemented as well considering two metal levels. Accordingly, a square background heater could be patterned on metal level one, covering the complete test chip surface. Several local heaters on the second metal level could than mimic hot-spots. Metallizations compatible with further processing steps during the dielectric deposition and die bonding have to be defined. The dielectric layers should adhere well and have to provide a high enough dielectric strength to support the applied voltage differences of maximal 30 V applied between the layers in the thermal test sites. In multi-cavity experiments the metal layers need also to be sealed from the water, to prevent shorting and water hydrolysis with the associated gas bubble formation, potentially blocking the fluid flow in microchannels (Figure 3.13 left).

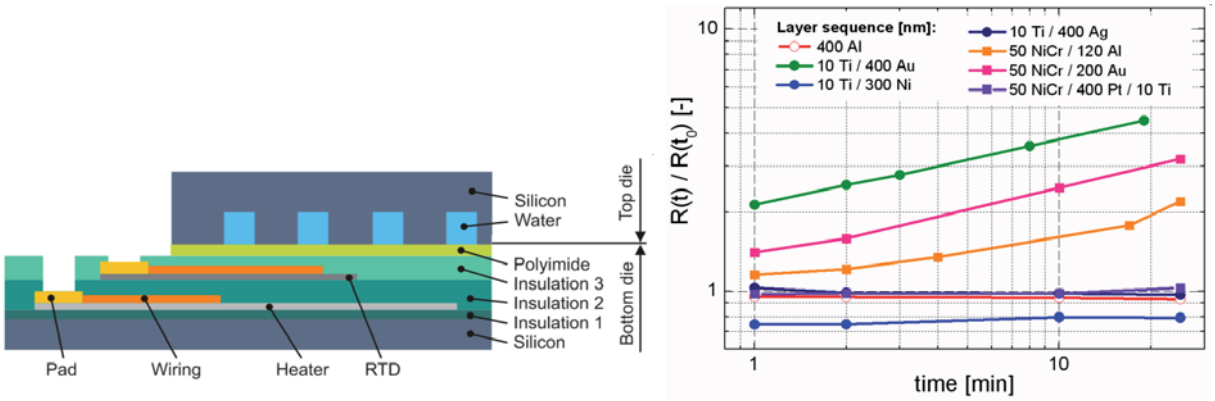


Figure 3.13: Sketch depicting the thin film layer sequence, in case of a double-metal-layer system, with polyimide bonded microchannels (left). Relative resistance change (with respect to the as deposited state $R(t_0)$) of metal layers after thermal excursions to 400 °C on a hot plate at ambient atmosphere (right).

A low temperature coefficient of resistance (TCR) of the heater film material reduces the temperature sensitifity of its power dissipation, resulting in improved heat flux uniformity and therefore low experimental uncertainty. Hence, NiCr 80/20 is especially suited as a heater metal with its low TCR of $\pm 50 \times 10^{-6}\,\frac{1}{\mathrm{K}}$. A high TCR, at a moderate resistivity is needed for RTDs to achieve a high sensitivity. The thermodynamic stability of the metal is also beneficial to minimize the resistive drift over the lifetime of the specimen. Therefore, noble metals such as platinum are widely used. Minimal resistivity and film stress are the main attributes to form low sheet resistance wires. Thick layers of aluminum and gold can be sputter deposited without the risk of delamination due to their ductile nature and low yield stress, with resulting stress relaxation during film growth and further processing. Electrical contacts with spring loaded probes or wire-bonds to the on-chip wiring is performed through dedicated metal pads. Minimal oxide layer thickness on the pads is especially important for reliable probe contacts. The scrubbing action during wire-bonding can handle a certain amount of oxide thickness on a soft metal such as aluminum. High bond quality is achieved using the same pad metallization as the wire material (Al and Au are most prominent), to minimize intermetallic compound formation. The properties of the mentioned metals are listed in Table 3.4. Intermediate thin layers (e.g. 50 nm) of titanium with the ability to form covalent as well as metallic bonds are used to improve the adhesion strength of metal films on dielectrics. The deposition and patterning of all the metals on the thermal test vehicles is performed by magnetron sputtering using lift-off resist technique.

METAL	RESISTIVITY [μΩ cm]	TCR [$10^{-6}\,\frac{1}{K}$]	PURPOSE
NiCr 80/20	108	±50	heater
Pt	11	3600	RTD
Al	2.7	4300	wire / pad
Au	2.2	4000	wire / pad

Table 3.4: Resistivity and temperature coefficient of resistance (TCR) and purpose of typical metals used on thermal test vehicles [99].

The thermal impedance of the dielectric layer (insulation 2 in Figure 3.13 left) needs to be less than $0.1\,\frac{K\,mm^2}{W}$ to keep the temperature difference between the heater and sensor layer on an acceptable level. The dielectric strength and the thermal conductivity of individual materials is listed in Table 3.5. The thermal resistance is within the specification for inorganic materials such as SiO_2, Si_3N_4, and Al_2O_3 with respect to a 30 V potential difference. These thermal resistance values are lower bounds considering perfect film quality. Stoichiometric variations, local defects, embedded particles, voiding, and water up-take can detoriate the brake-down voltage of the film dramatically. Thick inorganic layers are especially prone to delamination and crack formation due to high intrinsic stress and their brittle nature. Compared, the dielectric strength of organic materials is altered especially due to a high water up-take if exposed to water or humidity.

DIELECTRIC	DIELECTRIC STRENGTH [$\frac{kV}{mm}$]	THERMAL CONDUCTIVITY [$\frac{W}{mK}$]	MINIMAL LAYER THICKNESS [nm]	THERMAL RESISTANCE [$\frac{K\,mm^2}{W}$]
SiO_2	196	1.5	153	0.102
Si_3N_4	196	35	153	0.004
Al_2O_3	33	30	909	0.03
Polyimide	150 to 300	0.2	200 to 100	1.00 to 0.50

Table 3.5: Expected thermal resistance caused by the listed dielectric layers at a nominal voltage difference of 30 V between the metal layers [99].

Wet oxidation of silicon in a water vapor atmosphere at temperatures above 1000 °C results in a high quality SiO_2 layer. This dielectric was used on the silicon substrate as insulation 1 (Figure 3.13 left). Silicon-nitride is the preferred material between the metal layers (insulation 2). Plasma-enhanced chemical vapor deposition (PECVD) in an inductively coupled plasma (ICP) system at 300 °C with silane, diborane, and amonia precursers was used to deposit the silicon-nitride layers. Minimal film stress was achieved at a surplus of silane forming non-stoichiometric silicon-rich silicon-nitride to mitigate delamination and stress-induced defects.

Layers of 250 nm, 500 nm, and 1000 nm of silicon-nitride were grown on a continuous 200 nm thick Al coating to identify the minimal silicon-nitride layer thickness to achieve a void free dielectric layer. The samples were then exposed for 5 min to a liquid aluminum etchant (solution of phosphoric and nitric acid). Only the specimen with a 250 nm thick silicon-nitride layer resulted in etch pits. The experiment was repeated with lift-off patterned metallizations of 10 nm Ti / 200 nm Al layers. All specimens showed aluminum etch pits, independent of their silicon-nitride thickness. The pits were observed predominantly at the edge of the patterned metal and close to defects (e.g. particles). The limited conformality of the PECVD process with seam formation in the encapsulant at locations with reentrant geometries could be the cause of the etchant access to the metal. Hence, a highly conformal 100 nm thick Al_2O_3 layer was added by atomic layer deposition (ALD) from trimethylaluminium and water precursors at 250 °C onto the 1000 nm thick silicon-nitride layer, to seal the seams. No aluminum pits could be observed after the subsequent wet etching step, indicating a liquid tight encapsulation.

Two-metal-level specimens were fabricated to characterize the electrical quality of the dielectric film sequence. Both patterned metal levels consisted of 50 nm NiCr and 100 nm Al. The insulating dielectric was composed of 1000 nm PECVD silicon-rich SixNy and 100 nm ALD Al_2O_3 (Table 3.6 S1). A potential difference of 60V could be applied between the metal layers without electrical shorts and break-down after film deposition. Four of these specimen were exposed to 200, 250, 300, and 380 °C on a hot plate for 5 min each to mimic subsequent bonding processes. The electrical integrity was maintained up to 250 °C. Electrical shorts were observed for the samples exposed to 300 and 380 °C. Possibly, aluminum is diffusing through defects in the dielectric created by thermo-mechanical stress at elevated temperatures. The high melting point metal platinum is deposited onto or replacing the aluminum film, to prevent electrical shorts. To improve the adhesion to the subsequent silicon-nitride layer a 10 nm titanium film is deposited onto the platinum (Table 3.6 S2, S3). Both specimens did not show any electrical shorts up the maximal test temperature of 380 °C. However, a significant increase in resistivity was observed in case of sample S2.

#	Layer Sequence [nm]	Thermal Exposure up to 380 °C	Source
S1	Si / 100 SiO_2 / 50 NiCr / 100 Al / 1000 SixNy / ... ... 100 Al_2O_3 / 50 NiCr / 100 Al	- shorts $\geqq$ 300 °C	- crack - Al diffusion
S2	Si / 100 SiO_2 / 50 NiCr / 100 Al / 50 Pt / 10 Ti / ... ... 1000 SixNy / 100 Al_2O_3 / 50 NiCr / 100 Al	- no shorts - increased sheet resistance	- intermetallic
S3	Si / 100 SiO_2 / 50 NiCr / 300 Pt / 10 Ti / ... ... 1000 SixNy / 100 Al_2O_3 / 50 NiCr / 100 Al	- no shorts - invariant sheet resistance	

Table 3.6: Layer sequence of the two metal level samples and their electrical characteristic upon thermal exposure.

A series of unpatterned metal layers were sputter deposited on SiO_2 to study their relative resistance change upon exposure to 400 °C on a hot plate at ambient conditions. The sheet resistance was measured by a four-point-probe system (Jadel Engineering). A significant resistance increase was observed for the Ti / Au, NiCr / Al, and NiCr / Au system and could be caused by the formation of intermetallic compound between the metal layers. The Ti / Ni system resulted in a significant resistance drop of up to 20 % after one minute of thermal exposure. A resistance change below 5 % was observed for the Ti / Ag and NiCr / Pt / Ti system. Therefore the NiCr / Pt / Ti sequence with NiCr as heater and Pt as sensor, wiring, and pad metal was chosen as the preferred metal layer system for the two-metal-levels.

A 200 nm thick Al_2O_3 ALD layer was used as insulation 3 to seal the upper metallization from the water of the adjacent fluid cavity. A water drop was dispensed onto the alumina layer and heaters and senors were activated. At the nominal voltages (up to 30 V) no gas formation could be overserved in the water drop. Strong bubble formation due to hydrolysis was observed in the water without the insulating layer 3.

3.3 Single-Cavity, Double-Side Heated Test Vehicle

A single-cavity, double-side heated test vehicle represents one level of an interlayer-cooled chip stack. Its surfaces can be inspected by spatially resolved infrared imaging and is therefore preferentially used for detailed heat transfer analysis. A uniformly heated two-port cavity populated with channels or pin fins was designed to study the fundamental behavior of individual heat transfer geometries. Non-uniform power dissipation and heat transfer cavities, representing one quadrant of a four-port architecture, were designed, to validate the multi-scale modeleling concept. Both test sites were assembled in the multi-sample test-section to interface with the fluid loop and the electronics.

3.3.1 Test Vehicle with Uniform Power Dissipation and Heat Transfer Geometry

The double-side uniformly heated test vehicles with a uniform population of microchannels or pin-fins were used to experimentally characterize the most efficient heat-transfer geometry compatible with TSVs [117, 53]. They are composed of two thin film bonded silicon dies (subsection 3.1.3) with the outer dimension of 16 x 16 mm^2 (Figure 3.14 left). The top and bottom die are 325 µm and 525 µm in thickness, respectively. The cavity is etched into the bottom die with a height of 97 ± 6 or 189 ± 15 µm. The fluid is delivered through the bottom dies rectangular ports, with 1 mm wide and 10 mm long in- and outlets, for minimal fluid mal-distribution to the heat-transfer area of 1 cm^2. Discrete pressure drops along the fluid path can be resolved with four pressure ports implemented at positions 0, 1, 5, and 10 mm in downstream direction of the flow (Figure 3.14 right). Two identical rectangular resistive heaters (300 nm NiCr 80/20) were designed on the back-side of the lower and the upper silicon chip. The heaters are splitt to integrate a resistive temperature detector (300 nm Au) spanning the centerline transversal to the flow direction (Figure 3.15 left).

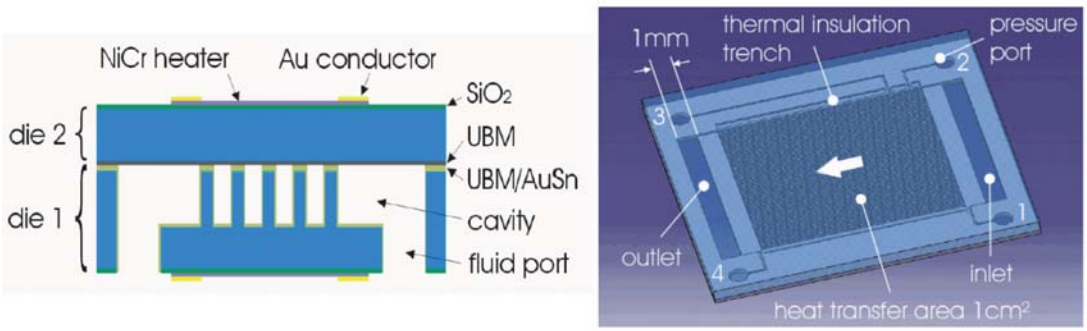


Figure 3.14: Cross-section through a test vehicle (TV) presenting silicon structure and individual thin-film layers. (UBM: under-bump metallization) (left). CAD model of the TV with four pressure ports and the rectangular fluid in-/outlets. The top chip is semi-transparent for illustrative purpose (right).

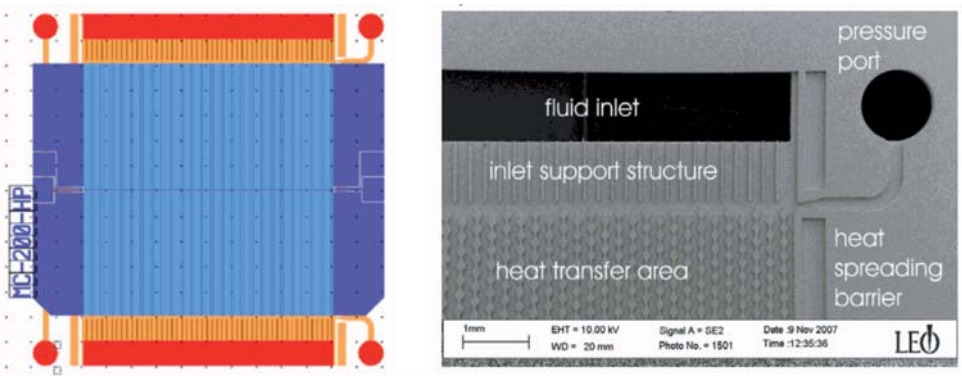


Figure 3.15: Layout of the fluid structure and the corresponding locations of the heater and temperature sensor (red: fluid connection, orange: fluid channels, light blue: resistive heater, dark blue: RTD and contact pad metallization) (left). SEM micrograph top view of the fluid cavity filled with perl chain geometries and the inlet and a pressure port (right).

3.3.2 Test Vehicle with Non-Uniform Power Dissipation and Heat Transfer Geometry

The double-side non-uniform heated test vehicles with a non-uniform population of microchannels or pin-fins were used to experimentally validate the multi-scale model approach proposed in chapter 2 [87]. They are composed of two thin film bonded (subsection 3.1.3) silicon dies with the outer dimension of 16 x 16 mm^2, equal to the uniform test vehicles (Figure 3.16 left).

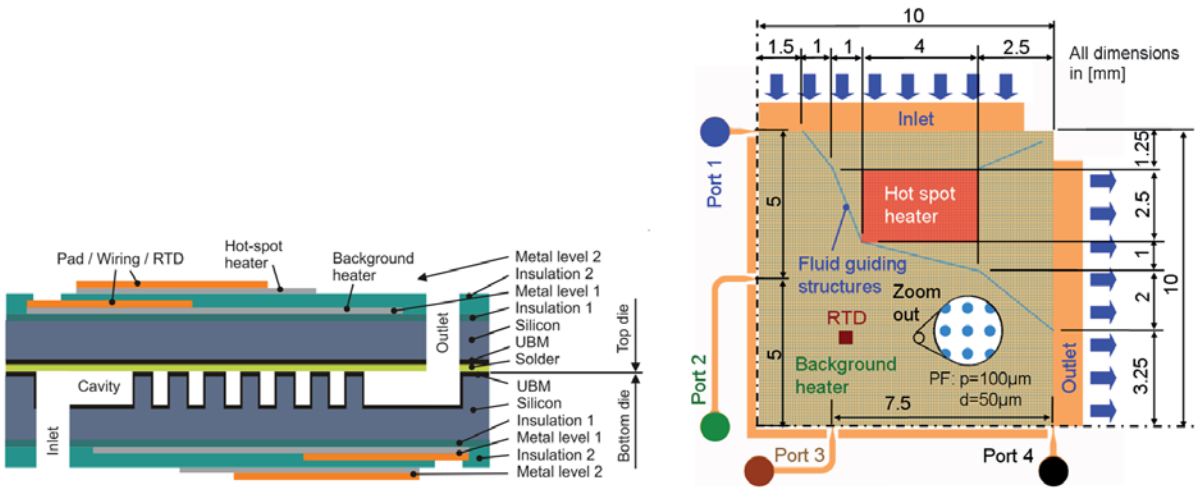


Figure 3.16: Cross-section through a non-uniform test vehicle (TV) along a flow line from inlet to outlet presenting silicon structure and individual thin-film layers (left). Layout of the non-uniform heated and populated four-port structure. The background heater (green) dissipates heat uniform across the heat transfer area. Non-uniform power dissipation is achieved with the additional rectangular 0.1 cm^2 sized hot-spot heater (red) located close to the fluid inlet. A RTD (brown) is also implemented. The inlet and outlet of the cavity are located at adjacent cavity edges and are 1 mm shorter than the cavity width. The fluid cavity is filled with microchannels or pin-fin arrays. Fluid guiding structures (blue) focus the fluid towards the hot-spot. Pressure ports along the cavity center lines allow pressure drop measurements (right).

However, the in- and outlet of the fluid cavity are located next to each other. They represent together with the 1 cm^2 cavity one quadrant of a four-port architecture with 4 cm^2 footprint (Figure 3.16 right). The top die is 325 µm and the bottom die with the 96 ± 3 µm deep fluid cavity is 525 µm thick. The fluid is delivered and drained through rectangular, 1 mm wide in- and outlets which are separately etched into the bottom and top die. This allows for the implementation of squared background heaters on metal level 1 of both dies (Figure 3.17 left).

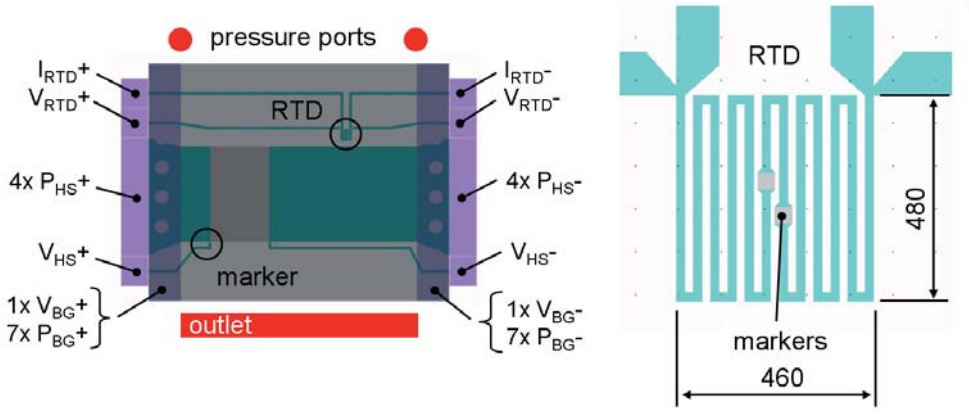


Figure 3.17: Metal layers and fluid connection of the top die. The background heater (light gray) with the eight pads for power delivery and voltage drop measurement (dark gray) are shown. Green layers represent the platinum RTD and the wiring to the hot-spot heater, which is connected through the pink pads of metal level two (left). The detailed geometry of the RTD reveals the implementation of local heaters, used as thermal alignment markers (right).

Again, four pressure ports are implemented at positions 0, 5, 12.5, and 20 mm in downstream direction from the inlet (Figure 3.16 right). Two metal levels are deposited on the back-side of the bottom and the top silicon die. A square, uniform background heater (50 nm NiCr 80/20) with 300 nm thick platinum pads and wiring is sputter deposited onto a 100 nm thick, wet grown SiO_2 layer and represents metal level 1. Additionaly, a 10 nm thick titanium layer is deposited onto the platinum, to improve the adhesion of the subsequent dielectric layer. Insulation 2 is composed of 1 µm SixNy and 100 nm Al_2O_3 as described in section 3.2. Metal level 2 (100 nm NiCr 80/20, 400 nm Pt) includes a rectangular hot-spot heater covering 10 % of the fluid cavity, one RTD used to calibrate the infrared camera, and four thermal alignment markers (Figure 3.17). Two markers are implemented in the RTD meander and two are placed at the lower corners of the hot-spot heater and electrically connected through the hot-spot voltage drop probes. The sheet resistance and the according heat flux dissipated from the markers (only NiCr layer) compared to the connecting wires (NiCr and Pt layer) is 39 times higher. The markers are electrically operated with a hundred times larger current, compared to the later sense current and result in local hot-spots, which can be used as geometrically reference for the infrared camera initial to the experiment. The RTD, background, and hot-spot heater are arranged congruent on the top and bottom die.

3.4 Multi-Cavity Test Vehicle: Pyramid-Chip-Stack

A multi-cavity thermal test vehicle is designed to demonstrate interlayer-cooling performance in a chip stack as described in [118]. Multiple hot-spot heaters on each tier can be addressed individually to study heat spreading and thermal cross-talk between different tiers and cavities. Individual hot-spot temperatures can be measured and can be used to validate the multi-scale modeling method at uniform and hot-spot dominant power maps.

All thermal demonstrator chip stacks include three power dissipating tiers and four heat removing fluid cavities (Figure 3.19). To reduce the process complexity, a pyramid-chip-stack configuration was realized with lateral electrical I/Os utilizing wire-bonds instead of TSVs (Figure 3.18 left). This also allows the integration of the fluid in- and outlets into the chip stack. Power can be dissipated independently in four hot spot heaters per tier on an area of 10 mm^2 each (2 x 5 mm^2 2-port / 3.33 x 3.33 mm^2 4-port). This results in a $\leqq$ 40% heat transfer area coverage. The heaters are distributed equidistant with a spacing of 0.42 mm for the 2-port and 0.92 mm for the 4-port, respectively. A meander design is used to meet a resistance specification of 30 Ω. The heater wire is divided into five parallel strips to reduce current crowding in the meander bends, resulting in a high heat flux uniformity. The hot-spot temperature (THS) is recorded with a four-point measurement of the resistive temperature device (RTD) located along the heater symmetry line (Figure 3.18 right).

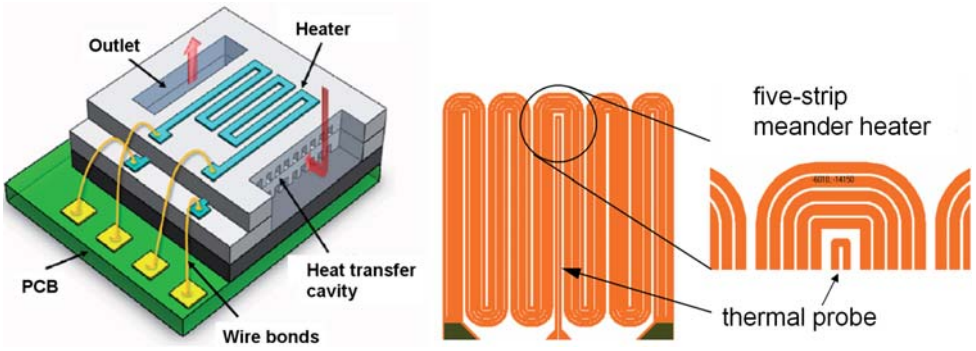


Figure 3.18: Sketch of the interlayer-cooled thermal demonstrator with the pyramid-chip-stack design and resulting lateral I/O (left). Five-strip hot spot heater design with integrated resistive thermal probe. Resistor metallization (orange), electrical leads (brown) (right).

The test vehicle fabrication sequence started with wafer level metal deposition onto 525 µm thick, 4″ diameter silicon substrates covered with 200 nm SiO_2 wet-oxide dielectrics. Aluminum (Al) strips with a thickness of 250 nm for the heaters and sensors followed by an additional 400 nm of Al acting as electrical leads and wire-bond pads are sputter deposited and patterned with lift-off technique. Atomic layer deposition (ALD) was used to cover the metal layers with a pinhole-free, 200 nm thick Alumina (Al_2O_3) insulation layer, to prevent hydrolysis in the water. The dielectric on bond pads was removed by buffered hydrofluoric acid using positive photo-resist masking. A 4 µm thick polyimide layer (HD3003, DuPont) was then spin-coated and structured in a oxygen plasma reactor with a positive photoresist mask. Cavities and ports were fabricated into the silicon die by double-side deep reactive ion etching. After a first electrical inspection the known-good-dies were singulated by wafer dicing. The alignment of the five silicon dies representing the chip stack was done with a brass stencil. This complete assembly was placed into a membrane oven. The polyimide bond was performed at 350 °C in 1 mbar vacuum under an applied load of 7 bar on the stack top surface through the oven membrane (Figure 3.19, 3.20a, 3.20b). Alignment accuracy was better than 10 µm, which is sufficient for the demonstrator. A leak test with water at 2 bar over pressure assured the bond line quality. The stack was then glued to the printed circuit board using a mechanically compliant silicon adhesive (Sylgard 577, Dow Corning) to minimize thermo-mechanical stress. Wedge – wedge wire bonding with 25 µm thick Al-wires was performed to support a maximal current load of 0.1 A (Figure 3.20c). To protect the wires, a UV curable epoxy (Norland 65, Optical Adhesives) was used as globe top. Finally, a PMMA manifold with fluid connections was attached to the stack with a underfill epoxy (EpoTek 302-3M, Epoxy Technologies)

at a defined gap of 60 µm forming the capillary (Figure 3.20d). The test vehicle silicon slab thickness is 425 µm instead of 50 µm compared to realistic chip stacks, to reduce wafer handling complexity. This will enhance the heat spreading capability in each layer. A realistic slab-thickness results from a maximal TSV height of typically 150 µm minus the cavity depth. Furthermore, we used a compliant polyimide layer for leak-tight bonding. This layer represents a thermal impedance of 20 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ and emulates the wiring levels of a real processor die with a typical thermal resistance of 7 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$. The offset of 13 $\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ needs to be considered in further discussion of the thermal performance.

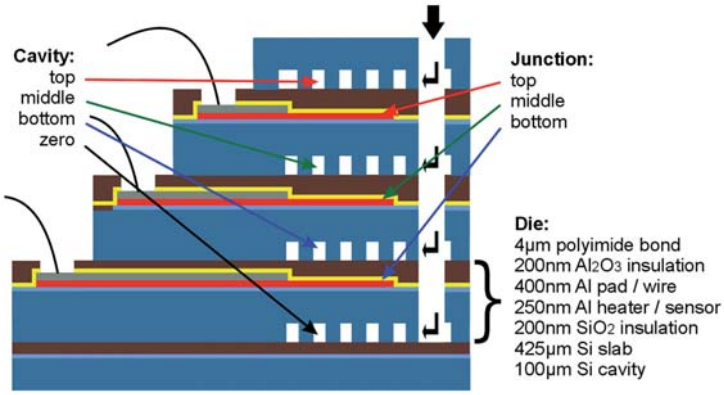


Figure 3.19: Pyramid-chip-stack cross-section showing layer and stacking sequence. No TSVs were implemented, to reduce complexity. Wire-bonds are used to power and read heaters and sensors instead. The bonding of the tiers is performed through a polyimide layer.

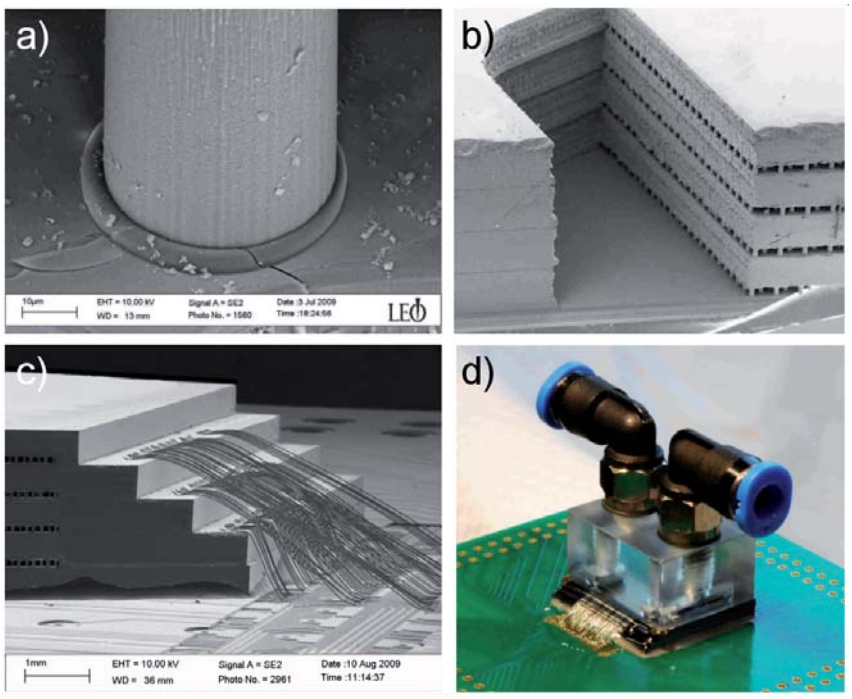


Figure 3.20: Scanning electron microscope and photographic close-ups of the pyramid-chip-stack: a) pin-fin bond showing polyimide meniscus, b) cross-section through fluid port and cavities, c) view at stack to board wire-bond I/Os, d) complete test vehicle mounted on the printed circuit board with fluid manifold and connection.

3.5 Single-Phase Test Loop with Spatially Resolved Infrared Imaging

The thermo-fluidic characterization of the test vehicles was performed on a single-phase fluid-loop with deionized water as coolant in the primary loop (Figure 3.21 left). The fluid temperature at the inlet is controlled (20 or $25 \pm 1\,^{\circ}\mathrm{C}$) through the secondary chiller loop (ProLine RP855, LAUDA) which is thermally coupled through a plate heat exchanger (LYTRON) to minimize the risk of contamination. The primary loop is operated with a magnetically coupled gear pump (FLUIDOTECH) and a 7 µm particle filter. Laminar pressure-gradient flowsensors FLR-1605A and FLR-1617A (OMEGA) with an accuracy of $<40\,\frac{\mathrm{ml}}{\mathrm{min}}$ and $<2\,\frac{\mathrm{ml}}{\mathrm{min}}$ were used for the uniform and non-uniform signle-cavity experiments. A Coriolis mass flow meter (MFS 3000-S03, KROHNE) with an accuracy of $<3\,\frac{\mathrm{ml}}{\mathrm{min}}$ was used for the pyramid-chip-stack instead. Differential pressure sensor (PD23-V-2, OMEGA, accuracy 2 mbar) were used to measure pressure drops between in- and outlet and between ports $1-4$, $2-4$ and $3-4$. The ambient, fluid in- and outlet temperatures were measured with T-type thermocouples. The junction temperature on the test vehicles was red from the deposited RTD sensors. The heaters are powered by multi-purpose DC power supplies. The dissipated power and the RTD resistance (Kelvin probes) is measured with a KEITHLEY 2701 multimeter and a KEITHLEY 7700 multiplexer card. The data acquisition and control was performed through a LabView platform.

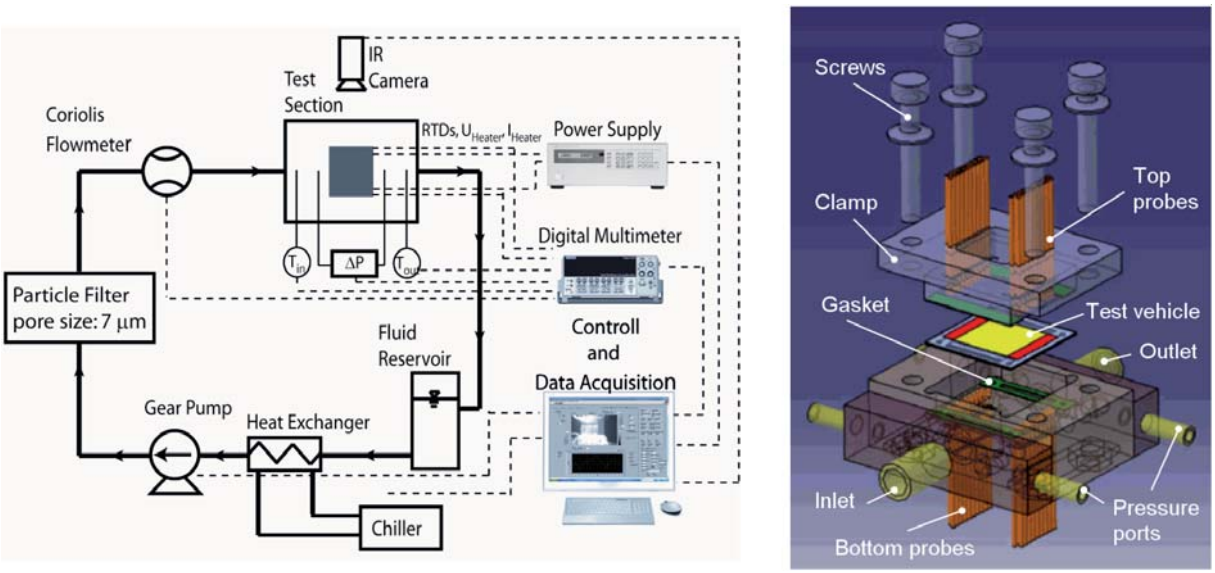


Figure 3.21: Schematic of single-phase test stand with data acquisition (left). CAD explosion drawing showing the multi-sample test section with probes and gaskets for the uniform heat transfer TVs. The test section for the non-uniform heat transfer TVs is build equivalent, though with the inlet from the lower support and the outlet on the upper clamp (right).

A multi-sample test section was developed for the single cavity test vehicles to easily interface them to the fluid loop and the electrical equipment (Figure 3.21 right). The test vehicle is clamped against customized gaskets connecting its fluid and pressure ports to the liquid loop and the pressure sensors. Fifteen spring-loaded electrical probes per heater side deliver a maximal current of 7.5 A and limit the total power dissipation per device to 400 W. Four additional probes are used to read out voltage and resistor values. The multi-sample test stand allows full optical access to the heat-transfer zone on both test vehicle sides. Accordingly, more detailed junction temperature information with a spatial resolution of 0.3 mm was obtained using an IR camera (Silver 420, CEDIP) with a noise-equivalent temperature difference (NETD) of 25 mK. Black electrical tape was attached to the test vehicle heater surface to achieve an emissivity of 0.95.

The TV holder was built with a low-thermal-conductivity glass ceramic (MACOR). This resulted in a junction-to-ambient parasitic thermal resistance of $16\,\frac{\mathrm{K}}{\mathrm{W}}$ at zero flow rate, and corresponds to a $<1.8\,\%$ underestimation of the thermal resistance. Sensible heat loss from fluid to ambient was $<0.7\,\%$. Other uncertainties are listed in Table 3.7.

PARAMETER	ERROR
Flow rate	$\pm 10\ \frac{\mathrm{ml}}{\mathrm{min}}$
Pressure drop measured	± 2 mbar
Pressure drop scaled: 4 cm^2	± 6 mbar
Fluid temperature	± 0.1 K
Junction temperature RTD	± 0.2 K
Junction temperature IR	± 0.3 K
Thermal resistance	± 1.4 %
Power dissipated to fluid	± 0.9 %
Heat flux at $\triangle T_{jmax-in} = 30$ K	± 1.4 %
Heat flux scaled: 1 cm^2, $\triangle T_{jmax-in} = 60$ K	± 1.4 %
Heat flux scaled: 4 cm^2, $\triangle T_{jmax-in} = 60$ K	± 5.0 %

Table 3.7: Experimental uncertainties derived from Gaussian law of error [119].

4 Experimental Results and Validation of Modeling Framework

This chapter reports on the experimental findings derived from the test vehicles described in chapter 3. Additionally, the multi-scale modeling concept proposed in chapter 2 is validated with respect to these results. The first set of tests with uniform heat transfer structures and power dissipation assesses the mass transfer and heat removal efficiency of different unit cell geometries (section 4.1). In section 4.2 building blocks such as fluid focusing, heat transfer structure modulation, and fluid delivery architectures are benchmarked against each other. Predictions from numerical modeling using the extended tensor-description of the periodic porous media are discussed with respect to their accuracy. Finally, chip stack temperatures including thermal cross-talk between individual active and heat removing layers are modeled and experimentally validated with the pyramid-chip-stack.

4.1 Uniform Single Cavity Experiment: Unit-Cell Shape Efficiency

The design of efficient heat transfer structures is a trade-off between bulk (fluid temperature increase) and convective thermal resistance [75, 53, 117]. The first parameter benefits from low friction factor of the cavity, whereas the second parameter improves with increasing fluid mixing (reduced boundary layer thickness). Typically, low friction factors are achieved with geometries resulting in low fluid mixing and vice versa. Hence, the optimal geometry is not trivial and balances both parameters with respect to the boundary condition defined by the system design. Pumping power is relevant if the cold plate pressure drop dominates in the fluid loop and would be prominent in desktop computers with only one chip stack to cool. In server racks, with more than hundred chip stacks, additional significant pressure drops are expected in fluid loop elements such as filters and quick-connects at a limited pressure head of the centrifugal pumps. Accordingly, the maximal pressure drop is a more stringent criteria. In this section, unit-cell geometries are experimentally assessed with respect to their heat removal efficiency in the double-side heated, single cavity test vehicles described in section 4.1.

STRUCTURE	ABBREVIATION	PITCH [µm]
Parallel plate	PP	
Microchannel	MC	200, 100, 50
Pin-fin in-line	PFI	200, 100, 50
Pin-fin staggered	PFS	200, 100, 50
Pin-fin staggered, distorted cell	PFS-dc	200
Pearl chain	PC	200
Drop-shaped pin-fin staggered	DPF	200

Table 4.1: Test vehicles assessed in the uniform heat transfer study.

Heat-transfer structures with uniform x and y pitch were implemented at full (fp) and half (hp) population. The pitch-to-pin-diameter or the pitch-to-fin-width ratio were 2:1. A parallel plate served as benchmark (Table 4.1, Figure 4.1). The PFS design with distorted cell (dc) was deducted from an in-line configuration. Every second column was shifted by half a pitch in flow direction. The pearl chain (PC) was built on the same pin locations and diameters, but the pins were connected with 20 µm wide channel walls. The drop-shaped pins were arranged as a 200 µm pitch PFS and completed with a 45° tail. Attributes of individual unit-cell geometries at 200 µm pitch are shown in Figure 4.2 (left). Further

characteristics are given in Figure 4.2 (right) for PFI as representative structure at 50, 100, and 200 µm pitch and at half population (hp).

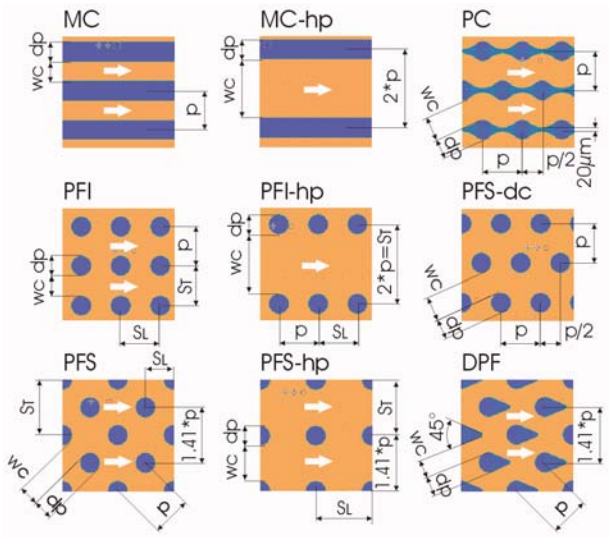


Figure 4.1: Unit-cell geometries compared in the experimental study.

4.1.1 Experimental Sequence and Extrapolation of Datasets

The aim of this work is to define the heat-removal capability of unit-cells for chip stacks with a footprint of 1 to 4 cm^2. Temperatures and pressure drops at a cavity size of 1 cm^2 are measured in the experiment. Hence, a linear extrapolation scheme was used to predict the performance for larger chip stacks. The test sequence started with pressure drop readings at a fluid linlet temperature T_{in} of 25 °C without any power dissipation. Later, the RTD sensors and the readings from the infrared camera were calibrated with respect to changing fluid inlet temperatures. In the subsequent thermal characterization, a control loop adjusted the input power to obtain $\triangle T_{jmax-in}$= 30 K. The temperature readings were corrected with respect to the reference silicon thickness of 100 µm, a $\triangle T_{jmax-in}$ of 60 K, and the optional linear extrapolation for large chip stacks, to compare the individual unit-cells at chip stack relevant conditions (Figure 4.3).

1. Correct temperature gradient due to silicon slab thickness difference between test setup ($t_{experiment}$ = 325 µm) and reference thickness of the 3D stack (t_{ref} = 100 µm).

2. Extrapolate readings in case of chip-size scaling from s_0 to s_1:

   - The heat flux and average fluid velocity are kept constant during extrapolation.
   - Flow rate:
   $$\dot{V}_1 = \dot{V}_0 \frac{s_1}{s_0} \tag{4.1}$$
   - Pressure drop:
   $$p_1 = (s_1 - s_0)\frac{dp}{ds} \mid_{s=s_0} + p_0 \tag{4.2}$$
   - Maximal junction temperature:
   $$\triangle T_{(jmax\,extr-in)0} = \triangle T_{(jmax-in)0} + \frac{\dot{Q}_0{}^{(s_1-s_0)}/_{s_0}}{\dot{V}_0 \cdot \rho \cdot c_p} \tag{4.3}$$

3. Calculate $\dot{q}_1$ by linear heat flux scaling for $\triangle T_{jmax-in}$ of 60 K:

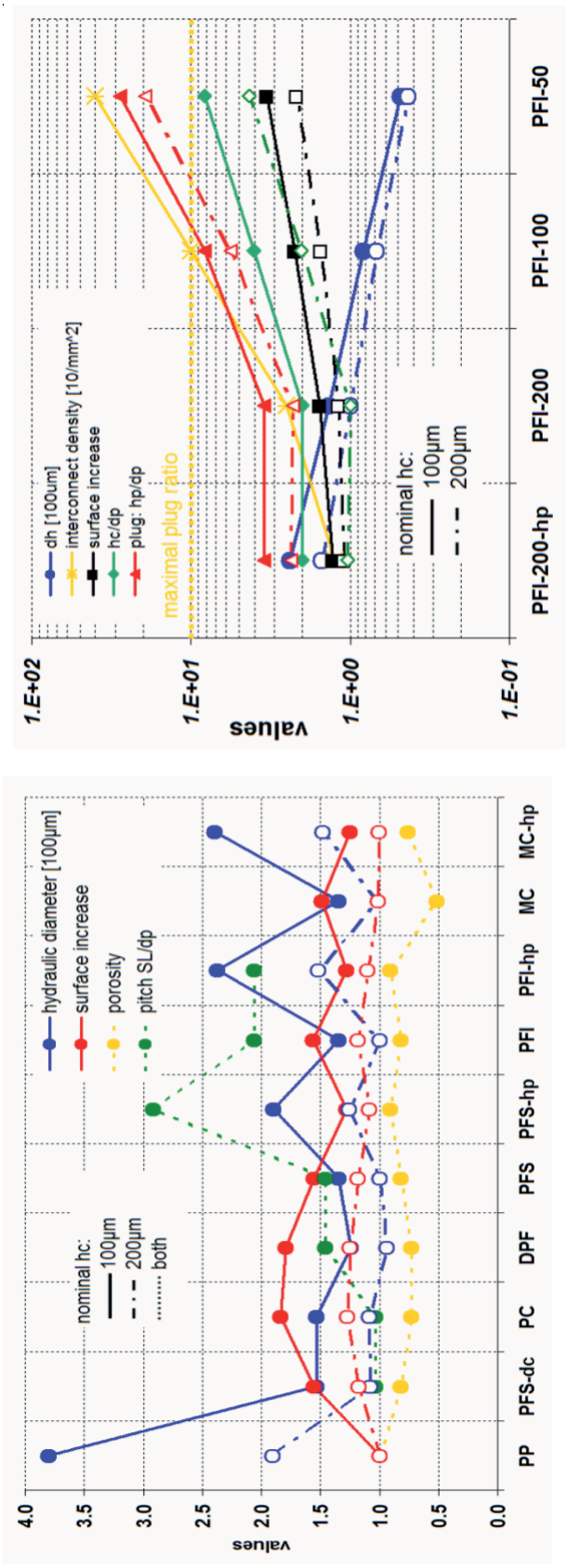


Figure 4.2: Comparison of relevant TV parameters at 200 µm pitch, such as the hydraulic diameter d_h, the relative wetted surface area $A_{wetted}/(2 \cdot A_{unit-cell})$, the porosity, and the ratio of pitch to pin diameter in flow direction (left). Characteristics of PFI structures at interconnect pitches of 200, 100, and 50 µm: such as d_h, the interconnect density, the increase of wetted surface area, the ratio of channel height to pin diameter, and the corresponding aspect ratio of the electroplated copper plug.

Pressure-drop and thermal-resistance readings were performed at different $\triangle T_{jmax-in}$ values to validate the method of linear heat-flux scaling in the range of interest. Uncertainties of the derived values are listed in Table 3.7.

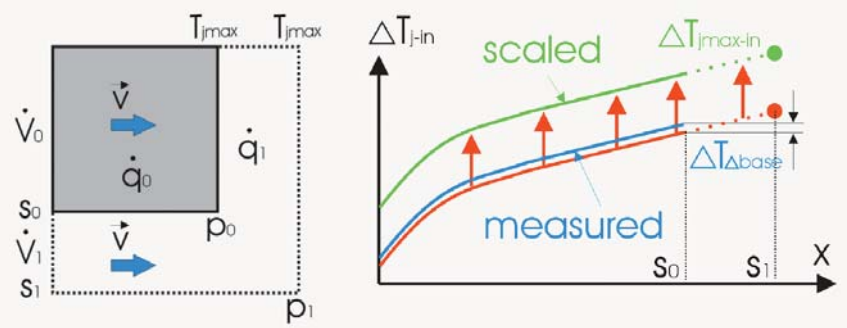


Figure 4.3: Chip-size, volumetric flow rate, pressuer drop and heat flux scaling from a die size of s_0 to s_1 at constant average velocity and T_{jmax} (left). Junction temperature correction due to base thickness difference (blue to red), optional: $\triangle T_{jmax-in}$ extrapolation (dotted red line and bullet), power scaling for reference $\triangle T_{jmax-in} = 60\,\mathrm{K}$ (right).

4.1.2 Hydrodynamic Characteristics of Unit-Cells

Local pressure-drop readings were validated with correlations derived from MUZYCHKA et al. [120] for developing thermal and hydrodynamic boundary layers in rectangular microchannels (h_c = 200 µm, w_c = 100 µm, l_c = 10 mm) with deionized water at 25 °C as described in [117, 53]. Experiment and prediction are in good agreement (Figure 4.6 left). The rate of pressure drop to volumetric flow for a 1 cm^2 chip stack is measured for all TVs at a pitch of 200 µm and channel heights of nominally 100 and 200 µm (Figure 4.4 left). Clearly, minimal pressure drop for a given flow rate is achieved by the parallel plate with the largest hydraulic diameter, without obstructions, but without surface enlargement. The other TVs can be categorized into three groups: As expected, PFS and DPF cause the highest pressure drops and nonlinear behavior because of fluid acceleration due to the inherent flow redirecting. The DPF has slightly lower pressure needs, even with its smaller hydraulic diameter and larger wetted surface area, but reduced flow separation, as discussed in KOSAR et al. [70]. The PFI, PFS-dc and PC outperformed the MC, and are the TSV-compatible fluid structures having the lowest pressure needs at $Re < 300$. In all three devices, the flow has to pass the minimal cross-section area only at one location per unit cell and can expand slightly afterward. Similarly to the MC, the flow in the PFI can pass in straight manner, whereas in PFS-dc and PC it gets redirected at a somewhat larger d_h. At the Re_{crit} of 300, a flow-regime change occurs, and the pressure gradient increases abruptly in PFI and PFS-dc. The PC and MC do not exhibit this threshold point at Re 300, most probably due to the smooth geometrical transition from pin to fin and the straight sidewalls. For TVs with reduced channel heights of 100 µm, the overall characteristics is equivalent, but no flow-regime change was detected in the pressure range tested. This could be attributed to damping effects caused by the end wall (bottom and top wall), as reported in PELES et al. [121]. TVs of PFI with different pitches and at half population for cavity heights of 100 and 200 µm (Figure 4.4 right) were also analyzed. At constant channel height, Re_{crit} was found to depend on the TV pitch. Again, endwall damping, which is less pronounced at small pitch to pin height values, could be the reason. The fanning friction factor for pin-fin structures is defined as follows (PRASHER et al. [94]):

$$f = \frac{\triangle P}{2 \cdot N_L \cdot \rho \cdot v_{max}^2}, \tag{4.4}$$

for a given pressure drop $\triangle P$, with the fluid density ρ and the number of pin rows in flow direction N_L. v_{max} is the maximal average velocity of the fluid found between the pins:

$$v_{max} = \frac{\dot{V}}{A_{min}} = \frac{\dot{V}}{h_c(b - d_p \cdot N_T)}, \tag{4.5}$$

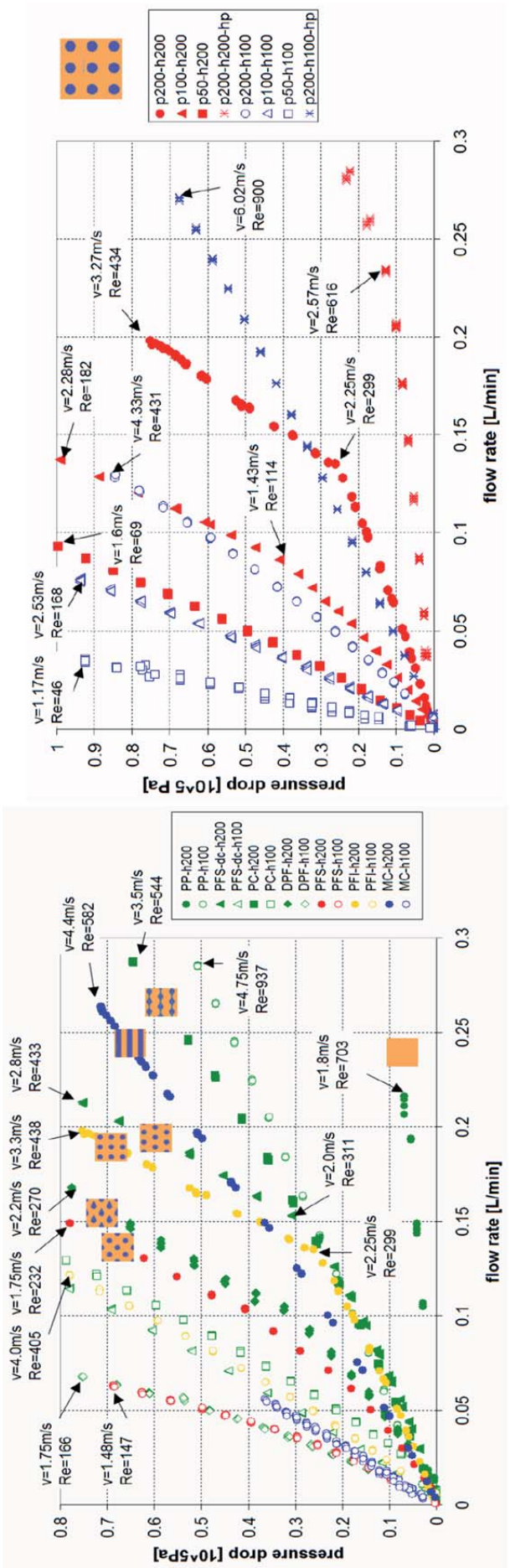


Figure 4.4: Comparison of fully populated TVs with 200 µm pitch and 100 and 200 µm channel height operated with $T_{in} = 25\,^{\circ}\mathrm{C}$ deionized water (left). Comparison of fully and half populated PFI TVs with channel heights of 100 and 200 µm and pitches of 50, 100 and 200 µm, operated with 25 °C deionized water (right).

with the cavity height h_c, heat-transfer area width b, pin diameter d_p and number of pins in a column N_T. In case of tube bundles with a height-to-diameter ratio > 10, Reynolds numbers are defined using the pin diameter. For short pin structures used in this study, the influence of the end wall becomes important, and the hydraulic diameter d_h of the minimal cross-section area between the pins is used as critical dimension for the Reynolds number:

$$Re = \frac{d_h \cdot v_{max}}{v} \; using \; d_h = \frac{2 h_c w_c}{h_c + w_c}, \tag{4.6}$$

at a fluid kinematic viscosity ν.

Experimental data of pin-fin in-line at values $< Re_{crit}$ was correlated assuming the following form:

$$f = c \left(\frac{h_c}{d_p}\right)^{\alpha_1} \left(\frac{S_L - d_p}{d_p}\right)^{\alpha_2} \left(\frac{S_T - d_p}{d_p}\right)^{\alpha_3} Re^{-m}, \tag{4.7}$$

for the friction factor f proposed by PRASHER et al. ([94]). All relevant parameters are included, such as the cavity height h_c, and the ratio of longitudinal S_L and transversal pitch S_T to pin diameter (Tables 4.2, 4.3). All data was found to be within a confidence level of 20 % (Figure 4.5 left).

The correlations for staggered pin-fin structures were done at a fixed S_T/S_L ratio of two. Therefore S_T and S_L are represented by the pin structure pitch p, and the equation has the following form:

$$f = c \left(\frac{h_c}{d_p}\right)^{\alpha_1} \left(\frac{p - d_p}{d_p}\right)^{\alpha_2} Re^{-m}. \tag{4.8}$$

The confidence level of the data at Re <100 is 25 % (Tables 4.2, 4.3). At Re > 100 the slope seems to change as indicated by PRASHER et al. [94] (Figure 4.5 right).

COEFFICIENT	c	α_1	α_2	α_3	m
PIN-FIN IN-LINE $Re < Re_{crit}$					
- Friction factor	13.044	-0.437	0.716	-0.199	0.912
- NUSSELT number	0.4862	-0.118	0.288	0.219	0.417
PIN-FIN STAGGERED					
- Friction factor at $Re < 100$	11.827	0.368	-4.102		0.774
- NUSSELT number$100 < Re < 1'000$	0.1012	0.444			0.838

Table 4.2: Values of coefficients for correlations developed for low-aspect-ratio pin-fin in-line and staggered arrangements.

RATIO	$\frac{h_c}{d_p}$	$\frac{S_L - d_p}{d_p}$	$\frac{S_T - d_p}{d_p}$	$\frac{p - d_p}{d_p}$
PIN-FIN IN-LINE $Re < Re_{crit}$				
- Friction factor	1 to 8	1 to 1.8	1 to 3	
- NUSSELT number	1 to 8	1 to 1.8	1 to 3	
PIN-FIN STAGGERED				
- Friction factor at $Re < 100$	1 to 8			1 to 1.5
- NUSSELT number$100 < Re < 1'000$	1 to 2			

Table 4.3: Validation range of correlations defined by the maximal and minimal geometrical ratios of the test vehicles used for data fitting.

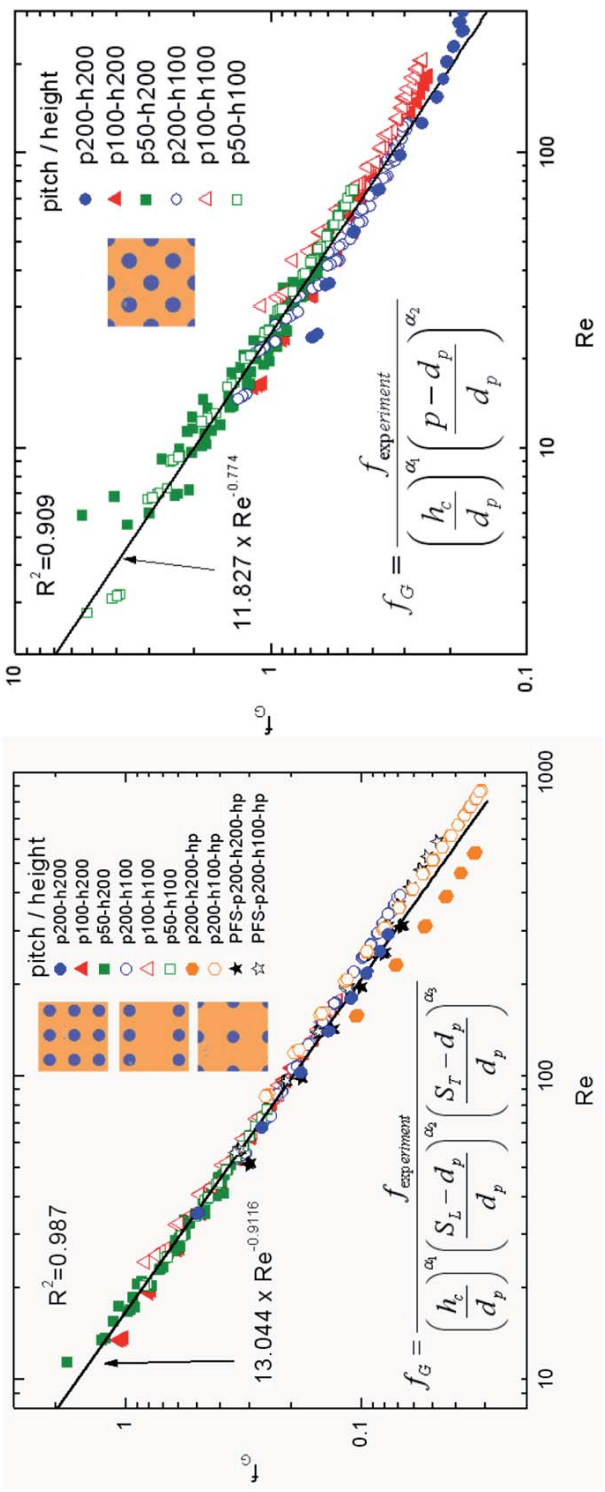


Figure 4.5: Comparison of friction factors obtained from experiment and correlation (Equation 4.7) developed by fitting the data of pin-fin in-line structures at low Re numbers, before the onset of regime transition ($Re < Re_{crit}$) (left). Comparison of the friction factors obtained from experiment and the correlation (Equation 4.8) developed by fitting the data of pin-fin staggered structures at REYNOLDS numbers <100 (right).

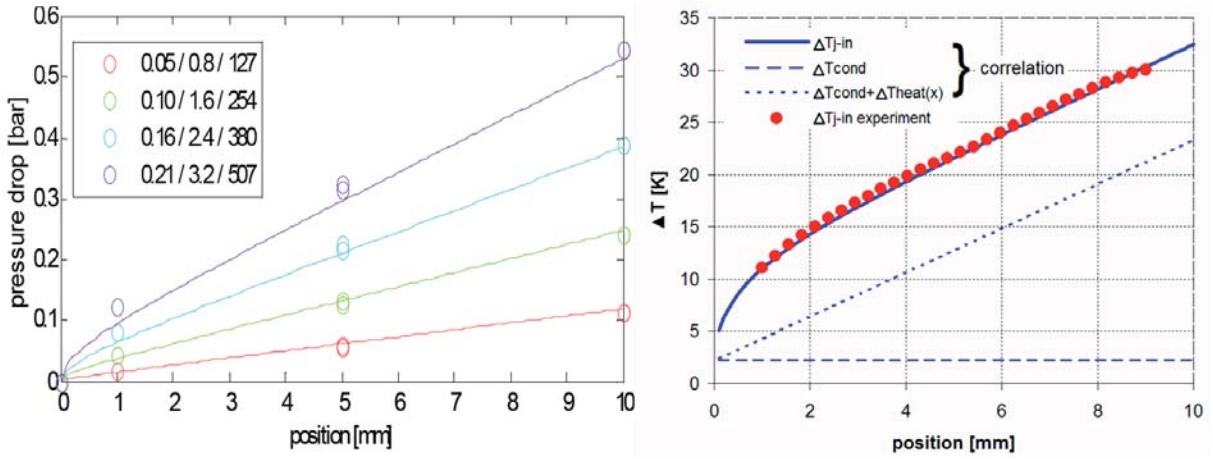


Figure 4.6: Validation of measured microchannel pressure-drop data (bullets) with correlations derived by MUZYCHKA (lines). Legend (inset): flow rate $[\frac{l}{min}]$ / mean velocity $[\frac{m}{s}]$ / REYNOLDS number (left). Validation of $\triangle T_{j-in}$ reading at combined entry region using correlation from MUZYCHKA et al. [120] for device MC-p100-h200 at 113 $\frac{ml}{min}$ and 166 W (right).

4.1.3 Heat Transfer Characteristics of Unit-Cells

To validate the thermal reading, the measured junction temperature was compared with predictions derived from the correlation reported in MUZYCHKA et al. [120] for the combined entry region of rectangular ducts and isoflux boundary conditions. The deviation was less than 5 % (Figure 4.6 right).

Figure 4.7 (left) compares the normalized thermal resistance, defined in subsection 1.3.2, Equation 1.7, at different volumetric flow rates of individual devices. Figure 4.7 (right) shows the percentage of the thermal gradient due to conduction and sensible heat absorption of the fluid. Note that the conduction part of all TVs is < 7 %. The PP without surface enhancement performs worst and is strongly dominated by the convective temperature gradient. At low flow rates of < 0.1 $\frac{l}{min}$, the PFS performs best at a pitch and structure height of 200 µm because of its fluid-mixing characteristic, closely followed by the DPF with reduced fluid separation. Third is the PFI, followed by PFS-dc and PC, which both perform eually as the MC at low flow rates. Devices with a low thermal resistance translates into a low $\triangle T_{conv}$ compared to $\triangle T_{cond} + \triangle T_{heat}$ at a constant flow rate. At a critical $\dot{V}$, the performance of TVs, such as PFS-dc, PFI, PC, PFI-hp, and PFS-hp, improves and reaches a value close to that of the PFS after the transition. Upon reduction of h_c from 200 to 100 µm, the size of the wetted surface decreases, but the performance at a given pitch is increased owing to the direct proportionality of R_{conv} to d_h as shown for PFI and MC. A minimal R_{conv} is measured for the PFI at a pitch of 50 µm and h_c of 200 µm. R_{conv} contributes less than 10 % to the total thermal gradient. The sum of the convective and conductive impedance is derived by deviding the temperature difference between the central junction temperature $T_{jcenter}$ and the average fluid temperature by the total power dissipated $\dot{Q}_{tot}$ by both heaters, multiplied by the heater area :

$$R_{conv} + R_{cond} = \frac{T_{jcenter} - (T_{out}+T_{in})/2}{\dot{Q}_{tot}/A_{heater}}. \tag{4.9}$$

The conductive impedance in case of double side heating and 1D heat flux with a silicon slab thickness t on both sides can be calculated as a network of two parallel resistors:

$$R_{cond} = \frac{t}{2k_{Si}}. \tag{4.10}$$

From the residual R_{conv} the heat-transfer coefficient h and Nu can be derived:

$$h = \frac{A_{heater}}{R_{conv} \cdot A_t} \; and \; Nu = \frac{hd_h}{k_f}, \tag{4.11}$$

with the wetted area being

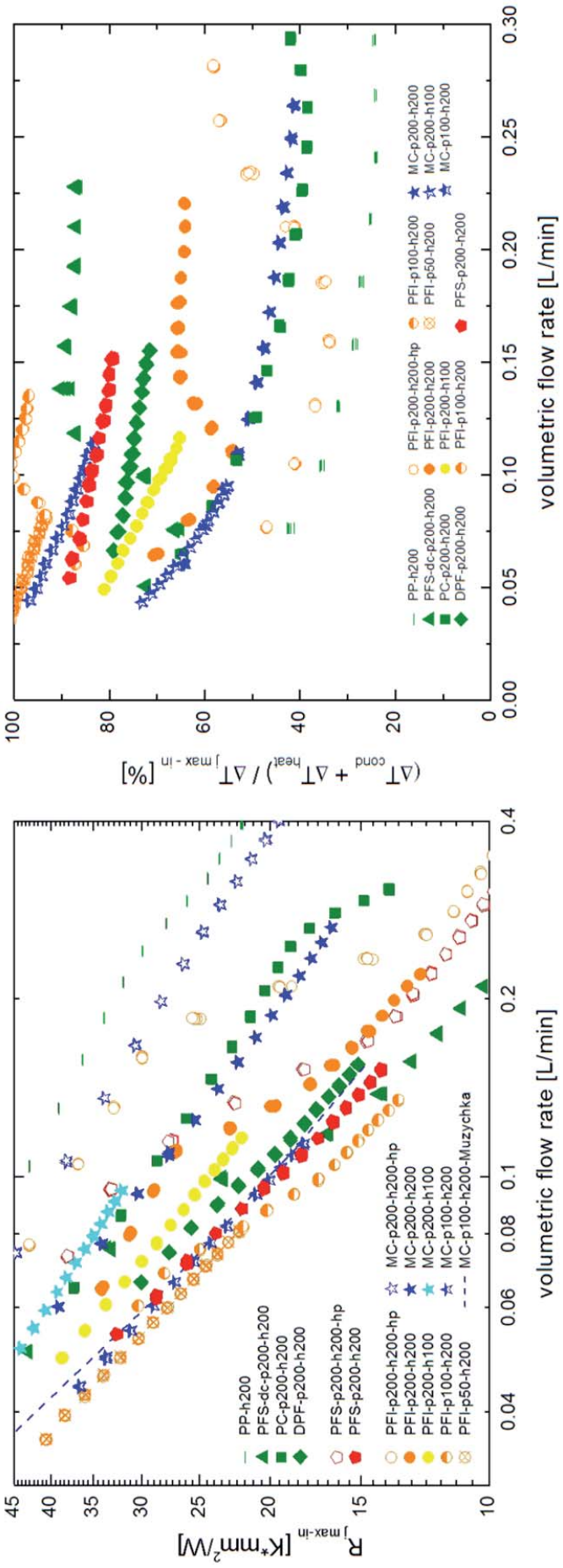


Figure 4.7: Normalized thermal resistance of individual TVs versus volumetric flow rate (left). Contribution of the thermal gradient to the total temperature gradient resulting from conduction and the fluid temperature increase (in percent) (left).

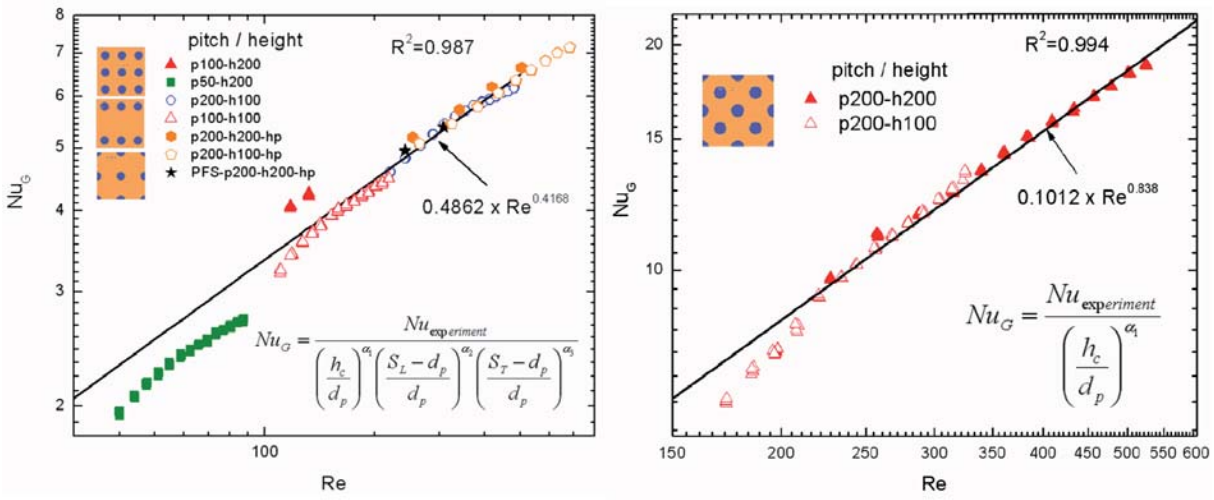


Figure 4.8: Comparison of NUSSELT numbers obtained from experiment and correlation, Equation 4.14, developed by fitting the data of pin-fin in-line structures at low Re numbers, before the onset of the flow regime transition ($Re < Re_{crit}$) (left). Comparison of NUSSELT numbers obtained from experiment and correlation, Equation 4.15, developed by fitting the data of pin-fin staggered structures at $Re < 100$ (right).

$$A_t = 2\left(A_{heater} - N_L \cdot N_T \cdot \frac{\pi d_p^2}{4}\right) + (\pi \eta_f h_c N_L N_T d_p). \tag{4.12}$$

The fin efficiency for equal temperature boundary conditions on both ends of the fin is given as

$$\eta_f = \frac{2}{m \cdot h_c} \cdot \frac{cosh(m \cdot h_c) - 1}{sinh(m \cdot h_c)} \; with \; m = \sqrt{\frac{hP_{fin}}{k_{Si} A_{fin}}}, \tag{4.13}$$

where P_{fin} is the perimeter and A_{fin} the cross-sectional area of the fin. Experimental data of pin-fin in-line at values $<Re_{crit}$ was correlated assuming the following form:

$$Nu = c\left(\frac{h_c}{d_p}\right)^{\alpha_1}\left(\frac{S_L - d_p}{d_p}\right)^{\alpha_2}\left(\frac{S_T - d_p}{d_p}\right)^{\alpha_3} Re^m, \tag{4.14}$$

proposed by PRASHER et al. [94]. All relevant parameters are included, such as the cavity height h_c and the longitudinal S_L and transversal pitch S_T to pin diameter ratio (Tables 4.2, 4.3). All data was found to be within a confidence level of 15 % (Figure 4.8 left). Only two samples of pin-fin staggered structures at two different pin heights, 100 and 200 µm, were measured. Therefore only two independent variables were considered in the data-fitting routine using the following equation (Tables 4.2, 4.3):

$$Nu = c\left(\frac{h_c}{d_p}\right)^{\alpha_1} Re^m. \tag{4.15}$$

The data was found to be within a confidence level of 10 % (Figure 4.8 right).

4.1.4 Flow Transition Augmented Heat Transfer

As discussed, abrupt pressure gradient changes and thermal resistance drops were observed at similar critical flow rates for structures with pin-fin in-line characteristic at low endwall damping (Figure 4.9 left) [117, 53]. Figure 4.9 (right) shows the normalized local pressure drop and normalized $\triangle T_{j-in}$ readings of the PFI-p200-h200-hp device with a transition regime between 160 and 280 $\frac{\text{ml}}{\text{min}}$. At flow rates below the transition, the pressure gradient is reduced towards the outlet, which can be attributed to developed boundary layers. In the transition regime, the outlet pressure gradient increases, whereas at larger flow rates the inlet pressure drop from position 0 to 1 mm is drastically reduced. Spatially resolved junction-temperature readings exhibit local temperature maxima moving from the fluid outlet towards the inlet at increasing flow rate in the transition phase (Figure 4.9 right, Figure 4.10 left). The local convective heat transfer at the outlet seems to improve. Note that the peak temperature even below the flow transition is not at position 10 mm, which corresponds to the fluid outlet, as there is some heat spreading in the silicon base. Detailed flow pattern observations using micro particle image velocimetry (μ- PIV) revealed vortex shedding as the source of the increased pressure drop and heat transfer [122]. Fluid mixing occurs after a critical distance downstream from the fluid inlet. With increasing flow rate, the transition front shifts towards the inlet. More detailed studies are needed to be able to predict the occurrence of the transition precisely. A detailed understanding would allow to engineer localized vortex shedding on hot-spots at a minimal pressure drop penalty.

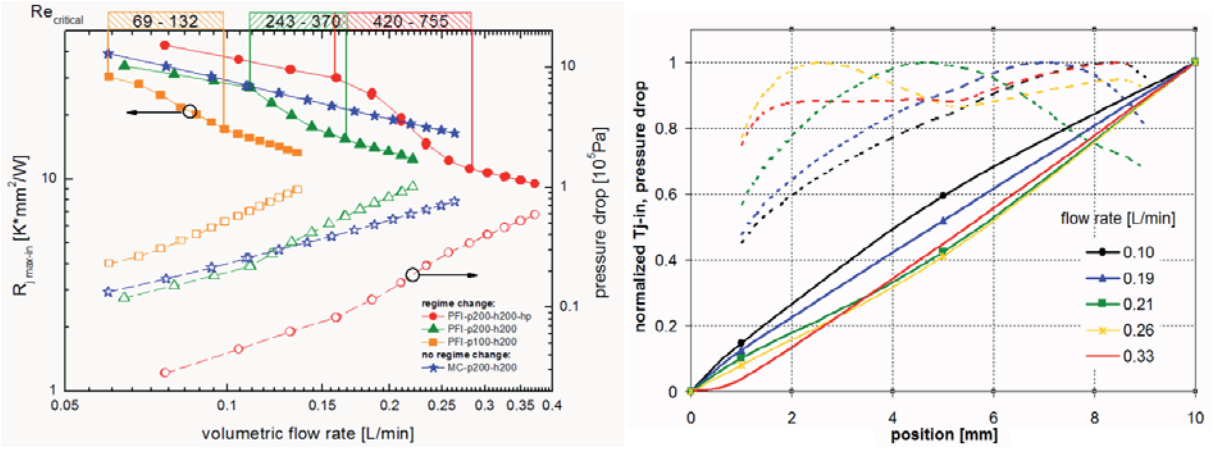


Figure 4.9: Lineup of thermal and hydrodynamic data for TVs with a distinct transition regime clearly shows the coincidence between pressure drop and thermal transition (left). Normalized local $\triangle P$ (solid) and normalized T_{j-in} (dashed) readings at different flow rates of device PFI-p200-h200-hp with a transition regime between 160 and 280 $\frac{\text{ml}}{\text{min}}$ (right).

4.1.5 Performance at Pressure or Pumping Power Constraints

Pressure Demand at a Given Heat Flux

The pressure drop is the most stringent metric in interlayer cooling, as centrifugal pump units used in servers are rather limited in pressure head than in volumetric flow rate. Therefore the TVs have to be evaluated for the pressure needed at a given heat flux and $\triangle T_{jmax-in}$ budget of 60 K. Of the fully populated TVs with 200 µm pitch and 200 µm channel height, device PFS-dc performs best, followed by PFI, PFS and DPF with the same performance as the MC and the PC with the worst performance (Figure 4.11 left). Both PFS-dc and PFI undergo the flow-regime transition and perform best. All devices perform significantly better at half population. The performance penalty due to the pitch reduction from 200 to 100 µm in the PFI is not severe if compared with the reduction at 50 µm pitch for a cavity height of 200 µm. A low performance is also measured for PFI TVs with 100 µm height. The measured data was scaled as described in subsection 4.1.1 to predict the pressure needs at given heat flux for a 4 cm^2 chip stack (Figure 4.11 right). The accuracy of the extrapolated data beyond the transition regime is unclear because its dependency on the area size has not yet been studied. Therefore devices undergoing a

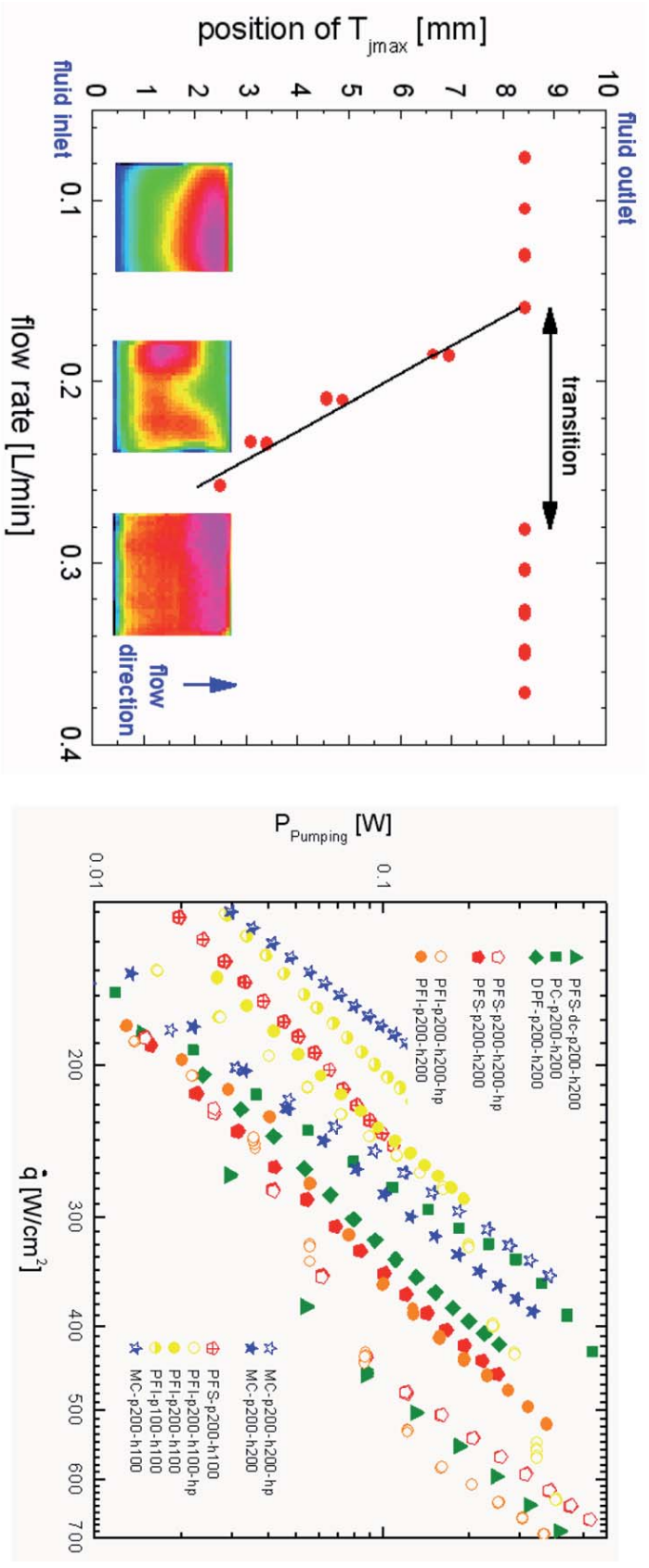


Figure 4.10: Position of maximal junction temperature measured from the fluid inlet at varying flow rates of device PFI-p200-h200-hp. Spatial resolved infrared images scaled to T_{jmax} for the individual flow rates show a transversal temperature non-uniformity during the transition. The color coding is from T_{jmax}, pink, to T_{jmin}, blue (left). Pumping power needs for a given uniform heat flux at a stack area of 1 cm^2 and a thermal budget of 60 K (right).

transition are marked by ''?''. For the larger chip size, the performance depends even more pronounced on the device friction factor. This causes the MC and PC designs to achieve a performance close to that of the PFI. The PFS TVs, with round and drop-shaped pins perform worst, accordingly. Table 4.4 compares experimental and extrapolated data of the best-performing devices. At $1\,\mathrm{cm}^2$ with pressure drops of 0.6×10^5 Pa, heat-flux levels $\geq 200\,\frac{\mathrm{W}}{\mathrm{cm}^2}$ can be removed at pitches $> 50\,\mu\mathrm{m}$ and $h_c = 200\,\mu\mathrm{m}$. At $h_c = 100\,\mu\mathrm{m}$, this level drops below $200\,\frac{\mathrm{W}}{\mathrm{cm}^2}$ for the $100\,\mu\mathrm{m}$ pitch. For $4\,\mathrm{cm}^2$ chip stacks, higher pressures, such as 1×10^5 Pa are needed to remove a heat flux $> 175\,\frac{\mathrm{W}}{\mathrm{cm}^2}$ at $100\,\mu\mathrm{m}$ pitch and $h_c = 200\,\mu\mathrm{m}$ or $200\,\mu\mathrm{m}$ pitch and $h_c = 100\,\mu\mathrm{m}$.

An assessment of different pin shapes at non-uniform power dissipation as a function of the power map contrast is presented in appendix A.

Pumping Power Demand at a Given Heat Flux

The pumping power need to operate a fluid cavity is calculated by multiplying the pressure drop and the flow rate at a given heat flux (Figure 4.10 right). The pumping power is in general reduced by an increasing hydraulic diameter of the unit-cell or by half population of pins in the dimension range considered in the experiment. Devices in pin-fin in-line type configuration profit from the regime transition. Moreover, for fully populated TVs, the pin-fin designs outperform the microchannel and pearl chain structures in general. The best-performing device is PFS-dc with distorted unit cell, going through the transition, and having the largest hydraulic diameter as well as fluid-mixing capability. This device already performed best in the pressure drop vs. heat flux comparison, but is the only one with nonuniform pitch in x- and y-direction. If technologically possible, the optimal design would have a minimal interconnect pitch in the stream-wise direction and a coarse transversal pitch to minimize the flow resistance.

TV/Pitch/Height/Population [µm]	Heat Flux [$\frac{\mathrm{W}}{\mathrm{cm}^2}$] $1\,\mathrm{cm}^2$ at 0.6×10^5 Pa	Heat Flux [$\frac{\mathrm{W}}{\mathrm{cm}^2}$] $4\,\mathrm{cm}^2$ at 1×10^5 Pa
PFS-dc/200/200/full	537	258?
PFI/200/200/half	681	425?
PFI/200/200/full	416	228?
PFI/100/200/full	394	177
PFI/50/200/full	200	92
PFI/200/100/half	300	186
PFI/200/100/full	223	123
PFI/100/100/full	156	74

Table 4.4: Performance of interlayer cooling at constant pressure drop. Devices marked with ''?'' are in the flow-regime transition with low extrapolation accuracy.

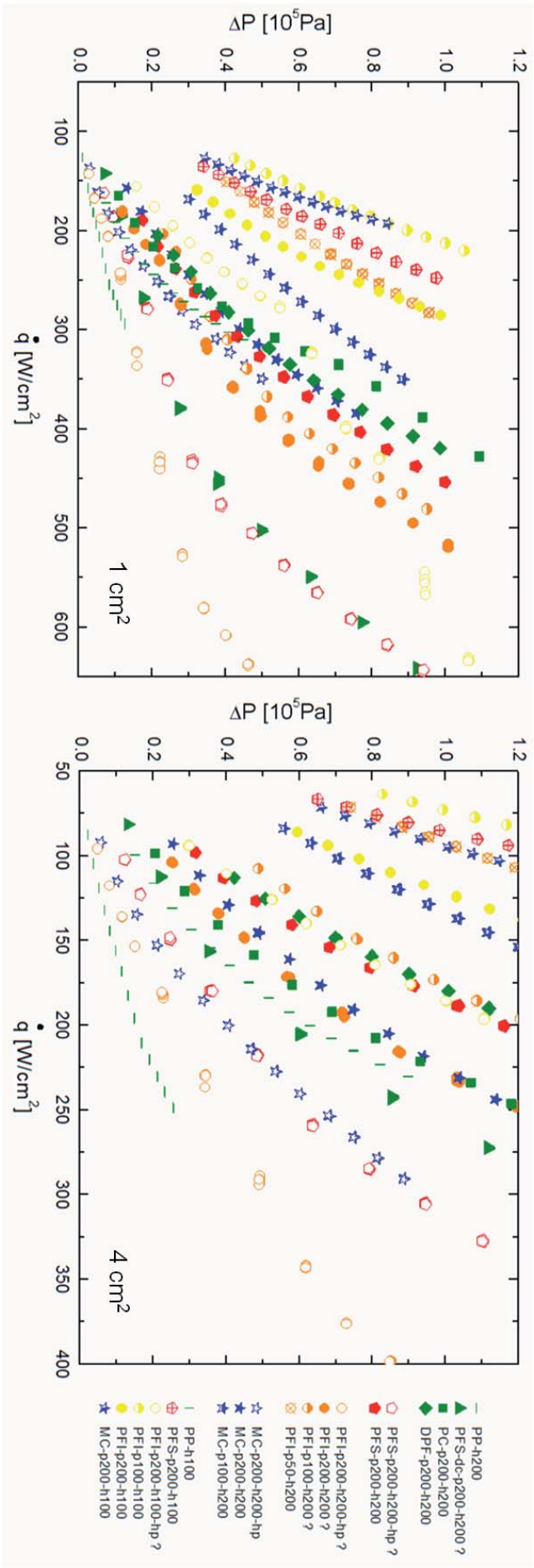


Figure 4.11: Pressure needs for a given uniform heat flux at a stack area of 1 cm^2 and a thermal budget of 60 K (left). Extrapolated pressure needs for a given uniform heat flux at a stack area of 4 cm^2 and a thermal budget of 60 K. Devices marked with "?" undergo a flow-regime transition, i.e., only data below the transition can be trusted (right).

4.2 Non-Uniform Single Cavity Experiment to Validate Multi-Scale Concept

A second set of experiments was performed to study the benefit of non-uniform heat removal building blocks such as fluid focusing, heat transfer structure modulation, and fluid delivery architectures [87]. The 1 cm^2 cavity of the non-uniformly heated test vehicle presented in section 3.3.2 was populated with six different designs (Figure 4.12).

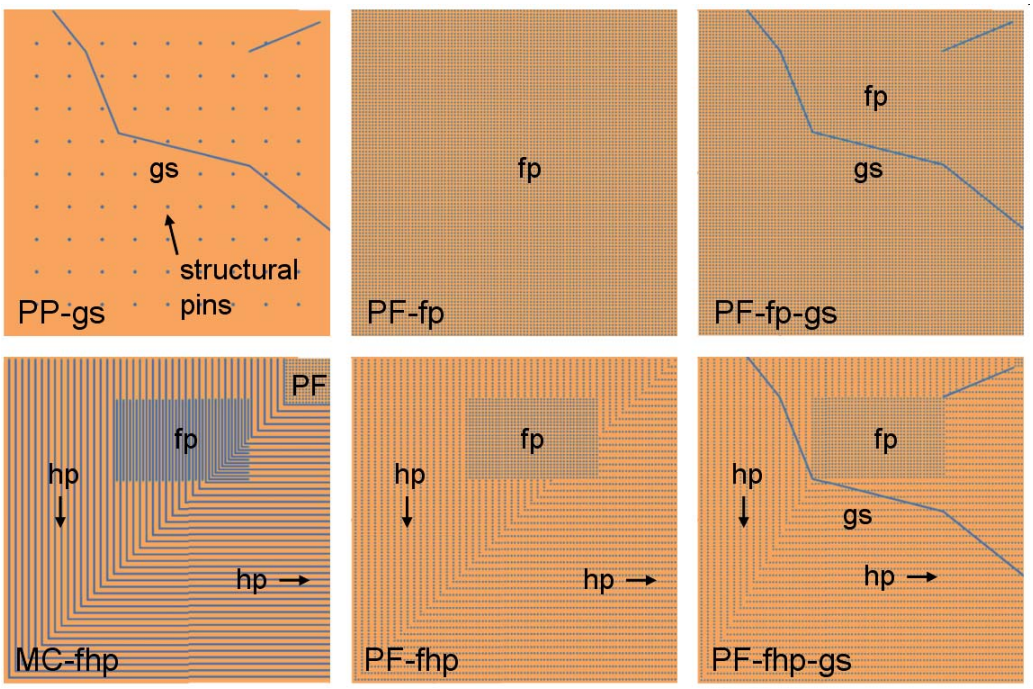


Figure 4.12: Design of the six heat transfer cavities with 100 µm height and unit-cell pitch. The pin diameter and channel width is half the pitch. Design PF-fhp-gs was also fabricated at a 50 µm unit-cell pitch. The acronyms are defined in Table 4.5.

Structure	Abbreviation
Parallel plate with guiding structure	PP-gs
Microchannel full-half populated	MC-fhp
Pin-fin full populated	PF-fp
Pin-fin full-half populated	PF-fhp
Pin-fin full populated with guiding structures	PF-fp-gs
Pin-fin full-half populated with guiding structures	PF-fhp-gs
Pin-fin full-half populated with guiding structures and 50 µm pitch	PF-fhp-gs-50

Table 4.5: Overview of cavity designs of implemented test vehicles and their abbreviation.

The 100 µm high cavity with inlet and outlet along top and right edge represents one quadrant of a 4 cm^2 fluid cavity with four-port fluid delivery architecture. A parallel plate device (PP) including fluid guiding structures (gs) made from silicon walls with a width of 100 µm to focus the coolant towards the hot-spot location is used as a reference. Structural pins were placed at a 1 mm pitch to support the fluid cavity and allow inlet pressures of 1×10^5 Pa without cracking of the silicon dies. Pin-fins (PF) at full population (fp) with a pitch and diameter of 100 and 50 µm respectively were aligned to the in- and outlet edge (PF-fp). The combination of pin-fins at full population and fluid guiding structures results in TV PF-fp-gs. Fluid structure modulation is implemented with areas of fully and half populated microchannels or pin-fins at hot-spot and background locations (MC-fhp, PF-fhp, PF-fhp-gs). The half

populated pin-fins are arranged orthogonal to the in- and outlet edge, resulting in two domains, with pins arranged in the main flow direction. A seventh design equivalent to PF-fhp-gs was additionally built, at a pitch of 50 µm (PF-fhp-gs-50) (Figure 4.13 left and right, Table 4.5).

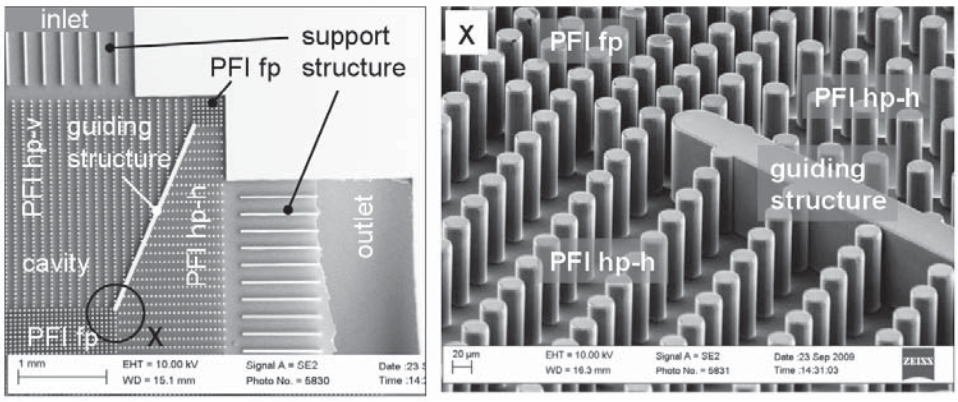


Figure 4.13: SEM photograph presenting the location and orientation of full and half populated pin-fins with guiding structure (PF-fhp-gs-50) (left). A close-up of the pin-fin array and the guiding structure is shown on the right.

4.2.1 Hydrodynamic and Spatially Resolved Thermal Measurements

The mass transport of each TV was characterized at a fluid inlet temperature T_{fin} of 25 °C without power dissipation. The calibration or the on-chip RTDs and the infrared camera was performed with changing T_{fin} from 10 °C to 70 °C. The infrared camera alignment was performed relative to the thermal alignment markers (section 3.3.2). Three different power maps were applied for a given pressure drop from cavity in- to outlet. First, both heaters were activated at a heat flux ratio of 5:1 (the hot-spot to background heater) on both TV sides. The power dissipation was adjusted to achieve a T_{jmax} of 60 °C at a T_{fin} of 25 °C. In the second and third reading, the hot-spot heater or the background heater was switched off, respectively. The maximal possible heat flux to be dissipated at a thermal budget of 60 K for the specific cavity was then linear extrapolated with respect to this readings.

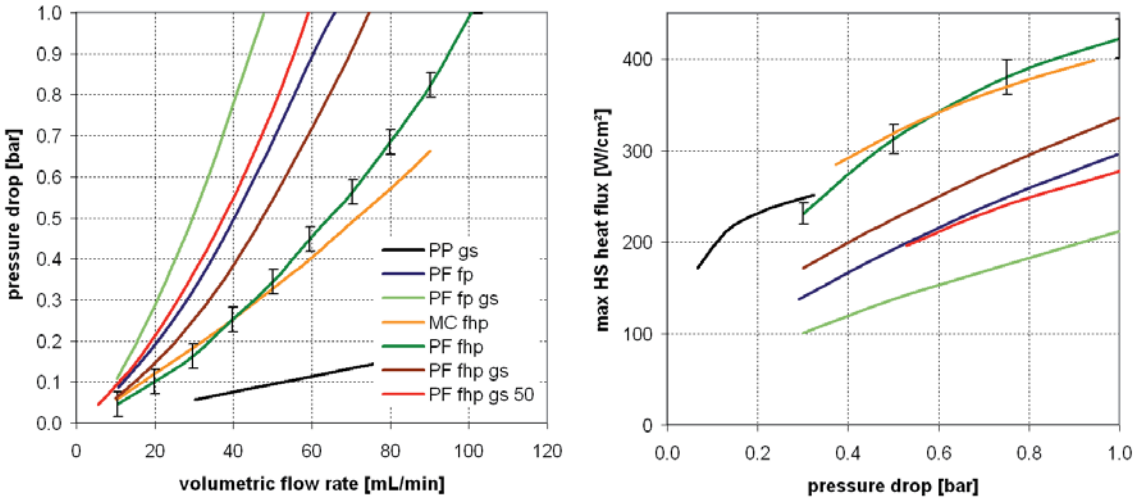


Figure 4.14: Pressure drop vs. volumetric flow rate of non-uniform cavities at a fluid temperature of 25 °C (left). Maximal hot-spot heat flux dissipated from both sides of the cavity at a given pressure drop, a thermal budget of 60 K, and a heat flux ratio of 5:1 (right).

The penalty in mass transport due to the fully populated pin-fin array in the cavity is clearly stated in Figure 4.14 (left), by comparing the PP-gs (black) with PP-fp-gs (light green). The overall flow resistance is somewhat relaxed without fluid guiding structures (PF-fp, blue) and improves even more, if unit-cell modulation is incorporated (PF-fhp (dark green), PF-fhp-gs (brown)). The pressure drop of the

PF-fhp is smaller compared to the MC-fhp (yellow) at low flow rates $< 40\,\frac{\text{ml}}{\text{min}}$. This is a consequence of the larger effective hydraulic diameter of pin-fins compared to microchannels (as discussed in subsection 4.1.2). This statement changes due to the non-linear characteristics at higher flow rates of the PF cavities, caused by fluid flow at arbitraty an angle-of-attack (compare with pressure drop characteristics of staggered pin-fins as shown in Figure 4.4 left). The advantage of unit-cell modulation is nicely demonstrated by comparing the fully populated pin-fin device (PF-fp, blue) with the modulated cavity (PF-fhp-gs, red) at a pin pitch of 100 µm and 50 µm, respectively.

The maximal hot-spot heat flux dissipated from both sides into the TV cavity at a fluid inlet to maximal junction temperature budget of 60 K and a heat flux ratio of 5:1 between hot-spot and background heater is depicted in Figure 4.14 (right). The TVs with unit-cell modulation but without guiding structures outperform all other TVs. The pin-fin arrangement (PF-fhp, green) is the preferred unit-cell geometry at larger pressures of > 0.6 bar thanks to increased fluid mixing, compensating for the lower flow rate, compared to microchannels (MC-fhp, yellow). The heat transfer capability of the parallel plate (PP-gs, black) is high at low pressures but enters the convective resistance dominated regime quite early as indicated by the reduced slope. The design of the fluid guiding structures focusing the coolant towards the hot-spot seams to be poor, resulting in a low power dissipation potential (brown, light green). Compared, the unit-cell modulation concept allows the extension of the interlayer cooling concept to 50 µm TSV pitches at high contrast power maps (PF-fhp-gs-50 (red) vs. PF-fp (blue)).

The detailed temperature response of the PF-fp at 1 bar pressure drop is depicted in Figure 4.15. The four-port fluid delivery architecture can mitigate high temperatures at the hot-spot heater location with five times higher heat flux compared to the background heater location (4.15 left). The maximal junction temperature occurs in the lower right corner, which can also be observed at uniform power dissipation (4.15 center). The streamline length from inlet to outlet towards the left and lower edge of the fluid cavity is monotonically increasing (Figure 4.20). This means, a minimal local coolant flow rate at a maximal sensible heat increase with the corresponding minimal convective heat transfer and maximal fluid temperature increase. A thermal wake caused by fluid temperature increase identifies the fluid flow with only the hot-spot heater active (4.15 right). The hot-spot heater area can be visually identified, but is not perfectly uniform due to heat spreading.

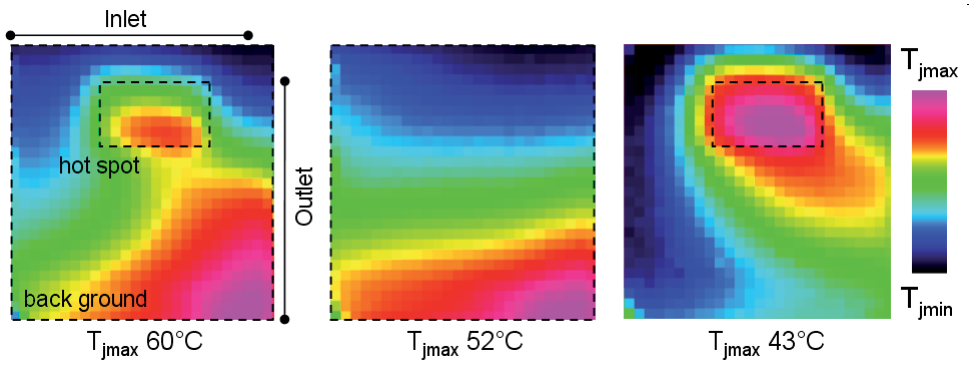


Figure 4.15: Temperature response measured by infrared imaging of the PF-fp cavity at 1 bar pressure drop and T_{fin} of 25 °C. Three different power maps were applied: active hot-spot and background heater at a heat flux ratio of 5:1 (left), uniform heat flux with background heater only (center), and high contrast heat flux with only the hot-spot heater active (right).

The use of half populated pin-fin or microchannel arrays in the background heater area improves the mass transport along the left and lower edge. Hence, the temperature maximum shifts from the lower right corner to the hot-spot heater area (PF-fp compared to PF-fhp and MC-fhp, Figure 4.16). Fluid focusing mitigates the junction temperature in the hot-spot compared to the background heater area. The increased coolant flow towards the hot-spot is compensated with a reduced fluid flow in the lower left background heater area, resulting in a more significant lower right peak junction maximum (PF-fp, PF-fhp compared to PF-fp-gs, PF-fhp-gs, Figure 4.16). In general, the design of fluid guiding structures to balance the junction temperature difference between hot-spot and background heater area has to be performed with respect to the heat transfer structure population present in the cavity. Devices

with a temperature gradient dominated by the convective thermal resistance, such as the parallel plate device (PP-gs), demand for aggressive fluid focusing (PP-gs, Figure 4.16). Whereas, modulated and fully populated pin arrays benefit from moderate and absent fluid focusing (PF-fhp-gs, PF-fp-gs, Figure 4.16).

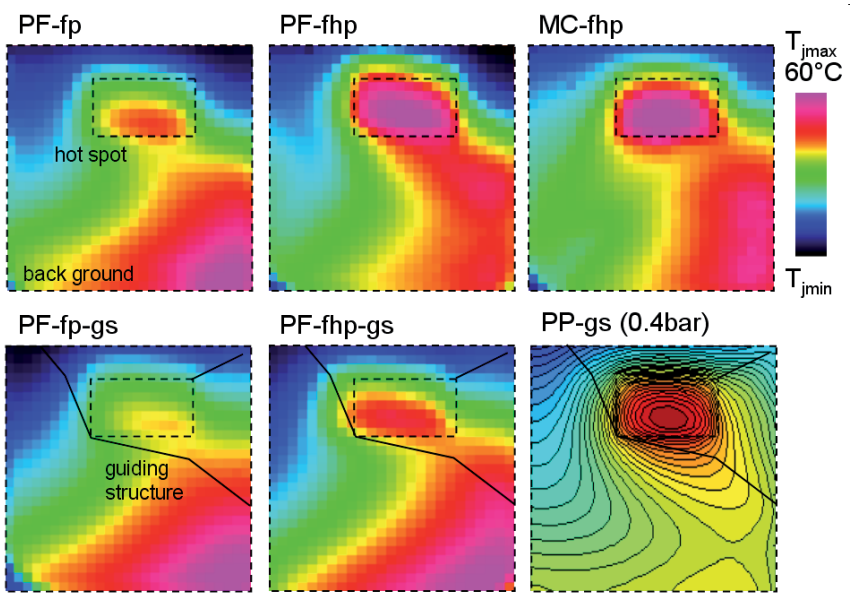


Figure 4.16: Temperature response measured by infrared imaging of the non-uniform TVs at 1 bar (PP-gs 0.4 bar) pressure drop and T_{fin} of 25 °C. The heat flux ratio between hot spot and background heater was set to 5:1 and the total power was adjusted to result in a $T_{jmax}$of 60 °C.

4.2.2 Field-Coupled Porous Media Model Representing Non-Uniform Cavities

The heat and mass transport in the non-uniform fluid cavities is modeled using the multi-scale modeling method proposed in chapter 2. The silicon dies are represented by 3D domains connected by field-coupling for the energy transport to the quasi-2D porous fluid-cavity domain in the finite volume model. The quasi-two dimensionality of the porous domain is achieved with slip-flow conditions at the upper and the lower cavity wall. Furthermore, the thermal conductivity of the coolant is set to be orthotropic with an infinite value normal to the cavity surface. The extended tensor and sinusoidal description were used for the permeability (Equation 2.50) and convective thermal resistance (Equation 2.51) respectively. Pressure boundary conditions are imposed on the inlet and outlet of the cavity domain (Figure 4.17, details of the implementation can be found in the appendix B).

The discretization is performed with an unstructured hexahedral dominant swept mesh provided by the meshing tool of ANSYS (Figure 4.17). For symmetry reasons, only one silicon slab and half of the cavity height were considered (Figure 4.17, inset). Only two volume element layers normal to the silicon slab surface are needed to represent the porous domain owing to the quasi-two dimensionality of the cavity domain. A mesh refinement in the upper right corner is performed where high velocity gradients are expected. The mesh representing one half of the test vehicle with up to 10,000 pin-fins consists of 43,273 nodes. A momentum-loss term is added to the NAVIER–STOKES equation as described in 2.19 to account for the porous nature of the cavity domain. The partial differential equations were solved with CFX 12 at a root mean square convergence criterion of 10^{-6}.

4.2.3 Validation of Multi-Scale Concept with Detailed Conjugate Heat and Mass Transfer Model of the Parallel Plate Test Vehicle

The validation of the flow rate and junction temperature predicted by the porous media model can be performed by a detailed conjugate heat and mass transfer model resolving the boundary layers with the use of thirty instead two volumes orthogonal to the cavity surface for the PP-gs TV (Figure 4.18

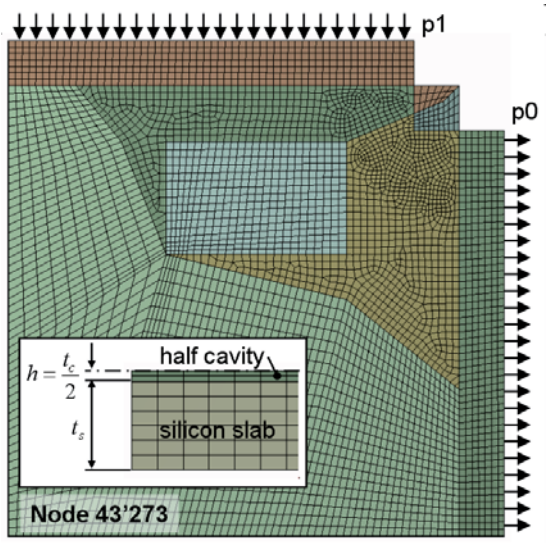


Figure 4.17: Top view of the unstructured hexahedral dominant mesh, with pressure boundary conditions at inlet and outlet. The fluid guiding structures are captured by sub-domains. The cross-section of the swept mesh representing one silicon slab and half the cavity is shown in the inset.

left). The pressure drop prediction of the porous model (triangles) with a isotropic permeability of $8.367 \times 10^{-10}\,\mathrm{m}^2$ are in close agreement with the detailed model (lines). Even recirculations behind guiding structures are captured correctly as shown in Figure 4.18 (right) (blue and red streamlines from porous and detailed modeling). A reliable prediction of recirculations in the design phase of guiding structures is important. Vortices larger than the heat spreading radius in the solid cause hot-spots due to their low exchange of coolant. The heat dissipation performance at a heat flux ratio of 5:1 and a thermal budget of 60 K is predicted conservatively by the porous model using the velocity independent convective thermal resistance for developed boundary layers of $40.5\,\frac{\mathrm{K\,mm^2}}{\mathrm{W}}$ for the solid-to-porous-media field coupling (red triangles and line from porous, detailed model). The accuracy improves if the velocity dependent apparent heat transfer coefficient for an average streamline length of 10 mm is applied (red diamonds).

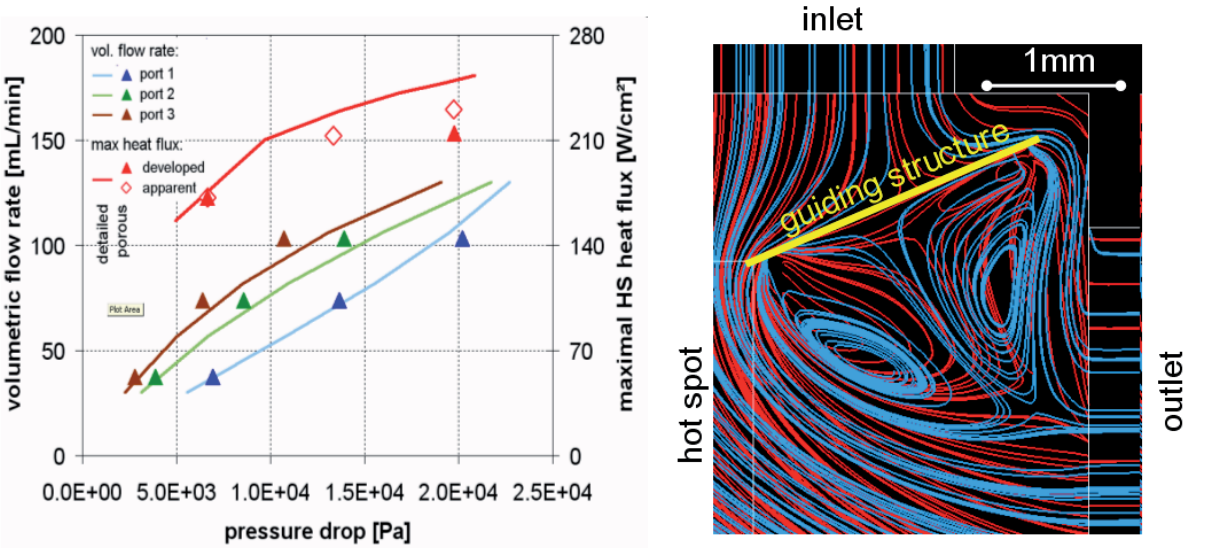


Figure 4.18: Predicted pressure values at the pressure ports for a given flow rate and maximal hot-spot heat flux at a hot-spot to background heat flux ratio of 5:1 and a thermal budget of 60 K derived from detailed and porous media modeling (left). Close-up of the upper right corner of the fluid cavity resolving streamlines of recirculations behind the fluid guiding structure (blue: porous, red: detailed) (right).

4.2.4 Validation of Multi-Scale Modeling Concept with Experimental Results

The mass transport prediction for pin-fin TVs was compared with experimental pressure port readings. The numerical results were obtained using the extended tensor, sinusoidal, or orthotropic permeability description for the porous domain as introduced in subsection 2.3.3 and are plotted against experimental results for the PF-fhp-gs device (Figure 4.19 left). Both, the sinusoidal and the orthotropic description result in an approximate 80 % overestimation in flow rate compared to a 18 % error for the extended tensor description (Figure 4.19 right, PF-fhp-gs). The sinusoidal and orthotropic estimation results in errors as high as 177 and 155 % for the TV with modulated pin-fin population but no fluid guiding structures (PF-fhp). The sinusoidal and extended tensor description perform equally conservative for TVs with fully populated pin-fin arrays (PF-fp, PF-fp-gs). The orthotropic description can be used reliably for microchannels and the parallel plate, with infinitely low permeability orthogonal to the microchannel direction and equal permeability in the two principle axis spanning the quasi-2D porous media (MC-fhp, PP). Based on this findings, it can be concluded that the basic sinusoidal permeability description is sufficient to describe fully populated pin-fin arrays, compared to half populated pin-fin arrays which are only correctly represented by the extended tensor description using look-up tables.

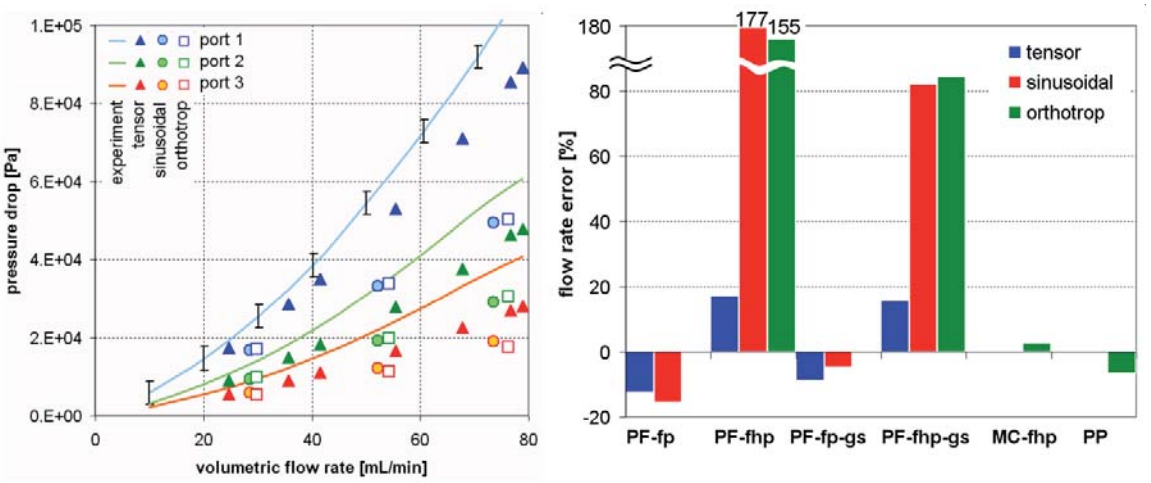


Figure 4.19: Pressure drop predictions from port 1, 2, and 3 with respect to port 4, using the extended tensor, sinusoidal, and orthotropic description of the permeability of the porous media, compared to experimental readings for FP-fhp-gs (left). Volumetric flow rate prediction error in percent with respect to the experimental value at 1 bar pressure drop for different TVs and permeability descriptions (right).

The modeled streamlines of the PF-fp device at 1 bar pressure drop using the sinusoidal or extended tensor description of the permeability are presented in Figure 4.20. A smooth flow direction change is observed for the sinusoidal case (left), whereas more pronounced flow in in-line (0°) and staggered (45°) direction results from the tensor description (right). This can be expected due to the local permeability maximum in these directions (Figure 2.29 left), causing this lock-in effect.

The DARCY velocities with the associated streamlines at 1 bar pressure drop for modulated heat transfer structures with and without fluid focusing are presented in Figure 4.21. An orthotropic permeability with an infinitely small value orthogonal to the channel direction was used to represent the microchannels with full (fp) and half population (hp), resulting in straight streamlines as expected (MC-fhp). Fluid by-passing at fp areas with low permeability has to be considered if communicating fluid structures like pin-fin arrays and domains with different pin densities are designed (PF-fhp, streamlines directed to the left and right of the hot-spot area). Fluid guiding structure can compensate for this, but possibly result in large velocity gradients (PF-fhp-gs). The streamlines of TV PF-fhp show a lock-in effect at a angle-of-attack of 37° and 18°. These values do not coincide with permeability maximum of the half populated pin-fin array (Figure 2.30 left). They could be the result of the long range influence of the fp domain and the change in directionality of the hp pin-fin array orientation. An interesting aspect is also the fluid velocity distribution, which appears to be more uniform for microchannels. Only the upper right half populated channels show a substantially higher flow rate due to their small length. In the hot-spot area,

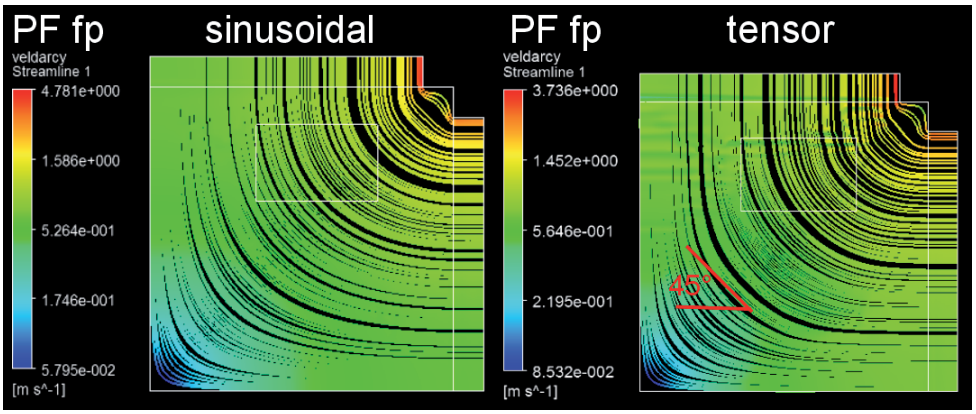


Figure 4.20: Modeled streamlines of device PF-fp at at pressure drop of 1 bar using the sinusoidal and extended tensor description for the porous media (left, right). The color coding reflects the local DARCY velocity.

the fp domain throttles the flow resulting in similar velocity compared to channels towards the lower left corner (MC-fhp). Almost a stagnating flow can be observed for the PF-fhp along the diagonal towards the lower left corner, which represents the symmetry point of the four-port architecture. These effect can be suppressed by fluid guiding structures (FP-pf-gs). The gs spanning from inlet to outlet, splits the fluid cavity into two pressure domains, where the static pressure is reduced monotonic (Figure 4.21 bottom right).

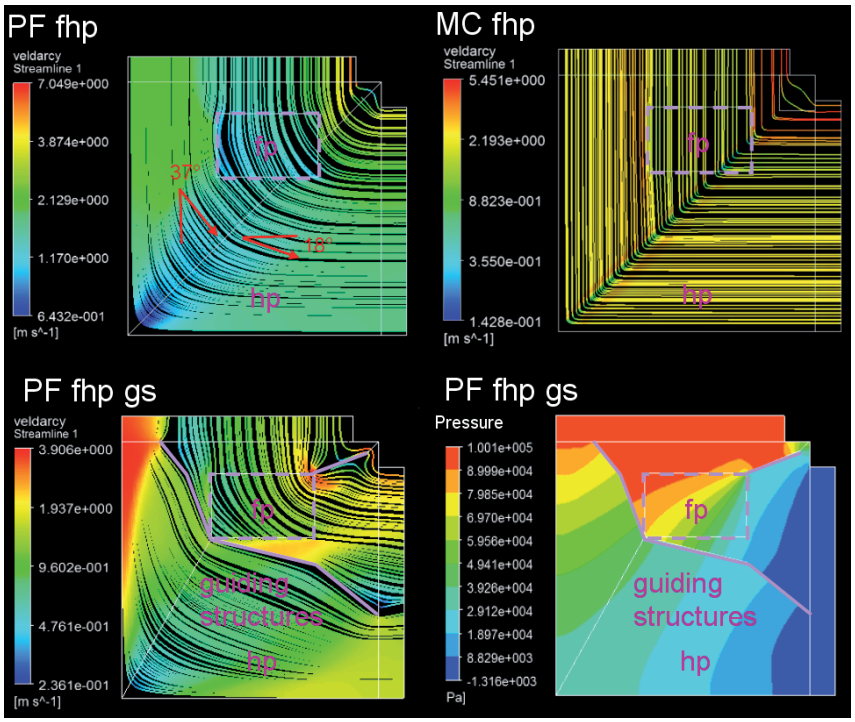


Figure 4.21: Modeled streamlines using the extended tensor and the orthotropic permeability description for the PF and the MC TV, respectively. The color coding reflects the local DARCY velocity at an applied pressure drop of 1 bar. The static pressure is plotted in the bottom right image for TV PF-fhp-gs.

The expansion (diffuser) and contraction (nozzle) of the streamlines from cavity inlet to diagonal and from diagonal to outlet results in coolant deceleration and acceleration with the associated exchange of dynamic and static pressure. The values used for the extended tensor description with the static pressure gradient as independent variable are derived from the sub-domain model with two-pairs of periodic boundary conditions representing an infinitely extended pin-fin array. Therefore, this method might not be applicable to predict fluid flows with significant transfer of dynamic to static pressure and vice

versa. Accordingly, the change in dynamic pressure ($dp_{dynamic}/ds$) is plotted relative to the static pressure gradient (dp_{static}/ds) along three different streamlines (L1 to L3) in the PF-fp cavity at 1 bar pressure drop (Figure 4.23 left). The velocity variation is smallest along L2 and results in a small relative transfer of pressure ($< \pm 0.5\,\%$). Largest absolute and relative velocity changes occur along L1 with a relative pressure transfer of $< \pm 1\,\%$. Due to the high velocities along L3 a relative pressure transfer of up to $-2\,\%$ was observed even at moderate velocity changes. The sign of the relative pressure transfer reflects in all cases the streamline section with decelerating and accelerating fluid (the sign might be counter-intuitive, but is the result of the negative pressure gradient in streamline direction). As a conclusion: the extended tensor description can be expected to hold at such low relative pressure transfers of less than 2 %, which is also supported by the pressure drop prediction validation (Figure 4.19 right).

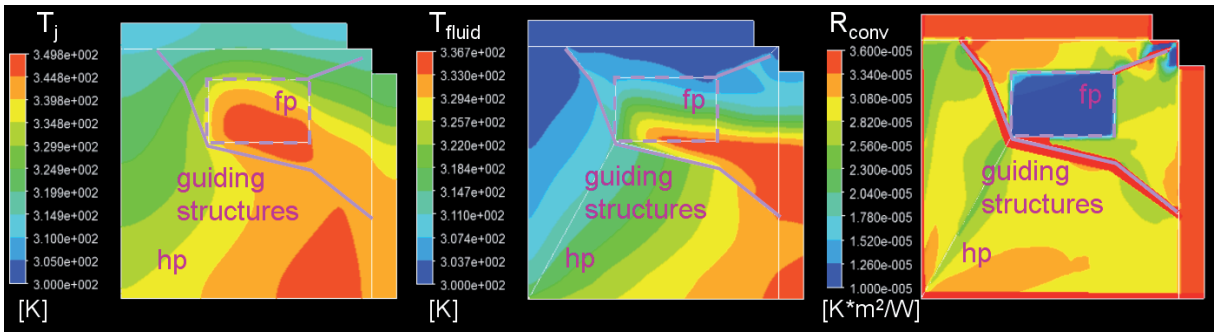


Figure 4.22: Modeling results of the TV PF-fhp-gs at 1 bar pressure drop using the extended tensor and the sinusoidal description for the mass and heat transport, respectively. The junction and fluid temperature field is depicted in the left and central contour plot. The velocity vector and pin-fin density dependent convective thermal resistance is shown on the right plot.

The predicted junction temperature field of the PF-fhp-gs TV at 1 bar pressure drop and a hot-spot, background heat flux of 250 and 50 $\frac{W}{cm^2}$, respectively, corresponds qualitatively with the experimental results (Figure 4.22 left and 4.16 bottom, center). The maximal junction temperature at hot spot and background heater locations are well balanced by heat transfer structure modulation and the resulting low convective thermal resistance in fully populated areas (Figure 4.22 right). The increase in fluid temperature of 36.7 K is substantial and results in viscosity variations of up to 50 % in the cavity (Figure 4.22 center and water properties Figure2.19). This does not only affect the overall flow rate to pressure drop dependency, but can also result in streamline changes with a positive feedback at locations of high power dissipation, where the viscosity drops. Accordingly, the viscosity change and the associated modified permeability are adjusted (needed for the pin-fin arrays as described in Figure 2.18 right) in the model. The effective permeability derived by the extended tensor description is defined by corresponding look-up tables for fluid temperatures of 25 and 60 °C and is linearly interpolated with respect to the local fluid temperature.

The apparent thermal resistance $R_{conv-apparent}$ defined as the average local thermal resistance present from the inlet to position z is plotted in Figure 4.23 (right) to assess the significance of thermally developing boundary layers on the heat transfer at hot-spot (width of 2.5 mm) or background locations (average length of approx. 5 mm). Correlations from MUZYCHKA [120] for developing hydrodynamic and thermal boundary layers in microchannels predict a 30 % drop in thermal resistance at a DARCY velocity of 1 $\frac{m}{s}$ (red, brown line). The entrance behavior of pin-fin structures at thermally developing boundary layers are derived by detailed numerical modeling. In a first step, the developed velocity field is derived at isothermal conditions of 25 °C with periodic hydrodynamic boundary conditions. Subsequently, the temperature field was computed for n cells, with n defined as the distance from inlet divided by the unit-cell pitch. The fluid inlet temperature was set to be isothermal for the first iteration. The resulting outlet fluid temperature field of the previous iteration was imposed as inlet temperature for all other iterations. The corresponding convective unit-cell resistance was finally computed in the post processor (Equation 2.39). The difference between the apparent and the developed convective thermal resistance for pin-fins is not as significant as for microchannels. This could be related to the increased fluid mixing in pin-fin structures, resulting in a fast converging flow and temperature field. As an extreme case, the

entrance length of staggered pin-fins was reported to be 5 pin rows only [123]. Accordingly, the convective thermal resistance adjustment for pin-fins is not as important as for microchannels. A reduction of less than 38 % needs to be considered for PF hp longitudinal and PF fp in-line domains, at 5 mm and 2.5 mm length, respectively and the expected DARCY velocities present in the cavity. So far, the discussion of developing boundary layers was only referring to flows in specific directions, along symmetry lines of the pin array. Further investigations would be needed to consider arbitrary angle-of-attack. It is also unclear, how to treat changing flow directions. The improvement in heat transfer might depend on the curvature radius of the stream line. In this study, the apparent convective resistance, if considered, was only a function of the pin-fin array, the sub-domain size (2.5 or 5 mm) and the DARCY velocity.

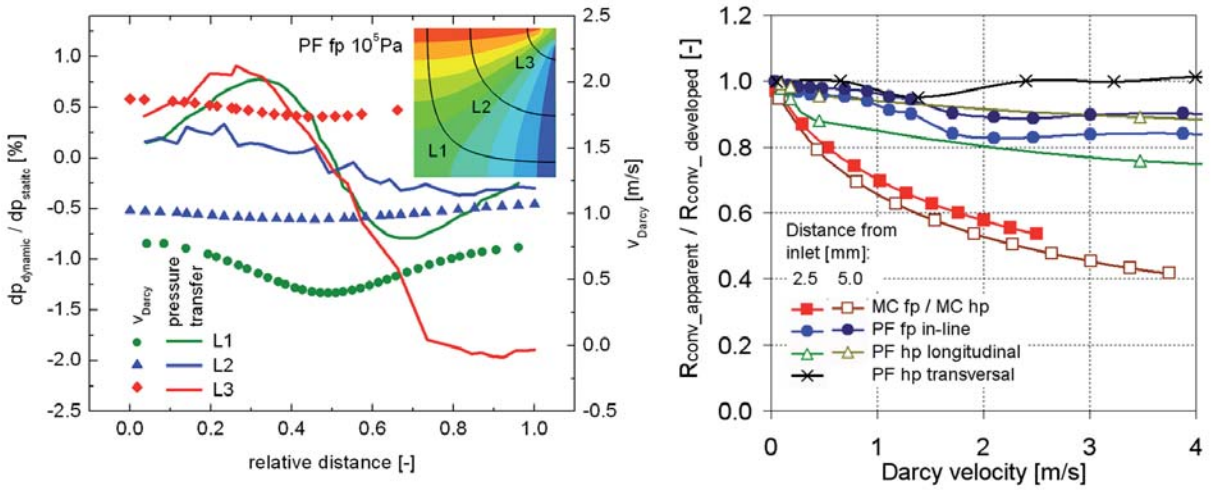


Figure 4.23: Transfer of dynamic to static pressure along streamlines (L1 to L3) in the PF-fp cavity at 1 bar pressure drop, relative to the local static pressure gradient. The corresponding DARCY velocity is also plotted. The static pressure and the streamlines of the cavity are shown in the upper right inset (left). Apparent to developed convective thermal resistance ratio for microchannels and pin-fins at a heat transfer length of 2.5 and 5 mm from the fluid inlet are plotted with respect to the DARCY velocity. The values for the microchannel and the pin-fins were derived from MUZYCHKAS correlations [120] and numerical modeling, respectively (right).

A local thermal resistance $R_{th}(x,y)$ is defined with respect to the fluid inlet temperature T_{fin} and the total power dissipated by the heaters P_{tot} , such as

$$R_{th}(x,y) = \frac{T_j(x,y) - T_{fin}}{P_{tot}}, \tag{4.16}$$

to be able to compare the modeling and the experimental results. In both cases a heat flux ratio of 5:1 was applied between the hot-spot and the background heater. In the model, heat-flux boundary conditions of 250 and 50 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ were imposed from the top and bottom die. Owing to current constraints, 120 and 24 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ were applied in the experiment. The limited optical access of the infrared camera to the test-vehicle surface resulted in artifacts at the periphery of the heat-transfer area. Therefore, these zones are excluded from the data analysis (white areas in Figure 4.24).

The thermal resistance maps of the PF-fhp-gs TV at the three different power maps (background only, hot-spot only, background and hot-spot heater) and a pressure drop of 1 bar derived by modeling and experiment are depicted in Figure 4.24. The model seams to capture the characteristics of the four-port fluid delivery with the high temperatures along the lower edge correctly based on the qualitative comparison. The effect of local temperature reduction caused by fluid focusing and the modulation of unit-cells seams to be more pronounced in the model (Figure 4.24 top). Heat spreading in the 300 µm thick silicon base can be clearly observed for the power map with hot-spot heater on only. The predicted and measured T_{jmax} locations are close to identical. Furthermore, the sensible heat increase and the flow direction of the coolant given by the fluid guiding structures becomes visible by the thermal wake magnitude and shape downstream of the hot-spot (Figure 4.24 middle). The superposition of the mentioned effects can be discussed with the background and hot-spot heater activated. In general, the consequence

of the introduced thermal building blocks are predicted well by the model. But, the thermal resistance is estimated conservatively, especially at the hot-spot heater location (Figure 4.24 bottom).

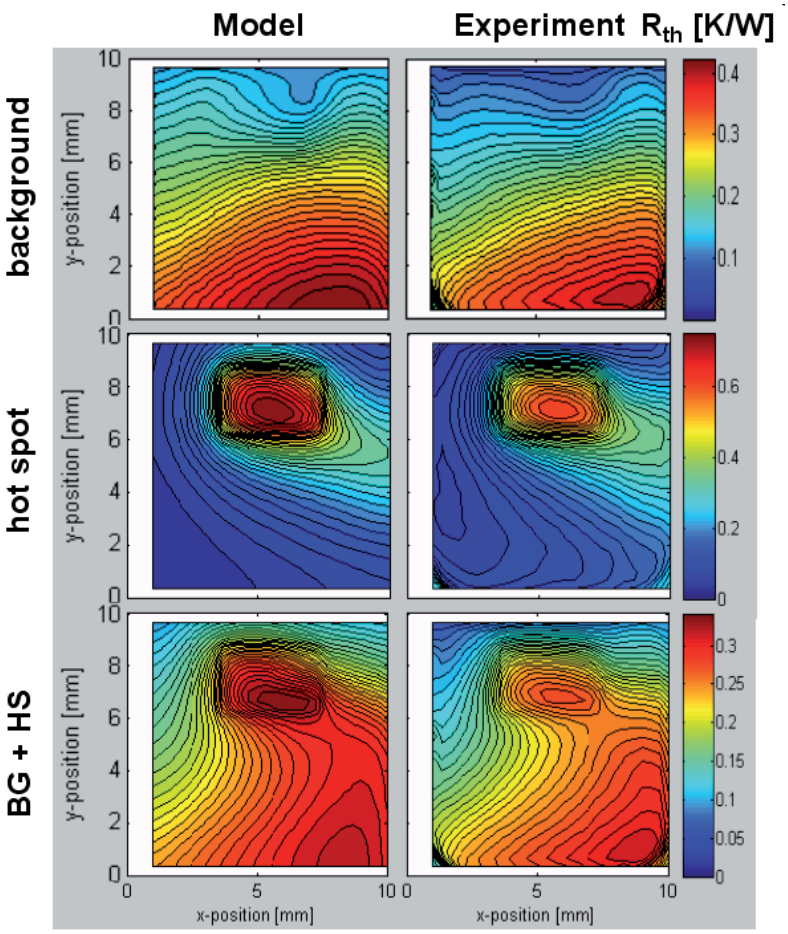


Figure 4.24: Spatial thermal resistance defined in Equation 4.16 from porous domain modeling considering temperature dependent coolant viscosity and apparent convective thermal resistances (left), as well as experimental investigation (right) for the PF-fhp-gs TV at three different power maps: background heater only (top), hot-spot heater only (center), and active background and hot-spot at a heat flux ratio of 5:1 (bottom) at a pressure drop of 1 bar. (White area of the heat transfer zone has not been incorporated into the data comparison, owing to experimental artifacts.)

The spatial thermal-resistance difference of the model with reference to the experiment for an active background and hot-spot heater is depicted in Figure 4.25 (left). The thermal resistance is predicted conservative predominantly towards the inlet and at the hot-spot location. This can be expected at an apparent thermal resistance defined uniformly across each cavity sub-domain. It results in an over- and underestimation of the local thermal resistance towards the in- and outlet. The relative error ϵ of the predicted R_{thCFD} with respect to the measured R_{thEXP} spatial thermal resistance is defined as follows:

$$\epsilon(x,y) = \frac{R_{thCFD}(x,y) - R_{thEXP}(x,y)}{R_{thEXP}(x,y)}. \tag{4.17}$$

The ϵ distribution computed for the background and the hot-spot active test case with a mean value μ of 0.174 and a standard deviation σ of 0.118 is plotted in Figure 4.25 (right). The R-square value of 0.947 indicates a high prediction quality. These three statistical values are listed in Table 4.6 for all the three power maps. The model overestimates the temperature by 17.5 % in average at uniform power dissipation. This value becomes negative (−8.5 %) and the standard deviation is increased from 11.4 to 27.9 % with only the hot-spot heater on. This is the result from the conservative, but optimistic prediction at hot-spot and background heater locations (Figure 4.24 middle).

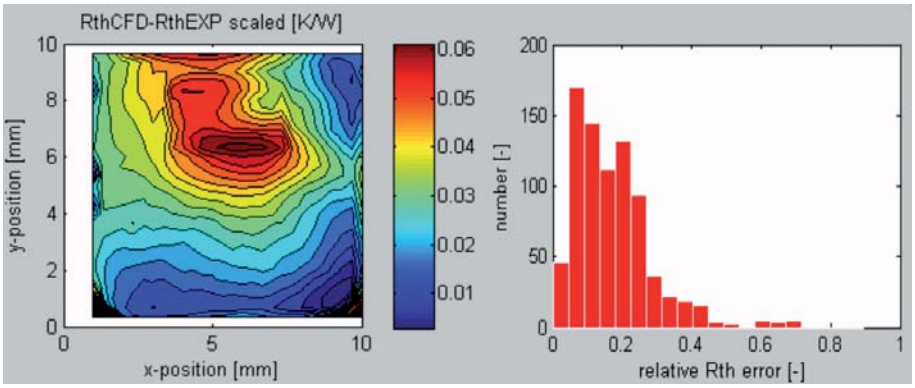


Figure 4.25: Spatial thermal resistance difference of the model with respect to the experiment at active background and hot-spot heaters and a pressure drop of 1 bar for the PF-fhp-gs TV. The model considers temperature dependent coolant viscosity and apparent convective thermal resistances (left). Distribution of the relative R_{th} error ϵ (Equation 4.17) for the same test case (right).

ACTIVE HEATER	R-SQUARE VALUE	STANDARD DEVIATION σ [%]	MEAN VALUE μ [%]
Background	0.981	11.4	17.5
Hot-spot	0.942	27.9	-8.5
Background and hot-spot	0.947	11.8	17.4

Table 4.6: Model prediction quality indicated by the R-square value and σ, and μ of the relative error ϵ distribution at the three power maps for the PF-fhp-gs TV considering temperature dependent coolant viscosity and apparent convective thermal resistances.

The predicted maximal heat flux per TV side with respect to a thermal budget of 60 K from fluid inlet to the maximal junction temperature at varying flow rates for the three power maps considering temperature dependent coolant viscosity and apparent convective thermal resistances is also validated by the experiment (Figure 4.26 left). The heat flux ratio of 5:1 is kept invariant for the test case with both the background and the hot-spot heater active. The heat transfer performance increases almost linearly with the flow rate, indicating fluid temperature dominant heat transfer. A slight saturation effect caused by the non-proportional convective heat transfer is visible. The difference in slope seams to be more pronounced at increased flow rates, indicating proper coolant temperature prediction, but increased error due to the entrance effects more prominent at high fluid velocities. Heat fluxes up to 90 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ per side can be dissipated even at uniform power dissipation, thanks to unit-cell modulation. Hot-spots with up to 480 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ in the corner of the four-port fluid-delivery architecture can be cooled, indicating again the advantage of this concept.

The prediction quality improvement from the introduction of temperature dependent coolant viscosity μ(T) and the apparent thermal resistance (apparent) can be quantified by the R-square value listed in Figure 4.26 (right) for the PF-fhp-gs TV at a constant flow rate for all models. A temperature dependent coolant viscosity results in a slight change of the flow field with a bias towards hot-spots and reduces, but improves the prediction quality at uniform compared to non-uniform power maps for the PF-fhp-gs test case. The use of an apparent convective thermal resistance improves the quality for uniform and high contrast power maps. The complexity of the model can now be adjusted depending on the afforded prediction quality based on this comparison.

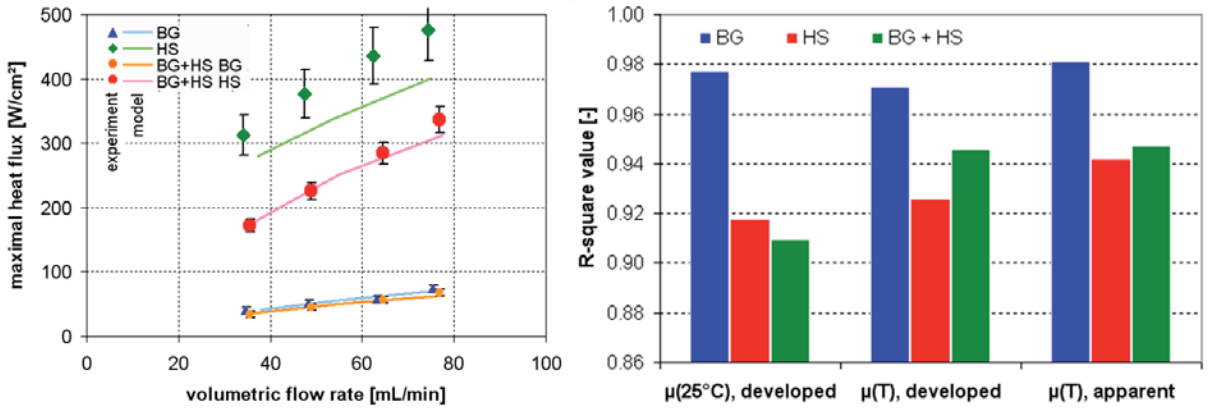


Figure 4.26: Maximal heat flux per TV side for a 60 K thermal budget from the fluid inlet to the maximal junction temperature at varying flow rates for the three different power maps for the PF fhp gs TV, considering temperature dependent coolant viscosity and apparent convective thermal resistances (left). Comparison of the model prediction quality indicated by the R-square value for the three power maps with and without the implementation of temperature dependent coolant viscosity (μ(T), μ(25 °C)) and apparent convective thermal resistances (developed, apparent) for the PF-fhp-gs TV at equal flow rates (right). (BG: background heater, HS: hot-spot heater)

The maximal heat flux per side at the 60 K thermal budget was finally estimated for all TV at 0.3 and 1 bar pressure drop considering temperature dependent viscosity and developed or apparent convective heat transfer (Figure 4.27). The confidence level for isotropic and orthotropic porous medias, such as PP and MC is high, with a deviation of +1 to −5 % to the experiment. Enhanced heat transfer structures, such as pin-fins are more delicate to describe. The extended tensor description is more general compared to the sinusoidal description and needs to be implemented if half populated pin-fin arrays are present as discussed in the beginning of this subsection. Following this guidelines, the heat flux estimate is mostly conservative with a deviation of +5 to −18 %.

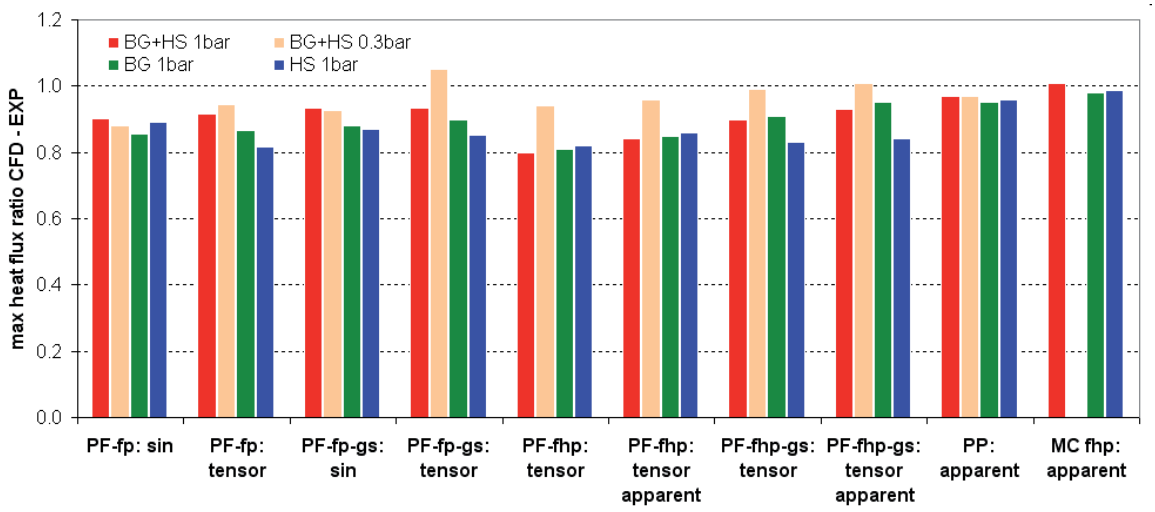


Figure 4.27: Maximal heat flux prediction quality of the porous-media model considering sinusoidal or extended tensor permeability description is depicted for individual test cavities. Non-developed thermal boundary layers are included as an apparent convective thermal resistance for a given domain if indicated. Otherwise, fully-developed boundary layers are considered.

4.3 Thermally Communicating Fluid Cavities: Interlayer Cooled Pyramid Chip Stack

The pyramid-chip-stack will be used to validate the porous-media modeling method with respect to thermal cross-talk between different active tiers and fluid cavities within a interlayer cooled chip stack as discussed in [118]. The fluid cavity spans a quadratic area of 1 cm^2 and is fully populated either with microchannel (CH) or pin-fin in-line (PF) heat transfer geometries (Table 4.7). The nominal channel and pin dimensions are listed in Table 4.8. Fluid in- and outlets with an aperture of 1.5 mm are arranged in two and four-port configuration. Again, the latter represents a single quadrant of a 4 cm^2 chip stack. One quadrant is sufficient to predict the total heat transfer performance due to symmetry reasons (Figure 4.28 left, middle). The pyramid-chip-stack is build with un-thinned silicon dies and a polyimide coating representing the back-end-of-line layers (wiring layers) compared to a product style chip stack with TSVs. The associated thermal resistances are compared and put into perspective to the convective and conductive (conduction from cavity top to bottom through the fins) thermal resistance of the porous domain (Figure 4.28 right).

The heat and mass transport experiments were initiated with a pressure drop vs. flow rate reading at 20 °C, followed by the RTD calibration at increasing fluid inlet temperature (20 to 70 °C) and were finalized by heat transfer measurements at different pressure drops and at different power maps with individual heaters powered at 12 W.

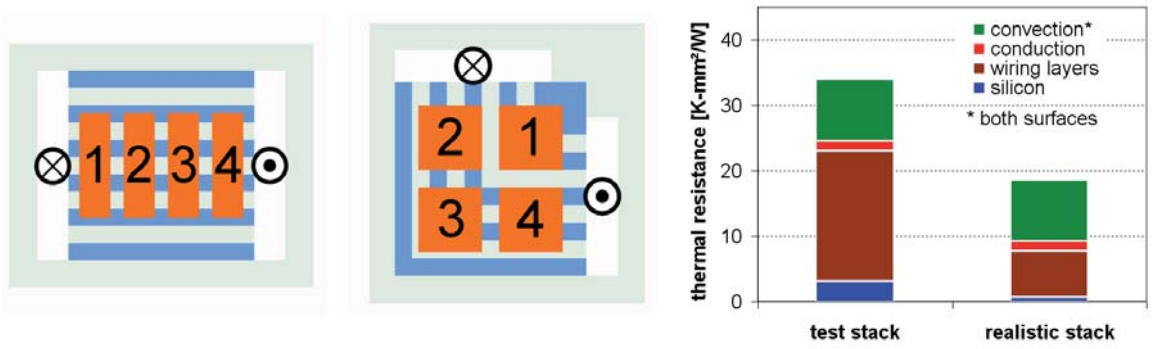


Figure 4.28: Top view of two-port (left) and four-port (middle) microchannel configuration. The four-port test vehicle represents only one quadrant of a 4 cm^2 cooled chip stack. Evenly distributed hot-spot areas (orange) are marked with identification numbers (left, middle). Thermal resistance value comparison of layers present in the thermal test stack compared to a realistic product chip stack. Most significant deviation can be noticed in the wiring layers (right).

Test Vehicle	In- / Outlet	Heat Transfer Geometry
2-port CH	2-port	microchannel
2-port PF	2-port	pin-fin in-line
4-port CH	4-port	microchannel
4-port PF	4-port	pin-fin in-line

Table 4.7: Pyramid chip stack test vehicles.

4.3.1 Field-Coupled Porous Media Model Representing Pyramid-Chip-Stack

The pyramid-chip-stack is modeled with all three tiers represented by the silicon slab, the wiring layers and the power map imposed at the contact surface between these two materials. The four cavities are represented with four quasi two-dimensional porous domains (only two nodes normal to the cavity surface) considering the solid and fluid phase. Heat spreading to other dies within the stack is performed through this solid phase representing the fin walls or the pins in the cavity, defined by its area fill factor and bulk thermal conductivity (which was set to be silicon, since no TSVs were present). The heat

PARAMETER	VALUE
Heat transfer area	10 x 10 mm^2
Heat transfer geometry:	
- channel / pin pitch	100 µm
- cavity height	100 µm
- channel wall width / pin diameter	50 µm
Hot-spot area per tier	4 times 10 mm^2

Table 4.8: Pyramid-chip-stack test vehicle dimensions.

transfer into the fluid is performed through field-coupling between the die and the fluid phase of the porous media as described previously (subsection 2.2.4). The orthotropic and the sinusoidal description of the permeability and the convective thermal resistance at developed boundary layer in case of microchannels and pin-fins at fully population was used. The modeling concept was implemented on a commercially available computational fluid dynamic platform (CFX V12, ANSYS) (Figure 4.29, details of the implementation can be found in the appendix B). The hexahedral dominant mesh representing all three tiers and four cavities with 10,000 pin-fins each consists of 32,368 nodes. A momentum-loss term is added to the NAVIER–STOKES equation as described in 2.19 to account for the porous nature of the cavity domain. The partial differential equations were solved with at a root mean square convergence criterion of 10^{-6}.

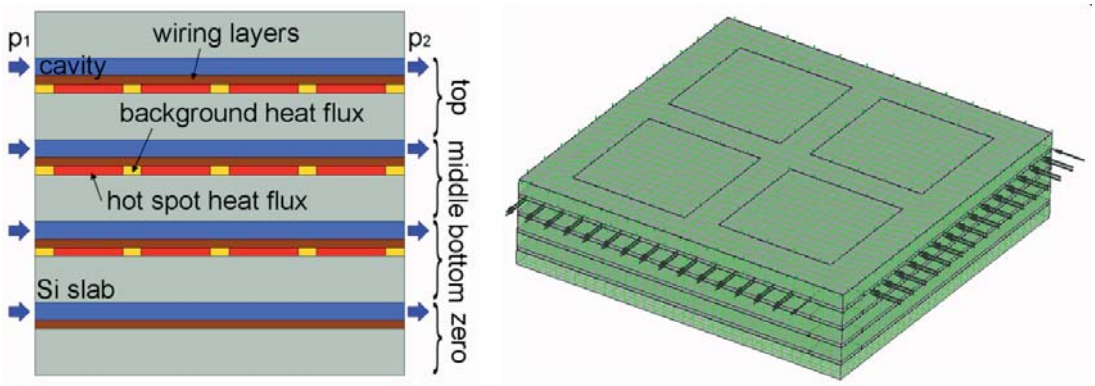


Figure 4.29: Cross-section of a two-port model presenting all implemented layers (left). Four-port model with indicated boundary conditions created with the ANSYS CFX pre-processor. The green lines represent the mesh (right).

4.3.2 Validation of Multi-Scale Modeing Concept with Experimental Results

A benchmark operating point was defined to compare the test vehicle performance at an applied pressure drop $\triangle p$ of 1 bar reasonable for server applications, fluid inlet temperatures T_{in} of 20 °C and a hot-spot power (PHS) of 12 W being the upper limit of reliable operation.

Mass Transfer Performance

Pressure drop measurements are presented in Figure 4.30 (left). The pin-fin permeability is higher compared to the microchannel as predicted from sub-domain modeling and results therefore in lower pressure drops at these low REYNOLDS numbers (< 124). Interestingly, the resulting flow rates of the two- and the four-port case nearly coincide for a given structure. This can be expected for the microchannels at laminar flow with a linear mass flow pressure drop relation, since the average channel length is the same. Non-linear behavior was detected for the pin-fin in four-port mode, where the fluid flow orientation from inlet to outlet is a smooth transition from in-line to staggered to in-line flow (Figure 4.30 center). The staggered flow is responsible for the non-linearity as derived from sub-domain modeling. The numerical results for the two-port PF test vehicle nicely represent the experiment. The deviation in case of the 4-port PF is −23 % compared to the experiment.

The velocity at the inlet and outlet and along the diagonal of the four-port are plotted in Figure 4.30 (right). The inlet velocity is increasing hyperbolically from less than 1 $\frac{\mathrm{m}}{\mathrm{s}}$ at the left half of the inlet and reaches a maximal value of 5 $\frac{\mathrm{m}}{\mathrm{s}}$ at the outer right position due to the linear reduction in streamline length. The velocity drops even to zero in the lower left corner in case of pin-fins. This stagnation point would also exist in a full four quadrant four-port, due to symmetry reasons. At this point hot-spots are problematic and have to rely on heat spreading. The velocities of the four-port populated microchannels are in general smaller, but do not drop to zero in the central symmetry point, due to fluid guiding of the channels.

Heat Transfer Performance

A test case with a random power map was computed to demonstrate the temperature response in the two-port CH device. Figure 4.31 (left) presents the result at benchmark conditions ($\triangle p$=1 bar, T_{in}=20 °C, P_{HS}=12 W) with hot-spots (HS) top HS 2, 3 / middle HS 1, 4 / bottom HS 2, 4 active (nomenclature according to Figure 4.28, 3.19). The non-uniform junction temperature response and the heat spreading are visible. The sensible heat accumulation of the fluid can also be noticed. The temperature normal to the cavity plane at the center of hot-spot two is plotted in Figure 4.31 (center) to identify the individual temperature gradients in the chip stack. As expected, the largest gradients are caused by the poor thermal conductivity of the polyimide layer (0.2 $\frac{\mathrm{W}}{\mathrm{m\,K}}$) and the convective heat transfer from the solid to the fluid.

The modeled hot-spot temperature defined as the average temperature along the sensor (T_{jla}) is compared with the measured hot-spot junction temperature (T_{HS}) to validate the solid-to-solid and solid-to-fluid field-coupling approach. The fluid temperature increase and the temperatures at the hot-spots are captured correctly. The predictions are conservative with a deviation ranging from zero to 21 % (Figure 4.31 right). The origin of this difference is a superposition of mainly three effects. First: an estimated 3.7 % of the total hot-spot power is dissipated in the lead wires. Second: a central gap in the hot-spot heater design of 200 µm width serving for the thermal probe placement interrupts the uniform power dissipation. This discontinuity in heat flux locally reduces the junction temperature. Third: the polyimide thickness is considered to be 4 µm. This is the case between the bonding areas were heat is dissipated through the polyimide into the fluid. However, the polyimide bond line thickness between heat transfer structure top and silicon slab is 3.2 µm thick. This results in an improved thermal coupling between the slab and the pin or channel wall. Without this parasitic effect in the experiment the estimated deviation would be $\pm$ 10 % which seems reasonable for device performance investigations and predictions. Further, the junction temperature T_j in the flow direction and in the center of the chip is plotted on Figure 4.31 (right) for each tier. The hot-spot contrast is still strong (remember hot-spot width of 2 mm) even with improved heat spreading capability due to the 425 µm silicon slab thickness. The validated modeling concept will be used in chapter 5 to derive interlayer cooling design-rules for electro-thermal co-designs.

4.3.3 Tier-to-Tie Thermal Crosstalk

The two-port test vehicles were operated at different regular hot-spot patterns and varying pressure drops to analyze the characteristics of the microchannel and the pin-fin structures. The maximal hot-spot temperatures are reported in Figure 4.32 (left). The hot-spot temperature of the CH devise is equal compared to the PF in case of a single active hot-spot despite its lower flow rate and higher convective thermal resistance. The reason is a stronger thermal coupling between tiers caused by its low porosity. It results in efficient heat distribution between the four cavities. This experimental finding was also confirmed by the model, with a constant offset of about 20 %. Heat spreading becomes asymmetric, if all hot-spots on the top layer are powered. In this case the spreading benefit of the CH is limited. The power dissipation pattern is quasi periodic with minimal heat spreading between cavities upon activation of all three hot-spots at position 2 (HS2) or the activation all hot-spots in the stack. The improved convective heat transfer and increased permeability of the PF results in lower stack temperatures in this modes.

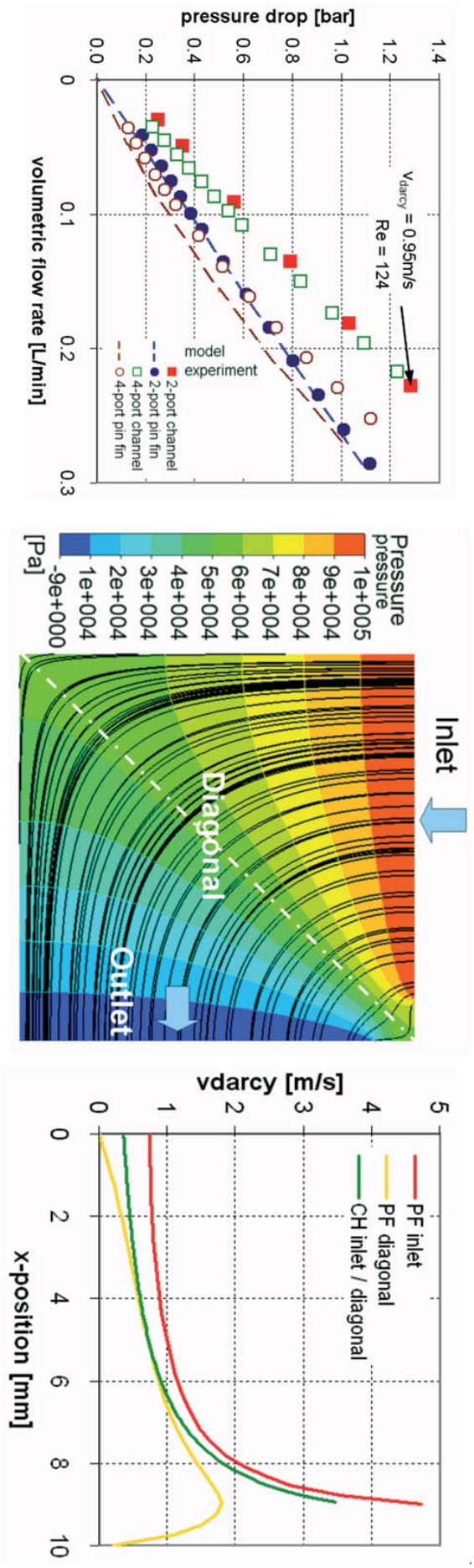


Figure 4.30: Comparison of pressure drop performance of all test vehicles measured. Dots representing experimental values, dashed lines porous-media model results (left). Pressure field and streamlines of the four-port PF device at 1 bar inlet pressure (center). Superficial velocity at the in- / outlet and the diagonal of the four-port PF and CH device at 1 bar inlet pressure are depicted (right).

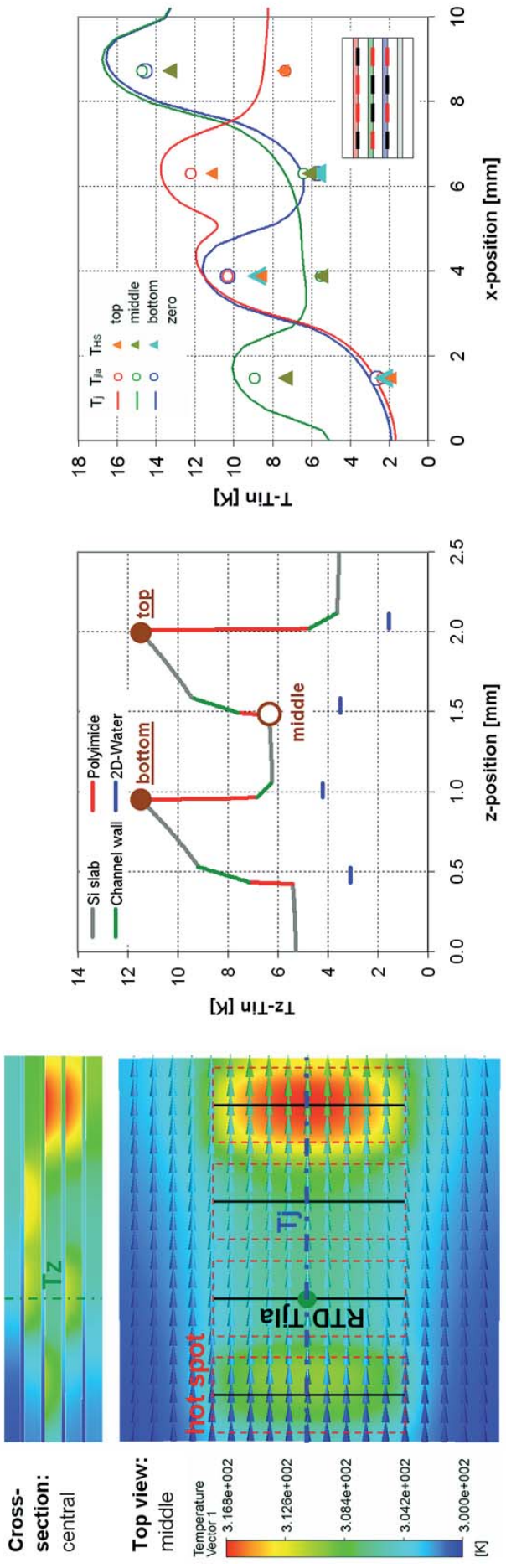


Figure 4.31: Central cross-section and top view to the middle junction of the random-powered test case showing the temperature map. Arrows represent the fluid flow direction with colors indicating the fluid temperature. Hot-spot areas are marked (red dashed squares) as well as the thermal probe location (black lines). Further model junction temperatures are depicted either from the center line T_j (blue, dash-dotted) or as the line average temperature T_{jla} representing the measured hot-spot temperature T_{HS} (left). Temperature T_z through the stack compared to the fluid inlet temperature T_{in} at the center of hot-spot 2, normal to the cavity plane is plotted. Filled brown dots indicate dissipating, empty dots represent inactive hot-spots. The line colors refer to individual materials in the chip stack (center). Measured T_{HS} (triangles) and modeled hot-spot temperature T_{jla} (circles) are presented together with the central junction temperature T_j (lines) of the two-port CH test vehicle at benchmark boundary condition and random hot-spot pattern: top HS 2, 3 / middle HS 1, 4 / bottom HS 2, 4 (right).

4.3.4 Benchmarking of Fluid Delivery Architecture: Two-port vs. Four-port

The strength of four-port fluid delivery is well demonstrated at benchmark operation and four active HSs on the top tier (Figure 4.32 right). In four-port flow only hot-spots 2, 3, 4 are thermally coupled through the fluids temperature, but not HS1. Furthermore, the coolant velocity at HS1 is highest (Figure 4.30 right). These are the reasons for the low temperature at HS1 and HS2 in case of the four-port. The temperature increase from HS2 to HS3 is most dominant due to the dramatic velocity drop towards the lower left corner (stagnation point). It is less pronounced for the CH device since the velocity does not drop to zero. Important to notice is the fact, that the four-port test demonstrates the cooling performance of a 4 cm^2 compared to a 1 cm^2 chip stack in case of the two-port.

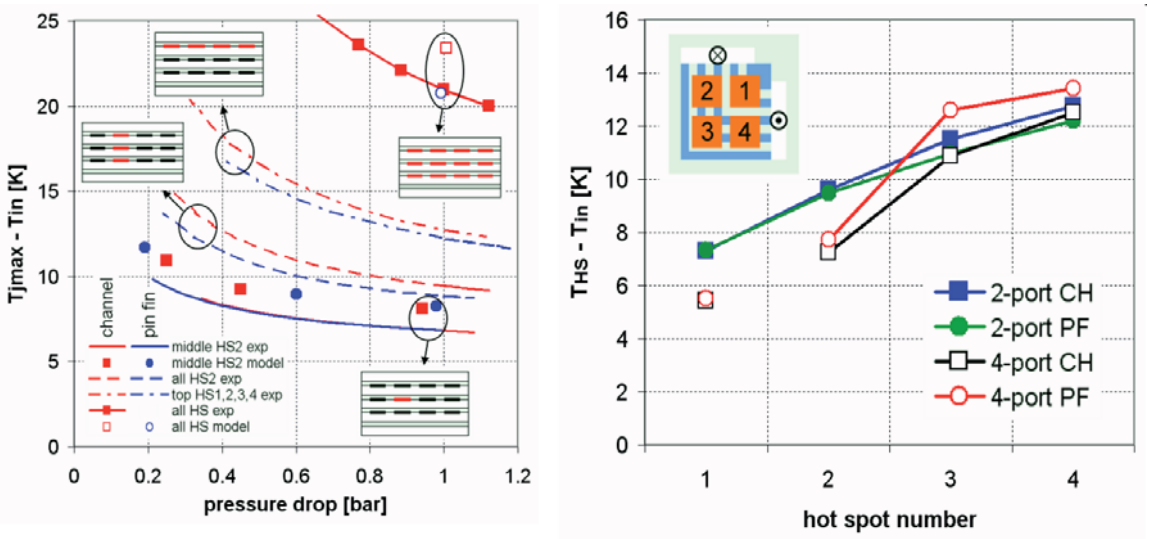


Figure 4.32: Experimental T_{HS} and modeled T_{jla} maximal hot-spot temperatures of a two-port CH and PF device at different power maps versus applied pressure drop. The deviation between model and experiment is ~20 % (left). Experimental hot-spot temperature comparison for two- and four-port fluid delivery and PF / CH heat transfer structures at benchmark operation and all top HS operational (right).

5 Interlayer Cooling Design-Rules, Conclusions and Outlook

It is now possible, with the experimental findings and the validation of the modeling concept, to predict the heat removal performance of interlayer cooling at realistic conditions, not compromising on test vehicle aspects. As a result, a benchmarking study with respect to back-side cooling is performed in the first section. Additionally, aspects such as flow maldistribution in the individual cavities, in-plane heat spreading efficiency, and temperature time constants are presented. The preferential use of individual unit-cell shapes and the pin-fin heat removal performance is analyzed in a scaling study. Recommendations on how and when to use individual heat transfer building blocks to enhance interlayer cooling performance, especially at non-uniform power dissipation, are listed.. Furthermore, opportunities for interlayer cooling aware floorplanning are discussed. Finally, future research tasks and opportunities are proposed.

5.1 Interlayer Cooled Chip Stacks: Performance and Characteristics

5.1.1 Benchmark: Uniform, Two-port Interlayer Cooling vs. Back-Side Heat Removal

A chip stack as reported in section 4.3 with a $1\,cm^2$ footprint, three active tiers including four hot-spots with an area of $10\,mm^2$ each and a TSV pitch of $100\,\mu m$ was considered for the interlayer vs. back-side cooling benchmark study. Realistic parameters such as a silicon slab thickness of $50\,\mu m$ and wiring layer thermal resistance of $7\,\frac{K\,mm^2}{W}$ (Figure 4.28 right) were considered. The aligned hot-spots are operated at a power level of $25\,W$ each, resulting in a heat flux of $250\,\frac{W}{cm^2}$, which is a typical value in core areas of high performance microprocessors. Furthermore, a heat flux of $50\,\frac{W}{cm^2}$ was imposed on the residual chip surface, representing the background power dissipation of the cache area. In total, $390\,W$ are dissipated on a $1\,cm^2$ footprint, corresponding to an average volumetric heat flow of $3.9\,\frac{kW}{cm^3}$ if a $1\,mm$ stack height is considered. Two-port fluid delivery at a $1\,bar$ pressure drop and pin-fins in in-line orientation with a height and diameter of $100\,\mu m$ and $50\,\mu m$ where considered, respectively.

The maximal junction temperature is reached in the middle tier towards the outlet in case of the interlayer cooled chips stack (Figure 5.1 left, dark green line). This tier has to share the two adjacent cavities with the top and bottom tier due to its central location. The values are well within typical temperature margins of $65\,K$. Interestingly, the top junction temperature is lower than the junction of the bottom tier. This is astonishing, since the bottom junction is more efficiently coupled to its own fluid cavity (zero), compared to the upper tier. This tier has to dissipate the heat through the low conducting wiring levels to its top cavity. The reason is the resulting asymmetry in heat flux. The fluid temperature in cavity zero is increasing more rapidly compared to the top cavity, indicating a heat flux crowding from the upper layers in the bottom section.

The junction temperature maximum of the pyramid test vehicle with its larger silicon thickness and wiring layer resistance (section 4.3, Figure 5.1 left, light green line) is comparable to the realistic product temperature. Enhanced heat spreading in the thicker silicon slab helps to mitigate hot-spot effects and suppresses maximum junction temperatures. This compensates for the increased temperature drop across the low conductive polyimide layer. The hot-spot contrast is much more dominant in the product example. An analysis of the thermal gradient ratio induced by thermal conduction and convection compared to the fluid temperature increase indicates again the significance of a low friction factor cavity design. This fact will be further accentuated at smaller interconnect pitches with cavities of reduced permeability due to reduced hydraulic diameters.

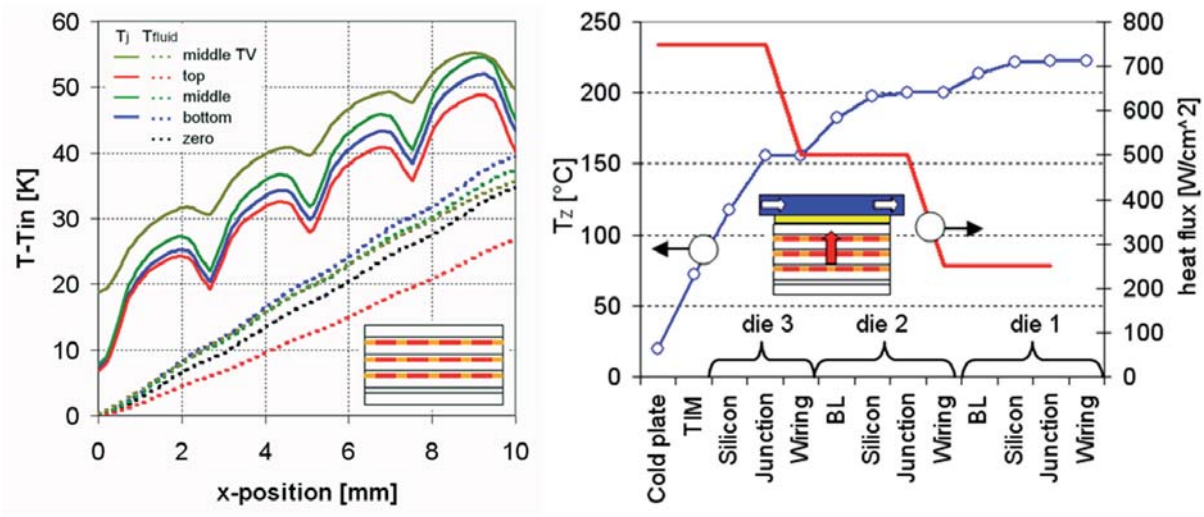


Figure 5.1: Benchmark study of the thermal response predicted for a product style chip stack with three active tiers, a $1\,\mathrm{cm}^2$ footprint and a $100\,\mu\mathrm{m}$ TSV pitch. A heat flux of $250\,\frac{\mathrm{W}}{\mathrm{cm}^2}$ and $50\,\frac{\mathrm{W}}{\mathrm{cm}^2}$ was assumed for core and cache areas respectively (insets, dark red: core, orange: cache). Three cavities populated with pin-fins are opperated at a $1\,\mathrm{bar}$ pressure drop for the interlayer cooled case. The junction (full lines) and fluid (dotted lines) temperatures from inlet to outlet are depicted relative to the fluid inlet temperature (blue, green, red: bottom, middle, top die). The middle tier junction temperature of the corresponding pyramid-chip-stack test vehicle with two-port PF is plotted in comparison (left). The heat flux and maximal junction temperature in the chip stack normal to its surface and in the center of the hot-spots is depicted, considering back-side cooling (right). A thermal resistance of 7 and $6\,\frac{\mathrm{K\,mm}^2}{\mathrm{W}}$ were assumed for the microchannel cold plate and thermal interface resistance. An unacceptable maximal junction temperature of $223\,^{\circ}\mathrm{C}$ is reached in the bottom tier (right).

The temperature response of the product style chip stack with a back-side cold plate and equal power dissipation, as the interlayer cooling case, is presented in Figure 5.1 (right). A high-performance microchannel coldplate and thermal interface is assumed with a thermal resistance of 7 and $6\,\frac{\mathrm{K\,mm}^2}{\mathrm{W}}$, respectively. A maximum junction temperature of $223\,^{\circ}\mathrm{C}$, which exceeds the $85\,^{\circ}\mathrm{C}$ limit by far, is reached in the bottom tier at hot-spot locations at a fluid inlet temperature of $20\,^{\circ}\mathrm{C}$. This is mainly caused by heat flux accumulation and demonstrates clearly the need of interlayer cooling for future high-performance microprocessor chip stacks.

5.1.2 Fluid Manifold Design: Pressure Penalty and Mass Flow Maldistribution

The manifold has to distribute the coolant from the copper tubing to the individual fluid cavities. Design-rules are needed to minimize coolant maldistribution between individual parallel arranged cavities. Therefore, a two dimensional mass transport model was performed, representing a interlayer cooled chip stack with two-port fluid delivery. The manifold structure was defined as a laminar flow fluid domain justified by its REYNOLDS number of 400 at $1.0\,\frac{\mathrm{m}}{\mathrm{s}}$ fluid flow. It is pressure coupled to the four porous domains, representing the microchannel cavities with $50\,\mu\mathrm{m}$ fin width, $100\,\mu\mathrm{m}$ height and pitch (Figure 5.2 left). Realistic velocity boundary conditions of 0.25 and $1.0\,\frac{\mathrm{m}}{\mathrm{s}}$ where imposed at the inlet of the manifold. The relative difference between the resulting maximal and minimal DARCY velocity in the cavities is plotted in Figure 5.2 (right, blue) for a 1 mm high, but 1.0 or 0.5 mm wide manifold structure (dark, light blue). In both cases, the flow non-uniformity induced by the manifold design is less than 1.1 %. This can be explained by the low permeability of the fluid cavities, which account for more than 97.5 % of the total pressure drop from manifold in- to outlet (Figure 5.2 right, green). In comparison, the flow maldistribution induced by a potential pin diameter manufacturing variation of $1\,\mu\mathrm{m}$ would be 4 % between cavities (Figure 2.18 left, blue) and thus much more significant.

The pressure of the fluid cavity with respect to the total pressure drop accounts still 92 % even for a $100\,\mu\mathrm{m}$ high parallel plate cavity, with maximal permeability and four-port fluid delivery (Figure 5.3 left). The pressure along the cavity in- and outlet is still close to isobar, even for such extreme cases, with

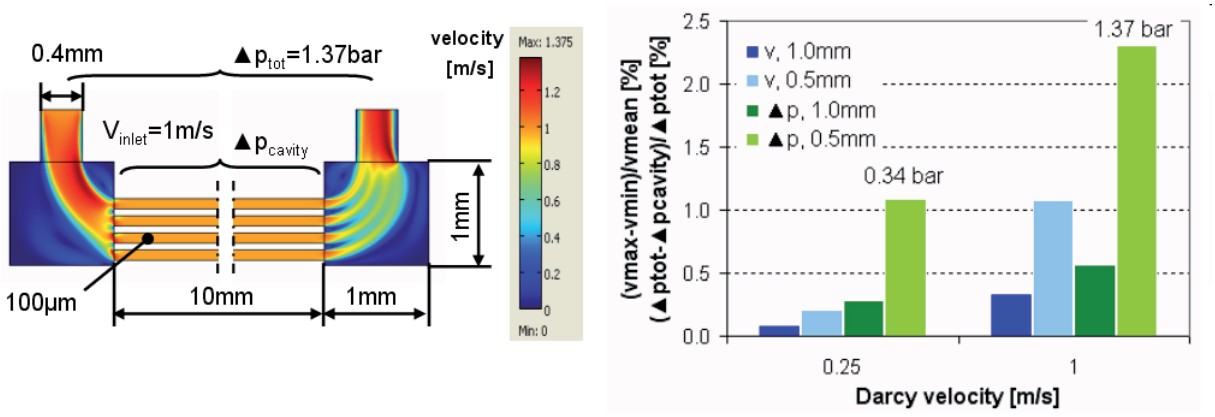


Figure 5.2: Predicted velocity field in a four-cavity chip stack including the laminar flow in the manifold structures. They are pressure coupled to the porous-domains representing the fluid cavities filled with microchannels at 50 µm fin width and 100 µm height and pitch (left). The resulting flow rate non-uniformity (maldistribution) induced by the manifold width is marginal (< 1.1 % at the relevant flow rates) due to the low permeability of the microchannels, responsible for more than 97.5% of the total pressure drop (right).

highly non-uniform fluid velocities in the cavity. Accordingly, the assumption of pressure boundary conditions at the porous-media periphery, as used in section 4.2.2 is justified.

5.1.3 Critical Hot-Spot Dimension: Hot-Spot Temperature Mitigation by Heat Spreading in Thin Dies

Heat spreading in the silicon slab is a further possibility to mitigate hot-spots in interlayer cooled chip stacks. This is efficient for high heat flux areas with small lateral dimensions, depending on the convective heat transfer and silicon slab thickness. The maximal conductive and convective junction temperature gradient resulting from a hot-spot heater with varying width was assessed relative to a uniform, but equal heat flux (Figure 5.3 right).

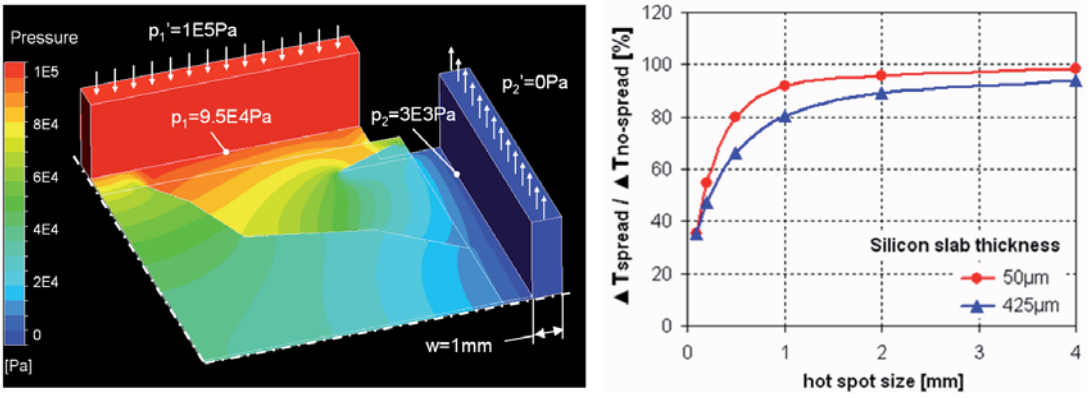


Figure 5.3: Modeled pressure field in manifold and parallel plate fluid cavity (100 µm height) with guiding structures at 1 bar pressure drop (left). Ratio of hot-spot temperature gradient caused by conduction and convection including in-plane heat spreading compared to a one-dimensional heat flux, for a 425 µm (test vehicle) and 50 µm (product) thick silicon slab at benchmark conditions (right).

A DARCY velocity of 1 $\frac{\mathrm{m}}{\mathrm{s}}$ and microchannel heat transfer structures with 50 µm fin width and 100 µm height and pitch were assumed. At a 425 µm thick silicon slab, present in the pyramid-chip-stack test vehicle, hot-spot mitigation becomes effective for hot-spot dimensions < 2 mm and is therefore called the critical hot-spot dimension. Heat spreading is less efficient in realistic chip stacks, with implemented TSVs, resulting in maximal silicon slab thickness of 50 µm. Accordingly, hot-spot mitigation becomes efficient only at hot-spot dimensions < 1 mm. The sum of the conductive and convective temperature gradient drops by a factor of two in case of 0.2 mm wide hot-spots, compared to 4 mm wide hot-spots.

Accordingly, high heat flux macros should be divided into small areas (below the critical hot-spot dimension) and arranged at an appropriate spacing, to increase the possible computing performance.

The critical hot-spot dimension is also the parameter, defining the minimal spatial power map resolution to be considered in the models, to derive all junction temperature details. The resulting junction temperature field stays close to constant at higher spatial power map resolutions, due to the heat spreading effect. At lower spatial power map resolutions a blurred junction temperature is the result (equivalent to low pass filtering) and does not include all the high spatial-frequency temperature changes.

5.1.4 Thermal Transients: Time Constant and Retardation

Time domain modulation of the air mass flow is used in today's system to optimize the fan power consumption. The rotation speed is reduced if the junction temperature drops below a critical temperate at low utilization of the microprocessors. This is especially efficient in future systems with implemented core gating. The shut-down of individual cores results in a strong power dissipation contrast between peak and idle load. A prerequisite of mass flow modulation is a faster response of the actuator element (e.g. fan or pump) compared to the thermal time constant of the junction. It is in the range of seconds for air cooled systems, with a high thermal capacity and resistance of the air heat sink and thermal interfaces present. Commercial fans and pumps can meet this dynamic requirement easily.

The thermal mass between fluid and junction is reduced substantially to a 50 µm silicon slab only, for interlayer cooling. A transient conjugate heat and mass transfer model was performed to derive the thermal time constants present in a chip stack, with three active tiers and four heat removing microchannel cavities at a length of 10 mm. The channel wall was 50 µm, the channel height and pitch 100 µm. A uniform power dissipation on all tiers of 50 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ was used as an initial condition with a coolant inlet temperature of 300 K and a bulk velocity of 1.6 $\frac{\mathrm{m}}{\mathrm{s}}$. The heat flux of one hot-spot heater with a width of 2 mm located on tier two at the inlet, middle, or outlet position was increased to 250 $\frac{\mathrm{W}}{\mathrm{cm}^2}$ at time zero. The junction temperature response was monitored on tier two at three locations (inlet, middle, outlet), on tier three at the inlet and in the fluid of cavity two at the outlet position (Figure 5.4). A local time constant τ of 1.6, 2.5, and 3.0 ms was measured at the left edge, in the center, and the right edge of the hot-spot heater at the cavity inlet, center (middle), and the outlet, respectively (inset Figure 5.4 right, dots). Advection is responsible for the retardation in temperature increase and results in increasing time constants, if measured at the hot-spot left edge, center, or right edge. It takes 0.6 ms for the coolant to travel 1 mm in distance at a bulk velocity of 1.6 $\frac{\mathrm{m}}{\mathrm{s}}$. This could explain the main part of the of the 0.9 and 1.4 ms difference in time constant at the three locations. The residual 0.3 and 0.2 ms could be attributed to the developing boundary layers with a resulting convective thermal resistance increase downstream from the fluid inlet.

Stack temperature controlled pump modulation is not applicable for interlayer cooling at a three order of magnitude reduced junction temperature thermal time constant. This due to the slow time response of pumps with their high impeller mass. The operating system would have to predict heat removal needs and has to change the pump speed with a time offset equivalent to the pump time constant prior to an increase in microprocessor activity.

Multiple studies discuss the advantage of dynamic work load allocation protocols to cap the maximal junction temperature by dynamic task allocation at frequencies above the inverse of the thermal time constant of the junction temperature. This method takes advantage of the heat capacity of the system [124]. The thread re-allocation comes with a computational overhead and is therefore only efficient up to certain frequencies. Accordingly, the effectiveness of dynamic work load allocation is reduced in case of interlayer vs. back-side cooling, with its low time thermal time constant. Additionally, advanced protocols have to be implemented accounting for retarded junction temperature increases caused by the advection as shown in Figure 5.4 (left and center). A 6 and 3 ms delay between temperature at the active hot-spot located at the inlet or middle position and the outlet junction temperature was identified.

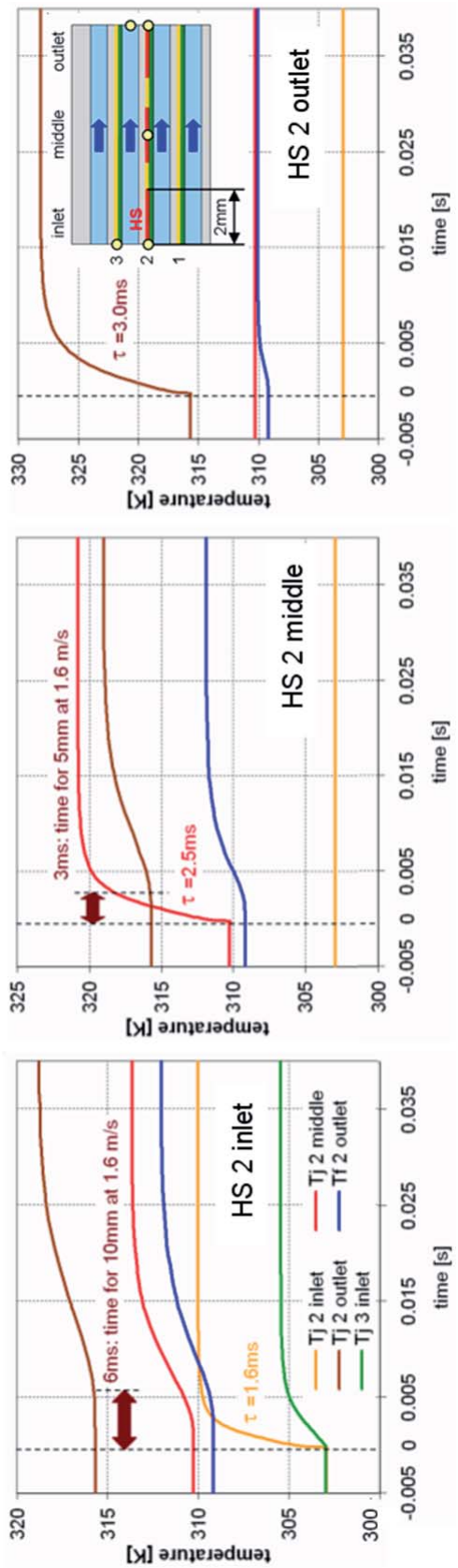


Figure 5.4: Temporal junction and fluid temperature response after applying a hot-spot heat flux of 250 $\frac{W}{cm^2}$ on the second tier at inlet, middle and outlet position (left, middle, right). Initially a uniform power dissipation on all three tiers of 50 $\frac{W}{cm^2}$ was applied. The four microchannel cavities with a length of 10 mm are operated at a pressure drop of 1 bar, resulting in a fluid bulk velocity of 1.6 $\frac{m}{s}$ with the resulting 6 ms fluid dwell time, from inlet to outlet. The four fluid cavities and three tiers, including hot-spot locations and sensor positions are shown in the inset (right).

5.2 Advanced Interlayer Cooling: Design-Rules and Recommendations

5.2.1 Unit-Cell Shape Menu and Interlayer Cooling Performance Scaling at Uniform Heat Flux

Recommendations on the selection of heat transfer unit-cell geometries can be given based on the experimental investigations at uniform heat flux (section 4.1). Pearl chain (PC) like geometries performed best with respect to pressure drop due to the increased effective hydraulic diameter compared to microchannels and the absence of vortex shedding at high REYNOLDS numbers, compared to pin-fin in-line unit-cells.

PERFORMANCE EVALUATION CRITERION	BEST PERFORMING DEVICE	USE
Pressure drop	- Perl Chain with largest hydraulic diameter and suppression of vortex shedding	- Area with low heat flux
Thermal resistance at given velocity	- Pin-fin staggered with maximal fluid mixing	- At hot-spot locations
Pressure drop need for given heat flux	- Pin-fin staggered distorted cell at non-uniform x-y pitch, pin-fin in-line with uniform x-y pitch, both devices leverage on vortex sheddingin the transition regime	- At uniform power maps
Pumping power need for given heat flux	- Pin-fin staggered distorted cell at non-uniform x-y pitch, maximal hydraulic diameter, followed by pin-fin in-line and pin-fin staggered	- At uniform power maps

Table 5.1: Best-performing heat transfer structure at a criterion of interest.

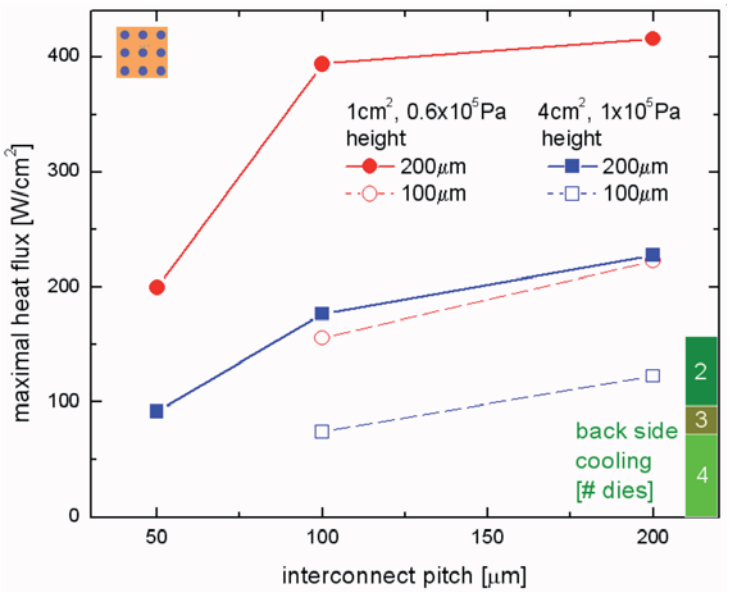


Figure 5.5: Maximal heat flux scaling at uniform power dissipation and the TSV pitch as an independent variable, assuming an infinite number of tiers and fluid cavities (results from the single-cavity experiment). The pin diameter of the PFI heat transfer structure is defined as half the TSV pitch. Cavity heights of 100 and 200 µm and stack footprints of 1 and 4 cm^2 at 0.6×10^5 Pa and 1×10^5 Pa are considered, respectively. The corresponding maximal heat flux considering back-side heat removal with 2, 3, and 4 tiers is compared (green bars).

Minimal thermal resistance at a given fluid velocity is achieved with pin-fins in staggered (PFS) orientation, due to maximal fluid mixing. The best compromise between mass transport and heat transfer at a given pressure or pumping power boundary condition is performed by the pin-fin in-line (PFI) arrangement, if uniform x-y TSV pitch is needed (Table 5.1). The flow transition caused by vortex shedding can also be used to reduce the junction temperature towards the cavity outlet. Maximal heat flux scaling for PFI heat transfer structures at changing interconnect pitches is presented in Figure 5.5, considering uniform power dissipation, representing devices with a low power map contrast. At a chip size of 1 cm^2 and a $\triangle T_{jmax-in}$ of 60 K, the heat-removal performance is $>200\ \frac{W}{cm^2}$ at pitches > 50 µm and a 0.6×10^5 Pa pressure drop to drive to coolant. The performance degrades significantly for a 4 cm^2 chip stack. Only TSVs with 200 and 100 µm pitch and h_c= 100 and 200 µm are able to remove heat fluxes of $>175\ \frac{W}{cm^2}$ at a pressure drop of 1×10^5 Pa. Additionally, the maximal heat flux to be dissipated at back-side heat removal for 2, 3, and 4 tiers is compared (green bars). The scalable interlayer cooling concept, which is independent of the strata number, outperforms the backside heat-removal approach for two or more CPU dies, but is limited to interconnect pitches > 50 µm at uniform heat removal.

5.2.2 Power Map aware Heat Removal: Heat Transfer Structure Modulation - Fluid Focusing - Fluid Delivery Architectures - Tier-to-Tier Crosstalk

Non-uniform heat transfer structures results in a more efficient use of the precious heat capacity of the water, especially at high heat flux contrast power dissipation, typically present on microprocessors. Heat transfer structure modulation (fully and half populated pin-fin domains) have proven to improve the heat removal performance by more than 50 % and extends the interlayer concept to 50 µm TSV pitch (Figure 4.14 right). Fluid focusing using fluid guiding structures together with pin-fin arrays has demonstrated to lead to improved balancing of hot-spot and background heater peak temperatures (Figure 4.24). Nevertheless, a careful design is important not to constrain the overall flow rate too much. Four-port fluid delivery is the architecture of choice for large chip stack footprints, with high heat fluxes imposed in the microprocessor corners. The resulting mass flow rate is four times improved and $600\ \frac{W}{cm^2}$ can be dissipated in the die corner. This at a 100 µm TSV pitch and cavity height of a 4 cm^2 die, instead of $60\ \frac{W}{cm^2}$ for the uniform two-port case. The use of pin-fins in four-port architectures deserves special care due to the stagnation point in the cavity center. Either, low central heat fluxes are imposed or additional fluid guiding structures need to be implemented. Otherwise, microchannels should be used for simplicity and result in a moderate flow even in the cavity center. Thermal crosstalk between cavities distributes the heat flux and mitigates hot-spot temperatures. This is true if hot-spots are not aligned between individual tiers (non-periodic arrangement). Microchannels have shown an improved tier-to-tier thermal coupling due to the increased silicon fill fraction compared to pin-fin arrays and are the heat transfer structure of choice, if dies with disparate power maps are stacked.

Method	Specialized For	Design Guidelines	Power Map
Modulation	Increased flow rate	- Eliminate pins if not needed for communication - Fluid might bypass fully populated zone - need for guiding structures	Any
Focusing	Hot-spot mitigation	- Careful design - guiding structures reduce overall flow rate	High heat flux contrast
Four-port	Increased flow rate	- Use of microchannels for marginal flow in cavity center (Pin-fins result in a stagnation point)	Hot-spot in die corner
Tier-to-tier crosstalk	Hot-spot mitigation	- Microchannels with large fill factor	Misaligned hot-spots on different tiers

Table 5.2: Recommendations on the use for building blocks for advanced interlayer cooling.

5.2.3 Electro-Thermal Co-Design: Interlayer Cooling Aware Floorplanning

A careful placement of high power dissipating macros should be considered during floor-planning, to reduce thermal excursions in interlayer cooled chip stacks. Temperatures of different hot-spot patterns were calculated for the realistic product like chip stack at operation conditions reported in Figure 5.1 (left). The maximal junction temperature with respect to the fluid inlet temperature and the total dissipated power is reported in Figure 5.6 (left). If four hot-spots have to be activated on one tier, they should be placed in the middle or top level of the chip stack. This due to the asymmetry of the tiers, with the low thermal resistance silicon slab on one side of the junction, compared to the high thermal resistance wiring-layers on the other side of it. If possible, the four hot-spots should be distributed randomly within the chip stack, to even further reduce the maximal junction temperature. Three hot-spots aligned on top of each other should be placed at the fluid inlet. The junction temperature is monotonically increasing if there location is moving from inlet to the outlet.

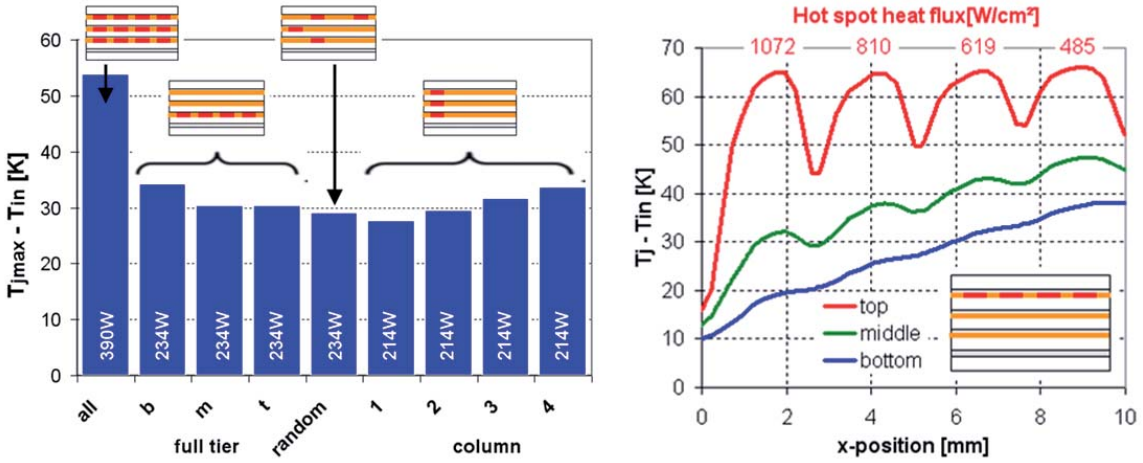


Figure 5.6: Maximal junction temperature with respect to the fluid inlet temperature of the product like chip stack reported in Figure 5.1 (left), operated at different hot-spot patterns. The white number represents the total dissipated power. The dark red lines indicate the activated hot-spots (left). Junction temperature response in case of four activated hot-spots on the top tier at benchmark condition is depicted. The hot-spot heat flux levels are individually adjusted to utilize the 65 K thermal budget, resulting in highest possible computational performance for the given hot-spot pattern (right).

Each hot-spot heat flux level should be adjusted to fully utilize the thermal budget of 65 K to maximize the computational performance of a chip stack. Accordingly, the heat flux levels of the hot-spots in the top tier were set to 1072, 810, 619 and 485 $\frac{W}{cm^2}$ and resulted in the maximal power dissipation of 440 W (Figure 5.6 right).

The maximal possible power dissipation for other hot-spot patterns, considering individually adjusted hot-spot heaters are listed in Table 5.3. Maximal aggregate power dissipation of 562 W is reached if all the hot-spots are activated. Maximal heat flux of 1081 $\frac{W}{cm^2}$ is acceptable if a single hot-spot is activated on the middle tier at location one, close to the fluid inlet.

Activated Hot Spots: Location And Number	Total Dissipated Power [W]	Peak Heat Flux [$\frac{W}{cm^2}$]
All, 12	562	746
Top tier all, 4	440	1072
Each tier first, 3	331	757
Middle tier second, 1	259	1081

Table 5.3: Chip stack performance for different hot-spot patterns and individually optimized hot-spot heat flux levels.

5.2.4 Interlayer Cooling Performance Evolution

First interlayer cooling packages will rely on simple and uniform heat transfer cavities, such as microchannels or pin-fins to mitigate risk during design and the later implementation. This early implementation step is highly important to learn about yield and reliability of such packages, fabricated in high-volume production lines. Additional elements, such as microchannel width modulation, hot-spot optimized fluid-networks, or the four-port architecture can be implemented later, as heat fluxes and the stack footprint are increasing. Further improvement demands for a electro-thermal co-design. The proposed thermal modeling framework needs to be coupled with the electrical design tools, resulting in optimized thermal-aware floorplanns. This means, hot-spots on different layers are misaligned, macros with high power dissipation are splitt into small submakros or are placed close to the fluid inlet. Additionally, temporal work-load allocation strategies can be added to this electrical design-rules. Finally, the limited mass flow rate could be improved by the implementation of superhydrophobic surfaces at low heat flux locations. It would result in an effective slip-length to mitigate the pressure drop in the cavity (Figure 5.7). A visualization of a four-port fluid delivery chip stack with non-uniform heat removal is presented in Figure 5.8.

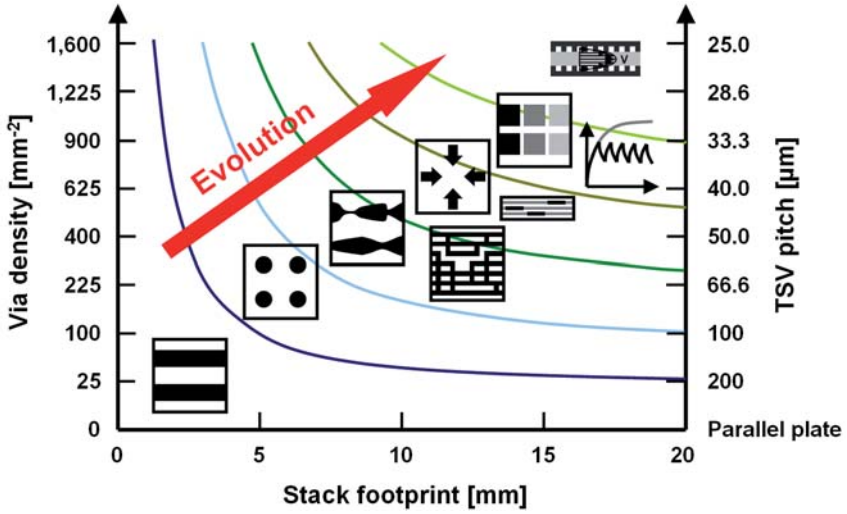


Figure 5.7: Interlayer cooling performance evolution: Hot-spot heat flux contours vs. via density and stack footprint at increasing interlayer cooling sophistication. Levels: i) microchannel, ii) pin fin array, iii) microchannel modulation or fluid network, iv) four-port fluid delivery architecture or misaligned hot-spots, v) high heat flux macros close to manifold inlet or temporal work-load allocation strategies, vi) slip-flow by superhydrophobic surfaces.

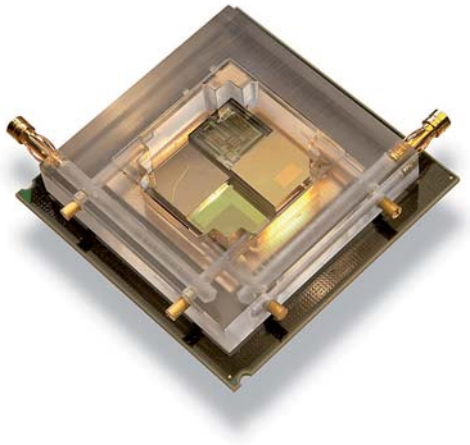


Figure 5.8: Photograph of a demonstrator visualizing an interlayer cooled chip stack with four-port fluid delivery and non-uniform heat removal cavities (dies are partially removed, to get visual access to individual layers).

5.3 Future Research Agenda

The focus in this study was to identify and investigate main heat transfer building-blocks for efficient interlayer cooling. Furthermore, a modeling framework was established to be able to efficiently predict the local convective heat transfer and thus, the chip stack temperature field. Additionally, technological grounds were prepared towards a reliable bonding process to perform fluid cavities without electrical shortings between TSVs in presence of water. There are still many aspects which have to be investigated to make interlayer cooling compatible with high-volume manufacturing at a high yield. Therefore, a research agenda is proposed addressing the next critical steps.

5.3.1 Interlayer Cooling Implementation: Reliable Packaging Technology

Patterned thin film bonding: The sealing-ring concept to electrically insulate individual TSVs was introduced in section 3.1 and is displayed in Figure 3.1. Up to this point, all test vehicles were bonded with unpatterned solder, due to the absence of TSVs. Hence, structured solder deposition by electroplating or lift-off sputtering needs to be investigated. The right amount of solder thickness at a large enough clearance between ring and pad structure has to be defined, to maximize the available hydraulic diameter in the cavity at a minimal risk of solder bridging or opens.

Solder Corrosion: The solder sealing ring is partially exposed to the water present in the fluid cavity. Electro-corrosion studies with appropriate corrosion inhibitors have to be performed for relevant solder alloys. Activation energies and corrosion rates need to be extracted from weight loss / gain or cyclic voltammetry measurements to be able to predict the reliability of the bond throughout its lifetime. Au-Sn seams to be a preferred candidate due to its high gold content as discussed in section 3.1.

Cavity to Fluid Loop Interface: The fluid containment and coolant distribution to individual cavities could be performed by a manifold bonded to a silicon interposer as sketched in Figure 1.8 (left). The use of the silicon interposer as a lower fluid containment does prevent water uptake of the underfill and therefore eliminates electro-corrosion effects at the solder balls, compared to a solution where the organic laminate takes the interposer functionality. Furthermore, the copper tubing from the fluid loop can be reliably attached by threads and sealing rings to the in- and outlet ports of the manifold part.

Thermo-Mechanical Integrity: One selection criteria of the manifold material is its thermo-mechanical compatibility with the silicon package on the organic laminate. Accordingly, a low coefficient of thermal expansion (CTE) and / or a low YOUNG's modulus should be targeted to mitigate mechanical stress in the package. Additionally, the manifold needs to be compatible with different processing steps, depending on the electrical interface from module to the main computing board. In case of ball-grid-arrays (BGAs) high-temperature stability up to 250 °C is important due to the additional reflow step. High mechanical loads have to be applied to the package when land-grid-arrays (LGAs) are used. LGA activation forces up to 1 kN demand for an increased manifold stiffness, to uniformly distribute the load to the individual springs of each contact.

High-Frequency Signal Transmission: Signal damping in the BEOL wiring and the TSVs needs to be investigated due to the proximity of water. Its permittivity of 81 and high loss-tangent are extraordinarily with respect to other materials used in an electronic package. Therefore, coaxial TSV designs or shielding concepts using metal layers in the fluid cavity might be important. The measurement could be performed by an odd and even number of ring-oscillators on one and another tier within the die-stack.

5.3.2 Advanced Heat and Mass Transfer

Vortex Shedding Prediction: Improved localized heat transfer through vortex-shedding in uniform pin-fin arrays was observed experimentally as discussed in subsection 4.1.4. This effect could be used to mitigate local hot-spots. Therefore, a detailed understanding of the phenomena has to be gained and a prediction method needs to be proposed. This should can be performed by additional experiments with micro-particle-velocimetry [122] and transient modeling of the fluid flow.

Slip-Flow by Superhydrophobic Surfaces: Several research teams investigated the effective slip-length of fluids on superhydrophobic surfaces [125, 126]. A surface with low surface-free-energy in conjunction with nano to micron-sized topographic results in a low fluid-to-solid contact ratio. More than 99 % of the fluid interfaces to gas trapped in surface cavities. The resulting fluid velocity at the gas-to-fluid interface can be significant due to the low dynamic viscosity of the gas. Slip-lengths larger than 100 µm were measured on surfaces with periodic pin topography at a pitch of several tens of micrometer [127]. The coolant flow rate in a microchannel with 100 µm hydraulic diameter and a superhydrophobic surface with a slip-length of 20 µm would double compared to a hydrophilic surface at a given pressure drop. The fluid penetrates into the gas cavity and wets the surface completely above a critical pressure. Hence, the slip-length is reduced to zero. This threshold depends on the pitch of the surface topology and is defined by the LAPLACE-pressure. Up to this point it is unclear if superhydrophobic surfaces with a pressure stability up to 1×10^5 Pa and a resulting slip-length larger than 20 µm could be designed. Proposed hierarchical topographies with a combination of nano and micron-sized patterns seams to be a promising direction to explore [128].

Energy Re-Use: The aggregate data center power consumption in the US in 2005 accounted for 2 % of the countries electrical power dissipation [129]. The power usage effectiveness (PUE) defined as ratio between the total data center power with respect to the power dissipated by the IT-equipment of traditional air-cooled data centers is approximately 2. This high losses are attributed to the inefficient air-cooling. The volumetric heat capacity of air is more than 2000 times smaller compared to water. Refrigeration loops can be eliminated in careful designed water cooled data centers with minimal and low thermal resistances in the heat path resulting in PUEs of 1.2. First concepts to re-use the dissipated heat at a quality of 50 °C in heating networks during the winter time in colder climates were demonstrated [130, 131]. This allows to reduce the total-cost-of-ownership of a data center, by selling the heat and can results in an effective PUE smaller than 1. The close proximity of the coolant to the junction with minimal thermal interfaces and the low flow rate with the large fluid temperature increase makes interlayer cooling a candidate for highly efficient data center cooling and energy-reuse. Therefore, specific designs should be proposed.

5.3.3 Extended Modeling Frame Work

Extended Porous Media Modeling and Database: The sub-domain model needs to be extended to derive the directional convective thermal resistance of pin-fins in a high accuracy. Hence, the incorporation of two periodic thermal boundary condition pairs has to be foreseen. Additional pin geometries should be investigated in the sub-domain model, to extend the database with directional and velocity dependent permeability and convective thermal resistance. From the datasets it might be possible to define a generic regression describing directional or velocity dependent parameters with respect to porous media parameters such as porosity and pore size. Ways to account for the developing hydrodynamic and thermal boundary layers should be proposed, to improve the accuracy of the porous-media models. This is especially important if strongly non-uniform cavities and power maps are implemented.

Compact Transient Thermal Modeling: Typical thermal design tools used during the electrical design phase are based on compact thermal modeling. They allow for a high spatial resolution of temporal effects in an IC-die, thanks to fast response of the method. Standard tools support heat conduction in the domains and heat transfer coefficients or temperatures at interfaces, relevant

for back-side cooling. Mass transport and convection has to be included in case of interlayer heat transfer, to account for the large fluid temperature increase. Fluid network modeling tools are available, but should be implemented into the compact thermal modeler to derive the local velocities. First concepts to represent convection as voltage dependent current source besides the capacitors and resistors responsible to predict the heat conduction in the solid were presented [73, 132]. This approach has also to leverage on extracted heat and mass transport parameters from sub-domain modeling. It is therefore compatible with the modeling framework presented in this thesis and allows to combine it with the tools used in electrical design phase. It also allows an efficient study of transient concepts such as thermal aware scheduling and algorithms therefor [133, 134].

Hot-Spot-Aware Fluid Network Algorithm: The manual design of hot-spot aware heat transfer cavities is a time-consuming process. Therefore, generic algorithms should be investigated to grow sophisticated fluid networks, minimizing pumping power for a given thermal budget. Fluid channels delivering coolant to low temperature areas would be redirected to locations exceeding the junction temperature limit in an iterative process. Appropriate rules need to be defined to guarantee a fast convergence towards the optimal fluid network.

A Power Map Contrast specific Pin Shape Optimization

A.1 Advanced Pin Shapes

In this section, the characteristics of various pin shapes are derived using the periodic boundary condition modeling concept discussed in subsection 2.3.1 and described in [86]. Unstructured meshes with hexahedral and prism-shaped elements for the solid and fluid domains are generated, respectively. The boundaries in the fluid domain are resolved with a prism inflation layer. The mesh quality in the unit-cell was validated in mesh refinement studies and resulted in typical node numbers of 44'000 and 131'000 for the solid and fluid domain, respectively. In the assessment, circular, square, super-elliptical, and drop-shaped (TSCHIRNHAUSEN cubic) pins were compared. Each pin shape was geometrically constrained by the need to embed a TSV with diameter of 40 µm. This resulted in a constant minimum width d_{min} of 50 µm between two pins transversal to the flow direction. The pitch and the cavity height t_c were set to 100 µm.

The super-elliptic pin shape is defined as:

$$\left|\frac{x}{a}\right|^m + \left|\frac{y}{a}\right|^m = 1, \tag{A.1}$$

where $a = 25\,\mu m$, $b = 25\,\mu m$, and $m = 1.5$. The square shape approximates a perfect square and is defined as a super-ellipse with $m = 4$. This represents a squared pin etched in silicon with its rounded corners resulting from the limited resolution of contact lithography.

The equation of the drop shape is

$$x^3 = 9a(x^2 - 3y^2), \tag{A.2}$$

where a equals 10 µm. The individual shapes are shown in Figure A.1. Their porosity, thermal conductive resistance, and ratio of wetted to projected area are reported in Table A.1. Also the parallel plate and the microchannel geometry is considered in the benchmark study.

Figure A.1: Top view of heat-transfer microchannel and pin-fin unit-cells. (a) microchannel, (b) circle, (c) square ($m = 4$), (d) super-ellipse ($m = 1.5$) and (e) drop.

The modified permeability (Equation 2.42) and convective thermal resistance (Equation 2.39) using water at 20 °C are computed for the pressure gradients of interest (1×10^6 to 1×10^7 $\frac{Pa}{m}$) and are compared with the experimental data from subchapter 3.3.1 as depicted in Figures A.2 (left, right). The flow is clearly laminar, even at high pressure gradients with REYNOLDS numbers of less than 150 for the pin-fins, and also for the parallel-plate with $Re = 1670$.

The permeability (and later also the convective thermal resistance) of the microchannel and the parallel plate are derived from correlations reported by SHAH and LONDON [82]. They are REYNOLDS-number-invariant, in contrast to those of the pin-fin structures, with a nonlinear behavior due to inertia effects. Pin shapes with a short flow constriction length relative to the unit-cell length result in an in-

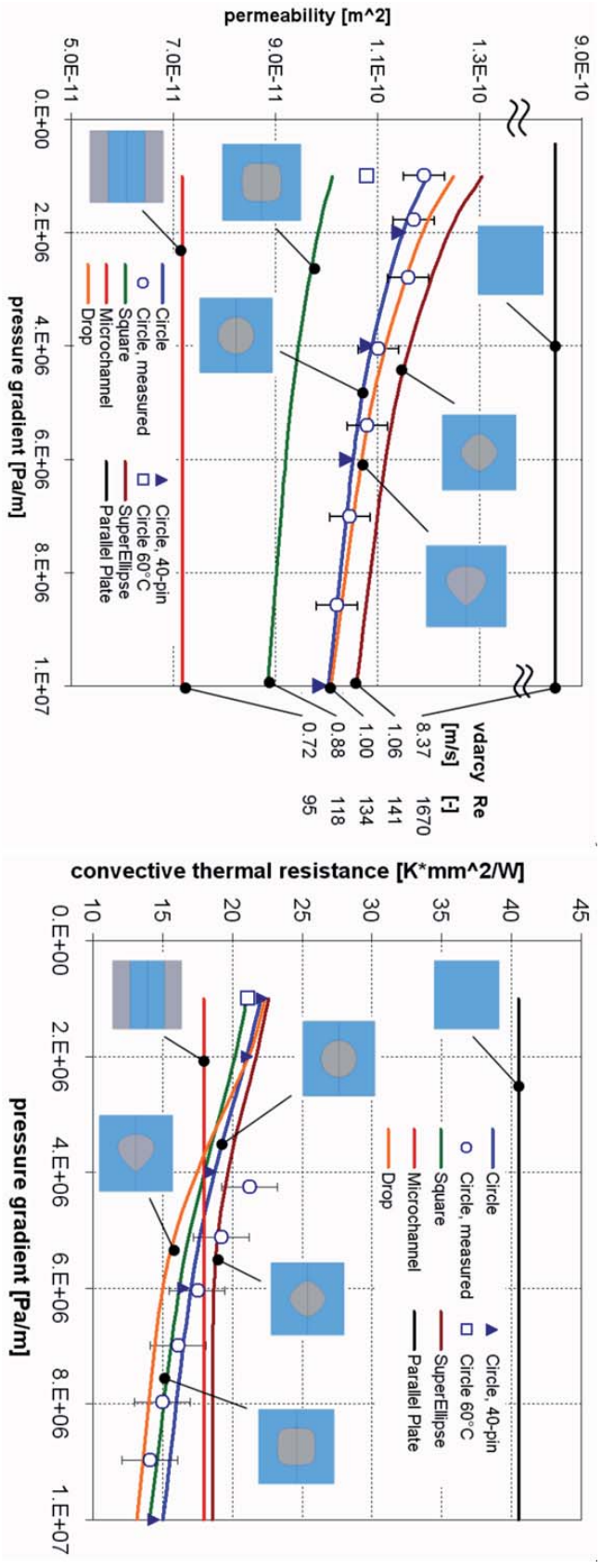


Figure A.2: Modified permeability of different pin shapes (at 20 °C fluid temperature), the microchannel and the parallel plate derived from numerical and correlation-based modeling. Experimental values from subsection 3.3.1 and numerical values from the 40-unit-cell model for the circular pin are also reported (Table 2.2) for validation purposes. The temperature dependency of the permeability of the circular pin shape is demonstrated at 60 °C fluid temperature (left). Convective thermal resistance of different pin shapes (at 20 °C average fluid temperature) at pressure gradient ranges of relevance for interlayer cooling are depicted. Numerical and experimental values from the 40-unit-cell model and from sub-chapter 3.3.1 have been added to validate the values for the circular pin. Furthermore, the temperature dependency of circular pins is demonstrated with an additional data-point at 60 °C average water temperature (right).

Pin Shape	Porosity [-]	Conductive Thermal Resistance $[\frac{\mathrm{K\,mm^2}}{\mathrm{W}}]$	Ratio of Wetted to Projected Area $[\frac{\mathrm{m^2}}{\mathrm{m^2}}]$
Parallel plate	1.000	inf	2.00
Microchannel	0.500	1.54	3.00
Circle	0.804	3.92	3.18
Square	0.769	3.33	3.29
Super-ellipse	0.829	4.49	3.14
Drop	0.826	4.43	3.18

Table A.1: Geometrical and conductive thermal characteristics of different pin shapes.

creased modified permeability (this ratio is 1 and 0.5 for the microchannel and the square pin shape, respectively). Hence, the super-ellipse outperforms the circle, square, and drop pin-shapes as well as the microchannel. The parallel plate with its 100 % porosity results in one order of magnitude increased permeability.

The pin shape performance sequence of the convective thermal resistance and the modified permeability is inverse at pressure gradients below $4 \times 10^6\ \frac{\mathrm{Pa}}{\mathrm{m}}$. The square and circular pin shapes start outperforming the microchannel above this pressure gradient value. This does not apply for the super-ellipse, which reaches a close to asymptotic value in this regime. Interestingly, the drop shape results in the highest heat-transfer capability. In general, these results indicate the difficulty to design pin shapes having high permeability and low convective thermal resistance at the same time.

In addition, the temperature dependency of the effective parameters is indicated with a data point for circular pins at a water temperature of 60 °C. For the convective thermal resistance, a change in permeability of -10 % and -4 % can be observed.

A.2 Pin Shape Selection with Respect to Power Map Contrast

The assessment of the individual pin shapes makes only sense in the context of their use. In this study, we consider a typical interlayer cooling cavity. The main parameters are the cavity length l_{cavity}, the power map, and the applied pressure difference from inlet to outlet of the cavity $\triangle p_{cavity}$. All these parameters influence the balance between the fluid temperature increase $\triangle T_{fluid}$ and the convective temperature drop $\triangle T_{conv}$. A compact thermal model assuming one-dimensional heat flux was used (resistor network in Figure 2.8) to derive the junction temperature in a 3D chip stack with an infinite number of tiers and equal power map. This approach was discussed in subsection 1.3.3 and is revered to as the "channel-walk-method".

A square chip stack with a 1 $\mathrm{cm^2}$ footprint, a fluid cavity height of 100 µm (as in the subdomain model), and a silicon slab thickness t_s of 50 µm is considered. A one-dimensional power map with a background $\dot{q}_{min}$ and a hot-spot $\dot{q}_{max}$ heat flux is imposed on each tier. The hot-spots are located at 2 to 3 mm and 7 to 8 mm downstream from the inlet position (Figure A.3 green line). The total power dissipated to each cavity is set to be 100 W for all cases. The benchmark pressure drop $\triangle p_{cavity}$ is set to 1×10^5 Pa, which is reasonable for server applications. The contrast of the power map is defined as the ratio R of maximum to minimum heat flux.

The fluid $\triangle T_f$, wall $\triangle T_w$, and junction temperature $\triangle T_j$ difference with respect to the fluid inlet temperature at a heat flux ratio R of 1.5 are shown in Figure A.3 for the circular pins as an example. The fluid temperature increase dominates the total thermal budget. The convective temperature drop contributes also significantly compared to the conductive temperature gradient, which can be neglected. The convective contribution at hot-spot locations becomes more pronounced at a power map contrast of $R = 5$. This strong dependency was the motivation to assess the pin shapes with respect to the power map contrast . The maximum junction temperature T_{jmax} with respect to the fluid inlet temperature T_{fin} is reported in Figure A.4.

The parallel plate results in the lowest junction temperatures at a low power map contrast of $R < 1.5$

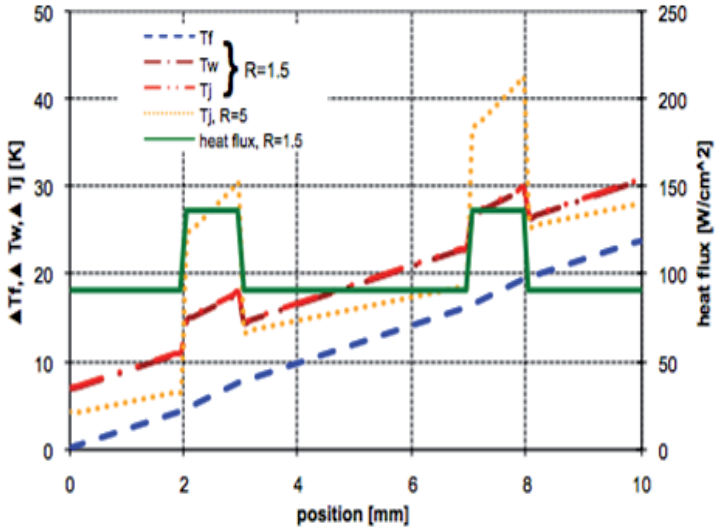


Figure A.3: Temperature response in case of circular pins with respect to the fluid inlet temperature and applied heat flux contrasts of $R = 1.5$ and 5, for a square chip stack with $1\,\mathrm{cm}^2$ footprint and a $1 \times 10^5\,\mathrm{Pa}$ pressure drop.

This thanks to its high permeability, with a resulting cavity flow rate of $0.5\,\frac{\mathrm{l}}{\mathrm{min}}$, roughly 10 times that of all other cavities. The poor convective heat transfer of the parallel plate alters its performance with increasing power map contrast. The drop shaped pins represent the best balance between permeability and convective thermal resistance from the area interconnect compatible heat-transfer structures, at both, low and high heat-flux ratios. They outperform the circular, square, and super-elliptic shapes and are a clear improvement over the microchannels. In general, the maximum junction temperature drops for heat flux ratios between 1 to 1.7. The location of T_{jmax} switches from the outlet (10 mm) to the end of the second hot-spot (8 mm) at the heat flux ratio of 1.7. This is depicted in Figure A.3 for $R = 1.5$, where both locations are quite balanced, and $R = 5$, where the hot-spot temperature dominates. The advantage of low-contrast power maps with reduced junction temperature becomes clear from Figure A.4. The fluid temperature increase from inlet to outlet $\triangle T_{fluid}$ relative to the maximum convective temperature gradient $max(\triangle T_{conv})$ for the circular pin-fins (dashed blue line) again demonstrates the importance of high-permeability cavities, especially for low-contrast power maps. This is even more pronounced at lower pressures (dashed orange line). Despite the drastic temperature increase caused mainly by the flow rate drop from $60\,\frac{\mathrm{ml}}{\mathrm{min}}$ for $1 \times 10^5\,\mathrm{Pa}$ to $32\,\frac{\mathrm{ml}}{\mathrm{min}}$ for $0.5 \times 10^5\,\mathrm{Pa}$, the local minimum shifts to heat flux ratios of 1.8. At larger chip-stack footprints, similar trends as discussed can be expected at lower pressure drops. This constrains interlayer cooling to chip-stack dimensions below $4\,\mathrm{cm}^2$.

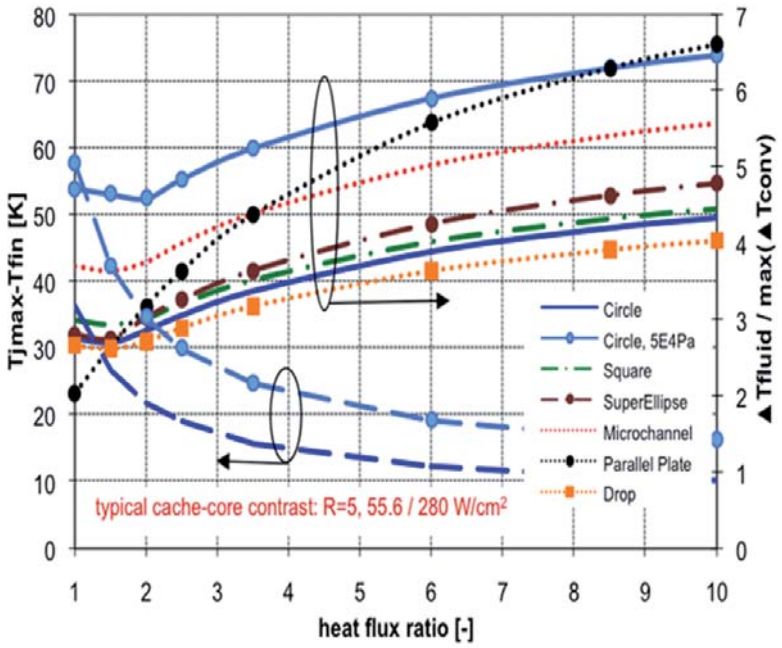


Figure A.4: Maximum junction temperature T_{jmax} with respect to the fluid inlet temperature T_{fin} for the individual heat-transfer geometries vs. the heat flux ratio, representing the contrast of the power map with two hot-spots as shown in Figure A.3. In addition, a square chip-stack with a $1\,\mathrm{cm}^2$ footprint at $1 \times 10^5\,\mathrm{Pa}$ pressure drop from cavity inlet to outlet and a total dissipated power of $100\,\mathrm{W}$ are assumed. Note, the pressure drop for the light blue lines is $0.5 \times 10^5\,\mathrm{Pa}$. The fluid temperature increase to the maximum convective temperature rise ratio is also plotted (dashed lines).

B Implementation of Porous-Media and Field-Coupling: ANSYS CFX

The standard ANSYS CFX V12 user interface does not provide all functions needed to implement the porous-media approach proposed and validated in sections 2.2, 4.2, and 4.3. Hence, a direct manipulation of the CFX-Solver input file created by CFX-Pre is needed. The coding has to be performed in the CFX command language (CCL), as described in the following sections.

B.1 Definition File Manipulation: CFX Command Line

The model geometry and the meshing of the individual domains in this study was performed by the ANSYS Design Modeler and ANSYS ICEM. CFX-Pre is then used to define domain types, material properties, boundary conditions, as well as the solver settings. The resulting definition file (*.def) includes all information needed for the CFX-Solver to compute the solution which can be analyzed with CFX-Post. All this steps are performed through a graphical user interface to minimize complexity at the expense of flexibility. The mentioned tools translate the user action into the CFX command language (CCL). The direct manipulation of this code results in additional flexibility, needed to solve non-standard problems.

The following commands are available in the CFX command line to read and write CCL code from and to the definition file:

```
cfx5cmds -read -def CFXfile.def -text ASCIIfile.ccl
cfx5cmds -write -def CFXfile.def -text ASCIIfile.ccl
```

The cfx5solve command can be used to start the CFX-Solver with the respective problem description:

```
cfx5solve -def CFXfile.def -double
```

Additional CCL code can be added by an extension of the cfx5solve command. Accordingly, porous rules (porous_rules.ccl available through the CFX support) needed to describe solid-to-solid heat transfer through the solid phase of the porous domain can be included:

```
cfx5solve -def CFXfile.def -ccl porous_rules.ccl -double
```

B.2 Mass Transport in the quasi Two-Dimensional Porous Domain

CFX does not support two-dimensional domains per se. A quasi two-dimensional porous domain is therefore represented by a minimal number of three mesh nodes (Figure 4.17 inset). Additionally, free-slip boundary conditions are imposed on the domain wall in CFX-Pre. The domain height is set equal to the cavity height, resulting in corresponding mass flow rates. The porosity has to be set for the considered heat transfer structure. The permeability of a domain in CFX-Pre is described utilizing the CFX expression language (CEL). Its implementation depends on the description method used.

Sinusoidal description

An isotropic loss model is chosen for the quasi-two dimensional domain in case of the sinusoidal description. The permeability is spatially defined as a function of the angle-of-attack $alpha$, fluid temperature T, and the DARCY velocity (superficial velocity) , as discussed in Table 2.4. The CEL code for the modified permeability of the pin-fin fully populated $kappaPFfp100$ case is presented as an example:

```
mysn = 1e-20[m/s]                                  # small number to pervent division by zero
myDynViscH2O = 2.414E-5[Pa s]*10^(247.8[K]/ \
    (T-140[K]))                                    # temp. dependent dynamic viscosity of water
vdarcy = truevelocity*Volume Porosity              # Darcy velocity
alpha = abs(atan(v/(u+mysn))                       # angle-of-attack
dpdxinline = -1.46E6[Pa*s^2/m^3]*vdarcy^2- \
    (-9.499E+04[Pa*s/m^2/K]*(T-273[K])+ \
    9.985E+06[Pa*s/m^2])*vdarcy                    # pressure gradient in in-line direction
dpdxstag = -9.55E6[Pa*s^2/m^3]*vdarcy^2- \
    (-1.225E+05[Pa*s/m^2/K]*(T-273[K])+ \
    1.065E+07[Pa*s/m^2])*vdarcy                    # pressure gradient in staggered direction
dpdx = (dpdxinline+dpdxstag)/2+ \
    (dpdxinline-dpdxstag)/2*cos(4*alpha)           # pressure gradient
kappaPFfp100 = -myDynViscH2O/dpdx*vdarcy           # modified permeability
```

Extended tensor description

A directional loss model is chosen for the quasi two-dimensional domain in case of the extended tensor description. The streamwise direction has to be defined and is used as local x-direction of individual porous domains. Two tables containing the x and y-permeability (κ_x, κ_y) as function of the pressure gradient (*dpds*, *gamma*) and the fluid temperature $tfluid$ need to be computed according to

$$\begin{pmatrix} \frac{dp}{dx} \\ \frac{dp}{dy} \end{pmatrix} = \nu \begin{pmatrix} \kappa_x^{-1} u \\ \kappa_y^{-1} v \end{pmatrix}, \tag{B.1}$$

from the velocity and pressure gradient field derived from detailed subdomain modeling. The tables are structured as follows:

```
[Name] kx

[Spatial Fields]
dpds    gamma    tfluid

[Data]
dpds[Pa/m]  gamma[rad]  tfluid[K]  kx[m^2]
5.00E+04    0.00E+00    2.50E+01   1.24E-10
...         ...         ...        ...
```

These values are maid available in CFX-Pre through User Functions. The directional loss is described e.g. for the pin-fin fully populated as follows:

```
mysnp = 1e-30 [Pa/m]                               # small number to pervent division by zero
dpdx = p.Gradient X                                # read local pressure gradient in x-direction
dpdy = p.Gradient Y                                # read local pressure gradient in y-direction
dpds = sqrt(dpdx^2+dpdy^2)                         # calculate local pressure gradient
gamma = abs(atan(dpdy/(dpdx+mysnp)))               # cal. orientation of local pressure gradient
kappaPFfp100x = kx.kx(dpds,gamma,T,0[m])           # interpolate permeability in x-dir. from table
kappaPFfp100y = ky.ky(dpds,gamma,T,0[m])           # interpolate permeability in y-dir. from table
```

Multiple porous domains

Multiple porous domains with individual permeabilities have to be defined analogous to the previous description for non-uniform heat removal cavities. The CEL function

```
inside()@domainname
```

results in a BOOLEAN response equal TRUE for locations inside the domain called “domainname”. It can be used in multiple IF-statements to impose a permeability expression to a specific domain:

```
if(inside()@domain1==1,kappa1,if(inside()@domain2==1,kappa2,kappa3))
```

B.3 Heat Transfer by Porous Domain Field-Coupling

An orthotropic fluid thermal conductivity with a close to infinite value normal to the fluid cavity has to be considered, to account for the quasi two-dimensionality of the porous domain (e.g. for water at constant temperature):

```
THERMAL CONDUCTIVITY:
        Option = Orthotropic Cartesian Components
        Thermal Conductivity X Component = 0.6069 [W m^-1 K^-1]
        Thermal Conductivity Y Component = 0.6069 [W m^-1 K^-1]
        Thermal Conductivity Z Component = 10000 [W m^-1 K^-1]
END
```

The field coupling between fluid-to-solid and solid-to-solid domains as described in subsection 2.2.4 is performed as follows:

Fluid-to-solid domain field-coupling

User Defined Wall Functions are available in CFX to describe the heat transfer in turbulent flows. In this study they are utilized to describe the field-coupling between the fluid phase in the porous domain and the adjacent solid domain. The following CCL code needs to be implemented for the respective surface pairs of the porous and solid domain to describe the heat transfer by the Wall Function Transfer Coefficient $twft$ and to achieve free-slip conditions:

```
HEAT TRANSFER:
        Option = Conservative Interface Flux
        TURBULENT WALL FUNCTIONS:
                Option = User Defined
                Wall Function Transfer Coefficient = twft
        END
END

MASS AND MOMENTUM:
        Option = No Slip Wall
        TURBULENT WALL FUNCTIONS:
                Option = User Defined
                Wall Function Transfer Coefficient = 1e-20[m/s]
        END
END
```

The local Wall Function Transfer Coefficient $twft$ depends on the local DARCY velocity and the angle-of-attack and is shown for the pin-fin fully populated case (according to Table 2.5):

```
mysn = 1e-20[m/s]                                   # number to pervent division by zero
porosityPFfp = 0.804                                # porosity
vdarcy = truevelocity*Volume Porosity               # Darcy velocity
alpha = abs(atan(v/(u+mysn))                        # angle-of-attack
Rconvinline = 2.527E-5[K*m^2/W]/((vdarcy+ \
    1.35[m/s])/(1[m/s]))^0.64+1.533E-6[K*m^2/W]     # in-line convective thermal resistance
Rconvstag = 2.527E-5[K*m^2/W]/((vdarcy+ \
    1.35[m/s])/(1[m/s]))^1.52+1.533E-6[K*m^2/W]     # staggered conv. thermal resistance
RconvPFfp100 = (Rconvinline+Rconvstag)/2+
    (Rconvinline-Rconvstag)/2*cos(4*alpha)          # convective thermal resistance
twftMain = 1/RconvPFfp100/porosityPFfp/Density/ \
    Specific Heat Capacityat Constant Pressure      # wall function transfer coefficient
```

The Wall Functions are only available if a turbulence model is selected for the individual porous domains. In reality, laminar flow exists in the cavities. Therefore, the EDDY viscosity and fractional intensity needs to be set to a insignificant value at the fluid inlet by:

```
TURBULENCE:
        Eddy Viscosity Ratio = 1e-20
        Fractional Intensity = 1e-20
        Option = Intensity and Eddy Viscosity Ratio
END
```

and the expert parameter "solve turbulence" needs to be set to FALSE:

```
EXPERT PARAMETERS:
        solve turbulence = f
END
```

Adiabatic conditions are imposed between the solid and the fluid phase in the porous domain by:

```
FLUID SOLID HEAT TRANSFER:
        Heat Transfer Coefficient = 0 [W m^-2 K^-1]
        Option = Heat Transfer Coefficient
END
```

This to comply with the field coupling definition discussed in subsection 2.2.4.

Solid-to-solid domain field-coupling

The heat dissipation between adjacent dies through heat conduction in the microchannel walls or the pin-fins is considered by the solid phase of the porous domain. An orthotropic thermal conductivity is defined reflecting the topology of the porous media. The effective thermal conductivity of the solid phase of a pin-fin array is only larger than zero normal to the cavity surface and is defined as:

```
THERMAL CONDUCTIVITY:
        Option = Orthotropic Cartesian Components
        Thermal Conductivity X Component = 0 [W m^-1 K^-1]
        Thermal Conductivity Y Component = 0 [W m^-1 K^-1]
        Thermal Conductivity Z Component = 140 [W m^-1 K^-1]
END
```

The effective thermal conductivity of the porous domain is then calculated from the porosity value and the thermal conductivity of the solid domain. CFX-Post allows to select between the phases (solid, fluid) in the porous domain to analyze the individual temperature fields of the solution.

C Abbreviations

Table C.1: A compilation for quick reference.

Abbreviation	Meaning
2D	Two dimensional
3D	Three dimensional
AES	Auger electron spectroscopy
ALD	Atomic layer deposition
AoA	Angle-of-attack
BC	Boundary condition
BCB	Benzocyclobutenes
BEOL	Back-end-of-line
BG	Background heater
BGA	Ball-grid-array
CCL	CFX command language
CFD	Computational fluid dynamics
CIS	CMOS image sensors
CMOS	Complementary metal–oxide–semiconductor
CMP	Chemical-mechanical polishing
COP	Coefficient-of-performance
CTE	Coefficient of thermal expansion
CVD	Chemical vapor deposition
D2D	Die-to-die
D2W	Die-to-wafer
DPF	Drop-shape pin-fin
DRAM	Dynamic random-access memory
DRIE	Deep reactive-ion etching
EDX	Energy-dispersive X-ray spectroscopy
F2B	Face-to-back
F2F	Face-to-face
FEOL	Front-end-of-line
fhp	Fully / half populated
fp	Fully populated
FVM	Finite volume method
gs	Guiding structures
HB-LED	High-brightness light-emitting diode
hp	Half populated
HS	Hot-spot heater
IC	Integrated circuit
ICP	Inductively coupled plasma

Abbreviation	Meaning
ICT	Information and communication technology
IDM	Integrated device manufacturer
ILD	Interlayer dielectric
IMC	Intermetallic compound
IMD	Intermetal dielectric
IO	Input / output
IP	Intellectual property
ITRS	International technology roadmap for semiconductors
KGD	Known-good-die
LGA	Land-grid-array
Ln	Level n
MC	Microchannel
MEMS	Microelectromechanical systems
MPU	Microprocessor unit
MTBF	Mean-time-between-failure
µ-PIV	Micro particle image velocimetry
OSAT	Outsourced assembly and test
PC	Perl-chain
PECVD	Plasma-enhanced chemical vapor deposition
PFI	Pin-fin in-line
PFS	Pin-fin staggered
PHS	Hot-spot power
PUE	Power usage effectiveness
REV	Representative elementary volume
RF	Radio frequency
RTD	Resistance thermal detector
SEM	Scanning electron microscope
SiP	System-in-package
SOC	System-on-a-chip
SOI	Silicon-on-insulator
SRAM	Static random-access memory
TCR	Temperature coefficient of resistance
THS	Temperature of hot-spot
TIM	Thermal interface material
TSV	Through-silicon via
UBM	Underbump metallization
W2W	Wafer-to-wafer
WLP	Wafer-level-package
XPS	X-ray photoelectron spectroscopy

D List of Symbols

Table D.1: Comprehensive list of symbols.

SYMBOL	DESCRIPTION	UNIT
Br	BRINKMAN number	[-]
f	FANNING friction factor	[-]
fr	DARCY friction factor	[-]
Nu	NUSSELT number	[-]
Pr	PRANDTL number	[-]
Re	REYNOLDS number	[-]
Re_{crit}	Critical REYNOLDS number	[-]
Re_K	REYNOLDS number based on pore diameter	[-]
γ	Pressure gradient direction	[°]
δ	$AuSn$ intermetallic compound	[-]
ϵ	Relative error	[%]
ζ'	Au_5Sn intermetallic compound	[-]
η_f	Fin-efficiency	[-]
θ	Flow direction	[°]
κ	Principal permeability	[m^2]
μ	Dynamic viscosity	[Pa s]
ν	Kinematic viscosity	[$\frac{m^2}{s}$]
ρ	Density	[$\frac{kg}{m^3}$]
σ	Standard deviation	[-]
τ	Time constant	[s]
φ	Porosity	[-]
a_n	Coefficients	[-]
AR_c	Aspect ratio of microchannel $= h_c/w_c$	[-]
A_{fs}	Wetted surface area	[m^2]
A_{heater}	Heater area	[m^2]
A_t	Wetted surface area	[m^2]
$A_{unit-cell}$	Unit-cell footprint	[m^2]
b	Heat transfer area width	[m]
C	Capacitance	[F]
c	Coefficient	[-]
c_f	Form-drag coefficient	[-]
c_p	Heat capacity	[$\frac{J}{kg\,K}$]
cp	Cold plate	[-]
d_h	Hydraulic diameter	[m]

SYMBOL	DESCRIPTION	UNIT
d_p	heat-transfer structure diameter	[m]
dp/ds	Pressure gradient	$[\frac{\mathrm{Pa}}{\mathrm{m}}]$
$\triangle P$	Pressure drop	[Pa]
$\triangle p_{cavity}$	Cavity pressure drop	[Pa]
$\triangle p_{cell}$	Unit-cell pressure drop	[Pa]
$\triangle p_{tot}$	Cavity and manifold pressure drop	[Pa]
$\triangle T_{cond}$	Conductive temperature increase	[K]
$\triangle T_{conv}$	Convective temperature increase	[K]
$\triangle T_{heat}$	Fluid temperature increase $= T_{fin}$–T_{fout}	[K]
$\triangle T_{jmax-fin}$	Maximal junction to fluid inlet temperature increase $= T_{jmax} - T_{fin}$	[K]
E_a	Activation energy	[J]
h	Heat transfer coefficient	$[\frac{\mathrm{W}}{\mathrm{m^2\,K}}]$
h_c	Channel height	[m]
i	Counter (integer)	[-]
K	Permeability	$[\mathrm{m}^2]$
k	Thermal conductivity	$[\frac{\mathrm{W}}{\mathrm{m\,K}}]$
k_m	Effective thermal conductivity	$[\frac{\mathrm{W}}{\mathrm{m\,K}}]$
k_{pf}	Thermal conductivity of fluid phase	$[\frac{\mathrm{W}}{\mathrm{m\,K}}]$
k_{ps}	Thermal conductivity of solid phase	$[\frac{\mathrm{W}}{\mathrm{m\,K}}]$
L	Length	[m]
l	Stack length	[m]
l_c	Channel length	[m]
L_{cell}	Cell length	[m]
$\dot{m}$	Mass flow rate	$[\frac{\mathrm{kg}}{\mathrm{s}}]$
m	Coefficient	[-]
n	Counter (integer)	[-]
N	Number of unit-cells	[-]
N_T	Pins in a column	[-]
p	Pressure	[Pa]
p_{in}	Inlet pressure	[Pa]
p_{out}	Outlet pressure	[Pa]
P_{active}	Active power	[W]
P_{BG}	Background heater power	[W]
P_{HS}	Hot-spot heater power	[W]
P_{tot}	Total dissipated power	[W]
$\dot{Q}$	Heat flow	[W]
$\dot{q}$	Heat flux	$[\frac{\mathrm{W}}{\mathrm{m^2}}]$
R	Electrical resistance	[Ω]
R_b	BEOL thermal resistance	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
R_{cond}	Conductive thermal resistance	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
R_{conv}	Convective thermal resistance	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
R_{cp}	Cold plate thermal resistance	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
R_{heat}	Effective thermal resistance representing $\triangle T_{heat}$	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$

SYMBOL	DESCRIPTION	UNIT
R_{si}	Silicon slab thermal resistance	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
R_{th}	Thermal resistance	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
R_{TIM}	Thermal interface resistance	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
R_{tot}	Total thermal resistance $= R_{cond} + R_{conv} + R_{heat}$	$[\frac{\mathrm{K\,m^2}}{\mathrm{W}}]$
$\vec{r}$	Position vector	[m]
s	Chip size	[m]
S_L	Longitudinal pitch	[m]
S_T	Transversal pitch	[m]
t	Thickness	[m]
$t_{experiment}$	Silicon slab thickness in experiment	[m]
t_{ref}	Silicon slab thickness in realistic chip stack	[m]
T	Temperature	[°C]
$T_f = T_{fluid}$	Fluid temperature	[°C]
$T_{fin} = T_{in}$	Fluid inlet temperature	[°C]
$T_{fout} = T_{out}$	Fluid outlet temperature	[°C]
T_j	Junction temperature	[°C]
T_{jla}	Line average junction temperature	[°C]
T_{jmax}	Maximal junction temperature	[°C]
T_{pf}	Temperature of fluid phase	[°C]
T_{ps}	Temperature of solid phase	[°C]
T_{reflow}	Reflow temperature	[°C]
T_s	Solid temperature	[°C]
T_w	Wall temperature	[°C]
v	Fluid velocity	$[\frac{\mathrm{m}}{\mathrm{s}}]$
v_{Darcy}	DARCY velocity $= v_{true} \cdot \varphi$	$[\frac{\mathrm{m}}{\mathrm{s}}]$
v_{mean}	Bulk fluid velocity	$[\frac{\mathrm{m}}{\mathrm{s}}]$
v_{true}	True velocity $= v_{mean}$	$[\frac{\mathrm{m}}{\mathrm{s}}]$
V_{dd}	Supply voltage	[V]
$\dot{V}$	Volumetric flow rate	$[\frac{\mathrm{m^3}}{\mathrm{s}}]$
V_f	Fluid volume	$[\mathrm{m^3}]$
V_{fs}	Volume of porous domain	$[\mathrm{m^3}]$
V_{tot}	Total volume	$[\mathrm{m^3}]$
w_c	Channel width	[m]
w_w	Channel wall width	[m]
x	Position in flow direction	[m]
y	Position transversal to flow direction	[m]

Bibliography

[1] G. E Moore. Cramming more components onto integrated circuits. *IEEE Solid-State Circuits Newsletter*, 20(3):33–35, 2006. 1

[2] K. J Kuhn. CMOS scaling beyond 32nm: Challenges and opportunities. In *Proc. of the 46th Annual Design Automation Conference (DAC), San Francisco, CA USA*, pages 310–313, 2009. 1

[3] F. P. Carson, Y. C. Kim, and I. S. Yoon. 3-D stacked package technology and trends. *Proceedings of the IEEE*, 97(1):31–42, 2009. 1

[4] P. Christie and D. Stroobandt. The interpretation and application of Rent's rule. *IEEE Transactions on Very Large Scale Integration(VLSI) Systems*, 8(6):639–648, 2000. 1

[5] Franz Laermer and Andrea Urban. Milestones in deep reactive ion etching. In *Proc. of the 13th International Conference on Solid-State Sensors and Actuators and Microsystems (TRANSDUCERS), Seoul, Korea*, volume 2 of *Digest of Technical Papers - International Conference on Solid State Sensors and Actuators and Microsystems, TRANSDUCERS '05*, pages 1118–1121, Seoul, Korea, Republic of, 2005. Institute of Electrical and Electronics Engineers Computer Society. Compendex. 1

[6] Andrea Urban and Franz Laermer. Bosch-process - DRIE success story, new applications and products. In *Proc. of the Materials Research Society Fall Meeting (MRS), Boston, CA USA*, volume 1112, pages 133–144, 2008. 1

[7] C. W. Kaanta, S. G. Bombardier, W. J. Cote, W. R. Hill, G. Kerszykowski, H. S. Landis, D. J. Poindexter, C. W. Pollard, G. H. Ross, and J. G. Ryan. Dual damascene: A ULSI wiring technology. In *Proc. of the 8th International VLSI Multilevel Interconnection Conference, Santa Clara, CA USA*, pages 144–152, 1991. 1

[8] Y. Xie. Processor architecture design using 3D integration technology. In *Proc. of the 23rd International Conference on VLSI Design, Bangalore, India*, pages 446–451, 2010. 2

[9] W. R Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. M Sule, M. Steer, and P. D Franzon. Demystifying 3D ICs: the pros and cons of going vertical. *IEEE Design and Test of Computers*, 22(6):498–511, 2005. 2

[10] R.H. Havemann and J.A. Hutchby. High-performance interconnects: An integration overview. *Proceedings of the IEEE*, 89(5):586–601, 2001. 2

[11] J. D Meindl, J. A Davis, P. Zarkesh-Ha, C. S Patel, K. P Martin, and P. A Kohl. Interconnect opportunities for gigascale integration. *IBM Journal of Research and Development*, 46(2-3):245–264, 2002. 2

[12] J. W Joyner, P. Zarkesh-Ha, and J. D Meindl. A stochastic global net-length distribution for a three-dimensional system-on-a-chip (3D-SoC). In *Proc. of the 14th IEEE International ASIC/SOC Conference, Arlington, VA USA*, pages 147–151, 2001. 2

[13] P. G. Emma and E. Kursun. Is 3D chip technology the next growth engine for performance improvement? *IBM Journal of Research and Development*, 52(6):541–552, 2008. 2

[14] B. Black, M. Annavaram, N. Brekelbaum, J. DeVale, L. Jiang, G. H Loh, D. McCaule, P. Morrow, D. W Nelson, and D. Pantuso. Die stacking (3d) microarchitecture. In *Proc. of the 39th IEEE/ACM International Symposium on Microarchitecture (MICRO), Orlando, FL USA*, pages 469–479, 2006. 2

[15] T. Kgil, S. D'Souza, A. Saidi, N. Binkert, R. Dreslinski, T. Mudge, S. Reinhardt, and K. Flautner. PicoServer: using 3D stacking technology to enable a compact energy efficient chip multiprocessor. *ACM SIGPLAN Notices*, 41(11):117–128, 2006. 2

[16] Kerry Bernstein, Paul Andry, Jerome Cann, Phil Emma, David Greenberg, Wilfried Haensch, Mike Ignatowski, Steve Koester, John Magerlein, Ruchir Puri, and Albert Young. Interconnects in the third dimension: design challenges for 3D ICs. In *Proc. of the 44th Design Automation Conference (DAC), San Diego, CA USA*, pages 562–567, San Diego, California, 2007. ACM. 2

[17] T. Kunio, K. Oyama, Y. Hayashi, and M. Morimoto. Three dimensional ICs having four stacked active device layers. In *International Electron Devices Meeting (IEDM) Technical Digest,*, pages 837–840, 1989. 3

[18] V. W. C. Chan, P. C. H. Chan, and M. Chan. Three-dimensional CMOS SOI integrated circuit usinghigh-temperature metal-induced lateral crystallization. *IEEE Transactions on Electron Devices*, 48(7):1394–1399, 2001. 3

[19] S. Jagar, M. Chan, M. C. Poon, H. Wang, M. Qin, P. K. Ko, and Y. Wang. Single grain thin-film-transistor (TFT) with SOI CMOS performanceformed by metal-induced-lateral-crystallization. In *International Electron Devices Meeting IEDM Technical Digest,*, pages 293–296, 1999. 3

[20] Y. Akasaka and T. Nishimura. Concept and basic technologies for 3-d ic structure. In *Proc. of International Electron Devices Meeting (IEDM)*, pages 488–491, 1986. 3

[21] A. W. Topol. Three-dimensional integrated circuits. *IBM Journal of Research and Development*, 50(4/5):506, 2006. 3

[22] T. R. Anthony. Forming electrical interconnections through semiconductor wafers. *Journal of Applied Physics*, 52:5340–5349, 1981. 4

[23] M. Koyanagi, T. Fukushima, and T. Tanaka. High-density through silicon vias for 3-D LSIs. *Proceedings of the IEEE*, 97(1):49–59, 2009. 4

[24] Philip Garrou, Christopher Bower, and Peter Ramm. *Handbook of 3D integration: Technology and applications of 3D integrated circuits*. Wiley-VCH, Weinheim, October 2008. 4

[25] C. K Tsang, P. S Andry, E. J Sprogis, C. S Patel, B. C Webb, D. G Manzer, and J. U Knickerbocker. CMOS-Compatible through silicon vias for 3-D process integration. In *Proc. of the Materials Research Society Symposium (MRS), San Francisco, CA USA*, volume 970, page 145, 2007. 4

[26] P. S. Andry, C. Tsang, E. Sprogis, C. Patel, S. L. Wright, B. C. Webb, L. P. Buchwalter, D. Manzer, R. Horton, and R. Polastre. A CMOS-compatible process for fabricating electrical through-vias in silicon. In *Proc. of the 56th Electronic Components and Technology Conference (ECTC), San Diego, CA USA*, pages 831–837, 2006. 4

[27] M. Shapiro, M. Interrante, P. Andry, B. Dang, C. Tsang, R. Liptak, J. Griffith, E. Sprogis, L. Guerin, and V. Truong. Reliable through silicon vias for 3D silicon applications. In *Proc. of the International Interconnect Technology Conference (IITC), Burlingame, CA USA*, pages 63 – 66, 2009. 4

[28] J. M. Jannou and J. Baron. 3d tsv technologies scenarios: Via first or via last? *Yole Developement Business Report*, 2010. 4, 5

[29] S. J. Koester, A. M. Young, R. R. Yu, S. Purushothaman, K. N Chen, D. C. La Tulipe, N. Rana, L. Shi, M. R. Wordeman, and E. J. Sprogis. Wafer-level 3D integration technology. *IBM Journal of Research and Development*, 52(6):583–597, 2008. 5, 6

[30] R. Puligadda, S. Pillalamarri, W. Hong, C. Brubaker, M. Wimplinger, and S. Pargfrieder. High-performance temporary adhesives for wafer bonding applications. In *Proc. of the Materials Research Society Symposium (MRS), Boston, MA USA*, volume 970, pages 239–252, 2007. 5

[31] J. Hermanowski. Thin wafer handling - study of temporary wafer bonding materials and processes. In *IEEE International Conference on 3D System Integration, 2009. 3DIC 2009*, pages 1–5. IEEE, September 2009. 5

[32] Robert Wieland, Karlheinz Bock, Erwin Hacker, Christof Landesberger, and Peter Ramm. Thin substrate handling by electrostatic force. In *2008 2nd European Conference & Exhibition on Integration Issues of Miniaturized Systems - MOMS, MOEMS, ICS and Electronic Components (SSI)*, pages 1–4. VDE, April 2008. 5

[33] S. Farrens. Vertical integration: A confederacy of alignment, bonding, and materials technologies. In *Proc. of the Materials Research Society Symposium (MRS), Boston, MA USA*, volume 970, page 215, 2007. 5

[34] C. S Tan, K. Chen, A. Fan, A. Chandrakasan, and R. Reif. Silicon layer stacking enabled by wafer bonding. In *Proc. of the Materials Research Society Symposium (MRS), San Francisco, CA USA*, volume 970, page 193, 2007. 6

[35] P. Enquist. High density bond interconnect (DBI) technology for three dimensional integrated circuit applications. In *Proc. of the Materials Research Society Symposium (MRS), Boston, MA USA*, volume 970, pages 13–24, 2007. 6

[36] K. Sakuma, P. S. Andry, C. K. Tsang, S. L. Wright, B. Dang, C. S. Patel, B. C. Webb, J. Maria, E. J. Sprogis, and S. K. Kang. 3D chip-stacking technology with through-silicon vias and low-volume lead-free interconnections. *IBM Journal of Research and Development*, 52(6):611–622, 2008. 6

[37] P. Ramm, A. Klumpp, R. Merkel, J. Weber, R. Wieland, A. Ostmann, and J. Wolf. 3D system integration technologies. In *Proc. of the Materials Research Society Symposium (MRS), Toulouse, France*, volume 766, pages 3–14, 2003. 6

[38] O.and Eigner M.and Gruber W.and Klumpp A.and Merkel R.and Ramm P.and Roth M.and Weber J. Huebner, H.and Ehrmann and R. Wieland. Face-to-face chip integration with full metal interface. In *Proc. of the Advanced Metallization Conference (AMC), San Diego, CA USA*, volume 18, pages 53–58, 2002. 6

[39] J. M. Jannou and J. Baron. 3d ic & tsv interconnects - business update. *Yole Developement Business Report*, 2010. 7

[40] J. U. Knickerbocker, P. S. Andry, B. Dang, R. R. Horton, C. S Patel, R. J. Polastre, K. Sakuma, E. S. Sprogis, C. K. Tsang, and B. C. Webb. 3D silicon integration. In *Proc. of the 58th Electronic Components and Technology Conference (ECTC), Orlando, FL USA*, pages 538–543, 2008. 7

[41] T. Brunschwiler and B. Michel. Handbook of 3D integration: Thermal management of vertically integrated packages. *Wiley-VCH, Weinheim*, Chap. 33:635–649, 2008. 7, 8, 10

[42] A. Rahman and R. Reif. Thermal analysis of three-dimensional (3-D) integrated circuits (ICs). In *Proc. of the IEEE International Interconnect Technology Conference (IITC), Burlingame, CA USA*, pages 157–159, 2001. 8

[43] W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti. Silicon CMOS devices beyond scaling. *IBM Journal of Research and Development*, 50(4/5):361, 2006. 8

[44] D. J Frank, R. H Dennard, E. Nowak, P. M Solomon, Y. Taur, and H. S.P Wong. Device scaling limits of si MOSFETs and their application dependencies. *Proceedings of the IEEE*, 89(3):259–288, 2001. 8

[45] H. F Hamann, A. Weger, J. A Lacey, Z. Hu, P. Bose, E. Cohen, and J. Wakil. Hotspot-limited microprocessors: Direct temperature and power distribution measurements. *IEEE Journal of Solid State Circuits*, 42(1):56, 2007. 8

[46] Y. S Lin, C. C Wu, C. S Chang, R. P Yang, W. M Chen, J. J Liaw, and C. H Diaz. Leakage scaling in deep submicron CMOS for SoC. *IEEE Transactions on Electron Devices*, 49(6):1034–1041, 2002. 8

[47] A. Jain and S. Ramanathan. Theoretical investigation of sub-ambient on-chip microprocessor cooling. In *Proc. of the 10th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), San Diego, CA USA*, pages 765–770, 2006. 9

[48] D. D Kandlur and T. W Keller. Green data centers and hot chips. In *Proc. of the 46th Annual Design Automation Conference (DAC), San Francisco, CA USA*, pages 888–890, 2009. 9

[49] J. G Koomey. *Estimating total power consumption by servers in the US and the world*. Citeseer, 2007. 9

[50] A. D. Kraus and A. Bar-Cohen. *Thermal analysis and control of electronic equipment*. McGraw-Hill, 1 edition edition, 1983. 9

[51] E. G. Colgan, B. Furman, M. Gaynes, W. S. Graham, N. C. LaBianca, J. H. Magerlein, R. J. Polastre, M. B. Rothwell, R. J. Bezama, and R. Choudhary. A practical implementation of silicon microchannel coolers for high power chips. *IEEE Transactions on Components and Packaging Technologies*, 30(2):218–225, 2007. 9

[52] R. J. Linderman, T. Brunschwiler, U. Kloter, H. Toy, and B. Michel. Hierarchical nested surface channels for reduced particle stacking and low-resistance thermal interfaces. In *Proc. of the 23th Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 87–94, 2007. 10

[53] T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann, and H. Reichl. Forced convective interlayer cooling in vertically integrated packages. In *Proc. 11th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), Orlando, FL USA*, pages 1114–1125, 2008. 10, 78, 85, 88, 95

[54] A. Bar-Cohen, T. Elperin, and R. Eliasi. Characterization of chippackages-justification, limitations, and future. In *Proc. of 5th Semiconductor Thermal and Temperature Measurement Symposium (SEMI-THERM), San Diego, CA USA*, pages 1–4, 1989. 10

[55] S. M. Sri-Jayantha, G. McVicker, K. Bernstein, and J. U. Knickerbocker. Thermomechanical modeling of 3D electronic packages. *IBM Journal of Research and Development*, 52(6):623–634, 2008. 10

[56] Eric Wong and Sung Kyu Lim. 3D floorplanning with thermal vias. In *Proc. of the Conference on Design, Automation and Test in Europe (DATE), Munich, Germany*, pages 878–883, 2006. 12

[57] H. Lee, S. Park, J. Baek, J. Lee, D. Lee, S. Oh, S. E Co, G. D LTD, and S. Korea. Thermal characterization of high performance MCP with silicon spacer having low thermal impedance. In *Proc. of the 21st Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 322–326, 2005. 12

[58] N. Popova, C. Schaeffer, C. Sarno, S. Parbaud, G. Moulin, and G. Kapelski. Thermal management for stacked 3D microelectronic packages. In *Proc. of the 36th Power Electronics Specialists Conference (PESC), Recife, Brazil*, pages 1761–1766, 2005. 12

[59] K. Takahashi, Y. Taguchi, M. Tomisaka, H. Yonemura, M. Hoshino, M. Ueno, Y. Egawa, Y. Nemoto, Y. Yamaji, and H. Terao. Process integration of 3D chip stack with vertical interconnection. In *Proc. of the 54th Electronic Components and Technology Conference (ECTC), Las Vegas, NV*, pages 601 – 609, 2004. 12

[60] X. Y. Chen, K. C. Toh, J. C. Chai, and D. Pinjala. Direct liquid cooling of a stacked multichip module. In *Proc. of the 4th Electronics Packaging Technology Conference (EPTC), Singapore*, pages 380–384, 2002. 12

[61] J. M Koo, S. Im, L. Jiang, and K. E Goodson. Integrated microchannel cooling for three-dimensional electronic circuit architectures. *Journal of Heat Transfer*, 127(1):49–58, 2005. 12

[62] B. Dang. *Integrated input/output interconnection and packaging for GSI*. PhD thesis, PhD thesis, Georgia Institute of Technology, Atlanta, USA, 2006. 12

[63] D. B. Tuckerman and R. F. W. Pease. High-performance heat sinking for VLSI. *IEEE Electron Device Letters*, 2(5):126–129, 1981. 14

[64] F. P Incropera. *Liquid cooling of electronic devices by single-phase convection*. Wiley-Interscience, 1999. 14

[65] J. Buongiorno, D. C Venerus, N. Prabhat, T. McKrell, J. Townsend, R. Christianson, Y. V Tolmachev, P. Keblinski, L. Hu, and J. L Alvarado. A benchmark study on the thermal conductivity of nanofluids. *Journal of Applied Physics*, 106(9):094312, 2009. 16

[66] R. Prasher, D. Song, J. Wang, and P. Phelan. Measurements of nanofluid viscosity and its implications for thermal applications. *Applied Physics Letters*, 89(13):133108–133108, 2006. 16

[67] W. Escher, T. Brunschwiler, N. Shalkevich, A. Shalkevich, T. Burgi, B. Michel, and D. Poulikakos. On the cooling of electronics with nanofluids. *Journal of heat transfer*, 133(5):051401, 2011. 16, 17

[68] T. Brunschwiler, H. Rothuizen, S. Paredes, B. Michel, E. Colgan, and P. Bezama. Hotspot-adapted cold plates to maximize system efficiency. In *Proc. ot the 15th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Leuven, Beglium*, pages 150–156, 2009. 17

[69] T. Brunschwiler, H. Rothuizen, U. Kloter, H. Reichl, B. Wunderle, H. Oppermann, and B. Michel. Hotspot-optimized interlayer cooling in vertically integrated packages. In *Proc. MRS fall meeting, Boston, MA USA*, pages 223–234, 2008. 17

[70] A. Kosar and Y. Peles. TCPT-2006-096. r2: Micro scale pin fin heat sink parametric performance evaluation study. *IEEE Transactions on Components and Packaging Technologies*, 30(4):855, 2007. 18, 41, 88

[71] N. Sahiti, A. Lemouedda, D. Stojkovic, F. Durst, and E. Franz. Performance comparison of pin fin in-duct flow arrays with various pin cross-sections. *Applied Thermal Engineering*, 26(11-12):1176–1192, 2006. 18

[72] W. A. Khan, J. R. Culham, and M. M. Yovanovich. The role of fin geometry in heat sink performance. *Journal of Electronic Packaging*, 128(4):324–330, December 2006. 18, 41

[73] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza. 3D-ICE: fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling. In *Proc. of the International Computer-Aided Design Conference (ICCAD), San Jose, CA USA*, pages 463–470, 2010. 22, 128

[74] T. Brunschwiler, H. Rothuizen, M. Fabbri, U. Kloter, B. Michel, R. J. Bezama, and G. Natarajan. Direct liquid jet impingement with micron-sized nozzle array and distributed return architecture. In *Proc. of the 20th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), San Diego, CA USA*, pages 196–203, 2006. 25

[75] W. Escher, B. Michel, and D. Poulikakos. A novel high performance, ultra thin heat sink for electronics. *International Journal of Heat and Fluid Flow*, 31:586–598, 2010. 25, 85

[76] N. T Nguyen and S. T Wereley. *Fundamentals and applications of microfluidics*. Artech House Inc., 2002. 25

[77] D. J. Acheson. *Elementary fluid dynamics*. Oxford University Press, New York, USA, 1990. 26

[78] Adrian Bejan, Ibrahim Dincer, Sylvie Lorente, Antonio Miguel, and Heitor Reis. *Porous and complex flow structures in modern technologies*. Springer-Verlag, New York, USA, 1 edition, April 2004. 29

[79] Massoud Kaviany. *Principles of heat transfer in porous media*. Springer, 1995. 29

[80] Donald A. Nield. *Convection in porous media*. Springer-Verlag, New York, USA, December 2009. 30

[81] F. Alfieri, M. K Tiwari, I. Zinovik, D. Poulikakos, T. Brunschwiler, and B. Michel. 3D integrated water cooling of a composite multilayer stack of chips. *Journal of Heat Transfer*, 132:121402, 2010. 31, 35

[82] R K Shah and Alexander L London. *Laminar flow forced convection in ductsa source book for compact heat exchanger analytical data*. Academic Press, New York, NY, 1978. 32, 34, 35, 129

[83] D. Ghosh Roychowdhury, Sarit Kumar Das, and T. Sundararajan. Numerical simulation of laminar flow and heat transfer over banks of staggered cylinders. *International Journal for Numerical Methods in Fluids*, 39(1):23–40, 2002. 35

[84] R. Wälchli, T. Brunschwiler, B. Michel, and D. Poulikakos. Combined local microchannel-scale CFD modeling and global chip scale network modeling for electronics cooling design. *International Journal of Heat and Mass Transfer*, 53(5-6):1004–1014, February 2010. 35

[85] K. Boomsma, D. Poulikakos, and Y. Ventikos. Simulations of flow through open cell metal foams using an idealized periodic cell structure. *International Journal of Heat and Fluid Flow*, 24(6):825–834, December 2003. 35

[86] G. Töral, R. Bender, Y. Leblebici, and T. Brunschwiler. Pin-shape assessment for interlayer-cooled chip stacks with periodic boundary condition modeling. In *Proc. 16th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Barcelona, Spain*, pages 1–10, 2010. 35, 129

[87] T. Brunschwiler, S. Paredes, U. Drechsler, B. Michel, B. Wunderle, and H. Reichl. Angle-of-attack investigation of pin-fin arrays in nonuniform heat-removal cavities for interlayer cooled chip stacks. In *Proc. 27th Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 116–124, 2011. 35, 48, 79, 99

[88] M. J Raw. *A new control-volume-based finite element procedure for the numerical solution of the fluid flow and scalar transport equations*. PhD thesis, PhD thesis, University of Waterloo Waterloo, Ontario, Canada, 1986. 38

[89] C. M. Rhie and W. L. Chow. A numerical study of the turbulent flow past an isolated airfoil with trailing edge separation. *Journal of American Institute of Aeronautics and Astronautics (AIAA)*, 21:1525–1532, 1983. 38

[90] M. Raw. Robustness of coupled algebraic multigrid for the Navier-Stokes equations. In *Proc. of the 34th Aerospace Sciences Meeting (AIAA), Reno, NV USA*, pages 96–0297, 1996. 38

[91] J. G. Ziegler and N. B. Nichols. Optimum settings for automatic controllers. *ASME Transactions*, 64:759–768, 1942. 40

[92] K. J. Åström and T. Hägglund. *Automatic tuning of PID controllers*. Instrument Society of America, 1995. 40

[93] Kwan-Soo Lee, Woo-Seung Kim, and Jong-Min Si. Optimal shape and arrangement of staggered pins in the channel of a plate heat exchanger. *International Journal of Heat and Mass Transfer*, 44(17):3223–3231, September 2001. 41

[94] R. S Prasher, J. Dirner, J. Y Chang, A. Myers, D. Chau, D. He, and S. Prstic. Nusselt number and friction factor of staggered arrays of low aspect ratio micropin-fins under cross flow for water as fluid. *Journal of Heat Transfer*, 129:141, 2007. 41, 88, 90, 94

[95] D. Steiner. *VDI-Wärmeatlas*. VDI-Verlag, Düsseldorf, Germany, 1997. 47

[96] H. R. Shanks, P. D. Maycock, P. H. Sidles, and G. C. Danielson. Thermal conductivity of silicon from 300 to 1400k. *Physical Review*, 130(5):1743, June 1963. 47

[97] D. E. Metzger, C. S. Fan, and S. W. Haley. Effects of pin shape and array orientation on heat transfer and pressure loss in pin fin arrays. *Journal of Engineering for Gas Turbines and Power*, 106(1):252–257, January 1984. 48

[98] V. I. Evenko and A. K. Anisin. Increasing efficiency of heat transfer by bundles of tubes in transverse flow. *Teploenergetika*, 23:21–24, 1976. 48

[99] Michael Pecht, Rakish Agarwal, F. Patrick McCluskey, Terrance J. Dishongh, Sirus Javadpour, and Rahul Mahajan. *Electronic packaging materials and their properties*. CRC Press, 1 edition, December 1998. 61, 62, 76

[100] T. B. Massalski, J. L. Murray, L. H. Bennett, and H. Baker. *Binary alloy phase diagrams. vol. i and ii*. American Society for Metals, Ohio , USA, 1986. 63

[101] G. Elger, R. Jordan, M. Suchodoletz, and H. Oppermann. Development of an low cost wafer level flip chip assembly process for high brightness LEDs using the AuSn metallurgy. In *Proc. of the 35th International Symposium on Microelectronics, Denver, Colorado USA*, pages 199–204, 2002. 63

[102] L. Dietrich, J. Wolf, O. Ehrmann, and H. Reichl. Wafer bumping technique using electroplating for high-dense chip packaging. 1998. 63

[103] C. Kallmayer, H. Oppermann, G. Eugelmann, E. Zakel, and H. Reichl. Self-aligning flip-chip assembly using eutectic gold/tin solder in different atmospheres. In *Proc. of the 19th International Electronics Manufacturing Technology Symposium (IEMT), Austin, TX USA*, pages 18–25, 2002. 63

[104] M. Hutter, F. Hohnke, H. Oppermann, M. Klein, and G. Engelmann. Assembly and reliability of flip chip solder joints using miniaturized Au/Sn bumps. In *Proc. of the 54th Electronic Components and Technology Conference (ECTC), Las Vegas, NV*, pages 49–57, 2004. 63

[105] M. Hutter, H. Oppermann, G. Engelmann, L. Dietrich, and H. Reichl. Precise flip chip assembly using electroplated AuSn20 and SnAg3. 5 solder. In *Proc. of the 56th Electronic Components and Technology Conference (ECTC), San Diego, CA USA*, pages 1–8, 2006. 63, 70

[106] A. Katz, C. H. Lee, and K. L. Tai. Advanced metallization schemes for bonding of InP-based laser devices to CVD-diamond heatsinks. *Materials Chemistry and Physics*, 37(4):303–328, 1994. 63, 70

[107] Goran S Matijasevic, Chin C Lee, and Chen Y Wang. AuSn alloy phase diagram and properties related to its use as a bonding medium. *Thin Solid Films*, 223(2):276–287, February 1993. 64, 66

[108] G. A Somorjai. *Chemistry in two dimensions: surfaces*. Cornell University Press, 1981. 64

[109] A. N Torgersen, L. Offernes, A. Kjekshus, and A. Olsen. The tin-rich part of the Au-Pt-Sn system. *Journal of Alloys and Compounds*, 314(1-2):92–95, 2001. 64, 70

[110] J. F Kuhmann, A. Preuss, B. Adolphi, K. Maly, T. Wirth, W. Oesterle, W. Pittroff, G. Weyer, and M. Fanciulli. Oxidation and reduction kinetics of eutectic SnPb, InSn, and AuSn: a knowledge base for fluxless solder bonding applications. *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, 21(2):134–141, 2002. 64, 65

[111] F. G Yost, F. M Hosking, and D. R Frear. *The mechanics of solder alloy wetting and spreading*. Van Nostrand Reinhold, 1993. 65

[112] D. R. Frear and D. M. Keicher. Fluxless soldering using activated acid vapors. In *Proc. of the Tech. Program NEPCON West, Anaheim, CA USA*, pages 1704–1715, 1992. 65

[113] H. Matsuki, H. Matsui, and E. Watanabe. Fluxless bump reflow using carboxylic acid. In *Proc. of the International Symposium on Advanced Packaging Materials: Processes, Properties and Interfaces, Braselton, GA USA*, pages 135–139, 2002. 65

[114] W. Lin and Y. C. Lee. Study of fluxless soldering using formic acid vapor. *IEEE Transactions on Advanced Packaging*, 22(4):592–601, 1999. 65

[115] G. Greitmann, H. Burkard, and J. Link. AuSn thin film solder layers for assembly of optoelectronic devices. In *Proc. of the 14th European Microelectronics and Packaging Conf., Friedrichshafen, Germany*, pages 1–5, 2003. 70

[116] J. Y. Tsai, C. W. Chang, Y. C. Shieh, Y. C. Hu, and C. R. Kao. Controlling the microstructure from the gold-tin reaction. *Journal of Electronic Materials*, 34(2):182–187, 2005. 70

[117] T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann, and H. Reichl. Interlayer cooling potential in vertically integrated packages. *Microsystem Technologies*, 15(1):57–74, 2009. 78, 85, 88, 95

[118] T. Brunschwiler, S. Paredes, U. Drechsler, B. Michel, W. Cesar, Y. Leblebici, B. Wunderle, and H. Reichl. Heat-removal performance scaling of interlayer cooled chip stacks. In *Proc. 12th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), Las Vegas, NV USA*, pages 1–12, 2010. 81, 111

[119] J. P Holman and W. J Gajda. *Experimental methods for engineers*. McGraw-Hill, New York, USA, 2001. 84

[120] Y. Muzychka and M. Yovanovich. Laminar forced convection heat transfer in the combined entry region of non-circular ducts. *Journal of Heat Transfer*, 126:54–61, 2004. 88, 92, 106, 107

[121] Y. Peles, A. Kosar, C. Mishra, C. J Kuo, and B. Schneider. Forced convective heat transfer across a pin fin micro heat sink. *International Journal of Heat and Mass Transfer*, 48(17):3615–3627, 2005. 88

[122] Adrian Renfer, Manish K. Tiwari, Thomas Brunschwiler, Bruno Michel, and Dimos Poulikakos. Hydrodynamics in a chip with micropin fin arrays for 3d integrated electronic cooling. In *Proc. of the 2nd European Conference on Microfluidics, Toulouse, France*, 2010. 95, 127

[123] A. Zukauskas and R. Ulinskas. *Heat transfer in banks of tubes in crossflow*. Taylor & Francis, 1 edition, April 1988. 107

[124] A. K Coskun, D. Atienza, T. S Rosing, T. Brunschwiler, and B. Michel. Energy-efficient variable-flow liquid cooling in 3D stacked architectures. In *Proc. of the Conference on Design, Automation and Test in Europe (DATE), Grenoble, France*, pages 111–116, 2010. 120

[125] C. H Choi, U. Ulmanella, J. Kim, C. M Ho, and C. J Kim. Effective slip and friction reduction in nanograted superhydrophobic microchannels. *Physics of Fluids*, 18:087105, 2006. 127

[126] Y. Kwon, N. Patankar, J. Choi, and J. Lee. Design of surface hierarchy for extreme hydrophobicity. *Langmuir*, 25(11):6129–6136, 2009. 127

[127] C. Lee, C. H Choi, and C. J.C.J Kim. Structured surfaces for a giant liquid slip. *Physical Review Letters*, 101(6):64501, 2008. 127

[128] C. Lee and C. J.C.J Kim. Maximizing the giant liquid slip on superhydrophobic microstructures by nanostructuring their sidewalls. *Langmuir*, 25(21):12812–12818, 2009. 127

[129] J. G Koomey. Worldwide electricity used in data centers. *Environmental Research Letters*, 3:034008, 2008. 127

[130] T. Brunschwiler, B. Smith, E. Ruetsche, and B. Michel. Toward zero-emission data centers through direct reuse of thermal energy. *IBM Journal of Research and Development*, 53(3):1–11, 2009. 127

[131] T. Brunschwiler, G. I Meijer, S. Paredes, W. Escher, and B. Michel. Direct waste heat utilization from liquid-cooled supercomputers. In *Proc. 14th International Heat Transfer Conference, Washington, DC USA*, pages 429–440, 2010. 127

[132] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza. Compact transient thermal model for 3D ICs with liquid cooling via enhanced heat transfer cavity geometries. In *Proc. of the 16th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Barcelona, Spain*, pages 1–6, 2010. 128

[133] M. M Sabry, A. Sridhar, D. Atienza, Y. Temiz, Y. Leblebici, S. Szczukiewicz, N. Borhani, J. R Thome, T. Brunschwiler, and B. Michel. Towards thermally-aware design of 3D MPSoCs with inter-tier cooling. In *Proc. of the Conference on Design, Automation and Test in Europe (DATE), Grenoble, France*, pages 1–6, 2011. 128

[134] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza. Compact transient thermal model for 3D ICs with liquid cooling via enhanced heat transfer cavity geometries. In *Proc. 16th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Barcelona, Spain*, pages 1–6, 2010. 128

List of Publications

Journal Articles

1 F. Alfieri, M. K. Tiwari, I. Zinovik, D. Poulikakos, T. Brunschwiler, and B. Michel. 3D integrated water cooling of a composite multilayer stack of chips. *Journal of Heat Transfer*, 132:121402, 2010.

2 T. Brunschwiler, U. Kloter, R. J. Linderman, H. Rothuizen, and B. Michel. Hierarchically nested channels for fast squeezing interfaces with reduced thermal resistance. *IEEE Transactions on Components and Packaging Technologies*, 30(2):226–234, 2007.

3 T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann, and H. Reichl. Interlayer cooling potential in vertically integrated packages. *Microsystem Technologies*, 15(1):57–74, 2009.

4 T. Brunschwiler, B. Smith, E. Ruetsche, and B. Michel. Toward zero-emission data centers through direct reuse of thermal energy. *IBM Journal of Research and Development*, 53(3):1–11, 2009.

5 W. Escher, T. Brunschwiler, B. Michel, and D. Poulikakos. Experimental investigation of an ultrathin manifold microchannel heat sink for Liquid-Cooled chips. *Journal of Heat Transfer*, 132:081402, 2010.

6 W. Escher, T. Brunschwiler, N. Shalkevich, A. Shalkevich, T. Burgi, B. Michel, and D. Poulikakos. On the cooling of electronics with nanofluids. *Journal of heat transfer*, 133(5):051401, 2011.

7 T. Morf, J. Weiss, L. Kull, H. Rothuizen, T. Toifl, M. Kossel, C. Menolfi, G. von Bueren, T. Brunschwiler, and M. Schmatz. Antenna coupled far-infrared THz detector in CMOS. *Electronics letters*, 45(25):1321–1323, 2009.

8 B. J. Offrein, D. Jubin, T. Koster, T. Brunschwiler, F. Horst, D. Wiesmann, I. Meijer, M. S. Petit, D. Webb, R. Germann, et al. Polarization-independent thermooptic phase shifters in silicon-oxynitride waveguides. *Photonics Technology Letters, IEEE*, 16(6):1483–1485, 2004.

9 A. Renfer, M. K. Tiwari, T. Brunschwiler, B. Michel, and D. Poulikakos. Experimental investigation into vortex structure and pressure drop across microcavities in 3D integrated electronics. *Experiments in Fluids*, pages 1–11, 2011.

10 G. Salis, S. F. Alvarado, M. Tschudy, T. Brunschwiler, and R. Allenspach. Hysteretic electroluminescence in organic light-emitting diodes for spin injection. *Physical Review B*, 70(8):085203, 2004.

11 R. Wälchli, T. Brunschwiler, B. Michel, and D. Poulikakos. Combined local microchannel-scale CFD modeling and global chip scale network modeling for electronics cooling design. *International Journal of Heat and Mass Transfer*, 53(5-6):1004–1014, 2010.

12 R. Wälchli, T. Brunschwiler, B. Michel, and D. Poulikakos. Self-contained, oscillating flow liquid cooling system for thin form factor high performance electronics. *Journal of Heat Transfer*, 132:051401, 2010.

Book Chapters

1 T. Brunschwiler and B. Michel. Handbook of 3D integration: Thermal management of vertically integrated packages. *Wiley-VCH, Weinheim*, Chap. 33:635–649, 2008.

Conference Papers

1 T. Brunschwiler, U. Kloter, R. J. Linderman, H. Rothuizen, and B. Michel. Hierarchically nested channels for fast squeezing interfaces with reduced thermal resistance. In *Proc. 21th Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 1–8, 2005.

2 T. Brunschwiler, G. I. Meijer, S. Paredes, W. Escher, and B. Michel. Direct waste heat utilization from liquid-cooled supercomputers. In *Proc. 14th International Heat Transfer Conference, Washington, DC USA*, pages 429–440, 2010.

3 T. Brunschwiler, S. Paredes, U. Drechsler, B. Michel, W. Cesar, Y. Leblebici, B. Wunderle, and H. Reichl. Heat-removal performance scaling of interlayer cooled chip stacks. In *Proc. 12th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), Las Vegas, NV USA*, pages 1–12, 2010.

4 T. Brunschwiler, S. Paredes, U. Drechsler, B. Michel, B. Wunderle, and H. Reichl. Angle-of-attack investigation of pin-fin arrays in nonuniform heat-removal cavities for interlayer cooled chip stacks. In *Proc. 27th Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 116–124, 2011.

5 T. Brunschwiler, H. Rothuizen, U. Kloter, H. Reichl, B. Wunderle, H. Oppermann, and B. Michel. Hotspot-optimized interlayer cooling in vertically integrated packages. In *Proc. MRS fall meeting, Boston, MA USA*, pages 223–234, 2008.

6 T. Brunschwiler, H. Rothuizen, S. Paredes, B. Michel, E. Colgan, and P. Bezama. Hotspot-adapted cold plates to maximize system efficiency. In *Proc. 15th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Leuven, Beglium*, pages 150–156, 2009.

7 A. K. Coskun, D. Atienza, T. S. Rosing, T. Brunschwiler, and B. Michel. Energy-efficient variable-flow liquid cooling in 3D stacked architectures. In *Proc. of the Conference on Design, Automation and Test in Europe, Leuven, Belgium*, pages 111–116, 2010.

8 M. Fabbri, A. Wetter, B. Mayer, T. Brunschwiler, B. Michel, H. Rothuizen, R. Linderman, and U. Kloter. Microchip cooling module based on FC72 slot jet arrays without cross-flow. In *Proc. 22th Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 54–58, 2006.

9 R. Linderman, T. Brunschwiler, B. Smith, and B. Michel. High-performance thermal interface technology overview. In *Proc. 13th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Budapest, Hungary*, pages 129–134, 2007.

10 B. J. Offrein, D. Wiesmann, G. Salis, M. Sousa, I. Meijer, F. Horst, R. Germann, T. Brunschwiler, D. J. Webb, and G. L. Bona. Advances in silicon oxynitride waveguides. In *Proc. of Integrated Photonics Research*, 2003.

11 A. Renfer, M. K. Tiwari, T. Brunschwiler, B. Michel, and D. Poulikakos. Hydrodynamics in a chip with micropin fin arrays for 3d integrated electronic cooling. In *Proc. of the 2nd European Conference on Microfluidics, Toulouse, France*, 2010.

12 M. M. Sabry, A. Sridhar, D. Atienza, Y. Temiz, Y. Leblebici, S. Szczukiewicz, N. Borhani, J. R. Thome, T. Brunschwiler, and B. Michel. Towards thermally-aware design of 3D MPSoCs with inter-tier cooling. In *Proc. of the Conference on Design, Automation and Test in Europe (DATE), Grenoble, France*, pages 1–6, 2011.

13 B. Smith, T. Brunschwiler, and B. Michel. Utility of transient testing to characterize thermal interface materials. In *Proc. 13th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Budapest, Hungary*, pages 6–11, 2007.

14 B. Smith, H. Rothuizen, R. Linderman, T. Brunschwiler, and B. Michel. Design of thermal interfaces with embedded microchannels to control bond line formation. In *Proc. 11th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), Orlando, FL USA*, pages 410–418, 2008.

15 A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza. 3D-ICE: fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling. In *Proc. of the International Computer-Aided Design Conference (ICCAD), San Jose, CA USA*, pages 463–470, 2010.

16 A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza. Compact transient thermal model for 3D ICs with liquid cooling via enhanced heat transfer cavity geometries. In *Proc. 16th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Barcelona, Spain*, pages 1–6, 2010.

17 G. Töral, R. Bender, Y. Leblebici, and T. Brunschwiler. Pin-shape assessment for interlayer-cooled chip stacks with periodic boundary condition modeling. In *Proc. 16th Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Barcelona, Spain*, pages 1–10, 2010.

18 R. Wälchli, R. Linderman, T. Brunschwiler, U. Kloter, H. Rothuizen, N. Bieri, D. Poulikakos, and B. Michel. Radially oscillating flow hybrid cooling system for low profile electronics applications. In *Proc. 24th Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 142–148, 2008.

Best Paper Awards

1 T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann, and H. Reichl. Forced convective interlayer cooling in vertically integrated packages. In *Proc. 11th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), Orlando, FL USA*, pages 1114–1125, 2008.

2 T. Brunschwiler, H. Rothuizen, M. Fabbri, U. Kloter, B. Michel, R. J. Bezama, and G. Natarajan. Direct liquid jet-impingment cooling with micron-sized nozzle array and distributed return architecture. In *Proc. 10th Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), San Diego, CA USA*, pages 196–203, 2006.

3 R. J. Linderman, T. Brunschwiler, U. Kloter, H. Toy, and B. Michel. Hierarchical nested surface channels for reduced

particle stacking and low-resistance thermal interfaces. In *Proc. 23th Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), San Jose, CA USA*, pages 87–94, 2007.

Issued Patents

1 T. Brunschwiler, K. Bernstein, and B. Michel. Assembly including plural through wafer vias method of cooling the assembly and method of fabricating the assembly, US0311826, 2009.

2 T. Brunschwiler, M. Despont, M. Lantz, B. Michel, and P. Vettiger. Semiconductor device with a high thermal dissipation efficiency, US7928565, 2010.

3 T. Brunschwiler, R. Holzner, J. Wagner, A. Dommann, K. Mundle, and J. Jutz. Optical waveguide coupler and method for producing same, US757914, 2001.

4 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, and H. Rothuizen. Heat spreader, US0017659, 2006.

5 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, and H. Rothuizen. Cooling device, US0119565, 2006.

6 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, and H. Rothuizen. Patterned structure for a thermal interface, US0263879, 2005.

7 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, and H. Rothuizen. Thermal interface with a patterned structure, US7282799, 2006.

8 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, and H. Rothuizen. Semiconductor chip assembly with flexible metal cantilevers, US7532475, 2006.

9 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, and H. Rothuizen. Method of fabricating a surface adapting cap with integral adapting material for single and multi chip module assemblies, US7547582, 2006.

10 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, H. Rothuizen, and R. Wälchli. Method and device for cooling a heat generating component, US0036538, 2008.

11 T. Brunschwiler, U. Kloter, R. Linderman, B. Michel, and E. Ruetsche. Coolant pumping system for mobile electronic systems, US0044005, 2008.

12 T. Brunschwiler, R. Linderman, B. Michel, and H. Rothuizen. Integrated circuit stack, US0290188, 2008.

13 T. Brunschwiler, R. Linderman, B. Michel, H. Rothuizen, and E. Ruetsche. Variable performance server system and method of operation, US7866173, 2008.

14 T. Brunschwiler, R. Linderman, B. Michel, and E. Ruetsche. Variable flow computer cooling system for a data center and method of operation, US7808780, 2008.

15 T. Brunschwiler, R. Linderman, H. Rothuizen, and R. Wälchli. Cooling system for an electronic component system cabinet, US7872867, 2008.

16 T. Brunschwiler, R. Linderman, and E. Ruetsche. One-dimensional hierarchical nested channel design for continuous feed manufacturing processes, US0154107, 2007.