ICCAD 2015 Contest in 3D Interlayer Cooling Optimized Network *

Arvind Sridhar
IBM Research -Zurich
Rüschlikon, ZH
Switzerland
rvi@zurich.ibm.com

Mohamed M. Sabry
Stanford University
Stanford CA
USA
msabry@stanford.edu

David Atienza
EPFL
Lausanne, VD
Switzerland
david.atienza@epfl.ch

ABSTRACT

Microchannel liquid cooling has been proposed since the late 2000s as a viable enabler for 3D integration of microprocessors to continue scaling of computing power in the face of increasingly reduced returns from CMOS technology scaling. Thermal and electrical demonstrations of microchannel liquid-cooled heat sinks on the back side of IC dies exist in the literature and the compatibility of its fabrication with the existing CMOS process has been shown. This compatibility also gives rise to the prospect of building of nearly an infinite variety of channel networks with no additional manufacturing cost. This ICCAD 2015 problem aims to identify methods to optimize such microchannel fluid networks, and to evaluate impact of different cooling networks on different computing architectures floorplans.

Categories and Subject Descriptors

1 [**System-level CAD**]: Dark Silicon and Power/Thermal Considerations

Keywords

Liquid-cooling, 3D ICs, Temperature-aware design

1. INTRODUCTION

Interlayer single-phase liquid cooling has been proposed as an effective cooling mechanism for the high-heat dissipated in high-performance 3D-stacked processing architectures [1]. Single-phase fluid, such as water, is injected into micro-scale channels (also called micro-channels) that are etched between two consecutive vertical tiers to carry the heat out from different layers in the 3D stack. This cooling mechanism has already been proven to be much more effective than conventional air-cooling in order to remove very high on-chip temperature and cool the computing system much faster (several orders of magnitude of improvement). Modeling methods for thermal simulation of ICs with microchannels already exist in in the literature [2]. These tools and methods are useful for evaluating designs.

However, the culminating point of the development of technologies and modeling methods is the realization of a scalable design method and optimizer for such inter-layer microchannel liquid-cooled heat sinks. Such optimizers must have at their heart the minimization of cooling energy expenditures if this technology is to be seen as a viable tool to address the challenges of dark silicon. Indeed a one-design-fits-all approach of using uniform micro-channels can easily lead to overcooling, which is not an energy-efficient operation, or non-uniform cooling that increases on-chip thermal gradients giving rise to reliability issues and shorter lifetimes. Thus, it is fundamental to optimize several parameters, such as the pressure drop across the channels, the number of fluid inlet and outlet ports, and the fluid network, to achieve optimized cooling in terms of energy efficiency, thermal gradient and peak temperature.

There have also been studies to optimize straight channels both at design-time employing methods such as modulating channel-widths [3] or by modifying heat conduction paths by judicious placement of TSVs [4]. There have also been methods proposed to address efficient run-time operation of liquid-cooled microchannels [5]. However, these methods do not take advantage of the nearly endless possibilities of designing microfluidic networks using the existing CMOS process with no additional manufacturing cost. Customized fluidic networks can be potentially created tailored to cool specific parts of a microprocessor utilizing the minimum possible cooling energy. Some initial studies in this direction, using four-port fluidic packages were explored in [6]. However, a systematic optimizing tool to create a customized microfluidic network doesn't exist.

The goal of this ICCAD 2015 problem is to bridge this gap in our knowledge by encouraging the contestants to evaluate the impact of different cooling networks on different computing architectures floorplans, when subjected to specific thermal performance requirements. We hope to motivate further research insights in liquid-based cooling, and potentially attract even more industrial and academic interest in this field through this contest.

2. THE ICCAD-2015 CONTEST OVERVIEW

The main objective of this problem is to find an optimized cooling network that minimizes a specific cost function while

*This work has been partially supported by the Nano-Tera YINS RTD project (no. 20NA21 150939), financed by the Swiss Confederation and scientifically evaluated by SNSF, and the EC FP7 GreenDataNet STREP project (Agreement No. 609000).

ICCAD 2015 Austin, Texas USA

Table 1: Terms used in the problem definition

f_i	Floorplan of layer i in the target 3D stacked architecture
m_{ji}	Architectural module j in the i^{th} layer
CN_i	Cooling network layer i in the target 3D stack
$T_{max}(m_{ji})$	Peak temperature observed in module m_{ji}
ΔP	Pressure drop observed in the entire cooling network
$\dot{V}$	Inlet flow rate

subjected to design and physical constraints. Before unfolding the target problem, we need to define several terms to avoid ambiguity. These terms are list in Table 1.

To reflect various operating conditions we have created a set of problems, which we elaborate in the following sections. As general definition, given a certain N-layer 3D-stacked computing architecture defined by floorplan $F = \{f_1, f_2, , f_N\}$, where each layer $i \subset [1, N]$ consists of a set of j modules $Mi = \{m_{1i}, m_{2i}, ..., m_{ji}\}$, provide the set of $N-1$ cooling network layers $CN = \{CN_1, CN_2, ..., CN_{N-1}\}$ (cooling network layer CN_i is placed between stacked layers i and $i+1$) that is a solution of:

1. **MinT:** minimizing the peak temperature (T_{max}) and thermal gradient (ΔT_{max}), subject to pressure drop constrain ($\Delta P \leq P_{max}$).

2. **MinCE:** minimizing cooling energy ($E = \Delta P \cdot \dot{V}$) subject to a given maximum peak temperature and peak thermal gradient.

Based on the problem definition, we will provide different case studies, where each case study will have the input defined as follows:

- Number of layers in the targeted 3D stack (i.e., the number N).
- The floorplan of each layer in the targeted 3D stack (F).
- The power traces of each component in the 3D stack ($M = \{M_1, M_2, , M_N\}$).
- The list of interlayer floorplan ($F_{IL} = \{fil_1, fil_2, ... , fil_{N-1}\}$), where each element of this list shows the locations where a microchannel placement is infeasible.

In addition to the mentioned inputs, we fix the channel width to the following value for all test cases:

- Channel width= 100μm.

Various teams will use the mentioned inputs to provide the following list of outputs, which are used for evaluation:

- The applied pressure drop to inject the fluid to the cooling network (ΔP) .
- The layout of the different cooling networks in the 3D stack (CN).
- The location of inlet and outlet ports.

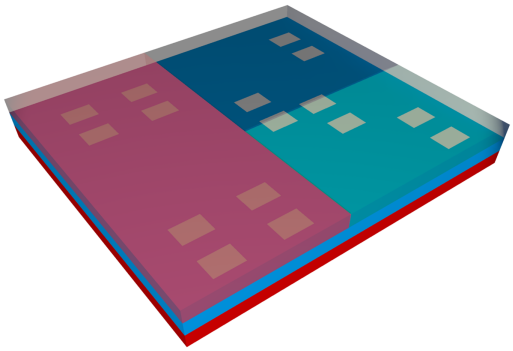

Figure 1: Schematic 3D diagram of an example 2-tier stack.

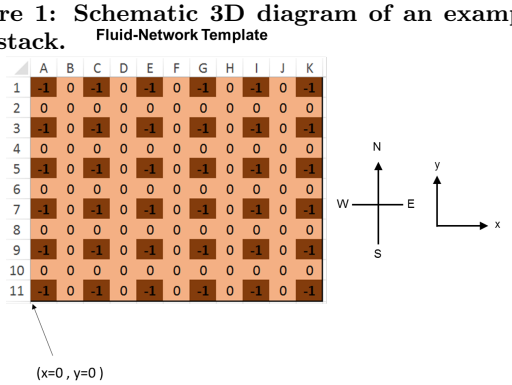

Figure 2: Template of a 11X11 Fluid-Network: 0 = silicon, 1= Fluid, -1=TSV (fixed), 2=inlet, 3=outlet.

Example: As an illustrative example, we provide the inputs of a 2-tier stack, which is depicted in Fig. 1. The inputs for this example is as follows:

$N = 2$
$F = \{f1, f2\}$
$f1 = \{m_{11} : [(0,0), (L,W)]\}$
$f2 = \{m_{12} : [(0,0), (L/2,W)], m_{22} :$
$[(L/2,0), (L/W/2)], m_{32} : [(L/2,W/2), [L,W)]\}$
$P(M2) = \{(0;0;0), (P1 : P2; P3), (P1; P2/2; P3)$
$, (P1/2; 0; P3/2), (0;0;0)\}$
$F_{IL} = \{fil_1\}$
$fil_1 = \{[L1,W1], [L2,W2], , [Lk,Wk]\}$

2.1 Evaluation Methodology

The proposed solutions by the different teams working in the competition, which is defined by the output list in the previous section, will be assessed based on:

1. Finding a feasible solution
2. Meeting the problem constraints
3. Minimizing the relevant metric to each problem

Detailed rules regarding the format of the problem statements provided, the mode of solving and the solutions obtained are as follows:

Rule #1: In order to help the contestants use an existing thermal model for performing their design evaluations, the description of the problem 3D stack, the floorplan and the

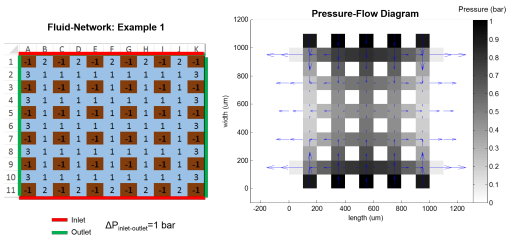

Figure 3: Fluid-Network Example 1 and the resulting Pressure-Flow Diagram

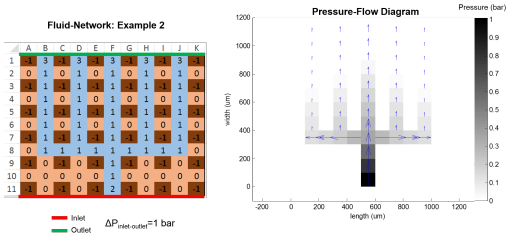

Figure 4: Fluid-Network Example 2 and the resulting Pressure-Flow Diagram

heat inputs will be given in the same format as the 3D-ICE input files (`.stk` and `.flp`). Please refer to the 3D-ICE User Guide available at [7]. Divide your cavity layer into square cells (from top-view) of 100μm. For a chip size of 10.1mmX10.1mm, this would give a grid of 101X101 cells. Note that some of the channel description part of the .stk file will be skipped in the input definition as it is part of the solution that you must compute.

Rule #2: Specify your solution fluid-network map solution in an excel file using an array of 0's and 1's. 0 indicates there is silicon in the cell, and 1 indicates channel/liquid. For the purpose of illustration, a 11X11 grid example template for a fluid-network representing a cavity layer is shown below in Fig. 2.

Rule #3: A 2D array of alternating cells cannot have fluid and have a -1 by default- to account for TSV fabrication. This is specified in the template using bold -1's in cells filled with brown background (Fig. 2).

Rule #4: The table in the excel file will be visualized in the same orientation as the top-view of the physical device. That is, the bottom-left (south-west) corner of the table will correspond to origin. The origin point and the cardinal directions considered are shown in Fig. 2. This is the same convention that is used for the input heat-flux map provided in the problem, and in the 3D-ICE model.

Rule #5: Your solution must have at least one inlet and one outlet. The inlet and outlets can only be at the edges of the microchannel layer. The edge fluid cells corresponding to inlet and outlet are specified using 2's and 3's respectively (see Examples 1-4, Fig. 3-6).

Rule #6: There can be multiple inlets and outlets along all four edges of your solution. However, to reduce the complexity of the packaging, for each side, there can be at most one "continuous" inlet and outlet, as illustrated in Example 3. There should be at least one traceable fluidic path from the inlet to the outlet of the system. If there are multiple channel layers in the problem, the inlet and outlet patterns

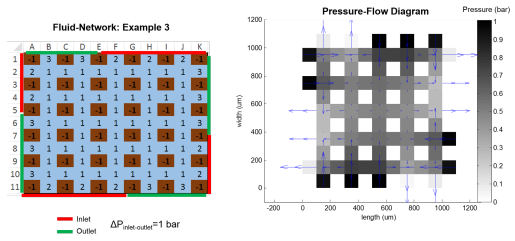

Figure 5: Fluid-Network Example 1 and the resulting Pressure-Flow Diagram

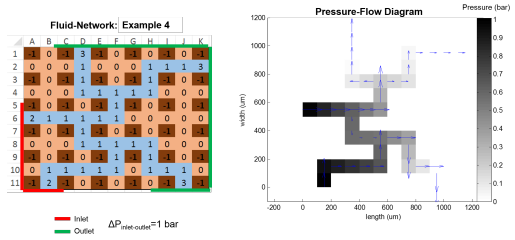

Figure 6: Fluid-Network Example 1 and the resulting Pressure-Flow Diagram

across the multiple layers must be identical. The individual fluid networks within each channel layer can however be different.

Rule #7: Provide your final solution excel file for each problem (separately), after you name the file in the following format:
<TeamID>_ProblemA_<testcase#>.xls
where <TeamID> refers to your Team Identification as you have been assigned by the ICCAD 2015 registration process <testcase#> refers to the serial number of the problem test case.

If there are multiple channel layers in a given test case, provide them as separate .xls files in the following format (channel layer counted from bottom to top):
<TeamID>_ProblemA_<testcase#>_<channellayer#>.xls

Rule #8: Along with your solution excel file, you must also provide ONE value of pressure drop between the inlet and outlet of your system (even if you have multiple inlets and outlets in the cavity, the pressure drop between each of them must be identical).

Rule #9: To all teams: please provide all your excel sheet submissions (named according to the rules) for all test cases in a SINGLE folder and all the pressure drops corresponding to each test case in a SINGLE text file inside the same folder.

Hints:

1. First, You must solve the flow rates into and out of each fluid cell from/to neighboring fluid cells based on the discretization above, and using the formula below (for simplicity of modeling, we assume Darcy-Weisbach friction factor for developed laminar flow):

   Here, subscript $i, i+1$ refer to the pressure drop between two neighboring cells with indices i and $i+1$, and the flow rate from cell i to the neighboring cell $i+1$ (please note that the indices are only illustrative,

and do not refer to which direction we are moving from the current cell to the neighboring cell- could be north, south, east or west). The value of the k or the method to compute is provided for each test case separately. Please see the test case description pdfs.

Essentially, you must construct and solve a simple "pressure-flow network" problem which looks like a resistive network where Pressures are voltage and flow rates are currents, according to the above formula and get all local flow rate for each fluid-fluid interface of each fluid cell in your network (see Examples 1-3). Describing an arbitrary network of microchannels here can be easily accomplished using Graph Theory, by building an "Directed Incidence matrix" Z where the rows indicate all the liquid cells/nodes and the columns containing +1/-1/0 indicate the presence/absence of a connection between 2 nodes (channel) and the assumed direction of flow (`https://reference.wolfram.com/language/ref/IncidenceMatrix.html`). This matrix Z would essentially give the relationship between the pressures in each cell/node (with respect to the reference pressure at the inlet) and the pressure drops across each edges/channel segments as follows:

By combining this matrix Z with the known linear relationship between pressure drop across each "edge" and the flow rate between them (Equation 1 above), you can easily solve the flow rates in the entire network using sparse matrix inversion.

2. Once the local fluid flow rates and directions are computed, you could tweak the 3D-ICE (4RM) model based on the same discretization of 100μm $\times$ 100μm for all layers. This will require some minor coding of the 3D-ICE program (if you plan to use it). For each fluid-solid interface in your problem, use the following formula to calculate the local convective conductance term in the equivalent thermal RC grid construction (see [7] for more details):

   $A_{interface}$ refers to the area of the wall interface of the fluid cell towards a particular neighboring solid cell. The value of/the method to compute the heat transfer coefficient h_{conv} will be provided for each test case separately. For simplicity of modeling, we assume Shah-London heat transfer correlation for fully developed laminar flow in all our problems.

3. Once you build the channel network flow rate and the heat transfer models as described above, your system model is ready. Now you use this model to search the design space of channel networks (by respecting the various constraints) using any method you like (Gradient descent methods, Monte-Carlo, Simulated Annealing, Genetic algorithms etc.) for an optimal design.

4. As a first attempt, if building the channel network model is too complex, you could simplify in various ways. For example, in the new heat transfer model, you could only consider the vertical thermal resistances and neglect the North-South-East-West resistances in the first pass, since bulk of the heat enters the channels from the Top and the Bottom. The optimal channel solution you obtain would be pretty close to the one you would find if you included these lateral resistances in a network. Once you find this approximate solution, you can refine your search to find the exact one.

5. There have been works done in the past on finding optimal microfluidic networks to minimize various cost functions, similar to this problem. Some of them are [8, 9, 10]. You could adapt the principles described in these papers to find the optimal design for this problem.

3. CONCLUSIONS

We hope that based on the submissions of the contestants, novel algorithms for searching the vast design-space of microfluidic channel networks can be identified, reinvigorating the research in the field of microchannel liquid cooling for 3D ICs and its implications for computer-aided design.

4. ACKNOWLEDGMENTS

The authors would like to thank Thomas Brunschwiler of IBM Research-Zurich for his valuable insights and suggestions that helped formulate this problem for ICCAD CAD Contest 2015.

5. REFERENCES

[1] T. Brunschwiler et al. Interlayer cooling potential in vertically integrated packages. *Microsyst. Technol.*, 15(1):57 – 74, 2009.

[2] A. Sridhar, A. Vincenzi, D. Atienza, and T. Brunschwiler. 3d-ice: A compact thermal model for early-stage design of liquid-cooled ics. *Computers, IEEE Transactions on*, 63(10):2576–2589, Oct 2014.

[3] M.M. Sabry, A. Sridhar, Jie Meng, A.K. Coskun, and D. Atienza. Greencool: An energy-efficient liquid cooling design technique for 3-d mpsocs via channel width modulation. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 32(4):524–537, April 2013.

[4] Bing Shi, A. Srivastava, and A. Bar-Cohen. Hybrid 3d-ic cooling system using micro-fluidic cooling and thermal tsvs. In *VLSI (ISVLSI), 2012 IEEE Computer Society Annual Symposium on*, pages 33–38, Aug 2012.

[5] A.K. Coskun, D. Atienza, T.S. Rosing, Thomas Brunschwiler, and Bruno Michel. Energy-efficient variable-flow liquid cooling in 3d stacked architectures. In *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, pages 111–116, March 2010.

[6] 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 *Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2011 27th Annual IEEE*, pages 116–124, March 2011.

[7] 3D-ICE. `http://esl.epfl.ch/3D-ICE`.

[8] Tijs Van Oevelen and Martine Baelmans. Numerical topology optimization of heat sinks. In *Proceedings of the 15th International Heat Transfer Conference*, pages 10–15, 2014.

[9] Yongcun Zhang and Shutian Liu. Design of conducting paths based on topology optimization. *Heat and Mass Transfer*, 44(10):1217–1227, 2008.

[10] Peng Xu and Boming Yu. The scaling laws of transport properties for fractal-like tree networks. *Journal of applied physics*, 100(10):104906, 2006.