A Virtual Integration Platform for 3DIC Design Space Exploration

William Wahby, Li Zheng, Yang Zhang, and Muhannad Bakir

***Abstract*—An integrated virtual platform for 3DIC system evaluation and design space exploration is presented. The virtual platform is implemented in MATLAB, and is composed of several simulation modules, including a compact 3DIC wire length distribution, a wire pitch and repeater insertion module, a 3DIC power supply noise estimation module, and a finite difference thermal simulator. The virtual platform is validated against several conventional 2D microprocessors, and predictions are made regarding 3D implementations of those processors.**

I. Introduction (1/2 page)

1) 3DICs are an attractive option for performance gains as 2D scaling slows due to physics and cost constraints
2) The 3DIC design space is not well understood
   a) There are many new tradeoffs that must be considered when designing a 3D system
      i) Signal delivery, power consumption, power delivery, and thermal management must all be considered simultaneously
   b) 3DICs encompass a broad spectrum of possible design choices and integration methodologies, as shown in Fig. 1 each with unique costs and strengths.
      i) Conventional 2D
      ii) 2.5D - die on interposer
      iii) 3D die stacking with TSVs
      iv) Monolithic 3D
         A) Blocks stacked on blocks
         B) Gates stacked on gates
         C) NFETs and PFETs on separate levels
   c) Additionally, different technologies must be evaluated for use in both 2D and 3D ICs
      i) Alternate wiring materials must be considered to improve RC delay, or to reduce electromigration concerns
      ii) Fludic cooling can be used to mitigate thermal challenges in high performance 2D and 3D ICs
3) In order to understand the 3D design space, all of these concerns must be modeled simultaneously
4) We present an integrated virtual platform for 3DIC co-design to enable rapid exploration of the 3DIC design space
5) The virtual platform consists of several simulation modules specially tailored for 3DICs, and interconnected to allow signal, power, and thermal codesign
   a) Novel compact stochastic wire length model which accounts for TSV-induced logic displacement
   b) Wire pitch and repeater insertion models which allow accurate determination of the number of metal layers
      i) These models are important for determining the delay and power consumption of the final system
   c) Power delivery models which account for the impact of simultaneous switching noise in 3DICs with TSVs
   d) A multiscale finite difference thermal module which rapidly and accurately determines the thermal profile in each layer of a 3DIC
6) To explore the capabilities of our simulation platform we will use it to investigate the impacts of advanced technologies and integration methodologies on a Sandy Bridge i7 processing core.

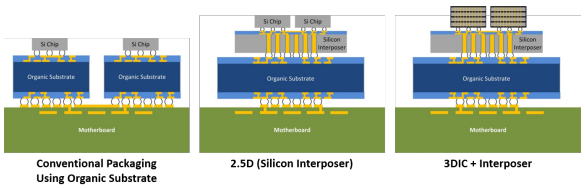

Fig. 1. There are many potential configurations for 3DICs, each with their own costs and advantages. Designers must manage the complexity of the 3DIC design space in order to achieve higher performance and lower cost systems.

II. Virtual Design Platform (1.5 pages total)

1) We have developed an integrated virtual platform for 3DIC design space exploration
2) The virtual platform consists of the following
   a) 3D wire length distribution which properly accounts for TSV area
   b) Metal layer pitch determination algorithms capable of handling alternate wiring materials
   c) A sub-optimal repeater insertion scheme
   d) Power supply noise models which account for power delivery in 3DICs
   e) A finite difference thermal module for analyzing the thermal impacts of 3D integration
3) The overall execution flow is shown in Fig. 2.
4) The 3DIC simulation platform can be used to compare the properties of 2D and 3D chips
   a) Stacked-die 3D and monolithic 3D can be considered, as shown in Fig. 3
5) Currently, the interconnections between blocks are not modeled when simulating heterogeneous 3D stacks.

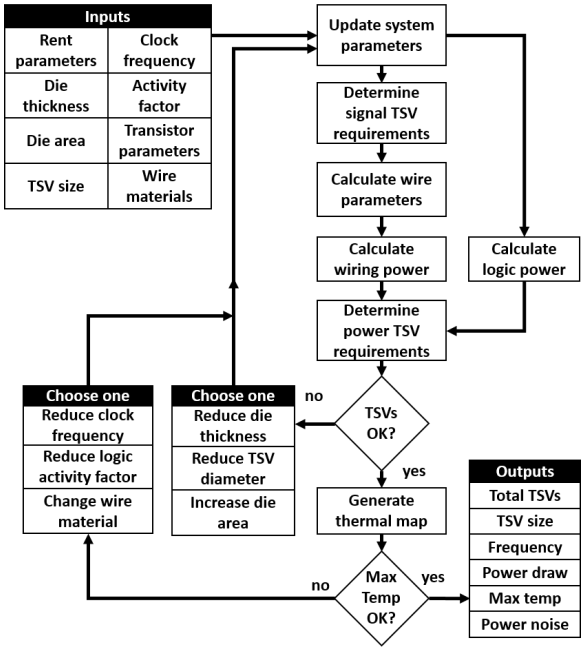

Fig. 2. Block diagram of the virtual platform execution flow.

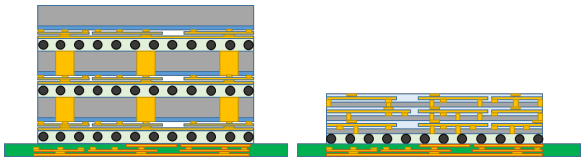

Fig. 3. 3DICs implemented with die stacking and TSVs (left) vs monolithic 3D technology (right).

A. Interconnect modeling (1/2 page)

1) Stochastic wire length distributions have been shown to be effective tools for the rapid prediction of interconnect properties in 2D and 3D ICs [1]–[5].
2) Efforts have been made to extend 2D wire length distributions to 3DICs [6]–[8].
3) We recently developed a compact model for the wire length distribution in 3DICs which properly accounts for the displacement of logic gates by TSVs [9].
4) Stochastic wire length distributions determine the number of wires of a particular length as

$$I_{idf}(l) = M(l)P_c(l) \tag{1}$$

5) Explanation of M and P_c
6) The general method of determining M and P in 2D ICs is laid out in [1].
7) The method was extended to consider 3DICs in [6], [7].
   a) The impact of TSV-induced gate blockage was not considered until [8] introduced a modified form of M which accounted for logic gate displacement.
   b) This method required brute-force calculation of the correction terms.
8) We recently developed a novel compact correction for M which includes the impact of TSV-induced gate displacement [9]
9) This work uses expressions for M and P from [7] with the compact correction developed in [9].
10) For a symmetric square chip, with a periodic square array of TSVs, the corrected gate pair function is [9]:

$$M_t^*(l) \cong M_t^o(l, N_s) - 2g * g + N_{tsv} M_t^0(l, N = w_{tsv}^2) \tag{2}$$

11) Explanation of terms
12) By substituting Eq. (2) for the 2D gate placement function in [7], the impact of TSV-induced logic displacement can be modeled
13) The wire pitch and number of metal layers are determined using a bottom-up wire scaling technique [3].
14) An optimal repeater insertion scheme is used to determine the size and number of repeaters to insert to meet the target delay.

B. Material considerations (1/8 page)

1) As interconnect dimensions continue to scale, the properties of wiring materials have begun to diverge from their bulk values.
2) Surface scattering and grain boundary scattering greatly enhance the resistivity of nanoscale copper wires [10].
3) In order to capture the impact of these effects on wire resistivity, we use a combined Mayadas-Shatzke and Fuchs-Sondheimer (MS+FS) model [10]–[13].
   a) Specularity of 0.55 and backscattering probability of 0.43 [10].
4) The grain size is approximated as the smallest dimension of the wire under consideration.

C. Power supply noise modeling (1/8 page)

1) Power supply noise must be suppressed to ensure reliable system operation
2) Power delivery in 3DICs is complicated by the limited area available for routing power interconnects between tiers.
3) We use the 3DIC power supply network models developed in [14], [15] to determine the maximum power supply noise in the 3D stack and the number of TSVs required for power delivery.

D. Thermal modeling (1/2 page)

1) Thermal issues are one of the greatest challenges in 3DIC design.
2) In order to design a thermally robust 3D system, the relationships between device technology, system performance area constraints, and packaging materials and technology must be explored.
3) We utilize a fast and accurate finite difference thermal model, as shown in Fig. 4

a) The heat transfer equation is

$$\nabla\left(K(x,y,z)\cdot\nabla\left(T(x,y,z)\right)\right)=P(x,y,z) \quad (3)$$

b) Following the scheme of [16], we discretize the heat transfer equation as

$$\frac{T_{i,j,k}-T_{i-1,j,k}}{\frac{x_1}{k_z l_y l_x}}+\frac{T_{i,j,k}-T_{i+1,j,k}}{\frac{x_2}{k_x l_y l_z}} + \frac{T_{i,j,k}-T_{i,j-1,k}}{\frac{y_1}{k_y l_x l_z}}+\frac{T_{i,j,k}-T_{i,j+1,k}}{\frac{y_2}{k_y l_x l_z}} + \frac{T_{i,j,k}-T_{i,j,k-1}}{\frac{z_1}{k_z l_x l_z}}+\frac{T_{i,j,k}-T_{i,j,k+1}}{\frac{z_2}{k_z l_x l_z}}=P_{tot} \quad (4)$$

c) A similar scheme can be derived for grid points along the boundary of the mesh.

d) In the case that one mesh has multiple materials, we assign the volume-weighted thermal conductivity of the surrounding volume to that grid point.

4) The accuracy of this finite difference module was assessed in [17], in which the performance of the finite difference scheme was compared against finite element ANSYS models of the same structure. The finite difference model was found to match the ANSYS results with a maximum error of 2.7%.

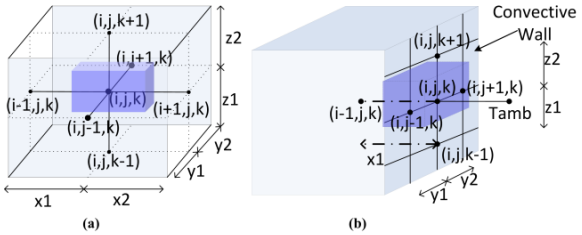

Fig. 4. Finite difference scheme. (a) Points inside the stack. (b) Boundary points at the face of the stack [16].

III. Validation (1/2 page)

1) The virtual platform was validated by comparing its predictions against published data for Intel processors ranging from the 65nm node to the 32nm node.
   a) The chip area, number of logic transistors, number of memory transistors, and size and shape of the cores and memory blocks were gathered from published data.
   b) Logic cores were simulated with a Rent exponent of 0.6, while memory cores used a Rent exponent of 0.4 [18].
   c) The expected wire pitch, number of metal layers, power consumption, and maximum junction temperature generated by the virtual platform have been compared in Table I.
2) The virtual platform shows good agreement with the high level design parameters for processors ranging from 65nm down to 32nm.
3) Wire pitch can be accurately predicted in each level, as shown in Fig. 5.

TABLE I. Comparison to actual data

Processor	Node	Wiring Tiers Actual	Wiring Tiers Predicted	TDP (W)	Predicted Power (W)
Core 2 Duo E6850	65nm	8	8	65	59.9
Core 2 Duo E8600	45nm	9[1]	8	65	59.5
Core i7 880	45nm	9[1]	7	95	103
Core i7 680	32nm	9[1]	8	73	59.2
Core i7 2700k	32nm	9[1]	8	95	116

[1] Includes one global metal layer for power distribution, not modeled here

4) The thermal module allows fine-grained simulation of the thermal distribution throughout a 2D or 3D stack, as shown in Fig. 6.

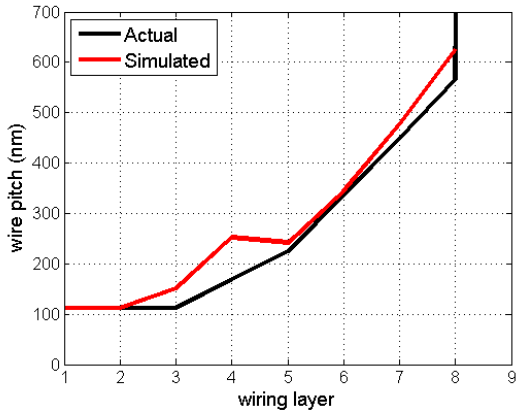

Fig. 5. Comparison of actual and expected wire pitch in a Core i7 2700k processor.

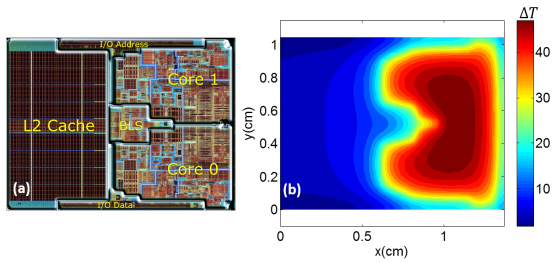

Fig. 6. Thermal simulation of a 65nm Merom processor. (a) Actual die. (b) Resulting thermal map after simulation with the virtual platform.

IV. 2D: Seeking improvements via materials innovation (3/4 page)

1) One path towards increasing system performance is to achieve improvements in the wiring materials
2) Wire resistivity and ILD permittivity determine the resistance and capacitance of on-chip wires
   a) Lower R and C can lower RC delay, enabling faster operation or reducing the need for power-hungry repeaters
   b) Lower wire capacitance reduces the power consumption of the wires themselves

3) Electromigration becomes a significant concern for system reliability at advanced process nodes (CITE)
4) In order to alleviate electromigration concerns at advanced process nodes, alternate materials may be required (CITE Leti)
5) Electromigration-resistant materials may have higher resistivities, potentially impacting signal performance, power consumption, and number of metal layers required
6) We will investigate the impacts of each of these effects on the wiring network of a single CPU core from an Intel Core i7 2700k
   a) Case 1: ILD permittivity reduced by X%
   b) Case 2: (Bulk) Metal resistivity reduced by Y% (Including size effects)
   c) Case 3: (Bulk) Metal resistivity increased by Z% for low-pitch layers (Including size effects)
7) In Fig. 7 we show the impact of material parameters on the number of metal levels in the system
8) In Fig. 8 we show the impact of material parameters on the power consumed by wires and repeaters in the system.
9) Discuss how much improvement in ILD permittivity is needed to offset increased wire permittivity

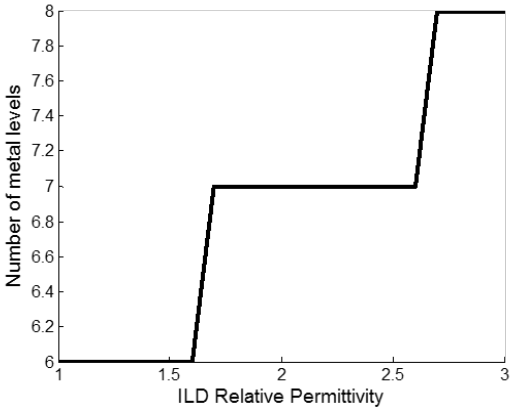

Fig. 7. Impact of material parameters on number of metal layers required to route the wires in a Sandy Bridge Core i7 2700k.

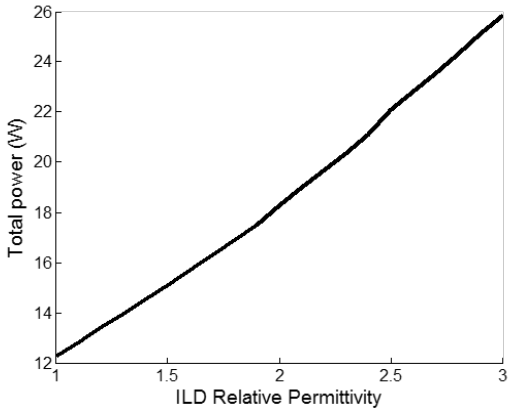

Fig. 8. Impact of material parameters on power consumed by wires and repeaters in a Sandy Bridge Core i7 2700k.

V. 3D: Finding equivalent power and performance gains without exotic materials (2 pages)

A. Reducing power consumption

1) 3D integration offers another potential route for realizing reductions in system power
   a) Reduction in wire length from 3D enables lower capacitance wiring networks
      i) Reduces overall delay, and therefore reduces the demand for repeaters
      ii) Directly reduces the power consumed by wires, potentially enabling increases in operating frequency
2) To examine the impact of 3D integration on power consumption, a single 18.5mm^2 Sandy Bridge processing core was simulated in different 3D configurations.
   a) TSV aspect ratio of 20:1
   b) The total area occupied by TSVs was limited to 10% of the overall die area
      i) Typical 3DIC designs focus on limiting the area consumed by TSVs to 1% or less to minimize their consumption of active area
      ii) Typical 3DIC designs currently focus on wide-io configurations as the large size of TSVs renders true point-to-point interconnection infeasible.
      iii) We are considering a more extreme case in order to understand the behavior of a monolithic or near-monolithic 3D design
      iv) The use of high aspect ratio TSVs and a large TSV area allocation allows the Sandy Bridge core to be implemented in 3D without extreme die thinning, as seen in Fig. 9b.
   c) The core is assumed to be partitioned into N equal pieces, which are then stacked vertically.
3) Significant power savings can be obtained by moving to a 3D design, as shown in Fig. 9a, at the cost of increased areal power density.
4) In order to realize the greatest power reduction from 3D

integration, the active layers should be as thin as posible

a) Thicker logic tiers reduce the wire length reduction achieved by moving to 3D
b) Monolithic 3D designs can significantly reduce the power consumed by the wiring network, as shown in Fig. 10.

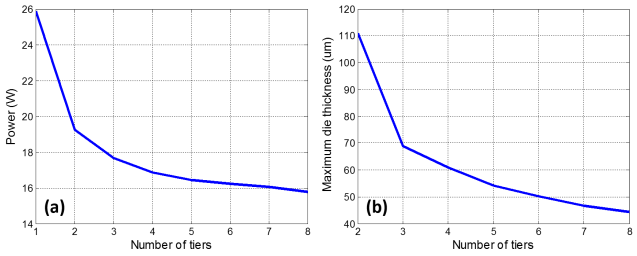

Fig. 9. (a) Total power consumption of a single Sandy Bridge (Core i7 2700k) processing core implemented with increasing levels of 3D integration. Significant power savings can be obtained by implementing this core in 3D. (b) Maximum allowable die thickness to meet the 10% die area limit on signal TSV usage with 20:1 aspect ratio TSVs. The total power consumption data in (a) was determined using the die thickness results from (b).

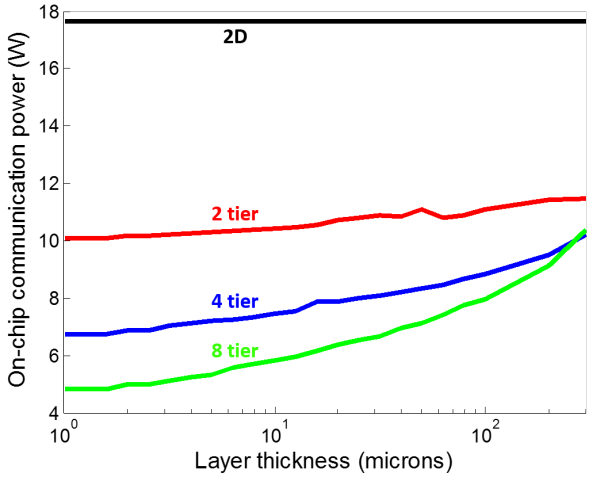

Fig. 10. Impact of die thickness on power consumed by wires in a single 32nm Sandy Bridge core implemented with increasing numbers of logic tiers. Very short distances between logic tiers can dramatically reduce the power consumed in the on-chip communication network.

B. Power Delivery

1) The total power consumption of a system can be reduced by moving to a 3D configuration
2) The areal power density increases as more tiers are stacked atop one another
3) As the power density increases, power delivery becomes more of a challenge
4) More TSVs may be needed to deliver power in 3D stacks
   a) Power TSVs compete with signal TSVs for already-limited silicon real estate
5) In order to mitigate the challenge of power delivery in 3DICs, thinner active tiers can be used, as shown in Fig. 11
   a) In this case we keep the TSV width constant and let the TSV length grow, which leads to unrealistic TSV dimensions for the thicker dice
   b) If we fixed the aspect ratio, the consumption of silicon area would be too extreme to result in a realizable device
   c) The power TSV requirements shown in Fig. 11 come in addition to the signal TSV reqirements in Fig. 9b.
6) Alternately, portions of the die or interposer could be allocated for decoupling capacitors
   a) This sets up a tradeoff between utilizing die area for decap our for power delivery
   b) The impact of decoupling capacitance on the number of TSVs required for power delivery is shown in Figs. 12 and 13.
   c) Interestingly, Fig. 13 shows a tradeoff between the amount of decoupling capacitance used and the number of TSVs required for the power delivery network.
      i) Initially, adding more decoupling capacitance actually degrades the performance of the power delivery network
      ii) After a certain point, adding more decoupling capacitance begins to reduce the number of TSVs required for power delivery.
   d) Can we minimize the area requirements with a mix of decap and power TSVs?

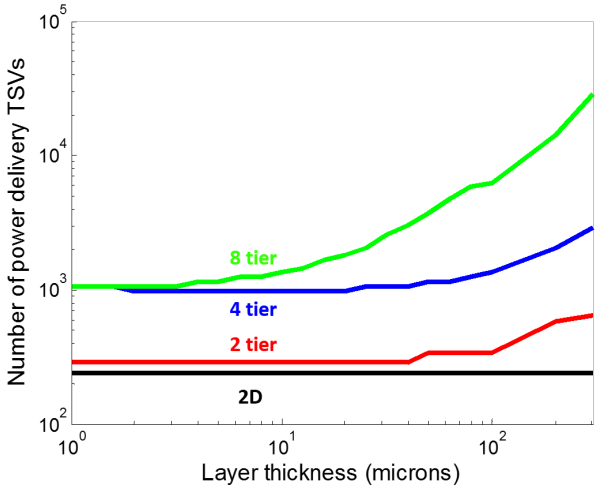

Fig. 11. Power TSV requirements in a single 32nm Sandy Bridge core implemented with increasing numbers of logic tiers. These TSVs are counted separately from the signal TSV requirements in Fig. 9b.

C. Thermal management

1) We have shown that implementing a Sandy Bridge processing core in 3D results in significant wire length and power savings

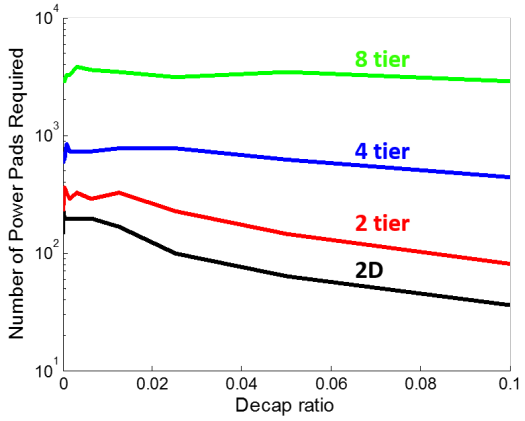

Fig. 12. Power TSV requirements as a function of the area allocated for decoupling capacitors in a single monolithic 3D 32nm Sandy Brige core implemented with 100 μm thick logic tiers. For the 4 and 8-tier stacks negligible improvements are seen with increased area allocation for decoupling capacitance, as the resistance and inductance of the very long power delivery TSVs is challenging to overcome.

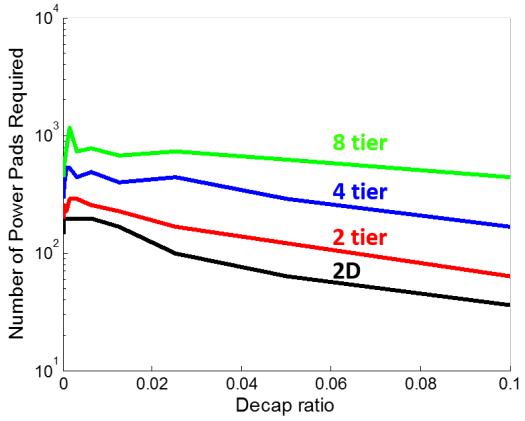

Fig. 13. Power TSV requirements as a function of the area allocated for decoupling capacitors in a single monolithic 3D 32nm Sandy Brige core implemented with 1 μm thick logic tiers. As additional silcion area is allocated for decoupling capacitors, the inductance of the power network is reduced and fewer power and ground TSVs are required to achieve an acceptable Vdd droop.

2) We have also shown that using ultrathin active layers can significantly ease the challenges of power delivery in 3D stacks
3) The key question for 3DICs is thermal management
   a) The total power dissipation is being reduced as the device is partitioned into more and more layers
   b) But the areal power density is increasing, as the power does not drop as fast as the 2D area
   c) This can lead to unreasonable temperatures.
4) We will investigate the impact of alternate cooling strategies on a single Sandy Bridge core implemented in several 2D and 3D configurations, cooled with both air, liquid heat sinks, and a combination of liquid heat sinks and microfluidically-cooled interposers. The different cooling configurations and their performance are shown in Figs. 14 to 16.
   a) 2D with air cooling
   b) 2D with liquid cooled heat sink
   c) die-stacked 3D with air cooling
5) In Figs. 14 to 16 the impact of water cooling on junction temperature for the different die configurations can be seen.
6) The combination of increased power density and reduced thermal contact area for 3DICs overwhelms conventional air-cooled heat sinks
7) By moving to liquid cooling the temperatures of the 3D stacks can be brought in line with acceptable maximum temperatures
8) In Fig. 19 we investigate the maximum operating frequency that can be achieved by the Sandy Bridge core in each configuration with water cooling.

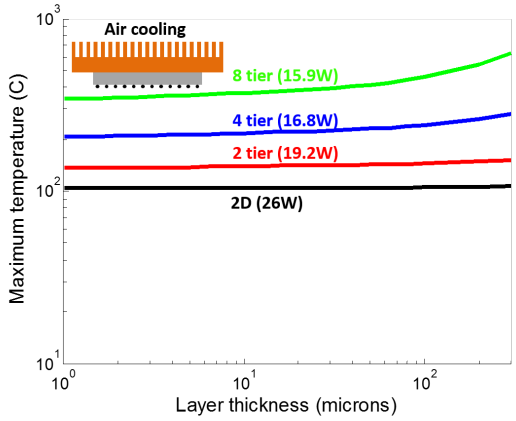

Fig. 14. Maximum temperature for a single Sandy Bridge CPU core implemented in 2D and monolithic 3D configurations, and cooled with a conventional air-cooled heat sink. Peak temperatures in each of the 3D cases significantly exceed feasible limits.

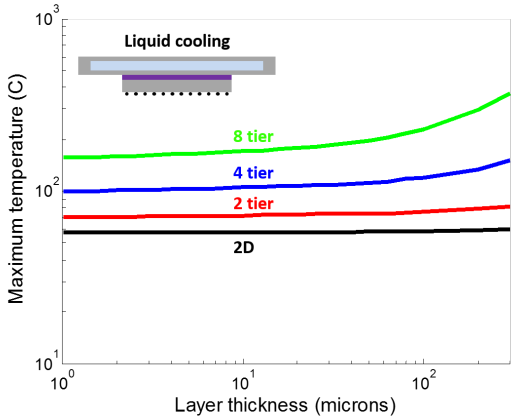

Fig. 15. Maximum temperature for a single Sandy Bridge CPU core implemented in 2D and monolithic 3D configurations, and cooled with a conventional water-cooled heat sink. Water cooling reduces the maximum temperature of the 2 tier 3D stack to acceptable levels, and monolithic 3D 4 tier stacks might be feasible.

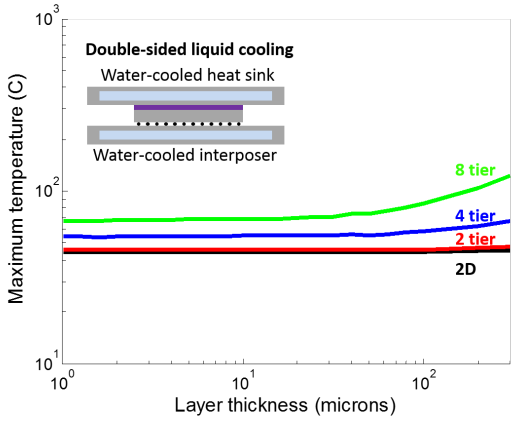

Fig. 16. Maximum temperature for a single Sandy Bridge CPU core implemented in 2D and monolithic 3D configurations, and aggressively cooled with a combination of a conventional water-cooled heat sink and a novel water-cooled interposer. In this configuration, 8 tier 3D stacks are feasible, though the 2D configuration runs much cooler due to its lower areal power density.

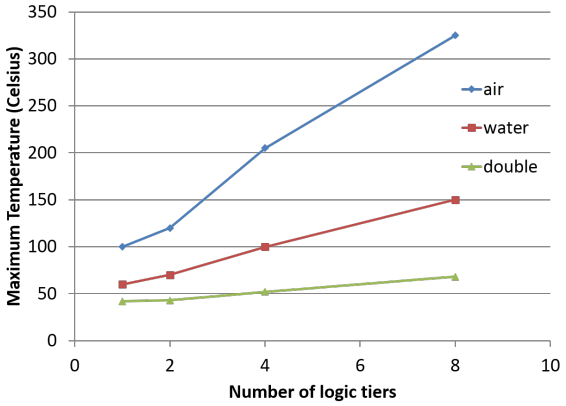

Fig. 17. Impact of alternate cooling strategies on maximum stack temperature. Active layer thickness is assumed to be 1 micron for this comparison.

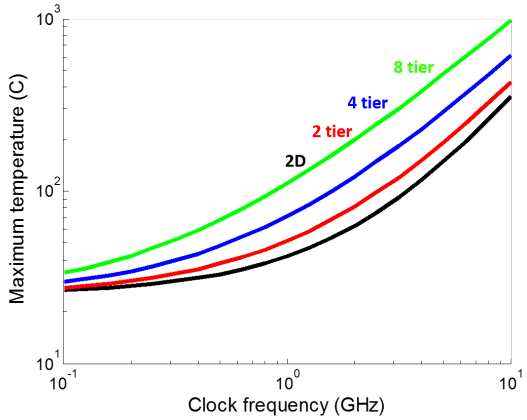

Fig. 18. Relationship between maximum junction temperature and operating frequency for a monolithic 3D 32nm Sandy Bridge CPU core with a conventional air-cooled heat sink.

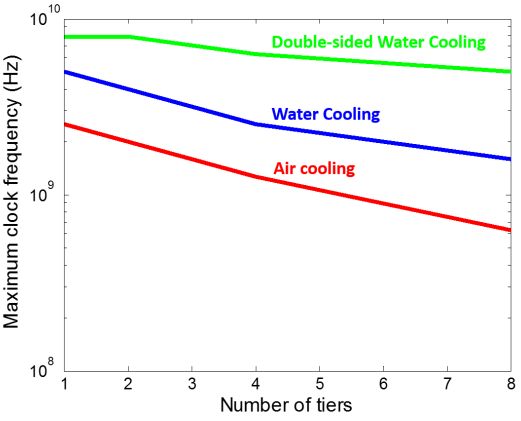

Fig. 19. Maximum operating frequency the Sandy Bridge processing core can achieve while still keeping maximum junction temperature below 90°C. Active layer thickness is assumed ot be 1 μm for this comparison.

VI. Conclusion

Models for signal delivery, power supply noise, and thermal performance in 3DICs were presented and linked into an integrated virtual platform. The virtual platform was validated against wire pitch, power, and thermal data for recent commercial microprocessors, and the benefits of 3D integration were examined for a 32nm processor core.

Acknowledgments

The authors gratefully acknowledge the support of the Semiconductor Research Corporation Global Research Collaboration (task ID 2254.001).

References

[1] J. Davis, V. De, and J. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI). I. Derivation and validation," *IEEE Transactions on Electron Devices*, vol. 45, no. 3, pp. 580–589, 1998.

[2] J. Davis, V. De, and J. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI). II. Applications to clock frequency, power dissipation, and chip size estimation," *IEEE Transactions on Electron Devices*, vol. 45, no. 3, pp. 590–597, 1998.

[3] R. Venkatesan, J. Davis, and J. Meindl, "Performance enhancement through optimal n-tier multilevel interconnect architectures," in *ASIC/SOC Conference, 1999. Proceedings. Twelfth Annual IEEE International*, 1999, pp. 19–23.

[4] J. Joyner, R. Venkatesan, P. Zarkesh-Ha, J. Davis, and J. Meindl, "Impact of three-dimensional architectures on interconnects in gigascale integration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 9, no. 6, pp. 922–928, 2001.

[5] D. C. Sekar, A. Naeemi, R. Sarvari, J. A. Davis, and J. D. Meindl, "IntSim: A CAD Tool for Optimization of Multilevel Interconnect Networks," in *Proceedings of the 2007 IEEE/ACM International Conference on Computer-aided Design*, ser. ICCAD '07. Piscataway, NJ, USA: IEEE Press, 2007, pp. 560–567.

[6] A. Rahman, A. Fan, J. Chung, and R. Reif, "Wire-length distribution of three-dimensional integrated circuits," in *Interconnect Technology, 1999. IEEE International Conference*, 1999, pp. 233–235.

[7] J. Joyner, P. Zarkesh-Ha, J. Davis, and J. Meindl, "A three-dimensional stochastic wire-length distribution for variable separation of strata," in *Interconnect Technology Conference, 2000. Proceedings of the IEEE 2000 International*, 2000, pp. 126–128.

[8] D. H. Kim, S. Mukhopadhyay, and S. K. Lim, "Through-silicon-via Aware Interconnect Prediction and Optimization for 3D Stacked ICs," in *Proceedings of the 11th International Workshop on System Level Interconnect Prediction*, ser. SLIP '09. New York, NY, USA: ACM, 2009, p. 8592.

[9] W. Wahby, A. Dembla, and M. Bakir, "Evaluation of 3DICs and fabrication of monolithic interlayer vias," in *3D Systems Integration Conference (3DIC), 2013 IEEE International*, Oct 2013, pp. 1–6.

[10] T. Sun, B. Yao, A. P. Warren, K. Barmak, M. F. Toney, R. E. Peale, and K. R. Coffey, "Surface and grain-boundary scattering in nanometric Cu films," *Phys. Rev. B*, vol. 81, p. 155454, Apr 2010.

[11] A. F. Mayadas and M. Shatzkes, "Electrical-Resistivity Model for Polycrystalline Films: the Case of Arbitrary Reflection at External Surfaces," *Physical Review B*, vol. 1, no. 4, pp. 1382–1389, Feb. 1970.

[12] K. Fuchs, "The conductivity of thin metallic films according to the electron theory of metals," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 34, pp. 100–108, 1 1938.

[13] E. Sondheimer, "The mean free path of electrons in metals," *Advances in Physics*, vol. 1, no. 1, pp. 1–42, 1952.

[14] L. Zheng, Y. Zhang, G. Huang, and M. Bakir, "Novel Electrical and Fluidic Microbumps for Silicon Interposer and 3-D ICs," *Components, Packaging and Manufacturing Technology, IEEE Transactions on*, vol. 4, no. 5, pp. 777–785, May 2014.

[15] G. Huang, M. Bakir, A. Naeemi, and J. Meindl, "Power Delivery for 3-D Chip Stacks: Physical Modeling and Design Implication," *Components, Packaging and Manufacturing Technology, IEEE Transactions on*, vol. 2, no. 5, pp. 852–859, May 2012.

[16] J. Xie and M. Swaminathan, "Electrical-Thermal Co-Simulation of 3D Integrated Systems With Micro-Fluidic Cooling and Joule Heating Effects," *Components, Packaging and Manufacturing Technology, IEEE Transactions on*, vol. 1, no. 2, pp. 234–246, Feb 2011.

[17] Y. Zhang, H. Oh, and M. Bakir, "Within-tier cooling and thermal isolation technologies for heterogeneous 3D ICs," in *3D Systems Integration Conference (3DIC), 2013 IEEE International*, Oct 2013, pp. 1–6.

[18] P. Christie and D. Stroobandt, "The interpretation and application of Rent's rule," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 8, no. 6, pp. 639–648, 2000.