A Simulation Tool for Rapid Investigation of Trends in 3-DIC Performance and Power Consumption

William Wahby, *Student Member, IEEE*, Li Zheng, Yang Zhang, *Student Member, IEEE*, and Muhannad S. Bakir, *Senior Member, IEEE*

***Abstract*— In order to compare the costs and the benefits of 2-D and 3-D integrated circuits (3-DICs) technologies, a compact simulation tool for 3-DIC system evaluation and design space exploration is presented. The simulator is implemented in the MATLAB, and is composed of several modules, including a compact 3-DIC wire-length distribution, a wire pitch and repeater insertion module, a 2-D and 3-DIC power supply noise estimation module, and a finite-difference thermal simulator. The simulator is validated against published data for several commercial 2-D processors at the 65-, 45-, and 32-nm nodes. In order to quantify the benefits of both 2-D and 3-D integration approaches, a 32-nm CPU core is modeled, and the impact of several technology parameters, including interlayer dielectric material, on-chip wire material, die thickness, and cooling solution, is explored. The results suggest that the 3-D integration may provide a significant power reduction for the 32-nm test case, but more aggressive cooling solutions must be employed to maintain the same clock frequency due to the increased areal power density of the 3-D CPU.**

***Index Terms*— 3-D integrated circuits (3-DICs), IC interconnections, IC modeling, power dissipation.**

I. Introduction

ECONOMIC and physical challenges to transistor scaling are driving interest in 3-D integration, but uncertainty regarding the fabrication costs and the system-level tradeoffs of 3-D integration complicates the design of 3-D integrated circuits (3-DICs). Projections of 3-DIC cost and performance are further impacted by the strongly coupled nature of communication, power delivery, and thermal management in 3-DICs. The 3-DIC design space is complex, as 3-DIC design encompasses a broad spectrum of possible design choices and integration methodologies, ranging from 2.5-D interposer-based integration all the way to finely grained monolithic 3-DICs, as shown in Fig. 1, each with unique costs and strengths. In addition, different technologies must be evaluated for use in both 2-D and 3-DICs. Low-k dielectrics can be used to reduce the parasitic capacitances in the wiring stack, simultaneously improving RC delay and reducing the power consumption of the wiring network. Alternate wiring materials are also being considered to improve the RC delay of on-chip interconnects, as well as to reduce the impact of electromigration [1]. Liquid cooling can be used to mitigate the thermal challenges in high-performance 3-DICs [2], [3]. In order to understand when a 3-D system might have advantages over a 2-D system, all of these factors must be modeled simultaneously. Performing these coupled simulations with high fidelity is computationally intensive, however, rendering thorough exploration of the 3-DIC design space challenging. Compact tools have been developed to estimate the 2-DIC performance [4], [5], but no analog exists for 3-DICs. In addition, the existing 2-D tools do not model power delivery or heat extraction, which are likely to be challenges for 3-DICs.

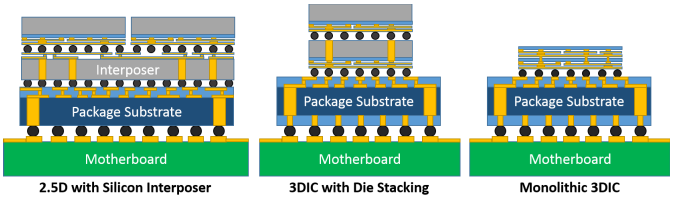


Fig. 1. There are many potential configurations for 3-DICs, each with their own costs and advantages. Designers must manage the complexity of the 3-DIC design space in order to achieve higher performance and lower cost systems.

The novel contribution of this paper is the development of a compact tool capable of rapidly simulating the properties of 3-DIC on-chip signal and power delivery networks. The simulator estimates the power consumption, the simultaneous switching noise (SSN), and the thermal profile of 2-DICs and 3-DICs, and enables the investigation of trends in performance and power consumption, as materials, devices, and integration methodologies are varied. The details of the on-chip wires, Through Silicon Vias, (TSVs), power delivery network, and steady-state thermal profile are all modeled in order to develop a self-consistent picture of the overall 3-D system performance. The simulator is intended to help answer what–if questions and to guide more precise studies of detailed 3-DIC physical design.

The organization of this paper is as follows. In Section II, the details of the models used and their interactions are presented. In Section III, the simulator is benchmarked against wire pitch and Thermal Design Power (TDP) data from several commercial processors. In Sections IV and V, the simulator

Manuscript received July 11, 2015; revised November 19, 2015; accepted November 21, 2015. Date of publication January 17, 2016; date of current version February 5, 2016. This work was supported in part by Semiconductor Research Corporation through the Global Research Collaboration Project under Grant 2254.001, and in part by the SRC Educational Alliance Intel Foundation Fellowship. Recommended for publication by Associate Editor E.-P. Li upon evaluation of reviewers' comments.

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: wwahby3@gatech.edu; lizheng@gatech.edu; steven.zhang@gatech.edu; muhannad.bakir@mirc.gatech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCPMT.2015.2511082

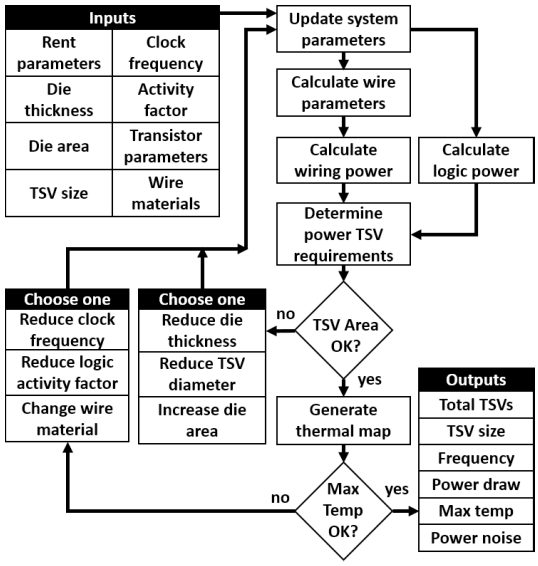


Fig. 2. Block diagram of the simulation platform execution flow.

is used to investigate the impacts of advanced technologies on a 32-nm CPU core. Specifically, the impacts of the interlayer dielectric (ILD), wire material, and 3-D stacking on the power consumption, power supply noise, and metal layers required for routing are investigated.

II. Simulation Framework

The simulation platform consists of the following.

1) 3-D wire-length distribution that accounts for TSV area.
2) Metal layer pitch determination algorithms capable of handling alternate wiring materials.
3) An optimal repeater insertion scheme.
4) A power supply noise model for 3-DICs.
5) A finite-difference thermal module for analyzing the thermal impacts of 3-D integration.

The simulation flow is shown in Fig. 2. First, the distribution of wire lengths in the system is estimated, which is in turn used to determine the number of wiring tiers required for signal routing, the wire pitch on each tier, and the number of repeaters needed to meet the delay constraint. In order to model the on-chip interconnects in a compact manner, the system is modeled as a homogeneous block of randomly interconnected logic. This assumption is commonly used with stochastic wire-length models for the rapid estimation of on-chip interconnect properties, at the cost of reduced insight into fine-grained design details [5]–[7]. In heterogeneous systems, each block is modeled separately, and the results are assembled into an overall power density map of each tier in the design to determine the thermal profile throughout the stack. In order to better predict the performance of such systems, the method of [8] can be used to homogenize a heterogeneous system.

Once the parameters of the on-chip interconnects are known, their power consumption can be estimated. The transistor dynamic and leakage power is also calculated to determine the overall power requirements for the design. The total power consumption is used in conjunction with the TSV and package pin resistance and inductance to estimate the SSN in the power delivery network. Additional power pads and TSVs are inserted until the noise drops to acceptable levels. At this point, the tool checks that the total TSV area does not exceed a user-specified limit; if TSV demand outstrips available area, the TSV diameter is reduced, and the interconnect and power modules are rerun.

Once the design passes the TSV area check, the thermal module is used to estimate the maximum temperature in the 2-D or 3-D design. In order to do this, the material parameters of the die, wiring tiers (including wires and ILD), TSVs, and interstitial layers are input into a finite-difference thermal simulator, along with a heat transfer boundary condition representing the heat sink. If the maximum temperature in the stack exceeds a user-defined limit, the clock frequency is reduced, and the previous modules are rerun until the maximum temperature is sufficiently reduced. As alternate means of power reduction, the logic activity factor can also be reduced or the wire or ILD materials be modified. The final design parameters are reported once the constraints on temperature and TSV area are satisfied.

The key limitations of this simulation platform are linked to the assumptions used to model the on-chip interconnects. Treating a design as a homogeneous block of randomly interconnected logic allows the rapid simulation of the aggregate properties of the on-chip wires, but limits visibility into the inner workings of the design. Off-chip IO and interconnections between blocks in heterogeneous systems are also not modeled, as design-specific information would be needed for accurate modeling. These omissions lead to the underestimation of the system power, and limit the range of validity of the simulator, but they also enable the consideration of a wide range of system types without detailed design information.

A. Interconnect Modeling

The on-chip interconnect network accounts for a significant fraction of the overall power consumption in many logic designs [5]; accordingly, the accurate determination of the parameters of the on-chip interconnect network is crucial for developing reasonable estimates of the overall power consumption and operation frequency of an IC. Specifically, the number of metal layers used for wiring must be estimated, as well as the pitch of the wires on each level, in order to determine the overall wiring capacitance, which is a key factor in the maximum wire delay and power consumption.

Stochastic wire-length models have been shown to be effective tools for the rapid prediction of interconnect properties in 2-DICs [5], [6] and 3-DICs [7], but the impact of TSV-induced gate-blockage in 3-DICs was not considered until Kim *et al.* [9] introduced a correction to account for finite TSV size. The method of [9] requires brute-force calculation, however, which is too computationally intensive for rapid simulation. To address this issue, a compact correction to the wire-length distribution was introduced in [10], which is used in this paper. Once the wire-length distribution is known, the wire pitch and the number of metal layers are determined using a bottom–up wire scaling technique [4], and a delay-optimal

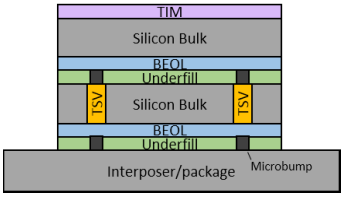


Fig. 3. Geometry used in the thermal module. TSVs are individually meshed, in order to better capture the impact of 3-D heat transport.

repeater insertion scheme is used to determine the size and the number of repeaters required to meet timing constraints [11].

The impact of surface scattering and grain-boundary scattering on wire resistivity is incorporated into the wire sizing algorithm with a combined Mayadas–Shatzke and Fuchs–Sondheimer model, with the specularity of 0.55 and the backscattering probability of 0.43 [12]. The metal grain size is approximated as the smallest dimension of each wire.

B. Power Supply Noise Modeling

Power supply noise must be suppressed to ensure reliable system operation, but power delivery in 3-DICs is complicated by the limited area available for routing power interconnects between tiers and by the additional parasitic resistance and inductance of the TSVs used for power delivery. In order to determine the maximum allowable TSV diameter, the number and size of the power delivery TSVs must be estimated. The analytic 3-DIC power supply model developed in [13] is used to estimate the SSN in the 3-D stack as a function of the number and size of the power delivery TSVs. The model uses the periodicity of the power grid to extrapolate the worst case SSN in the system from the detailed frequency-domain behavior of a single power delivery unit cell. Power is assumed to be delivered via a regular rectangular array of power and ground pads or TSVs connected by planar power delivery wires. An inverse Laplace transform is used to convert the frequency response of the power delivery network into the temporal response, from which the worst case noise can be easily extracted [14].

C. Thermal Modeling

Thermal issues are one of the greatest challenges in 3-DIC design. In order to design a thermally robust 3-D system, the relationships between device technology, system performance area constraints, and packaging materials and technology must be explored. To that end, we utilize a fast and accurate finite-difference thermal model, described in [15], with a nonconformal meshing strategy described in [16].

The thermal configuration considered in the simulation tool is shown in Fig. 3. One or more dice are assumed to be stacked vertically atop an interposer (which may be either a conventional organic package substrate, or a silicon or glass interposer). Each die is separated into three regions: 1) the bulk silicon; 2) the wiring stack, which is modeled as the volume-weighted average of the thermal conductivities of the wiring material and the ILD; and 3) the underfill material between each die. The power dissipated by each die is applied as an excitation at the boundary between the bulk silicon and the wiring stack. TSVs are modeled within the bulk of any dice below the top die in the stack. Each external boundary is modeled with a convective boundary condition; within-package boundaries are typically given a low heat transfer coefficient of 5 mW/m^{2}K, while the top and bottom surfaces are given higher values to reflect the cooling method used.

The accuracy of this finite-difference module was assessed in [16], 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%.

TABLE I
COMPARISON WITH ACTUAL DATA

Processor	Node	Signal Wire Tiers Actual	Signal Wire Tiers Predicted	TDP (W)	Predicted Power (W)
E6850	65nm	8	8	65	60.27 (-7.3%)
E8600	45nm	8[2]	8	65	63.62 (-2.1%)
i7 880	45nm	8[2]	7	95	105.56 (11.1%)
i7 680[1]	32nm	8[2]	6	73	52.74 (-27.8%)
i7 2700k	32nm	8[2]	8	95	91.80 (-3.3%)

[1] Multi-chip package with 32nm CPU die and 45nm gpu/support die.
[2] Design has one additional global metal layer for power distribution.

III. VALIDATION

The simulator was validated by comparing its predictions against published data for Intel processors ranging from the 65- to 32-nm node [17]–[24]. For each test case, the chip area, the number of logic transistors, the number of memory transistors, and the size and shape of the cores and memory blocks were gathered from published data. Logic cores were simulated with a Rent exponent of 0.6, while memory cores used a value of 0.4, and GPUs were simulated with a Rent exponent of 0.5 [25]. Each block was simulated separately to determine the number and pitch of metal levels required for routing and the total power consumption of each block. The pitch and the number of metal layers used for the overall design were then set by the block, which required the greatest number of wiring tiers (typically the CPU core). This information, along with the geometry and power requirements of each block, was then used by the thermal module to determine the maximum temperature in the system.

The expected wire pitch, number of metal layers, power consumption, and maximum junction temperature generated by the simulator have been compared in Table I. With the exception of the Core i7 680 test case, the tool shows reasonable agreement with the published data for these processors. The 45- and 32-nm test cases have a total of nine metal layers, but in all cases, the wires on the top layer are sized very large and used for power and clock delivery, rather than signal routing. Since the interconnect estimation module is used to estimate only the size and the pitch of signal wires, the top metal layer in these designs is not counted for the purpose of signal wire pitch validation. Instead, the upper wire tier is modeled in the power delivery simulation module of the

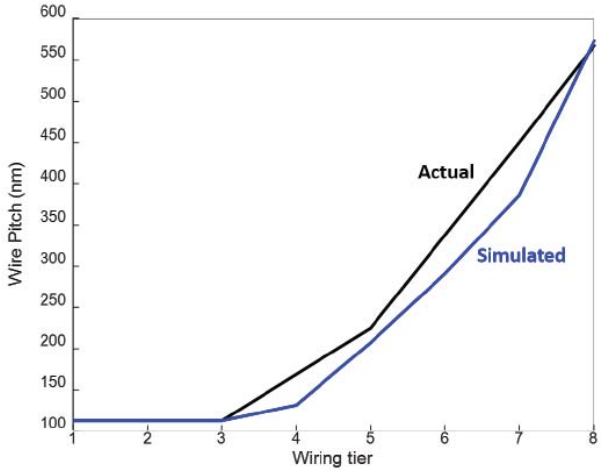


Fig. 4. Actual and expected wire pitch in a Core i7 2700k processor.

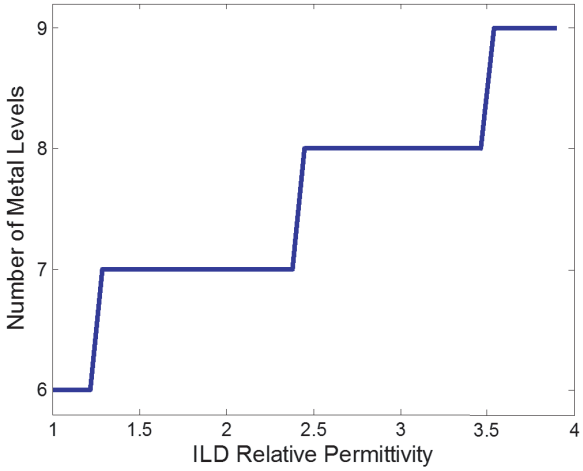


Fig. 5. Impact of ILD permittivity on the number of metal layers required to route the wires in a Sandy Bridge Core i7 2700k.

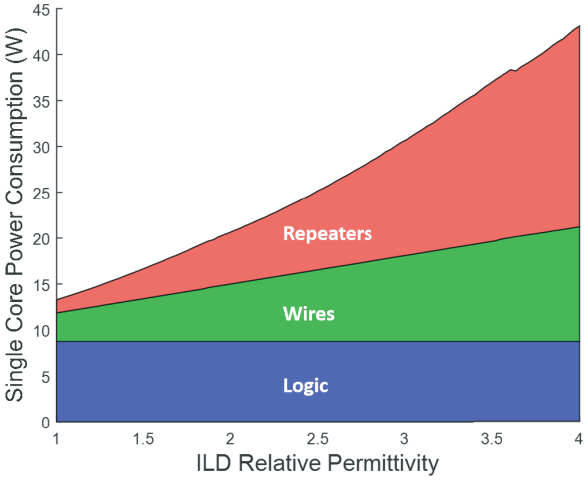


Fig. 6. Impact of ILD permittivity on the power consumed by wires and repeaters in a Sandy Bridge Core i7 2700k core.

simulator, but without data on the power noise margin and the number of power and ground pads used in these designs, the detailed validation of the power delivery module is not possible for these test cases. In addition to the number of metal levels, the simulator accurately predicts the wire pitch in each routing tier, as shown in Fig. 4.

The simulator underestimates both the number of wiring tiers and the overall power consumption for the Core i7 680 test case; this error could be due to the fact that the Core i7 680 is the only design implemented as a multichip module, with a 32-nm CPU die integrated with a 45-nm GPU die in the same package. The simulator currently does not estimate the power required for interblock communication in heterogeneous systems. While this will affect the power estimates for all test cases, the power required for communication between the two separate dice in the Core i7 680 is likely much larger than the power requirements for communication between GPUs and logic cores integrated on the same die.

The value of this simulation platform lies in its ability to consider many different effects very rapidly, enabling investigation of trends in the performance and power consumption of a reference design over a wide range of technologies and configurations. While the worst case error in these benchmarks is relatively high (27.8%), the typical error is much lower, and this level of accuracy is sufficient for the investigation of power and performance trends in 2-D and 3-DICs. In Sections IV and V the simulator is used, the simulator is used in this manner to investigate the impacts of material, technology, and packaging innovations on the power consumption and performance of 2-D and 3-DICs.

IV. 2-D: Impact of Materials Innovation

One path toward increasing system performance is to achieve improvements in the wiring materials. The permittivity of the ILD directly impacts the parasitic capacitance of the on-chip wires, which in turn impacts the wire *RC* delay and power consumption. In addition, decreasing the *RC* delay reduces the need for power-hungry repeaters.

In order to investigate the potential of ultralow-*k* (ULK) ILD materials, a 32-nm Sandy Bridge Core i7 was simulated with a range of different ILD permittivities, ranging from 3.9 (silicon dioxide), all the way down to 1 (vacuum). Fig. 5 shows that the number of metal layers required to fully route the Sandy Bridge processing core can be reduced from 8 to 6 if the relative dielectric constant of the ILD material can be brought below 1.3. Overall power consumption increases with ILD permittivity, as shown in Fig. 6, and significant power reductions are possible with ULK materials. The power reduction comes from both a reduction in wire power, as well as a reduction in the number and size of the repeaters needed to meet timing constraints.

As the critical dimensions of the smallest on-chip wires have decreased, electromigration has become a reliability concern at advanced process nodes [26], [27]. In order to address electromigration challenges at advanced process nodes, alternate materials may be required, potentially impacting signal performance, power consumption, and the number of metal layers required to fully route a design [1]. It is likely that only the lowest metal levels would use alternate materials [28].

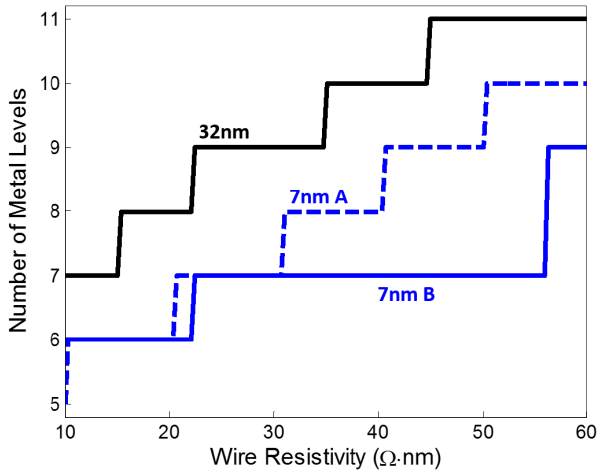


Fig. 7. Impact of wire resistivity on the number of metal layers required to route the wires in a Sandy Bridge CPU. Three cases are considered: a 32-nm Sandy Bridge core; 7 nm A, a hypothetical 7-nm Sandy Bridge core; and 7 nm B, in which only wires with width below 25 nm are modified.

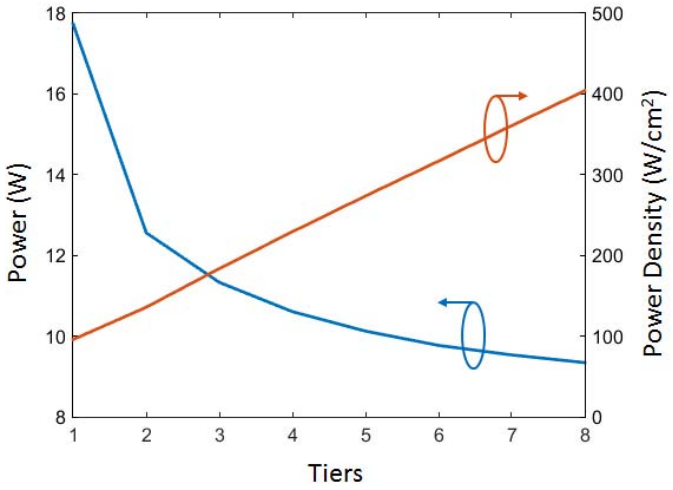


Fig. 8. Impact of block folding on power consumption and power density of a single 32-nm Sandy Bridge core.

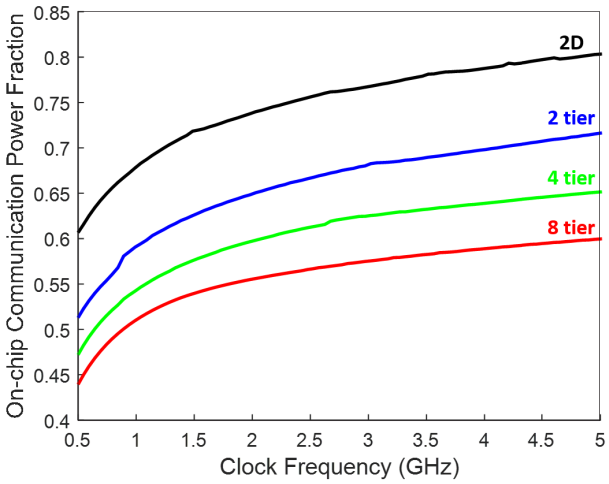


Fig. 9. Fraction of power consumed for on-chip communication as a function of 3-D configuration and operating frequency.

To investigate the impact of alternate metals on routing, a 7-nm Sandy Bridge CPU test case was constructed by scaling down the gate pitch, minimum wire pitch, transistor size, and all other lengths in the 32-nm Sandy Bridge by a factor of 4.57× (32/7). Two 7-nm test cases were considered: 7 nm A, in which all wires are composed of an alternate material, and 7 nm B, in which only wires thinner than 25 nm are replaced by the alternate material. The bulk resistivity of the alternate wiring material in both the 32- and 7-nm test cases was swept from 10 Ωnm (slightly lower than bulk Ag) to 60 Ωnm (slightly higher than bulk W). As shown in Fig. 7, the higher resistivity metals can significantly increase the number of metal levels required for signal routing, but this effect can be mitigated by restricting the use of alternate metals to the lowest wiring tiers. Since the greatest numbers of wires are routed in the lowest tiers, small changes to their dimensions can have large impacts on the wiring stack.

V. 3-D: Power Reduction Without Exotic Materials

A. Reducing Power Consumption

Implementing a design in 3-D can greatly reduce the average length of the on-chip interconnects, leading to reductions in the average delay and the power consumption of the signaling network [7]. In order to examine the impacts of 3-D integration, a single 18.5-mm^2 CPU core from a 32-nm Sandy Bridge Core i7 2700k is examined throughout this section. Unless otherwise noted, we assume a TSV aspect ratio of 20:1, and require that the TSVs use less than 10% of the total die area. Typically, the 3-DIC designs limit the TSV area to 1% or less to minimize the cannibalization of active area, but we have relaxed that limit here for illustrative purposes. In order to examine the impacts of 3-D integration, a single CPU core from a 32-nm Sandy Bridge Core i7 2700k is examined throughout this section. We consider a scenario in which logic gates and blocks can be placed on any tier, and in which TSVs are used as point-to-point interconnects. The core is assumed to be partitioned into N equal pieces, which are then stacked vertically.

Significant power savings can be obtained by moving to a 3-D design, as shown in Fig. 8, though the power reduction comes at the cost of increased areal power density, ultimately placing more stress on the heat sink. The design implications of the increased power density of 3-DICs will be discussed further in Section V-C. It is important to note that the 3-DICs reduce the on-chip communication power, fundamentally improving the energy efficiency of the system, as shown in Fig. 9.

In order to fully route a 3-DIC, space must be allocated on each tier for TSVs. Since TSVs consume space that could be used for logic, it is desirable to minimize the fraction of chip area consumed by TSVs. TSV diameter can be reduced by either increasing the TSV aspect ratio or by die thinning. Thicker logic tiers are attractive due to their higher mechanical stability, but they reduce the wire length advantages of 3-DICs. TSVs are typically limited to the diameters of 5–10 μm and aspect ratios between 5:1 and 20:1 [3], [29]. The impact of die thickness on 3-DIC power consumption is examined in Fig. 10.

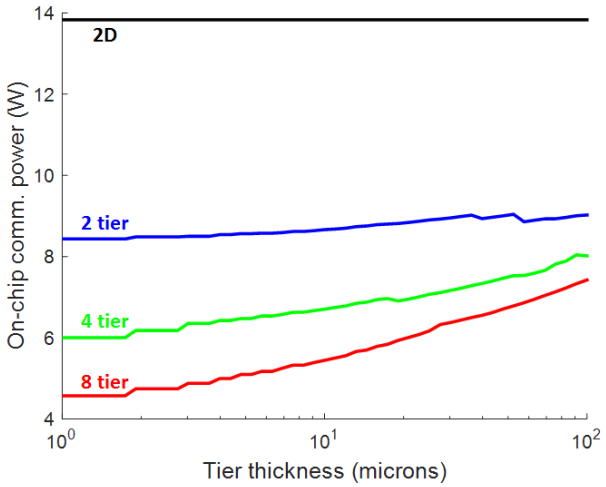


Fig. 10. Impact of die thickness on power consumed by wires and repeaters in a single 32-nm Sandy Bridge core implemented in 3-D.

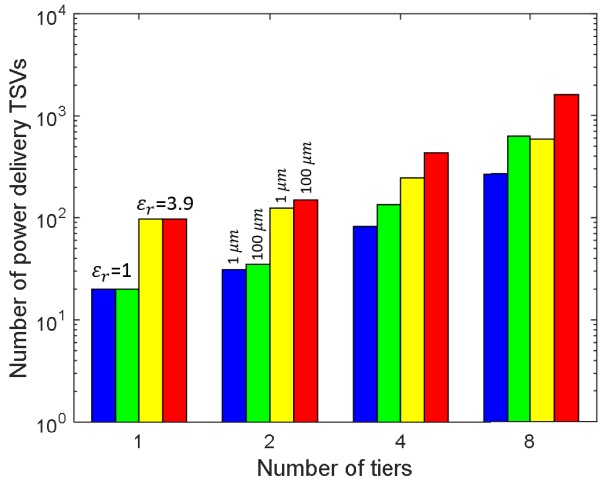


Fig. 11. Impact of ILD material, substrate thickness, and 3-D integration on power pad/TSV requirements in a 3-D Sandy Bridge CPU core.

In order to realize the greatest power reduction from 3-D integration, the active layers should be as thin as possible to maximize the number of long wires that can be shortened by block folding.

B. Power Delivery

Power delivery in 3-DICs is challenging, as a high-performance 3-DIC may have a significantly higher areal power density than an equivalent 2-D chip, while simultaneously having less space available for routing power delivery resources. In addition, power must be delivered to each tier through TSVs, increasing the parasitic resistance and inductance of the power delivery network. The number of power connections required by each tier is determined by the power draw and the power density of the system, which depends upon the dielectric properties and the substrate thickness (for 3-DICs). Both 2-D and 3-D systems benefit from the use of ULK ILDs, and the use of thin substrates in 3-D configurations can further reduce the demands on the power supply network.

In order to explore these effects, the Sandy Bridge test case was simulated with several substrate thicknesses and dielectric permittivities, in order to determine the number of power delivery pads or TSVs needed to reduce the SSN to below 15% of the nominal supply voltage. The results are presented in Fig. 11. For two-tier stacks, only a slight increase in power TSVs is observed over the 2-D case, but the eight-tier implementation requires roughly an order of magnitude more power connections than the 2-D design. The ILD permittivity has a strong impact on the power delivery requirements, as it directly affects the power consumption of the on-chip communication network. The thickness of the 3-DIC logic tiers is not a limiting factor for two-tier designs, but four- and eight-tier designs can realize nearly as much benefit from die thinning as from ULK dielectrics, due to the reduced power consumption of thin-tier 3-DICs (Fig. 10).

Another method to reduce the power supply noise is to integrate decoupling capacitors onto the die to compensate for the inductance of the power delivery network. While this practice

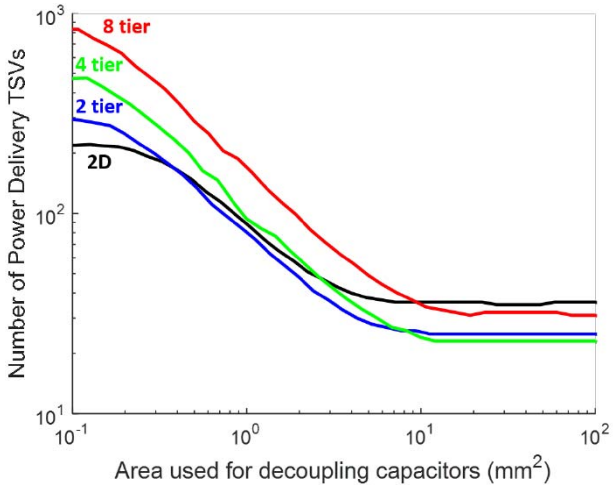


Fig. 12. Power TSV requirements for a single Sandy Bridge core as a function of 3-D configuration and area allocated for decoupling capacitors.

can improve power quality, it also sets up a tradeoff between utilizing die area for logic and power delivery. To explore this tradeoff, the 32-nm Sandy Bridge test case was simulated in the 2-D and 3-D configurations with varying amounts of silicon area allocated for decoupling capacitors. The power TSV diameter is assumed to be 10 μm and the thickness of each die in the 3-D stack is assumed to be 10 μm to investigate the potential of extreme die thinning. While handling thinned wafers can be challenging, alternate integration schemes in which wafers are bonded and subsequently thinned could enable the stacking of such thin layers without the need for modified wafer handling processes [30]. Alternately, the monolithic 3-DIC fabrication techniques could enable designs with extremely small intertier distances [31]. For simplicity, the impact of electromigration on power delivery TSVs is ignored. As shown in Fig. 12, increasing the decoupling capacitance can significantly reduce the number of power pads or TSVs, but achieving high decoupling capacitance densities could be challenging. Even with the use of decoupling capacitors, a more stringent lower bound on power TSV number may be set by the need to keep the current density carried by each TSV low enough to avoid electromigration.

TABLE II
MATERIAL PARAMETERS USED FOR THERMAL SIMULATION

Material	Thermal Cond. (W/mK)
Silicon	149
Copper	400
Underfill	0.3
Microbumps	60
Silicon Dioxide	1.38

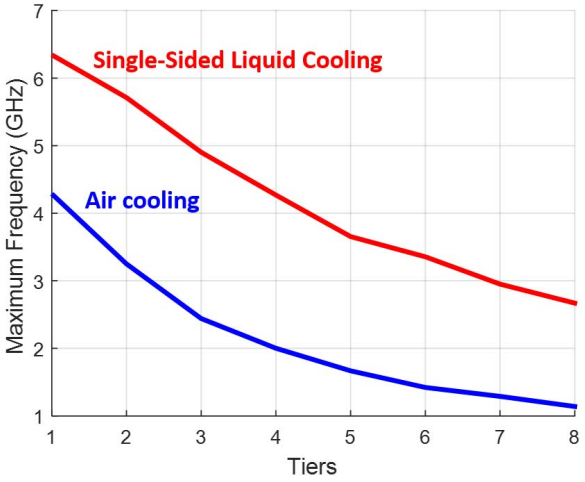


Fig. 13. Maximum clock frequency of 2-D and 3-D 32-nm Sandy Bridge CPU cores limited to 90 °C under air cooling (blue line) and liquid cooling (red line).

C. Thermal Management

Thermal management is a key challenge for 3-DICs. While the total power dissipation of an IC is expected to decrease as the system is partitioned into increasing numbers of layers (as shown in Fig. 8), the areal power density will still increase as tiers are stacked atop one another, leading to increased stress on the cooling system. In order to demonstrate the thermal capabilities of the simulation tool, and to quantify the thermal impact of 3-D stacking, the performance of a single 32-nm Sandy Bridge CPU core is examined in both 2-D and 3-D configurations, and with both air-cooled and liquid-cooled heat sinks. In all cases, the heat sinks are located on the backside of the top die. The air-cooled heat sink has a heat transfer coefficient of 1.83 W/cm^{2}K and the fluidic heat sink has a heat transfer coefficient of 4.63 W/cm^{2}K [32]. Boundaries internal to the package are assigned a heat transfer coefficient of 0.005 W/cm^{2}K. The thermal conductivities used for the materials in the stack are presented in Table II.

The CPU core was simulated in each configuration to find the maximum operating frequency which could be maintained while keeping the stack below 90 °C. In order to focus solely on the thermal aspects of 3-D stacking, we assumed that the system was thermally limited and ignored other factors, which could limit the operating frequency of the chip, such as clock distribution. The power consumed by the cooling solution is ignored in order to isolate intrinsic die-level effects from the details of the heat sink.

As shown in Fig. 13, the maximum frequency decreases steadily as the CPU is folded across more tiers, as the increased power density (Fig. 8) of the system increases the strain on the cooling system. In all cases, the microfluidically cooled cores can run faster than the air-cooled cores. With a more aggressive cooling solution, the thermally limited logic core considered here can be folded over up to four tiers while still achieving performance parity with an air-cooled 2-D implementation. It is important to note that the configuration considered here represents the most thermally challenging 3-D integration scenario: the stacking of high-performance logic. Systems that are not thermally limited will be able to take greater advantage of 3-D integration to increase performance and decrease power consumption. In addition, the performance of thermally limited 3-D stacks could be further improved by integrating microfluidic coolers into each die in the stack to thermally decouple each tier, allowing each cooler to efficiently extract heat dissipated in adjacent tiers.

VI. CONCLUSION

A compact simulation tool for 2-D and 3-D IC path finding was developed and used to examine the impacts of advanced technologies on system performance. The tool incorporates models for signal delivery, power supply noise, and thermal performance in 2-D and 3-D ICs, and was validated against wire pitch and power consumption data for recent commercial microprocessors. The simulation tool is available upon request. The simulation results suggest that the high-performance 3-DICs may require large numbers of power delivery TSVs, due to the TSV parasitics introduced into the power delivery network, as well as the increased power density of the 3-D cores. Die thinning and low-k dielectrics are expected to be effective tools for reducing the power consumption and power TSV requirements of high-performance 3-DICs. The results suggest that the 3-DICs should exhibit greater energy efficiency than their 2-D counterparts, though thermally limited 3-D designs may require more aggressive cooling solutions.

REFERENCES

[1] C. Adelmann *et al.*, "Alternative metals for advanced interconnects," in *Proc. IEEE Int. Interconnect Technol. Conf./Adv. Metallization Conf. (IITC/AMC)*, May 2014, pp. 173–176.
[2] D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," *IEEE Electron Device Lett.*, vol. 2, no. 5, pp. 126–129, May 1981.
[3] Y. Zhang, H. Oh, and M. S. Bakir, "Within-tier cooling and thermal isolation technologies for heterogeneous 3D ICs," in *Proc. IEEE Int. Conf. 3D Syst. Integr. (3DIC)*, Oct. 2013, pp. 1–6.
[4] R. Venkatesan, J. A. Davis, K. A. Bowman, and J. D. Meindl, "Optimal *n-tier* multilevel interconnect architectures for gigascale integration (GSI)," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, no. 6, pp. 899–912, Dec. 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 *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2007, pp. 560–567.
[6] J. A. Davis, V. K. De, and J. D. Meindl, "A stochastic wire-length distribution for gigascale integration (GSI). I. Derivation and validation," *IEEE Trans. Electron Devices*, vol. 45, no. 3, pp. 580–589, Mar. 1998.
[7] J. W. Joyner, R. Venkatesan, P. Zarkesh-Ha, J. A. Davis, and J. D. Meindl, "Impact of three-dimensional architectures on interconnects in gigascale integration," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, no. 6, pp. 922–928, Dec. 2001.
[8] P. Zarkesh-Ha, J. A. Davis, W. Loh, and J. D. Meindl, "On a pin versus gate relationship for heterogeneous systems: Heterogeneous Rent's rule," in *Proc. IEEE Custom Integr. Circuits Conf.*, May 1998, pp. 93–96.

[9] D. H. Kim, S. Mukhopadhyay, and S. K. Lim, "Through-silicon-via aware interconnect prediction and optimization for 3D stacked ICs," in *Proc. 11th Int. Workshop Syst. Level Interconnect Predict. (SLIP)*, 2009, pp. 85–92.

[10] W. Wahby, A. Dembla, and M. Bakir, "Evaluation of 3DICs and fabrication of monolithic interlayer vias," in *Proc. IEEE Int. 3D Syst. Integr. Conf. (3DIC)*, Oct. 2013, pp. 1–6.

[11] H. B. Bakoglu and J. D. Meindl, "Optimal interconnection circuits for VLSI," *IEEE Trans. Electron Devices*, vol. 32, no. 5, pp. 903–909, May 1985.

[12] T. Sun *et al.*, "Surface and grain-boundary scattering in nanometric Cu films," *Phys. Rev. B*, vol. 81, p. 155454, Apr. 2010.

[13] L. Zheng, Y. Zhang, G. Huang, and M. S. Bakir, "Novel electrical and fluidic microbumps for silicon interposer and 3-D ICs," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 4, no. 5, pp. 777–785, May 2014.

[14] G. Huang, M. S. Bakir, A. Naeemi, and J. D. Meindl, "Power delivery for 3-D chip stacks: Physical modeling and design implication," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 2, no. 5, pp. 852–859, May 2012.

[15] J. Xie and M. Swaminathan, "Electrical-thermal co-simulation of 3D integrated systems with micro-fluidic cooling and Joule heating effects," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 1, no. 2, pp. 234–246, Feb. 2011.

[16] Y. Zhang, Y. Zhang, and M. S. Bakir, "Thermal design and constraints for heterogeneous integrated chip stacks and isolation technology using air gap and thermal bridge," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 4, no. 12, pp. 1914–1924, Dec. 2014.

[17] P. Bai *et al.*, "A 65 nm logic technology featuring 35 nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 μm^2 SRAM cell," in *IEDM Tech. Dig.*, Dec. 2004, pp. 657–660.

[18] N. Sakran, M. Yuffe, M. Mehalel, J. Doweck, E. Knoll, and A. Kovacs, "The implementation of the 65 nm dual-core 64 b Merom processor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC), Dig. Tech. Papers*, Feb. 2007, pp. 106–107 and 590.

[19] K. Mistry *et al.*, "A 45 nm logic technology with high-k + metal gate transistors, strained silicon, 9 Cu interconnect layers, 193 nm dry patterning, and 100% Pb-free packaging," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2007, pp. 247–250.

[20] V. George *et al.*, "Penryn: 45-nm next generation Intel Core 2 processor," in *Proc. IEEE Asian Solid-State Circuits Conf. (ASSCC)*, Nov. 2007, pp. 14–17.

[21] R. Kumar and G. Hinton, "A family of 45 nm IA processors," in *IEEE Int. Solid-State Circuits Conf. (ISSCC)-Dig. Tech. Papers*, Feb. 2009, pp. 58–59.

[22] P. Packan *et al.*, "High performance 32 nm logic technology featuring 2$^{\text{nd}}$ generation high-k + metal gate transistors," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2009, pp. 1–4.

[23] N. A. Kurd *et al.*, "Westmere: A family of 32 nm IA processors," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2010, pp. 96–97.

[24] M. Yuffe, E. Knoll, M. Mehalel, J. Shor, and T. Kurts, "A fully integrated multi-CPU, GPU and memory controller 32 nm processor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2011, pp. 264–266.

[25] P. Christie and D. Stroobandt, "The interpretation and application of Rent's rule," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 6, pp. 639–648, Dec. 2000.

[26] N. L. Michael, C.-U. Kim, P. Gillespie, and R. Augur, "Electromigration failure in ultra-fine copper interconnects," *J. Electron. Mater.*, vol. 32, no. 10, pp. 988–993, Oct. 2003.

[27] Z. Tőkei, K. Croes, and G. P. Beyer, "Reliability of copper low-k interconnects," *Microelectron. Eng.*, vol. 87, no. 3, pp. 348–354, Mar. 2010.

[28] A. S. Oates, "Strategies to ensure electromigration reliability of Cu/low-k interconnects at 10 nm," *ECS J. Solid State Sci. Technol.*, vol. 4, no. 1, pp. N3168–N3176, 2015.

[29] J. H. Lau, "Evolution, challenge, and outlook of TSV, 3D IC integration and 3D silicon integration," in *Proc. Int. Symp. Adv. Packag. Mater. (APM)*, Oct. 2011, pp. 462–488.

[30] R. S. Patti, "Three-dimensional integrated circuits and the future of system-on-chip designs," *Proc. IEEE*, vol. 94, no. 6, pp. 1214–1224, Jun. 2006.

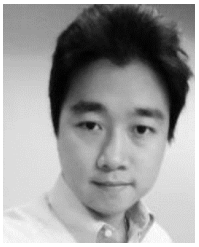
[31] M. Vinet *et al.*, "Monolithic 3D integration: A powerful alternative to classical 2D scaling," in *Proc. IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf. (S3S)*, Oct. 2014, pp. 1–3.

[32] Y. Zhang, A. Dembla, and M. S. Bakir, "Silicon micropin-fin heat sink with integrated TSVs for 3-D ICs: Tradeoff analysis and experimental testing," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 3, no. 11, pp. 1842–1850, Nov. 2013.

William Wahby (S'13) received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana–Champaign, computer engineering with the Georgia Institute of Technology, Atlanta, GA, USA.

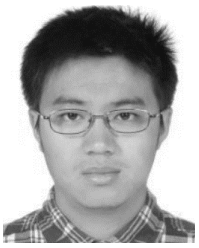
His current research interests include monolithic 3-D integration, chip-scale optical interconnects, and novel memory devices.

Mr. Wahby received the SRC Education Alliance Intel Foundation Fellowship in 2014.



Li Zheng received the B.S. degree from Zhejiang University, Hangzhou, China, in 2006, and the M.S. degrees in electrical and computer engineering from Shanghai Jiao Tong University, Shanghai, China, and the Georgia Institute of Technology, Atlanta, GA, USA, in 2009, where he is currently pursuing the Ph.D. degree in electrical and computer engineering.

His current research interests include embedded microfluidic cooling, power delivery modeling, and chip-to-chip signaling modeling for high performance 2.5-D and 3-D systems.



Yang Zhang (S'13) received the B.S. degree in microelectronics and mathematics from Peking University, Beijing, China, in 2012. He is currently pursuing the Ph.D. degree in electrical engineering with the Georgia Institute of Technology, Atlanta, GA, USA.

Muhannad S. Bakir (SM'12) received the B.E.E. degree from Auburn University, Auburn, AL, USA, in 1999, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2000 and 2003, respectively.

He is currently an Associate Professor with the School of Electrical and Computer Engineering, Georgia Tech. His current research interests include 3-D electronic system integration, advanced cooling and power delivery for 3-D systems, biosensors and their integration with CMOS circuitry, and nanofabrication technology.

Dr. Bakir was a recipient of the 2013 Intel Early Career Faculty Honor Award, the 2012 DARPA Young Faculty Award, and the 2011 IEEE Components, Packaging, and Manufacturing Technology (CPMT) Society Outstanding Young Engineer Award. He and his research group have received more than 15 conference and student paper awards, including five from the IEEE Electronic Components and Technology Conference, four from the IEEE International Interconnect Technology Conference, and one from the IEEE Custom Integrated Circuits Conference. His group received the 2014 Best Paper of the IEEE TRANSACTIONS ON COMPONENTS, PACKAGING, AND MANUFACTURING TECHNOLOGY in the area of advanced packaging. He was an Invited Participant in the 2012 National Academy of Engineering Frontiers of Engineering Symposium. In 2015, he was elected by the IEEE CPMT Society to serve as a Distinguished Lecturer, and was an Invited Speaker at the U.S. National Academies Frontiers of Sensor Science Symposium. He is an Editor of the IEEE TRANSACTIONS ON ELECTRON DEVICES and an Associate Editor of the IEEE TRANSACTIONS ON COMPONENTS, PACKAGING, AND MANUFACTURING TECHNOLOGY.