Dynamic Instruction Schedulers in a 3-Dimensional Integration Technology

Kiran Puttaswamy
Georgia Institute of Technology
School of Electrical and Computer Engineering
Atlanta, GA
kiranp@ece.gatech.edu

Gabriel H. Loh
Georgia Institute of Technology
College of Computing
Atlanta, GA
loh@cc.gatech.edu

ABSTRACT

We present the design of high-performance and energy-efficient dynamic instruction schedulers in a 3-Dimensional integration technology. Based on a previous observation that the critical path latency of a conventional dynamic scheduler is greatly affected by wire delay, we propose 3D-integrated scheduler designs by partitioning a conventional scheduler across multiple vertically-stacked die. The die-stacked organization reduces the lengths of critical wires thus reducing both latency and energy. Our simulation results show that a 20-entry (60-entry) instruction scheduler implemented in a 2-die stack achieves a 9% (15%) reduction in latency with simultaneous energy reduction as compared to a conventional planar design. The benefits are even larger when the instruction scheduler is implemented on a 4-die stack, with the corresponding latency reductions being 12% (22%).

Categories and Subject Descriptors

B.7 [**Integrated Circuits**]: Advanced technologies

General Terms

Design, Performance

Keywords

3D Technology, Instruction scheduler

1. INTRODUCTION

In modern out-of-order superscalar processors, the dynamic instruction scheduler is responsible for exposing instruction level parallelism by identifying instructions that can be executed in parallel. The capacity of the scheduler to identify opportunities for instruction level parallelism increases with the number of entries in the scheduler. Unfortunately, the latency and energy characteristics of the scheduler do not scale well with increasing number of entries [1]. Given that technology scaling has created an ever-widening gap between the relative delay of logic and wires [2, 3], increasing the number of entries in the dynamic scheduler worsens the wire delay, thus reducing the overall performance of the processor. 3-Dimensional integrated circuit (3D IC) technology can reduce the wire delay by vertically stacking multiple die and connecting the stacked die with a high-density, low latency die-to-die (D2D) interconnect interface [4]. In this paper, we discuss the implementation of dynamic instruction schedulers in a 3D technology.

The rest of the paper is organized as follows. Section 2 provides a short background on 3D technology. Section 3 describes a conventional planar dynamic scheduler. Section 4 explains our designs for dynamic schedulers implemented in 3D. Section 5 details our experimental framework. Section 6 presents the results and analysis of our 3D schedulers. Section 8 summarizes our contribution.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
GLSVLSI'06, April 30–May 2, 2006, Philadelphia, PA, USA.
Copyright 2006 ACM 1-59593-347-6/06/0004 ...$5.00.

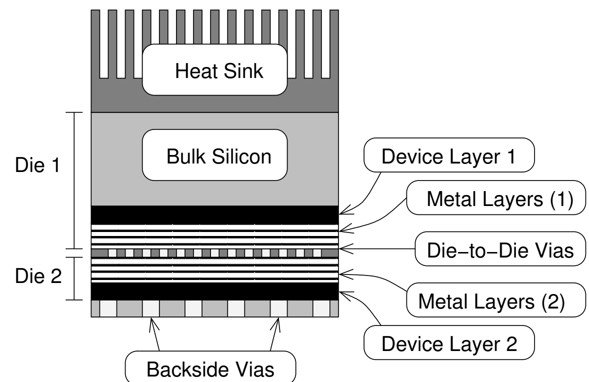

Figure 1: 2-die Stacked 3D Integrated Circuit. Dimensions not drawn to scale.

2. 3D INTEGRATION TECHNOLOGY

The semiconductor industry faces an increasing number of challenges that must be overcome to keep pace with Moore's Law [5] and industry projections [6]. Some of these problems include poor scaling of wire RC delays [7, 3] and increasing power consumption [8, 9]. 3-Dimensional stacked-die integration [10] has the potential to address many of these problems for future high-performance microprocessors. Processor companies are actively researching the technology [11,

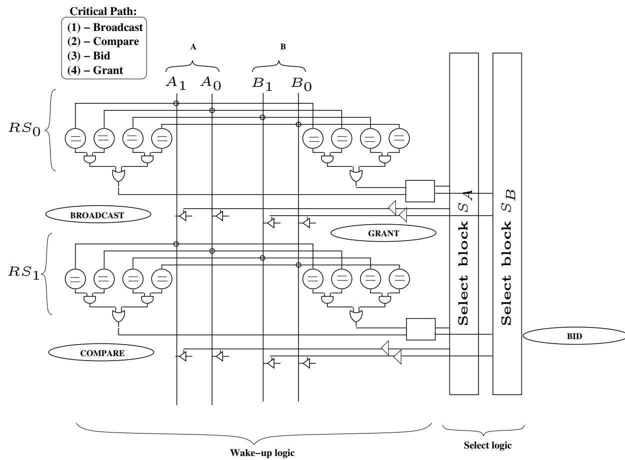

Figure 2: Conventional Dynamic Scheduler

12, 13, 14]. Academic research effort has focused on the circuit implementation, process technology, physical design and automated design tools [15, 16, 17, 18, 19, 20, 21, 22].

In this study, we consider a Cu-Cu bonding process for implementing the 3D circuits. Figure 1 shows the overall structure of a 2-die 3D stack. The ability to route signals in the vertical dimension enables long wires to be replaced by short die-to-die vias with significantly less capacitance and resistance. Two functional blocks connected by a long global route in a planar implementation can instead be vertically stacked to drastically reduce the communication distance by routing in the third dimension. Wire-dominated functional unit blocks can be folded on top of themselves to reduce the effects of intra-block wiring [19, 23, 24]. Reducing the amount of wire can also have a significant impact on power consumption [25].

3. PLANAR INSTRUCTION SCHEDULER

In this section, we describe the implementation of a dynamic instruction scheduler in a conventional planar technology. The dynamic scheduler consists of two components, namely the wakeup logic and the select logic.

The wakeup logic consists of a broadcast bus, a set of comparator circuits and buffers called reservation stations. The reservation station (RS) entries contain source operand addresses (called tags) of the instructions that have been dispatched from the in-order front-end of the processor. The instruction in each valid RS entry is ready for execution as soon as all its source operands become available. The wakeup logic requests functional units to execute the ready instructions. The select logic matches the requests from the wakeup logic to the available functional units and grants the functional units to the selected incoming requests. The selected instructions issue to the functional units and broadcast their destination tags on the broadcast buses. The wakeup logic simultaneously compares the source operand tags in the RS entries to the broadcast tags and generates a new set of ready instructions based on the tag comparison matches. The new set of ready instructions along with all unsuccessful (unselected) instructions from the previous cycle repeat their requests during the current cycle. Every cycle, all ready instructions participate in a bidding process by placing requests for functional units to the select logic and the select logic grants the available functional units to the selected requesting instructions. This chain of events that repeats every cycle is called the wakeup-select loop.

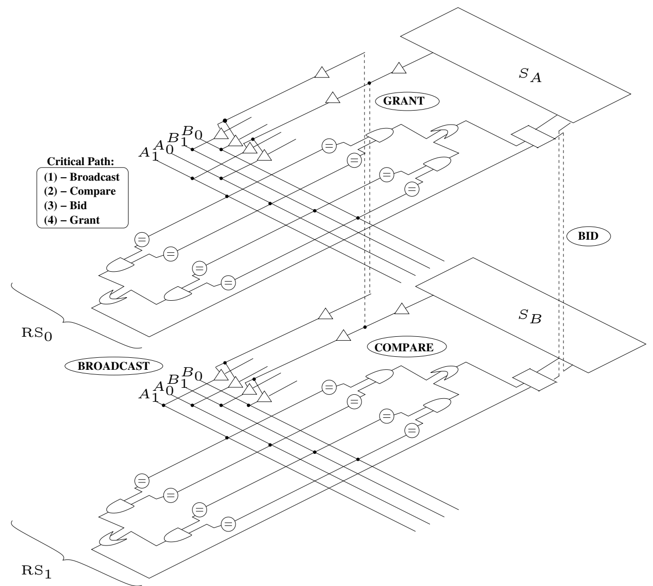

Figure 3: Entry Partitioned 3D Scheduler

Figure 2 shows a small, 2-issue dynamic scheduler with 2-bit tags and two RS entries. A and B denote the tag broadcasts for the two issues, with A_1A_0 and B_1B_0 being the 2-bit tags. The select logic consists of two blocks, S_A and S_B with each select logic block in charge of one issue (and one resultant tag broadcast). Figure 2 also shows the critical path through the wakeup-select loop starting at tag broadcast (1), compare logic (2), bid logic (3) and grant logic (4). For every additional RS entry, the length of the longest tag broadcast wire in the wakeup logic increases. Since the select logic height is pitch-matched to the wakeup logic height, the select logic suffers increased wire delays in addition to the increased logic circuitry to accommodate the requests originating from the additional entries. Thus, for increasing number of RS entries, the overall latency of the scheduler also increases.

The latency of the wakeup logic is influenced by two sources. One, the comparators connected to the tag broadcast bus increase the load capacitance, thus increasing the delay. Two, the critical path wire-length of the broadcast bus increases thus increasing both resistance and capacitance of the wire, thus contributing to the total delay. These two factors cause the wakeup-select loop to experience a rapidly diminishing frequency of operation as the scheduler size increases, thus negating the gains of a larger scheduler.

4. 3D SCHEDULERS

We propose designs of 3D-integrated schedulers to reduce the tag broadcast wire-lengths and comparator loads of the wakeup-select loop.

4.1 Entry Partitioned (EP) 3D Scheduler

We propose our first 3D design to reduce both the com-

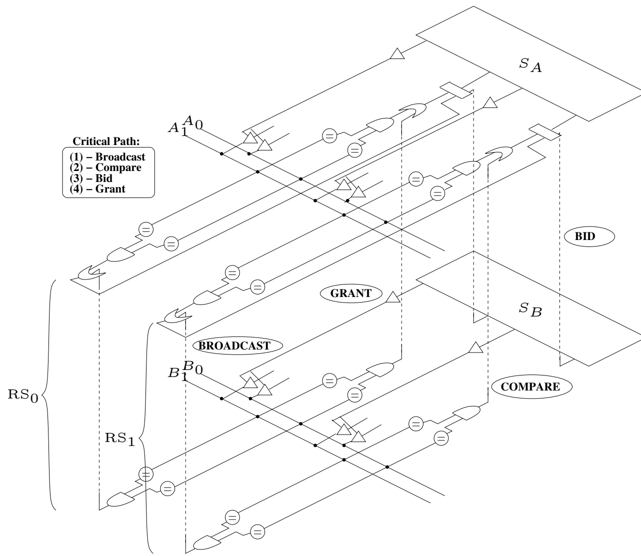

Figure 4: Tag Partitioned 3D Scheduler

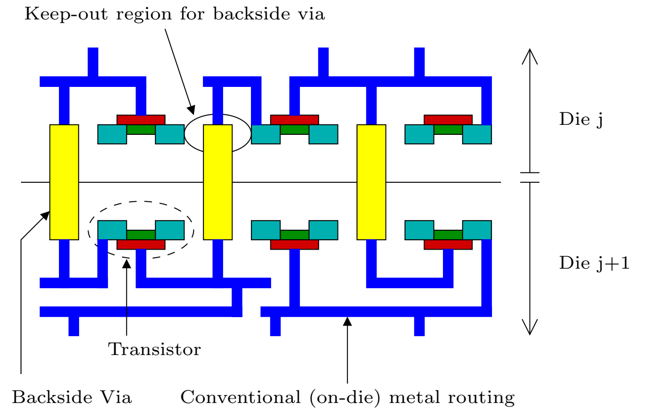

Figure 5: Backside via interface between two die. Dimensions not drawn to scale.

parator loading and the wire-length of the longest tag broadcast wire as compared to the planar scheduler. We place half the scheduler entries on the top die and the other half on the bottom die. In Figure 3, entries RS_0 and RS_1 are partitioned across the two die such that RS_0 is on the top die and RS_1 is on the bottom die. This design halves the wire-length and comparator loading of the individual tag broadcast buses by halving number of RS entries per die. The reduction in the comparator loading and the tag bus wire-length speeds up the tag broadcast, thus providing latency and power benefits. Note that the select logic is partitioned between the two vertically stacked die such that S_A is on the top die and S_B is on the bottom die. Since the select logic height is pitch-matched to the wakeup logic height, the select logic height also reduces, providing additional latency/energy savings.

There is a logic overhead of one extra driver per die. In addition to the additional driver, the 3D designs incur a minor overhead of two die-to-die (D2D) interconnects on the critical path, one each for the bid and the grant signals as shown in Figure 3. However, note that the additional latency and power consumption due to the additional driver and the D2D interconnects are more than compensated by the large reduction in wire-length and comparator loading on the critical path.

4.2 Tag Partitioned (TP) 3D Scheduler

With a tag-partitioned (TP) design, we partition half of the broadcast tags on the top die (A_1A_0 in Figure 4) and the other half on the bottom die (B_1B_0). This halves both the vertical wire-length and the horizontal wire-length of the tag broadcast bus. Since both the height and width reduce by half, the wakeup logic area gets reduced to a quarter of the planar wakeup logic. The comparator loading on the tag broadcast bus remains the same as in the planar implementation. The select logic blocks are also vertically stacked, further reducing the overall area footprint and providing additional savings in latency and energy.

The TP design incurs a minor latency overhead of two D2D vias on the critical path. However, the latency overhead of two D2D vias is insignificant when compared to the latency saving obtained from reducing the critical path wire-lengths. The comparator match signals are routed through D2D vias to each of the die. Bid signals are routed through D2D vias to enable the wakeup logic to bid on either of the select blocks S_A and S_B. However, the grant signals do not need to be vertically routed since the select logic block that enables the tag broadcasts on a particular die is co-located on the same die.

4.3 Extending to a 4-Die 3D Stack

So far, we have focused on the 2-die implementation of the scheduler. After bonding two die in a face-to-face topology as illustrated in Figure 1, the only remaining surfaces are the backsides of the silicon die thus necessitating backside vias. The backside vias are more challenging for the design of 3D ICs for two reasons. First, etching through the backside of the silicon will cause the pitch and length of the via to increase relative to the face-to-face D2D via. In practice, these backside vias are not much larger than the face-to-face D2D vias, with $<4\mu$m backside via pitches planned for production [26]. Second, the backside via must pass through the active region of the silicon die, which can disrupt the layout of transistors. The backside via along with some surrounding margin area is defined as a "keep-out" region (See Figure 5) where the circuit designer cannot place any active devices.

The 4-die entry-partitioned (EP) scheduler places one quarter of the RS entries on each die. Additional backside vias are required for routing the bid/grant signals to/from the select logic. Due to the keep-out regions of the backside vias, a naive routing can lead to substantial area overhead within the RS entries. However, Figure 6 shows a technique to insert some space between the RS entries and the select logic block to avoid disrupting the devices in either the RS entries or the select logic. This increases the wire-length slightly in the horizontal direction, however this is more than offset by the wire length reduction of the long tag broadcast buses.

The 4-die tag-partitioned (TP) scheduler is more challenging than the EP configuration since the inter-die via locations are spread over a large region within the wakeup logic. To avoid disruptions to the RS layout, one could place all

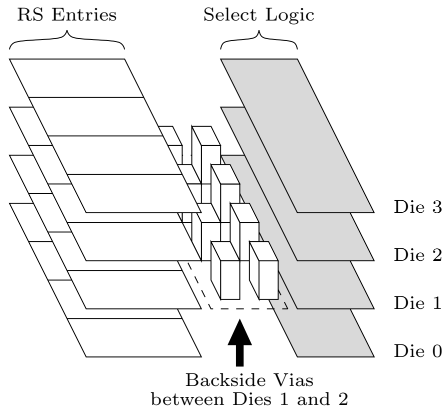

Figure 6: 4-die EP scheduler with extra space allocated for backside vias. Dimensions not drawn to scale.

Table 1: Scheduler latencies for planar, entry partitioned (EP) and tag partitioned (TP) 2-die 3D organizations.

RS entries	Latency (ps)			Saving	
	Planar	EP(2-die)	TP(2-die)	% EP	% TP
20	577	526	571	8.9	1.1
30	623	562	611	9.7	1.8
40	665	582	645	12.5	3.0
50	703	602	681	14.4	3.0
60	739	627	714	15.1	3.4

of the backside vias between the RS entries and the select logic in a fashion similar to the EP scheduler. This incurs some additional wire delay to route from the RS entry to the backside vias and then back to the RS entry logic.

5. CIRCUIT LATENCY AND ENERGY ESTIMATION

We use HSPICE simulations of custom circuit designs to evaluate the critical path latency and overall energy consumption of the dynamic instruction scheduler for planar and 3D implementations. For our HSPICE simulations, we use BSIM transistor models [27] for a 70nm technology. We derive the wire parameters by extrapolating to 70nm from a TSMC 180nm technology. While our absolute results are dependent on the technology parameters, the general trends and benefits of 3D circuits are likely to hold for a wide range of process and 3D parameters.

We use a distributed RC-ladder model for all wires in the circuits. The distance between the top metal layers on the two die is very small, and the pitch of the D2D vias are of the same order as the top level metal [28]. Therefore, we model the D2D interconnect as a 10μm (20μm) length of top-level metal for the face-to-face (backside) vias. This is a conservative assumption as previous work has indicated that the entire height of a thinned die can be less than 10μm [29]. To optimize our scheduler designs, we sweep through a range of transistor sizes and use the transistor sizes that minimize the overall scheduler delay.

6. RESULTS

The 3D schedulers provide simultaneous benefits in both latency and energy due to reduced wire RCs. We present the results for 2-die and 4-die 3D implementations in this section. For all simulations, we assume a scheduler that can broadcast up to four 7-bit tags and issue up to four instructions, per cycle. Such a configuration models the scheduler of a 4-wide superscalar out-of-order processor with 128 registers. We present our results for schedulers ranging from 20-entries to 60-entries to chart the trends for both the current generation and the future aggressive high-performance processors.

6.1 2-Die 3D Scheduler Results

Table 1 shows the latency results of our spice simulations for the planar and the 2-die 3D configurations for scheduler entries increasing from 20 to 120 entries. From Table 1, the 3D implementations have smaller overall latencies than the planar schedulers with identical number of RS entries. For large instruction schedulers, the overall latency is dominated by the tag broadcast component. Consequently, 3D technology provides greater relative benefit for larger sized schedulers. The latency savings of the EP topology is consistently greater than the latency savings of the TP topology. This is due to the fact that the EP has the benefit of both reduced wire-length and reduced comparator loading while the TP is benefited only by reduced wire-lengths. Although the TP provides drastic reductions in the wire-lengths (the TP footprint being a quarter of the original planar footprint), it does not reduce the comparator loading per broadcast wire. The results listed in Table 1 show that the scheduler latency can be reduced by 9-15% depending on the number of entries in the scheduler. If the instruction scheduling logic is the limiting factor of overall processor frequency [1], then this latency improvement can be directly translated into performance by increasing the 3D-processor clock speed. Alternatively, the 3D-processor frequency can be fixed to be the same as the planar processor frequency, while increasing the number of entries in the scheduler to expose more instruction level parallelism. In most situations, a scheduler with approximately twice as many entries can be implemented with the same latency as the original planar circuit. This is largely due to the fact that the tag broadcast latency scales quadratically with increasing wire length.

We measured the energy consumption of the scheduler using activity profiles collected from SPEC2000, MediaBench, and other benchmark applications. By combining the scheduler activity over the period of program execution and the circuit latency and power data obtained by HSPICE circuit simulations, we calculate the energy consumed by the scheduler for each program. For a 20-entry scheduler, we found that the 3D-EP configuration reduces the total scheduler energy consumption by 18% over all the benchmarks. For larger configurations such as the 60-entry case, the energy reduction can be as much as 45%. 3D designs reduce the switching energy due to the reduced resistance and capacitances of the wires.

6.2 4-Die 3D Scheduler Results

In this section, we present the results for the instruction scheduler implemented in a 4-die stack. We only consider the entry-partitioned (EP) organization because of the difficulties in laying out the tag-partitioned (TP) scheduler in

Table 2: Scheduler latencies for planar and entry-partitioned (EP) 4-die 3D organizations.

RS entries	Latency (ps) Planar (1-die)	EP (4-die)	Saving % EP
20	577	506	12.3
30	623	521	16.3
40	665	525	21.1
50	703	548	22.0

the presence of backside vias. Furthermore, the greater latency and energy benefits of the EP scheduler make it the clear winner between the two alternative 3D scheduler organizations. Table 2 shows the latencies of 4-die schedulers over a range of sizes. The planar latencies are repeated here for reference. Overall, the latencies are further improved as compared to the 2-die cases, ranging from 12-22%. The energy benefits for using a 4-die stack are similar, with a 20-entry scheduler providing up to 25% energy reduction and a 60-entry scheduler providing up to 67% reduction.

7. THERMAL DISCUSSION

3D die-stacking can potentially reduce the thermal reliability of the circuits. By stacking circuits, the worst-case power density may increase substantially. This in turn can lead to thermal hotspots [30]. A thermal analysis of a 3D scheduler is impractical since the worst-case temperature depends not only on the 3D layout of the scheduler, but also on the nearby components and their power-generation rates.

A naive analysis would lead one to believe that a 2-die (4-die) stack might double (quadruple) the power density. However, such an analysis assumes that the total power consumption of the 3D implementation is identical to the planar implementation. We have already seen that the total energy consumption of the scheduler reduces drastically when reimplemented in 3D. While the power density still increases, the total power of the 3D schedulers decreases. The total power has been shown to have a greater impact on the temperature than the power density [31]. In some situations, latency may be traded for further power reduction. For example, 3D implementation may provide enough latency slack to allow power-hungry dynamic logic circuits to be reimplemented with lower-power static logic circuits. At a microarchitectural level, the RS-entry allocation logic could steer instructions toward RS entries that are on the die adjacent to the heat sink and only use the other layers as necessary. In addition to these circuit and microarchitectural techniques, new cooling technologies may be needed to address the increased thermal effects of 3D structures [32].

8. SUMMARY

3D technology can provide significant benefits in terms of both performance and power consumption. We demonstrated how 3D technology can be applied to the design of a very important processor component to reduce the lengths of critical wire components. The reduction of wire length leads to substantial improvement in both latency and energy characteristics. In addition to the instruction scheduler, many other components in modern processors are dominated by wire delay and could benefit from a 3D implementation. Examples include on-chip caches, register files, the bypass network, and the branch misprediction notification datapath. Much research remains to understand how to best adapt these other blocks to exploit the strengths (density, vertical routing) of 3D technology. Apart from adapting conventional microarchitectures to a 3D technology, designing a new processor from the ground-up to exploit 3D may provide even greater performance and power benefits.

Acknowledgements

Funding and equipment for this project have been provided by Intel Corporation and a grant from the Microelectronics Advanced Research Corporation (MARCO).

9. REFERENCES

[1] S. Palacharla, "Complexity-Effective Superscalar Processors," Ph.D. dissertation, University of Wisconsin, 1998.

[2] V. Agarwal, M. S. Hrishikesh, S. W. Keckler, and D. Burger, "Clock Rate Versus IPC: The End of the Road for Conventional Microarchitectures," in *Proceedings of the 27th International Symposium on Computer Architecture*, Vancouver, Canada, June 2000, pp. 248–259.

[3] R. Ronen, A. Mendelson, K. Lai, S.-L. Lu, F. Pollack, and J. P. Shen, "Coming Challenges in Microarchitecture and Architecture," *Proceedings of the IEEE*, vol. 89, no. 3, pp. 325–340, March 2001.

[4] P. Morrow, M. J. Kobrinsky, S. Ramanathan, C.-M. Park, M. Harmes, V. Ramachandrarao, H. mog Park, G. Kloster, S. List, and S. Kim, "Wafer-Level 3D Interconnects Via Cu Bonding," in *Proceedings of the 21st Advanced Metallization Conference*, San Diego, CA, USA, October 2004.

[5] G. E. Moore, "Cramming More Components Onto Integrated Circuits," *Electronics*, April 1965.

[6] Semiconductor Industry Association, "The National Technology Roadmap for Semiconductors," 1999.

[7] S. Borkar, "Design Challenges of Technology Scaling," *IEEE Micro Magazine*, vol. 19, no. 4, pp. 23–29, July 1999.

[8] D. Brooks, P. W. Cook, P. Bose, S. E. Schuster, H. Jacobson, P. N. Kudva, A. Buyuktosunoglu, J.-D. Wellman, V. Zyuban, and M. Gupta, "Power-Aware Microarchitecture: Design and Modeling Challenges for Next-Generation Microprocessors," *IEEE Micro Magazine*, vol. 20, no. 6, pp. 26–44, November 2000.

[9] V. Srinivasan, D. Brooks, M. Gschwind, P. Bose, V. Zyuban, P. N. Strenski, and P. G. Emma, "Optimizing Pipelines for Power and Performance," in *Proceedings of the 35th International Symposium on Microarchitecture*, Istanbul, Turkey, November 2002, pp. 333–344.

[10] 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 Transactions on VLSI*, vol. 9, no. 6, pp. 922–928, Dec 2001.

[11] P. Reed, G. Yeung, and B. Black, "Design Aspects of a Microprocessor Data Cache using 3D Die Interconnect Technology," in *Proceedings of the International Conference on Integrated Circuit Design and Technology*, Austin, TX, USA, May 2005, pp. 15–18.

[12] B. Black, D. Nelson, C. Webb, and N. Samra, "3D Processing Technology and its Impact on IA32 Microprocessors," in *Proceedings of the 22nd International Conference on Computer Design*, San Jose, CA, USA, October 2004, pp. 316–318.

[13] D. Nelson, C. Webb, D. McCauley, K. Raol, J. R. II, J. DeVale, and B. Black, "A 3D Interconnect Methodology Applied to iA32-class Architectures for Performance Improvements through RC Mitigation," in *Proceedings of the 21st International VLSI Multilevel Interconnection Conference*, Waikoloa Beach, HI, USA, September 2004.

[14] K. W. Guarini, A. W. Topol, M. Ieong, R. Yu, L. Shi, M. R. Newport, D. J. Frank, D. V. Singh, G. M. Cohen, S. V. Nitta, D. C. Boyd, P. A. O'Neil, S. L. Tempest, H. B. Pogge, S. Purushothaman, and W. E. Haensch, "Electrical Integrity of State-of-the-Art 0.13μm SOI CMOS Devices and Circuits Transferred for Three-Dimensional (3D) Integrated Circuit (IC) Fabrication," in *Proceedings of the International Electron Devices Meeting*, December 2002, pp. 943–945.

[15] L. Cheng, L. Deng, and M. Wong, "Floorplan Design for 3-D VLSI Design," in *Proceedings of the Asia South Pacific Design Automation Conference*, Shanghai, China, January 2005.

[16] J. Cong and Y. Zhang, "Thermal-Driven Multilevel Routing for 3-D ICs," in *Proceedings of the Asia South Pacific Design Automation Conference*, Shanghai, China, January 2005.

[17] S. Das, A. Fan, K.-N. Chen, and C. S. Tan, "Technology, Performance, and Computer-Aided Design of Three-Dimensional Integrated Circuits," in *Proceedings of the International Symposium on Physical Design*, Phoenix, AZ, USA, April 2004, pp. 108–115.

[18] B. Goplen and S. Sapatnekar, "Efficient Thermal Placement of Standard Cells in 3D ICs Using a Force Directed Approach," in *Proceedings of the International Conference on Computer-Aided Design*, San Jose, CA, USA, November 2003, pp. 81–85.

[19] K. Puttaswamy and G. H. Loh, "Implementing Caches in a 3D Technology for High Performance Processors," in *Proceedings of the International Conference on Computer Design*, San Jose, CA, USA, October 2005.

[20] A. Rahman and R. Reif, "System Level Performance Evaluation of Three-Dimensional Integrated Circuits," *IEEE Transactions on VLSI*, vol. 8, no. 6, pp. 671–678, June 2000.

[21] M. Ekpanyapong, C. Ballapuram, M. Healy, S. K. Lim, H.-H. S. Lee, and G. H. Loh, "Thermal-Aware 3D Microarchitectural Floorplanning," 2004, *Submitted/under review*.

[22] S. K. Lim, "Physical design for 3D system on package," *IEEE Design & Test of Computers*, vol. 22, no. 6, pp. 532–539, Nov.-Dec 2005.

[23] Y.-F. Tsai, Y. Xie, N. Vijaykrishnan, and M. J. Irwin, "Three-Dimensional Cache Design Using 3DCacti," in *Proceedings of the International Conference on Computer Design*, San Jose, CA, USA, October 2005.

[24] K. Puttaswamy and G. H. Loh, "Implementing Register Files for High-Performance Microprocessors in a Die-Stacked (3D) Technology," in *Proceedings of the International Symposium on VLSI*, Karlsruhe, Germany, March 2006.

[25] J. W. Joyner and J. D. Meindl, "Opportunities for reduced power dissipation using three-dimensional integration," in *Proceedings of the*, 2002, pp. 148–150.

[26] S. Gupta, M. Hilbert, S. Hong, and R. Patti, "Techniques for Producing 3D ICs with High-Density Interconnect," in *Proceedings of the 21st International VLSI Multilevel Interconnection Conference*, Waikoloa Beach, HI, USA, September 2004.

[27] Y. Cao, T. Sato, D. Sylvester, M. Orshansky, and C. Hu, "New Paradigm of Predictive MOSFET and Interconnect Modeling for Early Circuit Design," in *Proceedings of the 2000 Custom Integrated Circuits Conference*, Orlando, FL, USA, May 2000, pp. 201–204.

[28] Y. Deng and W. Maly, "2.5D System Integration: A Design Driven System Implementation Schema," in *Proceedings of the Asia South Pacific Design Automation Conference*, Yokohama, Japan, January 2004, pp. 450–455.

[29] J.-Q. Lu and Y. Kwon and G. Rajagopalan and M. Gupta and J. McHaon and K.-W. Lee and R. P. Kraft and J. F. McDonald and T. S. Cale and R. J. Gutmann and B. Xu and E. Eisenbraun and J. Castracane and A. Kaloyeros, "A wafer-scale 3d ic technology platform using dielectric bonding glues and copper damascene patterned inter-wafer interconnects," in *Proceedings of the*, Troy, NY, USA, 2002, pp. 78–80.

[30] K. Skadron, M. R. Stan, K. Sankaranarayanan, W. Huang, S. Velusamy, and D. Tarjan, "Temperature-Aware Microarchitecture: Modeling and Implementation," *Transactions on Architecture and Code Optimization*, vol. 1, no. 1, pp. 94–125, March 2004.

[31] Y. Li, B. Lee, D. Brooks, Z. Hu, and K. Skadron, "CMP Design Space Exploration Subject to Physical Constraints," in *Proceedings of the 12th International Symposium on High Performance Computer Architecture*, 2006.

[32] R. S. Prasher, J.-Y. Chang, I. Sauciuc, S. Narasimhan, D. Chau, G. Chrysler, A. Myers, S. Prstic, and C. Hu, "Nano and Micro Technology-Based Next-Generation Package-Level Cooling Solutions," *Intel Technology Journal*, vol. 9, no. 4, November 2005.