Analyzing the Dark Silicon Phenomenon in a Many-Core Chip Multi-Processor under Deeply-Scaled Process Technologies

Alireza Shafaei Yanzhi Wang Srikanth Ramadurgam

Yuankun Xue Paul Bogdan Massoud Pedram

Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089
{shafaeib, yanzhiwa, ramadurg, yuankunx, pbogdan, pedram}@usc.edu

ABSTRACT

The impact of dark silicon phenomenon on multicore processors under deeply-scaled FinFET technologies is investigated in this paper. To do this accurately, a cross-layer framework, spanning device, circuit, and architecture levels is initially introduced. Using this framework, leakage and dynamic power consumptions as well as frequency levels of in-order and out-of-order (OoO) processor cores, and on-chip cache memories and routers in a network-on-chip-based chip multiprocessor system synthesized in 7nm FinFET technology and operating in both super- and near-threshold voltage regimes are presented. Subsequently, total power consumptions of multicore chips manufactured with (i) OoO and (ii) in-order processor cores are reported and compared. According to our results, for a 64-core chip and 15W thermal design power budget, 64% and 39% dark silicon are observed in OoO and in-order multicores, respectively, under super-threshold regime. These percentages drop to 19% and 0% for OoO and in-order multicores operating in the near-threshold regime, respectively. Furthermore, the highest energy efficiencies are achieved by operating in the near-threshold regime, which points to the effectiveness of near-threshold computing in mitigating the effect of dark silicon phenomenon under deeply-scaled technologies.

Categories and Subject Descriptors

B [**Hardware**]: General

Keywords

FinFET devices, dark silicon, near-threshold computing

GLSVLSI'15, May 20–22, 2015, Pittsburgh, PA, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3474-7/15/05 ...$15.00.
http://dx.doi.org/10.1145/2742060.2742096.

1. INTRODUCTION

Transistor dimensions have been shrinking in each technology generation, resulting in more transistors with faster switching speeds in successive technology nodes. On the other hand, in order to maintain the power density at a constant level, the supply voltage, V_{dd}, must also be scaled down by a similar factor as the feature size, which subsequently necessitates a reduced threshold voltage value, V_{th}. Decreasing the V_{th} induces an exponential increase in the OFF (leakage) current of the underlying devices, an effect which is not desired especially under nanometer technology nodes where the leakage power is leading the chip power consumption. As a result, V_{th}, and accordingly V_{dd}, are not scaling proportionally with the feature size. This phenomenon has in turn resulted in increased chip power density. However, our ability to remove heat from VLSI chips (using advanced packaging and cooling technologies) is rather limited. Hence, conventional processor performance scaling (which was mainly achieved by boosting the clock frequency of compute cores from one generation to next) has come to an end. Instead, system designers try to achieve higher computational capacity for their processors by integrating more cores onto the same chip, each core operating at the same peak frequency level as a previous-generation core. Clearly, achieving higher performance with respect to single-threaded applications has come to halt, although the performance on multi-threaded applications and ability to execute multiple concurrent applications on the same chip has been increasing.

Unfortunately, a new limitation (typically referred to as *dark silicon* phenomenon [11, 19, 17]) has arisen. This phenomenon refers to the fact that although we have the silicon real state on a die to integrate many cores onto the same chip, many of the integrated cores cannot be powered up at the same time, because the resulting power consumption will create power densities that will exceed the acceptable limits imposed on any die. The aforesaid limits are typically captured as a *thermal design power* (TDP)[1] for the chip. In other words, even though by scaling-down to new technology nodes and shrinking the transistor sizes, more cores can

[1] TDP is the maximum amount of power that chip can safely dissipate through the cooling system.

be packed on a same-area chip, only a subset of cores can be active at any time for a given TDP.

On the other hand, the chip industry is undergoing a technology shift from conventional planar CMOS transistors towards quasi-planar FinFET devices [2, 6, 1]. This is because of the improved (three-dimensional) gate control over the channel which diminishes source and drain controls, thereby reducing short channel effects [18]. Furthermore, a FinFET device offers higher immunity to random variations which mainly result from the undoped channel of FinFET devices [14, 20]. Additionally, the minimum energy point and the minimum energy-delay point of FinFET circuits occur at supply voltage levels lower than that of planar CMOS counterparts [13], enabling more aggressive voltage scalability in FinFET-based circuit designs. Because of these advantages, FinFET devices are currently recognized as a promising choice of device for deeply-scaled technologies, i.e., technology nodes beyond the 10nm regime [15].

It is predicted in [11] that regardless of the chip organization and topology, multicore scaling is power limited and a significant portion of the fixed-size chip needs to be powered-off, e.g., 50% in 8nm. This projection for future technology nodes is in fact based on 2010 release of the ITRS, which does not adequately consider the effect of the transition from bulk CMOS to FinFET process technologies. To mitigate this shortcoming, we present a device-circuit-architecture cross-layer framework to project multicore scaling and demonstrate the dark silicon phenomenon in deeply-scaled FinFET technologies. More precisely, at the device-level, we design and optimize FinFET devices with gate length of 7nm using advanced device simulators from Synopsys TCAD tool suite [5]. We then extract compact Verilog-A models in order to perform fast gate- and circuit-level simulations, and characterize a library of standard cells. Using this library of standard cells, we synthesize processor cores and *network-on-chip* (NoC) routers by using Synopsys Design Compiler, and report their frequency level and power consumption. Furthermore, characteristics of cache memories are derived from a modified version of CACTI with FinFET support [16].

In this paper, in order to study the effect of dark silicon in future technologies, we consider the following multicore platforms: (i) an out-of-order (OoO) multicore processor, where each core is a Nehalem-based OoO processor, and (ii) an in-order multicore processor, where each core is a LEON3 [3] microprocessor. According to our results, for a 64-core chip and 15W TDP budget, 64% (19%) and 39% (0%) dark silicon are observed in OoO and in-order processors, respectively, under super-threshold (near-threshold) regime. For a TDP of 20W, there is no dark silicon in both processors operating in the near-threshold regime, and the amount of dark silicon for super-threshold operation is reduced to 55% and 22% for the OoO and in-order processors, respectively.

The rest of this paper is organized as follows. The cross-layer design framework is introduced in Section 2, followed by the dark silicon prediction methodology for deeply-scaled technologies in Section 3. Prediction results are presented in Section 4, and finally, Section 5 concludes the paper.

2. CROSS-LAYER DESIGN FRAMEWORK

Our objective is to estimate the power consumption of major components of a multicore platform (i.e., processor cores, on-chip cache memories, and NoC routers) in future FinFET devices. Subsequently, we will use this information to predict the amount of dark silicon in processors manufactured with such deeply-scaled devices. To do this accurately, we adopt a cross-layer design framework, spanning device, circuit, and architecture levels, which is shown in Figure 1. Details of this framework are described next.

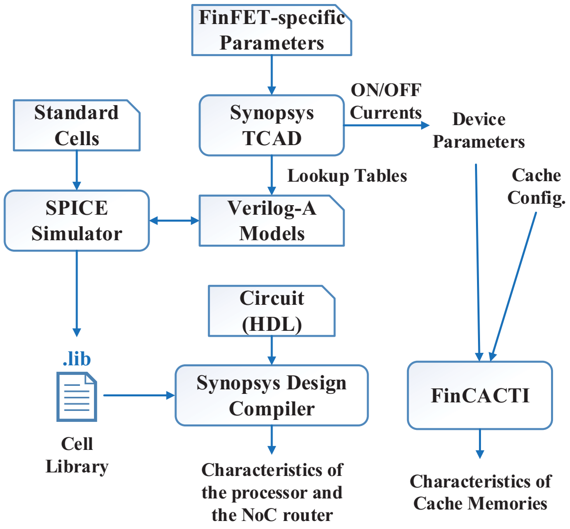

Figure 1: The cross-layer design framework for synthesizing/characterizing processor cores, NoC routers, and cache memories using FinFET devices.

2.1 Device-Level Design

FinFET devices [18] are currently viewed as the technology-of-choice beyond the 22nm regime [2, 6, 1], due to the improved gate control over the channel, and the reduced leakage current, sensitivity to process variations, and short-channel effects. As no industrial data is available, we build and simulate 7nm FinFET devices using Synopsys Sentaurus Device, the advanced multidimensional device simulator from the TCAD tool suite [5]. The gate length of our FinFET devices is 7nm, with 1.5nm gate underlap on each side, resulting in a channel length of 10nm. Furthermore, the nominal operating voltage is 0.45V, the threshold voltage is between 0.2V to 0.25V, and the subthreshold slope is ~80mV/dec, for both NFET and PFET devices.

In this paper, we consider the following two supply voltage operating modes: (i) super-threshold (ST) regime for high performance operation, and (ii) near-threshold (NT) regime for cases where the energy efficiency is the main concern. Characteristics of 7nm FinFETs [10], and for comparison purposes, 16nm PTM planar CMOS transistors [23] are reported in Table 1 for both ST and NT operations. In both regimes, the OFF current of 7nm FinFET devices is lower than that of 16nm PTM counterpart, which is approximately 12× smaller in the ST regime, but only 2× smaller in the NT operation mode. Basically, because of the negligible *drain induced barrier lowering* (DIBL) effect in FinFET devices, OFF currents of FinFET devices in ST and NT regimes are almost identical. However, for 16nm planar CMOS devices, the OFF current in the NT regime is ~5× lower than that of the ST regime.

Table 1: Characteristics of 7nm FinFET and 16nm planar CMOS (PTM) devices for super-threshold (ST) and near-threshold (NT) regimes.

Device	Operating	V_{dd}	ON Current (A/μm)		OFF Current (A/μm)		ON/OFF Current Ratio		Reference
Library	Mode	(V)	NFET	PFET	NFET	PFET	NFET	PFET	
7nm FinFET	ST	0.45	8.818e-04	5.504e-04	3.811e-08	5.782e-08	23,140	9,518	[10]
7nm FinFET	NT	0.3	1.494e-04	1.366e-04	3.497e-08	5.675e-08	4,272	2,408	
16nm PTM	ST	0.7	1.397e-03	9.839e-04	4.884e-07	6.084e-07	2,860	1,617	[23]
16nm PTM	NT	0.5	5.415e-04	3.621e-04	1.009e-07	9.081e-08	5,367	3,987	

2.2 Circuit-Level Design

Based on device simulations, we also extract compact Verilog-A models which serve as the interface between the SPICE engine and the device simulator. These SPICE-compatible Verilog-A models allow us to perform fast gate- and circuit-level simulations, compared with the extremely slow device-level simulations. An important application of the Verilog-A models is to characterize a library of standard cells which includes timing and power models as well as layout information for a set of combinational (e.g., INV, NAND, NOR, XOR) and sequential (e.g., latch and D-flip-flop) logic gates [22]. This information is then stored in the Liberty library format (.lib), and are later used in order to synthesize logic circuits, such as processor cores and NoC routers.

We also develop standard 6T and the more robust 8T SRAM cells made of the 7nm FinFET devices, which will be used as the main building blocks of cache memories. For both 6T and 8T SRAM cells, we (i) derive the cell area from the layout information in order to obtain the memory density, (ii) measure the *static noise margin* (SNM) in order to ensure the robust operation of the SRAM cell under deeply-scaled technology nodes, and (iii) calculate the leakage power of the SRAM cell. SNM and leakage power of SRAM cells are measured by SPICE simulations using the Verilog-A models.

2.3 Architecture-Level Evaluation

We synthesize the LEON3 [3] seven-stage processor (with cache memories excluded) and the Open Source NoC Router RTL [7] (with 128-bit link width, and two virtual channels per input port) based on the developed 7nm FinFET standard cell library. The LEON3 is a fully synthesisable VHDL model of a 32-bit processor based on the SPARC-V8 RISC architecture, and is selected as the in-order processor core in this paper. However, for an advanced OoO core, we adopt a Nehalem-based processor, and since we cannot find an RTL description for such processor, the frequency and power consumption are calculated using the McPAT tool [12]. Characteristics of the Nehalem-based processor core are derived under 45nm technology node, but are then scaled down to 7nm FinFET technology. In order to derive the appropriate technology scaling factors, we use Synopsys Design Compiler to synthesize several ISCAS bechmark circuits and processors using 45nm NanGate and 7nm FinFET standard cell libraries (under both ST and NT regimes).

In order to characterize FinFET-based cache memories, we use a modified version of CACTI tool [16], which also supports 7nm FinFET devices as well as the standard 8T SRAM cell. This modified CACTI tool also provides XML files for introducing new technologies and/or devices. By using these XML interfaces, we are able to characterize cache memories made of 7nm FinFET and 16nm planar CMOS devices. Testing results on a 16KB L1 cache demonstrates that 4.6× per-access energy reduction and 13× leakage power consumption reduction can be achieved when comparing 7nm FinFET and 16nm planar CMOS devices.

3. DARK SILICON PREDICTION METHODOLOGY

The methodology that we use in order to predict the percentage of dark silicon in future multicore processors made of deeply-scaled FinFET devices is described in this section.

3.1 Prediction Methodology for OoO Processors

We project OoO multicore scaling into 7nm gate-length FinFET technology using the following procedure. We adopt the Sniper [9] multicore simulator to execute various applications from the PARSEC [8] and SPLASH2 [21] benchmarks. We adopt a 64-core as the OoO multicore platform built in 45nm bulk CMOS technology, with Nehalem-based cores, individual L1 data and instruction caches with the following configurations: 32KB, 4-way, 3-cycle latency, 1-bank, LRU replacement policy, and individual L2 cache with the following configurations: 256KB, 4-way, 6-cycle latency, 2-bank, LRU replacement policy. The measured operating frequency of the OoO multicore platform is 2.6GHz.

When projecting into 7nm gate-length FinFET technology, we make the assumption that the projected multicore processor with 7nm FinFET technology uses the same core structure and L1/L2 cache configurations with the original processor with 45nm bulk CMOS technology. More specifically, the L1 instruction/data caches and L2 cache still take 3 cycles and 6 cycles, respectively. Since the propagation delay of logic circuits has better scalability with technology nodes than cache memories, the clock frequency of the 7nm FinFET processor will be dominated by cache latency. As a result, the clock frequency of the 7nm FinFET 64-core processor is determined from our modified CACTI tool and is ∼5GHz. For the data path using 7nm FinFET technology, we derive the dynamic energy consumption and leakage power consumption from the circuit synthesis results and power traces of the 64-core platform executing benchmarks. Please note that the processing core with 7nm FinFET technology may have larger portion of slack time in each clock cycle because the clock cycle is determined by the cache scaling, and thus, the leakage energy consumption may be more significant in this case.

Table 3: Frequency levels of the adopted processor cores, and the NoC router under 7nm FinFET technology, for super-threshold (ST) and near-threshold (NT) regimes. Frequencies are reported in GHz.

	OoO		In-Order	
Component	**ST**	**NT**	**ST**	**NT**
Processor	5	3.03	2.86	1.52
Router	4	3.03	2.86	1.52

3.2 Prediction Methodology for In-Order Processors

The in-order multicore platform is a 64-core processor, made of LEON3 cores, with individual L1 data and instruction cache memories with the following configurations: 16KB, direct-mapped, 1-cycle latency, 1-bank, LRU replacement policy, and individual L2 cache with the same configuration as that of the OoO multicore platform. In order to have a single-cycle L1 cache, the operating frequency of the in-order multicore platform is determined by the clock cycle of the L1 cache memory, which is 2.86GHz (measured by FinCACTI tool [16]). On the other hand, since the VHDL description of the LEON3 is available, we can synthesize it using our 7nm FinFET standard cell library. However, because LEON3 is a simple microprocessor, a scaling factor to translate the power consumption of the LEON3 to an actual processor is needed, which is described next.

We adopt a simple OoO processor, called mor1kx (Cappuccino implementation) [4], which is written in Verilog HDL, and synthesize it using Synopsys Design Compiler under our 7nm FinFET standard cell library. We then divide the leakage and dynamic power consumptions of the Nehalem-based core (obtained from McPAT) by the corresponding power component of the mor1kx core (obtained from Design Compiler) in order to derive the factor that scales the power consumption of a simple processor to an actual implementation. By multiplying this scaling factor by the LEON3 results, we can obtain the power consumption of a complex in-order processor. The reason that we are interested in in-order multicore processors is because such processors may be the future trend in many-core platforms.

4. RESULTS AND DISCUSSION

We assume that the multicore processor contains 64 tiles, where each tile is comprised of a processor core (OoO or in-order, depending on the type of the multicore processor), private L1 and L2 cache memories, and an NoC router. Routers are arranged in a two-dimensional 8×8 mesh topology. In this section, we adopt the following multicore platforms: (i) OoO - ST, (ii) OoO - NT, (iii) in-order - ST, and (iv) in-order - NT, where the first term denotes the type of core, and the second term indicates the operating mode of the multicore processor.

Table 2 reports the power consumptions of different components of the adopted multicore platforms for the 7nm FinFET technology. Frequency levels of the core and router for each platform are also shown in Table 3. The OoO processor consumes 1.9× (2.1×) more power compared with its in-order counterpart under the ST (NT) regime. On the other hand, the total power consumption of the OoO (in-order) tile has been reduced from 595mW (315mW) in the ST regime to 222mW (107mW) in the NT regime, resulting in 2.7× (3×) power reduction. In fact, using in-order cores and especially operating at the NT regime significantly reduces the power consumption, but this power reduction comes at the cost of performance degradation.

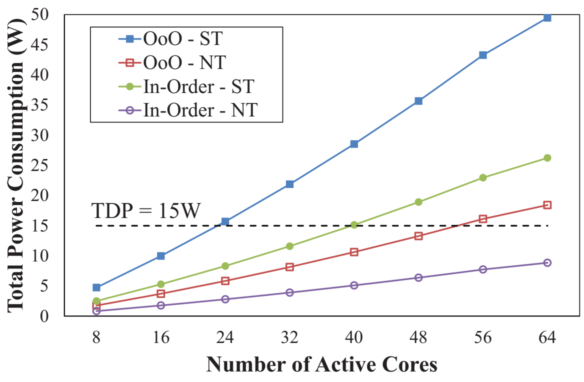

Figure 3: Total power consumption of different multicore platforms under 7nm FinFET technology vs. the number of active (turned-on) cores at the same time. TDP limits the number of cores that can be turned on at the same time, resulting in the dark silicon phenomenon.

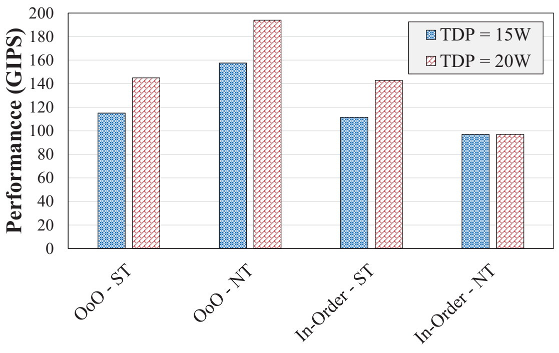

Figure 4: Performance of different multicore platforms under 7nm FinFET technology for TDP=15W and TDP=20W. Performance of a processor is defined as billion instructions per second (GIPS).

Using the results of Table 2, power breakdowns of the OoO and in-order tiles under ST and NT regimes are calculated and shown in Figure 2. Because of the architectural complexity of the OoO core, the processor is the main component of the power consumption of the OoO tile. However, by moving to a simpler in-order core, L2 cache (because of the high leakage power consumption) becomes the main component of the power consumption of the in-order tile. More precisely, the cache memory system (including L1 and L2 caches) dominates the total power consumption of the in-order processor.

We measure the total power consumption of the adopted multicore platforms, assuming that a subset of cores, varying from 8 to 64, are powered on (active) at the same time. For this purpose, a parallelism penalty of 0.625% is added to the total power consumption per core, which basically is

Table 2: **Power consumptions of the OoO and in-order cores, L1 and L2 caches, as well as the NoC router under 7nm FinFET technology, for super-threshold (ST) and near-threshold (NT) regimes. P_{dyn}, P_{leak}, and P_{tot} denote the average dynamic, leakage, and total power consumptions, respectively. No parallelism penalty is assumed for the reported power consumption of the 64-core platform in this table. All powers are reported in mW.**

	OoO - ST			OoO - NT			In-Order - ST			In-Order - NT		
Component	P_{dyn}	P_{leak}	P_{tot}	P_{dyn}	P_{leak}	P_{tot}	P_{dyn}	P_{leak}	P_{tot}	P_{dyn}	P_{leak}	P_{tot}
Processor	151	124	275	22	70	93	19	15	34	2	7	9
L1	73	37	110	7	21	29	82	19	102	8	11	19
L2	10	93	104	1	55	56	10	93	104	1	55	56
Router	39	67	106	6	38	44	28	48	76	3	19	22
Tile	273	321	595	36	185	222	140	176	315	15	92	107
64-Core			38,052			14,178			20,188			6,816

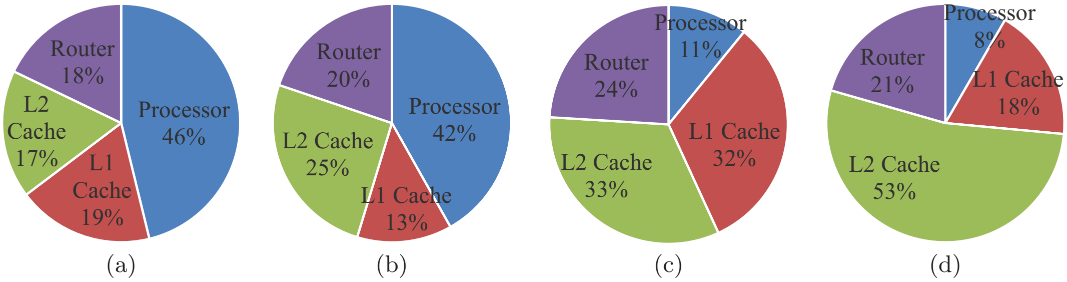

Figure 2: Power breakdowns of the OoO multicore processor in (a) super-threshold (ST) and (b) near-threshold (NT) regimes, and the in-order multicore processor in (c) ST and (d) NT regimes under 7nm FinFET technology.

due to the power overheads imposed by the cache coherency protocol and network congestions in cases where more cores are powered on. Results are illustrated in Figure 3. As can be seen, the in-order multicore platform operating in the NT regime results in the lowest power consumption, and does not experience any dark silicon effect even with TDP = 10W. In other multicore platforms, however, a portion of the core should be powered off, pointing to the existence of the dark silicon on chip.

In order to measure the effect of dark silicon in the adopted multicore platforms, we use TDP = 15W and TDP = 20W. Based on these TDP levels, the maximum number of cores that can be active at the same time, the total power consumption, and the percentage of cores that should be left dark in each multicore processor are reported in Table 4. Based on this table, we make the following observations. For TDP = 15W, 64% (19%) and 39% (0%) dark silicon are observed in OoO and in-order multicores, respectively, under the ST (NT) regime. However, increasing the TDP budget to 20W reduces the amount of dark silicon to 55% and 22% in OoO and in-order multicores operating in the ST regime, respectively, but leaves no dark silicon under the NT regime.

Finally, in order to make a conclusion, the performance as well as the energy efficiency of the multicore platforms are evaluated. For this purpose, performance of the processor is defined as billion instructions per second, and is denoted by $GIPS = N_{max} \times f_{clk}$, where N_{max} and f_{clk} represent the maximum number of active cores at any time and the clock frequency of the processor, respectively. On the other hand, the energy efficiency of the processor is defined as billion instructions per second per watt (or billion instructions per joule), and is denoted by $GIPS/W = GIPS\ /\ P_{total}$, where P_{total} is the total power consumption of the multicore processor when N_{max} cores are active at f_{clk}. Performance and energy efficiency values of the adopted multicore platforms are shown in Figure 4 and Figure 5, respectively. We can observe that for TDP=15W, the OoO processor operating in the near-threshold regime achieves the highest performance as well as the highest energy efficiency among all adopted multicore platforms. The highest energy efficiency under TDP=20W is obtained by using the in-order processor operating in the near-threshold regime, which is because of the extremely low power operation of this platform.

These results point to the effectiveness of the near-threshold operation in mitigating the effect of the dark silicon phenomenon under deeply-scaled FinFET technologies. In fact, the near-threshold operation enhances the performance by allowing more cores to be active at any time, and hence enabling more aggressive parallelism, and also increases the energy efficiency, because of operating in the minimum energy operation point of the system.

5. CONCLUSION

We studied the effect of dark silicon in future FinFET technologies for OoO and in-order multicore processors under ST and NT operating modes. For this purpose, a device-circuit-architecture cross-layer design and analysis framework has been introduced, which is adopted in order to derive the leakage and dynamic power consumptions as well frequency levels of OoO and in-order processor cores, on-chip L1 and L2 cache memories, and NoC routers. According to our results, for a 64-core chip and 15W thermal design power budget, 64% (19%) and 39% (0%) dark silicon are ob-

Table 4: Prediction of dark silicon in different multicore platforms in 7nm FinFET technology under 15W and 20W TDP values.

Multicore Platform	TDP = 15W			TDP = 20W		
	Max # of Active Cores	Total Power (W)	Dark Silicon	Max # of Active Cores	Total Power (W)	Dark Silicon
OoO - ST	23	14.957	64%	29	19.505	55%
OoO - NT	52	14.688	19%	64	19.140	0%
In-Order - ST	39	14.686	39%	50	19.912	22%
In-Order - NT	64	9.202	0%	64	9.202	0%

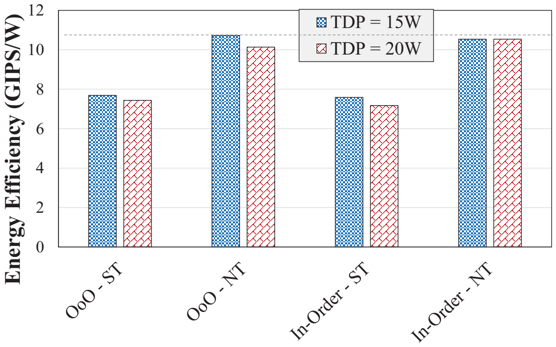

Figure 5: Energy efficiency of different multicore platforms under 7nm FinFET technology for TDP=15W and TDP=20W. Energy efficiency of a processor is defined as billion instructions per second per watt (GIPS/W).

served in OoO and in-order multicores, respectively, under the ST (NT) regime. Furthermore, performance and energy efficiency results of the multicore platforms point to the effectiveness of near-threshold computing in mitigating the effect of dark silicon phenomenon in deeply-scaled FinFET technologies.

6. ACKNOWLEDGMENTS

This research is supported by grants from the PERFECT program of the Defense Advanced Research Projects Agency and the Software and Hardware Foundations of the National Science Foundation.

7. REFERENCES

[1] GLOBALFOUNDRIES 14nm technology. [Online]. Available: http://www.globalfoundries.com/technology-solutions/leading-edge-technology/14-lpe-lpp

[2] Intel®22nm technology. [Online]. Available: http://www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html

[3] LEON3 processor. [Online]. Available: http://www.gaisler.com/index.php/products/processors/leon3

[4] mor1kx - an OpenRISC Processor IP Core. [Online]. Available: https://github.com/openrisc/mor1kx

[5] Synopsys Technology Computer-Aided Design (TCAD). [Online]. Available: http://www.synopsys.com/tools/tcad

[6] TSMC 16nm technology. [Online]. Available: http://www.tsmc.com/english/dedicatedFoundry/technology/16nm.htm

[7] D. U. Becker. *Efficient Microarchitecture for Network-on-Chip Routers*. PhD thesis, Stanford University, August 2012.

[8] C. Bienia. *Benchmarking Modern Multiprocessors*. PhD thesis, Princeton University, January 2011.

[9] T. Carlson, W. Heirman, and L. Eeckhout. Sniper: Exploring the level of abstraction for scalable and accurate parallel multi-core simulation. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, pages 1–12, Nov 2011.

[10] S. Chen, Y. Wang, X. Lin, Q. Xie, and M. Pedram. Performance prediction for multiple-threshold 7nm-FinFET-based circuits operating in multiple voltage regimes using a cross-layer simulation framework. In *IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, Oct. 2014.

[11] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger. Dark Silicon and the End of Multicore Scaling. In *38th International Symposium on Computer Architecture (ISCA)*, pages 365–376, 2011.

[12] S. Li, J.-H. Ahn, R. Strong, J. Brockman, D. Tullsen, and N. Jouppi. McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures. In *42nd International Symposium on Microarchitecture (MICRO-42)*, pages 469–480, Dec 2009.

[13] X. Lin, Y. Wang, and M. Pedram. Joint sizing and adaptive independent gate control for FinFET circuits operating in multiple voltage regimes using the logical effort method. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 444–449, Nov 2013.

[14] T. Matsukawa, S. O'uchi, K. Endo, Y. Ishikawa, H. Yamauchi, Y. X. Liu, J. Tsukada, K. Sakamoto, and M. Masahara. Comprehensive Analysis of Variability Sources of FinFET Characteristics. In *Symposium on VLSI Technology*, 2009.

[15] E. Nowak, I. Aller, T. Ludwig, K. Kim, R. Joshi, C.-T. Chuang, K. Bernstein, and R. Puri. Turning Silicon on its Edge [Double Gate CMOS/FinFET Technology]. *IEEE Circuits and Devices Magazine*, 20(1):20–31, 2004.

[16] A. Shafaei, Y. Wang, X. Lin, and M. Pedram. FinCACTI: Architectural Analysis and Modeling of Caches with Deeply-Scaled FinFET Devices. In *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2014.

[17] M. Shafique, S. Garg, J. Henkel, and D. Marculescu. The EDA Challenges in the Dark Silicon Era: Temperature, Reliability, and Variability Perspectives. In *51st Design Automation Conference (DAC)*, pages 185:1–185:6, 2014.

[18] S. Tang, L. Chang, N. Lindert, Y.-K. Choi, W.-C. Lee, X. Huang, V. Subramanian, J. Bokor, T.-J. King, and C. Hu. Finfet - a quasi-planar double-gate mosfet. In *IEEE International Solid-State Circuits Conference (ISSCC)*, 2001.

[19] M. B. Taylor. Is Dark Silicon Useful?: Harnessing the Four Horsemen of the Coming Dark Silicon Apocalypse. In *49th Design Automation Conference (DAC)*, 2012.

[20] X. Wang, A. Brown, B. Cheng, and A. Asenov. Statistical Variability and Reliability in Nanoscale FinFETs. In *IEEE International Electron Devices Meeting (IEDM)*, pages 5.4.1–5.4.4, Dec 2011.

[21] S. C. Woo, M. Ohara, E. Torrie, J. P. Singh, and A. Gupta. The splash-2 programs: Characterization and methodological considerations. In *Proceedings of the 22Nd Annual International Symposium on Computer Architecture (ISCA)*, ISCA '95, pages 24–36, 1995.

[22] Q. Xie, X. Lin, Y. Wang, M. Dousti, A. Shafaei, M. Ghasemi-Gol, and M. Pedram. 5nm FinFET Standard Cell Library Optimization and Circuit Synthesis in Near-and Super-Threshold Voltage Regimes. In *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2014.

[23] W. Zhao and Y. Cao. New Generation of Predictive Technology Model for Sub-45nm Early Design Exploration. *IEEE Transactions on Electron Devices*, 53(11):2816–2823, 2006.