

The EDA Challenges in the Dark Silicon Era

(Temperature, Reliability, and Variability Perspectives)

Muhammad Shafique
Chair for Embedded Systems
Karlsruhe Institute of Technology
muhammad.shafique@kit.edu

Siddharth Garg
Department of ECE
University of Waterloo
s6garg@uwaterloo.ca

Jörg Henkel
Chair for Embedded Systems
Karlsruhe Institute of Technology
henkel@kit.edu

Diana Marculescu
Department of ECE
Carnegie Mellon University
dianam@cmu.edu

Abstract—Technology scaling has resulted in smaller and faster transistors in successive technology generations. However, transistor power consumption no longer scales commensurately with integration density and, consequently, it is projected that in future technology nodes it will only be possible to simultaneously power on a fraction of cores on a multi-core chip in order to stay within the power budget. The part of the chip that is powered off is referred to as dark silicon and brings new challenges as well as opportunities for the design community, particularly in the context of the interaction of dark silicon with thermal, reliability and variability concerns. In this perspectives paper we describe these new challenges and opportunities, and provide preliminary experimental evidence in their support.

I. INTRODUCTION

Smaller feature sizes in every new technology node have enabled higher integration, faster switching and lower power consumption per transistor. By scaling supply voltage and threshold voltage by about the same factor as feature size, designers were able to obtain a commensurate decrease in switching power per transistor, such that the power density (power consumption per chip area) remained (approximately) constant from one technology node to another. This is referred to as the Dennard Scaling model. However, in leakage dominated, deep sub-micron technology nodes, reducing threshold voltage results in an exponential increase in leakage power. Hence, threshold voltage is no longer scaling, and, as a consequence, supply voltage cannot be scaled further without impacting performance. Thus, although we can still pack more transistors per area with technology scaling, the switching power per transistor *is not* scaling commensurately, and hence power density has been trending *upwards*. Coupled with the physical limits imposed by device packaging and cooling technology on the peak power and peak power density, this results in the so-called *Dark Silicon* era [7], [32], [16].

The new constraint imposed by dark silicon is that not all the transistors on the chip can be simultaneously powered on at full performance for a given *thermal design power* (TDP). The TDP is the maximum amount of power that can be supplied to the chip to ensure that the chip will operate within the safe range, i.e., below the thermal safe temperature, T_{safe} . In case the TDP is violated, the chip will generate heat at a faster pace than can be dissipated

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

DAC '14, June 01 - 05 2014, San Francisco, CA, USA
Copyright 2014 ACM 978-1-4503-2730-5/14/06\$15.00.
<http://dx.doi.org/10.1145/2593069.2593229>

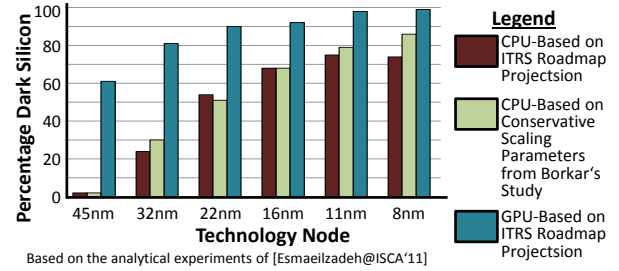


Fig. 1: Dark silicon trends for different technology nodes.

by the cooling system. Based on the technological data from ITRS and Intel, at the 8nm node, more than 50% of the chip area will be *dark* (see Fig. 1 (b)) [7], [16]. The question that we seek to address in this perspectives manuscript is the following: does the advent of dark silicon introduce new challenges and problems for the EDA community to address, or is it business as usual? Recent work has shown the need for architectural innovation for dark silicon chips, including the use of heterogeneous multi-cores and accelerator-rich architectures. Here we provide a different perspective, focused on the thermal management, reliability and variability concerns in dark silicon.

II. STATE-OF-THE-ART IN DARK SILICON

Given the abundance of transistors in dark silicon chips, the question is *if* and *how* they can be harnessed to improve performance within a power or peak temperature constraint. Much of the existing work in literature addresses this question, based on different design philosophies that include: (i) use of architectural heterogeneity and specialized cores; (ii) approximate computing; and (iii) so-called *dim silicon* that employs near-threshold voltages to enable a larger fraction of the chip to be powered on, albeit at lower voltages. We will discuss each approach separately.

Architectural Heterogeneity and Specialized Cores: Early research work has focused on exploiting the dark silicon area for designing specialized cores, incorporating heterogeneity, and/or application-specific, hardware accelerators [34], [11], [4]. In [34], *conservation cores* (c-cores) are synthesized to execute energy-intensive sections of an application, while the other sections of the application are executed by a general purpose core. The c-cores are powered-on only when their functionality is needed. Since c-cores are application-specific, a large number of c-cores are required to cover a wide range of applications. Therefore, quasi-specific cores (QsCores) [35] are proposed to target common templates in general-purpose applications, so that the same QsCores can be used by a wide range of applications.

Lyons et al. [22] and Cong et al. [5] propose memory hierarchy and

interconnect designs for accelerator rich architectures, respectively. Other work has focused on architectural abstractions that enable accelerator sharing and regulate contention for accelerator resources [4].

Another direction of research targets the design of general purpose heterogeneous multicore processors for dark silicon chips. Turakhia et al. [33] propose an architectural synthesis approach that determines the number of cores of each type to provision given an area and peak power constraint. [31] leverage device-level heterogeneity to provide trade-off points between energy-efficiency and performance.

Approximate Computing: Given that the power consumption of cores no longer scales quadratically with feature size, other methods of scaling power consumption must be found to alleviate the dark silicon challenge. One such approach is approximate computing, that relies on trading energy efficiency with accuracy, especially for error-tolerant applications like vision, machine learning, etc. [13], [9]. Approximate computing techniques at various levels of design abstraction have been discussed in the literature, ranging from circuit level techniques [3], to approximate data paths [36], [21], [24], [12] and programming language support [8].

Power Management for Dark and Dim Silicon: Recent research aims at run-time mechanisms to efficiently utilize the thermal design power (TDP) budget [25], [1] in order to maximize performance of cores that are either micro-architecturally heterogeneous, or homogeneous but synthesized with different power/performance targets. Note that these papers build upon a body of existing literature on run-time thermally constrained power management techniques for conventional processors (see for example [15], [14]), but the availability of dark silicon introduces new opportunities as we will highlight in this paper.

Computational Sprinting [29] leverages dark silicon to power-on many extra cores for a very short time period (hundreds of millisecond) to facilitate sub-second bursts of parallel computation. During this “sprint,” the active cores consume power that significantly exceeds the sustainable TDP budget, but the cores are immediately power-gated after the sprint. To handle the short temperature increase, a phase-change material is provided in the thermal packaging. Cooperative CPU/GPU performance boosting techniques have also been proposed [28]. Alternate methods are Intel’s Turbo Boost [30] and AMD’s Turbo CORE [27] technologies that leverage the temperature headroom to favor high-ILP applications by increasing the voltage/frequency of a core while power-gating other cores.

The recent work on *near-threshold computing* [19], [6], [23] presents an alternative approach to utilize dark cores, by turning on a larger fraction of the chip but at voltages close to the threshold voltage. These cores are referred to as *dim* cores. In [37], *dim silicon* with near threshold computing is integrated with the accelerator based c-cores (implemented as an ASIC or on FPGAs). *Dim silicon* works well for applications with high thread-level parallelism, but also exhibits high sensitivity to process variation and power supply fluctuations.

A. Open Challenges in the Dark Silicon Era

Much of the work discussed so far is *performance* focused, i.e., how to maximize performance within a power or temperature budget. However, technology scaling introduces new first-order concerns like reliability and variability. In addition, a number of factors that affect reliability, for example aging mechanisms, are highly correlated with peak temperature and thermal gradients, i.e., the spatio-temporal thermal profile of the chip. At the same time, manufacturing process induced variability results in core-to-core differences in leakage power and maximum frequency, which go on to affect all other important system quality metrics.

The broad open question that is unaddressed in literature is the impact of dark silicon on the spatio-temporal thermal profile, reliability and sensitivity to process variations of next generation general-purpose multicore systems.

III. OUR PERSPECTIVES

We now present our perspectives about new challenges and opportunities introduced by dark silicon in three specific, but related, contexts. Central to these ideas is the notion of **TDP diversity**: *the fact that, for dark silicon multicore processors, a multitude of different iso-power TDP modes are available*. A **TDP mode** is determined by: (i) the number of cores that are powered on, (ii) the operating mode (voltage and frequency) of each powered on core, and (iii) the location of the powered on cores. The power consumed for any TDP mode must, of course, be equal to the TDP specification of the chip.

We note that, in conventional multicore processors, there is only one TDP mode, i.e., when all cores are powered on at peak voltage and frequency (by definition). In contrast, for dark silicon multicore processors, each TDP mode results in a different thermal profile and impacts reliability differently, as we will see shortly. In addition, due to process variations, the same TDP mode on different chips can result in different thermal profiles and reliability impact. *Exploring the TDP mode induced differences in thermal behaviour and reliability impact, both within a chip and from chip-to-chip, form the unifying theme of our paper.*

The three perspectives we present in this paper are:

- P1: Thermal Management for Dark Silicon Multicores:** *“Dark silicon introduces new opportunities to optimize the thermal profile by choosing amongst one of many available TDP modes. Conversely, any reduction in peak temperature can be used to increase performance by providing more power to the chip, above and beyond the TDP.”*
- P2: TDP Diversity and Reliability Trade-offs in Dark Silicon Multicores:** *“Different TDP modes have starkly different behaviours from a reliability perspective. In addition, TDP modes expose natural trade-offs between transient and lifetime reliability mechanisms.”*
- P3: Leveraging and Exploiting Variability in Dark Silicon Multicores:** *“Although process variations have traditionally been viewed as detrimental effects, they can be exploited in dark silicon multicore processors to enhance quality metrics including peak temperature, thermal gradients and reliability.”*

In the following sub-sections, we elaborate on these perspectives in more depth and provide some early evidence supporting our ideas.

A. P1 - Thermal Management for Dark Silicon Multicores

The traditional TDP specification for conventional chips without dark silicon is less meaningful in the context of dark silicon because of the existence of multiple TDP modes. For conventional chips, a TDP specification corresponds to *only* one mode, i.e., when all the cores are active at full voltage/frequency, and consequently, a TDP specification also largely corresponds to a specific peak temperature. However, as we shall see, for dark silicon multicore processors, depending on the TDP mode, starkly different thermal profiles can be obtained. In the following, we will discuss how the selection of dark cores and run time application mapping on the powered-on cores can be used to generate different thermal profiles using the idea of dark silicon patterning (see Fig. 2).

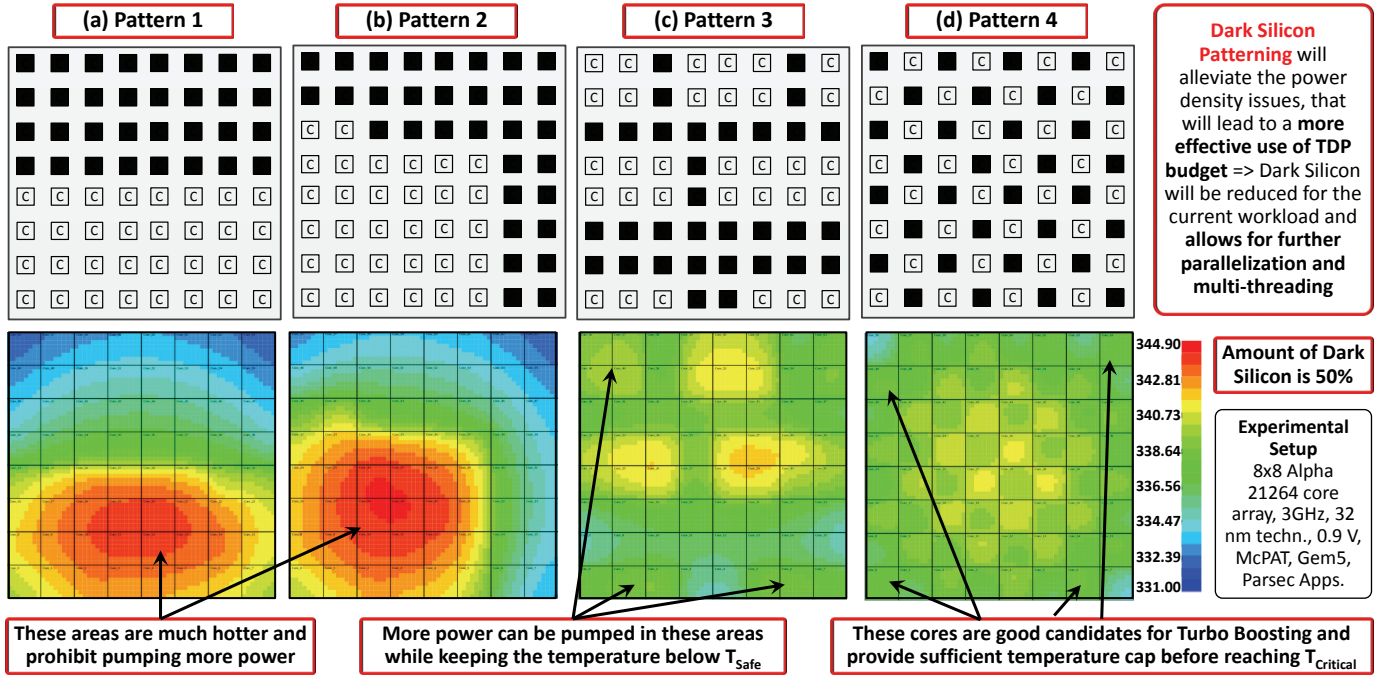


Fig. 2: Selection of dark cores impacts the chip power density and temperature profile and consequently the power budget utilization and the amount of total dark silicon.

Dark Silicon Patterning: “Dark silicon patterning determines the temporal and spatial shutdown of on-chip resources with the goal of minimizing peak temperature without violating TDP”.

Given a certain number of cores that need to be powered-on, a naive approach is to activate a set of contiguous cores, as shown in Fig. 2a. However, this may lead to scenarios of high power densities in certain on-chip regions, particularly under compute-intensive workloads. Due to high local power densities, temperature hot spots may occur on certain chip locations, much before the full chip’s average temperature exceeds $T_{critical}$. Moreover, thermal coupling may force the neighboring cores to throttle their voltage and frequency, thus degrading their performance.

To manage the dark silicon efficiently, our perspective is to determine appropriate *dark silicon patterns* (under temperature considerations) that will allow better thermal behavior by surrounding active cores with dark cores. Initial evidence of this perspective can be seen in our experimental thermal analysis for different patterns and the resulting thermal profiles as shown in Fig. 2.

However, one must be careful; for multi-threaded applications, increasing the distance between communicating threads may increase the network power dissipation and impact performance. Therefore, dark silicon patterning policies need to take effective patterning decisions, i.e., identifying the optimal TDP mode to activate at run time that jointly optimizes power density, performance and inter-core communication. These patterning decisions may also be exploited by the OS kernel to improve resource allocation and application mapping decisions.

Beyond a Single TDP Specification: Given the discussion on dark silicon patterning, and the existence of multiple TDP modes, the traditional notion of a TDP might be too pessimistic, because it might be possible to supply more power to the chip than dictated by the conventional TDP limit while still operating below the T_{safe} . Therefore, there is a strong need to devise new power metrics for run-time temperature-safe operation, i.e., how much power can a core

draw for a given pattern such that its temperature stays below T_{safe} . A motivational analysis for different cases is presented in Fig. 3.

Developing new power metrics requires analyzing the heat transfer between different cores in the presence of alternating dark and powered-on cores, and determining a relationship between temperature and power consumption. Determining the granularity at which such a model should be developed (i.e., for individual cores or a group of contiguous cores like 2x2 or 4x4) is an open research challenge. This will ensure that the system keeps operating under safe temperature range while the total power consumption exceeds TDP.

So far, the discussion has been in the context of cores running at full voltage/frequency. Of course, there exist a number of voltage/frequency operating points between a fully powered-on and a dark core. We call the cores that operate in this range *gray* cores. If patterning enables the chip to operate below T_{safe} at TDP, the available headroom can be used to activate additional cores in gray mode. An example scenario can be seen in Fig. 3.

Challenges and opportunities: To summarize, we list some new opportunities that arise from leveraging the available dark silicon:

- Leveraging dark silicon patterning to obtain desirable thermal behaviour, for example, reducing peak temperature or thermal gradients.
- Conversely, drawing more power than TDP while still keeping the temperature at any point on the chip below T_{safe} (see Fig. 2 c, d and Fig. 3).
- Higher temperature also aggravates the leakage power due to the thermal runaway issues. Efficient dark silicon patterning may allow for more cores to be activated due to the power savings from the improved thermal behaviour of dark silicon patterns.

B. P2 - Reliability and Dark Silicon Interactions

We will now highlight some new reliability related challenges and opportunities that are introduced by the availability of an abundance

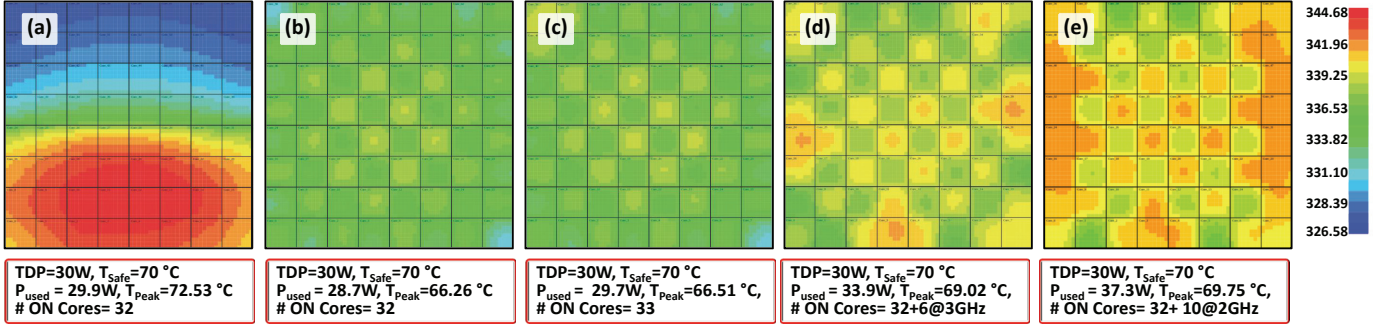


Fig. 3: The steady-state temperature of a 64-core chip for two different dark silicon patterns and different power budgets executing Parsec benchmarks. (a) corresponds to the pattern 1 of Figure 2(a) while (b-e) correspond to pattern 4 of Figure 2(d) for different number of active cores and used power. Since patterning reduces the peak temperature it enables activating more cores (c-e) and even selective boosting (d) while keeping the temperature below T_{Safe} .

| TDP Mode | 1 | 2 | 3 | 4 | 5 |
|-------------------|------|------|-------|-------|------|
| VDD (V) | 0.70 | 0.60 | 0.525 | 0.475 | 0.45 |
| Frequency (Ghz) | 4.5 | 2.95 | 1.8 | 1.11 | 0.83 |
| Num. Cores | 4 | 8 | 16 | 32 | 64 |
| Norm. Performance | 1 | 0.93 | 1.14 | 1.56 | 3.40 |

TABLE I: TDP modes and their corresponding voltages, frequencies, number of active cores and execution time for the parallel FFT benchmark.

of transistors on the chip. Of course, temperature itself has a strong impact on a number of reliability mechanisms, and therefore, the discussion in the previous sub-section is directly relevant to the issue of reliability, but as we shall see, there are several first-order reliability effects decoupled from temperature as well.

Let's begin by examining the concept of spatial graying in more detail: Table I shows five different spatial graying modes on a multicore processor with 64 (or more cores), all at the same TDP. The last row also shows the normalized execution time of a parallel FFT implementation from the SPLASH-2 benchmark suite on these different modes. Although the peak throughput for the 64 core configuration is higher, parallel applications consist of both serial and parallel sections, limiting their scalability with core count.

Each core configuration will likely have a different thermal map and different peak temperature. For some of the configurations with a fewer number of cores, patterning, as discussed in the prior section, could be used to further decrease peak temperature. Let's examine these configurations from a different perspective - soft-error reliability.

Soft errors in both caches and logic circuits (gates and latches) are an increasing concern with technology scaling [17]. The soft error rate is *directly* proportional to the exposed, active silicon area. This means that, all other things being equal, TDP mode 5 will have 16 \times the soft error rate compared to TDP mode 1 (see Table I for more details on each TDP mode). A number of factors exacerbate this even further.

First, the critical charge required to cause a single-event upset in a latch has been shown to decrease significantly with voltage scaling [2]. Thus, TDP mode 5 is likely to have an even higher soft error rate, because it runs at 0.45V (as opposed to 0.7V for TDP mode 1). Furthermore, for this example, TDP mode 5 also takes longer to execute the FFT benchmark, further increasing the likelihood of a particle strike occurring.

To help understand this effect quantitatively, we can model the soft

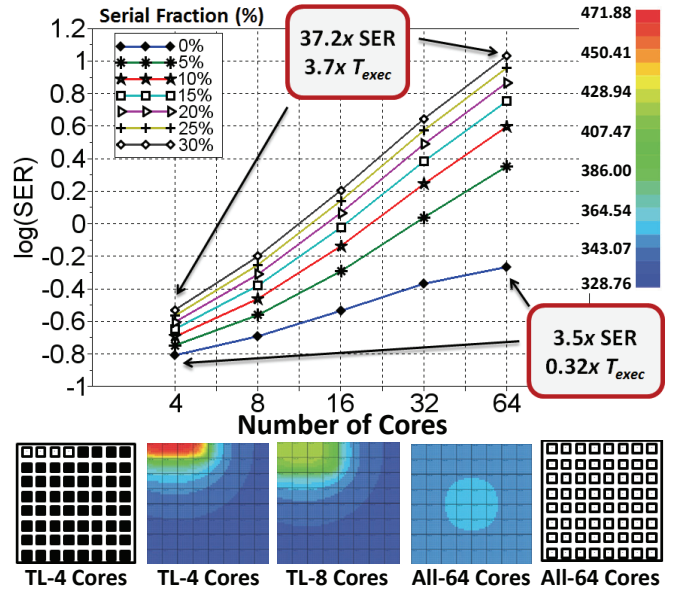


Fig. 4: Soft error rate as a function of number of active cores with varying serial fraction. For the thermal maps and floorplans in the figure, “TL” stands for “top left” and implies that contiguous cores starting from the top left corner are activated.

error rate during application execution as:

$$SER = H \times Pr\{Q_{strike} \geq Q_{crit}(V_{DD})\} \times N_{cores} \times T_{exec}$$

where H is the particle hit rate per unit area, Q_{hit} is the random variable of charge transferred on a particle hit, Q_{crit} is the V_{DD} dependent threshold beyond which a particle strike causes an upset, N_{cores} is the number of active cores, and T_{exec} is the execution latency of the benchmark.

Assuming a simple model for application execution time based on Amdahl's law, we can write T_{exec} as:

$$T_{exec} = \frac{T_{serial}}{f} + \frac{T_{par}}{f \times N_{cores}}$$

where f is the core frequency. Note that in this simplified model, performance is assumed to be directly dependent on frequency. The serial fraction of an application can be written as $\frac{T_{serial}}{T_{serial} + T_{par}}$. Finally, Chandra and Aitken [2] determine that the critical charge

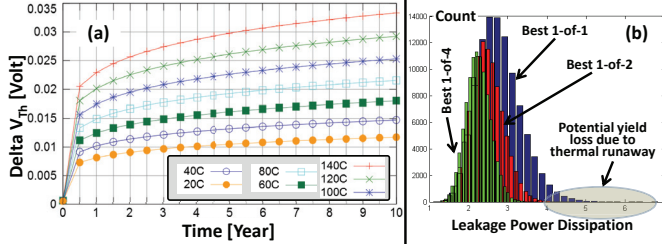


Fig. 5: (a) Impact of Temperature on the NBTI induced threshold voltage degradation. (b) Best 1-of-2 and 1-of-4 statistics for leakage power dissipation.

for a latch is an approximately linear function of V_{DD} , i.e.,

$$Q_{crit} = \alpha + \beta V_{DD}$$

Combining this, Figure 4 shows the logarithm of the soft error rate for the five different TDP modes with varying percentage of dark silicon and for applications with serial fraction ranging from 0 (perfectly parallel) to 30%. In this graph, we have assumed the charge distribution from a particle hit to be uniform, and the voltage/frequency data is from Table I. Note that for a serial fraction of 30%, the soft error rate for TDP mode 5 is more than 30 \times that of TDP mode 1. Even for perfectly parallel applications, TDP mode 5 still has 3.5 \times higher soft error rate compared to TDP mode 1. Note that, TDP mode 5 corresponds to the so-called *dim silicon* approach where cores are operated in near threshold voltages.

This raises the question, of course, of why one would use TDP mode 5 for applications with a high serial fraction. For one, as discussed in the previous section, in TDP mode 5, the same power is distributed more evenly across the chip compared to TDP mode 1. Thus, we would expect TDP mode 5 to have lower peak temperature. This is in fact the case, as illustrated in Figure 4, where thermal maps corresponding to three of the five TDP modes are shown. Observe the significant differences in peak temperature for TDP mode 1 versus TDP mode 5, although some of these differences might be alleviated by patterning.

Even if both TDP mode 1 and TDP mode 5 operate below T_{safe} , the temperature is strongly correlated with reliability. A difference between 10 $^{\circ}$ C and 15 $^{\circ}$ C can result in a 2x difference in MTTF and may increase interconnect delays by 5% [26]. Most of the major aging effects, like electromigration, stress migration, HCI, TDDDB, etc., are aggravated/accelerated by elevated temperatures. Fig. 5(a) illustrates the threshold voltage degradation (V_{th}) due to the Negative Bias Temperature Instability (NBTI) effect for a 22 nm PMOS transistor as a function of temperatures.

Challenges and opportunities: To summarize, dark silicon introduces some natural trade-offs between transient fault rates and lifetime reliability through aging mechanisms like NBTI, electro-migration, etc. In particular, we have noted that *dim silicon* modes of operation might have between 3X-30X higher soft error rates than modes in which there are large fractions of dark silicon. On the other hand, *dim silicon* modes are likely to run at reduced temperatures and cause less lifetime aging.

These natural trade-offs present the run-time manager with a rich design space of options to choose from, depending on the overall system quality targets. There is some work, notably the Bubble-wrap technique [18], that exploit different TDP modes, but only in the context of aging. How to integrate soft errors in this setting is an open question. In fact, recent work [20] has advanced the idea of leveraging the available dark silicon to provision a processor with

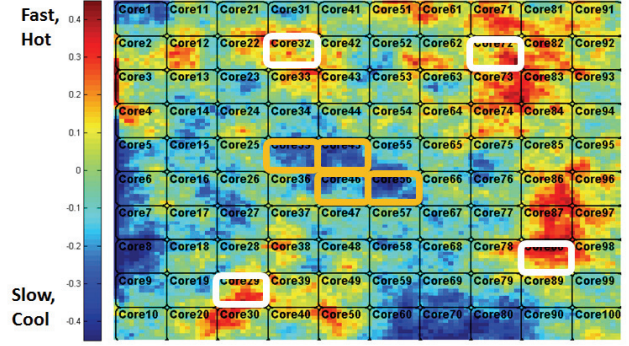


Fig. 6: Color map of process variations on a die with 100 cores. Two sets of four cores each are highlighted using the white (set 1) and orange (set 2) boxes. Depending on which set of cores is powered on, very different thermal profiles will be obtained.

multiple “reliability-heterogeneous cores” (i.e. ISA-compatible cores that provide a range of protection against different fault mechanisms at the expense of power and area), and a run-time system to manage reliability under TDP constraints.

C. P3 - Leveraging Variability in Dark Silicon Multicores

Process variations refer to the chip-to-chip and transistor-to-transistor variations in process parameters that result from non-idealities in semiconductor manufacturing. The magnitude of process variations increases with technology scaling since smaller transistors are more difficult to precisely manufacture.

At the micro-architecture level, process variations manifest as core-to-core differences in operating frequency and leakage power dissipation. For conventional multicore processors, process variations are problematic when the chip is running at TDP: the performance of a multi-threaded application is limited by the slowest thread, i.e., the core with the lowest frequency. In addition, since all cores are turned on, “leaky” cores can run significantly hotter than expected because of the feedback loop between leakage and temperature.

The redundancy introduced by dark silicon mitigates this problem. For example, “leaky” cores can be avoided (kept dark) to reduce peak temperature, and slow cores can be avoided to mitigate performance impact. This can be explained using best 1-of-K statistics: given the ability of pick the least leaky or highest frequency core out of K cores, how much benefit accrues in power/performance? Fig. 5(b) shows that with 75% of the chip dark, the tail of leakage power dissipation can be almost entirely removed by using a best 1-of-4 approach. Raghunathan et al. [10] have used similar ideas to demonstrate the performance benefits of being able to cherry-pick cores within a TDP constraint, but do not look at temperature or reliability.

Consider the variability map of a chip shown in Figure 6. The color bar indicates the leaky versus less leaky transistors on the die. Assume that we want to operate in TDP mode 1 from Table I, i.e., with four cores active. The four cores to activate can be chosen in many ways, two of which are shown in Figure 6. In the absence of variability, based on the discussions on patterning for peak temperature reduction (see Figure 2), the set of cores in the white boxes would result in lower peak temperature if turned on, compared to those in the orange boxes. However, note that the cores in the former set also have high leakage power dissipation, while the cores in the latter set have low leakage. Thus, in the presence of variability, the optimal decision must be made taking into account a *superposition* of temperature and variability maps. Similar arguments can be made with respect to aging as well.

Challenges and opportunities: Dark silicon transforms variability from a concern to an opportunity that can be exploited, at least in part. In the presence of variability, even homogeneous cores behave heterogeneously — this heterogeneity, coupled with techniques such as dark silicon patterning, might enable more reliable operation with more desirable thermal behaviour. However, a significant challenge is that, in the presence of variations, each fabricated chip is unique and different. This introduces issues with respect to how variability information can be effectively communicated to higher levels of abstractions (such as the operating system and application software) so as to make variability, reliability and thermally aware decisions.

IV. CONCLUSION

We have provided new perspectives on the challenges and opportunities that the emergence of dark silicon introduces in the context of thermal management, reliability (both transient and permanent), and variability. In particular, we note that dark silicon introduces a new run-time knob — which cores should be powered-on and which should be dark for a given TDP — which was absent before the dark silicon problem arose. Our preliminary empirical results indicate that tuning this knob can result in starkly different peak temperatures, soft error rates and lifetime aging, and the differences are further heightened in the presence of process variations.

ACKNOWLEDGMENTS

This work was partly supported by the German Research Foundation (DFG) as part of the Transregional Collaborative Research Centre "Invasive Computing" (SFB/TR 89); <http://invasic.de>, and the National Sciences and Engineering Research Council of Canada (NSERC). The authors would like to thank Heba Khdr (CES, KIT) and Bharathwaj Ragunathan (ECE, Waterloo) for assistance with experiments.

REFERENCES

- [1] J. Allred et al. Designing for dark silicon: a methodological perspective on energy efficient systems. In *Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, 2012.
- [2] V. Chandra and R. Aitken. Impact of technology and voltage scaling on the soft error susceptibility in nanoscale cmos. In *IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems*, 2008.
- [3] M. Choudhury and K. Mohanram. Approximate logic circuits for low overhead, non-intrusive concurrent error detection. In *Proceedings of the EDAA Conference on Design Automation and Test in Europe (DATE)*, 2008.
- [4] J. Cong et al. Architecture support for accelerator-rich cmps. In *Proceedings of the ACM 49th Annual Design Automation Conference (DAC)*, 2012.
- [5] J. Cong and B. Xiao. Optimization of interconnects between accelerators and shared memories in dark silicon. In *Proceedings of the 32nd IEEE/ACM International Conference on Computer-Aided Design*, 2013.
- [6] R.G. Dreslinski et al. Near-threshold computing: Reclaiming moore's law through energy efficient integrated circuits. *Proceedings of the IEEE*, 98(2):253–266, 2010.
- [7] H. Esmailzadeh et al. Dark silicon and the end of multicore scaling. In *Computer Architecture (ISCA)*, 2011 38th Annual International Symposium on, pages 365–376, 2011.
- [8] H. Esmailzadeh et al. Architecture support for disciplined approximate programming. In *Proceedings of the 17th International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 301–312, 2012.
- [9] H. Esmailzadeh et al. Neural acceleration for general-purpose approximate programs. *Micro, IEEE*, 33(3):16–27, 2013.
- [10] Raghunathan et al. Cherry-picking: exploiting process variations in dark-silicon homogeneous chip multi-processors. In *Proceedings of the Conference on Design, Automation and Test in Europe*, pages 39–44, 2013.
- [11] N. Goulding-Hotta et al. The greendroid mobile application processor: An architecture for silicon's dark future. *Micro, IEEE*, 31(2):86–95, 2011.
- [12] V. Gupta et al. Impact: Imprecise adders for low-power approximate computing. In *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED)*, pages 409–414, 2011.
- [13] J. Han and M. Orshansky. Approximate computing: An emerging paradigm for energy-efficient design. In *Proceedings of the 18th IEEE European Test Symposium (ETS)*, 2013.
- [14] V. Hanumaiah et al. Performance optimal online dvfs and task migration techniques for thermally constrained multi-core processors. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 30(11):1677–1690, 2011.
- [15] V. Hanumaiah and S. Vrudhula. Energy-efficient operation of multi-core processors by dvfs, task migration and active cooling. *IEEE Transactions on Computers*, 99(1):1, 2012.
- [16] N. Hardavellas et al. Toward dark silicon in servers. *Micro, IEEE*, 31(4):6–15, 2011.
- [17] J. Henkel et al. Reliable on-chip systems in the nano-era: Lessons learnt and future trends. In *Proceedings of the 50th Annual Design Automation Conference*, pages 99:1–99:10, 2013.
- [18] U.R. Karpuzcu et al. The bubblewrap many-core: popping cores for sequential acceleration. In *42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 447–458, 2009.
- [19] U.R. Karpuzcu et al. Energysmart: Toward energy-efficient manycores for near-threshold computing. In *19th IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2013.
- [20] F. Kriebel et al. Aser: Adaptive soft error resilience for reliability-heterogeneous processors in the dark silicon era. In *DAC*, 2014.
- [21] P. Kulkarni et al. Trading accuracy for power with an underdesigned multiplier architecture. In *Proceedings of the 24th International Conference on VLSI Design (VLSI Design)*, pages 346–351, 2011.
- [22] M. Lyons et al. The accelerator store: A shared memory framework for accelerator-based systems. *ACM Trans. Archit. Code Optim.*, 8(4):48:1–48:22, 2012.
- [23] D. Markovic et al. Ultralow-power design in near-threshold region. *Proceedings of the IEEE*, 98(2):237–252, 2010.
- [24] D. Mohapatra et al. Design of voltage-scalable meta-functions for approximate computing. In *Proceedings of the Conference on Design, Automation Test in Europe Conference Exhibition (DATE)*, 2011.
- [25] T. Muthukaruppan et al. Hierarchical power management for asymmetric multi-core in dark silicon era. In *Proceedings of the 50th Annual Design Automation Conference (DAC)*, pages 174:1–174:9, 2013.
- [26] V. Narayanan and Y. Xie. Reliability concerns in embedded system designs. *Computer*, 39(1):118–120, 2006.
- [27] S. Nussbaum. Amd trinity apu. *HotChips '12*, 2012.
- [28] I. Paul et al. Cooperative boosting: needy versus greedy power management. *SIGARCH Computer Architecture News*, 41(3):285–296, 2013.
- [29] A. Raghavan et al. Computational sprinting. In *Proceedings of the 2012 IEEE 18th International Symposium on High-Performance Computer Architecture (HPCA)*, pages 1–12, 2012.
- [30] E. Rotem et al. Power-management architecture of the intel microarchitecture code-named sandy bridge. *IEEE Micro*, 32(2):20–27, 2012.
- [31] K. Swaminathan et al. Steep-slope devices: From dark to dim silicon. *Micro, IEEE*, 33(5):50–59, Sept 2013.
- [32] M. Taylor. Is dark silicon useful?: harnessing the four horsemen of the coming dark silicon apocalypse. In *Proceedings of the 49th ACM Annual Design Automation Conference (DAC)*, pages 1131–1136, 2012.
- [33] Y. Turakhia et al. Hades: Architectural synthesis for heterogeneous dark silicon chip multi-processors. In *Proceedings of the 50th ACM Design Automation Conference (DAC)*, 2013.
- [34] G. Venkatesh et al. Conservation cores: reducing the energy of mature computations. In *Proceedings of the 15th Symposium on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 205–218, 2010.
- [35] G. Venkatesh et al. Qscores: trading dark silicon for scalable energy efficiency with quasi-specific cores. In *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, 2011.
- [36] A.K. Verma et al. Variable latency speculative addition: A new paradigm for arithmetic circuit design. In *Proceedings of the Design, Automation and Test in Europe Conference*, pages 1250–1255, 2008.
- [37] L. Wang and K. Skadron. Implications of the power wall: Dim cores and reconfigurable logic. *IEEE Micro*, 33(5):40–48, 2013.