Towards Performance and Reliability-Efficient Computing in the Dark Silicon Era

Jörg Henkel, Santiago Pagani, Heba Khdr, Florian Kriebel, Semeen Rehman, Muhammad Shafique
Chair for Embedded Systems (CES), Karlsruhe Institute of Technology (KIT), Germany
Corresponding Authors: {henkel, muhammad.shafique}@kit.edu

***Abstract*—This paper discusses the power density and temperature induced issues in modern on-chip systems due to the high integration density and roadblock on the voltage scaling. First, the emerging dark silicon problem is discussed, and the corresponding critical research challenges in future chips are enumerated. Afterwards, we present an overview of some key research efforts and concepts that leverage dark silicon for performance and reliability optimization of on-chip systems under power or temperature constraints. The summarized works account for heat transfer inside a chip, as well as the varying performance and power trade-offs of gray silicon, that is, the potential benefits of operating at lower-than-nominal voltage and frequency levels. Besides realizing reliability-heterogeneous architectures, reliability of an on-chip system is enhanced by exploiting dark silicon for aging deceleration and resilience-driven resource management to mitigate soft-errors. Several of the tools discussed in this paper are available for download at http://ces.itec.kit.edu/download.**

I. Introduction

High transistor integration density and roadblock on the voltage scaling in the next-generation technology nodes result in increasing power densities and temperatures on a chip [1], [2], as seen in Figure 1(a) and Figure 1(b). A direct consequence of increasing local power densities and temperatures is the emerging dark silicon problem [1], [2], [3], [4], that is, using the same cooling technology as in previous scaling generations, all parts of a chip cannot be simultaneously active at nominal operating conditions for a given Thermal Design Power (TDP)[1]. These effects can drastically slow down the current trends in performance gains between generations, as well as aggravate several reliability threats (like aging and soft errors) [6], [7], which in turn challenges the viability of further cost-effective technology scaling.

The work in [1] presents a detailed survey of the state-of-the-art techniques involving dark silicon. Generally, most of the state-of-the-art techniques use TDP as a per-chip power budget constraint in order to model and predict the future percentages of dark silicon or to propose optimizations. For example, when operating at nominal conditions, that is, voltage and frequencies, the work in [3] predicts that at 8 nm up to 80% of the chip area will have to remain dark in order to satisfy the TDP power budget. However, as stated above, dark silicon is not a power consumption issue, rather a temperature-related issue caused by high local power densities that cannot be effectively cooled down (e.g., in a cost efficient way). Therefore, the work in [8] predicts new trends in dark silicon which are much lower than the original estimates from [3], as shown in Figure 1(c).

In this paper, we highlight the key research challenges on future dark silicon chips. Furthermore, we present an overview of our research efforts for performance and reliability optimization of on-chip systems under power or temperature constraints. In our works, we account for the heat dissipation characteristics inside a chip, as well as varying performance and power trade-offs of gray silicon, that is, the potential benefits of operating at lower-than-nominal voltage and frequency levels by taking advantage of the convex relation between voltage/frequency and the resulting power consumption. Several of the tools discussed in this paper are available for download at http://ces.itec.kit.edu/download.

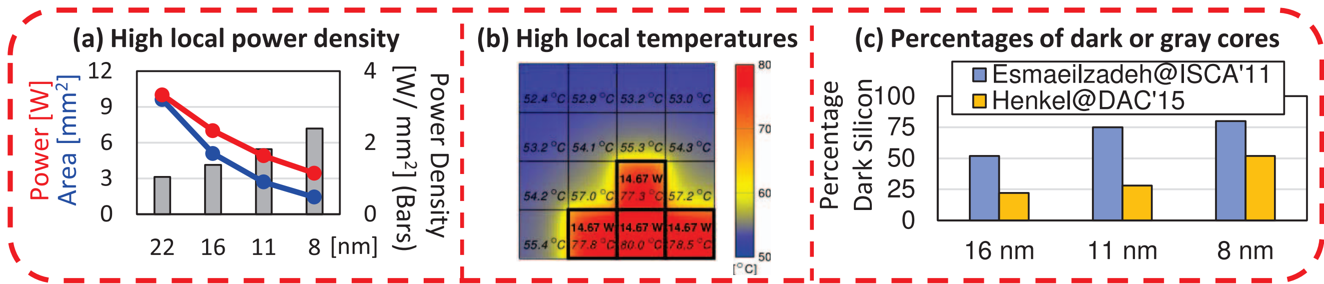
A. Challenges in the Dark Silicon Era

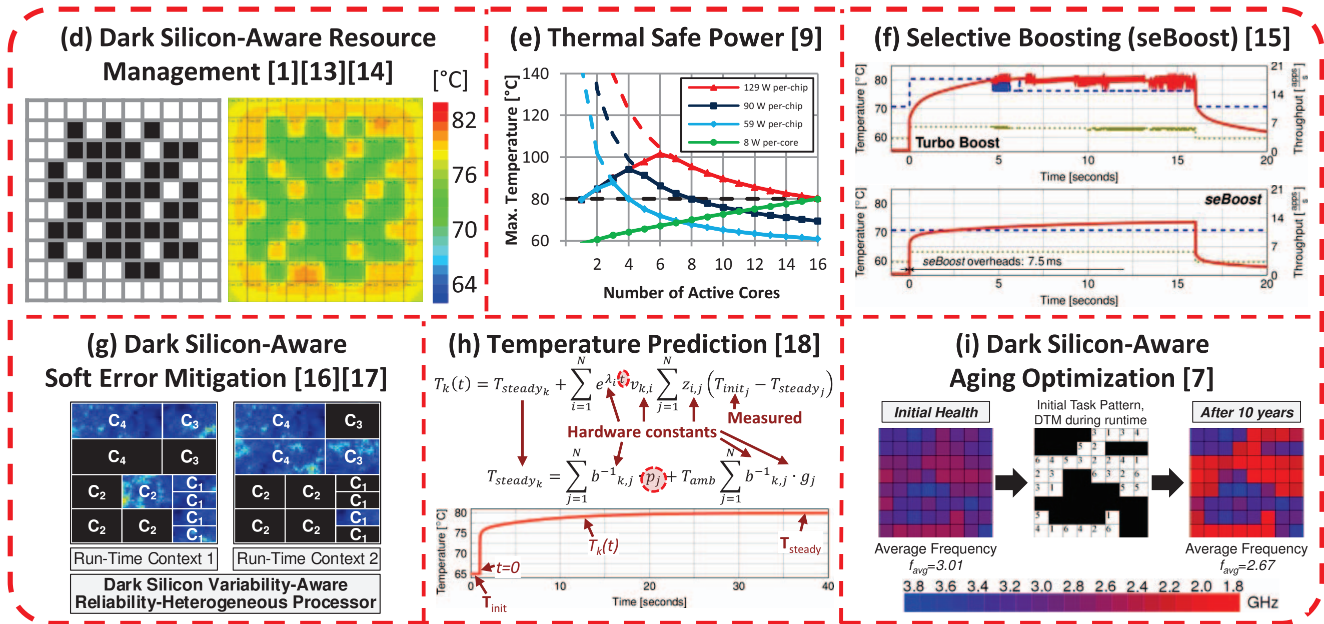
Developing and managing dark silicon manycore processors that operate under power density and thermal constraints involves several challenges, both in terms of performance and reliability optimizations. Some of the key challenges are highlighted as follows:

- **Roadblock on the Performance Improvements:** In future dark silicon chips, the usage of on-chip resources needs to be power- and temperature-aware. One promising approach for exploiting dark transistors is to develop heterogeneous computing fabrics, which can best accommodate different types of applications. Nevertheless, determining the number of active, dark, and gray cores should consider the diversity of performance and power, as well as the Thread Level Parallelism (TLP) and Instruction Level Parallelism (ILP) nature of every application. Furthermore, the thread-to-core mapping which determines the location of active, dark, and gray cores, should focus on alleviating local power density and thermal hot spots, which will further assist in boosting the performance.
- **Pessimism in Power Budgeting:** The recent studies in [1], [9] have effectively shown that, when the objective is to execute under a thermal constraint, using TDP as a per-chip or per-core power budget as an abstraction can be pessimistic or thermally unsafe, depending on the runtime scenario and the value of TDP, as seen in Figure 1(e). Therefore, in order to accurately define dark silicon constraints, it is crucial to investigate novel thermal-aware power budgeting techniques.
- **Lightweight Transient Temperature Prediction:** Runtime resource management decisions (such as voltage/frequency scaling, task migration, or core activations/deactivations) change the power consumption throughout the chip, which can in turn result in transient temperatures much higher than any steady-state scenarios. Therefore, it is crucial to

[1]As defined in [5], TDP "is the highest expected sustainable power while running known power intensive real applications". Thus, it should be a safe power level to run the system, and therefore manufacturers recommend to design the cooling solution to dissipate TDP. In this way, without incurring in unnecessary high cooling costs, running the system at this power level should not cause thermal problems. However, TDP is not the maximum achievable power.

978-3-9815370-6-2/DATE16/©2016 EDAA





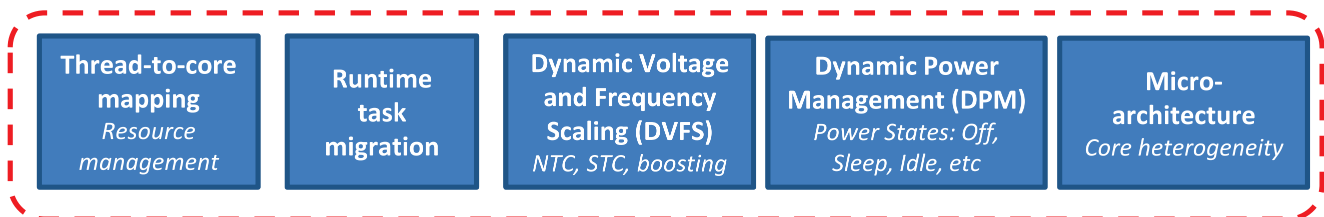


Fig. 1: Overview of the Dark Silicon Problem and Our Proposed Solutions.

have methods that allow for lightweight run-time prediction of transient temperature peaks before making runtime decisions.

- **Exploiting Dark Silicon by using Boosting Techniques:** By executing at different voltage and frequency levels, there may exist several gray levels (including Near Threshold Computing [8], [10]) and several boosting levels of operation. Furthermore, executing some cores at high voltage and frequency levels allows the system to exploit the existing thermal headroom on the chip in order to satisfy runtime requirements surges. However, choosing the voltage and frequency levels of boosted and non-boosted cores is a very challenging process, particularly due to the non-linear transient thermal behavior on the chip and the heat transfer among cores, and also because this will have a direct impact on the number of dark, gray, and fully lit cores, as shown in Figure 2.
- **Reliability Considerations:** High temperatures aggravate several reliability threats. Therefore, as dark silicon is caused by high power densities and on-chip temperatures, it poses several challenges in terms of reliability. For example, the abundance of transistors can potentially be leveraged

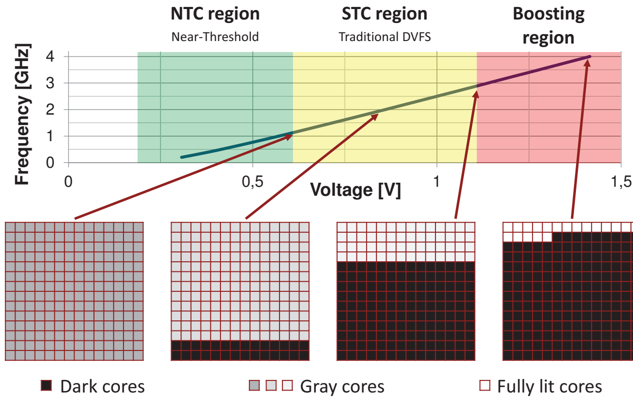


Fig. 2: Trade-offs between the number of dark, gray, and fully lit cores, and the voltage/frequency settings: Near Threshold Computing (NTC), Super-Threshold Voltage Computing (STC), and boosting operation.

to design reliability-heterogeneous processors that support diverse hardening levels which provide trade-offs between reliability and power consumption. Similarly, dark silicon can be leveraged to decelerate aging or improve the aging profile of a chip, that is, balance aging of different cores.

In order to address the challenges above, we presented several techniques for the design and run-time management of dark silicon chips under power and thermal constraints, as briefly outlined in the rest of this paper and conceptually shown in Figure 1.

B. *Overview of Dark Silicon and Our Proposed Solutions*

Figure 1 shows an overview of dark silicon, our proposed design and management techniques/solutions, and some underlying basic hardware and software methods that are commonly present in standard chip designs and are exploited by our techniques. Particularly, the new constraints and effects imposed by dark silicon are depicted in Figures 1(a), 1(b), and 1(c). Figure 1(a) shows the power consumption and area trends of different technology scaling generations, and the corresponding power density, which is shown to be increasing. Directly related, Figure 1(b) (from [9]) shows temperature simulations conducted in HotSpot [11] for a system with 16 cores (Alpha 21264 cores in 45 nm, simulated with McPAT [12]), in which we can observe the resulting high local temperatures present on the chip when 4 cores consume 14.67 Watts each. Furthermore, related to the estimations of dark silicon on future chips, Figure 1(c) shows the estimated dark silicon percentages presented in [3] and [8] for operating at nominal conditions.

Our proposed techniques for mitigating and managing dark silicon are outlined in Figures 1(d)-1(i). Figure 1(d) shows how dark silicon can be exploited by resource management techniques (e.g. [13] and [14]) to improve the thermal profile of the chip. Figure 1(e) motivates the Thermal Safe Power (TSP) [9] power budgeting approach by showing HotSpot simulations in which we can observe the resulting maximum temperatures in a chip (with DTM deactivated) when equally distributing constant per-chip power budgets among different numbers of active cores. Assuming a temperature for triggering DTM of 80°C, for a given power budget and number of active cores, when the resulting maximum temperature is below (above) 80°C it implies that the given power budget is pessimistic (not thermally safe) for that specific scenario. Figure 1(f) shows transient temperature simulations conducted with HotSpot for using two different boosting techniques, particularly, a simple closed-loop control-based boosting technique and *seBoost* [15]. Figure 1(g) shows an architectural template of a reliability-heterogeneous processor with two run-time contexts. A reliability-heterogeneous processor exhibits several iso-ISA cores (i.e. cores having the same instruction set architecture) that feature distinct hardening features (e.g., either pipeline and cache is protected using hardware-/architecture-level support, or only cache, or triplicated pipeline with voter, or any combination of different hardware/architecture-level protection mechanisms for different parts of a core). Each core of a particular type exhibits distinct power, area, and reliability properties. This can be exploited to optimize the soft error resilience of a system considering a set of concurrently executing applications under a given power budget. The figure shows two run-time contexts, i.e. how different cores are selected for two different sets of concurrently executing applications [16] and [17]. Figure 1(h) presents some mathematical details of the MatEx [18] transient temperature prediction tool, specifically, how the transient temperature of node k at future time t can be easily computed by knowing some hardware constants, the temperatures measured throughout the chip, and the expected steady-state temperatures if the power on the chip would remain constant. The expected steady-state temperatures can be computed by knowing the ambient temperature and the power consumption throughout the chip. Finally, Figure 1(i) shows that how a sophisticated selection of dark and powered-on cores can be exploited to improve the health map of a chip, i.e. aging state of different cores [7].

II. Optimizations to Improve Performance

A. *Dark Silicon-Aware Resource Management*

The performance on dark silicon chips is limited by the number of active and dark cores on the chip. However, these numbers significantly vary according to the workload executing on the chip, because different applications exhibit different power profiles. Therefore, resource management in dark silicon chips needs to consider power and performance profiles of the applications while determining the number of active and dark cores, in order to obtain the maximum possible performance out of the available resources. In addition to the number of active and dark cores, voltage and frequency (v/f) levels of the active cores play a significant role in determining the amount of dark cores on the chip and thereby the resulting performance.

As a result, in [13], we proposed a dark silicon-aware resource management technique, called *DsRem*, that distributes the chip's resources; cores and power; among different applications under a thermal constraint. In particular, it determines the number of active cores and their v/f level for each application, in order to maximize the overall system performance. Moreover, *DsRem* adapts the number of active and dark cores considering the TLP and the ILP of the applications, while maximizing the performance (see Figure 3). To increase their performance, high TLP applications require more active cores and lower v/f levels in comparison with high ILP applications that require less active cores and higher v/f

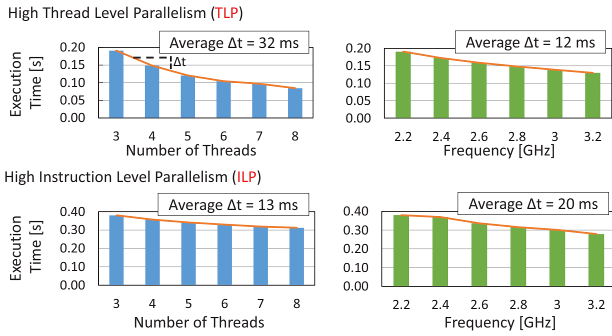
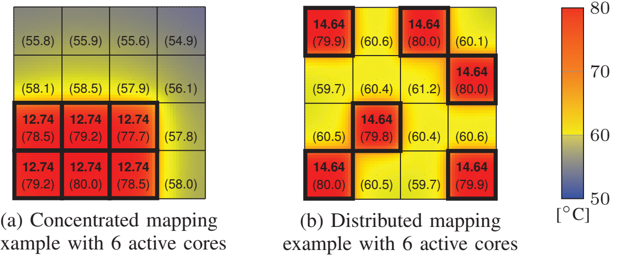


Fig. 3: The impact of TLP and ILP on the application performance.



(a) Concentrated mapping example with 6 active cores

(b) Distributed mapping example with 6 active cores

Fig. 4: Example of TSP for two different mappings for a maximum temperature of 80°C (figure from [9]). Top numbers are the power consumptions (in Watts) of each active core (boxed in black). Bottom numbers in parenthesis are the temperatures in the center of each core (in °C). Detailed temperatures are shown according to the color bar.

levels. Our evaluation shows that *DsRem* can outperform the state-of-the-art techniques by an average of 35%.

Our technique *DaSiM* [14] enables variability-aware dark silicon management. It exploits the so-called *dark silicon patterning*, along with the core-to-core leakage power variations, to optimize the system performance under peak temperature constraints. The dark silicon patterns denote various spatio-temporal decisions for the cores' power state (i.e. when to make a core dark and at which location). This directly influences the resulting chip thermal profile due to improved heat dissipation, and thereby enables (1) powering-on more cores to facilitate high-TLP applications, and/or (2) boosting certain cores to facilitate high-ILP applications. An improved thermal profile also reduces the number of Dynamic Thermal Management (DTM) events that, in turn, further improves the system performance. In order to enable run-time optimizations, *DaSim* employs a light-weight online temperature prediction mechanism. It estimates the chip thermal profile for a given candidate solution. This way, the thermal impact of different thread mappings is investigated. Our evaluations illustrate that *DaSim* improves the system performance (i.e. collective performance improvement for a set of concurrently executing applications) by up to $1.8x$ when compared to state-of-the-art dark-silicon unaware mapping techniques.

B. Thermal Safe Power (TSP)

In order to have an abstraction from thermal issues, generally, system designers use TDP or some other value as a per-chip or per-core power budget. This idea is motivated by the notion that running the chip under such a power budget will be thermally safe, that is, DTM will not be triggered. However, temperature is directly related to local power densities, and therefore using a *single* and *constant* value as per-chip or per-core power budget for manycore systems, e.g., TDP, is either pessimistic or not thermally safe, as seen in Figure 1(e). A major step towards dealing with the dark silicon problem [1], [2] is thus through *efficient* power budgeting techniques.

Therefore, we proposed a new power budget concept, called Thermal Safe Power (TSP) [9], which is a thermal-aware abstraction that provides safe power constraints as a function of the number of active cores. Running cores at power levels below TSP results in a higher total system performance than standard power budgeting solutions, while the maximum temperature among all cores remains below the threshold level that triggers DTM. For a specific floorplan and ambient temperature, TSP can be computed to obtain safe power constraints for the worst-case mappings, that is, concentrated mappings that promote hot spots, as shown in the example of Figure 4(a), thus allowing the system designers to abstract from mapping decisions. Moreover, TSP can also be computed at run-time for a particular mapping of active cores and ambient temperature, allowing the system to optimize the power budgets for dispersed core mappings, for example, a mapping like the one shown in Figure 4(b). When the number of active cores grows, the TSP values decrease, which in turn means executing cores at lower voltage and frequency levels. Namely, TSP provides a simple relation between the number of active cores and their maximum allowed voltage and frequency settings, which directly shows the number of dark/gray cores in a chip for different scenarios.

C. Selective Boosting (seBoost)

Boosting techniques have been widely adopted in commercial multicore and manycore systems, mainly because they provide means to satisfy performance requirements surges, for one or more cores, at run-time. Current boosting techniques select the boosting levels (for boosted cores) and the throttle-down levels (for non-boosted cores) either arbitrarily or through step-wise control approaches. Although simple to implement and potentially efficient, these methods might result in unnecessary performance losses for the non-boosted cores, in short boosting intervals, in failing to satisfy the required performance surges, or in unnecessary high power and energy consumption. Therefore, we presented an efficient and lightweight run-time boosting technique based on transient temperature estimation [18], called *seBoost* [15]. This technique guarantees meeting run-time performance requirements surges, by executing the boosted cores at the required frequencies for the entire boosting intervals, while throttling down the non-boosted cores. In order to minimize the performance losses for the non-boosted cores, the throttling down levels are chosen such that the maximum frequencies among all cores reaches the critical temperature precisely when the boosting is expected to expire. Our experiments show that *seBoost* can in fact guarantee the required performance surges, while the state-of-the-art control techniques fail to constantly satisfy the run-time requirements.

D. Transient Temperature Prediction (MatEx)

Typically, run-time scheduling decisions, such as voltage/frequency scaling, task migration, core activations/deactivations, etc., are used to optimize the resource usages on a chip. These run-time decisions change the power consumption throughout the chip, which can in turn result in transient temperatures much higher than any steady-state scenarios. In order to be thermally safe, it is important to evaluate the transient peaks before making resource management decisions. Hence, we developed a method for computing these transient peaks in just a few milliseconds. Our technique, called MatEx [18], works for any compact thermal model consisting in a system of first-order differential equations, like the RC thermal networks used in HotSpot [11]. Instead of using regular numerical methods, MatEx is based on an *analytical* solution of the differential equations using matrix exponentials and linear algebra. This results in a mathematical expression which can easily be analyzed and differentiated to compute the maximum transient temperatures. Furthermore, MatEx can also be used to efficiently compute any future transient temperatures without accuracy losses or the need to compute all step-wise partial temperatures, as is the case in standard numerical methods.

III. Optimizations to Improve Reliability

In addition to the performance and thermal aspects of the technology scaling, reliability becomes an important design concern in the dark silicon era. In the following it will be shown how dark silicon can be leveraged to address three main reliability threats, namely soft errors [19] (i.e., transient faults in the hardware caused by high energy particles manifesting as bit flips), manufacturing process variations [20], [21] (i.e., different parameters of identical transistors on the same chip and across different chips due to manufacturing difficulties for smaller transistors) and NBTI-induced aging [22] (i.e., a slow-down of the circuit due to a threshold voltage shift caused by stressing the PMOS transistors).

A. Dark Silicon-Aware Soft Error Mitigation

In order to address the soft error and process variation problems in the era of dark silicon, two main challenges have to be met: (1) How an architecture/a chip should be designed considering area and dark silicon constraints, and a set of representative benchmark applications while accounting for process variation and reliability enhancements. (2) How such an architecture can be managed at run time in order to improve the system's reliability.

The design-time challenges are addressed in [16], [17] by developing a library of core types with different area, power and reliability properties, which is used to customize a chip by selecting a set of iso-ISA reliability-heterogeneous cores. That means that the chip can also be over-provisioned with cores of the same type to account for process variation problems, as even cores of the same type can have different power characteristics. Such architectures are referred to as *variability-aware, reliability-heterogeneous processors*. At run time the characteristics of each individual chip as well as the application characteristics are considered. As shown in [17] different applications exhibit diverse vulnerabilities and error masking properties, so that not all of them need the same level of reliability. Moreover, the set of concurrently executing applications may vary. While taking into account the dark silicon constraint, applications are mapped to a core type that satisfies their performance and reliability requirements. During this mapping, the different power characteristics of cores due to process variation are also accounted for, i.e., the cores with the lowest power consumption of a specific type are selected first.

The results of [16], [17] demonstrate that dark silicon, process variations and reliability are important design aspects of future systems. A key aspect is to leverage the distinct reliability and performance characteristics of different applications in reliability-heterogeneous manycore processors. Compared to state-of-the-art [17] achieves 58%-96% improvements of the system reliability. Furthermore, [16] shows that ignoring the process variation across different chips and identical core types on the same chip leads to a violation of the dark silicon constraint (up to 60% of the chips). However, conservative solutions such as guardbanding result in a considerably higher vulnerability as they consider safe margins which leads to the selection of less protected cores.

B. Dark Silicon-Aware Aging Optimization

In [7], we developed a technique called *Hayat* that harnesses the available dark silicon and manufacturing process induced core-to-core variability in frequency and leakage power to decelerate and/or balance the chip-wide temperature-dependent aging effects. It accounts for the thermal profile and core usage when taking temperature-dependent aging optimization decisions. Core usage is important for early-stage aging (i.e. within the first two years of core usage), while temperature effects are relatively more dominating in aggravating the aging rates during the later years (i.e. aging beyond the 2nd year of operation). Hayat also considers the leakage power variations to optimize the health map of a chip (i.e. aged state of different cores) when executing many applications concurrently.

Our run-time aging management system selects a subset of cores for a set of executing applications such that their performance requirements are satisfied and the overall chip aging rate is minimized. In order to proactively achieve this, *Hayat* performs two key operations: (1) determine a Dark Silicon Pattern such that chip aging is decelerated considering the lateral heat dissipation effects and core usage; and (2) determine *variation-aware thread-to-core mapping* such that a balanced aging profile is achieved while meeting application's throughput constraints. To do so, it schedules high temperature-generating threads to the so-called *healthy cores* (i.e. less-aged and faster cores) and vice versa. *Hayat* continuously optimizes the dark silicon patterns to decelerate the aging rates of faster-aging cores by leveraging frequency/leakage power variations and workload properties, while respecting thermal constraints (i.e., $T_{peak} < T_{safe}$).

In order to realize proactive aging optimization, *Hayat* requires **an online but lightweight chip's health map estimation technique**. This is based on offline generated $3D$-aging tables, obtained through processor aging estimation techniques of [dTUNE], which provide health degradation for a core at a given temperature and usage profile (i.e. duty cycle). The health map prediction is done in a two step process. (1) First,

for a given candidate solution, thermal profile of the chip and corresponding duty cycles for given workloads are predicted. (2) The predicted temperature and duty cycles are used to make a lookup in the $3D$-aging tables.

In the optimization step, the health degradation map for each candidate solution is predicted, and the one that minimizes the accumulated health degradation of all cores is selected. Since chip aging is a long-term phenomenon (i.e., over years), we also devised the notion of *"aging epoch"* to account for different runtime effects, and to facilitate accelerated aging evaluation. This considers task migration events, workload variations experienced by different cores over time, and changing thermal variations and stress levels on different cores in precise thermal simulations. However, coarse-grained aging epochs determine the granularity of health monitoring and aging evaluation, and in turn exploit multiple thermal simulation cycles. Our evaluations for numerous process variation maps illustrate that *Hayat* not only decelerates the chip by 6 months – 5 years (depending upon the required lifetime constraint) compared to state-of-the-art techniques, but also achieves a balanced aging profile.

Furthermore, our work in [23] illustrates how sub-μs voltage switches may lead to transient errors due to aging, highlighting for the first time, the short-term effects of aging. To prevent the occurrence of transient errors, we propose to design on-chip systems with A-GEAR, that is, adaptive guardbands protecting against the long-term *and* short-term effects of aging at minimal overhead.

IV. Conclusions

In this paper we have discussed some aspects of the emerging dark silicon problem and the corresponding critical research challenges on future chips. Furthermore, we presented an overview of some of our key research efforts and concepts that leverage dark silicon for performance and reliability optimization of on-chip systems under power or temperature constraints. We show how the dark silicon problem can be turned into an opportunity for performance and reliability optimization.

Acknowledgements

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 as part of the priority program *Dependable Embedded Systems* (SPP 1500 – http://spp1500.itec.kit.edu).

References

[1] 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)*, 2014, pp. 185:1–185:6.

[2] M. Shafique, S. Garg, T. Mitra, S. Parameswaran, and J. Henkel, "Dark silicon as a challenge for hardware/software co-design: Invited special session paper," in *International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, 2014, pp. 13:1–13:10.

[3] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger, "Dark silicon and the end of multicore scaling," in *the 38th International Symposium on Computer Architecture (ISCA)*, 2011, pp. 365–376.

[4] R. Perricone, X. S. Hu, J. Nahas, and M. Niemier, "Can beyond-cmos devices illuminate dark silicon?" in *Proceedings of the 19th IEEE/ACM Design, Automation and Test in Europe (DATE)*, March 2016.

[5] Intel Corporation, "Dual-core intel xeon processor 5100 series datasheet, revision 003," August 2007.

[6] J. Henkel, L. Bauer, N. Dutt, P. Gupta, S. R. Nassif, M. Shafique, M. B. Tahoori, and N. Wehn, "Reliable on-chip systems in the nano-era: lessons learnt and future trends," in *DAC*, 2013.

[7] D. Gnad, M. Shafique, F. Kriebel, D. Sun, and J. Henkel, "Hayat: Harnessing dark silicon and variability for aging deceleration and balancing," in *52nd Design Automation Conference (DAC)*, 2015.

[8] J. Henkel, H. Khdr, S. Pagani, and M. Shafique, "New trends in dark silicon," in *Proceedings of the 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, June 2015, pp. 119:1–119:6.

[9] S. Pagani, H. Khdr, W. Munawar, J.-J. Chen, M. Shafique, M. Li, and J. Henkel, "TSP: Thermal safe power - efficient power budgeting for many-core systems in dark silicon," in *the 9th IEEE/ACM International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, Oct 2014, pp. 10:1–10:10.

[10] A. Pahlevan, J. Picorel, A. P. Zarandi, D. Rossi, M. Zapater, A. Bartolini, P. G. Del Valle, D. Atienza, L. Benini, and B. Falsafi, "Towards near-threshold server processors," in *Proceedings of the 19th IEEE/ACM Design, Automation and Test in Europe (DATE)*, March 2016.

[11] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. R. Stan, "Hotspot: A compact thermal modeling methodology for early-stage vlsi design," *IEEE Transactions on VLSI Systems*, vol. 14, no. 5, pp. 501–513, May 2006.

[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 IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2009, pp. 469–480.

[13] H. Khdr, S. Pagani, M. Shafique, and J. Henkel, "Thermal constrained resource management for mixed ILP-TLP workloads in dark silicon chips," in *Proceedings of the 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, June 2015, pp. 179:1–179:6.

[14] M. Shafique, D. Gnad, S. Garg, and J. Henkel, "Variability-aware dark silicon management in on-chip many-core systems," in *Design, Automation and Test in Europe (DATE)*, 2015.

[15] S. Pagani, M. Shafique, H. Khdr, J.-J. Chen, and J. Henkel, "seBoost: Selective boosting for heterogeneous manycores," in *the 10th IEEE/ACM International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, Oct 2015, pp. 104–113.

[16] F. Kriebel, S. Rehman, M. Shafique, S. Garg, and J. Henkel, "Variability- and reliability-awareness in the age of dark silicon," *IEEE Design and Test*, 2015.

[17] F. Kriebel, S. Rehman, D. Sun, M. Shafique, and J. Henkel, "ASER: Adaptive soft error resilience for reliability-heterogeneous processors in the dark silicon era," in *51st Design Automation Conference (DAC)*, 2014.

[18] S. Pagani, J.-J. Chen, M. Shafique, and J. Henkel, "MatEx: Efficient transient and peak temperature computation for compact thermal models," in *Proceedings of the 18th Design, Automation and Test in Europe (DATE)*, March 2015, pp. 1515–1520.

[19] R. Baumann, "Soft errors in advanced computer systems," *Design and Test of Computers, IEEE*, vol. 22, no. 3, pp. 258–266, May 2005.

[20] Y. Turakhia, B. Raghunathan, S. Garg, and D. Marculescu, "Hades: architectural synthesis for heterogeneous dark silicon chip multi-processors," in *The 50th Annual Design Automation Conference 2013, DAC*, 2013, pp. 173:1–173:7.

[21] B. Raghunathan, Y. Turakhia, S. Garg, and D. Marculescu, "Cherry-picking: exploiting process variations in dark-silicon homogeneous chip multi-processors," in *Design, Automation and Test in Europe, DATE*, 2013, pp. 39–44.

[22] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *Journal of Applied Physics*, vol. 94, no. 1, pp. 1–18, 2003.

[23] V. M. van Santen, H. Amrouch, N. Parihar, S. Mahapatra, and J. Henkel, "Aging-aware voltage scaling," in *Proceedings of the 19th IEEE/ACM Design, Automation and Test in Europe (DATE)*, March 2016.