Dark Silicon Considered Harmful: A Case for Truly Green Computing

Erik Brunvand
School of Computing
University of Utah
elb@cs.utah.edu

Donald Kline, Jr
Electrical and Comp. Engineering
University of Pittsburgh
dek61@pitt.edu

Alex K. Jones
Electrical and Comp. Engineering
University of Pittsburgh
akjones@pitt.edu

***Abstract*—As individuals and researchers approach the challenge of green computing it is natural to consider the energy consumption of computational devices and their supporting systems during their use phase (i.e., after they are deployed into service). However, for computing to be truly green, all phases of the system life-cycle, from manufacturing to disposal, must be considered. In particular there is limited awareness to the considerable fraction of the total life-cycle environmental impacts of computing systems that result from the fabrication of the integrated circuits (ICs) that are used in those devices. Ironically, the trend toward dark silicon accelerators, often targeted at improving operational energy efficiency, may be counterproductive for holistic energy reduction of computing systems. The increased chip area that results from a large percentage of dark silicon may exacerbate the fabrication impacts to the point that overall sustainability is actually decreased. In this paper, we explore some properties of manufacturing and operational energy efficiency and make a case that truly green computing must carefully consider the tradeoffs involved.**

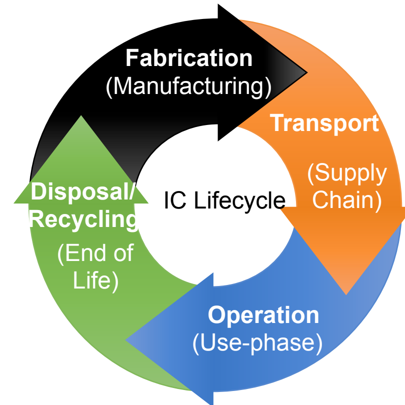


Figure 1. Life-cycle to consider for holistic integrated circuit (IC) sustainability costs related to green computing.

I. Introduction

Taken as a whole, computing in all its forms including personal and mobile electronics, desktop computing, and cloud-based compute services represent an ever growing portion of our overall energy and resource use. As researchers approach this challenge it is natural to consider the energy consumption of computational devices and their supporting systems during their *use phase* (*i.e.,* after they are deployed into service). This includes reducing energy consumption in processors, memory systems, peripheral devices, cooling systems and a host of other components that are used in deployed systems. However, for truly green computing, all phases of the system life-cycle, illustrated in Figure 1, must be considered. In particular there is limited awareness to the considerable fraction of the sustainability impacts due to the fabrication of integrated circuits (ICs) used in computing devices. Studies have shown that the energy and environmental costs of IC manufacture can actually significantly outpace their use-phase costs [1, 2]. Moreover, IC manufacturing has been demonstrated to have a significant and rising contribution to environmental impacts of more overtly impactful domains such as building construction [3].

Due to challenges of scaling CMOS technologies to sub 10nm feature sizes, new fabrication solutions have been employed such as multi-patterning lithography, exotic transistor shapes (*e.g.*, Fin and GAA (gate-all-around) FETs), and increasingly diverse materials (*e.g.,* III-V gate channels, low-κ dielectrics, high conductivity metals, etc.). As a result, fabrication effort is accelerating much more rapidly than in the past, outstripping the density improvements provided by the next technology node [4].

These trends in IC fabrication, combined with the use of more transistors for increasing core counts, for larger memories and storage, or for so-called "dark silicon" system structures *may actually be harmful to overall sustainability* at deeply scaled geometries. The high-level explanation is that to increase holistic sustainability, additional time and energy spent in fabrication due to a larger die or for a more exotic manufacturing process must amortized via decreased use-phase energy over the lifetime of the chip.

For infrequently used segments of the chip, entirely leaving out the dark silicon to be handled in some other more general purpose fashion may only come at a nominal increase in use-phase energy. However, the reduced chip area that results from saving a large percentage of dark silicon may significantly decrease the fabrication impacts to the point that overall sustainability is actually improved. As it is imperative to identify methods for environmentally responsible usage of extremely scaled technologies for next generation computing, in this paper we examine some implications of manufacturing and operational energy efficiency and make a case that truly

978-1-5386-7466-6/18 $31.00 © 2018 IEEE

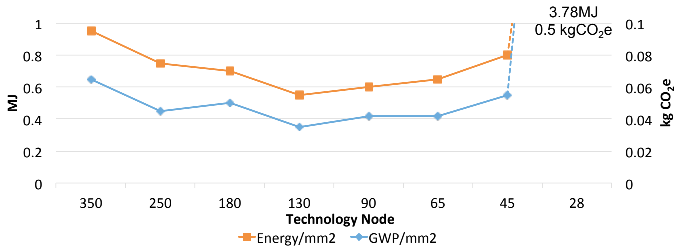


Figure 2. IC fabrication energy and global warming potential of CMOS fabrication at different process nodes [2, 11, 12].

green computing must carefully consider these tradeoffs.

II. Background and Related Work

It is important to consider carefully what is meant by "green" computing. It is common to confuse the idea of green computing with conservation, recycling, or energy efficiency. These are all important considerations for responsible product development, but to evaluate the greenness of a product, the total environmental impact from "cradle to grave" must be considered. This can include quantifying everything from acquiring raw materials for manufacturing, to the product disposal method after its use.

The study of overall sustainability is typically called *Life Cycle Assessment* (LCA) for which many standards exist [5–8]. An LCA framework is the necessary underpinning to truly understand and engineer how processes and products use materials, water, and energy resources throughout their lifetimes. LCA provides a comprehensive and quantitative analysis of the environmental impacts of a product or process throughout its entire life cycle.

One of the most difficult aspects of studying holistic green computing systems from an IC perspective is obtaining detailed environmental data from IC fabrication to support LCA analysis. A limited number of environmental studies have been reported of computing systems [9–11] and semiconductor fabrication [12, 13] for CMOS ranging from 0.35μm to 32nm, as well as DRAM and Flash technologies from similar generations [12]. Some studies report detailed impacts [14] and economic costs [15] per process step at particular technology nodes (*i.e.,* 130nm), while others provide aggregated data for an example or typical process at that node [12].

One analysis (shown in Figure 2) shows trends in fabrication energy and carbon emissions, the latter noted as global warming potential (GWP) measured in CO_2 equivalent (CO_2e) of Si-CMOS ICs. Another way to interpret these results is that during the period of Dennard Scaling, environmental impacts per chip area were also improving. At the end of Dennard Scaling (circa 2005) [16], the environmental impacts per chip area started to increase, although the impacts per transistor were still improving due to the density benefits. There is now evidence that the final benefit of scaling, reduced cost per transistor, was lost at the 28nm feature size (circa 2011) [4]. This is also seen as the spike in environmental impacts after 45nm [2] and follows from a tenant of LCA that economic cost trends tend to mirror environmental impact trends [17].

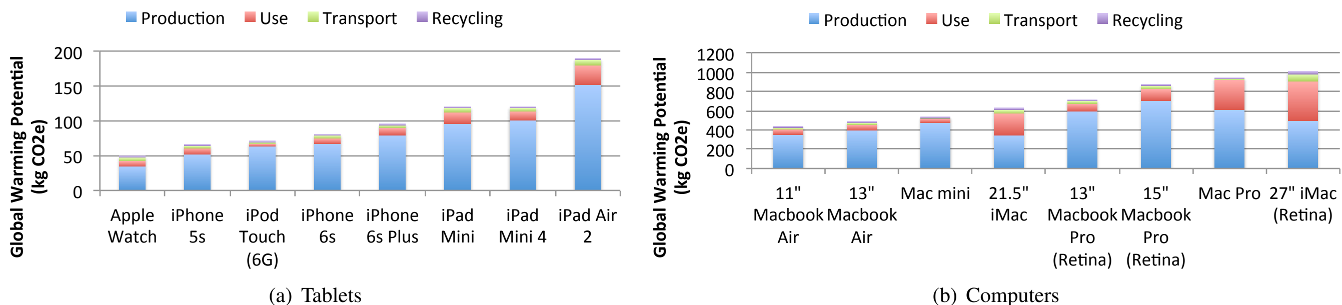
From a technology perspective, fabrication trends also support this. Current optical lithography, already one of the most expensive steps of fabrication, has a physical limit at around 30nm. To descend further, given that extreme ultraviolet (EUV) lithography has yet to become cost effective [18], multiple-patterning approaches are required such that at 22nm essentially double the number of lithography steps, and at $\leq$10nm potentially an order of magnitude more steps are required [19]. Other new process steps are required to make increasingly small transistors (e.g., FinFETs). This escalation of processing steps is not limited to transistors for example additional steps are required to achieve sufficiently low-κ dielectrics for increasingly thin metal interconnects [19]. Researchers have also studied the additional steps and environmental impacts incurred when augmenting CMOS with spin-transfer torque memory cells (STT-MRAM) and incorporated a model of those costs in the popular tool NVSim [20].

The trends of increasing fabrication energy cost could be somewhat mitigated if IC die area per system reduced with the descent in feature size. With shrinking transistors and increased device density, ISO architecture systems should decrease in overall die area as smaller process nodes are used. However, particularly from 130nm to 45nm, the opposite trend has been observed. Systems tend to have *more* IC area increasing from 750 to 1200 mm^2 between 2001 and 2010. This is largely due to trends of including more functionality as systems migrate to smaller technology nodes. These trends include adding more processor cores, embedded memory, accelerators such as graphics processing units, and solid state storage [9], the initial steps of the movement towards dark silicon. The increasing economic cost and dramatically increasing environmental impacts for sub 32/28nm nodes are alarming [2, 4]. To address this requires new strategies for holistic energy reduction in light of these emerging trends.

III. Why Green IC Fabrication?

Optimizing computing energy in the *use phase* of an IC is compelling because it helps to address thermal density issues of deeply scaled CMOS, maximizes battery-life of mobile computing platforms, while also addressing sustainability. However, our industry has been so successful at reducing use-phase energy that the primary sustainability issue is shifting to manufacturing [2] as shown in Figure 3 [11]. This is due to trends in fabrication techniques for increasingly small geometries, such as increasing photo-lithography and metrology costs and these trends show no signs of abating [4].

The substantial effort required to create devices at these small geometries also requires considerable adjustments to avoid losses in yield that would otherwise occur when reducing feature sizes [12]. As new technology nodes are introduced and tested, it is becoming increasingly difficult to achieve desired yields [21–23], which greatly increases the monetary cost and development effort of these technologies [16]. Unfortunately, the environmental impacts from manufacturing at these deeply



(a) Tablets

(b) Computers

Figure 3. Impact of manufacturing/production (dominated by ICs and displays) in "use-phase energy" optimized systems from Apple Computer. Use-phase impacts are calculated based on a two-year operational lifetime to match the expected market lifetime [11].

scaled nodes follows the same negative trends, requiring a dramatic amount of energy to be invested before the operational energy analysis of a chip in mass production can even be considered. Moreover, dark silicon accelerators and increased die size in general exacerbate this reduction in yield [24, 25], continuing to compound the increased manufacturing costs.

A more insidious trend is the vulnerability to runtime faults at these increasingly smaller geometries. From the increased vulnerability to multi-bit faults from radiation [26], to memory faults including reduced retention time [27, 28] and wordline crosstalk [29], continued scaling is exacerbating vulnerabilities to these faults. The current solution to these problems concurs with the dark silicon ideology, to devote increasingly large areas to error correcting techniques, ranging from traditional ECC [30], to fault maps of vulnerable cells [31–34], to various encoding tables for correction [35, 36] in addition to required auxiliary bits. Thus, desired real-estate for dark-silicon accelerators at these low technology nodes often must also plan to reserve space for required corresponding error correction, greatly increasing the required chip area and risking further reductions in yield.

To encapsulate the sustainability costs of IC fabrication we will co-opt a term from LCA on buildings: *embodied energy*. This is a term used to describe the energy consumed by all of the processes associated with the production of a building, from the mining and processing of natural resources to manufacturing, transport and product delivery. For the embodied energy of an IC we include the energy consumed by all the processes associated with the fabrication of the IC, from the refining and processing of materials used to the significant energy used in the fabrication process itself.

Given the rising energy impact of computing described in the introduction, and growing consumer awareness of sustainability and the impacts of energy use on our environment, similar to other green products from eco-friendly disposable cups to hybrid and electric automobiles, future consumers might make a choice of computing products based both on their performance and their sustainability. That is, an environmentally aware consumer may choose a product with slightly lower performance, but significantly higher sustainability. This could be thought of as the computing equivalent of the "Energy Star" use-phase energy ratings for consumer appliances [37].

The development of a holistic rating system like Energy Star for computing systems requires understanding the impacts involved in fabricating more sustainable ICs. Thus, models and tools that enable designer-in-the-loop optimization methods to improve sustainability under performance, power, and cost constraints are essential. Computer-aided design tools create useful abstractions of the fabrication process that bridge the gap between details of the process and the hardware description from the IC designer. A careful analysis and characterization of the sustainability impact of these choices will inform both the designer and the tools used in the design process. However, designers currently do not have a way to visualize at design time the effect their decisions have on the sustainability and environmental cost of IC fabrication. In the particular case of dark silicon, including additional transistors for functionality that is infrequently enabled has potentially significant, and negative implications to sustainability that are not reported by existing tools, and therefore are not communicated to the designer.

There is a significant investment in tools to evaluate sustainability in certain domains. For example, commercial tools like SimaPro [38] and GaBi [39] estimate life-cycle impacts of many processes, and while they do include categories that include semiconductors, their analysis is typically tuned for other domains such as the built environment. A more detailed parametric approach to studying semiconductors has been proposed for 130nm technology [14] and is the fundamental approach for gathering data about CMOS processes for which detailed process studies have been completed to 32nm [12]. However, a carefully parameterized model that scales to state of the art technology nodes (e.g., circa 14nm and beyond) is essential to evaluate holistic sustainability at modern technology levels. Further, using such a model, detailed green computing optimization techniques may apply the model at each of the layers of the design process from the system level, to the design level, and down to fundamental process level.

IV. Tools for Holistic Green Computing

While there is a considerable body of work to optimize and model use-phase energy consumption of modern computer architectures, the existing art in sustainability for ICs is still in its infancy. For example, the McPAT use-phase power estimation

flow [40] provides detailed operational phase power estimations for different hypothetical processors targeting different feature sizes. CACTI [41] provides a similar estimation tool for different processor cache configurations. There are also tools for graphics processing units [42] and many others. All of these tools estimate only the operational power consumption, and while some report details of chip area, they do not report details on embodied energy or other manufacturing impacts. In this section we introduce a few recent advances in more holistic energy evaluation.

A. Parameterized, Scalable Process Model

Determining the impact of the changing process technology at different and shrinking geometries or other process trends such as 3D CMOS and hybrid CMOS with post-CMOS extensions such as sensors, emerging memories, etc., requires a method to break down the environmental impact at each stage of the process. By changing the number of steps to reflect a particular process, the aggregate impact can be estimated. Using relative weight of the different process steps [15] and scaling to normalize the aggregate impacts to the published data [12], a first-order parameterized model of impacts for different technology nodes is possible.

Recent work developed such a model [43] to determine the embodied energy contributed by each process step down to the 32nm process node. Based on this parameterized model concept, we propose a model that introduces energy per process step for 20nm, 15nm, 10nm and 7nm shown in Figure 4. The figure reports the relative weight, in terms of fabrication energy, of fundamental process steps at different technologies. For technologies between 130nm and 32nm the model reports expected increases in process steps such as lithography, metrology, deposition, and etching. However, lithography and metrology are increasing at a faster rate than the others, which matches industry trends.

Using the principle that economic cost is an indicator for environmental impacts [17], published economic cost data, and based on the reported process changes from ITRS [19] such as the switch to litho-etch-litho-etch (LELE), self-aligned double, triple, and quadruple patterning at the 22/20, 14/16, 10, and 7nm nodes, respectively are integrated into our new model. Due to the use of different forms of multiple patterning lithography, deposition and etching steps required for multiple patterning are included in lithography and broken down by the optical contribution (solid bar) and the non-optical steps (patterned bar). As with the original published model [43], which provided data down to the 32nm process node, this extended model is capable of reporting energy, carbon emissions, carcinogenic chemicals, volatile organic compounds (VOCs), and wastewater in addition to energy.

This model provides the basis for sustainability explorations of all levels of design and fabrication for ICs. Designer choices at all levels, from system to circuit level, can refer back to the model to quantify the impact of these choices. In particular, green computing choices, including dark silicon, that have traditionally targeted improved operational energy must be considered in the context of how much area they consume. The basic premise of dark silicon is that area used for accelerators or augmented system structures such as caches that are not powered up most of the time is relatively benign, but this is in terms of use-phase energy consumption. The extra area also contributes to the embodied energy of the IC.

In this case we consider an expanded definition of dark silicon area. A traditional definition would include 2D area increases due to the addition of the dark silicon structures. If we consider other possible extensions to the base IC such as the introduction of hybrid non-CMOS techniques, those additional fabrication layers can be thought of as dark silicon in the vertical dimension. That is, the extra fabrication steps are applied to the entire chip, but are only used in the particular structures that require those steps. One example of this type of vertical dark silicon is spin-transfer torque magnetic memory (STT-MRAM) which adds a number of process steps. This provides an important third axis to consider in light of the increasing per-step fabrication cost.

B. Evaluation Flows

The model from the previous section is the heart of evaluation tools to evaluate and compare the manufacturing and use phase impacts of new computing systems. One recent tool is the GreenChip simulation flow [2]. GreenChip first simulates the behavior of a mixture of workloads on a proposed architecture to generate performance statistics. The simulator output, architecture specification, and technology node are then fed into a use phase power estimation flow, which is based on the myriad of existing tools to evaluate operational energy. Notably, GreenChip also includes a manufacturing environmental cost estimator that uses a combination of the technology node impacts per area and the predicted area of the IC or ICs. The area is combined with a simple manufacturing model similar to the cost per area of CMOS logic shown in Figure 2 to determine overall manufacturing cost. Because GreenChip includes area in its manufacturing cost estimates, it is an example of a tool that can be used to explore the relative sustainability of proposed systems that include extra functionality manifesting as dark silicon.

An important part of GreenChip is that it provides a method to compare holistic sustainability. GreenChip uses indifference point [44] and break even analyses of total energy footprint. Indifference point analysis is a common economic metric to determine the point at which there is no difference in cost (in this case sustainability) between two alternatives. The break even time is the point when a new system's operational impact savings will overcome its manufacturing impacts.

Using a tool like GreenChip, it will be possible to explore tradeoffs between operational energy consumption, and manufacturing energy when considering dark silicon augmentation. If the augmentation is in the form of hybrid circuits such as STT-MRAM, a different tool, a manufacturing-saavy version of NVSIM, has been developed that calculates the additional manufacturing impact (including embodied energy) using a similar parameterized model for SRAM and STT-MRAM [20].

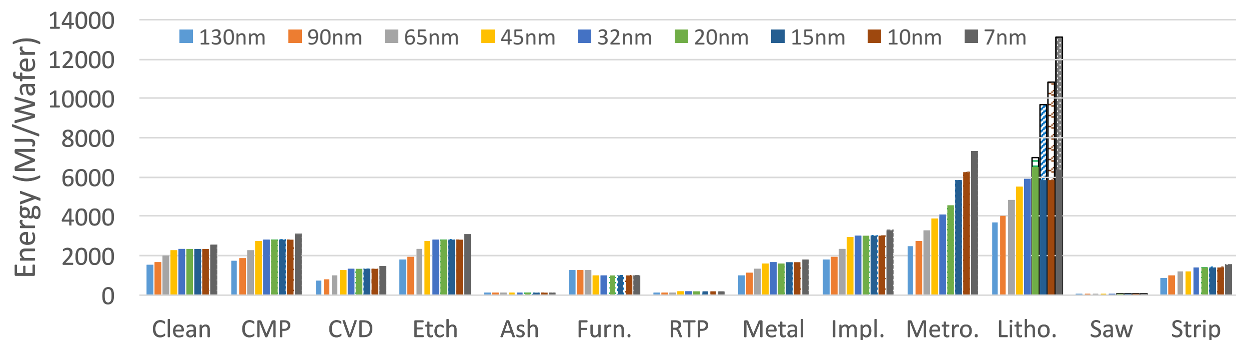


Figure 4. Energy used per process step calculated using the process model.

Table I
MANUFACTURING COSTS FOR CHIPS AT DIFFERENT PROCESS NODES FOLLOWING PRODUCT TRENDS (PSEUDO ISO-AREA) [12, 45, 46].

Process Node (nm)	90	65	45	28	22	14
Core Count	1	2	4	8	8	10
LLC size (MB)	1	2	4	8	20	25
Area (mm^2)	207	227	207	158	356	294
Embodied Energy (MJ)	124	148	164	598	631	721

This approach can quantify the tradeoff both in terms of 2D and 3D dark silicon structures.

V. RESULTS

To quantify these trends we provide two case studies. In the first case study, we examine the tradeoff of using additional semiconductor real estate in standard processor components such as core count and cache size as well as examining the impact of hybrid post-CMOS extensions. In the second study we consider the required energy savings from adding a dark silicon accelerator in terms of holistic energy consumption.

In both studies we consider the baseline of a typical high-end architecture across process technologies shown in Table I. As the technology node descends, the core count and cache size increases. For the later technology nodes, cache size increases much faster than core count, as achieving sufficient parallelism to improve performance beyond a few cores is difficult in commodity systems.

Also in these studies we considered different usage scenarios, as these significantly impact the relationship of the embodied to operational energy usage. Scenarios were selected with different activity and sleep ratios representing the load experienced by a cloud server (Server) that is typically online but often underloaded, a high-performance machine (HPC) that is typically constantly online and heavily loaded, a desktop machine (Desktop) that is used often, but lightly during the working day, and a mobile device (Mobile) that is mostly asleep, but when it wakes up is heavily loaded based on ratios from the literature [12, 47, 48].

A. Case Study 1: Processor Configurations

As processors have descended below the 90nm node, clock frequencies have become relatively fixed to manage thermal

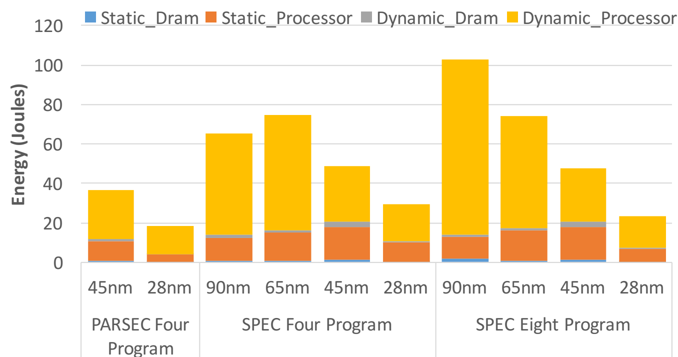


Figure 5. Joules Per 10 billion instructions for Parsec and SPEC multiprogram workloads at different nodes. All runs are with the same chip area as part of the iso-area comparison.

concerns. Thus, the post Dennard scaling method for leveraging the additional density per die to achieve performance improvements by increasing the number of processor cores and on-chip cache sizes. Of course, as we have previously discussed the more area and the smaller the geometry the larger the embodied energy cost of the IC. For example, the pseudo ISO-area processors from Table I presents the highest capability commercial products available across a span of several years. Using GreenChip we analyzed the IPC, energy, and cache misses per kilo-instruction (MPKI) of a mix of benchmarks including the Parsec [49] and SPEC-CPU2006 [50] multi-program workloads. The Parsec workloads are multi-threaded, while the SPEC workloads are single threaded. In this study we were only able to model processors down to 28nm due to limitations in the GreenChip tool. The use phase energy, shown in Figure 5, shows how increasing the core count and cache size can dramatically reduce use phase energy. However, the the appropriate environmental design choice requires a combination of both manufacturing and use phase energy trends. The average indifference points and break even times are shown in Figure 6 for the four scenarios. The results indicate that HPC systems are typically the only ones where moving to a new technology node makes sense from a sustainability point of view.

Another potential trade-off is replacing the SRAM caches

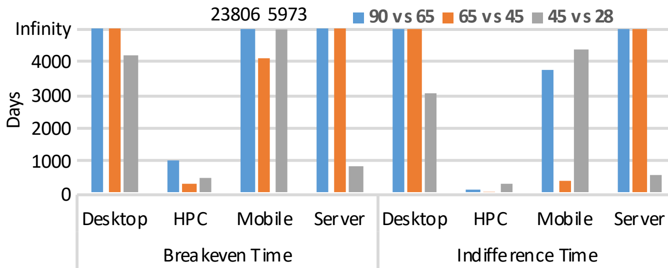


Figure 6. Average break even times and indifference points across all benchmarks for pseudo iso-area comparison.

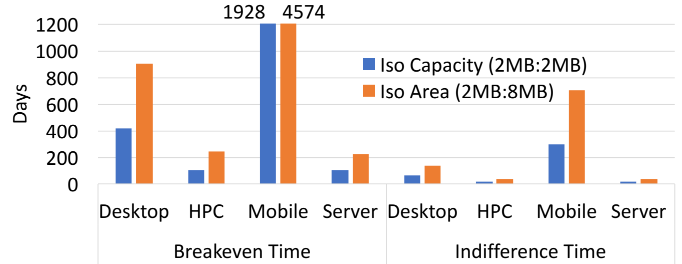


Figure 7. Time to ISO-energy considering the trade-off of manufacturing versus operational energy.

Table II
SRAM AND STT-MRAM CACHE PARAMETERS.

Cache Type	read lat. (ns)	read pwr. (W)	write lat. (ns)	write pwr. (W)	Stat. pwr W	Manufacturing Energy (MJ)	Manufacturing GWP ($kgCO_2e$)
2M SR	1.71	0.26	1.71	0.26	2.95	9.20	0.53
2M STT	3.78	0.23	11.90	0.058	0.3	10.87	0.63
8M STT	4.03	0.25	11.97	0.068	1.01	11.49	0.66

with STT-MRAM to save operational energy. Table II provides a comparison of the operational and embodied energy of the different cache choices from a modified NVSim [20]. ISO-capacity STT-MRAM caches can actually reduce die area, but at the expense of additional process steps, which actually increase the embodied energy. Thus, both the 2D and 3D manufacturing decisions must be considered. ISO-area STT-MRAM increases cache capacity in the same die area but further increases embodied energy due to the more complicated process. In terms of operational phase impacts, the product of energy and cycles per instruction (CPI) reveals that ISO-capacity replacement reduces energy primarily due to reduced static power. It also maintains similar performance to the SRAM cache. The miss rate is unchanged and the access latency penalty compared to SRAM is nominal as the applications are dominated by reads and even the slower writes do impact the critical path. The ISO-area replacement provides a further improvement to performance due to a reduced miss rate and only nominal latency and power increase over the ISO-capacity replacement, achieving a 20% operational phase improvement overall.

Similarly to Figure 6, Figure 7 reports the time required before the operational energy savings makes up for the manufacturing energy overhead from STT-RAM replacement. While in this comparison we are considering only the cache component of the chip isolation, as these memories are caches, the overheads of including STT-RAM would apply to the entire area of the die for the processor, exacerbating the 3D impact of the hybrid fabrication choice. At 45nm, the Desktop, HPC, and Server scenarios all have indifference times under one year, indicating that both the ISO-capacity and ISO-area STT are the more sustainable than SRAM for these scenarios. For the desktop ISO-area comparison, the upgrade from SRAM to the ISO-area STT requires a longer time investment (three years) to recover the sustainability costs. For the Mobile scenario, the breakeven times are over five and twelve years for the ISO-capacity and ISO-area STT, respectively, indicating upgrading to STT does not actually save in sustainability for usage times lower than those intervals.

B. Case Study 2: Dark Silicon Accelerators

Using tools like GreenChip allows us to estimate the tradeoff between two design choices for both operational and embodied energy. We used this tool to examine the impact of adding dark silicon accelerators by considering the operational phase energy improvement required to offset the embodied energy increase for different sized accelerators at different fabrication technologies. For each of the usage scenarios we determined the energy required to operate for one year based on the usage scenario. Then we added dedicated area for hypothetical dark silicon hardware accelerators that consumed an additional 5%, 10%, 25%, and 50% of the processor area. Figures 8 and 9 report the required operational energy savings in terms of a percentage of overall energy necessary to overcome the extra manufacturing energy due to the accelerators. Figure 8 shows this comparison for server scenarios in an HPC and cloud setting and Figure 9 shows this for commodity machines such as desktop machines and mobile computers such as tablets. In this study, the operational energy for active, idle, and sleep for processors without detailed models was reported from processor data sheets [45]. Also note, to provide the best data resolution for the different scenarios, Figure 8 has a different scale (y-axis goes to 15%) than Figure 9 (y-axis goes to 100%).

Clearly, the usage scenario is very important, with HPC machines that are nearly 100% active quickly benefiting with relatively low operational energy savings and mobile devices requiring dramatic benefits to make the dark silicon worthwhile. Another important observable trend is that as the technology node descends, the operational energy savings required to recover the embodied energy increases. For example, even for a small accelerator at 14nm geometries, accelerators much achieve significant reductions for the mobile and desktop devices (14% and 23%, respectively). For larger amounts of dark silicon the required savings for mobile and desktop devices quickly increases, even becoming unrecoverable in one year. For HPC and cloud servers, dark silicon can be more

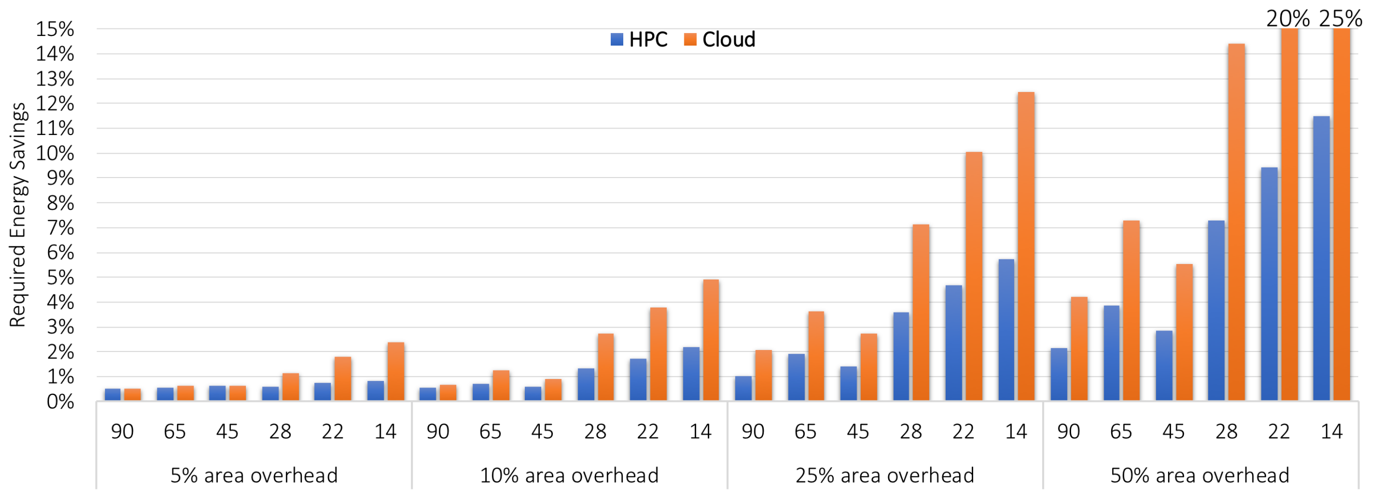


Figure 8. Energy reduction required in HPC and Cloud systems when adding a “dark silicon” accelerator to recover the embodied energy cost in one year.

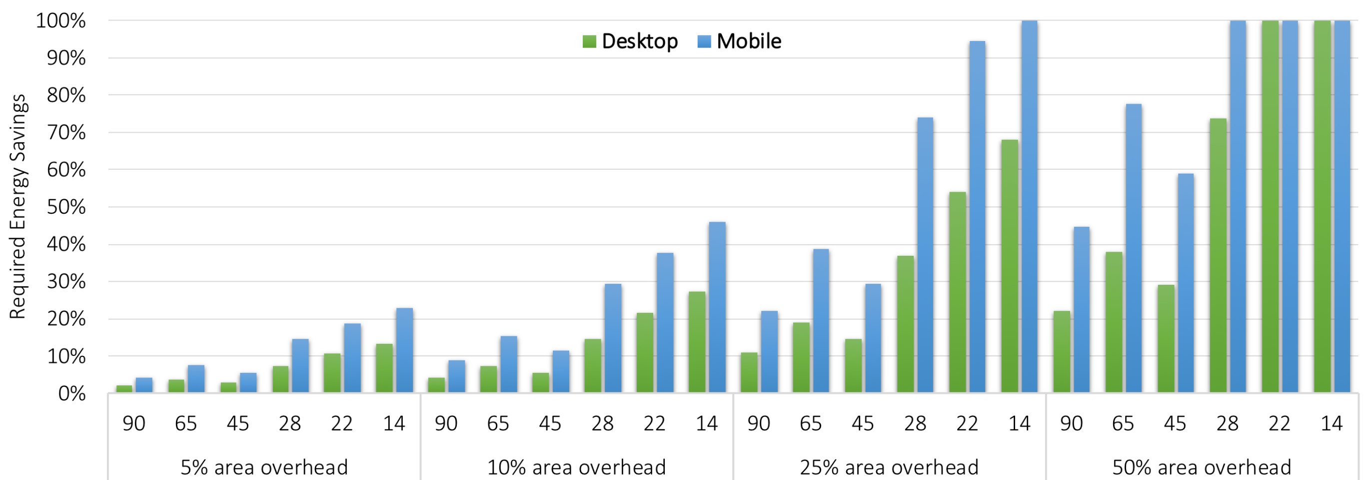


Figure 9. Energy reduction required in Desktop and Mobile systems when adding a “dark silicon” accelerator to recover the embodied energy cost in one year.

useful, but there is still a tradeoff. For significant chip real estate dedicated to dark silicon it may not provide the expected advantage. Adding 25% additional overhead (not dissimilar from adding a GPU), it can require 6% and 13% savings on operational energy at 14nm for the HPC and cloud server case, respectively, to payoff. Thus, dark silicon in some cases provides the opportunity for a systemic advantage, but in many cases, the embodied energy cost should be carefully considered prior to including it in future systems, particularly for mobile devices.

VI. Conclusion

A dark silicon approach that increases chip size by including extra transistors (and chip area), or extra fabrication steps/materials for seldom-used functionality is potentially the wrong choice for sustainability. As new designs and architectures are created and scaled, it feels natural to increase cache sizes, cores, and accelerators, especially if they are powered up only occasionally. With extreme scaling, it has become comfortable to use whatever IC real-estate was necessary to gain use-phase energy reductions. However, based on our analysis, holistic sustainability can result in optimizations that appear counter-intuitive for traditional metrics of performance and use-phase energy. Unfortunately, a system designer likely does not currently have the tools to make these choices. We propose that including IC fabrication sustainability and the concept of embodied energy into a design flow can help the designer understand how system choices impact sustainability.

These tools could also allow consumers to understand the holistic sustainability of a computing product. We suggest that a change in thinking can result in new marketing opportunities at the consumer level that allow for dramatic increases in sustainability. For example, a product with an improved sustainability score might have a strong market if small,

intelligent sacrifices in performance allow for considerable sustainability gains over a competitor. The tools described and extensions proposed here can help quantify and tune the sustainability score of future products. In a future with looming energy and environmental issues, we believe that understanding holistic sustainability is a grand challenge that must be addressed by the EDA community.

References

[1] Jones, A., Chen, Y., Collinge, W., Xu, H., Schaefer, L., Landis, A., and Bilec, M. Considering fabrication in sustainable computing. In *ICCAD* (2013).

[2] Kline Jr., D., Parshook, N., Ge, X., Brunvand, E., Melhem, R., Chrysanthis, P. K., and Jones, A. K. Holistically evaluating the environmental impacts in modern computing systems. In *IGSC* (2016).

[3] Jones, A., Liao, L., Collinge, W., Xu, H., Schaefer, L., Landis, A., and Bilec, M. Green computing: A life cycle perspective. In *IGCC* (June 2013).

[4] Or-Bach, Z. Moore's law has stopped at 28nm. *Solid State Technology: Insights for Electronics Manufacturing* (2014).

[5] ISO. Environmental management – life cycle assessment – requirements and guidelines. Tech. Rep. ISO 14044, 2006.

[6] Vigon, B. W., Tolle, D. A., Cornaby, B. W., et al. Life-cycle assessment: Inventory guidelines and principles. Tech. Rep. EPA/600/R-92/24, Cincinnati, OH, Feb. 1993.

[7] Fava, J., Denison, R., Jones, B., et al., Eds. *A Technical Framework for Life-Cycle Assessment*. SETAC Foundation for Environmental Education, Washington, DC, 1991.

[8] UNEP/SETAC. Life cycle approaches: The road from analysis to practice. Tech. rep., Paris, 2005.

[9] Yao, M. a., Higgs, T. G., Cullen, M. J., Stewart, S., and Brady, T. a. Comparative assessment of life cycle assessment methods used for personal computers. *Env. Sci. & Tech. 44*, 19 (Oct. 2010), 7335–46.

[10] Teehan, P. Lca studies of tablets; embodied co2 of tablets; comparison with similar products. In *Green Electronics Council Slates/Tablets Workshop* (Dallas, TX, 2013).

[11] Apple Inc. Environmental report. [Available Online]: http://www.apple.com/environment/reports/, 2015.

[12] Boyd, S. B. *Life-Cycle Assessment of Semiconductors*. Springer, 2012.

[13] Boyd, S. B., Horvath, A., and Dornfeld, D. Life-cycle energy demand and global warming potential of computational logic. *Env. Sci. & Tech. 43*, 19 (2009), 7303–7309.

[14] Murphy, C. F., Kenig, G. A., Allen, D. T., Laurent, J.-P., and Dyer, D. E. Development of parametric material, energy, and emission inventories for wafer fabrication in the semiconductor industry. *Env. Sci. & Tech. 37*, 23 (2003).

[15] Jones, S. W. Understanding the costs of mems products. IC Knowledge LLC.

[16] Bahar, R. I., Jones, A. K., Katkoori, S., et al. Workshops on extreme scale design automation (ESDA). Tech. rep., CCC, 2014. http://www.cra.org/ccc/files/docs/esda/CCC_ESDA\%20Report.pdf.

[17] Matthews, H. S., and Small, M. J. Extending the boundaries of life-cycle assessment through environmental economic input-output models. *Jouranl of Industrial Ecology 4*, 3 (2001).

[18] Courtland, R. The molten tin solution. *IEEE Spectrum 53*, 11 (November 2016), 28–41.

[19] International technology roadmap for semiconductors. Tech. rep. [Available online] http://www.itrs.net/reports.html.

[20] Bayram, I., Eken, E., Kline, D., Parshook, N., Chen, Y., and Jones, A. K. Modeling stt-ram fabrication cost and impacts in nvsim. In *IGSC* (2016).

[21] Clarke, P., and McGrath, D. *Foundries have 28nm yield issues, say execs*, 2011. https://www.eetimes.com/document.asp?doc_id=1260527.

[22] Eassa, A. *Intel Corporation's 14-Nanometer Yields Still Have Room for Improvement*, 2017. https://www.fool.com/investing/2017/02/18/intel-corporations-14-nanometer-yields-still-have.aspx.

[23] Shilov, A. *Intel Delays Mass Production of 10nm CPUs to 2019*, 2018. https://www.anandtech.com/show/12693/intel-delays-mass-production-of-10-nm-cpus-to-2019.

[24] Sydow, M. Compare logic-array to asic-chip cost per good die. *Chip Design Magazine* (2006).

[25] Hennessy, J. L., and Patterson, D. A. *Computer architecture: a quantitative approach*. Elsevier, 2011.

[26] LaBel, K. A., and Cohn, L. M. Radiation testing and evaluation issues for modern integrated circuits. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050243591.pdf.

[27] Liu, J., Jaiyen, B., Kim, Y., Wilkerson, C., and Mutlu, O. An experimental study of data retention behavior in modern dram devices: Implications for retention time profiling mechanisms. In *ACM SIGARCH Computer Architecture News* (2013), vol. 41, ACM, pp. 60–71.

[28] Liu, J., Jaiyen, B., Veras, R., and Mutlu, O. Raidr: Retention-aware intelligent dram refresh. In *ACM SIGARCH Computer Architecture News* (2012), vol. 40, IEEE Computer Society, pp. 1–12.

[29] Kim, Y., Daly, R., Kim, J., Fallin, C., Lee, J. H., Lee, D., Wilkerson, C., Lai, K., and Mutlu, O. Flipping bits in memory without accessing them: An experimental study of dram disturbance errors. In *ACM SIGARCH Computer Architecture News* (2014), vol. 42, IEEE Press, pp. 361–372.

[30] Hamming, R. W. Error detecting and error correcting codes. *Bell Labs Technical Journal 29* (1950), 147–160.

[31] Nair, P. J., Kim, D.-H., and Qureshi, M. K. Archshield: Architectural framework for assisting dram scaling by tolerating high error rates. In *ACM SIGARCH Computer Architecture News* (2013), vol. 41, ACM, pp. 72–83.

[32] Kline, D., Melhem, R., and Jones, A. K. Sustainable fault management and error correction for next-generation main memories. In *Green and Sustainable Computing Conference (IGSC), 2017 Eighth International* (2017), IEEE, pp. 1–6.

[33] Schechter, S., Loh, G. H., Strauss, K., and Burger, D. Use ecp, not ecc, for hard failures in resistive memories. In *ACM SIGARCH Computer Architecture News* (2010), vol. 38, ACM, pp. 141–152.

[34] Zhang, J., Kline, D., Fang, L., Melhem, R., and Jones, A. K. Yoda: Judge me by my size, do you? In *ICCD* (2017), IEEE, pp. 395–398.

[35] Fan, J., Jiang, S., Shu, J., Zhang, Y., and Zhen, W. Aegis: Partitioning data block for efficient recovery of stuck-at-faults in phase change memory. In *MICRO* (2013), pp. 433–444.

[36] Zhang, J., Kline Jr, D., Fang, L., Melhem, R., and Jones, A. K. Dynamic partitioning to mitigate stuck-at faults in emerging memories. In *ICCAD* (2017), IEEE Press, pp. 651–658.

[37] Environmental Protection Agency. Energy Star. https://www.energystar.gov/.

[38] Pre Consultants. *SimaPro Life Cycle Analysis version 7.2 (software)*. Amersfort, The Netherlands.

[39] GaBi LCA software.

[40] Li, S., Ahn, J. H., Strong, R. D., Brockman, J. B., Tullsen, D. M., and Jouppi, N. P. McPAT: an integrated power, area, and timing modeling framework for multicore and manycore architectures. In *MICRO* (2009).

[41] Jouppi, N. P., and Wilton, S. J. An enhanced access and cycle time model for on-chip caches. Tech. Rep. TR-93-5, Compaq, 1994.

[42] Leng, J., Hetherington, T., ElTantawy, A., Gilani, S., Kim, N. S., Aamodt, T. M., and Reddi, V. J. Gpuwattch: Enabling energy optimizations in gpgpus. *SIGARCH Comput. Archit. News 41*, 3 (June 2013), 487–498.

[43] Kline, D., Parshook, N., Johnson, A., Stine, J. E., Stanchina, W., Brunvand, E., and Jones, A. K. Sustainable ic design and fabrication. In *Green and Sustainable Computing Conference (IGSC), 2017 Eighth International* (2017), IEEE, pp. 1–8.

[44] Scott, M. J., and Antonsson, E. K. Using indifference points in engineering decisions. In *Proc. of ASME Des. Eng. Tech. Confs.* (2000).

[45] Intel Corporation. Intel core i7 (5960x,6950x). Datasheet. Available Online, Accessed December 2016.

[46] Borkar, R., Bohr, M., and Jourdan, S. Advancing moore's law – the road to 14nm. Slideshow, 2014.

[47] Pieper, S. M., Paul, J. M., and Schulte, M. J. A new era of performance evaluation. *IEEE Computer 40*, 9 (2007).

[48] Suckling, J., and Lee, J. Redefining scope: the true environmental impact of smartphones? *Intl. Jour. of Life Cycle Assess. 20*, 8 (2015).

[49] Bienia, C., Kumar, S., Singh, J. P., and Li, K. The parsec benchmark suite: Characterization and architectural implications. In *PACT* (2008).

[50] Henning, J. L. SPEC CPU2006 benchmark descriptions. *SIGARCH Comput. Archit. News 34*, 4 (2006).