
STEEP-SLOPE DEVICES: FROM DARK TO DIM SILICON

ALTHOUGH THE SUPERIOR SUBTHRESHOLD CHARACTERISTICS OF STEEP-SLOPE DEVICES CAN HELP POWER UP MORE CORES, RESEARCHERS STILL NEED CMOS TECHNOLOGY TO ACCELERATE SEQUENTIAL APPLICATIONS BECAUSE IT CAN REACH HIGHER FREQUENCIES. DEVICE-LEVEL HETEROGENEOUS MULTICORES OFFER THE BEST OF BOTH WORLDS, BUT THEY NEED SMART RESOURCE MANAGEMENT TO REALIZE THIS PROMISE. THIS ARTICLE DISCUSSES DEVICE-LEVEL HETEROGENEOUS MULTICORES AND RESOURCE-MANAGEMENT SCHEMES FOR REACHING HIGHER ENERGY EFFICIENCY.

Karthik Swaminathan
Emre Kultursay
Vinay Saripalli
Vijaykrishnan Narayanan
Mahmut T. Kandemir
Suman Datta
Pennsylvania State
University

..... While each technology generation lets us pack more cores on the same die, thermal and power delivery constraints have precluded any scaling in the power budget available to these cores. This forces cores to operate at very low voltages (“dim silicon”) to stay within the allotted power budgets. Unfortunately, low-voltage operation of silicon CMOS technology is extremely energy-inefficient. As the supply voltage approaches the threshold voltage, the transistor delay increases rapidly, resulting in a drop in the clock frequency. To avoid inefficient low-voltage operating points, CMOS multicores typically power on only a subset of the available cores and turn off the remaining cores (“dark silicon”).

Steep-slope devices show much better subthreshold characteristics than CMOS devices because they are not limited by the same thermal subthreshold slope barrier. These devices thus have the potential to attain much better energy efficiencies, making them great candidates to address the “dark silicon apocalypse”¹ through the design of dim-silicon-optimized multicores. Because

existing drive currents in higher-voltage steep-slope devices are limited as compared to CMOS devices, their maximum operating frequency is also limited. Hence, cores that use steep-slope devices fail to reach the energy efficiency and performance of CMOS cores in a dark silicon setting. A general-purpose multicore, on the other hand, must operate well on both ends of the dark and dim silicon spectrum to serve various workloads efficiently.

A heterogeneous multicore comprising cores of different device technologies can enable efficient execution in both dark and dim silicon configurations. In this article, we use a combination of steep-slope devices and CMOS devices in the design of heterogeneous multicores. We adopt the most promising steep-slope device candidate, interband tunnel field-effect transistors (TFETs), and evaluate a heterogeneous CMOS-TFET multicore. (For more information, see the “Related Work in Steep-Slope Devices and Energy-Efficient Architectures” sidebar.) Although device-level heterogeneous multicores can achieve much better energy efficiency

Related Work in Steep-Slope Devices and Energy-Efficient Architectures

Although innovations such as multigate transistor technology (Si FinFETs) have somewhat improved the near-threshold characteristics of CMOS, the physical device limitations are still prevalent at a subthreshold level. These days, many researchers are working on steep-slope switches to overcome the thermal energy limitation-imposed 60 mV/decade subthreshold slope. Recently proposed nano-electro-mechanical switches (NEMS) or NEM relay devices achieve this by utilizing movable electrodes and the instability points between electrical and mechanical forces.¹ Another alternative is to reduce the body factor in gate capacitance to less than 1 by using negative-capacitance FET (NC-FET) and exploiting the nonlinear energy-dependent polarization in ferroelectric gates.² The Tunnel FET (TFET) technology we use in this work uses gate-controlled band-to-band tunneling mechanisms.³

Various efforts in architecture have attacked the dark silicon problem by improving the power consumption of multicores. Examples include using architecturally asymmetric and dynamically morphable multicores.^{4,5} The benefits that can be obtained from these methods are orthogonal to those that can be achieved by moving to a new technology that does not have the same subthreshold slope restrictions as CMOS technology.

References

1. K.A.-M. Ionescu et al., "Ultra Low Power: Emerging Devices and Their Benefits for Integrated Circuits," *Proc. IEEE Int'l Electron Devices Meeting (IEDM 11)*, IEEE CS, 2011, pp. 16.1.1-16.1.4.
2. S. Salahuddin and S. Datta, "Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nano-scale Devices," *Nano Letters*, vol. 8, no. 2, 2008, pp. 405-410.
3. A.C. Seabaugh and Q. Zhang, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proc. IEEE*, vol. 98, no. 12, 2010, pp. 2095-2110.
4. R. Kumar et al., "Heterogeneous Chip Multiprocessors," *Computer*, vol. 38, no. 11, 2005, pp. 32-38.
5. E. Ipek et al., "Core Fusion: Accommodating Software Diversity in Chip Multiprocessors," *Proc. 34th Ann. Int'l Symp. Computer Architecture (ISCA 07)*, ACM, 2007, pp. 186-197.

and performance levels than an equivalent homogeneous multicore, it is difficult to efficiently exploit the power and performance trade-offs between the cores implemented using different technologies. Toward this goal, this article proposes and evaluates various static and dynamic application-aware mapping and scheduling techniques that could significantly improve energy efficiency and performance. Although the techniques this article discusses can be applied in the context of any heterogeneous multicore, using them on device-level heterogeneous multicores gives us new insights on heterogeneous computing. Furthermore, the unparalleled near-threshold and subthreshold performance of TFETs result in hitherto unexplored cost functions and operating points.

Background

The ever-increasing power density problem forces us to search for new ways to save power. As Figure 1a shows, it's no longer possible to reduce the supply voltage of conventional CMOS transistors without leading to unacceptable off-state leakage currents or significantly impacting the switching performance of transistors. For a technology

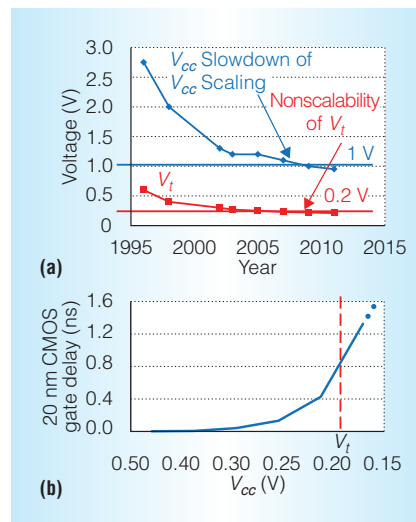


Figure 1. Scaling of CMOS supply voltage (V_{cc}) and threshold voltage (V_t) (a), and subthreshold delay behavior of CMOS (b). We observe that, with subsequent technology generations, the gap between the supply and threshold voltage for CMOS transistors goes on reducing, resulting in a sharp rise in delay.

node with a fixed threshold voltage, reducing the supply voltage further pushes devices into the subthreshold regime on account of reduced $V_{cc} - V_t$. Figure 1b shows that, in the subthreshold region, CMOS device delay grows exponentially. This degradation is due to the intrinsic 60 mV/decade minimum subthreshold slope of CMOS devices, which leads to very low subthreshold drive currents. This, in turn, causes CMOS circuits to operate at extremely low frequencies. Consequently, CMOS multicores show poor performance in a low-voltage, dim silicon configuration, and prefer operating at a higher-voltage, dark silicon configuration.

The use of alternative device technologies has been proposed in order to overcome this barrier imposed by CMOS technology. Recently, various new steep-slope devices have emerged that can implement energy-efficient low-voltage circuits. These devices' physics let them achieve sub-60 mV/decade subthreshold slopes. This leads to higher I_{on}/I_{off} ratios at low voltages, which translates into higher drive currents and lower off-state leakage currents. At near-threshold and subthreshold voltages, steep-slope devices have the potential to achieve superior performance with energy efficiencies that are orders of magnitude higher than CMOS devices. One instance of steep-slope devices, the interband TFET, is a promising slope device candidate due to its superior operation stability and better fabrication compatibility than other alternatives. TFETs show tremendous potential for scaling supply voltages and reducing power consumption. Researchers have already demonstrated logic and memory applications using TFET devices operating at 0.1 V,² and processors designed using TFETs are projected to be in production by 2020.

Although steep-slope devices are promising, they have some disadvantages compared to CMOS technology. For instance, TFET energy efficiency is superior to CMOS only at low voltages. As the supply voltage increases, the inherent limitation in the TFET charge-carrying mechanism causes the current to saturate above a certain operating voltage. This escalates power consumption rapidly and restricts maximum operating frequency. Although

a multicore processor made solely of TFETs can achieve much better performance on a dim silicon configuration, it will not be able to reach high frequencies on a dark silicon configuration. On the other hand, different applications prefer different configurations; for instance, scalable applications prefer using all cores and exploit thread-level parallelism (TLP), but unscalable applications, or applications with large sequential regions, better benefit from using a few cores at higher frequencies. Clearly, neither a homogeneous CMOS nor a homogeneous TFET multicore can serve both purposes.

In this article, we consider a heterogeneous multicore comprising a few CMOS cores (that are particularly useful for accelerating sequential or unscalable codes) and many TFET cores (that are optimized to operate efficiently at low voltages to cater to highly parallel workloads). Figure 2 compares a homogeneous CMOS multicore with a heterogeneous CMOS-TFET multicore in both dark and dim silicon configurations. The heterogeneous multicore can match the dark silicon performance of the homogeneous configuration because it can activate the same number of CMOS cores at high frequencies. In addition, it can outperform the homogeneous processor in a dim silicon setting because it employs low-voltage optimized TFET cores. The heterogeneous multicore can thus use the same power budget to either turn on more cores at the same frequency or use the same number of cores at higher frequencies.

Although a CMOS-TFET heterogeneous multicore can operate efficiently on both dark and dim silicon configurations, an application could prefer one configuration over another owing to factors such as peak instruction throughput and thread and core scalability. Therefore, mapping applications on a heterogeneous system poses several interesting questions. Given a number of applications to execute, how many cores of each type and how much power should each application be allocated? How should the applications' threads share these resources? In order to answer these questions, we can formulate an optimization problem by reducing power consumption under performance

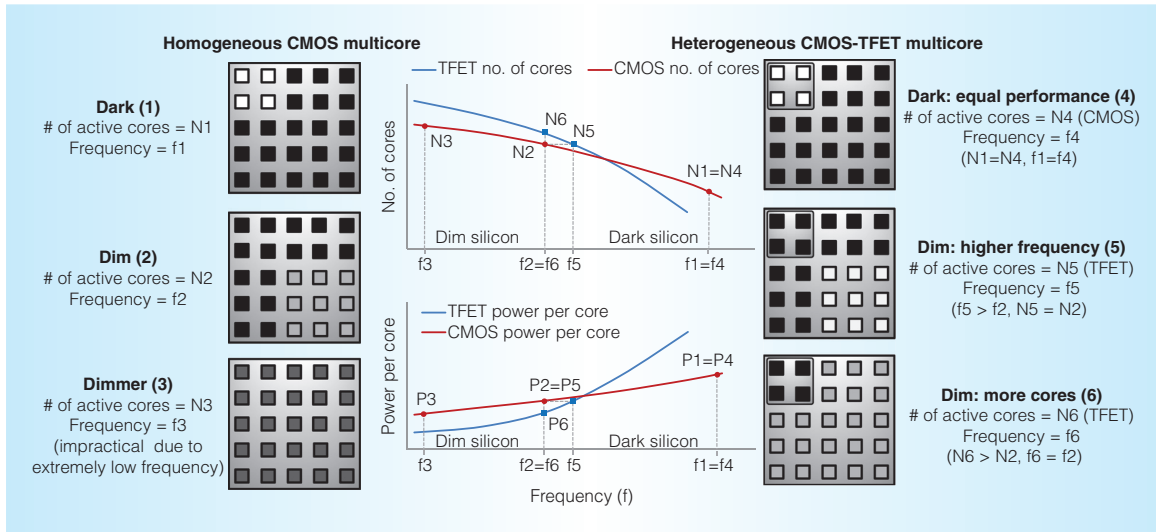


Figure 2. A homogeneous CMOS multicore (left) and a heterogeneous CMOS-TFET (tunnel field-effect transistor) multicore (right) operating at dark and dim silicon settings because of limited available power. The graphs (center; not to scale) show the frequency and number of cores, and frequency and power per core trade-offs, between the two types of cores. In a dark silicon setting (fewer cores, higher voltage), the heterogeneous multicore can match the homogeneous multicore's performance as long as it contains enough CMOS cores (1 versus 4). In a dim silicon setting (more cores, smaller voltage), the heterogeneous multicore can outperform the homogeneous multicore either by using the same number of TFET cores at a higher frequency (2 versus 5) or by using more TFET cores at the same frequency (2 versus 6). Further dimming the CMOS multicore can enable more cores to be turned on but forces these cores to operate at extremely low frequencies, leading to very poor performance (3).

constraints, and improving performance under power constraints.

In this article, we approach this problem from these two perspectives and propose various static and dynamic application scheduling and mapping schemes. We demonstrate our schemes' effectiveness in improving the energy efficiency of heterogeneous CMOS-TFET multicores to well above equivalent homogeneous multicores.

A dark silicon approach

Dark silicon causes large parts of a multicore to be left without power. This can be considered as a motivation to incorporate diversity into the multicore. This gives us the opportunity to identify and use the configuration that best suits the workload being executed, and also dynamically adapt to the application's needs as well.

Modern multicores already apply dynamic energy-saving methods such as dynamic voltage and frequency scaling (DVFS). One version of DVFS reduces the

core voltage and frequency in memory-bound application phases to reduce energy while minimizing performance degradation, using a metric known as the energy-delay product (EDP). Another DVFS method addresses a problem specific to multithreaded applications, where a thread arriving at a synchronization barrier must wait until all other threads also reach the barrier, consuming power without doing any useful work. Barrier-aware DVFS reduces fast cores' voltage and frequency to minimize barrier wait times and improve energy efficiency.

When these DVFS techniques are employed, cores frequently operate at voltages and frequencies lower than the maximum. Therefore, on a CMOS-TFET heterogeneous multicore, a dynamic thread migration scheme can complement these techniques. Whenever a CMOS core's frequency falls below the CMOS-TFET crossover frequency f_c (that is, the frequency at which the most energy-efficient core type switches from CMOS to TFET) as a result of DVFS,

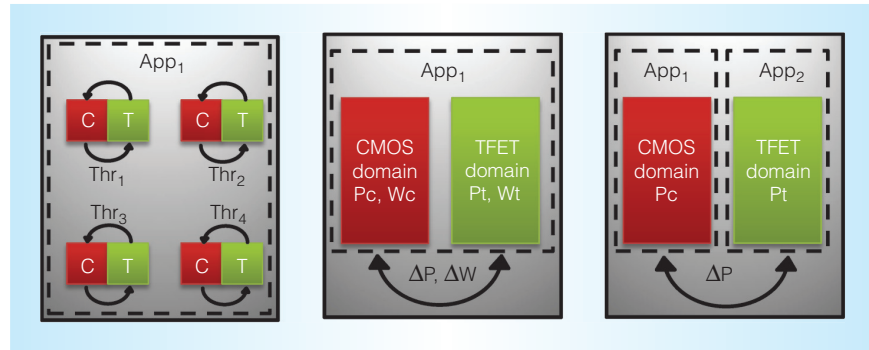


Figure 3. Dynamic optimizations proposed for improving the performance and energy efficiency of heterogeneous CMOS-TFET multicores. Migrating threads across CMOS and TFET cores (a); power and work partitioning across threads belonging to a single application (b); and power and resource partitioning across threads belonging to different applications (c).

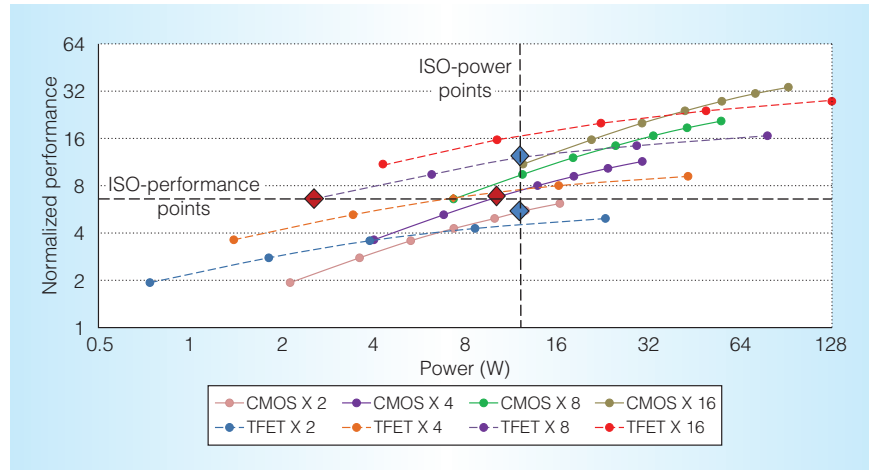


Figure 4. Performance versus power characteristics of a sample benchmark for various numbers of CMOS and TFET cores. Each curve is plotted over the full range of operating frequencies for each device type.

migrating the thread running on the CMOS core to the TFET core coupled with this CMOS core puts the thread into a more energy-efficient execution state,³ as Figure 3a shows. Similarly, an increase in TFET core frequency above f_c triggers a thread migration from the TFET core to the corresponding CMOS core. This scheme enables each tile in the heterogeneous multicore to search among the iso-performance configurations (Figure 4) and use the one with the smallest power consumption. Hence, each tile acts as a low-voltage optimized core at low voltages

and as a high-voltage optimized core at high voltages. To analyze this method's benefits, we simulated an eight-core processor with four CMOS cores and four TFET cores. We assumed that, due to power limitations, the maximum number of cores that can simultaneously be powered on is restricted to 4, but the choice of the exact CMOS/TFET core combination to use can vary dynamically. The baseline homogeneous system and the proposed heterogeneous system both employ the two energy-saving DVFS mechanisms: EDP-aware DVFS and barrier-aware DVFS.

To determine our thread migration scheme's impact on the heterogeneous system's energy efficiency, we measured its EDP while running multithreaded *Splash-2* benchmarks. Our results, presented in Figure 5, show that, on average the heterogeneous multicore has 20 percent better EDP than the homogeneous multicore. As the impact of the DVFS schemes varies on each benchmark, the major source of benefits with thread migration also varies. For instance, in *lu*, we observed the biggest EDP improvements when we used EDP-aware DVFS allied with thread migration, whereas in *water-spa*, thread migration makes the biggest impact when used in conjunction with barrier-aware DVFS. These results indicate that significant energy savings can be obtained by exploiting the energy-efficient behavior of TFET cores at low voltages. As a generalization, we can treat this 4-CMOS/4-TFET configuration as one tile of a larger many-core, and we can use the energy saved on one tile to turn on or ramp up the power budget of other tiles in the system.

A dim silicon approach

An alternate formulation of the energy-efficiency optimization problem is to maximize performance under a fixed power budget. Here, instead of letting large parts of the multicore remain unused because of dark silicon, we adopt a dim silicon approach. We now have more cores sharing the available power budget, and to find an energy-efficient runtime configuration, we exploit the applications' characteristics when distributing resources.

Based on the type of workload we are running on the heterogeneous multicore, we classify this resource distribution problem as

- a multithreaded application executing alone, or
- two (or more) multithreaded applications sharing the cores and power budget.

For both types of workloads, we designed and evaluated static and dynamic schemes that map the available cores and redistribute the available power to threads and applications on a heterogeneous multicore to improve performance. (See the "Experimental Setup" sidebar for additional details on our simulation infrastructure and experiments.)

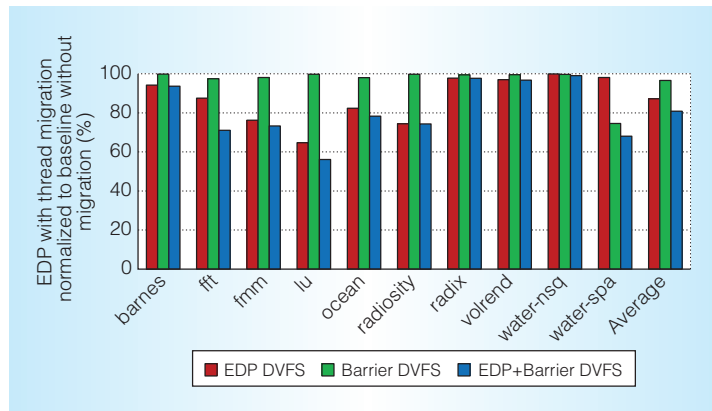


Figure 5. The reduction in energy-delay product (EDP) obtained by enabling thread migration on the heterogeneous multicore when EDP-aware dynamic voltage and frequency scaling (DVFS), barrier-aware DVFS, and both EDP-aware and barrier-aware DVFS are used. We observe that the EDP improves by nearly 20 percent when both DVFS techniques are implemented.

A single multithreaded application on a heterogeneous multicore

Given an application to be executed on a heterogeneous multicore, we can consider two possible thread-to-core mapping schemes:

- using only one type of core (either CMOS or TFET, exclusively) at any time (homogeneous mapping), and
- using both types of cores simultaneously (heterogeneous mapping).

In homogeneous mapping, cores of the unused type are left dark, allowing the active cores to use the entire power budget. In heterogeneous mapping, all cores will share the total available power budget, and the application threads will be mapped to both CMOS and TFET cores. Because cores of different device types have different V/f characteristics, these cores will run at different operating points although they are allocated equal per-core power budgets. This will result in different types of cores having unequal performance. We thus employ a dynamic load-balancing scheme (for example, using dynamic parallel loop scheduling) to avoid any inefficiencies that could arise due to equal work partitioning across application threads. After load balancing, the core type that operates at a more energy-efficient point would complete more work in the same time frame because all cores have an equal per-core power budget

Experimental Setup

We performed our experiments using the Simics full-system simulator. For our thread migration study, we simulated a 4-CMOS homogeneous and a 4-CMOS/4-TFET heterogeneous multicore.¹ For our power and work partitioning experiments, we modeled the multicores listed in Table 2 in the main article.² Our Si-FinFET and TFET cores are architecturally similar to the Intel Atom Z520.³ These cores were also equipped to run DVFS with a 1-ms epoch. In our dynamic work partitioning study, we modified the SPEC-OMP 2001 benchmarks to incorporate dynamic loop scheduling. For our experiments with multiple multithreaded applications, we built eight workloads by randomly pairing Parsec benchmarks. In our workloads, each application is associated with a user-defined weight that represents its relative importance. Further details on our experimental setup and simulation parameters can be found elsewhere.²

References

1. K. Swaminathan et al., "Improving Energy Efficiency of Multithreaded Applications Using Heterogeneous CMOS-TFET Multicores," *Proc. Int'l Symp. Low Power Electronics and Design (ISLPED 11)*, IEEE CS, 2011, pp. 247-252.
2. E. Kultursay et al., "Performance Enhancement under Power Constraints Using Heterogeneous CMOS-TFET Multicores," *Proc. Int'l Conf. Hardware/Software Codesign and System Synthesis (CODES+ISSS 12)*, ACM, 2012, pp. 245-254.
3. *Intel Atom Processor Z5xx Series Datasheet* Intel, tech. report 319535-003US, June 2010.

Table 1. Evaluated schemes for single workloads in a dim silicon approach.

Processor	Thread mapping	Workload partitioning across threads	Power partitioning across threads	Code
32 CMOS or 32 TFET	CMOS or TFET	Equal	Equal	<i>BestBase</i>
8 CMOS and 24 TFET	CMOS and TFET	Equal	Equal	<i>Hetero-BestManual</i>
	CMOS and TFET	Equal	Equal	<i>Hetero-Naive</i>
	CMOS and TFET	Dynamic	Equal	<i>Hetero-DynWork</i>
	CMOS and TFET	Dynamic	Dynamic	<i>Hetero-DynWork-DynPow</i>

(iso-power points in Figure 4). In this case, repartitioning the available power across cores, as in Figure 3b, can further improve the multicore's overall performance.

We implemented and evaluated a power partitioning scheme that treats CMOS and TFET cores as two independent, homogeneous power domains and redistributes the total chip power among the domains using a *perturb-and-observe* method.⁴ This scheme periodically transfers a small amount of power from one domain to the other and observes the resulting performance improvement or degradation. Depending on the outcome, it either continues to transfer power in the same direction, or reverses the direction of the power transfer. By combining power partitioning with heterogeneous application mapping and dynamic loop scheduling, we automatically optimize performance at runtime.

Tables 1 and 2 list the mapping techniques we evaluated for this study. Starting with two baseline homogeneous 32-core processors that are all-CMOS and all-TFET, we first determined the configuration that shows the better performance for each application (*BestBase*). We then analyzed our heterogeneous 8-CMOS/24-TFET multicore with a homogeneous mapping. An application can prefer running on 8 CMOS or 24 TFET cores on the basis of its scaling behavior. We assume that, in our baseline heterogeneous configuration (*Hetero-BestManual*), the user selects the best-performing option. To explore the benefits of using both types of cores in the heterogeneous processor simultaneously, we first evaluated a naive technology substitution scheme without any runtime mapping or scheduling optimizations (*Hetero-Naive*). We then enabled

Table 2. Evaluated schemes for multiprogrammed workloads in a dim silicon approach.

Processor	Thread mapping	Workload partitioning across threads	Power partitioning across threads	Code
32 CMOS or 32 TFET	CMOS or TFET	Equal	Static	<i>BestBase-StaticPow</i>
	CMOS or TFET	Equal	Dynamic	<i>BestBase-DynPow</i>
8 CMOS and 24 TFET	CMOS and TFET	Equal	Static	<i>Hetero-StaticPow</i>
	CMOS and TFET	Equal	Dynamic	<i>Hetero-DynPow</i>

dynamic work partitioning (*Hetero-DynWork*) and dynamic power partitioning (*Hetero-DynWork-DynPow*).

Figure 6 shows our results with three heterogeneous multicore experiments normalized to *BestBase*. Homogeneous mapping (*Hetero-BestManual*) results in a 5 percent performance degradation. A simple technology substitution (*Hetero-Naive*) yields only a 4 percent improvement. Adding dynamic work partitioning (*Hetero-DynWork*) brings an additional 12 percent performance improvement. Our combined scheme (*Hetero-DynWork-DynPow*) performs best, achieving 21 percent better performance than the baseline.

Multiple multithreaded applications sharing a heterogeneous multicore

We propose static and dynamic optimizations to improve the performance of a power-constrained multicore when two applications are running concurrently. To simplify the problem, we use a homogeneous application-to-core mapping, where each application is assigned to either CMOS or TFET cores. In our static scheme, we first examine the relative scalability of applications using static profiling. Working with two applications scheduled to run together, the application that scales better with the number of cores is mapped to TFET cores and the application that scales better with frequency runs on CMOS cores. The total power budget is partitioned among the two applications (that is, the CMOS and TFET domains) based on the ratio of the user-defined application weights.

Because power and core allocation in our profile-based scheme is fixed throughout the entire execution of the workload, it cannot capture the changing behavior of applications. Hence, we propose a dynamic scheme that starts with the initial power allocation

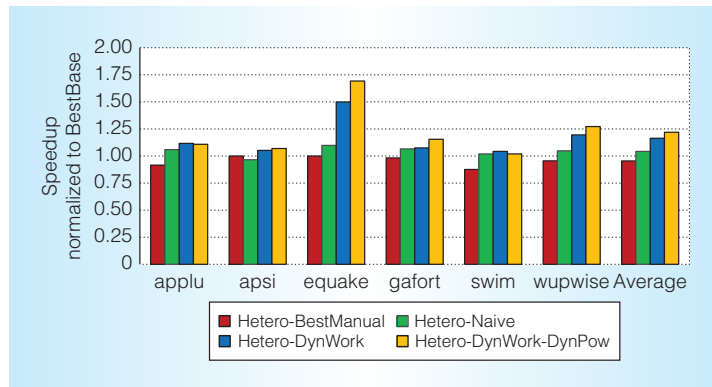


Figure 6. Performance improvement obtained with the CMOS-TFET multicore using homogeneous mapping, heterogeneous mapping, dynamic work partitioning, and dynamic work and power partitioning combined. Our combined scheme (*Hetero-DynWork-DynPow*) achieves 21 percent better performance than the baseline.

identified by the static scheme and dynamically repartitions power based on the energy efficiency the applications achieve (see Figure 3c). It uses a scheme similar to the *perturb-and-observe* mechanism. In this case, because each application runs on only one type of core, we distribute the power allocated to each domain equally across its cores. To address fairness concerns, we limit the maximum performance degradation that an application can suffer to 10 percent.

Tables 1 and 2 list the configurations we evaluated for this study. For our baseline, we choose the best-performing homogeneous configuration out of two 32-core processors with all-CMOS and all-TFET cores when using static power partitioning (*BestBase-StaticPow*). The amount of power allocated to each application is decided statically based on the weights and the power budget. The number of cores to use for each application is selected using this power allocation and profile-based scaling information. Each

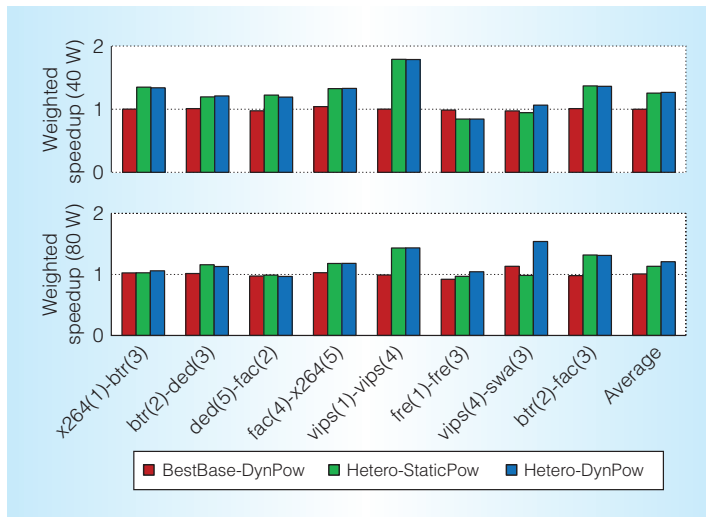


Figure 7. Weighted speedups obtained with static and dynamic power-partitioning methods when two applications are executed together (normalized to *BestBase-StaticPow*). Results are obtained for homogeneous and heterogeneous multicores under 40 W (top) and 80 W (bottom) power budgets. Application weights are given in parentheses.

core uses DVFS while staying within the allocated per-core power budget. The heterogeneous equivalent of this static scheme (*Hetero-StaticPow*) considers what type of core would be better for each application when making application-to-core mapping decisions. The dynamic versions of these two schemes use the adaptive power partitioning discussed earlier (*BestBase-DynPow* and *Hetero-DynPow*).

Figure 7 shows our results obtained with two different power budgets (40 W and 80 W), normalized to our baseline (*BestBase-StaticPow*). Dynamic power partitioning has a negligible impact on the homogeneous system because equal power partitioning is sufficient when all cores are identical. In the 40 W case (top), the small per-core power budget results in the power transferred in our dynamic schemes being too small to cause transitions across DVFS levels. Therefore, our dynamic schemes yield small benefits. We still observe significant performance improvements due to our heterogeneous multicore (25 percent with *Hetero-StaticPow* and 27 percent with *Hetero-DynPow*). For the 80 W case (bottom), our dynamic schemes yield higher speedups than the static schemes. This time, our static and dynamic schemes enable the heterogeneous multicore to bring 13 and 21 percent performance

improvement, respectively. Note that the benefits we get from the technology change decrease when we have more power, because higher per-core power budgets reduce the TFET cores' advantage over CMOS cores.

The inherent physical limitations of CMOS transistors at near-threshold and subthreshold operating voltages has necessitated researchers to search for new device technologies and examine the adoption of device-level heterogeneous processors for next-generation architectures. Processors designed using steep-slope transistors, especially TFETs, have the potential to achieve unprecedented performance and energy efficiencies over a wide range of operating points.

There are, however, several challenges that remain before steep-slope transistor-based processors can become a reality. Steep-slope devices have yet to emulate the peak performance capabilities of state-of-the-art high-performance CMOS transistors. In addition, although the TFET manufacturing process steps are compatible with those of CMOS processors, it is still too early to examine the effects of process variation and yield on the process. Although steep-slope devices have been used to realize basic combinational and sequential circuits, realizing entire processors from these devices is still under ongoing research.

MICRO

Acknowledgments

This work was supported in part by the Center for Low Energy Systems Technology (LEAST), one of six centers supported by the STARnet phase of the Focus Center Research Program (FCRP), an SRC program sponsored by MARCO (Microelectronics Advanced Research Corporation) and DARPA. It was also supported by NSF awards 1205618 and 1213052. Karthik Swaminathan was supported by the IBM PhD fellowship for the 2012-2013 academic year. The authors also thank Huichu Liu for her valuable input.

References

1. M.B. Taylor, "Is Dark Silicon Useful? Harnessing the Four Horsemen of the Coming Dark Silicon Apocalypse," *Proc. 49th Ann.*

- Design Automation Conf. (DAC 12)*, ACM, 2012, pp. 1131-1136.
2. S. Mookerjee et al., "Experimental Demonstration of 100 nm Channel Length In_{0.53}Ga_{0.47}As-based Vertical Inter-band Tunnel Field Effect Transistors (TFETs) for Ultra Low-Power Logic and SRAM Applications," *Proc. IEEE Int'l Electron Devices Meeting (IEDM 09)*, IEEE CS, 2009, doi:10.1109/IEDM.2009.5424355.
 3. K. Swaminathan et al., "Improving Energy Efficiency of Multithreaded Applications Using Heterogeneous CMOS-TFET Multicores," *Proc. Int'l Symp. Low Power Electronics and Design (ISLPED 11)*, IEEE CS, 2011, pp. 247-252.
 4. E. Kultursay et al., "Performance Enhancement under Power Constraints Using Heterogeneous CMOS-TFET Multicores," *Proc. Int'l Conf. Hardware/Software Codesign and System Synthesis* ACM, 2012, pp. 245-254.

Karthik Swaminathan is a PhD student in the Computer Science and Engineering Department at Pennsylvania State University. His research focuses on power-aware computer architectures. He is currently working on leveraging emerging device technologies in the architectural domain to improve performance, power, and reliability. Swaminathan has a bachelor's and master's degree in electrical engineering from the Indian Institute of Technology, Madras.

Emre Kultursay is a PhD student in the Computer Science and Engineering Department at Pennsylvania State University. His research interests include compiler optimizations for high-performance computing systems, energy-efficient multicore architectures, and heterogeneous processors and systems. Kultursay has a BS in electrical engineering and computer engineering from the Middle East Technical University, Turkey.

Vinay Saripalli is a design, technology, and CAD services engineer at Intel. His research interests involve development of CAD automation flows for analog circuit design productivity improvement, and process

migration methods for analog circuits. Saripalli has a PhD in computer science and engineering from Pennsylvania State University, where he completed the work for this article.

Vijaykrishnan Narayanan is a professor in the Departments of Computer Science and Engineering and Electrical Engineering at Pennsylvania State University. His research interests include power-aware and reliable systems, embedded systems, nanoscale devices and interactions with system architectures, reconfigurable systems, networks-on-chips, and domain-specific computing. Narayanan has a PhD in computer science and engineering from the University of South Florida.

Mahmut T. Kandemir is a professor in the Computer Science and Engineering Department at Pennsylvania State University. His research interests include optimizing compilers, runtime systems, embedded systems, I/O and high-performance storage, and power-aware computing. Kandemir has a PhD in computer science and engineering from Syracuse University.

Suman Datta is a professor in the Department of Electrical Engineering at Pennsylvania State University. His research interests include exploring new materials, novel nanofabrication techniques, and device structures for CMOS enhancement and replacement for future energy-efficient, high-performance information processing systems. Datta has a PhD in electrical and computer engineering from the University of Cincinnati.

Direct questions and comments about this article to Karthik Swaminathan, 111, Information Sciences and Technology Building, Pennsylvania State University, University Park, PA 16802; kvs120@cse.psu.edu.



Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.