

📡 (<http://semiengineering.com/feed>)

f (<https://www.facebook.com/SemiEngineering>)

🐦 (<http://www.twitter.com/SemiEngineering>)

in (<http://www.linkedin.com/company/semiconductor-engineering>)

📺 (<http://www.youtube.com/user/SperlingMediaGroup>)

MENU ☰

SYSTEM-LEVEL DESIGN (/CATEGORY-MAIN-PAGE-SLD/)

# Darker Silicon

*MRAMs offer less volatile cache to address the dark silicon dilemma. What happened to Dennard's Law?*

JANUARY 29TH, 2015 - BY: **KATHERINE DERBYSHIRE**  
([HTTP://SEMIENGINEERING.COM/AUTHOR/KATHERINE/](http://semiengineering.com/author/katherine/))



For the last several decades, integrated circuit manufacturers have focused their efforts on **Moore's Law** ([http://semiengineering.com/kc/knowledge\\_center.php?kcid=74](http://semiengineering.com/kc/knowledge_center.php?kcid=74)), increasing transistor density at constant cost. For much of that time, Dennard's Law also held: As the dimensions of a device go down, so does power consumption. Smaller transistors ran faster, used less power, and cost less.

As most readers already know, however, there was a limit. Smaller devices with thinner dielectrics and shorter channels are more prone to leakage. Indeed, leakage, negligible for much of the industry's history and ignored in Dennard's **original paper** (<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=4785543>), now approaches the same order of magnitude as the circuit's dynamic power. Advances such as the introduction of high dielectric constant gate dielectric materials helped, but leakage-limited transistor structures are now a fact of life. Switching a transistor at a lower threshold voltage requires a

## SPONSORS

 ( <a href="http://www.mentor.com/">http://www.mentor.com/</a> )	 ( <a href="http://www.cadence.com/">http://www.cadence.com/</a> )
 ( <a href="http://www.synopsys.com/">http://www.synopsys.com/</a> )	 ( <a href="http://www.open-silicon.com/">http://www.open-silicon.com/</a> )
 ( <a href="http://www.esilicon.com/">http://www.esilicon.com/</a> )	 ( <a href="http://www.arteris.com/">http://www.arteris.com/</a> )
 ( <a href="http://www.sonicsinc.com/">http://www.sonicsinc.com/</a> )	 ( <a href="http://www.accelera.com/">http://www.accelera.com/</a> )
 ( <a href="http://si2.org">http://si2.org</a> )	

## NEWSLETTER SIGNUP

Email:

- Interests:
- System-Level Design
  - Low Power-High Performance
  - Manufacturing, Design & Test

thinner gate dielectric, but leakage constraints place a lower bound on dielectric thickness. As a result, while feature sizes have continued to shrink, threshold voltage has not.

### Plenty of transistors, not enough power

This failure of Dennard scaling has introduced the era of what designers call “dark silicon.” If the number of transistors doubles, but the power budget for the circuit as a whole stays the same — or goes down, thanks to the proliferation of mobile devices — then the available power for each transistor is cut in half. If threshold voltage stays the same, then the number of transistors that can operate at one time is also cut in half. These non-operational transistors are dark silicon, measured as a fraction of the chip’s total area.

Calculating the power consumption of a generic chip is difficult. It depends on a wide range of factors, from dielectric thickness and process variation to the workload of different parts of the chip. Still, as Greg Yeric, senior principal engineer at **ARM** (<http://semiengineering.com/kc/entity.php?eid=22186>), explained in a short course at the 2014 IEEE Electron Device Meeting, projections estimate the dark silicon fraction will be about one-third of total area in the 20nm technology node (including 16/14nm finFETs), increasing to as much as 80% by the 5nm node. Real products are likely to achieve better results, but clearly power consumption imposes an increasingly severe design constraint.

At that point manufacturers may be tempted to ask why they are putting so much effort into making smaller transistors if designers aren’t planning to use them. Part of the answer is that “dark” silicon is not “useless” or “wasted” silicon. In any design, many circuit paths will be “dark” at any given moment. Some elements, such as specialized logic and cache memory, are particularly “dark-silicon friendly,” in that they contribute to overall IC performance while consuming power only in special situations.

### Specialized cores help...somewhat

In fact, this insight led to the integrated circuit industry’s current focus on multicore designs. If a problem can be broken into parallel components, then several cores running at a relatively low speed can still deliver better overall

IoT & Security

Subscribe

### POPULAR TAGS

2.5D ([HTTP://SEMIENGINEERING.COM/TAG/2-5D/](http://semiengineering.com/tag/2-5d/))

3D-IC ([HTTP://SEMIENGINEERING.COM/TAG/3D-IC/](http://semiengineering.com/tag/3d-ic/))

14NM ([HTTP://SEMIENGINEERING.COM/TAG/14NM/](http://semiengineering.com/tag/14nm/))

20NM ([HTTP://SEMIENGINEERING.COM/TAG/20NM/](http://semiengineering.com/tag/20nm/))

ANSYS ([HTTP://SEMIENGINEERING.COM/TAG/ANSYS/](http://semiengineering.com/tag/ansys/))

ANSYS APACHE ([HTTP://SEMIENGINEERING.COM/TAG/ANSYS-APACHE/](http://semiengineering.com/tag/ansys-apache/))

APACHE DESIGN ([HTTP://SEMIENGINEERING.COM/TAG/APACHE-DESIGN/](http://semiengineering.com/tag/apache-design/))

APPLE ([HTTP://SEMIENGINEERING.COM/TAG/APPLE/](http://semiengineering.com/tag/apple/))

APPLIED MATERIALS ([HTTP://SEMIENGINEERING.COM/TAG/APPLIED-MATERIALS/](http://semiengineering.com/tag/applied-materials/))

ARM ([HTTP://SEMIENGINEERING.COM/TAG/ARM/](http://semiengineering.com/tag/arm/))

ARTERIS ([HTTP://SEMIENGINEERING.COM/TAG/ARTERIS/](http://semiengineering.com/tag/arteris/))

ATRENTA ([HTTP://SEMIENGINEERING.COM/TAG/ATRENTA/](http://semiengineering.com/tag/atrenta/))

BUSINESS ([HTTP://SEMIENGINEERING.COM/TAG/BUSINESS/](http://semiengineering.com/tag/business/))

CADENCE ([HTTP://SEMIENGINEERING.COM/TAG/CADENCE/](http://semiengineering.com/tag/cadence/))

EDA ([HTTP://SEMIENGINEERING.COM/TAG/EDA/](http://semiengineering.com/tag/eda/))

ESILICON ([HTTP://SEMIENGINEERING.COM/TAG/ESILICON/](http://semiengineering.com/tag/esilicon/))

EUV ([HTTP://SEMIENGINEERING.COM/TAG/EUV/](http://semiengineering.com/tag/euv/))

FD-SOI ([HTTP://SEMIENGINEERING.COM/TAG/FD-SOI/](http://semiengineering.com/tag/fd-soi/))

FINFETS ([HTTP://SEMIENGINEERING.COM/TAG/FINFETS/](http://semiengineering.com/tag/finfets/))

GLOBALFOUNDRIES ([HTTP://SEMIENGINEERING.COM/TAG/GLOBALFOUNDRIES/](http://semiengineering.com/tag/globalfoundries/))

IBM ([HTTP://SEMIENGINEERING.COM/TAG/IBM/](http://semiengineering.com/tag/ibm/))

INTEL ([HTTP://SEMIENGINEERING.COM/TAG/INTEL/](http://semiengineering.com/tag/intel/))

INTERNET OF THINGS ([HTTP://SEMIENGINEERING.COM/TAG/INTERNET-OF-THINGS/](http://semiengineering.com/tag/internet-of-things/))

IOT ([HTTP://SEMIENGINEERING.COM/TAG/IOT/](http://semiengineering.com/tag/iot/))

IP ([HTTP://SEMIENGINEERING.COM/TAG/IP/](http://semiengineering.com/tag/ip/))

LITHOGRAPHY ([HTTP://SEMIENGINEERING.COM/TAG/LITHOGRAPHY/](http://semiengineering.com/tag/lithography/))

performance than a single core running at high speed. Many problems, and in particular many computation-intensive problems — digital photography, video rendering, database searching, etc. — are readily parallelizable. Moreover, the availability of parallel processing allows designers and software engineers to address larger problems with larger data sets in the same amount of time.

The tradeoff between parallel and sequential processing is still hotly debated in the software and design worlds. (Readers interested in the argument might start [here](http://research.cs.wisc.edu/multifacet/papers/ieeecomputer08_amdahl_multicore.pdf) ([http://research.cs.wisc.edu/multifacet/papers/ieeecomputer08\\_amdahl\\_multicore.pdf](http://research.cs.wisc.edu/multifacet/papers/ieeecomputer08_amdahl_multicore.pdf).) As frequency and voltage scaling have become more difficult, devices with multiple general-purpose cores have proliferated. As power constraints become more severe, though, the fundamental design assumption that silicon area is expensive and should be conserved has been turned on its head.

In an expensive silicon paradigm, it makes sense to design general-purpose logic that can be re-used by many different problems. In the dark silicon era, though, transistors are readily available, but power is very limited. Thus, as UC San Diego professor Michael Taylor [explains](http://cseweb.ucsd.edu/~mbtaylor/papers/taylor_landscape_ds_ieee_micro_2013.pdf) ([http://cseweb.ucsd.edu/~mbtaylor/papers/taylor\\_landscape\\_ds\\_ieee\\_micro\\_2013.pdf](http://cseweb.ucsd.edu/~mbtaylor/papers/taylor_landscape_ds_ieee_micro_2013.pdf)) it, designers can “spend” transistors in order to “buy” power efficiency. For example, a circuit might have many different special-purpose cores that perform one task very efficiently but are dark the rest of the time.

Along those lines, Taylor’s group has [proposed](http://cseweb.ucsd.edu/~mbtaylor/papers/taylor_landscape_ds_ieee_micro_2013.pdf) ([http://cseweb.ucsd.edu/~mbtaylor/papers/taylor\\_landscape\\_ds\\_ieee\\_micro\\_2013.pdf](http://cseweb.ucsd.edu/~mbtaylor/papers/taylor_landscape_ds_ieee_micro_2013.pdf)) “GreenDroid,” a power-optimized approach to the popular Android phone and tablet platform. They found that 43,000 static instructions accounted for 95% of the typical Android device’s workload, and estimate that only 7 mm<sup>2</sup> of silicon in a 45nm process is needed to accommodate those instructions. In place of general-purpose cores, the proposed GreenDroid design uses many different “conservation cores” optimized for specific key functions.

As a design paradigm, this approach is problematic. While the

MENTOR GRAPHICS  
([HTTP://SEMIENGINEERING.COM /TAG/MENTOR-GRAPHICS/](http://semiengineering.com/tag/mentor-graphics/))

MOORE’S LAW  
([HTTP://SEMIENGINEERING.COM /TAG/MOORES-LAW/](http://semiengineering.com/tag/moores-law/))

OPEN-SILICON  
([HTTP://SEMIENGINEERING.COM /TAG/OPEN-SILICON/](http://semiengineering.com/tag/open-silicon/))

QUALCOMM  
([HTTP://SEMIENGINEERING.COM /TAG/QUALCOMM/](http://semiengineering.com/tag/qualcomm/))

RAMBUS  
([HTTP://SEMIENGINEERING.COM /TAG/RAMBUS/](http://semiengineering.com/tag/rambus/))

REAL INTENT  
([HTTP://SEMIENGINEERING.COM /TAG/REAL-INTENT/](http://semiengineering.com/tag/real-intent/))

SAMSUNG  
([HTTP://SEMIENGINEERING.COM /TAG/SAMSUNG/](http://semiengineering.com/tag/samsung/))

SECURITY  
([HTTP://SEMIENGINEERING.COM /TAG/SECURITY/](http://semiengineering.com/tag/security/))

SEMI ([HTTP://SEMIENGINEERING.COM /TAG/SEMI/](http://semiengineering.com/tag/semi/))

SEMICO RESEARCH  
([HTTP://SEMIENGINEERING.COM /TAG/SEMICO-RESEARCH/](http://semiengineering.com/tag/semico-research/))

SOFTWARE  
([HTTP://SEMIENGINEERING.COM /TAG/SOFTWARE/](http://semiengineering.com/tag/software/))

SONICS  
([HTTP://SEMIENGINEERING.COM /TAG/SONICS/](http://semiengineering.com/tag/sonics/))

STACKED DIE  
([HTTP://SEMIENGINEERING.COM /TAG/STACKED-DIE/](http://semiengineering.com/tag/stacked-die/))

STMICROELECTRONICS  
([HTTP://SEMIENGINEERING.COM /TAG/STMICROELECTRONICS/](http://semiengineering.com/tag/stmicroelectronics/))

SYNOPSYS  
([HTTP://SEMIENGINEERING.COM /TAG/SYNOPSYS/](http://semiengineering.com/tag/synopsys/))

SYSTEM-LEVEL DESIGN  
([HTTP://SEMIENGINEERING.COM /TAG/SYSTEM-LEVEL-DESIGN/](http://semiengineering.com/tag/system-level-design/))

TENSILICA  
([HTTP://SEMIENGINEERING.COM /TAG/TENSILICA/](http://semiengineering.com/tag/tensilica/))

TSMC  
([HTTP://SEMIENGINEERING.COM /TAG/TSMC/](http://semiengineering.com/tag/tsmc/))

VERIFICATION  
([HTTP://SEMIENGINEERING.COM /TAG/VERIFICATION/](http://semiengineering.com/tag/verification/))

---

## RECENT COMMENTS

- Todd on First Look: 10nm  
(<http://semiengineering.com/first-look-10nm/#comment-806>)
- striderx on New Challenges For Wearables  
(<http://semiengineering.com/new-challenges-for-wearables-2/#comment-805>)
- Brian Bailey on The Danger of Using Patents  
(<http://semiengineering.com/the-danger-of-using-patents/#comment-804>)
- Sam Appleton

conservation cores can be automatically generated, based on statistical measures of the target workload, the design also needs to be able to dynamically switch between specialized and general-purpose blocks, depending on which tasks a particular piece of software requires. Over-reliance on specialized cores risks creating a **"Tower of Babel" ([http://cseweb.ucsd.edu/~mbtaylor/papers/taylor\\_dark\\_silicon\\_horsemen\\_dac\\_2012.pdf](http://cseweb.ucsd.edu/~mbtaylor/papers/taylor_dark_silicon_horsemen_dac_2012.pdf))** situation, in which a core cannot be used for even closely-related computations, and software developers are slow to adopt new hardware because of the difficulty of programming for it.

Meanwhile, all of these factors conspire to substantially increase hardware complexity and therefore the demands placed on human designers and programmers. And even after all that effort, specialized cores, like general-purpose cores, will only keep the dark silicon problem at bay for so long. Ultimately, the overhead involved in switching between cores will itself consume a substantial fraction of the available power.

### **Can non-volatile cache memory change the game?**

This is where the manufacturing side of the house comes in. In broad terms, any kind of multicore design approach makes use of "power gating." The parts of the chip that are not in use are powered off completely, eliminating both static leakage losses and dynamic power consumption. However, as Takahiro Hanyu and colleagues at Tohoku University explained at December's IEEE Electron Device Meeting (paper #28.2), switching to the "off" state requires a "backup" step to store the logic state to memory, and a "bootup" step to restore it. These operations consume both power and time, while the cache memory used for storage also uses power. DRAMs are difficult to scale to very small dimensions, difficult to integrate with CMOS processes, and require a regular "refresh" operation. **SRAM ([http://semiengineering.com/kc/knowledge\\_center.php?kcid=92](http://semiengineering.com/kc/knowledge_center.php?kcid=92))s**, currently used for on-chip cache memory, are prone to leakage. Existing non-volatile memory technologies, like flash, are too slow and require too much write power.

However, as K. Ikegami and colleagues at Toshiba pointed out (IEDM2014, paper #28.1), non-volatile cache memories don't

(<http://www.ausdia.com>) on The Danger of Using Patents (<http://semiengineering.com/the-danger-of-using-patents/#comment-803>)

- xcore on What's Next For Memory? (<http://semiengineering.com/whats-next-for-memory/#comment-802>)
  - xcore on What's Next For Memory? (<http://semiengineering.com/whats-next-for-memory/#comment-801>)
  - Nandini Jayakumar on (Low) Power Predictions 2015 (<http://semiengineering.com/low-power-predictions-2015/#comment-798>)
  - O Boy on Digital TV: The Need For Speed (<http://semiengineering.com/digital-tv-the-need-for-speed/#comment-796>)
  - FuturePlus on Will There Be A DDR5? (<http://semiengineering.com/will-there-be-a-ddr5/#comment-794>)
  - DarthVaderMentor on An Inside Look At The GlobalFoundries-IBM Deal (<http://semiengineering.com/inside-the-globalfoundries-ibm-deal/#comment-793>)
-

need to store data “forever” in the way that bulk storage components like flash disks do. Rather, they only need to store data for some multiple of the cache refresh rate, which is long enough to ensure that it is no longer needed. According to the Toshiba group, a data retention time of a second or so should be more than adequate for most applications.

This observation creates a potential opening for “less volatile” memory elements, with retention times longer than conventional **DRAM** ([http://semiengineering.com/kc/knowledge\\_center.php?kcid=93](http://semiengineering.com/kc/knowledge_center.php?kcid=93))s or SRAMs, but lower write current requirements than conventional non-volatile memories. In **spin-transfer torque MRAMs** ([http://www.eetimes.com/author.asp?section\\_id=36&doc\\_id=1323466](http://www.eetimes.com/author.asp?section_id=36&doc_id=1323466)), Ikegami explained, the write current depends on the length of the write pulse: smaller cells are slower, but require less power. In processor simulations, write times of 3 to 4ns appear to be fast enough for mobile processor cache access, and offer write currents of less than 45 microamps. Meanwhile, retention time depends on the thermal stability factor,  $\Delta$ , a measure of the stability of the element’s magnetic behavior.

As the Toshiba group showed, thermal stability factors of 70 or more are needed for “permanent” storage, but a  $\Delta$  of only 60 is adequate for cache memory. This level of performance was demonstrated in an MTJ-Last process, where the magnetic tunnel junctions used by STT-**MRAM** ([http://semiengineering.com/kc/knowledge\\_center.php?kcid=95](http://semiengineering.com/kc/knowledge_center.php?kcid=95))s were integrated after CMOS metal fabrication. In benchmark studies, these devices reduced energy requirements by 60% while suffering only a 7% performance degradation relative to SRAM cache memory.

While integrating non-volatile cache memory presents many process challenges, it is conceptually easy to visualize because the circuit logic will behave in the same way and be approachable with the same design tools. But more radical solutions to the dark silicon challenge also have been proposed.

One, to be considered in a future article, is neuromorphic computing. In biological brains, the “retention time” of a

neural pathway depends on the frequency with which it is activated. For instance, you may remember how to get to your childhood home more easily than a restaurant you visited last month. Neuromorphic computing sees the brain — a powerful low-voltage, low-frequency computational system — as, Taylor **wrote ([http://cseweb.ucsd.edu/~mbtaylor/papers/taylor\\_landscape\\_ds\\_ieee\\_micro\\_2013.pdf](http://cseweb.ucsd.edu/~mbtaylor/papers/taylor_landscape_ds_ieee_micro_2013.pdf))**, “an existence proof of highly parallel, reliable, and dark operation,” and a potential model for human-designed systems that avoid the constraints of conventional serial, Boolean logic.

TAGS: [20NM \(HTTP://SEMIENGINEERING.COM/TAG/20NM/\)](http://semiengineering.com/tag/20nm/)

[5NM \(HTTP://SEMIENGINEERING.COM/TAG/5NM/\)](http://semiengineering.com/tag/5nm/)

[ARM \(HTTP://SEMIENGINEERING.COM/TAG/ARM/\)](http://semiengineering.com/tag/arm/)

[DARK SILICON \(HTTP://SEMIENGINEERING.COM/TAG/DARK-SILICON/\)](http://semiengineering.com/tag/darker-silicon/)

[DENNARD'S LAW \(HTTP://SEMIENGINEERING.COM/TAG/DENNARDS-LAW/\)](http://semiengineering.com/tag/dennards-law/)

[DRAM \(HTTP://SEMIENGINEERING.COM/TAG/DRAM/\)](http://semiengineering.com/tag/dram/)

[IEDM \(HTTP://SEMIENGINEERING.COM/TAG/IEDM/\)](http://semiengineering.com/tag/iedm/)

[MEMORY \(HTTP://SEMIENGINEERING.COM/TAG/MEMORY/\)](http://semiengineering.com/tag/memory/)

[MOORE'S LAW \(HTTP://SEMIENGINEERING.COM/TAG/MOORES-LAW/\)](http://semiengineering.com/tag/moores-law/)

[MRAM \(HTTP://SEMIENGINEERING.COM/TAG/MRAM/\)](http://semiengineering.com/tag/mram/)

[NEUROMORPHIC COMPUTING \(HTTP://SEMIENGINEERING.COM/TAG/NEUROMORPHIC-COMPUTING/\)](http://semiengineering.com/tag/neuromorphic-computing/)

[RETENTION TIME \(HTTP://SEMIENGINEERING.COM/TAG/RETENTION-TIME/\)](http://semiengineering.com/tag/retention-time/)

[SPIN-TRANSFER TORQUE MRAM \(HTTP://SEMIENGINEERING.COM/TAG/SPIN-TRANSFER-TORQUE-MRAM/\)](http://semiengineering.com/tag/spin-transfer-torque-mram/)

[SRAM \(HTTP://SEMIENGINEERING.COM/TAG/SRAM/\)](http://semiengineering.com/tag/sram/)

[TOSHIBA \(HTTP://SEMIENGINEERING.COM/TAG/TOSHIBA/\)](http://semiengineering.com/tag/toshiba/)

[UC SAN DIEGO \(HTTP://SEMIENGINEERING.COM/TAG/UC-SAN-DIEGO/\)](http://semiengineering.com/tag/uc-san-diego/)



**Katherine Derbyshire** (all posts)

(<http://semiengineering.com/author/katherine/>)

Katherine Derbyshire is a technology editor at Semiconductor Engineering.

**Comments** **Community** **Login** ▾

Sort by Best ▾

Share  Favorite 

Be the first to comment.

---

## SEMICONDUCTORENGINEERING (1)

### ABOUT

About us (<http://semiengineering.com/corp>)

Contact us (<http://semiengineering.com/corp>)

Advertising on SemiEng (<http://semiengineering.com/marketing>)

### NAVIGATION

Homepage (<http://semiengineering.com/>)

System-Level Design (<http://semiengineering.com/category-main-page-sld/>)

Low Power-High Perf (<http://semiengineering.com/category-main-page-lphp/>)

Manufacturing Des & Test (<http://semiengineering.com/category-main-page-manufacturing-design/>)

IoT & Security (<http://semiengineering.com/category-main-page-iot-security/>)

Knowledge Centers (<http://semiengineering.com/kc>)

Startup Corner (<http://semiengineering.com/startup-corner>)

Bus & Marketing Strategies (<http://semiengineering.com/bussiness-marketing-strategies/>)

### CONNECT WITH US

Facebook (<https://www.facebook.com/SemiEngineering>)

Twitter ([@SemiEngineering](http://www.twitter.com/SemiEngineering))

LinkedIn (<http://www.linkedin.com/company/semiconductor-engineering>)

YouTube (<http://www.youtube.com/user/SperlingMediaGroup>)

---

Copyright ©2013-2015 SMG | **Terms of Service** (<http://semiengineering.com/terms-of-service/>) | **Privacy Policy** (<http://semiengineering.com/privacy/>)