http://www.eejournal.com/archives/articles/20131105-hardceiling/

© by Kevin Morris November 5, 2013 (bold marked by Reiner)

The Hard Ceiling

Power Plays the Death Card

Moore's Law is a maddening mistress. As our engineering community has collectively **held the tail of this comet** for the past forty-seven years, we've desperately struggled to divine its limits. Where and why will it all end? **Will lithography run out of gas,** brining the exponential curve of semiconductor progress to a halt? Will packaging and IO constraints become so tight that more transistors would make no difference? Or, will economics **bring the whole house of cards crashing down** - putting us in a situation where there is just **no profit in pushing the process envelope?**

These are still questions that keep many of us employed - predicting, prognosticating, and pontificating from our virtual pedestals - trying to read the technological tea leaves and triangulate a trend line that will serve up that special insight we seek. We want to know the form of the destructor. When the exponential constants of almost fifty years **make a tectonic shift** and our career-long assumptions change forever, we'd appreciate some forewarning. We want to **look the end of an era** in the eye.

I am now ready to make a call.

When the grim reaper rides in and swipes his sickle through the silky fabric of fifty years of unbridled technological progress - the likes of which has never been experienced in the history of the human race - the head beneath the hood will **be the flow of coulombs through the catacombs.**

It is power that will be our ultimate undoing.

As transistors get ever smaller, we can constantly cram more of them into the same physical space. However, for the most part, those physical spaces are not changing. The size of a smartphone has varied slightly over the years, but it isn't likely to change significantly in the foreseeable future. That, in turn, means that **the size of a battery isn't likely to expand dramatically**, and we all know that battery density is not on any progress curve that even remotely resembles Moore's Law. Therefore, the amount of power available for smartphone-related activities will remain fairly flat. We can shove in as many logic gates as we like, but they'll have to divvy up the same amount of power they use today.

On the other end of the scale, **server farms are packed into buildings at the maximum density** for which the local utilities can supply the enormous amounts of juice required to feed **the ravenous appetites of tens of thousands of von Neumann machines** whirring away at multi-gigahertz speeds, along with **massive amounts of memory needed** to deal with the data explosion produced by those prolific processors. Then, another good part of the power budget is used to run **giant air-conditioning** systems that remove the excess heat produced by all that gear. The bottom line? No matter how many transistors we can stuff into the building, they'll still be competing for the same size power mains coming in from the electric company.

In fact, at just about every scale you can name, the hard ceiling at which our design exploration space ends is the total power budget. Whatever we design to make our machines do more work cannot increase the power consumed. Therefore, as system engineers, we are no longer in the performance business.

We are in the efficiency business.

With planar CMOS processes, each passing technology node has made us more efficient with dynamic power consumption (the power each transistor uses doing actual work) primarily because of lower supply voltage requirements. However, with each of those improvements has come **an increase in the amount of leakage current** that flows through each transistor when it is not operating. Thus, over time**, leakage current has become the boogeyman**, while active current has tracked Moore's Law improvements.

FinFET (or, in Intel parlance, "Tri-Gate") technology has pushed our power problems back a bit - delivering the most toggles for the fewest watts we've ever seen. However, even with that impressive discontinuity - computational power efficiency is not improving as fast as semiconductor transistor density. The upshot of it all? We can add as many transistors to our design as we want, we're just not going to have the power to use them all. Think of the world's population of transistors more than doubling every two years, but the supply of power to feed them remaining fixed. Eventually, we're going to have a time of famine.

In a way**, the end of Moore's Law could be refreshing.** It might even re-invigorate the art of engineering itself. The tidal force of exponential progress over the decades has been so great that it overwhelmed most clever or subtle advances. After all, what's the point of spending a couple of years with a novel idea that will gain your team 20-30% when a process node leap in the same time frame is going to double your competition's performance? Moore's Law is a "go big or go home" proposition, where the giant companies with the resources to pour tens to hundreds of millions of dollars into bringing their design to the next process node will always win out over more clever but smaller efforts that can't stay on the bleeding edge of semiconductor technology. **If Moore's Law is no longer the trump card,** the playing field for innovation could be leveled once again, and we could see a renaissance of invention replace the current grind of spend, shrink, and pray.

The first target for this post-apocalyptic revolution should be our computing architecture itself. **von Neumann machines are** very efficient in their use of gates, but **very inefficient in their use of power** for each computation. When gates were precious and power was cheap, that was a great tradeoff. However, with gates now being almost free, and power being the scarce resource, the venerable von Neumann may have run his course.

Vanquishing von Neumann, however, will not be an easy undertaking. ***A half-century of software* sits squarely on top of the sequential assumptions imposed by that architecture.*** While the computationally power-efficient future clearly is based on parallel processing and datapath structures, **the means to move existing software to such platforms has been elusive at best.** Taking something like plain-old legacy C code and efficiently targeting a heterogeneous computing machine with multiple processors and some kind of **programmable logic** fabric is the job of some **fantastic futuristic compiler** - the likes of which we have only begun to imagine. This non-existent compiler is, however, the key to **unlocking a quantum leap in computing efficiency** - allowing us to take advantage of a much larger number of transistors and therefore to create an exponentially-faster computer without hitting the aforementioned power ceiling.

The engineering for this new era will most likely not be in the semiconductor realm. While there are certainly formidable dragons to slay - double and triple patterning, EUV, nanotubes and more - software and compiler technology may take an increasing share of the spotlight in the next era. **It will be fascinating to watch.**

*) also see http://hartenstein.de/EIS2/#mistake for the **Reconfigurable Computing Paradox** see http://xputer.de/RCpx/