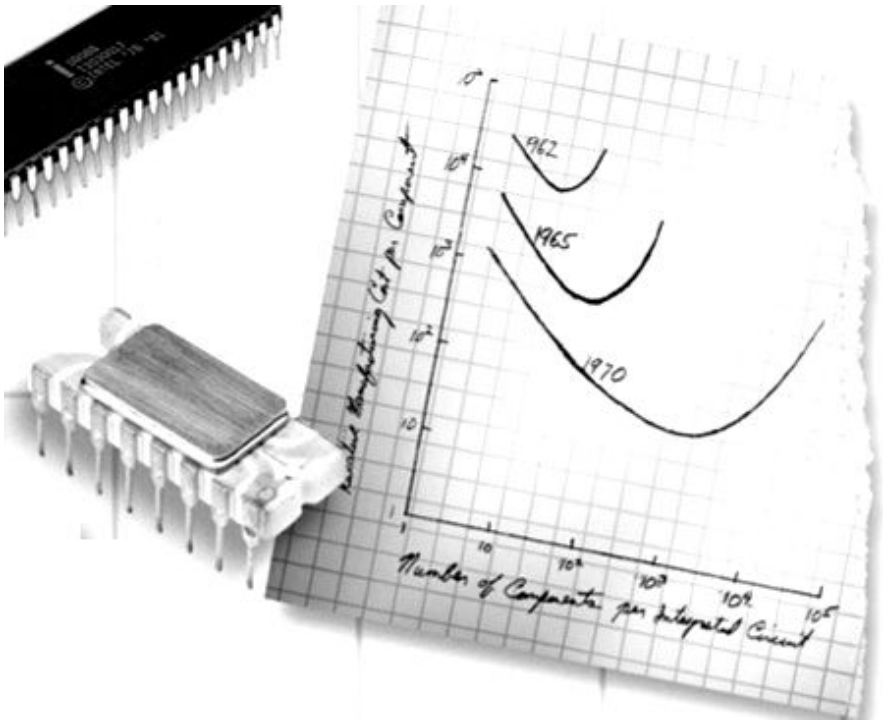
# TSMC THINKS IT CAN UPHOLD MOORE'S LAW FOR DECADES

September 13, 2019 Michael Feldman

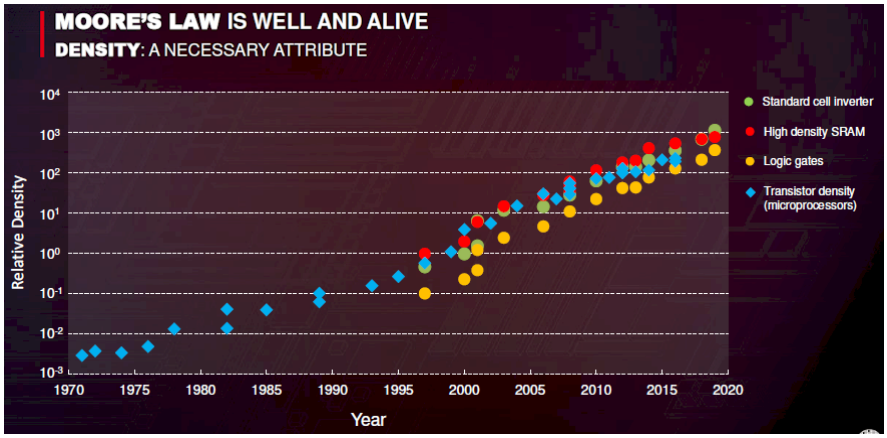



If you thought the gang at Intel were Moore's Law biggest devotees, you probably haven't heard Philip Wong expound on the subject. Wong, who is vice president of corporate research at Taiwan Semiconductor Manufacturing Corp, gave

ference where he and well, but with main viable for the

"It's not dead," he told the Hot Chips attendees. It's not slowing down. It's not even sick."

In Wong's telling, the only thing that matters to sustain Moore's Law is to keep improving density. And while he admitted that clock speeds have plateaued with the death of Dennard scaling, it is transistor density that will enable better performance and energy efficiency.



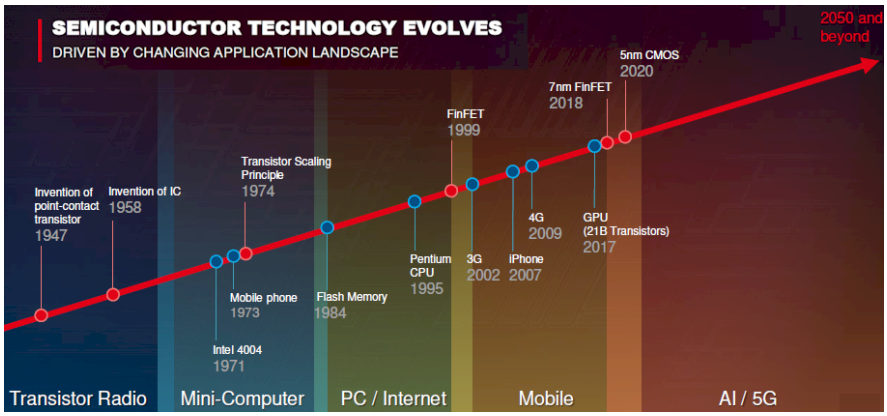

Ultimately, it doesn't matter how those higher densities are achieved. According to Wong, as long as companies can keep delivering more transistors in smaller spaces with better energy efficiency, that's all that matters. In the short-term, this will likely be achieved the old-fashioned way, namely by improving CMOS process technology so that transistors with smaller gate lengths can be manufactured.

TSMC is currently etching 7 nanometer transistors and is on the way to 5 nanometers. Wong said the design ecosystem of the 5 nanometer node is ready now and they have already begun risk production – that is, the process node and design tools are complete and it is producing viable wafers. On its last earnings call, TSMC said it plans to start volume production of 5 nanometer chips in the first half of 2020. Apparently, there is even a 3 nanometer node in the pipeline at TSMC.

ling planar chips
ll eventually stop
ling, we're down
e'll run out of
atoms," he explained.

But that doesn't mean the end of density. He noted there have been a number of innovations in semiconductor manufacturing that kept density on an upward curve even after Dennard scaling ended. In particular, the use of strained silicon and high-k metal gate technology came along, followed by FInFet (Fin Field Effector Transistor), which introduced 3D structures. And now a technology known as DTCO (Design Technology Co-Optimization) is being explored to push transistors below 7 nanometers.

All of these innovations took place because new computing platforms had to be developed for applications that demanded faster and more energy-efficient hardware. That evolution spanned minicomputers in the 1970s, the PC in the 1980s, the Internet in the 1990s, and now mobile computing. Each one drove demands for greater density through refinements in semiconductor manufacturing. Wong thinks the next big push will come from artificial intelligence and 5G.




So what innovations are going to be required to keep the Moore's Law party going?

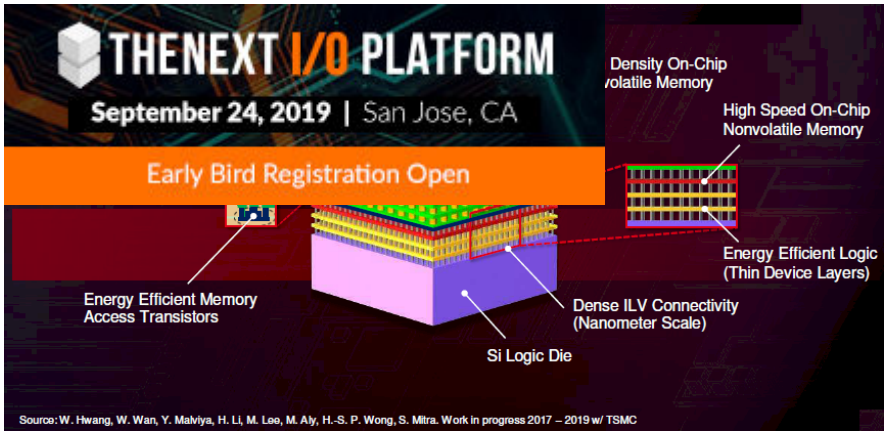
In the near-term, the use of chiplets to build multi-chip packages in 2.5D structures will increase overall compute and memory density, even if the chips themselves don't getting any denser. Wong said this is already making the

important than
n the same

TSMC has its own version of 2.5D packaging with its Chip-on-Wafer-on-Substrate (CoWoS) technology. (Intel's Embedded Multi-die Interconnect Bridge, shortened to EMIB, is a competing packaging technology.) CoWoS enables multichip packages to be constructed by mounting chiplets and suitable memory devices on top of a silicon interposer and connecting them using through-silicon vias (TSVs). For readers of this publication, the most noteworthy CoWoS implementation is Nvidia's Tesla V100 GPU accelerator, which packages a GV100 GPU with high bandwidth memory (HBM) modules. But more extensive integration, with larger numbers of chiplets is on its way from Intel, AMD, and Xilinx.

But 2.5D will only take you so far density-wise. A more scalable solution will require a true 3D packaging technology. For that, Wong said our best bet is N3XT, a 3D monolithic design based on new nanomaterials and fine-grained integration of memory and logic. N3XT, which stands for Nano-Engineered Computing Systems Technology, has been kicking around academia since 2015, but with companies like TSMC apparently taking it seriously, it has a decent chance of being commercialized.

Wong threw up a slide showing a schematic for what such a device might look like. It's comprised of layers of energy-efficient logic (yellow), high speed memory (red), and high capacity non-volatile memory (green), stacked together in an interleaved fashion. All of this sits atop a conventional silicon logic die (purple).




The key to this is connecting these different components with something called ILV, which stands for Inter-Layer-Via. Unlike a micron-scale TSV, an ILV can be formed at nanoscale dimensions. Although this is an extremely important piece of the N3XT technology, Wong didn't have much to say about it. Apparently though, ILV is something TSMC has been working on and has a number of patent applications around it.

For these 3D packages, interleaving memory and logic is rather important because it reduces the distance between them, which will make it possible to deliver the kind of high bandwidth, low latency communication that will be needed for applications like AI and 5G. With CMOS, interleaving memory with the logic is not possible because logic transistors need temperatures of around 1,000 degrees C to be etched properly, which would destroy the adjacent components during manufacturing. Rather, you need something that can be laid down at around 400 degrees C.

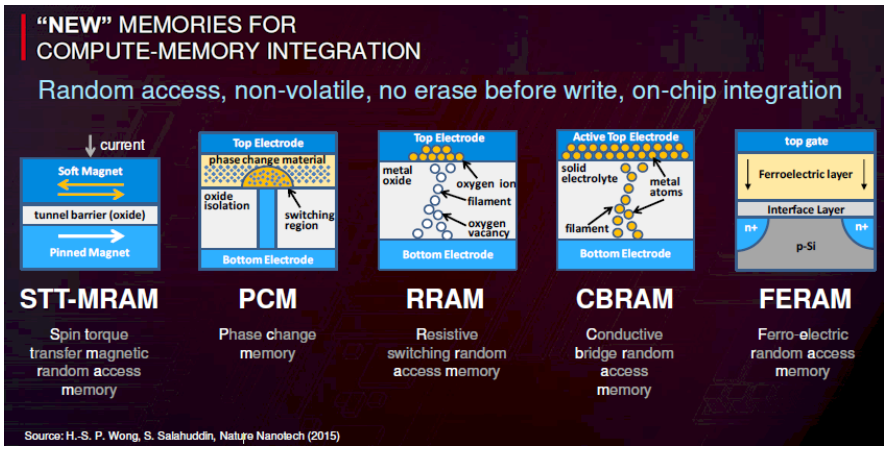
It just so happens that there are new materials that have been researched over the last several years that appear to be suitable for manufacturing high performance transistors at relatively low temperatures. Unlike the bulk silicon-based materials that are used for semiconductors today, they are transition metal dichalcogenides (TMDs), which are based on elements such as molybdenum, tungsten, and selenium.

TMD materials also display high carrier mobility – that is, the ability to move electrons through them easily – but with thin channels. Conveniently, those are the attributes you want if you're building transistors smaller than two to three

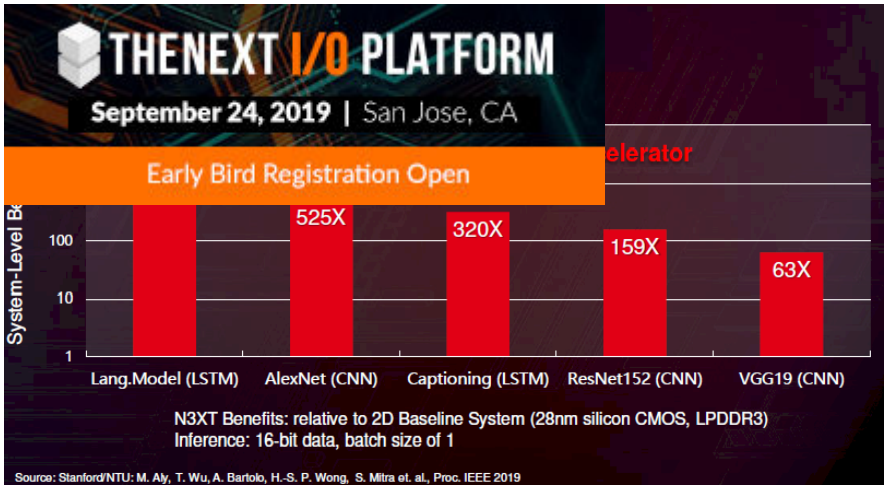
nental TMD wafers

nanotubes. Experimental versions have been manufactured that demonstrate good semiconductor behavior, said Wong. In fact, working prototypes of carbon nanotube–based logic and SRAM devices have been built, including, most recently, a RISC-V implementation by MIT researchers.

On the memory side, Wong said the most likely new candidates for 3D integration are spin torque transfer magnetic RAM (SST–MRAM), phase change memory (PCM), resistive RAM (ReRAM), conductive bridge RAM (CBRAM) and ferro–electric RAM (FERAM). All of them have the key attributes of being random access, non–volatile devices that don't require an erase before a write. Some of these are already commercially available, including Everspin's MRAM, Samsung's embedded MRAM, Crossbar's ReRAM, and Intel's 3D XPoint (which most believe is a variant of PCM).




Researchers have simulated the performance of N3XT devices and compared them against traditional two–dimensional chips similarly configured with regard to logic and memory capacities. According to their studies, using a variety of machine learning inference benchmarks, the N3XT devices were between 63X and 1,971X as efficient as their 2D competition.




All of which sounds promising. But Wong didn't elaborate on exactly how these technologies will sustain a Moore's Law rate of improvement for the next 30 years. Switching to new nanomaterials, for example will essentially provide a one-time bump in transistor density with regard to the 2D components. Eventually you'll run up against the atomic limits here as well.

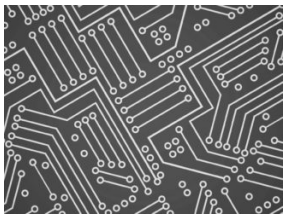
Theoretically, if you could double the stack height of 3D devices every 18 months, you could at least achieve the density improvement from a real estate perspective. Of course, for mobile and other embedded devices this is going to get unwieldy rather quickly, and even for datacenter computers, it will only take seven or eight generations to bump up against a 12-foot ceiling.

For this to work for multiple decades, other technology innovations that weren't mentioned in Wong's talk will have to be developed to keep densities on a Moore's Law trajectory. But if you're a chip manufacturer like TSMC, you have to believe that researchers will be delivering a steady stream of such candidate technologies that just need the impetus of new, more demanding applications to spur commercialization. And if history is any guide, those applications are certain to come.

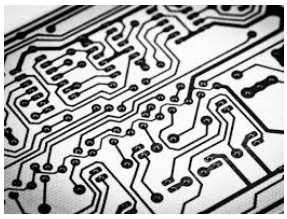
he week directly from

## RELATED ARTICLES

THE DATACENTER IMPACT OF THE GLOBALFOUNDRIES 7 NANOMETER SPIKE



GIVING WAFERSCALE PROCESSORS ANOTHER SHOT



DESIGNING CUSTOM CHIPS IN-HOUSE IS THE NEW NORMAL

## 5 COMMENTS

**gary** says:

SEPTEMBER 13, 2019 AT 6:18 PM

I don't follow where "seven or eight generations to bump up against a 12-foot ceiling." comes from. He is talking about atomic dimensions going 3D. You can double atomic dimensions for a very long time before you even see them.

REPLY

**Bryan Lagos** says:

SEPTEMBER 14, 2019 AT 3:00 AM

Yes, I was confused too. I thought I missed something. Perhaps this is just humor?

REPLY

**Stan Ford** says:

SEPTEMBER 13, 2019 AT 7:20 PM

was asked on how and Resistive RAM e, had actually world technology integration. This question was artfully dodged. Did the journalist check on how many of prior "contributions" of Wong and Mitra (leaving aside their Stanford banner for a moment) actually made any impact to the world we know ?

REPLY

**DSavic** says:

SEPTEMBER 14, 2019 AT 4:44 AM

If we strictly follow Wikipedia definition: "Moore's law is the observation that the number of transistors in a dense integrated circuit doubles about every two years" then I would agree with Dr. Phillip Wong . However, industry progress was really driven by the fact that every two years (or 18 months, as per Intel's executive David House) we doubled number of transistors and chip performance, keeping more–less the same power and cost. That was the main driving force behind industry success in the last few decades, not just transistor's density. That statement stopped being true a decade ago. I have no doubt that TSMC, Intel and other companies will keep inventing new processes (or improving existing ones), there will be new materials, etc, but those days of doubling performances FOR A GIVEN POWER AND COST are definitively gone. The industry needs new driving force and that is (chip–level and system –level) architecture change, driven by more efficient use of transistors (accelerators vs gen purpose) and data movement–related energy reduction, combined with advanced packaging and new materials.
I see this being, in one way or another, covered in Dr. Wong's presentation, but without clear distinction, which can confuse people to believe Moore Law still works.

REPLY

ut incorporating the economics aspect are being intentionally disingenuous – it's marketing sleight of hand. Fabrication is expensive and getting more so all the time. Density, cost and volume – you need all 3. Density at too high of a cost means reduced volume and an unprofitable node.

REPLY

## LEAVE A REPLY

Your email address will not be published.

Comment

Name *

Email *

Website

POST COMMENT

This site uses Akismet to reduce spam. Learn how your comment data is processed.

**TSMC THINKS IT CAN UPHOLD MOORE'S LAW FOR DECADES**

**KEEPING PACE WITH RAPID STORAGE SOPHISTICATION**

**NETWORKING NEEDS MORE THAN INCREMENTAL INNOVATION**

**WHEN IT COMES TO FILE SERVICES, THE CLOUD IS CATCHING UP**

## ABOUT

*The Next Platform* is published by Stackhouse Publishing Inc in partnership with the UK's top technology publication, *The Register*.

It offers in-depth coverage of high-end computing at large enterprises, supercomputing centers, hyperscale data centers, and public clouds. Read more…

## SIGN UP TO OUR NEWSLETTER

Featuring highlights, analysis, and stories from the week directly from us to your inbox with nothing in between.

SUBSCRIBE NOW

## RECENT ARTICLES