Moore's Law: Where are we and which way are we going?

Greg Yeric greg.yeric@arm.com

Fellow

ARM Research



Moore's Law at ARM

ARM Research Future Silicon Technology Group

- Predictive technology modeling
- Disruptive technologies: Assessment, Incubation
- Future hardware resilience





Has cost per transistor scaling reversed?



Or is everything just fine?

Answer:

Cost/transistor still going down post-28nm, but not as straightforward as you might think, so it deserves an asterisk and some discussion



Cost Reduction On Track



Standard Cell scaling...proper shrink



Fig. 2 A: 9T N14 AOI showing pin access points in green, B: 9T AOI in N10 showing triple-patterned M1 in blue, red, brown, C: 7.5T N10 AOI showing that adequate pin access can be preserved.

B" node "N+1" Proper scaling 0.7x linear, But if that hits a cliff in litho costs, might want to try something smarter

A: node "N"

L. Liebmann et al, VLSI, p.112 (2016).

Standard Cell scaling: Affordable shrink...



...and we'll make it up somewhere else



Standard Cell scaling... with track reduction a.k.a. fin depopulation



Can get to desired transistor density at less than 0.7x linear shrink, if we can take tracks (height) out of the standard cells. This usually comes at the cost of few fins per cell (black)

Standard Cell scaling... 4, 3, 2 fins per transistor





OK, so now what?

Fin depopulation strategy quickly comes to an abrupt end. Must look for other tactics.



Fig. 5 A: grating-and-cut compliant 6T N7 cell, B: N5 cell maintains unidirectional layout style even with the arrival of EUV lithography, C: N5 with over-scaled M0 as proposed by IMEC.

L. Liebmann et al, VLSI, p.112 (2016).

N7 to N5 Scaling

N5 is emerging to be a bitter-sweet node: the delay in the much anticipated transition to EUV forces layouts to remain unidirectional, Fig. 5B. Additionally, the critical pitches bump up against hard 'non lithography' limits for poly and fin pitch scaling as was shown in Table 1. Compensating for modest pitch scaling through further cell height reduction requires substantial and risky process innovation such as the ability to contact poly-gates over active fins. While these solutions are being actively explored, alternatives such as over-scaling only the lowest level of metal, Fig. 5C, as proposed by researchers at IMEC [2], could be called upon to maintain an overall 0.56x area scale factor for one more finFET node.

As traditional pitch-based Moore's Law advancement wanes, process technologists are pursue "scaling boosters" to make up the slack. Analogous to design accelerators.



Example "scaling booster": Self-Aligned gate contact

One potential density improvement : Enabling gate contacts (in red highlighted area) to extend over active fins without contacting the S/D



The majority of wafer cost increase is in the patterning of conventional wires Previously laughable ideas for transistor integration might become rational options Other scaling booster options: single diffusion breaks, fully self-aligned vias, super vias → Requires design-technology co-optimization (DTCO). Routabililty, e.g.

A word of caution on cost prediction math

7nm steppers will be 50% faster than 28nm steppers

- > 250 wafers per hour = < 15 seconds per wafer</p>
- Chuck moves > Im/second (develops > 10G)
- 3 nm X and Y accuracy*

*drop England onto the globe, align it to a precision one brick, 5 times per second

are your circuits worthy?

Continuous improvement from all suppliers: resist, light source, mask, etch....

Contraction of the second seco



Cost per transistor

- Lithographic pitch scaling as easy path for Moore's Law has passed*
 - Heroic efforts throughout the ecosystem are mitigating the rate of slowing
 - *5nm with a good EUV would be a welcome exception. (3nm may require EUV double patterning)
- "scaling boosters" required to make up the slack (ex. fin depopulation)
 - Future scaling boosters could get 5nm to desired cost, might get us to 3nm**
 - Self-aligned gate contact, fully self-aligned via, super via, single diffusion break **Without breaking physical design. Design-breaking boosters exist, and may be required.
 - We move from long-lived lithographic scaling paradigm to two-node and even one-time boosts (quicker, bigger steps)
 - Managing this complexity will be key to getting to 3nm on time and on budget
 - Requires thorough benchmarking (DTCO) to down-select and find the best path

Moore's Law and Dennard Scaling

From Moore's 1965 paper

"shrinking dimensions on an integrated structure makes it possible to operate the structure at higher speed for the same power per unit area"

	Table 1Scaling Results for	This doesn't happen automatically anymore.
	Device or Circu Device dimensio Doping concentr Voltage V Current I Capacitance $\epsilon A/\epsilon$	From 28nm, we are often in the Reverse Dennard Scaling era: Smaller gate pitch can often mean slower (and larger!) chips PPA improvement? Surepick any 2. Pressure to bifurcate processes becomes stronger
	Delay time/circu Power dissipation Power density V	it VC/I $1/\kappa$ n/circuit VI $1/\kappa^2$ I/A 1





After 7nm: lots of options, no clear path



ARM

Example of the materials revolution: 2D materials



Graphene opened the Pandora's box of 2D materials.



BEOL Scaling summary:



http://www.zyvexlabs.com/EIPBNuG/2005MicroGraph.html

Half of your performance and power is going here

- Future transistors will struggle with severe parasitic load
 - \rightarrow we're already in reverse-Dennard scaling
- Low-k (airgap), low-R (cobalt, ruthenium): mitigate, but don't solve
- Problems are not just lateral, but more and more vertical (Via R)
- Power/Performance scaling may be more challenged than cost scaling



Beyond transistors and wires

Two technologies that could really help the system roadmaps:

- Find a better memory
 - PCM, RRAM, MRAM: Nothing yet is a "super" memory. (from an on-die cache perspective. Storage class memory is great for systems, but another topic)
 - Endurance will be a key system factor (i.e., can we work with a memory less robust than SRAM)
 - Non-volatility is a big wild card, both in traditional cache hierarchy and power states
- Utilize 3DIC
 - Look at additional partitioning options to optimize PPAC.
 - There is a wide variety of present and future technologies to choose from.
 - Future technologies need design tools (chicken and egg problem)



The 3DIC Roadmap: homogenous SiP to transistors

	3D-SIC		3D-SOC		3D-IC	
3D-Wiring level	Global	Semi-global	Intermediate	Local	FEOL	
Partitioning	Die	blocks of stan	dard cells	Standard cells	Transistors	
3D Technology	Die stacking Die-to-Wafer stacks Die-to-Si-interposer	Parallel FEOL	wafer processing afer bonding	Sequential FEOL processing Active layer bonding or deposition		
2-tier stack schematic				2 nd FEOL after stacking	Multi-tier FEOL	
Characteristic	Known Good Die 3D stacks or Si-interposer stacking	BEOL between 2 FEC Overlay 2 nd tier defined by W2W alignment/bonding		DL layers Overlay 2 nd tier defi alignment	FEOL/FEOL stack ned by litho scanner	
Contact Pitch <i>Relative density:</i>	$\begin{array}{c} 40 \Rightarrow 20 \Rightarrow 10 \mu m \Rightarrow 5 \mu m \\ \frac{1}{16} \Rightarrow \frac{1}{4} \Rightarrow 1 \qquad \Rightarrow 4 \end{array}$	5 ⇒1µm 4 ⇒100	2 µm ⇒ 0.5 µm 50 ⇒ 400	200 ⇒ 100 nm 5000 ⇒ 10000	< 100 nm > 10000	

E. Beyne, IEEE Design & Test, May/June 2016



heterogeneous SiP: "3D-SOC" \rightarrow partitioning



E. Beyne, IEEE Design & Test, May/June 2016

Block-level partitioning: 3D-SoC optimizes PPAC



#ARMTechCon

ARM

TechCor

High density 3DIC

Tools that can bond wafers at this density are in the pipeline. This would allow folding inside the cores themselves



E. Beyne, IEEE Design & Test, May/June 2016

Gate-level partitioning: 3D-IC

ARM / Georgia Tech study to be published at ICCAD. There's a benefit, but is it a large enough one time bump to justify?





High density 3DIC at the transistor level



E. Beyne, IEEE Design & Test, May/June 2016



Pressure on R&D costs in the industry



Industry has its hands full with 5nm

The needs and complexities of solutions for 3/2 nm solutions will be unprecedented



Future technology pipeline



Research

ARM

TechCon

Examples of ARM activities along this theme:



ARM Cortex[™]-M0 DesignStart



> 150 university downloads!

ARM

[echCor

IP and Development Boards

ARM has announced the availability of the ARM® Cortex™-M0 DesignStart Processor via the <u>ARM DesignStart™ online IP portal</u> as a download after a simple click-through EULA. The processor is configured as a synthesizable, obfuscated verilog netlist designed for academics, start-ups and ad-hoc technology teams looking to teach or prototype with a real ARM processor. Implementation is possible in almost any FPGA to enable SoC design courses and projects around a real, instruction set compatible ARM core. It can also be manufactured according to the EULA. From there, anything from embedded systems courses or microprocessor applications can be taught using the same platform. Through this online access model, ARM will accelerate the proliferation of ARM technology in university curriculums and research projects.

The Official ARM Cortex™-M0 DesignStart Example Design Kit (EDK) is Now Freely Downloadable!

The ARM Cortex-M0 DesignStart Example Design Kit (EDK) helps to start designing complete systems (hardware + software) with the Cortex-M0 DesignStart Processor on an FPGA board. It includes a selection of simple AHB-Lite peripherals, AHB-Lite bus infrastructure components, example systems usable as design templates, software examples based on the popular Keil MDK-ARM software development tools, and associated quickstart guides and documentation. It's everything you need to get an example system based on the M0_DS running out-of-box in a Xilinx-based Digilent Nexsys3 FPGA board!

https://www.arm.com/support/university/researchers/soc-fpga/index.php



2015: ARM joins Semiconductor Research Corporation

Semiconductor Research Corporation	Search 👂					
Home My SRC	Programs	Liaisons	Student Center	Calendar	Librar	
Home » SRC » Member » Roster SRC Members GRC STARnet NRI						
Advanced Micro Devices, Inc.		Analog D	evices, Inc.			
ARM, Incorporated	Freescal	Freescale Semiconductor, Inc.				
GLOBALFOUNDRIES		IBM Corp	IBM Corporation			
Intel Corporation	Mentor 0	Mentor Graphics Corporation				
Microsoft	Mubadal	Mubadala Technology				
Qualcomm	RTI Inte	RTI International				
Texas Instruments Incorporated	Tokyo El	Tokyo Electron Limited				

GRC THRUSTS

AMS-CSD

Analog/Mixed-Signal Circuits, Systems, and Devices

CADT

Computer-Aided Design and Test

EP3C

Efficiency and Performance for Connectivity Constrained Computing

ESH Environment, Safety, and Health

13T Innovative and Intelligent Internet of Things

LMD

Logic and Memory Devices

N M P Nanomanufacturing Materials and Processes

P K G Packaging

SemiSynBio

Semiconductor Synthetic Biology

SLD

System Level Design

T 3 S Trustworthy and Secure Semiconductors and Systems

GRC CENTERS

A C E 4 S ATIC-SRC Center of Excellence for Energy Efficient Electronics Systems

CAIST

The New York Center for Advanced Interconnect Science and Technology

EBSM

SRC Engineering Research Center for Environmentally Benign Semiconductor Manufacturing

T x A C E Texas Analog Center of Excellence



2nd Annual ARM Research Summit: Sept 11-13, 2017

ARM Research Summit Computing for the next decade

15-16 September 2016 | Churchill College, Cambridge, UK

Over 70 talks

Example "Moore's Law" area papers:

Scaling and Variability of Multi-Gate Non-Planar III-V	Karol Kalna	Swansea University
Optical Computing: A technology Whose Time Has Come?	C David Wright	University of Exeter
Design Technology Co-Optimisation in Advance Technology Nodes	Asen Asenov	University of Glasgow
3D IC Revolution: Where Are We Now, and What is Next?	Sung-Kyu Lim	Georgia Tech



Broad ecosystem DTCO: ARM / ASML collaboration



2016: ARM joins IMEC Logic INSITE program



ARM is contributing by helping IMEC improve benchmarking of "scaling boosters"



2016: ARM joins Stanford SystemX Alliance

Stanford ENGINEERING SystemX Alliance

Bio Interfaces: Transformative technologies for healthcare | LEAD: Prof. Shan Wang and Prof. Tom Soh | DETAILS

- Design Productivity: Improving critical area productivity | LEAD: Prof. Mark Horowitz | DETAILS
- Energy/Power Management Systems: Efficient energy power supply | LEAD: Prof. Juan Rivas-Davila | DETAILS
- Heterogeneous Integration: Bridging everything into anything | LEADS: Prof. Eric Pop and Prof. H.S. Philip Wong | DETAILS
- Internet of Everything: Universal connectivity of devices and systems | LEAD: Prof. Amin Arbabian and Prof. Thomas Lee | DETAILS
- Photonic and Quantum Technologies: More efficient devices and systems; optics LEAD: Prof. Jelena Vuckovic | DETAILS
- Computation for Data Analytics: Optimizing application, architecture & technology LEAD: Prof. Chris Ré and Prof. Subhasish Mitra | DETAILS

To be presented at IEDM (December)





Fig. 1. Integrated design flow for technology modeling and VLSI system implementation.

Fig. 2. FET device structures of (a) FinFET, (b) NWFET, and (c) 2D-FETs. All labeled numbers are in the unit of nm. The Fin and NW widths are both 5 nm.



device candidates in a full core

design context, to help get more

accurate projections of value



ASU ASAP 7nm predictive PDK



2.

3.

4.

5.



University eNVM incubation project

Find a better memory

Too early to understand if this particular material can be successfully integrated (and the road is littered with well-intentioned NVM development projects), but this is a good example of the materials-to-design expediting theme. Perhaps tune in to TechCon 2017 to find out.

- PCM, RRAM, MRAM: Nothing yet is a "super" memory. (for on-die cache. SCRM is great, but another topic)
- Endurance will be a key system factor (i.e., can we work with a memory less robust than SRAM)
- Non-volatility is a big wild card, both in traditional cache hierarchy and power states
- Utilize 3DIC
 - Look at additional partitioning option
 - There is a wide variety of present and
 - · Future technologies need design to

Lab-to-fab proof of concept

material to design: what a company like ARM can provide:

- Design domain knowledge
- Test chip design / test / analysis





Summary

- Traditional Moore's Law cost reduction is slowing down
 - Broad ecosystem progress plus scaling boosters can achieve 5nm (3nm?) with manageable disruption
 - Slowing of traditional scaling means more radical scaling options will have opportunity
- Power/performance scaling may be more difficult than cost (transistors + wires)
 - Scaling obstacles will be increasingly require novel integration, materials, devices, and physics
- Heterogeneity: challenges and opportunities
 - Different materials for NMOS / PMOS, LVT / HVT, local wires vs. global wires. Varying pitches.
 - $3D-SiP \rightarrow 3D-SOC \rightarrow 3DIC$: Huge design challenges. DTCO \rightarrow STCO
 - Novel memories, photonics, (3D-SOC, moving to 3D-IC), other wild cards
- Continued progress will require managing the rate and level of change
 - Increasing "top to bottom" DTCO including semiconductor tool manufactures, EDA, etc.
 - Accelerate novel technology from universities and research consortia: "lab to fab"
 - Benchmarking is a key topic. DTCO \rightarrow STCO



Among the topics I did not talk about:

- The IoT:
 - Reducing cost/function (more Moore) vs. enabling new functions (more than Moore)
 - Moore's Law progress enabled the Internet of Things
 - Vanishing cost of compute, plus reasonable cost of big compute (data analytics)
 - Ubiquitous internet, and vanishing cost to connect
 - Children of Moore's law: Energy harvesting, sensors, printed electronics
 - These new capabilities are heavily leveraging the progress of the semiconductor industry
 - I0um to nm progress in 50 years: sensors can span from cell to virus to protein / molecule
 → Materials "functionalization", add smell and taste, beyond human capabilities. (medical)
- Wacky things that could re-write the scaling math
 - Bending physics with nm-scale capabilities: Plasmonics, energy filtering, metamaterials, ...
 - Spintronics, non-volatile transistors, BEOL eNVM as FPGA, neuromorphic devices, ...



Thank you

