Where is semiconductor technology heading?
A view from industry and implications on computational nanotechnology

Carlos H. Díaz
R&D, Taiwan Semiconductor Manufacturing Company, Ltd., 121, Park Ave. 3, Hsinchu, Taiwan. Email: chdiaz@tsmc.com

Abstract **–** The quest for sustainable growth in computing performance and expanded functional capabilities of information technology and communication (ICT) products requires energy-efficiency improvements of underlying technologies in devices, systems, architectures, algorithms and software, and information representation and processing. This paper discusses emerging transistors, memories, and interconnect fabrics. The paper highlights open research areas, the necessary completeness of the metrics, and associated modeling challenges to identify viable alternatives to state-of-the art technologies and their projected evolutionary paths.

Introduction

Bridging the gap between existing silicon nano-technology and future VLSI needs innovations in devices and interconnect fabrics, each and all-cohesively enabling higher integration-density, performance improvements, and capabilities at lower power consumption cross-generations. Continued growth in computing capacity requires significant improvements in energy efficiency for it to be sustainable as illustrated in Fig.1 adapted from the SRC-SIA Decadal Plan of Semiconductors [1]. Artificial intelligence (AI) is a key element in the fourth industrial revolution. Increased cognitive capabilities are key to next generation AI, attaining them while decreasing the power consumption will be critical regardless of application domain as illustrated in Fig. 2. The quest for the necessary energy efficiency requires innovations at all levels from the basic technology structures and building blocks to the system architectures and algorithms. Research and development monumental efforts on silicon-based CMOS technology scaling continuously raise the bar that emerging devices and interconnects need to meet for value [2]. This paper overviews progress and research challenges for some of the leading emerging devices and interconnect fabrics.

Transistors

Power supply scaling is a critical knob in boosting energy efficiency from generation to generation, capacitance being the other. The optimal nominal operating voltage (minimum) for a given speed goal is bounded to the left by leakage power and to the right by active power as illustrated in Fig 3. To retain or improve switching speeds while also reducing power supply, materials with significantly better transport properties than silicon are needed as shown in Fig.3. Transistor structures such as stacked gate-all around channels will enable improved electrostatic control / steeper subthreshold slope than present Fin-FETs resulting in significant reduction of minimum operating voltages as shown in Fig. 4 [2]. Exploratory work on low-dimensional materials such as transition metal di-chalcogenides [3], graphene nanoribbons [4,5], or carbon nanotubes [6,7] seeks to demonstrate higher than state-of-the art drive current capability per unit footprint while also reducing transistor capacitances. Wafer-scale device-quality channel material synthesis - placement - patterning, reliable ultra-thin gate-dielectrics, gate-stacks supporting multiple work-functions, and thermally stable ultra-low resistivity contacts remain critical challenges. Figures 5 to 8 illustrate some of the best results to date for low dimensional materials. Despite these advancements, significant theoretical and experimental work sill needed to identify true platform-viable alternatives to silicon-based CMOS transistors. Comprehensive and predictive fundamental transport models that can realistically project on and off-state capabilities including thermally and mechanically stable low-resistance contacts remain imperative. Also key is accelerating the turn-around time (TAT) for fundamental screening of new materials synthesis and processing concepts through modeling.

Memory Elements and Arrays

A representative compute memory hierarchy of a computing system is shown in Fig. 9. Emerging memory devices in a given level of the memory hierarchy must outperform incumbent technologies on critical indexes to be considered promising alternatives. Those critical indexes include density, energy efficiency, speed, endurance, retention, environmental robustness, controllability, and complexity as proxy to cost/bit. SOT-MRAM is a potential alternative being explored as an SRAM alternative, demonstrating materials and cell structures with low-operating current while supporting tight write-error rates and magnetic immunity requirements remain challenging as illustrated in Fig. 10. Ferroelectric memories are also subject of active research for their high density and energy-efficiency potential [13], progress in understanding and resolution of endurance has recently being reported [14] as shown in Fig. 11. Emerging memory research and development demands ever increasing modeling capabilities to enable accurate, predictive, and fast TAT mapping of the design space including process variability, bit-error-rate, retention, and endurance metrics that along with power-performance-area (PPA) indicators are key to assert alternative memory cells across the memory hierarchy.

Interconnect fabrics

The resistance of vias, via-line interfaces, and lines represent a continuous challenge to the attainable chip-level performance and energy efficiency in advanced nodes. The search for materials with the goal of 2x or larger resistance reduction over elemental state-of-the-art solutions is a top challenge; yet when successful significant power-performance benefits are expected at the corresponding inception node as illustrated in Fig. 12. Inter-chip data movement (e.g. between external memory and processing units) is also an area of significant opportunities for elevating the system-level performance and energy efficiency. Scalable 3D-interconnect fabrics enabling increasingly higher intra and cross-die connection density as shown in Fig. 13 will be instrumental to denser VLSI systems supporting very high memory bandwidths [15,2].

Conclusion

Systems with increased levels of performance, functionality, and density will require increasingly more significant energy-efficiency innovations from software to process technology. Significant progress continues to be made in exploratory devices and interconnects. Yet challenges remain to attain proofs of concept which meet complete sets of critical metrics asserting their potential over evolutionary state-of-the art silicon-based pathways. To this end experimental efforts compounded with a robust computational modelling framework remains imperative to efficient and effective research and pathfinding as illustrated in Fig. 14 [16].

References

[1] SRC Decadal Plan for Semiconductors.
[2] M. Liu, IEEE ISSCC, 2020.
[3] T.-A. Chen et al., Nature, 2020.
[4] A. Way et al., J. Phys. Chem. Lett., 2019.
[5] Z. Mutlu et al., ACS Nano, 2021.
[6] J. Wang et al., Nature Catalysis, 2018.
[7] G. Pitner et al., IEEE IEDM, 2020.
[8] Tang et al., VLSI Tech. Symp., 2017.
[9] G. Pitner et al., Nano Letter, 2019.
[10] Q. Cao et al., Science, 2015
[11] A. Franklin, Nat. Nano, 2010
[12] S. Lin et al., IEEE IRPS, 2021.
[13] A. Tan et al., arXiv:2103.08806v1.
[14] P. J. Liao et al., IEEE IRPS, 2021.
[15] D. Yu, IEEE ISSCC, 2021.
[16] J. Wu, et al., IEEE SISPAD 2021.

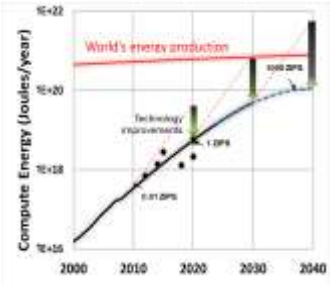


Fig. 1. Technology improvements will become increasingly important to sustain computational capacity growth over the next decades [1].

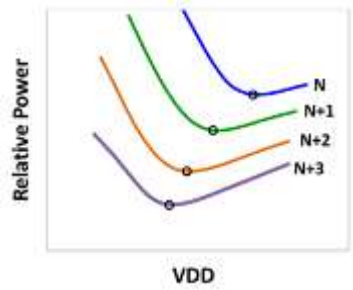


Fig. 2. Power supply scaling critical knob for cross-generation enhancements in energy-efficiency.

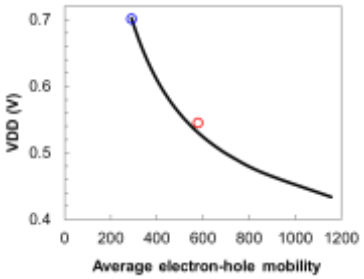


Fig. 3. Channel materials with significantly better transport properties critical to boost drive strength and circuit speed while scaling power supply.

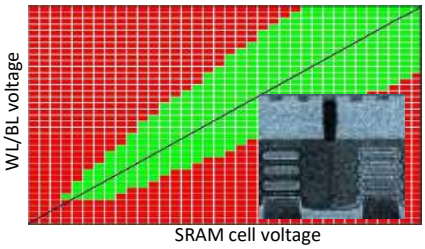


Fig. 4. Stacked gate-all-around channel structures to enable V_{DD} scaling beyond Fin-FETs, c.f. [2].

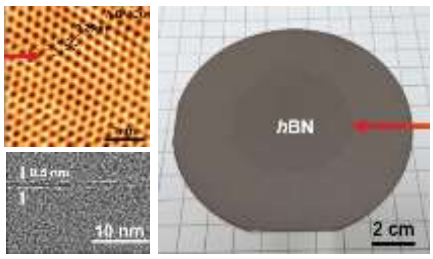


Fig. 5. Wafer-level synthesis demonstration of 2D channel materials and interlayer gate dielectrics [3].

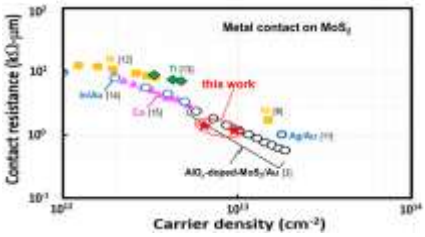
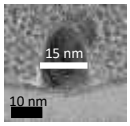


Fig. 6. Progress in low-resistance contacts of 2D TMDs [4]. Aside additional improvements in Rc, thermal stability, and CMOS capability also critical.



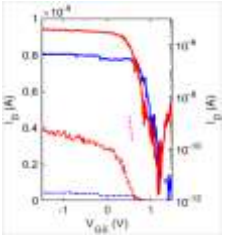


Fig. 7. Top-gated CNTs with ALD interlayer dielectric and HK attaining 65mV/dec @ Lg ~ 15 nm [7]. Progress on the synthesis of oriented and regularly spaced CNTs also continues to make inroads [6].

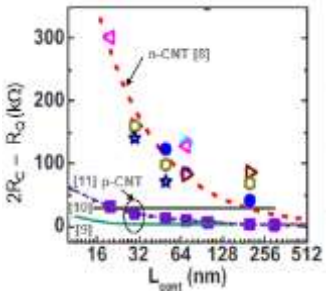


Fig. 8. Thermally stable low-resistance contacts for n/p-CNT transistors are subjects of active research.

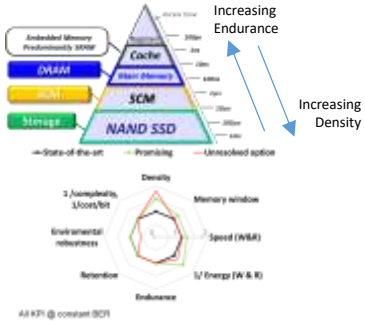


Fig. 9. Memory hierarchy and key research metrics.

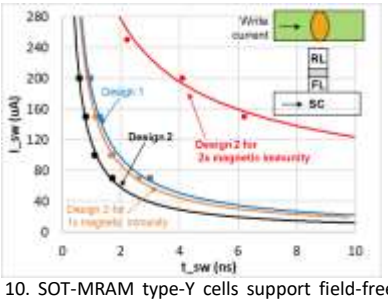


Fig. 10. SOT-MRAM type-Y cells support field-free operation. Low write current under density, magnetic immunity, write speed, and tight error requirements remain challenging.

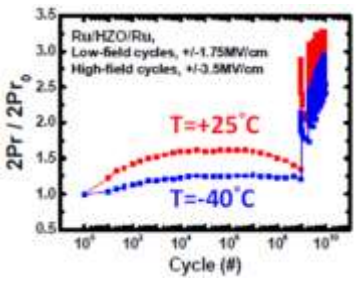


Fig. 11. Progress in fundamental understanding and approaches to high endurance ferroelectric memory cells has been reported, c.f. [14].

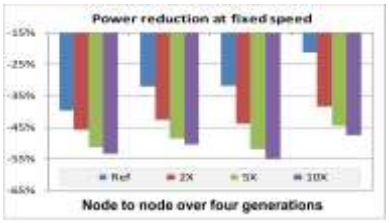


Fig. 12. Complex materials beyond elemental interconnect solutions continue to be searched with the goal of seeking 2x or larger via and line resistance reduction for significant chip-level power-performance benefits at inception node.

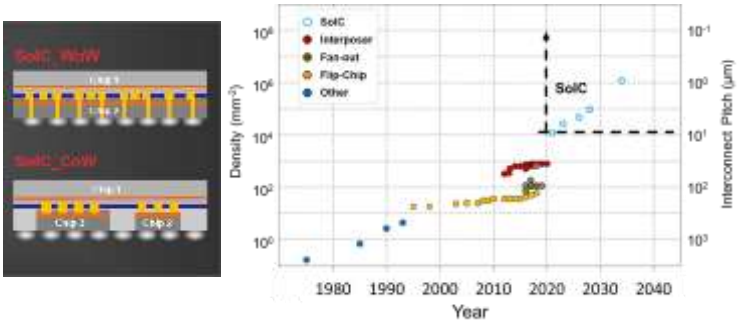


Fig. 13. New 3D stacking and interconnect fabrics will support sustainable I/O density increases .

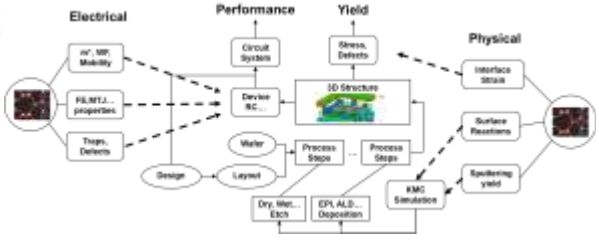


Fig. 14. A Virtual-Fab modeling framework.