

Application Challenges Resulting From Technologies of the Future

Alan Gara

Intel Fellow

Exascale Chief Architect



Legal Disclaimer

Today's presentations contain forward-looking statements. All statements made that are not historical facts are subject to a number of risks and uncertainties, and actual results may differ materially.

NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL® PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. INTEL PRODUCTS ARE NOT INTENDED FOR USE IN MEDICAL, LIFE SAVING, OR LIFE SUSTAINING APPLICATIONS.

Intel does not control or audit the design or implementation of third party benchmarks or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmarks are reported and confirm whether the referenced benchmarks are accurate and reflect performance of systems available for purchase.

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. See www.intel.com/products/processor_number for details.

Intel, processors, chipsets, and desktop boards may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel, Intel Xeon, Intel Core microarchitecture, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others

Copyright © 2011, Intel Corporation. All rights reserved.



The next step in perf/\$

Typical DRAM Memory Die (2016) ~ 8Gb will be about 100 mm² (as always)

Processor floating point unit ~0.03 mm² (2 Flops/cycle) (see below)

Even if the core is 100x bigger than the FPU, At 1.0 GB/core we have >100x more silicon in memory than processing. This is not cost balanced.

Threading gives us a mechanism to increase bandwidth to support much higher throughput

New memory architectures

- For cost balance we need to either,
- 1) Use much less memory per compute
- or
- 2) Make physical size of capacity much smaller

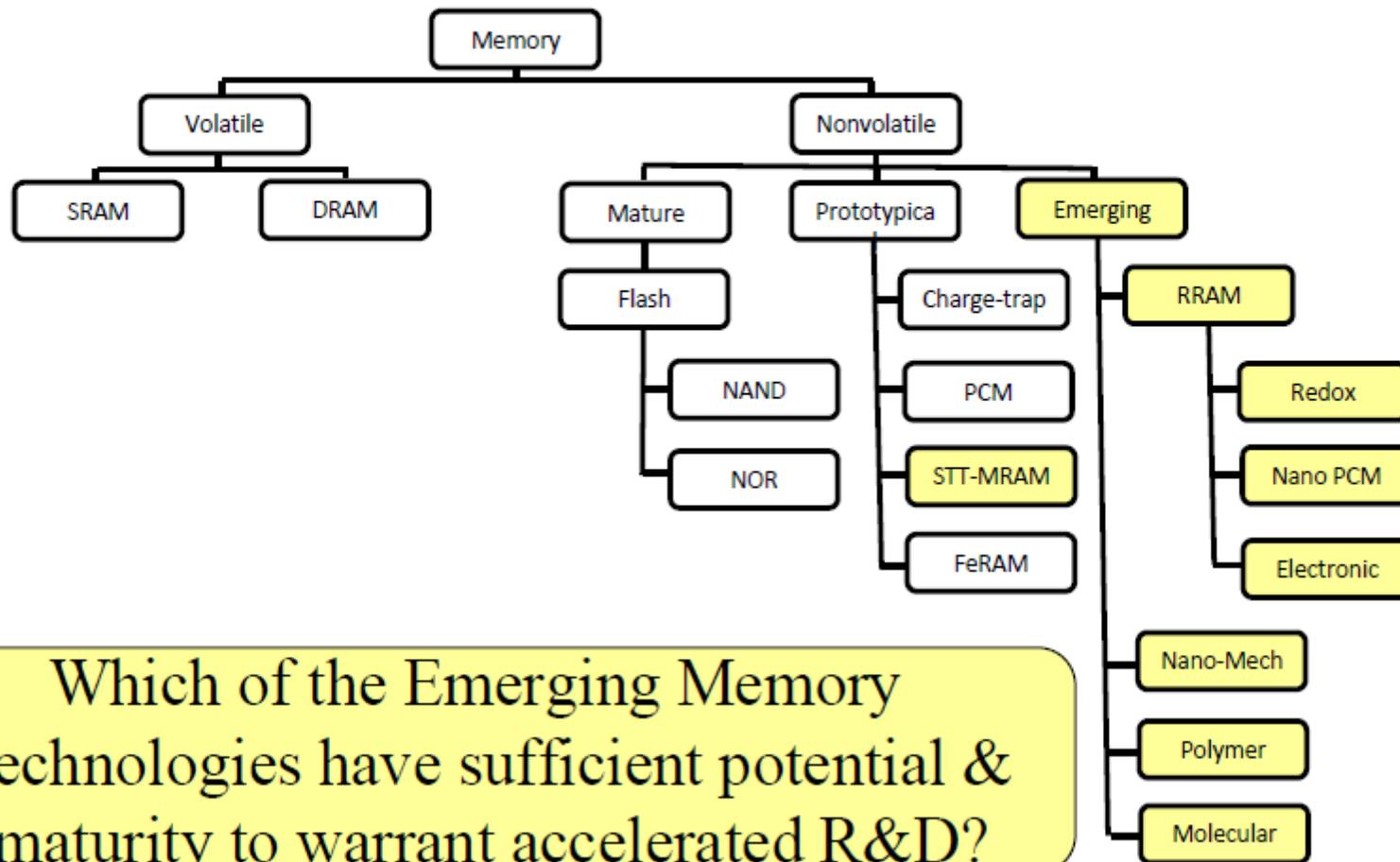
Year	Tech (nm)	V	Area (mm ²)	Power (W)	Performance (Flops/W)	Area (mm ²)	Performance (Flops/W)	
2004	90	1.10	0.50	1.0	0.50	0.50	0.50	
2007	65	1.10	0.26	1.0	0.26	0.26	0.26	
2010	45	1.00	0.13	0.45	0.13	0.09	0.09	
2013	32	0.90	0.06	0.29	0.06	2.9E+07	0.08	
2016	22	0.80	0.03	0.18	0.03	4.09	1.8E+07	0.07
2019	16	0.70	0.02	0.11	0.02	5.63	1.1E+07	0.06

Table 7.4: Expected area, power, and performance of FPUs with technology scaling.

New device technologies

- Many emerging memory technologies
 - Some now moving to production... it will happen
- Spin Torque Transfer devices
 - One of the potential silicon based variations of the future
- Silicon Photonics for communications
 - It is here and will improve rapidly (just in time)

Taxonomy of Emerging Memory Technologies



Which of the Emerging Memory Technologies have sufficient potential & maturity to warrant accelerated R&D?

Candidate Emerging Memory Technologies

Emerging Research Memory Technology Entry	Stand-Alone	Embedded
Ferroelectric-gate FET	X	
Nanomechanical RAM	X	X
Spin Transfer Torque RAM		X
Nanowire Phase Change Memory	X	X
Redox Memory (including Fuse/Antifuse Memory)	X	X
Electronic Effects Memory (Charge trapping, Mott transition, Ferroelectric barrier effects)		X
Macromolecular memory	X	X
Molecular memory	X	X

Courtesy of James Hutchby SRC



Two different directions

New memory technologies
replace/augment DRAM

DRAM the remains
dominant load-store
memory technology



Memory capacity per compute
5x-10x better than DRAM

Memory capacity per
performance drops 10x to 20x
from current levels.

Modest need for threading
when new technologies
available.

Aggressive threading is
commonplace/necessary.

Program model changes focus
on increasing task scaling.

Program model changes focus
on thread scaling.

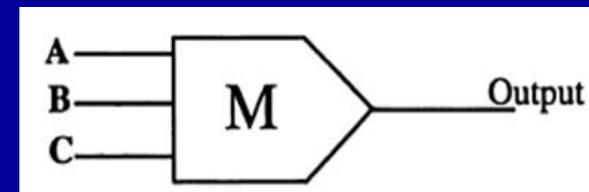
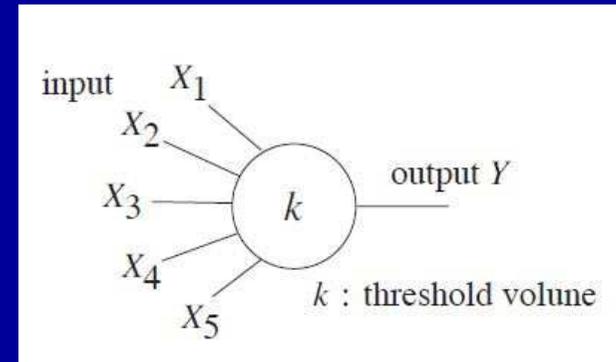
More powerful “unit” devices

- Greater functionality per device - less interconnect per function
- Interconnects losses are a large fraction of total power*

- Use more powerful “unit logic devices” –multi input threshold gate

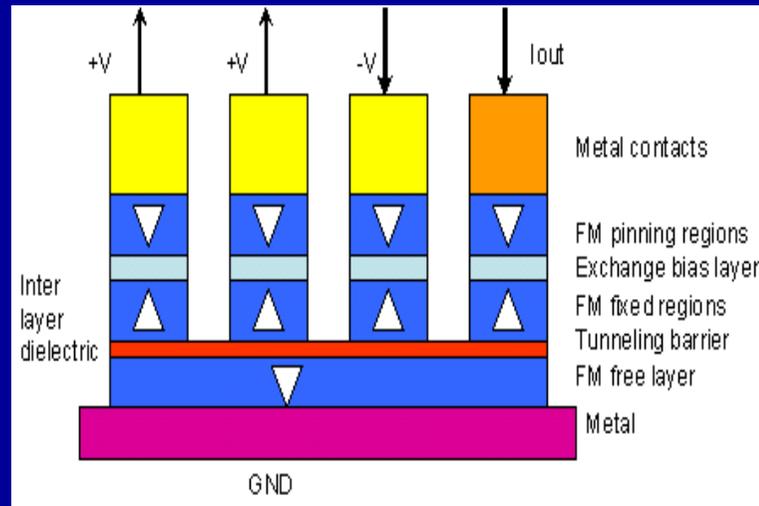
- 3 input majority gate is specific example

- Can be implemented using spin torque transfer technology

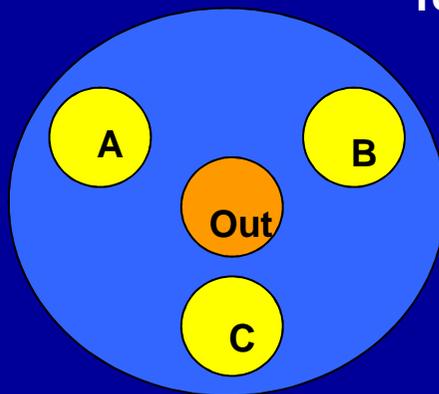


* Magen et. al. Interconnect-Power Dissipation in a Microprocessor

Spin torque transfer majority gate (STTMG)



Top view



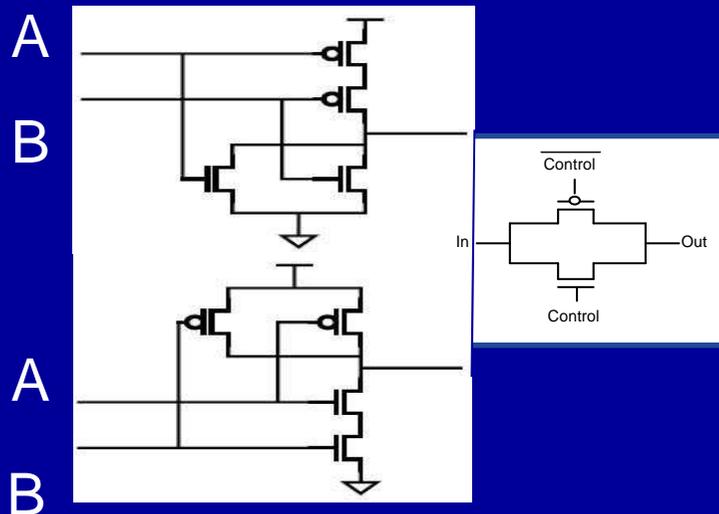
- Four stacks of ferromagnetic materials, similar to perpendicular MTJs
- Three stacks - inputs. One stack - output
- Free layer is common to all four stacks
- Polarity of free layer can be controlled by polarity of voltage applied to 2 of the 3 input stacks
- Polarity of free layer can be sensed by magnetoresistance of forth stack

Courtesy of George Bourianoff (Intel)

Majority Gate Equivalent circuit functionality

CMOS

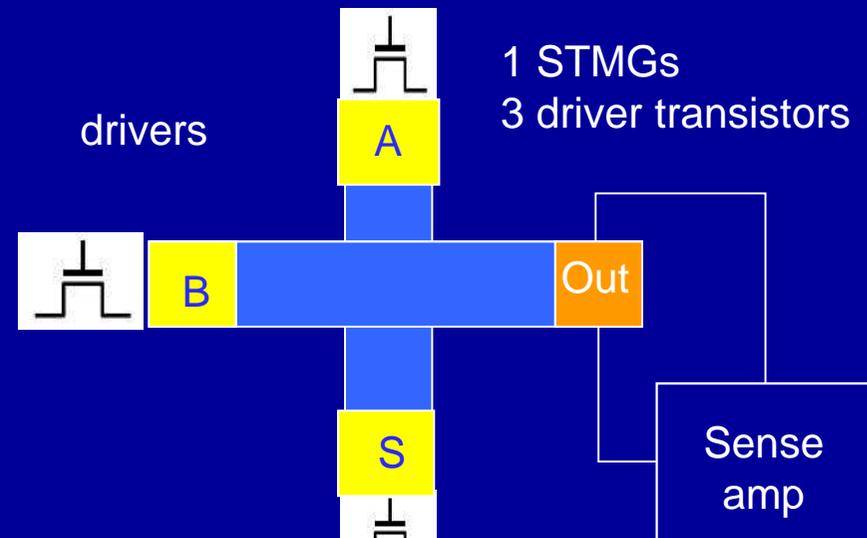
12 transistors



Average area per transistor in a MPU,
according to ITRS
Requires 12 transistors

$$A_{\text{gate}} \sim 12 * 72 F^2 = 824 F^2$$

STMG



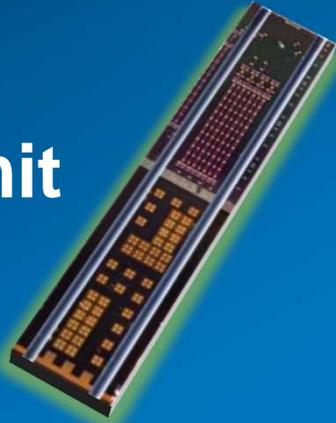
Magnetic circuits in the metal layers.
Drive transistors no additional area.
Scalable with process feature size.

$$A_{\text{gate}} \sim 36 F^2$$

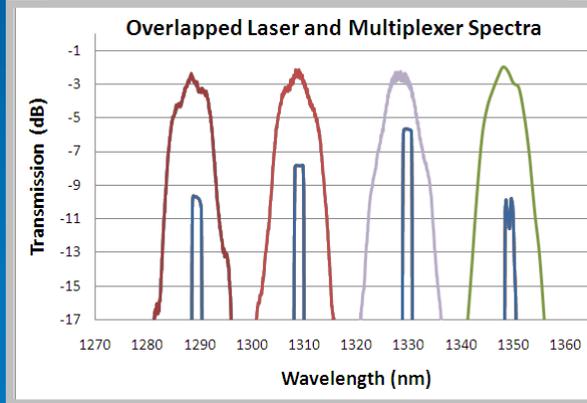
Courtesy of George Bourianoff (Intel)

Measured Data

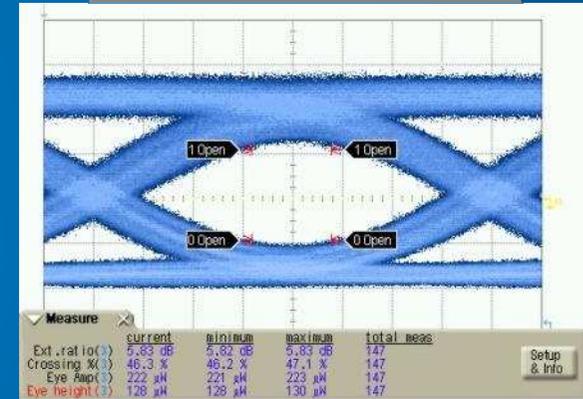
Transmit



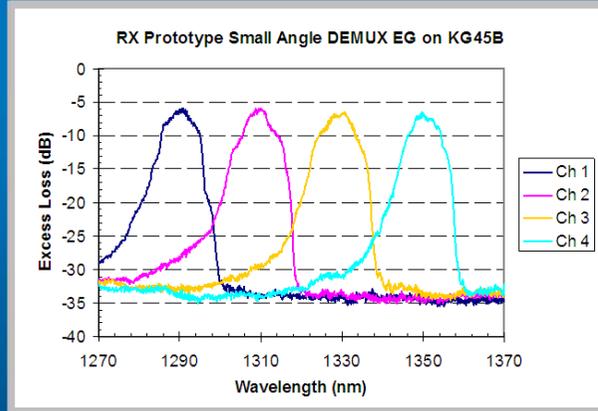
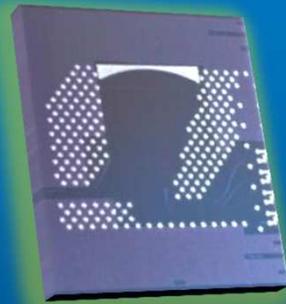
4 hybrid Silicon Laser Outputs



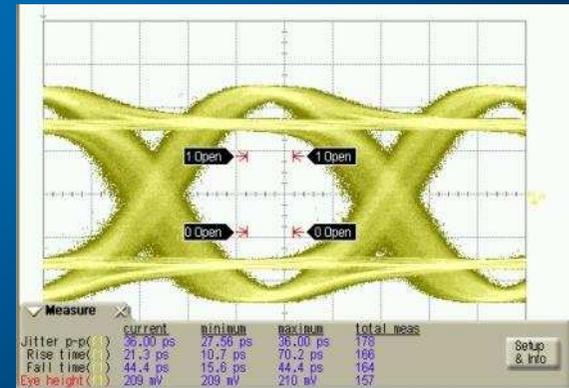
12.5Gbps data output per channel



Receive



De-Multiplexer separates wavelengths



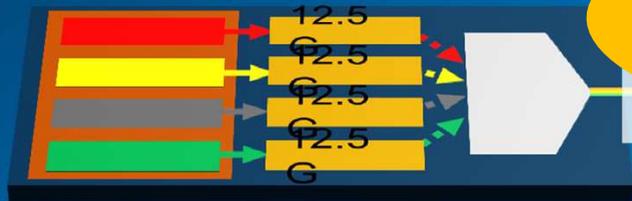
Electrical Output From Receiver

We ran link for more than 3 days with no errors (>3 Petabit)
 Translates to Bit-Error-Rate (BER) of $< 3e^{-15}$



The Path to Tera-scale Data Rates

Today: 12.5 Gbps x 4 = 50Gbps



Need to drive up bandwidth per fiber as the fiber and fiber connectors will become dominant cost.

Scale

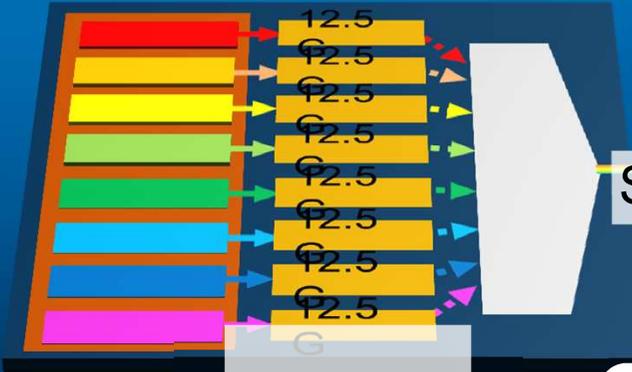
100G..160G



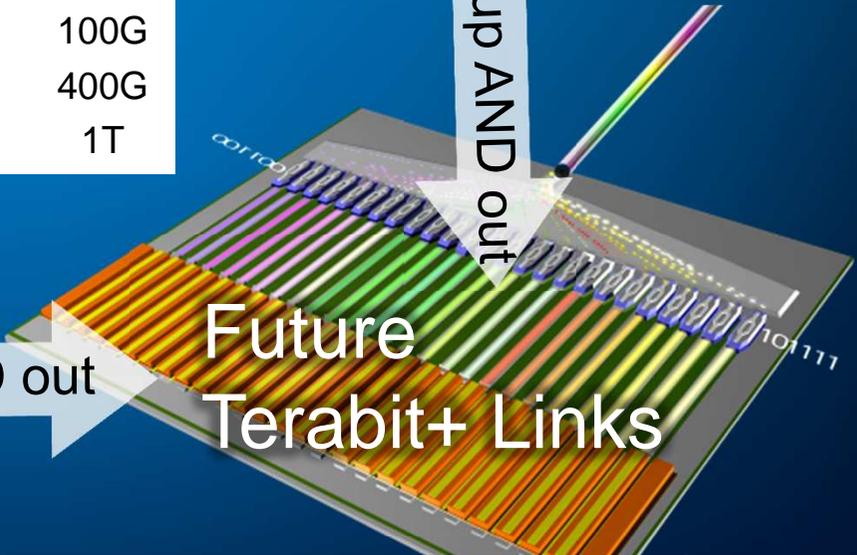
Speed	Width	Rate
12.5	x4	50G
12.5	x8	100G
25	x16	400G
40	x25	1T

Scale up AND out

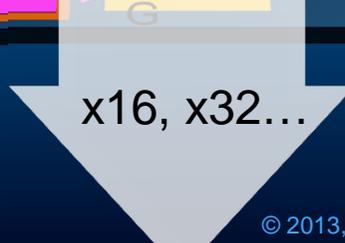
12.5 Gbps x 8 = 100Gbps



Scale up AND out



Future Terabit+ Links



Could enable cost-effective high speed I/O for data-intensive applications



Challenges for Optical

Power Efficiency vs Electrical



- Electrical links continue to improve performance/watt
- Until we have optimized circuits and link for optical we will always have power penalty for E-O-E conversion. There is no free lunch....

Take away: put optical where it brings value not everywhere ... (yet)

Many challenges ahead (we all know what they are)

Resilience

Extreme parallelism

Limited memory capacity and bandwidth

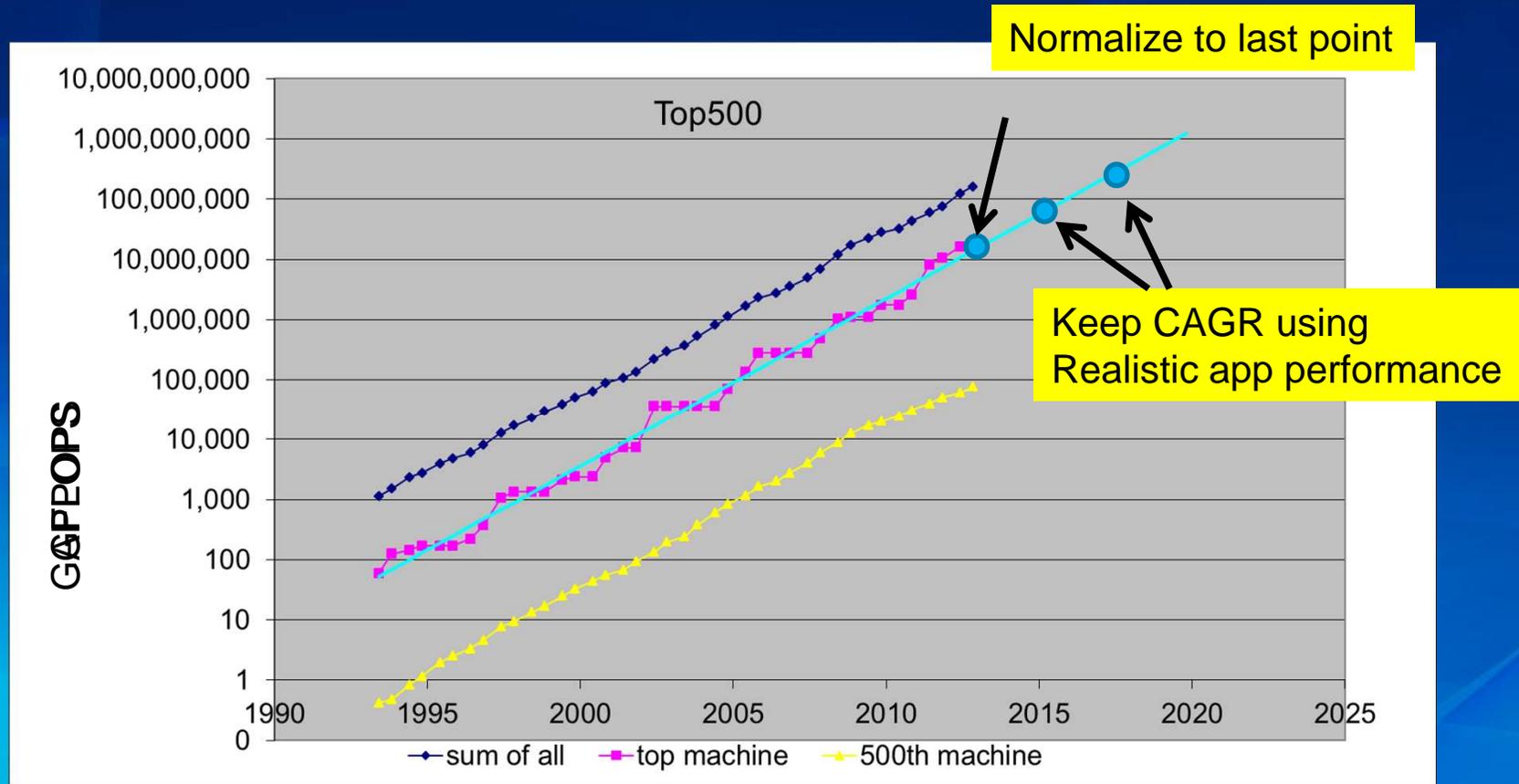
Cost

·
Power

But what is the definition of success?



A new measurement to define “best” is needed *(Need to hold onto the historic rate of progress)*



Transition to realistic application performance benchmarks



System Power is NOW critical

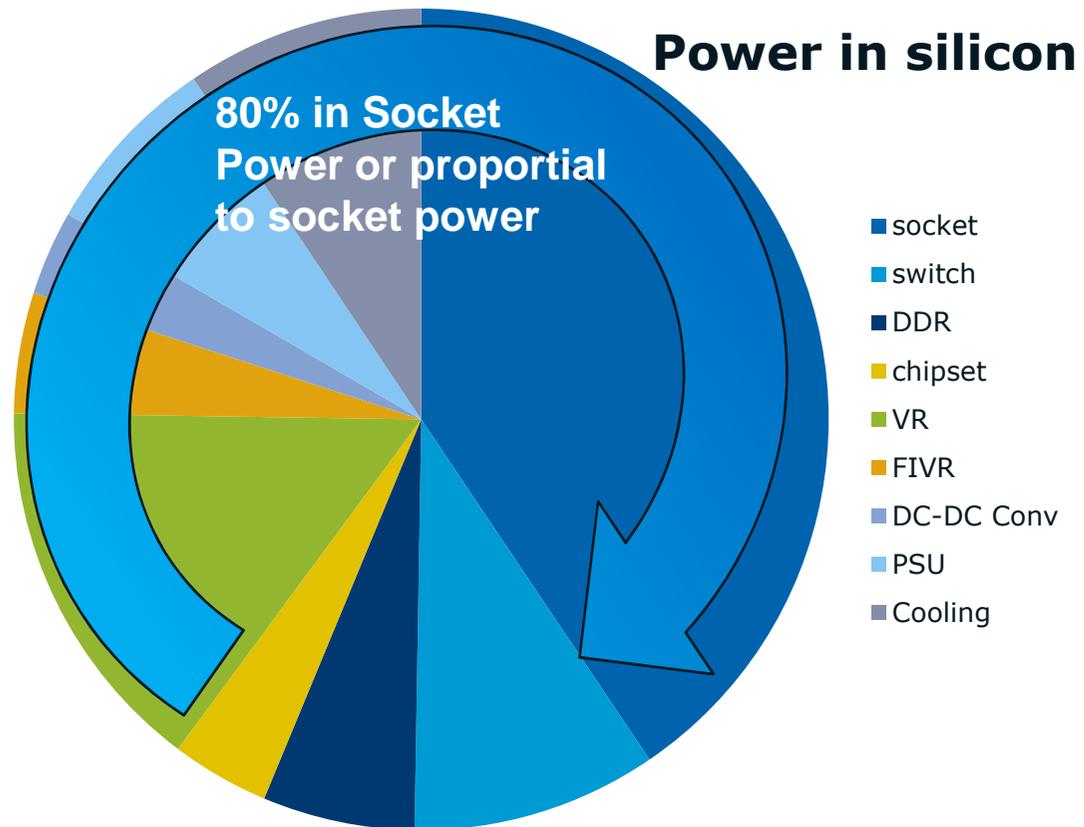
Socket silicon power and power losses amount to ~ 80% of total power

We have run out of runway on this.

Need to focus on all aspects of system power.

Socket power and power distribution and cooling dominates.

Solution likely to be four 25% impacts that together gets us the additional factor of 2 we need beyond silicon tech scaling.



Baseline architectural improvements will focus on thread scalability

Limitations to thread scalability

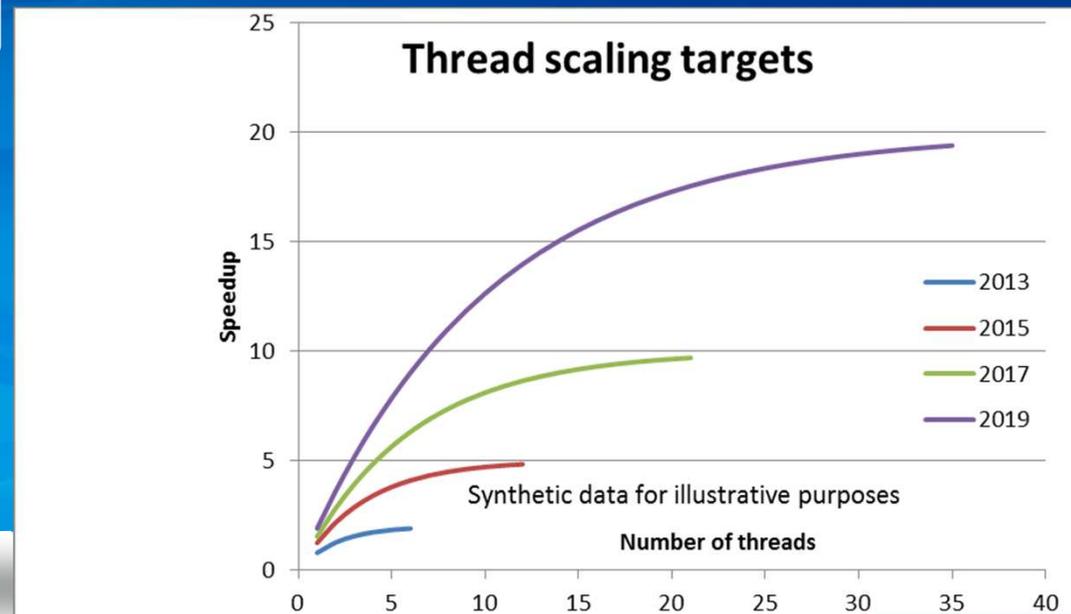
- Load/execution imbalance
- False cache sharing
- Start up overheads
- Synchronization overheads
- Reductions
- Amdahl's law limits

There are hardware components to each of these limitations

Revolutionary hardware changes will allow evolutionary user access to performance.

Will improve performance through increasing thread performance and increasing thread scalability.

Will allow for true **energy optimization** for both fine grain and coarse grain.



Focusing on Power...

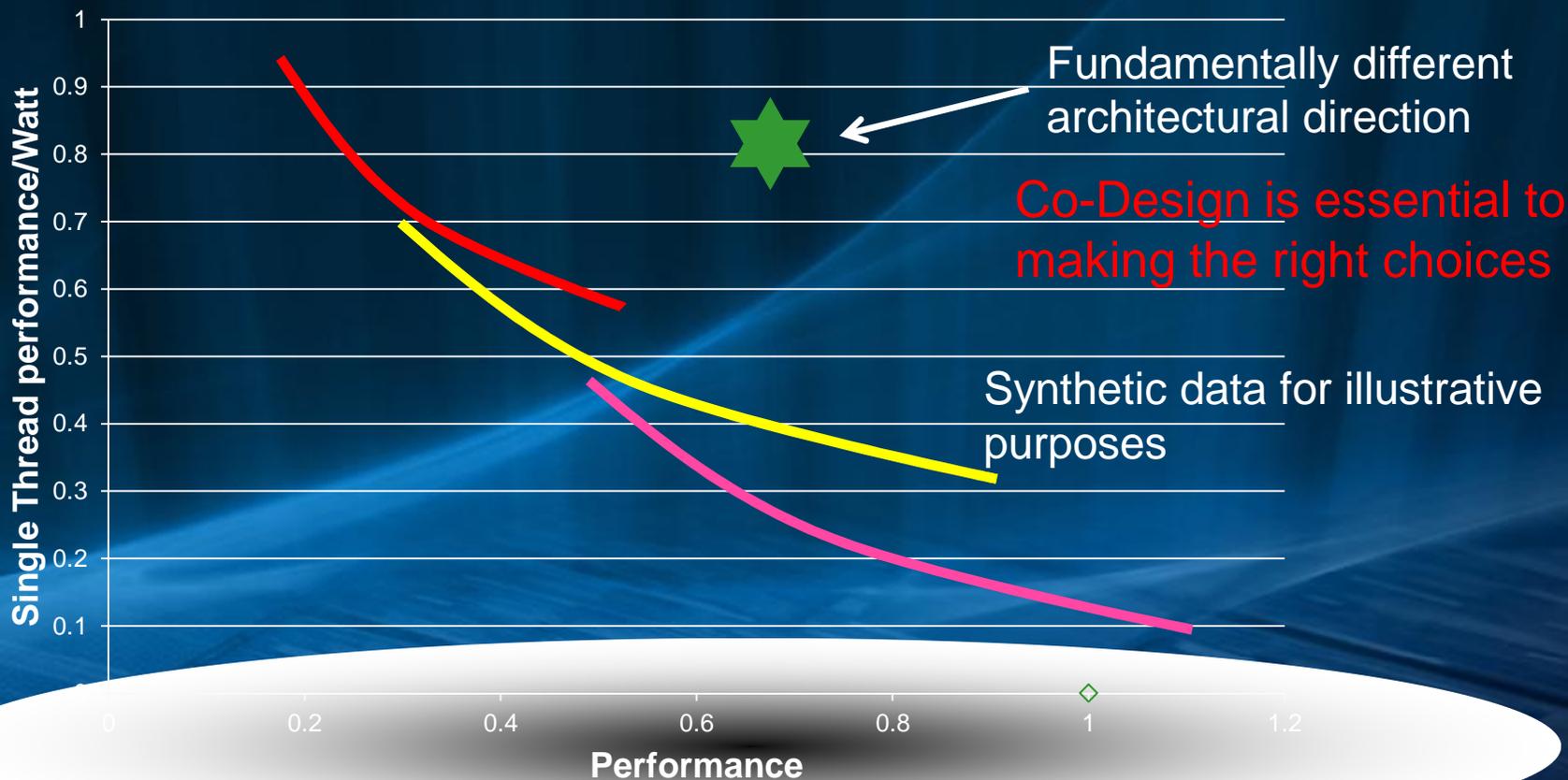
(Some non-conventional directions ...)

- 1) Come at this from a program model perspective?
- 2) Architect thinking about energy differently?



We know that small cores can be more energy efficient

Single thread Perf/Watt versus Single thread Perf

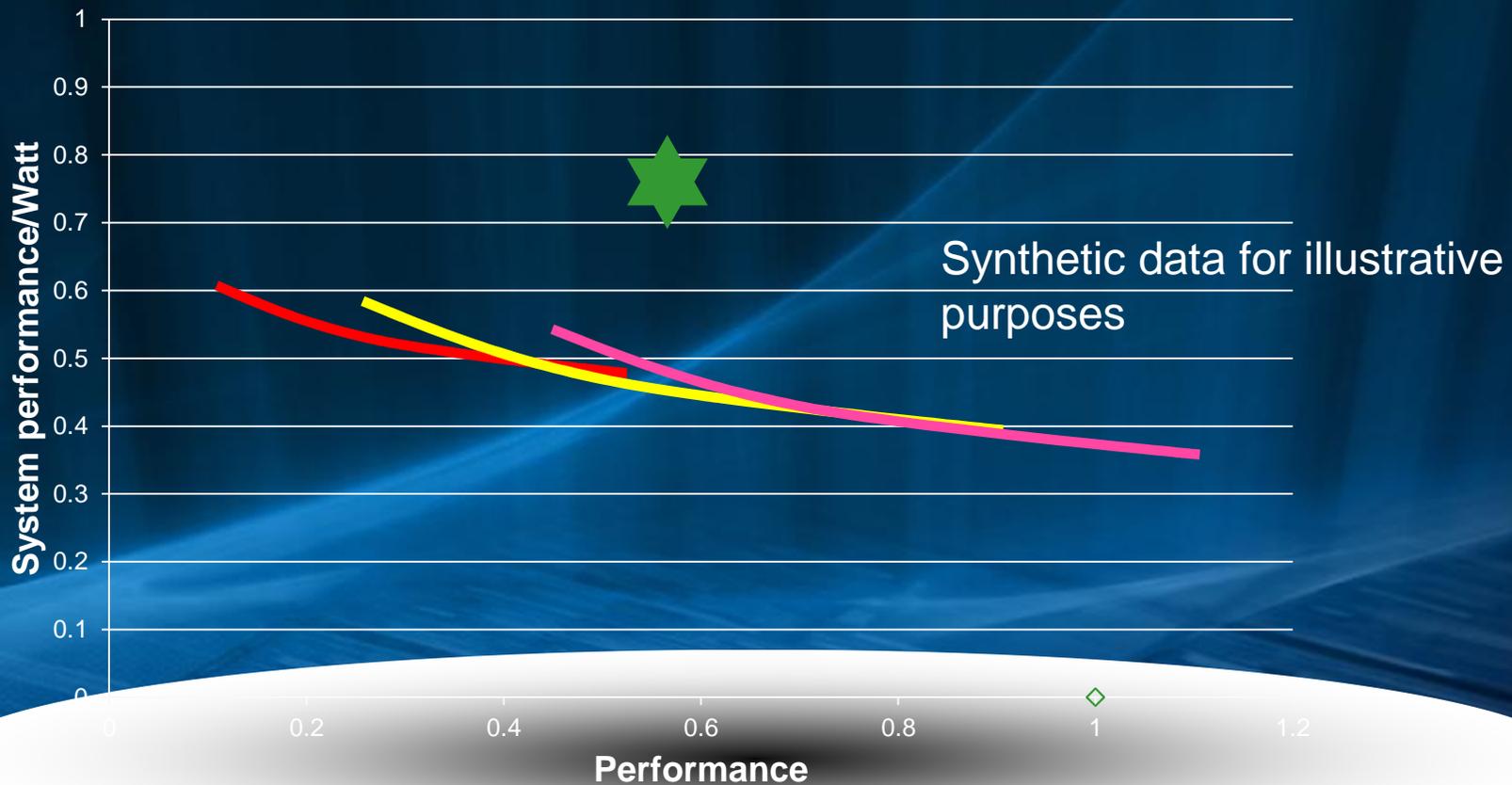


Different core complexities lead to a family of curves



Mapping to system performance

System Perf/Watt versus Single thread Perf



When thread scaling issues are accounted for, core differences often get washed out



Coming at power from a program model perspective (2)

Maybe some heresy is needed...

BAD! You want to make it harder to program? Are you talking functional programming?

A MORE constrained program model.

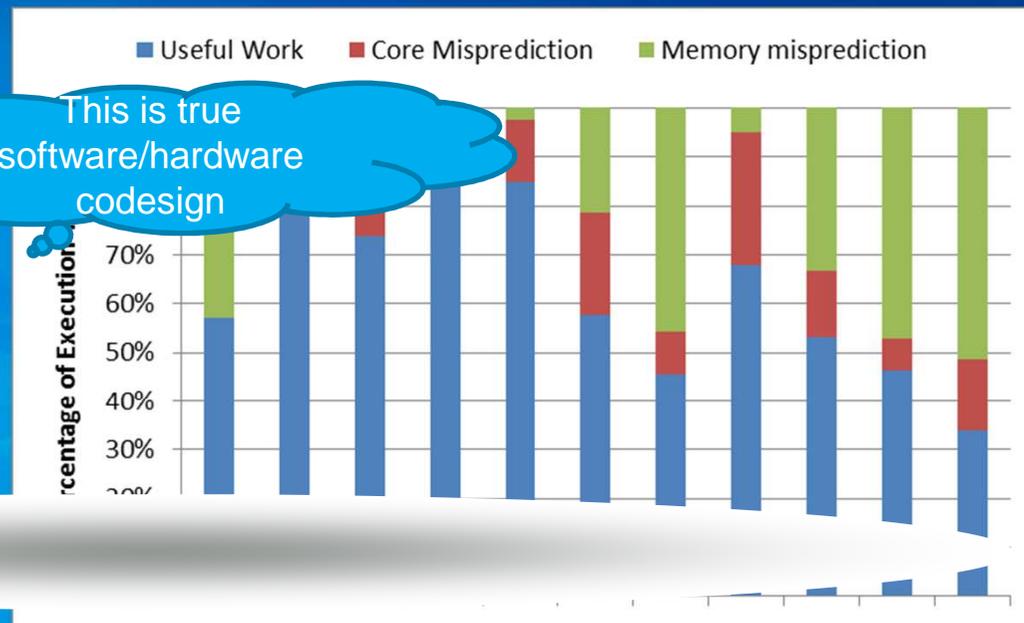
Could lead to ...

- Reduce/eliminate branch misses
- Reduce/eliminate cache misses
- Allow for a simpler core.
- Sweet spot for freq goes higher
- Simpler memory/cache hierarchy
- Compiler improvements in auto parallelization. (could force trivial)
- Compiler improvements in ILP

Which could result in ...

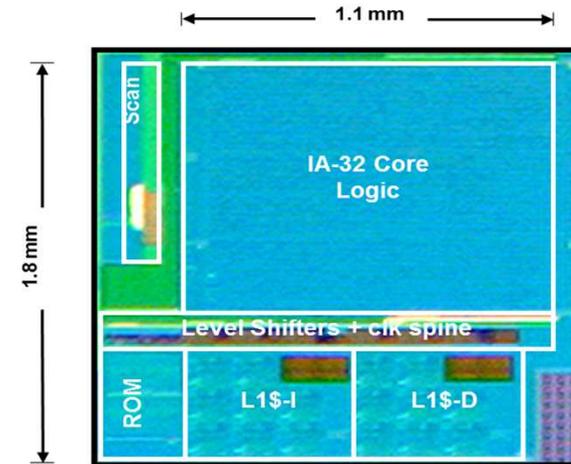
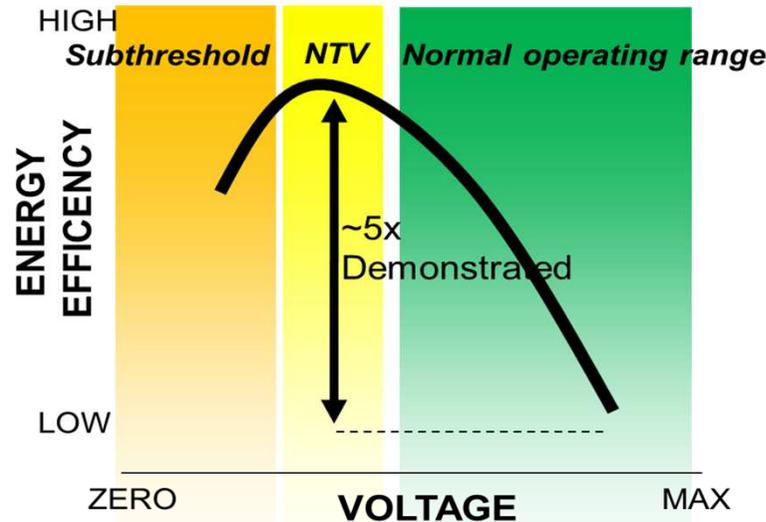
- A code with much more "user single thread" performance.
- Significantly improved perf/W

This is true software/hardware codesign



Addressing the Power Challenge

Near Threshold Voltage Operation Demonstrated (conventional view)



Ultra-low Power	Energy Efficient	High Performance
280 mV	0.45 V	1.2 V
3 MHz	60 MHz	915 MHz
2 mW	10 mW	737 mW
1500 Mips/W	5830 Mips/W	1240 Mips/W

Technology	32nm High-K Metal Gate
Interconnect	1 Poly, 9 Metal (Cu)
Transistors	Core:6M
Core Area	2mm ²
Package	951 Pins FCBGA11

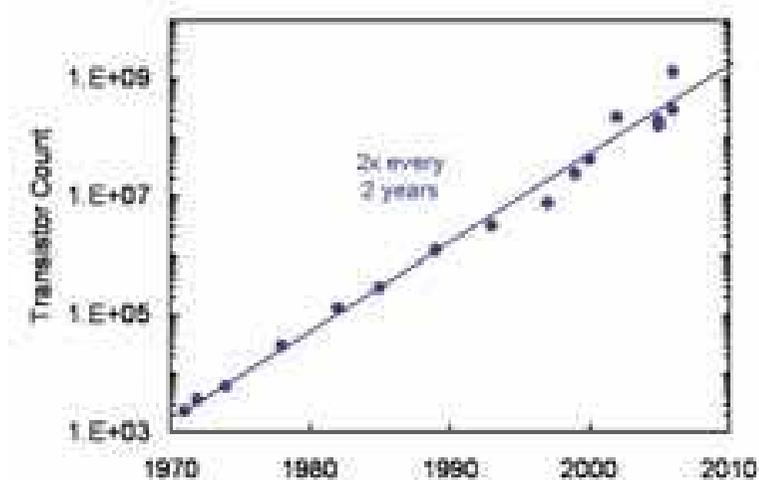
Wide Dynamic Range

S. Jain, S. Khare et. al. "A 280mV-to-1.2V Wide-Operating-Range IA-32 Processor in 32nm CMOS," ISSCC 2012

Voltage has been a big knob but we need more than voltage and technology scaling

Do we need to cheat Maxwell?

(Thinking about energy efficiency differently)



James Clerk Maxwell: (1831-1879)



Transistor count doubles every 2 years.
Performance of 2x / 2 years would seem a stretch goal.

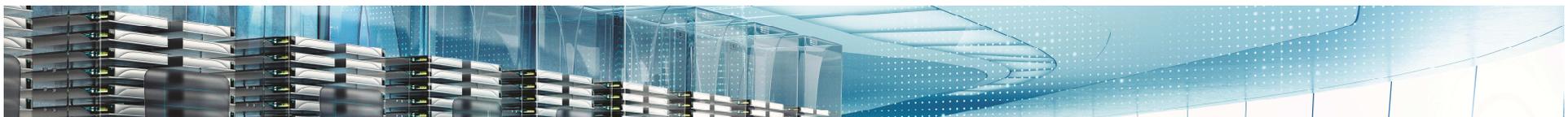
So can we expect to do much better?

Should not try to beat physics but scaling is not performance. (note performance is not one of Dennard's rules)

Performance is information processing

Dennard's scaling laws

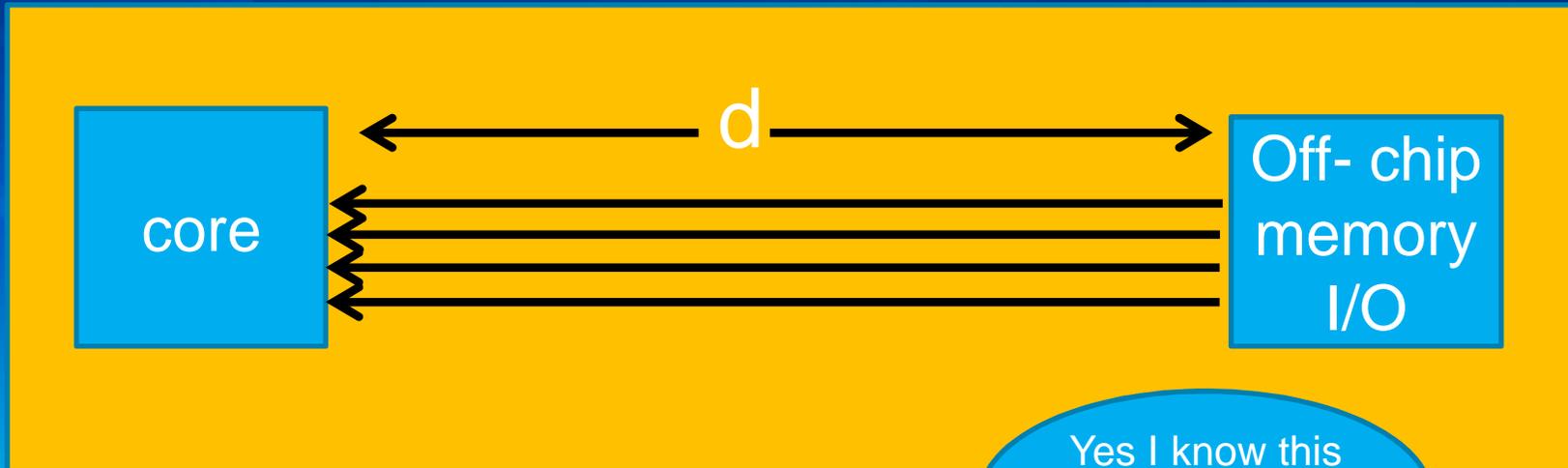
Device or Circuit Parameter	Scaling Factor
Device dimension tox, L, W	$1/\kappa$
Doping concentration Na	κ
Voltage V	$1/\kappa$
Current I	$1/\kappa$
Capacitance $\epsilon A/t$	$1/\kappa$
Delay time/circuit VC/I	$1/\kappa$
Power dissipation/circuit VI	$1/\kappa^2$
Power density VI/A	1



Architect as if there is no silicon scaling (we will get there)

A simple example provides foot for thought...

Energy is needed to move information...



$$P = Cv^2 f \text{ (per wire)}$$

$$E = Cv^2 \text{ (frequency independent)}$$

Yes I know this seems like kindergarten

Only way to improve is to reduce capacitance or voltage?

Utilizing time for encoding information



Does this work for logic? (homework)

16 possible transition times
encode 4 bits of information



Total energy is equivalent to one wire changing
And C_{ap} is actually an average assuming neighboring tracks.



So what is the limit ?

Energy limit arising from entropy considerations “Landauer limit “ Rolf Landauer IBM 1961

Minimum energy for the computation of one bit of information is ...

$$=kT \ln(2) \quad k = \text{Boltzmann's constant} \quad T = \text{absolute temperature}$$

If we associate each stage of a floating point multiplication with an irreversible bit

Energy for double precision multiply is $\sim 1.6 * 10^{-5}$ pJ
(Assuming simple full adder tree and each adder results in 2 outputs)

Minimum energy for 1EF/ is ~ 16 Watts (which happens to be the power used in a human brain)

1 ZettaFlop/s could be done with as little as 16KW

1 YottaFlop/s could be done with 16MW ?



Summary

- There are many new technologies in flight. These will have a profound impact on how we program systems in the future. (we are on the right path for at least the next 10 years)
- Power is and will remain the biggest challenge. We need to no longer improve performance faster than energy efficiency improvements.
- Silicon scaling for density is alive and well. We will need to explore architectural and technology solutions to the energy efficiency challenge in the future.
- We need to have the right measures of goodness to navigate directions.



