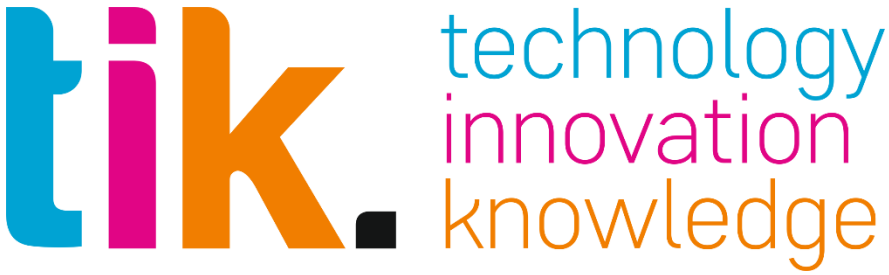
Energy Efficiency in Computing (1)

CERN Academic Training – May 2016

Andrzej Nowak

http://tik.services


tik.
technology
innovation
knowledge

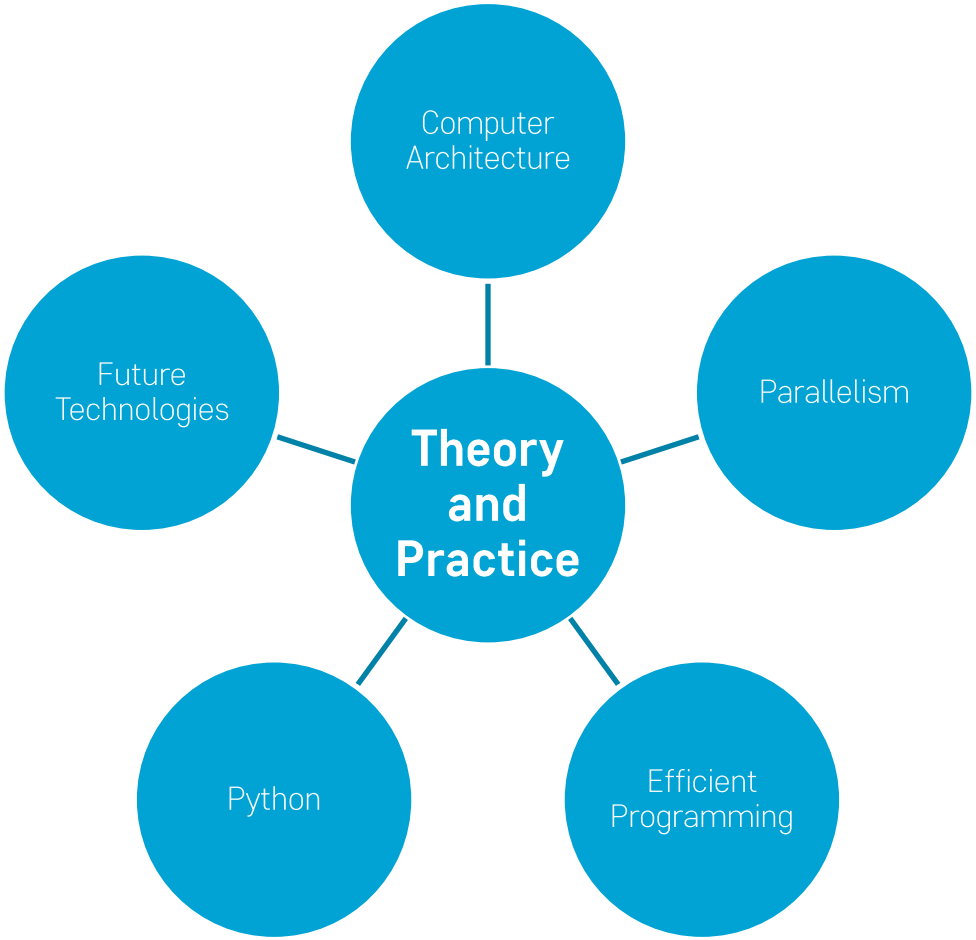
TIK Services

Technology

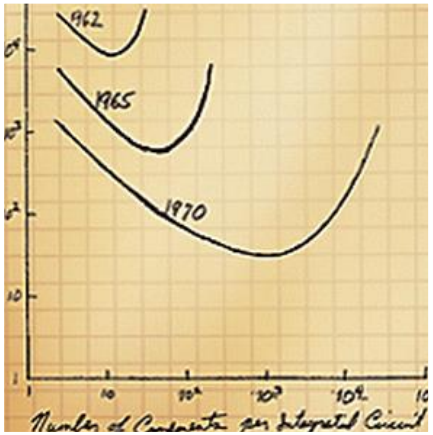
Innovation

Knowledge

TIK Training



Outline

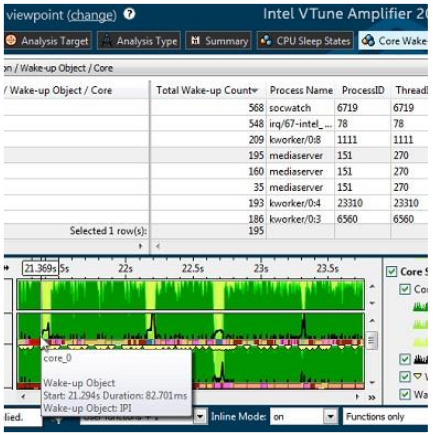


Day 1:

Silicon, hardware

Day 2:

Datacenters, software, future technologies

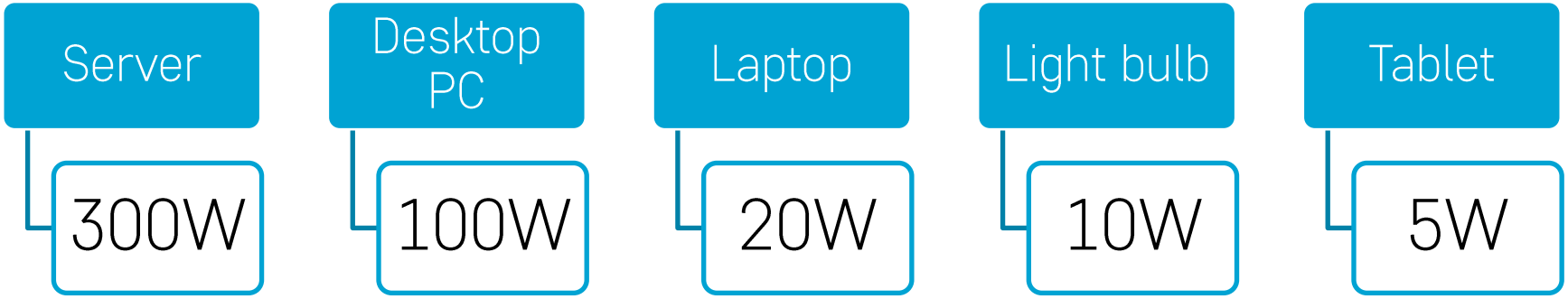


Basics

1 Watt = 1 Joule / second

(power = energy / second)

Everyday devices



Practical considerations

How much does it cost to charge the iPhone 6 and iPhone 6 Plus?

iPhone 6
$0.47/year (3.8 kWh)

iPhone 6 Plus
$0.52/year (4.2 kWh)

Opower 2014

Image: opower

Practical considerations

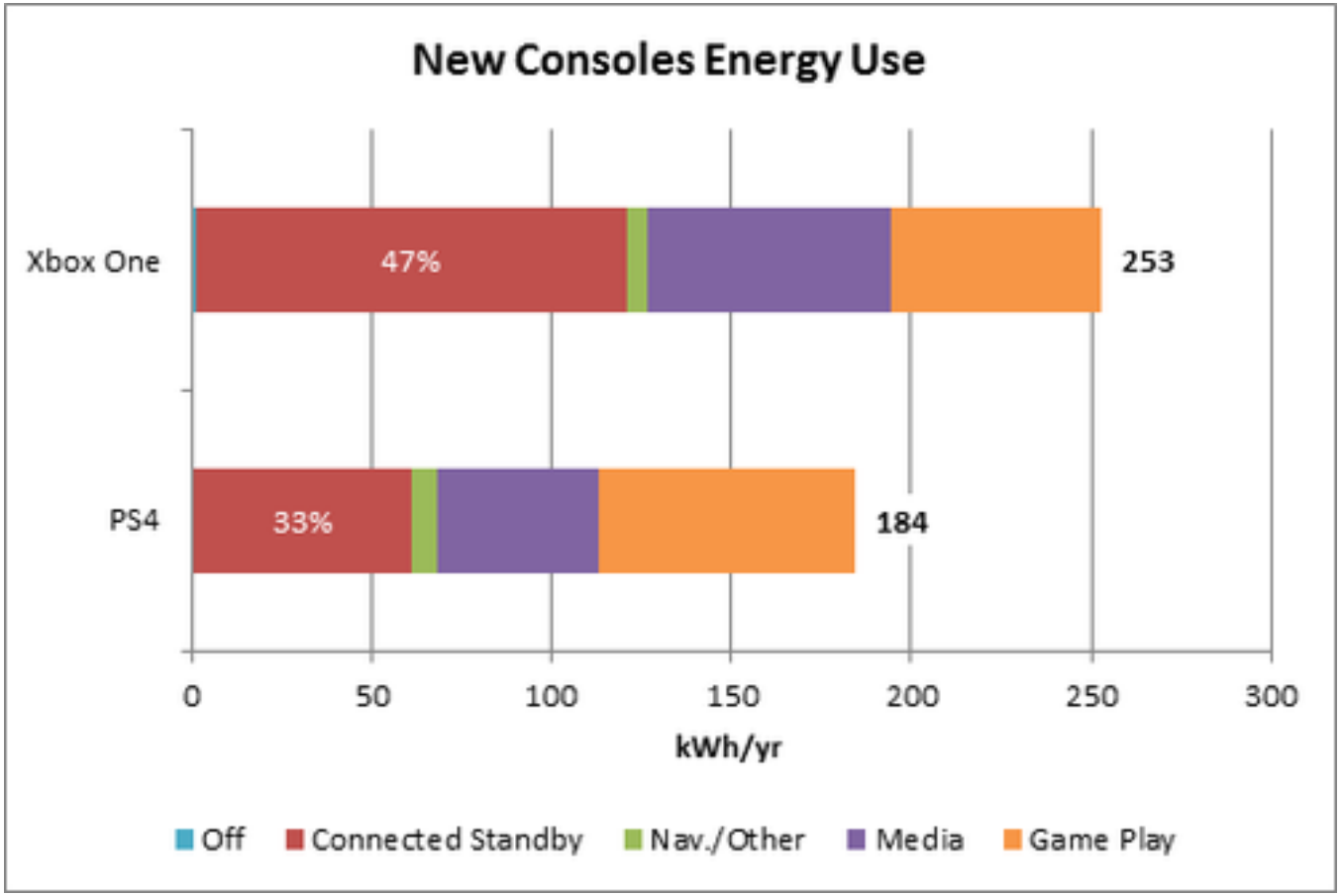


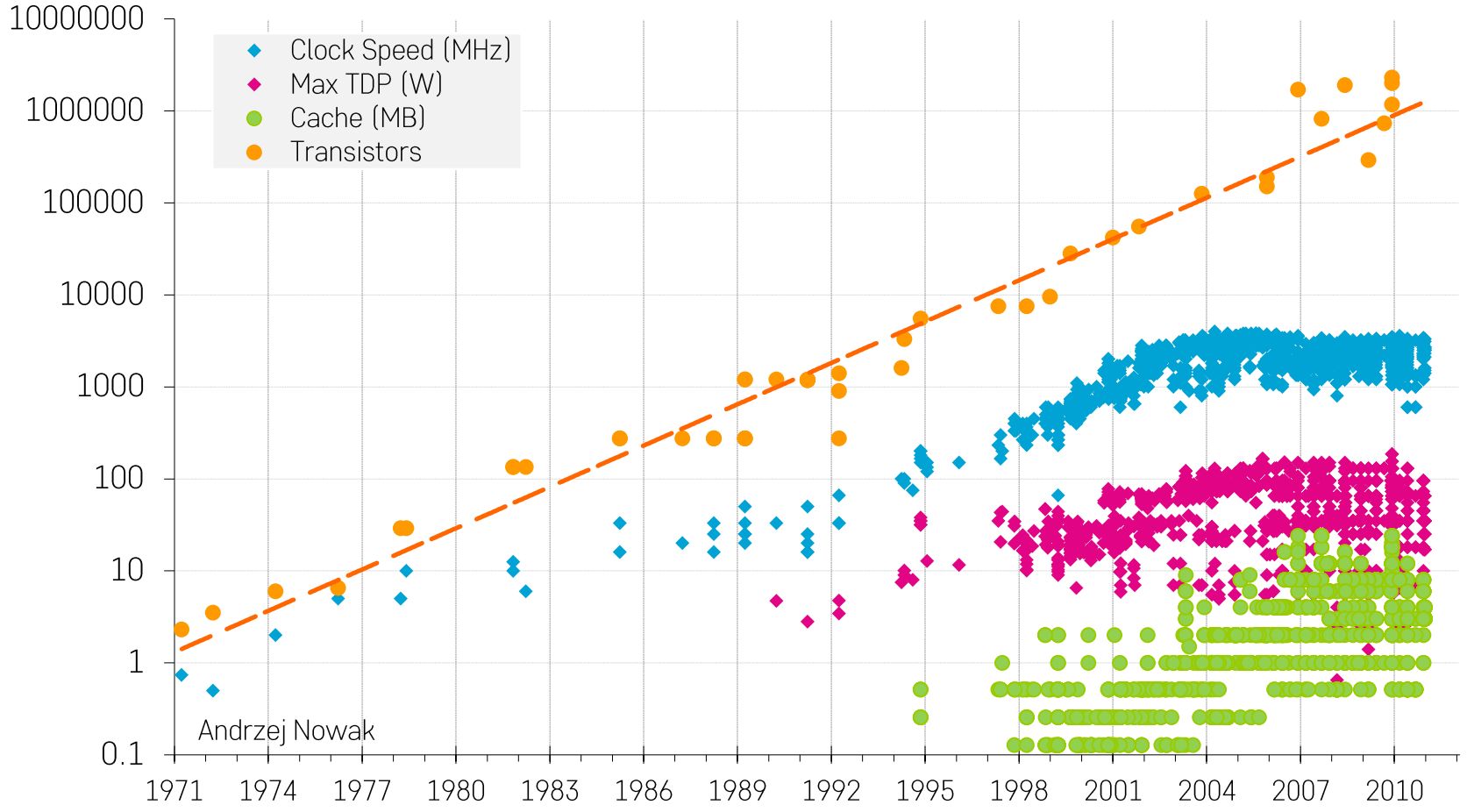
Image: NRDS

Energy-efficient Silicon

tik.

Moore's Law and power

Since 1971 – based on 1'700 Intel CPUs



Gain (and Pain)

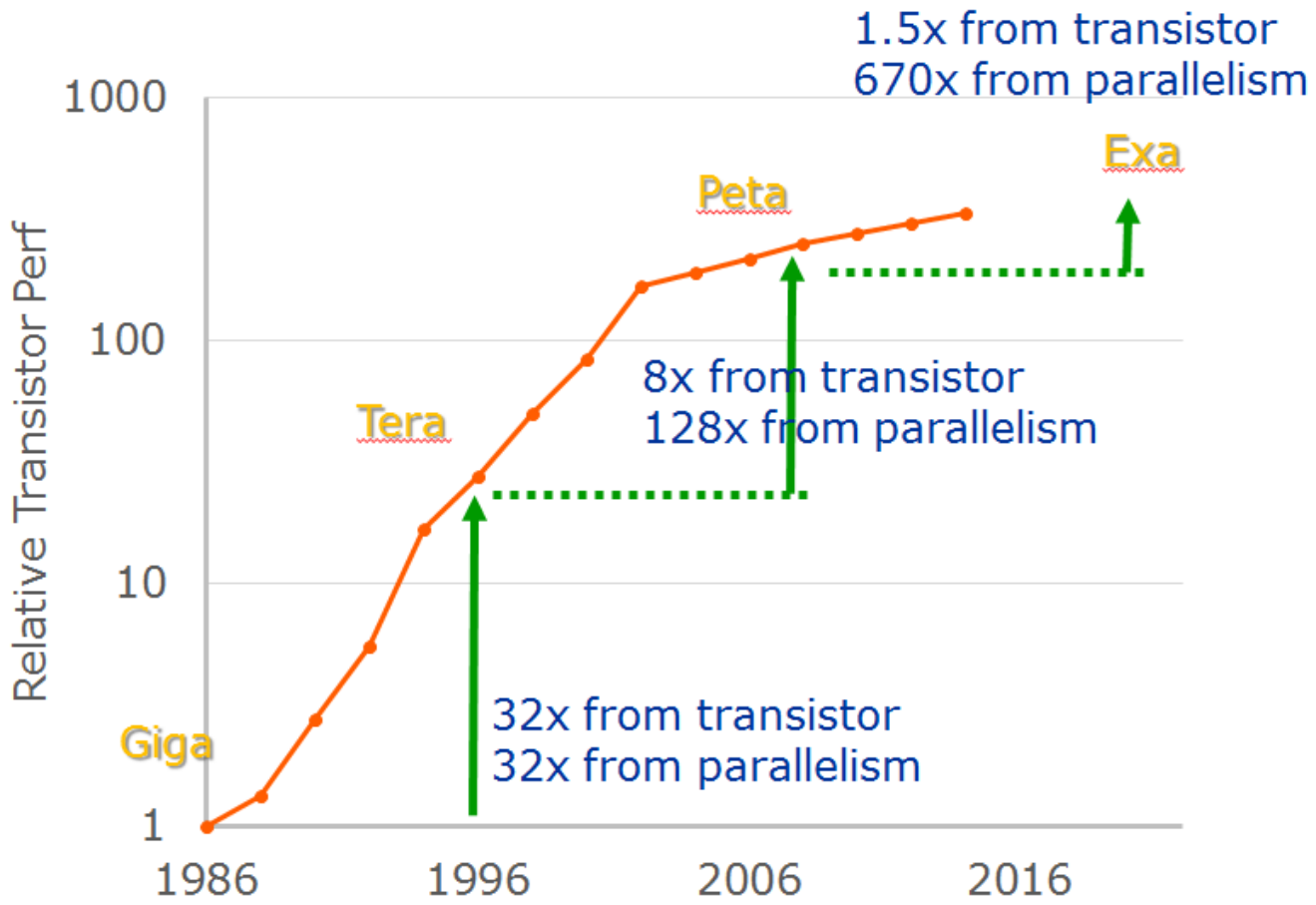


Image: S. Borkar/Intel

Moore's Law and Top500

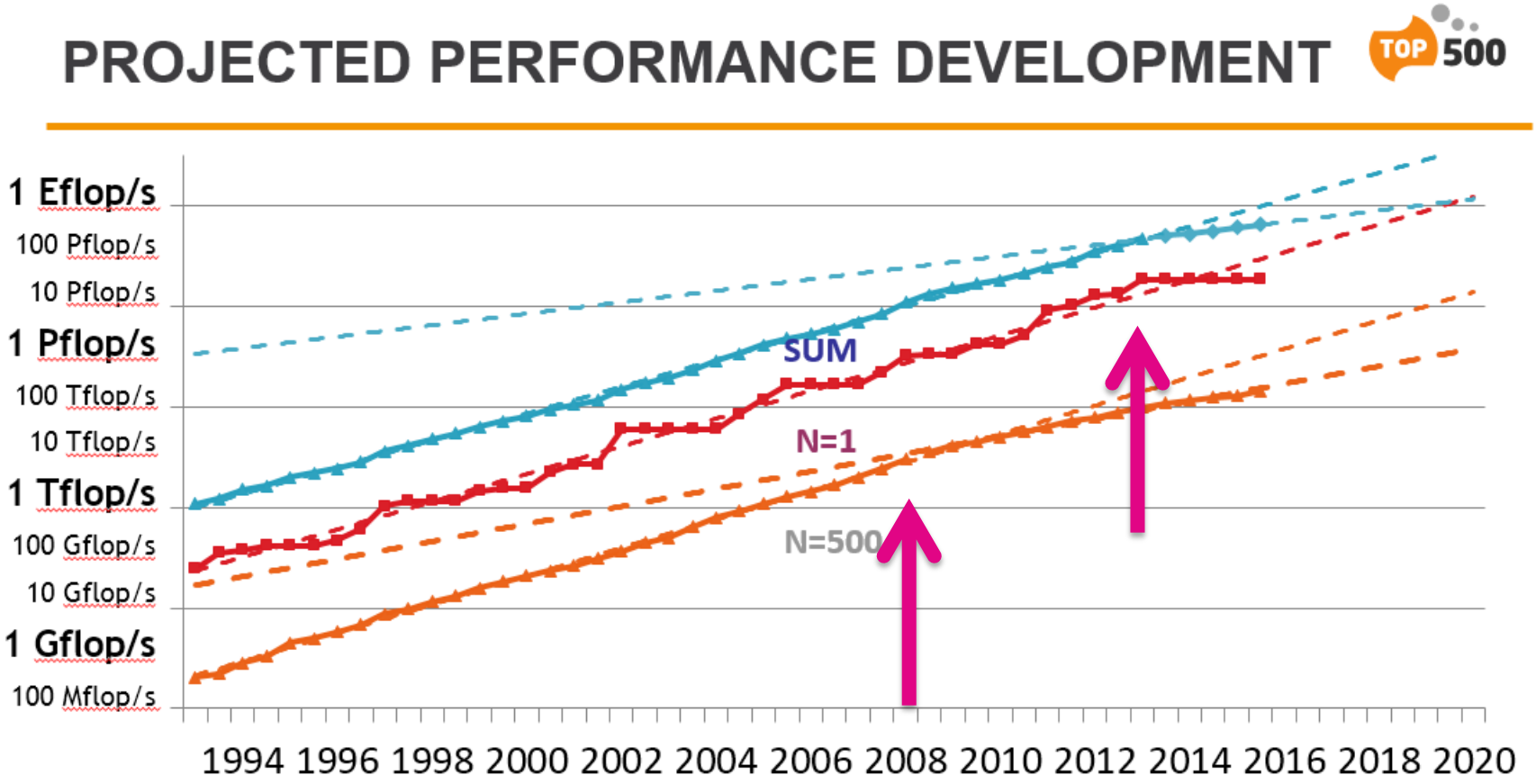
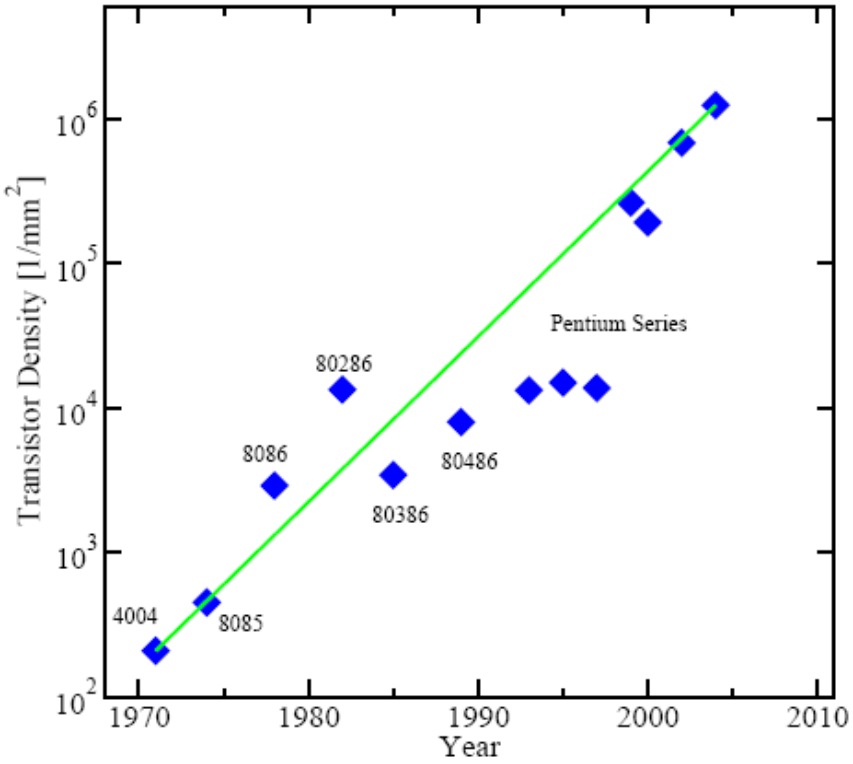
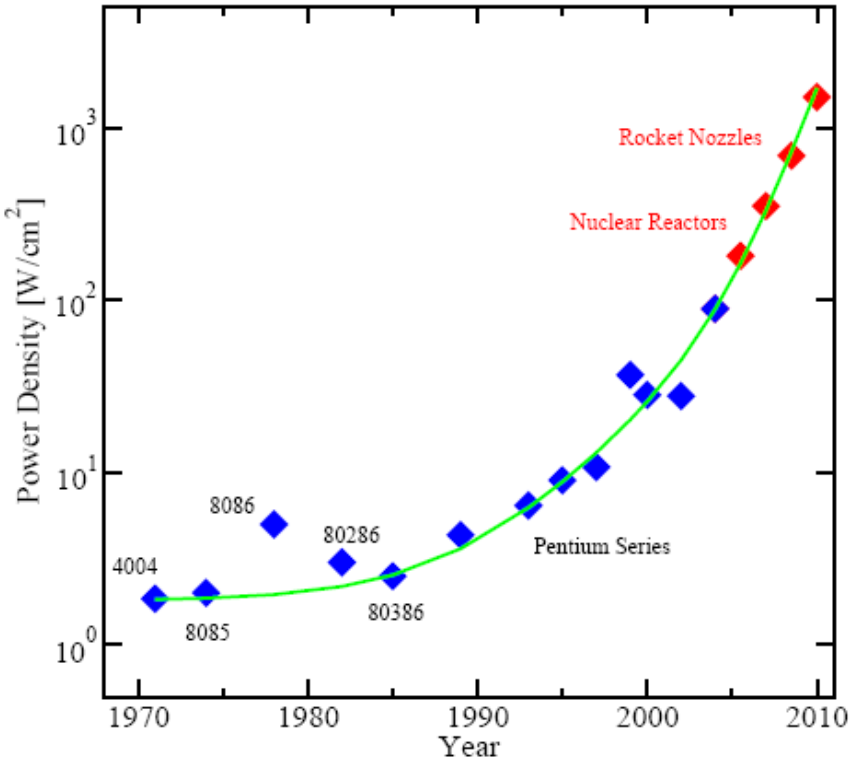


Image: E. Strohmaier / HPC Wire

The power problem



(a) Transistor integration density per die



(a) Power loss density per die

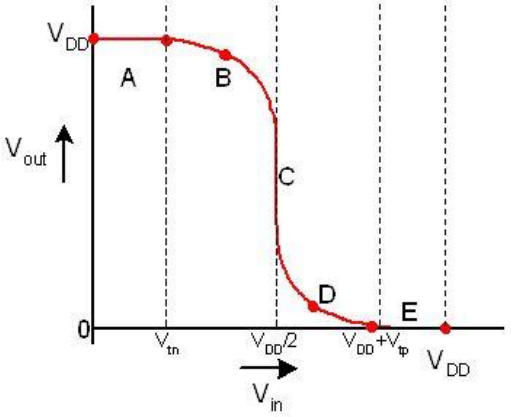
Image: TU Wien

Transistor operation

Operating Regions

- Revisit transistor operating regions

Region	nMOS	pMOS
A	Cutoff	Linear
B	Saturation	Linear
C	Saturation	Saturation
D	Linear	Saturation
E	Linear	Cutoff



CMOS VLSI Design Slide 24

Image: Internet

Transistor power

$$P = ACV^2 f + tAVI_{short} f + VI_{leak}$$

$$f_{max} = \frac{(V - V_{threshold})^2}{V}$$

Transistor power facts

Dynamic power scales with the square of the voltage

- Need to reduce V as much as possible

Transistors have a gate threshold voltage for switching from off to on

- Can't reduce V too much

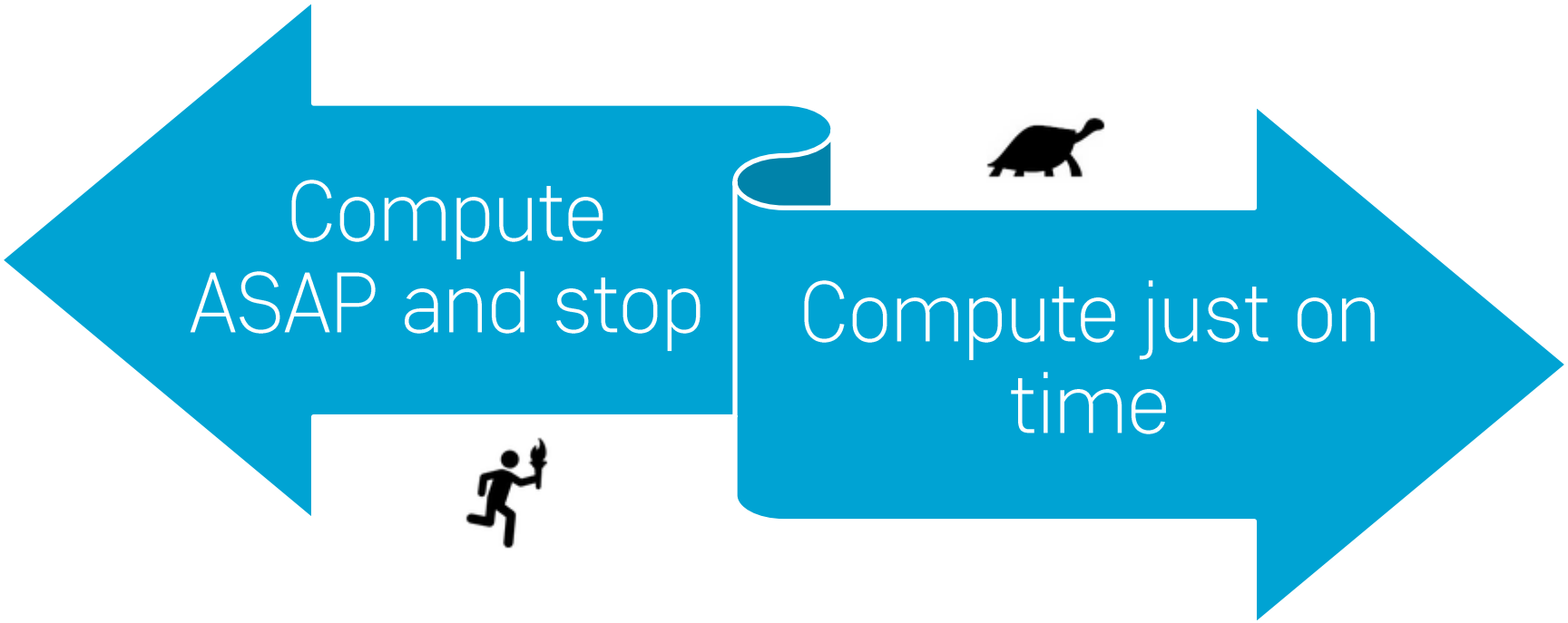
Switching speed ~inversely proportional to V

- Can't reduce V too much

Transistor power problems

- Too much power lost through “high” voltage
  - Dynamic voltage scaling
- Power consumption vs frequency goes with the power law
  - Reduce frequency statically
  - Reduce frequency dynamically
- Switching delays. Solutions:
  - Make wires shorter
  - Make gates smaller
  - Use faster materials

Computation strategies



Dark Silicon

- Systems increasingly limited by power consumption, not number of transistors
- ➔ "**Dark Silicon**" : Most of the chip will be OFF to meet thermal limits

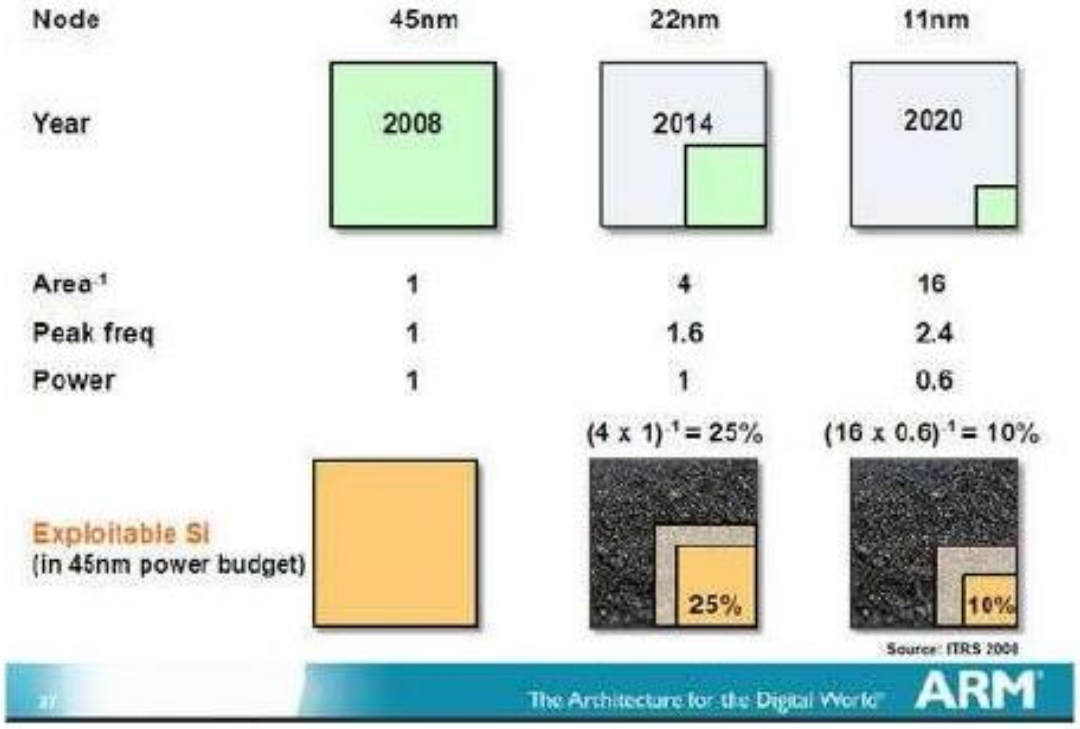


Image: ARM/Cambridge U

Alleviating the problem

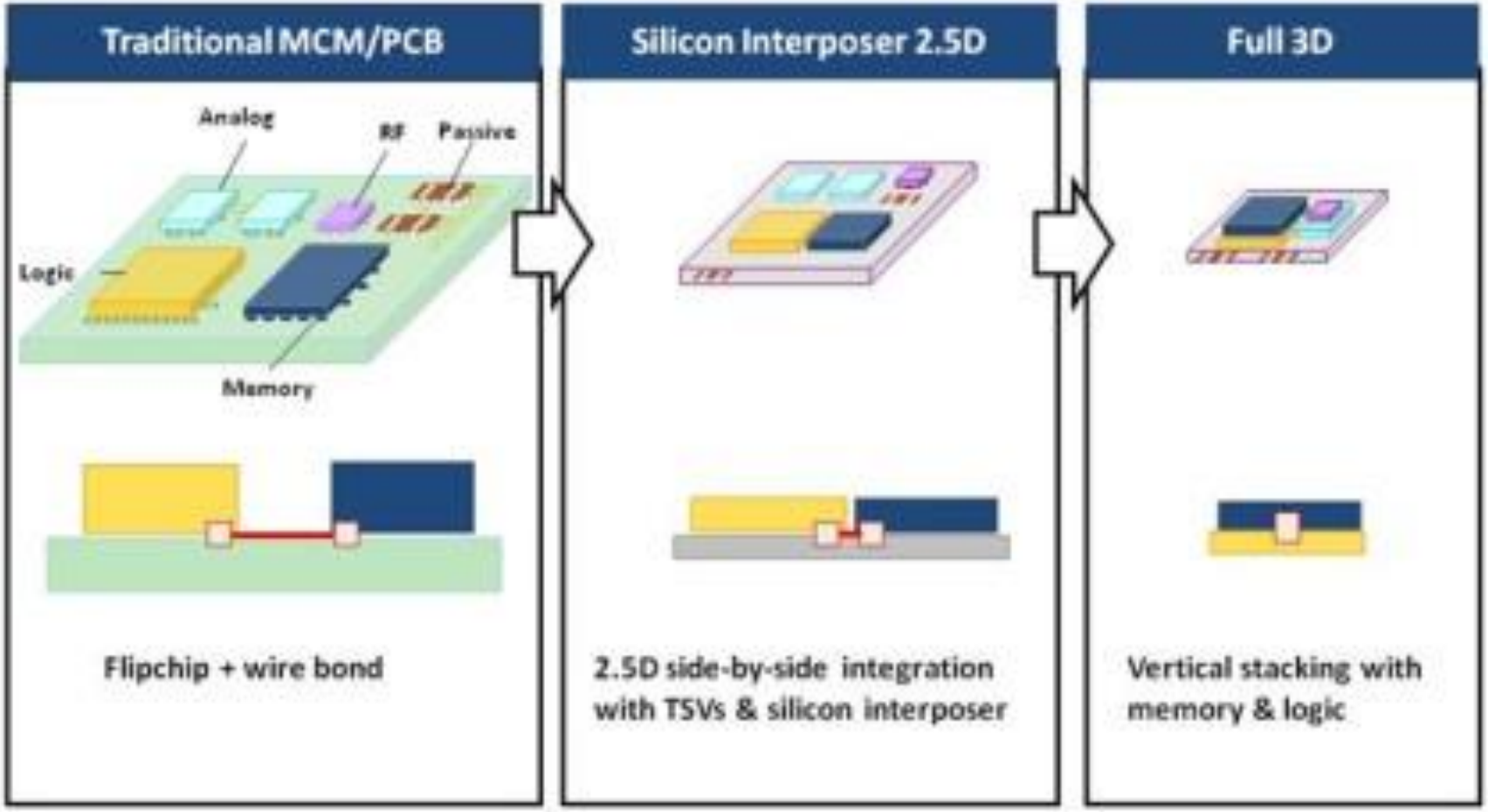


Image: EE Times

Energy-efficient Components

Reconfigurable logic
FPGAs

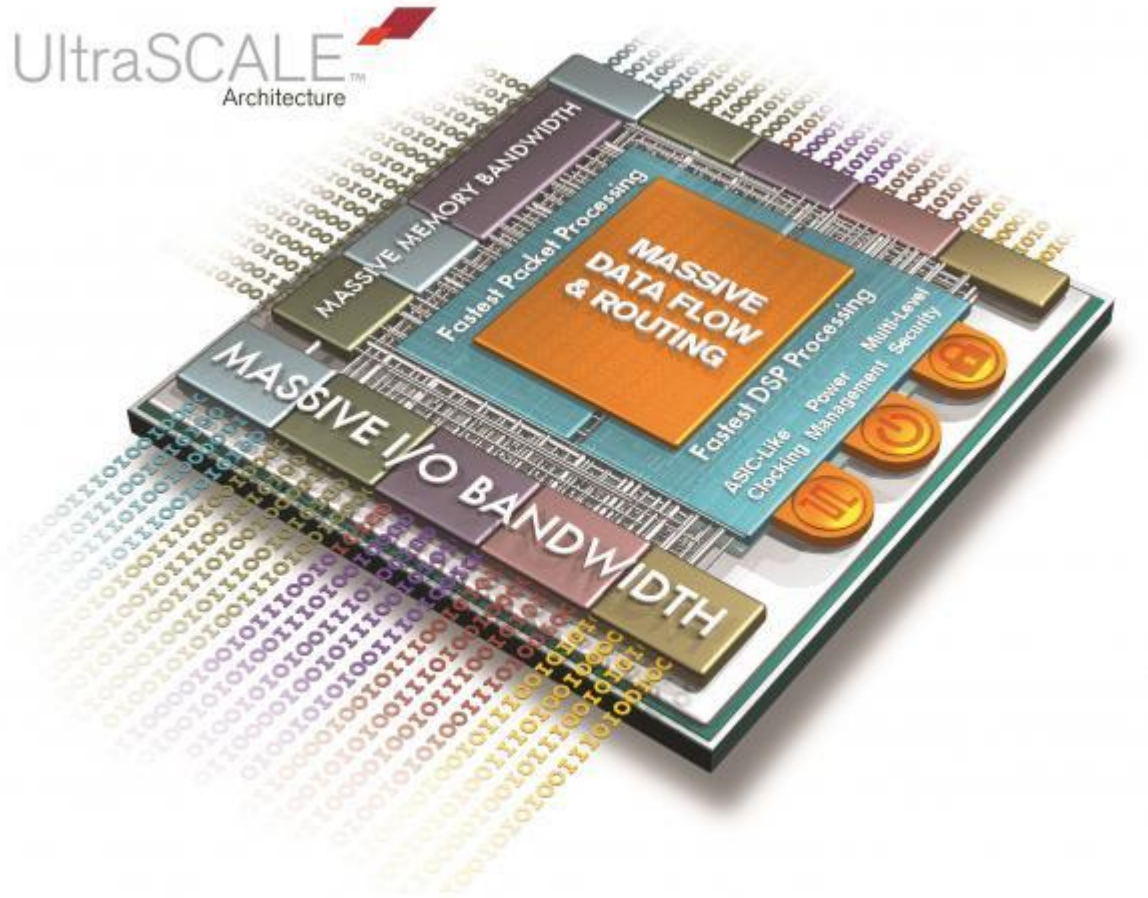
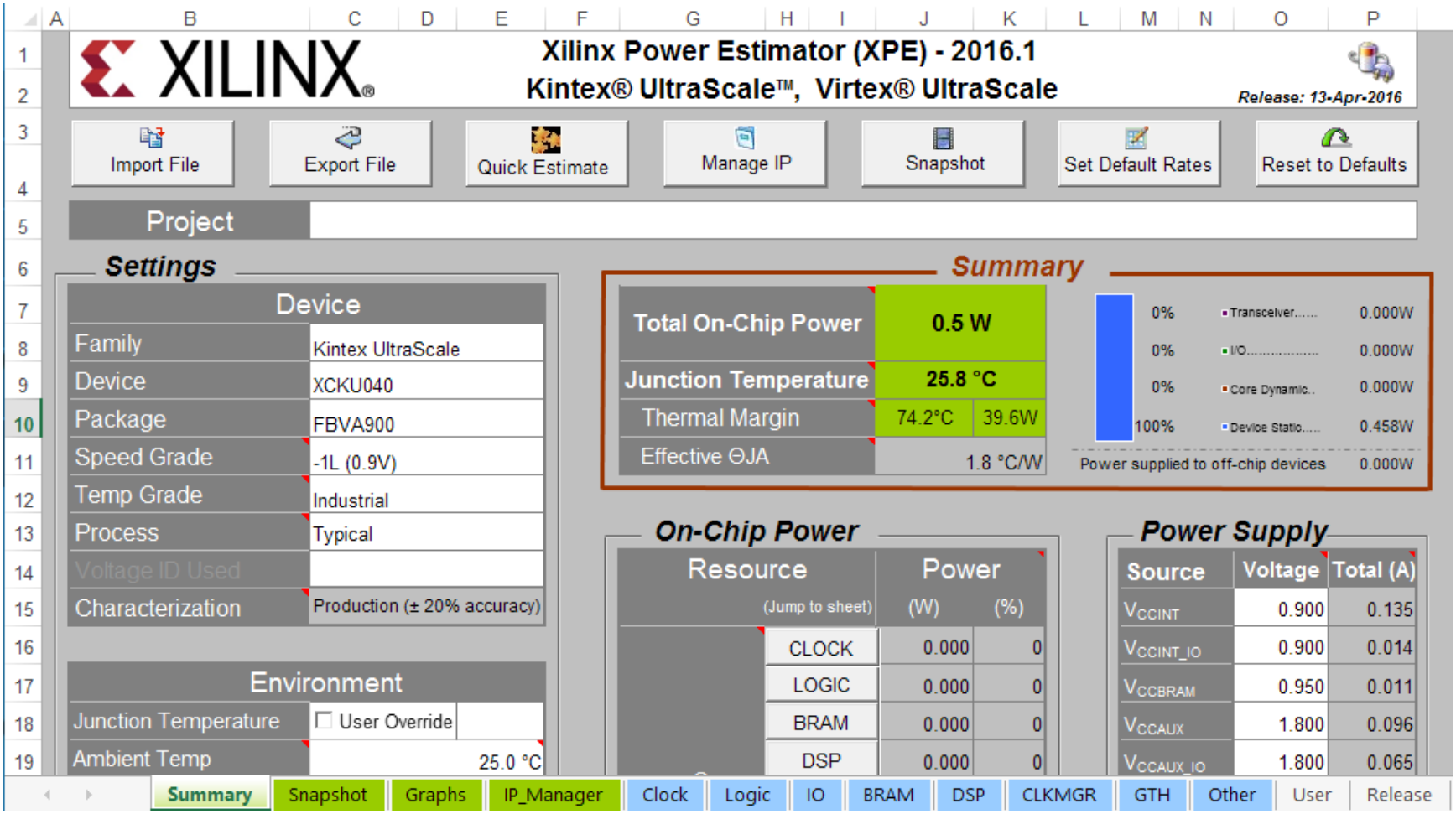


Image: Xilinx

Reconfigurable logic
Power constraints as seen by Xilinx



Reconfigurable logic
Environmental constraints as seen by Xilinx

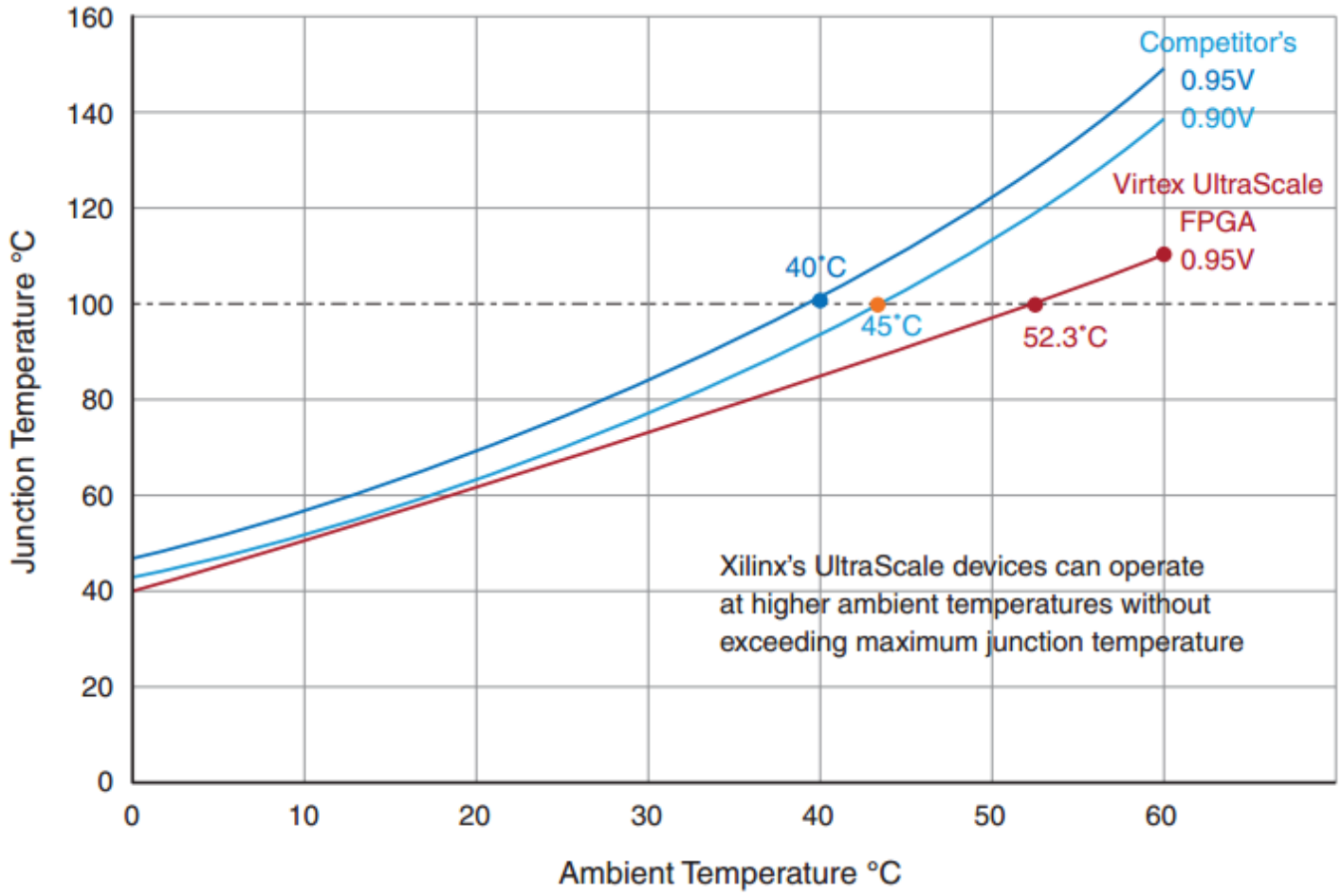


Image: Xilinx

Reconfigurable logic

Environmental constraints as seen by Xilinx

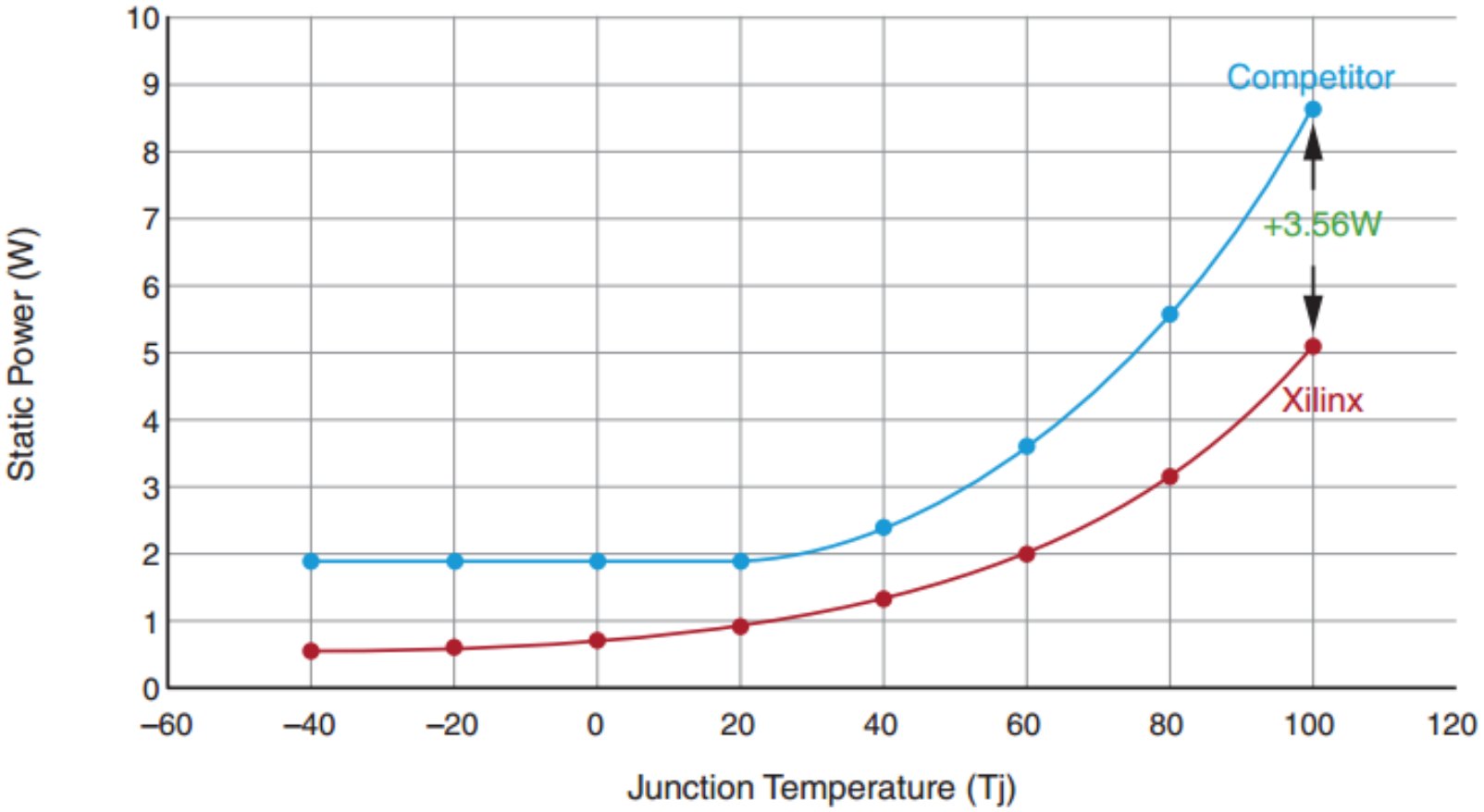


Image: Xilinx

Reconfigurable logic

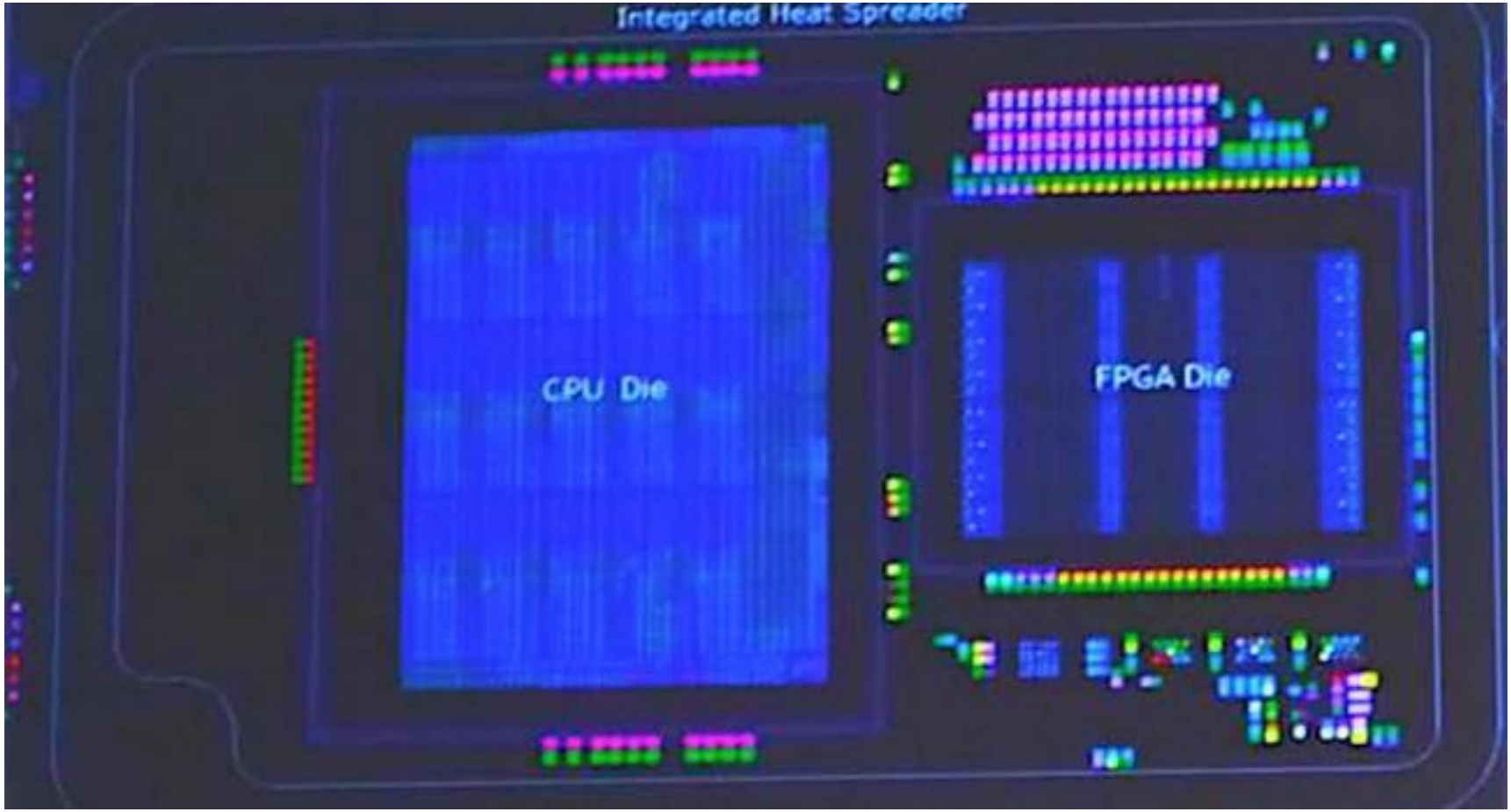


Image: The Next Platform

SoC

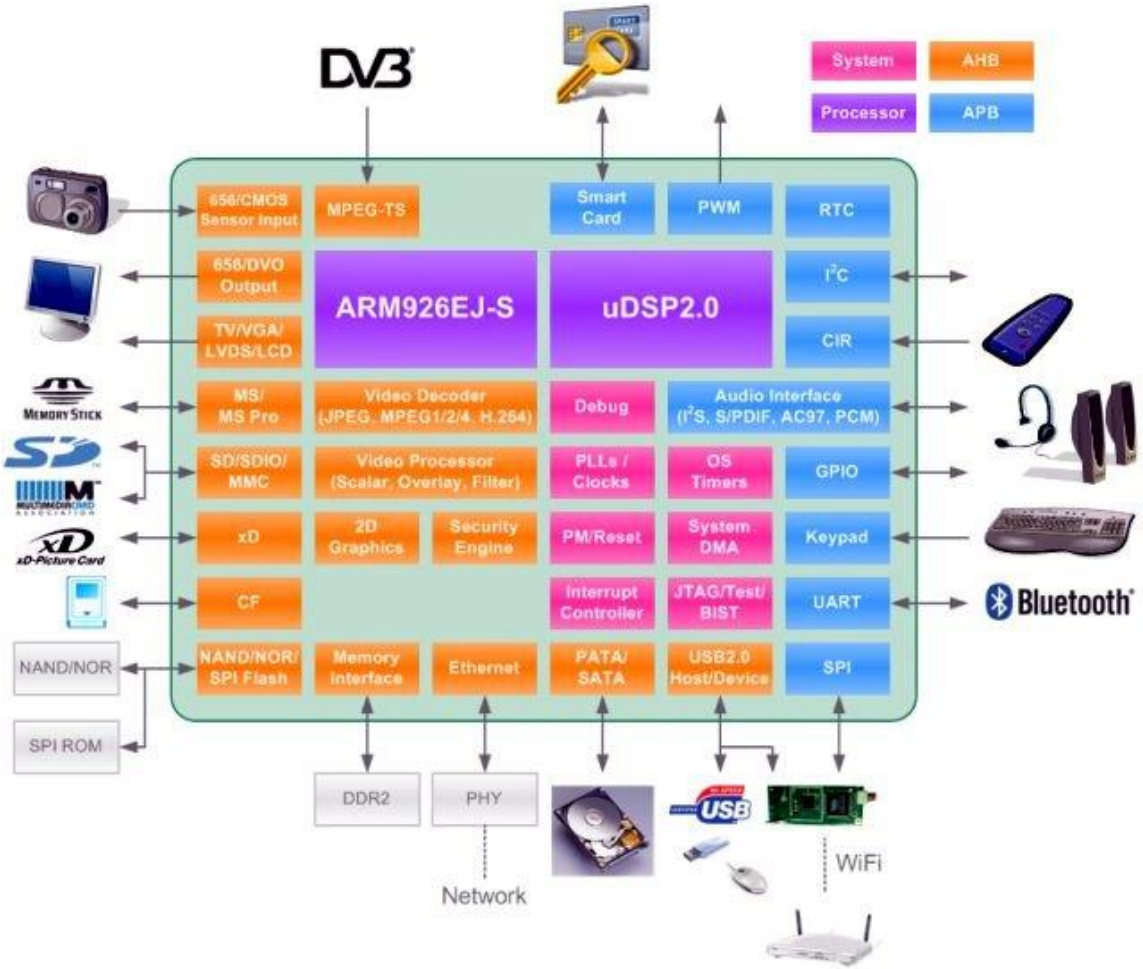
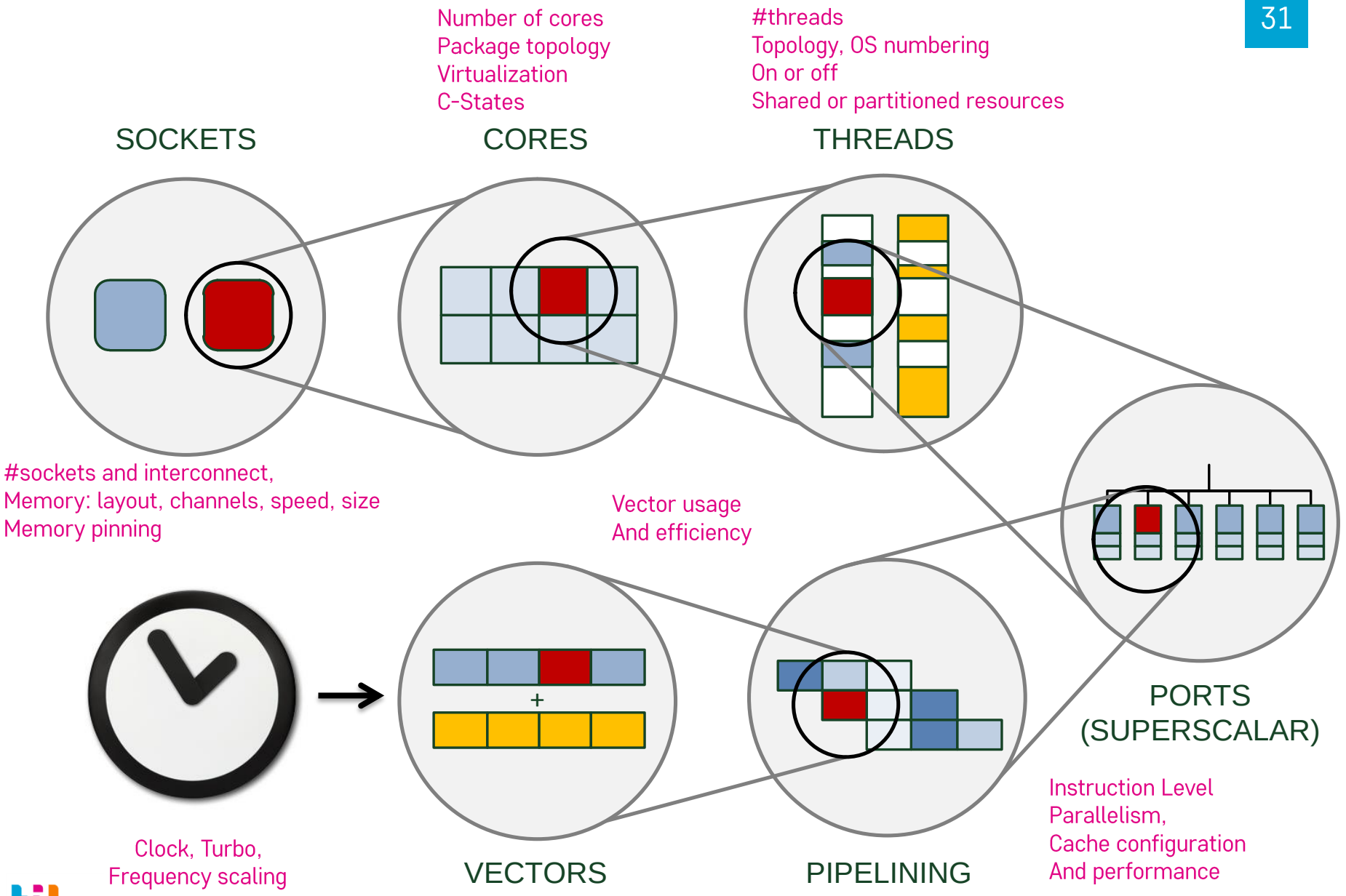


Image: ARM

Rough CPU energy breakdown

- Clock distribution ~10%
- L1/L2 ~20%
- Exec ~35-40%
- Routing/Links ~25%
- Other ~10%


Number of cores
Package topology
Virtualization
C-States
#threads
Topology, OS numbering
On or off
Shared or partitioned resources
SOCKETS
CORES
THREADS
#sockets and interconnect,
Memory: layout, channels, speed, size
Memory pinning
Vector usage
And efficiency
PORTS
(SUPERSCALAR)
Instruction Level
Parallelism,
Cache configuration
And performance
Clock, Turbo,
Frequency scaling
VECTORS
PIPELINING

CPU C-States
Idle power and latency

Table 1: CPU Idle Power States in Nexus 4.

Idle State	Name	Idle System Power (mW)	Latency (μS)†
C0	Wait for Interrupt	433	1
C1	Retention	390	415
C2	Power Collapse Standalone	330	1300
C3	Power Collapse	200	2000
Without entering idle states		1,060	0

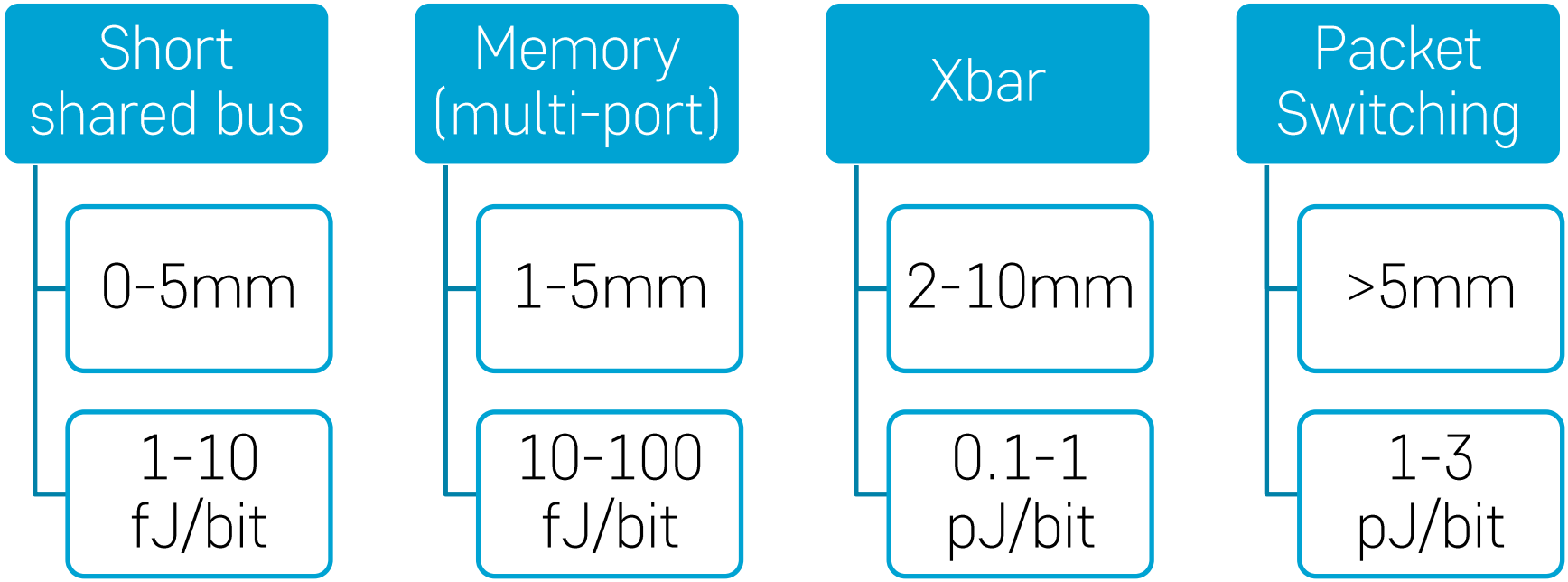
†: The data is obtained from the Nexus 4 kernel source code.

Image: Microsoft

Energy per operation

- Theoretical FLOPS
  - Intel Westmere: 1.7nJ/flop
  - Intel Haswell: 110pJ/flop
  - NVIDIA Fermi: 225pJ/flop
  - ARM (Cortex-A7): 90-150pJ/flop
  - ARM (Cortex-A15): 250-1200pJ/flop
  - Exascale target for ops: 20pJ/flop (= Exa @20MW)
- Communication :
  - Core-to-core: ~10pJ/byte
  - Chip-mem: ~150pJ/byte
  - Chip-chip: ~100pJ/byte

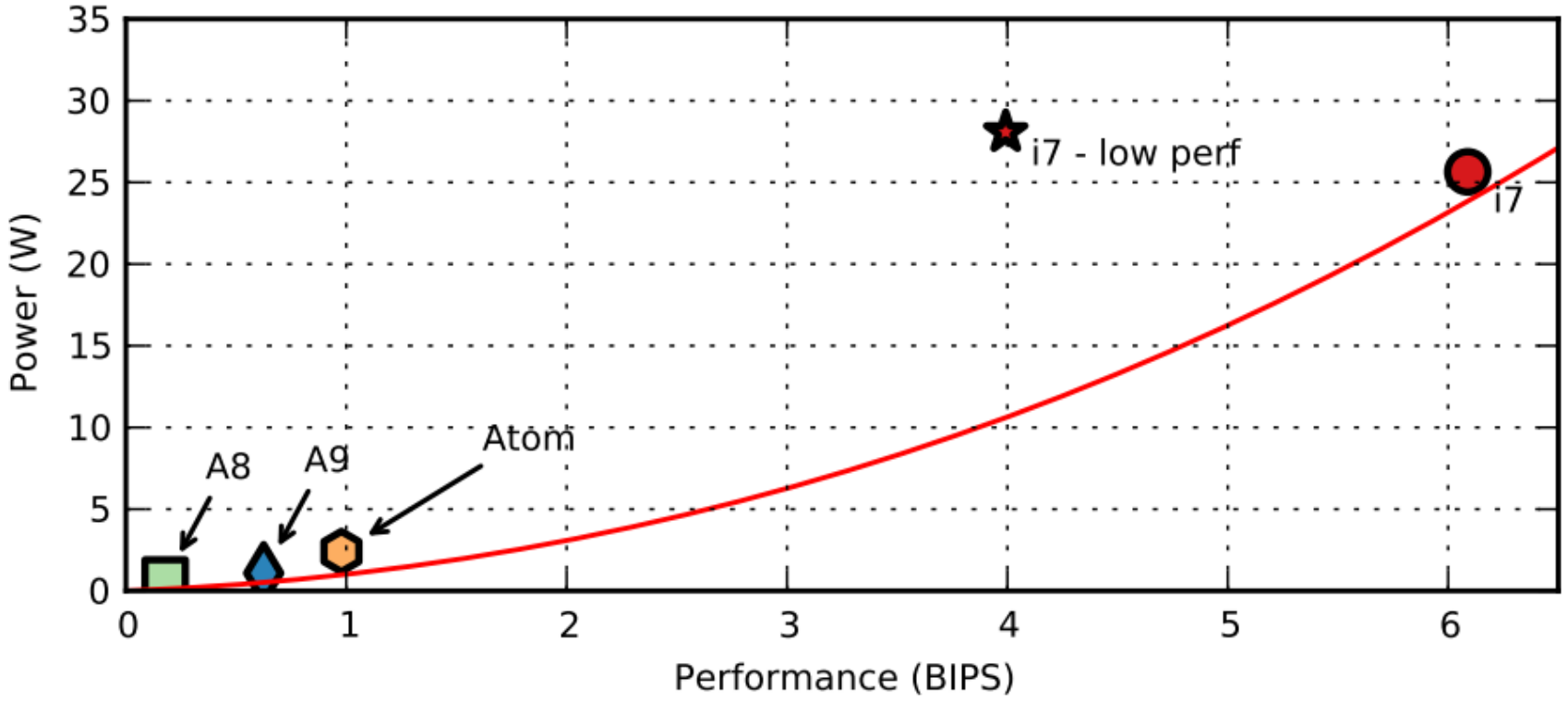
Interconnect energy



Storage energy cost

	SRAM	DRAM	FLASH	Disk
Energy/bit	1 pJ/bit	100 pJ/bit	500 pJ/bit	100nJ/bit

Intel Architecture vs. ARM
Power



RAM energy cost
All about density

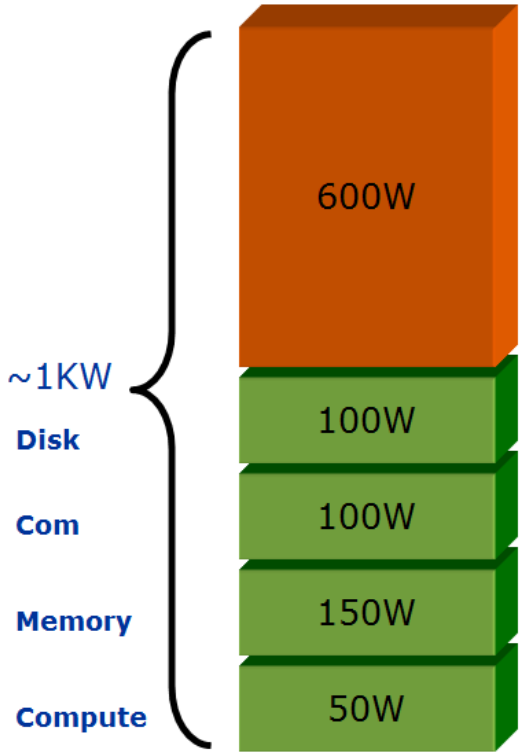
Table 2: Energy Consumption of Memory Modules (Source: ATP)

Memory Module	Capacity (GB)	Memory Type	Power Consumption (W)	Power Consumption (W/GB)
AL48M72F4GKF8S	16	DDR3, Registered, ECC, 4 rank	8.710	0.54
AL24M72E4BKH9S	8	DDR3, Registered, ECC, 2 rank	6.132	0.77
AL12M72B8BKH9S	4	DDR3, Registered, ECC, 2 rank, memory chips with 256Mx8 organization (2-Gb chips)	2.934	0.73
AL56M72B8BJH9S	2	DDR3, Registered, ECC, 2 rank, memory chips with 128Mx8 organization (1-Gb chips)	5.132	2.57
AL28M72A8BJH9S	1	DDR3, Registered, ECC, 1 rank	2.241	2.24
AQ12M72E8BKH9S	4	DDR3, Unbuffered, ECC, 2 rank, memory chips with 256Mx8 organization (2-Gb chips)	2.214	0.55
AQ28M72D8BJH9S	1	DDR3, Unbuffered, ECC, 1 rank	1.387	1.39
AJ28K72F8BJE6S	1	DDR2, Unbuffered, ECC	1.872	1.87
AP56K72G4BHE6S	2	DDR2, FB-DIMM, ECC	13.683	6.84

Image: ATP

Energy spending in a server
Today

Teraflop system today



Decode and control
Address translations…
Power supply losses
Bloated with inefficient architectural features

10TB disk @ 1TB/disk @10W

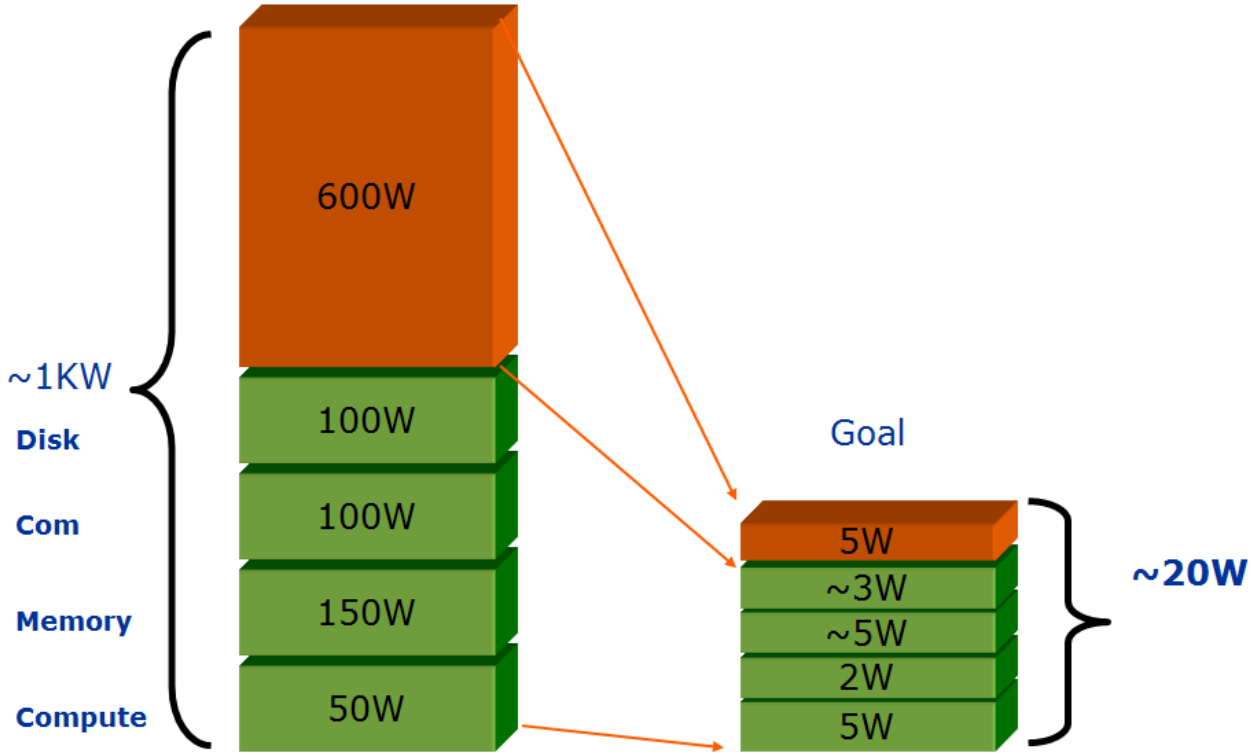
100pJ com per FLOP

0.1B/FLOP @ 1.5nJ per Byte

50pJ per FLOP

From S. Borkar / Intel

Energy spending in a server
Goal



From S. Borkar / Intel

Thank you

e-mail: an@tik.services

http://tik.services