Energy Efficiency in Computing (2)

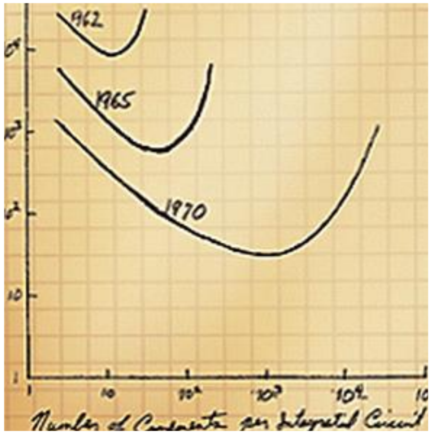
CERN Academic Training – May 2016

Andrzej Nowak

http://tik.services

tik.
technology
innovation
knowledge

Outline

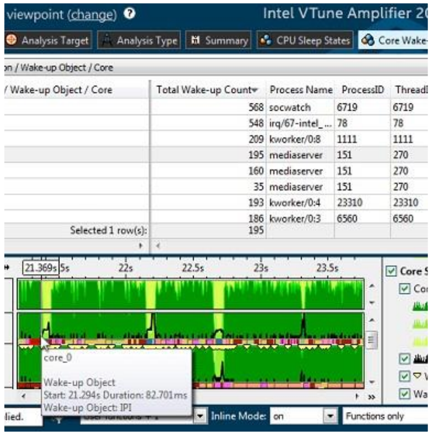


Day 1:

Silicon, hardware

Day 2:

Datacenters, software, future technologies



Energy-efficient Datacenters

Top500 power efficiency

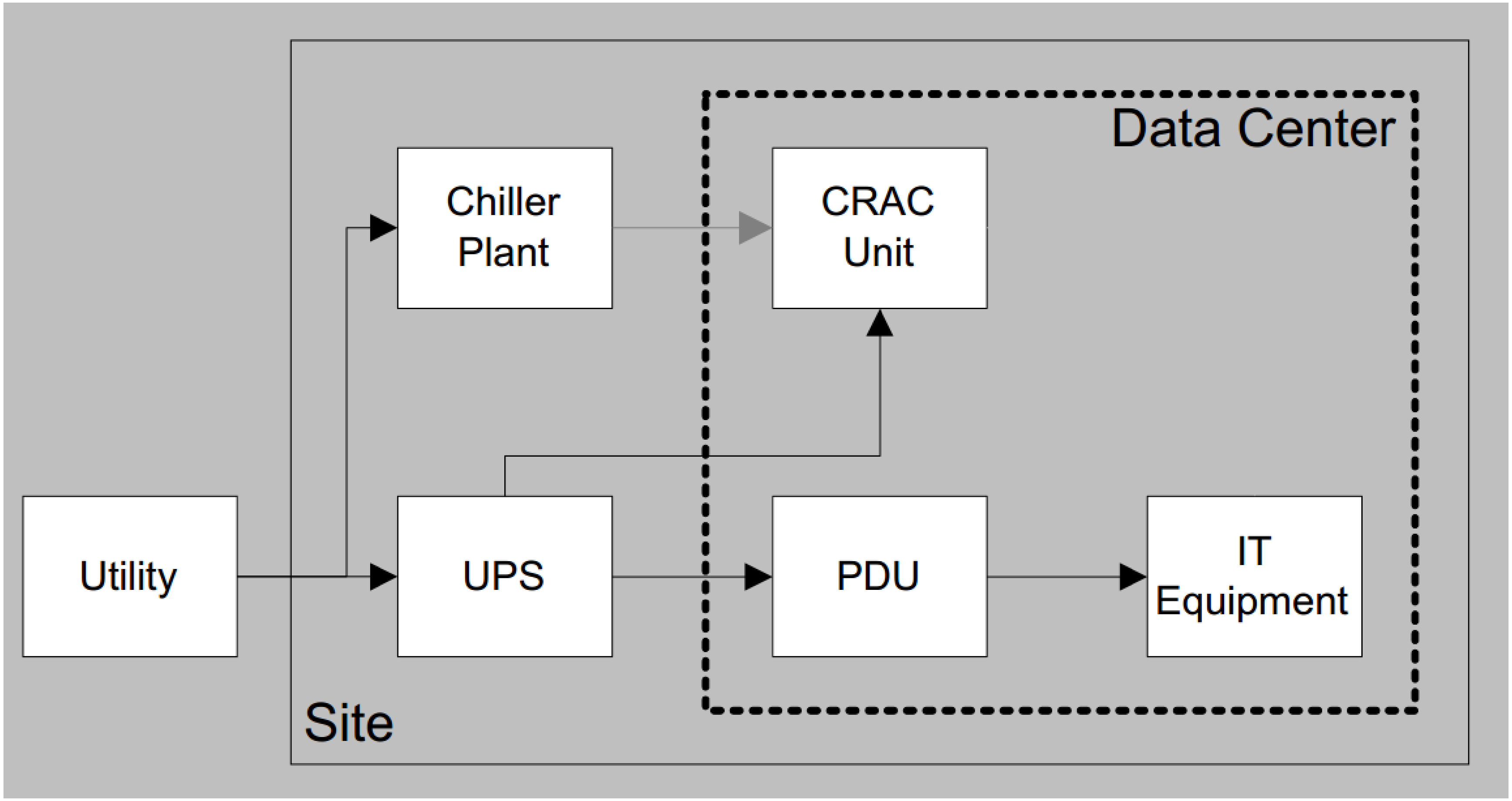
MOST POWER EFFICIENT ARCHITECTURES TOP 500

Computer	Rmax/Power
Tsubame KFC/DL, NEC, Xeon 6C 2.1GHz, IB FDR, **NVIDIA K80**	**4,856**
Sugon Cluster W780I, Xeon 8C 2.6GHz, IB QDR, **NVIDIA K80**	**4,778**
Inspur TS10000 HPC Server, Xeon 6C 2.4GHz, 10GE, **NVIDIA K40** (Multiple)	**4,497** (best)
Suiren, Xeon 10C 2.2GHz, IB FDR **PEZY-SC**	**4,044**
Taurus GPUs, Bull R400, Xeon 12C 2.5GHz, IB FDR, **NVIDIA K80**	**3,277**
Sango, Supermicro, Xeon 12C 2.5GHz, IB FDR, **Intel Phi**	**3,223**
XingGui, Dell, Xeon 10C/8C 2/2.6GHz, IB FDR, **NVIDIA K40m/K20m**	**3,187**
Romeo, Bull Cluster, Xeon 8C 2.6GHz, IB FDR, **NVIDIA K20x**	**3,131**
Sekirei-ACC, SGI ICE XA, 12C 2.5GHz, IB FDR, **NVIDIA K40**	**3,045**
HA-PACS TCA, Cray Cluster, Xeon 10C 2.8GHz, QDR, **NVIDIA K20x**	**2,980**
SANAM, Adtech, ASUS, Xeon 8C 2.0GHz, IB FDR, **AMD FirePro**	**2,973**

[Mflops/Watt]

Image: E. Strohmaier / HPC Wire

Datacenters



From M.K. Patterson/Intel

Datacenters
Metrics that matter

Power Usage Effectiveness

$$PUE = \frac{Total\ Power}{IT\ Power}$$

Server-PUE

$$ITUE = \frac{Infrastructure\ Burden + Compute}{Compute}$$

Total Usage Effectiveness

$$TUE = PUE\ x\ ITUE$$

Datacenter choices
From M.K. Patterson/Intel

- Air or liquid cooling? What kind? Where does it come from?
- Hot- or cold-aisle?
- What kind of floor, is it raised?
- Modular or not?
- What kind of UPS?
- What kind of rack density?
- Material vs. TCO cost

Image: CERN

Thermal control

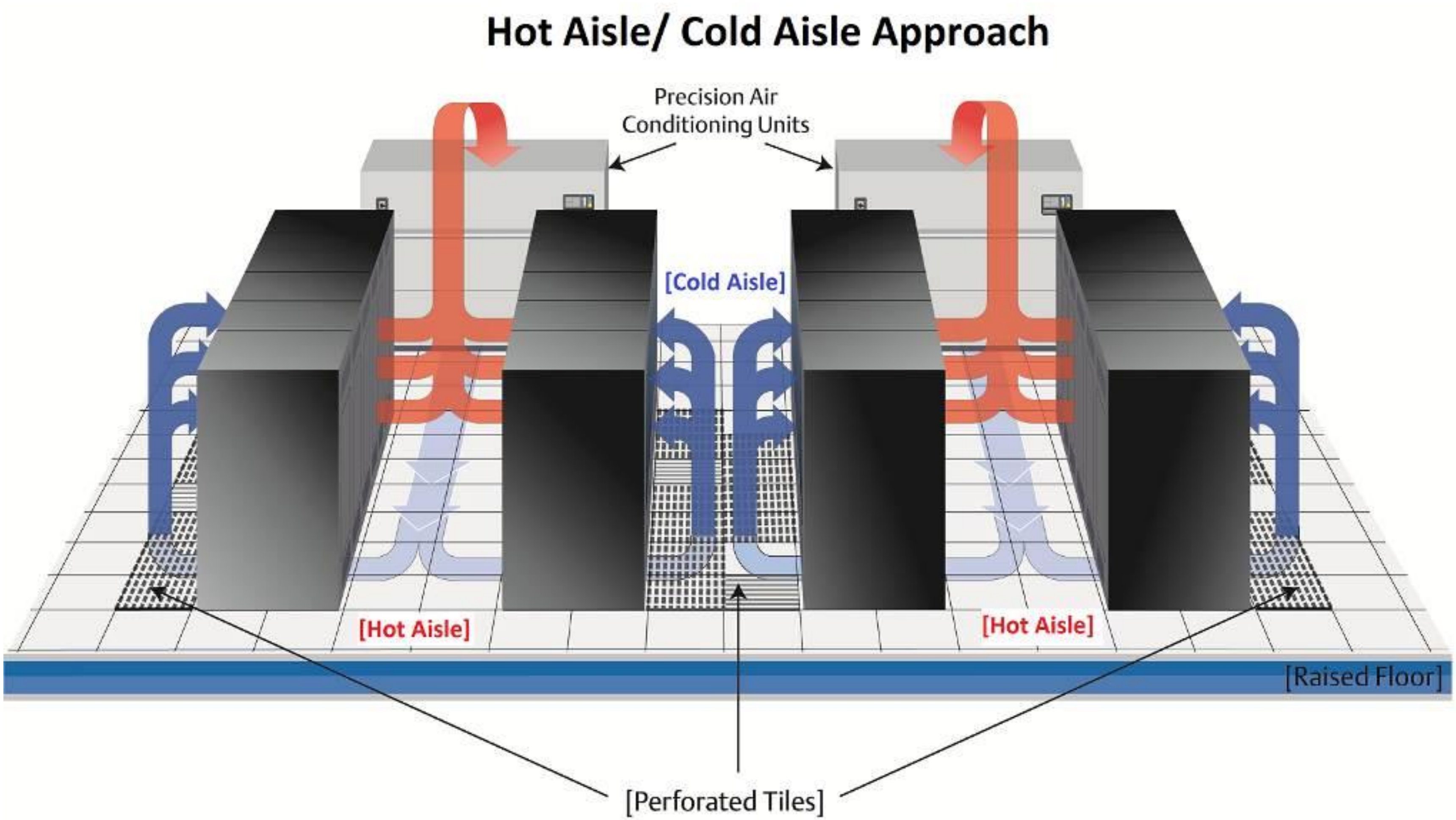


Image: CW

Thermal debugging

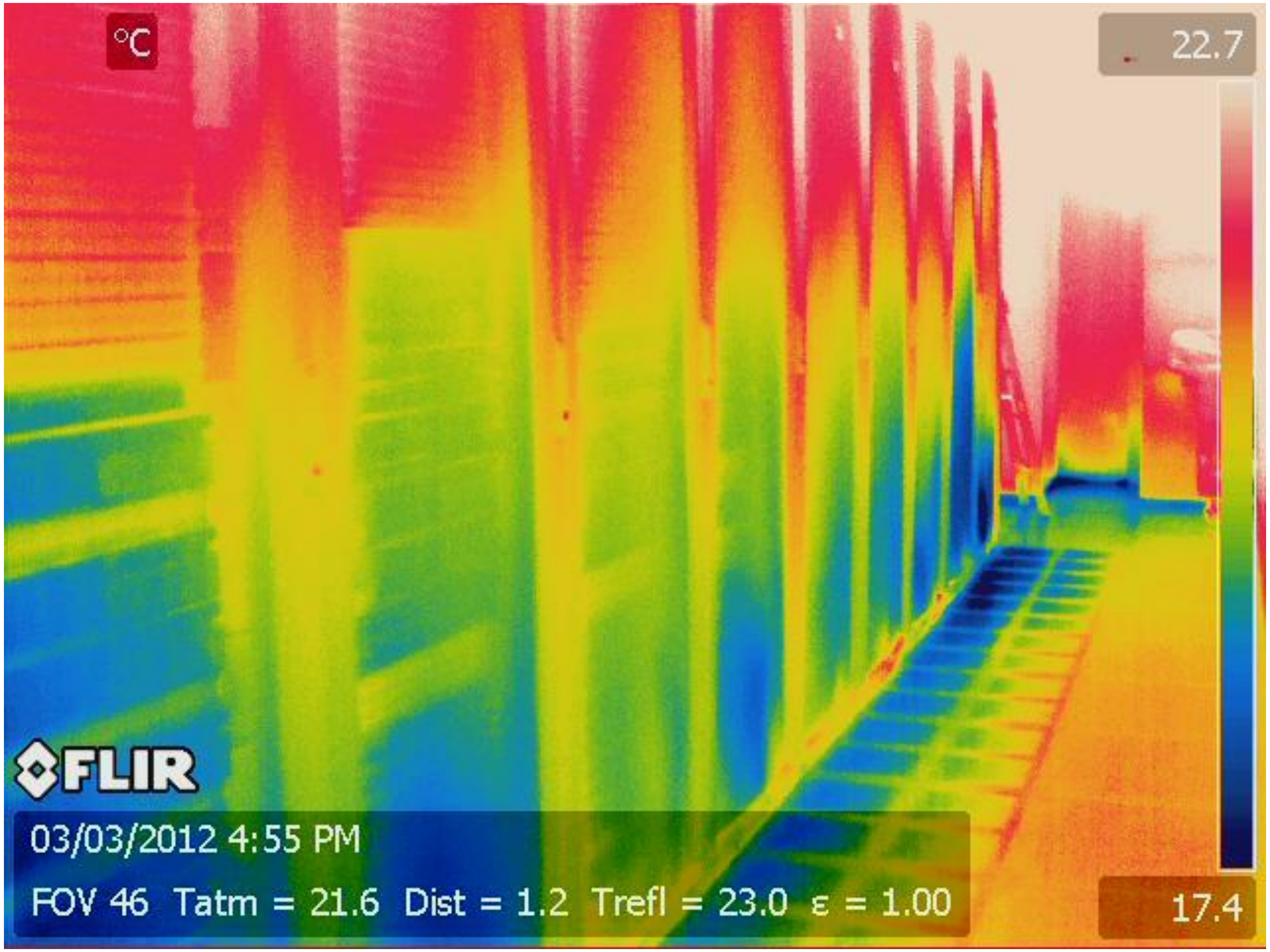


Image: IR Imaging Services

Thermal debugging

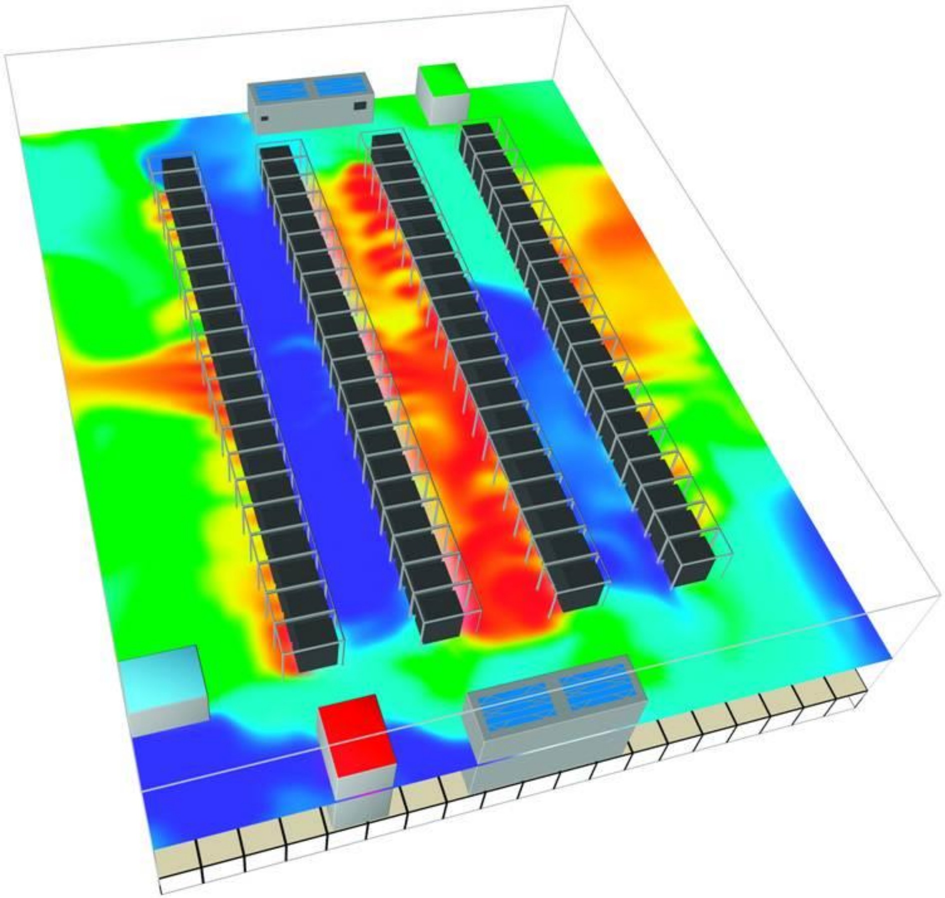


Image: CW

Creative solutions for datacenters
Submersion cooling

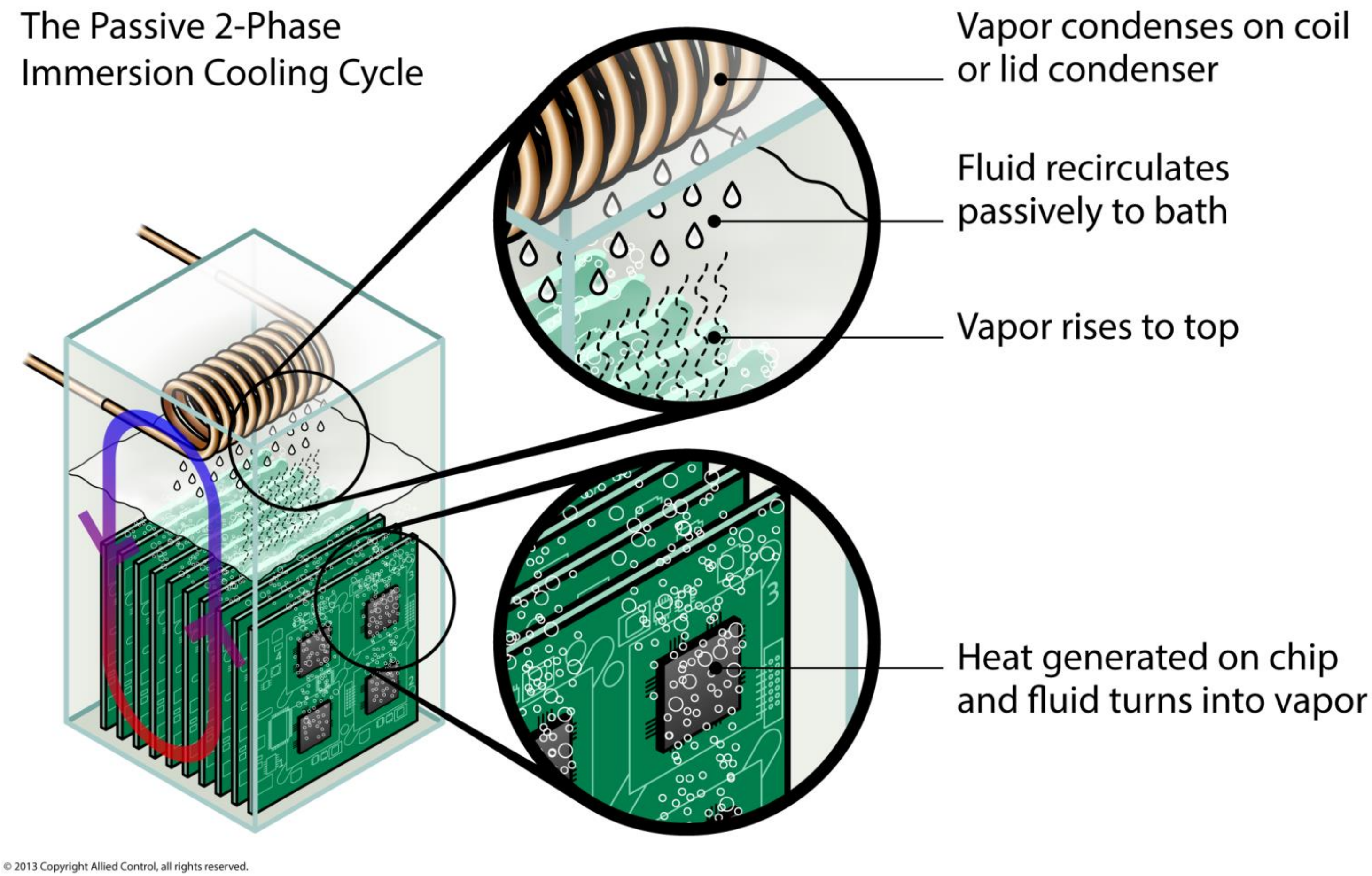


Image: Allied Control

Creative solutions for datacenters

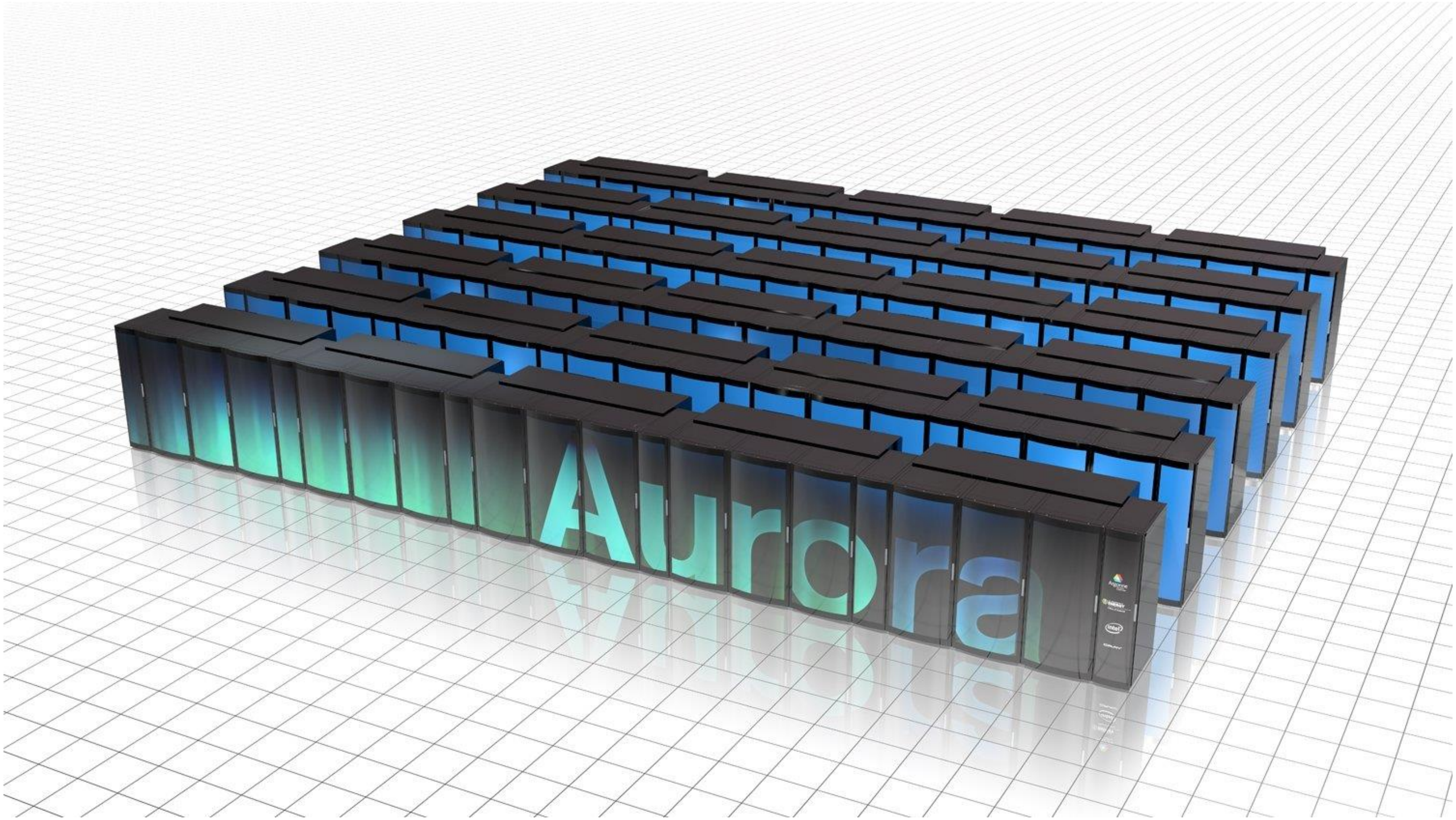


Image: Cray/Intel

Creative solutions for datacenters

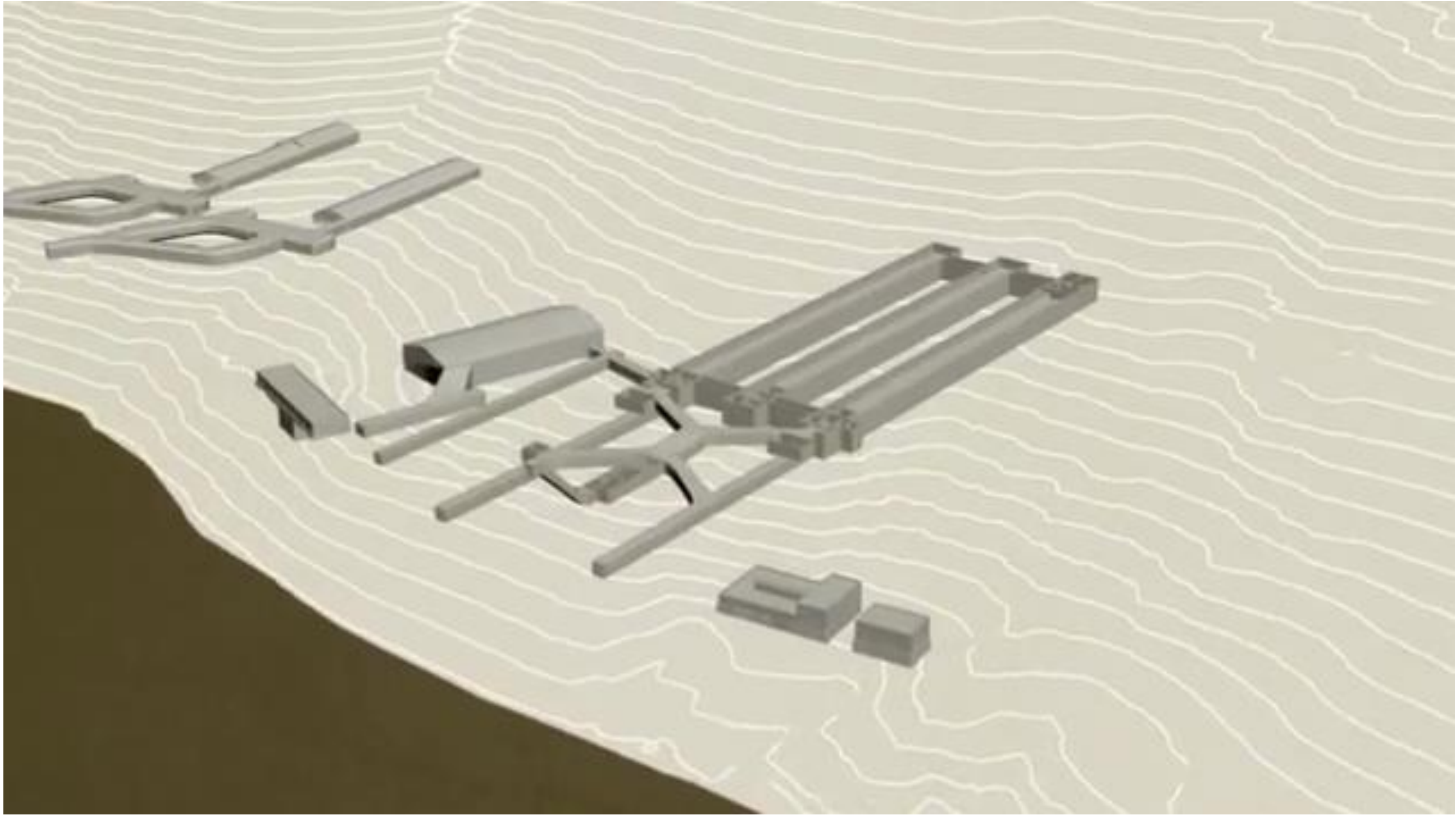


Image: Green Mountain

Not-so-creative solutions?

- Ultimately - density:
  - In-package memory, stacked (2.5D or 3D)
  - Integrated fabric/networking
  - Higher package integration
  - Switching closer to compute
  - Si-Ph – cost benefits, but power performance a question
- As well as:
  - Metrics and research
  - Power
  - Cooling optimization

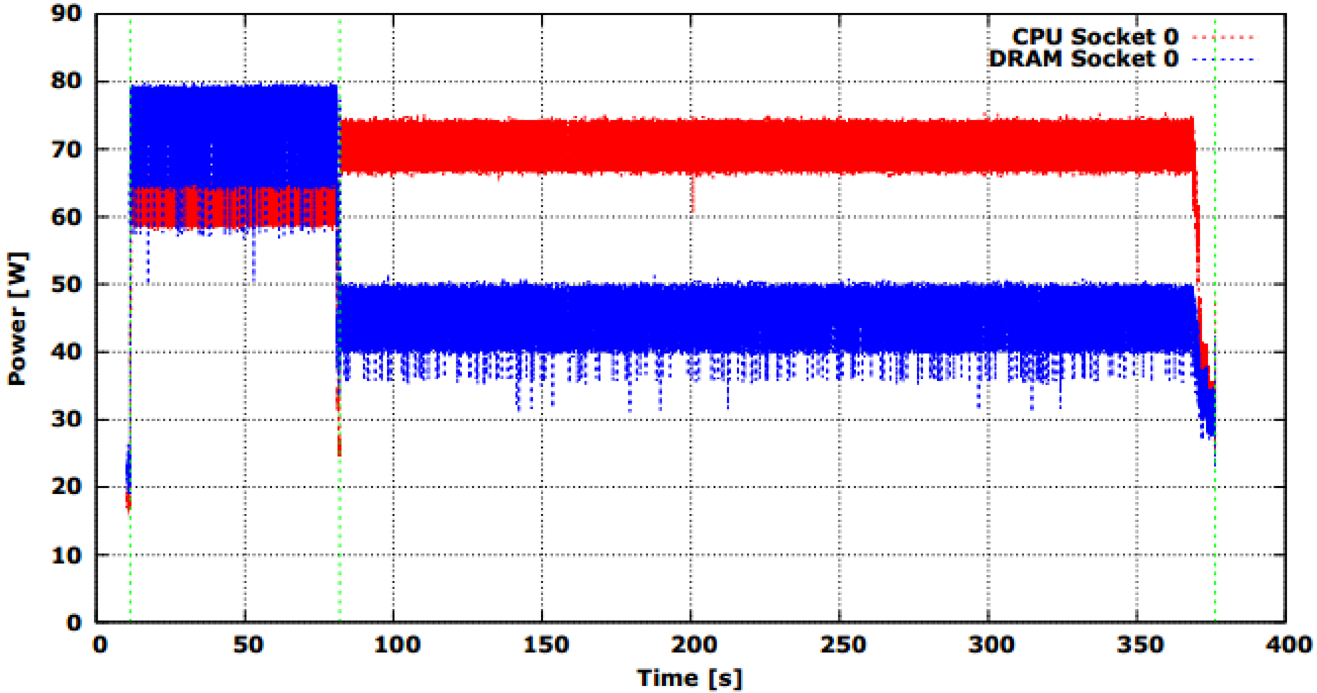
Energy-efficient Software

Energy and code

Image: E. Calore / INFN / COLA

Energy and code

Acquired data example using RAPL counters



Intel Haswell CPU energy counters acquired at 100Hz and converted in Watt; acquisition performed with a custom developed wrapper to the PAPI library.

Image: E. Calore / INFN / COLA

Energy per instruction?
Example study

Instruction	Cortex-A7		Cortex-A15	
	min EPI	max EPI	min EPI	max EPI
Simple Integer	50	80	200	450
Simple Float/Double	90	200	250	1500
Multiplication	80	340	360	1730
Division	150	1200	1270	1960
Load (L1 hit)	150	195	450	450
Store (L1 hit)	185	195	680	750
Store (L1 miss)	200		700	
Load (L1 miss)	270		1000	

Table 7.1: Minimum (w/o RAW) and maximum (w/ RAW) Energy per Instruction (pJ) at 1GHz

Image: E. Vasilakis / FORTH

Energy per instruction?
Example study

Cortex-A7 Energy Per Instruction (pJ)

Instr. \ Freq. MHz	500	600	700	800	900	1000	1100	1200
add	63	62	61	64	72	82	94	105
and	54	53	52	54	61	69	79	89
eor	55	55	54	56	63	72	81	92
mul	116	114	112	116	128	146	166	189
orr	55	55	54	56	63	72	81	92
rsb	63	62	62	65	72	83	93	105
sub	64	63	62	65	73	83	94	105
div	178	174	170	177	195	221	251	286

Table 7.4: Integer logic and arithmetic instructions with 3 register operands with RAW dependencies

Image: E. Vasilakis / FORTH

Energy profiling

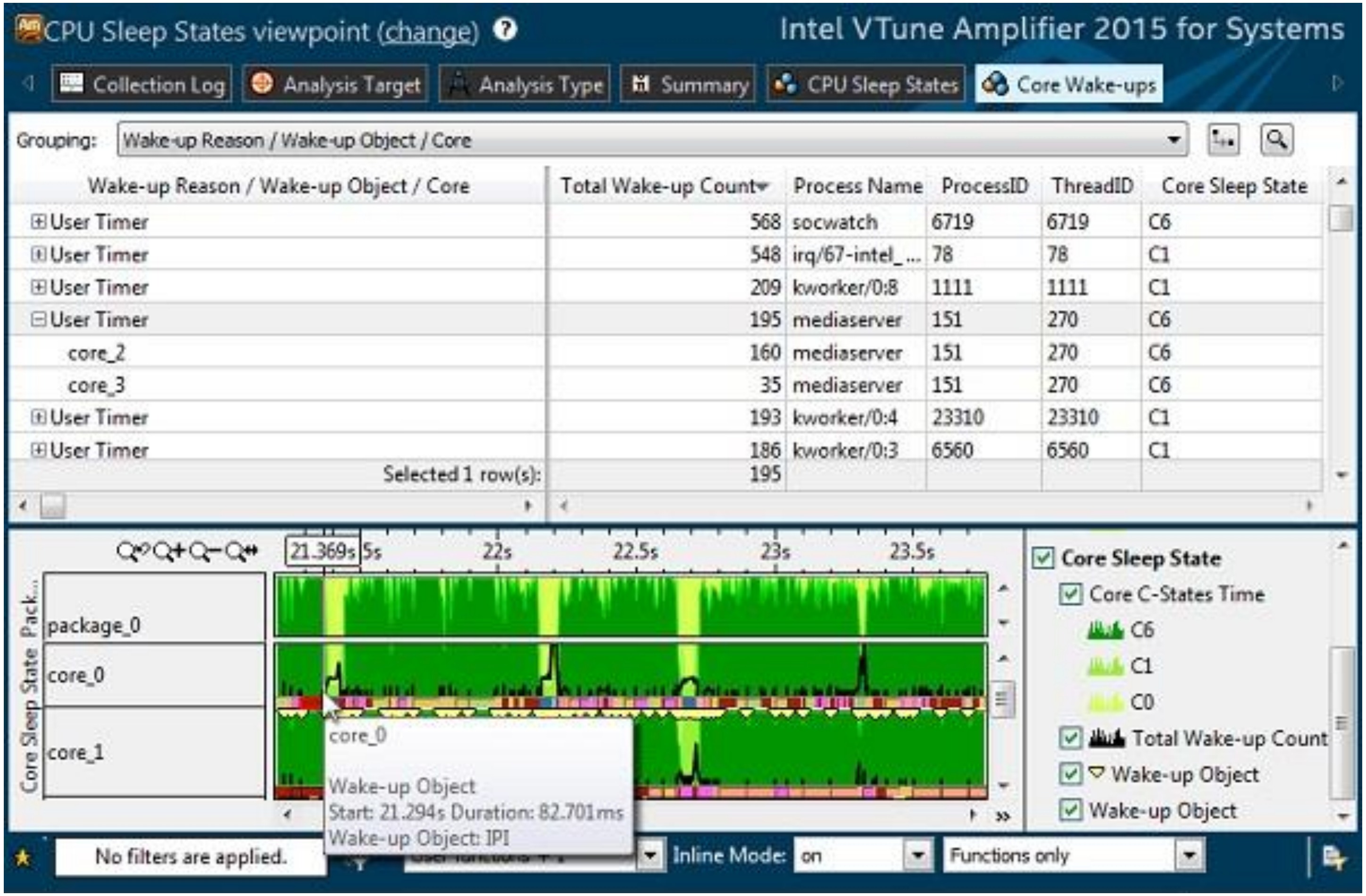


Image: Intel

Energy-aware scheduling
"Energy to solution"

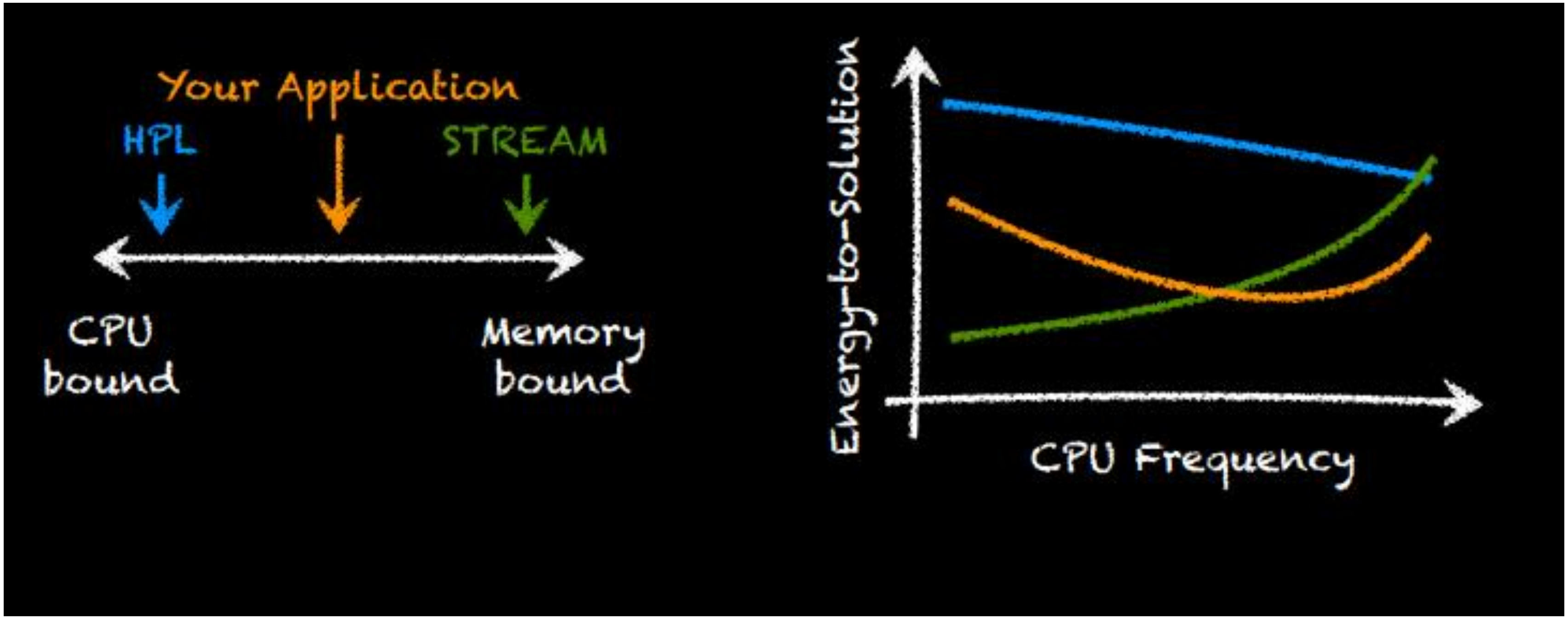


Image: LRZ

Energy-aware scheduling

"Energy to solution"

Conclusion

- LRZ Policy on SuperMUC is now:
  - No application tag: run @ default frequency (2.3 GHz)
  - With application tag:
    - Execute at 2.4 GHz if performance gain > 2.5%
    - Execute at 2.5 GHz if performance gain > 5%
    - Execute at 2.6 GHz if performance gain > 8.5%
    - Execute at 2.7 GHz if performance gain > 12%
- Applies to all jobs on SuperMUC
- Estimated energy savings ~5 %
- Big incentive for scientists to improve their codes!

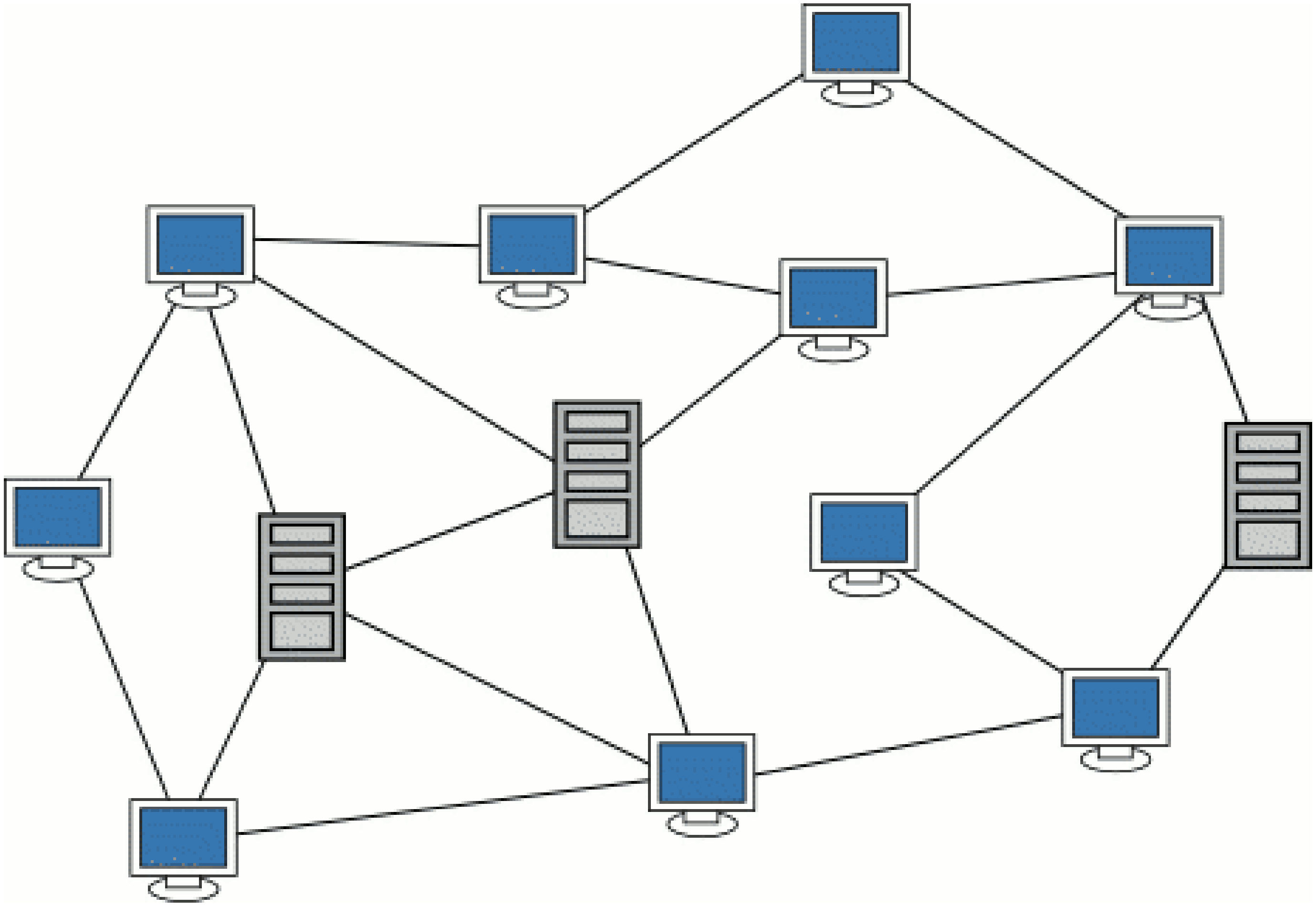
Image: LRZ

Future technologies, applications

From mainstream to exotic

“The Internet of Things”

Mesh networking/computing



"The Internet of Things"
Communication and energy

	Wi-Fi	Zigbee	Bluetooth Low Energy
Sleep	10 µW	4 µW	8 µW
Receive (Rx) Power	90 mW	84 mW	28.5 mW
Transmit (Tx) Power	350 mW	72 mW	26.5 mW
Average Power for 10 Messages Per Day	500 µW	414 µW	50 µW

Image: RFID journal

"The Internet of Things"
802.15.4 example

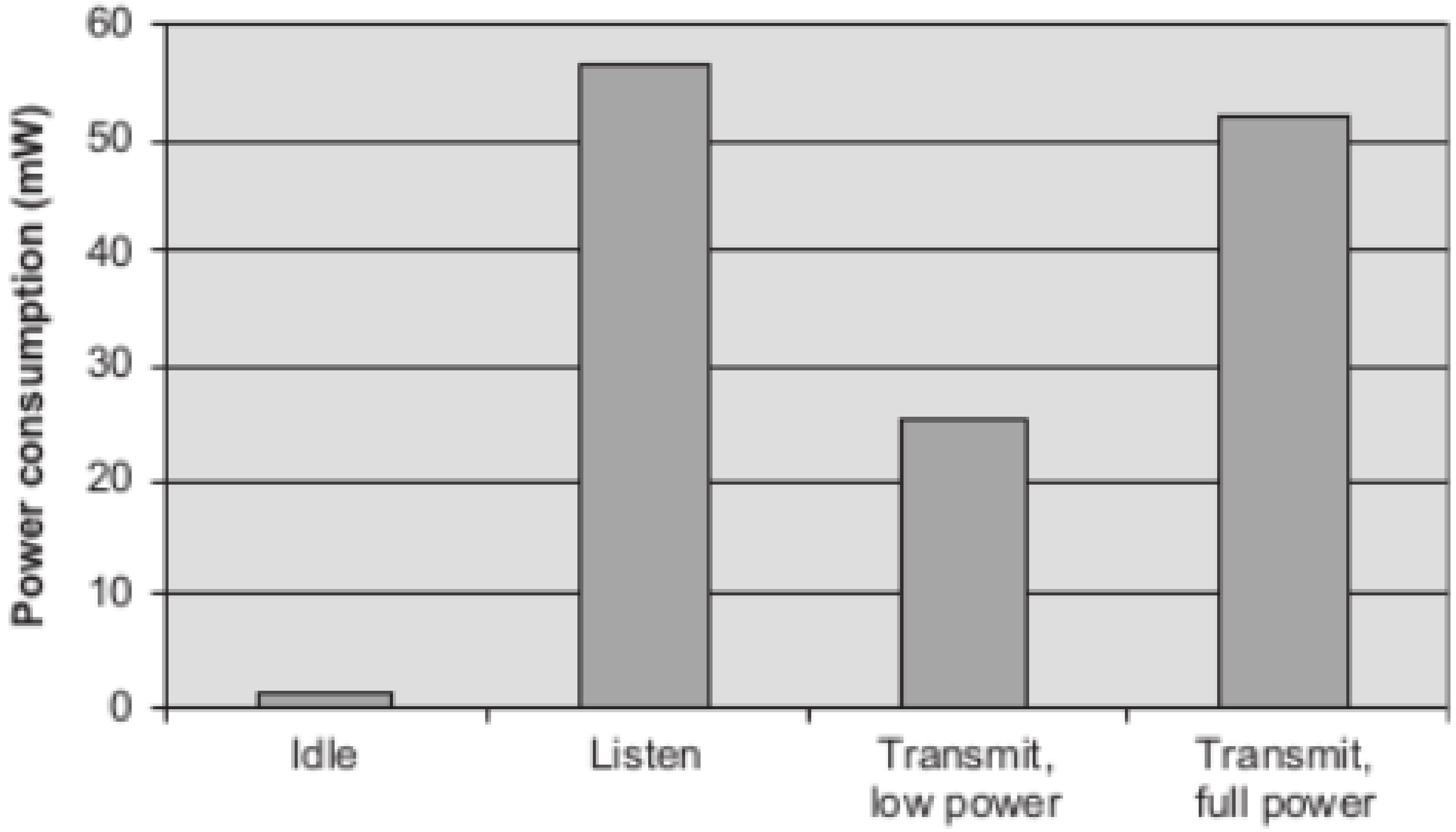


Image: JP Vasseur

"The Internet of Things"

Wi-Fi example

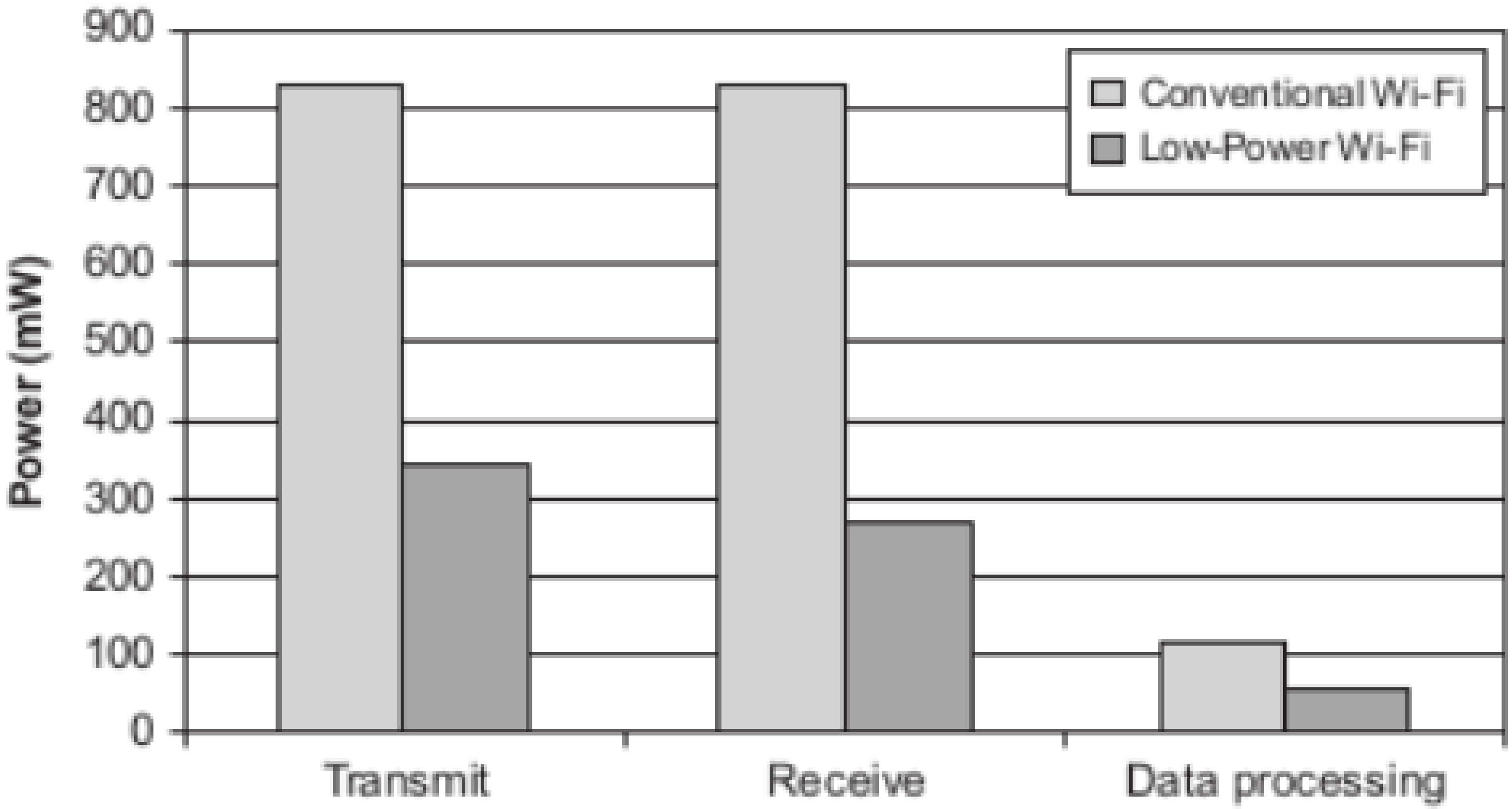


Image: JP Vasseur

“The Internet of Things”

BLE example – standard 600mAh battery

		Broadcasting power		
		-30 dBm [low]	-4 dBm	+4 dBm [high]
Advertising interval	2000 ms [long]	3.3 years	3 years	2.3 years
	1000 ms	1.9 years	1.7 years	1.3 years
	600 ms	1.2 years	1 year	300 days
	200 ms	160 days	140 days	104 days
	50 ms [short]	40 days	35 days	26 days

Image: Estimote

Sensors

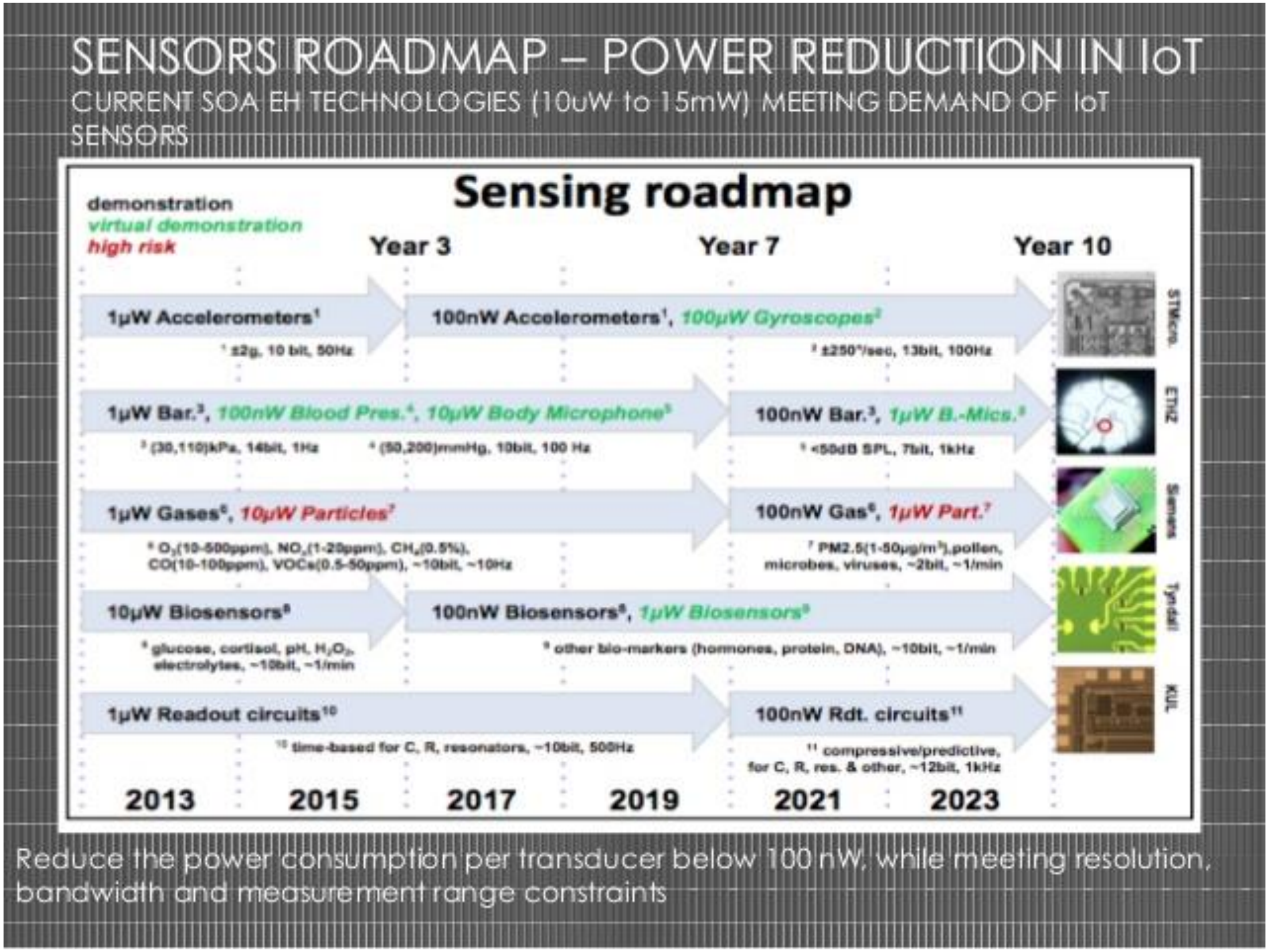


Image: J. Funk

“The Internet of Things” recapped
Energy harvesting

Source	Energy source	Source power	Harvested power
Photovoltaic			
Indoor	Energy Harvester at office environment	0.1mW/cm2	10uW/cm2
Outdoor	Energy Harvester outside in a sunny day at noon	100mW/cm2	10mW/cm2
Vibration	Human walking with harvester in their shoes	0.5+1m/s@1+50Hz	4uW/cm2
Thermal	Human body at ambient air	20mW/cm2	25uW/cm2
RF			
GSM 900MHz	RF harvester at a city restaurant	0.3 to 0.03uW/cm2	0.1uW/cm2
GSM 1800MHz		0.1 to 0.01uW/cm2	

Image: RFID journal

Accelerators
According to Intel

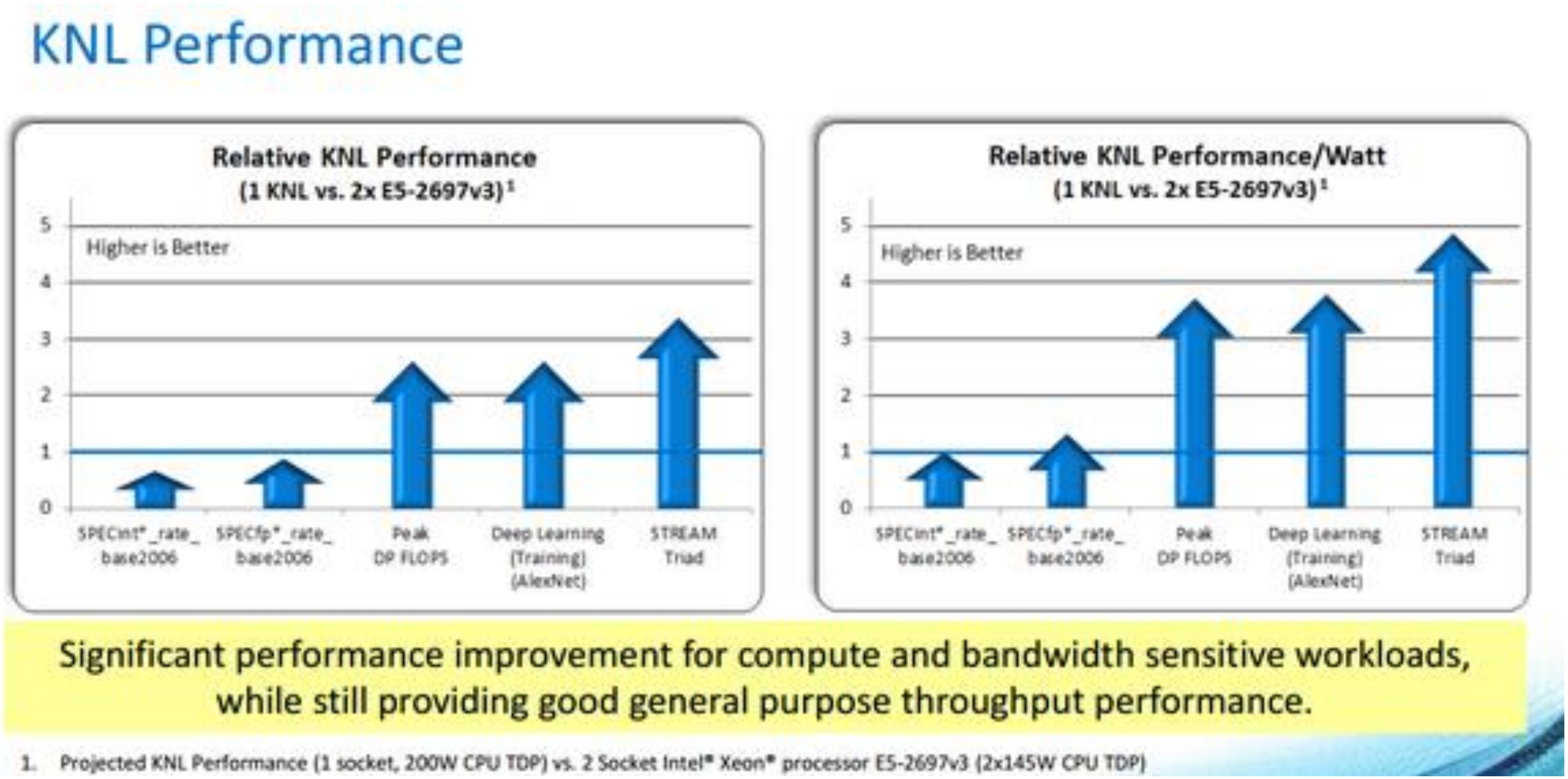


Image: Intel

Accelerators
According to NVIDIA

Tesla Model	K10	K20	K20X	K40	K80	M4	M40
GPU	2 * GK104	GK110	GK110	GK110B	2 * GK210B	GM206	GM200
CUDA Cores	2 * 1,536	2,496	2,688	2,880	4,992	1,024	3,072
Base Core Clock Speed	745 MHz	706 MHz	732 MHz	745 MHz	560 MHz	872 MHz	948 MHz
GPU Boost Clock Speed	–	–	–	875 MHz	875 MHz	1,072 MHz	1,114 MHz
SMXs or SMMs	2 * 8	13	14	15	2 * 13	8	24
Base SP, Teraflops	4.58	3.52	3.95	4.29	5.6	*	*
Peak SP, Teraflops	4.58	3.52	3.95	5.0	8.74	2.2	7.0
Base DP, Teraflops	0.19	1.17	1.31	1.43	1.87	*	*
Peak DP, Teraflops	0.19	1.17	1.31	1.66	2.91	0.06	0.20
GDDR5 Memory	8 GB	5 GB	6 GB	12 GB	24 GB	4 GB	12 GB
Memory Clock Speed	2.5 GHz	2.6 GHz	2.6 GHz	3.0 GHz	2.5 GHz	2.75 GHz	3.0 GHz
Memory Bandwidth	320 GB/sec	208 GB/sec	250 GB/sec	288 GB/sec	480 GB/sec	88 GB/sec	288 GB/sec
Power Draw	225 W	225 W	235 W	235 W	300 W	50 W – 75 W	250 W
SP Efficiency (Gigaflops/Watt)	20.4	15.6	16.8	21.3	29.1	29.3	28.0

* *Base SP and DP teraflops unknown*

Table: NVIDIA

"Mini-accelerators"
NVIDIA M4

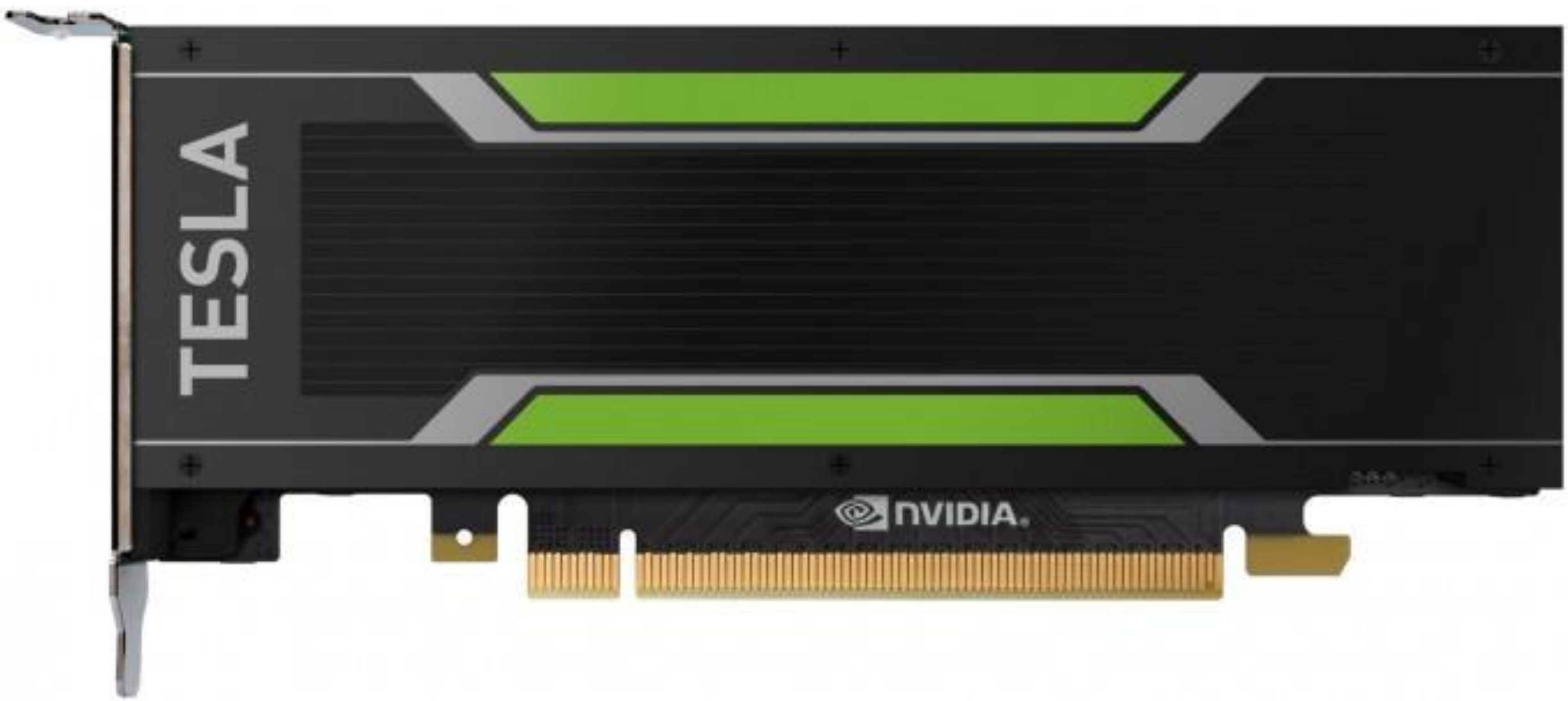


Image: NVIDIA

Heterogeneous… accelerators

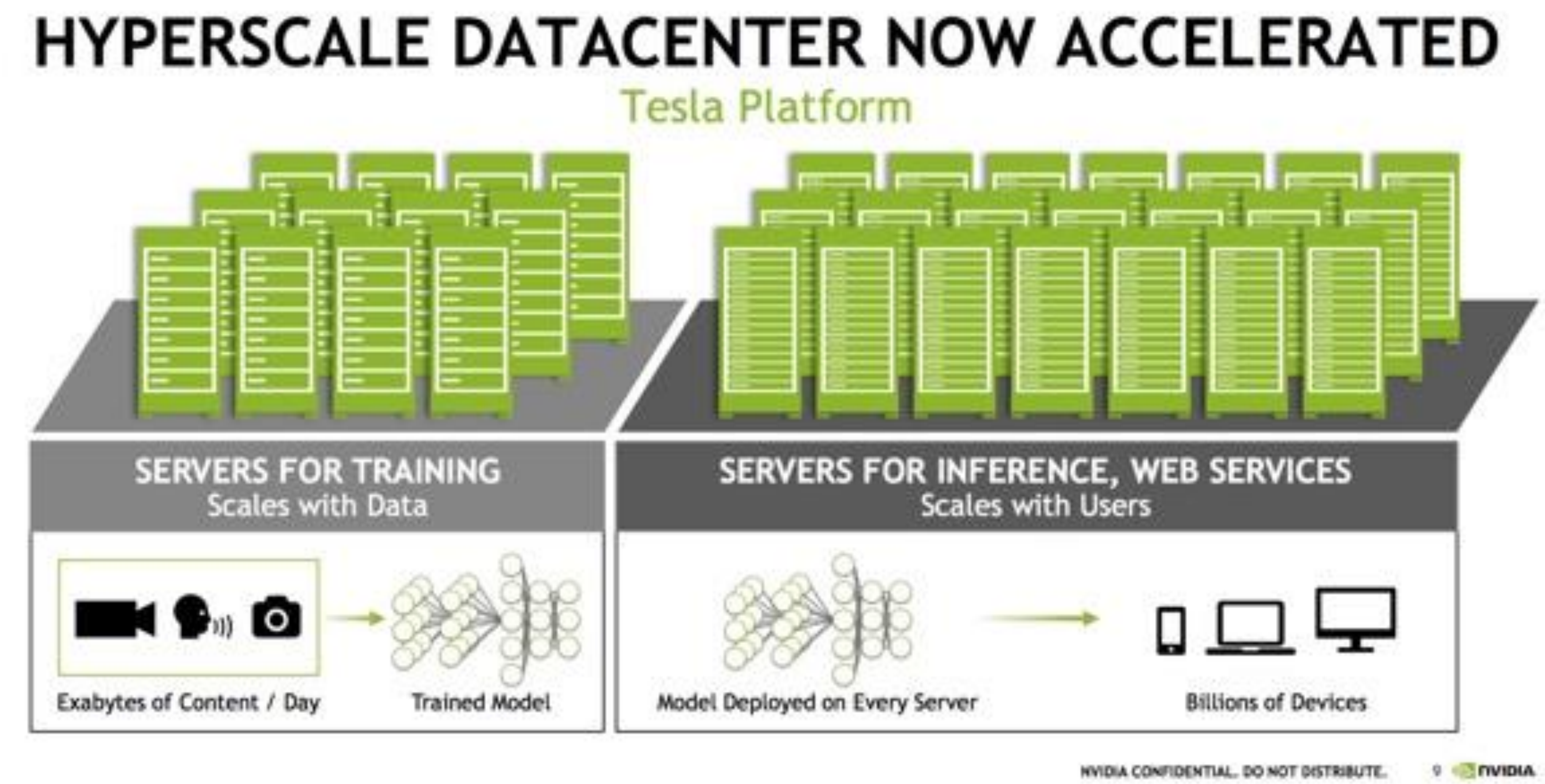


Image: NVIDIA

Combos

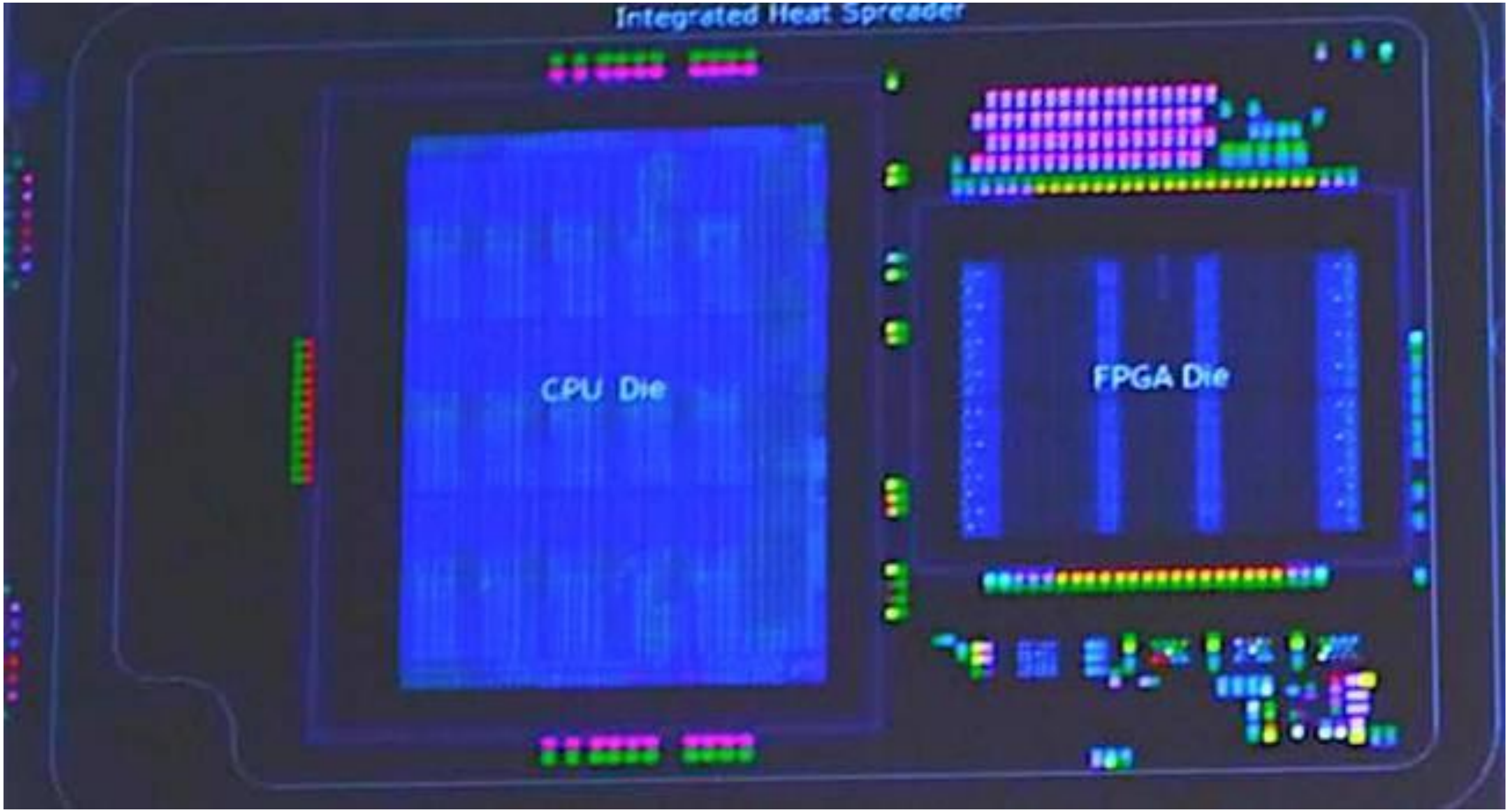
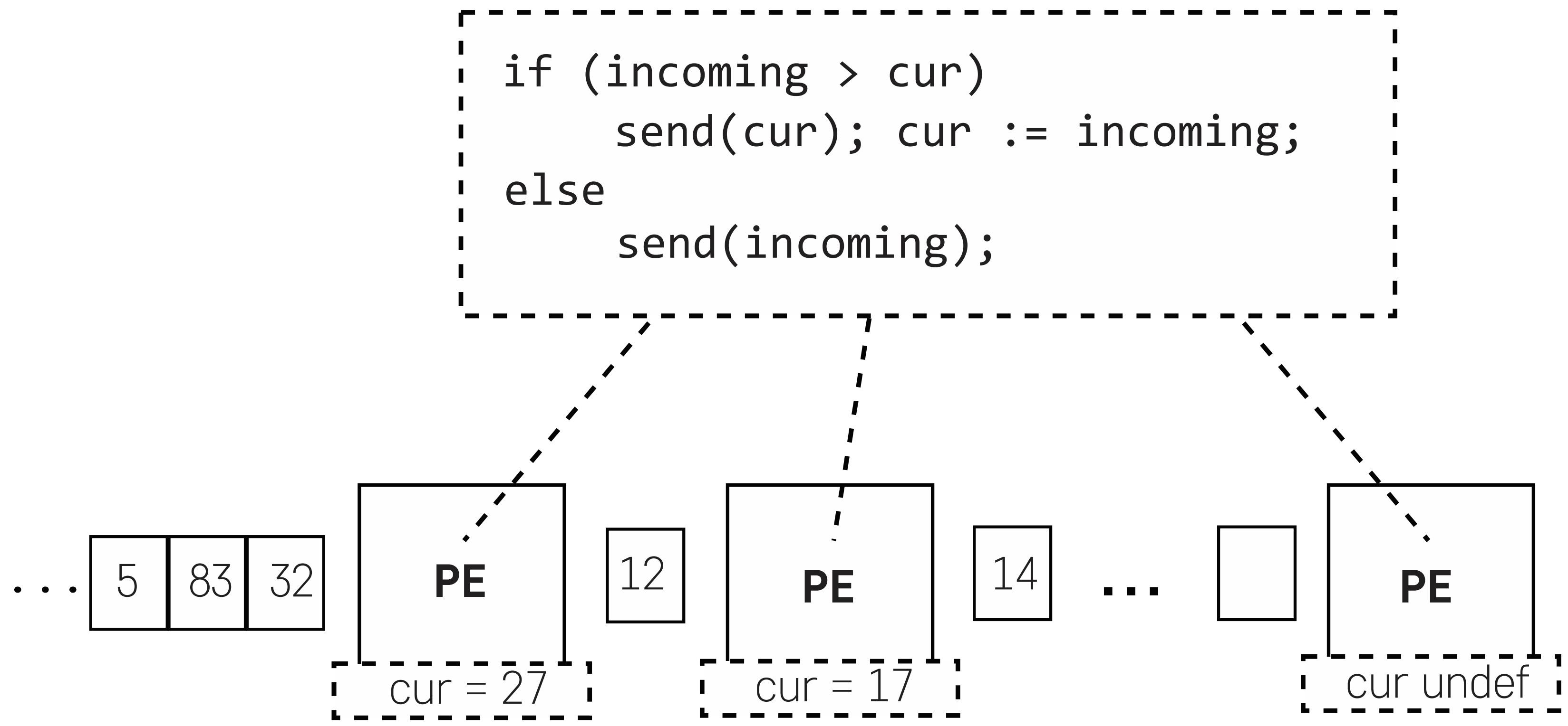


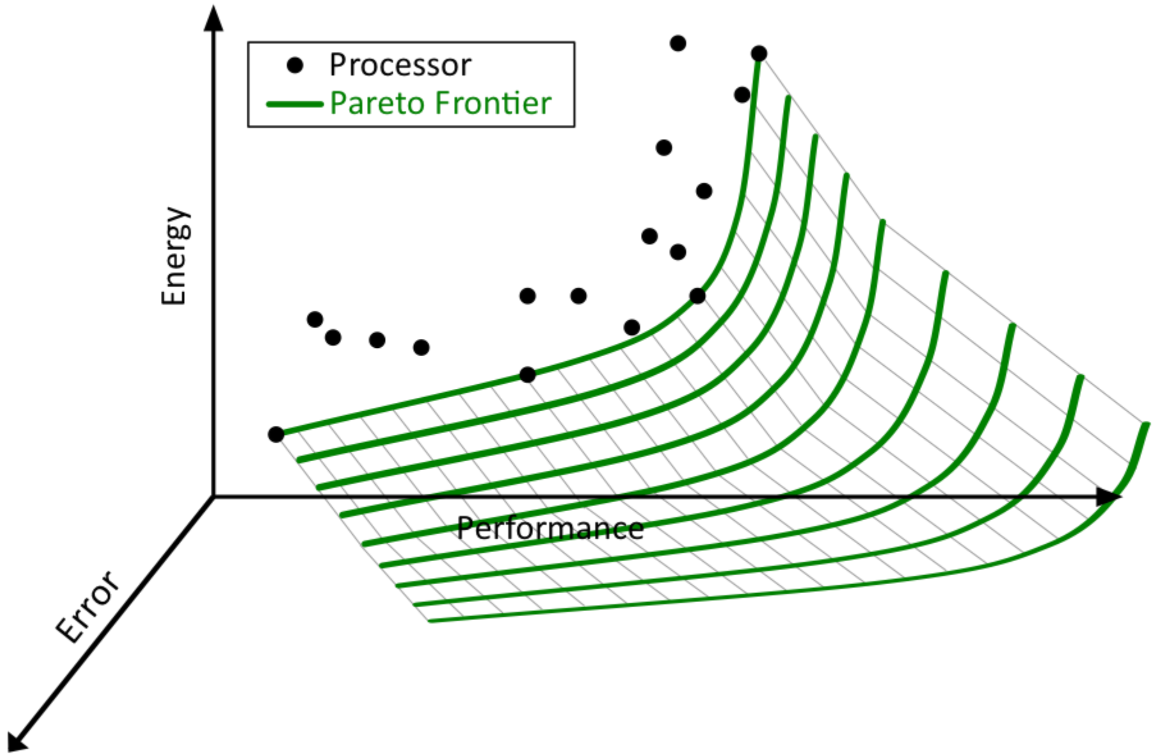
Image: The Next Platform

Spatial architectures
Triggered instructions



Adapted from IEEE Micro/MIT

Approximate computing



Source; IEEE Rebooting Computing

"Imprecise" computing

MIT Technology Review

Log in / Register Search

Subscribe

Topics+ The Daily Magazine Business Reports More+

Computing

Why a Chip That's Bad at Math Can Help Computers Tackle Harder Problems

DARPA funded the development of a new computer chip that's hardwired to make simple mistakes but can help computers understand the world.

by Tom Simonite April 14, 2016

Your math teacher lied to you. Sometimes getting your sums wrong is a good thing.

So says Joseph Bates, cofounder and CEO of Singular Computing, a company whose computer chips are hardwired to be incapable of performing mathematical calculations correctly. Ask it to add 1 and 1 and you will get answers like 2.01 or 1.98.

The Pentagon research agency DARPA funded the creation of Singular's chip because that fuzziness can be an asset when it comes to some of the hardest problems for computers, such as making sense of video or other messy real-world data. "Just because the hardware is sucky doesn't mean the software's result has to be," says Bates.

A chip that can't guarantee that every calculation is perfect can still get good results on many problems but needs fewer circuits and burns less energy, he says.

Bates has worked with Sandia National Lab, Carnegie Mellon University, the Office of Naval Research, and MIT on tests that used simulations to show how the S1 chip's inexact operations might make certain tricky computing tasks more efficient. Problems with data that comes with built-in noise from the real world, or where some

Approximate computing ctd.

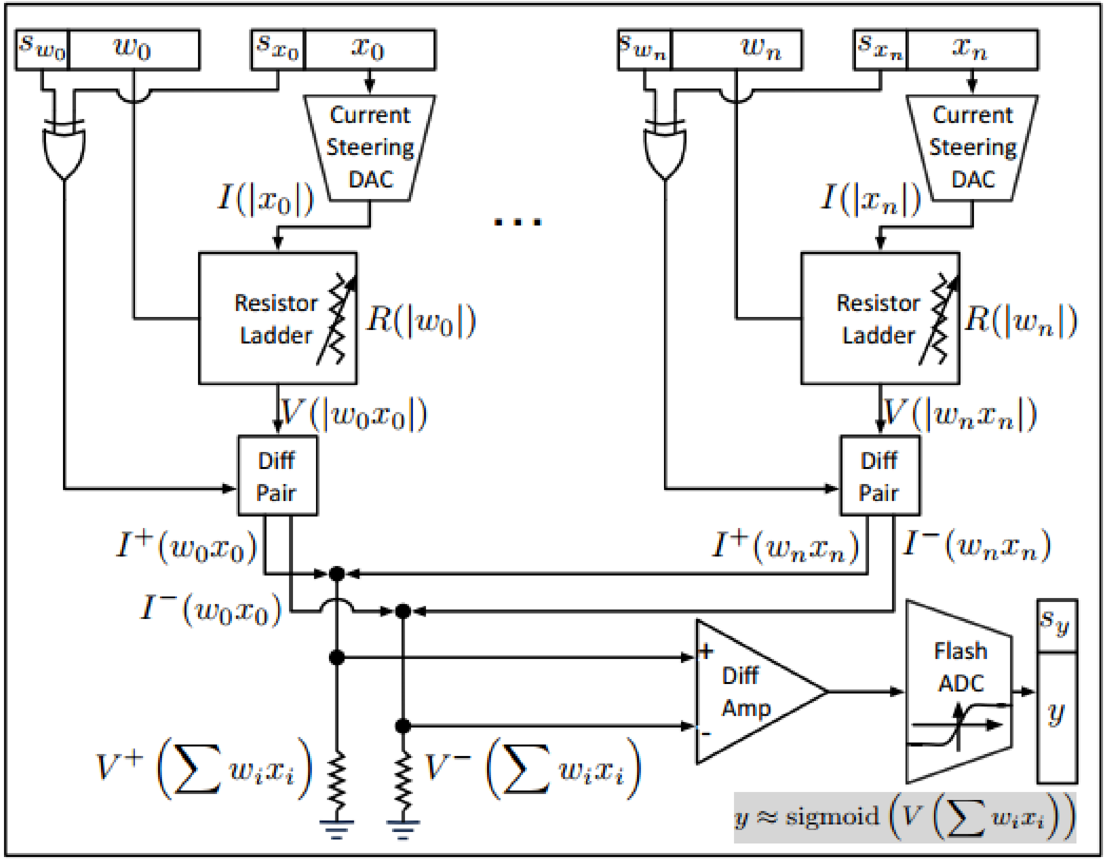
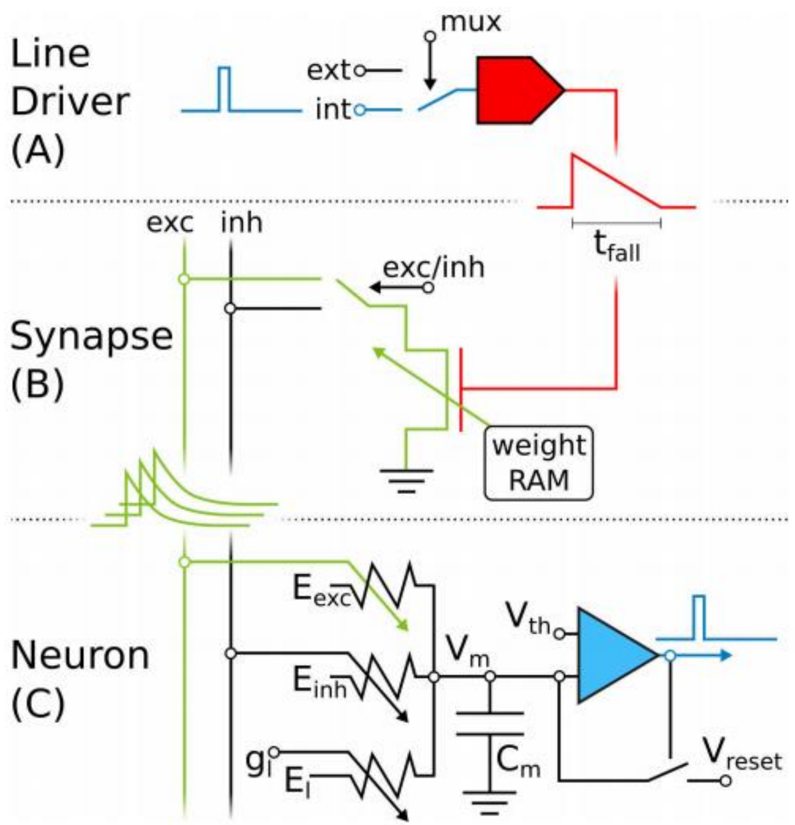


Figure 3: A single analog neuron (ANU).

Source: St. Amant et al, ISCA 2014

Neuromorphic computing (1)

- Pattern detection, probabilistic inference
- Massive parallelism
- Storage and computation coupled and distributed
- Built on simple blocks (neurons)
- Analog operation – spiking networks



Neuromorphic computing (2)
Qualcomm

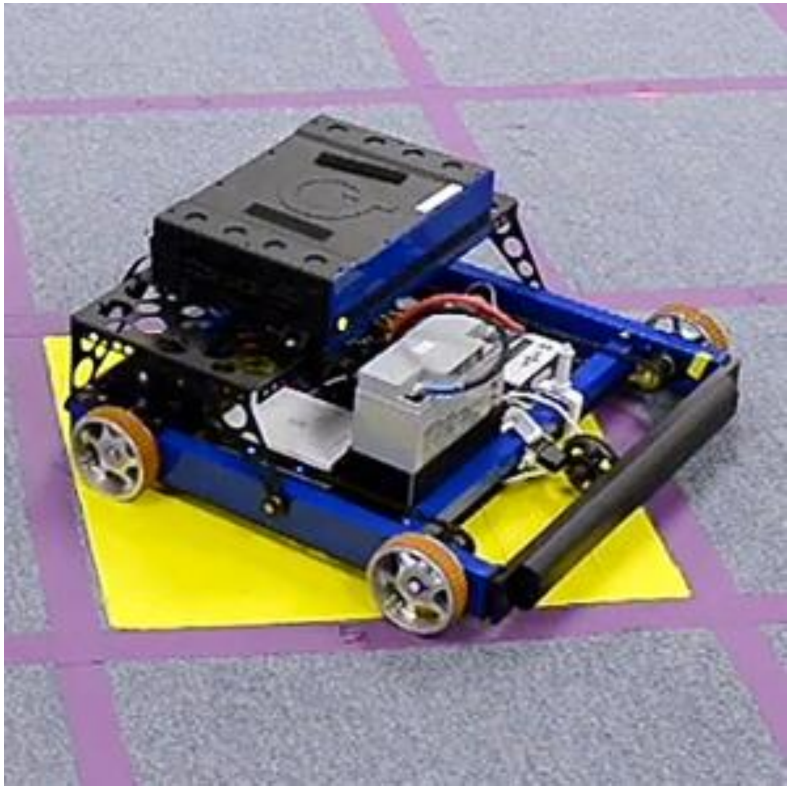
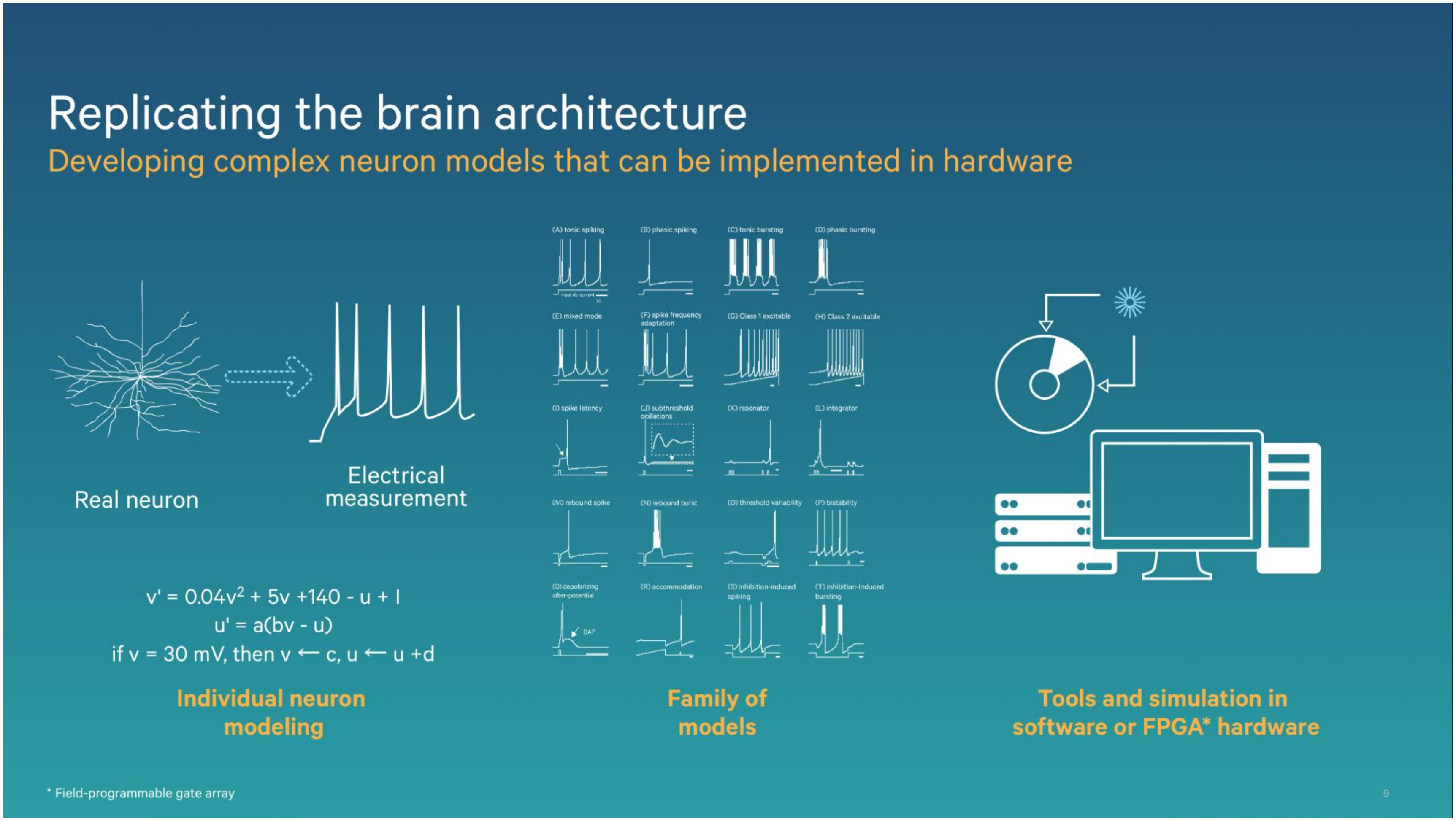
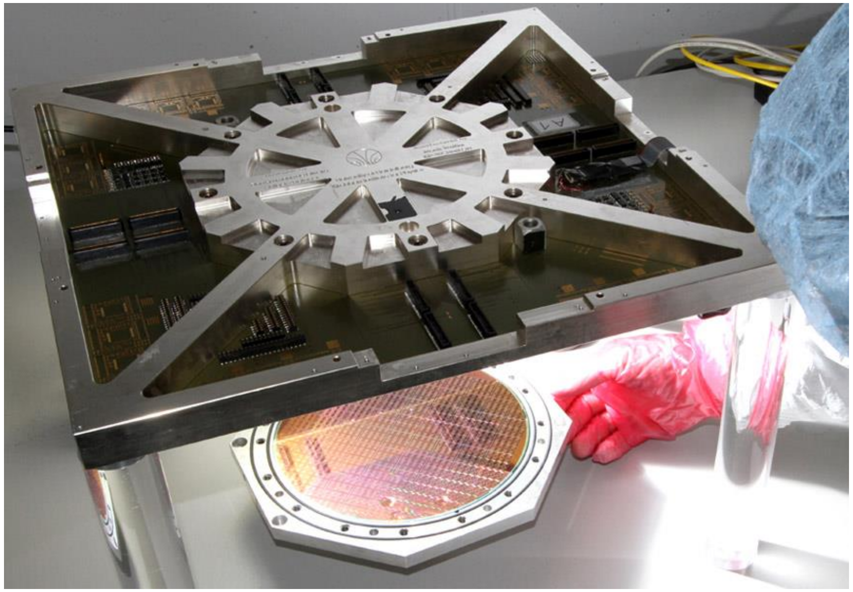
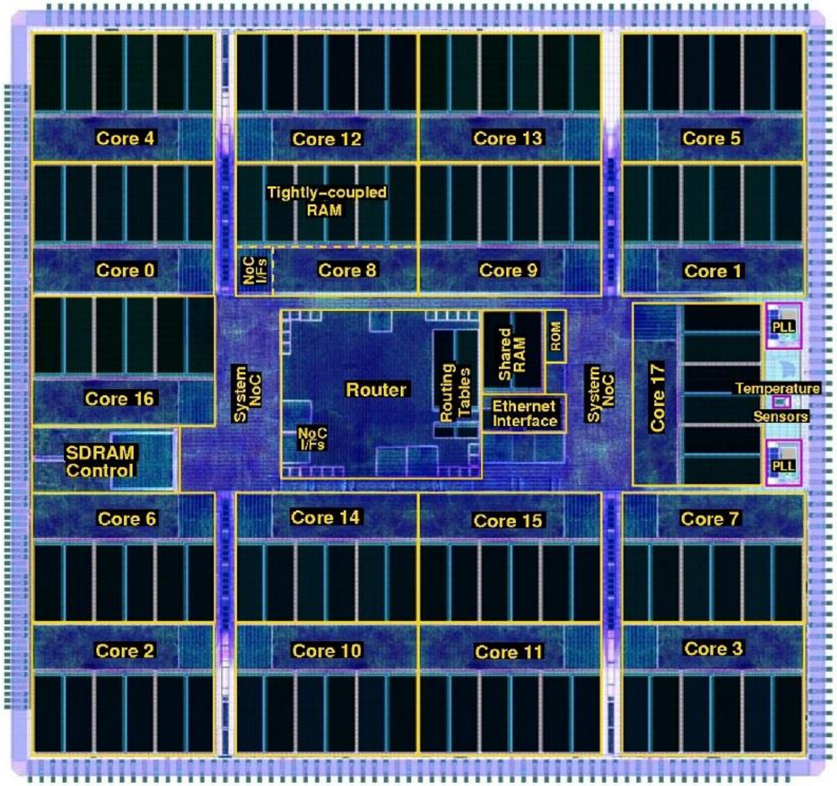


Image: Qualcomm

Neuromorphic computing (3)
Qualcomm






Core 4
Core 12
Core 13
Core 5
Tightly-coupled RAM
Core 0
NoC I/Fs
Core 8
Core 9
Core 1
PLL
Shared RAM
ROM
System NoC
Router
Routing Tables
System NoC
Core 17
Temperature
Sensors
Core 16
Ethernet Interface
NoC I/Fs
SDRAM Control
PLL
Core 6
Core 14
Core 15
Core 7
Core 2
Core 10
Core 11
Core 3

Neuromorphic computing (5)

Processing Powers

	What they do well	What they're good for
Neuromorphic chips	Detect and predict patterns in complex data, using relatively little electricity	Applications that are rich in visual or auditory data and that require a machine to adjust its behavior as it interacts with the world
Traditional chips (von Neumann architecture)	Reliably make precise calculations	Anything that can be reduced to a numerical problem, although more complex problems require substantial amounts of power

MIT Technology Review

Energy efficiency – bottom line

Infrastructure and casing

- Minimum power overheads

Hardware

- Optimized for performance/Watt

Operating system

- Energy aware, actively optimizing

Software

- Energy aware

Thank you

e-mail: an@tik.services

http://tik.services