Energy Efficient Computing
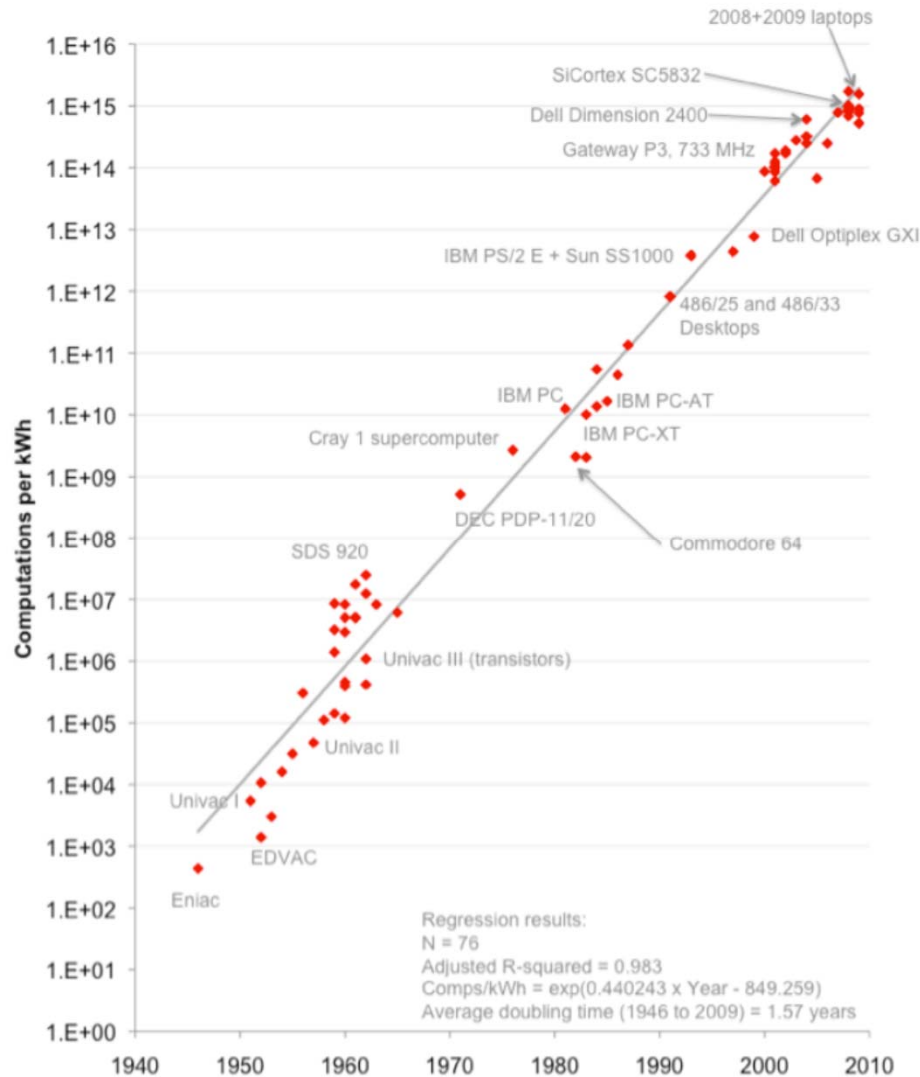
Lennart Johnsson
Advanced Computing Research Lab

UNIVERSITY *of* **HOUSTON**

What we do?

- Work with various vendors to enhance/advance architectures and platforms for HPC
- Validation of designs through energy efficiency and performance benchmarks
- Develop algorithms and software tools

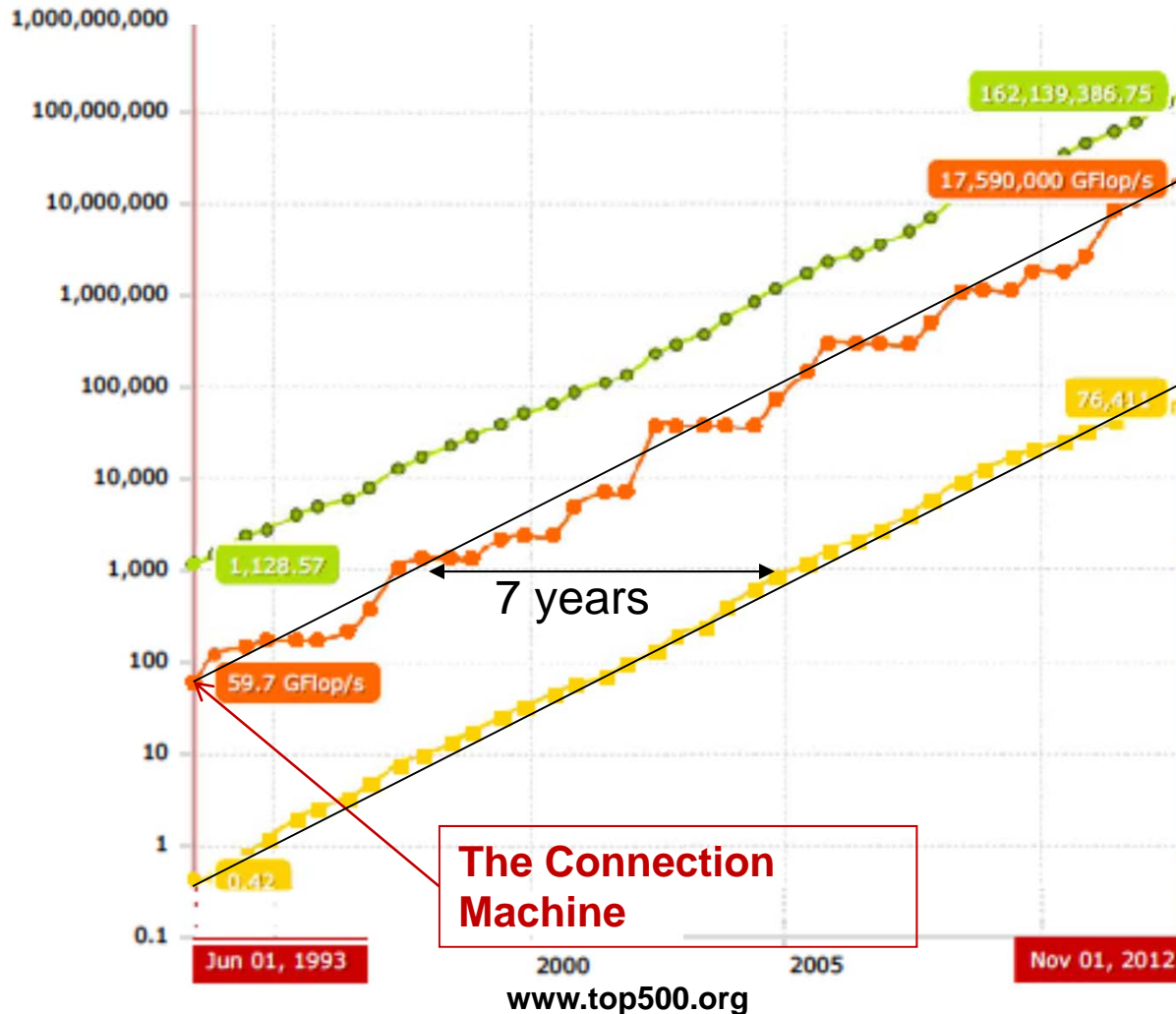# Why is energy efficiency important?

Energy efficiency evolution

Energy efficiency doubling every 18.84 months on average measured as computation/kWh

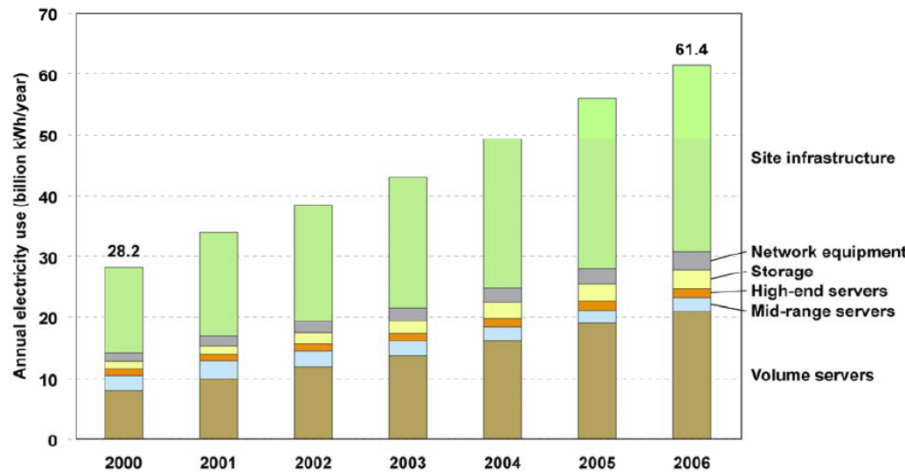# Top500 system performance evolution

Performance doubling period on average:

No 1   – 13.64 months

No 500 – 12.90 months

The Gap

The energy efficiency improvement as determined by Koomey does not match the performance growth of HPC systems as measured by the Top500 list

The Gap indicates a growth rate in energy consumption for HPC systems of about 20%/yr.

EPA study projections: 14% - 17%/yr

Uptime Institute projections: 20%/yr

PDC experience: 20%/yr

Report to Congress on Server and Data Center Energy Efficiency", Public Law 109-431, U.S Environmental Protection Agency, Energy Star Program, August 2, 2007, http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf

"Findings on Data Center Energy Consumption Growth May Already Exceed EPA's Prediction Through 2010!",
K. G. Brill, The Uptime Institute, 2008,
http://uptimeinstitute.org/content/view/155/147

End use component	2000 Electricity use (billion kWh)	2000 % Total	2006 Electricity use (billion kWh)	2006 % Total	2000 – 2006 electricity use CAGR
Site infrastructure	14.1	50%	30.7	50%	14%
Network equipment	1.4	5%	3.0	5%	14%
Storage	1.1	4%	3.2	5%	20%
High-end servers	1.1	4%	1.5	2%	5%
Mid-range servers	2.5	9%	2.2	4%	-2%
Volume servers	8.0	29%	20.9	34%	17%
Total	28.2		61.4		14%

UNIVERSITY *of* HOUSTON

Evolution of Data Center Energy Costs (US)

The Cost to Power & Cool a Server Has Exceeded the Cost of the Server…
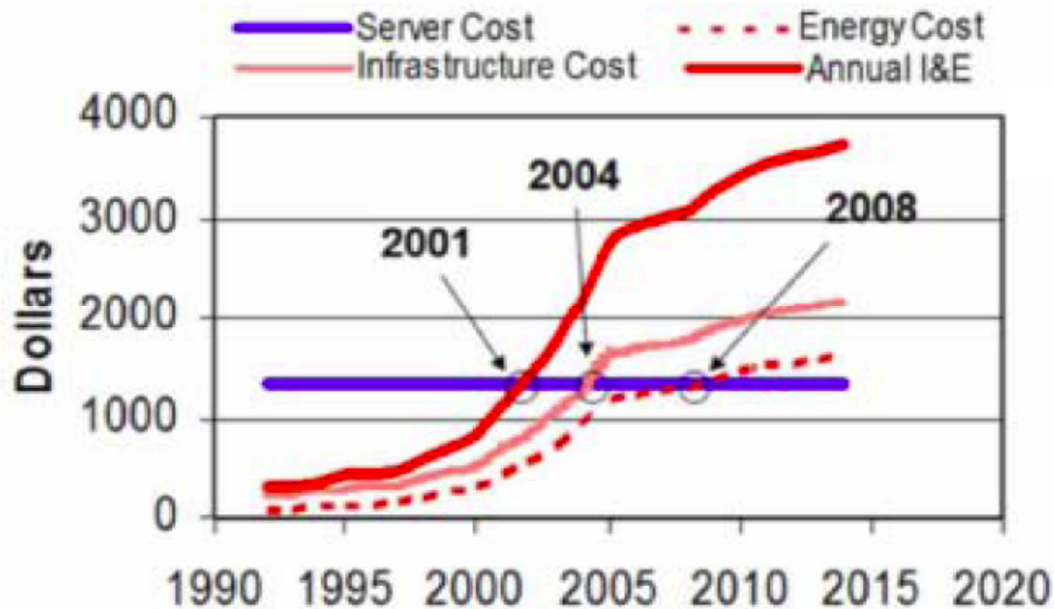
• Annual Amortized Costs in the Data Center for a 1U Server

• (3 yr server life, 10 yr infrastructure life)

• HP has invested heavily to reduce customers' I&E costs

Source: Belady, C., 2007, "In the Data Center, Power and Cooling Costs More than IT Equipment it Supports", Electronics Cooling Magazine (Feb issue).

Source: Tahir Cader, Energy Efficiency in HPC – An Industry Perspective, High Speed Computing, April 27 – 30, 2009

UNIVERSITY *of* HOUSTON

Exa-scale Data Centre Challenges

DOE E3 Report: Extrapolation of existing design trends to Exascale in 2016 Estimate: **130 MW**

DARPA Study: More detailed assessment of component technologies
Estimate: **20 MW** just for memory alone, **60 MW** aggregate extrapolated from current design trends

Nuclear power plant: 1–1,5 GW

The current approach is not sustainable!
More holistic approach is needed!
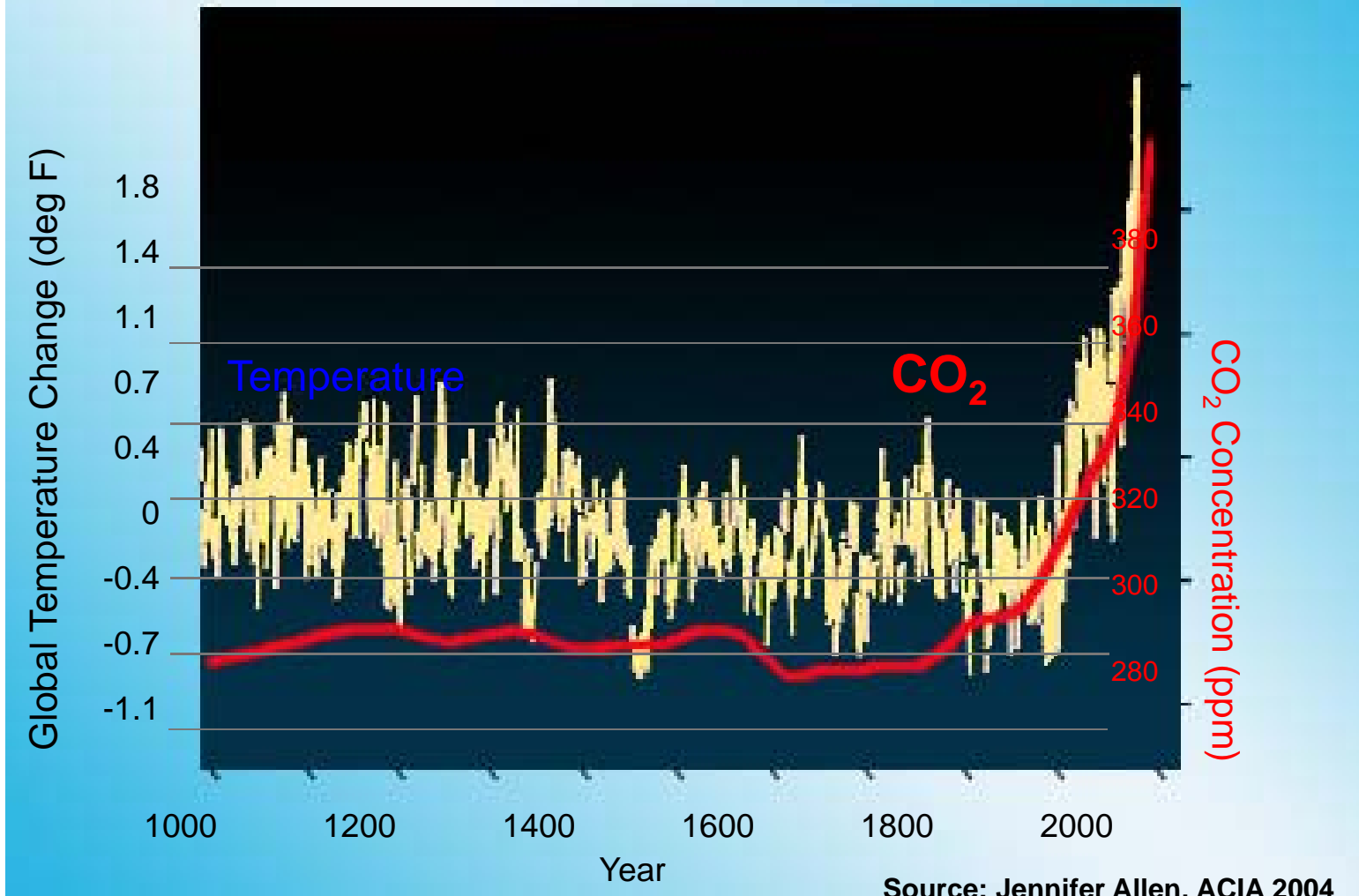
Rule of thumb: 1 MW = $1M/yr in electricity cost

A large data center (Google, Microsoft, Facebook, ….) consumes 100+ MW!

UNIVERSITY *of* **HOUSTON**

1000 Years of CO_2 and Global Temperature Change

Global Temperature Change (deg F)

1.8
1.4
1.1
0.7
0.4
0
-0.4
-0.7
-1.1

Temperature

CO_2

CO_2 Concentration (ppm)

380
360
340
320
300
280

1000   1200   1400   1600   1800   2000

Year

Source: Jennifer Allen, ACIA 2004

Source: http://alaskaconservationsolutions.com/acs/images/stories/docs/AkCS_current.ppt
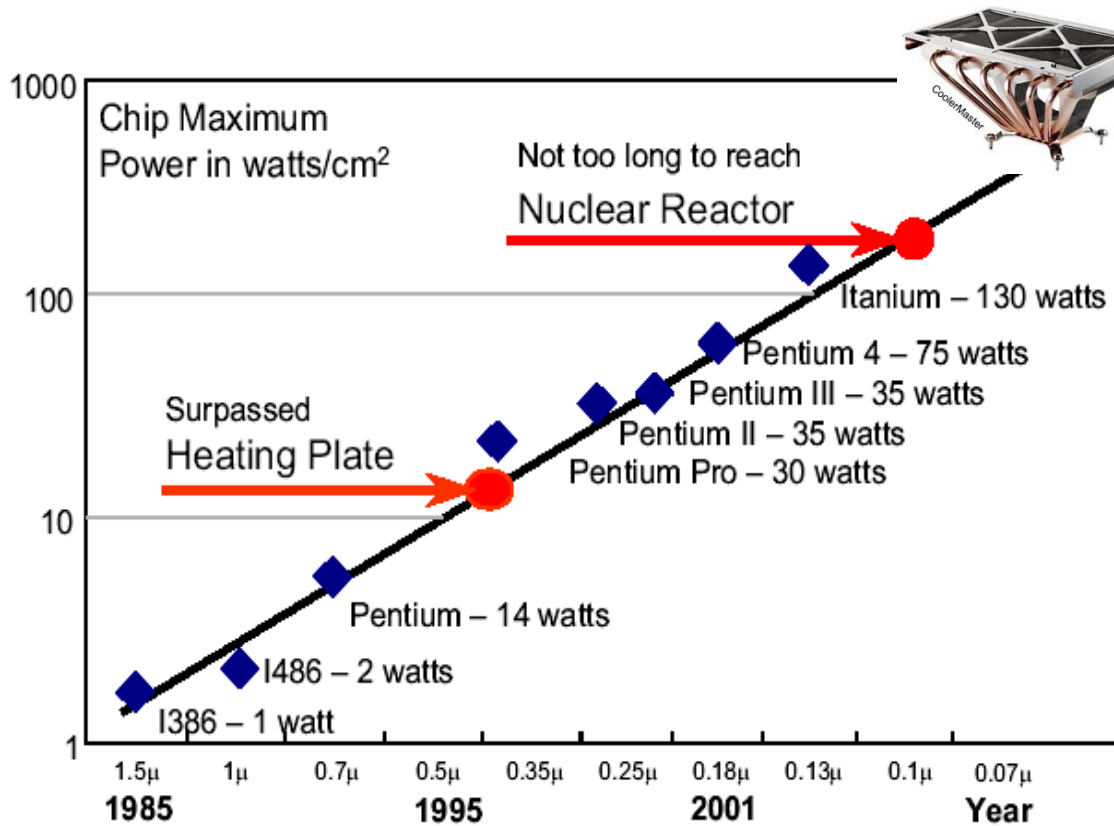
UNIVERSITY of HOUSTON

An inefficient truth
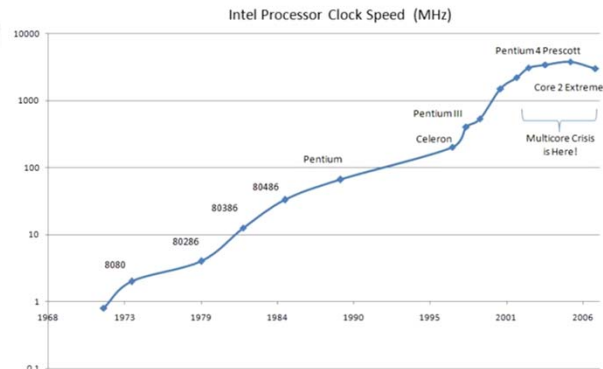# ICT impact on CO_2 emissions*

- It is estimated that the ICT industry alone produces CO_2 emissions that is equivalent to the carbon output of the entire aviation industry. Direct emissions of Internet and ICT amounts to 2-3% of world emissions and is expected to grow to 6+% by the end of the decade
- ICT emissions growth fastest of any sector in society; expected to double every 4 to 6 years with current approaches
- One small computer server generates as much carbon dioxide as a SUV with a fuel efficiency of 15 miles per gallon

UNIVERSITY *of* **HOUSTON**

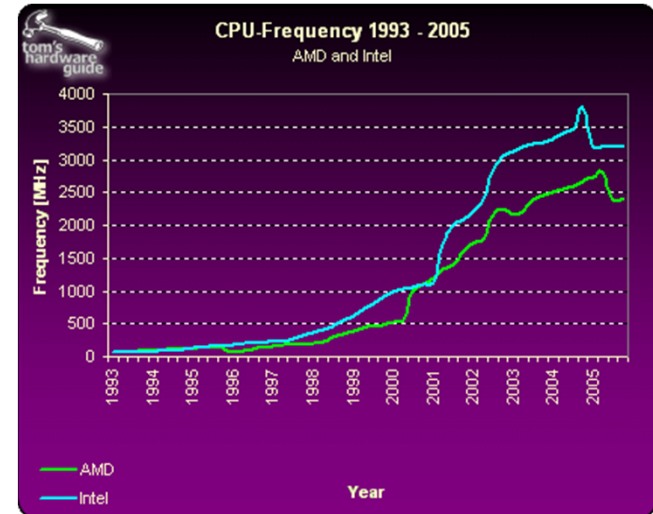# Despite remarkable transistor energy efficiency improvement CPUs got hotter

Heat density of Intel CPUs, Source Shekhar Borkar, Intel

Intel Processor Clock Speed (MHz)), from
http://smoothspan.files.wordpress.com/2007/09/clockspeeds.jpg

http://www.tomshardware.com/reviews/mother-cpu-charts-2005,1175.html

How to improve energy efficiency?

How to improve energy efficiency?

- Reduce energy consumption
- Energy recovery

What type of Architecture?

Reducing Waste

Mark Horowitz 2007: "Years of research in low-power embedded computing have shown only one design technique to reduce power: **reduce waste**."

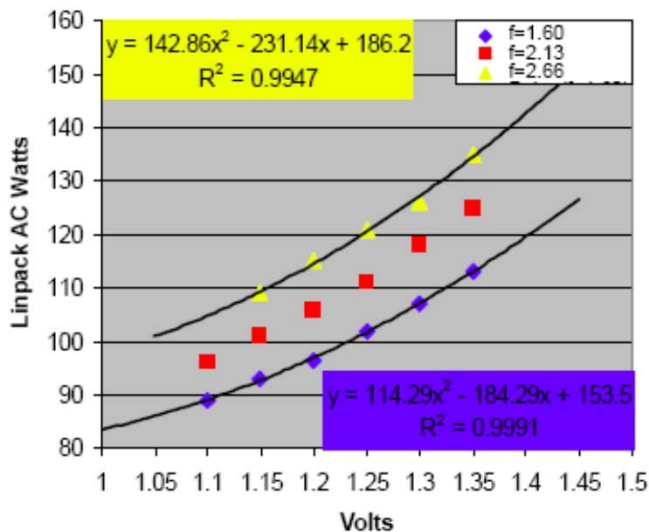Seymour Cray 1977: "Don't put anything in to a supercomputer that isn't necessary."

23

Exascale Computing Technology Challenges, John Shalf
National Energy Research Supercomputing Center, Lawrence Berkeley National Laboratory
ScicomP / SP-XXL 16, San Francisco, May 12, 2010

UNIVERSITY *of* HOUSTON

What type of Architecture?

UNIVERSITY *of* **HOUSTON**

Energy Consumption

"We are on the Wrong side of a Square Law" Fred Pollack 1999

New goal for CPU design: "Double *Valued Performance* every 18 months, at the same power level", Fred Pollack

Pollack, F (1999). *New Microarchitecture Challenges in the Coming Generations of CMOS Process Technologies.* Paper presented at the Proceedings of the 32nd Annual IEEE/ACM International Symposium on Microarchitecture, Haifa, Israel.

$y = 142.86x^2 - 231.14x + 186.2$
$R^2 = 0.9947$

$y = 114.29x^2 - 184.29x + 153.5$
$R^2 = 0.9991$

f=1.60
f=2.13
f=2.66

Linpack: $15f(V-0.2)^2 + 45V + 19$
STREAM: $5f(V-0.2)^2 + 50V + 19$

Product	Normalized Performance	Normalized Power	EPI on 65 nm at 1.33 volts (nJ)
i486	1.0	1.0	10
Pentium	2.0	2.7	14
Pentium Pro	3.6	9	24
Pentium 4 (Willamette)	6.0	23	38
Pentium 4 (Cedarmill)	7.9	38	48
Pentium M (Dothan)	5.4	7	15
Core Duo (Yonah)	7.7	8	11

Ed Grochowski, Murali Annavaram Energy per Instruction Trends in Intel® Microprocessors.
http://support.intel.co.jp/pressroom/kits/core2duo/pdf/epi-trends-final2.pdf

Energy Cost of Operations

Operation	Energy (pJ)
64b Floating FMA (2 ops)	100
64b Integer Add	1
Write 64b DFF	0.5
Read 64b Register (64 x 32 bank)	3.5
Read 64b RAM (64 x 2K)	25
Read tags (24 x 2K)	8
Move 64b 1mm	6
Move 64b 20mm	120
Move 64b off chip	256
Read 64b from DRAM	2000

What kind of architecture (core)

How Small is Small
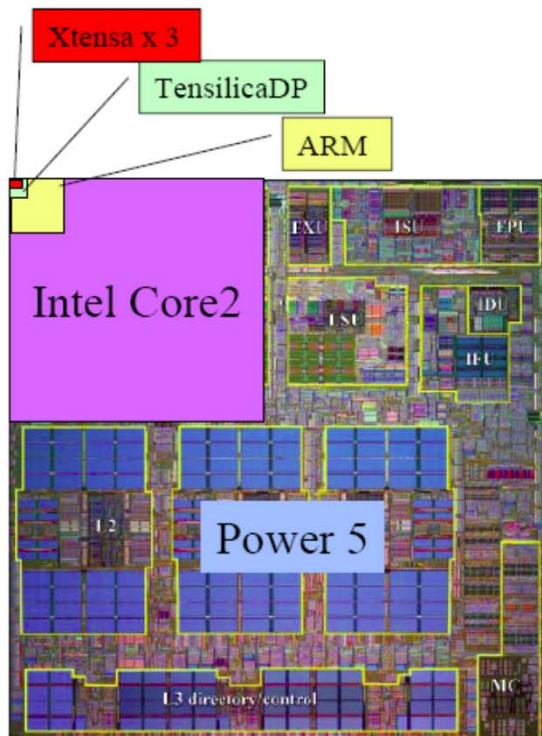
- Power5 (server)
  - 389mm^2
  - 120W@1900MHz
- Intel Core2 sc (laptop)
  - 130mm^2
  - 15W@1000MHz
- ARM Cortex A8 (toaster oven)
  - 5mm^2
  - 0.8W@800MHz
- Tensilica DP (cell phones)
  - 0.8mm^2
  - 0.09W@600MHz
- Tensilica Xtensa (Cisco Rtr)
  - 0.32mm^2 for 3!
  - 0.05W@600MHz

Each core operates at 1/3 to 1/10th efficiency of largest chip, but you can pack 100x more cores onto a chip and consume 1/20 the power
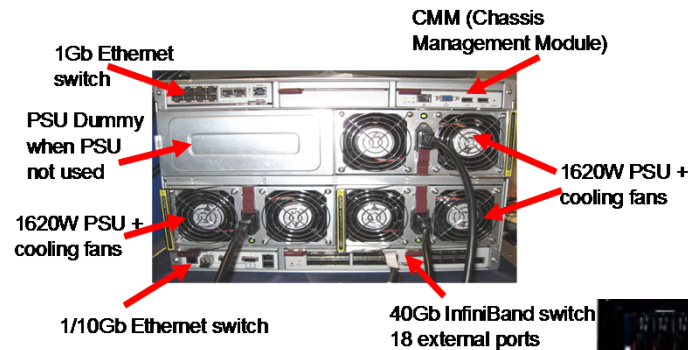
- Cubic power improvement with lower clock rate due to V^2F

- Slower clock rates enable use of simpler cores

- Simpler cores use less area (lower leakage) and reduce cost

- Tailor design to application to reduce waste

http://www.csm.ornl.gov/workshops/SOS11/presentations/j_shalf.pdf

UNIVERSITY *of* HOUSTON

SNIC/KTH PRACE Prototype I

CMM (Chassis Management Module)
1Gb Ethernet switch
PSU Dummy when PSU not used
1620W PSU + cooling fans
1620W PSU + cooling fans
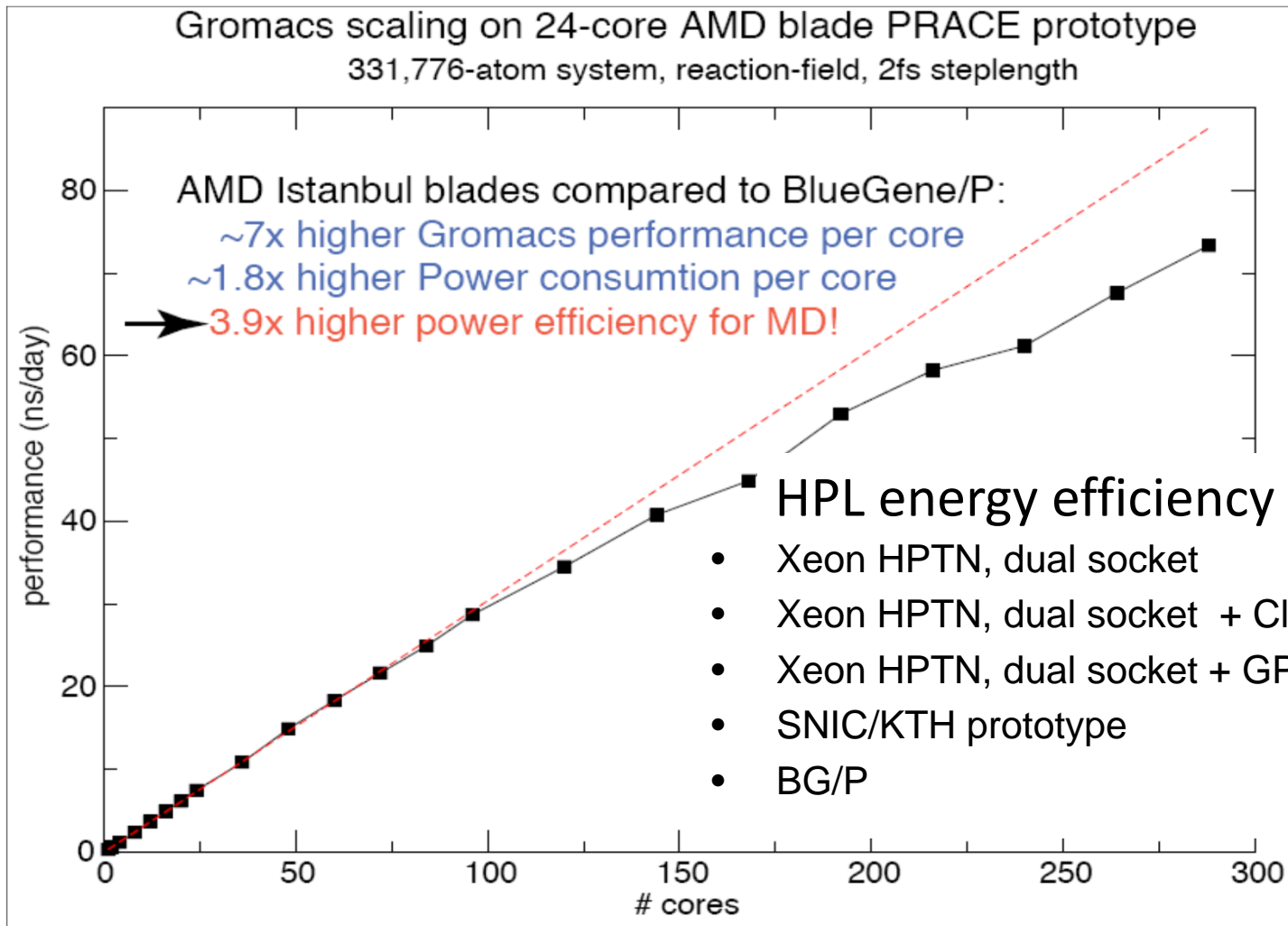1/10Gb Ethernet switch
40Gb InfiniBand switch 18 external ports

- New 4-socket blade with 4 DIMMs per socket supporting PCI-Express Gen 2 x16
- Four 6-core 2.1 GHz 55W ADP AMD Istanbul CPUs, 32GB/node
- 10-blade in a 7U chassis with 36-port QDR IB switch, new efficient power supplies.
- 2TF/chassis, 12 TF/rack, 30 kW (6 x 4.8)
- 180 nodes, 4320 cores, full bisection QDR IB interconnect

Network:
- QDR Infiniband
- 2-level Fat-Tree
- Leaf level 36-port switches built into chassis
- Five external 36-port switches

SNIC/KTH/PRACE Prototype I

Gromacs scaling on 24-core AMD blade PRACE prototype
331,776-atom system, reaction-field, 2fs steplength

AMD Istanbul blades compared to BlueGene/P:
~7x higher Gromacs performance per core
~1.8x higher Power consumtion per core
→ 3.9x higher power efficiency for MD!

HPL energy efficiency

- Xeon HPTN, dual socket — 240 MF/W
- Xeon HPTN, dual socket + Clearspeed — 326 MF/W
- Xeon HPTN, dual socket + GPU — 270 MF/W
- SNIC/KTH prototype — 344 MF/W
- BG/P — 357 MF/W

Nominal Energy Efficiency of Mobile CPUs, x86 CPUs and GPUs

ARM Cortex-9			ATOM			AMD 12-core			Intel 6-core			ATI 9370		
Cores	W	GF/W	Cores	W	GF/W	Cores	W	GF/W	Cores	W	GF/W	Cores	W	GF/W
4	~2	~0.5	2	2+	~0.5	12	115	~0.9	6	130	~0.6	1600	225	~2.3

nVidia Fermi			TMS320C6678			IBM BQC			ClearSpeed CX700		
Cores	W	GF/W	Cores	W	GF/W	Cores	W	GF/W	Cores	W	GF/W
512	225	~2.2	8	4	~ 15	16	55	3.7	192	10	~10

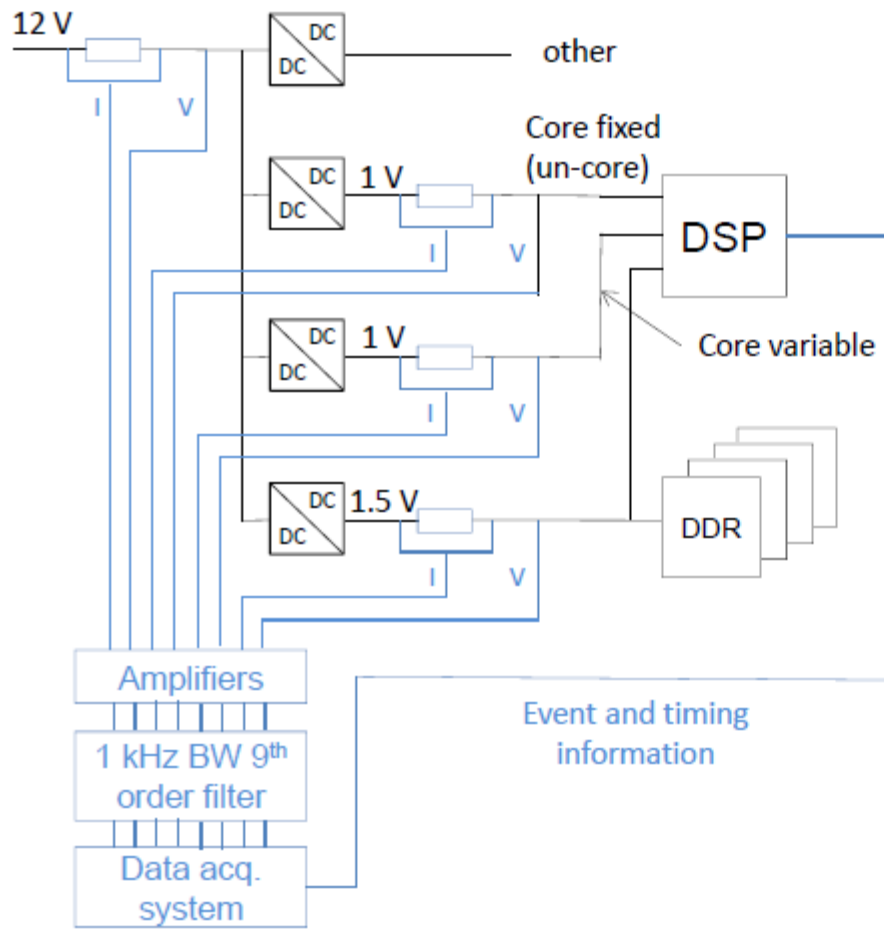Very approximate estimates!!

KTH/SNIC/PRACE Prototype II

KTH/SNIC/PRACE DSP HPC node
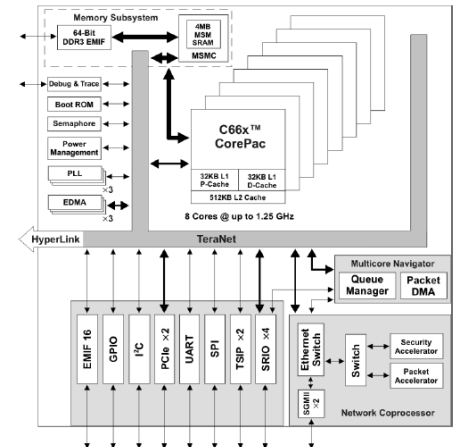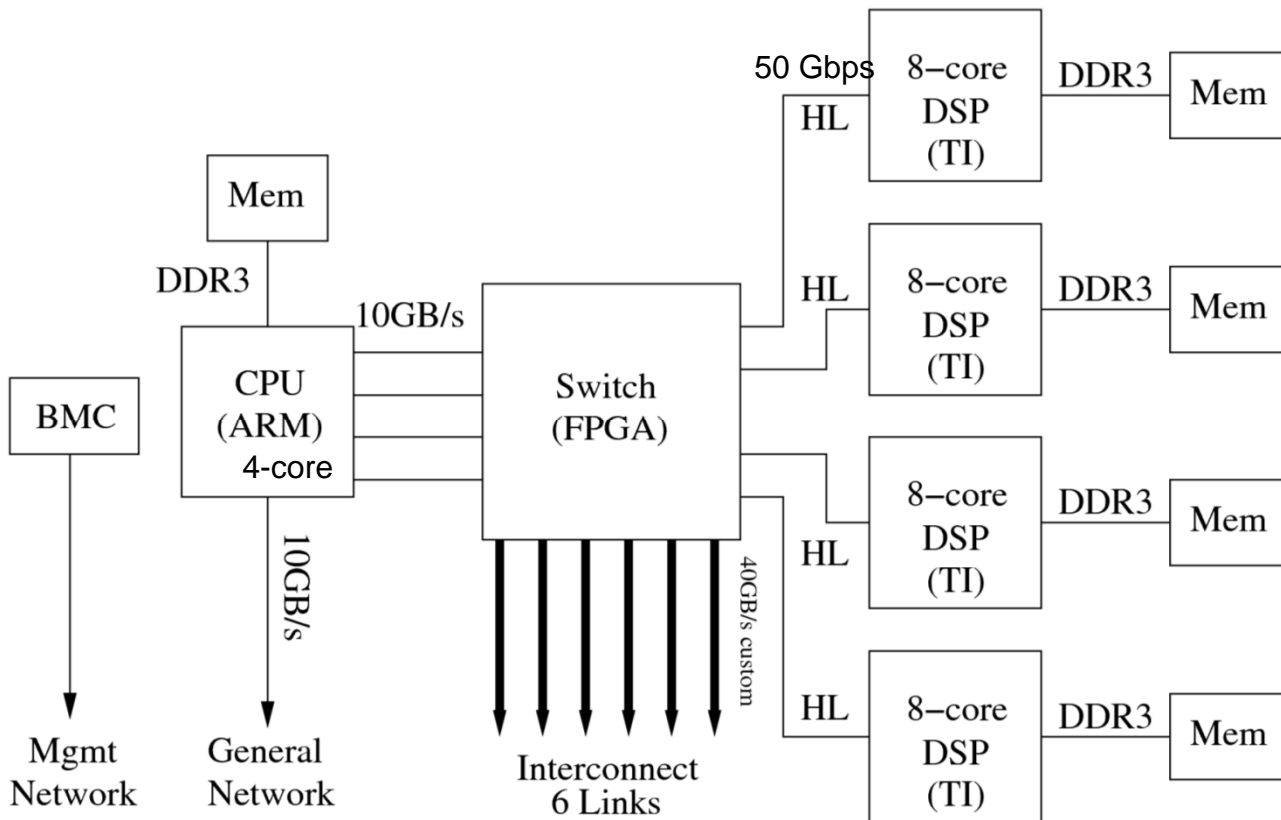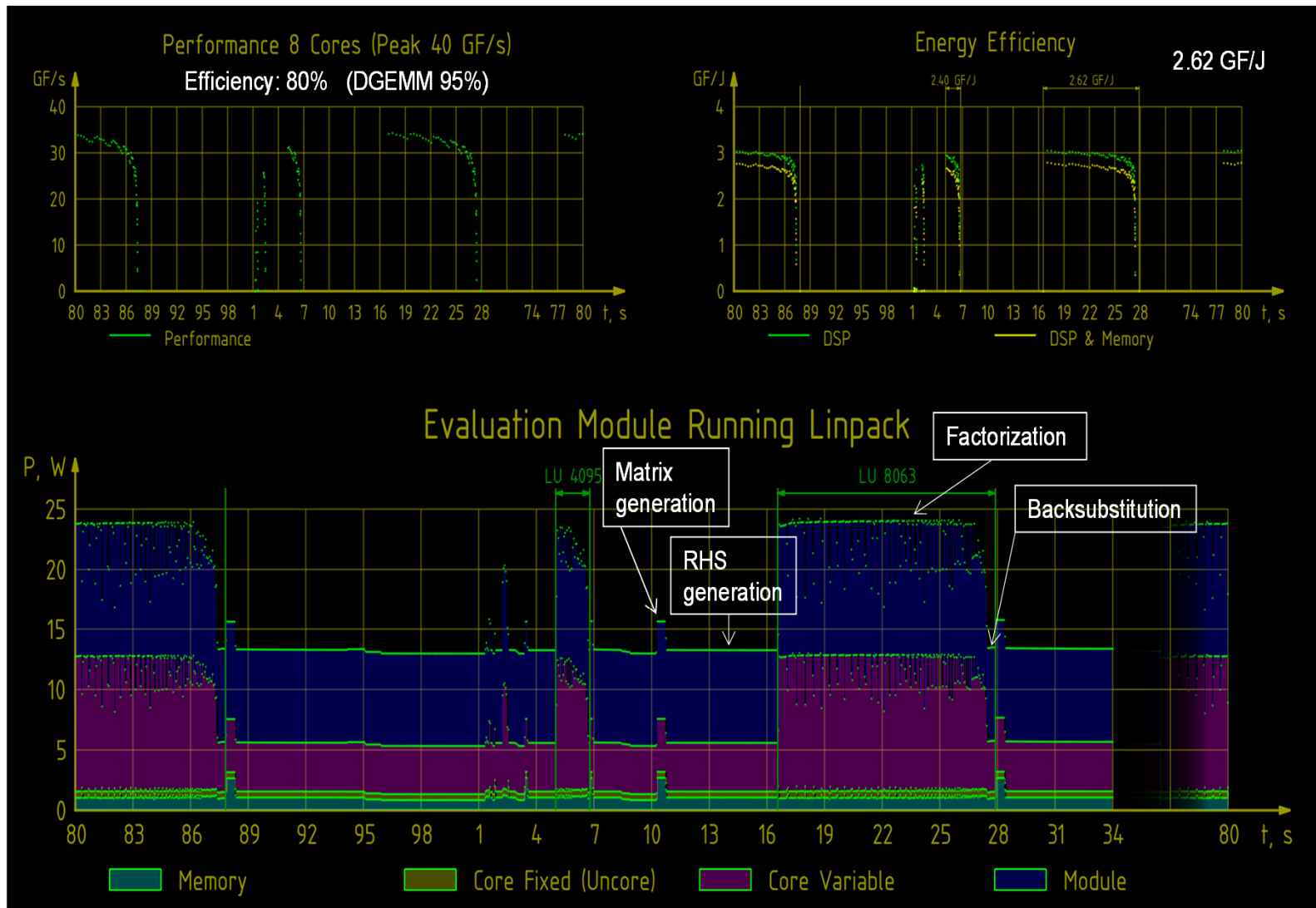

Target:
15 – 20W
32 GB
2.5 GF/W Linpack

Instrumentation of the C6678 Module

- Four differential channels for Current
- Four differential channels for Voltage
- Sampling rate 125 kHz, 125/8 kHz per channel
- Accuracy better than 1%

Linpack on TI 6678 EVM

DSP HPL Intermediate Results

Size	GF/s	Eff. %	Cores(W)	Mem (W)	Other (W)	Total (W)	Cores+Mem (MF/J)	Total (MF/J)
127	1.3	4	6.0	1.26	6.87	14.08	176	90
255	2.8	9	4.8	0.99	5.17	10.95	493	260
511	6.0	19	6.4	1.12	6.58	14.09	796	425
1023	11.3	35	8.0	1.19	7.65	15.86	1230	672
2047	16.9	53	9.2	1.10	8.13	18.40	1649	920
4095	22.0	69	10.3	1.03	8.70	20.03	1939	1097
8063	25.6	80	11.2	0.99	9.20	21.39	**2097**	1195

Imagine the impact... TI's KeyStone SoC + HP Moonshot

HP PROJECT MOONSHOT

Software Defined Servers

2013-04-19. Last week, market leader Hewlett Packard announced a huge change in the server landscape with its recent Moonshot announcement. .....

..... "TI's KeyStone II-based SoCs, which integrate fixed- and floating- point DSP cores with multiple ARM® Cortex™A-15 MPCore processors, packet and security processing, and high speed interconnect, give customers the performance, scalability and programmability needed to build software-defined servers."

HP Project Moonshot is dedicated to designing extreme low-energy server technologies. HP expects data center efficiencies to reach new heights for select workloads and applications, consuming up to 89% less energy .

We are pursuing HPC cartridges with HP and TI .....

77% less costly

89% less energy

80% less space

97% less complex

UNIVERSITY of HOUSTON

Next Prototype – Enahanced Mobile Video CPU

Fragrak 28nm Platform by Movidius

SW Controlled I/O Multiplexing

MEBI | SEBI | SDIO x3 | SPI x3 | FLSH | MIPI DSI 2x | MIPI CSI 2x | USB2 OTG | I²C x2 | I²S x2 | JTAG | UART x2

TS | GPS | TIM | RISC | NAL

64

CMX 256kB / SHAVE 04
CMX 256kB / SHAVE 08
CMX 256kB / SHAVE 12
CMX 256kB / SHAVE 16

ICB | ICB | ICB | ICB

XCB

Bridge

Main Bus

128

Stacked 256/512MB SDRAM die

64

DDR3 LP

L2 512kB

**450GFLOPS/W
(IEEE 754 SP)**

Movidius IP
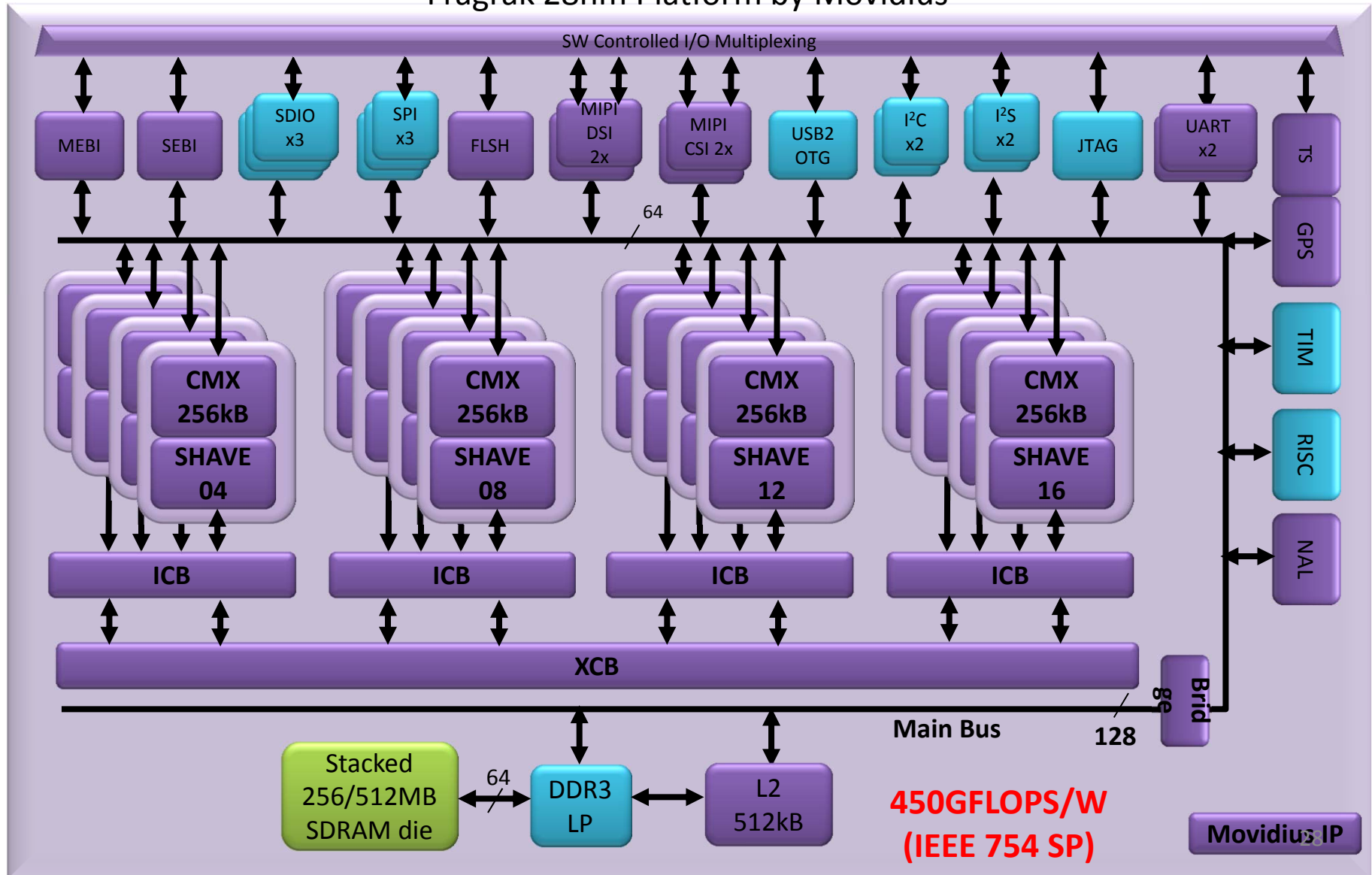
Dynamic Voltage and Frequency Scaling

The Case for Energy-Proportional Computing

Luiz André Barroso and Urs Hölzle

Google

Figure 1. Average CPU utilization of more than 5,000 servers during a six-month period. Servers are rarely completely idle and seldom operate near their maximum utilization, instead operating most of the time at between 10 and 50 percent of their maximum utilization levels.
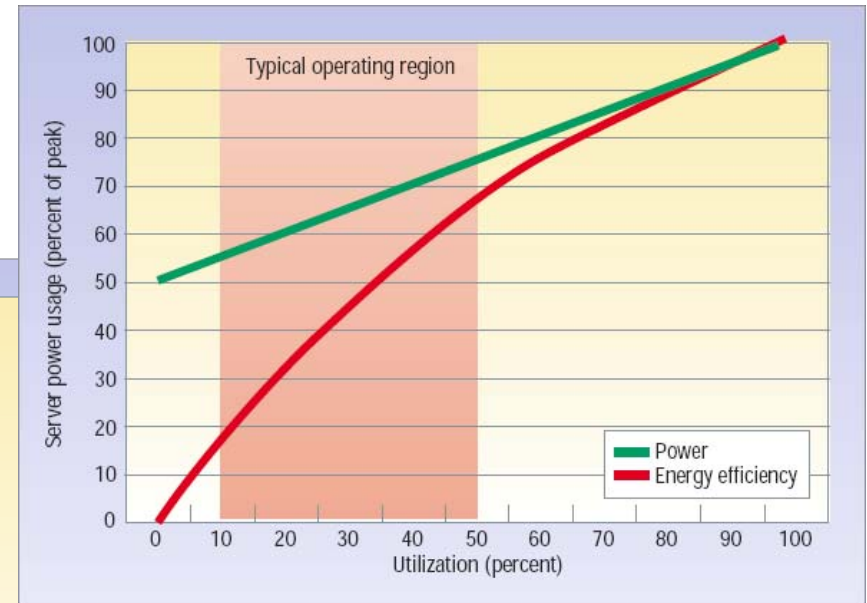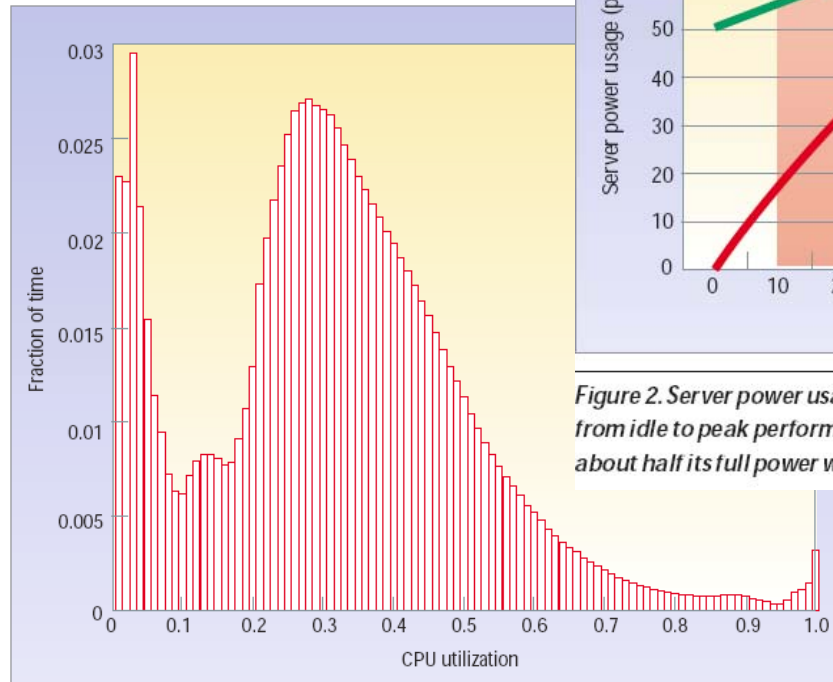
Figure 2. Server power usage and energy efficiency at varying utilization levels, from idle to peak performance. Even an energy-efficient server still consumes about half its full power when doing virtually no work.

"The Case for Energy-Proportional Computing", Luiz André Barroso, Urs Hölzle, *IEEE Computer*, vol. 40 (2007). http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//pubs/archive/33387.pdf

UNIVERSITY *of* **HOUSTON**

PDC Energy Recovery Project
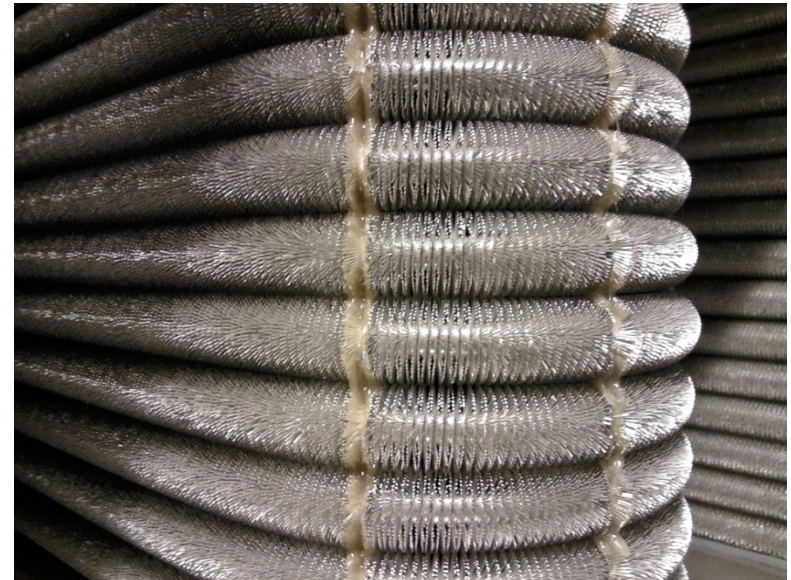
Liquid Cooling – Submersion

http://www.grcooling.com

PDC is evaluating this technology

- Server: Supermicro H8QG6 with four 6274 AMD Opteron processors with 128GB of LV DDR3 8GB DIMMs.
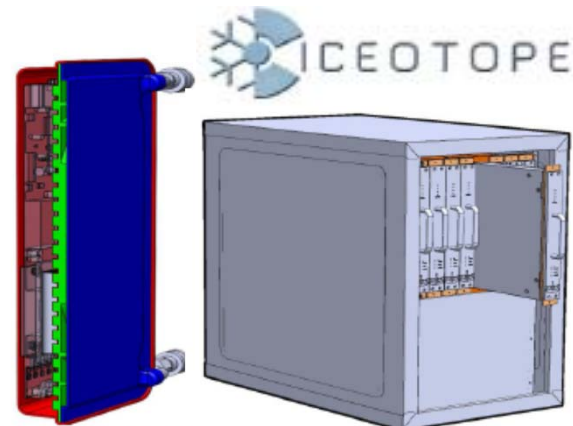- Evaluation still in progress; currently operated with coolant C and water

http://www.iceotope.co.uk

Figure 6: Iceotope modul (cross-section)

Figure 7: Standard 19" mount chassis (10U)
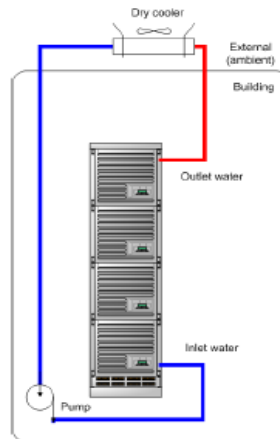
Figure 8: Example deployment

http://www.hardcorecomputer.com

New Students Welcome!!!

UNIVERSITY *of* **HOUSTON**