PDC Summer School,
Aug 26 2010
Lennart Johnsson

ADVANCED COMPUTING RESEARCH LABORATORY

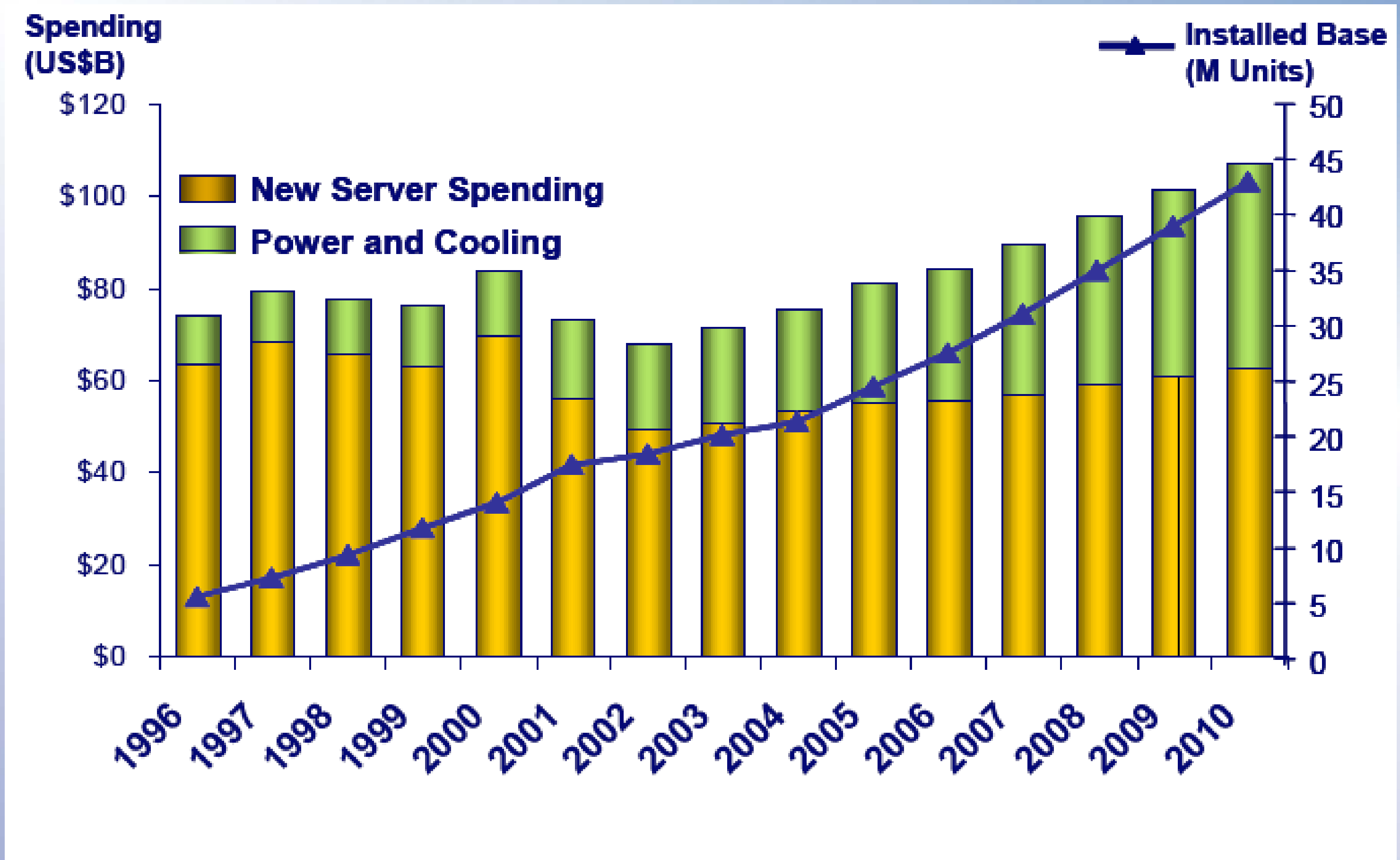
Green Computing:
A Case Study

Lennart Johnsson

Hugh Roy and Lillie Cranz Cullen Distinguished University Chair
University of Houston
Director, Texas Learning and Computation Center

Professor, School of Computer Science and Communications
KTH, Stockholm
Director, PDC

Worldwide Server Installed Base, New Server Spending, and Power and Cooling Expense



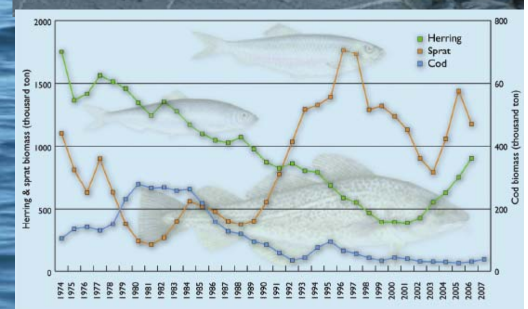
Source: IDC, 2006

Power per rack has increased from a few kW to 30+ kW

Estimate for recent purchase: 4 yr cooling cost ~1.5 times cluster cost

The US: By 2010, for every dollar spent on servers $0.70 will be spent on cooling and power
Europe and Japan: Energy cost typically 50% higher. By 2010 cooling and power cost dominate


Ozone



http://www.riksdagen.se/upload/Dokument/utskotteunamnd/200809/200809RFR_4_eng.pdf

magazine covers
courtesy of Chris Rose

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

Earth's Climate is Rapidly Entering a Novel Realm Not Experienced for Millions of Years

"Global Warming" Implies:

- Gradual,
- Uniform,
- Mainly About Temperature,
- and Quite Possibly Benign.

What's Happening is:

- Rapid,
- Non-Uniform,
- Affecting Everything About Climate,
- and is Almost Entirely Harmful.

John Holdren, Director Office of Science and Technology Policy June 25, 2008

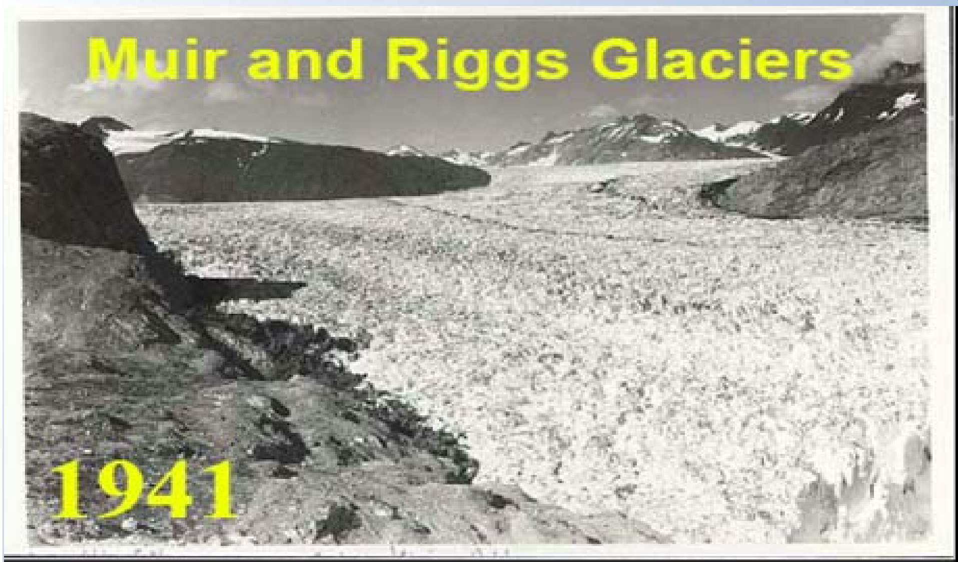
A More Accurate Term is 'Global Climatic Disruption'

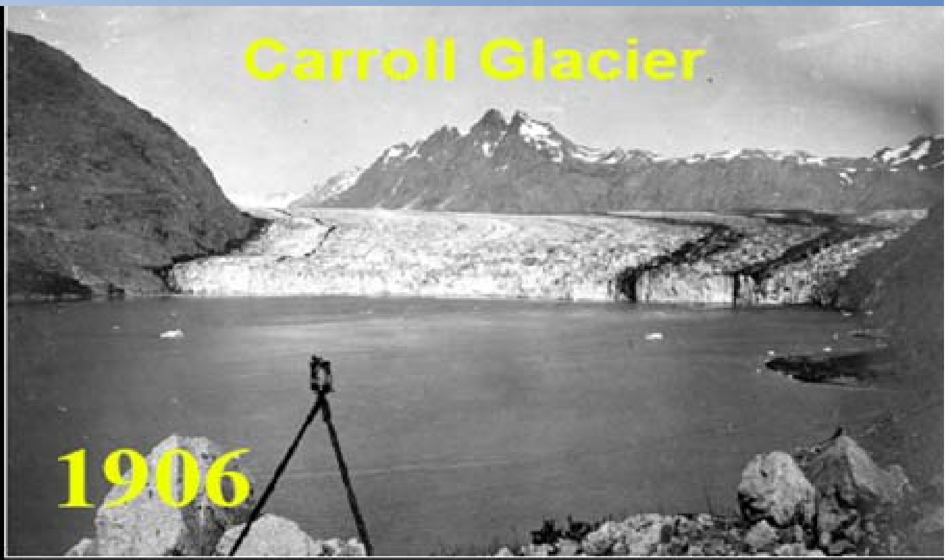
This Ongoing Disruption Is:

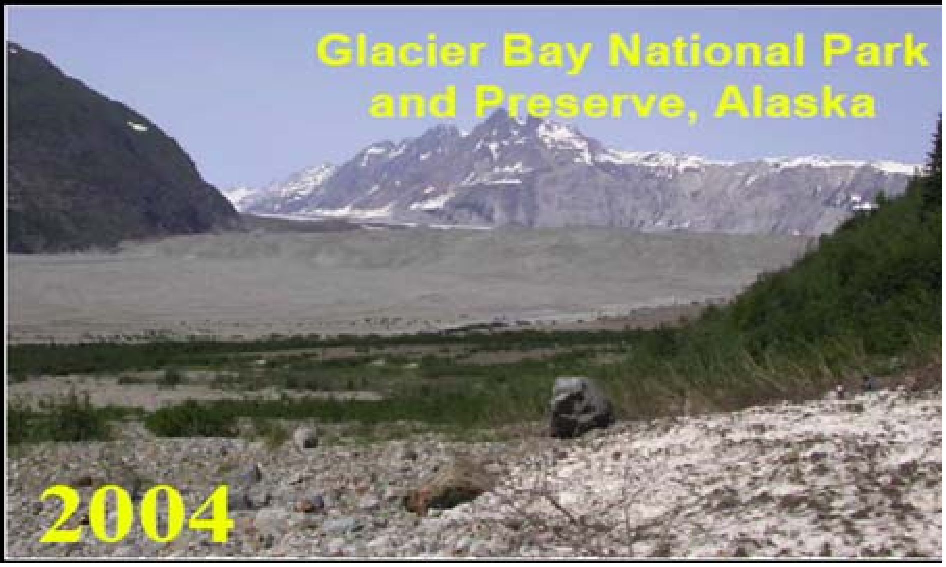
- Real Without Doubt
- Mainly Caused by Humans
- Already Producing Significant Harm
- Growing More Rapidly Than Expected"

http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/UCSB_041410_final.ppt

Retreating Glaciers







Source: Iskhaq Iskandar, http://www.jsps.go.jp/j-sdialogue/2007c/data/52_dr_iskander_02.pdf

PDC Summer School,
Aug 26 2010
Lennart Johnsson

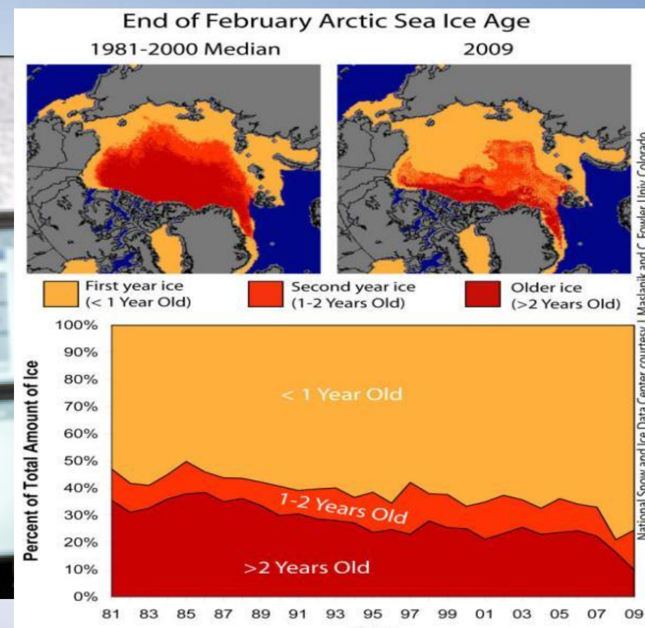
ADVANCED COMPUTING RESEARCH LABORATORY

Climatic Disruption: Decreasing Arctic Ice

"We are almost out of multiyear sea ice in the northern hemisphere--I've never seen anything like this in my 30 years of working in the high Arctic."

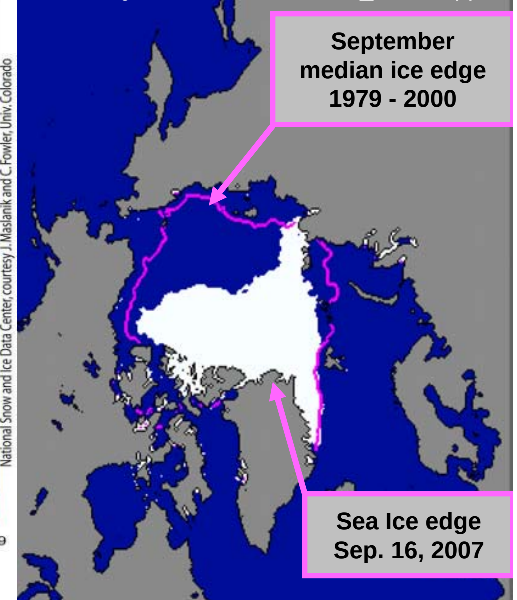
--David Barber, Canada's Research Chair in Arctic System Science at the University of Manitoba
October 29, 2009

http://news.yahoo.com/s/nm/20091029/sc_nm/us_climate_canada_arctic_1



http://news.cnet.com/8301-11128_3-10213891-54.html

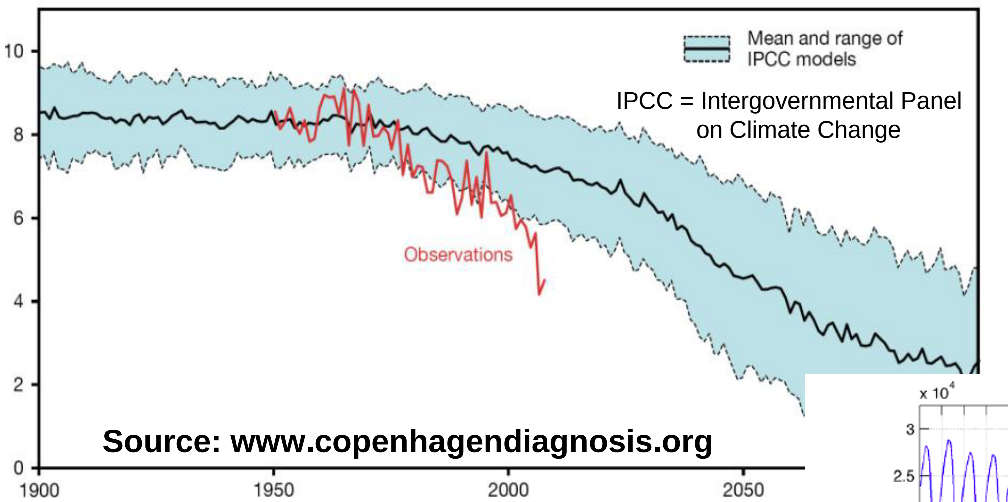
http://alaskaconservationsolutions.com/acs/images/stories/docs/AkCS_current.ppt



http://alaskaconservationsolutions.com/acs/images/stories/docs/AkCS_current.ppt

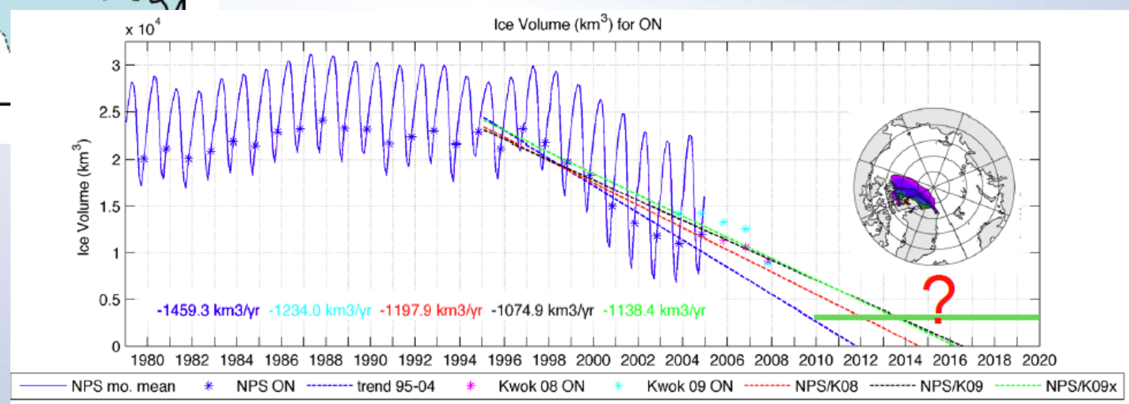
ADVANCED COMPUTING RESEARCH LABORATORY

Arctic Summer Ice Melting Accelerating



Ice area millions km², September minimum

Ice Volume km³



Observational estimates (cyan / purple stars):
- Obs Fall (ON) '07 volume <9000 km³ , ~20% uncertainty,
- Negative volume trends: 1197 - 1234 km³/yr
- Combined (95-07) model / data linear volume trend projects ice-free fall by 2016
- Same trend with extended K09 (assuming the constant ON volume for 07 - 09)
- Some (?) sea ice will remain beyond due to increased ridging of thinner ice
- Uncertainty (95-07) is ±3yrs and not all volume must disappear

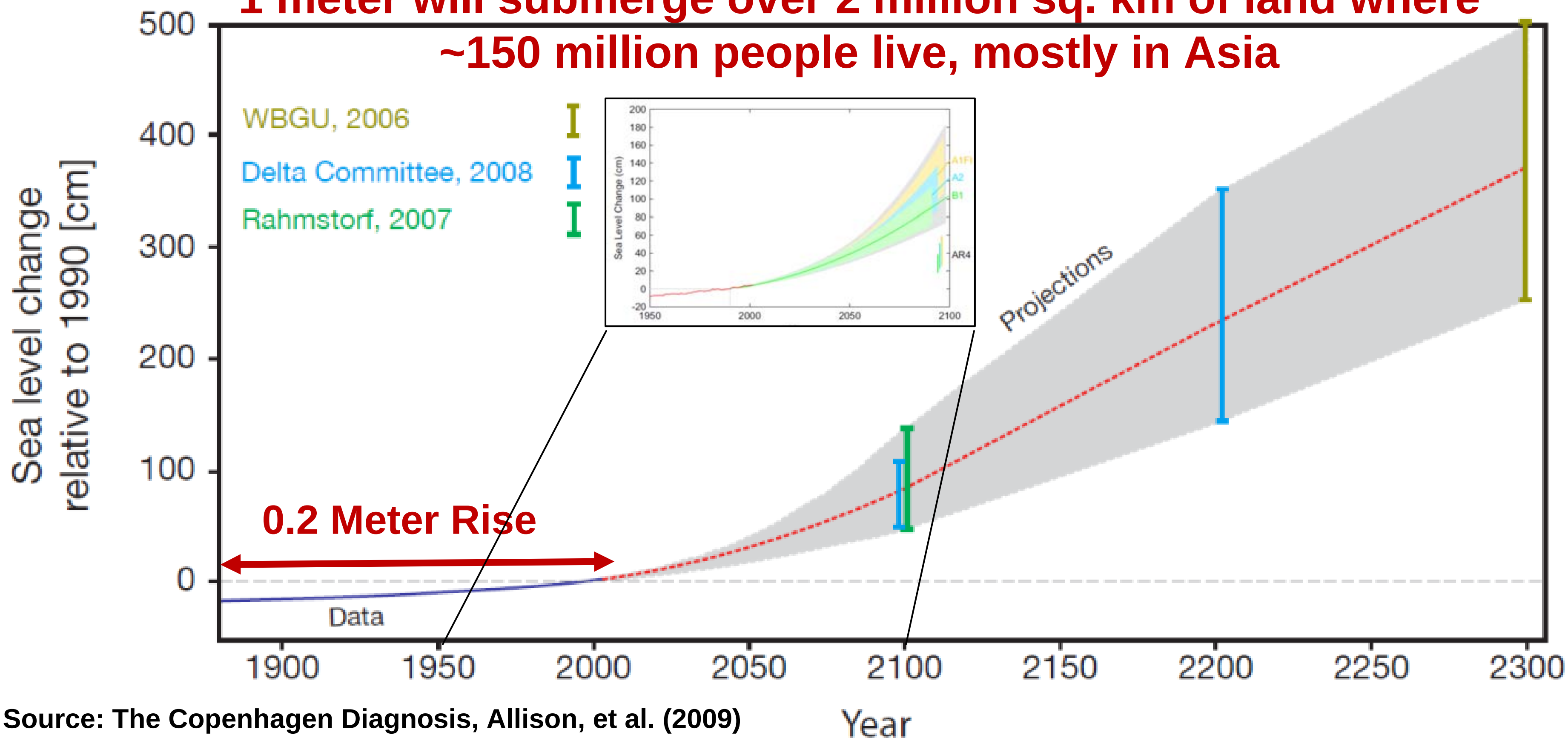
Kwok et al., JGR 2009, Kwok & Cunnigham, JGR 2008

ADVANCED COMPUTING RESEARCH LABORATORY

Human Induced Sea Level Rise Will Continue for Centuries

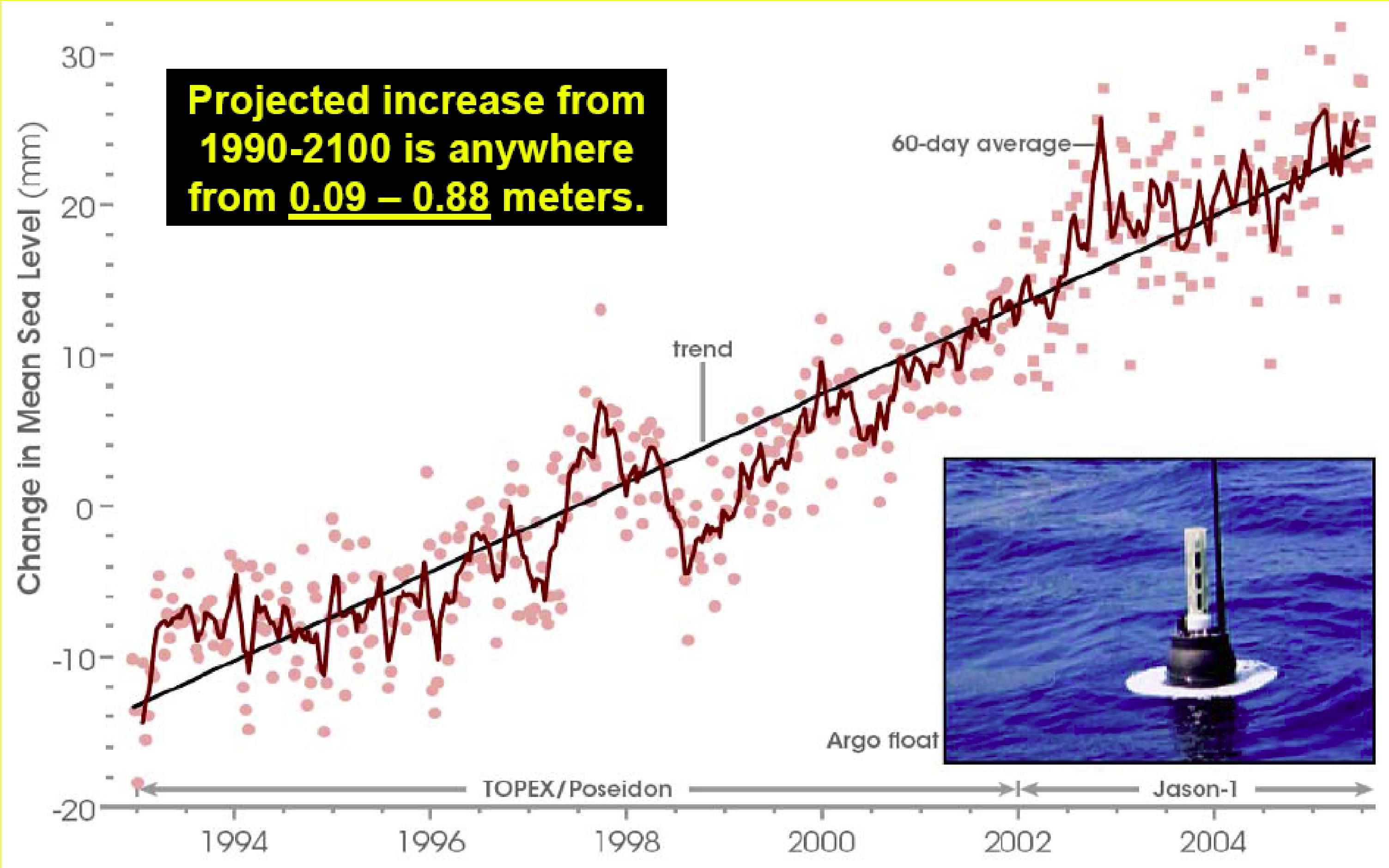
CO_2 Emissions are an Impulse to Earth Climate System—Equilibrium Response will Take Centuries

1 meter will submerge over 2 million sq. km of land where ~150 million people live, mostly in Asia



Source: The Copenhagen Diagnosis, Allison, et al. (2009)

Sea level has increased about 3 mm/yr between 1993 and 2005



1/3[rd] due to melting glaciers

2/3[rd] due expansion from warming oceans

Source: Trenberth, NCAR 2005

Source: Iskhaq Iskandar, http://www.jsps.go.jp/j-sdialogue/2007c/data/52_dr_iskander_02.pdf

Cataclysmic Global Consequences: Inundation

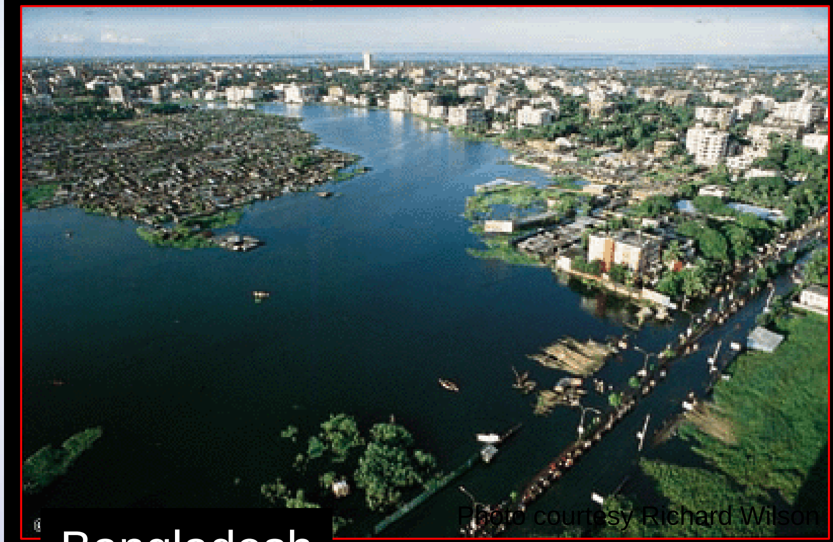


Photo courtesy Richard Wilson

Bangladesh

- Bangladesh: More than 17 million people within 3 feet of sea level
- Tuvalu: Island nation, highest elevation 15 ft; mostly less than 1m
- Lohachara: First inhabited island (10,000 people) submerged

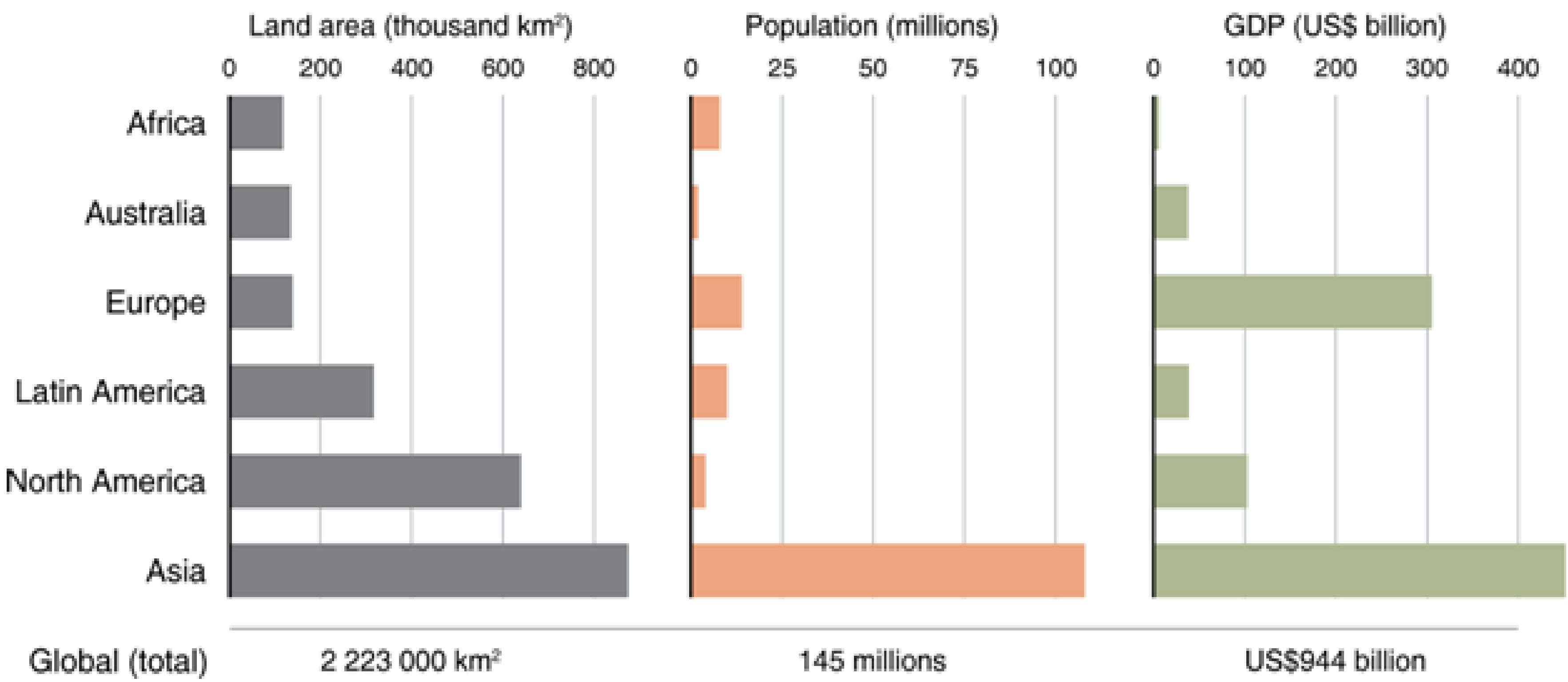
Independent, 12/06


Photo courtesy ourbangla.com


Photo © Gary Braasch

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

Population, Area and Economy Affected by a 1 m Sea Level Rise



http://maps.grida.no/go/graphic/population-area-and-economy-affected-by-a-1-m-sea-level-rise-global-and-regional-estimates-based-on-

"Global sea level linked to global temperature,"
Martin Vermeer and Stefan Rahmstorf, PNAS, v. 106, 21527–21532 (2009)

UNEP G·R·I·D ARENDAL
Environmental Knowledge for Change

Severe Weather

Katrina – August 28, 2005

Andrew: ~$27B (1992)

Charley: ~$ 7B (2004)

Hugo: ~$ 4B (1989) ($6.2B 2004 dollars)

Frances: ~$ 4B (2004)

Katrina: > $100B (2005)

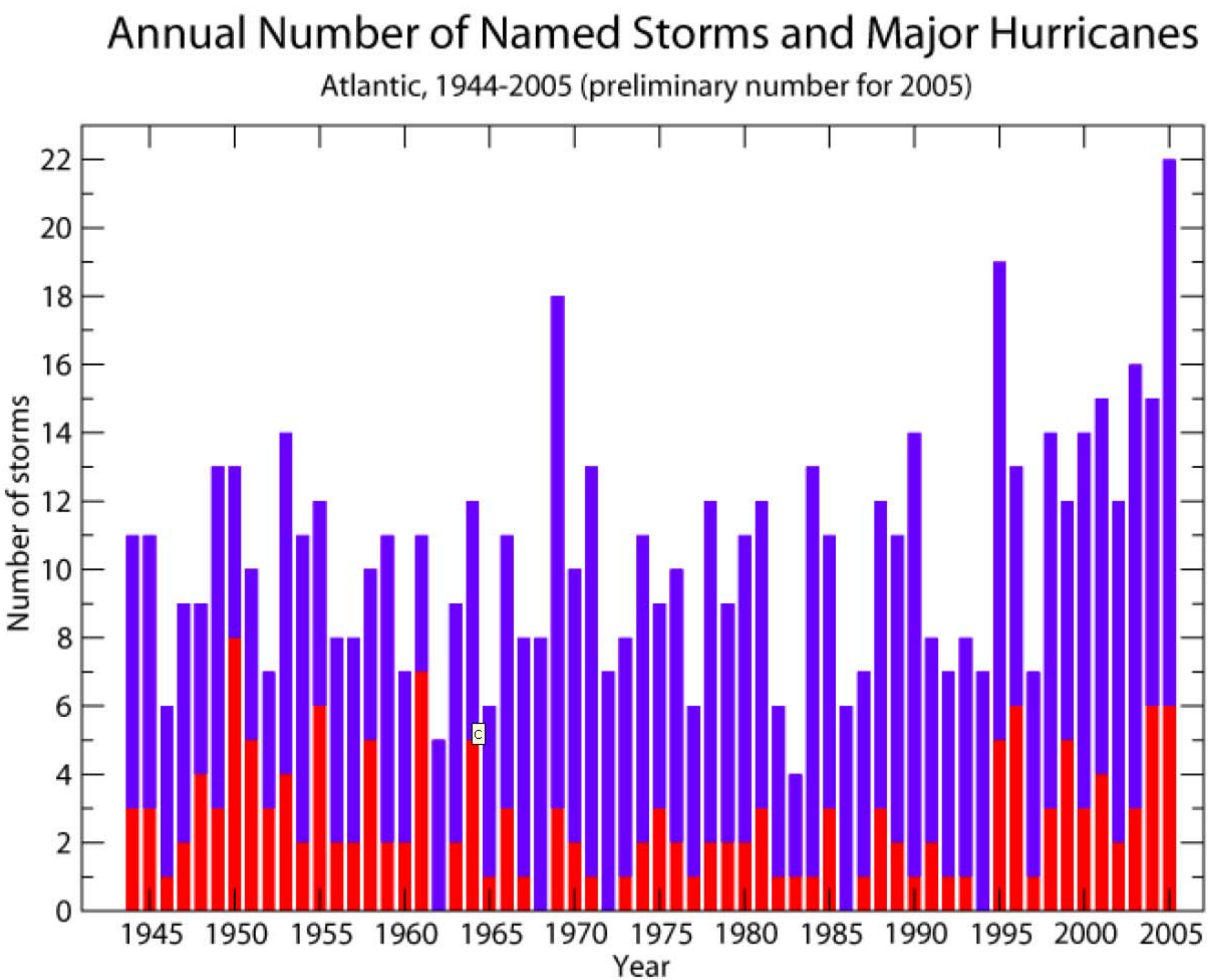
Hurricanes:

For 1925 - 1995 the US cost was $5 billion/yr for a total of 244 landfalls. But, hurricane Andrew alone caused damage in excess of $27 billion.

The US loss of life has gone down to <20/yr typically. The Galveston "Great Hurricane" year 1900 caused over 6,000 deaths.

Since 1990 the number of landfalls each year is increasing.

Warnings and emergency response costs on average $800 million/yr. Satellites, forecasting efforts and research cost $200 – 225 million/yr.



http://www.geo.umass.edu/climate/global_warming_update.pdf

PDC Summer School,
Aug 26 2010
Lennart Johnsson

ADVANCED COMPUTING RESEARCH LABORATORY

Tornados


http://en.wikipedia.org/wiki/File:Dimmit_Sequence.jpg

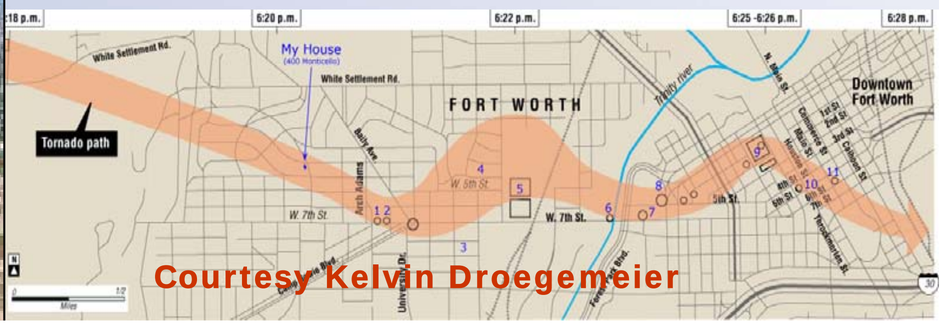

http://www.miapearlman.com/images/tornado.jpg


http://www.crh.noaa.gov/mkx/document/tor/images/tor060884/damage-1.jpg


http://en.wikipedia.org/wiki/Tornado
Anadarko, Oklahoma


www.drjudywood.com/.../spics/tornado-760291.jpg


http://g.imagehost.org/0819/tornado-damage.jpg



Courtesy Kelvin Droegemeier

PDC Summer School,
Aug 26 2010
Lennart Johnsson

ADVANCED COMPUTING RESEARCH LABORATORY

Russia may have lost 15,000 lives already, and $15 billion, or 1% of GDP, according to Bloomberg.
The smog in Moscow is a driving force behind the fires' deadly impact, with 7000 being killed already in the city. Aug 10, 2010

Wildfires

In a single week, San Diego County wildfires killed 16 people, destroyed nearly 2,500 homes and burned nearly 400,000 acres. Oct 2003

http://legacy.signonsandiego.com/news/fires/weekoffire/images/mainimage4.jpg

img.ibtimes.com/www/data/images/full/2010/08/

Russia Wildfires 2010

http://topnews.in/law/files/Russian-fires-control.jpg

http://msnbcmedia1.msn.com/j/MSNBC/Components/Photo/_new/100810-russianFire-vmed-218p.grid-6x2.jpg

http://www.tolerance.ca/image/photo_1281943312664-2-0_94181_G.jpg

Floods

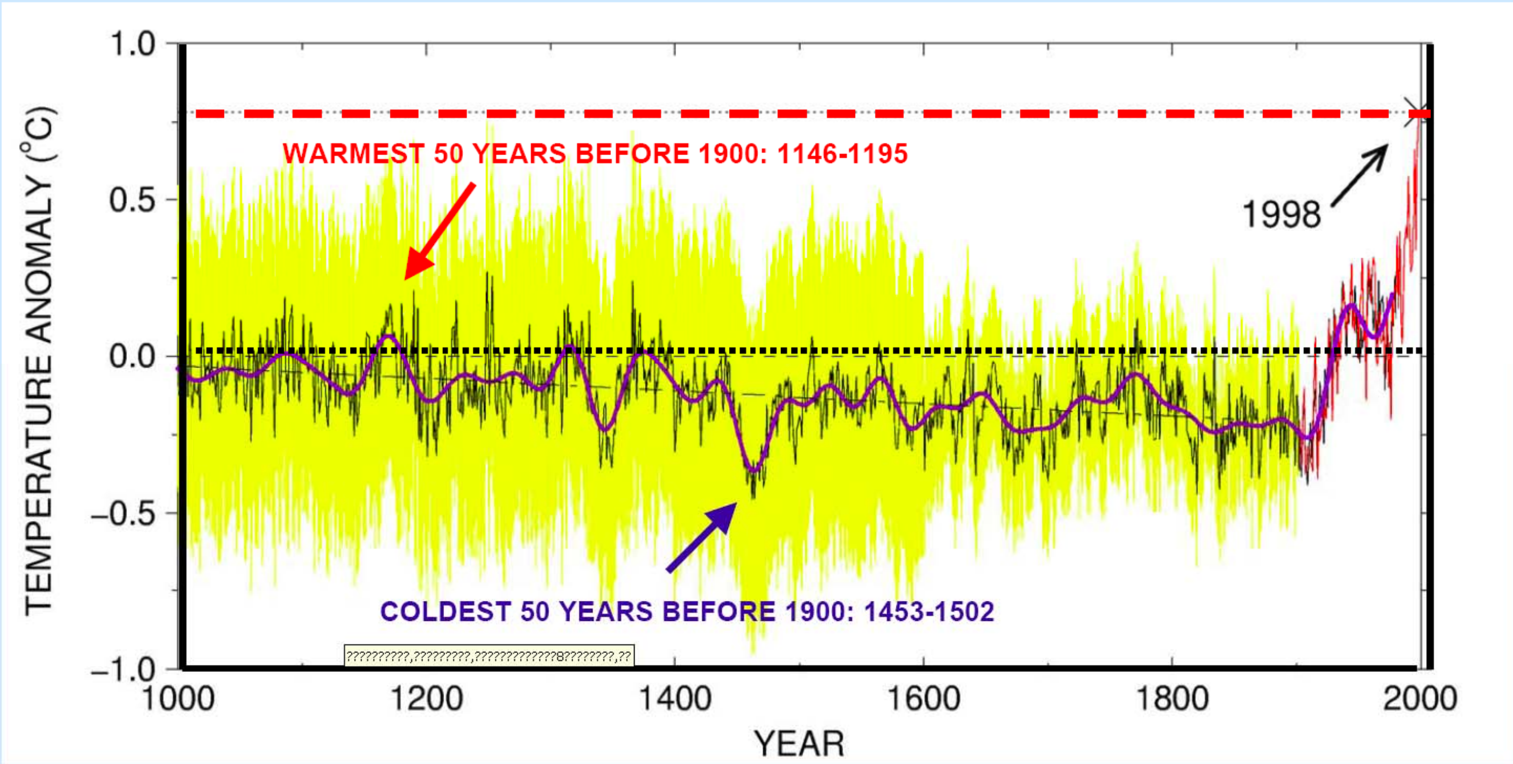
April 30 – May 7, 2010
TN, KY, MS
31 deaths. Nashville Mayor Karl Dean estimates the damage from weekend flooding could easily top $1 billion.

UK June – July 2007
13 deaths
more than 1 million affected
cost about £6 billion

China,Bloomberg Aug 17,2010
1450 deaths through Aug 6
Aug 7 1254 killed in mudslide with 490 missing

"Northern Hemisphere temperatures during the past millennium: *inferences, uncertainties and limitations*"

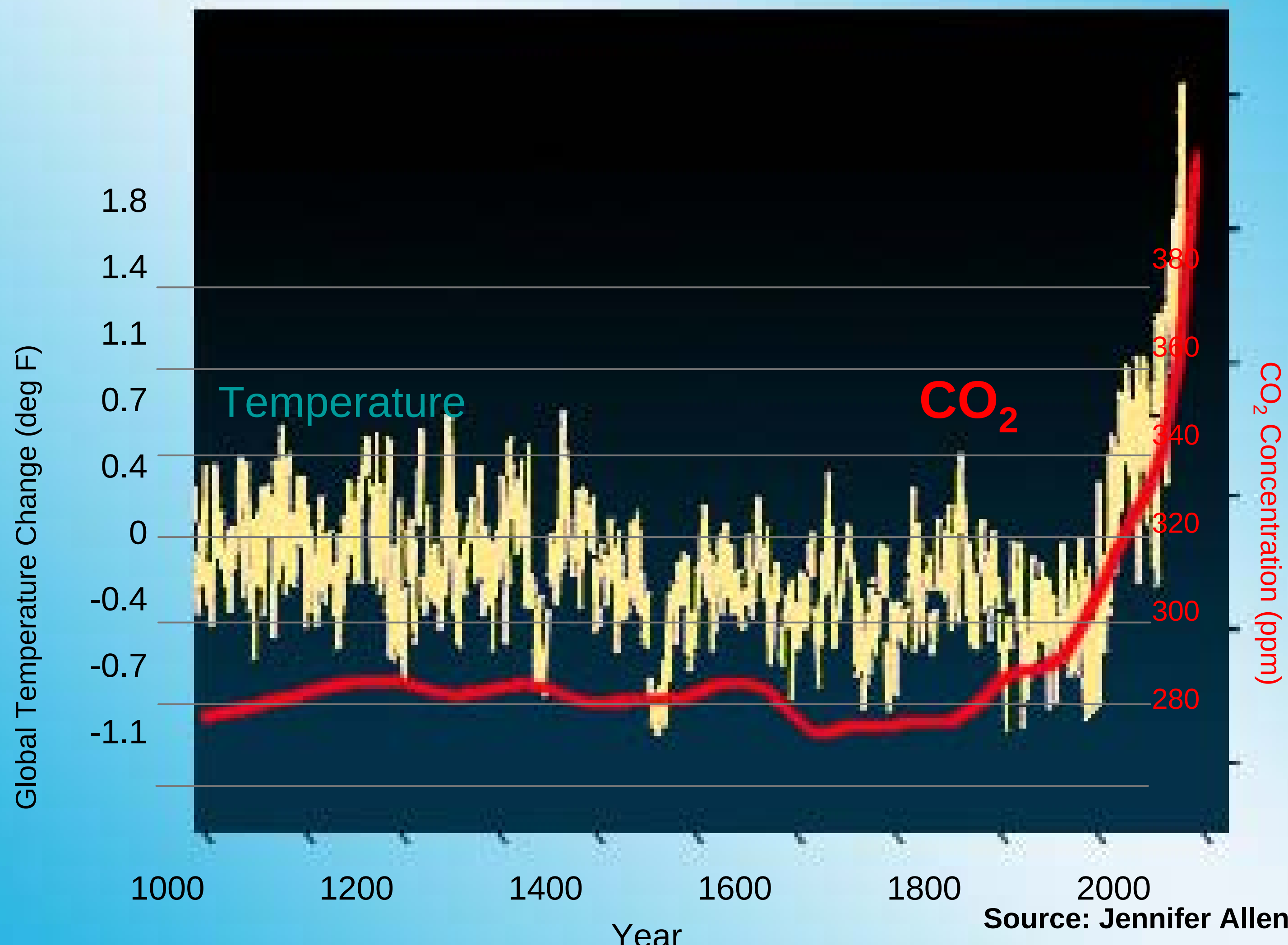
M. Mann, R. Bradley & M. Hughes, 1999



"*Mann[et al] effectively erased the well-known phenomena of the Medieval Warming Period--when, by the way, it was warmer than it is today--and the Little Ice Age…*" J. Inhofe 2005

http://www.geo.umass.edu/climate/global_warming_update.pdf

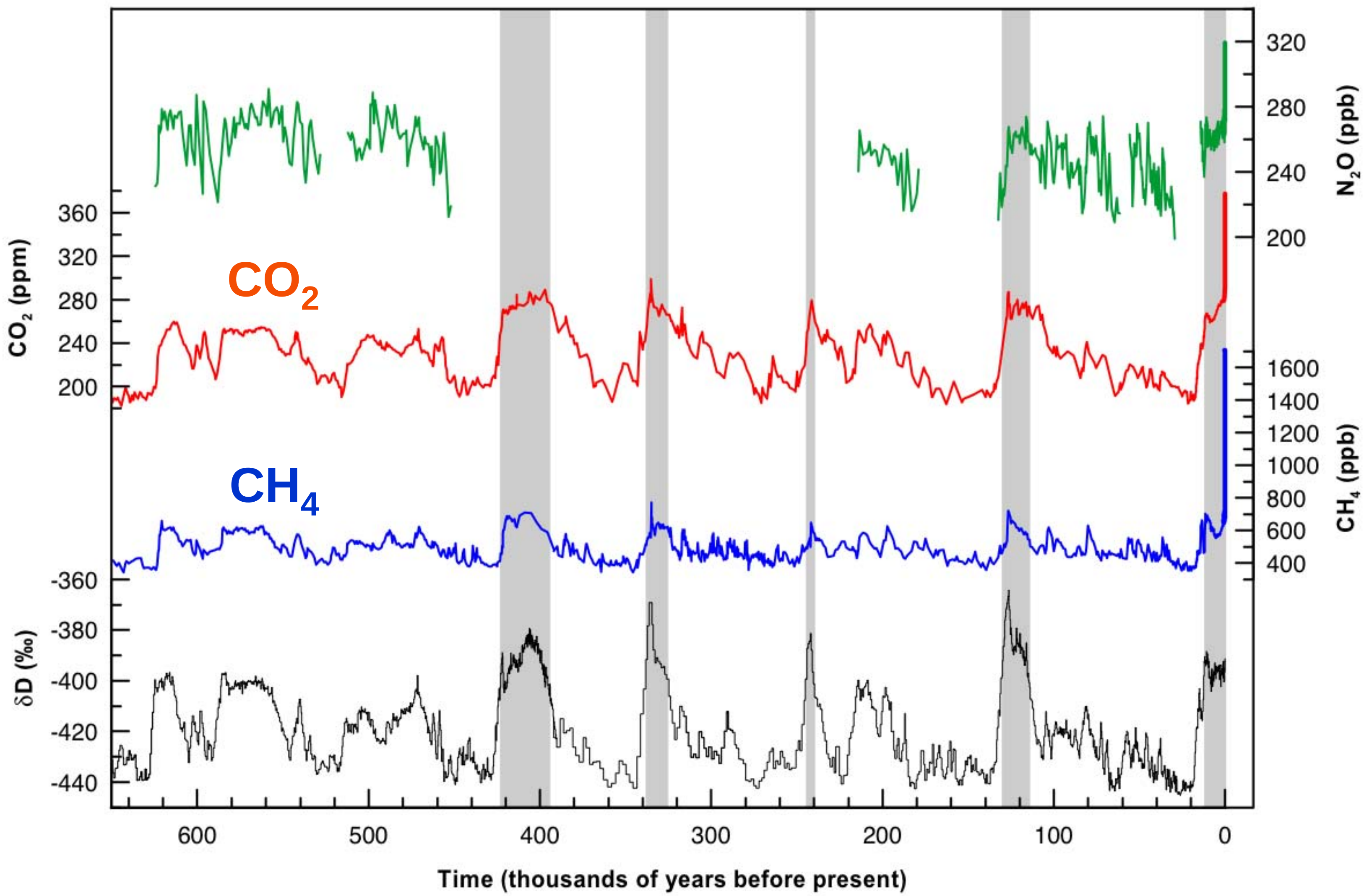
1000 Years of CO_2 and Global Temperature Change



Source: Jennifer Allen, ACIA 2004

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

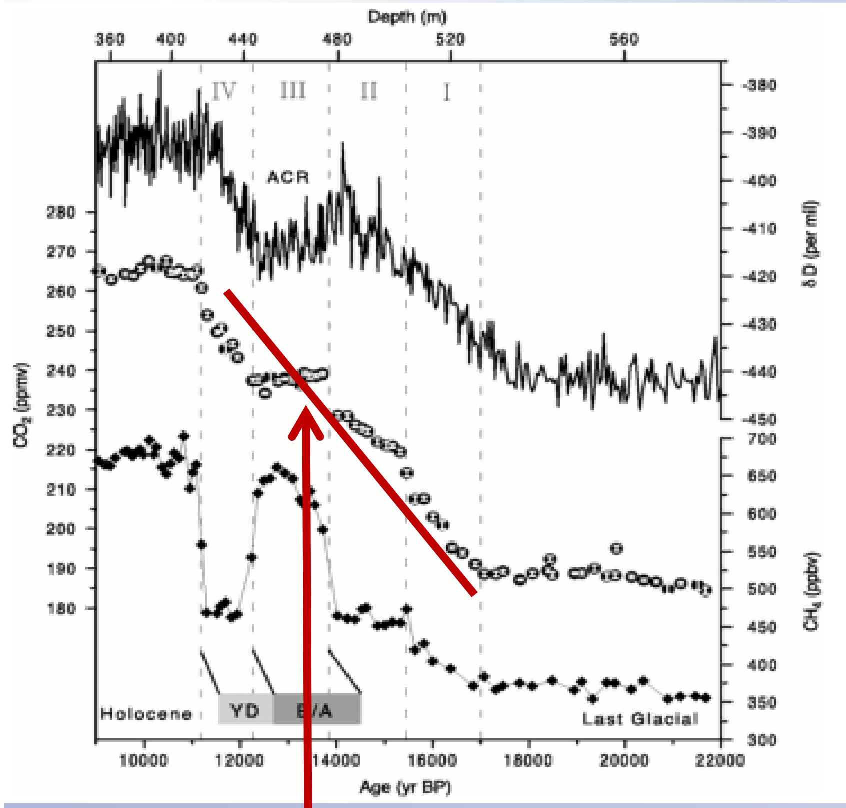
Glacial-Interglacial Ice Core Data



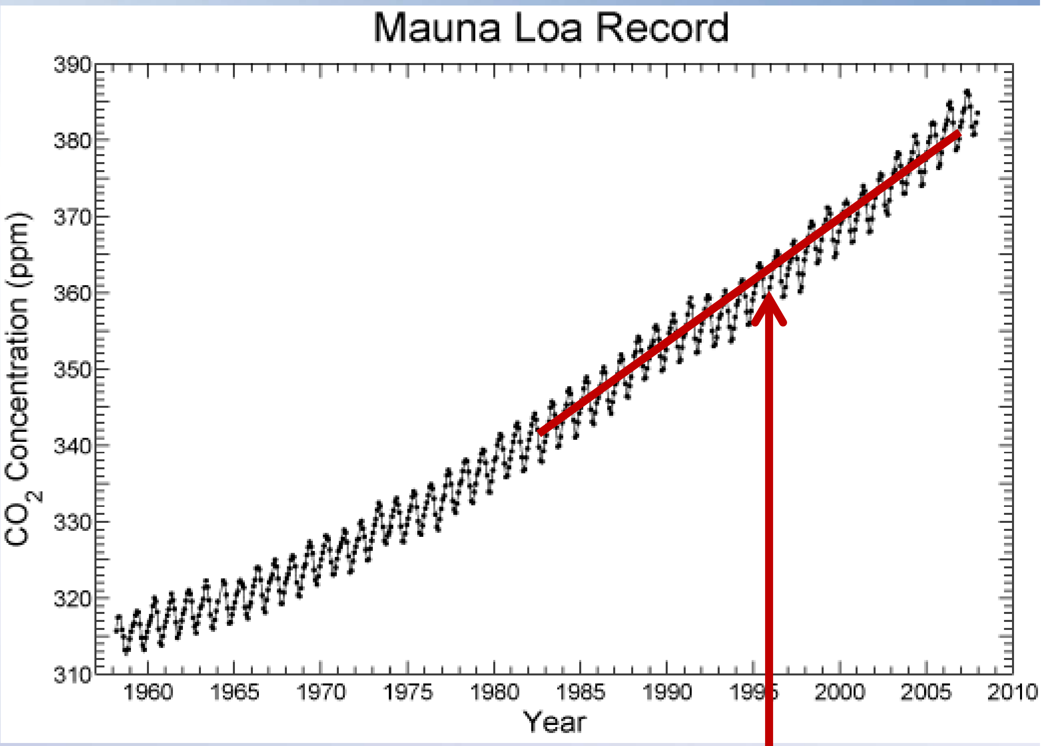
Source: http://alaskaconservationsolutions.com/acs/images/stories/docs/AkCS_current.ppt

ADVANCED COMPUTING RESEARCH LABORATORY

The Earth is Warming Over 100 Times Faster Today Than During the Last Ice Age Warming!



http://scrippsco2.ucsd.edu/program_history/keeling_curve_lessons.html

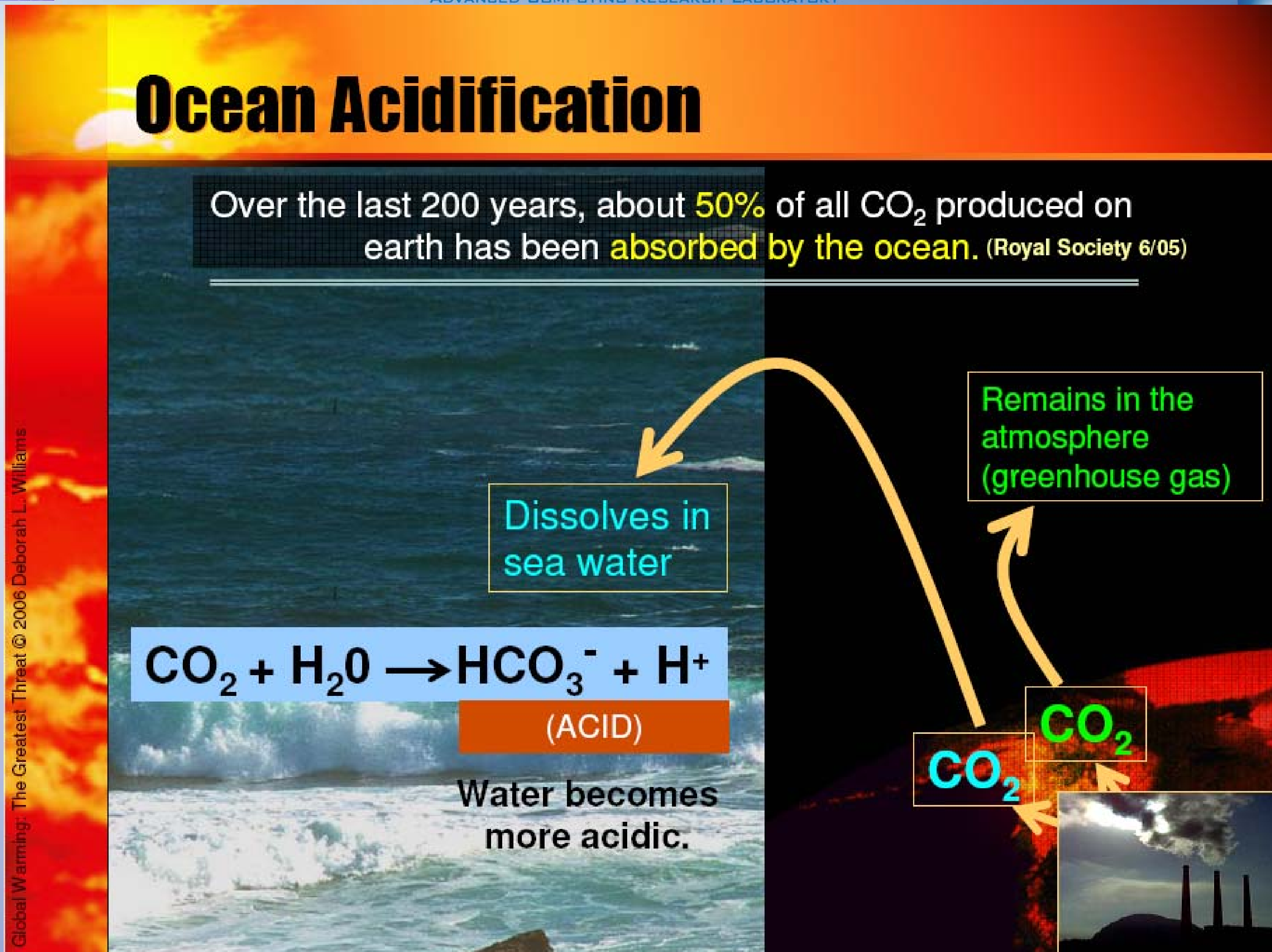


CO_2 Rose From 185 to 265ppm (80ppm) in 6000 years or 1.33 ppm per Century

CO_2 Has Risen From 335 to 385ppm (50ppm) in 30 years or 1.6 ppm per Year

http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/UCSB_041410_final.ppt

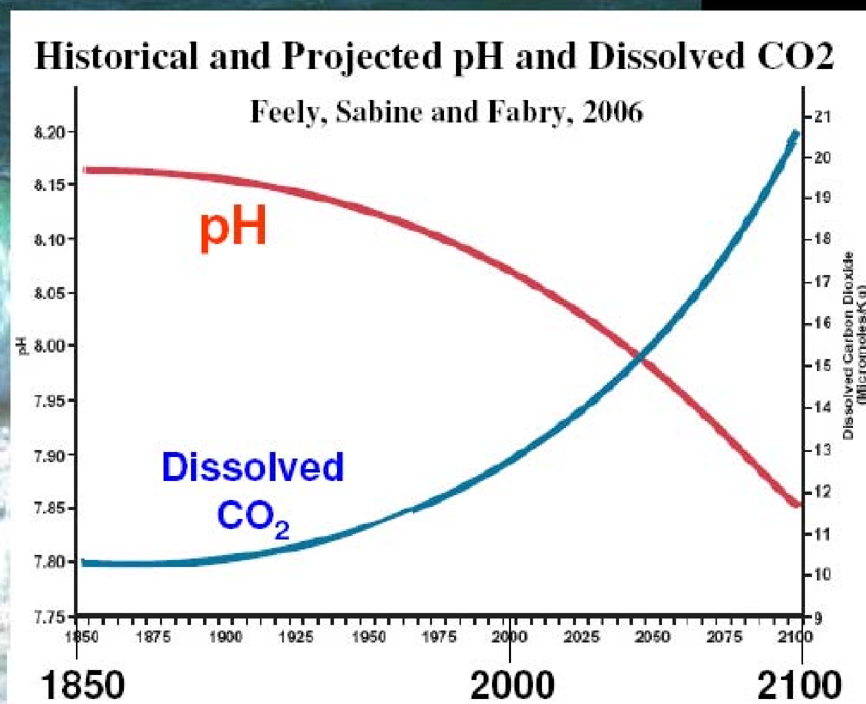
Global Cataclysmic Concerns



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

Ocean Acidification

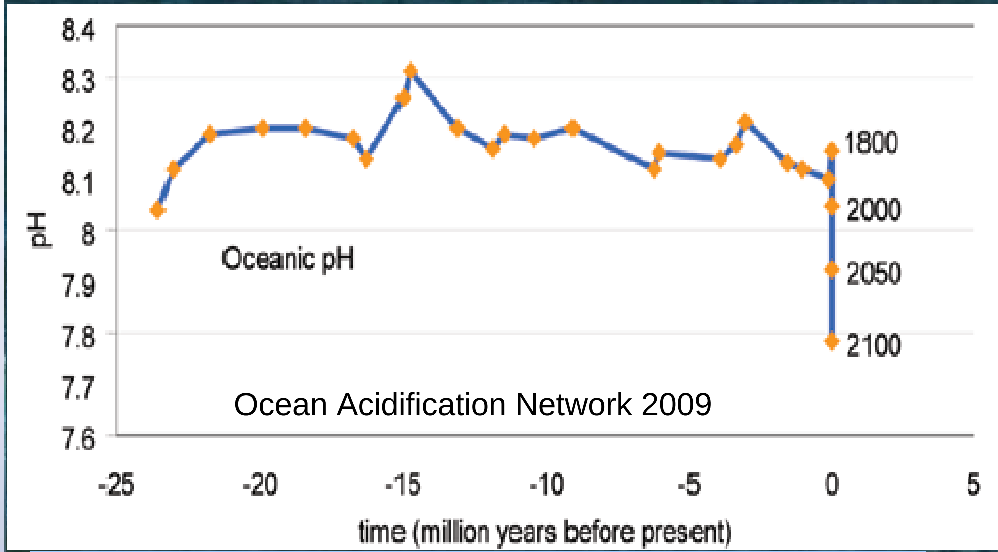
Lower pH = MORE ACID



- Since 1850, ocean pH has decreased by about 0.1 unit (30% increase in acidity). (Royal Society 2006)
- At present rate of CO_2 emission, acidity predicted to increase by 0.4 units pH (3-fold increase in H ions) by 2100.
- Carbonate ion concentrations decrease.

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

Ocean Acidification



$$H^+ + CO_3^{2-} \longrightarrow HCO_3^-$$

Carbonate

Bicarbonate

Less Carbonate

$$Ca + CO_3^{2-} \longrightarrow CaCO_3$$

- Hydrogen ions combine with carbonate ions in the water to form bicarbonate
- This removes carbonate ions from the water making it more difficult for organisms to form the **$CaCO_3$** they need for their shells.
- Carbonate ion concentrations decrease
- Aragonite, critical for most shells and coral is one of two polymorphs of $CaCO_3$

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

Ocean Acidification

- Animals with calcium carbonate shells -- corals, sea urchins, snails, mussels, clams, certain plankton, and others -- have trouble building skeletons and shells can even begin to dissolve. "Within decades these shell-dissolving conditions are projected to be reached and to persist throughout most of the year in the polar oceans." (Monaco Declaration 2008)


Pteropod


Squid

- Pteropods (an important food source for salmon, cod, herring, and pollock) likely not able to survive at CO_2 levels predicted for 2100 (600ppm, pH 7.9) (Nature 9/05)
- Coral reefs at serious risk; doubling CO_2, stop growing and begin dissolving (GRL 2009)
- Larger animals like squid may have trouble extracting oxygen
- Food chain disruptions


Clam

All photos this page courtesy of NOAA

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

Coral Bleaching

- Corals damaged by higher water temperatures and acidification
- Higher water temperatures cause bleaching: corals expel *zooxanthellae* algae
- Corals need the algae for nutrition

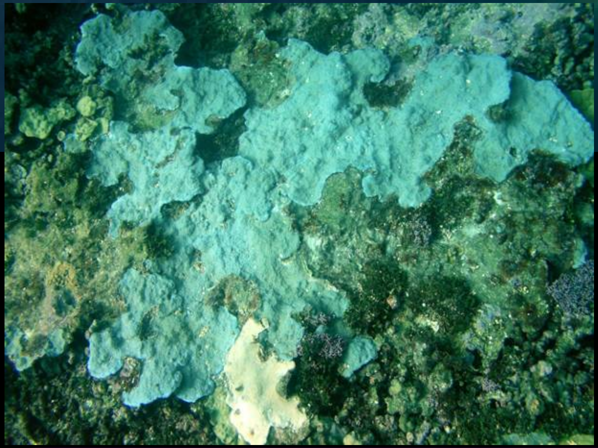
Healthy staghorn coral

Bleached staghorn coral (algae expelled)

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

Coral Bleaching

- Belize estimated 40% loss since 1998 (Independent, 6/06)
- Seychelles 90% bleached in 1998, now only 7.5% cover; 50% decline in fish diversity (Proceedings of the National Academy of Sciences, 5/06)
- If warming continues, Great Barrier Reef could lose 95% of living coral by 2050 (Ove Hoegh-Guldberg/ WWF, 2005)
- Disease followed bleaching in Caribbean Reefs in 2005/06 (Proceedings of the National Academy of Science, 8/06)



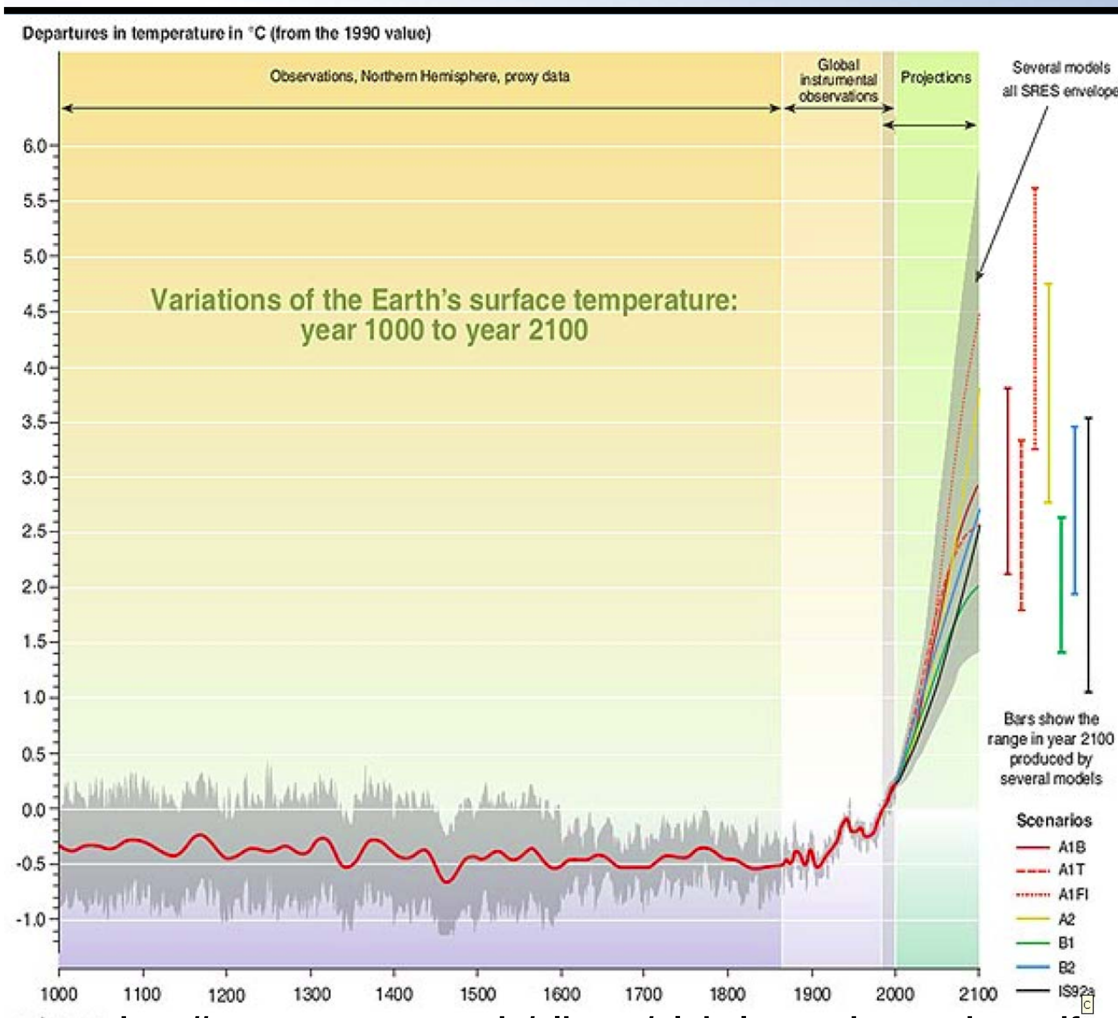
Source: http://alaskaconservationsolutions.com/acs/images/stories/docs/AkCS_current.ppt

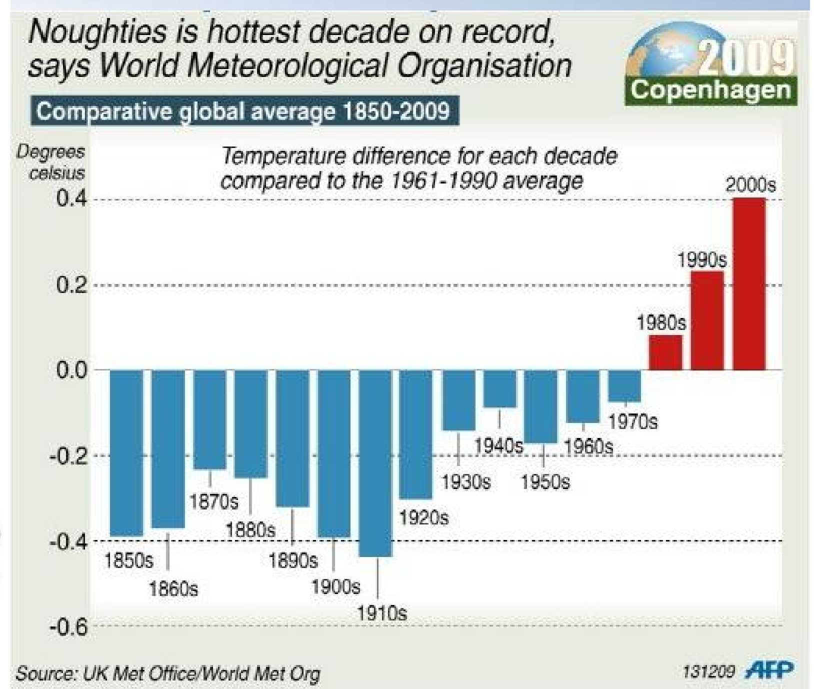
International Health Impacts

- Increased epidemics of malaria in Africa; new cases in Turkey and elsewhere
- Increased cerebral-cardiovascular conditions in China
- Increased heat wave deaths on Europe (52,000 in 2003), typhoid fever, Vibrio vulnificus, Ostreopsis ovata, Congo Crimea hemorrhagic fever
- Dengue fever in SE Asia
- More mercury release, flooding, storms
- WHO 150,000 deaths and 5 million illnesses per year attributable to global warming; numbers expected to double (Nature, 2005)

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

Temperature predicted to rise in all scenarios

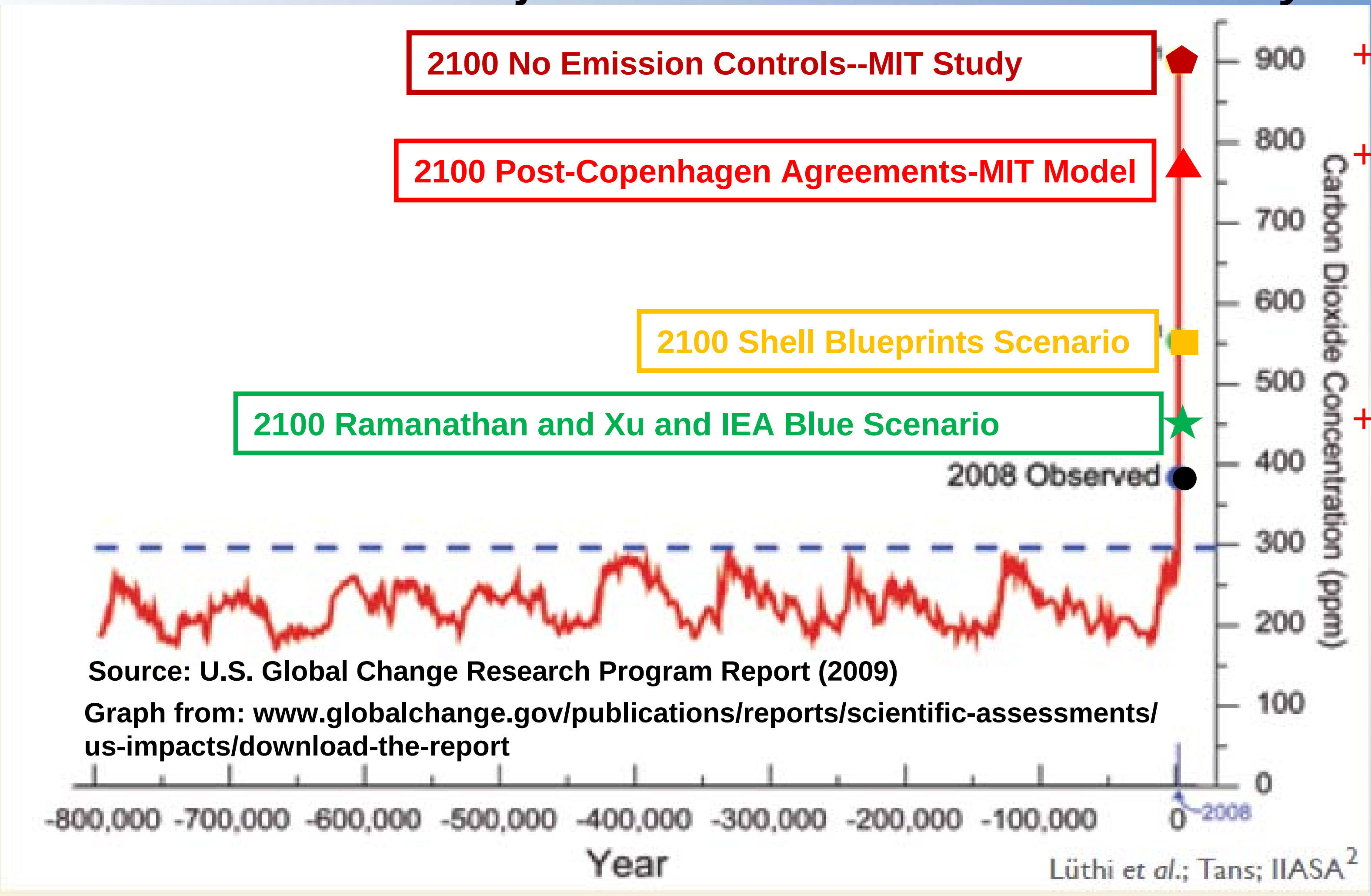




http://www.geo.umass.edu/climate/global_warming_update.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

Atmospheric CO_2 Levels for Last 800,000 Years and Several Projections for the 21st Century



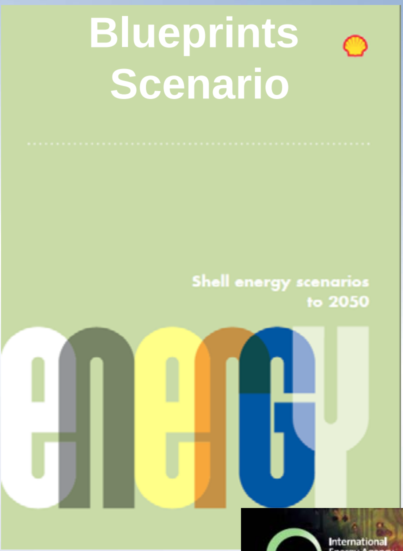
Source: U.S. Global Change Research Program Report (2009)

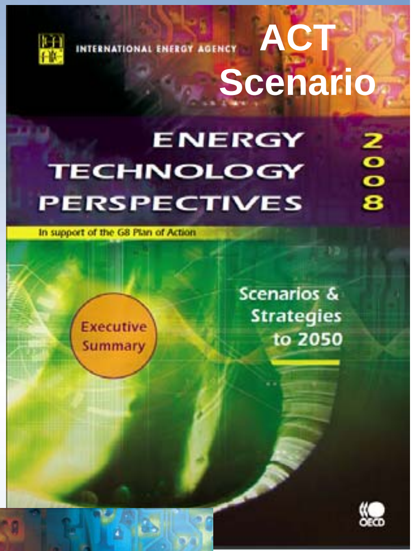
Graph from: www.globalchange.gov/publications/reports/scientific-assessments/us-impacts/download-the-report

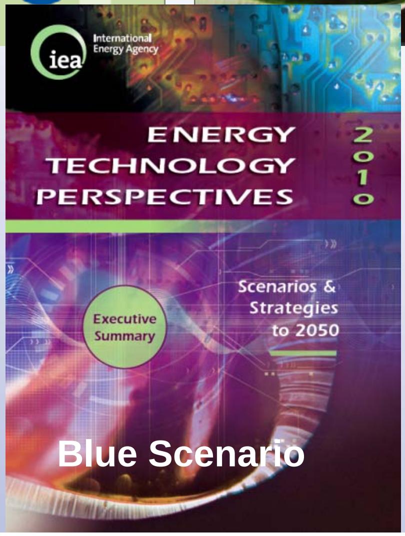
http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/AALD_Infrastructure_071610.final.ppt

What Changes to the Global Energy System Must be Made by 2050 To Limit Climate Change?

- Two Targets
  - 550 ppm
    - Shell Oil Blueprints Scenario
    - International Energy Agency ACT Scenario
    - Bring CO_2 Emissions by 2050 Back to 2005 Levels
  - 450 ppm
    - Ramanathan and Xu Reduction Paths
    - IEA Blue Scenario
    - Bring CO_2 Emissions by 2050 to 50% Below 2005 Levels



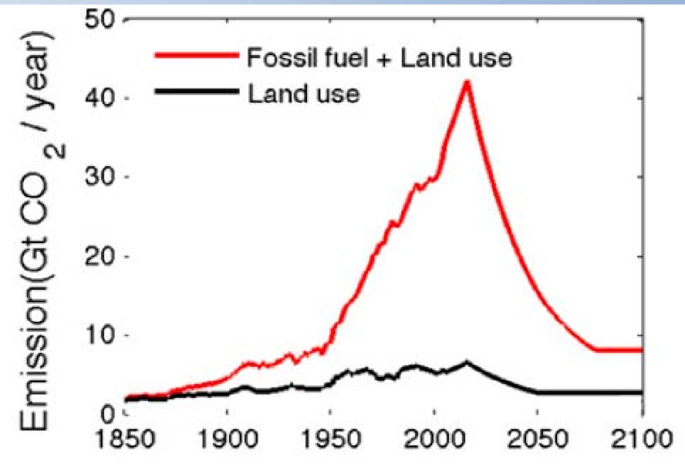




http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/NCAR_072310.ppt

Urgent Actions Required to Limit Global Warming to Less Than 2 Degrees Centigrade

- Three Simultaneous Actions
  - Reduce annual CO_2 emissions 50% by 2050—keep CO_2 concentration below 441 ppm
  - Balance removing cooling aerosols by removing warming black carbon and ozone
  - Greatly reduce emissions of short-lived GHGs-Methane and Hydrofluorocarbons



John Sterman, Jay W. Forrester Professor in Computer Science Professor of System Dynamics Director, MIT System Dynamics Group

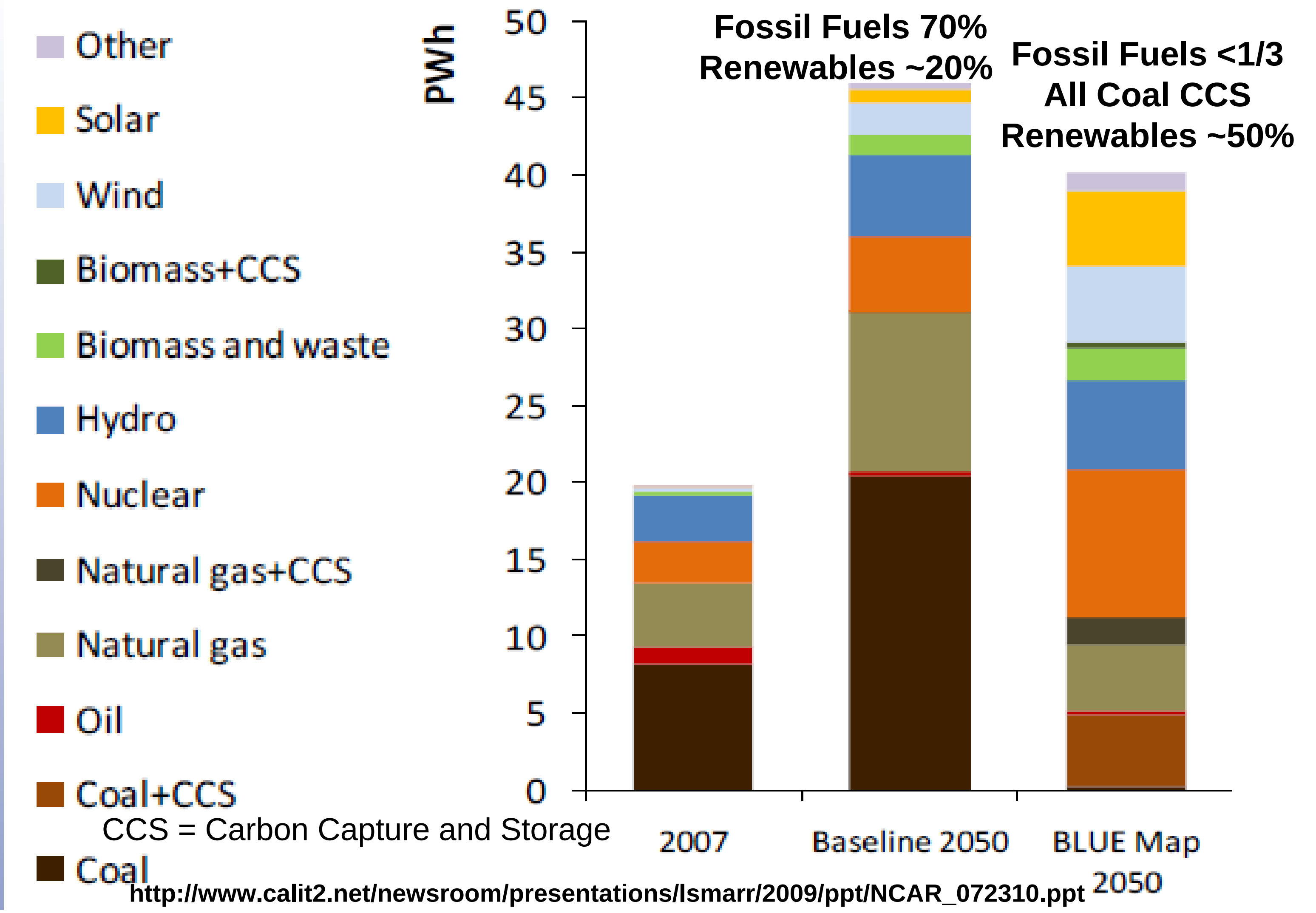
To stabilize atmospheric concentrations of greenhouse gases Sterman says that emissions must peak before 2020 and then fall at least 80% below recent levels by 2050

- Alternative Energy Must Scale Up Very Quickly
- Carbon Sequestration Must be Widely Used for Coal

"The Copenhagen Accord for limiting global warming: Criteria, constraints, and available avenues," PNAS, v. 107, 8055-62 (May 4, 2010) V. Ramanathan and Y. Xu, Scripps Institution of Oceanography, UCSD

http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/Shaffer_Class_MGT166_060110_final.ppt

IEA Blue Map Requires Massive Decarbonising of the Electricity Sector



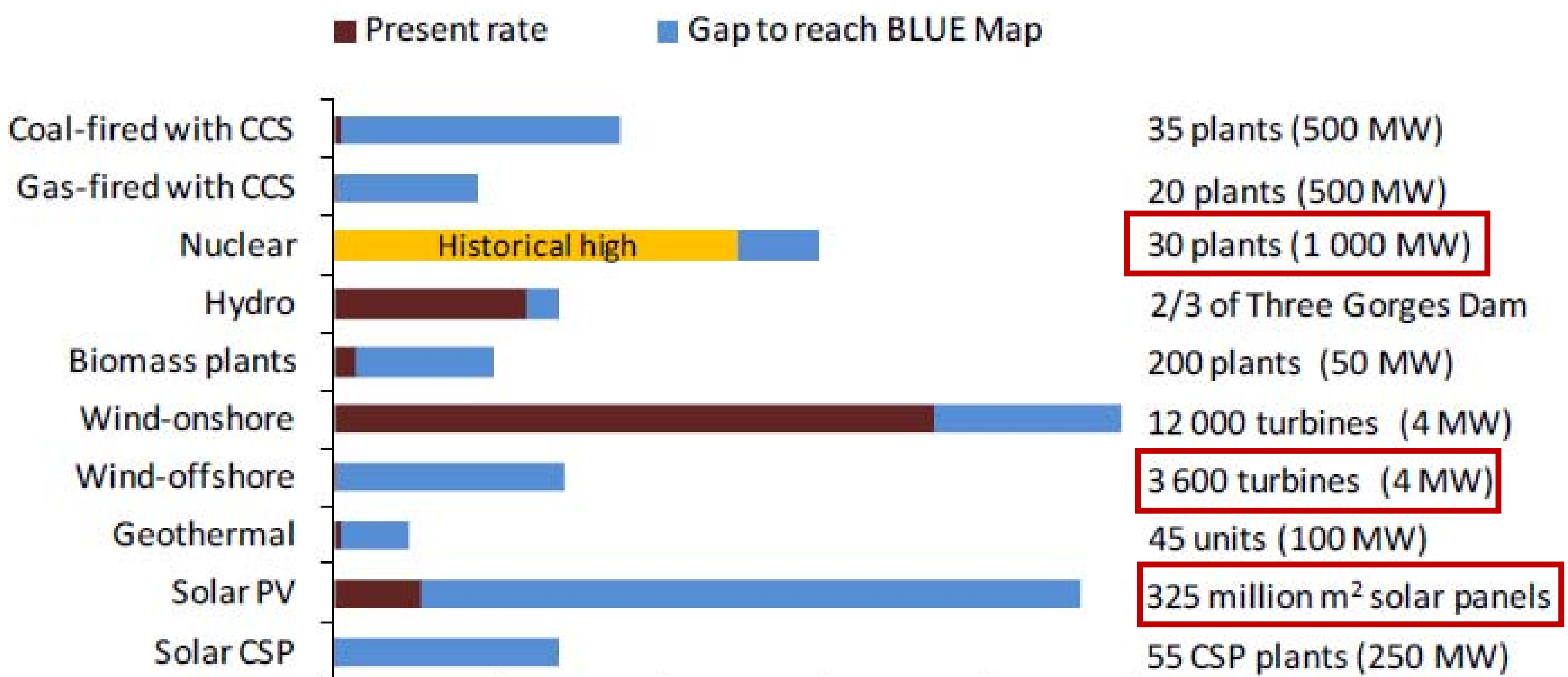
CCS = Carbon Capture and Storage

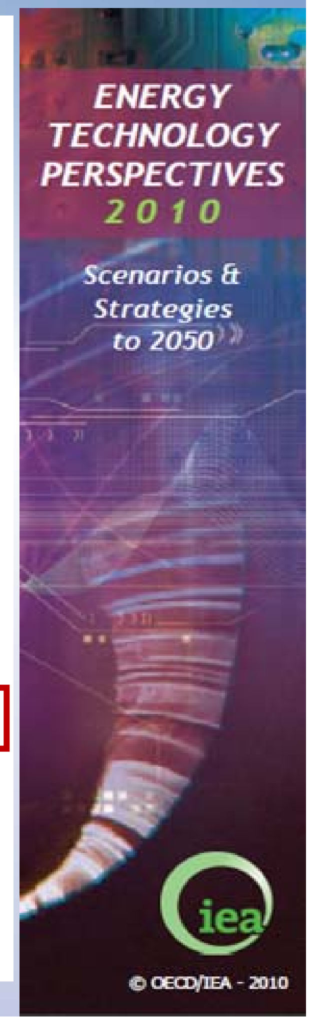
http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/NCAR_072310.ppt

ADVANCED COMPUTING RESEARCH LABORATORY

Average Annual Electricity Capacity Additions To 2050 Needed to Achieve the BLUE Map Scenario

Well Underway with Nuclear, On-Shore Wind, and Hydro,
Massive Increases Needed in All Other Modes





Solar PV =
Solar Photo-Voltaics

Solar CSP =
Concentrated Solar Power

Google Buys Wind Energy ….

May 4, 2010. Google has purchased a tax equity stake in two South Dakota wind farms, capable of generating a combined 169.5 megawatts of electricity.

Google has announced its investment in the Ashtabula 2 and Wilton Wind 2 wind farms, both with enough output to power 55,000 homes. The farms are located in North Dakota, in one of the United States' best fields for wind power generation. While these farms may not see a direct connection to Google's data centers to start, the 169.5 megawatts of power generated by these farms will be used to offset tax costs incurred by Google in other areas.

July 20, 2010. "On July 30 we will begin purchasing the clean energy from 114 megawatts of wind generation at the NextEra Energy Resources Story County II facility in Iowa at a predetermined rate for 20 years… This power is enough to supply several data centers." Urs Hoelzle, Google's SVP Operations

But new renewable energy sources are unreliable …..

New challenges and opportunities in operations, unless effective storage solutions can be found

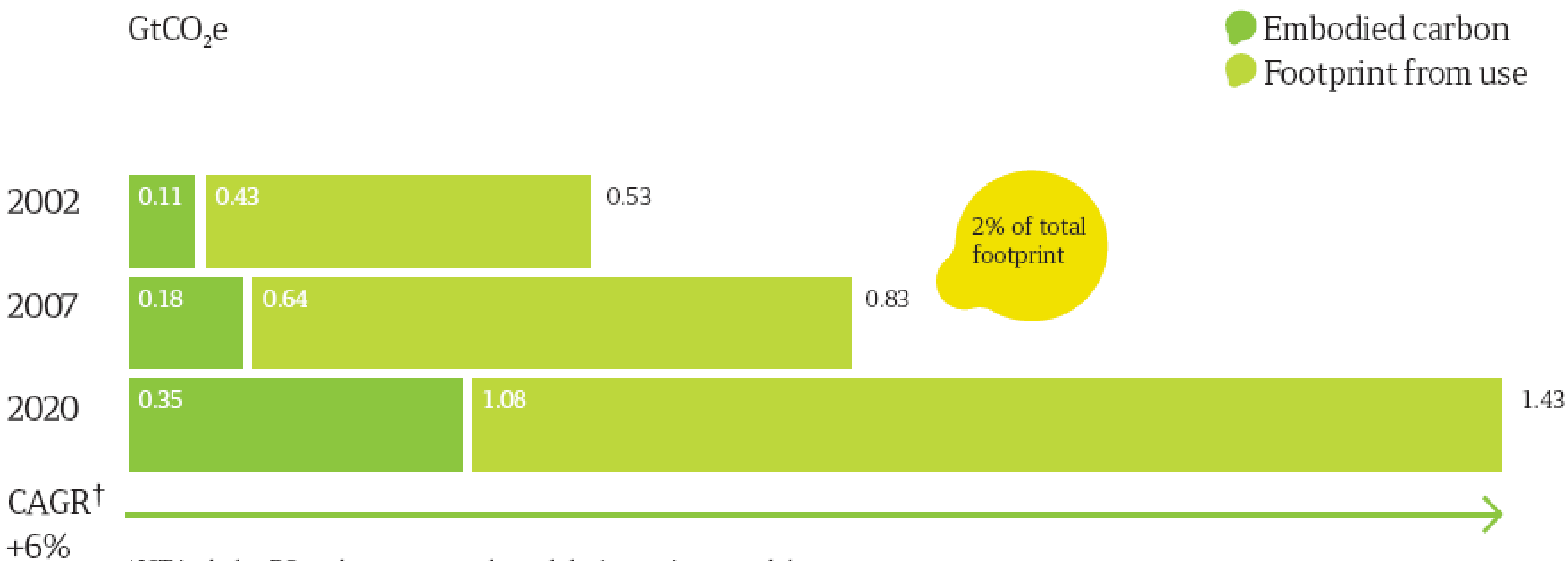
ADVANCED COMPUTING RESEARCH LABORATORY

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
- 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

***An Inefficient Tuth: http://www.globalactionplan.org.uk/event_detail.aspx?eid=2696e0e0-28fe-4121-bd36-3670c02eda49**

The Projected ICT Carbon Emission



*ICT includes PCs, telecoms networks and devices, printers and data centres.
†Compounded annual growth rate.

the assumptions behind the growth in emissions expected in 2020:

- **takes into account likely efficient technology developments that affect the power consumption of products and services**
- **and their expected penetration in the market in 2020**

http://www.smart2020.org/_assets/files/02_Smart2020Report.pdf

ICT's incredible improvement in energy efficiency

	1978	2008	Energy-efficiency improvement
Automobiles	14.3 miles per gallon of gas	20.0 miles per gallon of gas	40 percent
Passenger Airliners	22.8 revenue passenger miles per gallon	50.4 revenue passenger miles per gallon	121 percent
Agriculture	0.63 units of output per unit of energy use	1.46 units of output per unit of energy use	132 percent
Steel Manufacturing	63 pounds of steel per MBtu	167 pounds of steel per MBtu	167 percent
Lighting	Incandescent light bulb—13 lumens per watt	Compact fluorescent bulb—57 lumens per watt	339 percent
Computer Systems	1,400 instructions per second per watt	40,000,000 instructions per second per watt	2,857,000 percent

The American Council for an Energy-Efficient Economy report for the Technology CEO Council, http://www.techceocouncil.org/images/stories/pdfs/TCCsmartgreen2-1.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

ICT can enable a big reduction in the rate of climate change

- American Council for an Energy-Efficient Economy (ACEEE) studied this issue and concluded:
  - "For every extra Kwh of electricity that has been demanded by ICT, the US economy increased its overall energy savings by a factor of about 10…" (2008)
- The Climate Group and the "Global e-Sustainability Initiative" published a report entitled, "*Smart 2020: Enabling the Low Carbon Economy in the Information Age*" (2008), concluding:
  - Smart 2020 concludes that ICT strategies could reduce **up to 15% percent** of global emissions in 2020 against a "business as usual" baseline
- US Addendum to *Smart 2020* report, prepared by Boston Consulting Group indicates that ICT strategies could reduce US carbon emissions by **up to 22 percent** by 2020 vs. business-as-usual
- **TAKE AWAY**: ICT strategies offer huge potential for addressing climate challenge

Source: Stephen Harper

1 09 Senate DESC Briefing

5

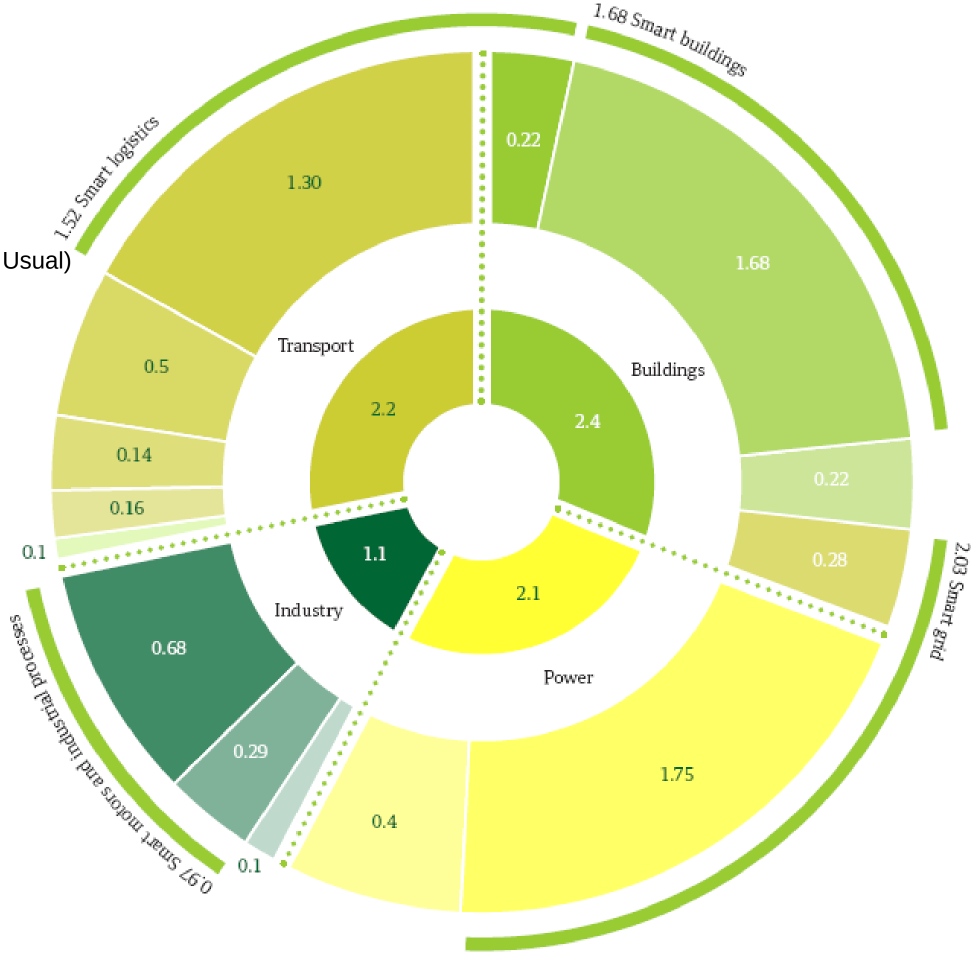
ase.org/uploaded_files/presentations/HARPER%201%2009%20DESC%20ASE.pdf

ICT potential CO_2 reduction in other sectors

$GtCO_2e$

7.8 $GtCO_2e$ of ICT-enabled abatements are possible out of the total BAU emissions in 2020 (51.9 $GtCO_2e$) (BAU = Business As Usual)

The SMART opportunities including dematerialisation were analysed in depth



Industry
- Smart motors
- Industrial process automation
- Dematerialisation* (reduce production of DVDs, paper)

Transport
- Smart logistics
- Private transport optimisation
- Dematerialisation (e-commerce, videoconferencing, teleworking)
- Efficient vehicles (plug-ins and smart cars)
- Traffic flow monitoring, planning and simulation

Buildings
- Smart logistics†
- Smart buildings
- Dematerialisation (teleworking)
- Smart grid‡

Power
- Smart grid
- Efficient generation of power, combined heat and power (CHP)

*Dematerialisation breaks down into all sectors except power. See detailed assumptions in Appendix 3.
†Reduces warehousing space needed through reduction in inventory. See Appendix 3.
‡Reduces energy used in the home through behaviour change. See Appendix 3.

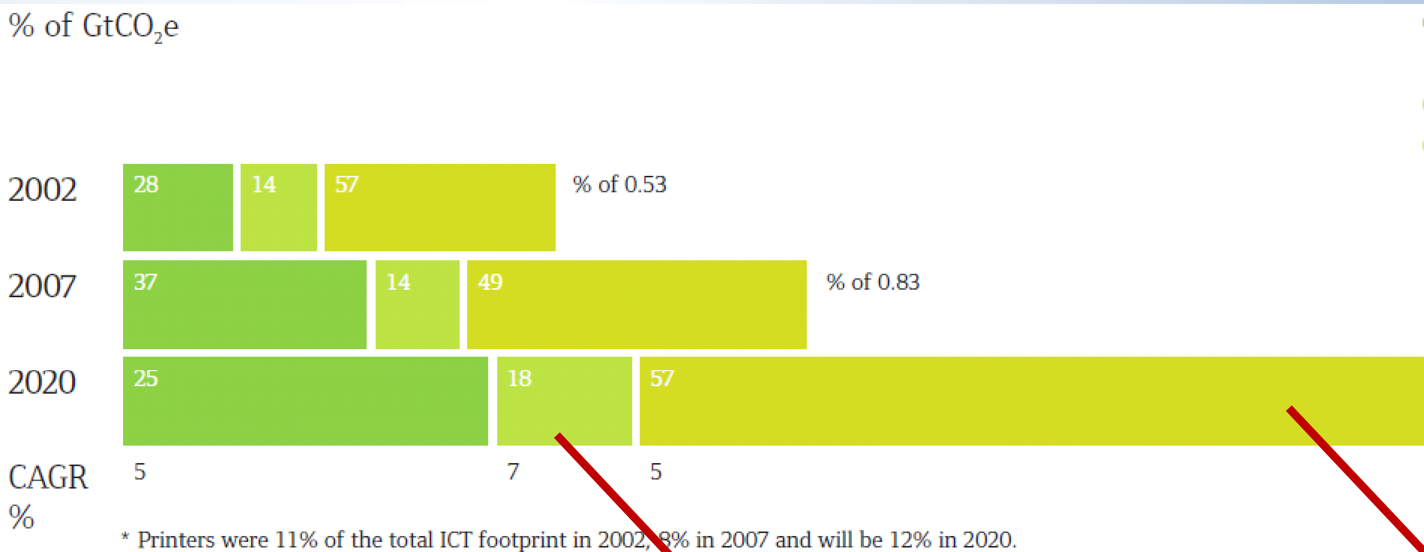
http://www.smart2020.org/_assets/files/02_Smart2020Report.pdf

PDC Summer School,
Aug 26 2010
Lennart Johnsson

ADVANCED COMPUTING RESEARCH LABORATORY

The Projected ICT Carbon Emission by Subsector

The Number of PCs (Desktops and Laptops) Globally is Expected to Increase from 592 Million in 2002 to More Than Four Billion in 2020



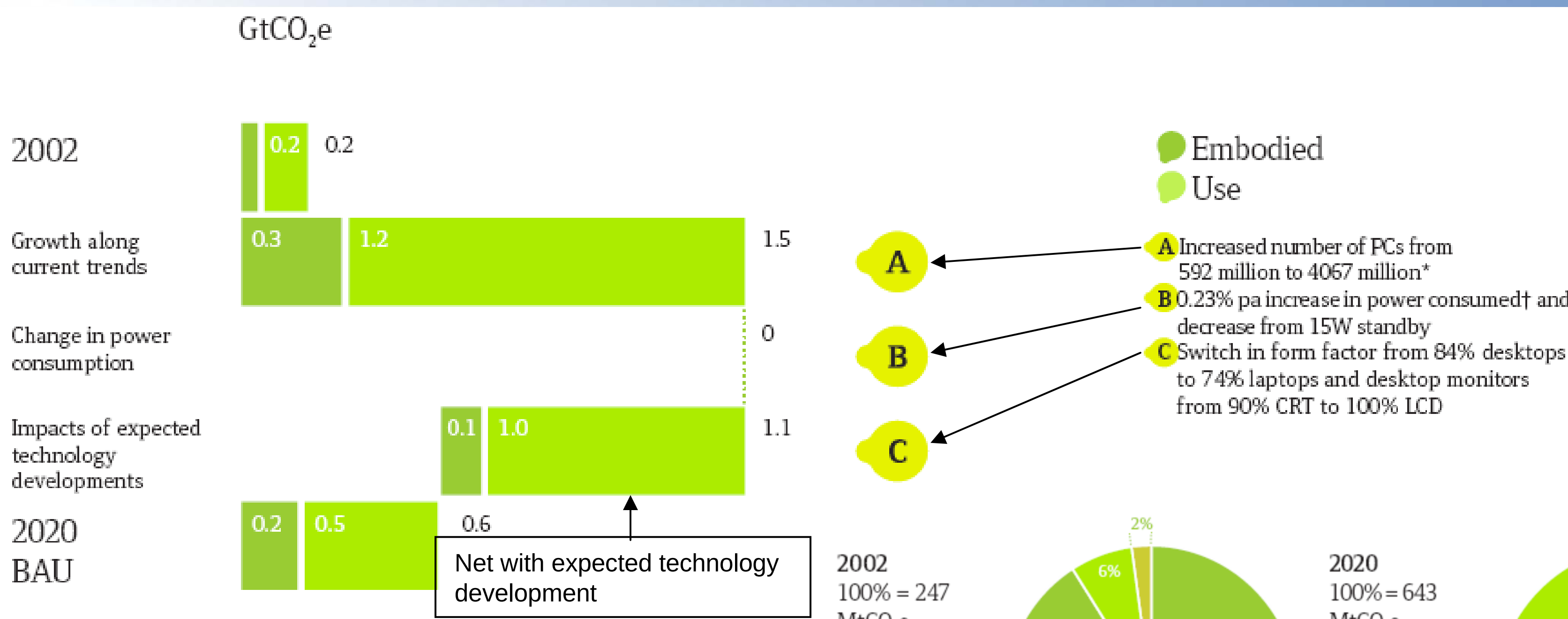
Data Centers Are Rapidly Improving

PCs Are Biggest Problem

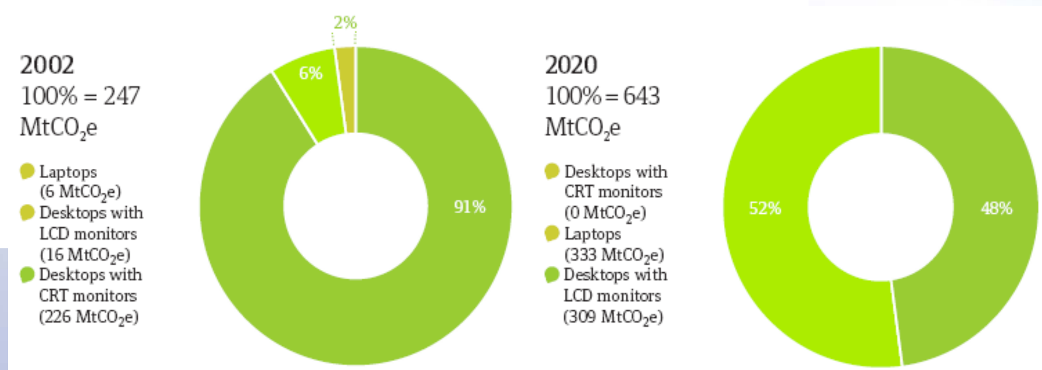
http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/SCI_Utah_043010.ppt

http://www.smart2020.org/_assets/files/02_Smart2020Report.pdf

Projected evolution of PC related CO_2 emissions



* Based on Gartner estimates until 2011 and trend extrapolation to 2020.
† Based on McManus, T. (2002), *Moore's Law and PC Power,* presentation to Tulane Engineering Forum.



Desktops with CRT monitors represented 44% of the total ICT footprint (91% of 49%).

Desktops with LCD monitors and laptops represented 4% of the total ICT footprint (8% of 49%).

Laptops will represent 22% of the total ICT footprint (52% of 42%).

Desktops with LCD monitors will represent 20% of the total ICT footprint (48% of 42%).

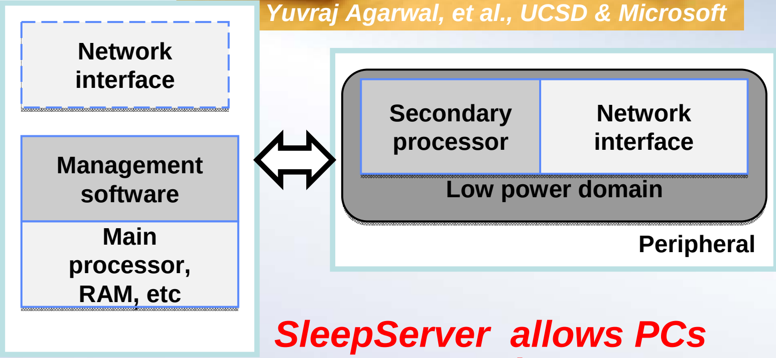
http://www.smart2020.org/_assets/files/02_Smart2020Report.pdf

PC Power Savings with *SleepServer:* A Networked Server-Based Energy Saving System

Dell OptiPlex 745 Desktop PC State	Power
Normal Idle State	**102.1W**
Lowest CPU Frequency	**97.4W**
Disable Multiple Cores	**93.1W**
“Base Power”	**93.1W**
Sleep state (ACPI State S3) Using SleepServers	2.3W

(ACPI = Advanced Configuration and Power Interface)

Yuvraj Agarwal, et al., UCSD & Microsoft



- Power drops from 102W to < 2.5W
- Assuming a 45 hour work week
  - 620kWh saved per year for each PC
- Additional application latency: 3s - 10s across applications
  - Not significant as a percentage of resulting session

SleepServer allows PCs to “Suspend to RAM” to maintain their network and application level presence

Source: Rajesh Gupta, UCSD CSE, Calit2

UCSanDiego | Energy Dashboard

http://energy.ucsd.edu/device/meterdisplay.php?meterID=3091420330&mode=pastyear

energy.ucsd.edu

kW-Hours:488.77 kW-H Averge Watts:55.80 W
Energy costs:$63.54
Estimated Energy Savings with Sleep Server: 32.62%
Estimated Cost Savings with Sleep Server: $28.4

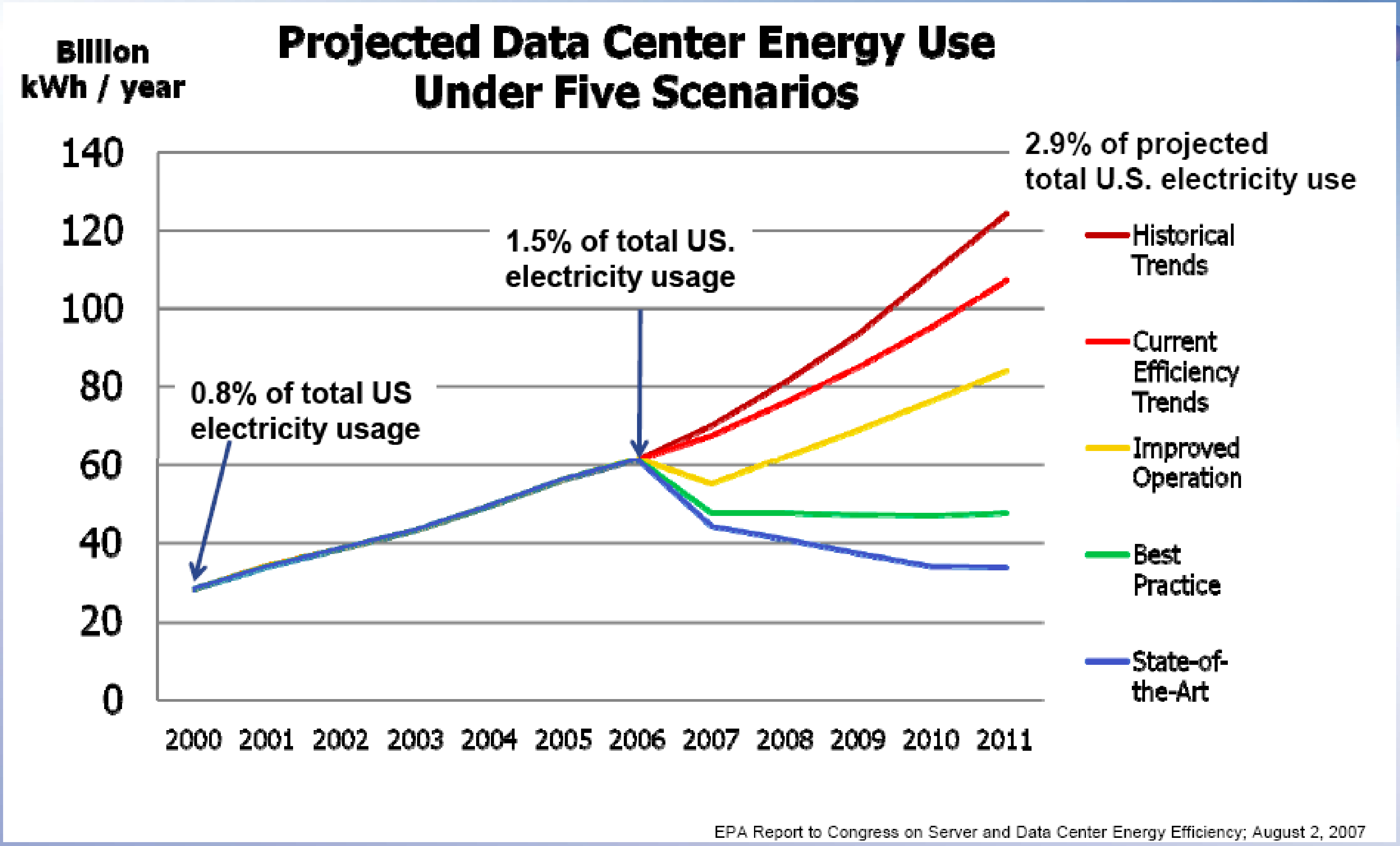
Past year Power consumption for device #3091420330

From:	Apr, 12, 2009 12:28:50 PM	**Resolution:**	Every six hours (averaged)
To:	Apr, 12, 2010 12:28:50 PM	**Timespan:**	1 year

http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/SCI_Utah_043010.ppt

Zoom in this graph (for better data resolution, generate a new graph with a smaller time duration): 10D 1M 3M 1Y YTD MAX

Projected US Data Center Energy Needs



http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

Cyber-infrastructure: the scary facts

- 50% of today's Data Centers have insufficient power and cooling;*
- By 2010, half of all Data Centers will have to relocate or outsource applications to another facility.*
- During the next 5 years, 90% of all companies will experience some kind of power disruption. In that same period one in four companies will experience a significant business disruption*
- Cyber-infrastructure is often the 2nd largest consumer of electricity after basic heat and power on university campuses
- Demand for cyber-infrastructure is growing dramatically

In the US, the growth in power consumption between 2008 and 2010 is equivalent to 10 new power plants!

*Sourcce: http://www.nanog.org/mtg-0802/levy.html

Revolutionizing Data Center Efficiency – Key Analysis, McKinsey & Co, April 2008

ADVANCED COMPUTING RESEARCH LABORATORY

Location .....

$/kWh	Location	Possible reason
0.036	Idaho	Local hydro power
0.10	California	Electricity transmitted long distance; Limited transmission capability; No coal fired power plants allowed in California
0.18	Hawaii	No local energy source. All fuel must be imported.

From http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

ADVANCED COMPUTING RESEARCH LABORATORY

Greening the Data Center

RACKFORCE DYNAMIC DATACENTER SERVICES

[redacted]

- Locate datacenters near green power sources
- Hydropower is the best source at this time
- Combine solar and wind with hydropower for best result
- Green Result - CO_2 reduced by:

10 to 50X

http://www.rackforce.com/documents/summit09-green_to_the_core_ii.pdf

11 © 2009 RackForce Networks Inc.

"Unpredictable"

GreenStar NETWORK

"Most forms of renewable energy are not reliable – at any given location. But Canada's Green Star Network aims to demonstrate that by allowing the computations to follow the renewable energy across a large, fast network, the footprint of high-throughput computing can be drastically reduced."

International Science Grid This Week, April 4, 2010

ADVANCED COMPUTING RESEARCH LABORATORY

HPC at the next level: Exa-scale

LBNL IJHPCA Study for ~1/5 Exaflop for Climate Science in 2008
- Extrapolation of Blue Gene and AMD design trends
- Estimate: **20 MW** for BG and **179 MW** for AMD

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

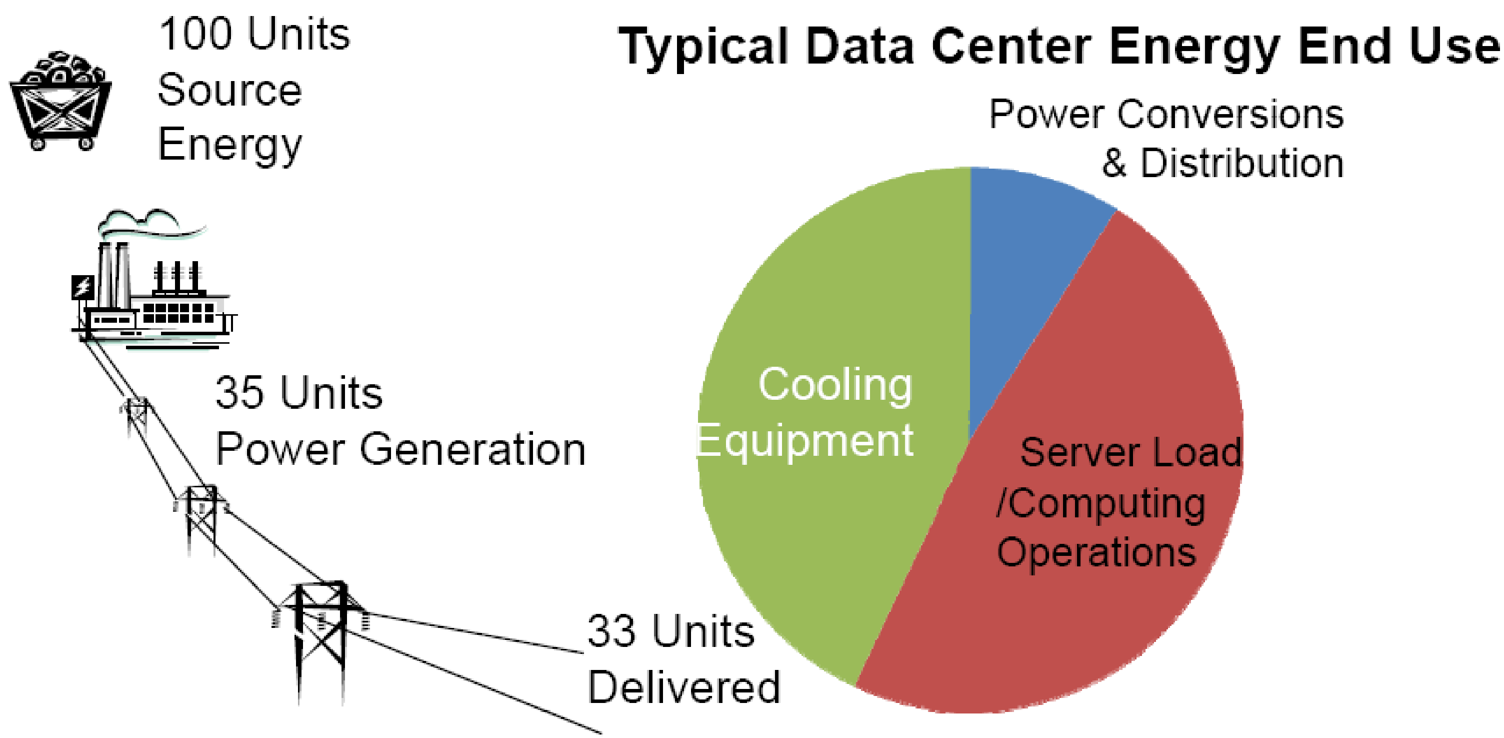
The current approach is not sustainable!

More holistic approach is needed!

Total Data Center Energy Efficiency

Data Center Energy Efficiency = 15% (or less)

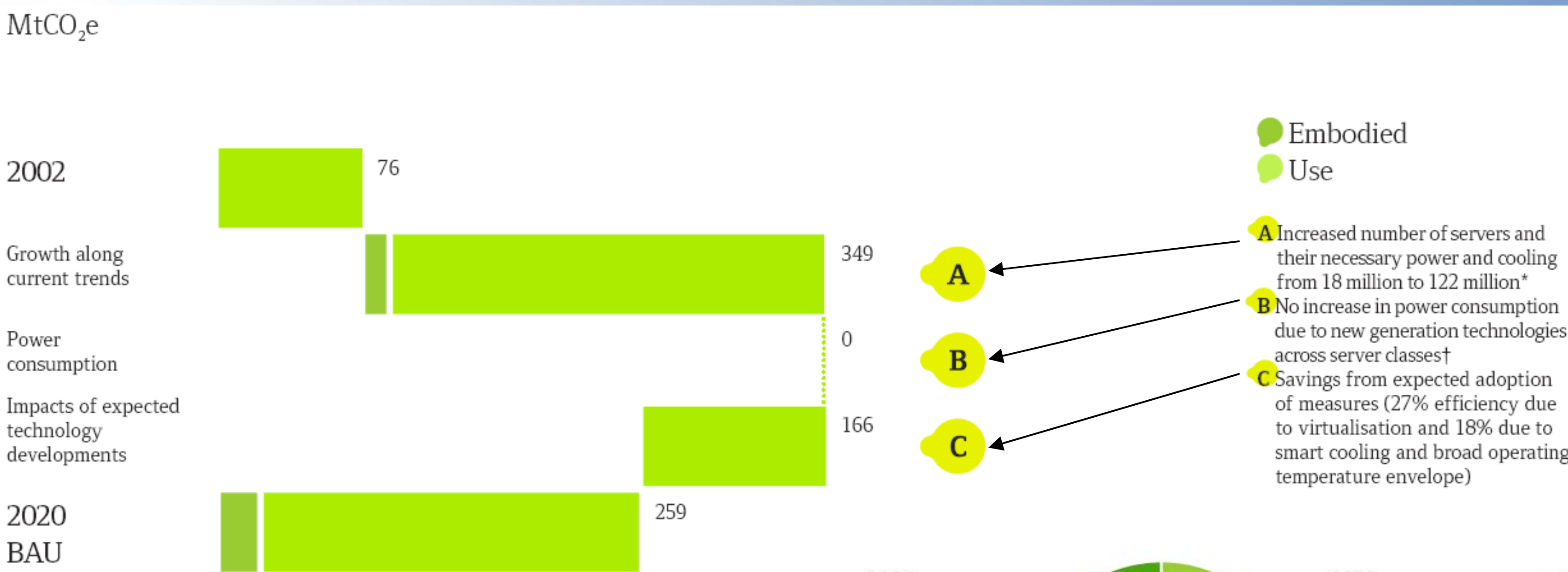
(Energy Efficiency = Useful computation / Total Source Energy)



Source: Paul Scheihing, U.S. Department of Energy, Energy Efficiency and Renewable Energy

http://www1.eere.energy.gov/industry/saveenergynow/pdfs/doe_data_centers_presentation.pdf

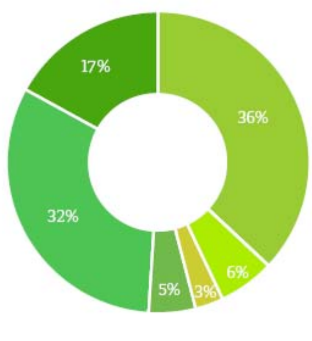
Data Center CO_2 emission projections



*Based on IDC estimates until 2011 and trend extrapolation to 2020, excluding virtualisation.
†Power consumption per server kept constant over time.

2002
100% = 76 $MtCO_2e$

- Volume servers (27 $MtCO_2e$)
- Cooling systems (24 $MtCO_2e$)
- Power systems (13 $MtCO_2e$)
- Mid-range servers (5 $MtCO_2e$)
- Storage systems (4 $MtCO_2e$)
- High-end servers (2 $MtCO_2e$)

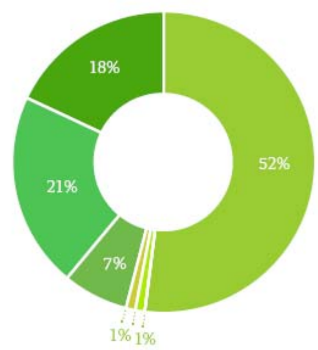


Volume servers represented 5% of the total ICT footprint (36% of 14%).

Data centre cooling systems represented 4% of the total ICT footprint (32% of 14%).

2020
100% = 259 $MtCO_2e$

- Volume servers (136 $MtCO_2e$)
- Cooling systems (70 MtCO2e)
- Power systems (62 $MtCO_2e$)
- Storage systems (18 $MtCO_2e$)
- High-end servers (5 $MtCO_2e$)
- Mid-range servers (2 $MtCO_2e$)



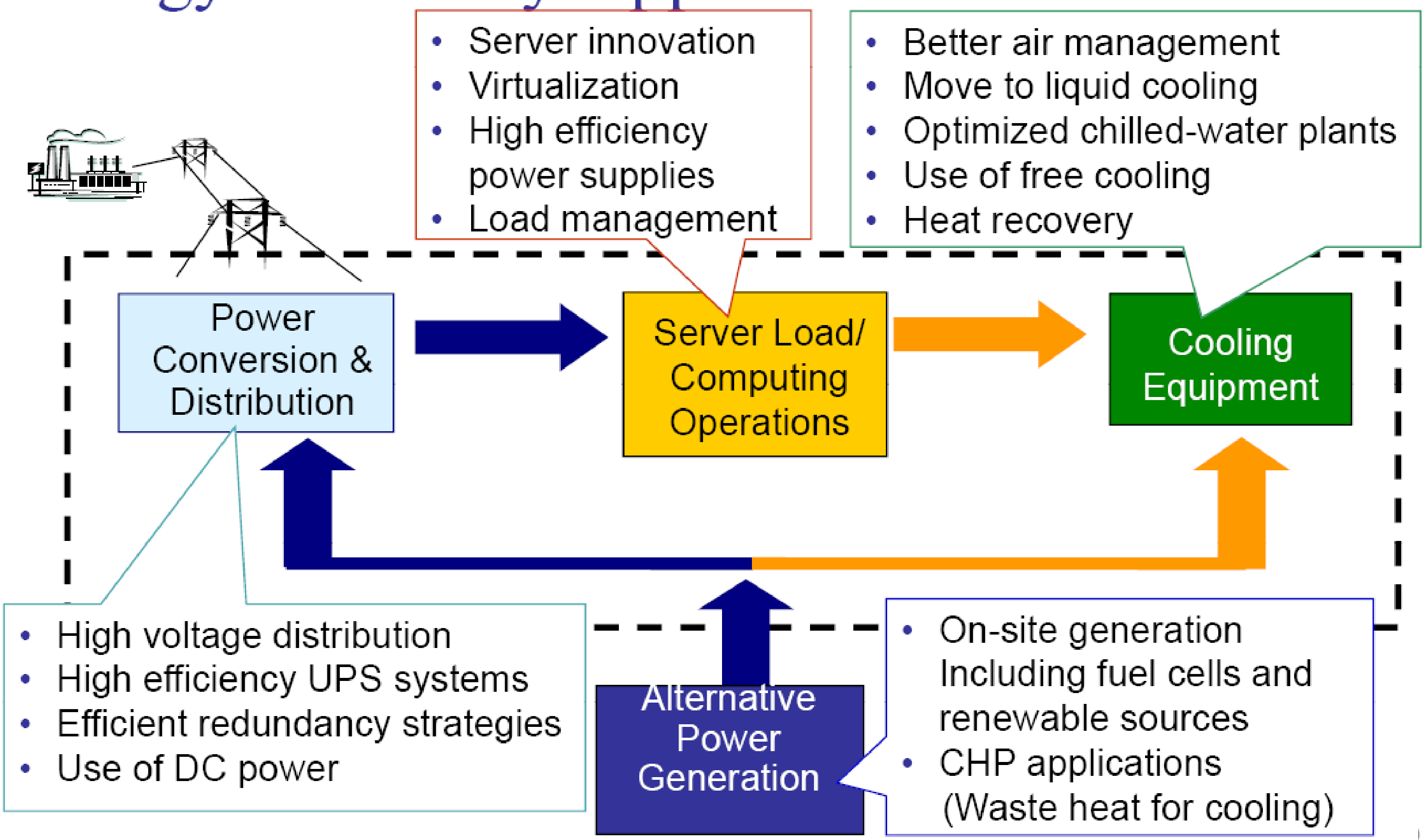
Volume servers will represent 9% of the total ICT footprint (52% of 18%).

Data centre cooling systems will represent 4% of the total ICT footprint (21% of 18%).

http://www.smart2020.org/_assets/files/02_Smart2020Report.pdf

Data Center Energy Efficiency

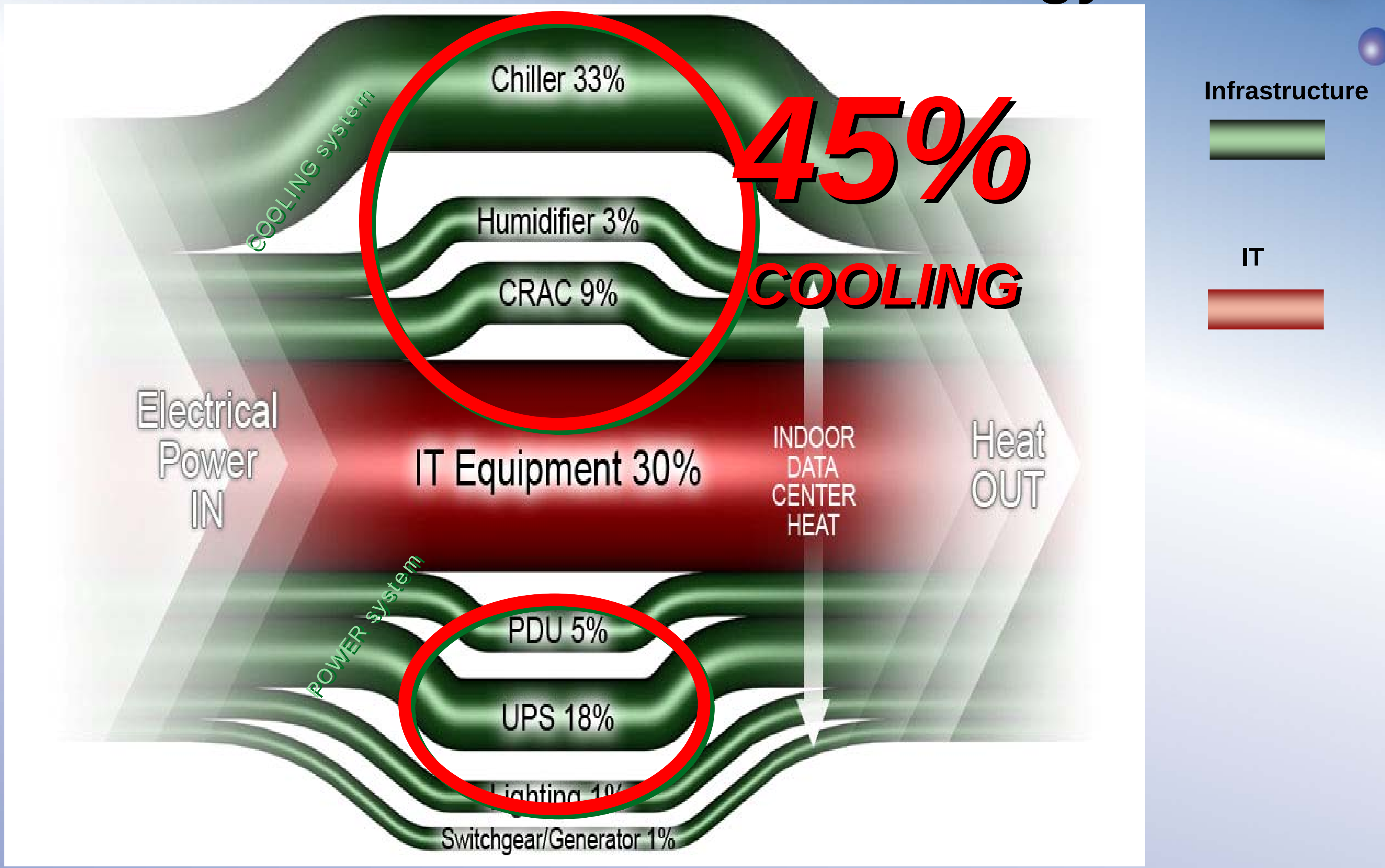
Energy Efficiency Opportunities



Source: Paul Scheihing, U.S. Department of Energy, Energy Efficiency and Renewable Energy
http://www1.eere.energy.gov/industry/saveenergynow/pdfs/doe_data_centers_presentation.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

Traditional Data Center Energy Use



http://www.slideshare.net/fabioce/apc-il-moderno-data-center-4453903

ADVANCED COMPUTING RESEARCH LABORATORY

Data Center Efficiency Measures

PUE/DCiE: The Green Grid Metrics For Facilities

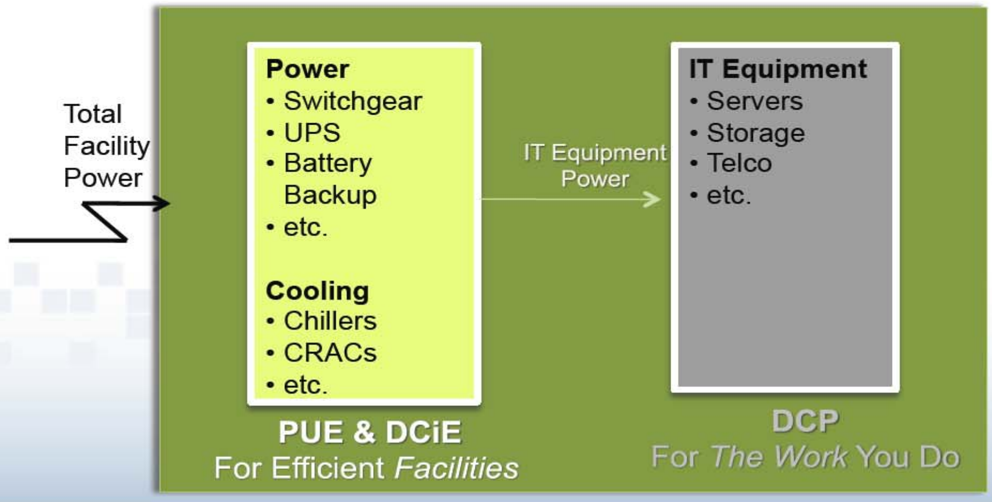
the green grid SM
get connected to efficient IT

$$\textbf{PUE} = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

How efficient is my facility in delivering power to IT equipment?

$$\textbf{DCiE} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}} \times 100\%$$

What % of facility power is delivered to my IT equipment?



6

PUE = Power Usage Efficiency

DCiE = Data Center infrastructure Efficiency

Slide Courtesy Michael K Patterson, Intel, 1st European HPC Infrastructure Workshop, September 2009

ADVANCED COMPUTING RESEARCH LABORATORY

Data Center Efficiency Measures

New PUE/DCiE Guidelines

the green grid™ SM
get connected to efficient IT

$$\text{PUE} = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

How efficient is my facility in delivering power to IT equipment?

$$\text{DCiE} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}} \times 100\%$$

What % of facility power is delivered to my IT equipment?

- **What were the environmental conditions?**
- **When were the measurements taken?**
- **How often were measurements taken?**

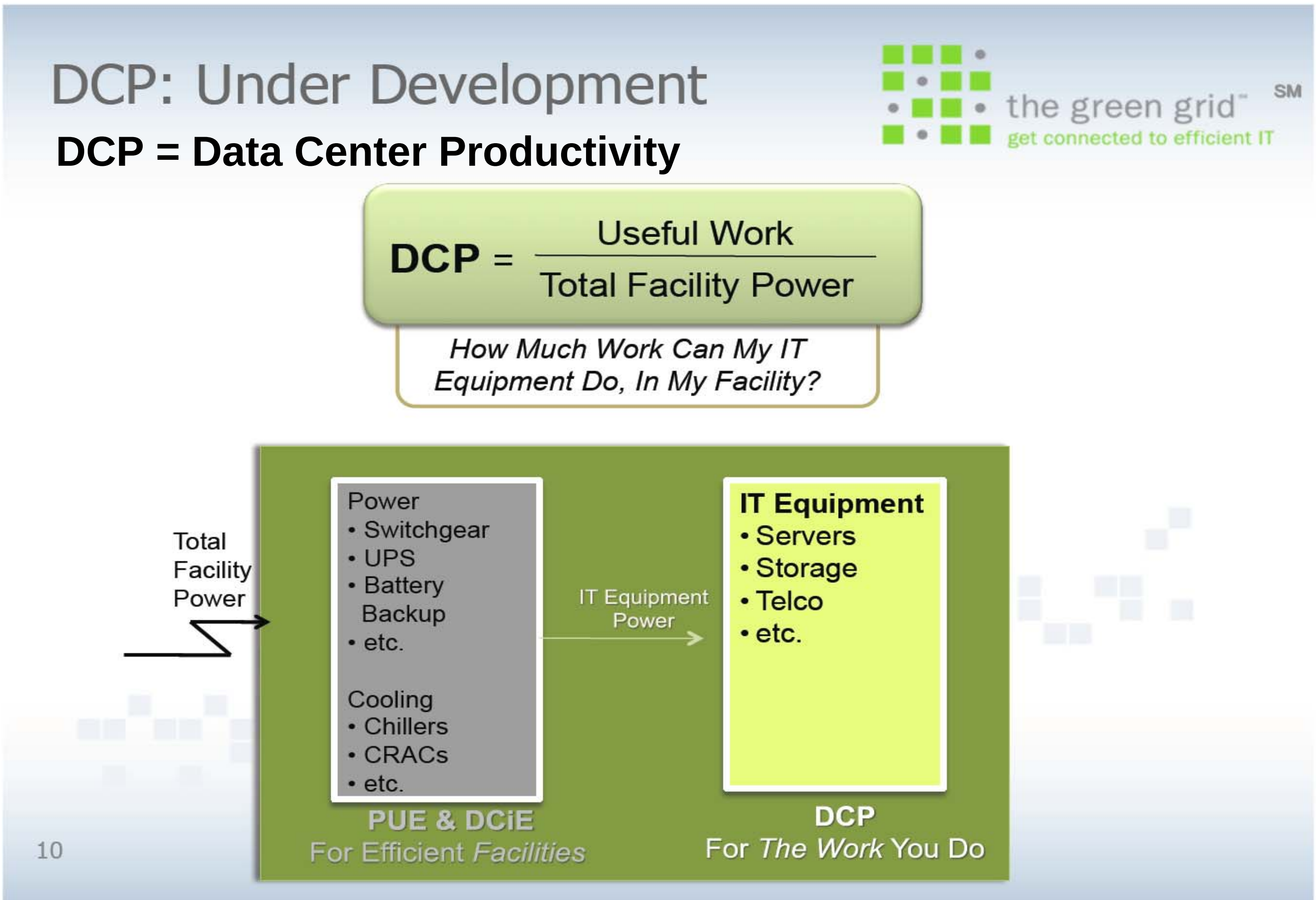
How to compare results?

7

Slide Courtesy Michael K Patterson, Intel, 1st European HPC Infrastructure Workshop, September 2009

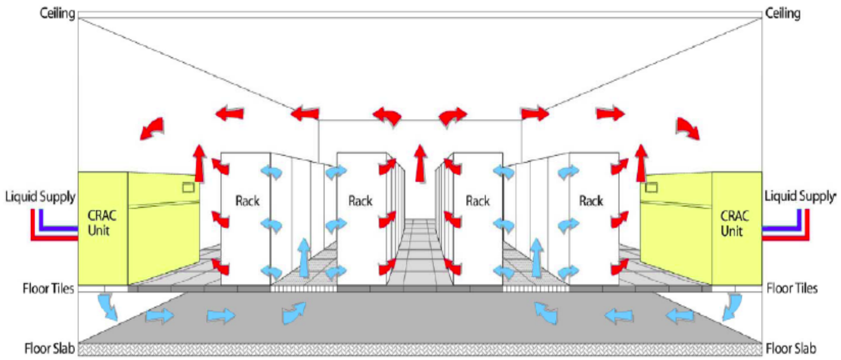
ADVANCED COMPUTING RESEARCH LABORATORY

Data Center Efficiency Measures



Slide Courtesy Michael K Patterson, Intel, 1st European HPC Infrastructure Workshop, September 2009

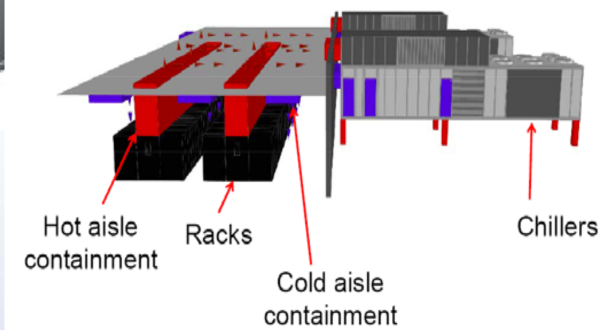
Data Center Air Management



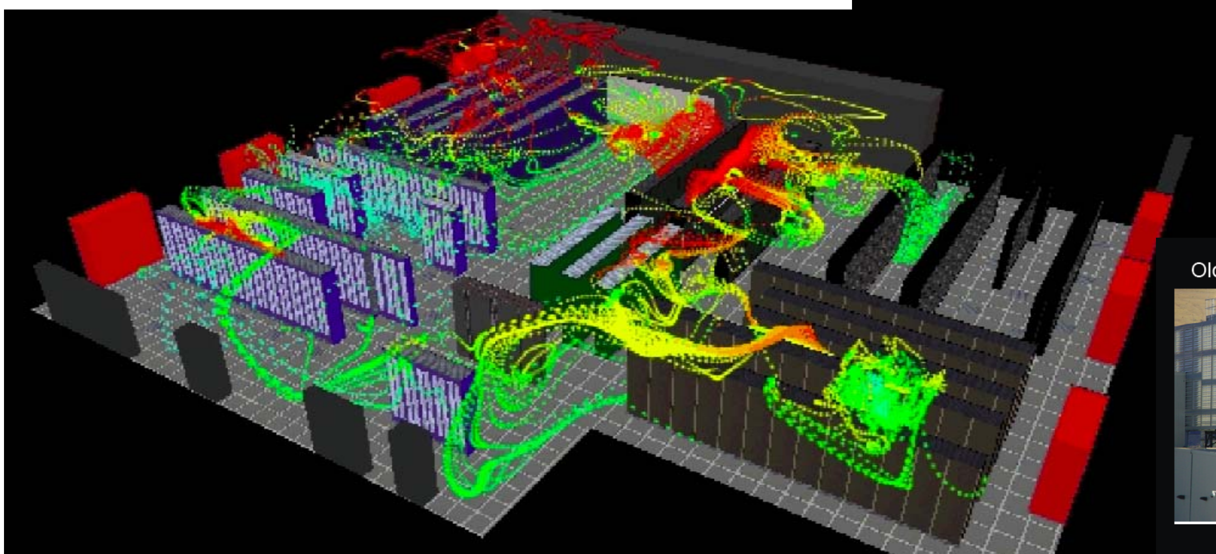
Font: Luiz Andre Barroso, Urs Hoelzle, "The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines", 2009.
(Image courtesy of DLB associates , ref [23] of the book)

Switch Communications data center

See video at…http://www.switchnap.com/pages/video.php



Source: Switch Communications, www.switchnap.com



http://www.infn.it/CCR/riunioni/presentazioni/pres_ott_05/michelotto_summary_workshop.pdf

Old "Always ON" Chiller Design

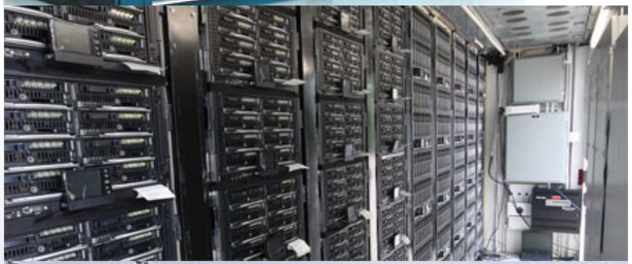
Free Cooling Chiller Design

Efficiency Improvement – 2x

The Containerized Data Center

Microsoft 500,000sqft M$550 data center in Northlake, IL

The 40-foot shipping containers packed with servers can deliver a Power Usage Effectiveness (PUE) energy efficiency rating of 1.22. The 40-foot CBlox containers can house as many as 2,500 servers (Max 22 racks @ 27kW/rack)



The company says it will pack between 150 and 220 containers in the first floor of the Chicago site, meaning the massive data center could house between 375,000 and 550,000 servers in the container farm.

Illustration Courtesy of Microsoft

http://www.datacenterknowledge.com/archives/2008/10/20/microsoft-pue-of-122-for-data-center-containers/

Google ….

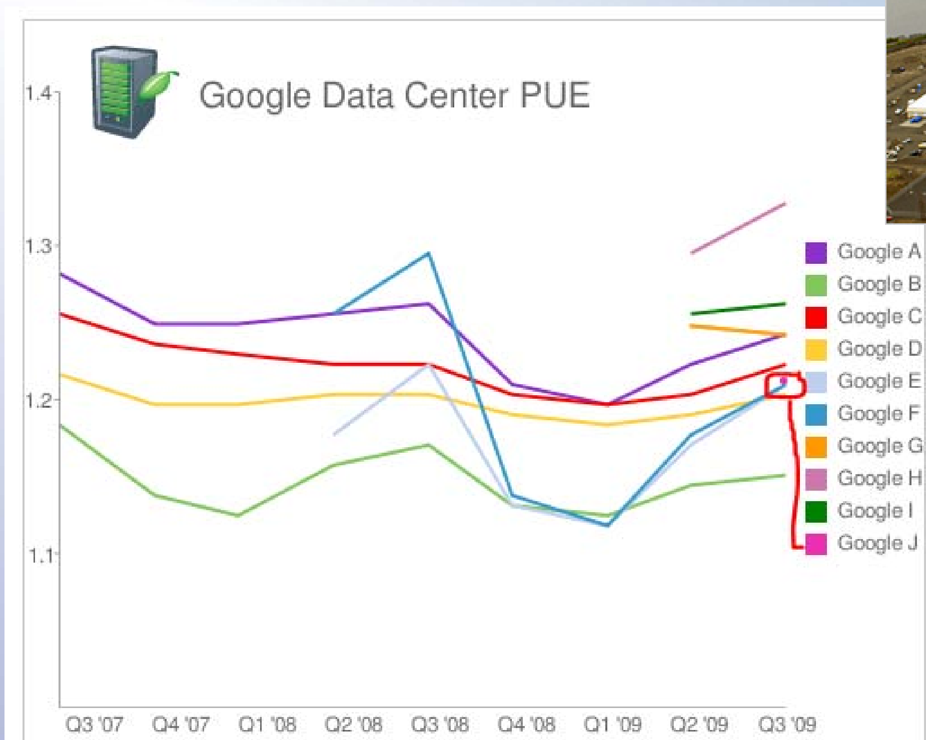


Figure 1: PUE data for ten large-scale Google data centers

$$PUE = \frac{E_{US1} + E_{US2} + E_{TX} + E_{HV}}{E_{US2} + E_{Net1} - E_{CRAC} - E_{UPS} - E_{LV}}$$

- **EUS1** Energy consumption for type 1 unit substations feeding the cooling plant, lighting, and some network equipment
- **EUS2** Energy consumption for type 2 unit substations feeding servers, network, storage, and CRACs
- **ETX** Medium and high voltage transformer losses
- **EHV** High voltage cable losses
- **ELV** Low voltage cable losses
- **ECRAC** CRAC energy consumption
- **EUPS** Energy loss at UPSes which feed servers, network, and storage equipment
- **ENet1** Network room energy fed from type 1 unit substitution

http://www.greenm3.com/2009/10/google-releases-q3-2009-pue-numbers.html

Google ….

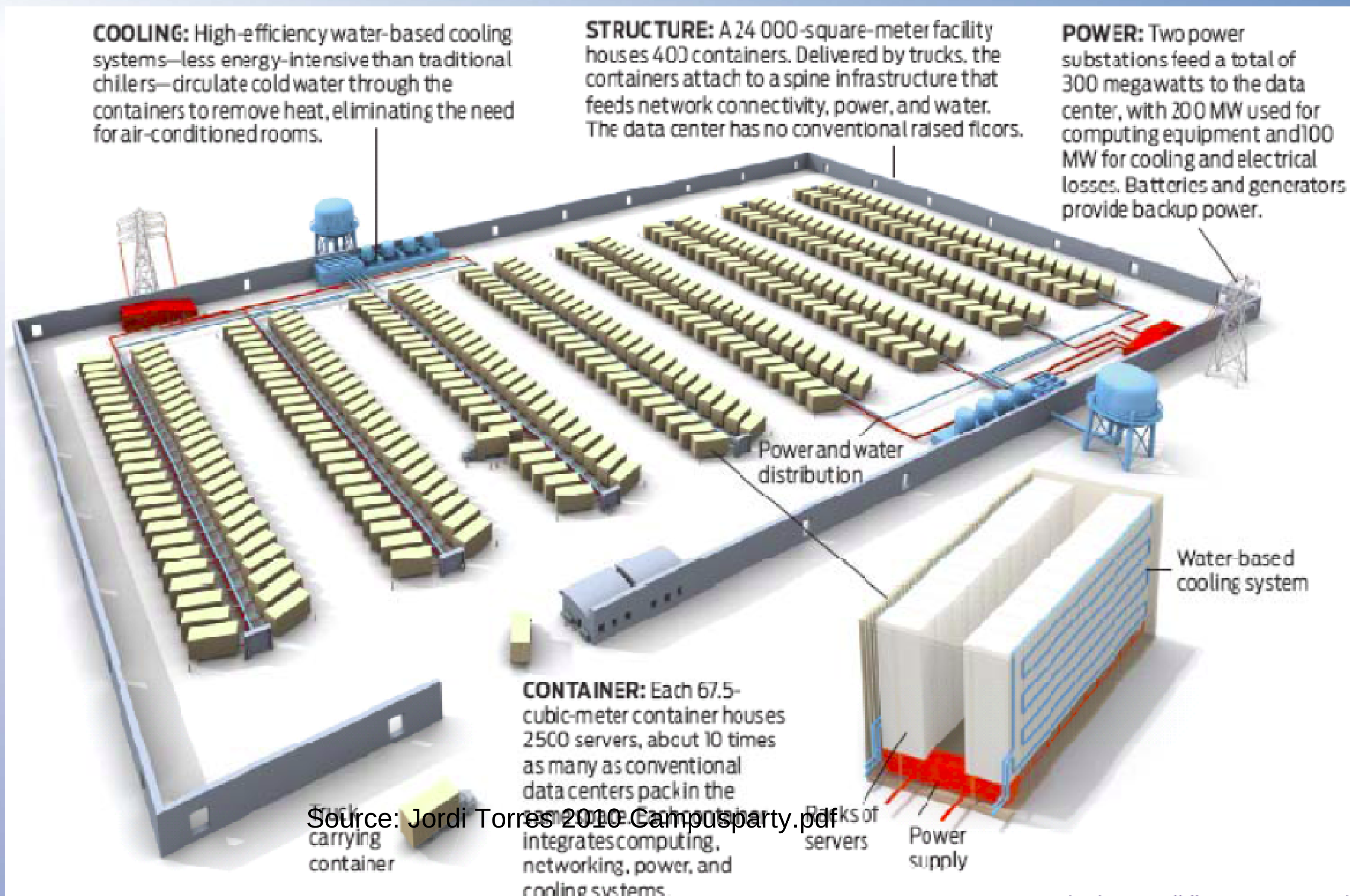
"Google has taken the "free cooling" (the use of fresh air from outside the data center to support the cooling system) strategy to the next level. Rather than using chillers part-time, the company has eliminated them entirely in its data center near Saint-Ghislain, Belgium, which began operating in late 2008 and also features an on-site water purification facility that allows it to use water from a nearby industrial canal rather than a municipal water utility."

"So what happens if the weather gets hot? On those days, Google says it will turn off equipment as needed in Belgium and shift computing load to other data centers." … "Google's Vijay Gill hinted that the company has developed automated tools to manage data center heat loads and quickly redistribute workloads during thermal events"

PDC Summer School,
Aug 26 2010
Lennart Johnsson

ADVANCED COMPUTING RESEARCH LABORATORY

Next Generation Data Center

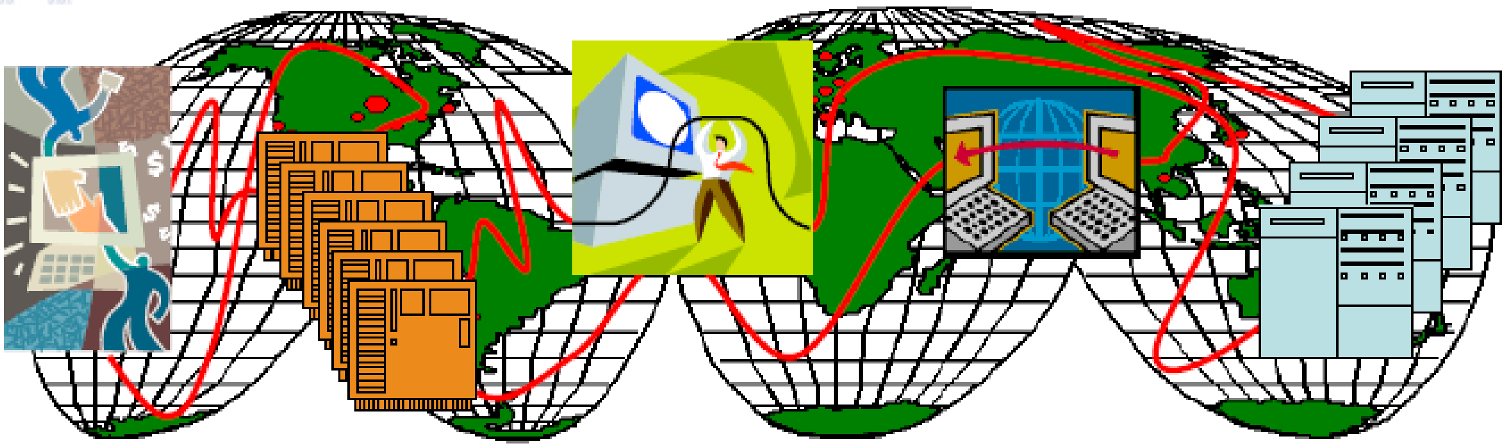


Source: Jordi Torres 2010 Campusparty.pdf

Source: Tech Titans Building Boom , Randy H. Katz.
IEEE Spectrum, February 2009
http://spectrum.ieee.org/green-tech/buildings/tech-titans-building-boom

- *"I think there is a world market for about five computers"* — Remark attributed to Thomas J. Watson (Chairman of the Board of International Business Machines) – 1943
- *"... In a sense, says Yahoo Research Chief Prabhakar Raghavan, there are only five computers on earth. He lists Google, Yahoo, Microsoft, IBM, and Amazon. Few others, he says, can turn electricity into computing power with comparable efficiency ..."*

From Google and the wisdom of clouds, by Steven Baker - BusinessWeek.com

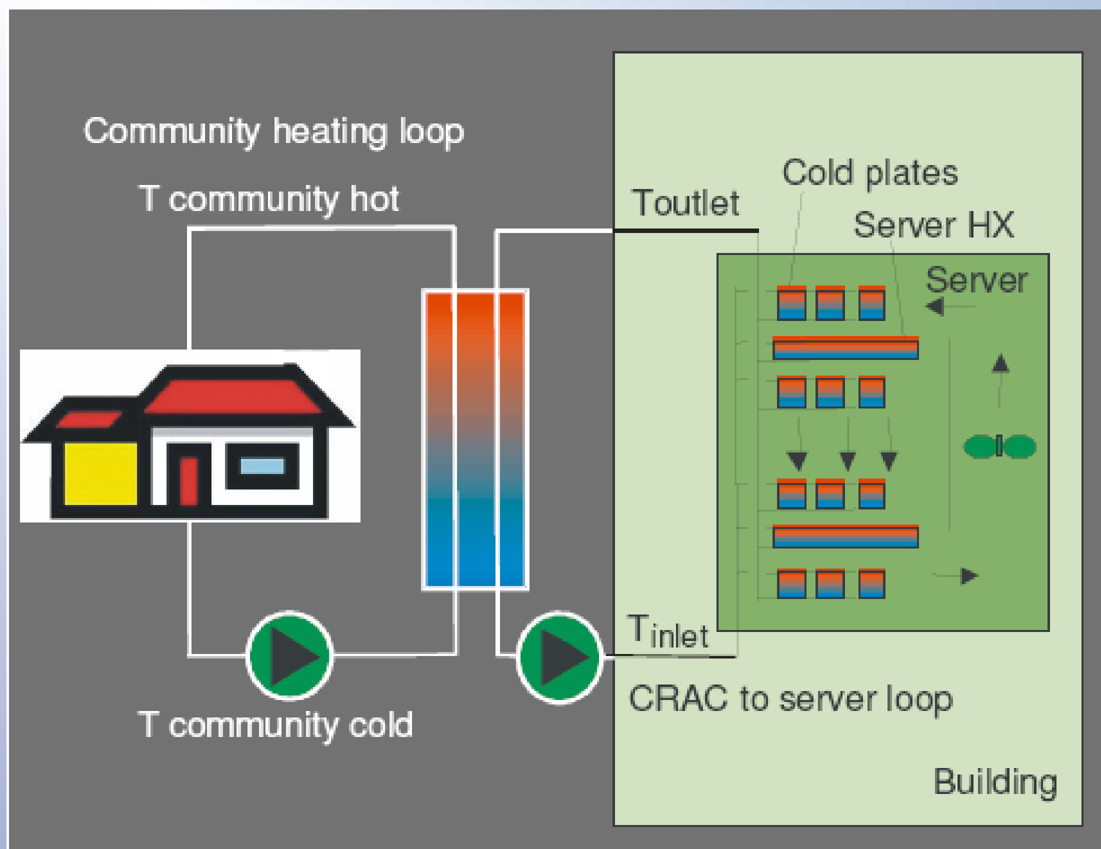


Source: Stefan Beco, CSFI May 30, 2008,
www.reservoir-fp7.eu/twiki/pub/Reservoir/PresentationsPage/RESERVOIR_CSFI08.ppt

ADVANCED COMPUTING RESEARCH LABORATORY

Zero Emission Data Center

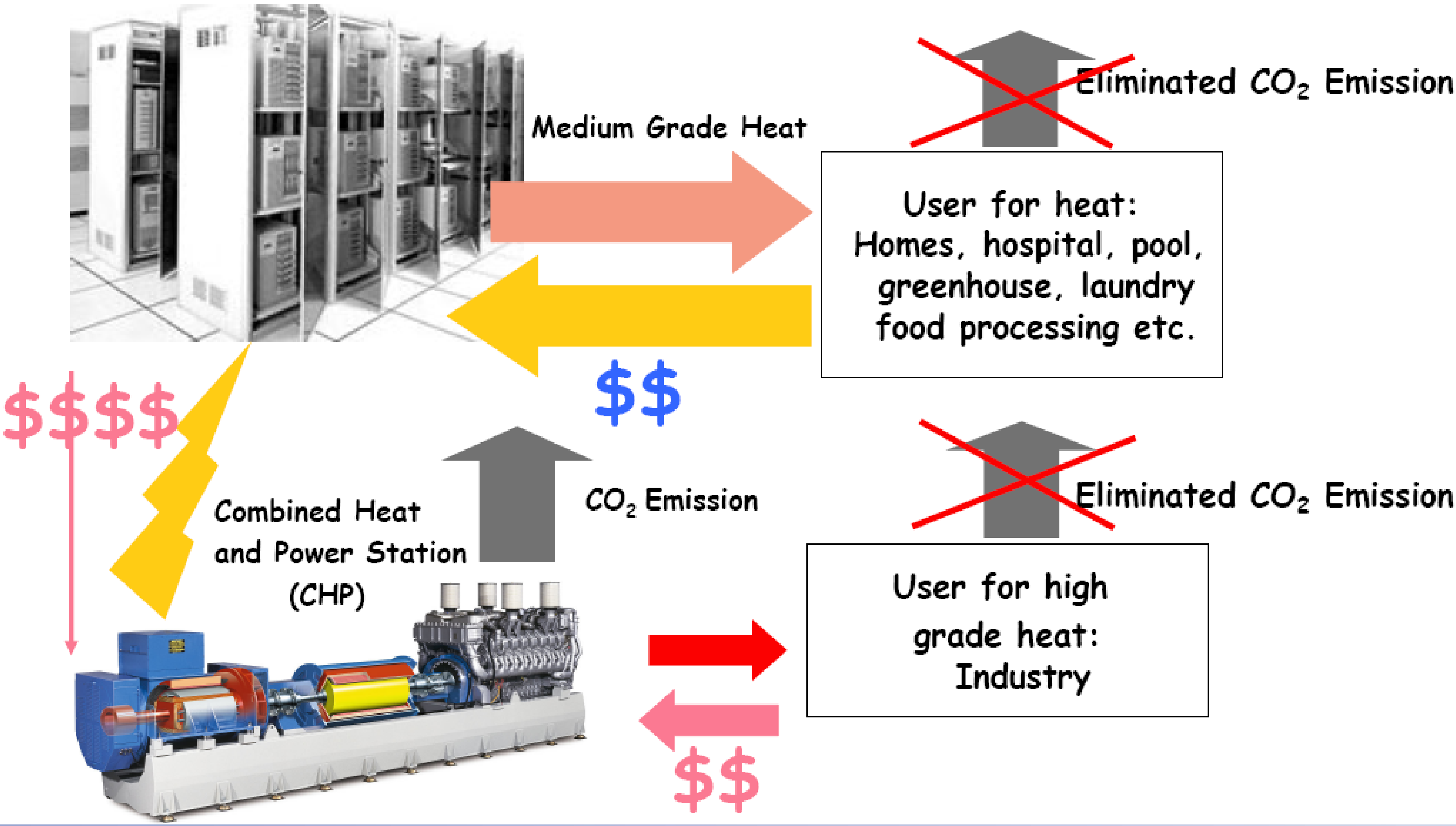
IBM Research – Cool with hot water!



Source: T. Brunschwiler, B. Smith, E. Ruetsche and B. Michel, IBM Research, Zurich

Zero Emission Data Center

DC4

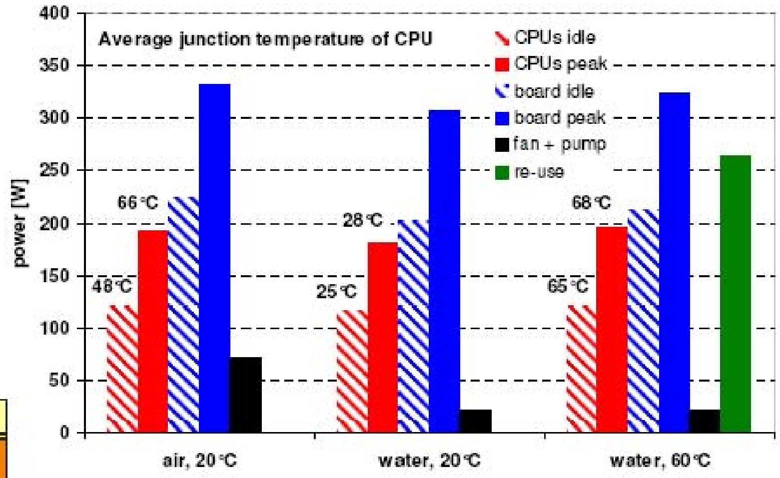


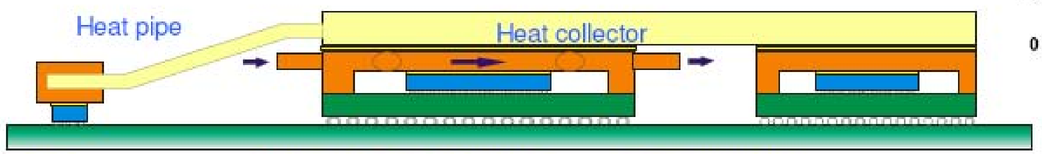
Source: T. Brunschwiler, B. Smith, E. Ruetsche and B. Michel, IBM Research, Zurich

First Prototype at IBM Rüschlikon

- Reduce cooling energy by tailored water cooling system
  - *Cooling the chip with "hot" water (up to 60 °C)*
  - *Free cooling: no energy-intensive chillers needed*
- Reuse waste heat for remote heating
  - *The prototype reuses 75% of the energy for remote heating*
  - *Obtain recyclable heat (60 °C) for remote heating.*
  - *Best in a cold climate with dense population*
- Prototype
  - *Similar Power of CPU and main board for air / liquid 60 °C cooled version*
  - *Large fan power reduction*
  - *Liquid pump much more efficient and can vary flow at the rack level*

Direct attached / integrated micro-channel cold plate with one interface





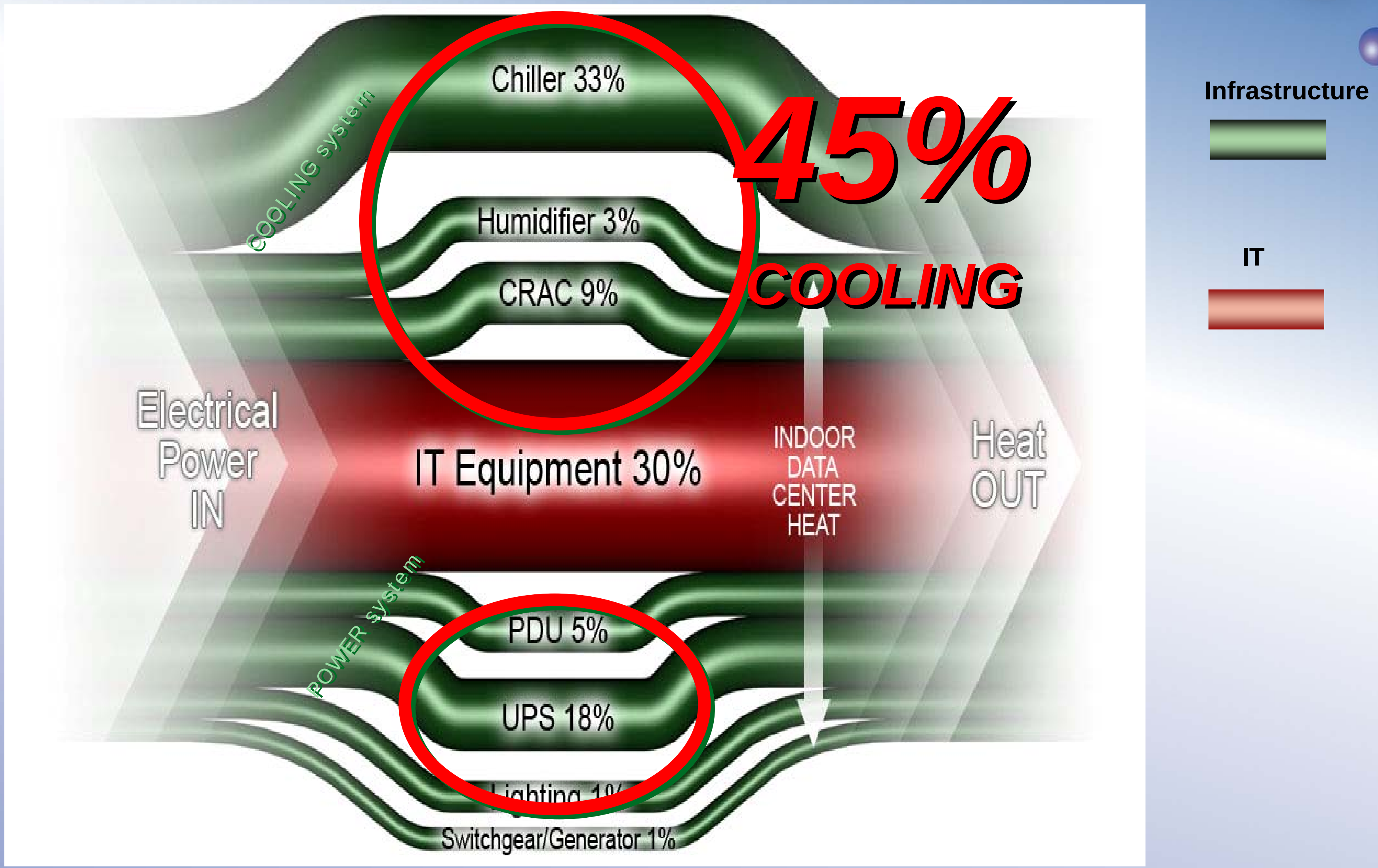
Experimental validation:
Inlet temperatures up to 60 °C

Source: T. Brunschwiler, B. Smith, E. Ruetsche and B. Michel, IBM Research, Zurich

PDC is moving towards increased energy reuse, but not quite as far as the IBM zero-emission data center vision

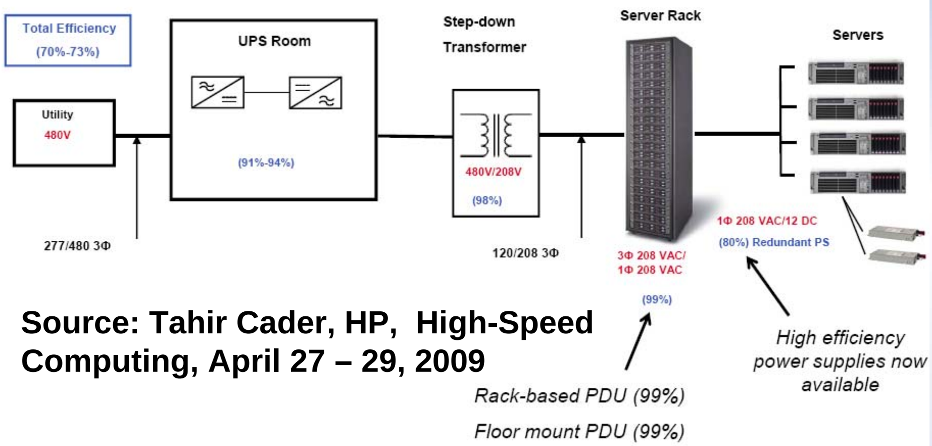
ADVANCED COMPUTING RESEARCH LABORATORY

Recall - Traditional Data Center Energy Use



http://www.slideshare.net/fabioce/apc-il-moderno-data-center-4453903

Data Center Power Efficiencies



Source: Tahir Cader, HP, High-Speed Computing, April 27 – 29, 2009

	5 yrs. ago	2010	2015
PUE	2, 3, Higher	1.1 Great	?
UPS (Part of PUE)	94%	98%+	?
PS	75%	94%+	?
Fan Power	60+ W	2-10 W (< 1%)	?

Source: Richard Kaufmann, HP, SC09

Top500 dominated by blades

www.hp.com

www.supermicro.com

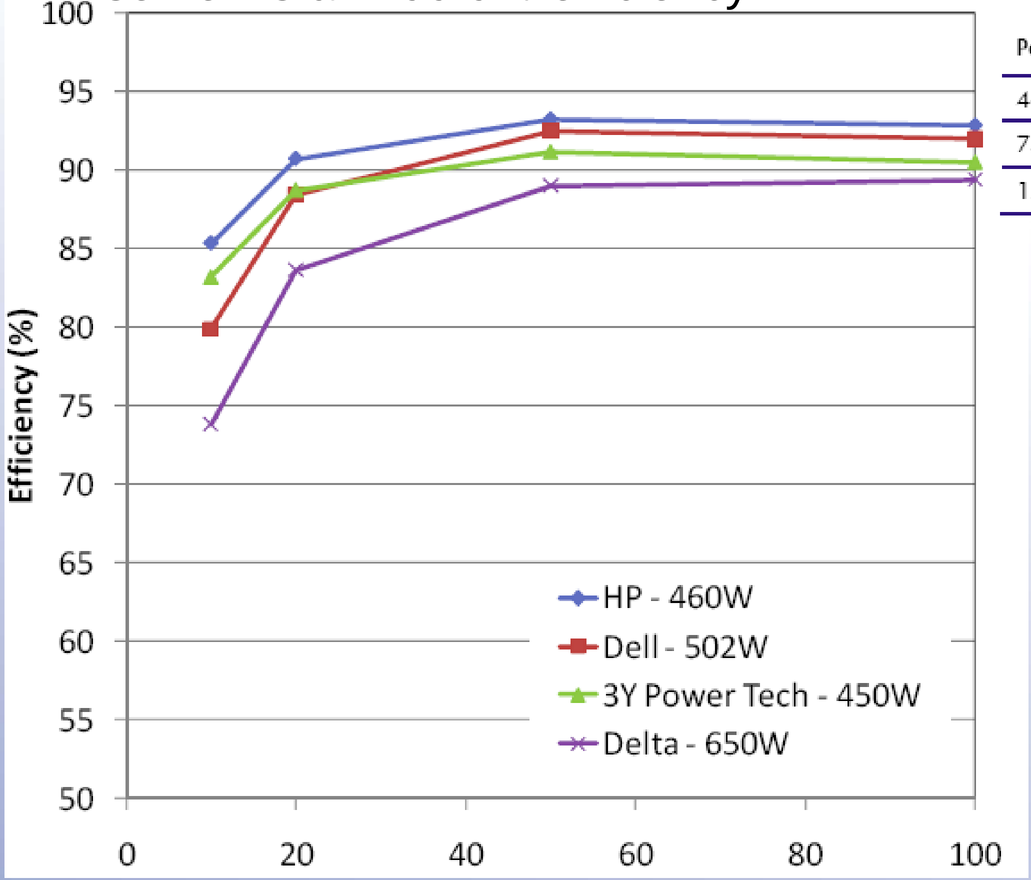
www.ibm.com

www.bull.com **www.dell.com**

ADVANCED COMPUTING RESEARCH LABORATORY

Power Supplies

Google: PS with battery built into server. Claim 99.9% efficiency



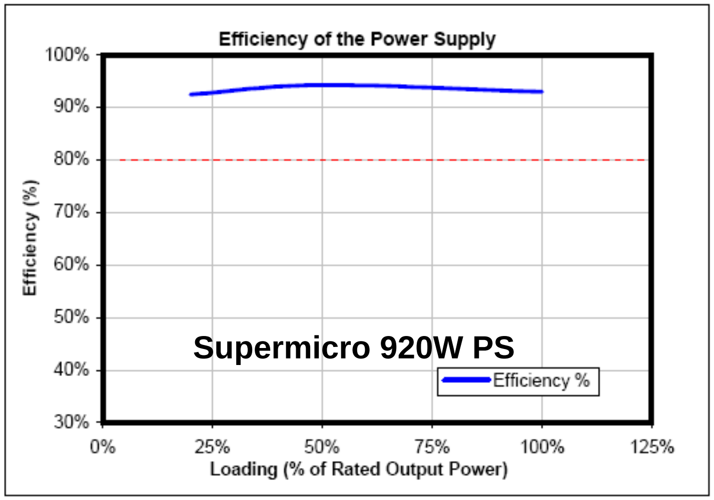
Source: Tahir Cader, HP, Energy Efficiency in HPC – An Industry Perspective, High Speed Computing,April 27 – 30, 2009

Example1: HP Proliant Power Supplies

Power supply type	Percent of efficiency @ 20% load	@ 50% load	@ 100% load	80 PLUS Certification
460-watt	90.70%	93.20%	92.81%	Gold
750-watt	91.33%	94.58%	92.57%	Gold
1200-watt (AC)	86.84%	91.75%	91.19%	Silver

Example2: HP Blade Chassis Power Supplies 94.6% peak (see www.80plus.org) and 90.x% at 10% load

http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00816246/c00816246.pdf



http://www.supermicro.com/products/powersupply/80PLUS/80PLUS_PWS-920P-1R.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

GreenLight Experiment: Direct 400V DC-Powered Modular Data Center

- Concept—avoid DC To AC To DC Conversion Losses
  - Computers Use DC Power Internally
  - Solar & Fuel Cells Produce DC
  - Can Computers & Storage Use DC Directly?
  - Is DC System Scalable?
  - How to Handle Renewable Intermittency?
- Prototype Being Built in GreenLight Instrument
  - Build DC Rack Inside of GreenLight Modular Data Center
    - 5 Nehalem Sun Servers
    - 5 Nehalem Intel Servers
    - 1 Sun Thumper Storage Server
  - Building Custom DC Sensor System to Provide DC Monitoring
  - Operational August-Sept. 2010

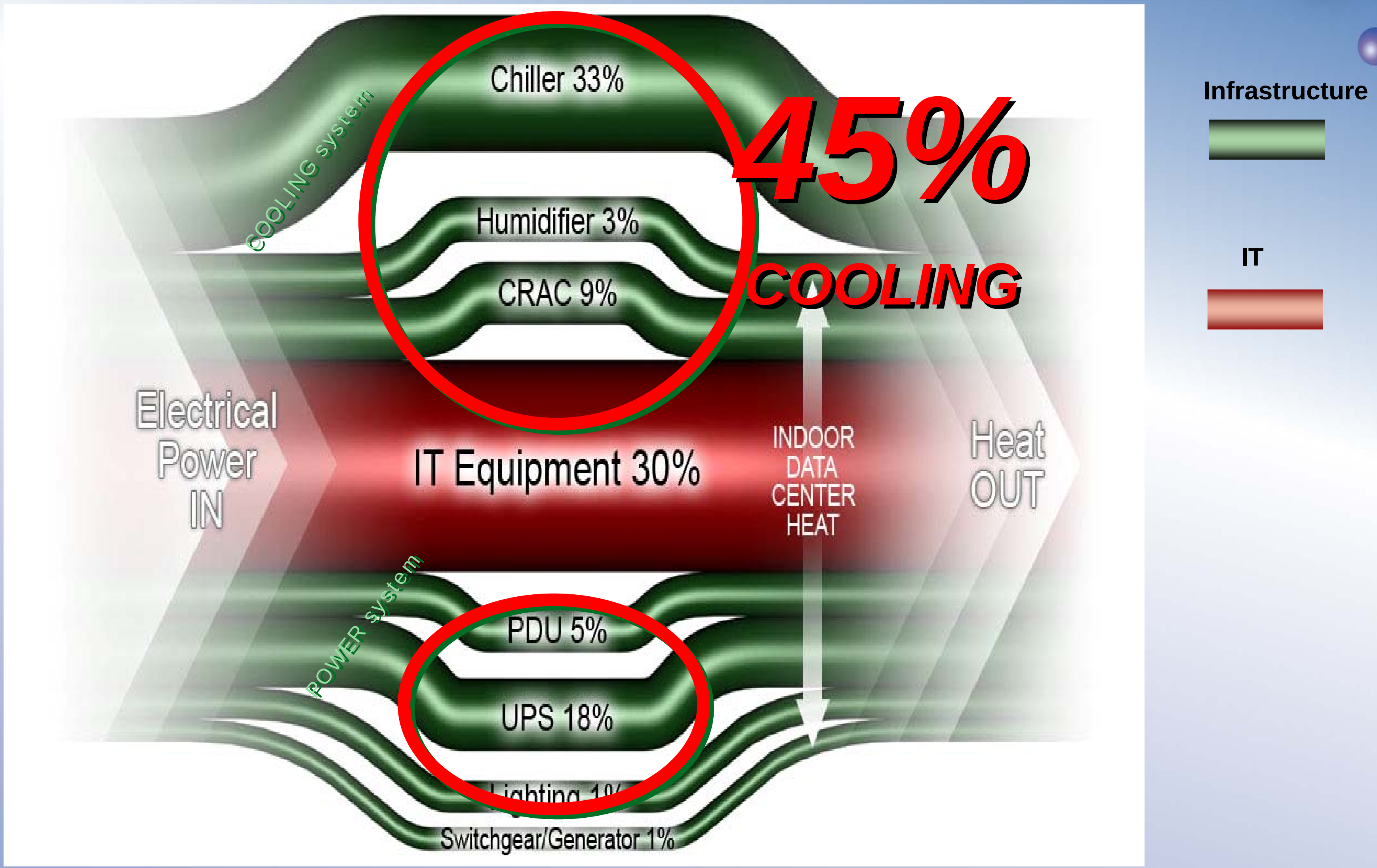
UCSD DC Fuel Cell 2800kW
Sun MDC <100-200kW

All With DC Power Supplies

Next Step: Couple to Solar and Fuel Cell

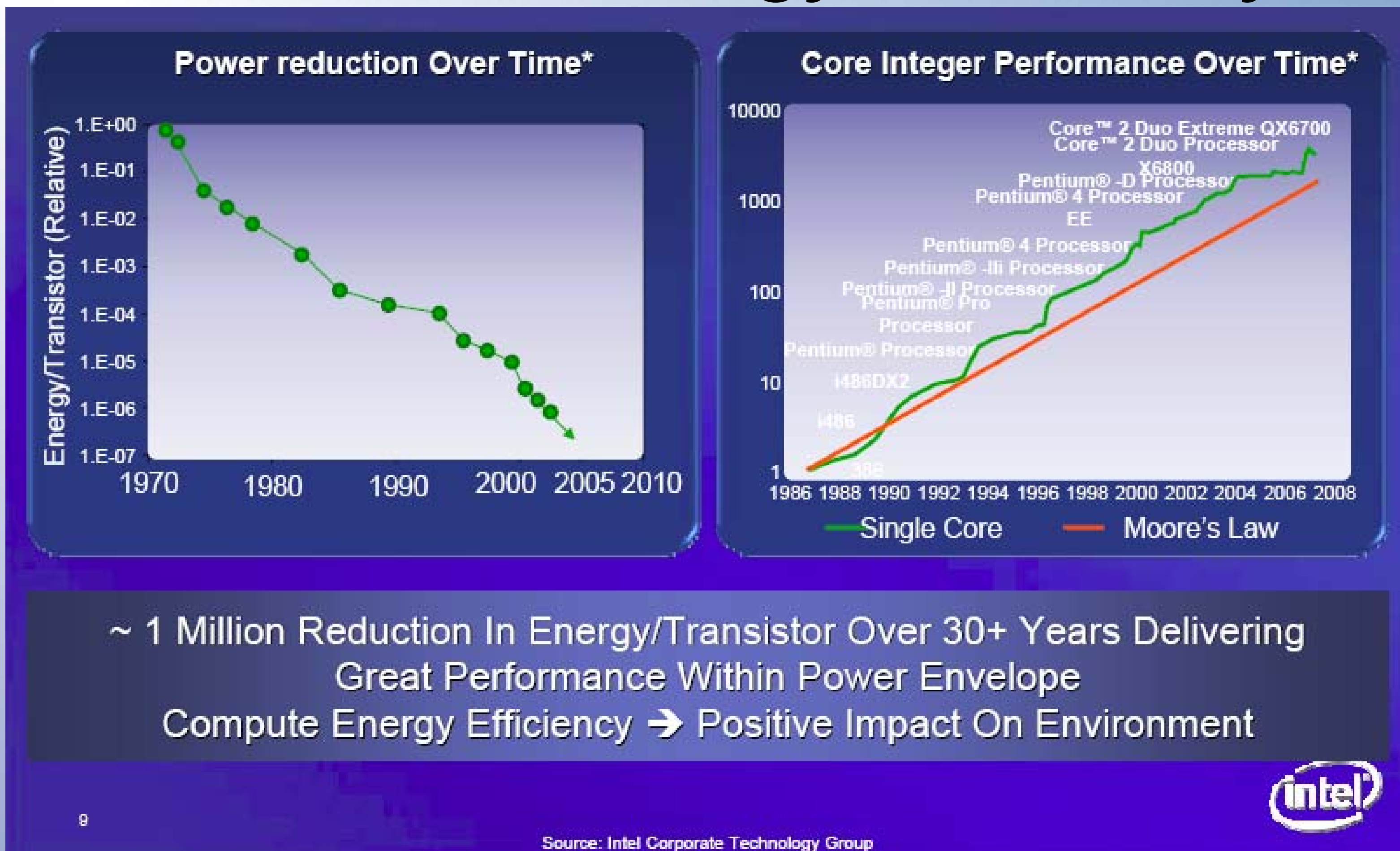
Source: Tom DeFanti, Greg Hidley, Calit2; Tajana Rosing, UCSD CSE
http://www.calit2.net/newsroom/presentations/lsmarr/2009/ppt/NCAR_072310.ppt

Recall - Traditional Data Center Energy Use



http://www.slideshare.net/fabioce/apc-il-moderno-data-center-4453903

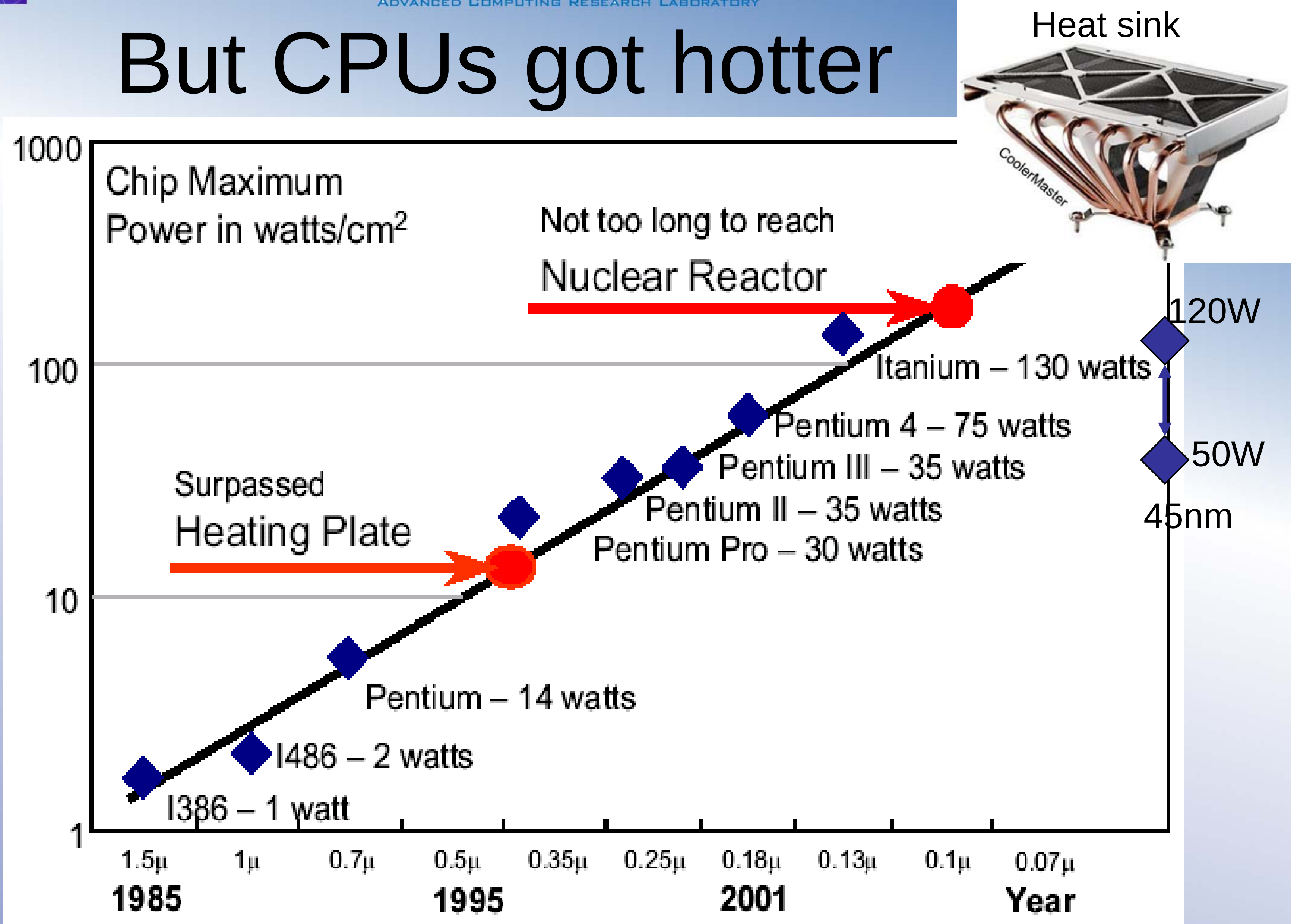
Incredible Improvement in Device Energy Efficiency



Source: Lorie Wigle, Intel, http://piee.stanford.edu/cgi-bin/docs/behavior/becc/2008/presentations/18-4C-01-Eco-Technology - Delivering Efficiency and Innovation.pdf

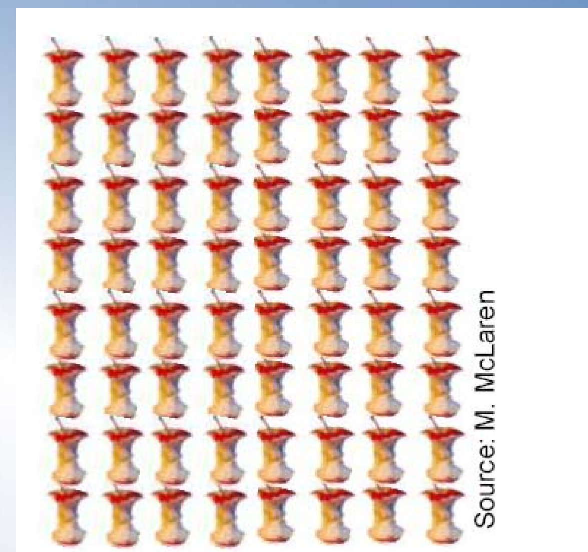
ADVANCED COMPUTING RESEARCH LABORATORY

But CPUs got hotter



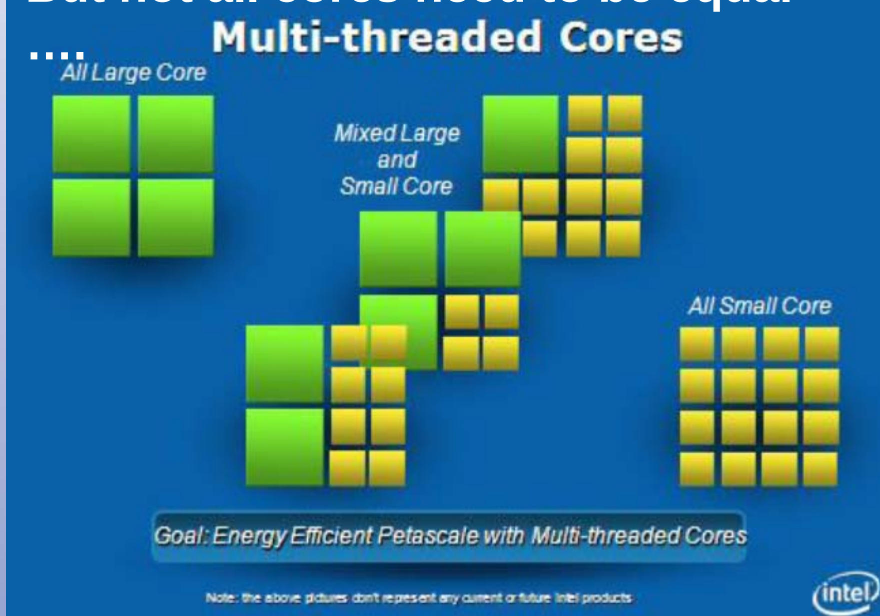
We got multi-core ….

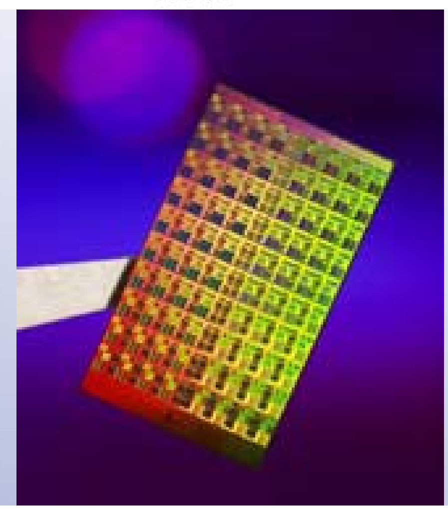
Today 6 – 12 cores



2015?

But not all cores need to be equal ….





IT Equipment: Where does the energy go?

- CPUs
- Memory
- Interconnect
- Fans
- Motherboards
- …….

Component	Peak Power	Count	Total
CPU	40 W	2	80 W
Memory	9 W	4	36 W
Disk	12 W	1	12 W
PCI Slots	25 W	2	50 W
Mother Board	25 W	1	25 W
Fan	10 W	1	10 W
System Total			213 W

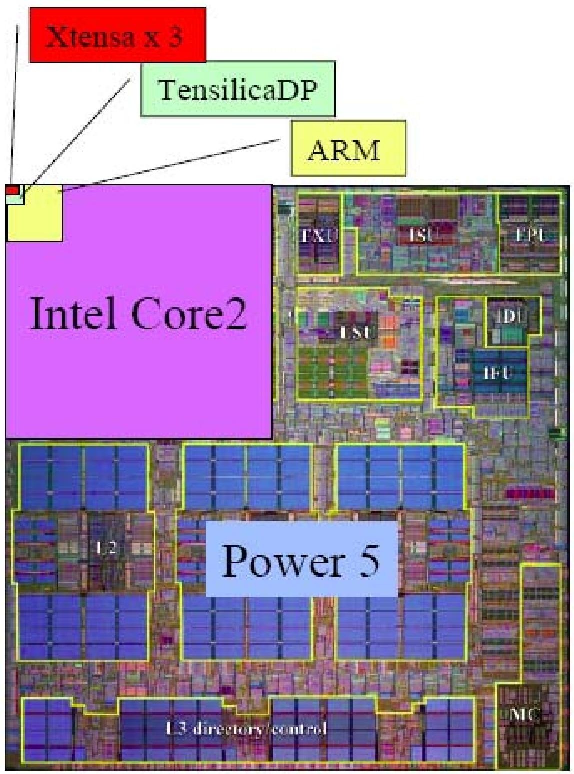
X. Fan, W-D Weber, L. Barroso, "Power Provisioning for a Warehouse-sized Computer," ISCA'07, San Diego, (June 2007).

How to improve energy efficiency of HPC systems?

- Alternative architectures ➔ programming model
- Dynamic control of systems
- Improved algorithms and software – energy aware

What kind of core - size

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

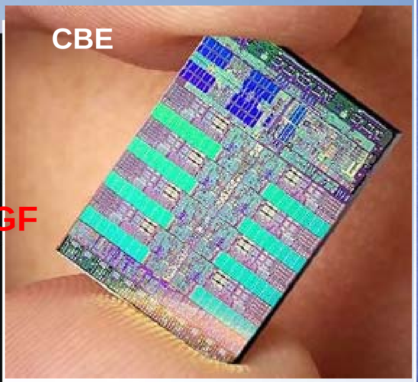
Each core operates at 1/3 to 1/10th efficiency of largest chip, but you can pack 100x more cores onto a chip and consume 1/20 the power

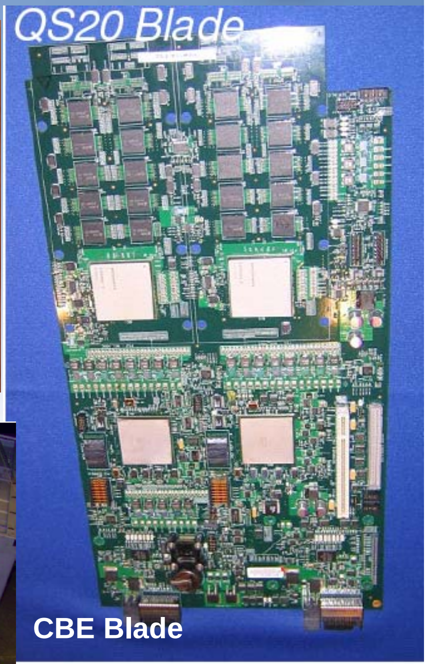
- **Cubic power improvement with lower clock rate due to V^2F**
- **Slower clock rates enable use of simpler cores**
- **Simpler cores use less area (lower leakage) and reduce cost**
- **Tailor design to application to reduce waste□□**

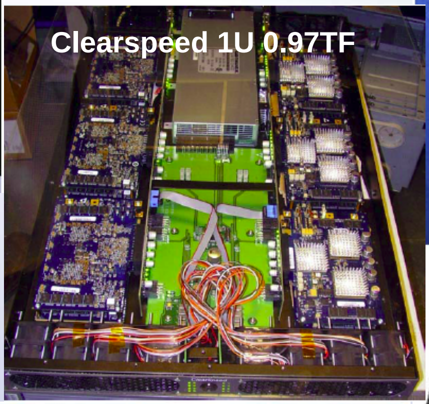
http://www.csm.ornl.gov/workshops/SOS11/presentations/j_shalf.pdf

What kind of core - accelerators

	Cell BE	Nvidia G80 C1060	ClearSpeed CSX600
32-bit FP	200+ GFLOPS	360+ GFLOPS	25+ GFLOPS
64-bit FP	20+ GFLOPS 100+GF	78 GF	25+ GFLOPS 96 GF
Clock frequency	3.2 GHz	575 MHz	210 MHz
Transistors/ chip	~ 241M	~ 681M	~ 128M
Power	~ 110 Watts	~ 145 W (for GeForce 8800 GTX board)	~ 10W 25W board

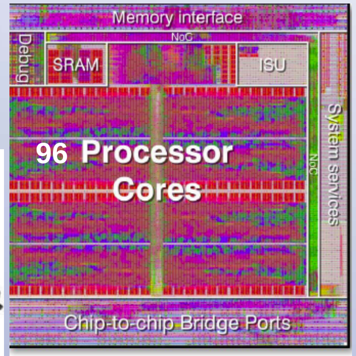






nVidia GF8800GTX

nVidia Tesla S870 1U 1.3 TF



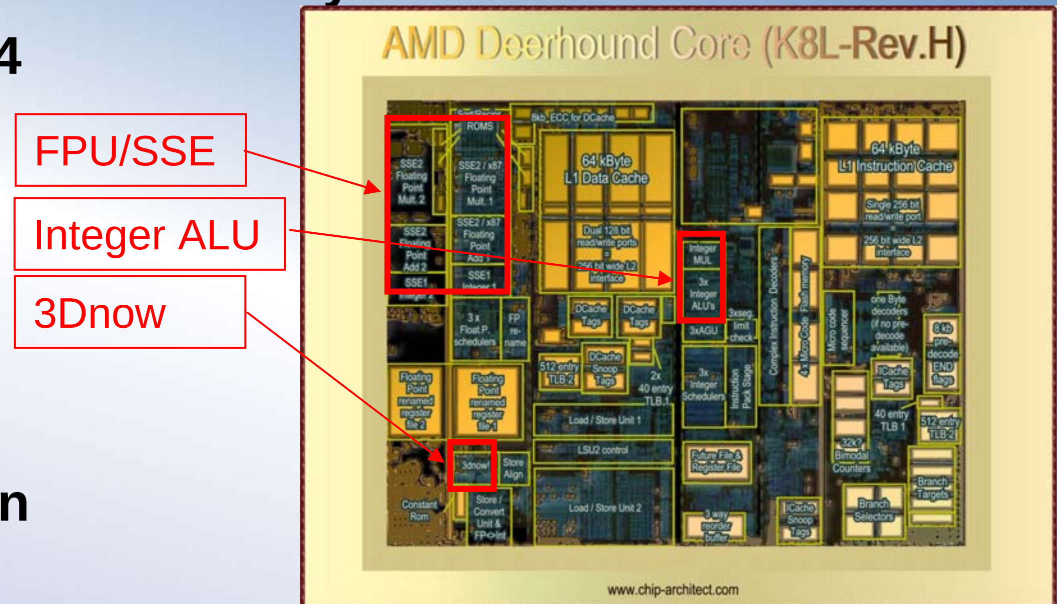
ClearSpeed CSX600

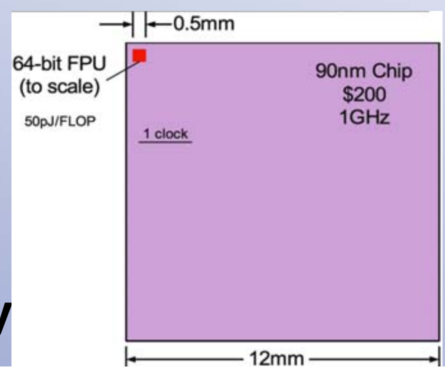
Clearspeed PCI-X board

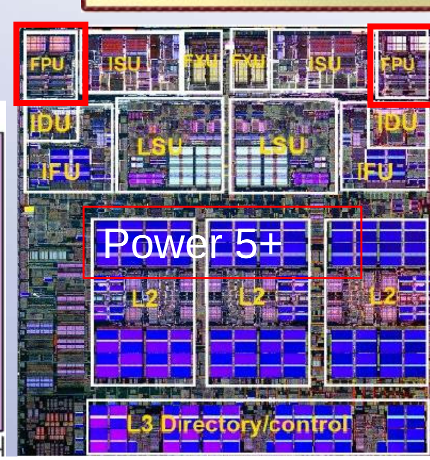
http://gamma.cs.unc.edu/SC2007/Sarkar-SC07-Multicore-Workshop.pdf

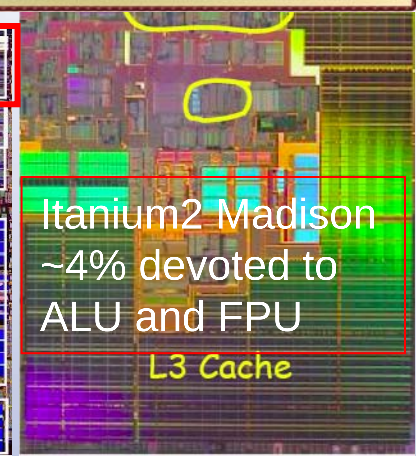
Why this performance difference?

- **Standard processors are optimized for a 40 year old model**
  - **Efficiency as defined in 1964**
  - **Heavy weight threads**
  - **Complex control**
  - **Big overhead per operation**
  - **Parallelism added as an afterthought**
  - **A large fraction of the silicon devoted to**
    - **Address translation**
    - **Instruction reordering**
    - **Register renaming**
    - **Cache hierarchy**









ADVANCED COMPUTING RESEARCH LABORATORY

Power fundamentals – Exascale

Processor

- Modern processors being designed today (for 2010) dissipate about 200 pJ/op total.

  This is ~**200W/TF 2010**

- In 2018 we might be able to drop this to 10 pJ/op

  ~ **10W/TF 2018**

- This is then **16 MW** for a sustained HPL Exaflops
- This does not include memory, interconnect, I/O, power delivery, cooling or anything else

Memory

- Cannot afford separate DRAM in an Exa-ops machine!
- Propose a MIP machine with Aggressive voltage scaling on 8nm
- Might get to 40 KW/PF –

  60 MW for sustained Exa-ops

Source: William J Camp, Intel, http://www.lanl.gov/orgs/hpc/salishan/pdfs/Salishan%20slides/Camp2.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

Power fundamentals - Exascale

Interconnect

- For short distances: still Cu
- Off Board: Si photonics
- Need ~ 0.1 B/Flop Interconnect
- Assume (a miracle) 5 mW/Gbit/sec

~ 50 MW for the interconnect!

I/O

- Optics is the only choice:
- 10-20 PetaBytes/sec
- ~ a few MW (a swag)

Power and Cooling

Still 30% of the total power budget in 2018!
Total power requirement in **2018**: **120—200 MW!**

Source: William J Camp, Intel, http://www.lanl.gov/orgs/hpc/salishan/pdfs/Salishan%20slides/Camp2.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

Power Fundamentals - Exascale

IBM

Extrapolating an Exaflop in 2018

Standard technology scaling will not get us there in 2018

	BlueGene/L (2005)	Exaflop Directly scaled	Exaflop compromise using evolutionary technology	Assumption for "compromise guess"
Node Peak Perf	5.6GF	20TF	20TF	Same node count (64k)
hardware concurrency/node	2	8000	1600	Assume 3.5GHz
System Power in Compute Chip	1 MW	3.5 GW	25 MW	Expected based on 30W for 200 GF with 6x technology improvement through 4 technology generations. (Only compute chip power scaling, I/Os also scaled same way)
Link Bandwidth (Each unidirectional 3-D link)	1.4Gbps	5 Tbps	1 Tbps	Not possible to maintain bandwidth ratio.
Wires per unidirectional 3-D link	2	400 wires	80 wires	Large wire count will eliminate high density and drive links onto cables where they are 100x more expensive. Assume 20 Gbps signaling
Pins in network on node	24 pins	5,000 pins	1,000 pins	20 Gbps differential assumed. 20 Gbps over copper will be limited to 12 inches. Will need optics for in rack interconnects. 10Gbps now possible in both copper and optics.
Power in network	100 KW	20 MW	4 MW	10 mW/Gbps assumed. Now: 25 mW/Gbps for long distance (greater than 2 feet on copper) for both ends one direction. 45mW/Gbps optics both ends one direction. + 15mW/Gbps of electrical Electrical power in future: separately optimized links for power.
Memory Bandwidth/node	5.6GB/s	20TB/s	1 TB/s	Not possible to maintain external bandwidth/Flop
L2 cache/node	4 MB	16 GB	500 MB	About 6-7 technology generations with expected eDRAM density improvements
Data pins associated with memory/node	128 data pins	40,000 pins	2000 pins	3.2 Gbps per pin
Power in memory I/O (not DRAM)	12.8 KW	80 MW	4 MW	10 mW/Gbps assumed. Most current power in address bus. Future probably about 15mW/Gbps maybe get to 10mW/Gbps (2.5mW/Gbps is c*v^2*f for random data on data pins) Address power is higher.
QCD CG single iteration time	2.3 msec	11 usec	15 usec	Requires: 1) fast global sum (2 per iteration) 2) hardware offload for messaging (Driverless messaging)

~0.02 B/s:F/s
Power, cost and packaging driven

~1/20 B/Flop bandwidth
Power and packaging driven

Source: Dave Turek, IBM, CASC 20 yr Anniversary, September 22 – 23, 2009, http://www.casc.org/meetings/09sept/Dave_Turek.ppt

ADVANCED COMPUTING RESEARCH LABORATORY

DARPA Exascale study

- Last 30 years:
  - "Gigascale" computing first in a single vector processor
  - "Terascale" computing first via several thousand microprocessors
  - "Petascale" computing first via several hundred thousand cores
- Commercial technology: *to date*
  - Always shrunk prior "XXX" scale to smaller form factor
  - Shrink, with speedup, enabled next "XXX" scale
- Space/Embedded computing has lagged far behind
  - Environment forced implementation constraints
  - Power budget limited both clock rate & parallelism
- "Exascale" now on horizon
  - But beginning to suffer similar constraints as space
  - And technologies to tackle exa challenges ***very relevant***

Especially Energy/Power

http://www.ll.mit.edu/HPEC/agendas/proc09/Day1/S1_0955_Kogge_presentation.ppt

Green Flash Strawman System Design

Three different approaches examined (in 2008 technology)
Computation .015ºX.02ºX100L: 10 PFlops sustained, ~200 PFlops peak

- AMD Opteron: **Commodity approach, lower efficiency for scientific applications offset by cost efficiencies of mass market**
- BlueGene: **Generic embedded processor core and customize system-on-chip (SoC) to improve power efficiency for scientific applications**
- Tensilica XTensa: **Customized embedded CPU w/SoC provides further power efficiency benefits but maintains programmability**

Processor	Clock	Peak/ Core (Gflops)	Cores/ Socket	Sockets	Cores	Power	Cost 2008
AMD Opteron	2.8GHz	5.6	2	890K	1.7M	179 MW	$1B+
IBM BG/P	850MHz	3.4	4	740K	3.0M	20 MW	$1B+
Green Flash / Tensilica XTensa	650MHz	2.7	32	120K	4.0M	3 MW	$75M

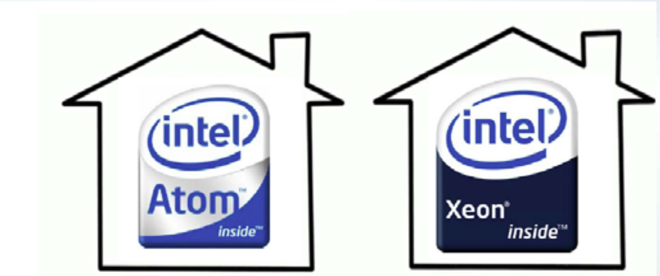
Slide courtesy Horst Simon, NERSC, http://www.cs.berkeley.edu/~demmel/cs267_Spr09/Lectures/SimonPrinceton0409.ppt

ADVANCED COMPUTING RESEARCH LABORATORY

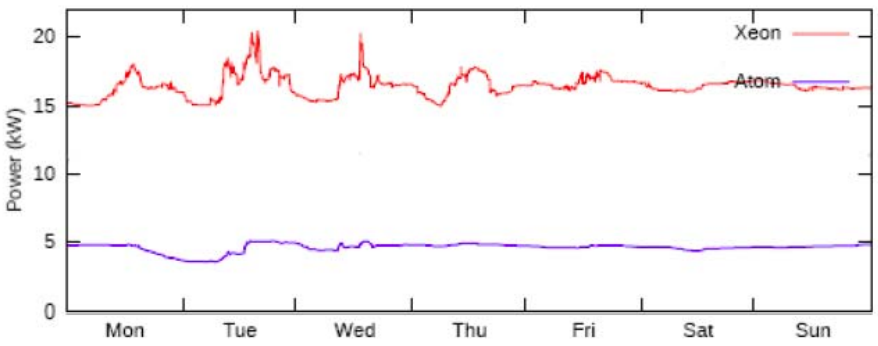
SGI Molecule (SC08 Concept)

- Intel ATOM processor ~2W
- 360 cores on 180 nodes in 3U
- 5040 cores in standard rack with 10TB
- 2GB/s per core
- ~1.6 GHz
- ~15kW/rack

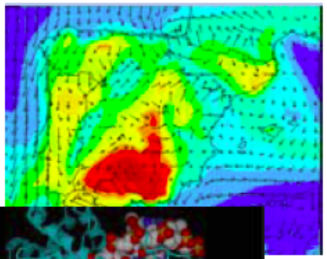
Low Power CPUs vs "Standard" CPUs



- **Good approach for transactional workloads (webs)**
  - Experiments at BSC [1]:



- **However the numerical applications are not like that**
  - Experiments at BSC with 4 HPC tasks[1]:
    - 1 Xeon * 1 hour → 317 Watts
    - 2 Atom * 5 hours → 398 Watts
  - **The most energy-efficient approach in this environment is running jobs very fast and then power the system off.**



Source: Jordi Torres 2010 Campusparty.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

ARM based server …..

The New York Times

Business Day
Technology

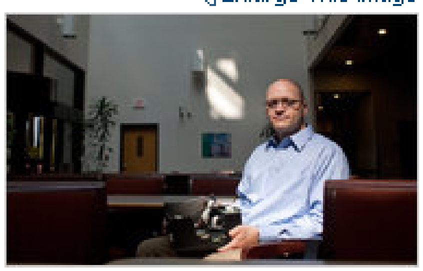
WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH

Start-Up Aims to Slay Chip Goliath

By ASHLEE VANCE
Published: August 15, 2010

A group of investors, including companies from the United States, Europe and the United Arab Emirates, has formed in a bid to disrupt one of Intel's most lucrative franchises.

Enlarge This Image



Ben Sklar for The New York Times

Barry Evans is chief of Smooth-Stone, a name that refers to David's weapon in the Bible.

The companies have put $48 million into Smooth-Stone, a start-up based in Austin, Tex., betting that it can modify low-power smartphone chips to run servers, the computers in corporate data centers. If successful, Smooth-Stone would undermine Intel's server-chip business and offer companies, especially those with vast data centers like Google, Amazon.com, Facebook and Microsoft, cost savings.

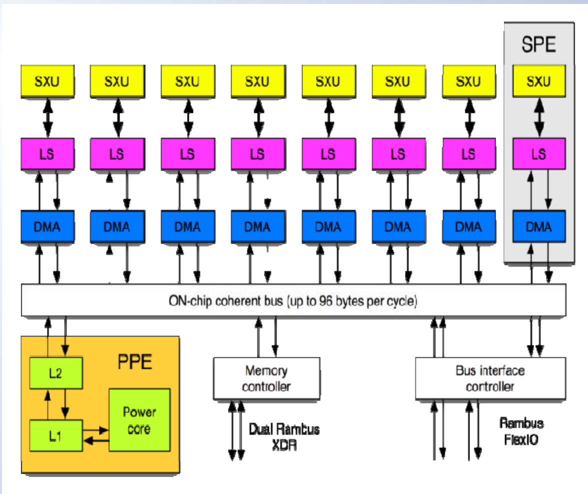
PRACE Technology Prototypes

Prototypes	Installation Site	Targeted Components
eQPACE	JSC, Germany	Interconnects, energy efficiency and density
RapidMind	BAdW-LRZ, Germany	Programming models for hybrid systems
LRZ-CINES 1	CINES, France BAdW-LRZ, Germany	Intel Nehalem-EP, ClearSpeed and QDR Infiniband
LRZ-CINES 2	BAdW-LRZ, Germany	Intel Nehalem-EX, Numalink5, Intel Larrabee
Hybrid Technology	CEA, France	GPGPU, HMPP
Maxwell FPGA	EPCC, UK	FPGA, energy efficiency and programing
PGAS Compiler	CSCS, Switzerland	PGAS programming model
ClearSpeed	NCF, Netherlands	ClearSpeed
XC4-IO	CINECA, Italy	I/O and File System perf/, SSD for metadata,
Accelerator efficiency	PSNC, Poland, SFTC, UK	Power consumption, porting of applications
PGAS Programming	CSC, Finland	Performance of UPC and CAF
Parallel GPU	CSC, Finland	Parallelizing CUDA, porting CUDA to OpenCL
SNIC-KTH	KTH, Sweden	Energy efficient computing

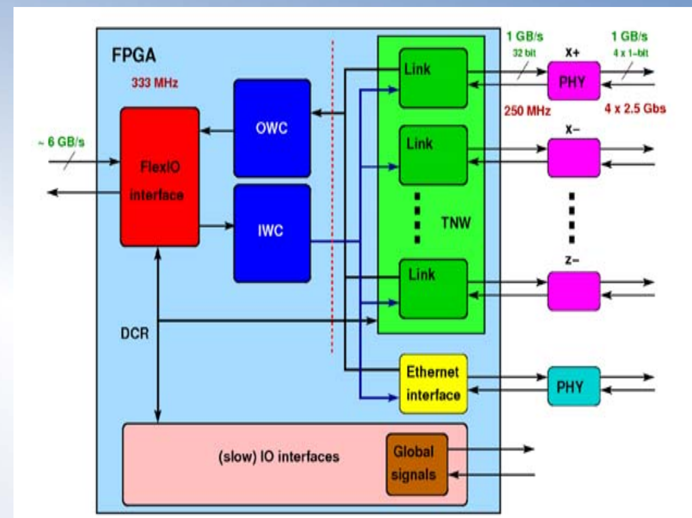
How Green is Green HPC?

Green500 Rank	MFLOPS/W	Computer June 2010 List	Power (kW)
1	773.38	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.74
1	773.38	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.74
1	773.38	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.74
4	492.64	Nebulae	2580.00
5	458.33	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz , IB	276.00
5	458.33	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz , IB	138.00
7	444.94	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz , Voltaire IB	2345.50
8	431.88	Mole-8.5 Cluster Xeon L5520 2.26 Ghz, nVidia Tesla, IB	480.00
9	418.47	iDataPlex, Xeon X56xx 6C 2.8 GHz, IB	72.00
10	397.56	iDataPlex, Xeon X56xx 6C 2.66 GHz, IB	72.00

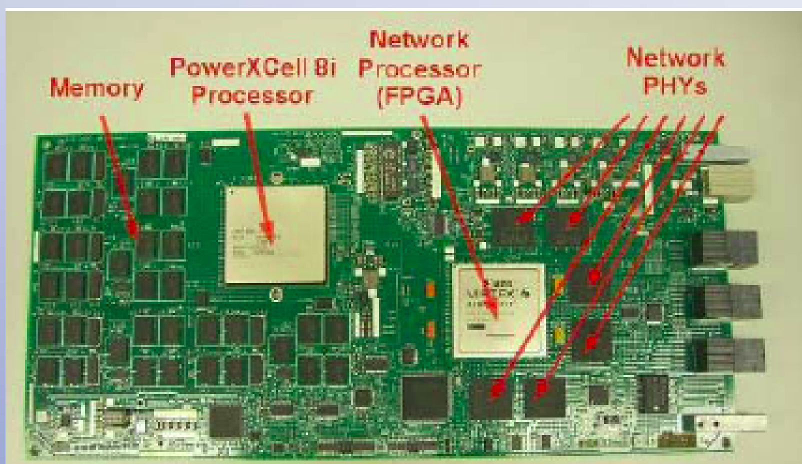
eQPACE (extended QCD PArallel computing on Cell)



Cell processor PowerXCell 8i



eQPACE FPGA network processor (extension of QPACE)



eQPACE board

eQPACE with frontend at JSC

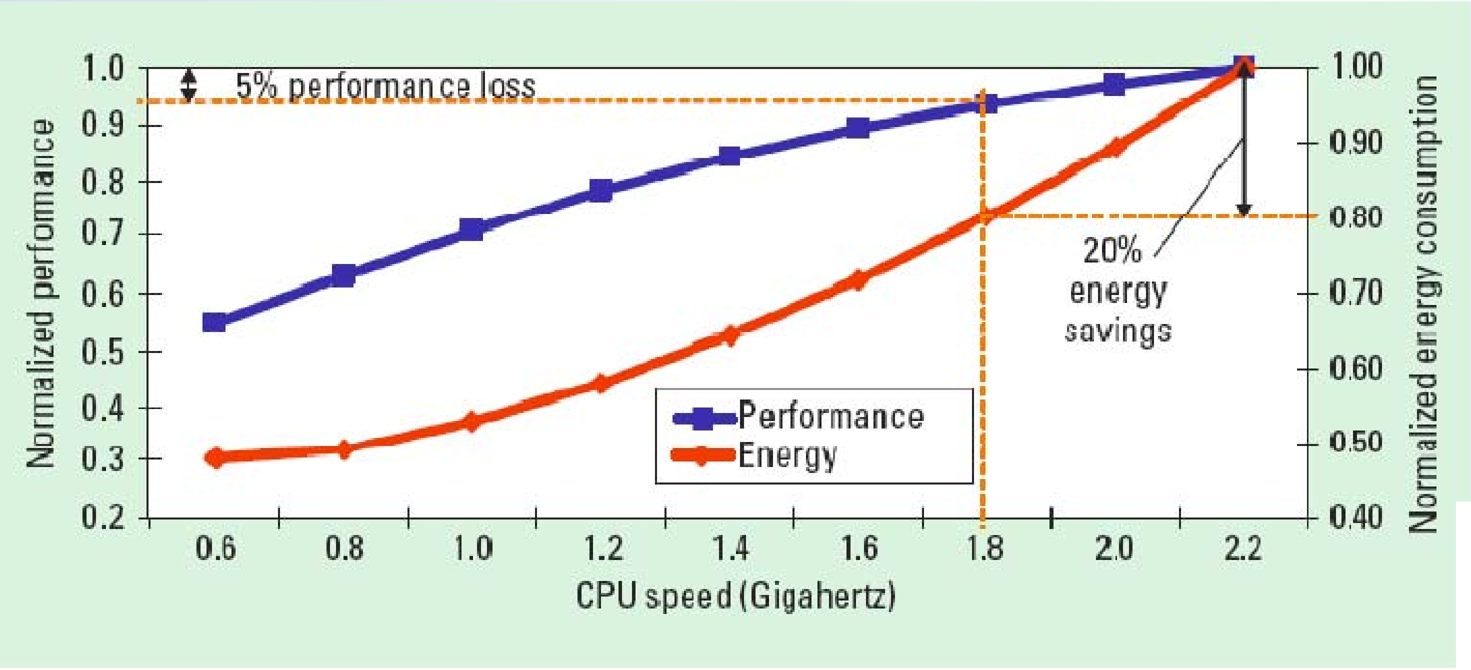
Dynamic Control of systems

Dynamic Voltage and Frequency Scaling

Have been used in mobile computing for several years, but relatively new in HPC

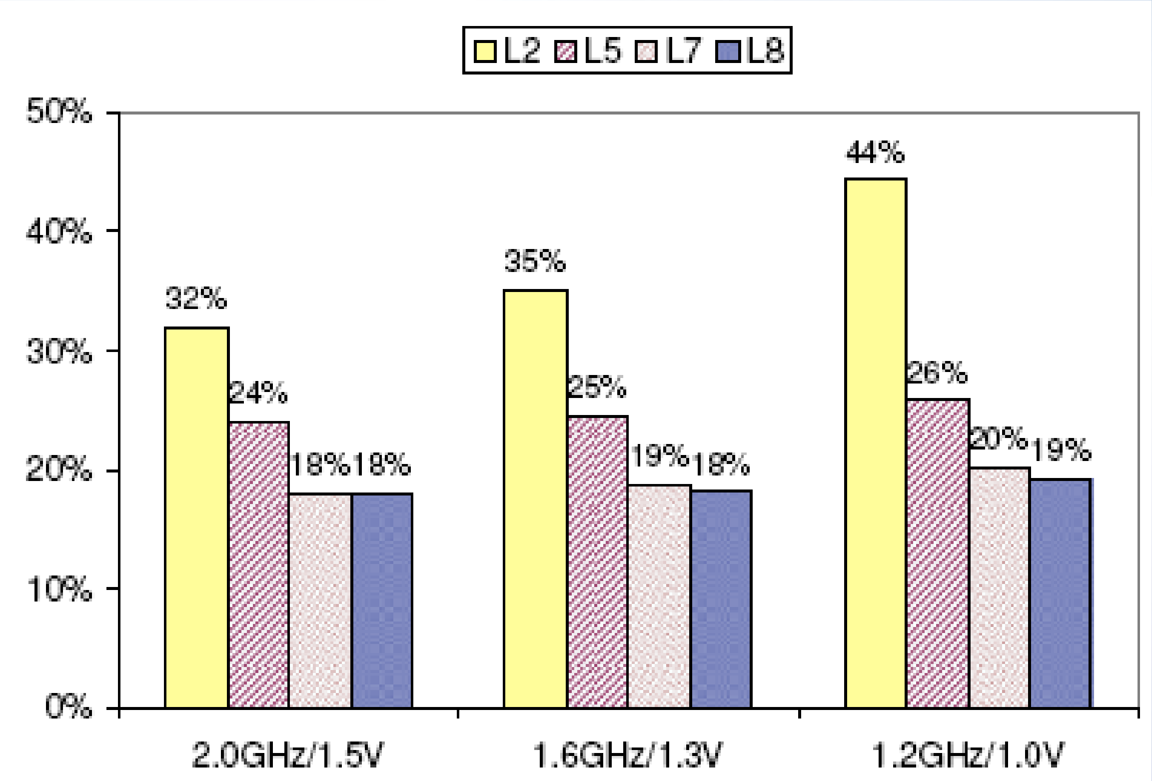
Power Performance trade-off

- Power ∝ Voltage^2 x frequency (V^2f)
- Frequency ∝ Voltage
- Power ∝ Frequency^3



Source: The Case for Energy-Proportional Computing, Borroso, Holze, IEEE Computer, December 2007

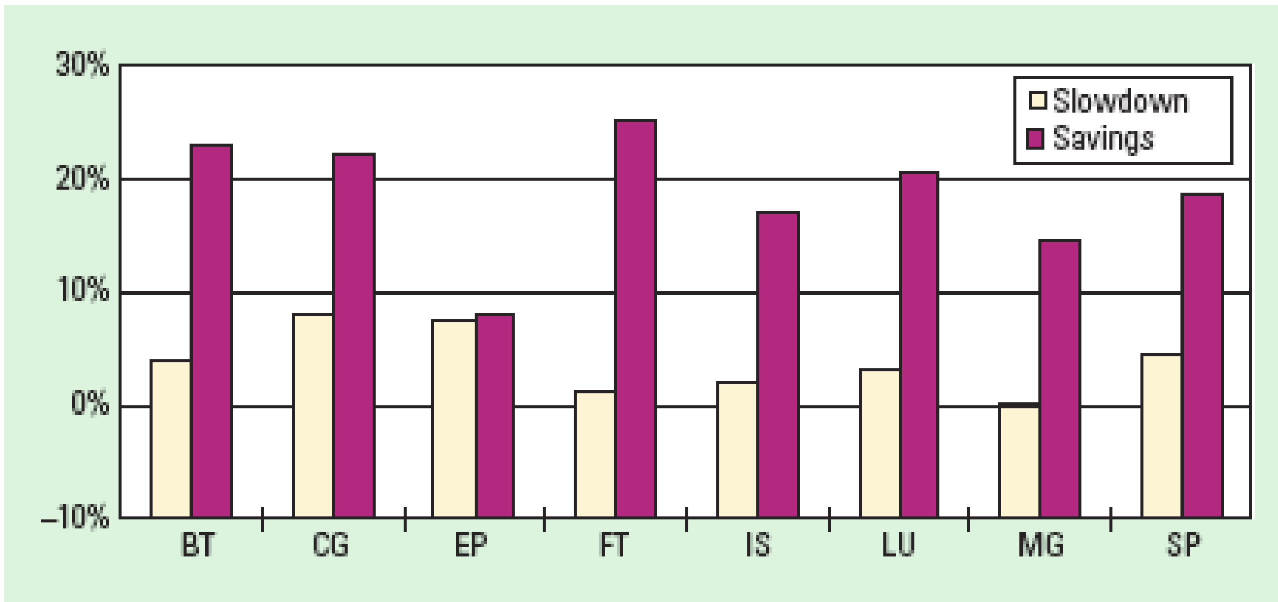
Example: Tomcatv (mesh generation code in the SPEC benchmark suite)



Source: A Power-Aware Run-Time System for HighPerformance Computing, Chunghsing Hsu and Wuchun Feng

ADVANCED COMPUTING RESEARCH LABORATORY

NAS Parallel Benchmarks on Opteron Cluster



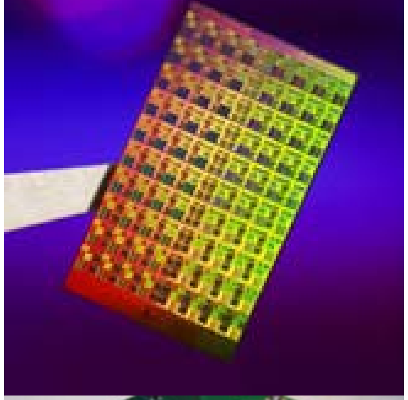
3 Energy savings and performance slowdown of the EnergyFit DVFS algorithm. Overall, we observed an average of 20 percent energy savings and 3 percent performance slowdown. For the MG benchmark, we observed a 15 percent energy savings and 1 percent performance speedup.

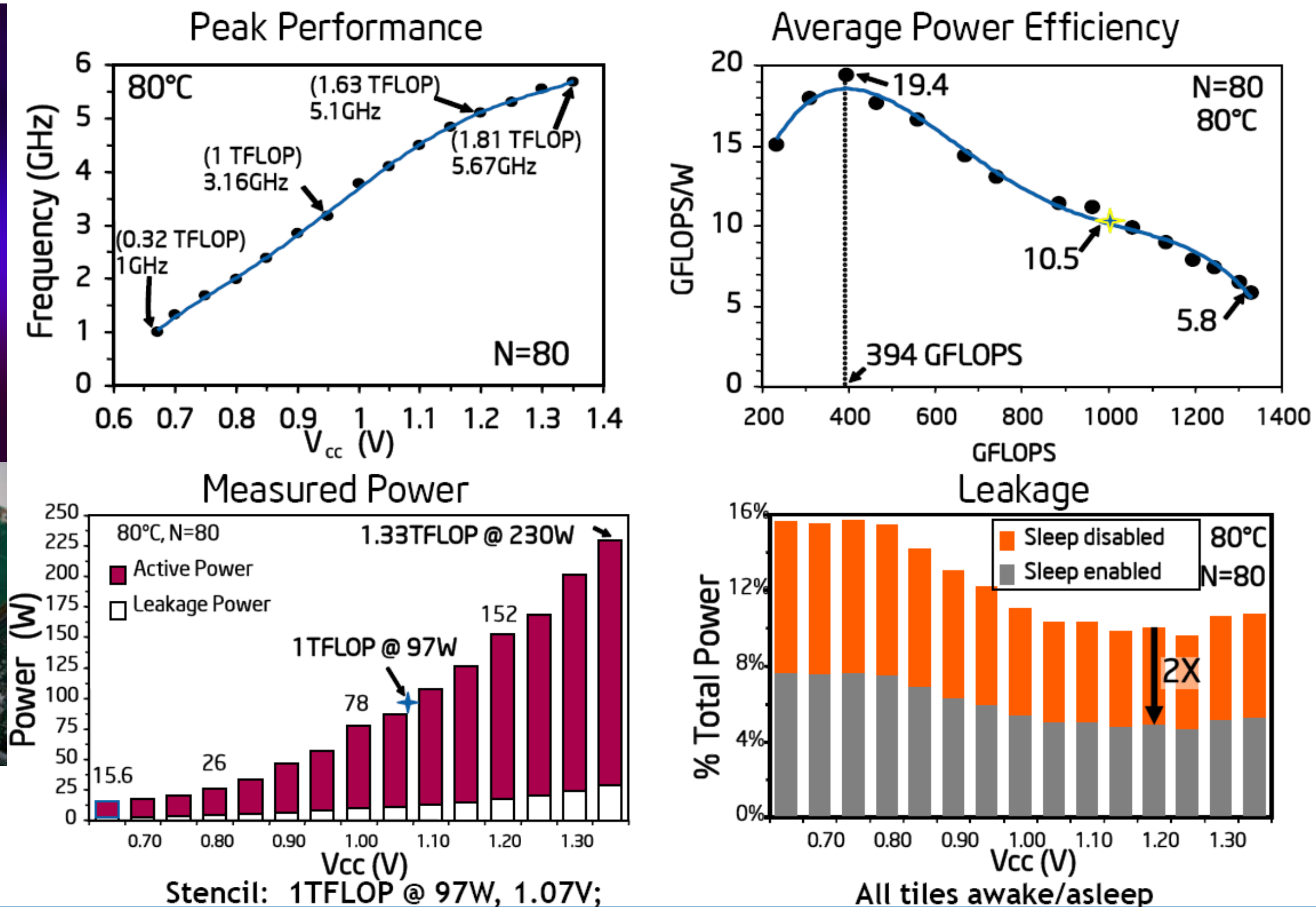
Source: A Power-Aware Run-Time System for HighPerformance Computing, Chunghsing Hsu and Wuchun Feng

ADVANCED COMPUTING RESEARCH LABORATORY

Power Performance Results

Polaris 80-core

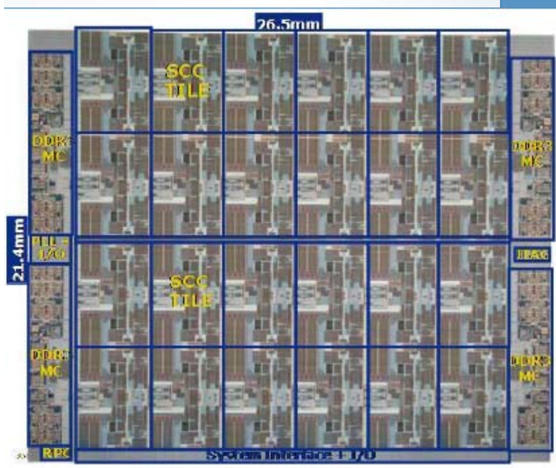


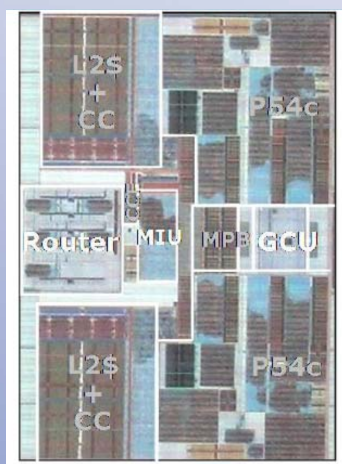


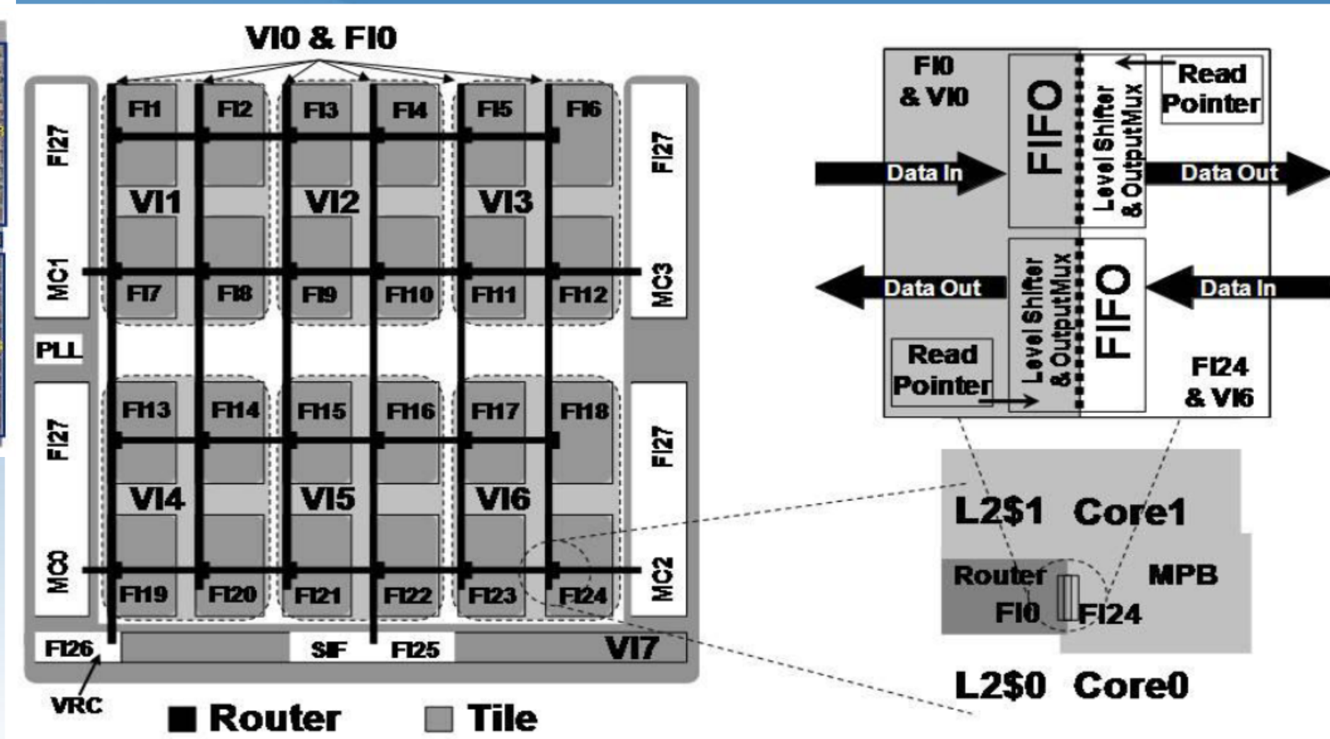
Vangal, S., et al., "An 80-Tile 1.28TFLOPS Network-on-Chip in 65 nm CMOS," in *Proceedings of ISSCC 2007(IEEE International Solid-State Circuits Conference)*, Feb. 12, 2007.

Intel's Single-Chip Cloud Computer

Voltage and Frequency islands







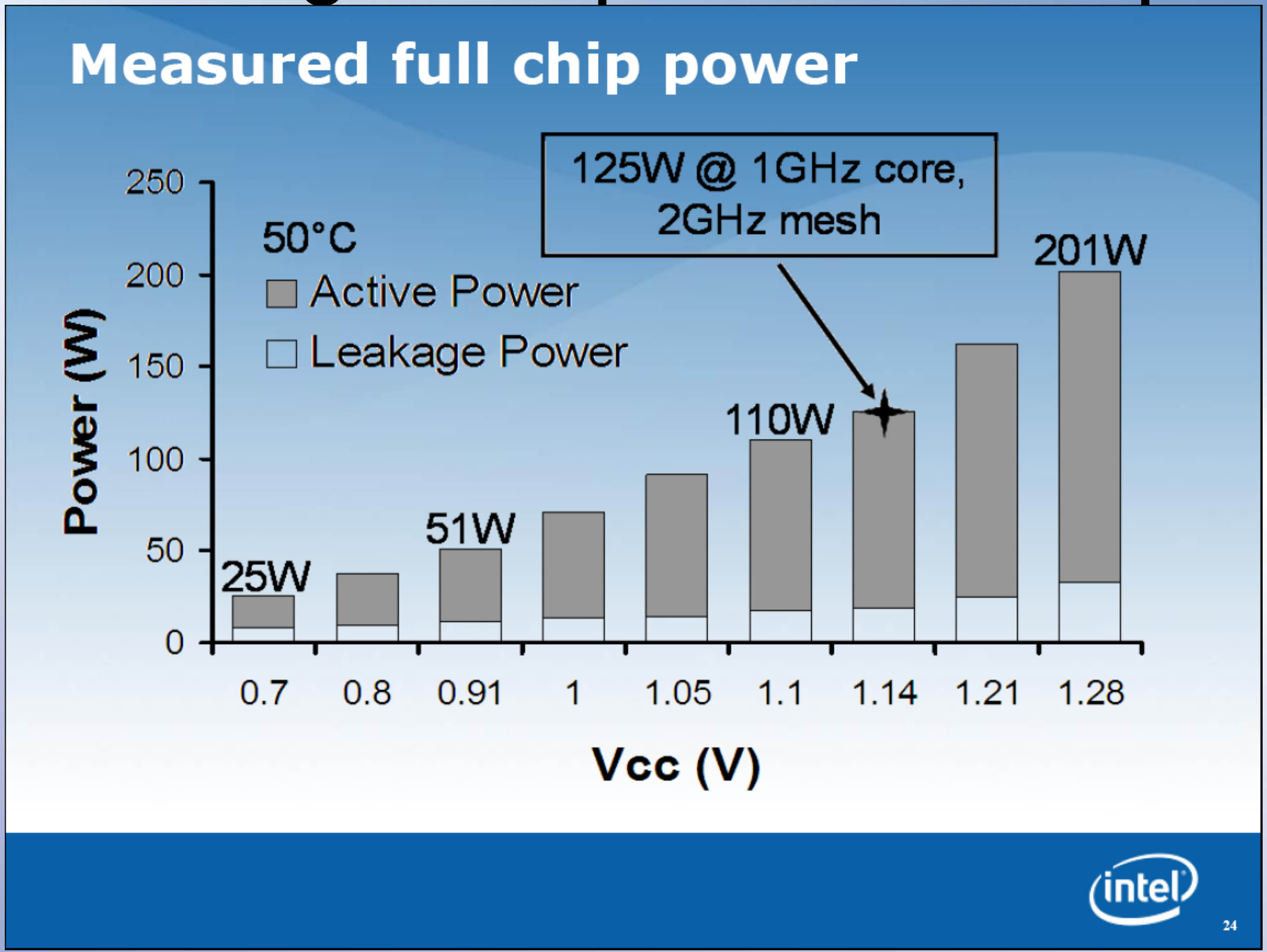
28 Frequency Islands (FI) 8 Voltage Islands (VI)

21

http://techresearch.intel.com/articles/Tera-Scale/1836.htm

ADVANCED COMPUTING RESEARCH LABORATORY

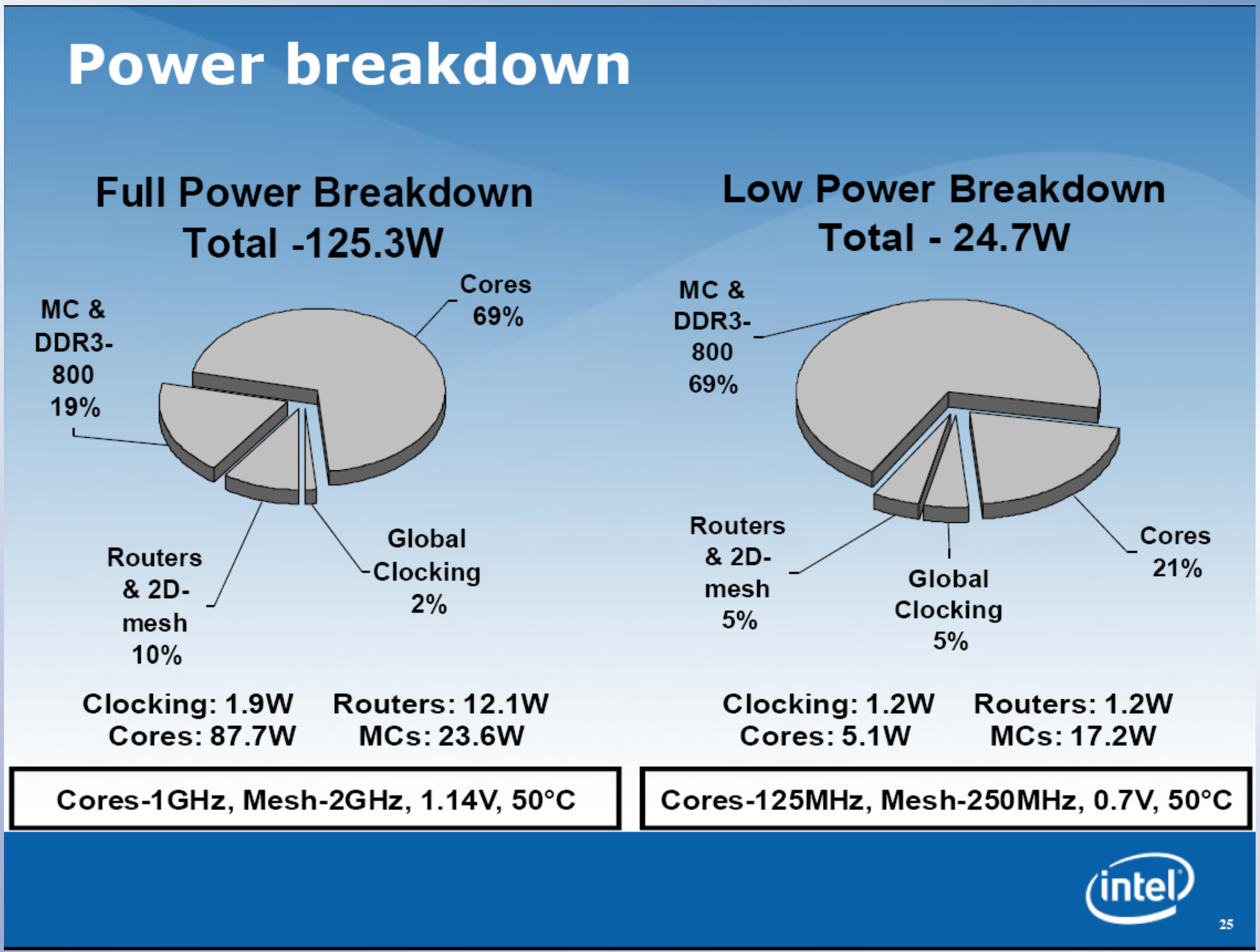
Intel's Single-Chip Cloud Computer



http://techresearch.intel.com/articles/Tera-Scale/1836.htm

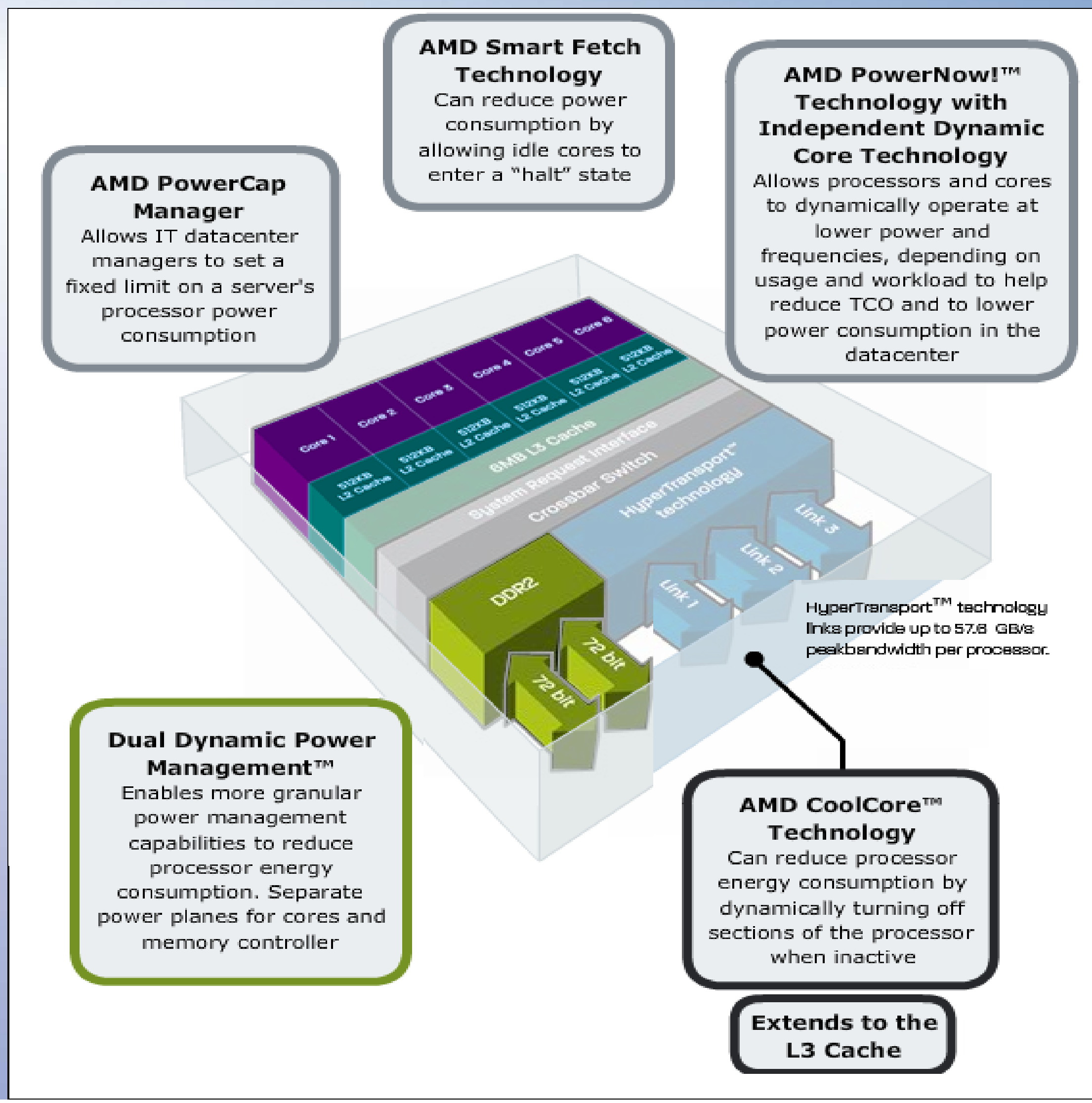
PDC Summer School,
Aug 26 2010
Lennart Johnsson

Intel's Single-Chip Cloud Computer

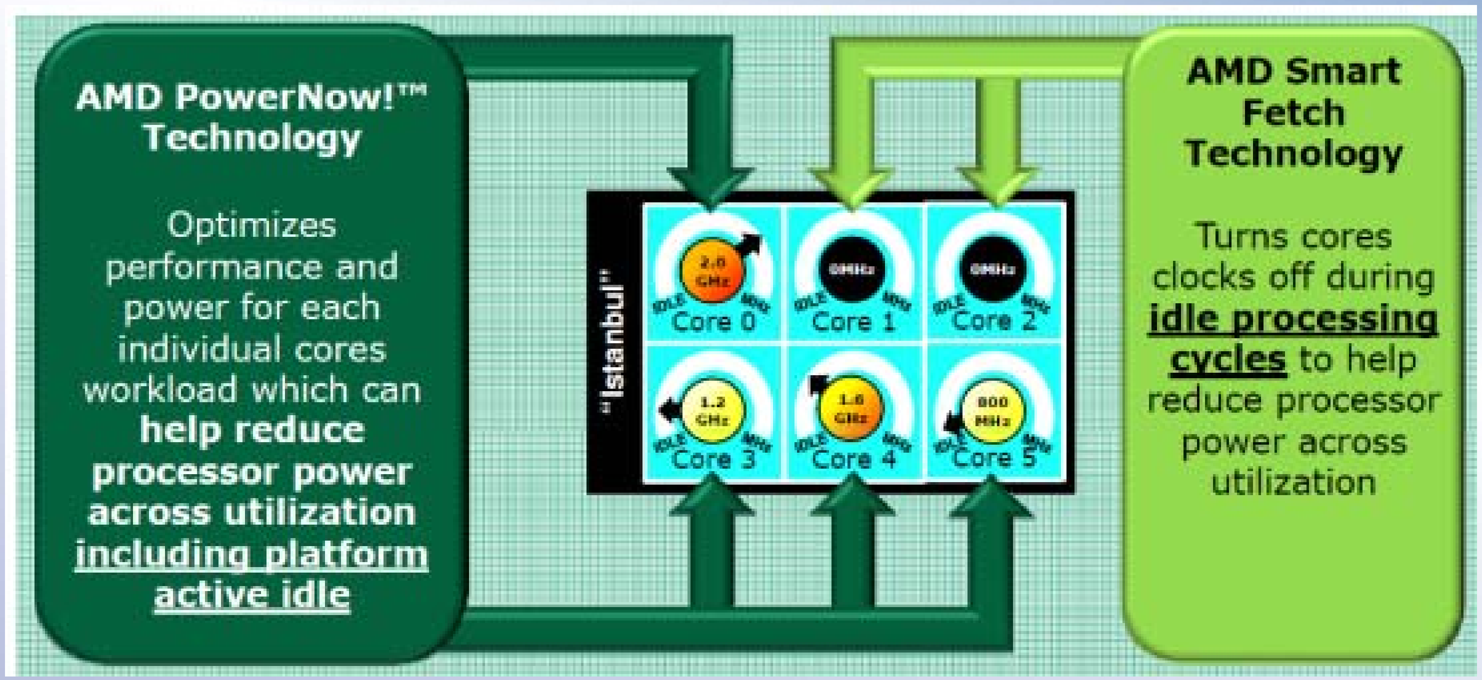


http://techresearch.intel.com/articles/Tera-Scale/1836.htm

AMD Power Management Technologies



AMD Power Management Technologies



Each core can have its own frequency

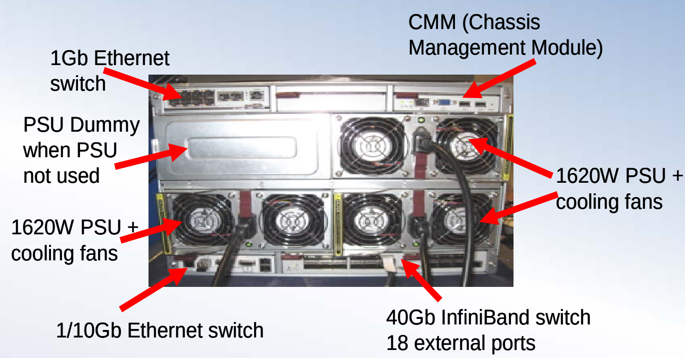
Two voltage planes: once for all cores, one for memory controller

AMD Power Management Technologies

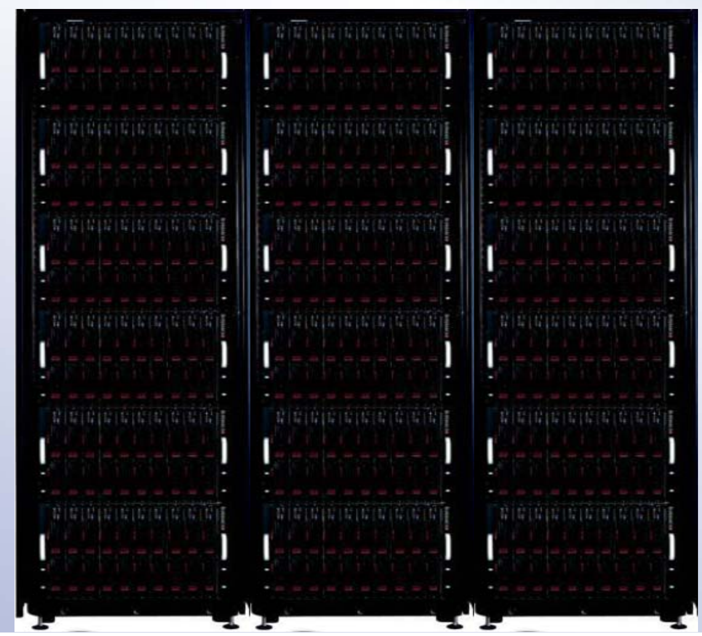
CoolCore technology can automatically turn off sections of the core logic and memory controller to reduce power consumption. Power for these sections can be turned off or on very fast - within one clock cycle.
This feature was introduced in AMD K10 micro-architecture.

Smart Fetch Technology allows cores to enter a "halt" state during idle processing times, causing them to draw less power. Before entering the halt state, data from the L1 and L2 caches are transferred to the shared L3 cache so that the contents of the idle cores can be retrieved.

SNIC/KTH PRACE Prototype

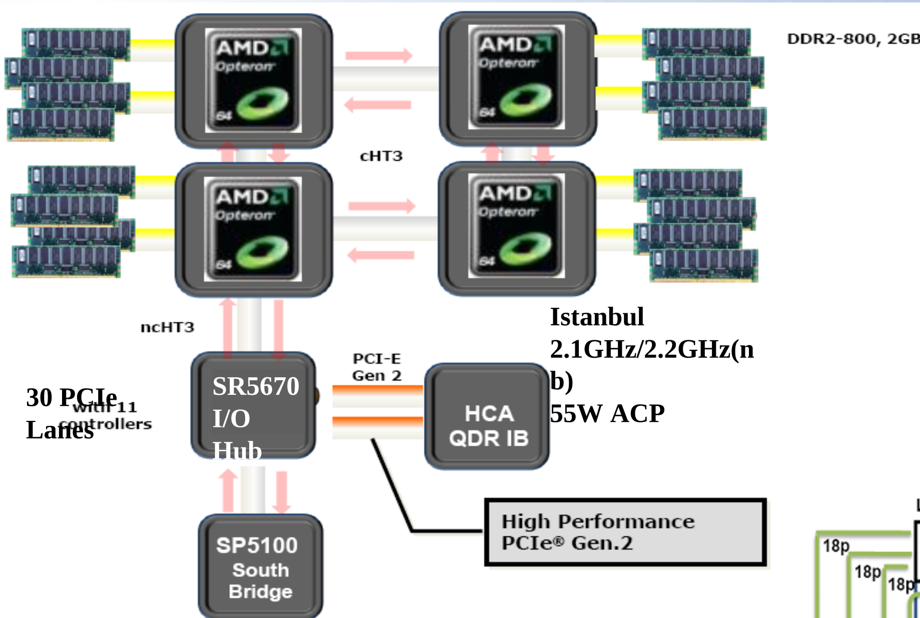


- New 4-socket blade with 4 DIMMs per socket supporting PCI-Express Gen 2 x16
- Four 6-core 2.1 GHz 55W ADP AMD Istanbul CPUs, 32GB/node
- 10-blade in a 7U chassis with 36-port QDR IB switch, new efficient power supplies.
- 2TF/chassis, 12 TF/rack, 30 kW (6 x 4.8)
- 180 nodes, 4320 cores, full bisection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



ADVANCED COMPUTING RESEARCH LABORATORY

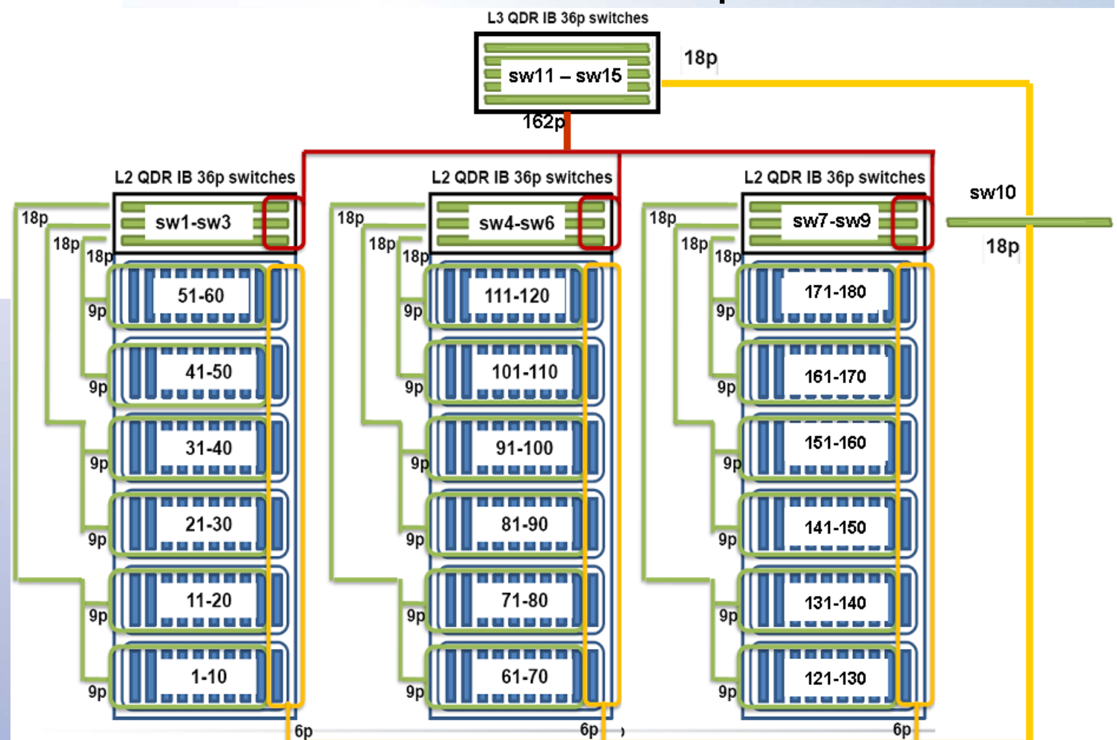
SNIC/KTH PRACE Prototype



Node

Network:

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



ADVANCED COMPUTING RESEARCH LABORATORY

Density - Examples

	Sockets/ rack	Cores/ rack	GF/ core	TF/ rack	kW/ rack	TF/ m^2	kW/m^2	TF/ kW	Linpack TF/kW	Linpack Eff.
BG/P	2048	4096	3.4	13.9	40	20.6	59	0.35	0.36 – 0.37	0.80
HP 2x blades (HTN, 80W)	256	1024	12 (3GHz)	12.3 (3GHz)	45	19.1	70	0.27	0.22	0.79
SGI Molecule	192	384	1.6	0.6	2	0.9	3	0.3	0.16	0.5
SiCortex	972	5832	1.4	8.2	22	2.2	8.5	0.37	0.22	0.58
SiCortex 2H09	972	11664	2.8	32.7	30	12.5	11.6	1.09	0.63	0.58
Supermicro PRACE prop	240	1440	9.6	13.8	32	19.2 – 21.5	44.4 – 49.8	0.43	0.36	0.84
Twin servers (quad core)	160	640	12 (3GHz)	7.7 (3GHz)	22	12	34.3	0.35	0.24	0.86-0.88

kW/rack: estimated or nominal claimed, not measured peak
TF/m^2: HP 2x 220c and Twin blades assuming 0.6x1.07m^2 racks
kW/m^2: not including cooling and service areas
Linpack TF/kW: BG/P from Top500 Nov 2008, SiCortex from company info,
Twin server from SGI ICE from Top500 Nov 2008
SGI Molecule: based on PRACE prototype offer (not concept presented at SC08)

ADVANCED COMPUTING RESEARCH LABORATORY

PRACE Prototype – Data movement

	Node (memory)					Link BW GB/s	TF/s / TB/s chassis	TF/s / TB/s C - C	TF/s / TB/s R - R
	GF	GB	GB/core	GB/s	GF/s / GB/s				
BG/P	13.6	2	0.5	13.6	1	0.425	32	43	51
HP 2x blades (HTN)	96	16	2	5.3	18	2.5			
SGI Molecule	3.2	2	1	4.2	0.76	0.125			
SiCortex	8.4	8	1.25	2.1	3.8	1.6			60
SiCortex 2H09	33.6	12	1	4.3?	7.7?	3.5?			121
Supermicro PRACE	230.4	32	1.33	8*6.4	4.5	5	46	46	46
Twin servers (quad core)	96	24	3	5.3?	18?	2.5	38		

HP 2x 220c: 4 DIMM slots/node, table assumes 4GB DIMMs
SiCortex: 2 DIMMs/socket, GB/s assume 533 MHz DDR2 (533/800 MHz DDR2 conflicting info from vendor)
2H09: 3 DIMM slots/node, table assumes 4GB DDR3 DIMMs @1066MHz, no vendor response
Supermicro: 4x4 DIMM slots/node
Twin servers: 3 DIMM slots/node assumed (Supermicro), 4GB DIMMs

ADVANCED COMPUTING RESEARCH LABORATORY

Memory

- Exascale system
  - CPUs 16 MW
  - **Memory 60 MW**
  - Interconnect 50 MW
  - I/O

SNIC/KTH/PRACE Prototype

Component	Power (W)	Percent (%)
CPUs	2,880	56.8
Memory 1.3 GB/core	**800**	**15.8**
PS	355	7.0
Fans	350	6.9
Motherboards	300	5.9
HT3 Links	120	2.4
IB HCAs	100	2.0
IB Switch	100	2.0
GigE Switch	40	0.8
CMM	20	0.4
Total	5,065	100.0

HPL observed: Max 4,647 Avg 4,625 W Stream observed: 3,620 W

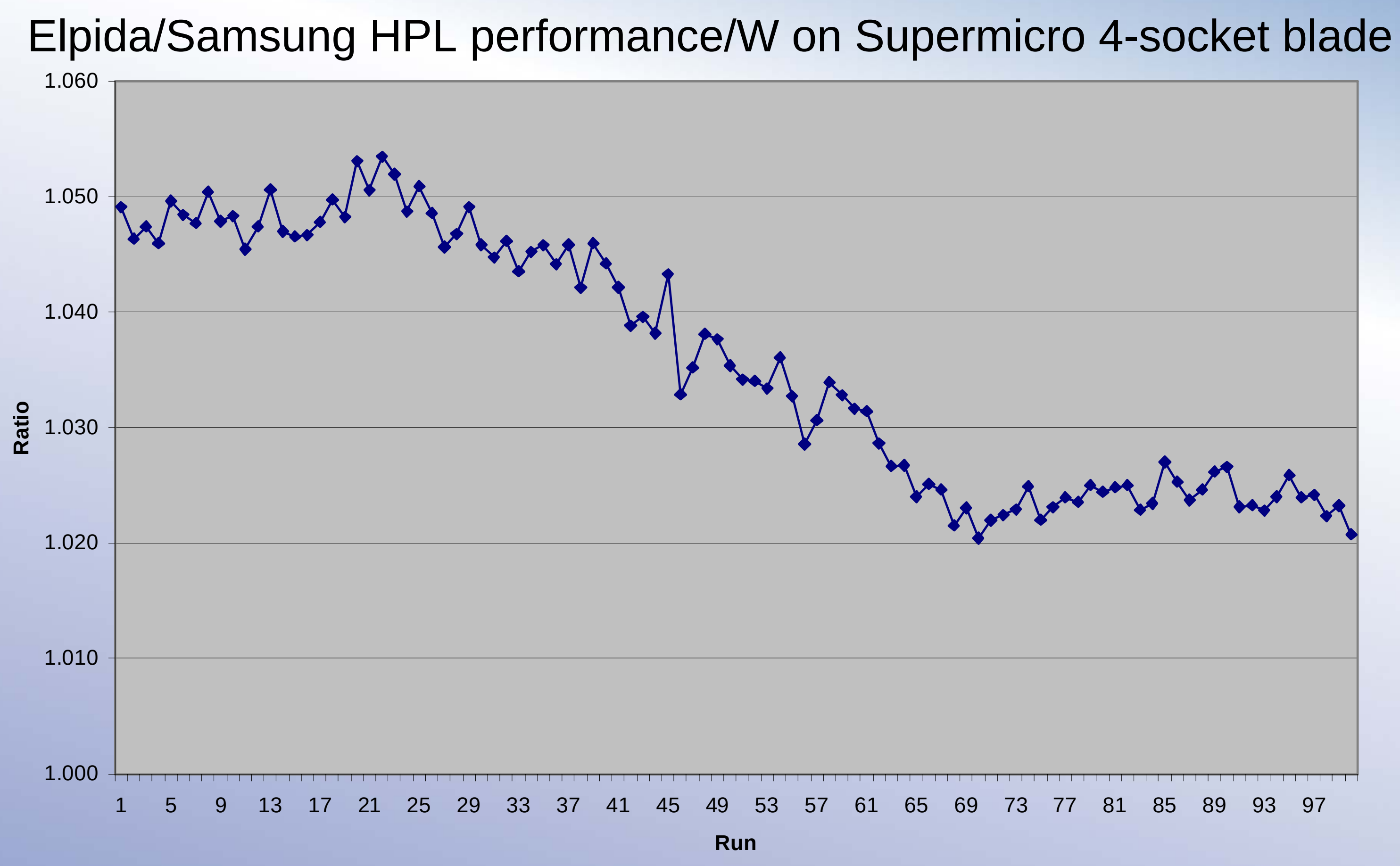
ADVANCED COMPUTING RESEARCH LABORATORY

Memory study

- Elpida, Hynix, Micron, Samsung power consumption for DIMMs estimated using public tools and published chip specs
- Measurements carried out with Elpida, Hynix and Samsung DIMMs (on “old” motherboard and chipset, Istanbul 75W ADP CPUs)

PDC Summer School,
Aug 26 2010
Lennart Johnsson

Elpida and Samsung relative HPL performance/W



ADVANCED COMPUTING RESEARCH LABORATORY

Memory selection

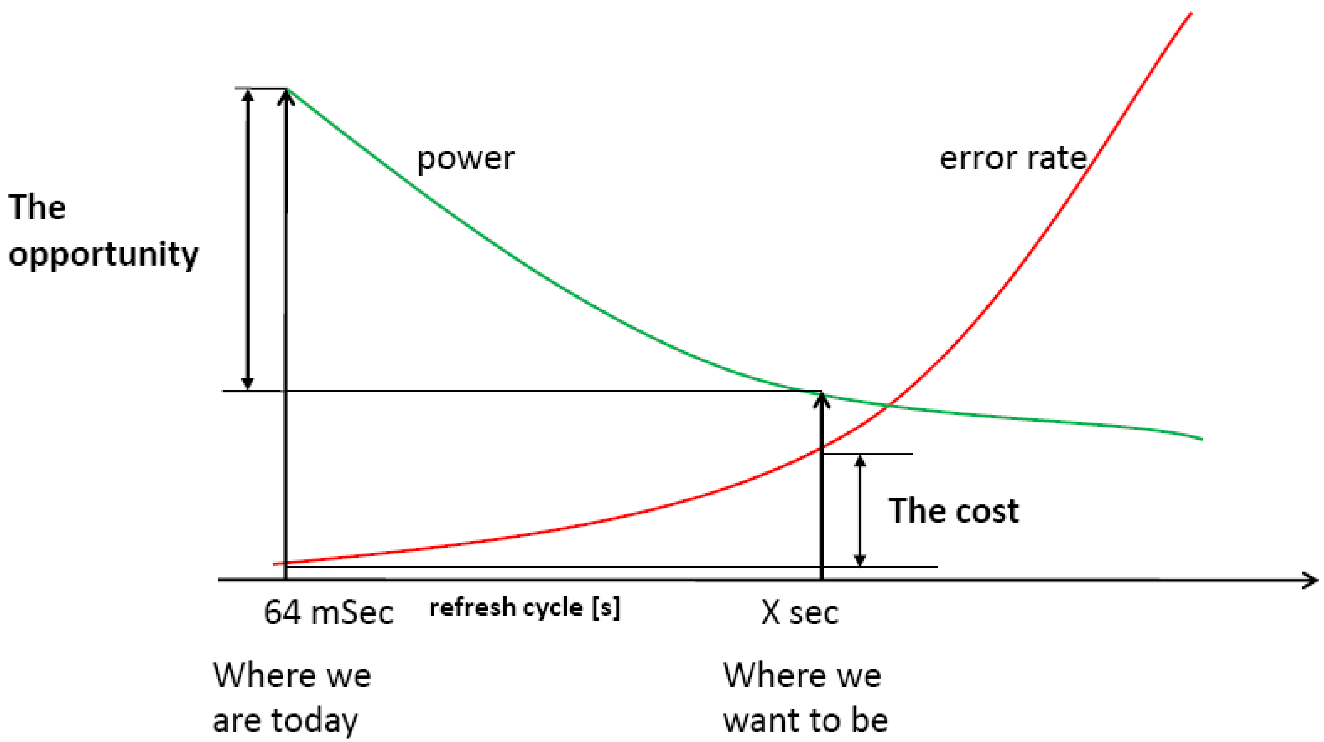
- Elpida vs Hynix on Phase-II motherboards
  - Hynix 97.6% power consumption of Elpida for HPL
  - Hynix 99.7% of Elpida HPL performance
  - Hynix 107.9% of Elpida Stream performance

Hynix selected

ADVANCED COMPUTING RESEARCH LABORATORY

Memory:
Reducing Power Consumption

Motivation: DRAM Refresh



If software is able to tolerate errors, we can lower refresh rates pretty drastically to save power

http://www.sigops.org/sosp/sosp09/slides/liu-slides-sosp09wip.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

Memory: Reducing Power Consumption

Flicker: Contributions

- **First software technique to intentionally lower hardware reliability for energy savings**
- Minimal changes to hardware – based on PASR mode in existing DRAMs
- No modifications required for legacy applications – incremental deployment
- **Reduced overall DRAM power by 20-25% with negligible loss of performance (< 1 %) and reliability across wide range of apps**

http://www.sigops.org/sosp/sosp09/slides/liu-slides-sosp09wip.pdf

ADVANCED COMPUTING RESEARCH LABORATORY

HPL Summary

Efficiency

Reference platform	91%
Reference platform + Clearspeed	61.8%
Reference platform + GPU	52.5%
SNIC/KTH prototype	79% (preliminary)
eQPACE (Cell)	79.9%

ADVANCED COMPUTING RESEARCH LABORATORY

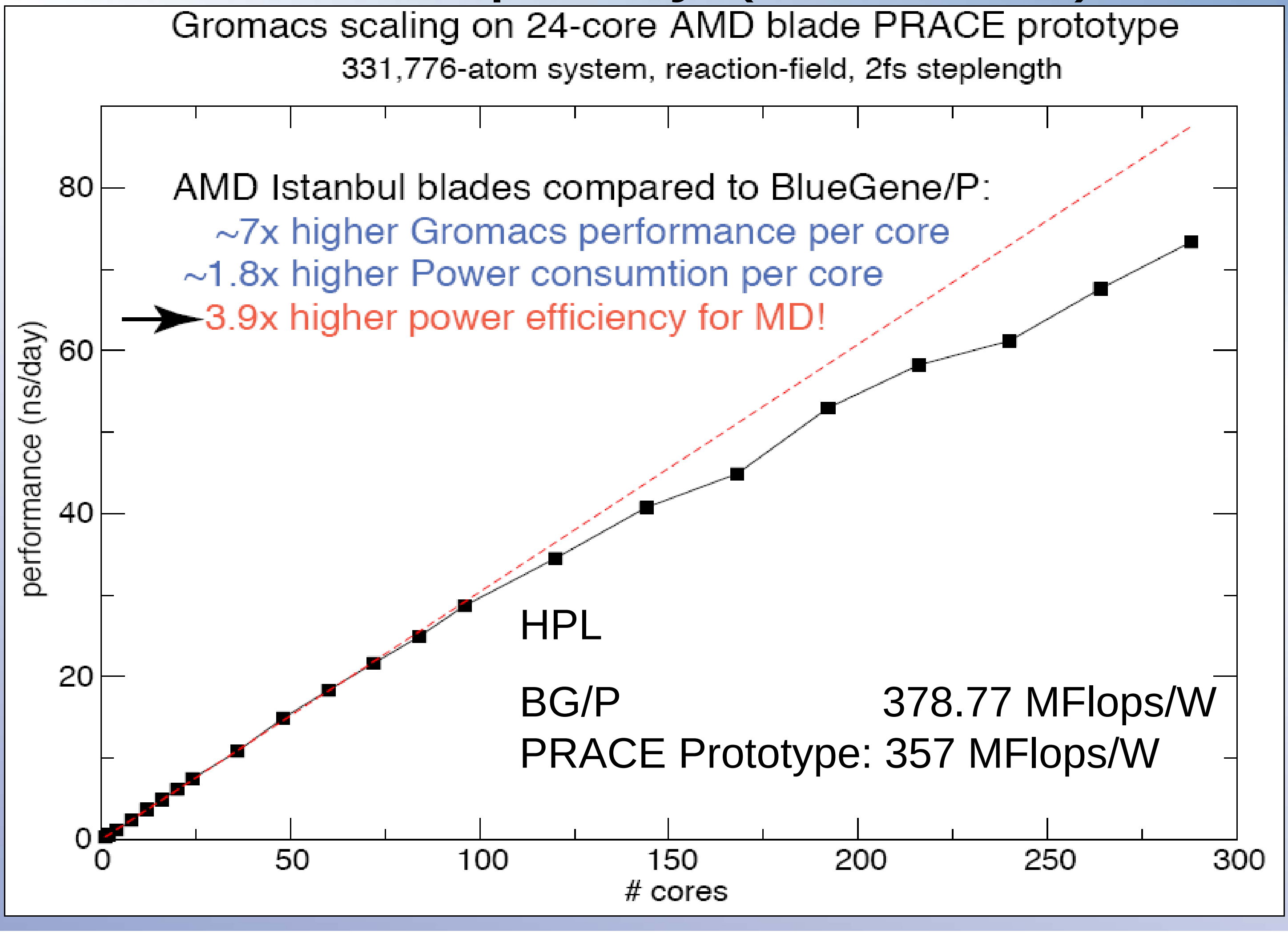
HPL Summary

Power Efficiency

Reference platform	240 MF/W
Reference platform + Clearspeed	326 MF/W
Reference platform + GPU	270 MF/W
SNIC/KTH prototype	344 MF/W(prelim.)
eQPACE (Cell)	773 MF/W

ACRL
ADVANCED COMPUTING RESEARCH LABORATORY

Standard CPU (AMD) vs Low Frequency (PowerPC)



Energy Aware Algorithms and Software

Examples: Scheduling

"Most forms of renewable energy are not reliable – at any given location. But Canada's [Green Star Network] aims to demonstrate that by allowing the computations to follow the renewable energy across a large, fast network, the footprint of high-throughput computing can be drastically reduced."

International Science Grid This Week, April 4, 2010

Move load to data centers based among other things cooling capability (example data center near Saint-Ghislain, Belgium)

Energy Aware Algorithms and Software

Examples: Scheduling

Research

CNet, Aug 18, 2009,
Energy-aware Internet routing coming soon

Cutting the Electric Bill for Internet-Scale Systems,
Asfandyar Qureshi, MIT CSAIL
Rick Weber, Akamai Technologies,
Hari Balakrishnan, MIT CSAIL,
John Guttag
MIT CSAIL, Bruce Maggs,
Carnegie Mellon University

Presented at SigComm 2009

Bounded Slow Down Threshold Driven Parallel Job Scheduling for Energy Efficient HPC centers

Maja Etinskiy, Julita Corbalany;z,
Jesus Labartay, Mateo Valero

Barcelona Supercomputing Center and
Department of Computer Architecture
Technical University of Catalonia

capinfo.e.ac.upc.edu/PDFs/dir10/file003490.pdf

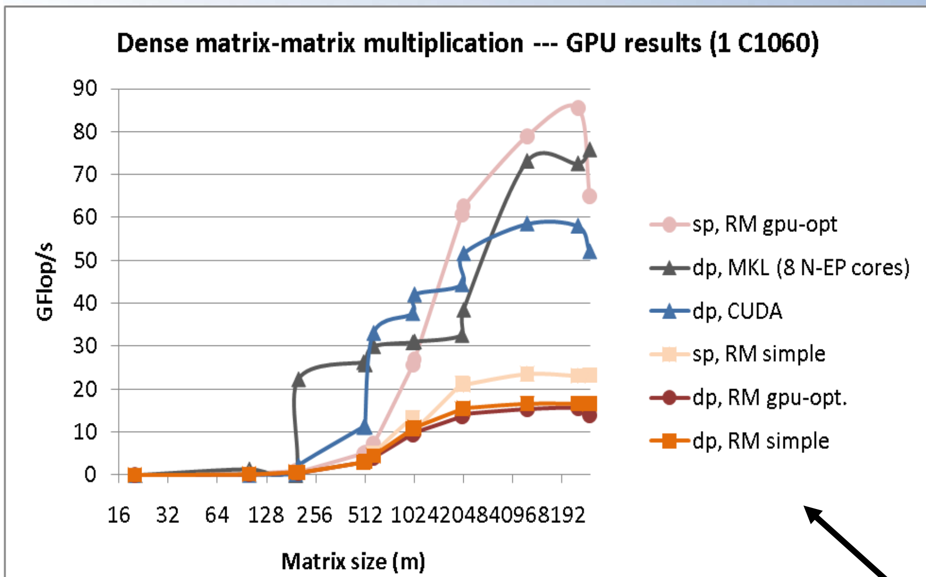
Energy Aware Algorithms and Software

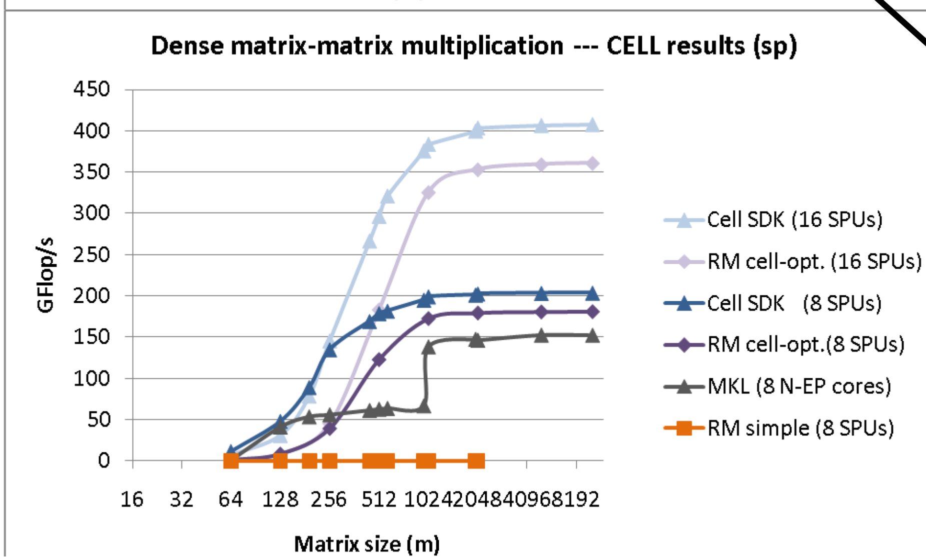
Efficiency of many codes <10%!!!

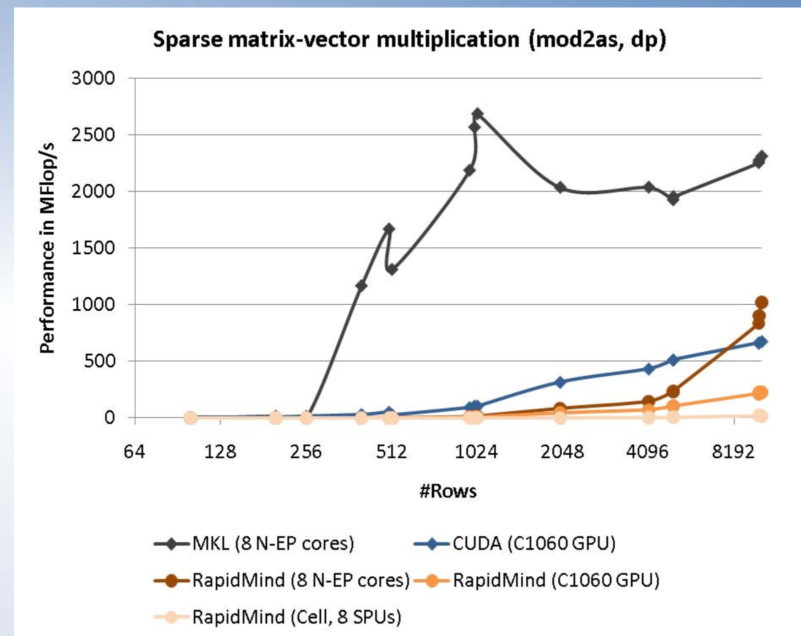
Great opportunity!!!

ADVANCED COMPUTING RESEARCH LABORATORY

Programming of other PRACE prototypes







mod2as

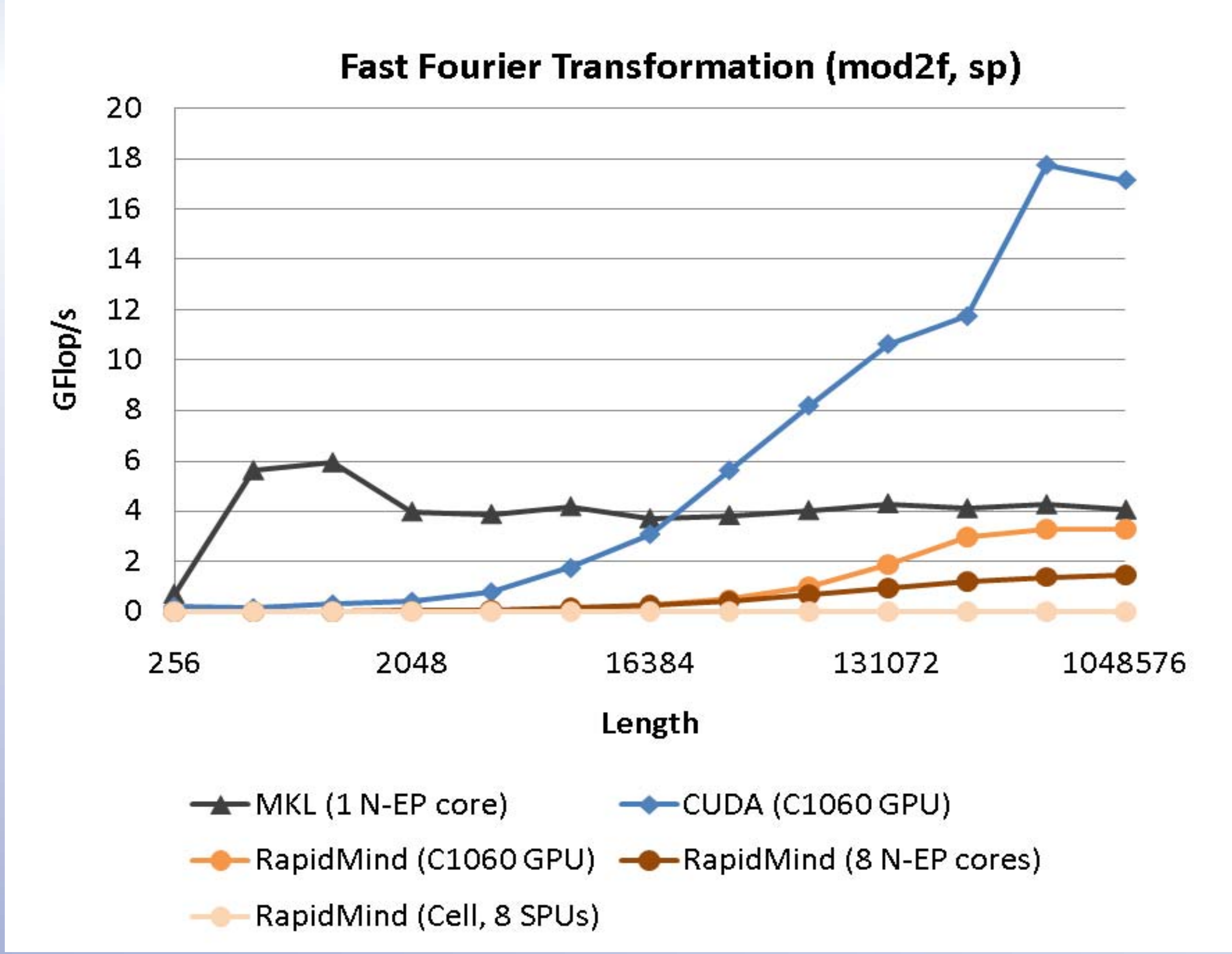
mod2am on Nehalem-EP vs nVidia C1060

mod2am on Nehalem-EP vs (CELL)

PRACE EU Deliverable D8.3.2

ADVANCED COMPUTING RESEARCH LABORATORY

Programming of other PRACE prototypes



mod2f on Nehalem-EP vs Cell vs nVidia C1060

PRACE EU Deliverable D8.3.2

Programming PRACE prototypes

Programming heterogeneous systems is still difficult and tools need significant improvement

SNIC/KTH PRACE Prototype

Ready for research in energy efficient computing.
For access contact:

Daniel Ahlin, dah@pdc.kth.se

Gilbert Netzer, noname@pdc.kth.se

ADVANCED COMPUTING RESEARCH LABORATORY

Summary

- Climate change is real!
- Reducing rate of change is urgent!
- ICT has the potential to improve energy efficiency of other sectors 5 – 10 fold its own energy consumption!
- Significant increase in renewable energy sources poses new challenges (unreliable)
- Great progress has been made in improving infrastructure
- Many opportunities and challenges in computer systems design and operations
- Software challenges at least as severe as ever before
- Measurement “standards” can help drive energy efficiency