

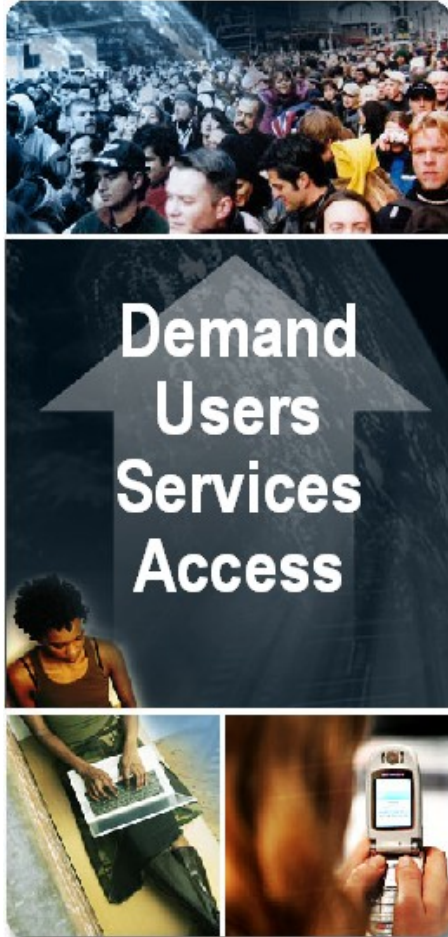


# The End of Redundancy

**Alan Wood – Sun Microsystems**  
**May 8, 2009**



# Growing Demand, Shrinking Resources



“By 2008, 50% of current data centers will have insufficient power and cooling capacity to meet the demands of high-density equipment.”  
---Gartner \*

Energy bills traditionally have accounted for less than 10% of an overall IT budget but soon could account for more than half.” --- Gartner \*

IDC estimates that 60% of data centers are already out of power, space, and cooling.



\*Source: <http://www.gartner.com/it/page.jsp?id=499090>

# EPA Report to Congress on Server and Data Center Energy Efficiency

- Data center energy more than doubled from 2000 to 2006.
- The power and cooling infrastructure accounts for 50% of data center total energy consumption.
- The energy used by the nation's servers and data centers in 2006:
  - > 61 billion kilowatt-hours (kWh)
  - > 1.5% of total U.S. electricity consumption
  - > Total electricity cost of about \$4.5 billion
  - > Equal to 5.8 million average US Households



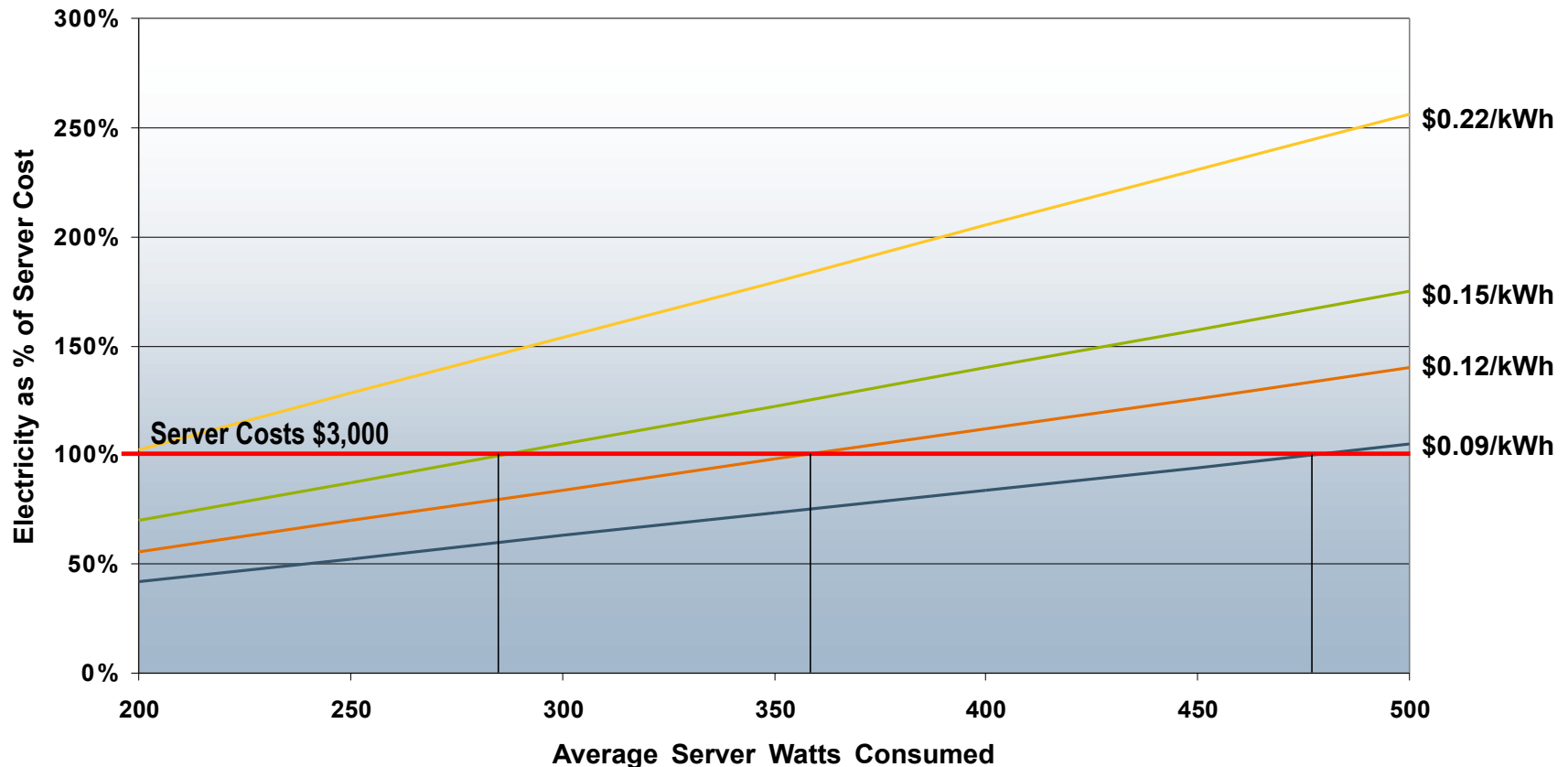
Source: EPA Report to Congress on Server and Data Center Energy Efficiency

# Agenda

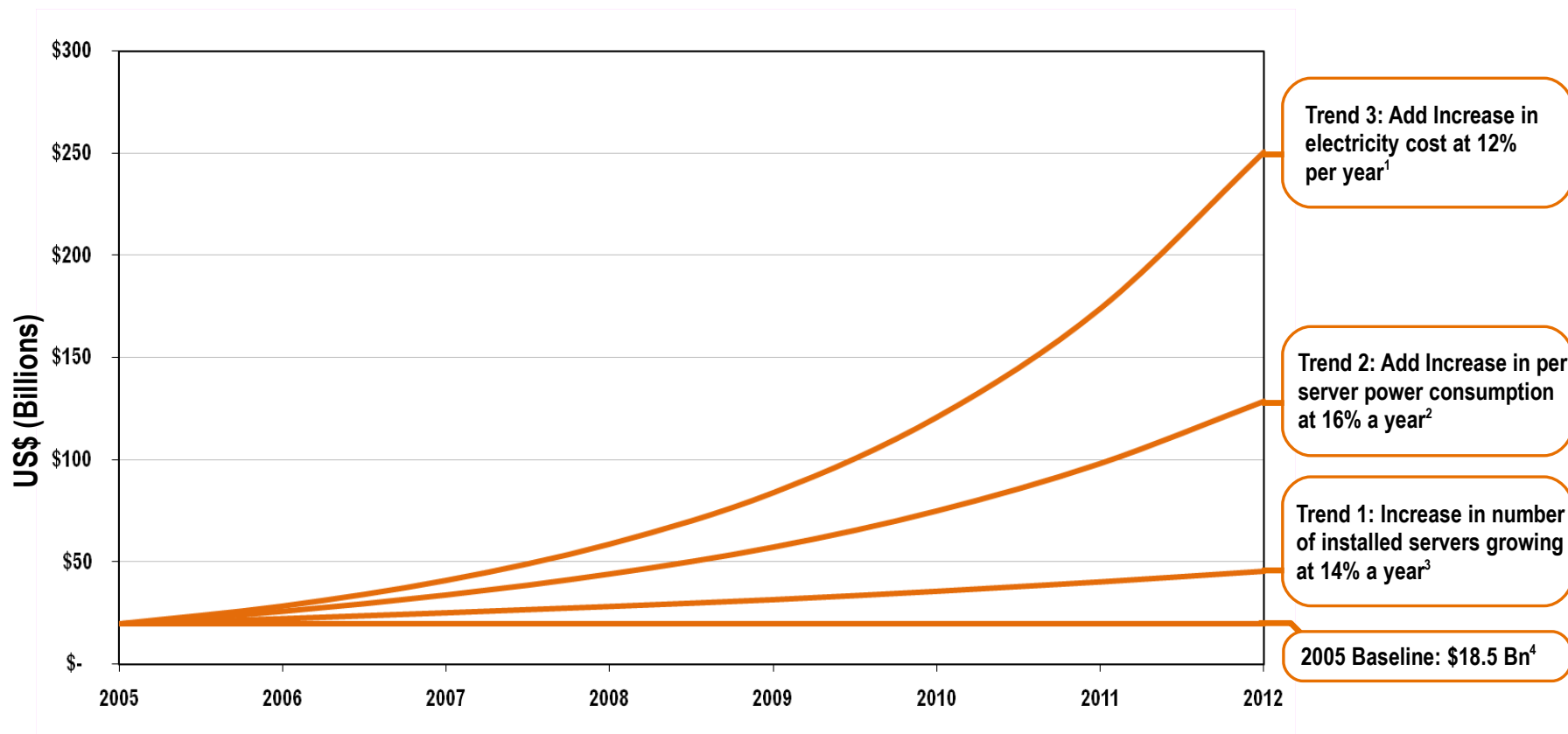
- Power costs more than hardware (Lifetime volume server power consumption costs more than the server)
- What industry is doing to lower power consumption
- The potential impact on dependability
- How the dependability community should respond

## When OPEX Exceeds CAPEX

### Electricity OPEX as % of Server CAPEX Over Server Lifetime of 4 Years



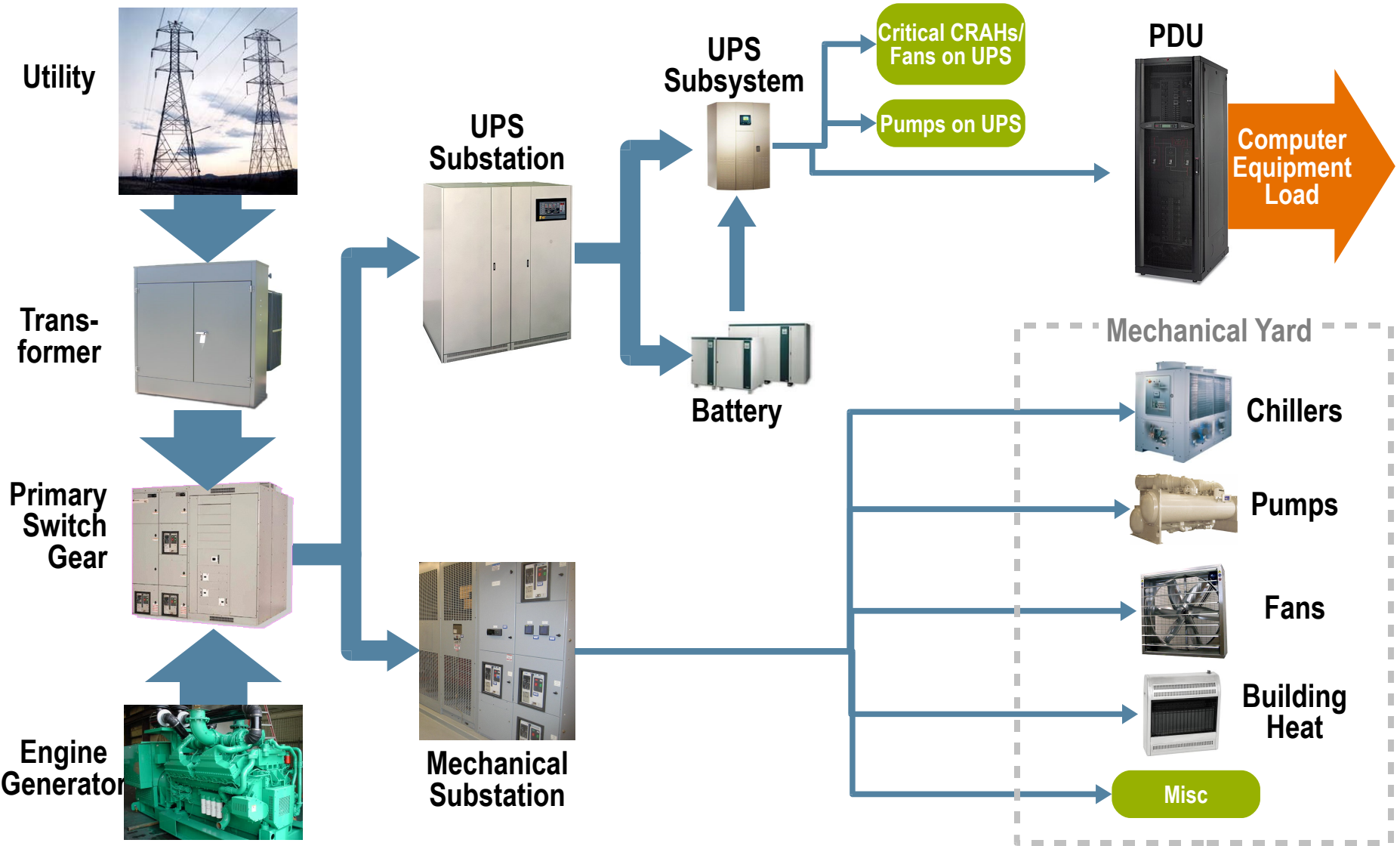
## The Cubing Effect



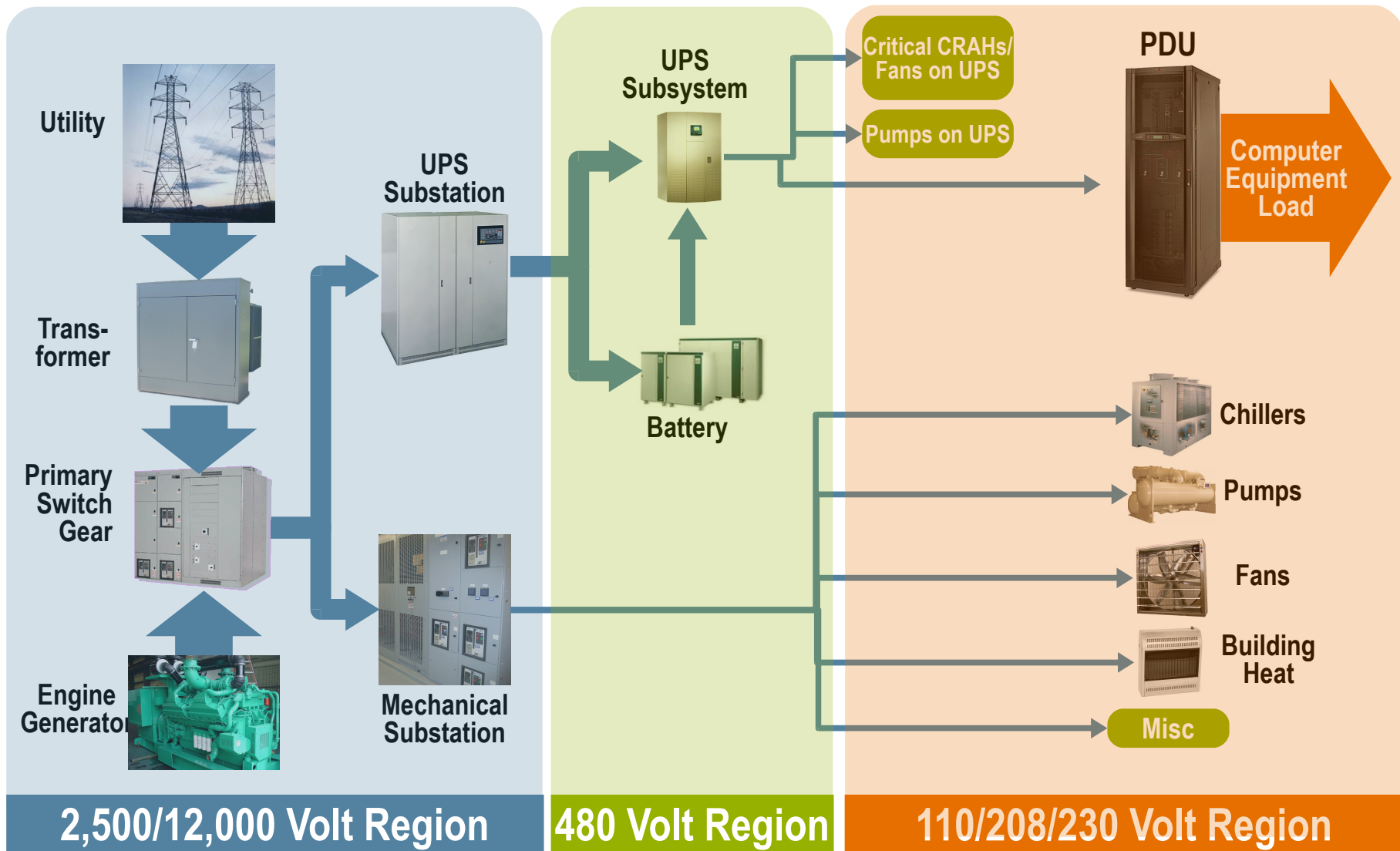
**By 2012 data center power consumption costs could grow to \$250B worldwide – demanding proactive energy management solutions**

1.U.S. Energy Information Administration (www.eia.doe.gov)  
2.Sun primary research  
3.IDC#34867 U.S. and Worldwide Server Installed Base 2006-2009 Forecast (February 2006)  
4.IDC Worldwide Server Power and Cooling Expense 2006-2010 Forecast

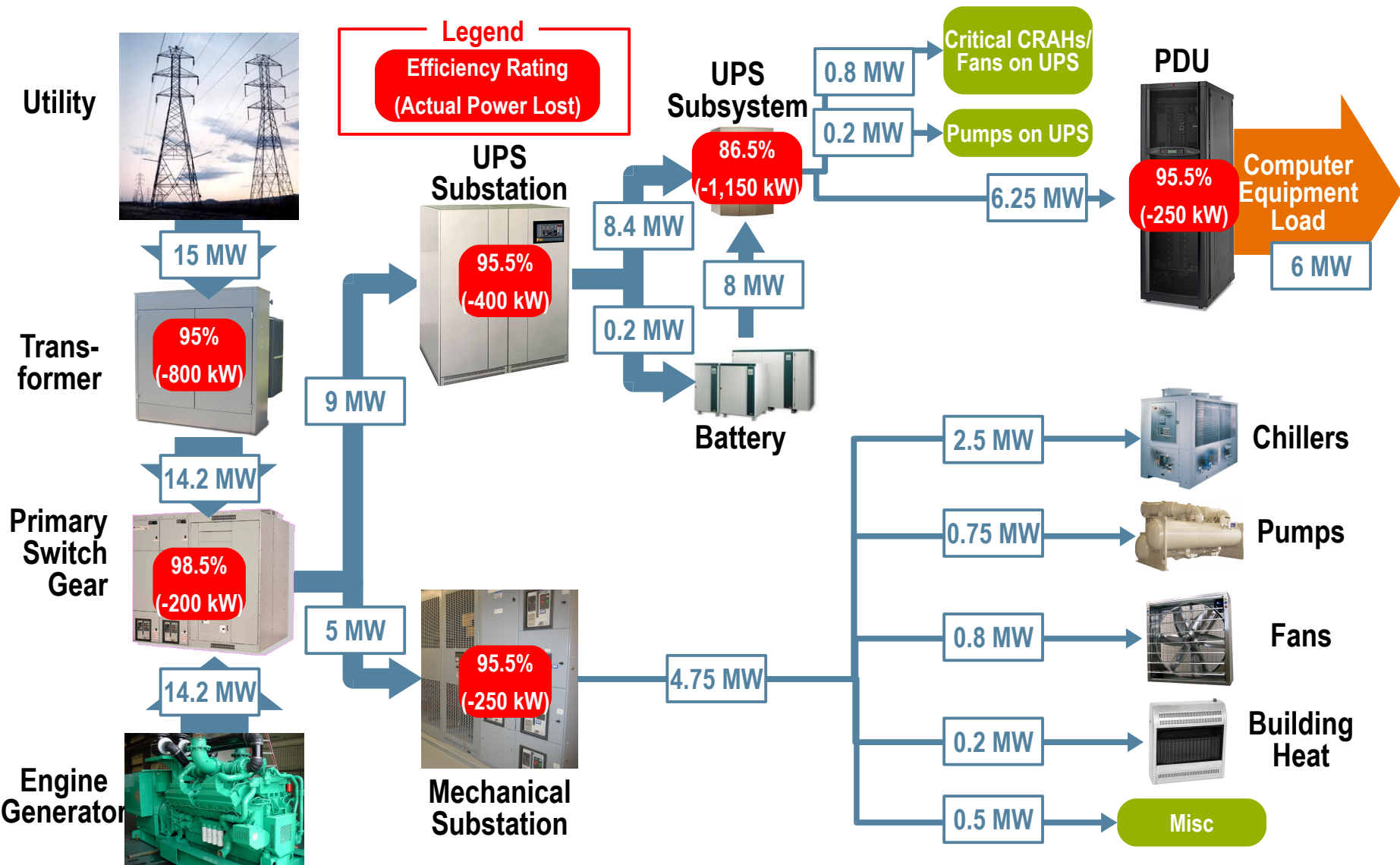
# Data Center Power Flow



# Data Center Voltage Regions



## Typical Power Flows in a 15MW Data Center



# Data Center Efficiency Metrics

- DCIE: Data Center Infrastructure Efficiency

$$\frac{\text{Power That Reaches the IT Equipment}}{\text{Total Power to the Data Center}}$$

- PUE: Power Use Efficiency

$$\frac{\text{Total Power to the Data Center}}{\text{Power That Reaches the IT Equipment}}$$

- Emissions Factor

> CO<sub>2</sub> emission factor = the average emission rate of carbon dioxide for a given source of energy

$$\frac{\sum_{\text{All Sources}} (\text{CO}_2\text{EF for each source}) \times (\text{kWh of energy delivered by that source})}{\sum (\text{Total kWh delivered by the utility})}$$

# Eco in the Data Center

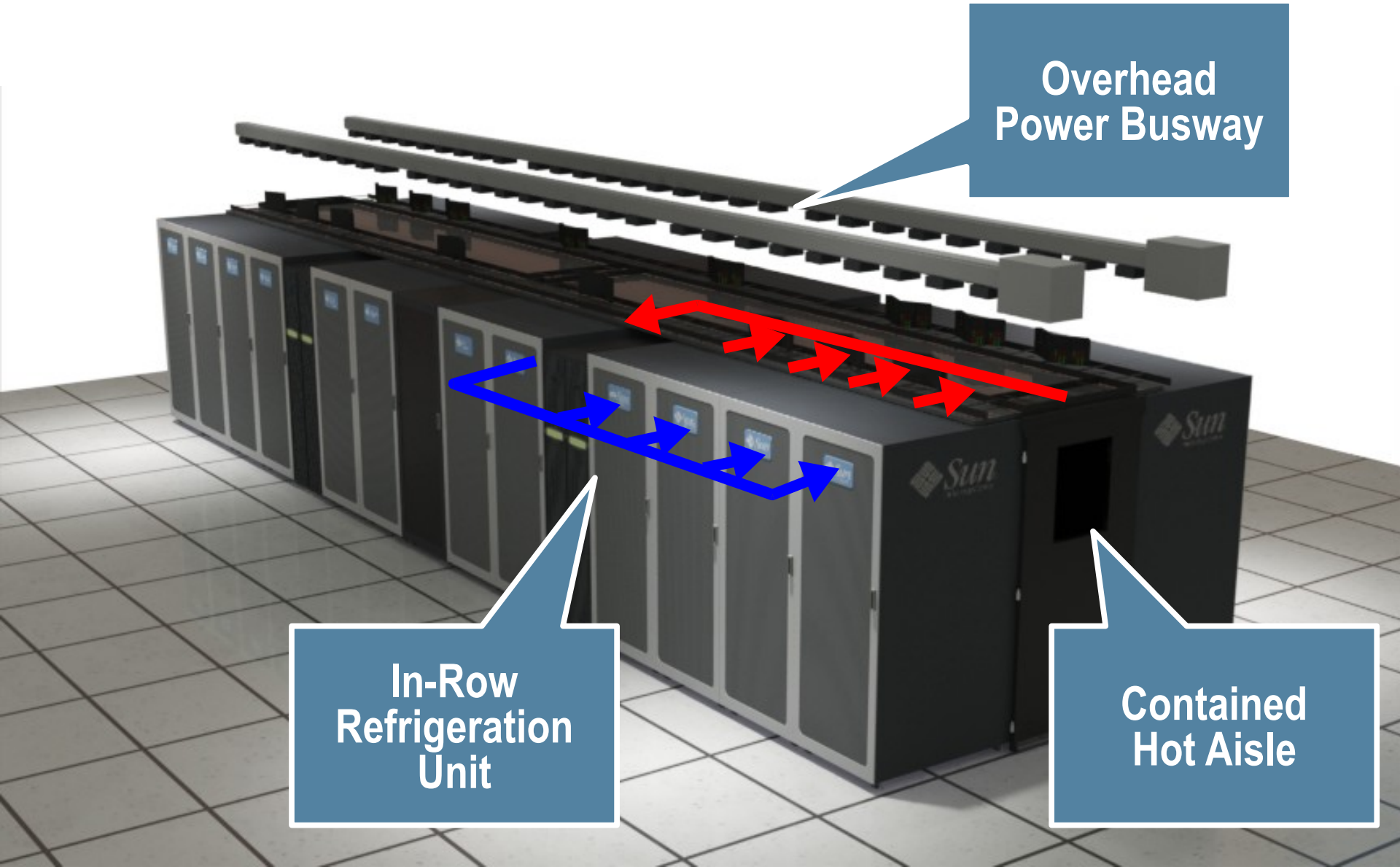
## Reconciling Conflicting Goals

- Last Year... AVAILABILITY AND PERFORMANCE MATTER
- Last Month... ENERGY EFFICIENCY MATTERS
- The Truth... THEY ALL MATTER
- This is a complex relationship, not an either/or situation
- IT equipment serves critical functions

# What is Industry Doing?

- At the data center level
  - > Economizers, air-side and water-side
  - > Better data center layouts
    - > More efficient power and cooling
    - > Consolidation/Virtualization
  - > Turn off everything not being used
  - > Throttle to match required performance
- At the server level
  - > Throttle/idle to match required performance
    - > New architectures to allow idle CPUs and memory
  - > Fan and power supply efficiency and settings
  - > New architectures, e.g., flash as a memory tier

# Pods with Hot Aisle Containment



# What is Industry Doing? - 2

- At the chip level
  - > Lower power, e.g., multi-core, multi-threaded microprocessors
  - > Decreased leakage current
  - > Throttle to match required performance (idle/sleep states)
- At the OS level
  - > Virtualization
  - > Power-aware resource management and scheduling
- At the application level
  - > Very limited research

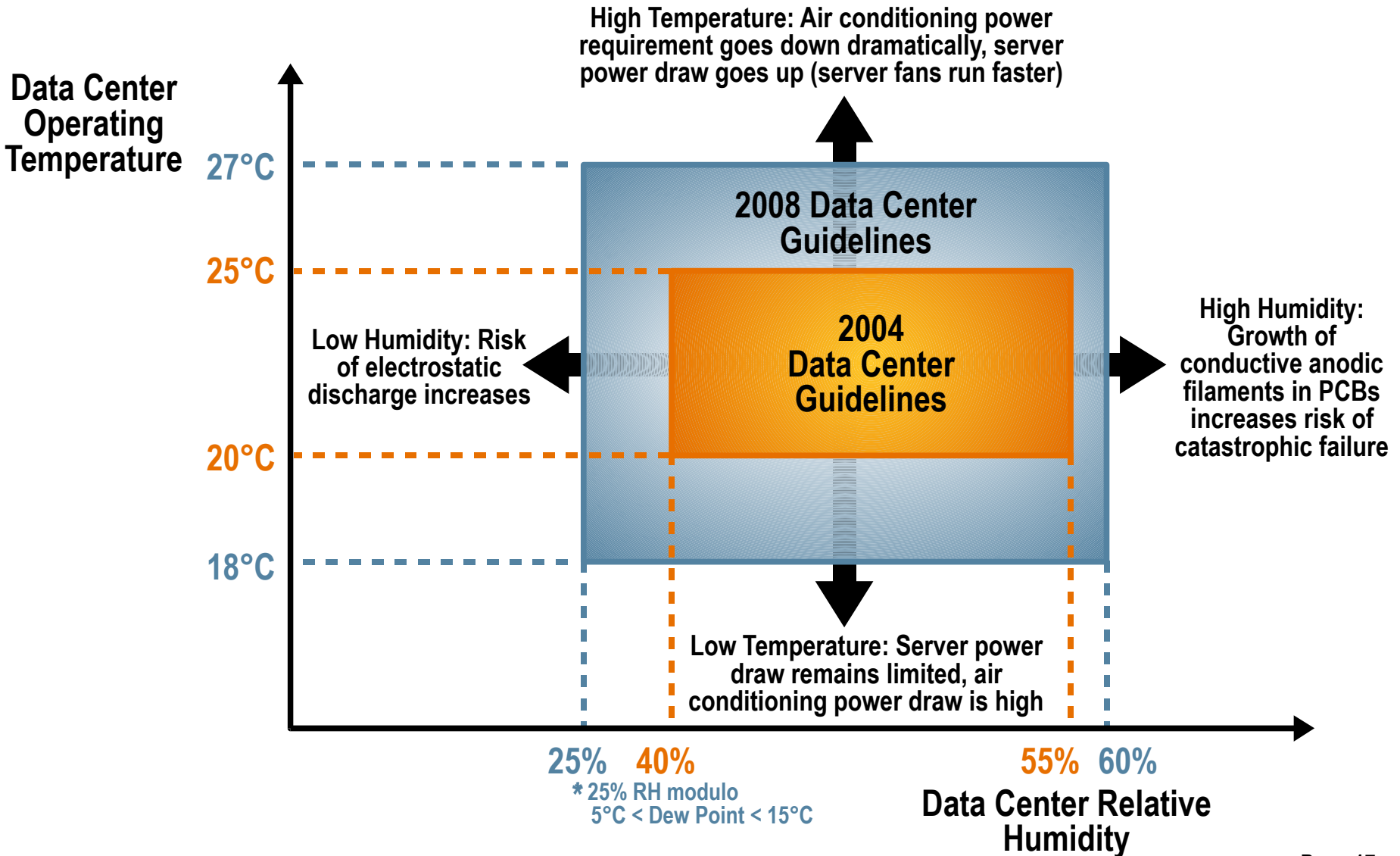
# The Energy Efficient Stack

<b>Monitoring, Management, and Services</b>	Energy efficiency through giving customers visibility into actual server power consumption, managing power draw across multiple servers, and providing optimization services
<b>Operating System</b>	Energy efficiency through advising hypervisor of application workload requirements and supporting hypervisor directives to localize allocated resources
<b>Virtualization Layer</b>	Energy efficiency through control of primary allocation of system resources (CPU, memory, I/O) to multiple guest OSs depending on need
<b>Platform</b>	Energy efficiency through control of memory DIMMs, I/O links, disks, and intelligence in firmware service processors
<b>Microprocessor</b>	Energy efficiency through control of clock frequency, cores, threads, instruction pipeline and memory interfaces

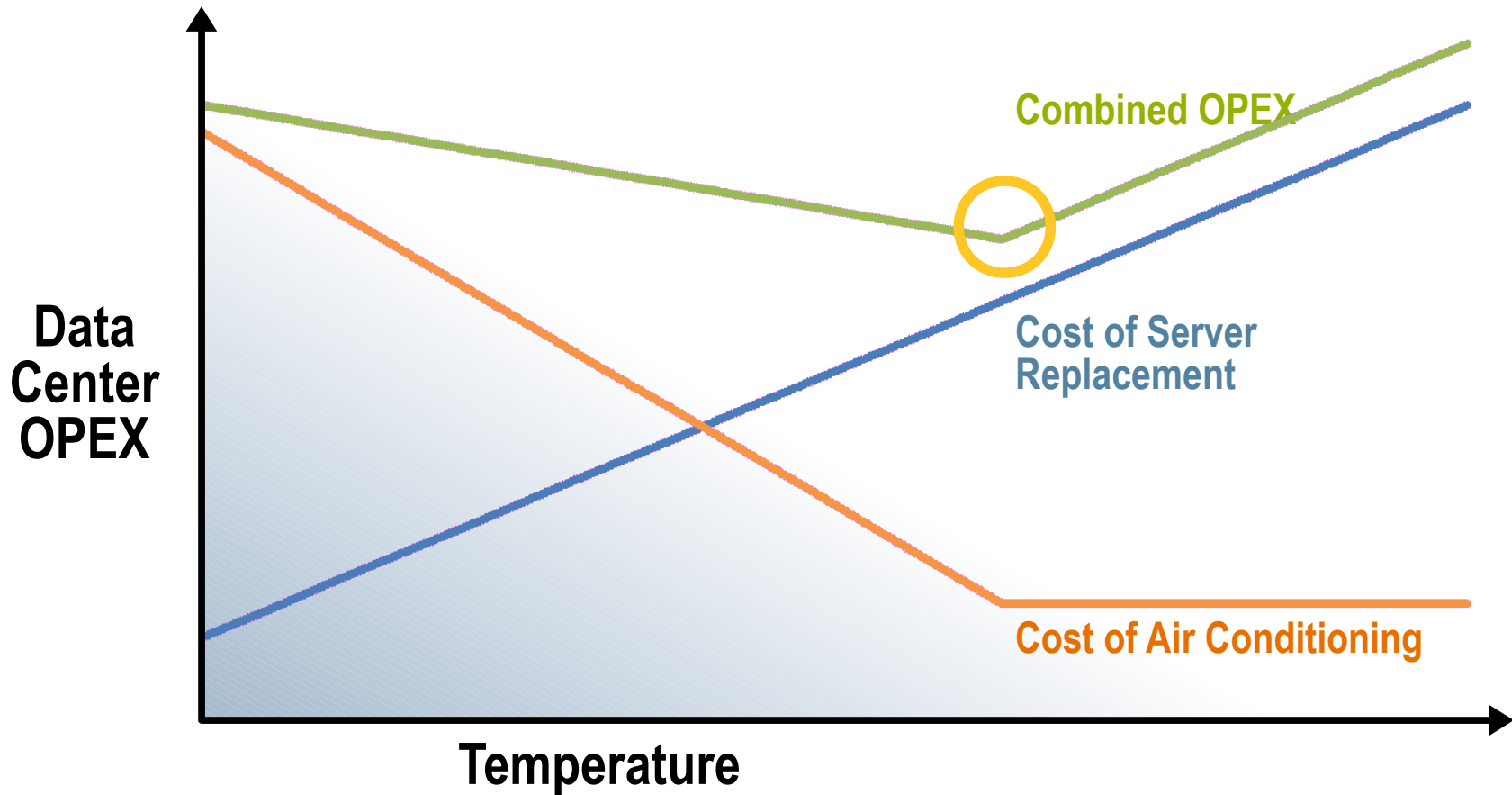
# Dependability Impact

- Data centers operate at higher temperature and humidity
- Increased power and temperature cycling
  - > From turning off equipment
  - > From throttling performance
- Lower voltage/power increases soft error vulnerability
- Many new states to verify
  - > Reduced power states
  - > Software control of hardware
- Derating, hot spots, wearout, ... may all change

## 2008 ASHRAE Guidelines



# Minimize Total Cost



# Dependability Community Guidance From Industry

- Don't hurt performance
  - > A few percent at most
- Don't use much chip area
  - > 10% max from SELSE 2 (includes arrays)
- And now, don't use any extra power either
  - > And allow throttling

**And, by the way, our customers still want  
the same levels of dependability**

# The End of Redundancy?

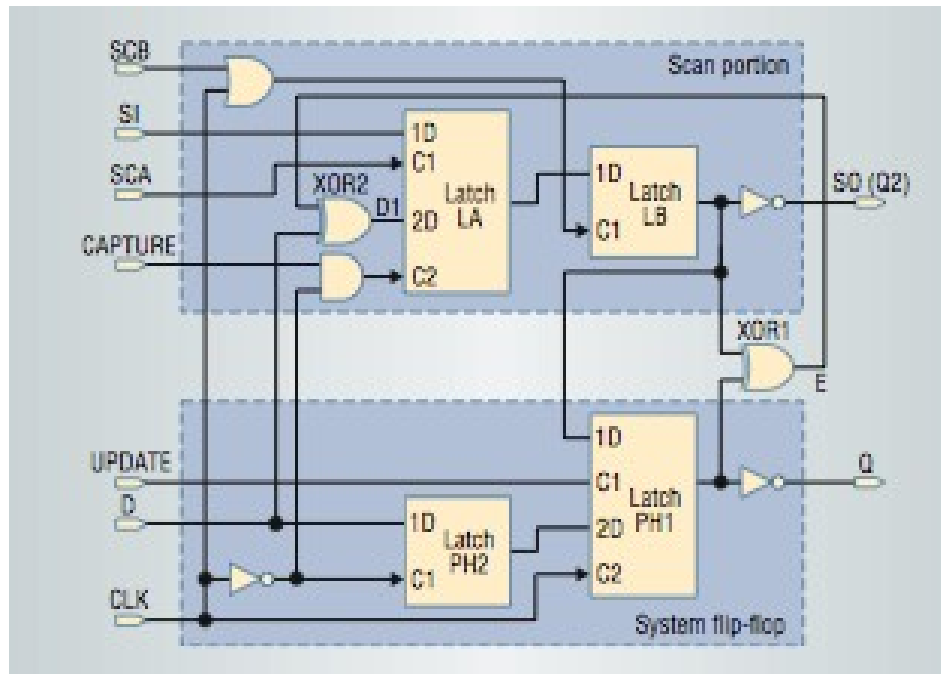
- Many classical fault-tolerance techniques assume redundancy
- Active redundant hardware consumes too much power
  - > Standby redundancy consumes too much power today (leakage current)
  - > Even non-powered redundant hardware takes space that equates to power
  - > Fans and power supplies may be exceptions

# How Should the Dependability Community Respond?

- Focus on techniques that minimize power
  - > Information redundancy (on larger blocks of data)
  - > Self-checking logic, e.g., state machines
  - > Assertion checking
- Tradeoff power consumption with recovery latency
  - > Retry instead of active redundancy/in-line recovery
- Consider new architectures
  - > Store data in flash instead of DRAM/disk
  - > Independent electrical zones on chips

# Example Low Power, Self-Checking Logic

- Uses existing scan chain logic

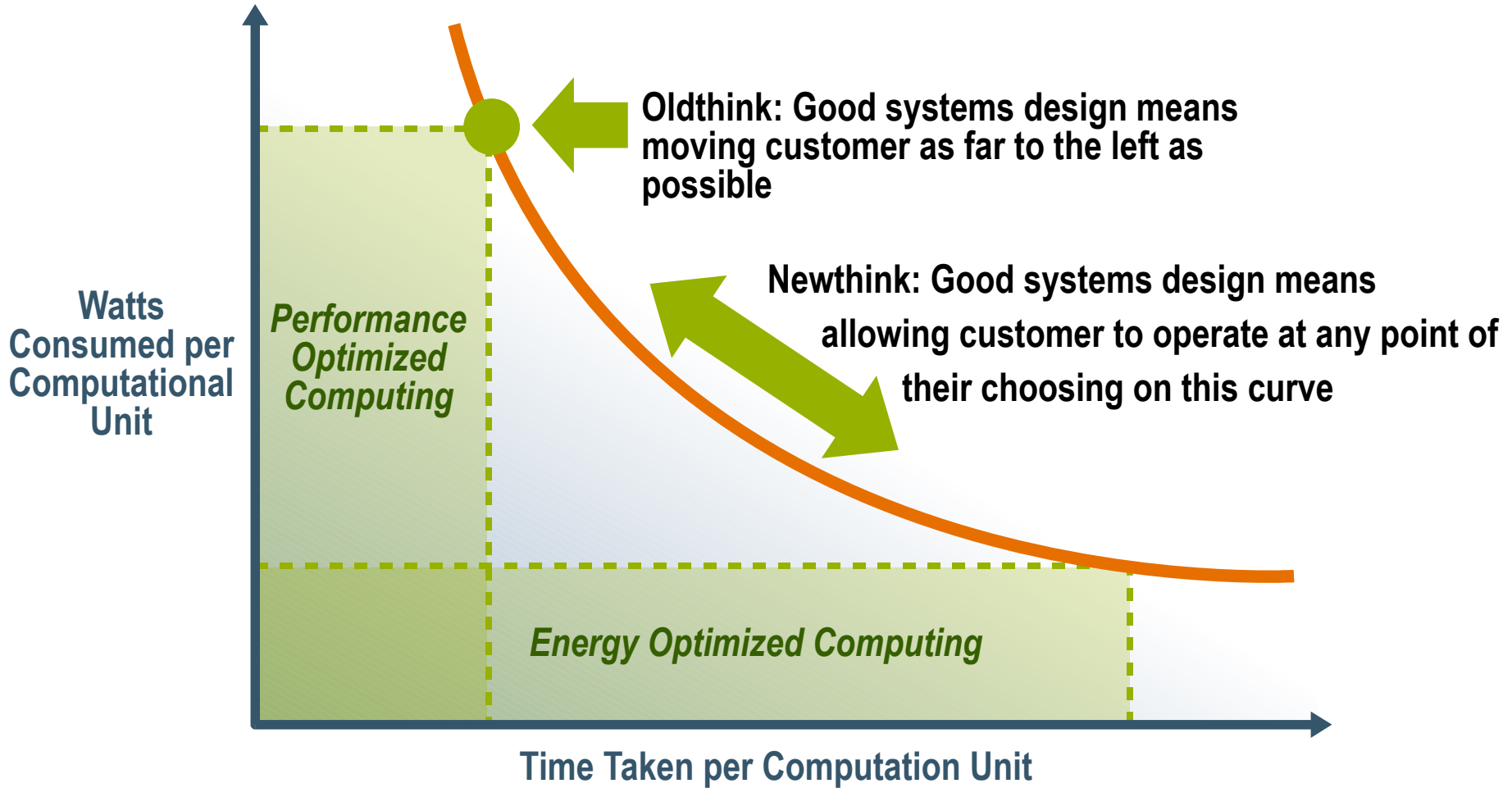


Reference: S. Mitra, et al, "Robust System Design with Built-in Soft-Error Resilience", IEEE Computer, Feb., 2005, pp 43-52

# How Should the Dependability Community Respond? - 2

- Learn from low-power computing – laptops, embedded systems
  - > Laptops have lots of power cycles but still work
- Do it in software
  - > Reliable protocols, end-to-end checksums, process pairs
- Power-aware middleware and applications
- Focus on the entire system

# Newthink vs. Oldthink in System Design



# The Good News

- What a great research opportunity!
- What a great opportunity to have a positive impact on society!
- What a great way to recruit environmentally-conscious engineers!