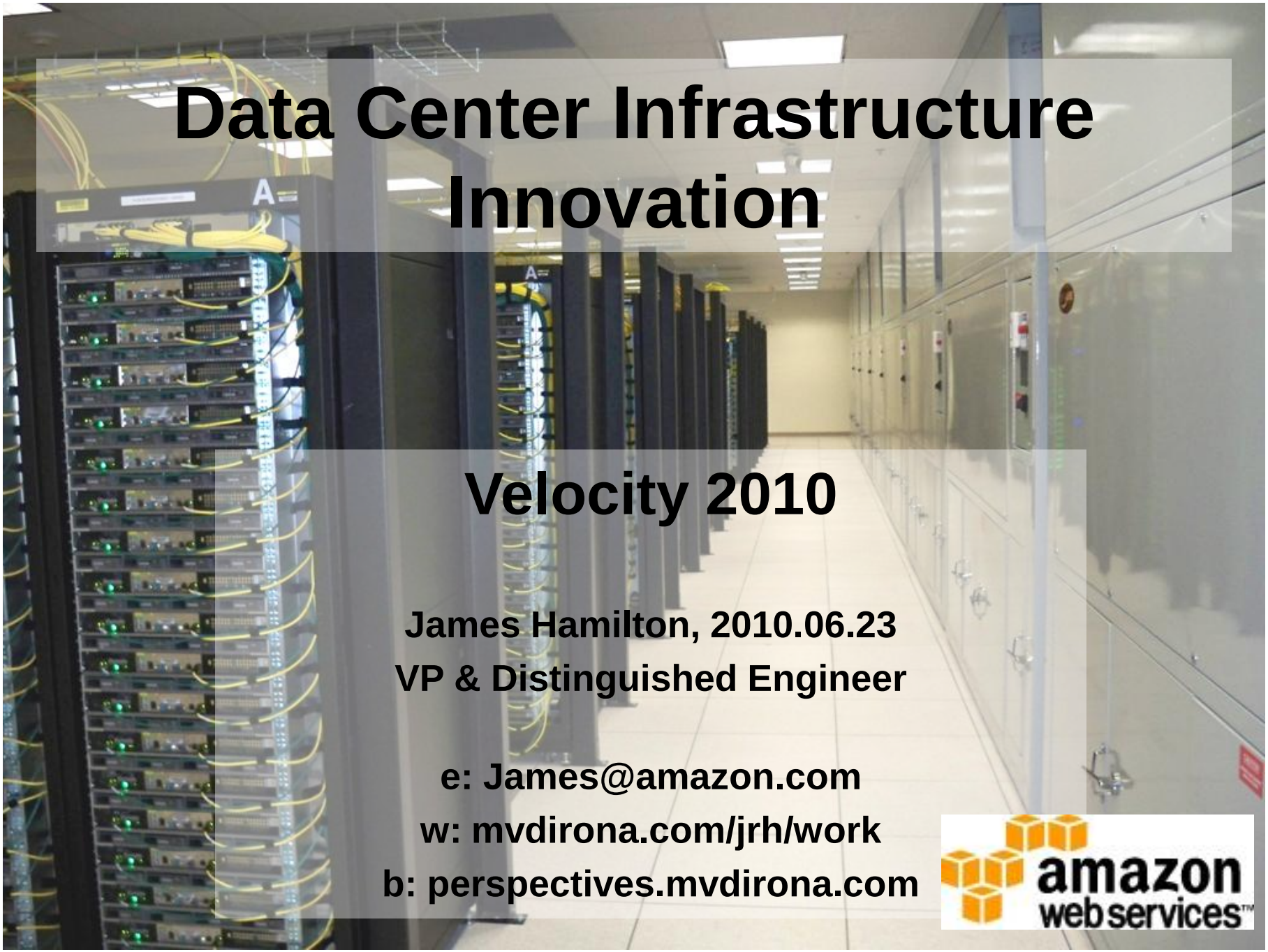


Data Center Infrastructure Innovation

Velocity 2010

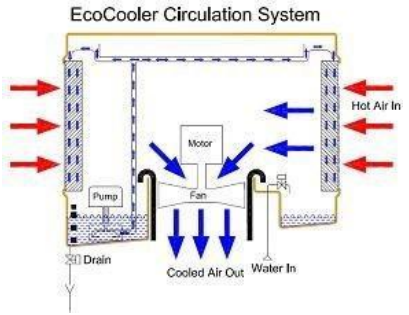
James Hamilton, 2010.06.23
VP & Distinguished Engineer

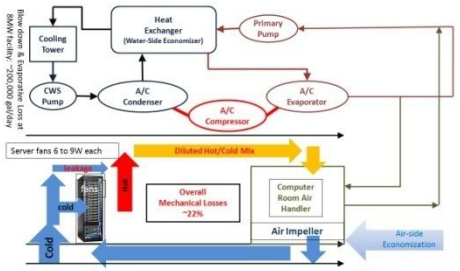
e: James@amazon.com
w: mvdirona.com/jrh/work
b: perspectives.mvdirona.com

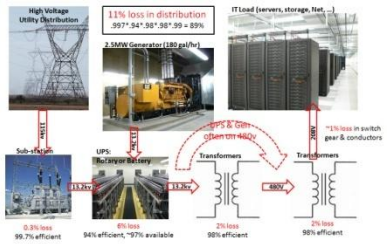
amazon web services™

Outline

- Quickening Pace of Datacenter Infrastructure Innovation
- Where does the money go?
- Power Distribution Infrastructure
- Mechanical Systems
- Sea Change in Net Gear
- Server Innovations







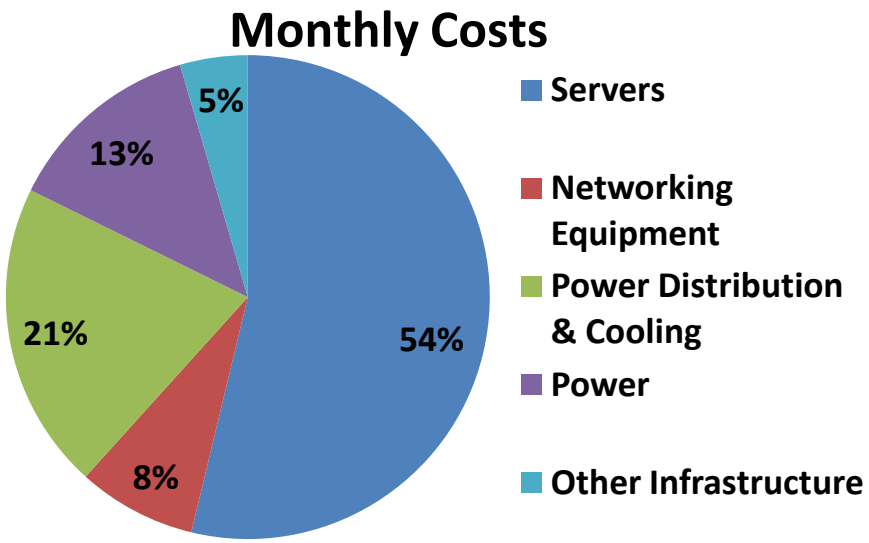
Talk does not necessarily represent positions of current or past employers

Pace of Innovation

- Datacenter pace of innovation increasing
  - Driven by cloud service providers and very high scale internet applications like search
  - Cost of datacenter & H/W infrastructure dominates
  - Not just a cost center
- High focus on infrastructure innovation
  - Driving down cost
  - Increasing aggregate reliability
  - Reducing resource consumption footprint

Where Does the Money Go?

- **Assumptions:**
  - Facility: ~$88M for 8MW critical power
  - Servers: 46,000 @ $1.45k each
  - Server power draw at 30% load: 80%
  - Commercial Power: ~$0.07/kWhr
  - Power Usage Effectiveness: 1.5



3yr server, 4yr net gear, & 10 yr infrastructure amortization

- **Observations:**
  - 34% costs functionally related to power (trending up while server costs down)
  - Networking high at 8% of costs & 19% of total server cost (many pay more)

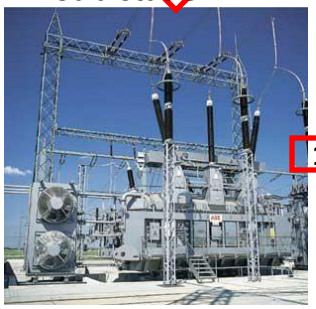
Updated from: http://perspectives.mvdirona.com/2008/11/28/CostOfPowerInLargeScaleDataCenters.aspx

Power Distribution

High Voltage Utility Distribution

115kv

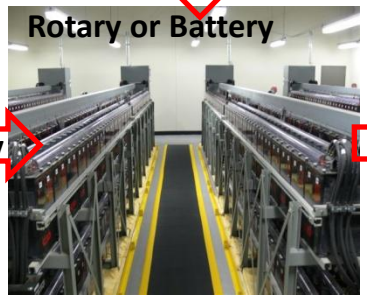
Sub-station



0.3% loss
99.7% efficient

13.2kv

UPS:
Rotary or Battery



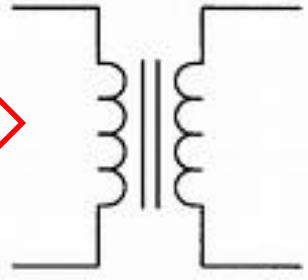
6% loss
94% efficient, ~97% available

Generators

13.2kv

13.2kv

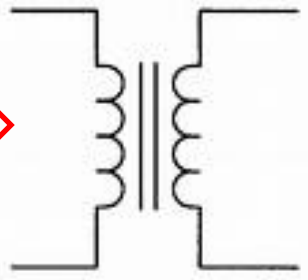
Transformers



2% loss
98% efficient

480V

Transformers



2% loss
98% efficient

208V

IT Load (servers, storage, Net, …)

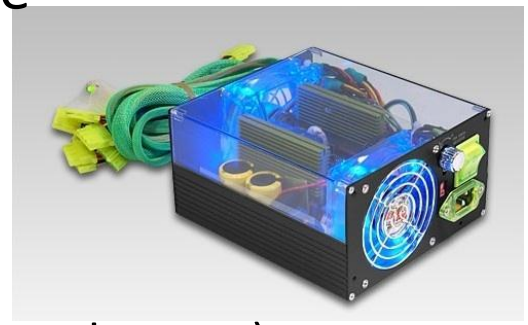
UPS & Gen
often on 480v

~1% loss in switch gear & conductors

11% lost in distribution
.997*.94*.98*.98*.99 = 89%

Power Distribution Efficiency Summary

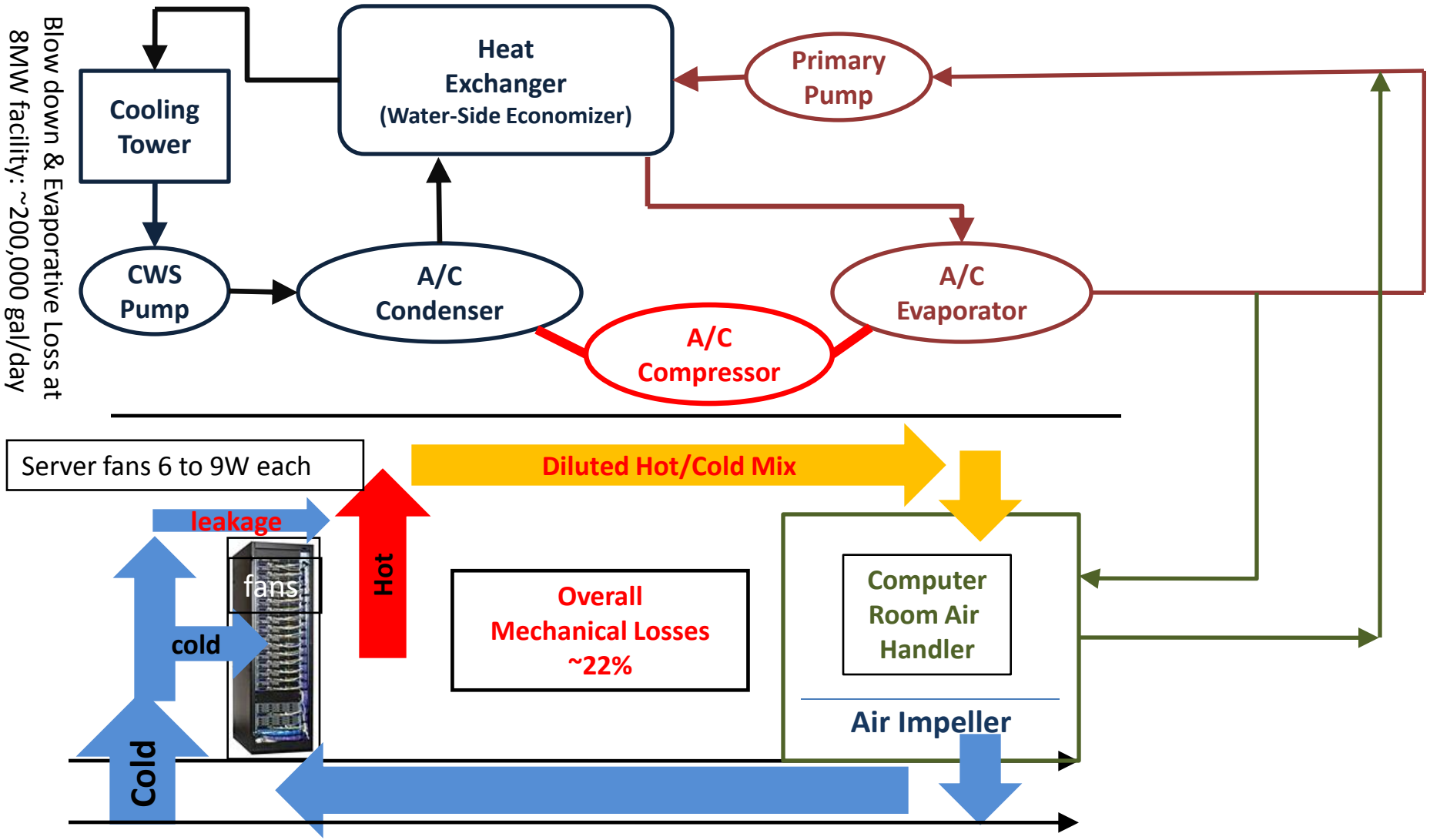
- Note: 2 more power conversions after last slide
  5. Power Supply: often <80% at typical load
  6. On board voltage regulators (VRMs or VRDs)
     - <80 Common & ~95% efficient available



- Rules to minimize power distribution losses:
  - Oversell power (more potential load than provisioned power)
  - Avoid conversions (fewer transformer steps & more efficient UPS)
  - Increase efficiency of conversions
  - High voltage as close to load as possible
  - Size voltage regulators to load & use efficient parts
  - High voltage direct current a small potential gain

- But power distribution improvements bounded to 11%

Mechanical Systems



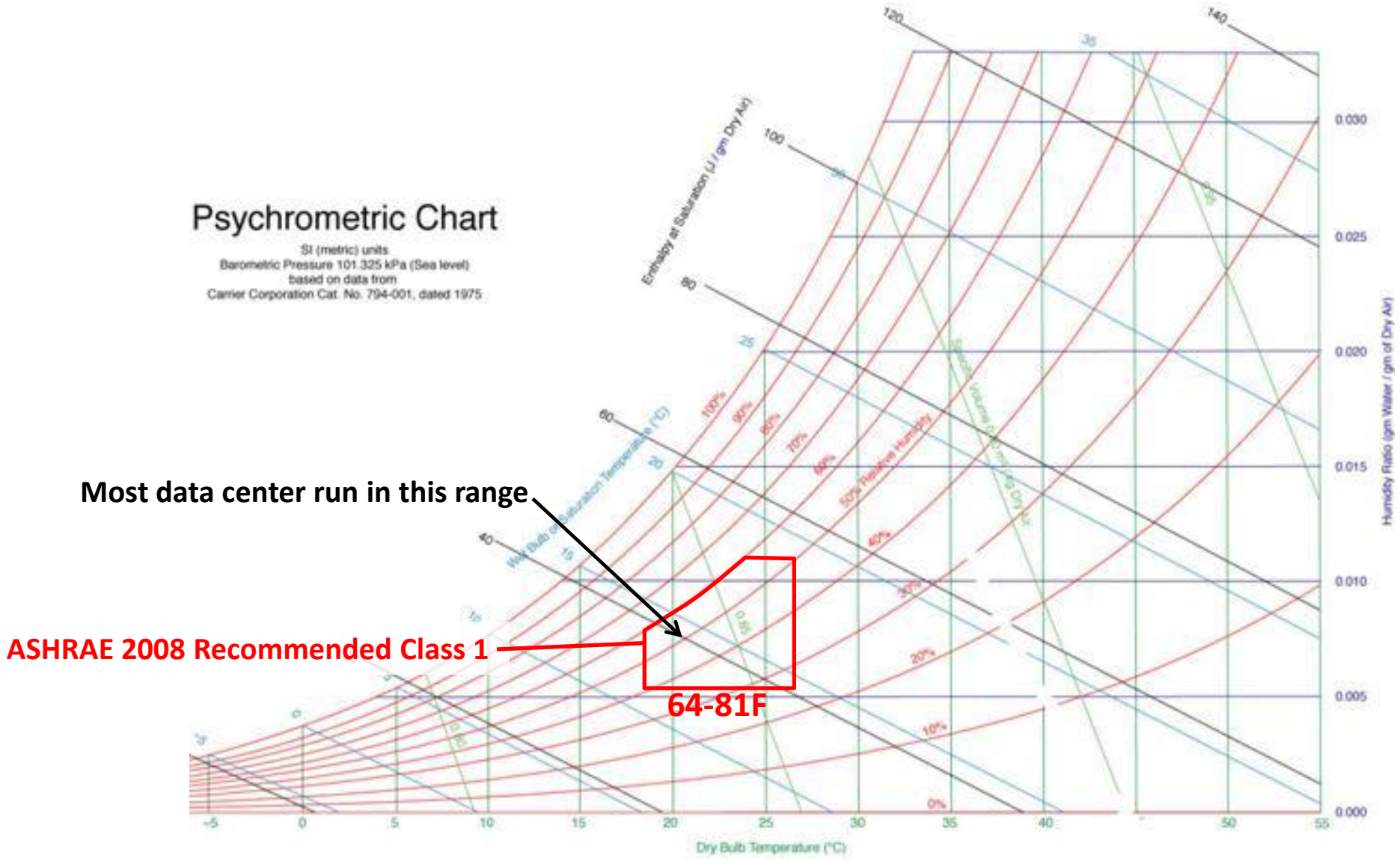
Hot Aisle/Cold Aisle Containment

WriteLine

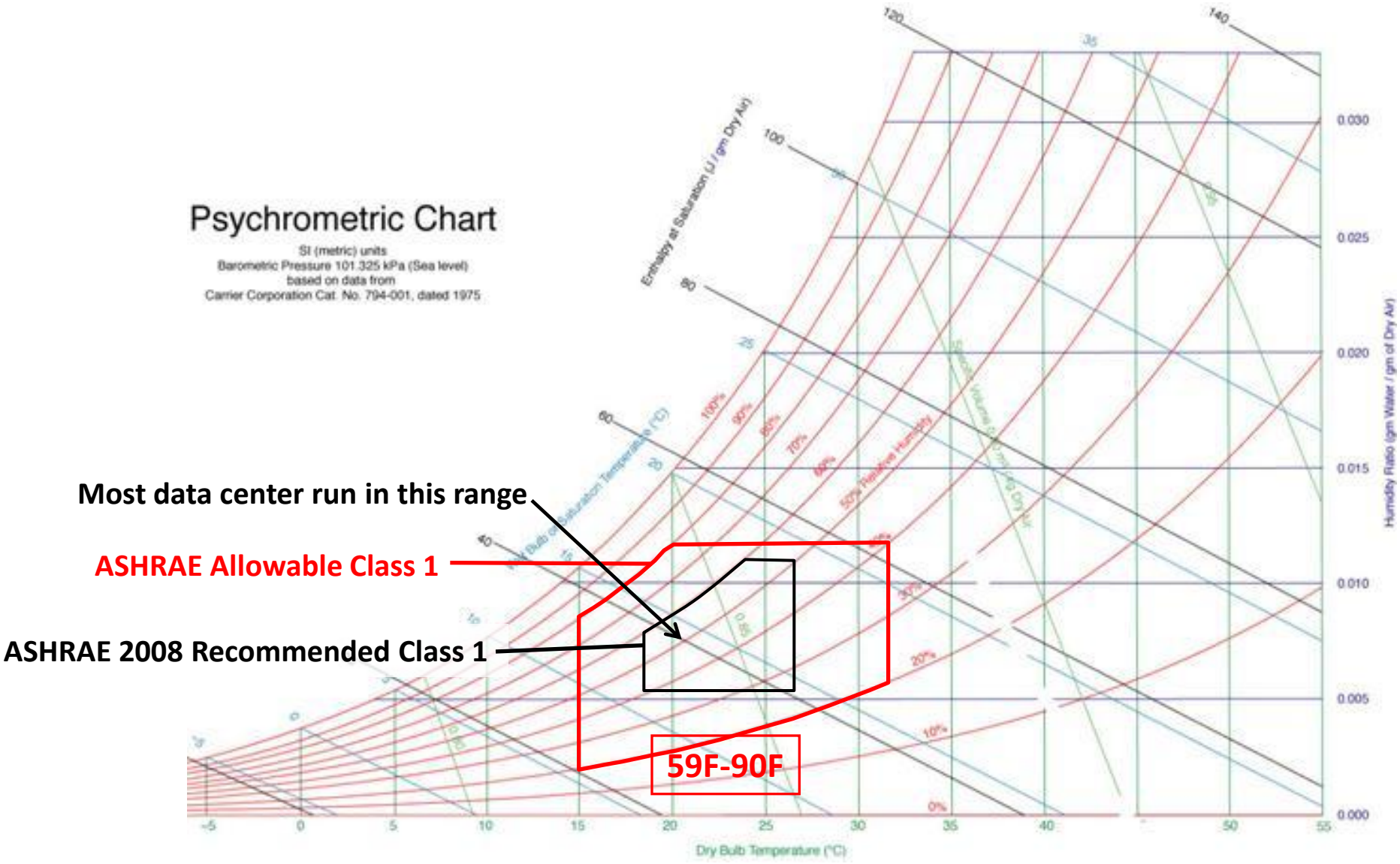
Intel

Intel

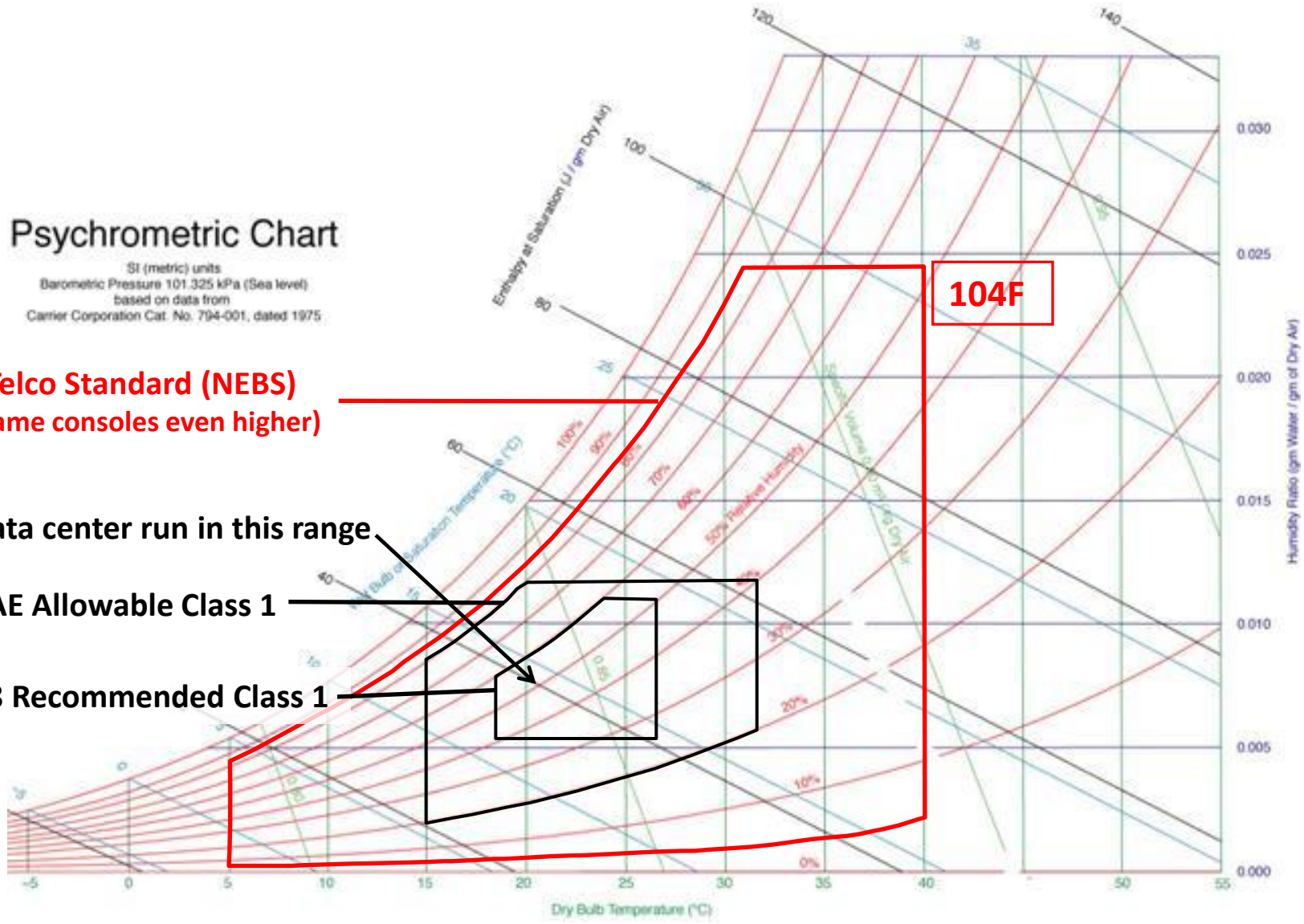
ASHRAE 2008 Recommended



ASHRAE Allowable

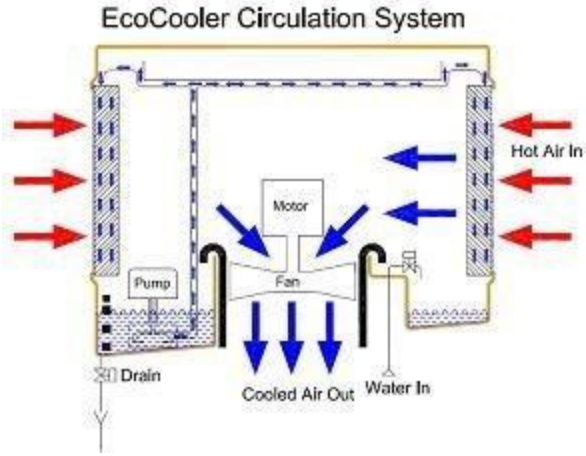


NEBS (Telco) & Rackable Systems



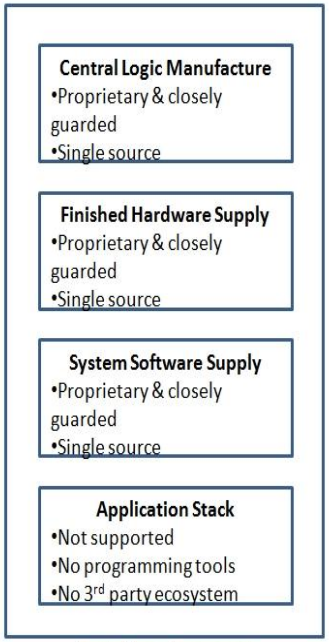
Air-Side Economization & Evaporative Cooling

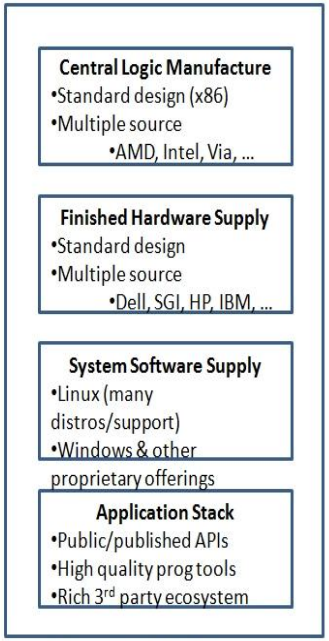
- Limiting factors to high temp operation
  - Higher fan power trade-off
  - More semiconductor leakage current
  - Possible negative failure rate impact
- Avoid direct expansion cooling entirely
  - Air side economization
  - Higher data center temperatures
  - Evaporative cooling
- Requires Filtration
  - Particulate & chemical pollution

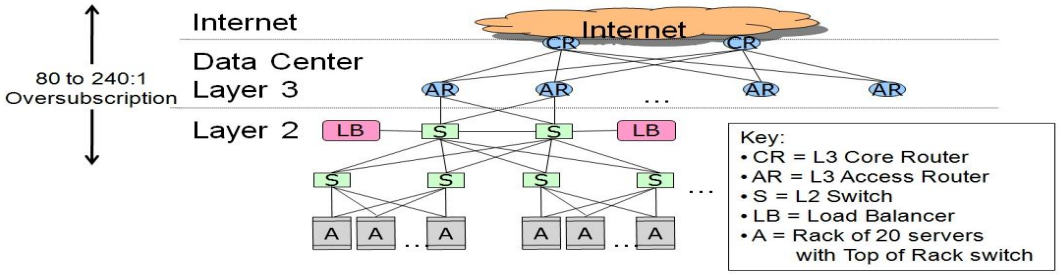


Sea Change in Net Gear

- Current networks over-subscribed
  - Forces workload placement restrictions
  - Goal: all points in datacenter equidistant
- Mainframe model goes commodity
  - Competition at each layer rather than vertical integration
- OpenFlow: open S/W platform
  - Distributed control plane to central control
  - E.g. VL2, Portland, and others







Server Innovation

- Shared Infrastructure Racks
  - Shared redundant PSUs & fans
  - e.g. Dell Fortuna & Rackable CloudRack
- Next Level: Multi-server on board
  - Intel Atom: SeaMicro
  - ARM: SmoothStone
- Very Low-Cost, Low-Power Servers
  - ARM, Atom, client & embedded CPUs
  - Cold storage (reduce CPU $ to GB)
  - Highly partitionable workloads: Web services, memcached
- Low utilization is still the elephant in room

More Information

- **These Slides:**
  - http://mvdirona.com/jrh/TalksAndPapers/JamesHamilton_Velocity20100623.pdf
- **Power and Total Power Usage Effectiveness**
  - http://perspectives.mvdirona.com/2009/06/15/PUEAndTotalPowerUsageEfficiencyTPUE.aspx
- **Berkeley Above the Clouds Paper**
  - http://perspectives.mvdirona.com/2009/02/13/BerkeleyAboveTheClouds.aspx
- **Degraded Operations Mode**
  - http://perspectives.mvdirona.com/2008/08/31/DegradedOperationsMode.aspx
- **Cost of Power**
  - http://perspectives.mvdirona.com/2008/11/28/CostOfPowerInLargeScaleDataCenters.aspx
  - http://perspectives.mvdirona.com/2008/12/06/AnnualFullyBurdenedCostOfPower.aspx
- **Power Optimization**
  - http://labs.google.com/papers/power_provisioning.pdf
- **Cooperative, Expendable, Microslice Servers**
  - http://perspectives.mvdirona.com/2009/01/15/TheCaseForLowCostLowPowerServers.aspx
- **Power Proportionality**
  - http://www.barroso.org/publications/ieee_computer07.pdf
- **Resource Consumption Shaping:**
  - http://perspectives.mvdirona.com/2008/12/17/ResourceConsumptionShaping.aspx
- **Email & Blog**
  - James@amazon.com & http://perspectives.mvdirona.com