

**CS528**

**Data Center  
Energy Consumption Modeling**

A Sahu

Dept of CSE, IIT Guwahati

# Energy Efficient System: Design and Management

- **Point to consider**

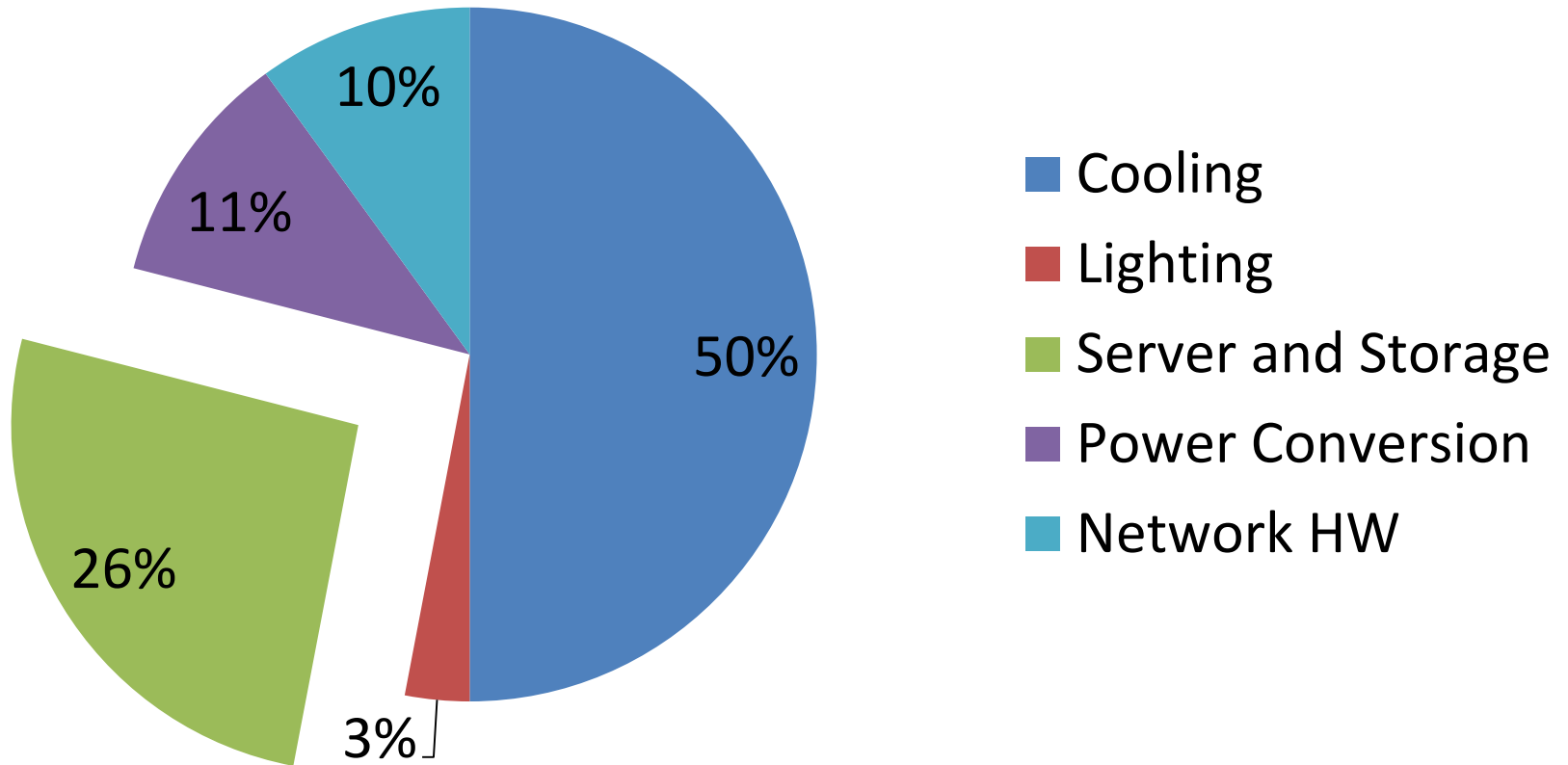
1. Energy efficient Infrastructure
2. Energy Model of Infrastructure
  - Blades/Server Machine CPU, Memory
3. Energy Efficient Scheduling
  - How to manage the Jobs

# **Data Center Energy Consumption Modeling**

**Ref: Dayaratna et. al., IEEE Comm. Survey, 2016**

# Energy Consumption Breakdown

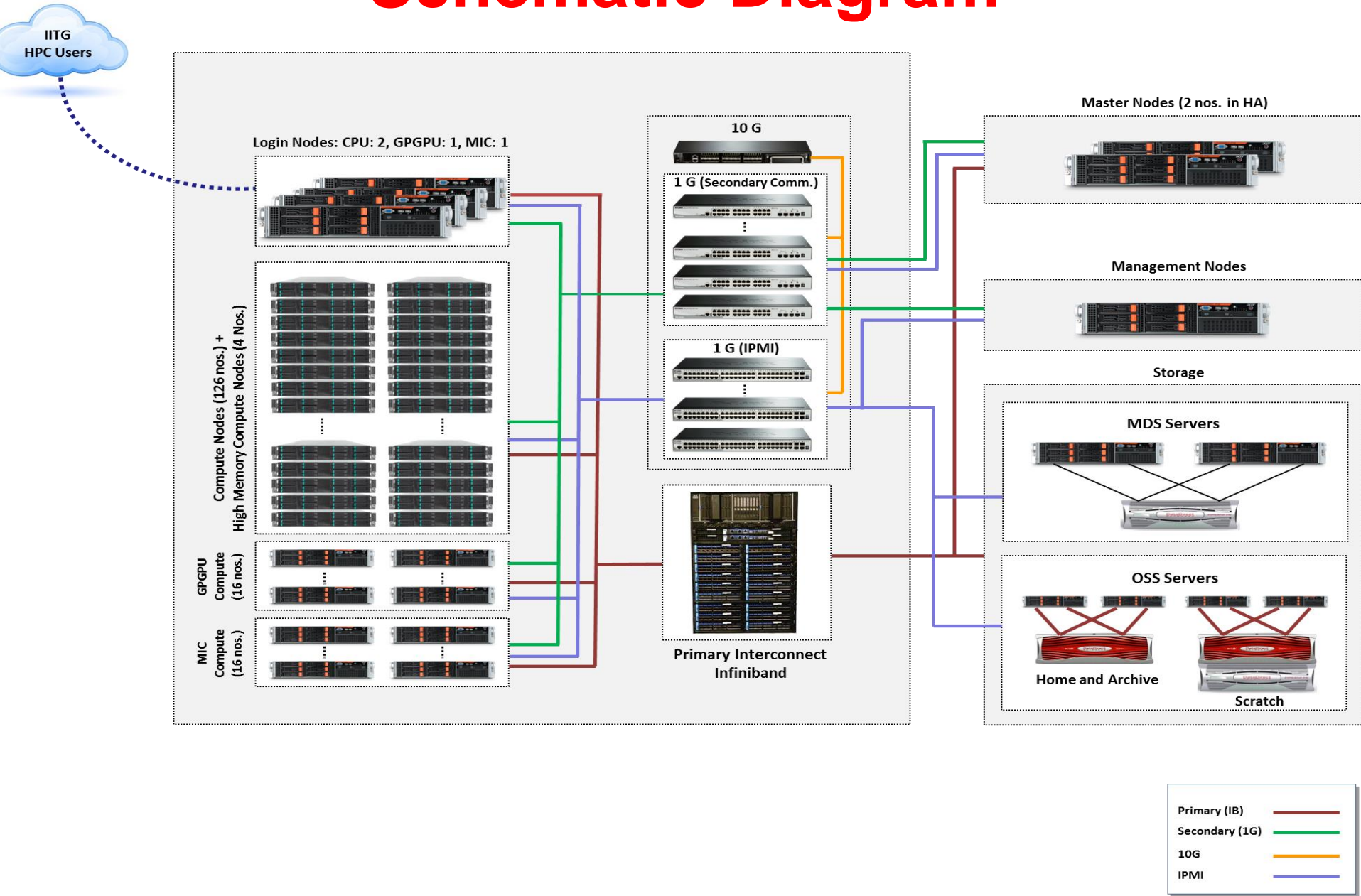
EC Breakdown



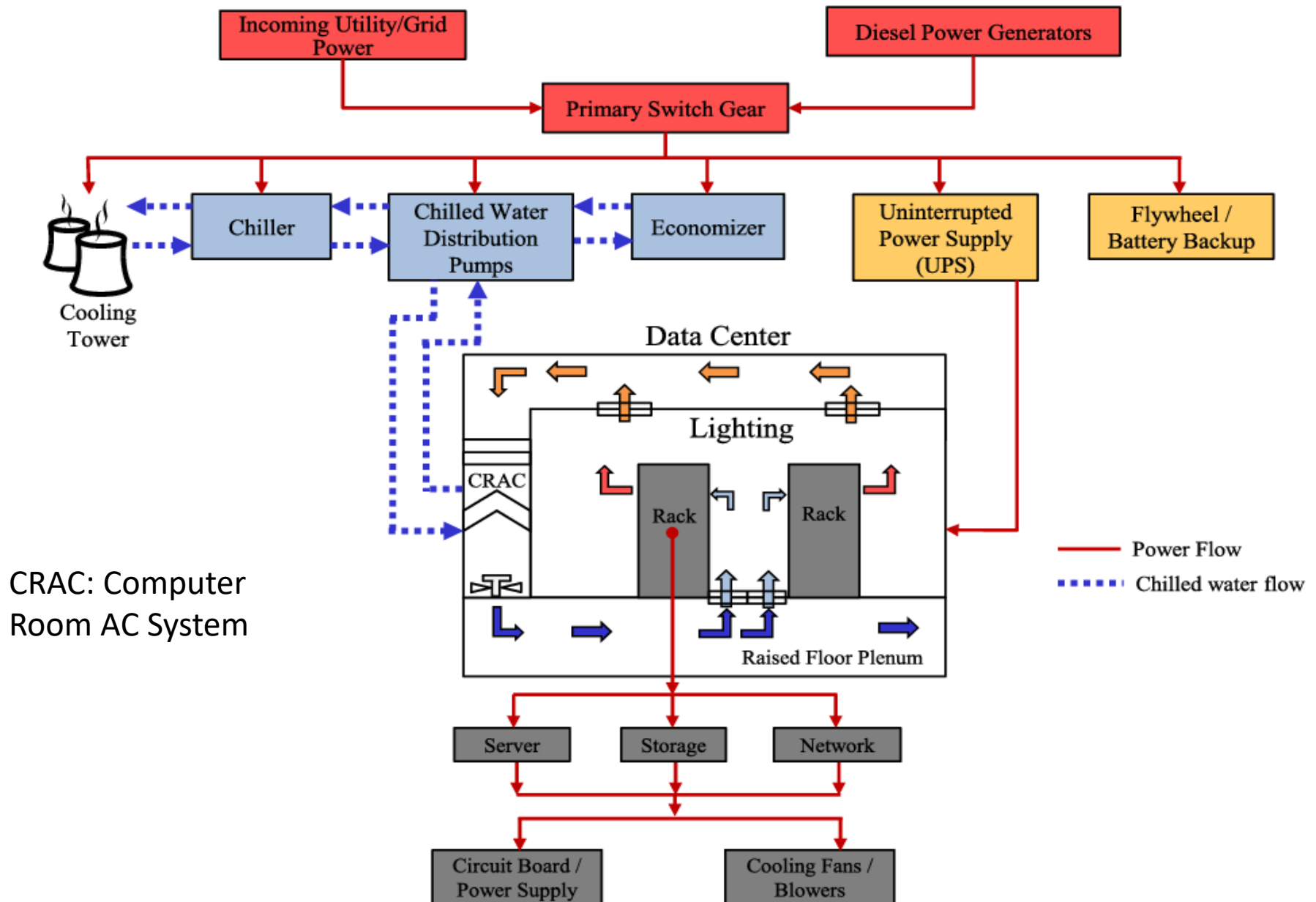
# PARAM ISHAN



# Schematic Diagram



# Power Flow in typical Data Center



# Metrics for Data Center Efficiency

- Data energy consumption : IT and Non-IT
- IT Parts
  - Server, Storage and Networking
- Non-IT Parts
  - Cooling, Lighting and Loss Power Conversion and Transmission



# Metrics for Data Center Efficiency

- Most widely used DC energy efficiency metric is *Power Usage Effectiveness (PUE)*

- The PUE

$$\eta_{\text{PUE}} = \frac{\text{Total data center annual energy}}{\text{Total IT annual energy}}$$

- The term  $\eta_{\text{PUE}} \geq 1$ , since data centers draw considerable amount of power as non-IT power
- A higher PUE : A greater portion electricity spent on cooling and the rest of the infrastructure

# Metrics for Data Center Efficiency

- *Data Center Infrastructure Efficiency (DCiE)*

$$\eta_{\text{DCiE}} = 1/\eta_{\text{PUE}} = \frac{\text{IT Device Power Consumption}}{\text{Total Power Consumption}} \times 100\%$$

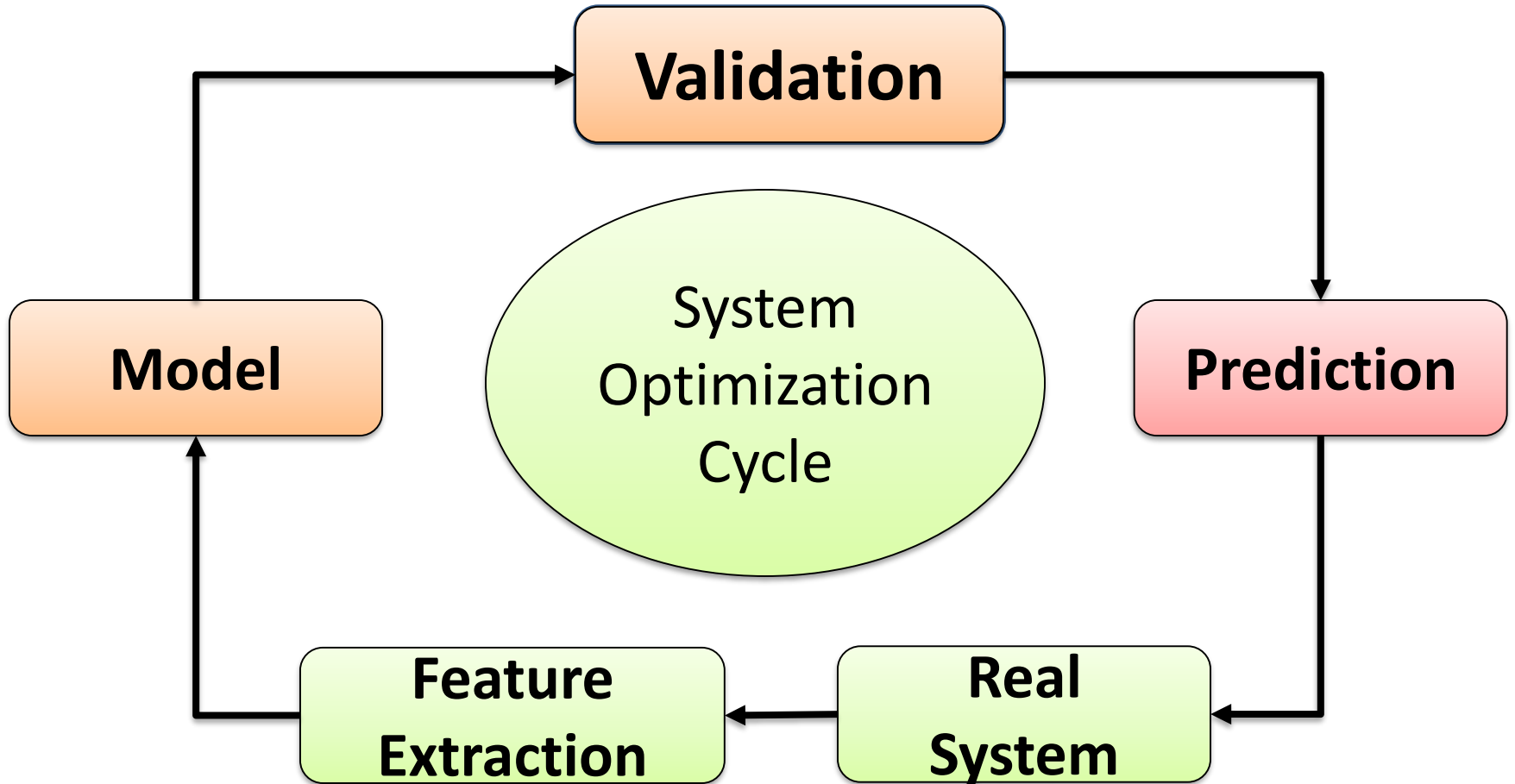
- *Data center Performance Per Energy (DPPE)*

$$\eta_{\text{DPPE}} = \frac{\text{Throughput at Data Center}}{\text{Energy Consumption}}$$

- *Data Center Green Energy Coefficient*

$$\eta_{\text{GEC}} = \frac{\text{Energy from Green Source (solar, wind, etc)}}{\text{Energy Consumption}}$$

# EC modeling and prediction in DC



# Organizational Framework for Power Models

- Instantaneous Power Consumption at time  $t$

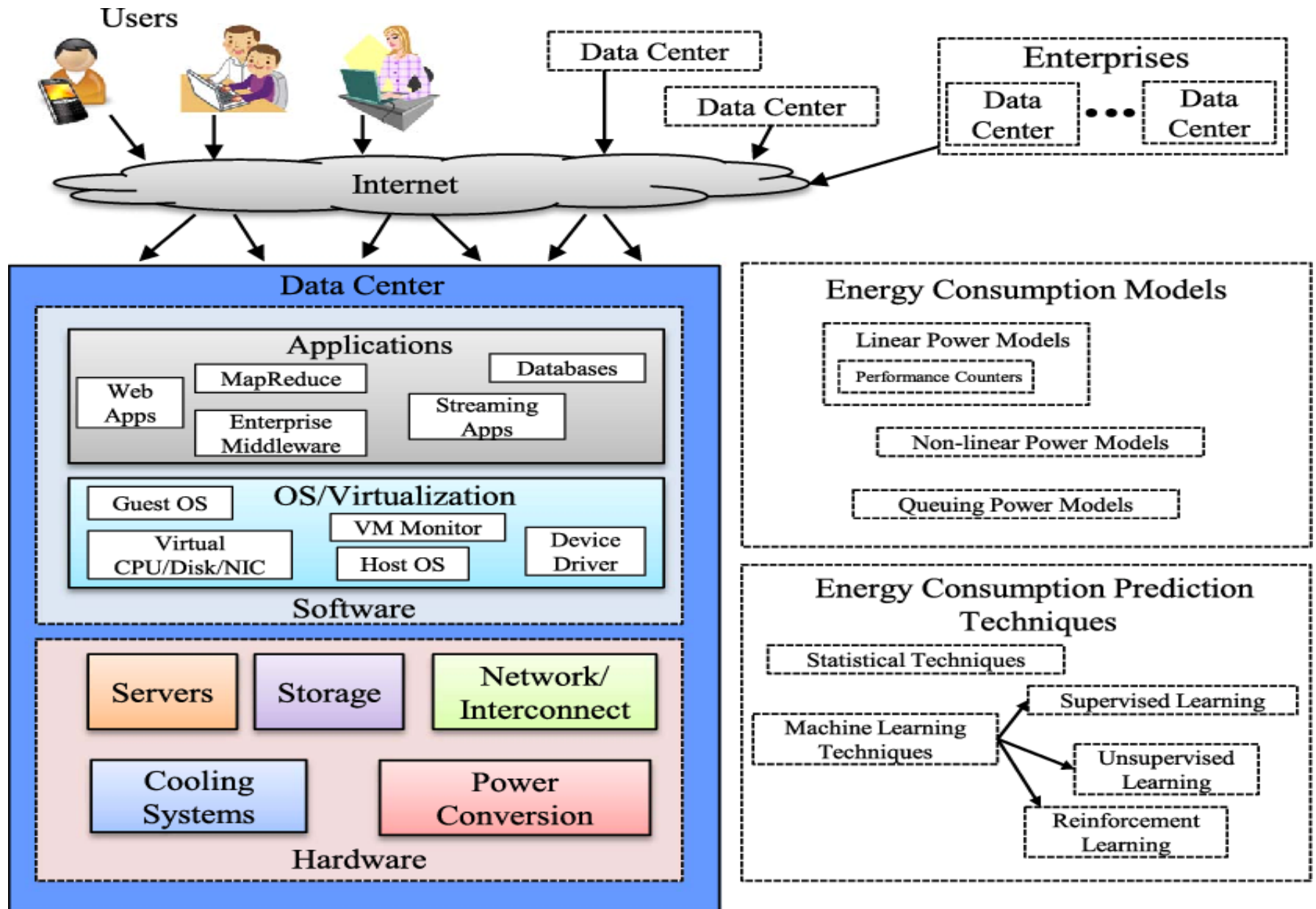
$$P_t = f(S_t, A_t, E_t)$$

- $S_t$  : internal state at time  $t$ 
  - Physical HW, OS, Application
  - HW config of processor, amount of mem, disk, NIC
  - Raw power measurement and Perf. CTRs at time  $t$
- $A_t$  : input to application at time  $t$ , request rate
- $E_t$  : Execution and Scheduling Strategy
  - Control CPU freq, Power Off/On, Software uses, Apps  $\rightarrow$  Core, Load balancing rule **at time  $t$** ,
- Prediction :  $P_{t+1} = f(S_t, A_t, E_t)$

# EC modeling and prediction in DC

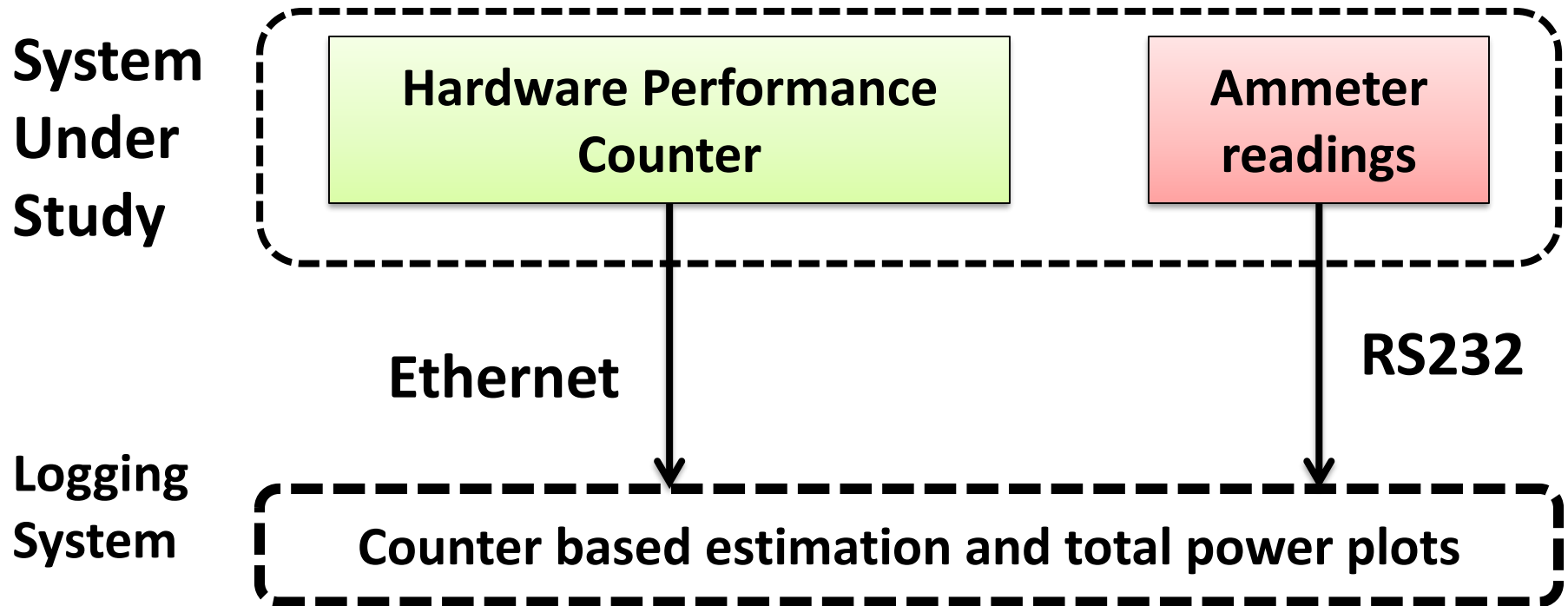
- A systematic view of the energy consumption modeling and prediction process.
- The data center system optimization cycle consists of four main steps
  - feature extraction
  - model construction
  - model validation,
  - and usage of the model.

# Holistic view of EC modeling and Prediction in DC



# Hybrid Approach for system PC Estimation

- Performance CTRs and Ammeter Reading
- PC Estimation

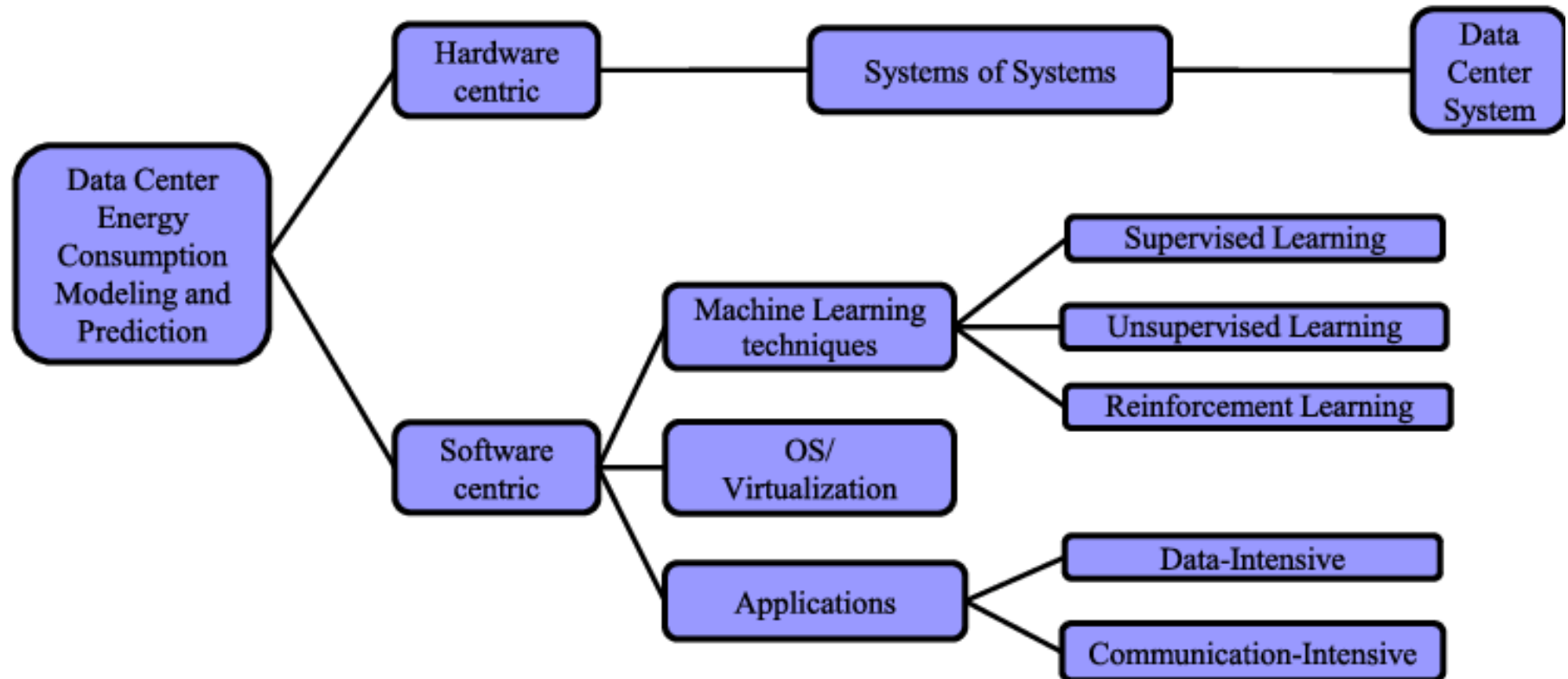


# Taxonomy of EC Modeling and Pred in DC

- EC Modeling & Prediction in DC
- Software centric
  - ML Technique
    - Supervised/Unsupervised/Reinforcement ML
  - OS/Virtualizations
  - Applications
    - Data Intensive and Compute Intensive
- Hardware Centric



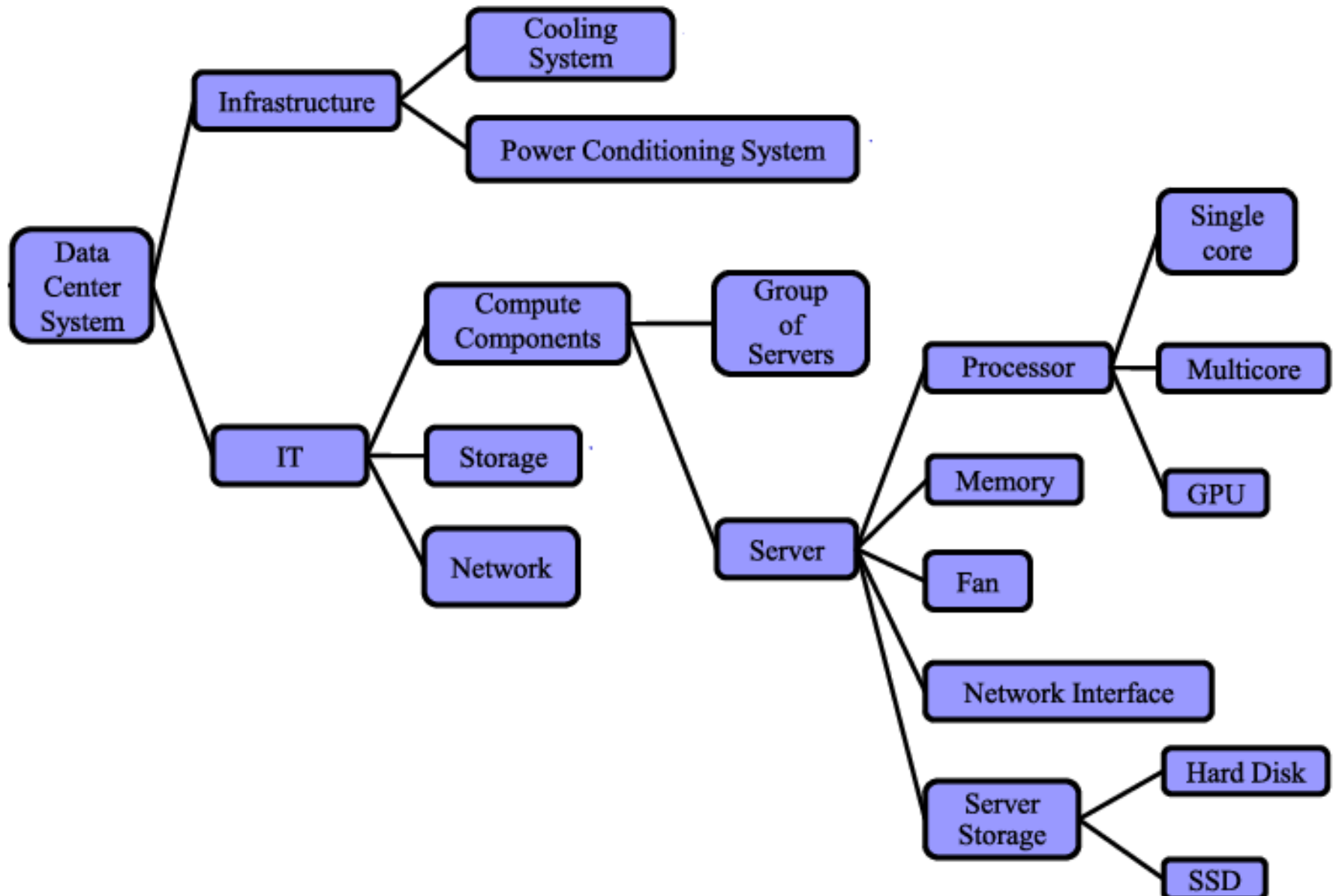
# Taxonomy of EC Modeling and Pred in DC



# Taxonomy of EC Modeling and Pred in DC

- Hardware Centric: Infrastructure and IT
- Infrastructure
  - Cooling System and Power Conditioning System
- IT
  - Compute Components
    - Group of Server
    - Server: Processor(Single, Multi), Memory, FAN, NIC, Disk (HDD, SDD)
  - Storage
  - Network

# Taxonomy of EC Modeling and Pred in DC



# Organizational Framework for Power Models

- Instantaneous Power Consumption at time  $t$

$$P_t = f(S_t, A_t, E_t)$$

- $S_t$  : internal state at time  $t$ 
  - Physical HW, OS, Application
  - HW config of processor, amount of mem, disk, NIC
  - Raw power measurement and Perf. CTRs at time  $t$
- $A_t$  : input to application at time  $t$ , request rate
- $E_t$  : Execution and Scheduling Strategy
  - Control CPU freq, Power Off/On, Software uses, Apps  $\rightarrow$  Core, Load balancing rule **at time  $t$** ,
- Prediction :  $P_{t+1} = f(S_t, A_t, E_t)$

# Aggregate View of Server Energy Model

- Additive Server Power Model is the Simplest Power Model
- Energy consumption of CPU and Mem while running the Algorithm A

$$E(A) = E_{\text{cpu}}(A) + E_{\text{memory}}(A)$$

- More detailed including  $E_{\text{I/O}}$

$$E_{\text{total}} = E_{\text{cpu}} + E_{\text{mem}} + E_{\text{I/O}}$$

- More detailed elaborated I/O

$$E_{\text{total}} = E_{\text{cpu}} + E_{\text{mem}} + E_{\text{disk}} + E_{\text{NIC}}$$

# Aggregate View of Server Energy Model

- More detailed using P.T of individual components

$$E_{\text{total}} = P_{\text{comp}} T_{\text{comp}} + P_{\text{NIC}} T_{\text{NIC}} + P_{\text{net\_dev}} T_{\text{net\_dev}}$$

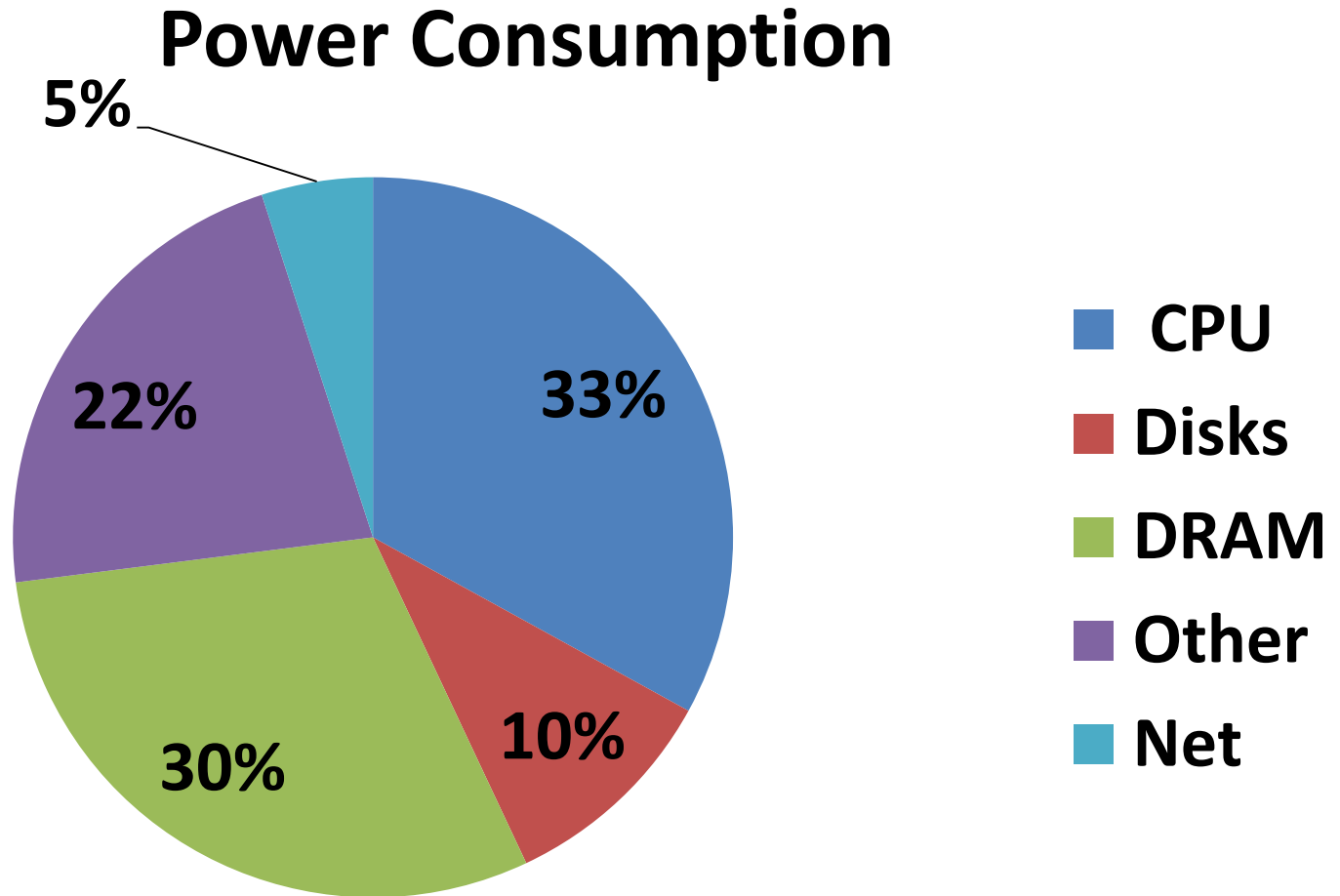
- Power model using resource utilization at time t

$$P_t = C_{\text{cpu}} U_{\text{cpu},t} + C_{\text{mem}} U_{\text{mem},t} + C_{\text{disk}} U_{\text{disk},t} + C_{\text{nic}} U_{\text{nic},t}$$

- Energy model of Entire system including board and electro-memchanical components

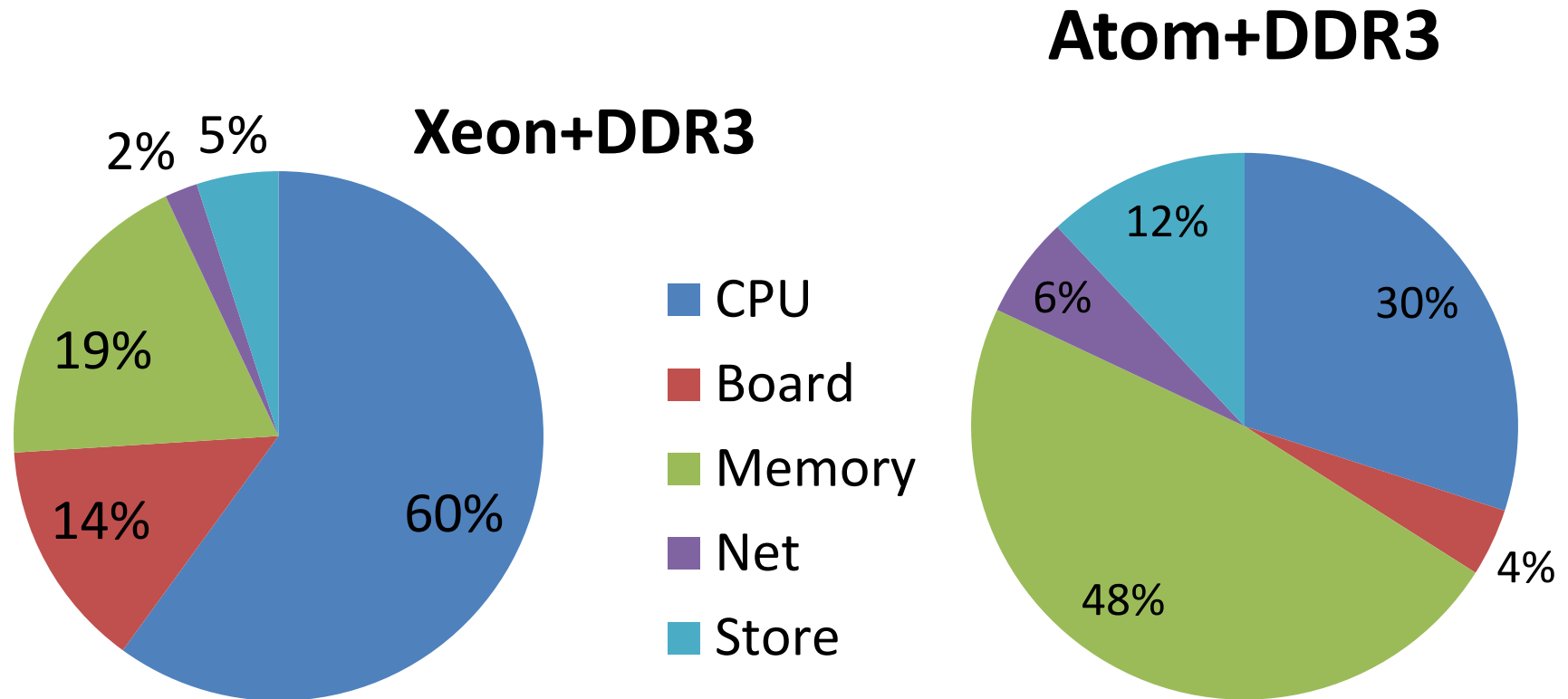
$$E_{\text{system}} = A_0 (E_{\text{proc}} + E_{\text{mem}}) + A_1 E_{\text{em}} + A_2 E_{\text{board}} + A_3 E_{\text{hdd}}$$

# Approximate distribution by components of WSC at Google 2007



**PC of Components of Warehouse Scale Computer**

# Power breakdown of Atom vs Xeon Server



**PC breakdown across the components of two servers**  
**(a) Intel Xeon based server, (b) Intel Atom based server**



# PC Model Virtualized Server

- Assuming n number of VM running in a server

$$P_{\text{server}} = P_{\text{baseline}} + \sum_{i=1}^n P_{\text{vm}}(i)$$

- Expanded power modeling using usage of VMs

$$P_{\text{server}} = P_{\text{baseline}} + P_{\text{ne}} + \alpha \sum_{i=1}^n U_{\text{cpu}}(i) + \beta \sum_{i=1}^n U_{\text{mem}}(i) + \gamma \sum_{i=1}^n U_{\text{io}}(i)$$

# System Utilization based Power Model

- CPU is largest power consumer

$$P = c_0 + c_1 f^3$$

- PC of Blade server linear model

$$P_{\text{blade}} = 12.45 + 0.236 * u_{\text{cpu}} - 4.4 \times 10^{-8} * u_{\text{mem}} \\ + 0.0281 * u_{\text{dsk}} + 3.1 \times 10^{-8} * u_{\text{net}}$$

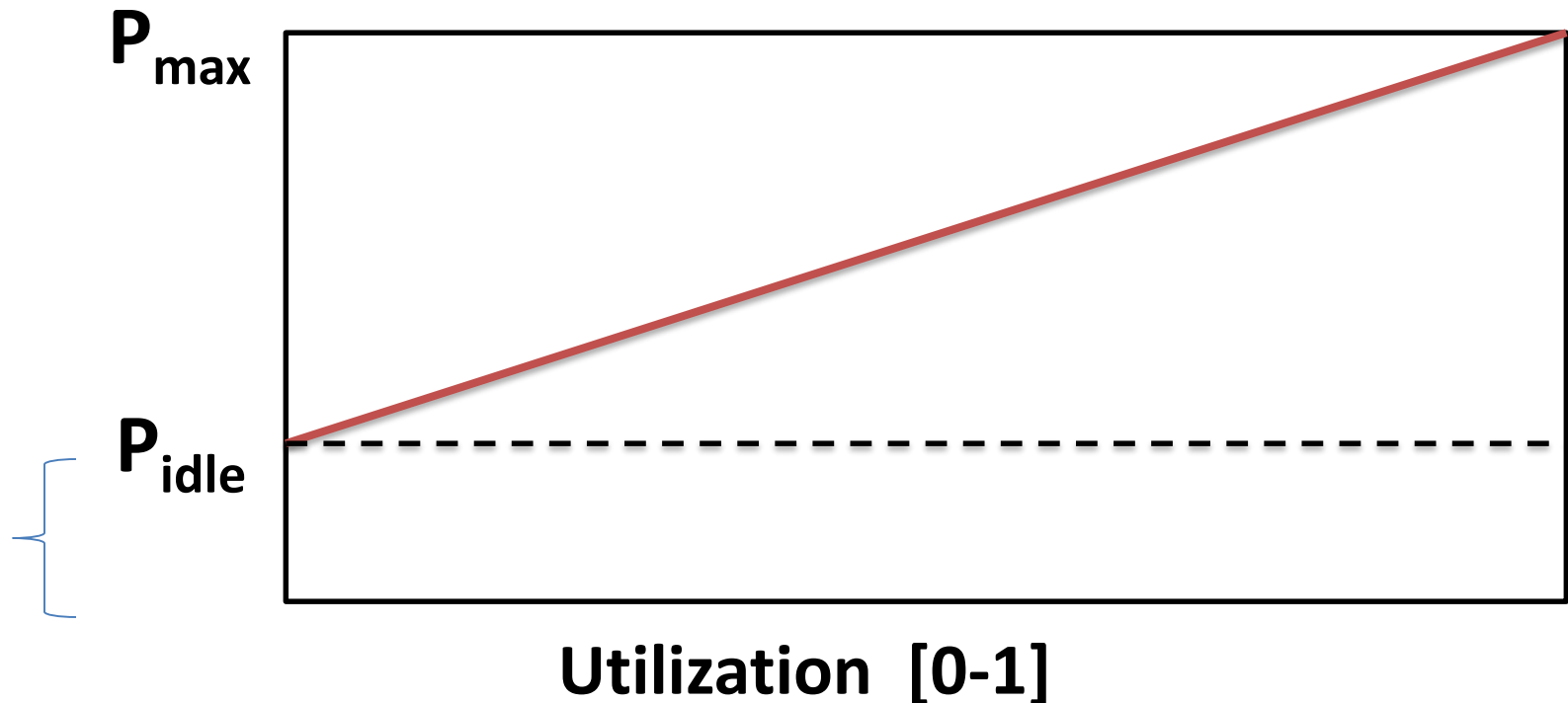
- Most used PC model of CPU/Server, PC of server at utilization  $u$

$$P_u = (P_{\text{max}} - P_{\text{idle}}) u + P_{\text{idle}}$$

# System Utilization based Power Model

- Most used PC model of CPU/Server, PC of server at utilization  $u$

$$P_u = (P_{\max} - P_{\text{idle}}) u + P_{\text{idle}}$$

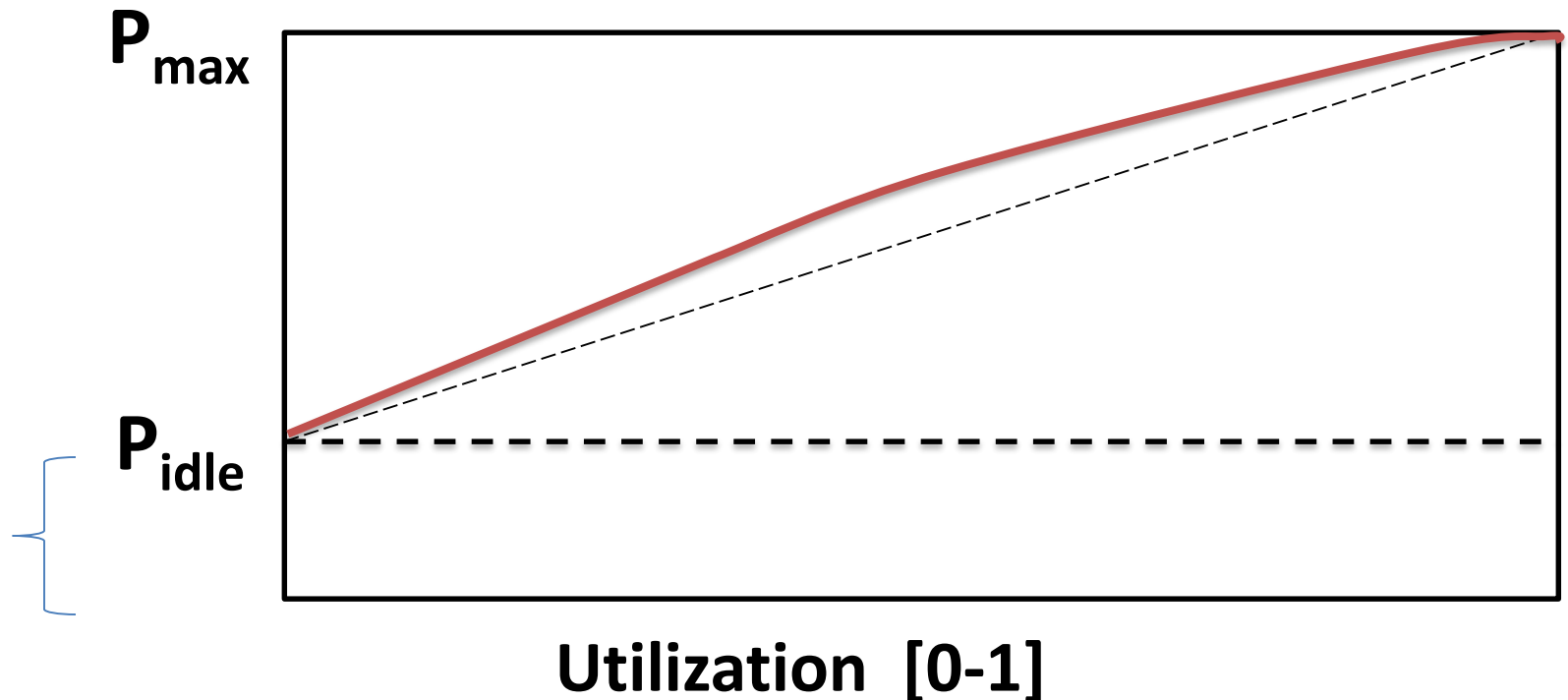


# System Utilization based Power Model

- Another PC model of at utilization  $u$

$$P_u = (P_{\max} - P_{\text{idle}}) (2u - u^r) + P_{\text{idle}}$$

- The value of  $r \approx 1.4$  Experimentally



# Normalized Utilization based Power Model

- Normalized PC

$$P_{\text{norm}} = (P_{\text{sys}} - P_{\text{idle}}) / (P_{\text{busy}} - P_{\text{idle}})$$

–  $P_{\text{busy}}$  is PC when  $U = 1$

- Another model

$$P(t) = P_{\text{idle}} + (P_{\text{full}} - P_{\text{idle}}) * \alpha U(t)^\beta$$

– Where  $\alpha, \beta$  are server dependent parameter

- **Modified Popular PC model**

$$P = P_{\text{idle}} + b_i(t)^\alpha / A \Rightarrow P = P_s + k f^3$$

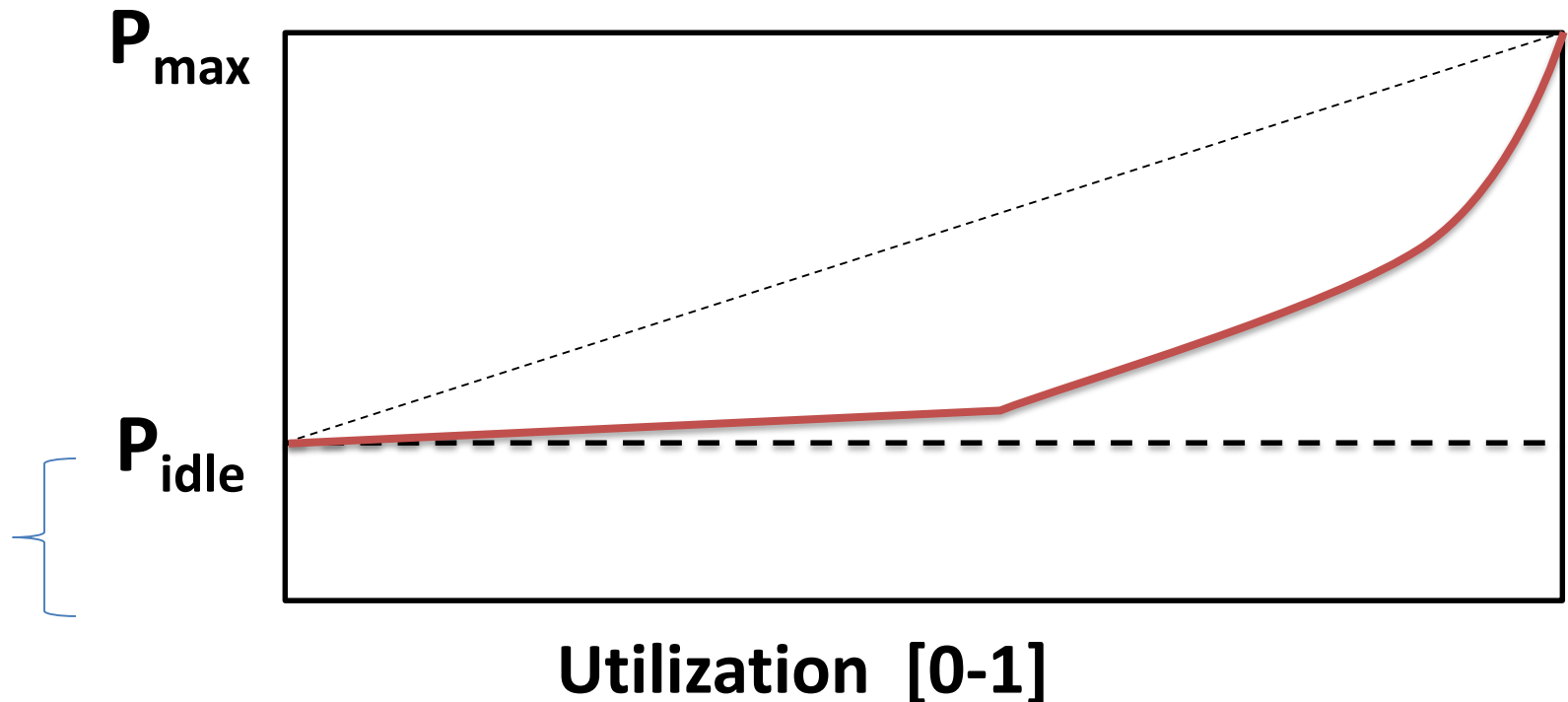
–  $A, P_{\text{idle}}, \alpha$  are constant depend on server,  $\alpha = 3$

– The term  $b_i(t)$  is rate of utilization

# System Utilization based Power Model

- Most used PC model of CPU/Server, PC of server at utilization  $u$

$$P_u = P_{idle} + k f^3$$



# PC Model Server: Multicore Processor

- Additive PC model : simple  $P_n = \sum_{j=1}^n P_c(j)$
- Another model with diff levels of core speed

$$P_{\text{core}} = \rho \cdot s^\alpha = (\lambda/m) \cdot R \cdot s^{\alpha-1}$$

- s is speed,  $\lambda$  arrival rate of task, m number of core
- R is avg number of instr. to execute,  $\rho$  = utilization
- Total PC of Server  $P = m \cdot \rho \cdot s^\alpha = \lambda \cdot R \cdot s^{\alpha-1}$ 
  - Where  $m \cdot \rho = \lambda \cdot x_{\text{avg}}$  is avg number of busy core in S
- Total PC including Static part

$$P = m \cdot \rho \cdot s^\alpha + P_{\text{static}} = \lambda \cdot R \cdot s^{\alpha-1} + P_{\text{static}}$$

$$\text{if } \rho=1 \rightarrow P = m \cdot s^\alpha + P_{\text{static}}$$

# PC Model Server: Multicore Processor

- Additive PC model :  $P_{proc} = P_{mc} + P_{dies} + P_{intd}$ 
  - Mc, dies, intd are PC of chip level mandatory components, constituent die, inter-die comm.
- Another model with diff levels of core speed

$$P_{proc} = P_{base} + (C.f^3 + D) + (\sum_{i=1}^3 g_i L_i) + g_m M$$

- $L_i$ ,  $i$ th level cache miss,  $i=1,2,3$ ,  $M$  is memory access

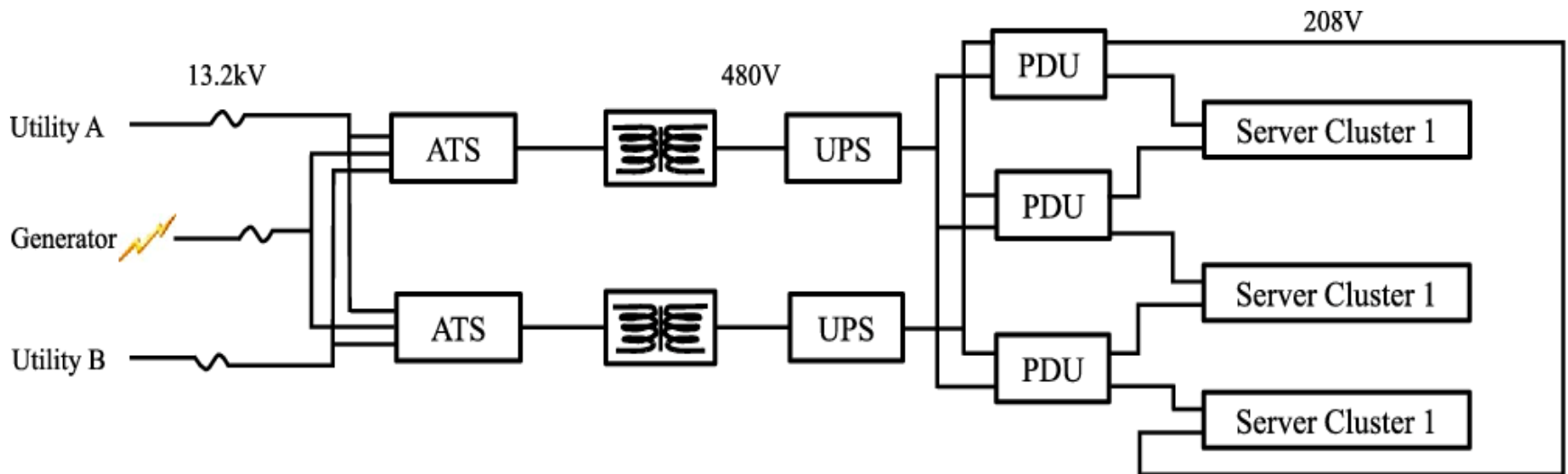


# Modeling PC of Memory

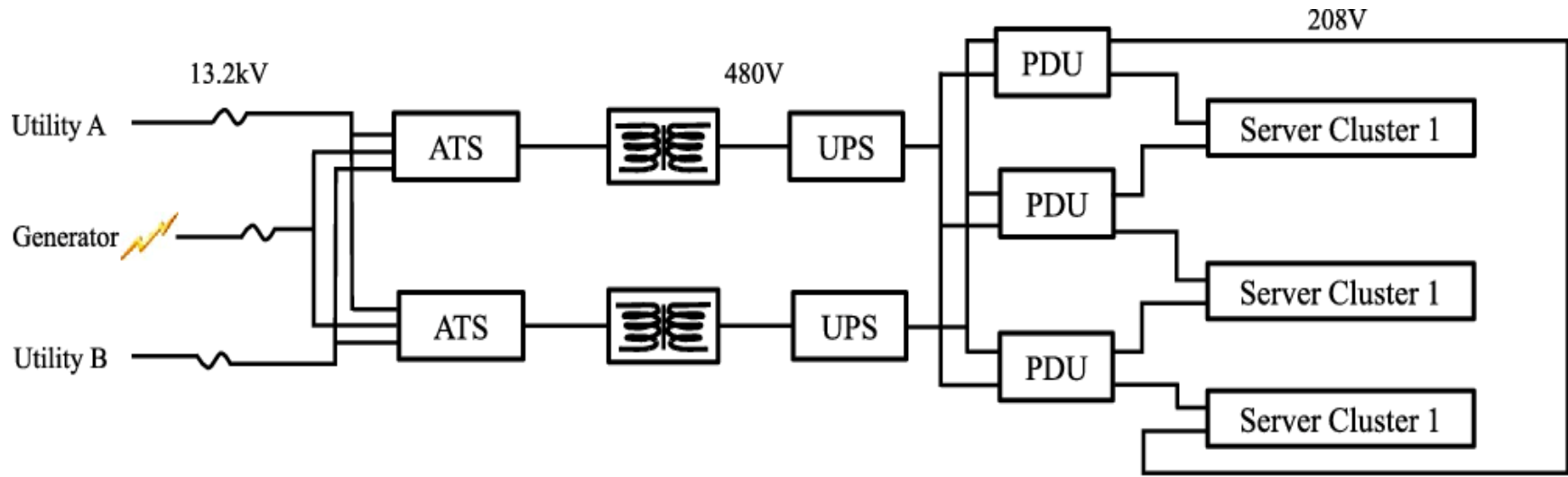
- Memory DRAM:  $P_{dm} = P_{st\_dm} + \alpha_1 \cdot N_{rd} + \alpha_2 \cdot N_{wr}$
- $P_{mem} = D \cdot S \cdot R \cdot \sigma + E_{rw} \rho_{rw} + D \cdot E_{ap} f_{ap}$ 
  - D chip/subset, S subset/rank, R rank/Chanel
  - $\sigma$  is static PC of DRAM,  $\rho_{rw}$  rd/wr BW/channel,  $E_{rw}$  is Energy per rw/bit,  $f_{ap}$  is freq of active recharge,  $E_{ap}$  is Energy/pre-charge

# Modeling EC of Power Conditioning System

- Example power delivery system of DC with redundant dist. Path
- PDU: Power Dist. Unit., ATS: Auto Trans Switch

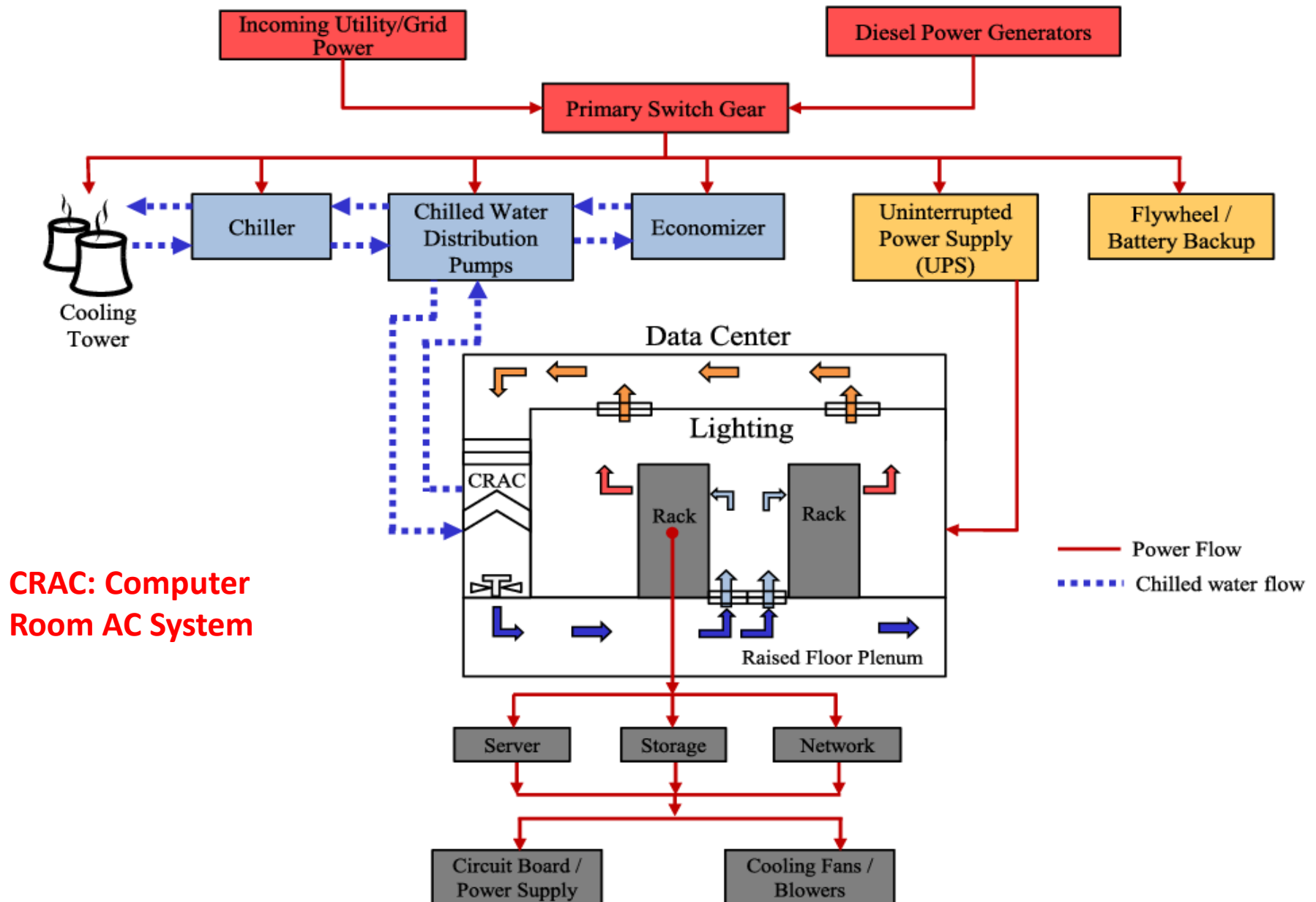


# Power Conditioning System



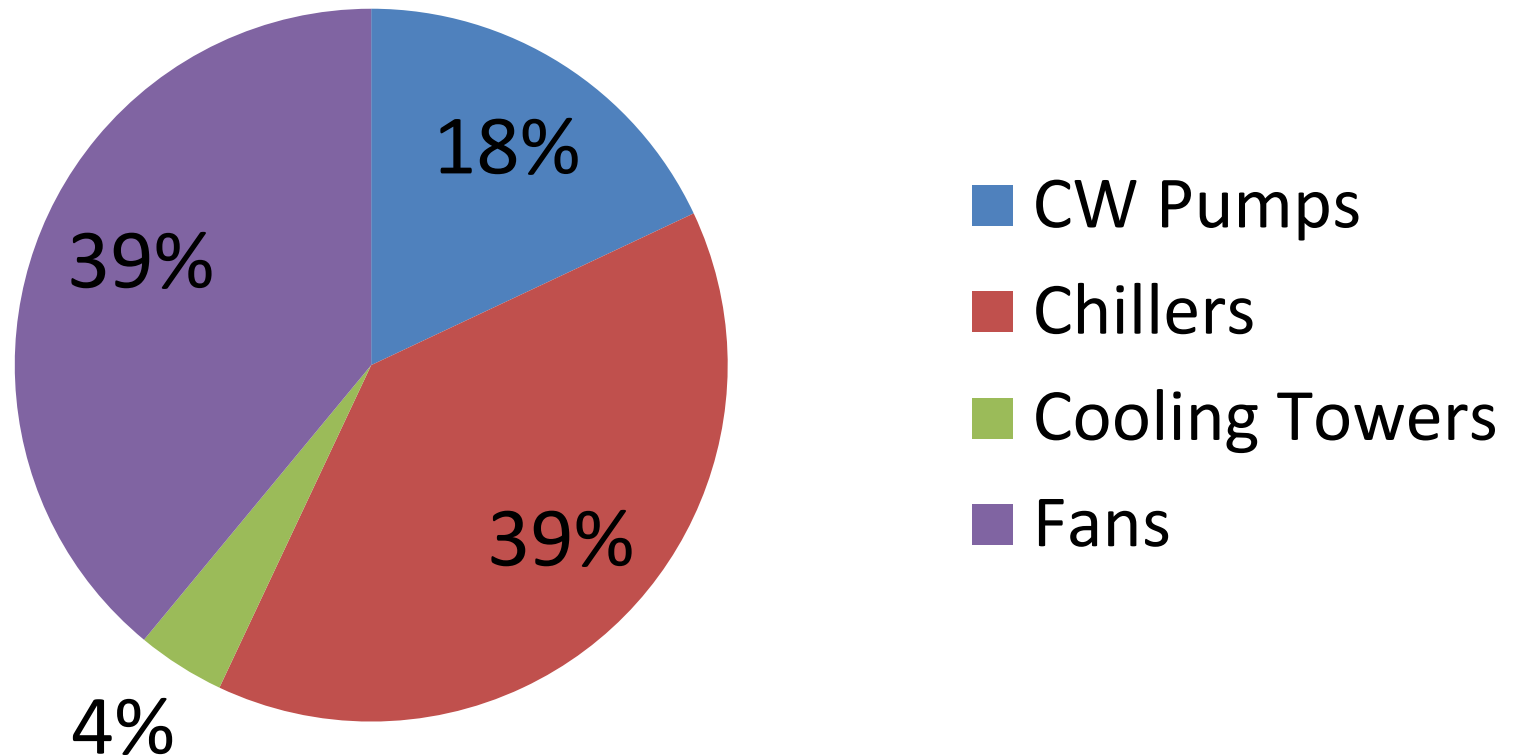
- $P_{\text{pdu\_loss}} = P_{\text{pdu\_idle}} + \pi_{\text{pdu}} (\sum_N P_{\text{srv}})^2$   
–  $\pi_{\text{pdu}}$  is power loss co-efficient
- $P_{\text{ups\_loss}} = P_{\text{ups\_idle}} + \pi_{\text{ups}} (\sum_M P_{\text{pdu}})$
- PDU waste 3%, where UPS waste 9%

# Power Flow in typical Data Center



# Typical Power Break of HVAC

A break down of Power Consumption of  
HVAC of a Data Center



HVAC :Heating, Ventilation and Cooling System