CS528
Data Center Energy Consumption Model

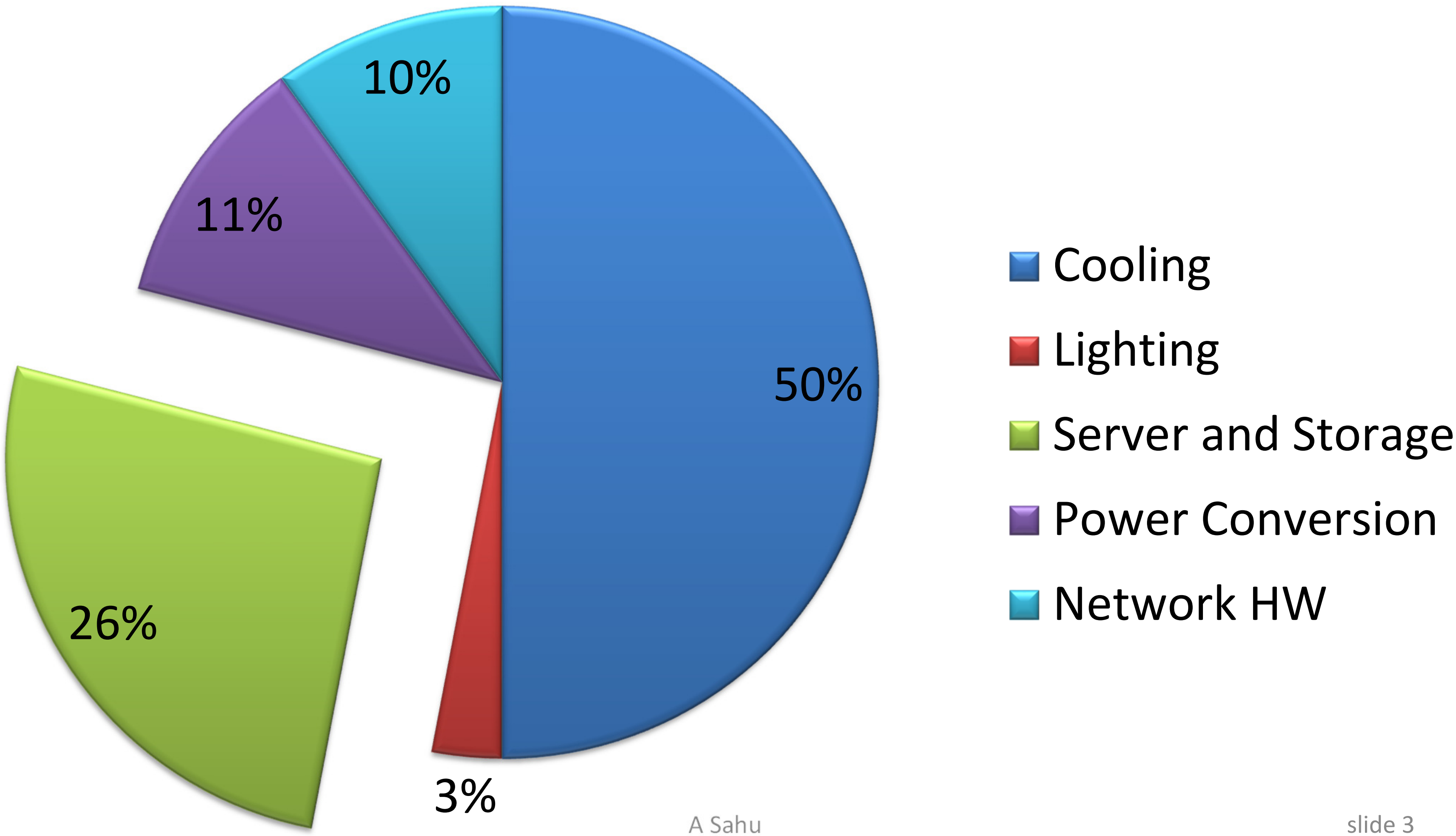
A Sahu

Dept of CSE, IIT Guwahati

Data Center Energy Consumption Modeling

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

Energy Consumption Breakdown of 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

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

- The term $\eta_{PUE} \geq 1$, 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)*

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

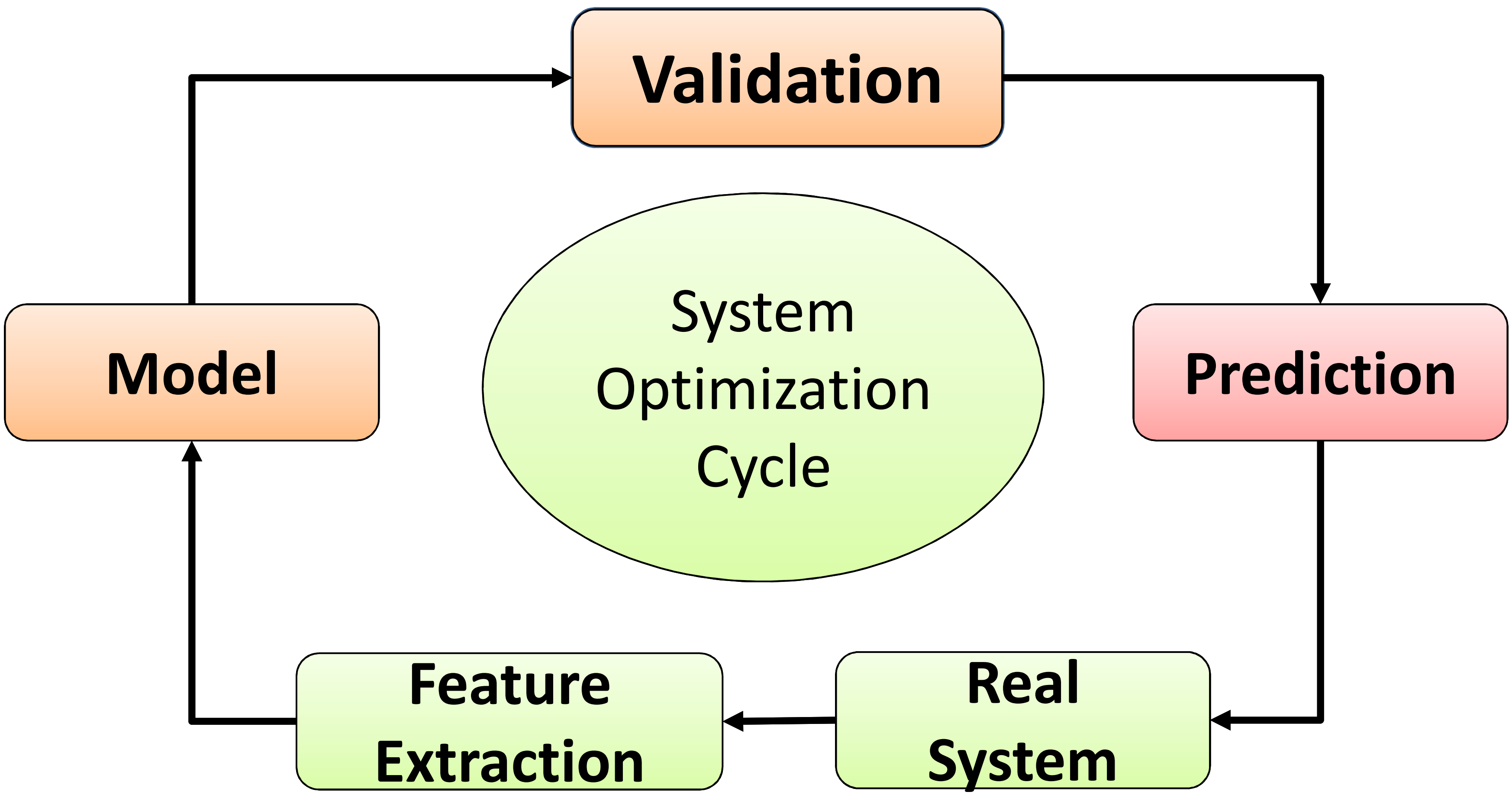
- *Data center Performance Per Energy (DPPE)*

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

- Data Center Green Energy Coefficient

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

EC modeling and prediction in DC



Organizational Framework for Power Models

- Instantaneous Power Consumption at time t

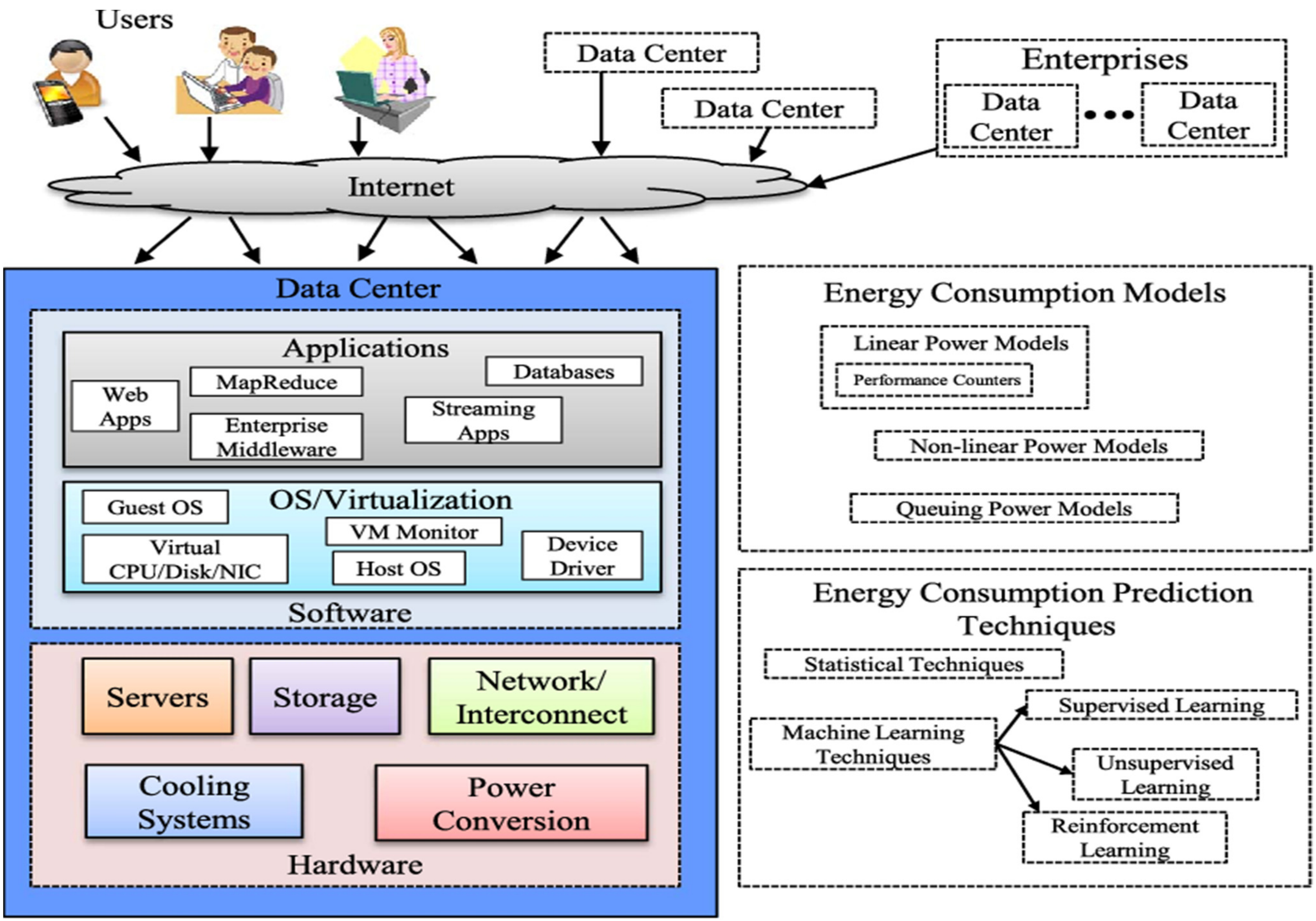
  $P_t = f(S_t, A_t, E_t)$

- 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→Core, Load balancing rule **at time t**,
- Prediction : $P_{t+1} = f(S_t, A_t, E_t)$

EC modeling and prediction in DC

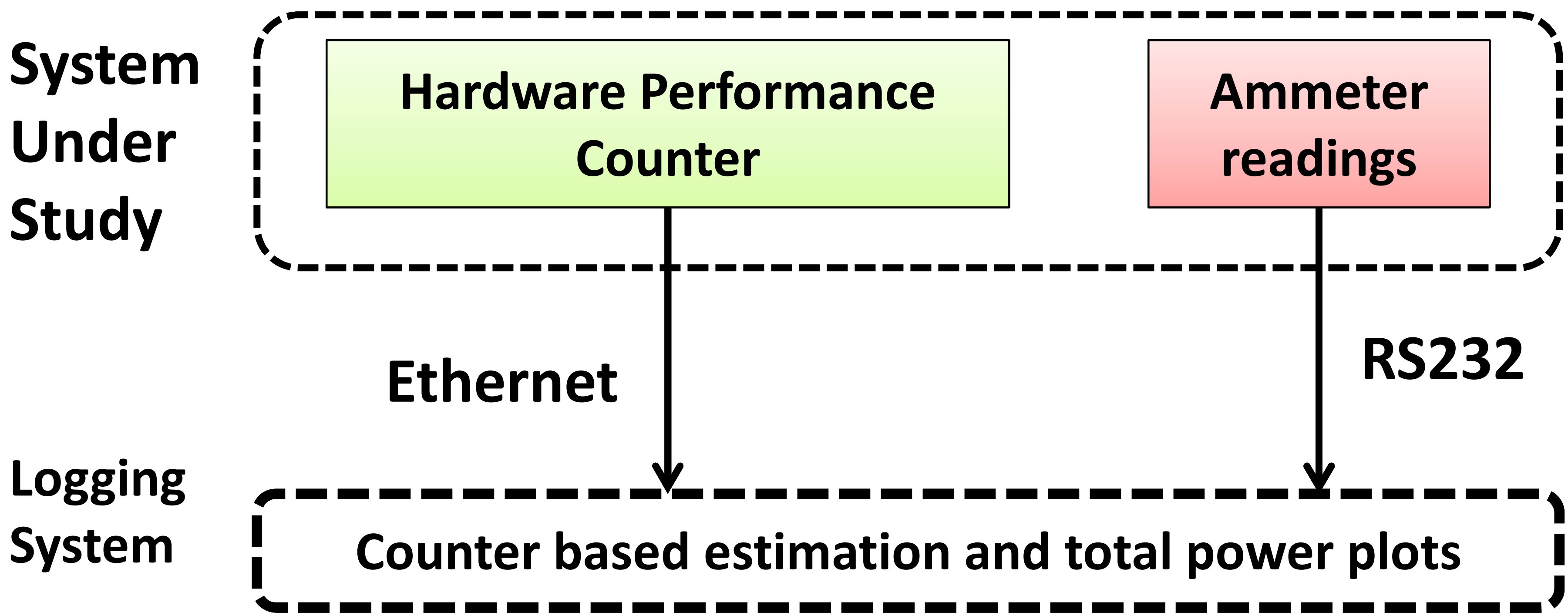
- A systematic view of the energy consumption modeling and prediction process.
- The data center system optimization cycle consists of four main steps
  - feature extraction
  - model construction
  - model validation,
  - and usage of the model.

Holistic view of EC modeling and Prediction in DC



Hybrid Approach for system PC Estimation

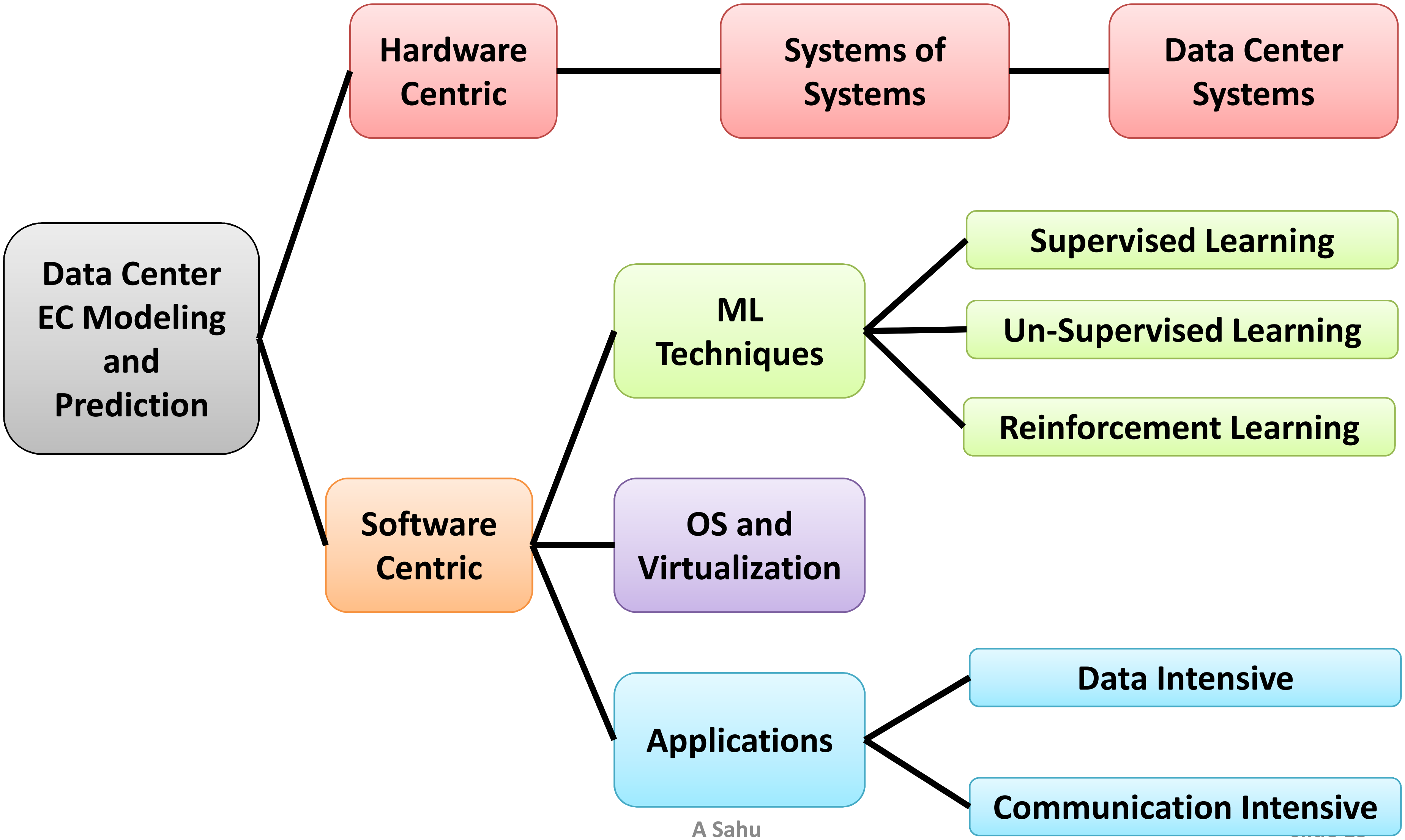
- Performance CTRs and Ammeter Reading
- PC Estimation



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

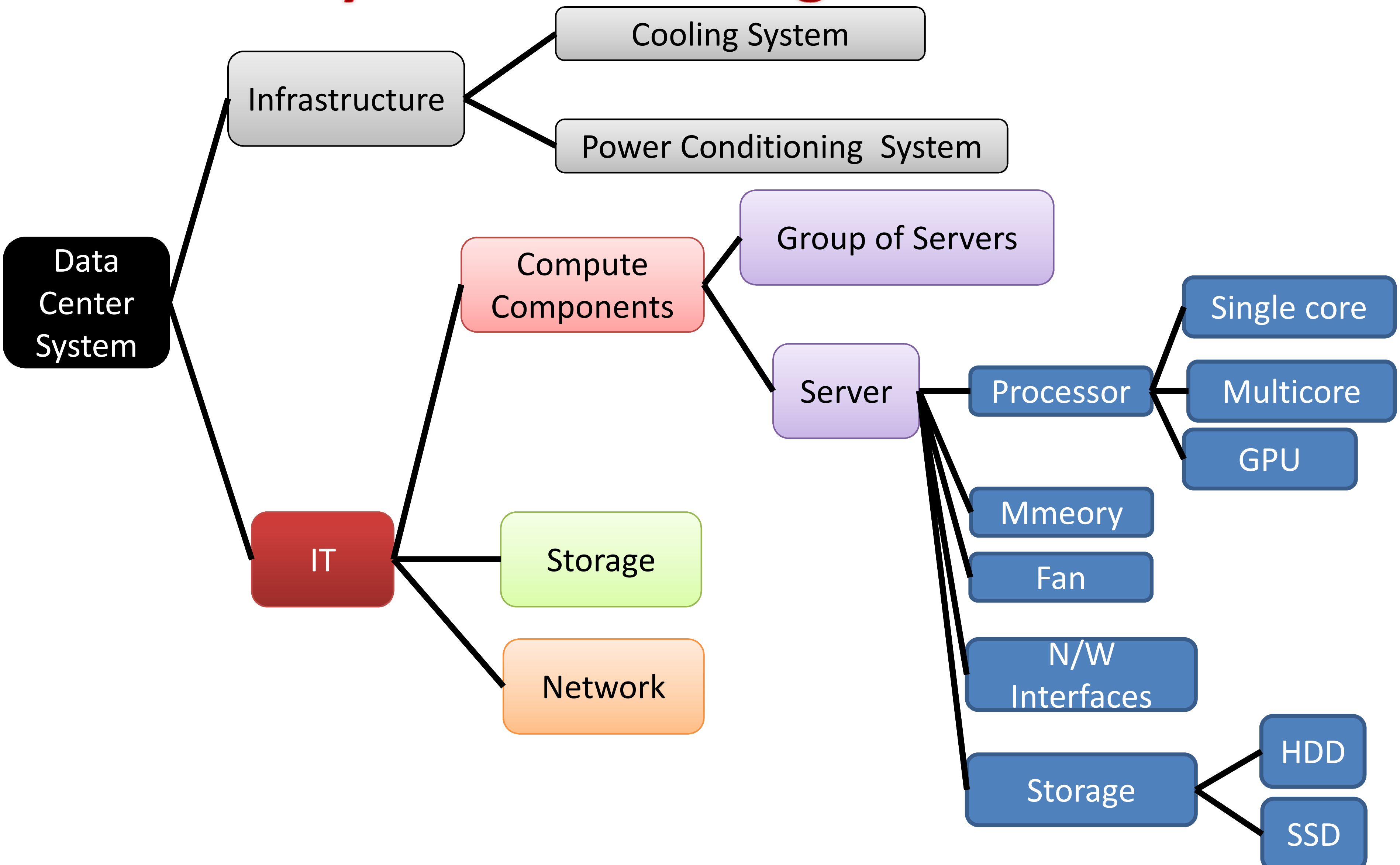
Taxonomy of EC Modeling and Prediction in Data Center



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



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

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

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

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

- More detailed elaborated I/O

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

Aggregate View of Server Energy Model

- More detailed using P.T of individual components

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

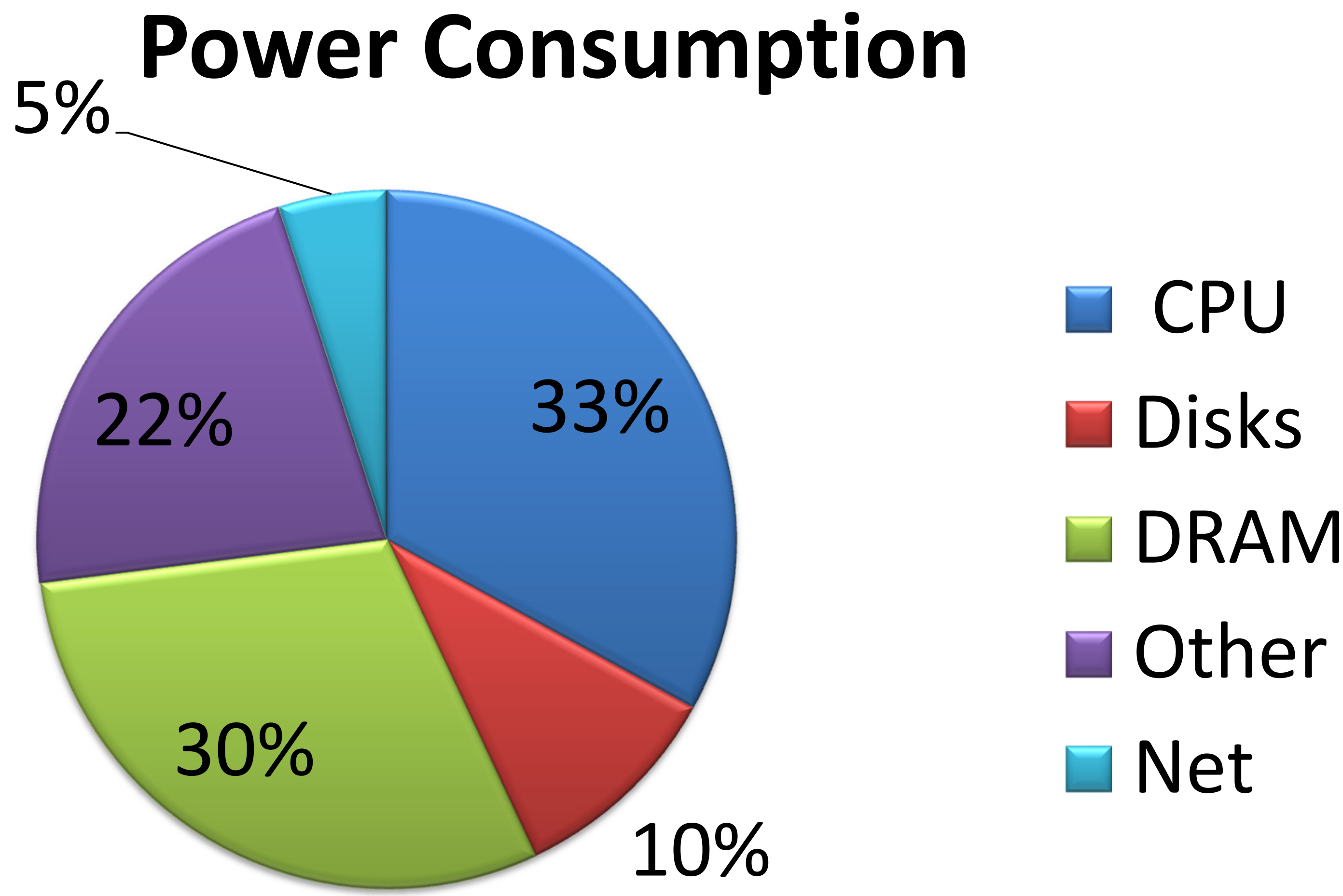
- Power model using resource utilization at time t

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

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

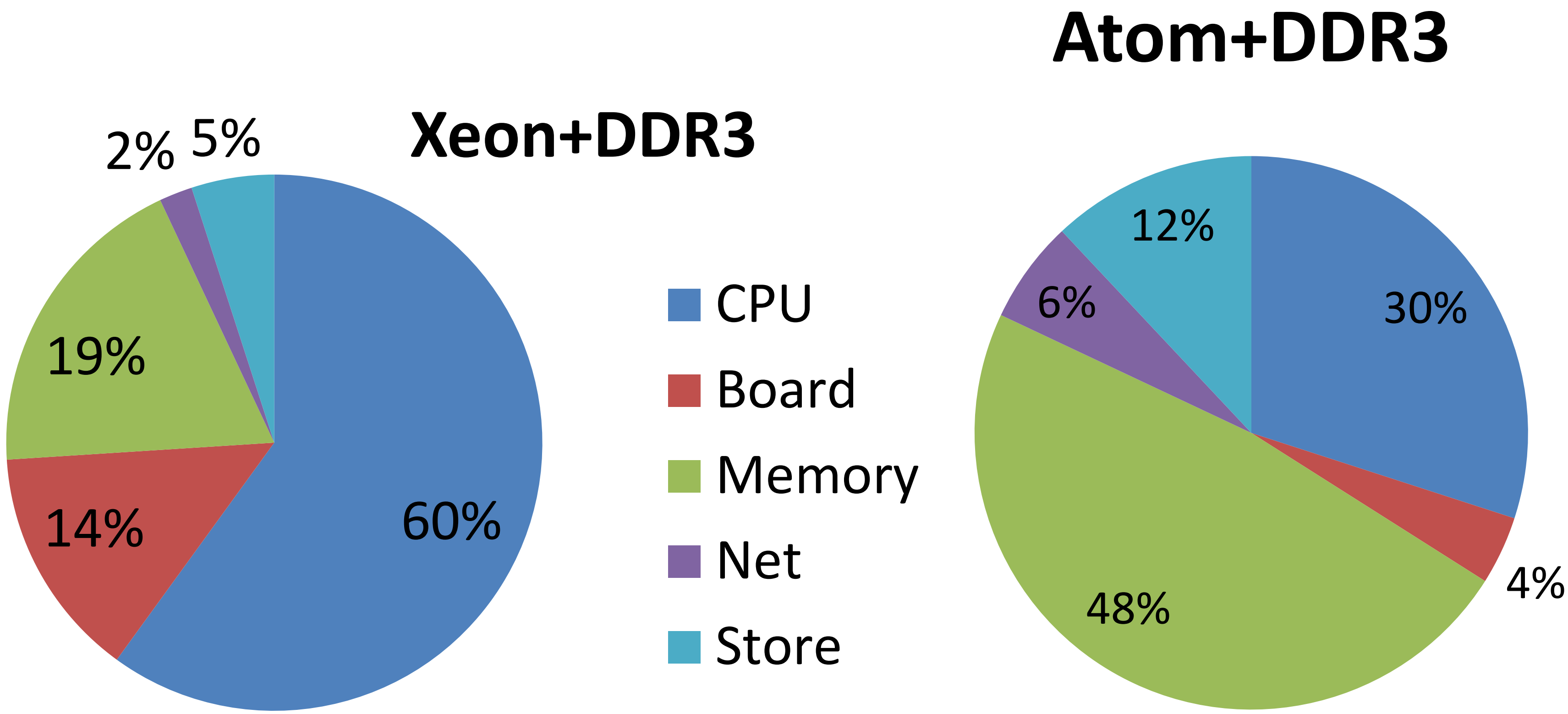
$$E_{system} = A_0(E_{proc} + E_{mem}) + A_1E_{em} + A_2E_{board} + A_3E_{hdd}$$

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

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

- Expanded power modeling using usage of VMs

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

System Utilization based Power Model

- CPU is largest power consumer

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

- PC of Blade server linear model

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

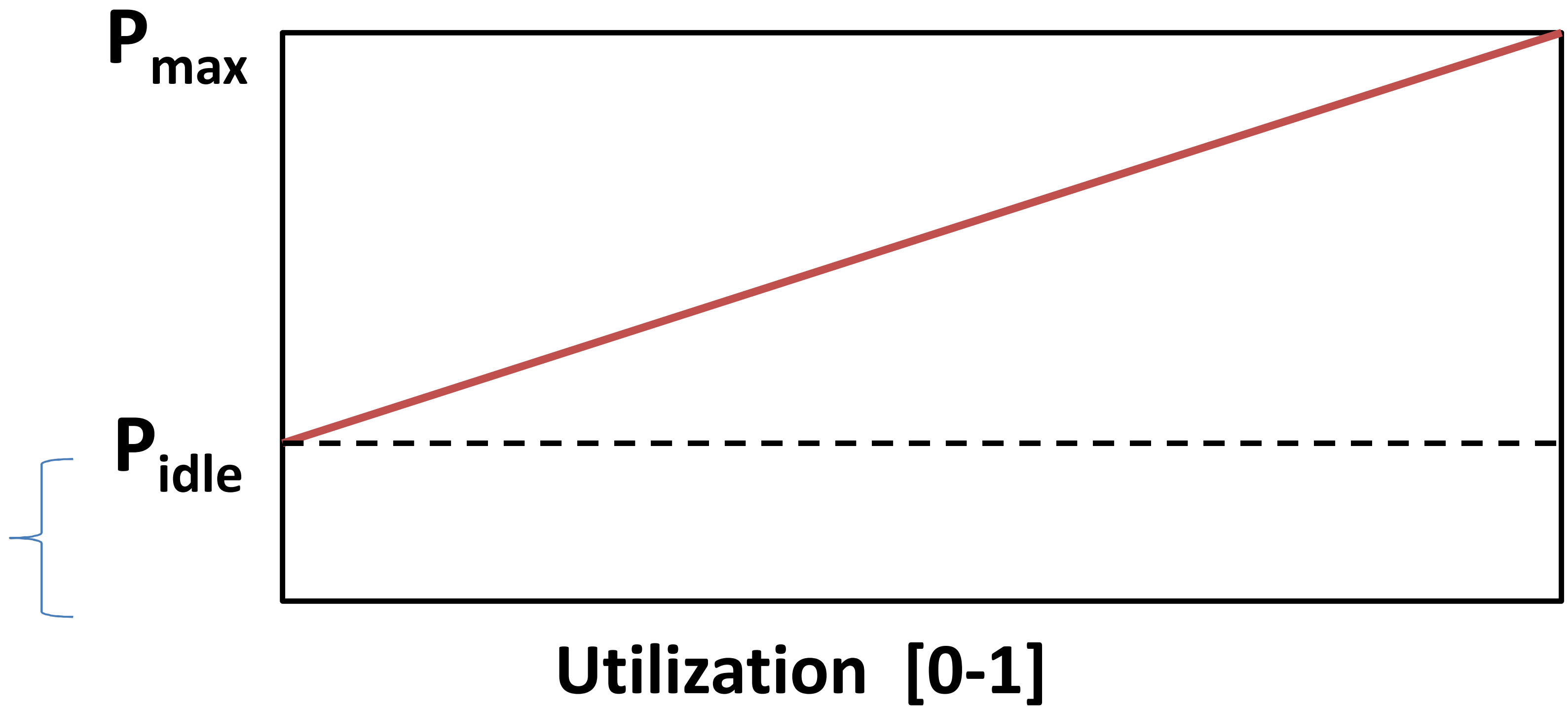
- **Most used PC model of CPU/Server, PC of server at utilization u**

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

System Utilization based Power Model

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

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

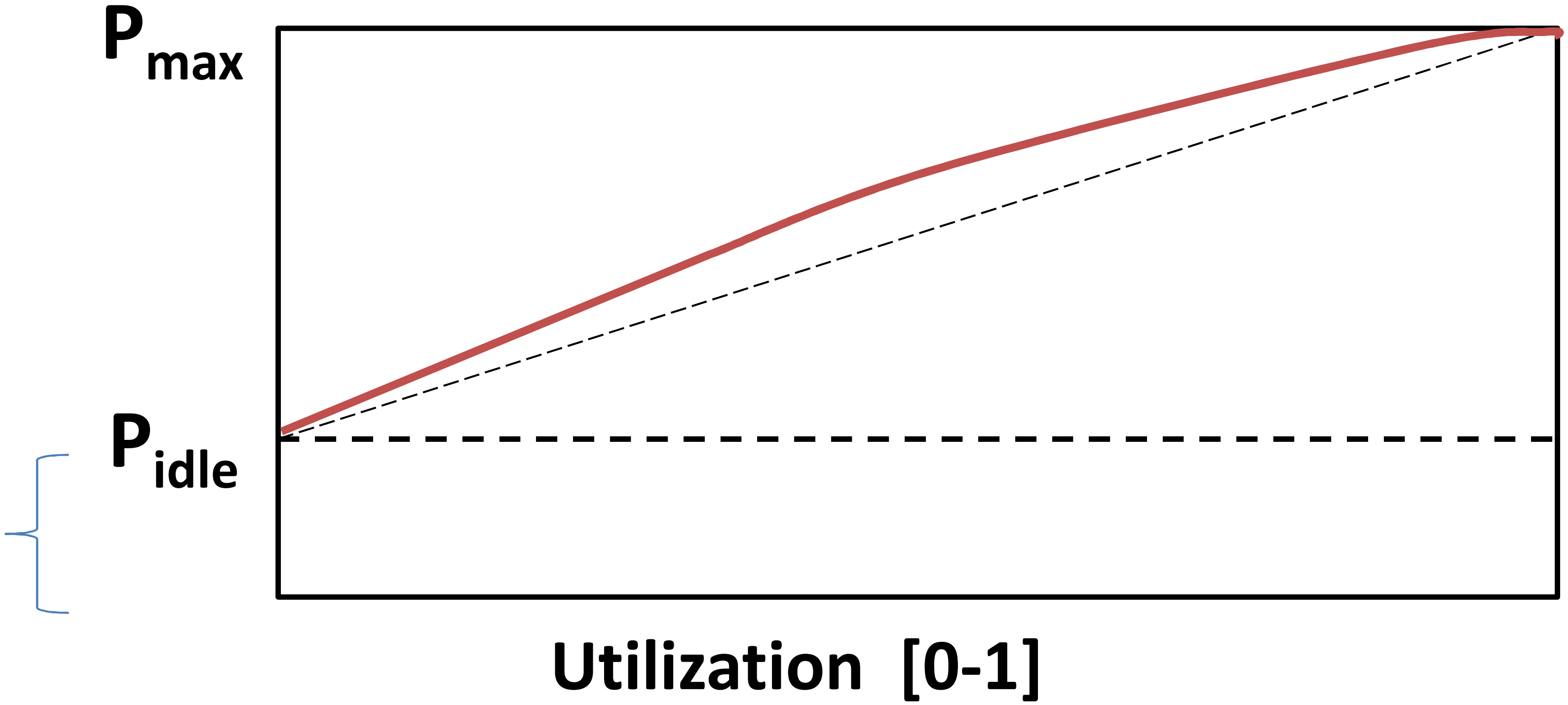


System Utilization based Power Model

- Another PC model of at utilization u

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

- The value of r≈1.4 Experimentally



Normalized Utilization based Power Model

- Normalized PC

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

  - **P_{busy}** is PC when U =1

- Another model

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

  - **Where** α, β are server dependent parameter

- **Modified Popular PC model**

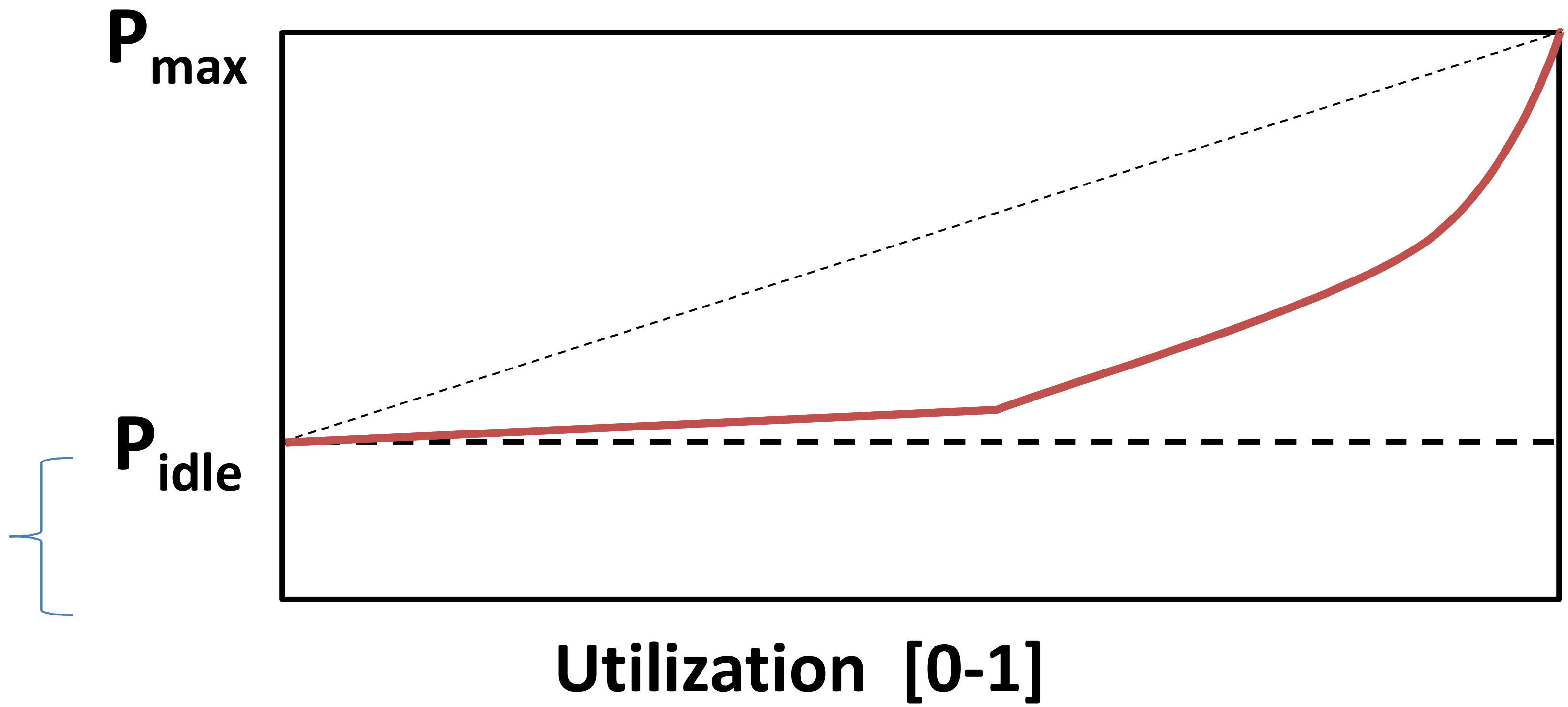
$$P = P_{idle}+ b_i(t)^{\alpha}/A \implies P= Ps + k f^{3}$$

  - A, P_{idle}, α are constant depend on server, α =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 $P_n=\Sigma_{j=1}^{n} P_c(j)$
- Another model with diff levels of core speed

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

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

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

if $\rho=1$ ➔ $P = m.s^{\alpha} + P_{static}$

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

$$P_{proc}= P_{base}+ (C.f^3+D) +(\Sigma_{i=1}^{3}g_iL_i)+g_mM$$

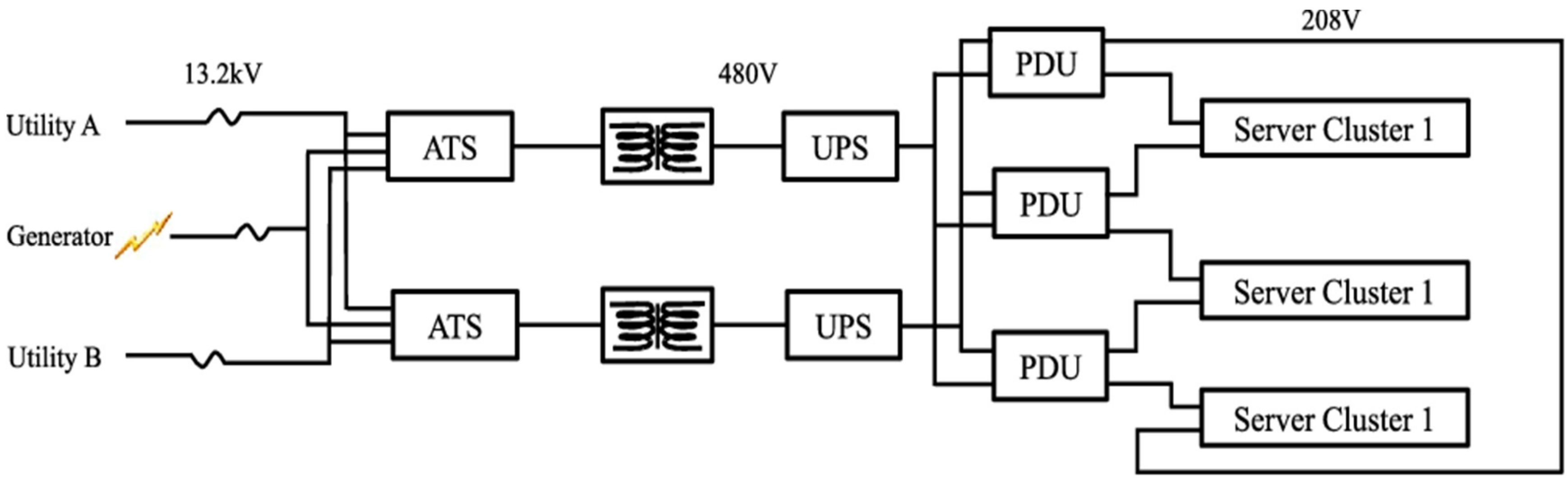
  - Li, ith level cache miss, i=1,2,3, M is memory access

Modeling PC of Memory

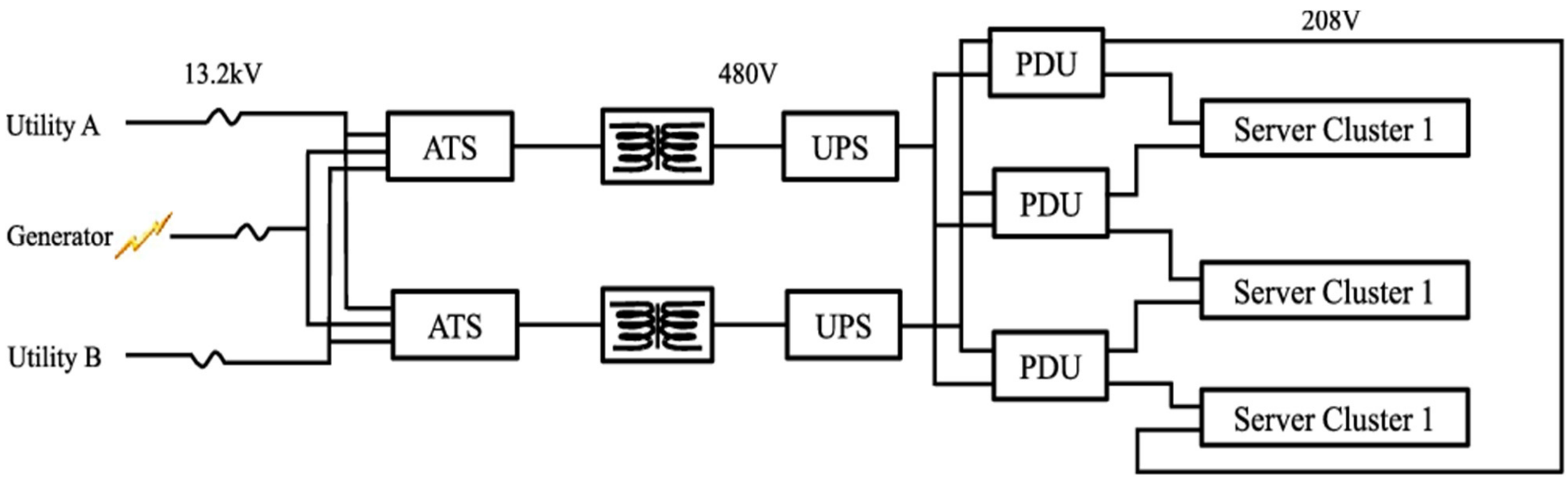
- Memory DRAM: $P_{dm}=P_{st\_dm}+\alpha_1.N_{rd}+\alpha_2.N_{wr}$
- $P_{mem}=D.S.R.\sigma+E_{rw}\rho_{rw}+D.E_{ap}f_{ap}$
  - D chip/subset, S subset/rank, R rank/Chanel
  - σ is static PC of DRAM, ρ_{rw} rd/wr BW/channel, E_{rw} is Energy per rw/bit, f_{ap} is freq of active recharge, E_{ap} is Energy/pre-charge

Modeling EC of Power Conditioning System

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

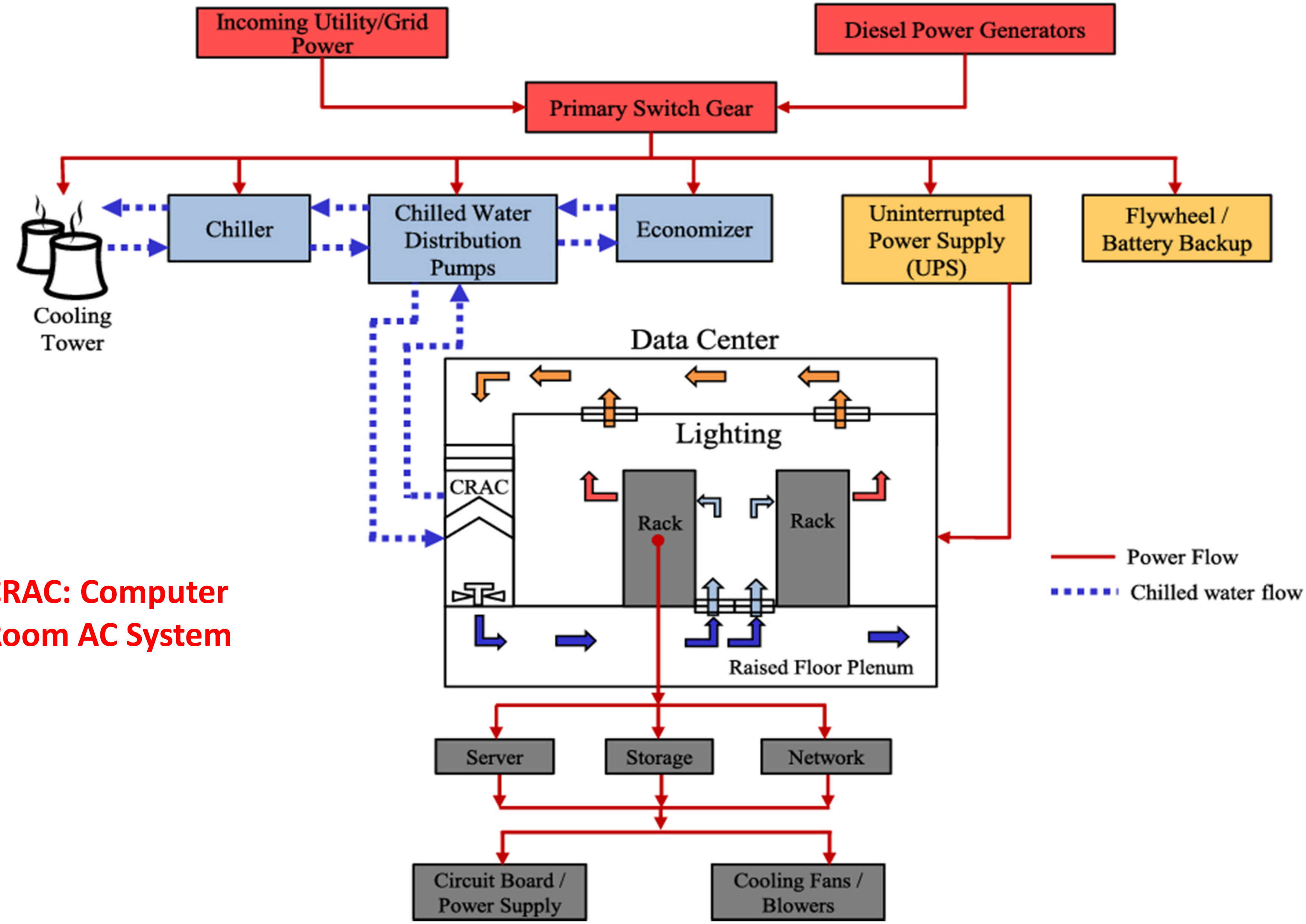


Power Conditioning System



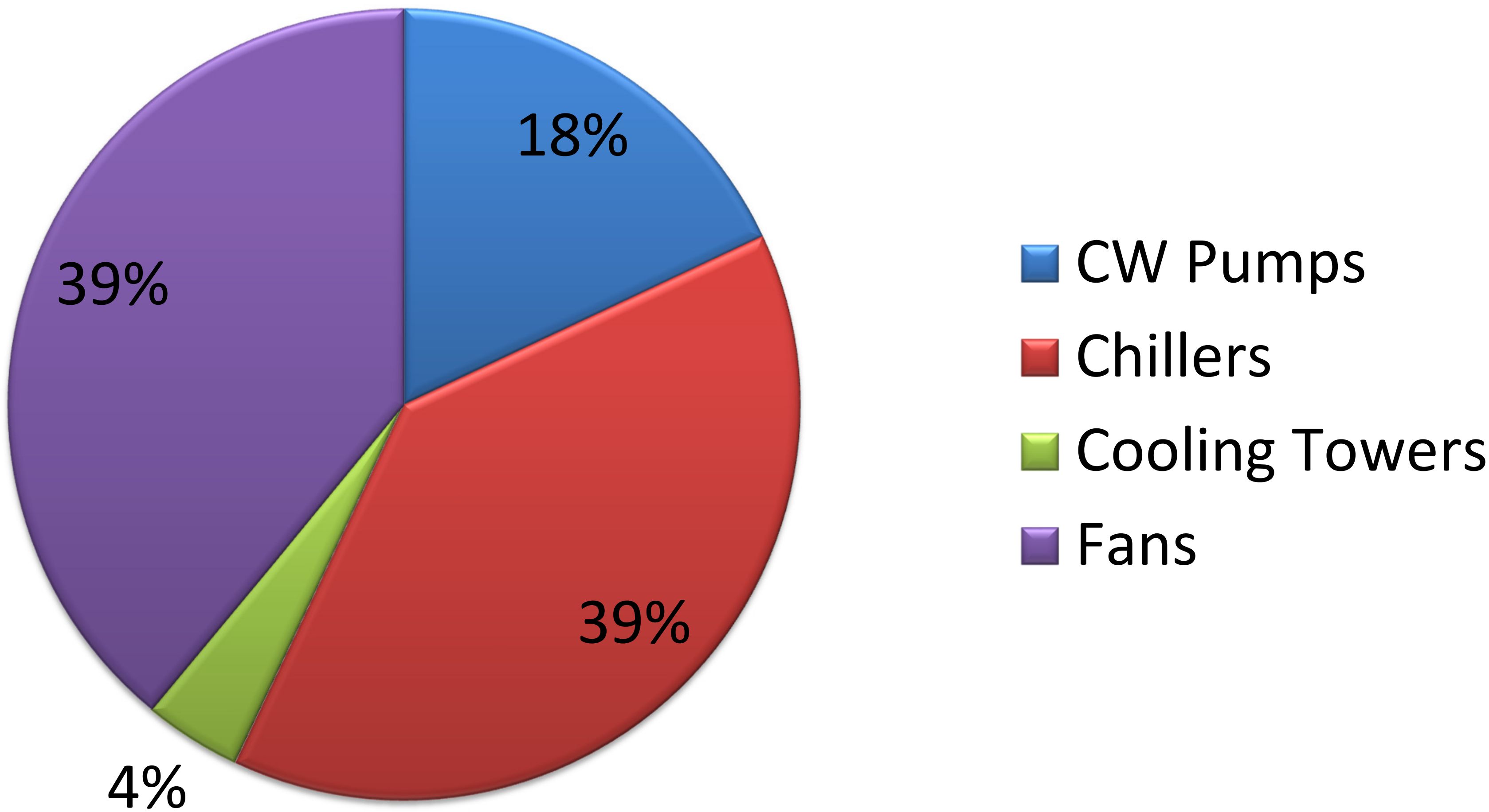
- $P_{pdu\_loss} = P_{pdu\_idle} + \pi_{pdu}(\Sigma_N P_{srv})^2$
  - π_{pdu} is power loss co-efficient
- $P_{ups\_loss} = P_{ups\_idle} + \pi_{ups}(\Sigma_M P_{pdu})$
- PDU waste 3%, where UPS waste 9%

Power Flow in typical Data Center



CRAC: Computer Room AC System

Typical Power Breakdown of HVAC of Data Center



HVAC :Heating, Ventilation and Cooling System

Energy Efficient System: Design and Management

- **Point to consider**

✓ Energy efficient Infrastructure

✓ Energy Model of Infrastructure

  – Blades/Server Machine CPU, Memory

1. **Energy Efficient Scheduling**
   – **How to manage the Jobs**