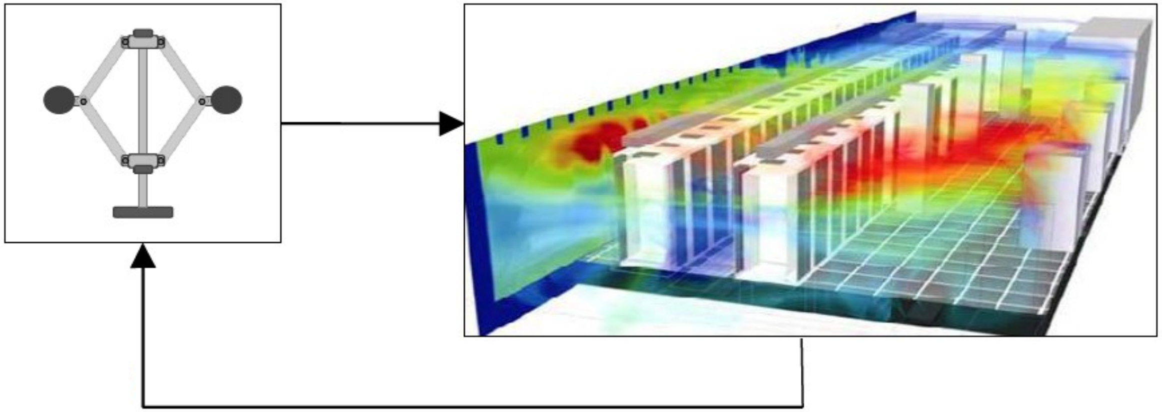
Models and Control Strategies for Data Center Energy Efficiency

Luca Parolini
Bruce Krogh
Bruno Sinopoli
Electrical and Computer Engineering
Carnegie Mellon University

Special Thanks to Zhikui Wang, HP Labs



LCCC Workshop
Lund, Sweden
December 5-7, 2012

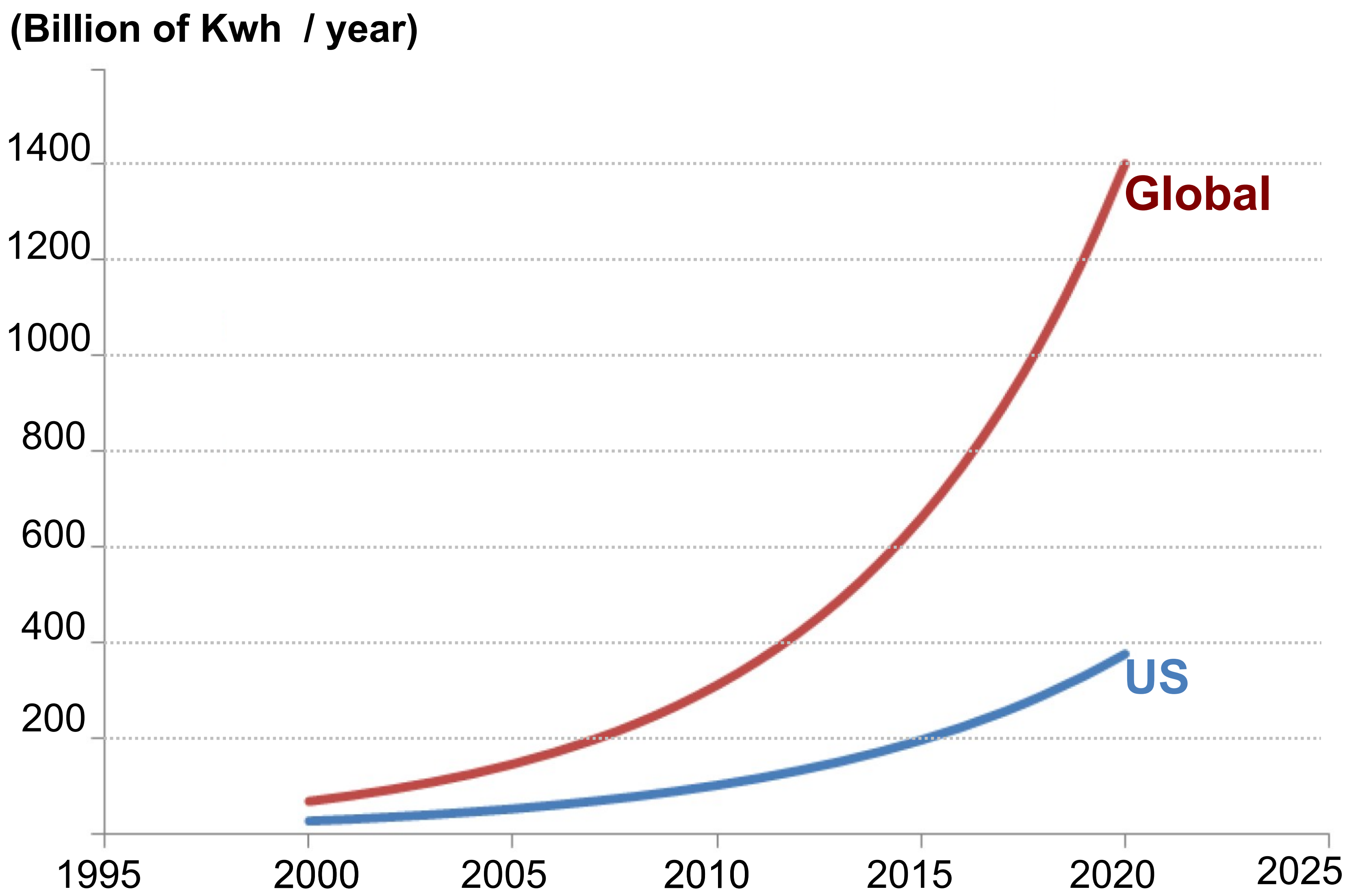
Data center examples

- **Facebook's data center in North Carolina, US**
  - 450$ million project
  - ~28.000 m^2 (300.000 ft^2)
  - Operated by 35 - 45 full-time employees
- **Racks**
  - Contain 42 (1U) servers in a rack
  - 1 server: 480mm x 800mm x 44mm

	Idle power	***Peak power***
Server	200 W	350 W
Rack	8.4 kW	~15 kW

http://blog.cluttr.be/2011/01/26/the-case-for-data-center-power-efficiency/
http://www.datacenterknowledge.com/

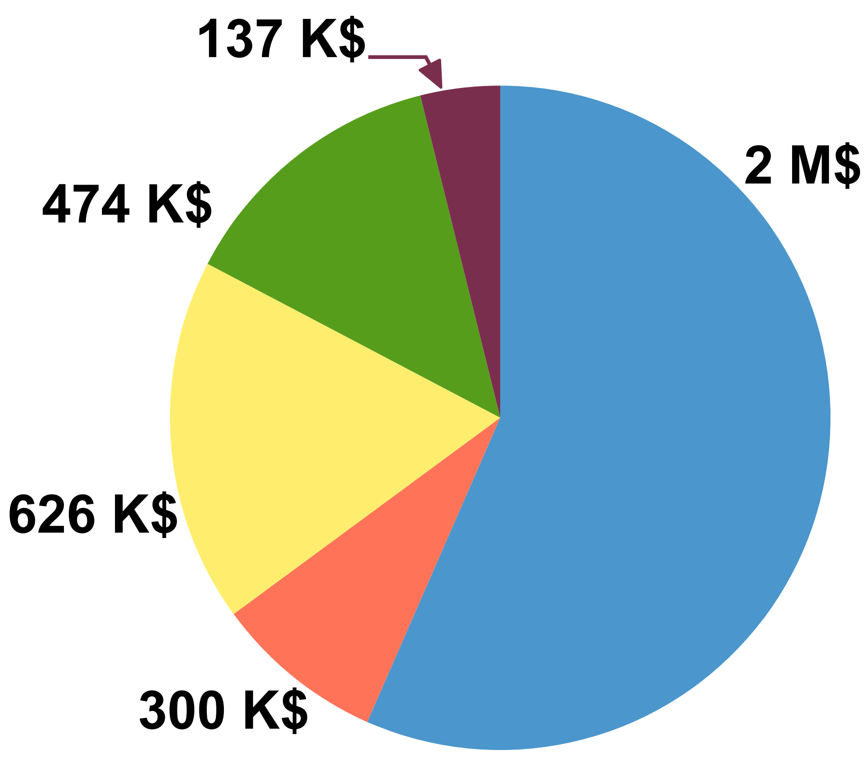
Electricity consumption



C. L. Belady, Projecting annual new datacenter construction market size, Microsoft, 2011

Monthly operating cost

- **Large-scale facility, 50K servers**
  - Facility cost amortized over 10 years
  - Server cost amortized over 3 years
  - Servers account for 70% of total power consumption

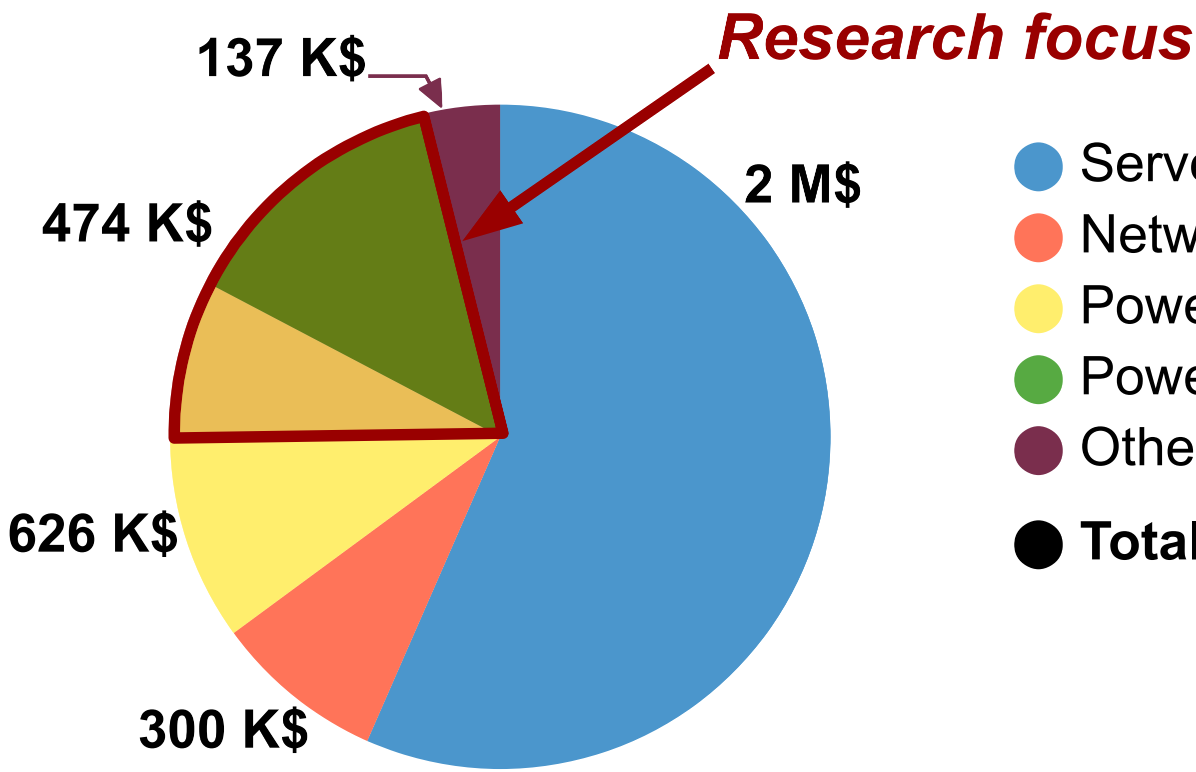


Server (hardware cost)	*57%*
Networking equipment	*8%*
Power distribution and cooling	*18%*
Powering cost	*13%*
Other infrastructure	*4%*

- **Total:** ***3.5 M$***

Monthly operating cost

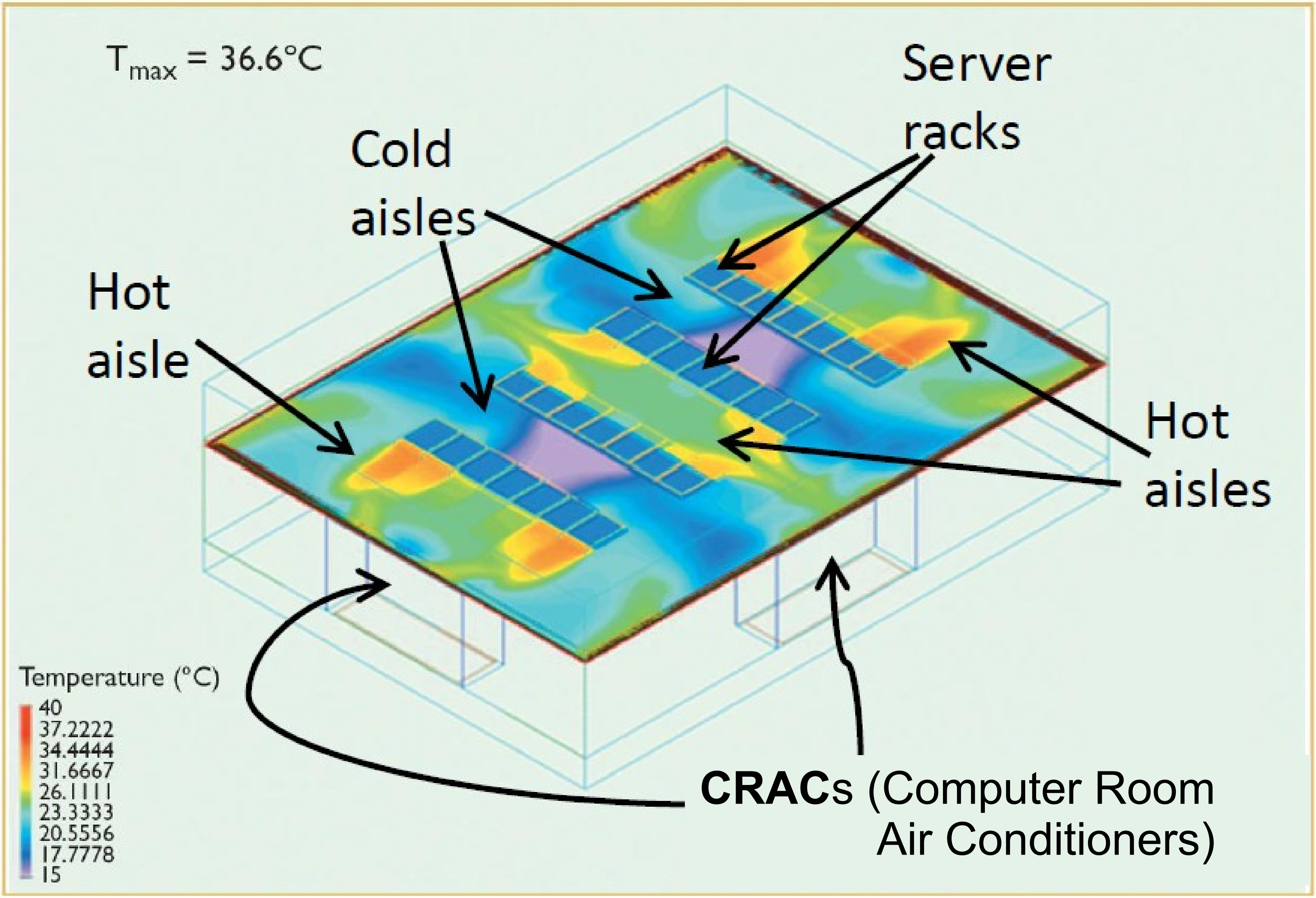
- **Large-scale facility, 50K servers**
  - Facility cost amortized over 10 years
  - Server cost amortized over 3 years
  - Servers consume 70% of total power consumption



Server (hardware cost)	*57%*
Networking equipment	*8%*
Power distribution and cooling	*18%*
Powering cost	*13%*
Other infrastructure	*4%*

Total: ***3.5 M$***

Temperature distribution

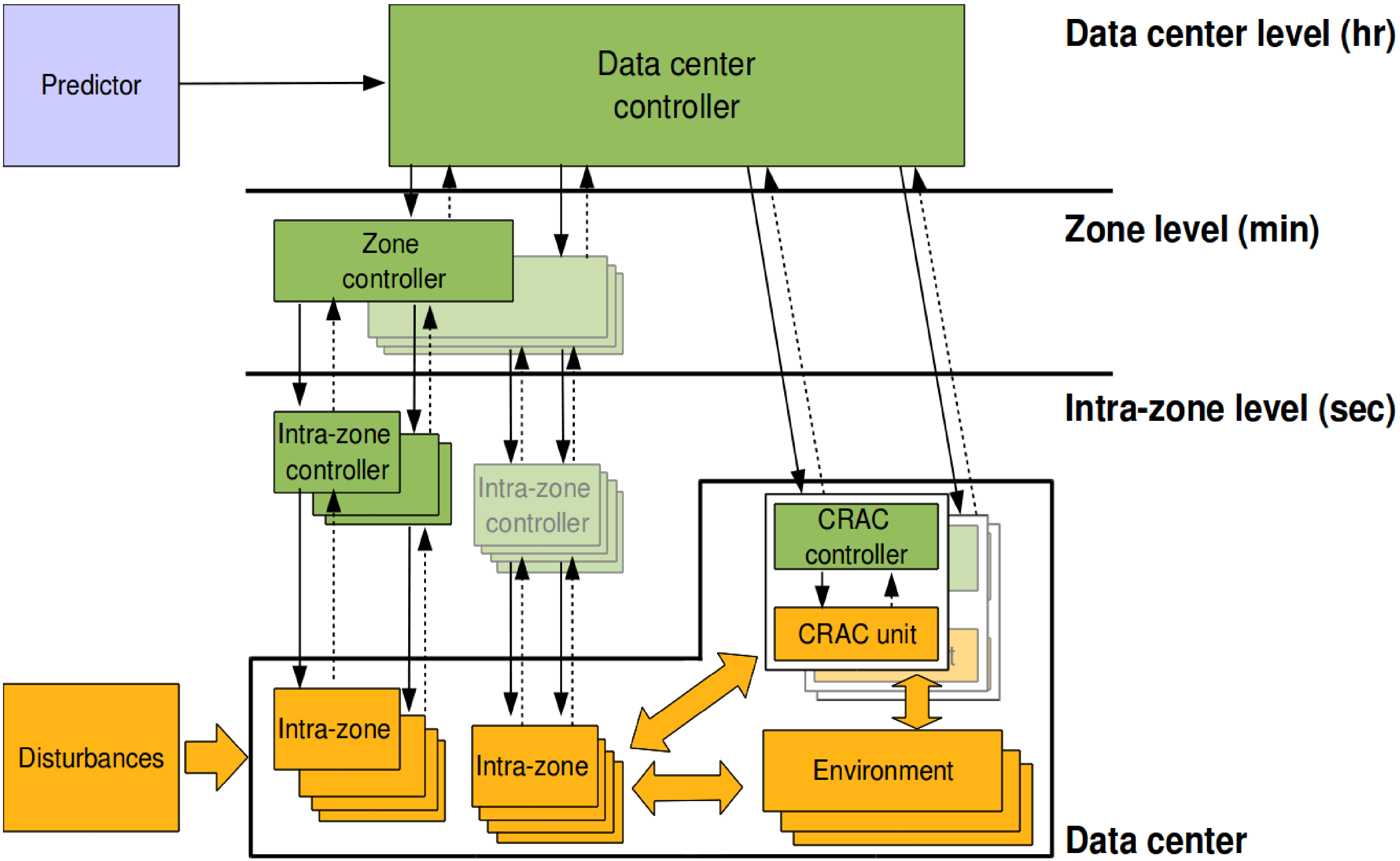


R. K. Sharma *et al.* “Balance of Power: Dynamic Thermal Management for Internet Data Centers”, IEEE Internet Computing 2005

Outline

- Introduction
- **Control-oriented model**
- Control strategies
- Simulation results
- Conclusion and future work

Hierarchical control approach



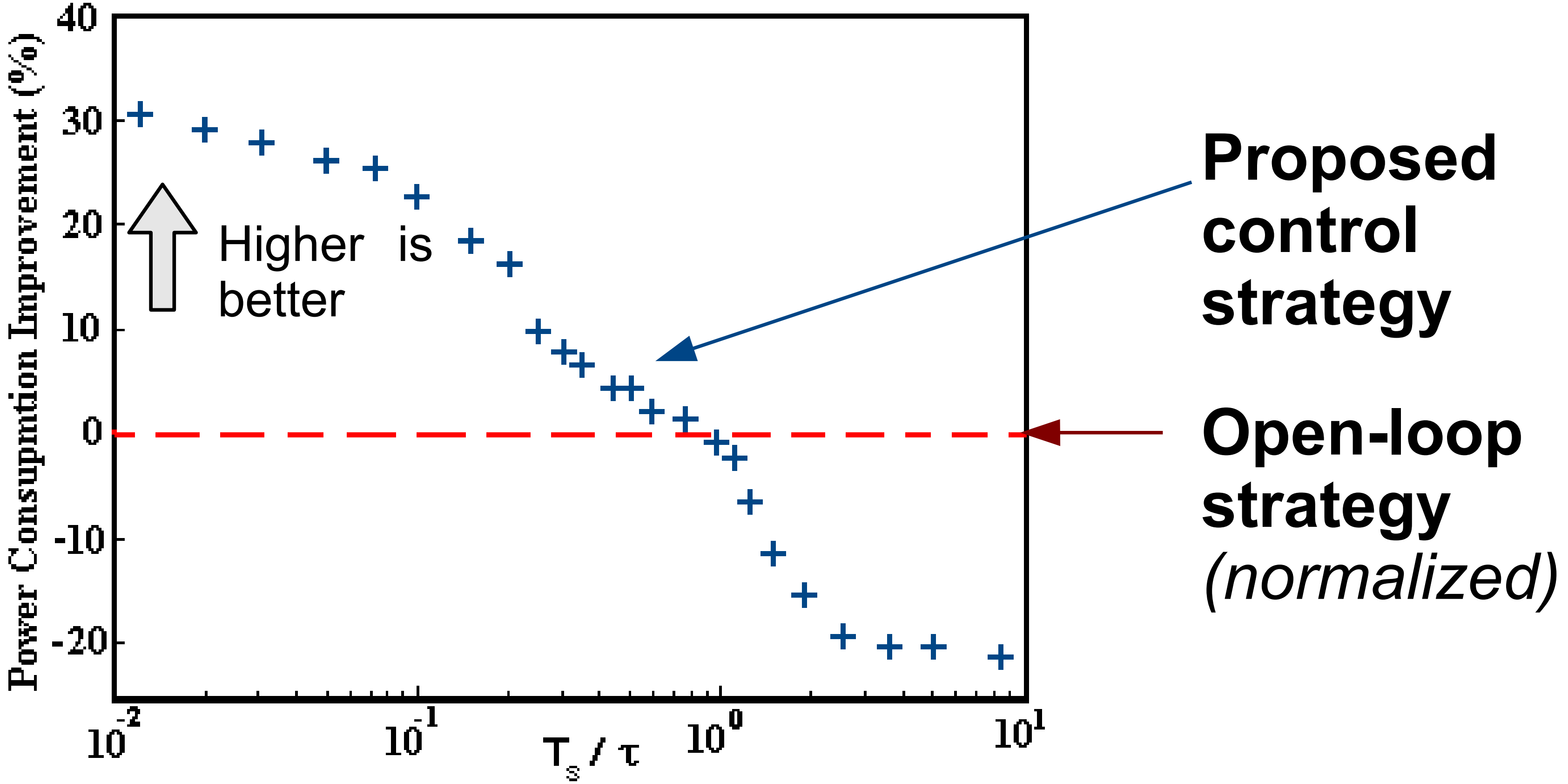
Zone level control

- **Operates at a faster time-scale**
- **Decides how many servers in the zone should be turned on**
- **Control actions based on**
  - Desired workload execution rate (predictive control)
  - Current resource use in the zone (reactive control)
- **Considers**
  - Time to turn servers on: T_S
  - Variability of workload arrival rate

Aghajani, *et al.* "Dynamic Power Allocation in Server Farms: a Real Time Optimization Approach", CDC 2010

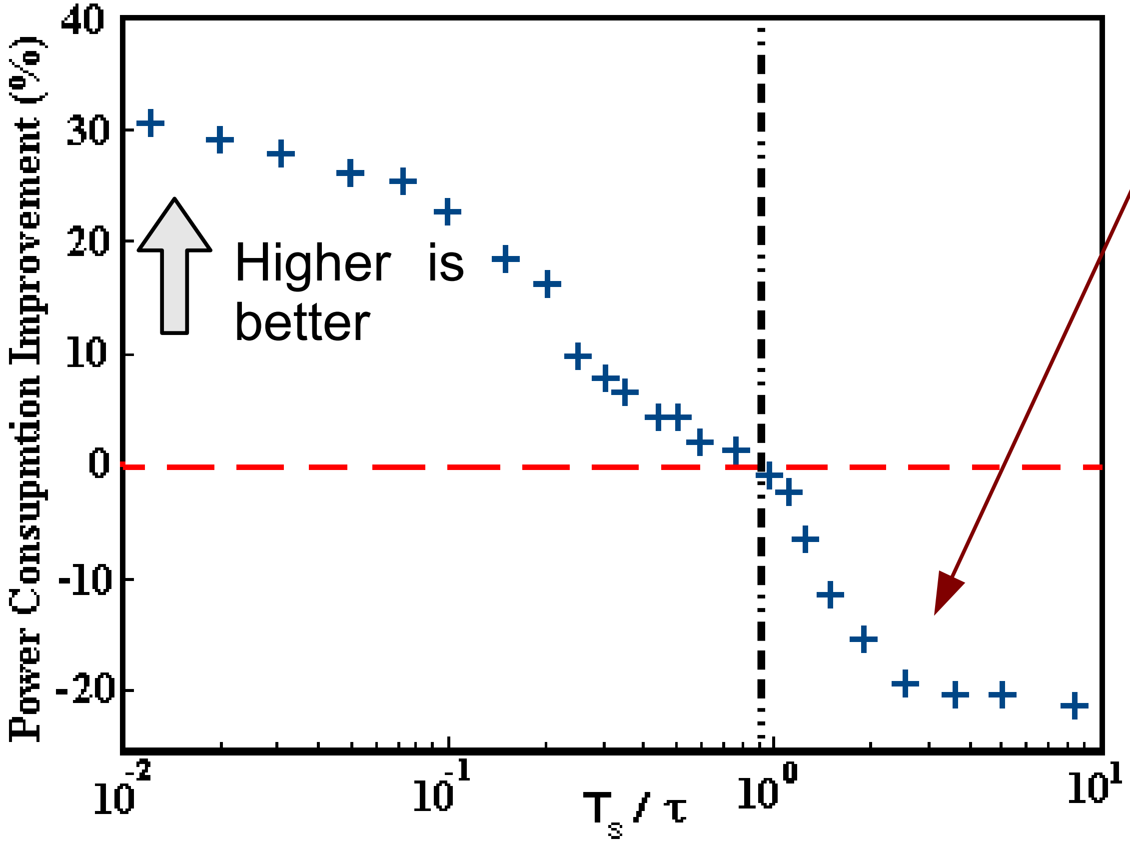
Zone level control - simulation results

- **Compare proposed control approach against optimal open-loop strategy**
  - *τ:* expected time interval over which the workload arrival rate is constant



Zone level control - simulation results

- **Compare proposed control approach against optimal open-loop strategy**
  - *τ:* expected time interval over which the workload arrival rate is constant



- **Workload arrival rate varies too fast with respect to the time to turn servers on**
- **Best solution is to never turn servers on and off**

Modeling approach: Data Center Level

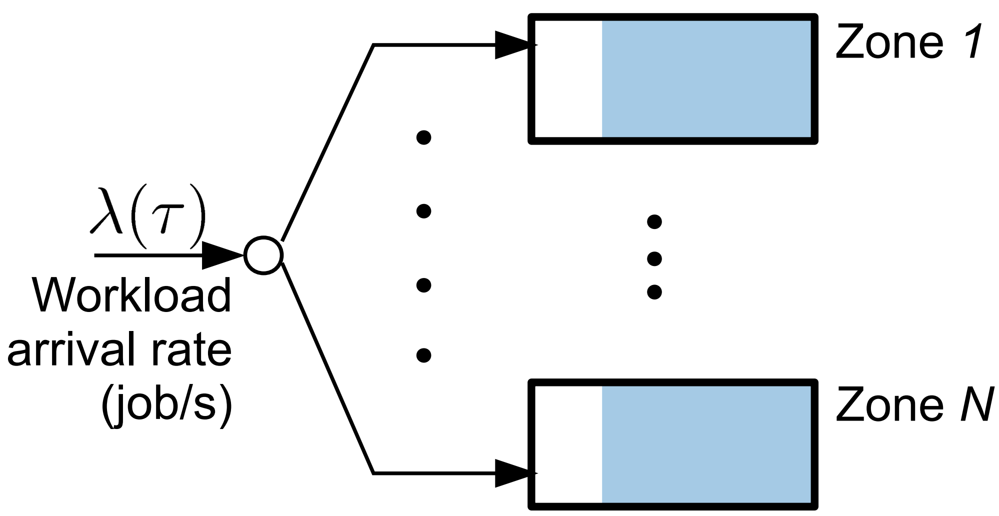
- **Focuses on processes in the hours time scale**

- **Groups servers into zones**
  - Power consumption of a zone is proportional to the amount of workload executed
  - Data are always available
  - Neglect the time to turn servers On and Off
    - Much shorter than the controller sampling-time

- **Considers**
  - Computational and thermal dynamics
  - Nonlinear efficiency of the CRAC units
    - Service level agreements (SLAs) with users and the power-grid

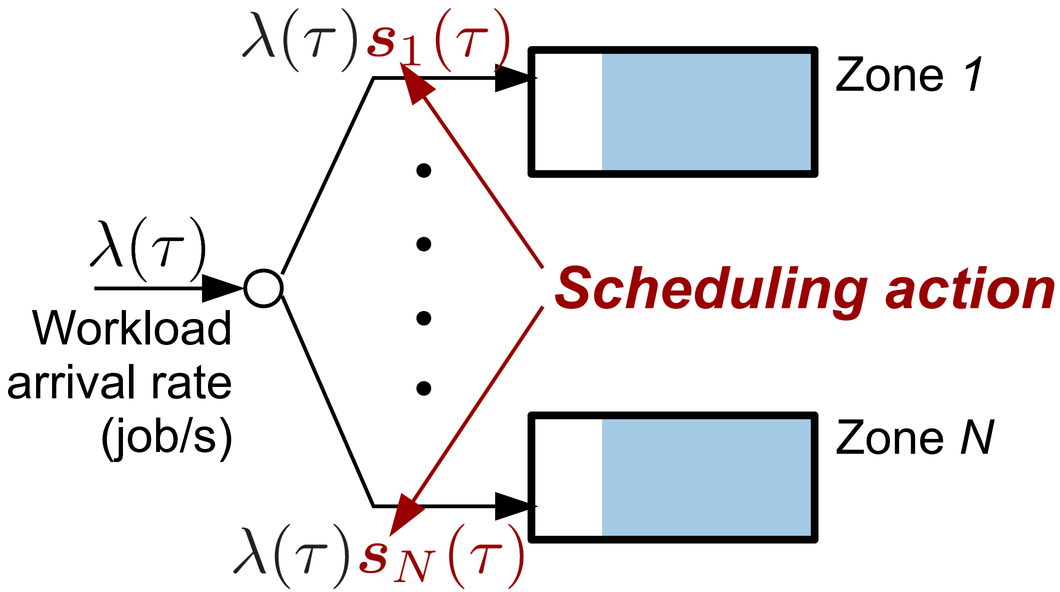
Computational network

- **Describes the evolution of resource usage in the data center**
- **Based on a fluid approximation of the workload execution and arrival processes**
  - First-order approximation of a queuing system



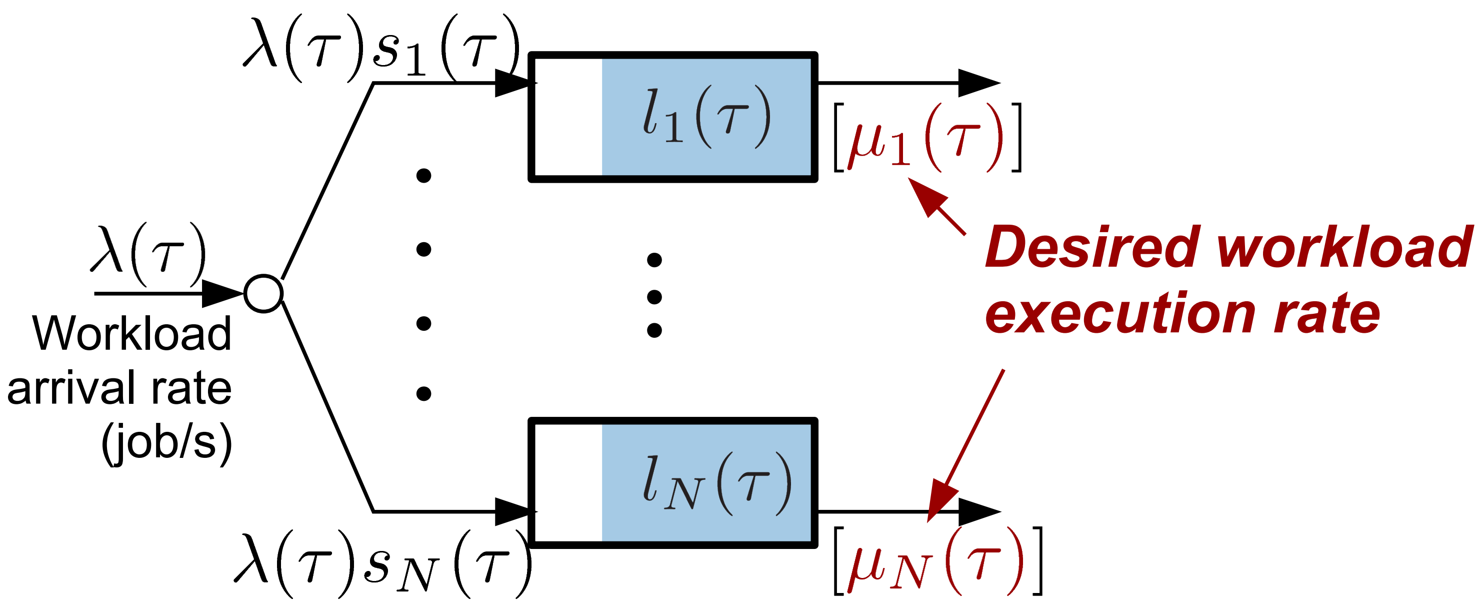
Computational network

- **Describes the evolution of resource usage in the data center**
- **Based on a fluid approximation of the workload execution and arrival processes**
  - First-order approximation of a queuing system



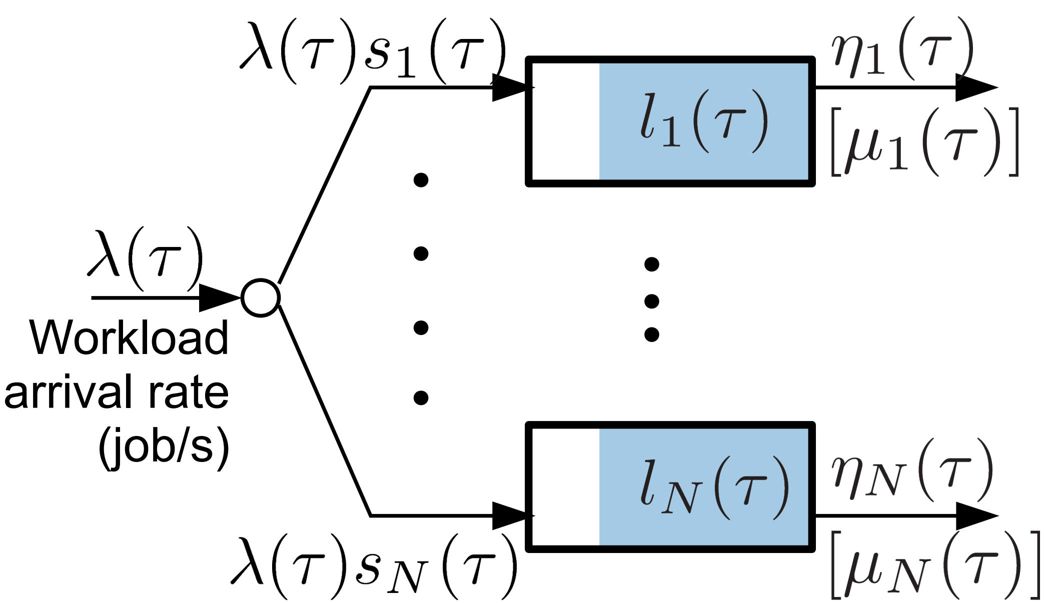
Computational network

- **Describes the evolution of resource usage in the data center**
- **Based on a fluid approximation of the workload execution and arrival processes**
  - First-order approximation of a queuing system



Computational network

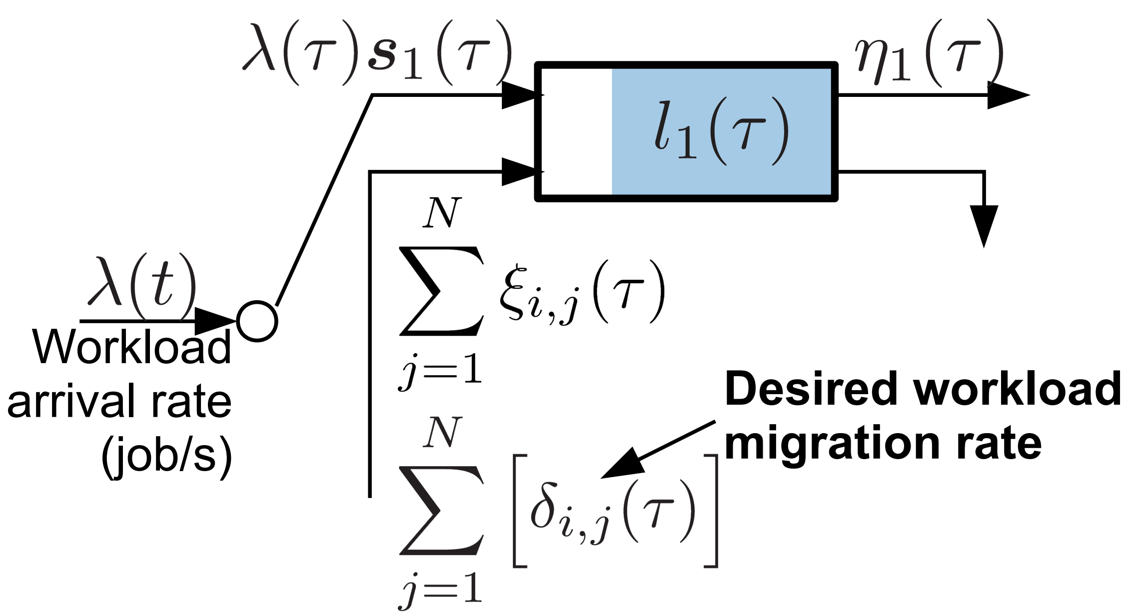
- **Describes the evolution of resource usage in the data center**
- **Based on a fluid approximation of the workload execution and arrival processes**
  - First-order approximation of a queuing system



$$a_1(\tau) = \lambda(\tau)s_1(\tau)$$
$$d_1(\tau) = \eta_1(\tau)$$
$$\dot{l}_1(\tau) = a_1(\tau) - d_1(\tau)$$
$$\eta_1(\tau) = \begin{cases} \mu_1(\tau) & \text{if } l_1(\tau) > 0 \\ & \text{or } a_1(\tau) > \mu_1(\tau) \\ a_1(\tau) & \text{otherwise} \end{cases}$$

Computational network

- **Describes the evolution of resource usage in the data center**
- **Based on a fluid approximation of the workload execution and arrival processes**
  - First-order approximation of a queuing system



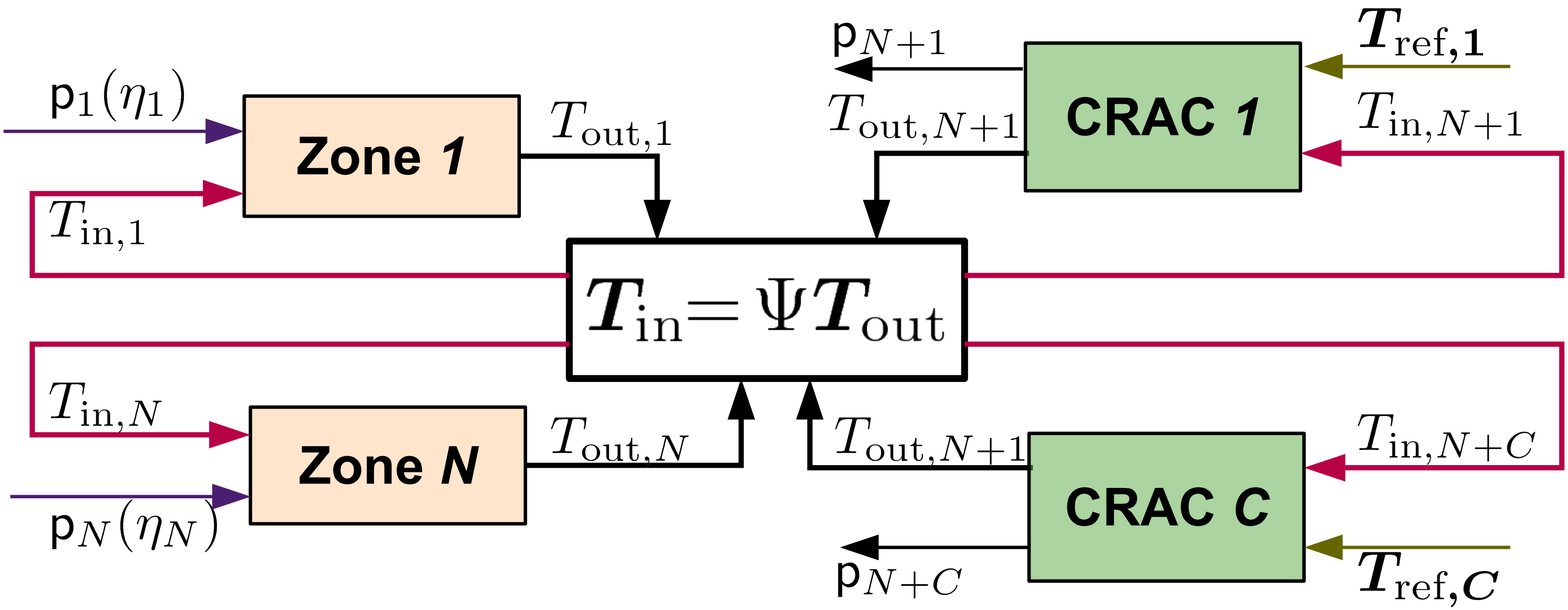
$$a_1(\tau) = \lambda(\tau)\boldsymbol{s}_1(\tau) + \sum_{j=1}^{N} \xi_{1,j}(\tau)$$

$$d_1(\tau) = \eta_1(\tau) + \sum_{j=1}^{N} \xi_{j,1}(\tau)$$

$$\dot{l}_1(\tau) = a_1(\tau) - d_1(\tau)$$

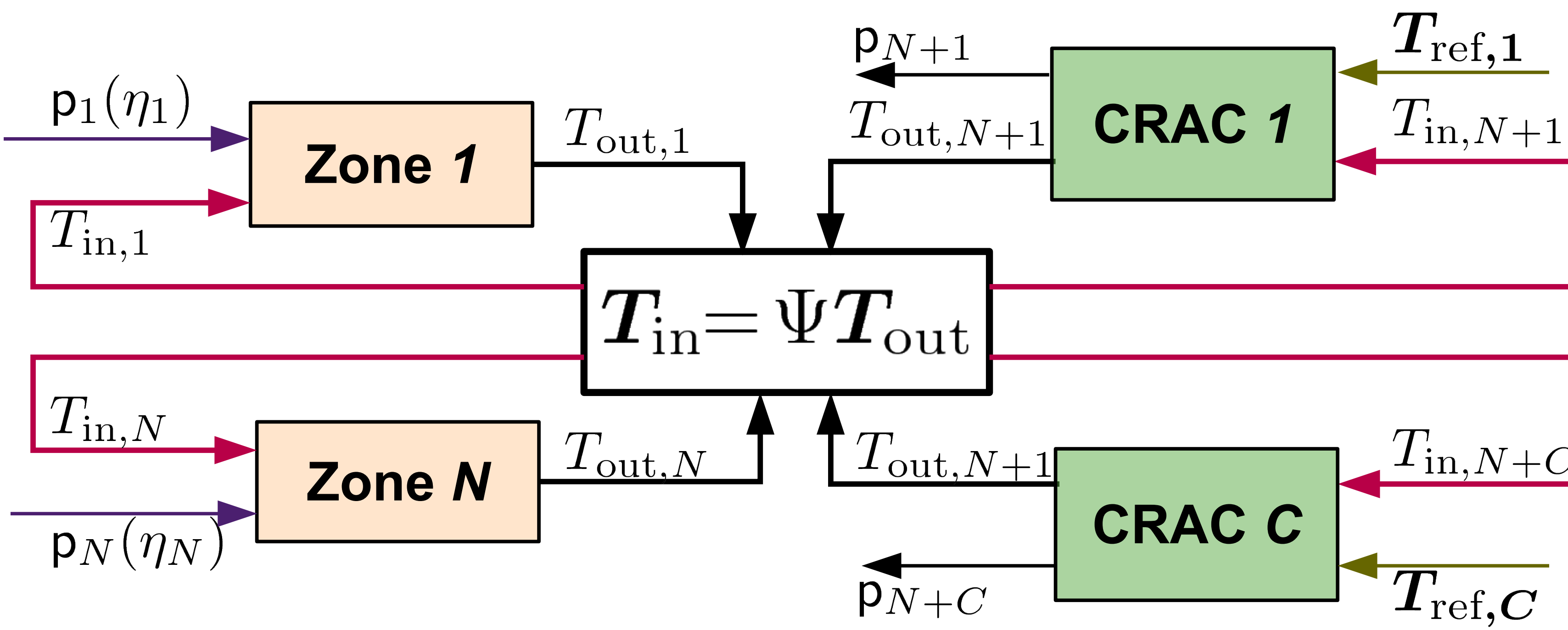
$$\eta_1(\tau) = \begin{cases} \mu_1(\tau) & \text{if } l_1(\tau) > 0 \\ & \text{or } a_1(\tau) > \mu_1(\tau) \\ a_1(\tau) & \text{otherwise} \end{cases}$$

Thermal network



Q. Tang *et al.* "Sensor-based fast thermal evaluation model for energy efficient high-performance data centers", 2006

Thermal network



- **Inlet temperature constraint** $\boldsymbol{T}_{\mathrm{in}}(\tau) = \Psi \boldsymbol{T}_{\mathrm{out}}(\tau) \leq \overline{\boldsymbol{T}_{\mathrm{in}}}$

Heat removed rate (W)

- **CRAC power consumption** $p_i(t) = \dfrac{\dot{Q}_i(t)}{COP_i(T_{\mathrm{out},i}(t))}$

Q. Tang *et al.* "Sensor-based fast thermal evaluation model for energy efficient high-performance data centers", 2006

Variables at step *k*

			Variables
Input	Controllable	Job scheduling	$\boldsymbol{s}(k)$
		Resource allocation	$\boldsymbol{\mu}(k)$
		Job migration	$\boldsymbol{\delta}(k)$
		CRAC unit reference temperature	$\boldsymbol{T}_{\mathrm{ref}}(k)$
	Uncontrollable	Job arrival	$\boldsymbol{\lambda}(k)$
Output	Zone power consumption		$\boldsymbol{p}_N(k)$
	Power consumption of CRAC nodes		$\boldsymbol{p}_C(k)$
	Input temperatures of zones		$\boldsymbol{T}_{\boldsymbol{in}}(k)$
State	Number of jobs in every zones		$\boldsymbol{l}(k)$
	Output temperatures of CRACs and zones		$\boldsymbol{T}_{out}(k)$

Outline

- Introduction
- Control-oriented model
- **Control strategies**
- Simulation results
- Conclusion and future work

Control strategies

- **Baseline**
  - Open-loop strategy
  - Sets control variables for the worst-case scenario
- **Uncoordinated**
  - Manages the computational and the thermal resources independently
  - Neglects the thermal-computational coupling in the data center
- **Coordinated**
  - Manages the computational and the thermal resources in a single optimization problem

Baseline & uncoordinated approaches

- **Baseline**

$$\boldsymbol{\mu}(\tau) = \overline{\boldsymbol{\mu}} \qquad \boldsymbol{\delta}(\tau) = \mathbf{0} \qquad \boldsymbol{s}(\tau) = \mathbf{1}\frac{1}{N} \qquad \boldsymbol{T}_{\text{ref}}(\tau) = \underline{\boldsymbol{T}}_{\text{ref}}$$

- **Uncoordinated controller**

$$\begin{aligned}
\min_{\mathcal{M},\mathcal{S},\mathcal{D}} \quad & \sum_{h=k}^{k+\mathcal{T}-1} \mathbf{1}^T \hat{\mathbf{p}}_{\mathcal{N}}(h|k) \\
\textbf{s.t.} \quad & \hat{\boldsymbol{l}}(k|k) = \boldsymbol{l}(k) \\
& \text{for all } h = k, \ldots, k+\mathcal{T}-1 \\
& \text{computational dynamics} \\
& \text{QoS constraints} \\
& \mathbf{0} \leq \hat{\boldsymbol{\mu}}(h|k) \leq \overline{\boldsymbol{\mu}}, \quad \hat{\boldsymbol{\delta}}(h|k) = \mathbf{0} \\
& \mathbf{0} \leq \hat{\boldsymbol{s}}(h|k) \leq \mathbf{1}, \quad \mathbf{1}^T \hat{\boldsymbol{s}}(h|k) \leq 1
\end{aligned}$$

Minimize expected zone power consumption

- Computational dynamics
- Quality of service (QoS) constraints
- Control constraints

$$\mathcal{D} = \left\{ \hat{\boldsymbol{\delta}}(k|k), \ldots, \hat{\boldsymbol{\delta}}(k+\mathcal{T}-1|k) \right\} \qquad \mathcal{T}_{ref} = \left\{ \hat{\boldsymbol{T}}_{\text{ref}}(k|k), \ldots, \hat{\boldsymbol{T}}_{\text{ref}}(k+\mathcal{T}-1|k) \right\}$$

$$\mathcal{M} = \left\{ \hat{\boldsymbol{\mu}}(k|k), \ldots, \hat{\boldsymbol{\mu}}(k+\mathcal{T}-1|k) \right\} \quad \mathcal{S} = \left\{ \hat{\boldsymbol{s}}(k|k), \ldots, \hat{\boldsymbol{s}}(k+\mathcal{T}-1|k) \right\}$$

Baseline & uncoordinated approaches

- **Baseline**

$$\boldsymbol{\mu}(\tau) = \overline{\boldsymbol{\mu}} \qquad \boldsymbol{\delta}(\tau) = \mathbf{0} \qquad \boldsymbol{s}(\tau) = \mathbf{1}\frac{1}{N} \qquad \boldsymbol{T}_{\text{ref}}(\tau) = \underline{\boldsymbol{T}}_{\text{ref}}$$

- **Uncoordinated controller**

$$\begin{aligned} \min_{\mathcal{M},\mathcal{S},\mathcal{D}} \quad & \sum_{h=k}^{k+\mathcal{T}-1} \mathbf{1}^T \hat{\mathbf{p}}_{\mathcal{N}}(h|k) \\ \textbf{s.t.} \quad & \hat{\boldsymbol{l}}(k|k) = \boldsymbol{l}(k) \\ & \text{for all } h = k, \ldots, k+\mathcal{T}-1 \\ & \text{computational dynamics} \\ & \text{QoS constraints} \\ & \mathbf{0} \leq \hat{\boldsymbol{\mu}}(h|k) \leq \overline{\boldsymbol{\mu}}, \quad \hat{\boldsymbol{\delta}}(h|k) = \mathbf{0} \\ & \mathbf{0} \leq \hat{\boldsymbol{s}}(h|k) \leq \mathbf{1}, \quad \mathbf{1}^T \hat{\boldsymbol{s}}(h|k) \leq 1 \end{aligned}$$

$$\xrightarrow{\{\hat{\mathbf{p}}_{\mathcal{N}}(h|k)\}}$$

$$\begin{aligned} \min_{\mathcal{T}_{ref}} \quad & \sum_{h=k}^{k+\mathcal{T}-1} \mathbf{1}^T \hat{\mathbf{p}}_{\mathcal{C}}(h|k) \\ \textbf{s.t.} \quad & \hat{\boldsymbol{T}}_{\text{out}}(k|k) = \boldsymbol{T}_{\text{out}}(k) \\ & \text{for all } h = k, \ldots, k+\mathcal{T}-1 \\ & \text{thermal dynamics} \\ & \underline{\boldsymbol{T}}_{\text{ref}} \leq \hat{\boldsymbol{T}}_{\text{ref}}(h|k) \leq \overline{\boldsymbol{T}}_{\text{ref}} \\ & \hat{\boldsymbol{T}}_{\text{in}}(h+1|k) \leq \overline{\boldsymbol{T}}_{\text{in}} \end{aligned}$$

$$\mathcal{D} = \left\{ \hat{\boldsymbol{\delta}}(k|k), \ldots, \hat{\boldsymbol{\delta}}(k+\mathcal{T}-1|k) \right\} \qquad \mathcal{T}_{ref} = \left\{ \hat{\boldsymbol{T}}_{\text{ref}}(k|k), \ldots, \hat{\boldsymbol{T}}_{\text{ref}}(k+\mathcal{T}-1|k) \right\}$$

$$\mathcal{M} = \left\{ \hat{\boldsymbol{\mu}}(k|k), \ldots, \hat{\boldsymbol{\mu}}(k+\mathcal{T}-1|k) \right\} \qquad \mathcal{S} = \left\{ \hat{\boldsymbol{s}}(k|k), \ldots, \hat{\boldsymbol{s}}(k+\mathcal{T}-1|k) \right\}$$

Coordinated approach

- **Considers the computational and the thermal dynamics in the same optimization problem**

$$\min_{\mathcal{M},\mathcal{S},\mathcal{D},\mathcal{T}_{ref}} \left(\sum_{h=k}^{k+\mathcal{T}-1} \mathbf{1}^T \hat{\mathbf{p}}_{\mathcal{N}}(h|k) + \mathbf{1}^T \hat{\mathbf{p}}_{\mathcal{C}}(h|k) \right)$$

Minimize expected zone and CRAC power consumption

$$\begin{aligned}
\textbf{s.t.}\quad & \hat{\boldsymbol{l}}(k|k) = \boldsymbol{l}(k), \quad \hat{\boldsymbol{T}}_{\text{out}}(k|k) = \boldsymbol{T}_{\text{out}}(k) \\
& \text{for all } h = k, \ldots, k + \mathcal{T} - 1 \\
& \text{computational dynamics,} \\
& \text{thermal dynamics,} \\
& \text{QoS constraints,} \\
& \mathbf{0} \le \hat{\boldsymbol{\mu}}(h|k) \le \overline{\boldsymbol{\mu}}, \quad \hat{\boldsymbol{\delta}}(h|k) = \mathbf{0} \\
& \mathbf{0} \le \hat{\boldsymbol{s}}(h|k) \le \mathbf{1}, \mathbf{1}^T \hat{\boldsymbol{s}}(h|k) \le 1, \\
& \underline{\boldsymbol{T}_{\text{ref}}} \le \hat{\boldsymbol{T}}_{\text{ref}}(h|k) \le \overline{\boldsymbol{T}_{\text{ref}}}, \quad \hat{\boldsymbol{T}}_{\text{in}}(h+1|k) \le \overline{\boldsymbol{T}_{\text{in}}}, \\
& \hat{\mathbf{p}}(h|k) = \mathsf{B}_{\eta} \hat{\boldsymbol{\eta}}(h|k)
\end{aligned}$$

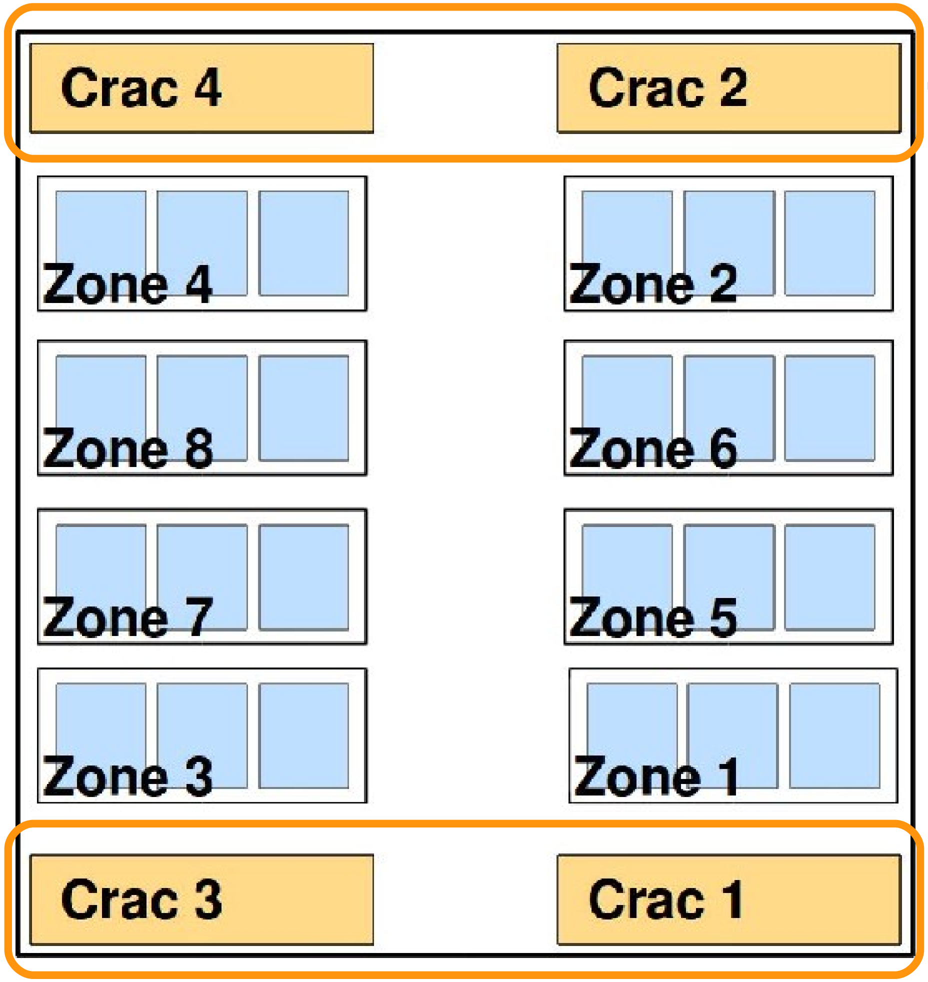
Considers both thermal and computational dynamics

Thermal-computational coupling

Outline

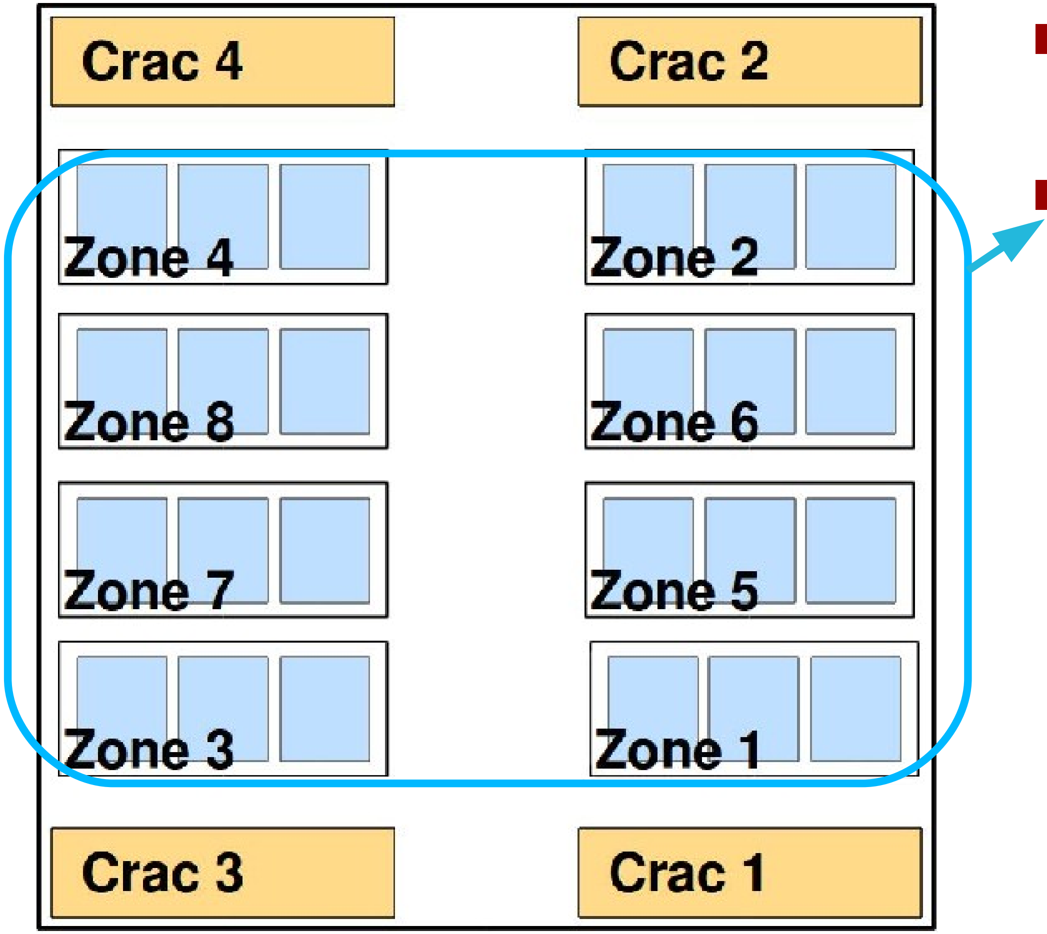
- Introduction
- Control-oriented model
- Control strategies
- **Simulation results**
- Conclusion and future work

Simulation parameters



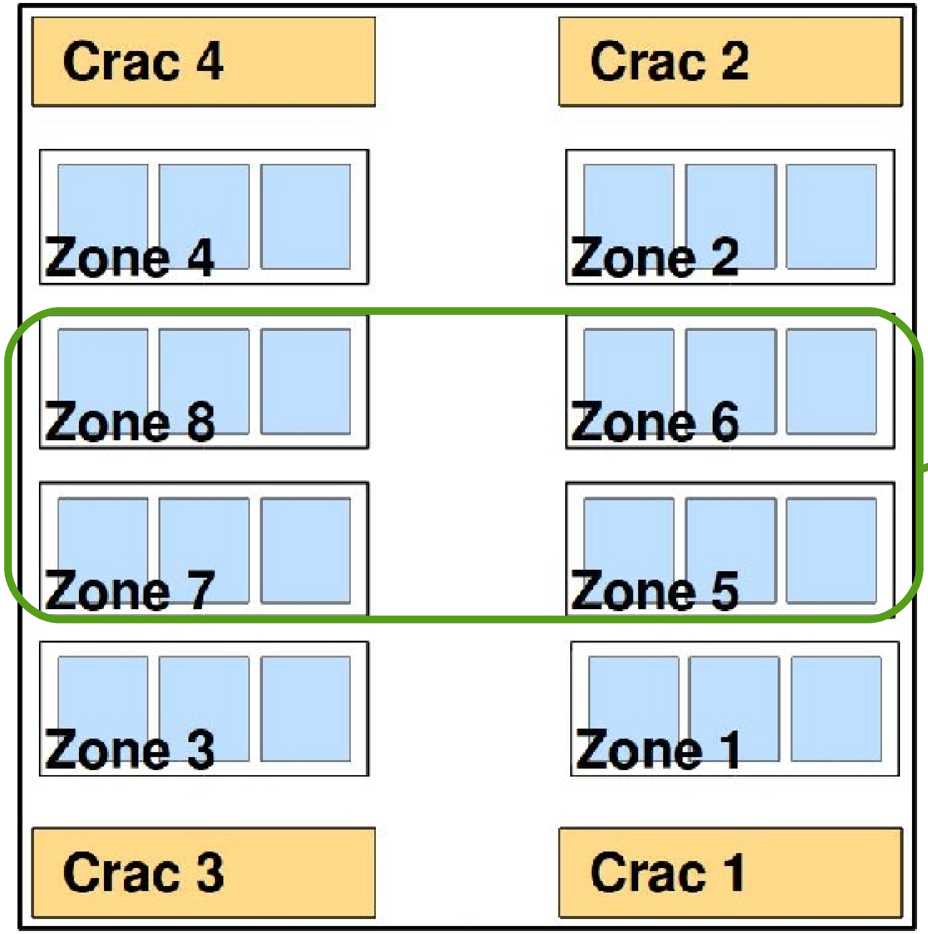
- **4 CRAC units**
  - Identical to each other

Simulation parameters



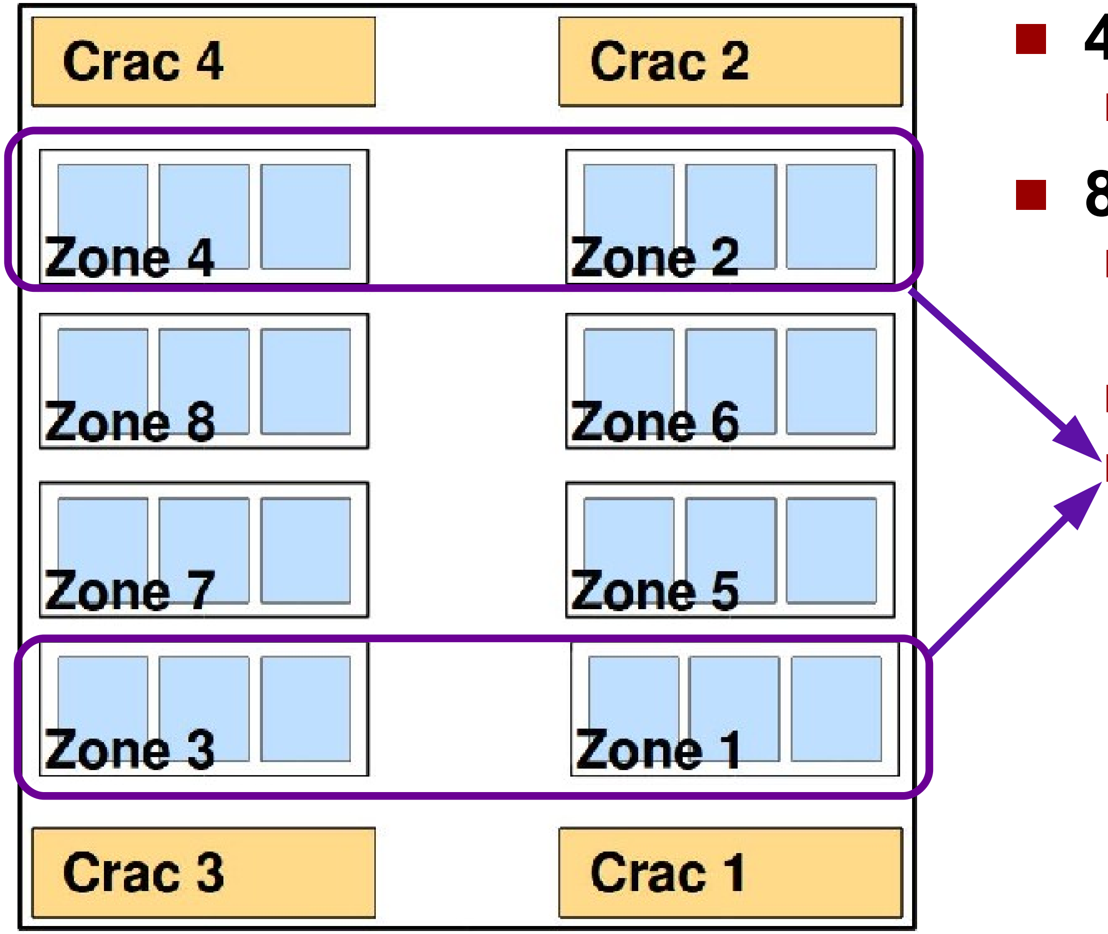
- **4 CRAC units**
  - Identical to each other
- **8 Zones**
  - 3 Racks each (126 servers per zone)

Simulation parameters



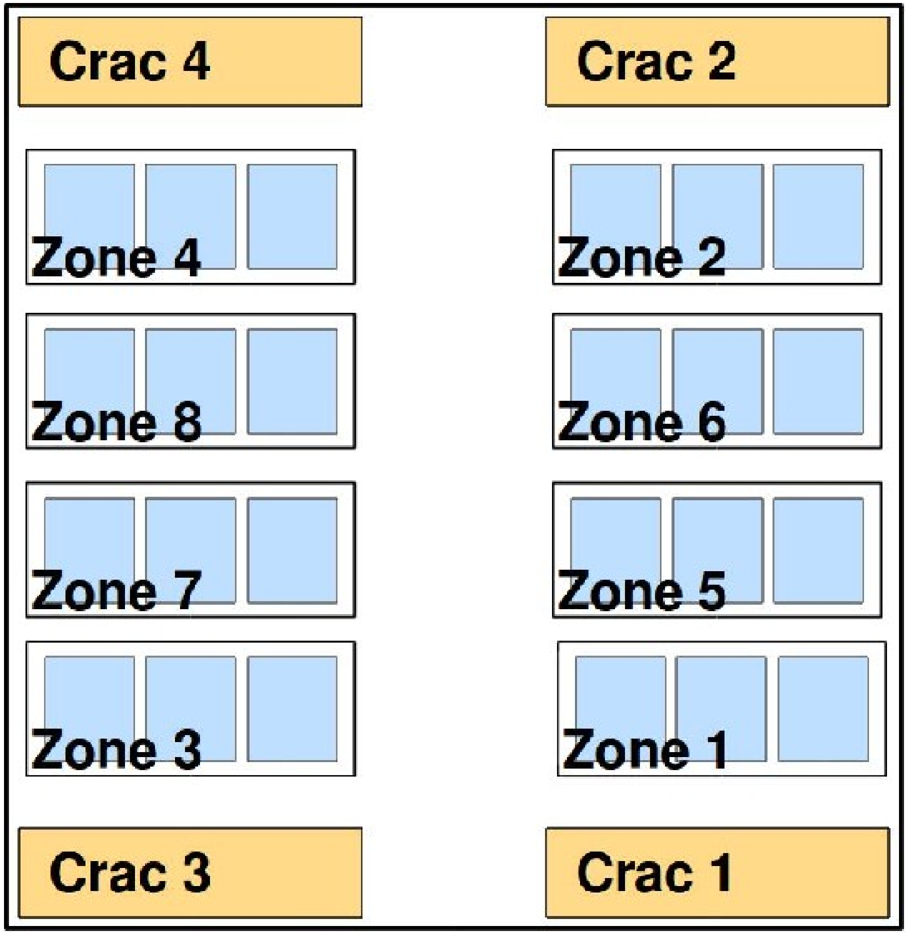
- **4 CRAC units**
  - Identical to each other
- **8 Zones**
  - 3 Racks each (126 servers per zone)
  - Energy efficient servers

Simulation parameters



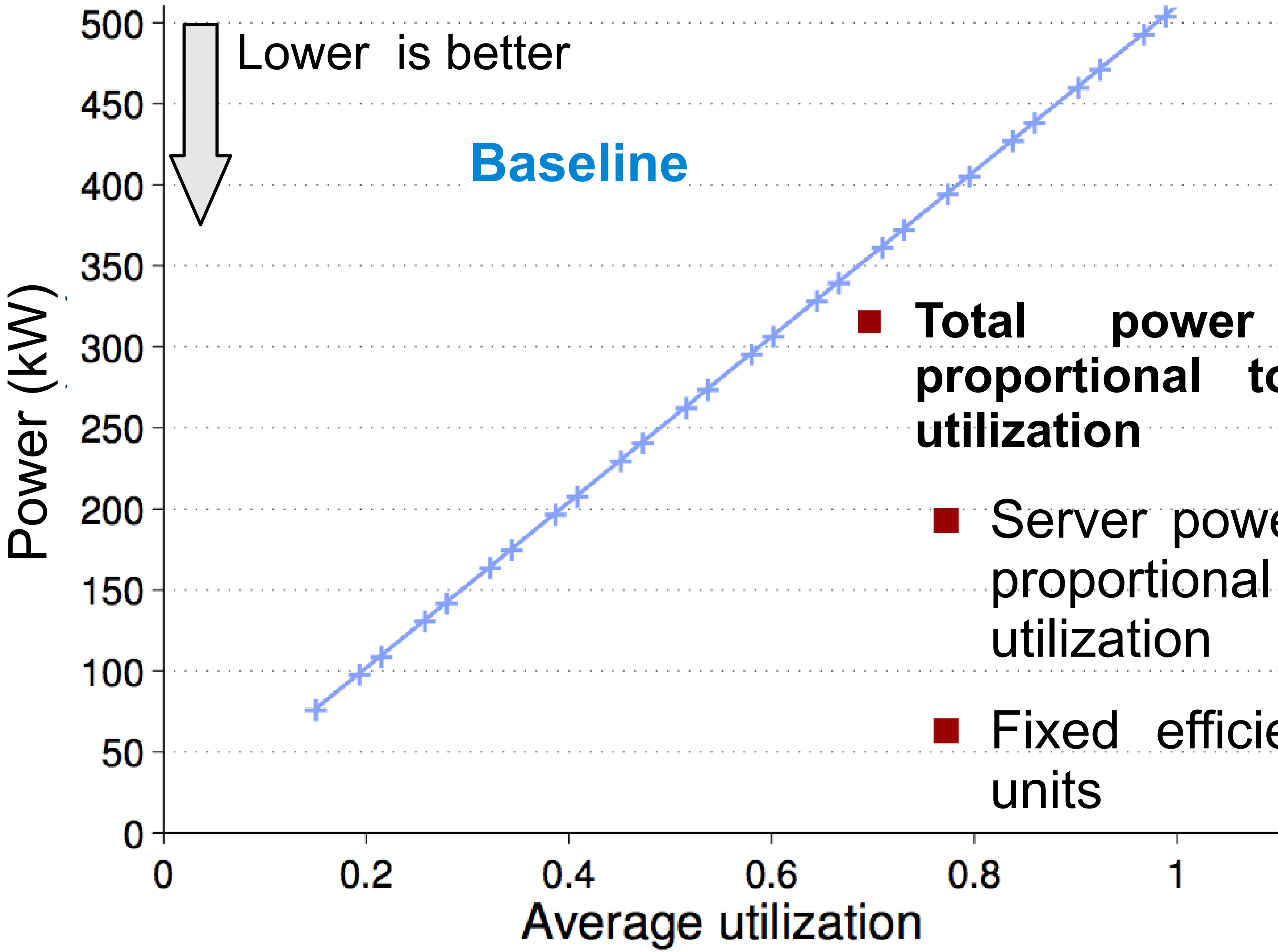
- **4 CRAC units**
  - Identical to each other
- **8 Zones**
  - 3 Racks each (126 servers per zone)
  - Energy efficient servers
  - Efficiently cooled servers

Simulation parameters



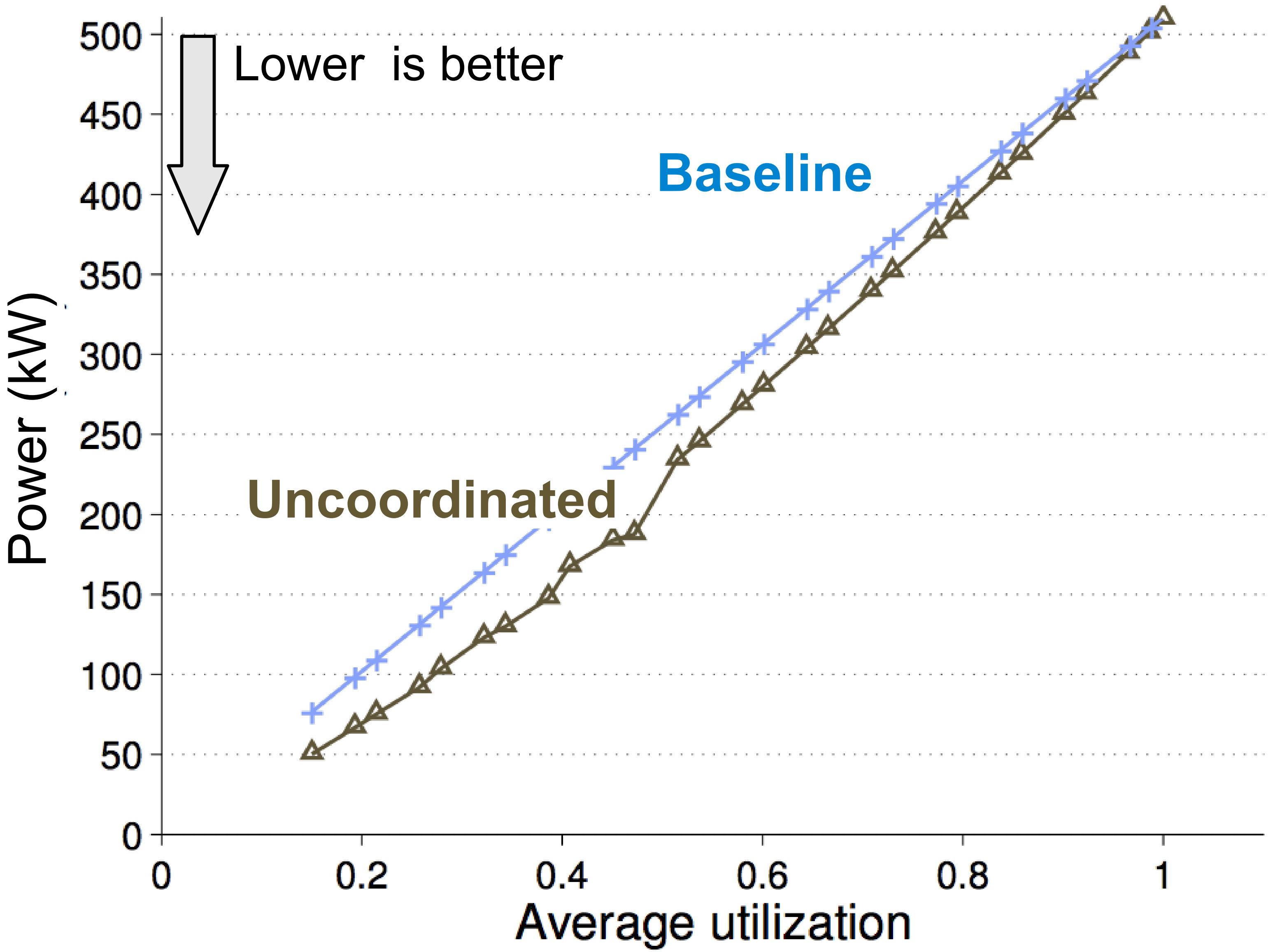
- **4 CRAC units**
  - Identical to each other
- **8 Zones**
  - 3 Racks each (126 servers per zone)
  - Energy efficient servers
  - Efficiently cooled servers
- **Analyze the average power consumption for different workload arrival rates**
- **Modeling language: TomSym**
- **Solver: KNITRO 7.0**

Total power consumption

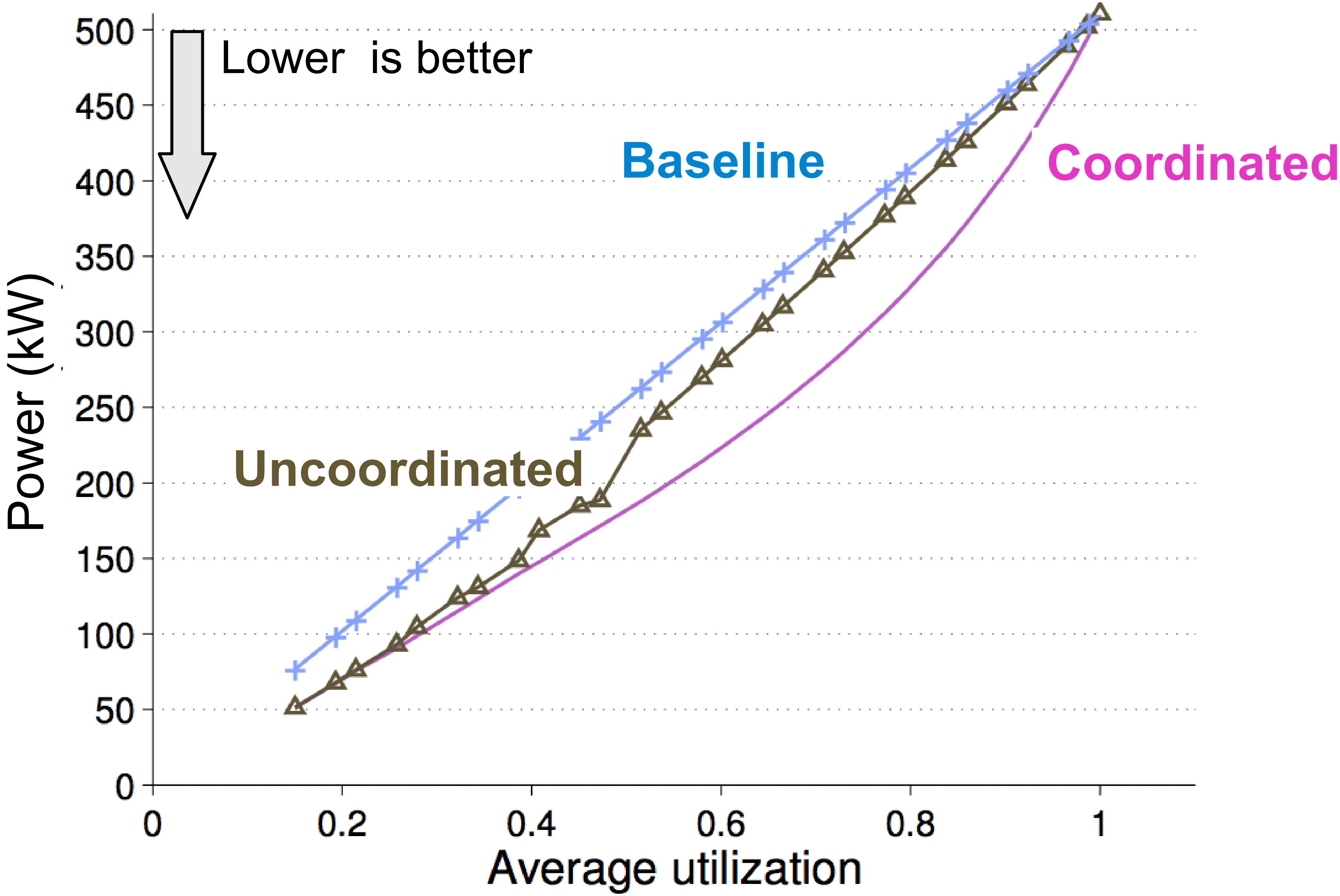


- **Total power consumption proportional to the average utilization**
  - Server power consumption proportional to the average utilization
  - Fixed efficiency of CRAC units

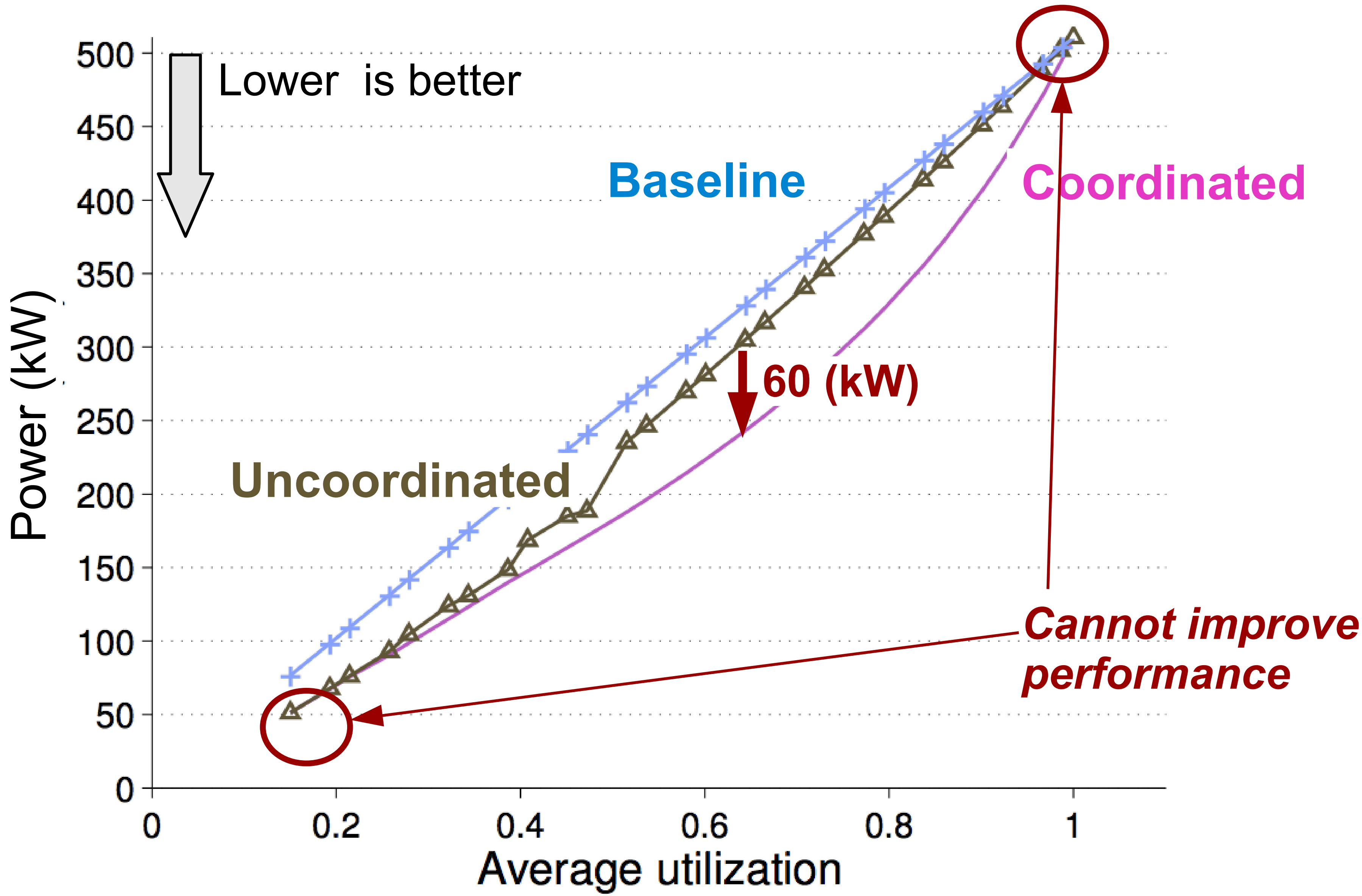
Total power consumption



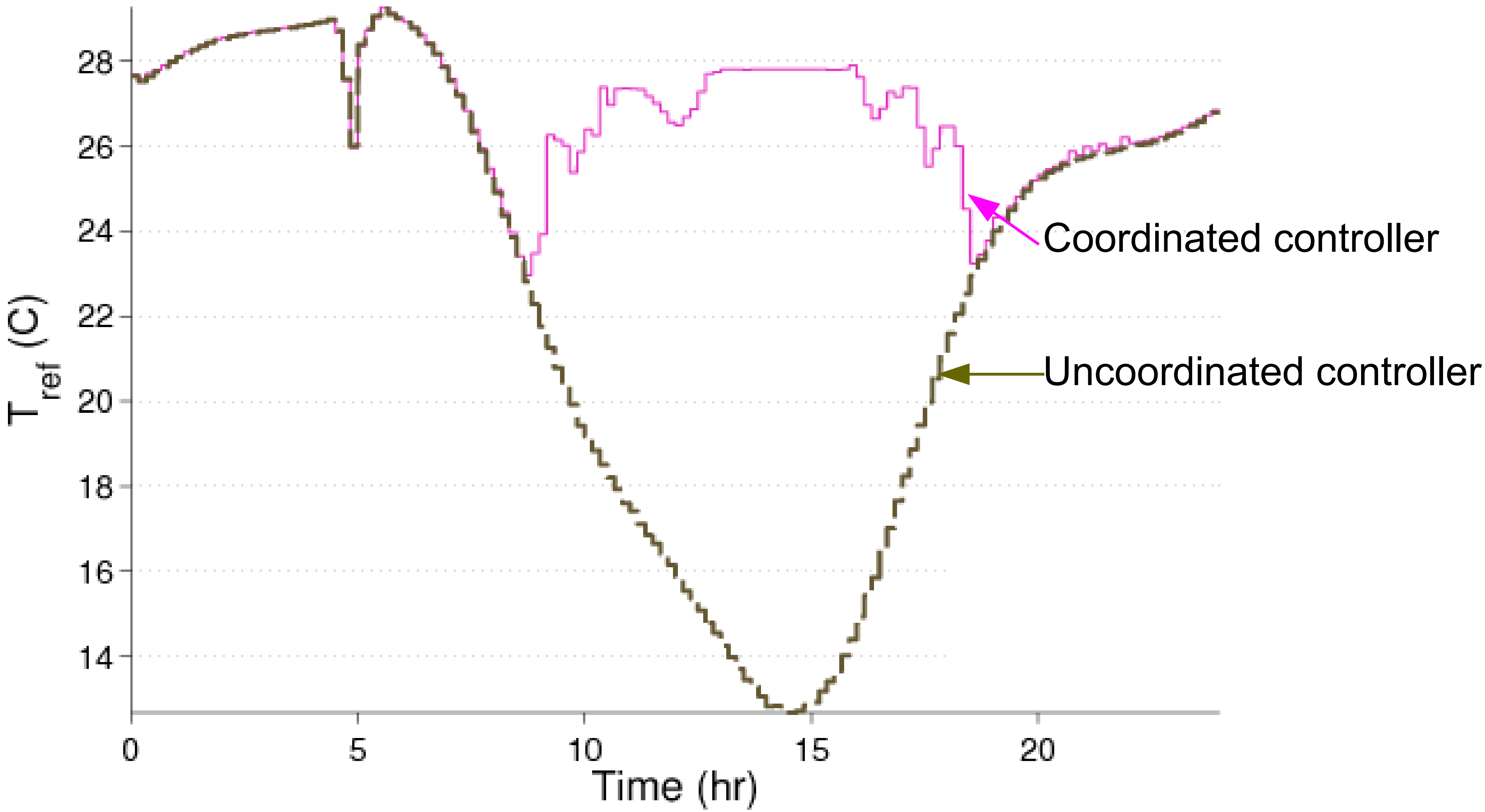
Total power consumption



Total power consumption

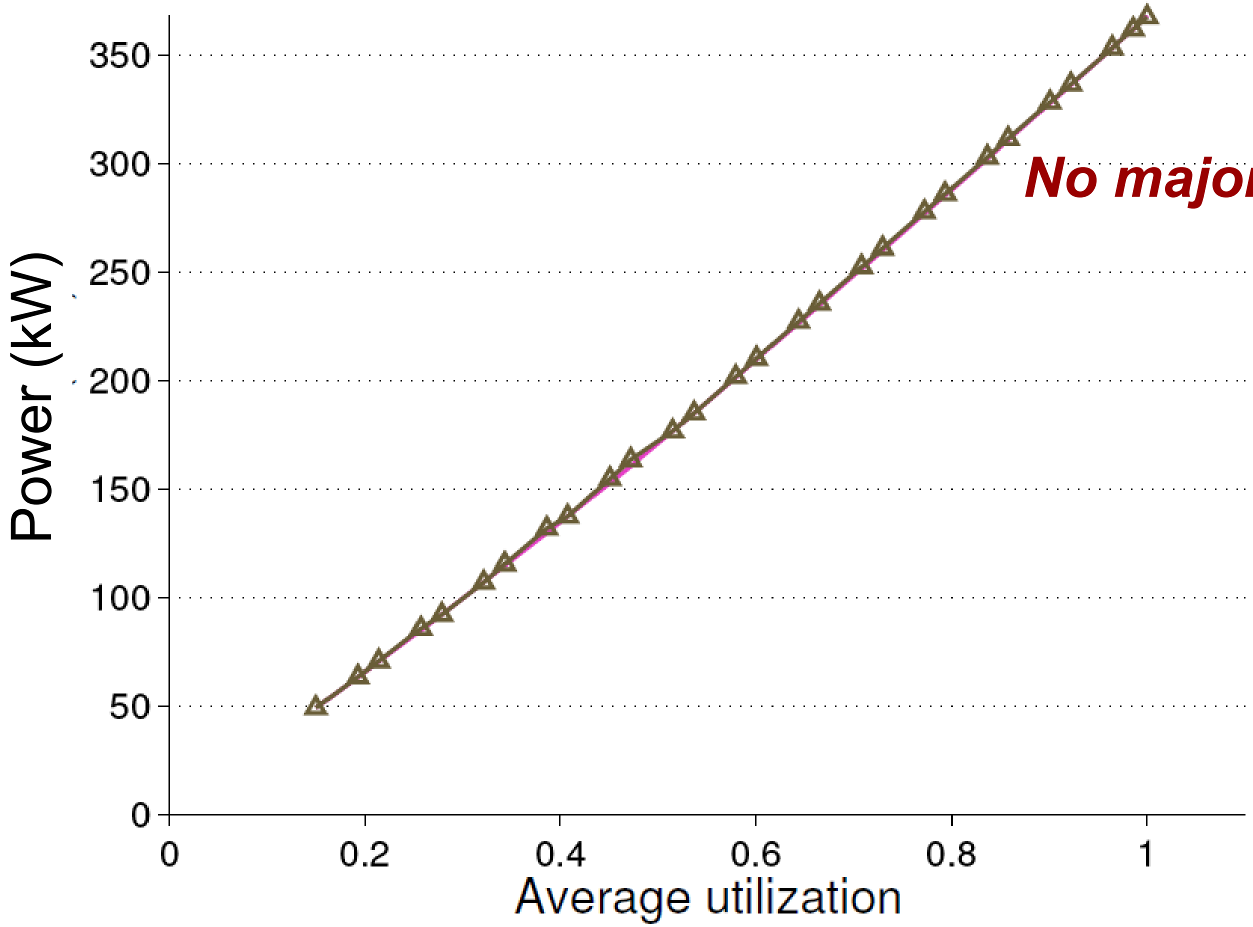


Average reference temperatures



Total power consumption

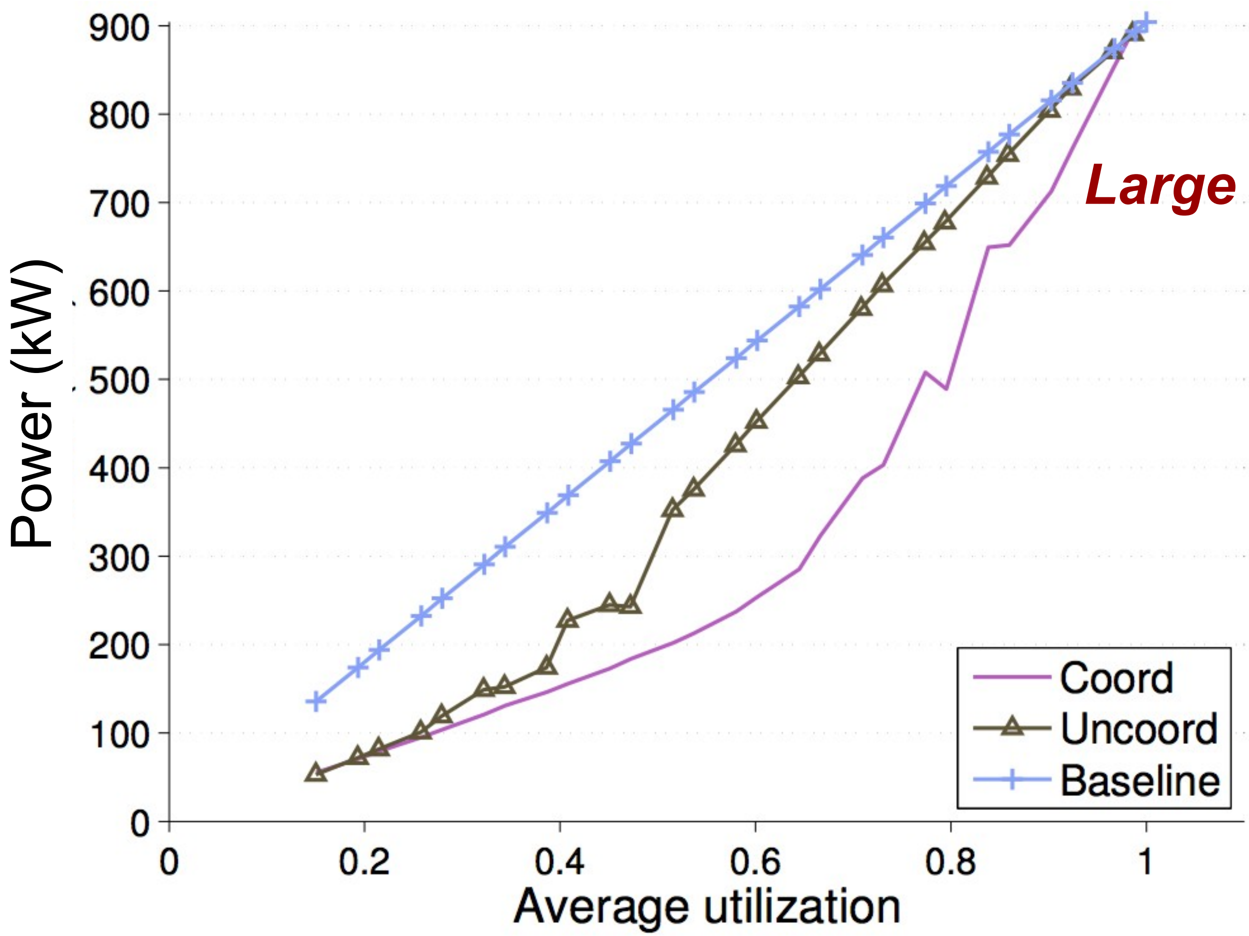
- **How do the controllers perform when all of the zones are efficiently cooled?**



No major improvements

Total power consumption

- **How do the controllers perform when large variability exists among the zone efficiency cooling?**



Large improvements

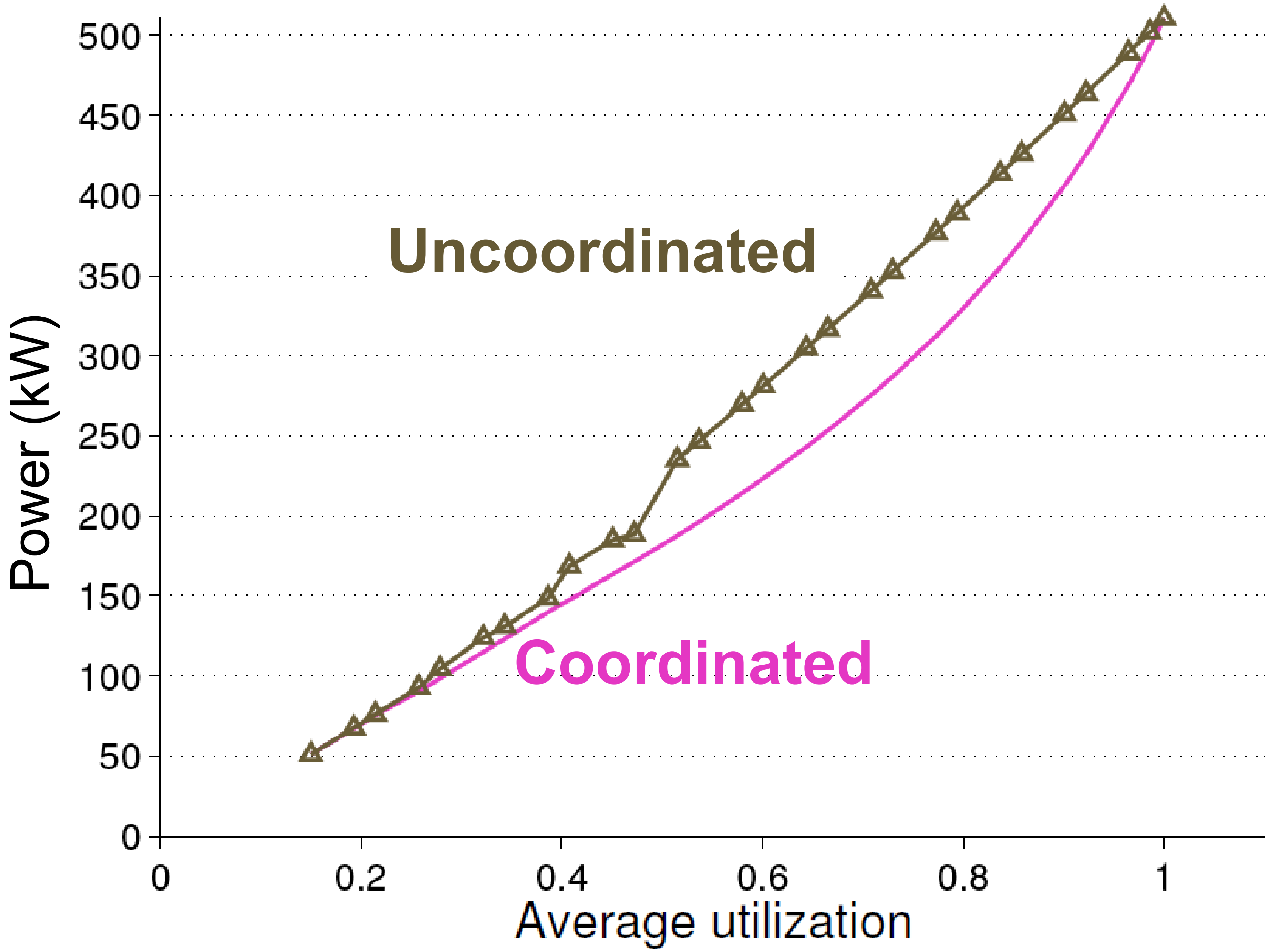
Cyber-Physical index

- **Given a data center**
  - How much energy can be saved by a coordinated controller, with respect to an uncoordinated controller?

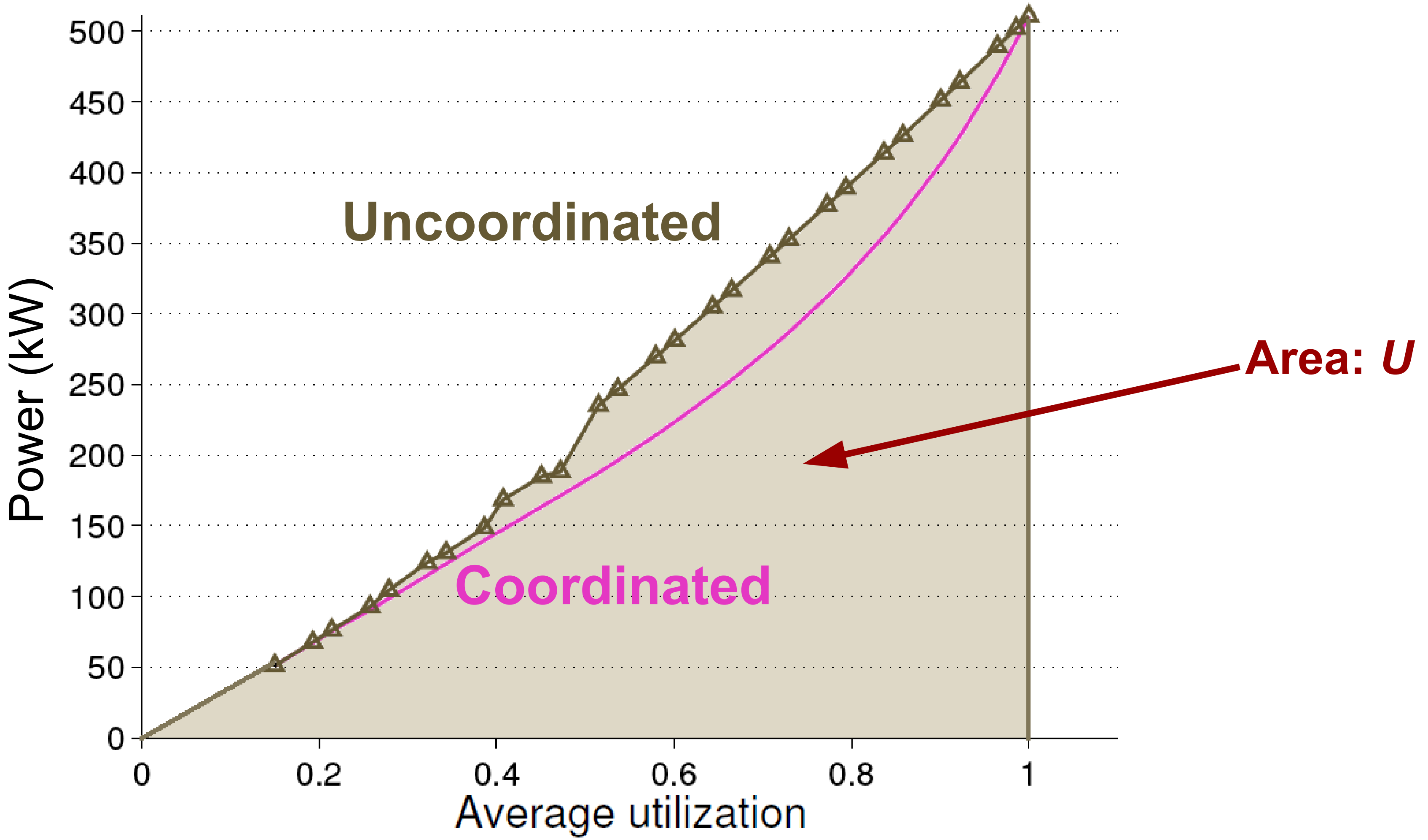
Cyber-Physical index

- **Given a data center**
  - How much energy can be saved by a coordinated controller, with respect to an uncoordinated controller?
- **Cyber-Physical index (CPI), values in [0,1]**
  - When CPI is close to 1, then a coordinated approach is advisable
  - When CPI is close to 0, then an uncoordinated approach tends to be as efficient as a coordinated approach
- **CPI is function of the sensitivity of the zones with respect to variations of the workload departure rate and of reference temperatures**

Relative efficiency

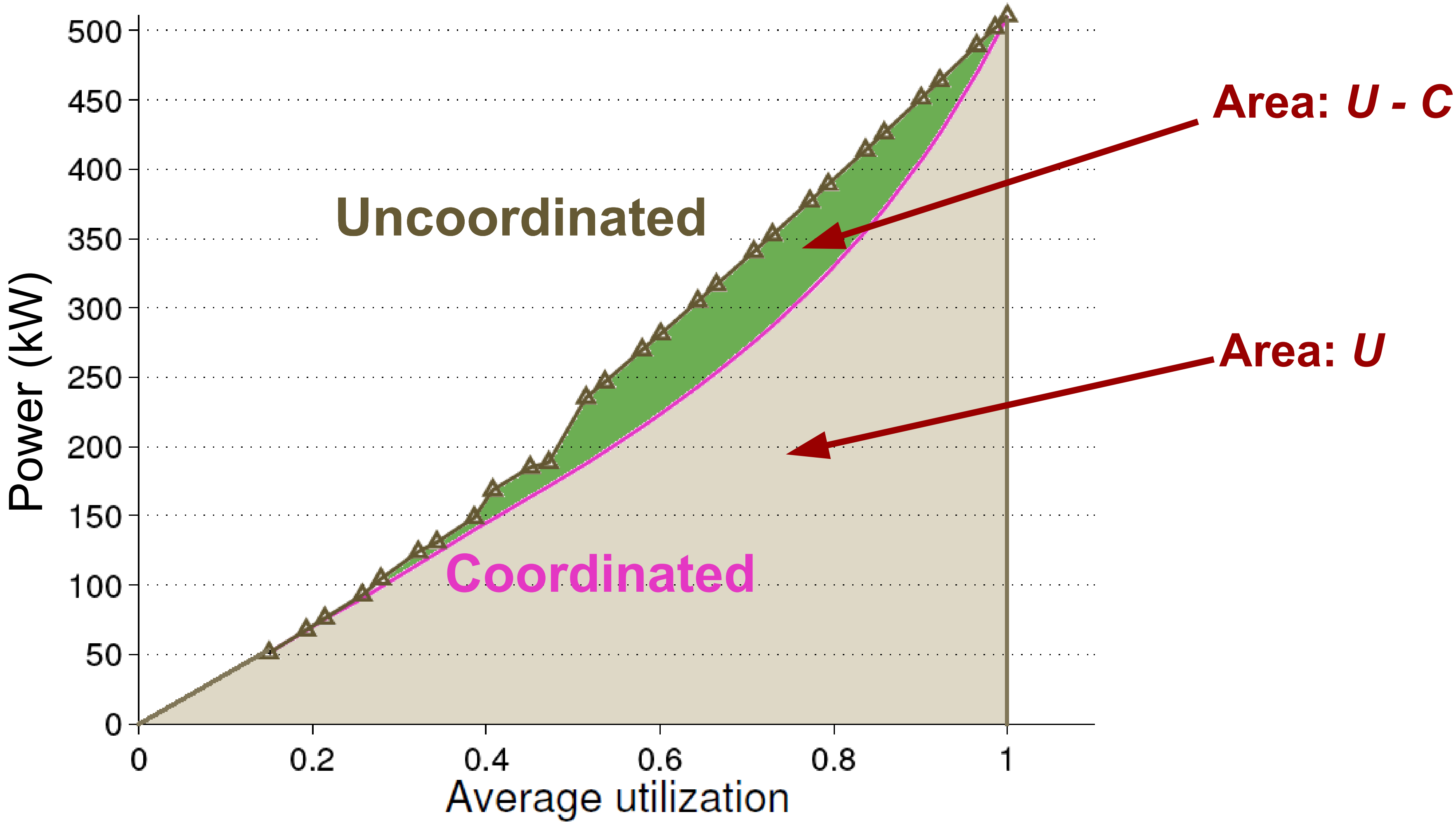


Relative efficiency



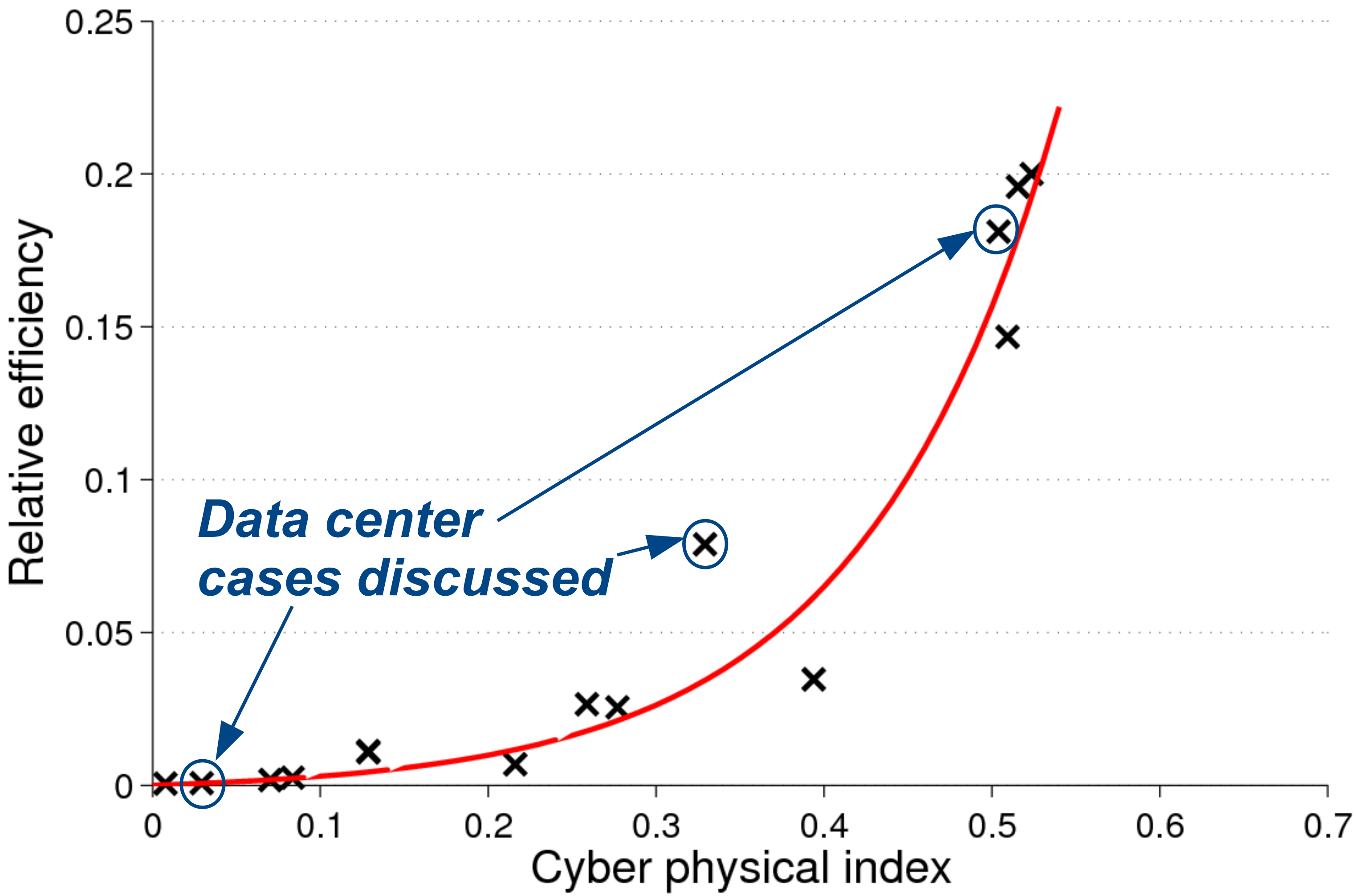
Relative efficiency

- **Relative efficiency =** $\dfrac{U - C}{U}$



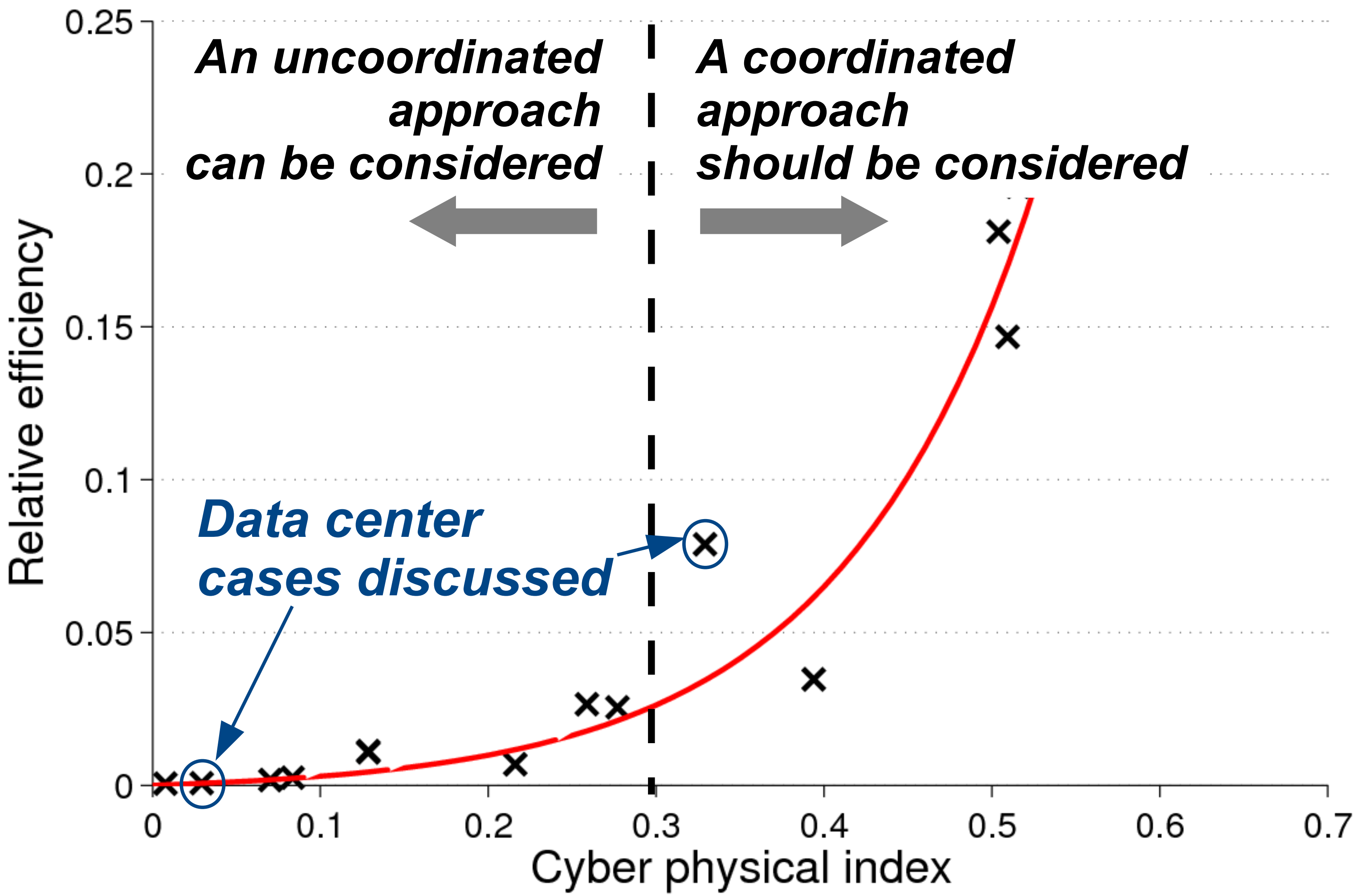
Relative savings

- Every point represents a different data center



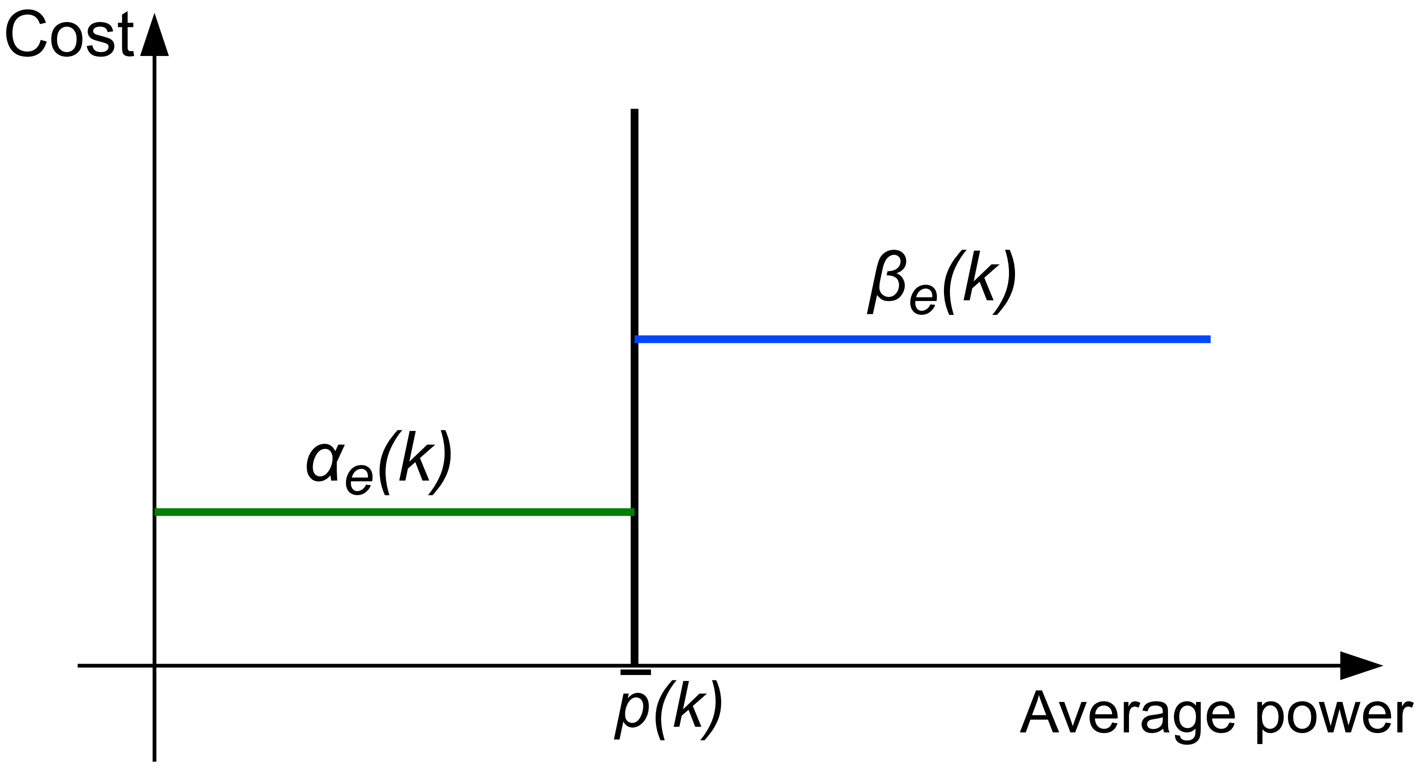
Relative savings

- Every point represents a different data center



Interaction with the smart grid

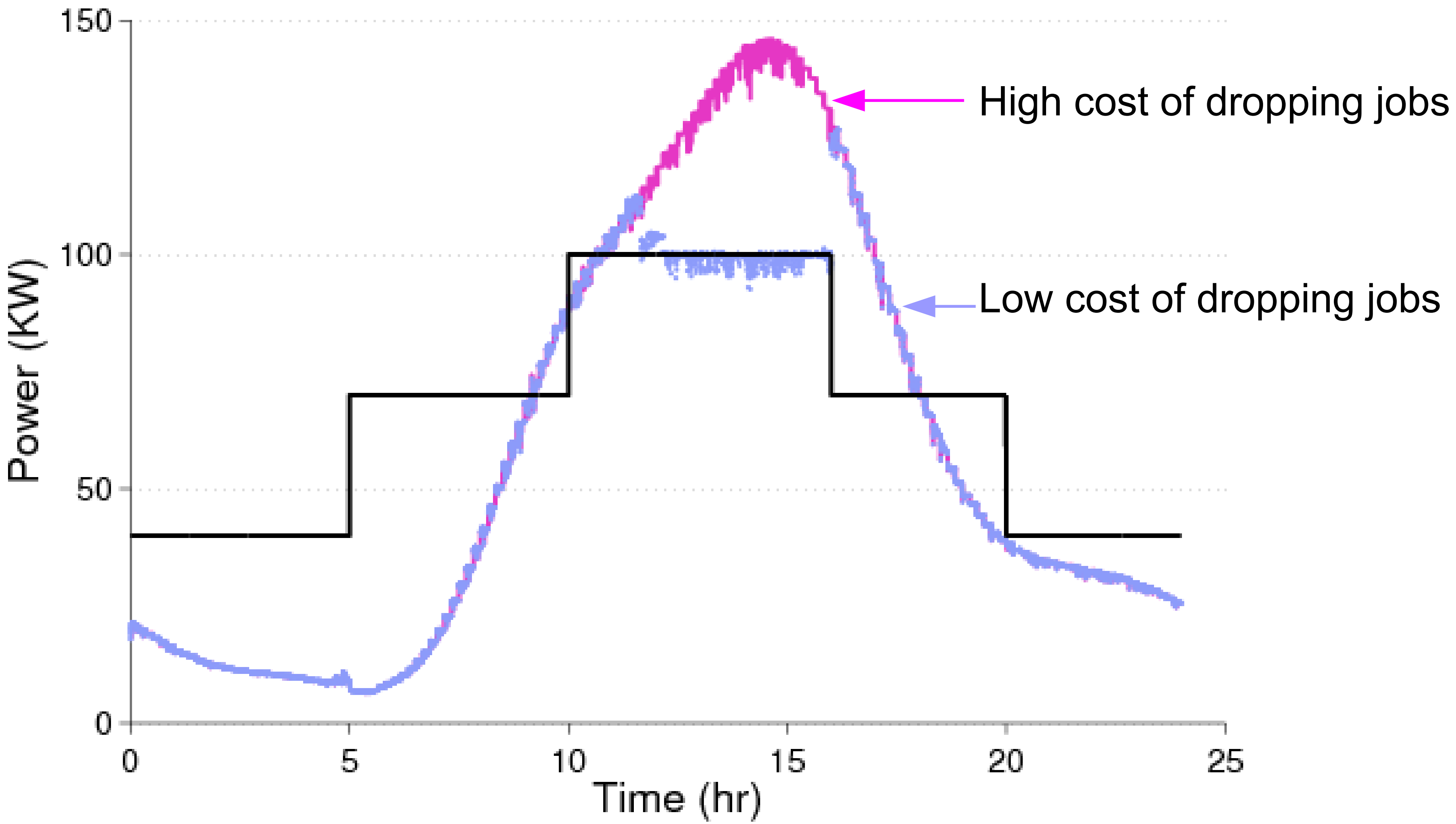
- **Time-varying electricity price**
  - Used by the smart-grid to cap the average power consumption of the data center



Run-time cost

- **Difference between income due to the workload processing and the cost of powering the data center**
- **Depends on two service level agreements (SLAs)**
  - SLA_U: sets the income based on the quality of service (QoS)
    - Approximated by the ratio between required and assigned hardware resources
  - SLA_G: sets the data center's powering cost
    - The energy cost is time-varying and power consumption dependent

Total data center power consumption



Conclusion

- **Discussed a control-oriented data center model**
  - Considers both the computational and the physical characteristics of a data center
  - The model can be extended to consider electricity cost (interaction with the smart-grid) and other data center equipment
- **Introduced two control strategies**
  - Representative of different approaches to data center control
- **Compared the performance of the control strategies under the same scenario**
- **Proposed a cyber-physical index**
  - First attempt to characterize the thermal and computational characteristics of the data center within a single index
- **Analyzed the impact of time-varying electricity prices**

Future work

- **Controllers can take advantage of service level agreements (SLAs) with both the users and the power grid**
  - Uncoordinated control approach can be as optimal as the coordinated approach
  - Depending on the SLA with the grid, the data center may induce large variations on the real-time electricity price
- **Given a data center, where should we locate its server so as to reduce its CPI?**
- **Feasibility and stability of coordinated and uncoordinated controller**
  - The coordinated controller is always feasible, but does not lead to a stable equilibrium point
  - How should the cost function be formulated so that the closed-loop system has a stable, economically optimal, equilibrium point?