

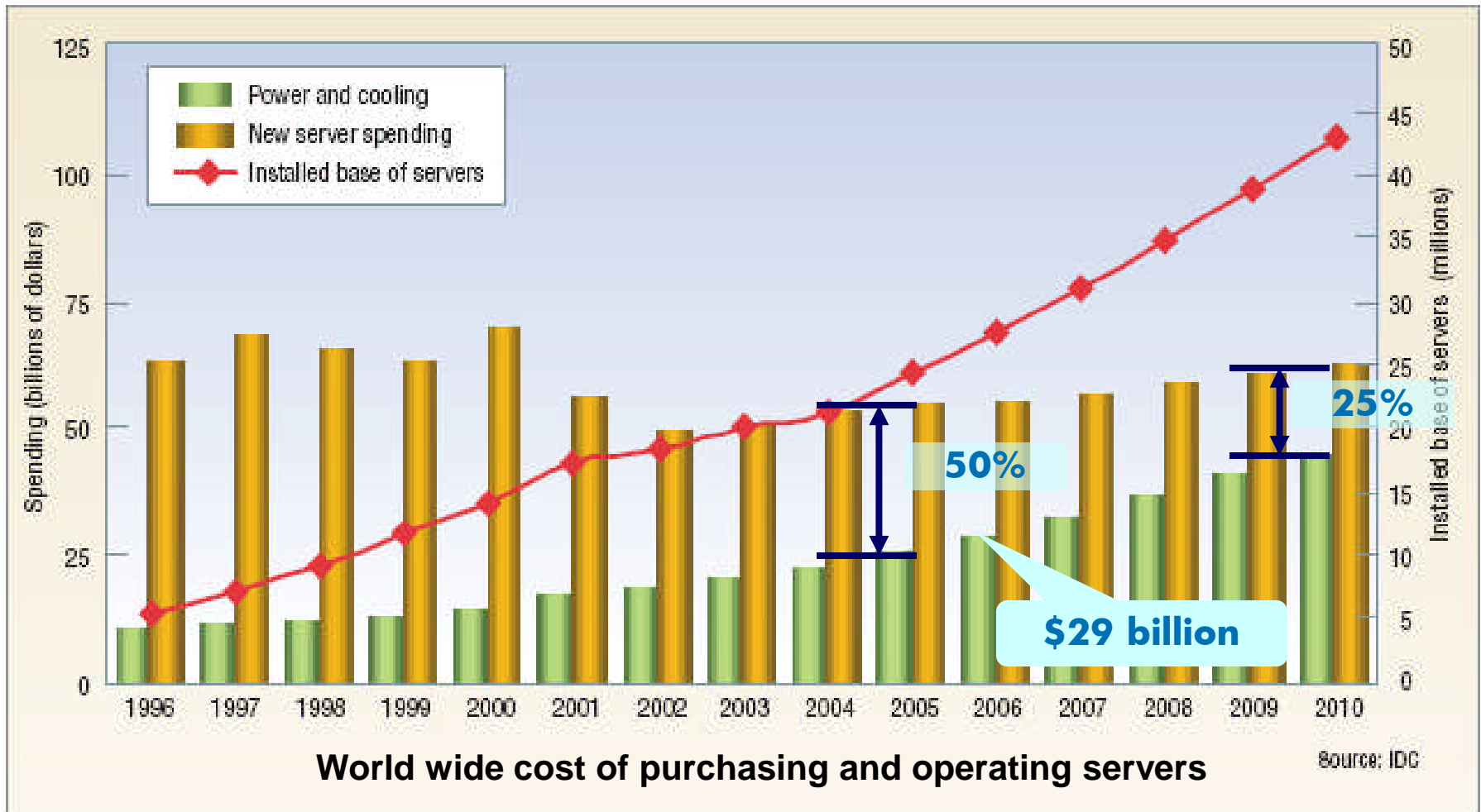
Green Clouds: Power Consumption as a First Order Criterion?

Vanish Talwar

HP Labs

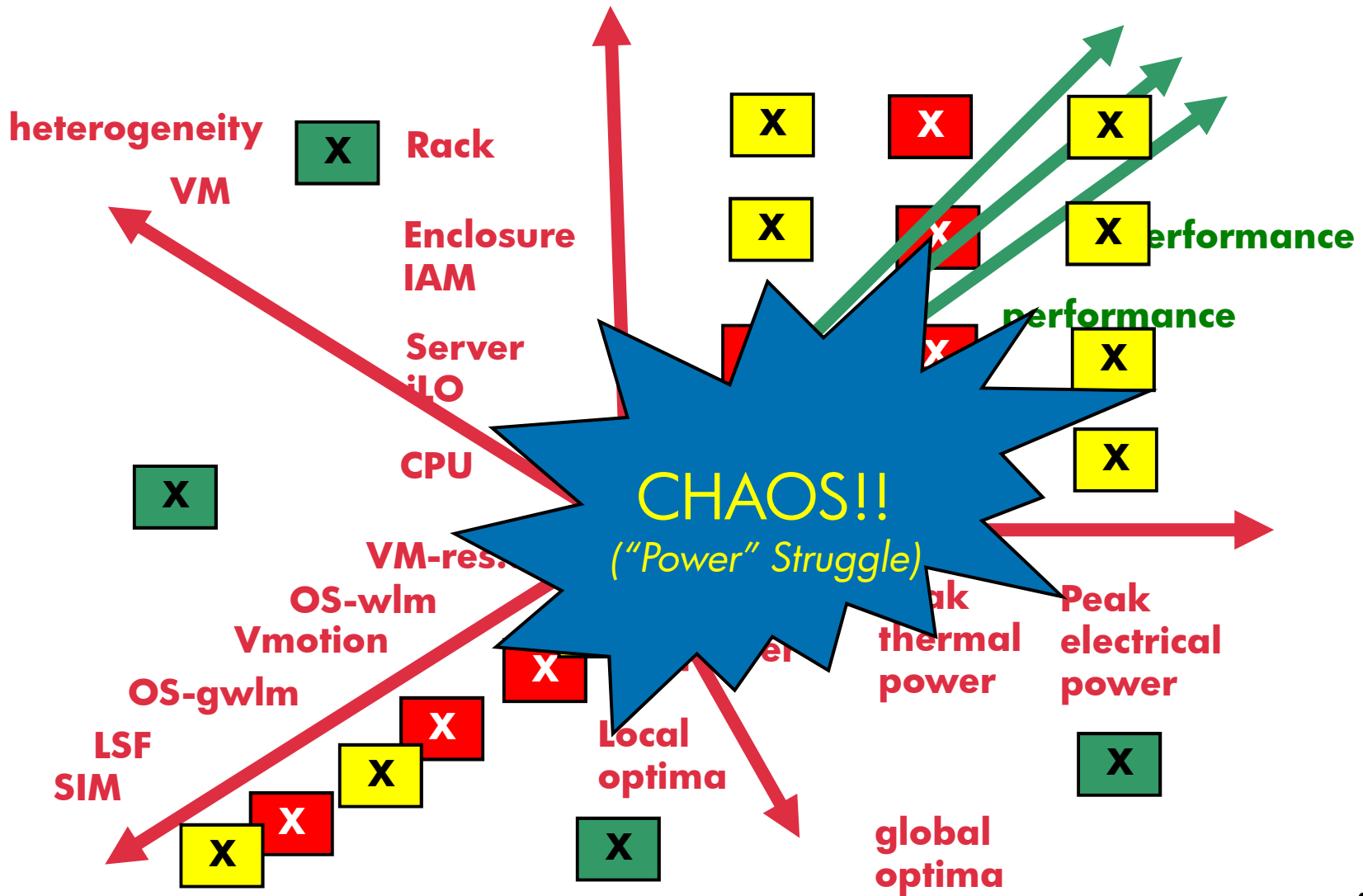


Motivation



environmental impact, heat & density, reliability

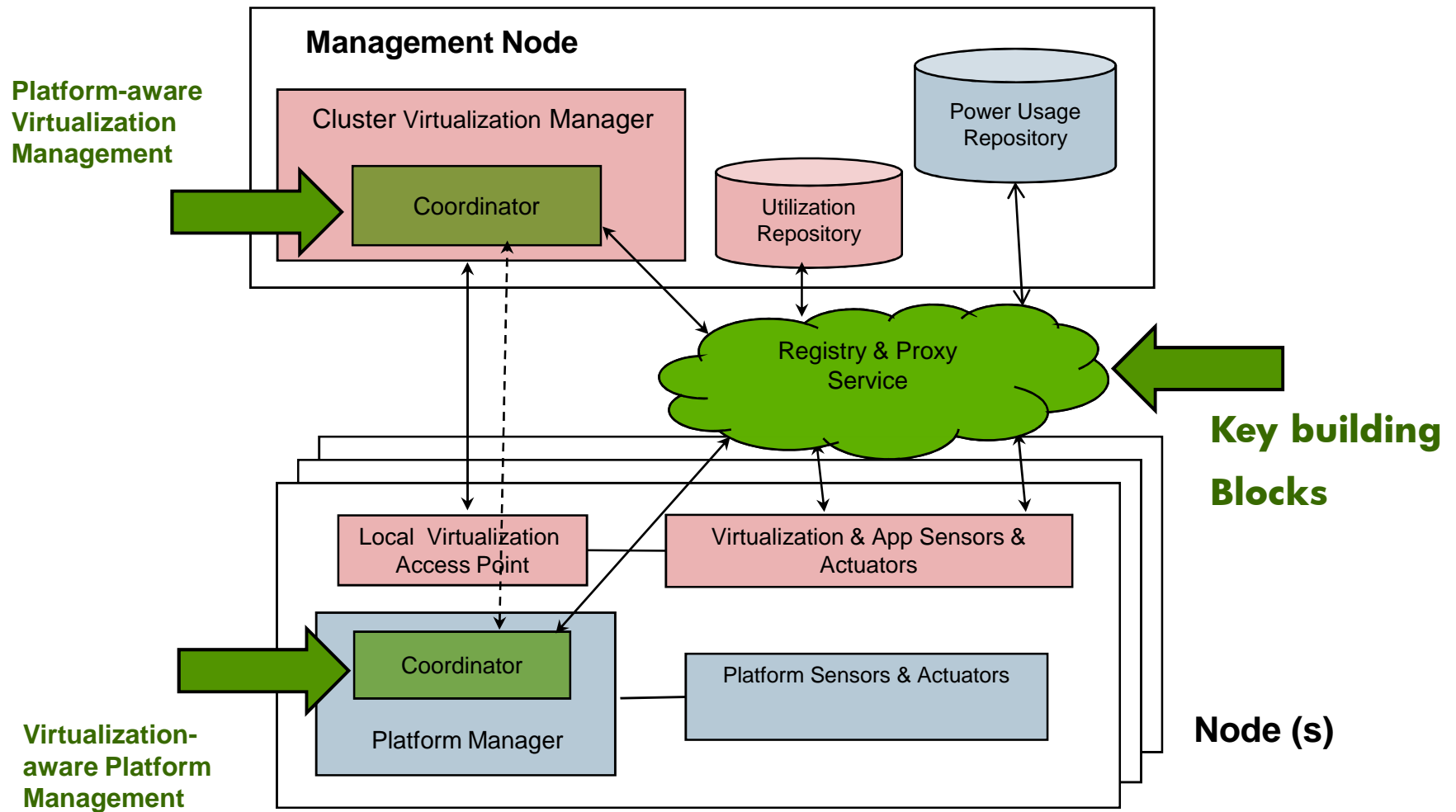
The Problem



Research Questions

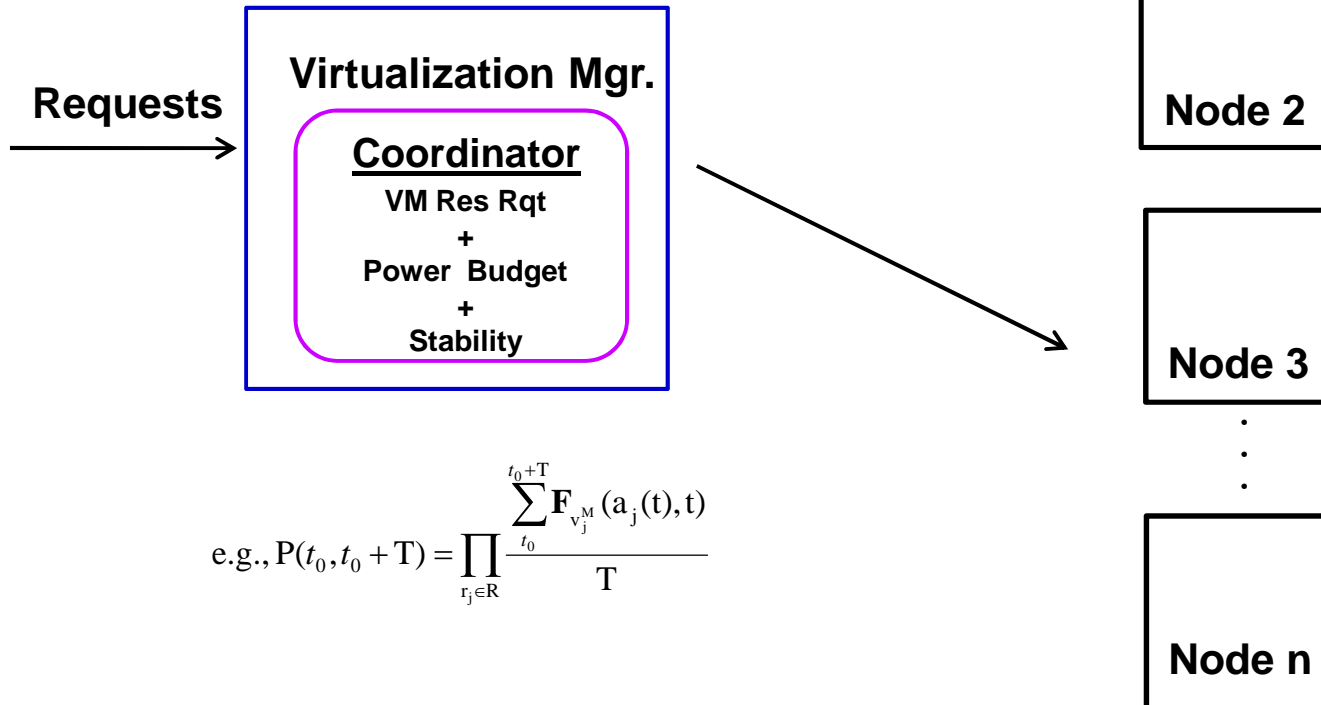
- Co-ordination Design
 - How to ensure correctness, stability, efficiency?
 - How to make local decisions with incomplete global info?
 - How to build in support for dynamism?
- Implications of Co-ordination
 - Can we simplify or consolidate controllers?
 - Do we revisit policies and mechanisms of the controllers?
 - How sensitive is the design to apps and systems considered?

Architecture Issues



Coordinated policies

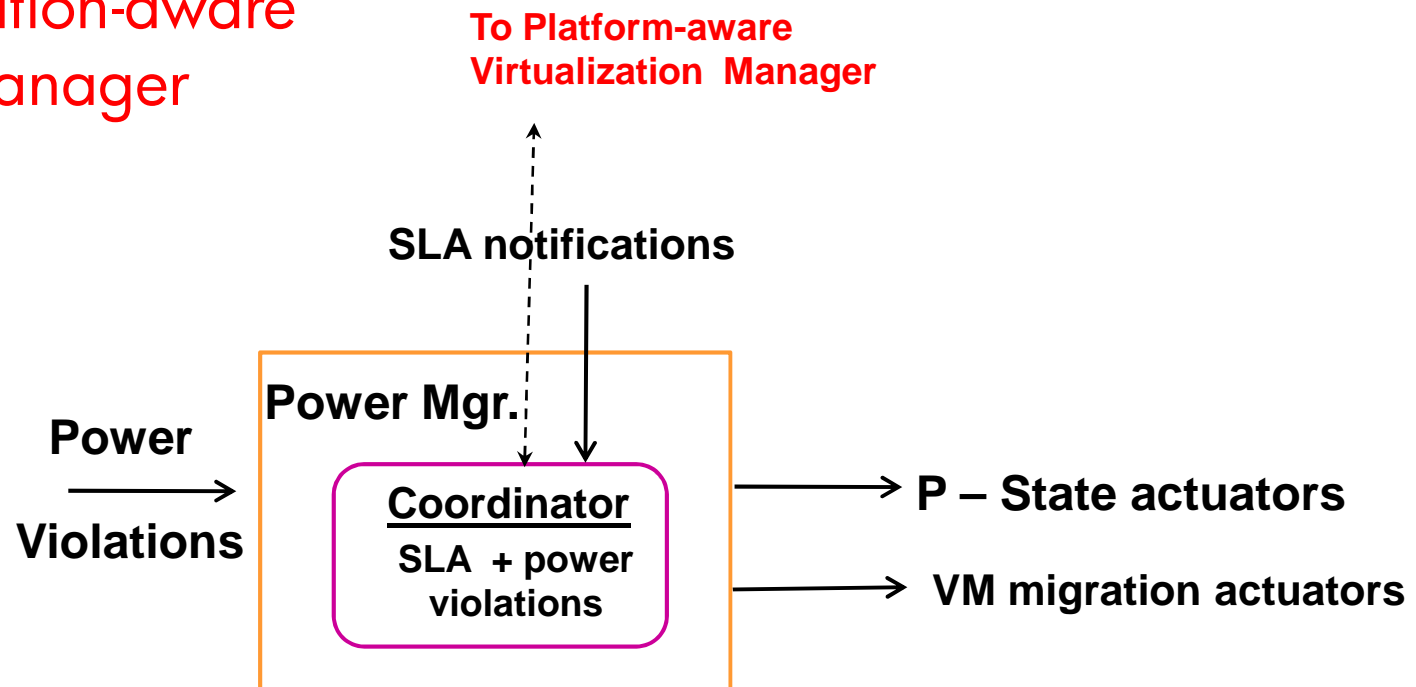
Platform-aware Virtualization Manager



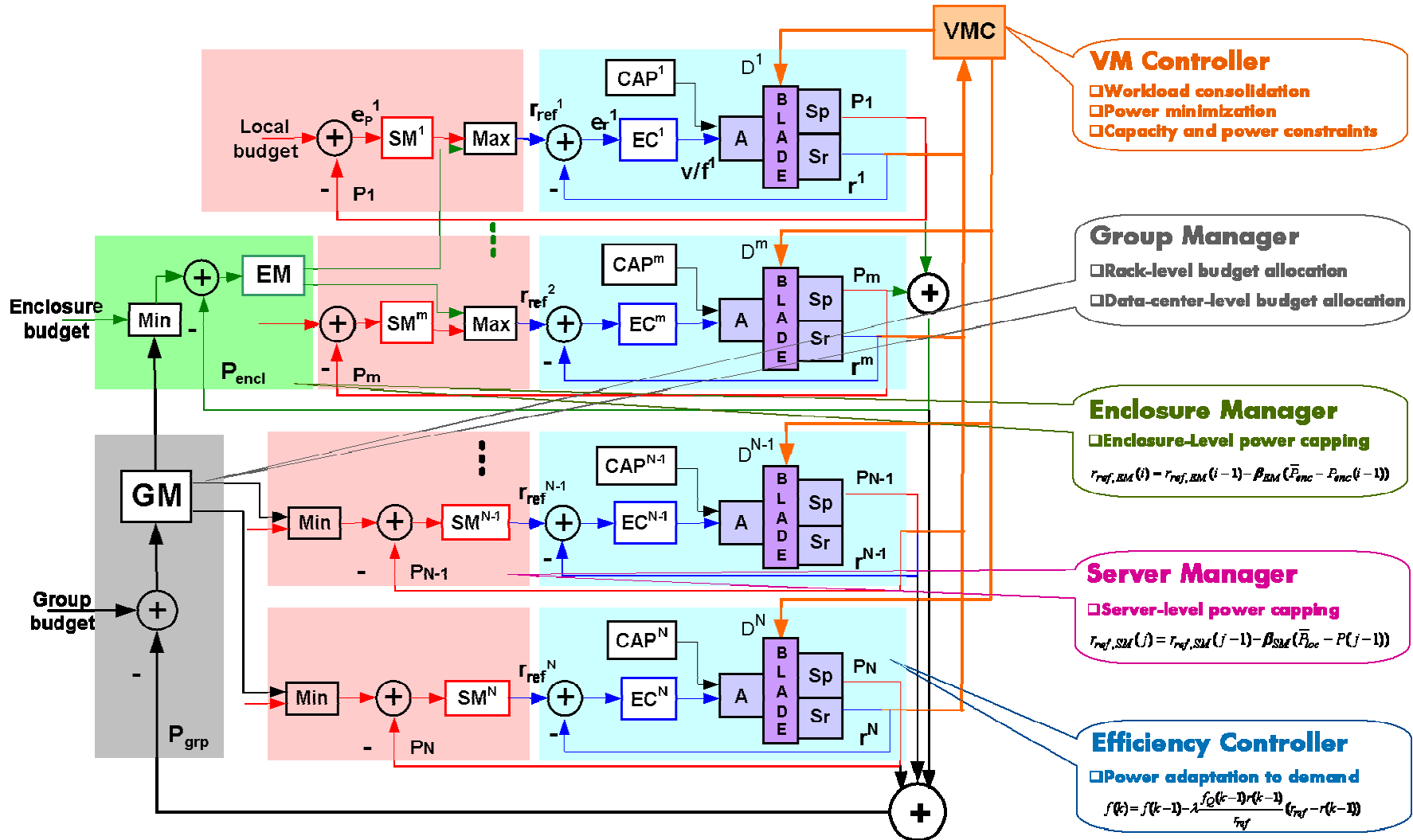
$$\text{e.g., } P(t_0, t_0 + T) = \prod_{r_j \in R} \frac{\sum_{t_0}^{t_0+T} \mathbf{F}_{v_j^M}(a_j(t), t)}{T}$$

Coordinated policies

Virtualization-aware Power Manager



Extending for Scale



VM Controller

- Workload consolidation
- Power minimization
- Capacity and power constraints

Group Manager

- Rack-level budget allocation
- Data-center-level budget allocation

Enclosure Manager

- Enclosure-Level power capping

$$r_{ref,EM}(i) = r_{ref,EM}(i-1) - \beta_{EM} (\bar{P}_{enc} - P_{enc}(i-1))$$

Server Manager

- Server-level power capping

$$r_{ref,SM}(j) = r_{ref,SM}(j-1) - \beta_{SM} (\bar{P}_{oc} - P(j-1))$$

Efficiency Controller

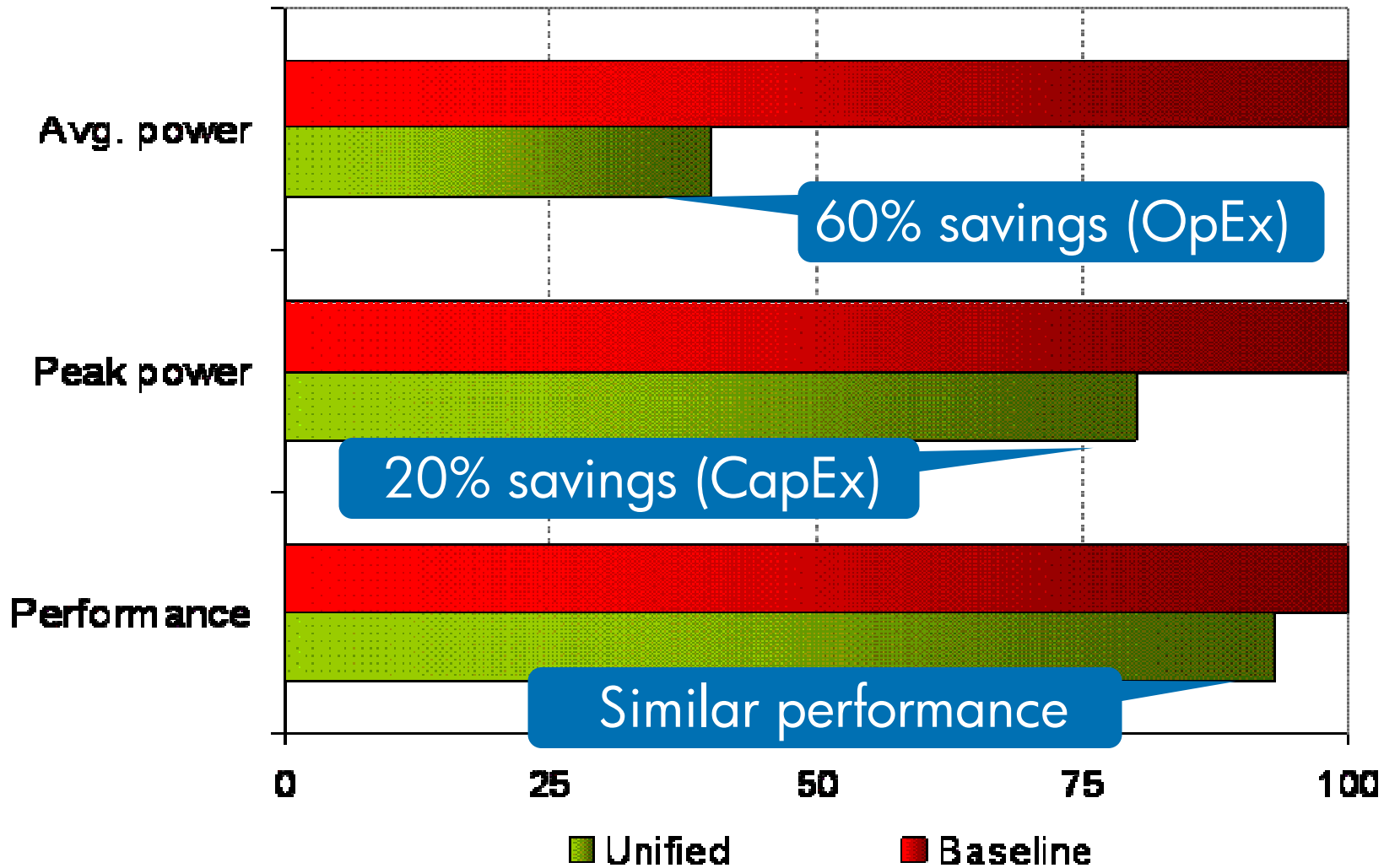
- Power adaptation to demand

$$f(k) = f(k-1) - \lambda \frac{f_0(k-1)r(k-1)}{r_{ref}} (r_{ref} - r(k-1))$$

Results (1)

- Improved power savings (10%)
- Significantly better QoS (71%)
- Better stability (54%)

Results (2)



Open Issues – Designing for Clouds

- **Break silos:** Multi-metric monitoring at large scale
 - platform (power, cooling), virtualization, application levels
 - light-weight anomaly detector on these multiple metrics
 - composability and dynamic configuration of monitoring functions
- **Address scale:** Federated policies
 - should be able to scale-up and scale-down
 - multi-level translation of objectives from data center to node
 - discovery and registration of distributed coordinators/controllers; degree of coordination/coupling