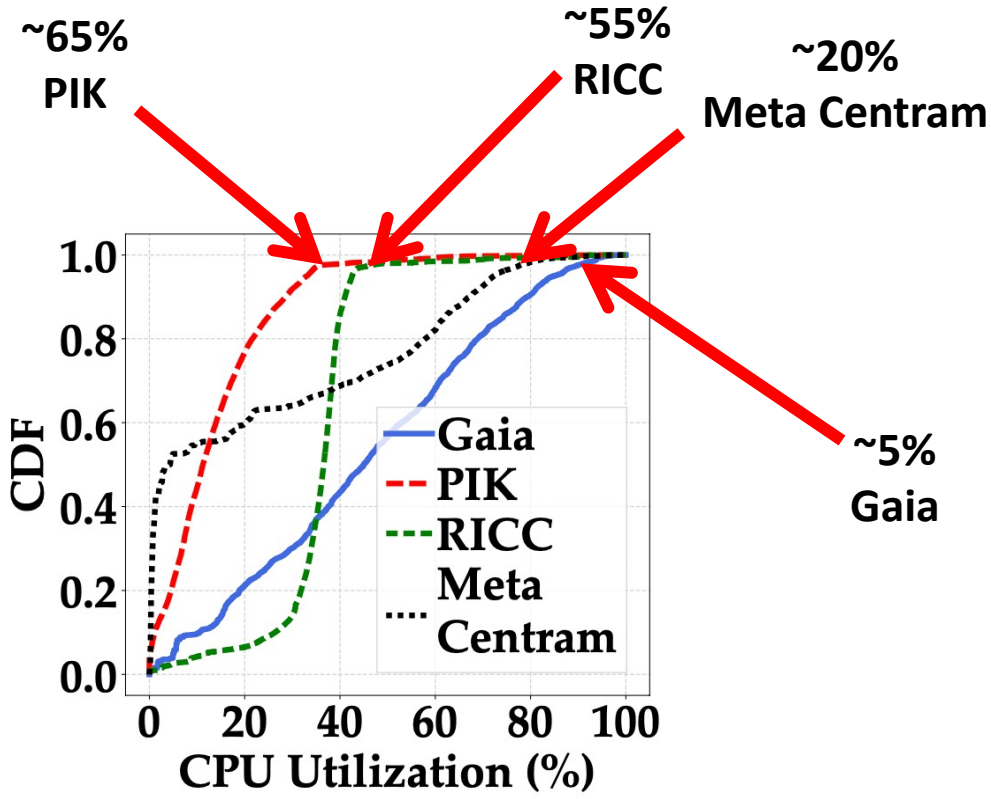
Market Mechanism-Based User-in-the-Loop Scalable Power Oversubscription for HPC Systems

Md Rajib Hossen[1], **Kishwar Ahmed[2]**, Mohammad A Islam[1]

[1]The University of Texas at Arlington, [2]The University of Toledo

Acknowledgement: This work is supported in parts by the NSF under grants EECS-2452357 and CNS-2300124

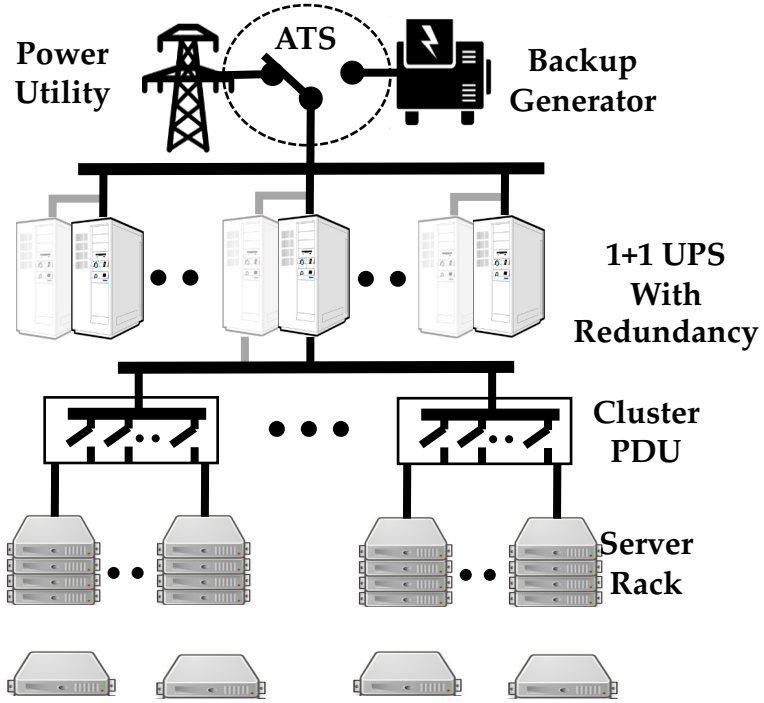
HPC System Utilization



HPC systems are highly underutilized

Power Oversubscription

Increase utilization via power oversubscription



Keep this infrastructure fixed

Add new server racks

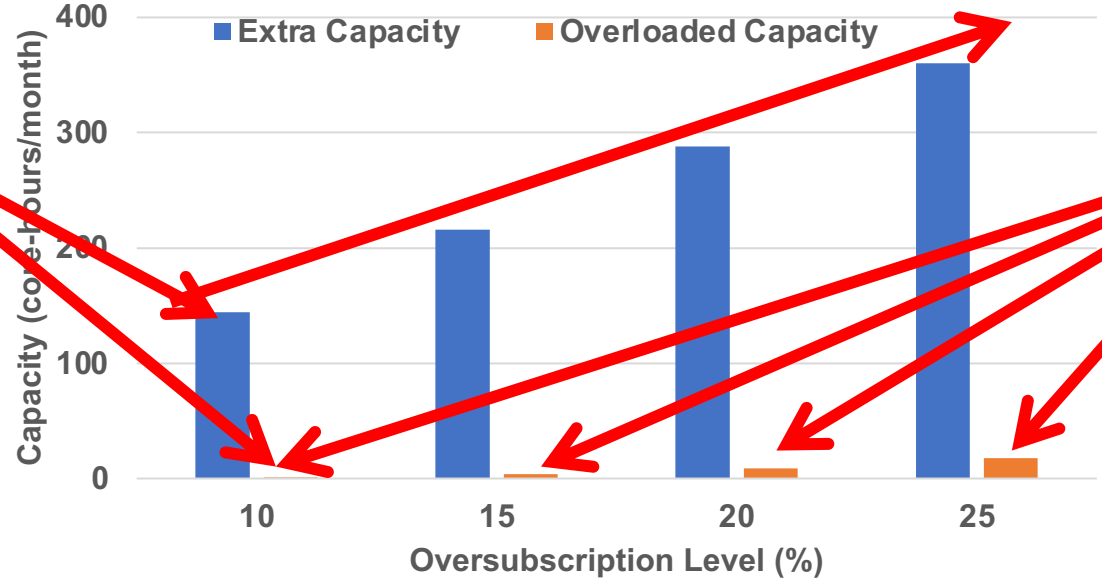
Additional server racks (without additional infrastructure investment)

Benefits of Power Oversubscription

144K added

vs

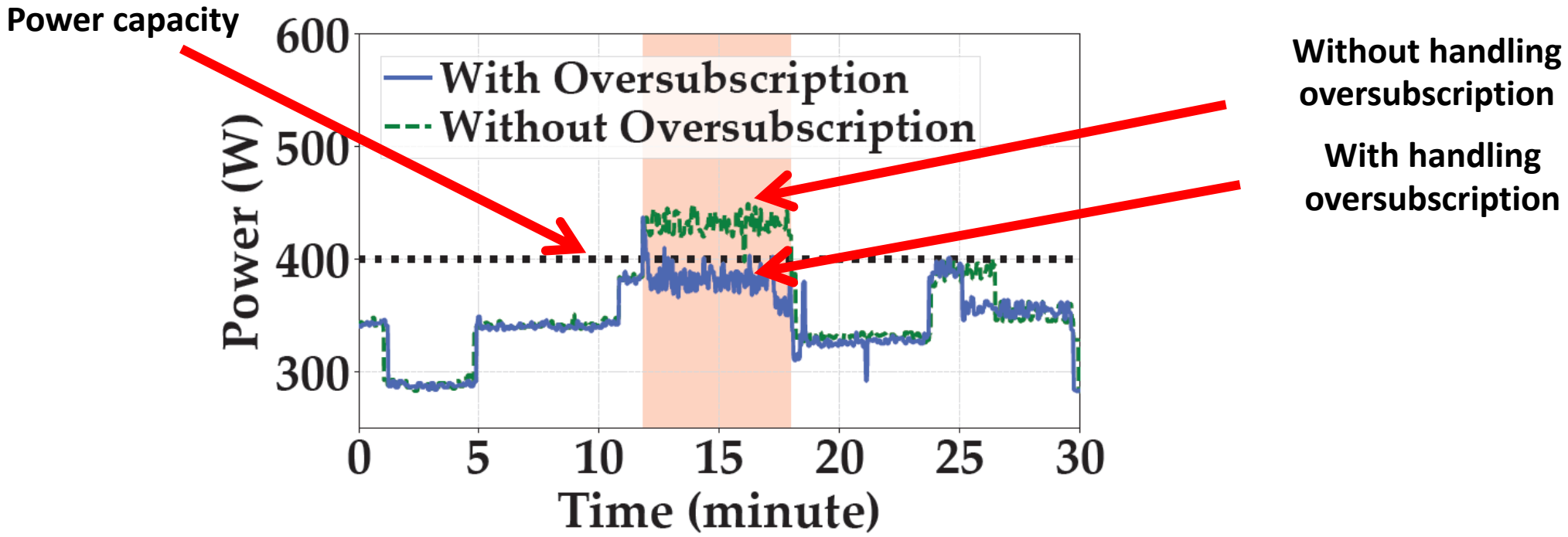
1.25K cut



Cut back to avoid overload

We can add far more core-hours than we have to cut back!

Challenges of Oversubscription



We propose a user-in-the-loop reactive approach to manage oversubscribed HPC

Handling Power Overloads in HPC

Performance degradation

OPT: $\underset{\delta_m}{\text{minimize}} \sum_{m=1}^{M(t)} \mathcal{L}_m(\delta_m)$

Target power reduction

subject to $\sum_{m=1}^{M(t)} \mathcal{P}(\delta_m) \geq P(t) - C$

Challenging for HPC manager to determine the performance impact

Supply Function Bidding

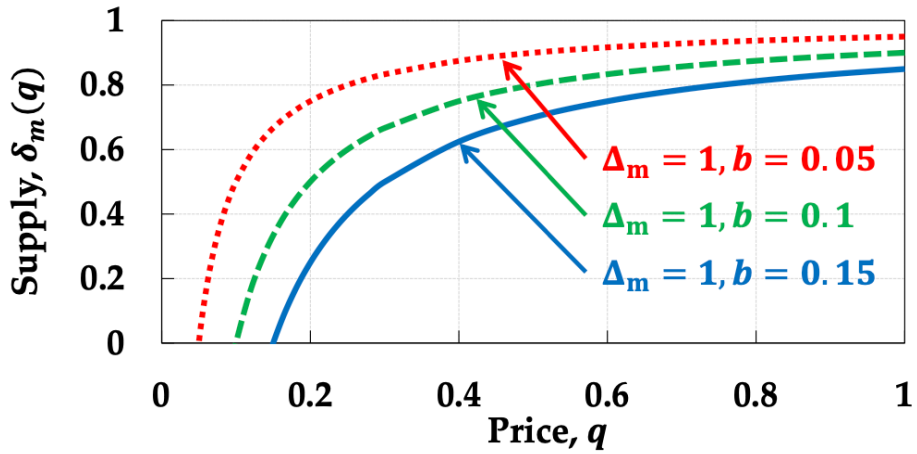
Engage HPC users in the power reduction decision

Maximum resource reduction

Bidding parameter

$$\delta_m(q) = \left[\Delta_m - \frac{b_m}{q}\right]^+$$

Price



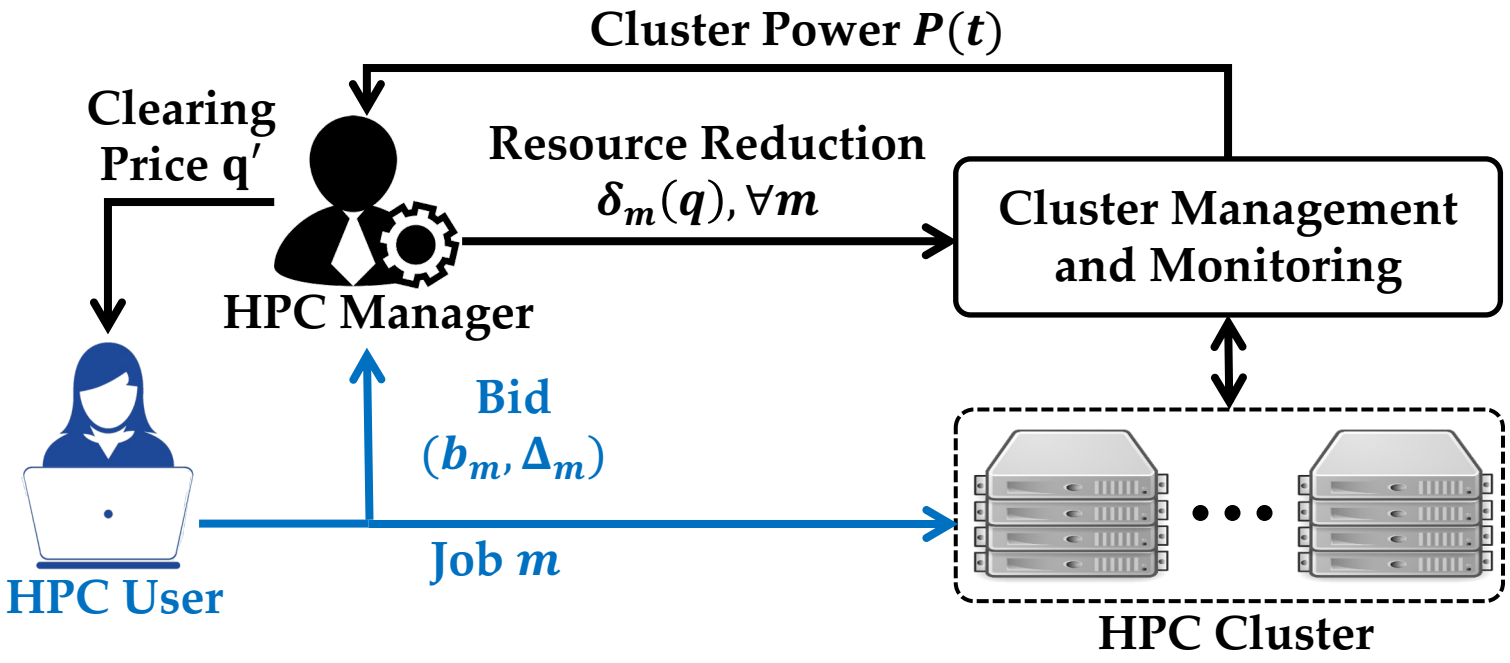
Power Reduction during Overload

MClr: $$\underset{q}{\text{minimize}} \sum_{m=1}^{M} q \cdot \delta_m(q)$$

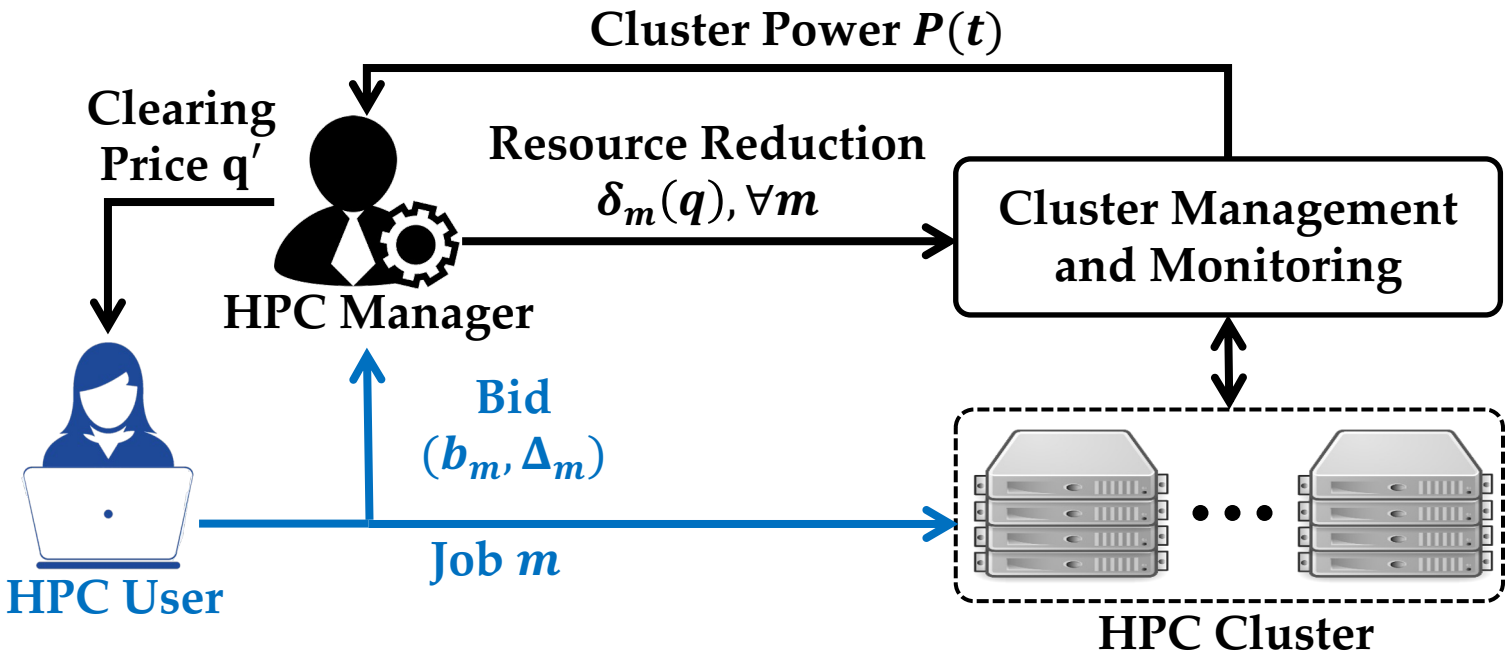
subject to $$\sum_{m=1}^{M(t)} \mathcal{P}(\delta_m(q)) \geq P(t) - C$$

HPC manager no longer needs to determine performance impact!

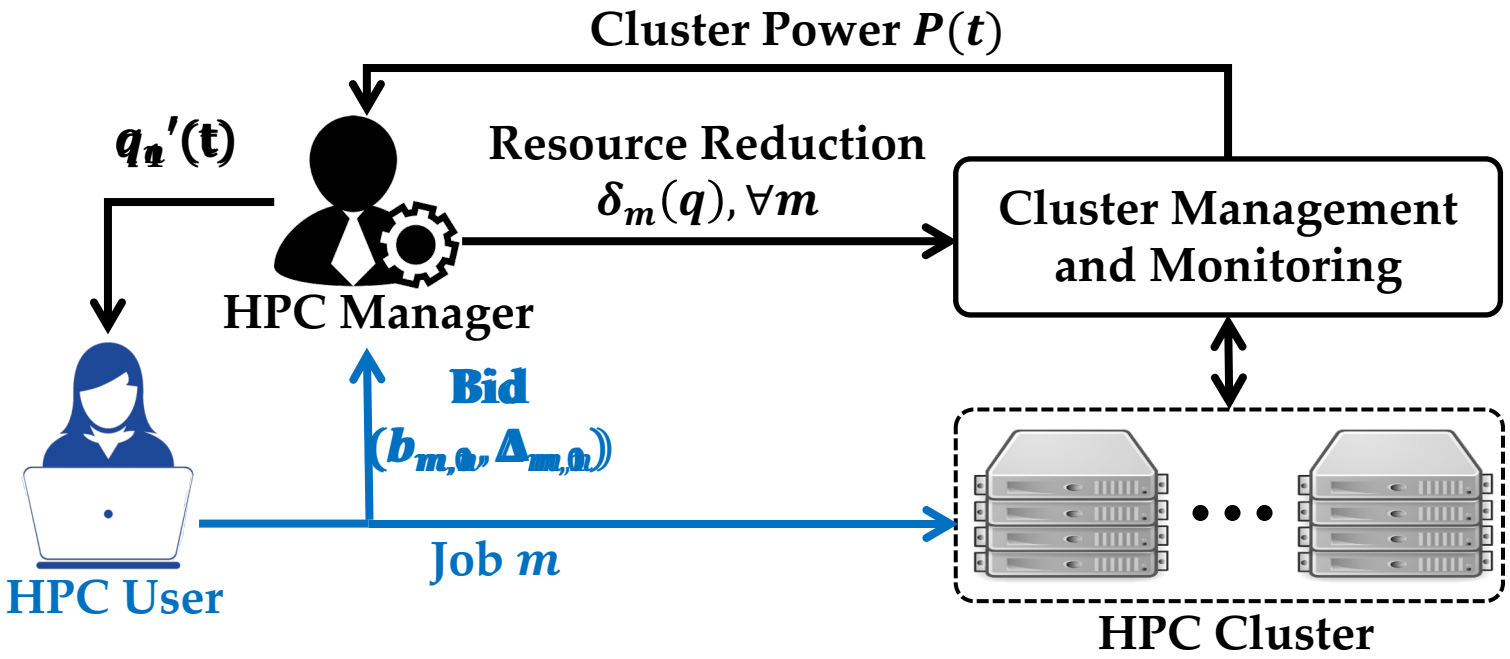
Market-based Power Reduction (MPR)



MPR with Static Bidding (MPR-STAT)

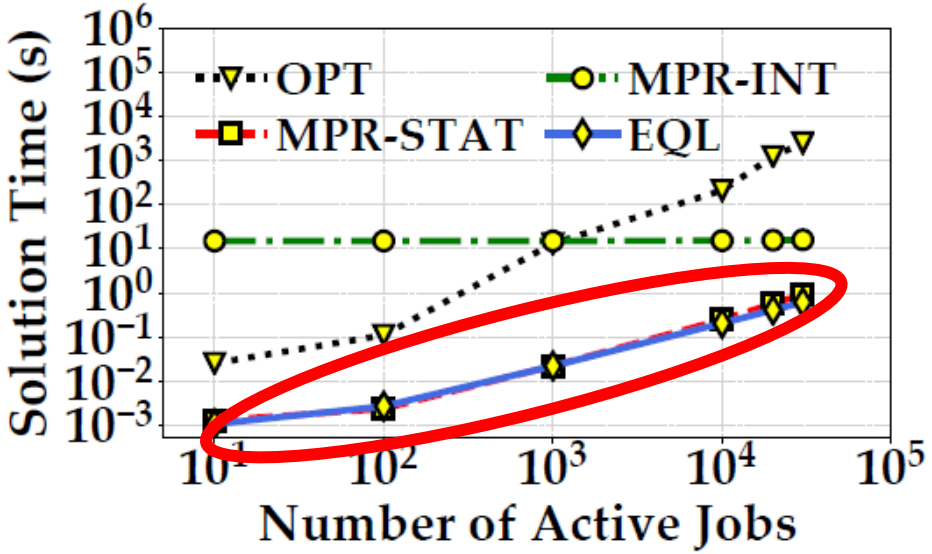


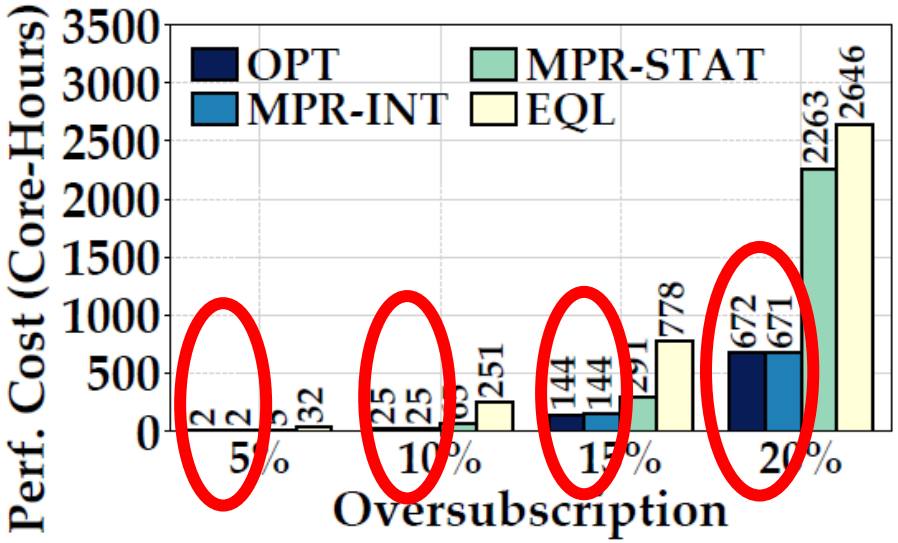
Interactive MPR (MPR-INT)



Until clearing price converges

MPR-STAT vs MPR-INT





MPR-STAT

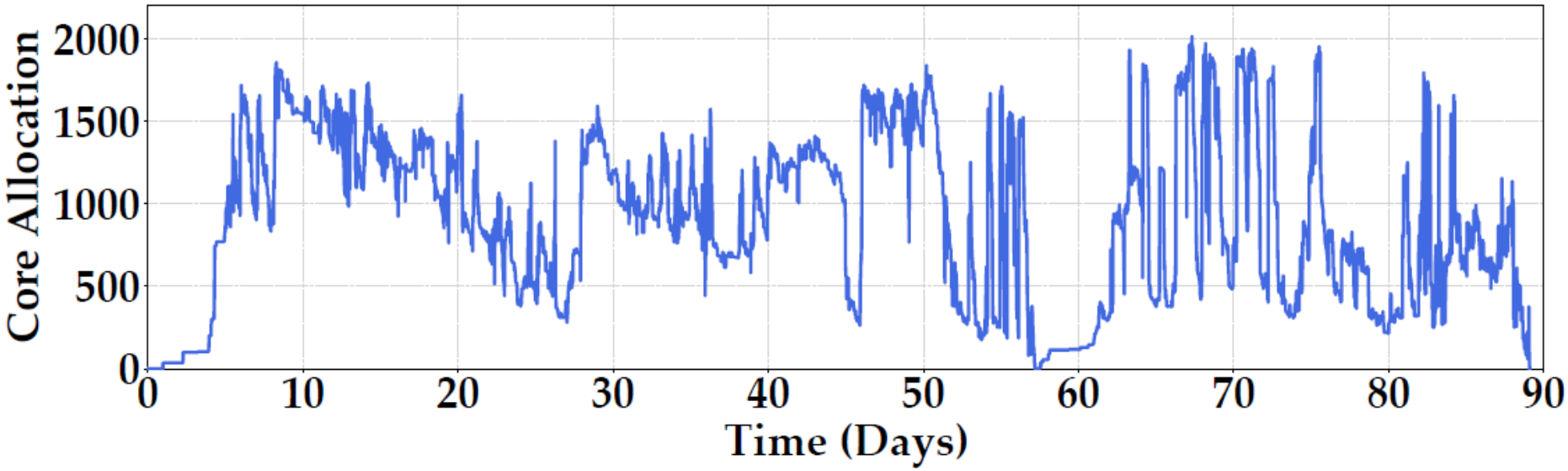
Quick solution time

MPR-INT

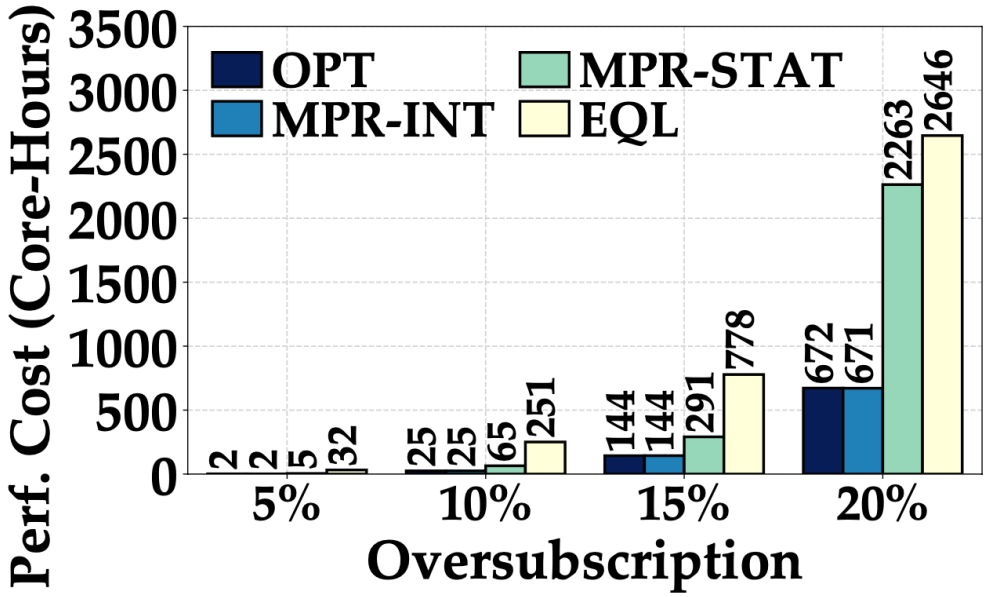
Very good performance

Evaluation

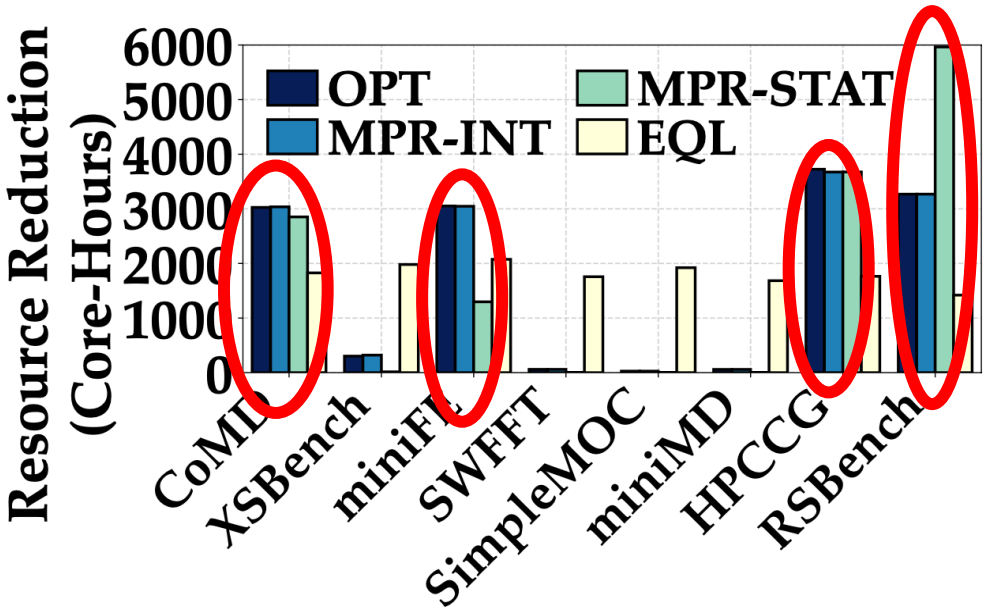
- Real-world workload traces
- Benchmarks: OPT vs EQL vs MPR



Benchmark Comparison



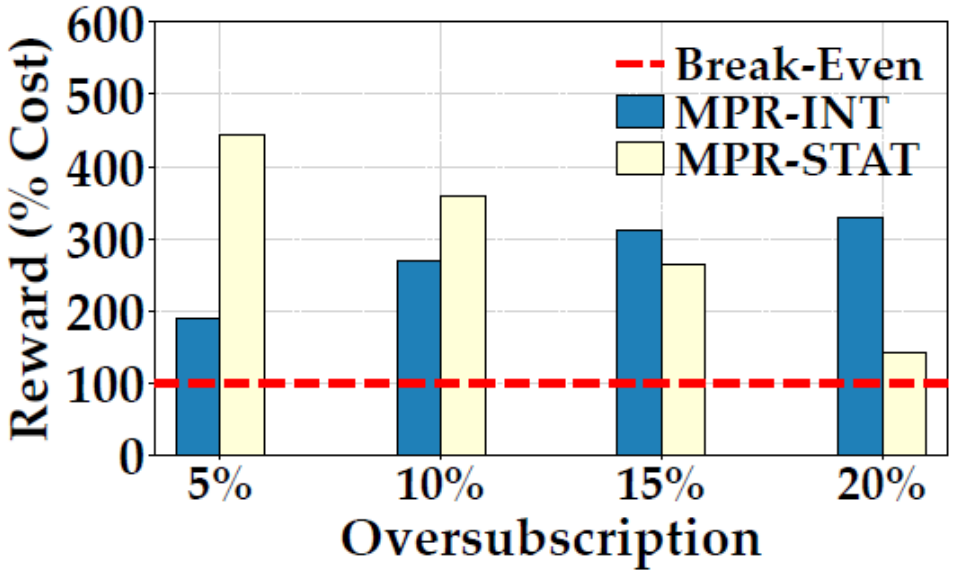
Performance cost



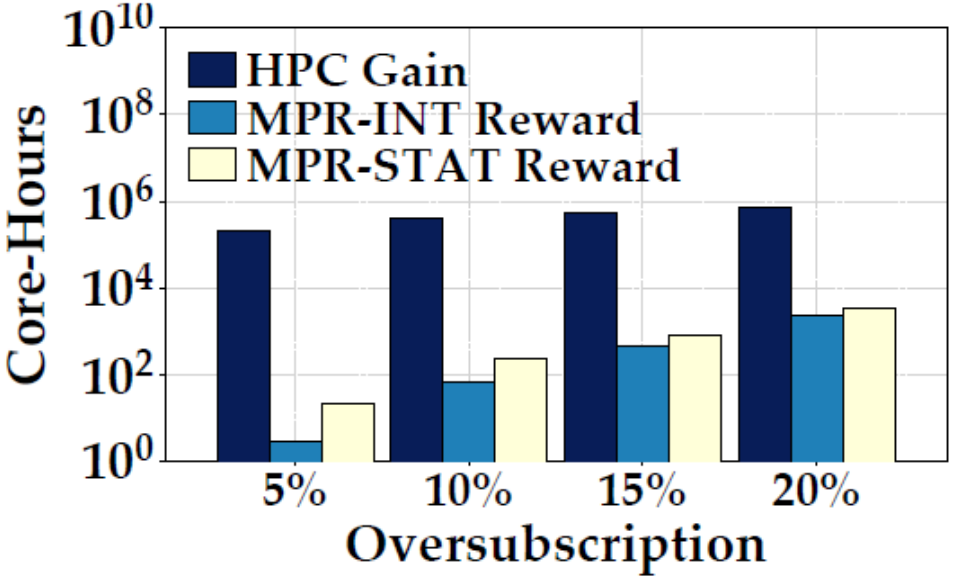
Profile-wise resource reduction

MPR and OPT incurs lower performance cost

Market Performance



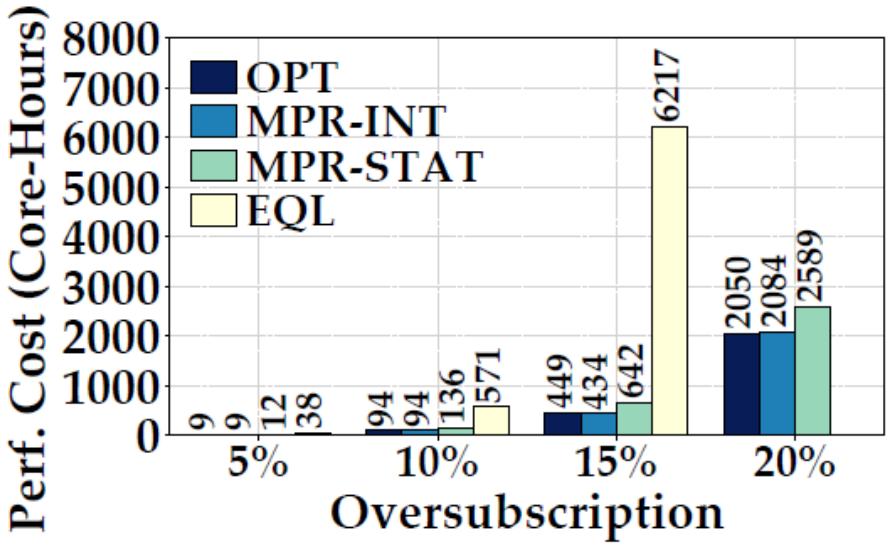
User's reward

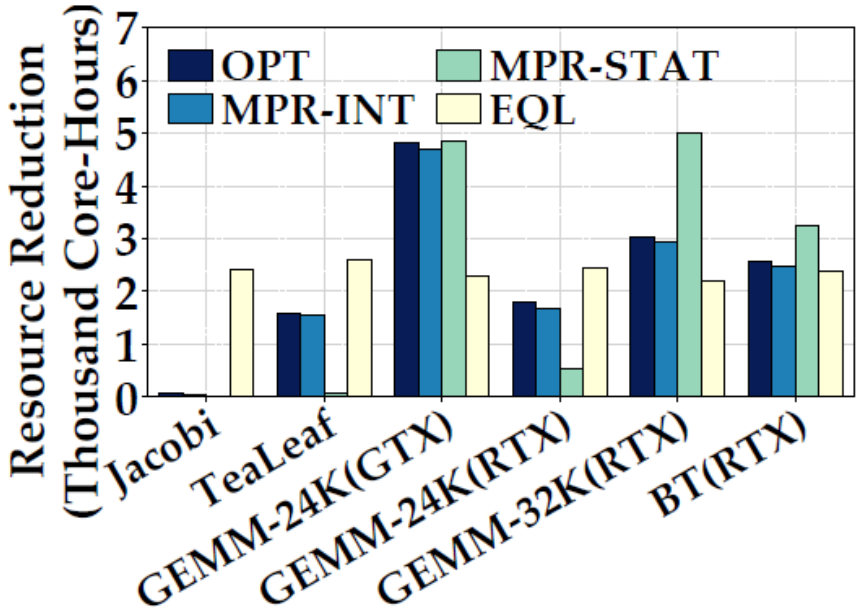


HPC manager's gain

Both HPC Users and HPC manager benefits from their participation

Heterogeneous System Performance





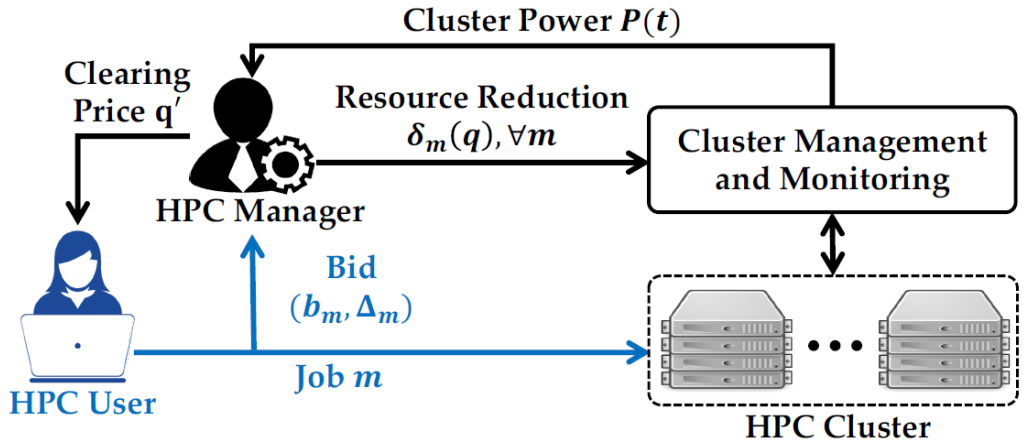
Performance cost

Profile wise resource reduction

MPR and OPT incurs lower performance cost

Key Take Away

- MPR: A market-based approach to managing oversubscribed HPC
  - Does not require job-wise power estimation and tracking
  - User-in-the-loop
  - Highly-scalable management solution
  - Go beyond power subscription – carbon reduction, demand response



Thank You!

Questions?