Do Predictors for Resource Overcommitment Even Predict?

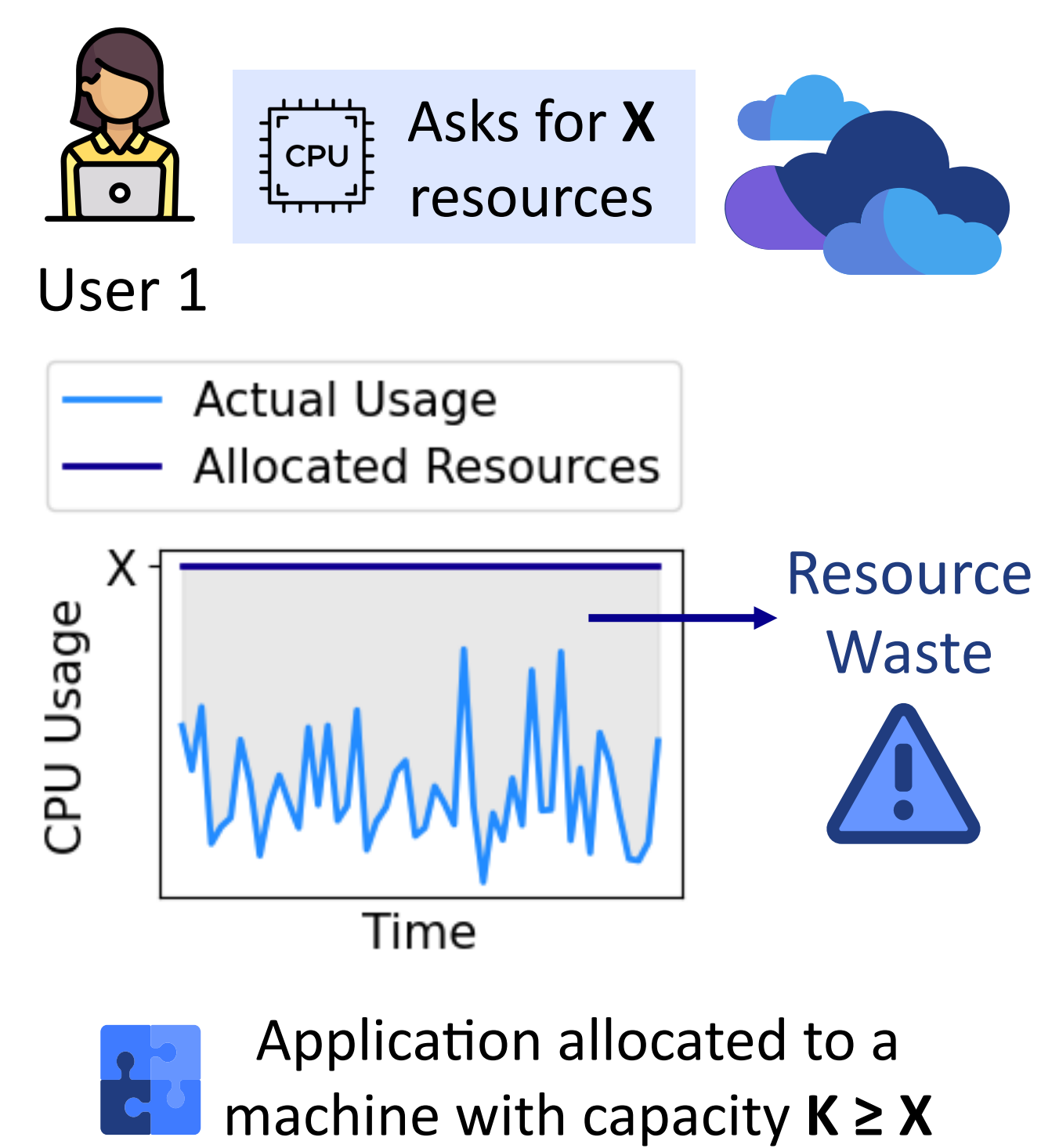Georgia Christofidi   Thaleia Dimitra Doudali

IMDEA Software Institute, Madrid, Spain
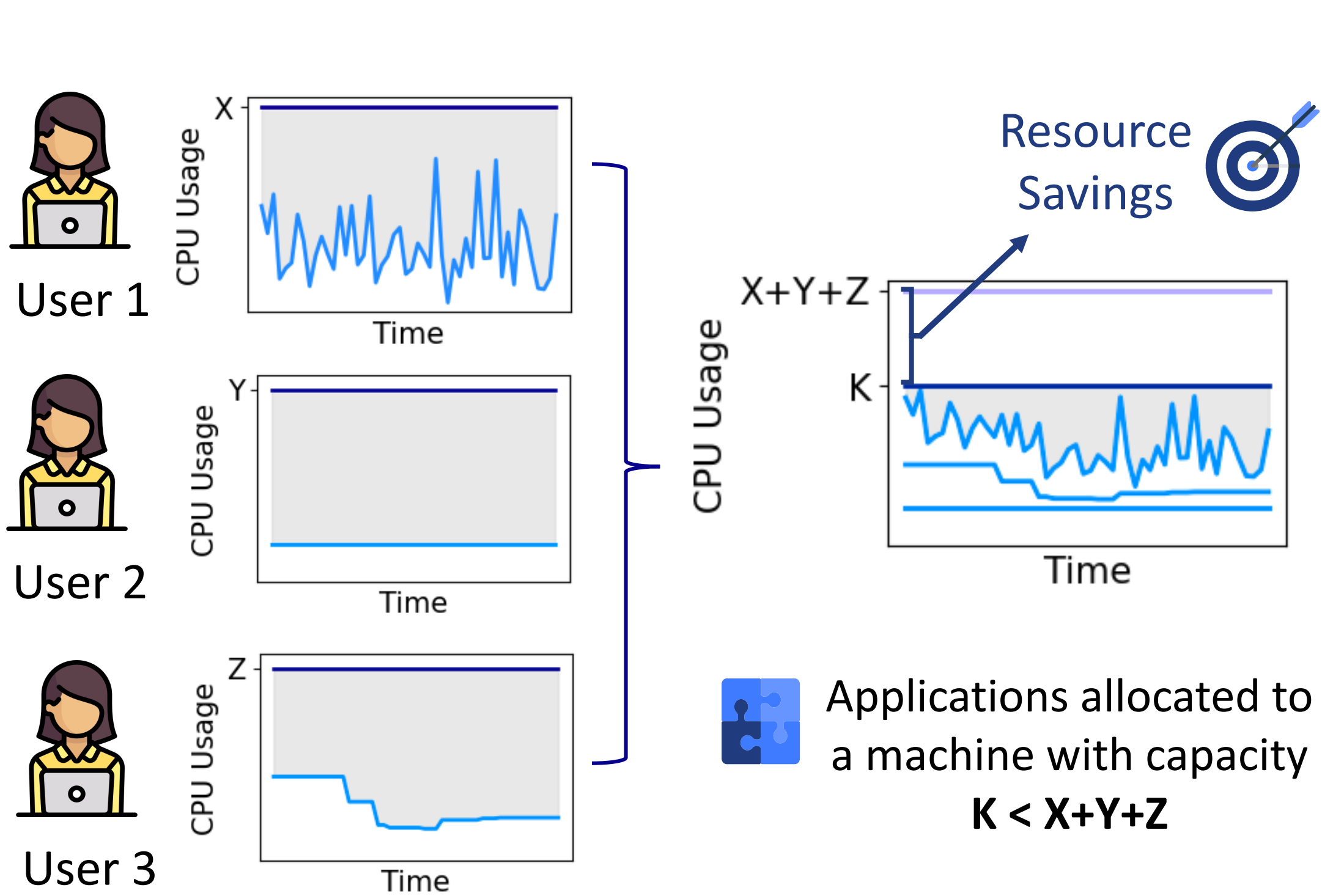
1. Problem Space

1. **Traditional** Resource Allocation

Asks for **X** resources

User 1

- Actual Usage
- Allocated Resources

Resource Waste

Application allocated to a machine with capacity **K ≥ X**

2. Resource Allocation with **Overcommitment**

User 1
User 2
User 3

Resource Savings

X+Y+Z
K

Applications allocated to a machine with capacity **K < X+Y+Z**

3. **Benefits** & **Risks** of Resource Overcommitment

✓ Minimize resource waste

✓ Improve resource and cost efficiency

✗ Resource usage underestimations

↓ Application Performance
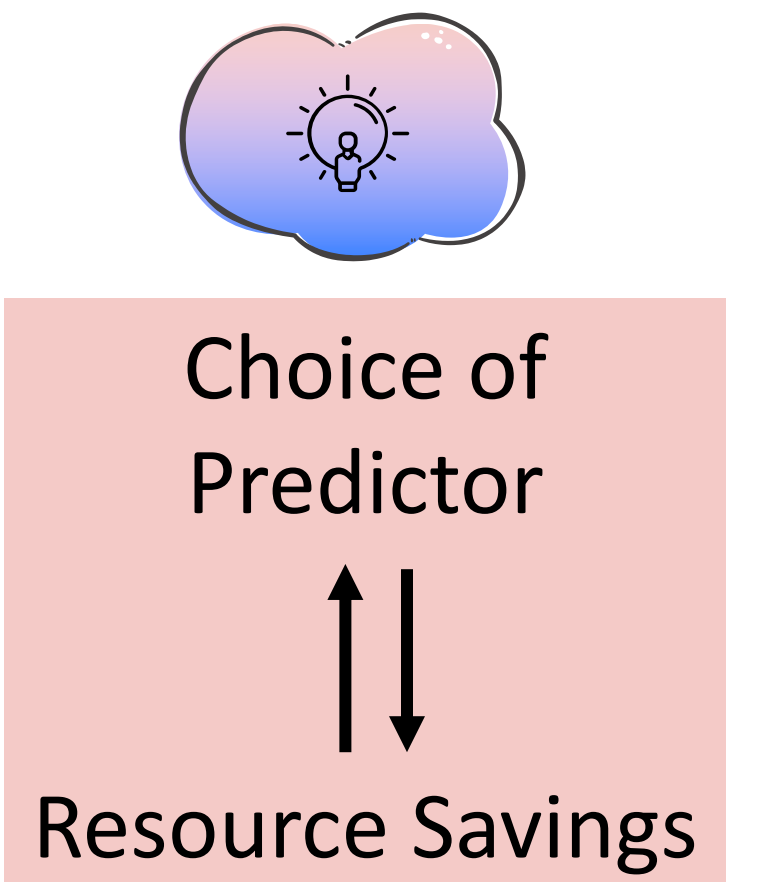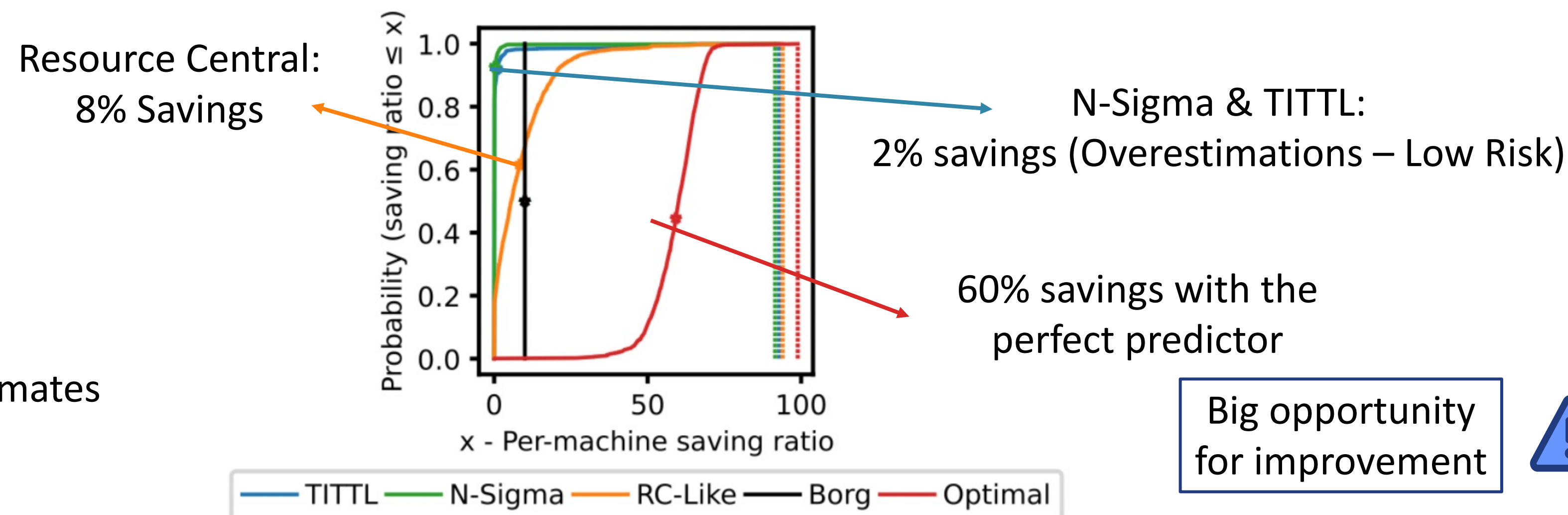↓ User Experience
Competing for resources
SLO Violations

How to realize the benefits despite the risks? **Predict** future resource usage!

2. Existing Approaches

Current Predictors

Future **U**sage =

1. Borg
90% * **U**

2. Resource Central
sum of the k-th %-ile

3. N-Sigma
$U + N * std(U)$

4. TITTL
Max(1, 2, 3)

- Conservative
- Often overestimates to avoid **risk**

Resource Central: 8% Savings

N-Sigma & TITTL: 2% savings (Overestimations – Low Risk)

60% savings with the perfect predictor

Big opportunity for improvement
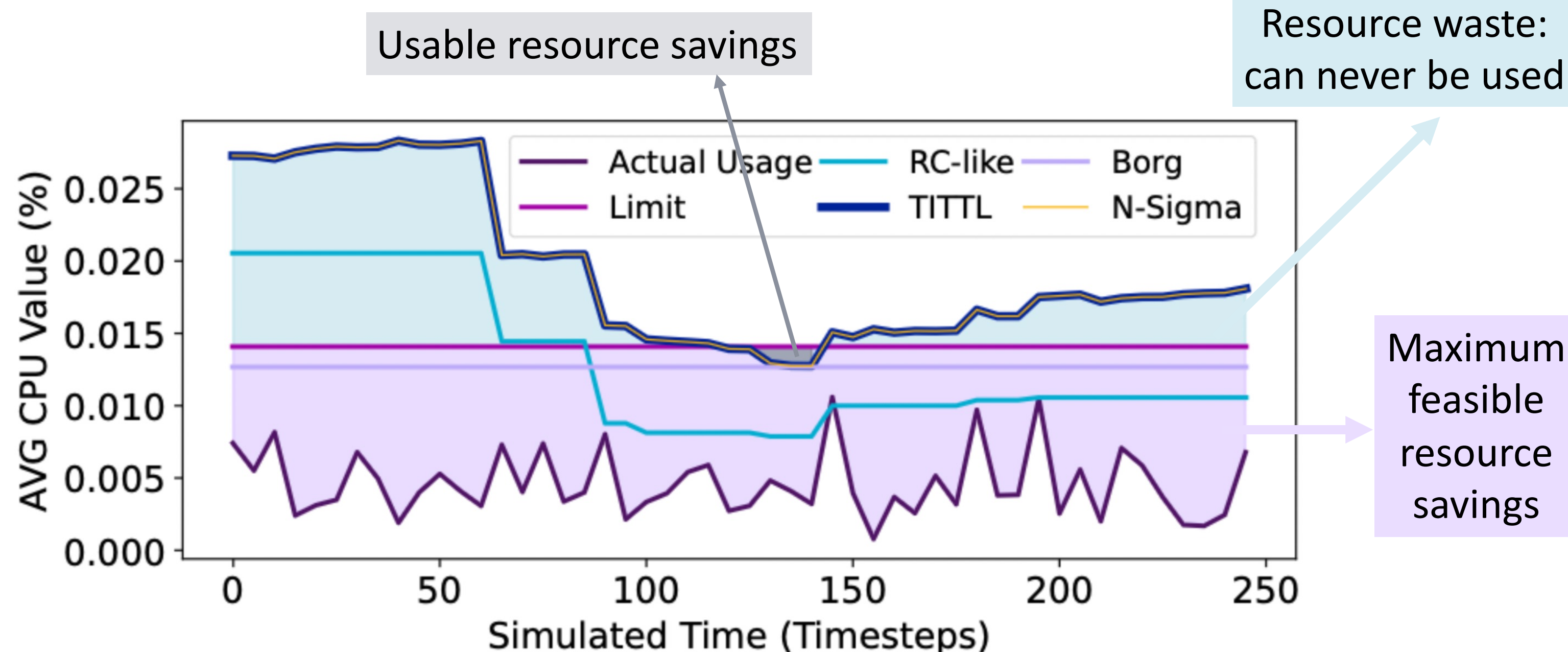
Choice of Predictor ↑↓ Resource Savings

3. Experimental Analysis

Our Insight:

Predictions >> Actual Usage

Predictions > User-asked Resources (Limit) for **94%** of times
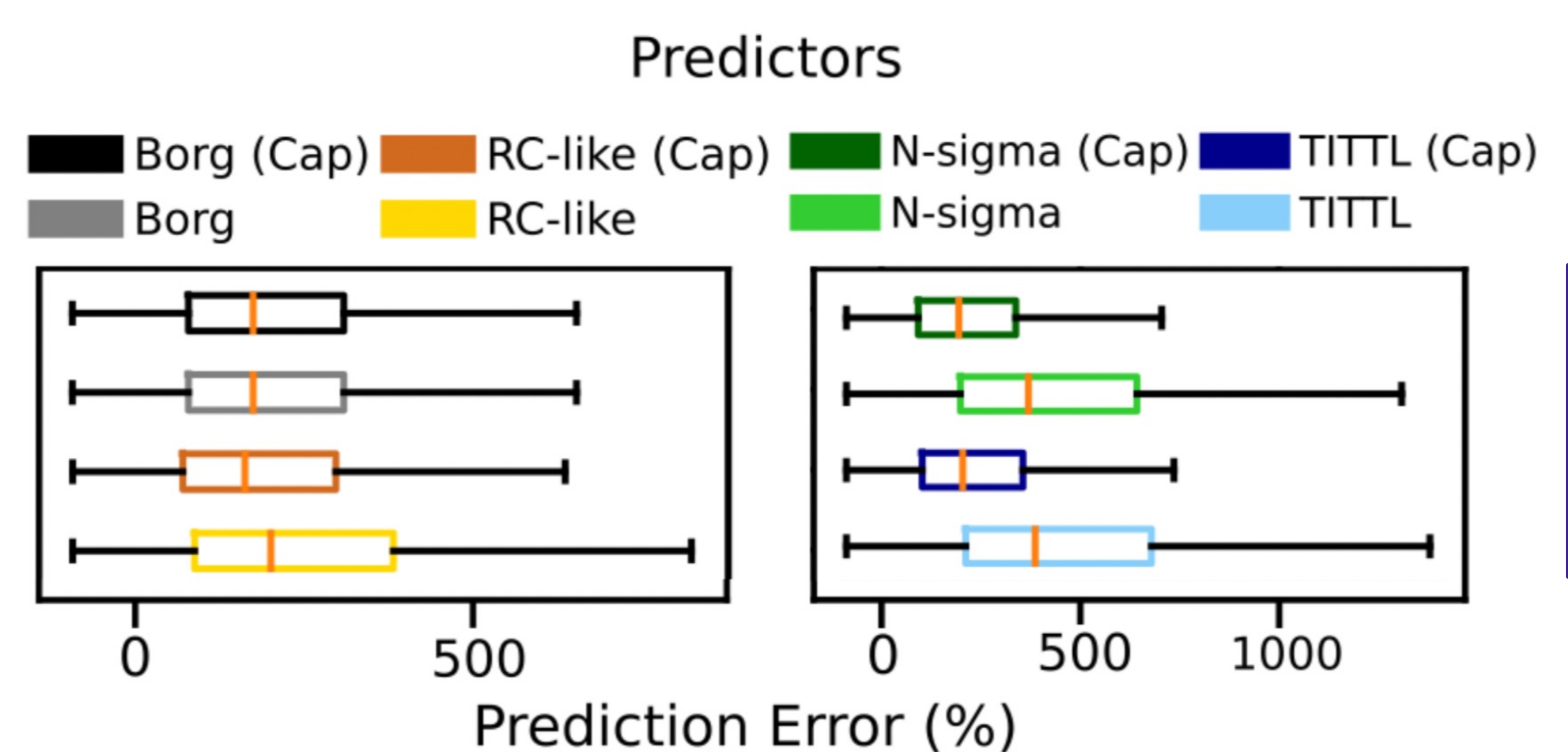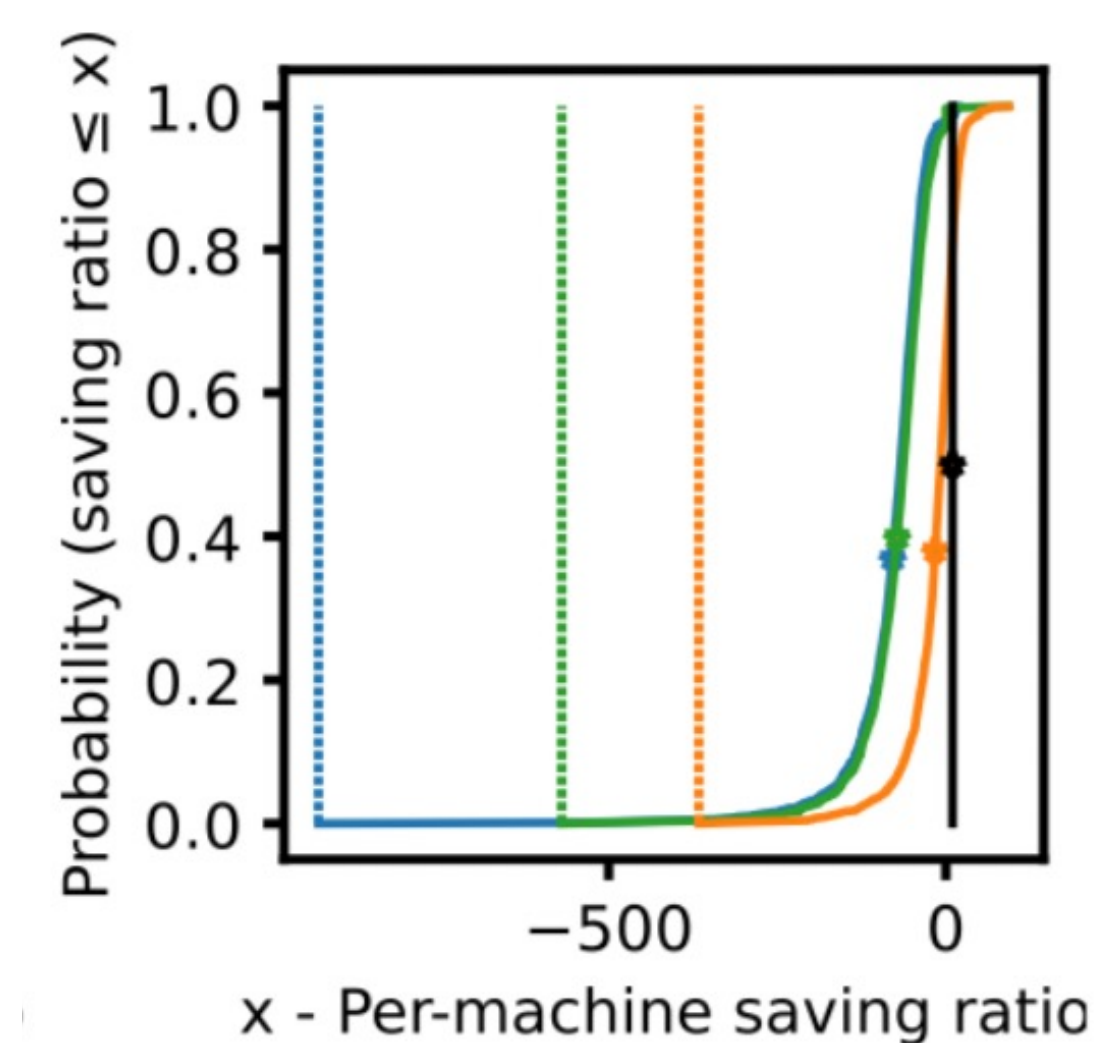
94% of times **no Overcommitment**

No Benefit!

Does the predictor even predict at all?

Threshold: Limit

If *Prediction > Limit:* then *Prediction = Limit*

Usable resource savings

Resource waste: can never be used

Maximum feasible resource savings

Savings without the threshold

Negative values → Resource **Waste** instead of Savings

Predictors

- Borg (Cap)
- RC-like (Cap)
- N-sigma (Cap)
- TITTL (Cap)
- Borg
- RC-like
- N-sigma
- TITTL

568% Average Relative Error

Current predictors severely **overestimate** resource usage.

4. Summary

Current predictors for resource overcommitment have low effectiveness.

- Resource usage overestimation
- Very little resource savings
- Overcommitment rarely achieved

A highly accurate predictor can enable **59% higher resource savings**, on average, compared to current approaches.

Future Research

What predictive model to use?

🔍 Lightweight solutions

🔍 Practical Integration of ML & simple heuristics

🔍 High Prediction Accuracy

We need a predictor that accurately predicts and closes the existing gap in attainable resource savings.

Scan for code and paper: