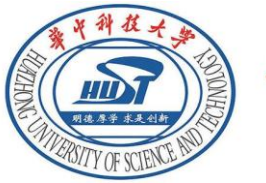


華中科技大學
HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

PreMatch: An Adaptive Cost-Effective Energy Scheduling System for Data Centers

Daping Li, Jiguang Wan, Nannan Zhao, Duo Wen, Chao Zhang, Fei Wu and Changsheng Xie

Wuhan National Laboratory for Optoelectronics, HUST, China

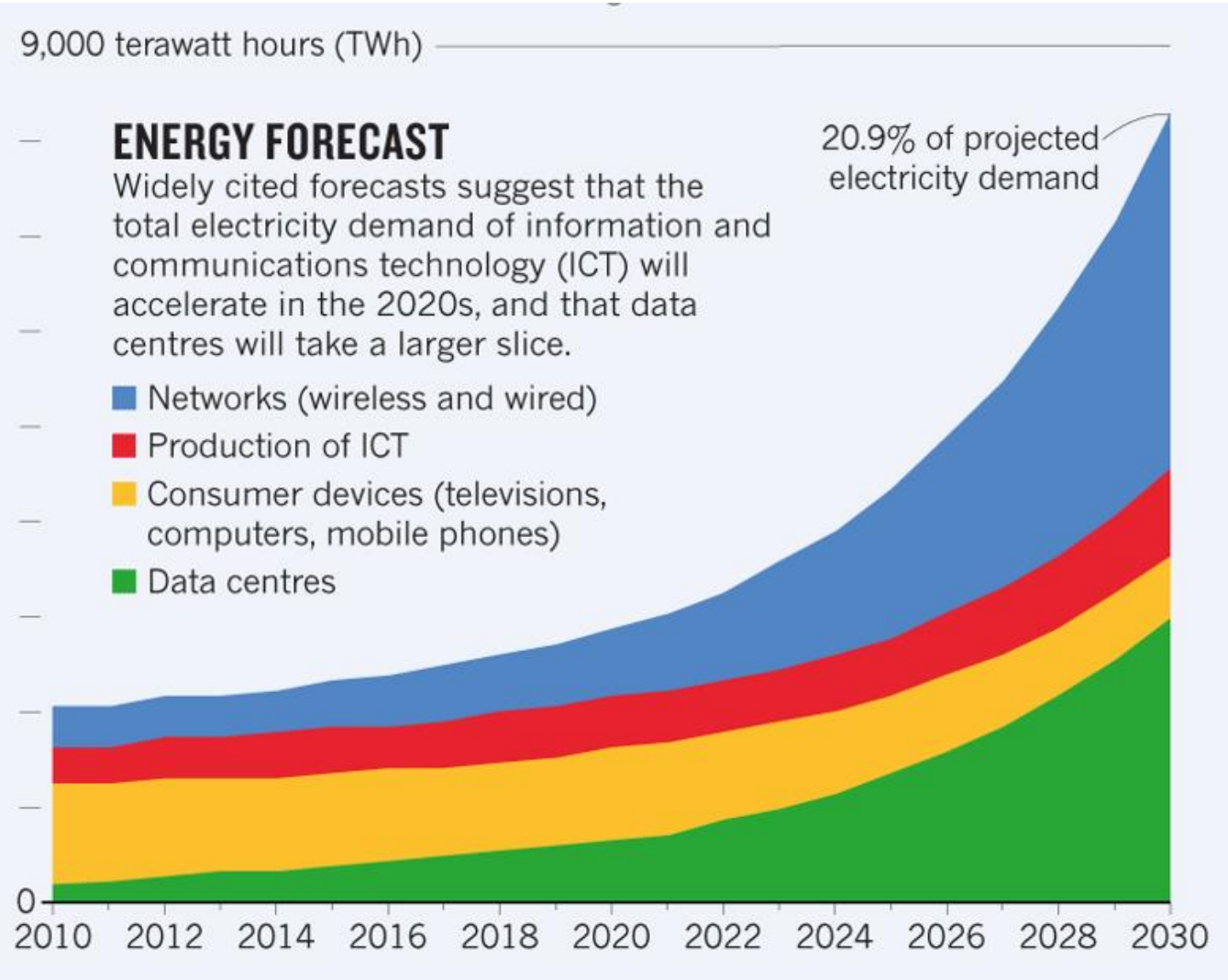
MSST 2020

CONTENTS

01

Background and problem

Part 1 Background



Source: How to stop data centres from gobbling up the world's electricity, Nature, 2018

- **Electricity demand of ICT**
  - 2020, about 7% of total electricity, flat
  - Worst in 2030, about 21% of total electricity
  - accounts for more than 2% of global emissions
- **Data centers take 1/3 of ICT**
  - Storage takes about 1/5 to 1/2

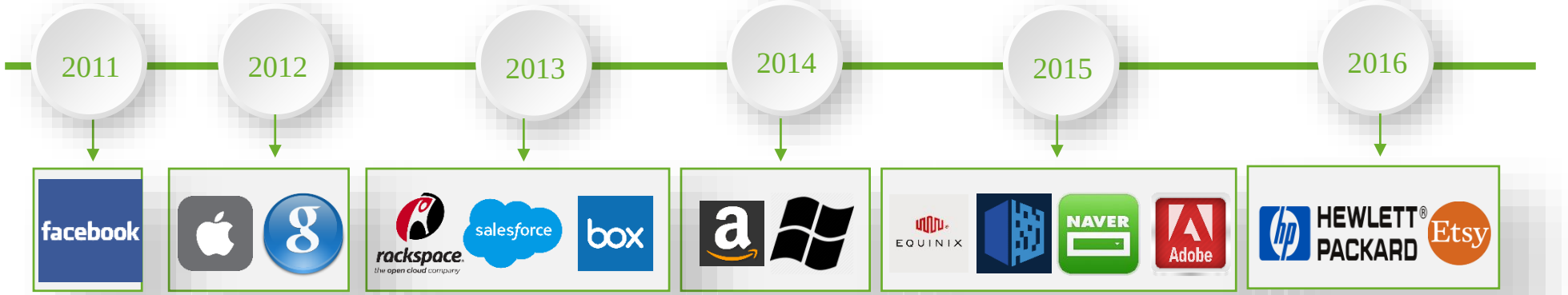
Part 1 Background

- **Why green energy?**
  - Reduce the carbon footprint
    - ✓ slow down global warming
  - Business sense
    - ✓ Green electricity will be cheaper than traditional electricity by 2027
- **Many companies are using green energy**
  - Facebook, Apple, and Google have firstly made 100% renewable commitments

Apple

Google

facebook



Source: Click Clean Renewable Energy Tracker (January 2017) www.clickclean.org

Part 1 Problem

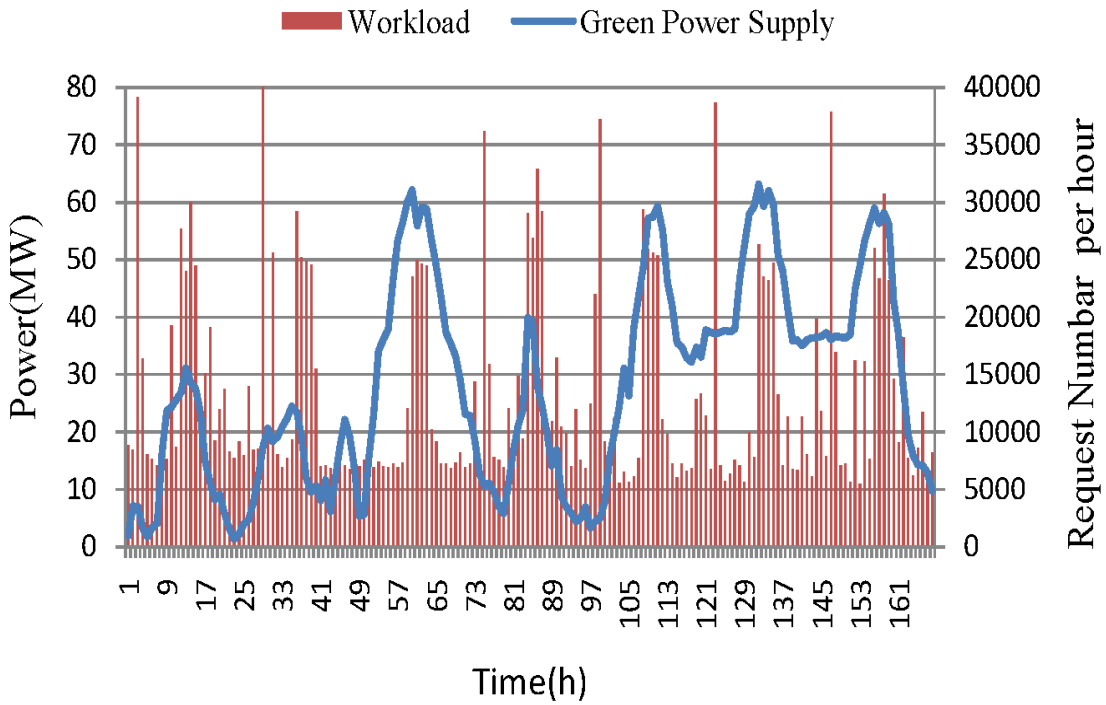
workload

mismatch

Green energy

uncontrollable

variable



Part 1 Existing methods

Workload-driven

- **Main method**
  - Workload variations → active devices
  - Redundant green energy will be wasted or stored in the energy buffer units
  - Traditional energy as a supplement
- **Advantages**
  - Satisfactory performance
- **Disadvantages**
  - Low green energy utilization
  - High cost and environmental contamination

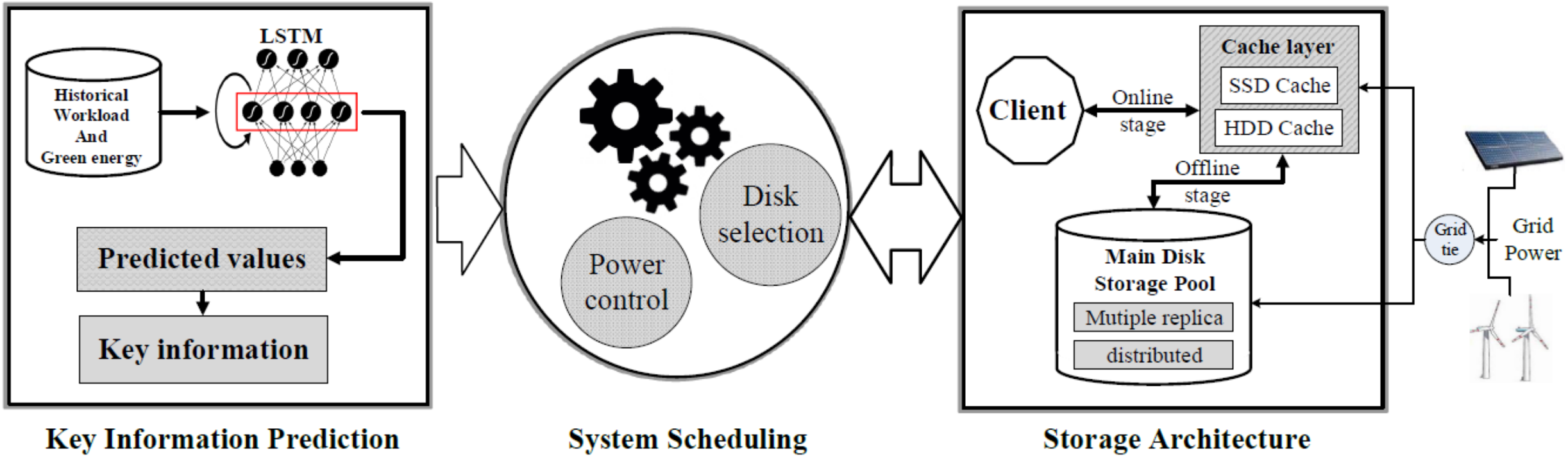
Supply-oriented

- **Main method**
  - Green energy variations → active devices
  - Delay most of the latency-insensitive tasks
  - Migrate workloads between devices
  - High-performance devices as the cache
- **Advantages**
  - High green energy utilization
- **Disadvantages**
  - Low performance

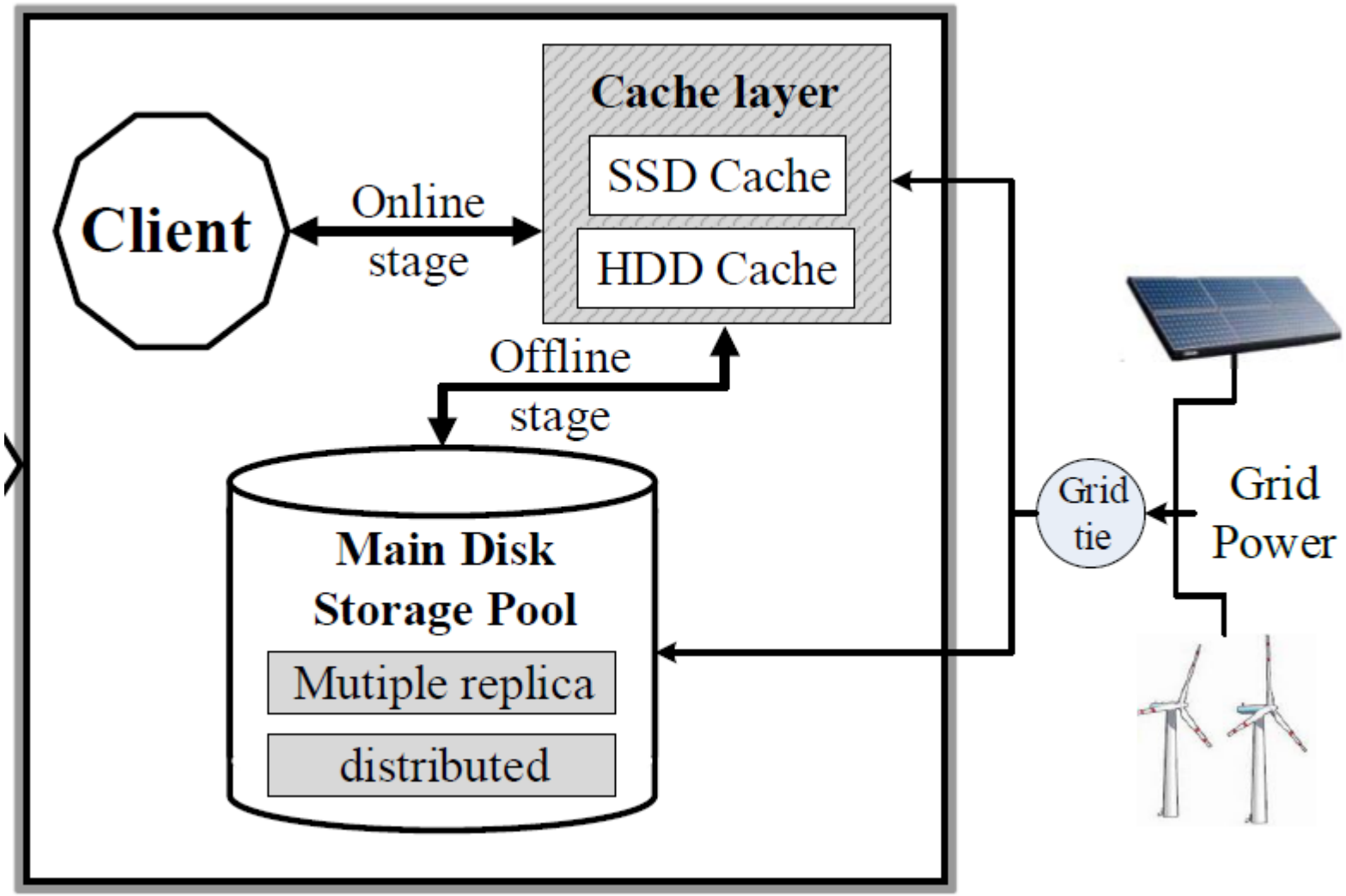
02

PreMatch

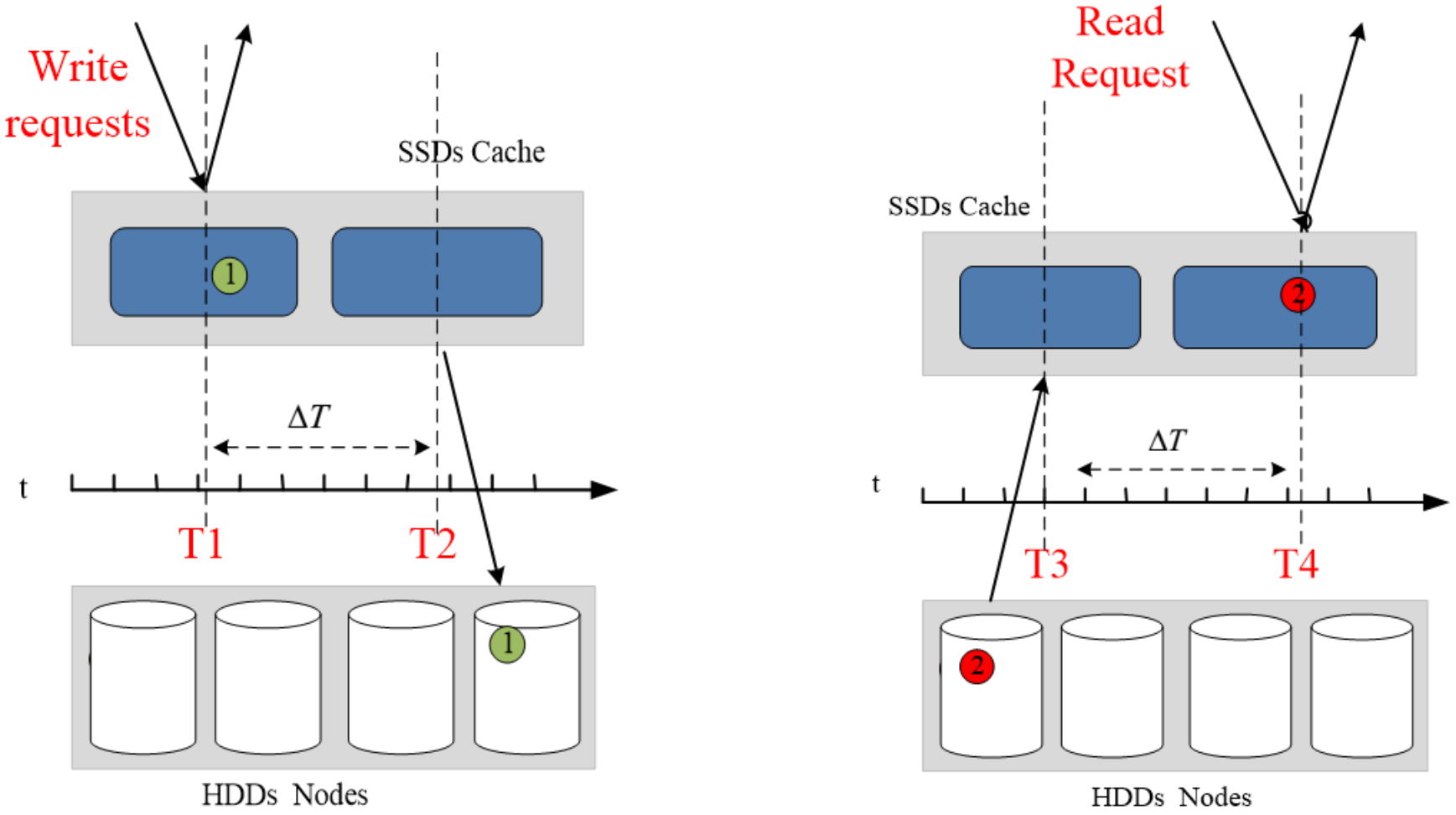
Part 2 Architecture



Part 2 Storage architecture

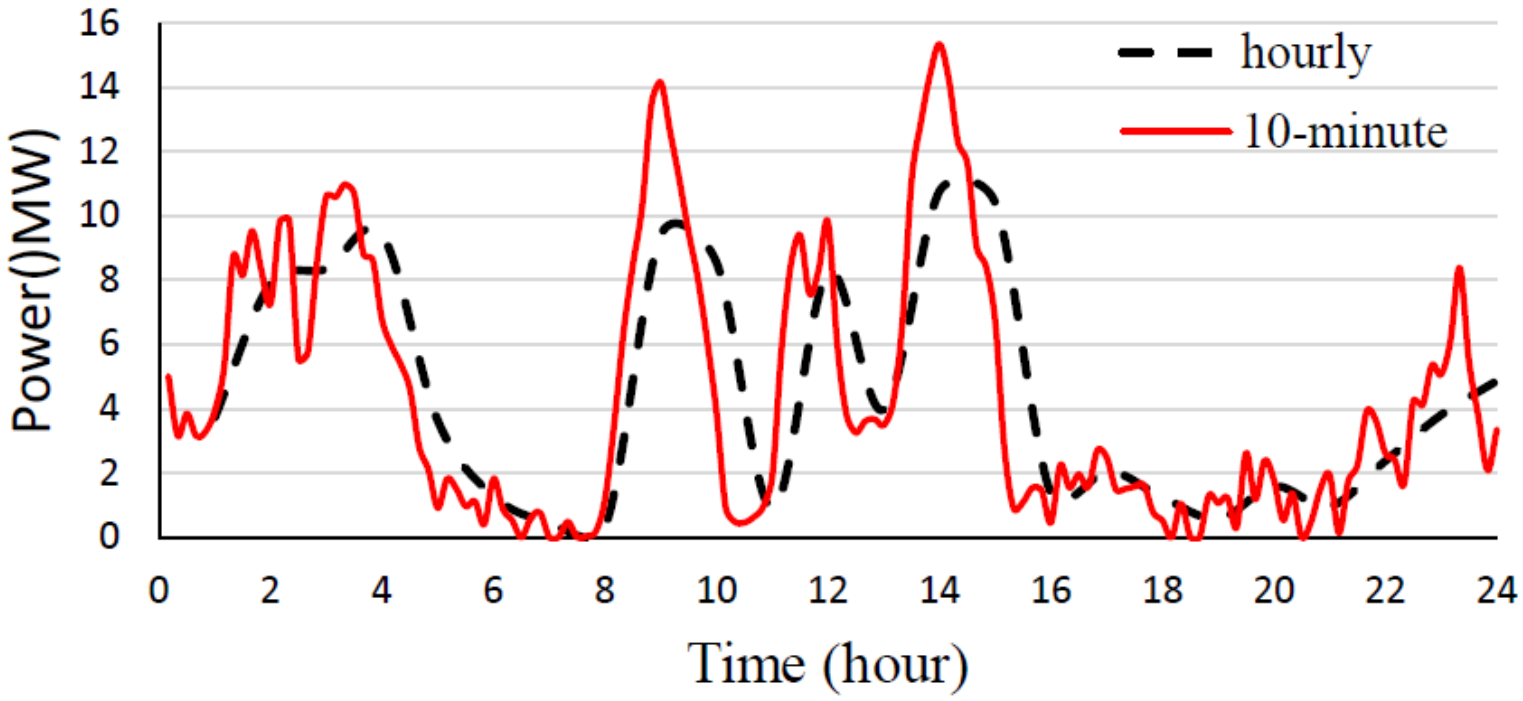


Part 2 Storage architecture—two stages



Part 2 Key information prediction—problem analysis

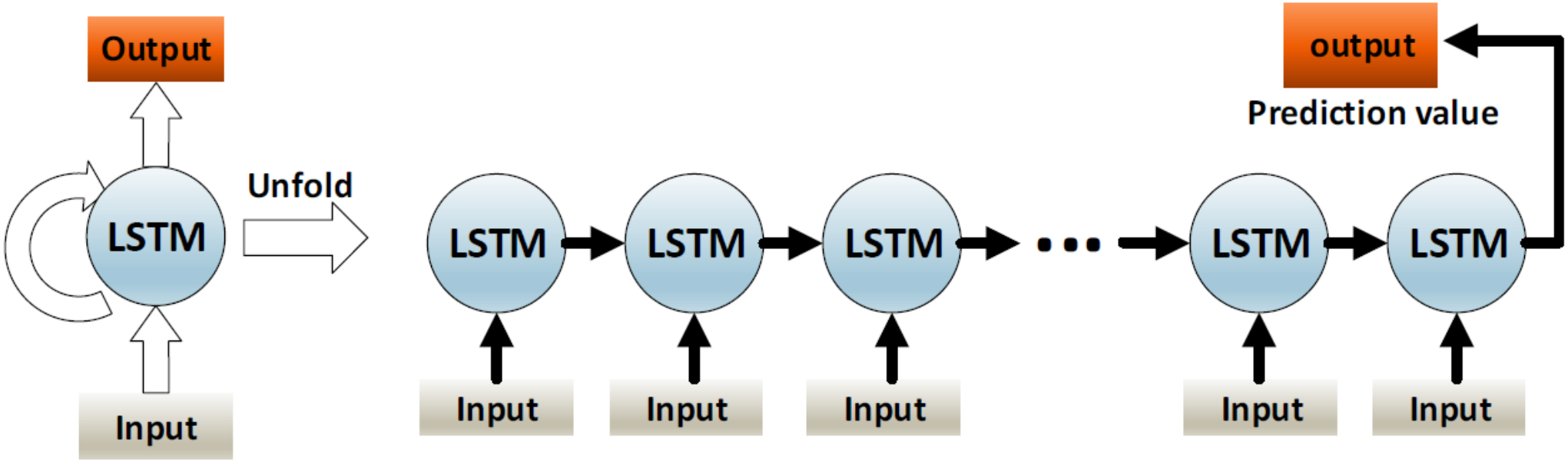
- Workload and green energy is variable



the hourly and 10-minute time interval data curve of green energy supply

Part 2 Key information prediction—prediction technology

- Long Short-Term Memory(LSTM)
  - ➢Recurrent Neural Network
  - ➢Retain information from the historical data



Part 2 Key information prediction—what we need?

- **Direct prediction:**
  - ➢The accurate values of the following local and long-range cycle
    - ✓The accurate value of workload is labeled 1(heavy) and 0(light)
  - ➢Should be predicted by the LSTM from the historical data
- **Indirect prediction:**
  - ➢The upward (1) and downward (0) variation trend
  - ➢Obtained from the historical and directly predicted data

Part 2 System scheduling –power control

- Control the number of active devices
  - The SSD cache and the HDD cache will always be active
  - Green energy will supply the primary replica firstly
  - The non-primary replicas are always off to save energy
- The active P-disks ↔ dominant one of the green energy and workload
  - workload heavy, powering on all the primary P-disks
  - workload light, control the number of active P-disks to match the green energy variation
- Considering the local and long-range variation trends to avoid frequent switching
  - For example, if workload is downward in the local cycle but upward in the long-range cycle, nothing will be done.

Part 2 System scheduling –disk selection

- Primary replica
  - ➢The hottest P-disks will be activated first
  - ➢The coldest P-disks will be turned off first
- Non-primary replica
  - ➢The P-disks with most dirty data will be activated first
  - ➢The P-disks with least dirty data will be turned off first

03

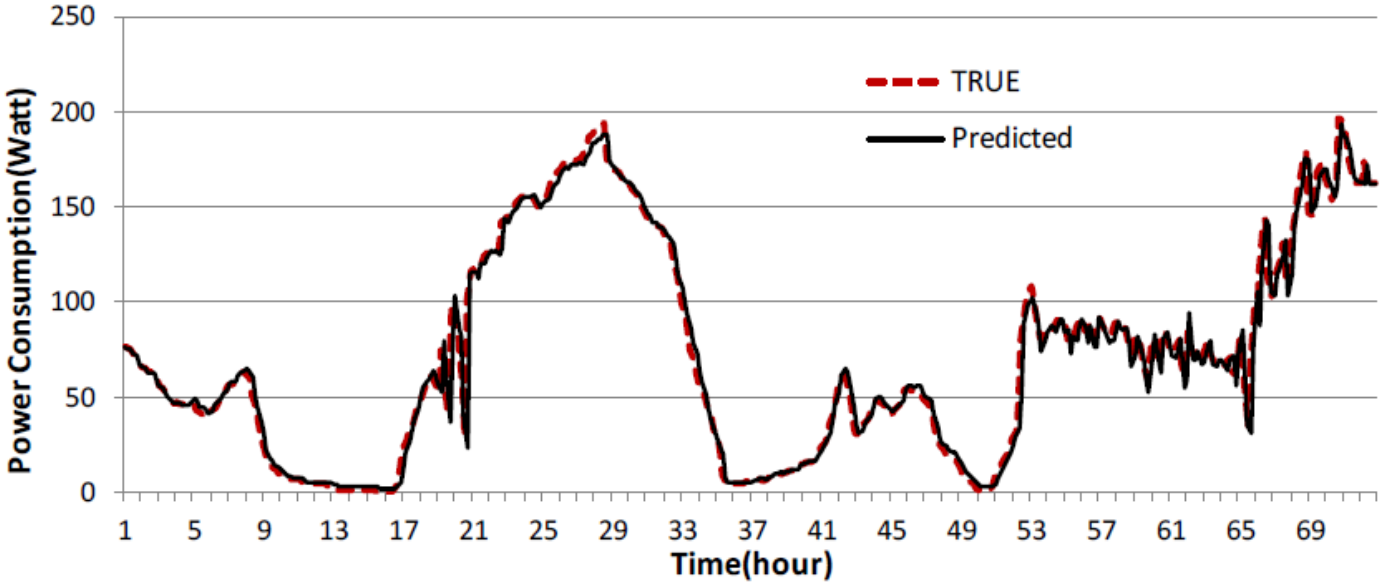
Experiment

Part 3 Experiment setup

- Four configurations used for comparison
  - Standard
    - ✓ An SSD-cache based data center without any power management policy
  - WDS(WorkLoad-Driven Scheme)
    - ✓ Keep the cache and the primary replica active
  - PreMatch
    - ✓ Keep the cache active, using the predicted information
  - PreMatch-T
    - ✓ Keep the cache active, using the real-world information
- 1 SSD cache: 3 HDD cache: 30 P-disks
- Simulation

Part 3 Experiment results—Accuracy of Prediction

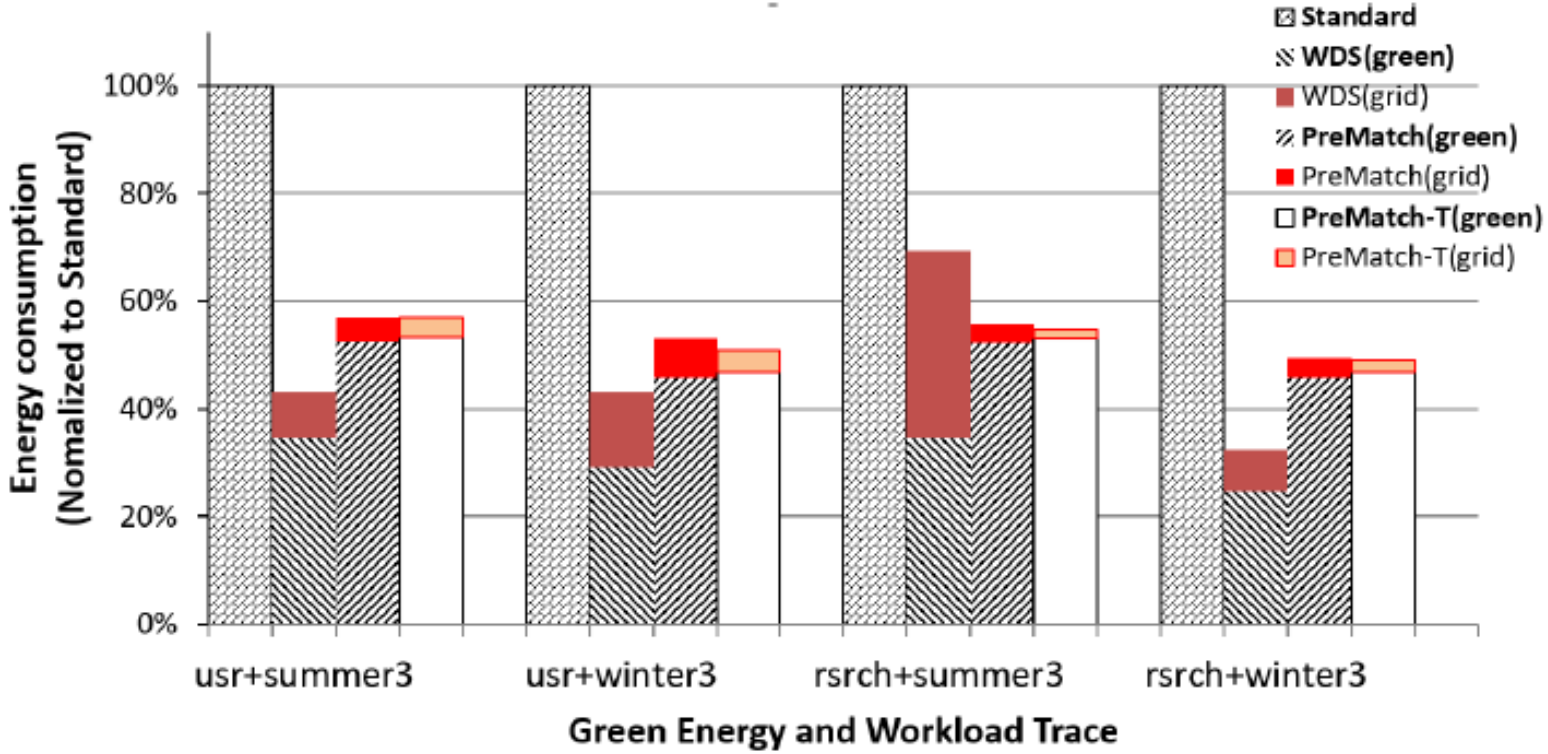
workload	local mode	local variation	long-range variation
usr	97.22%	95.35 %	97.14%
rsrch	97.47%	95.59%	95.71%
Green energy	**RMSE of local value**	**local variation**	**long-range variation**
Winter3	1.54	86.77%	80.28%
Summer3	1.84	80.26%	79.05 %



- most of the accuracy values of workload are above 95%.
  - ✓ As only 1(heavy or up) and 0 (light or down) is predicted
- accuracy of green energy is above 80%
  - ✓ predicted curve is almost coincided with the true winters trace

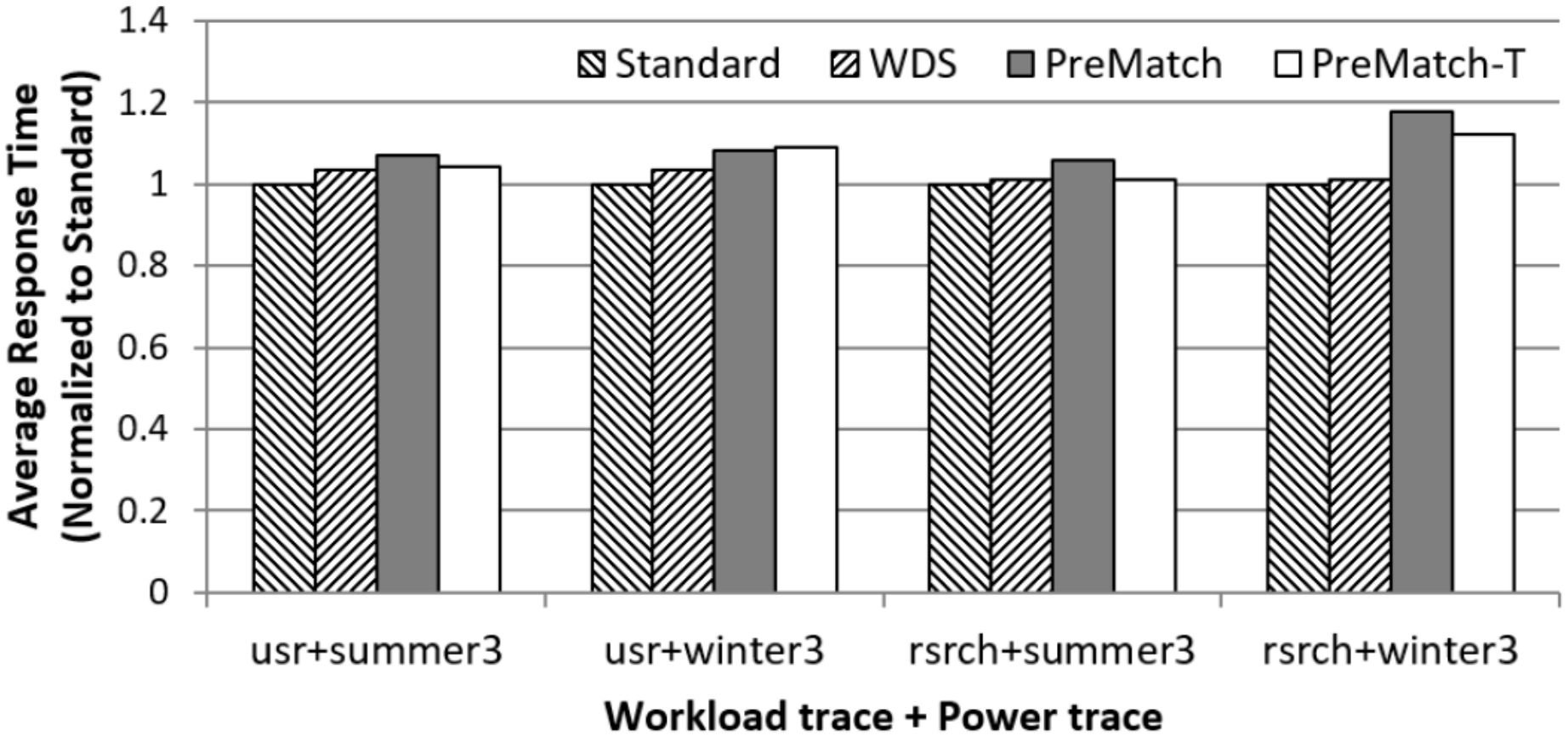
Part 3 Experiment results—energy consumption

- Standard mode consumes 100% grid energy
- PreMatch reduce grid energy up to 98.5% when compared to the standard method
- PreMatch consumes only the half grid energy of WDS at most
- PreMatch consumes more traditional grid energy than PreMatch-T



Part 3 Experiment results—performance

- PreMatch has little performance degradation

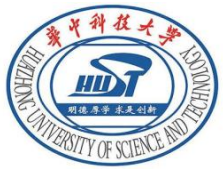


04

conclusion

Part 4 Conclusion

- Problem
  - ➢Mismatch between the workload and green energy
- The main idead of PreMatch
  - Low-energy and low-latency online stage (lack of green energy)
  - High-energy and high-latency online stage (sufficient green energy)
  - Key information prediction
  - active P-disks are made proportional to the periodically dominant one of the green energy and workload dynamically
- Effect
  - Compared with WDS, PreMatch can improves 35% additional green energy utilization ratio and reduces at least 50% grid energy with little performance degradation.



華中科技大學

HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Thanks!