Enabling Datacenter Servers to Scale Out Economically and Sustainably

Chao Li, Yang Hu, Ruijin Zhou, Ming Liu, Longjun Liu, Jingling Yuan, Tao Li

Presented by **Chao Li**

MICRO-46

Dec 10, 2013, Davis, CA

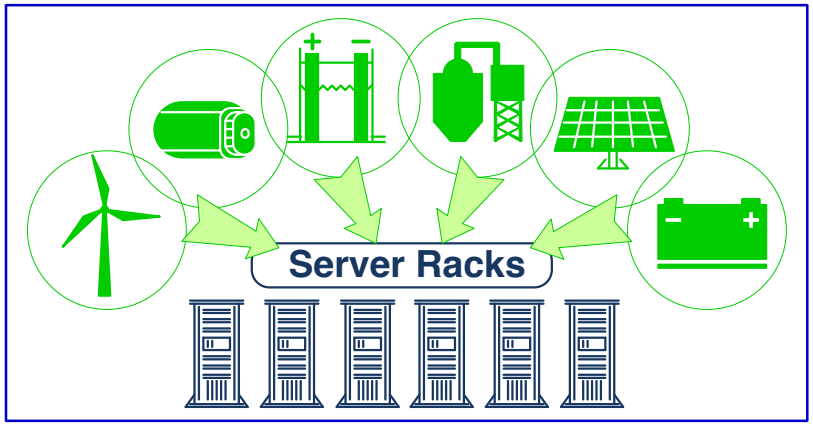
IDEAL (Intelligent Design of Efficient Architectures Laboratory)

Department of Electrical and Computer Engineering

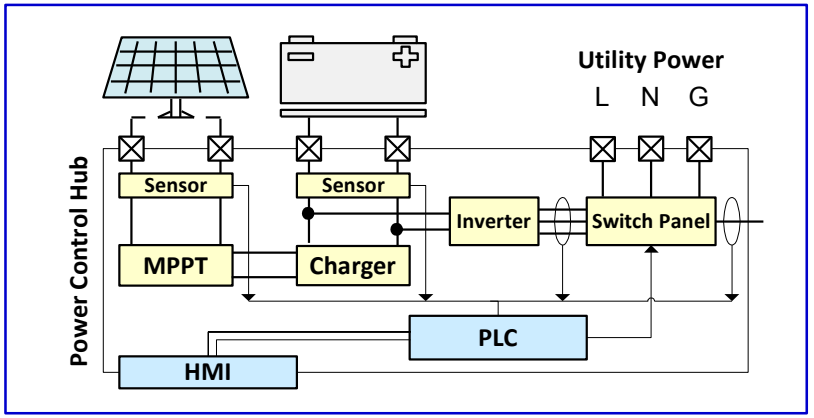
University of Florida

Talk Overview

1. Background and Motivation



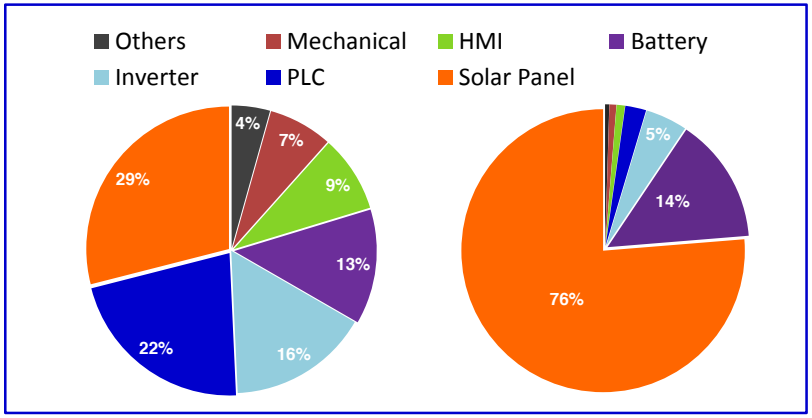
2. Oasis: Design and Prototype



3. Optimized Oasis Operation

$P_{Load} > P_{Renewable}$				
1	R	DB > 0	upsC > 60%	Switch = 'N', cpuFreq = 'Highest'
2			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
3		DB = 0	upsC > 60%	Switch = 'N', cpuFreq = 'Low'
4			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
5	U	DB > 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Highest'
6			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'
7		DB = 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Lowest'
8			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'

4. Evaluation and Discussion



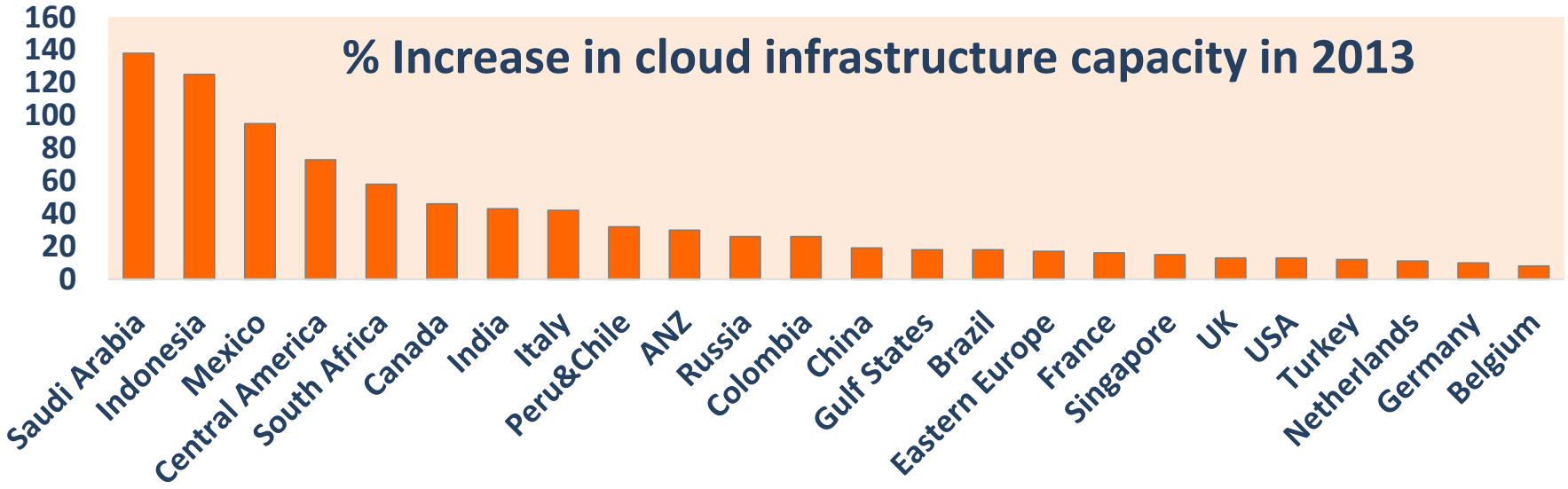
Datacenter Footprint Continues to Expand

Everything is in the Cloud

Ever-increasing user data

Endless data processing

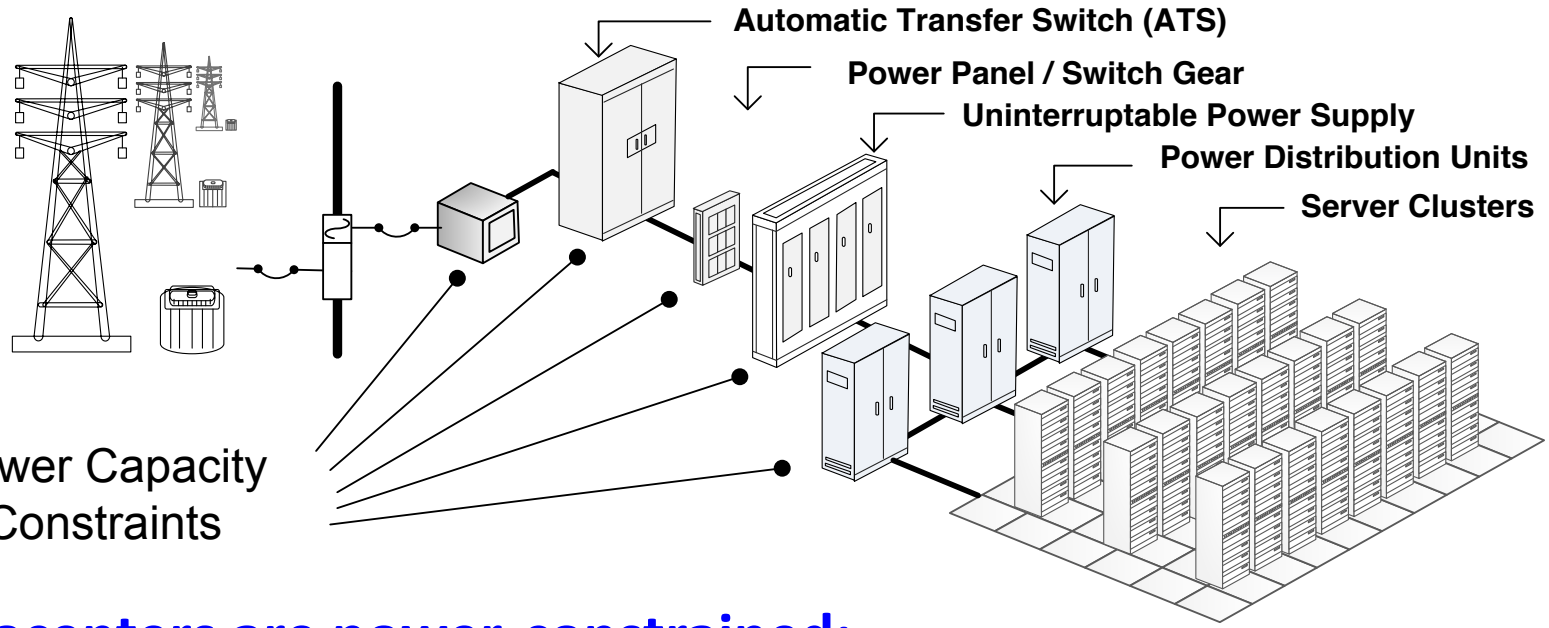
More servers are needed!



- **Horizontal scaling (scale out) has gained increasing attention**

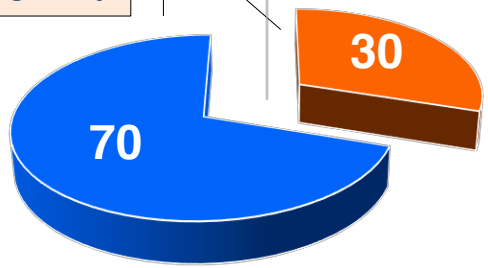
[1] DCD Industry Census 2012: Energy, http://www.dcd-intelligence.com/

The Power Provisioning Capacity Problem

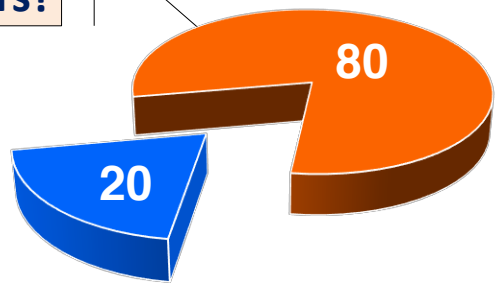


- **Datacenters are power-constrained:**
  - Limited power capacity headroom

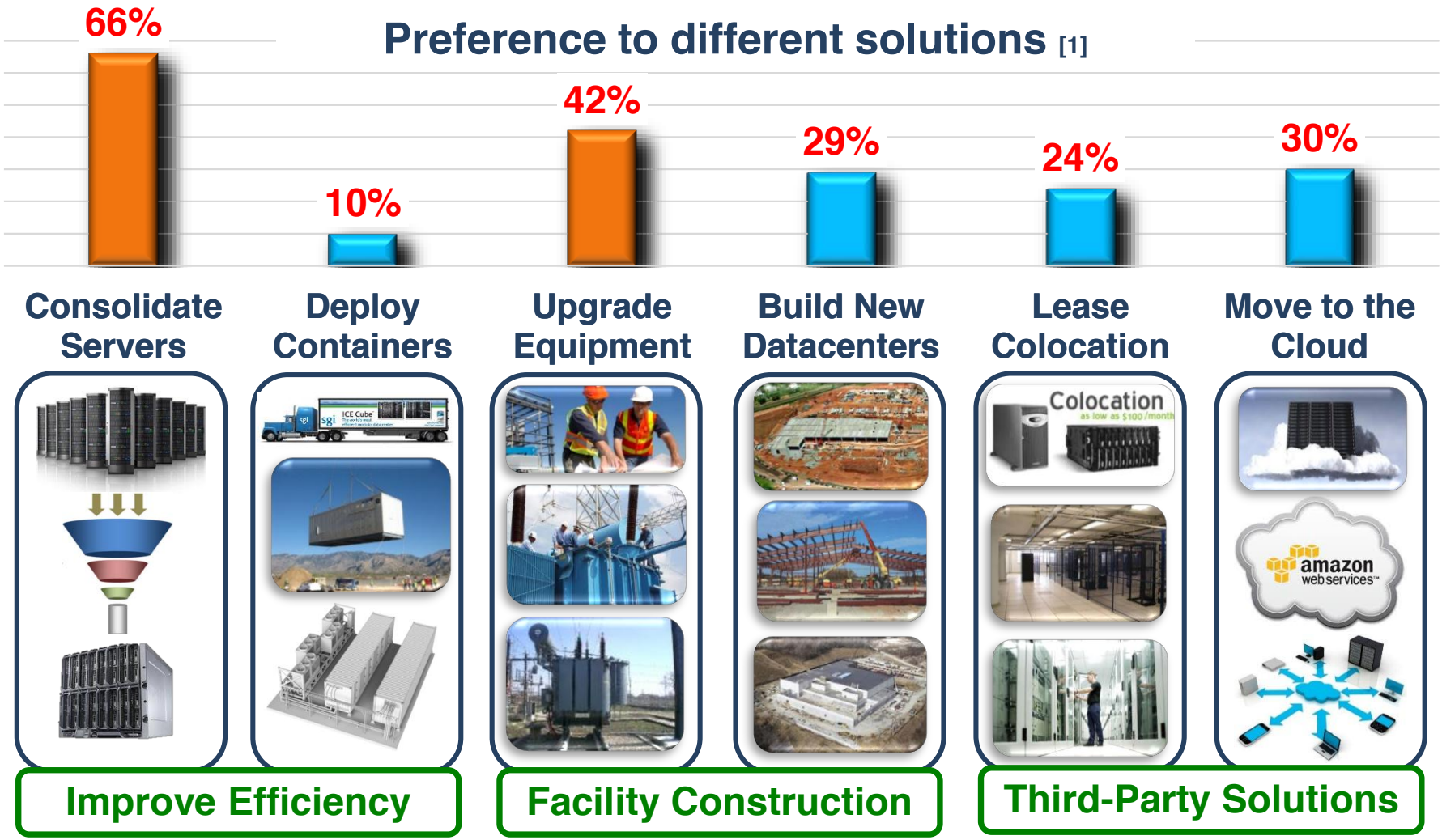
Run out of power capacity in 2012 ?



Capacity expanded in the last 5 years?



Existing Solutions

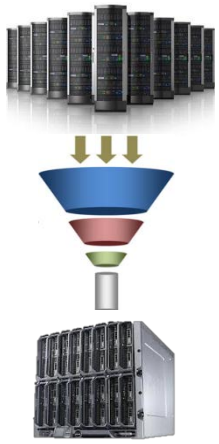


[1] the Uptime Institute 2012 Data Center Industry Survey, 2012

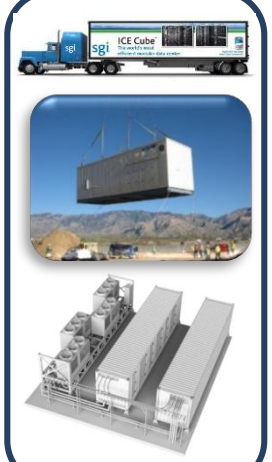
Existing Solutions

Schemes	Problems
Improve Efficiency	Power under-provisioning issue and low performance
Facility Construction	High capital investment and long construction lead time
Third-Party Solutions	Not suitable for large-scale enterprise datacenters

Consolidate Servers



Deploy Containers



Upgrade Equipment

Build New Datacenters

Lease Colocation

Move to the Cloud

Improve Efficiency

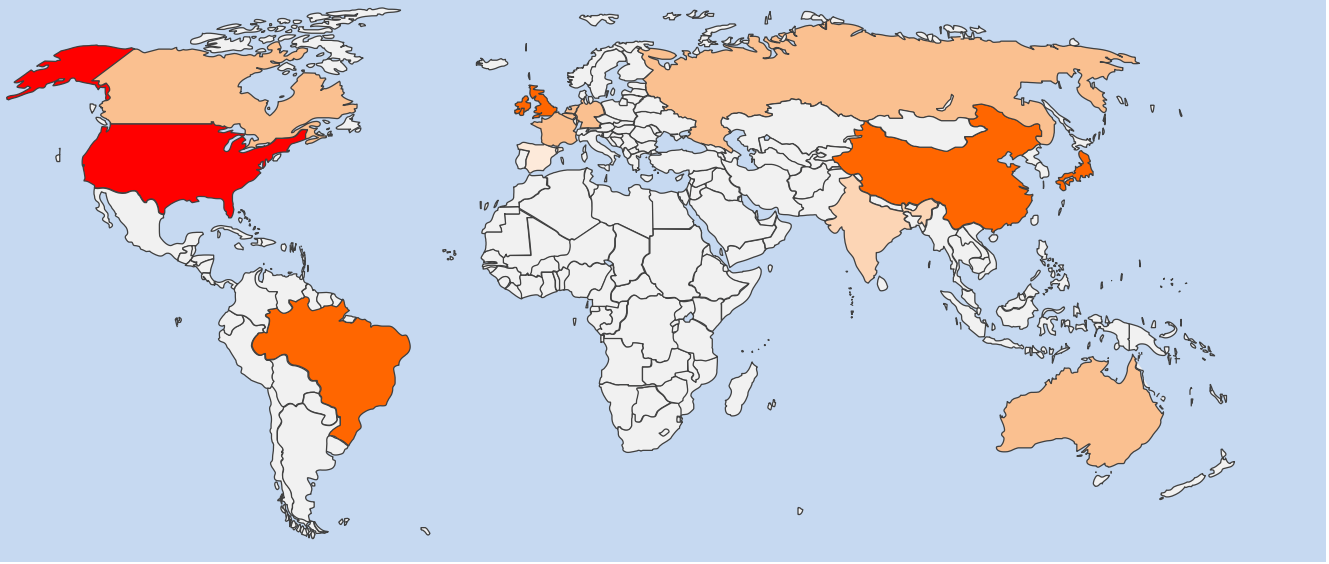
Facility Construction

Third-Party Solutions

Energy and Environmental Problems

8 TWh	USA
3 TWh	China
2 TWh	U.K.
2 TWh	Japan
2 TWh	Brazil
1 TWh	France
1 TWh	Benelux
1 TWh	Canada
1 TWh	Germany
1 TWh	Russia
1 TWh	Australia
1 TWh	India

The increase in server energy demand (2012 - 2013) [2]



- **Server energy consumption:**
  - 1.8% of global electricity usage
  - Might triple within 8 years [1]

300 ~ 400 TWh in 2012

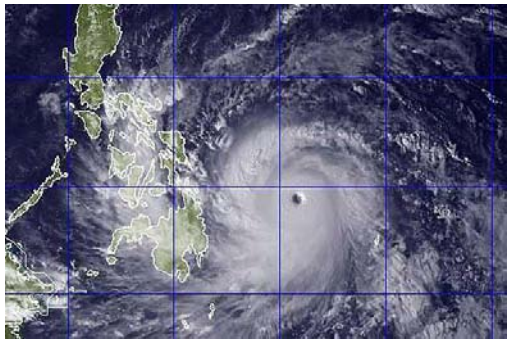
1000 ~ 1400 TWh in 2020

[1] C. Belady, Projecting Annual New Datacenter Construction Market Size, Global Foundation Services, 2011

[2] DCD Industry Census 2012: Energy, http://www.dcd-intelligence.com/

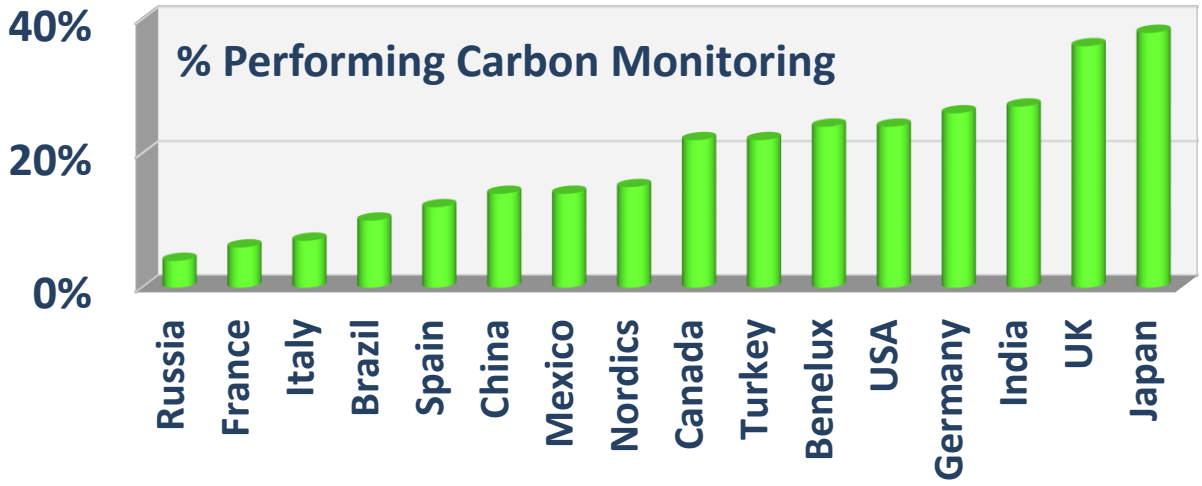
Energy and Environmental Problems

Hurricane Sandy, 2012
(Northeastern US)

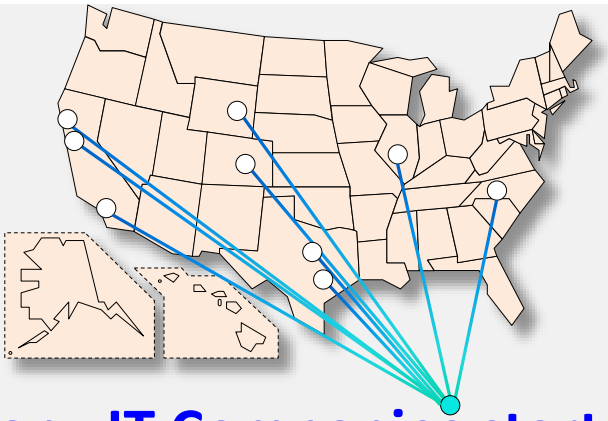


Typhoon Haiyan, 2013
(Southeast Asia)

- **The greenhouse effect and climate change**
- **1MW data center → 10~15 Kt CO_2 yearly**
- **Datacenters are carbon-constrained:**
  - Must cap carbon emissions



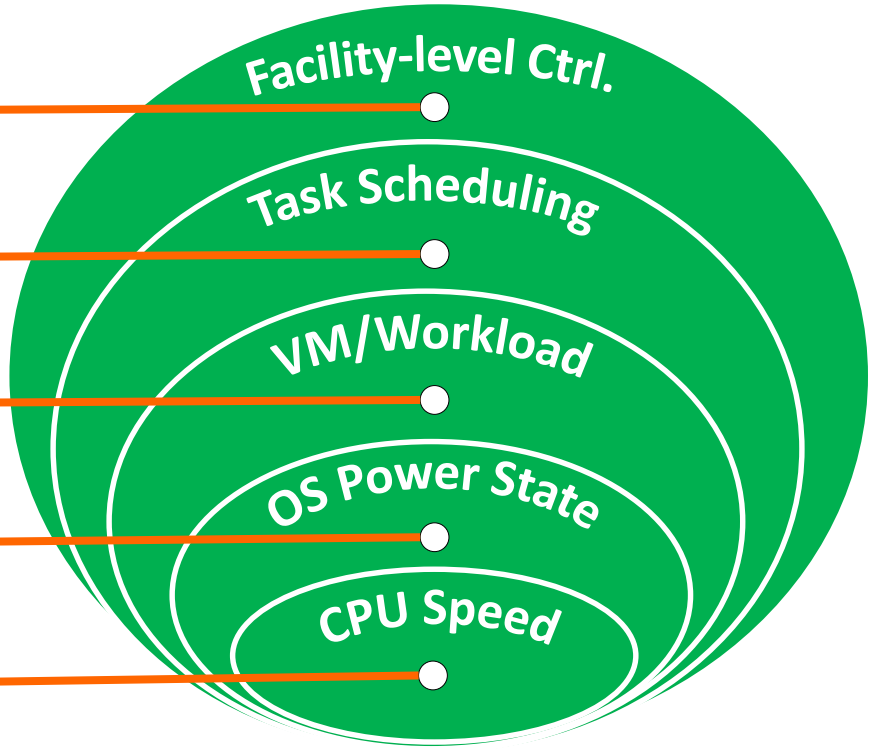
Renewable Energy Powered Systems



Many IT Companies start to integrate non-conventional clean energy solutions

Green Computing - Related Work

- **Mainly focus on managing solar/wind**
  - Supply/Load co-scheduling **[ASPLOS'13, HPCA'13]**
  - Supply-aware job scheduling **[Eurosys'12]**
  - Supply-driven load migration **[ISCA'12]**
  - Avoid shedding critical load **[ASPLOS'11]**
  - Optimal power allocation **[HPCA'11]**

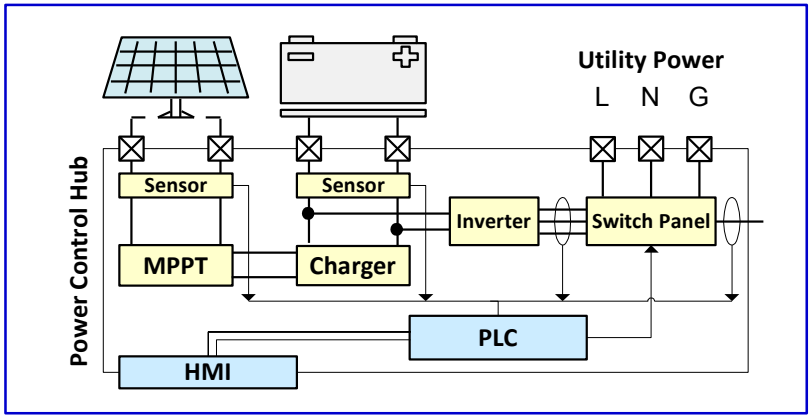


- **We explore carbon-conscious capacity expansion schemes**
  - Scalable, sustainable, and economical power provisioning

Talk Overview

1. Background and Motivation

2. Oasis: Design and Prototype

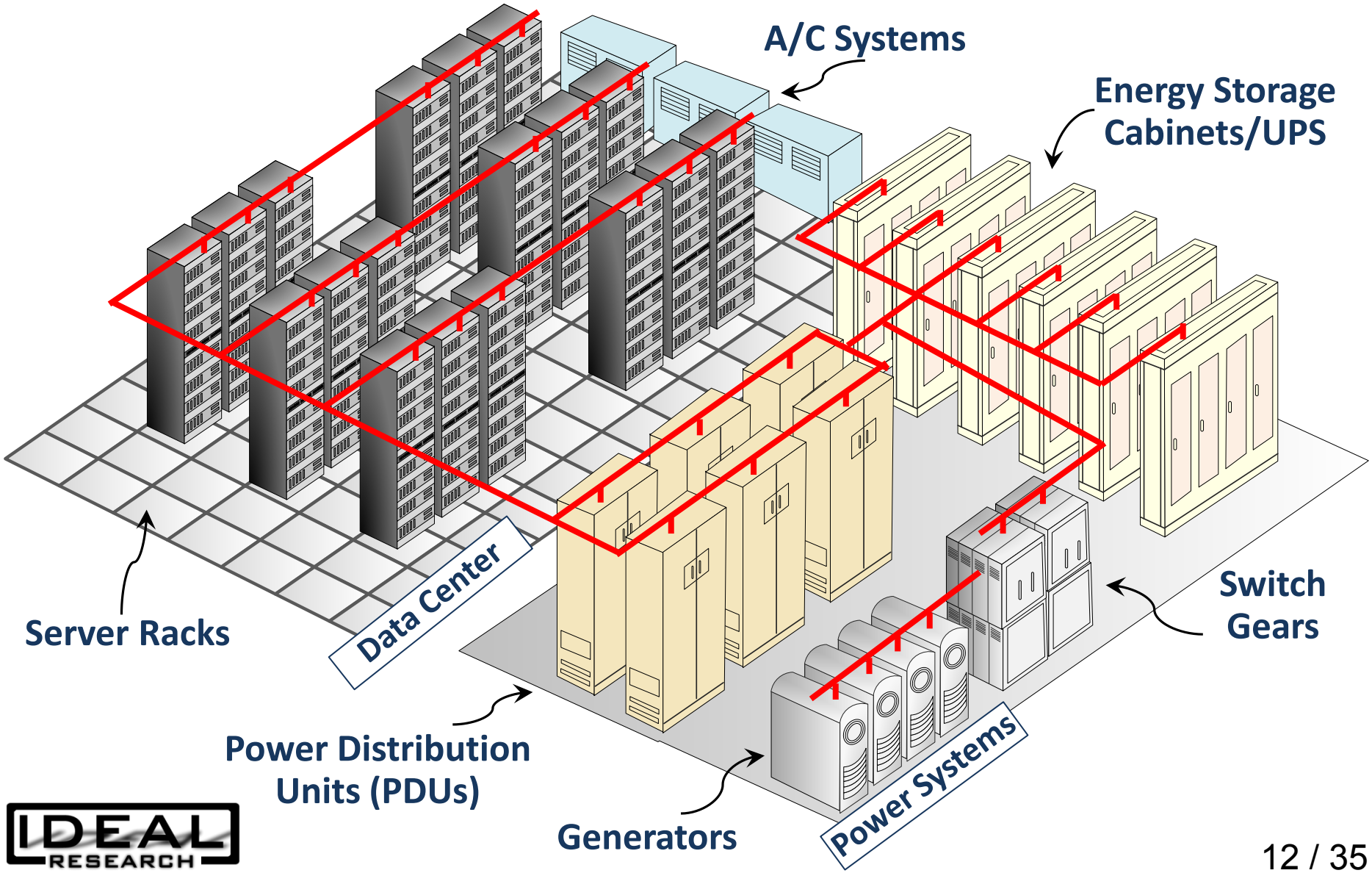


3. Optimized Oasis Operation

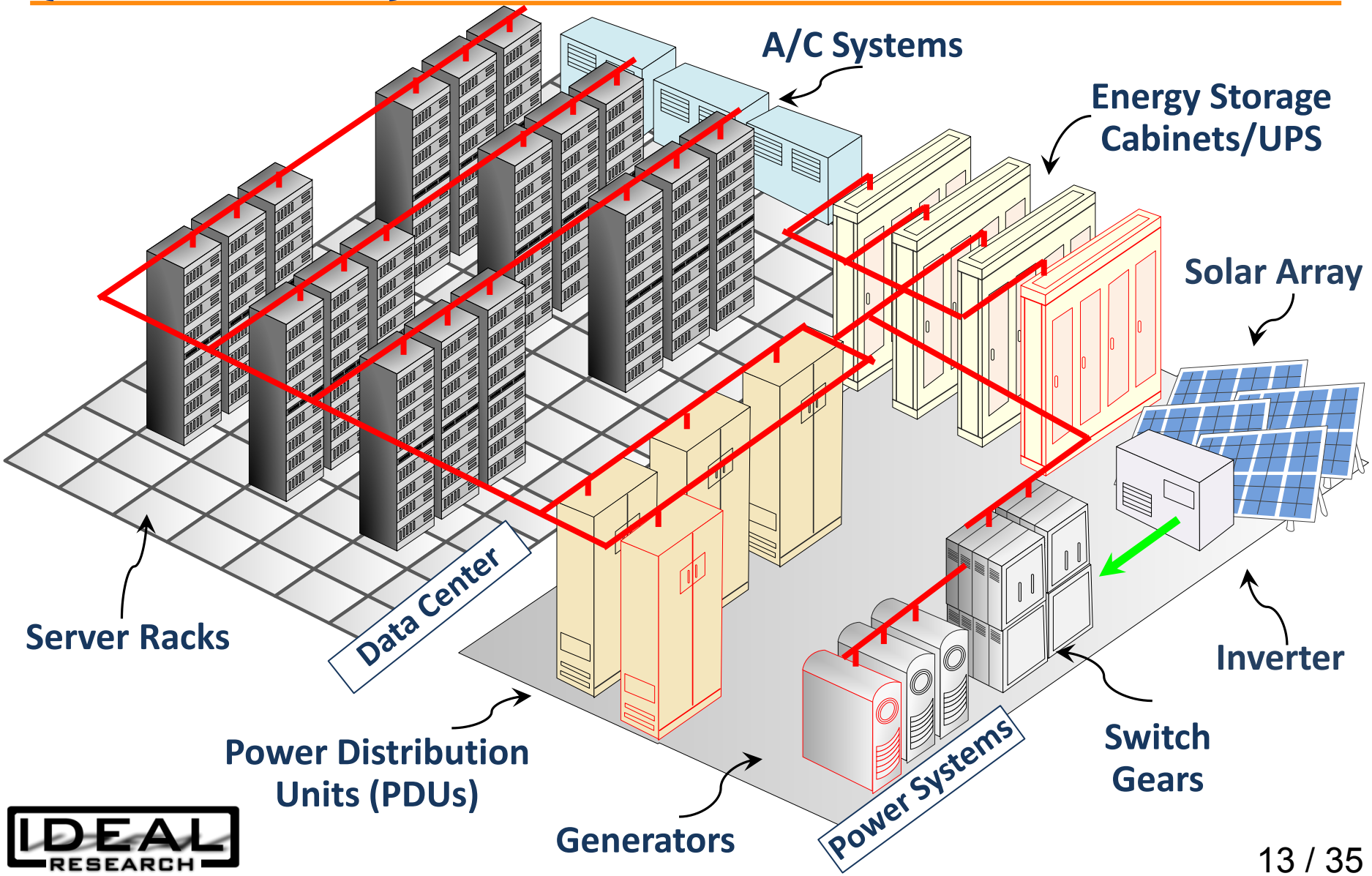
$P_{Load} > P_{Renewable}$				
1	R	DB > 0	upsC > 60%	Switch = 'N', cpuFreq = 'Highest'
2			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
3		DB = 0	upsC > 60%	Switch = 'N', cpuFreq = 'Low'
4			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
5	U	DB > 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Highest'
6			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'
7		DB = 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Lowest'
8			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'

4. Evaluation and Discussion

Utility Power Over-Provisioning (Conventional)



Centralized Power Capacity Expansion (Conventional)



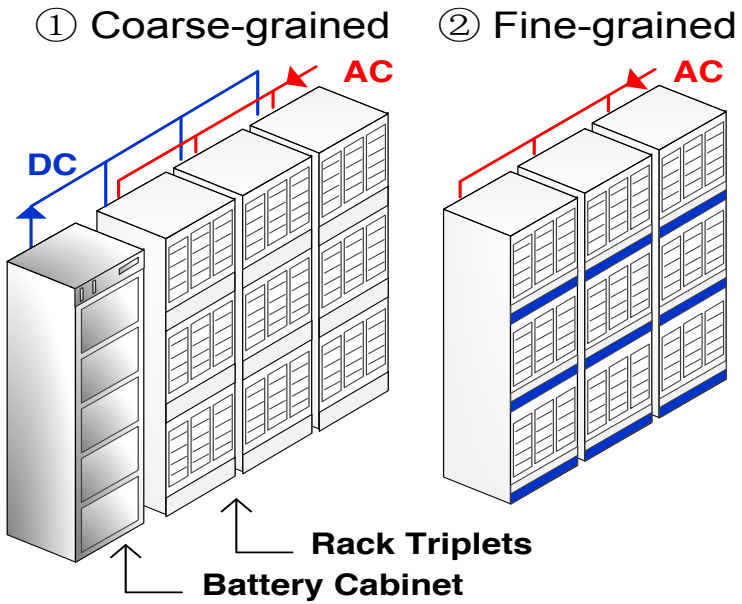
Scale-Out Models

Models \ Metrics	Carbon Emission	Capacity Scalability	Cost of Utility Power	Cost of Green Power
Utility Over-Provisioning	Poor	Poor	High	N/A
Centralized Expansion	Good	Poor	Reduced	High
Ideal Power Provisioning	Good	Good	Reduced	Reduced

- **Oasis: green energy solutions + pay-as-you-grow model**
  - Adds green power budget directly to server racks
  - Gradually increases green power capacity

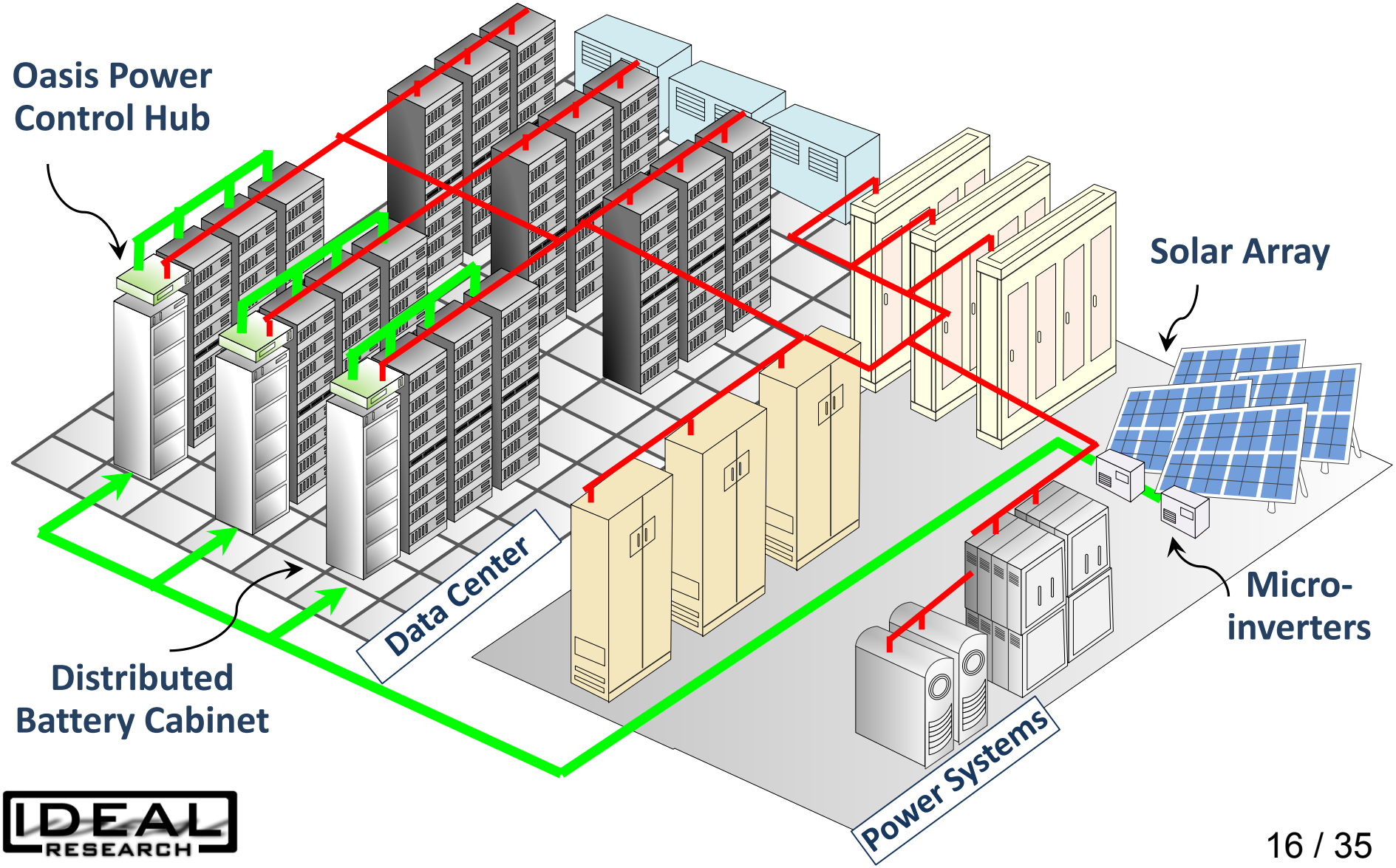
We Leverage Modular Power Sources

- **Distributed Battery System**

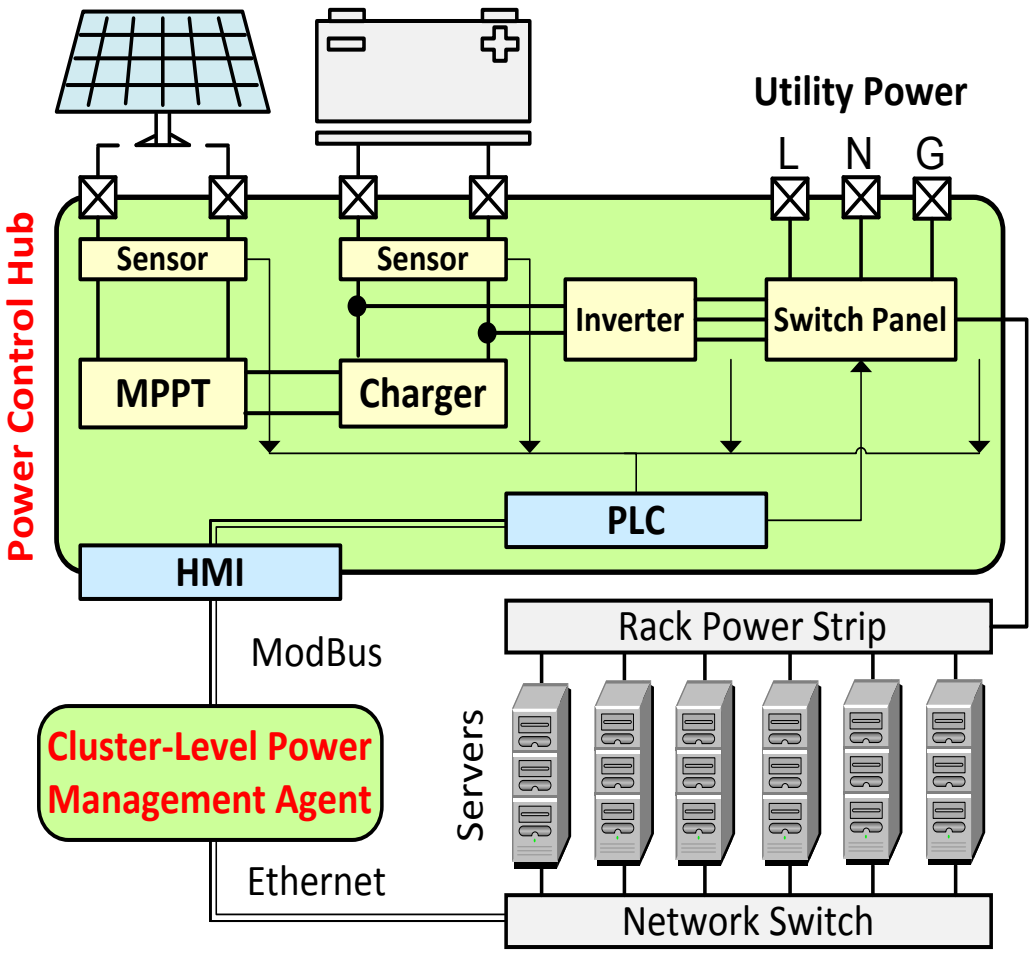


- **Solar Module with Micro-inverters**

Distributed Incremental Integration (Architecture of Oasis)

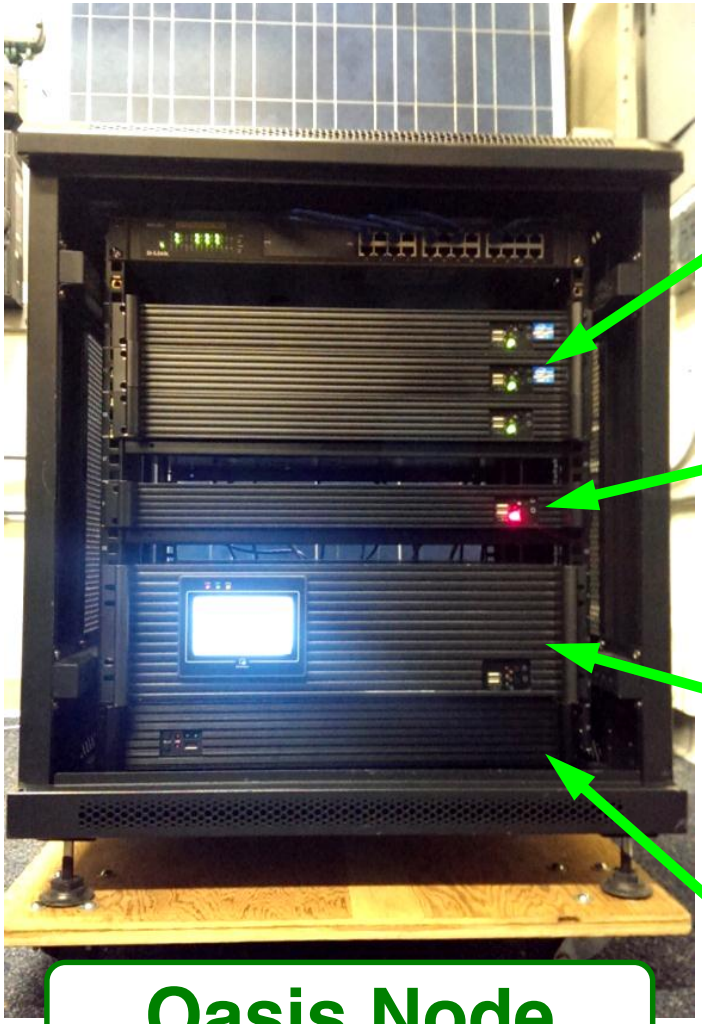


Oasis Implementation: An Overview



- **Power Ctrl. Hub**
  - PLC
    - Manages sensors and switchgears
  - HMI
    - Communication gateway of Oasis
- **Power Mgmt. Agent**
  - Send/Receive power management signals
  - Coordinates power supply and server load

Oasis Implementation: An Overview



Oasis Node

Server Nodes

- Intel i-7 based micro server

Power Mgmt. Agent

- AMD low power node (25W)

Power Ctrl. Hub

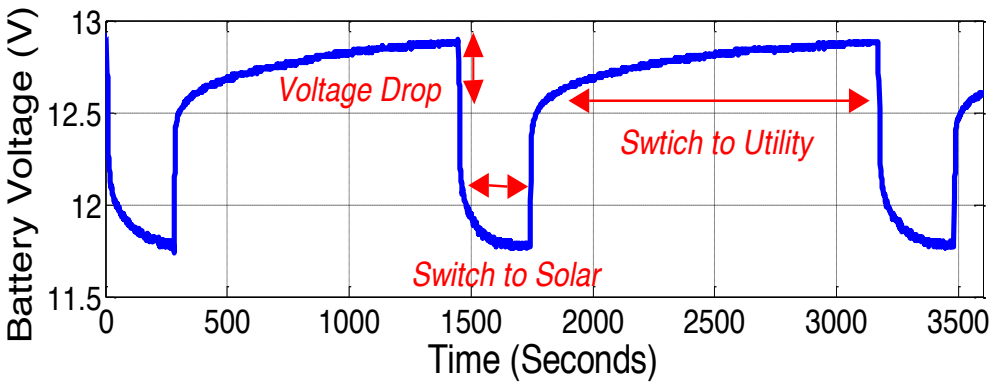
- PLC + HMI + Inverter + …

Battery Chassis

- Nine 2 Ah lead acid batteries

Hybrid Power Supply Scheme

Roof-mounted solar panels in our lab

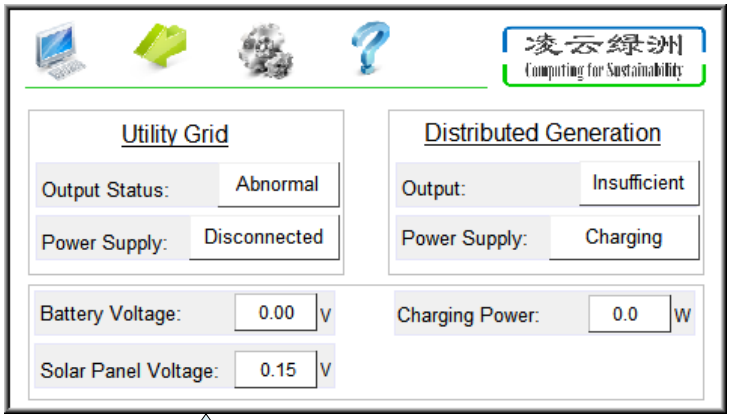


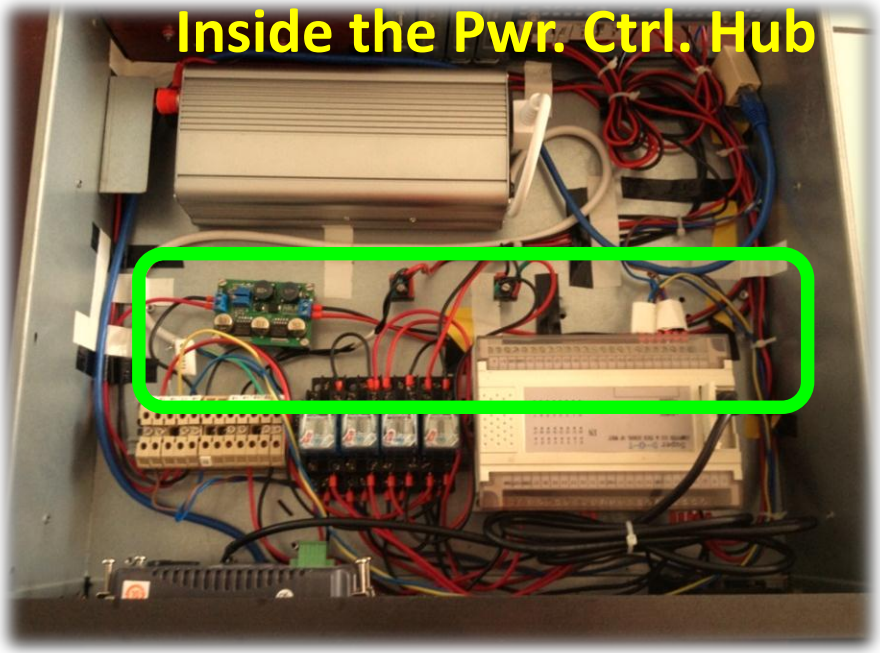
Battery charging and discharging scenarios

- **Stored solar energy**
  - Release solar energy when batteries are fully charged
  - Charge batteries with solar power when the SOC is low
- **Utility power supply**
  - The primary energy source in cloudy days or at night

Power Control Hub - I

- **Monitors power supply status**
  - Emergency alert
  - Battery capacity check
  - Health status assessment



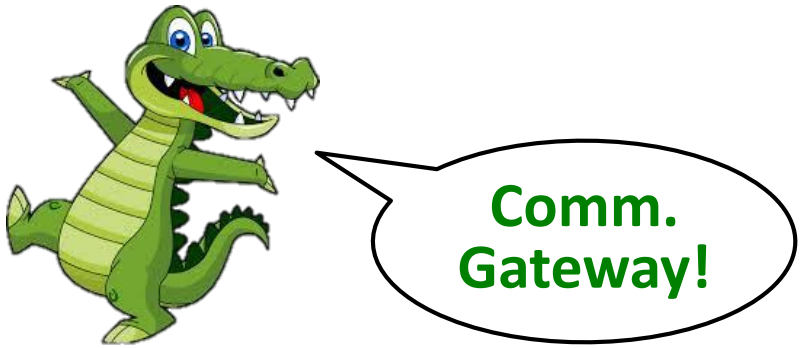


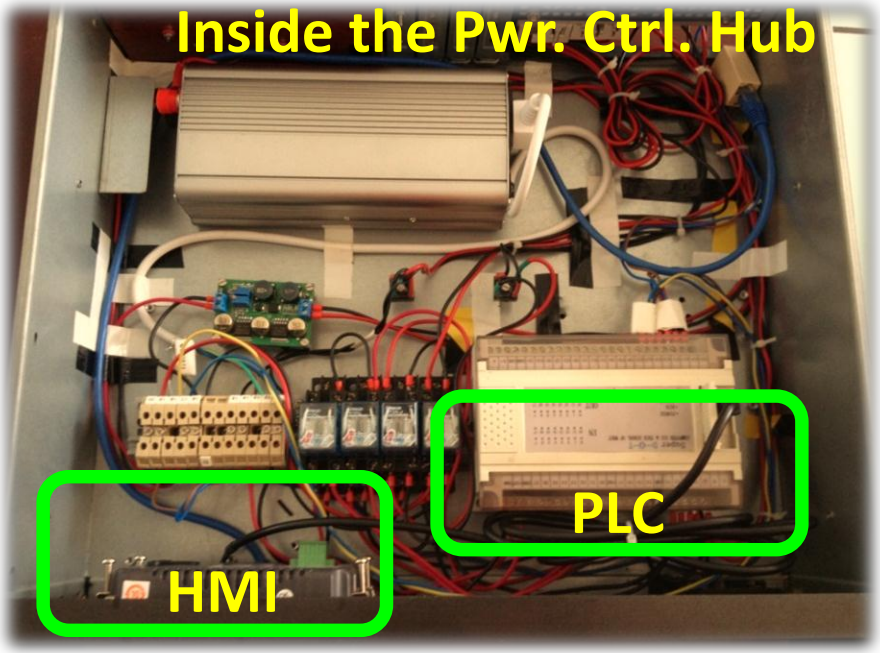
to HMI touch screen display

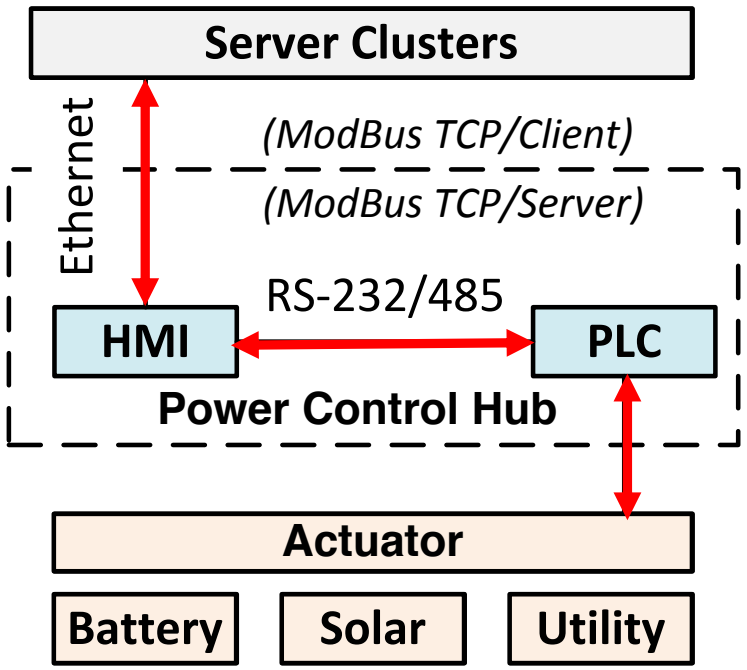
to the power mgmt. agent

Power Control Hub - II

- **Bridges power supply and load**
  - Send/Receive control signals
  - Send/Store monitored data

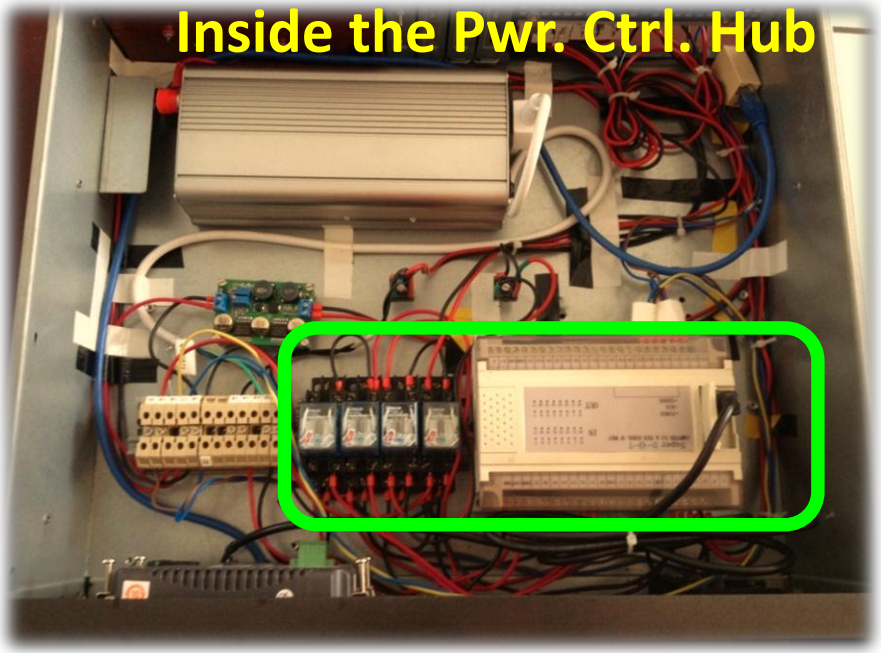






Power Control Hub - III

- **Performs Power Supply Switch**
  - Switch between solar power and utility power
  - Leverage high-voltage relay array controlled by a PLC

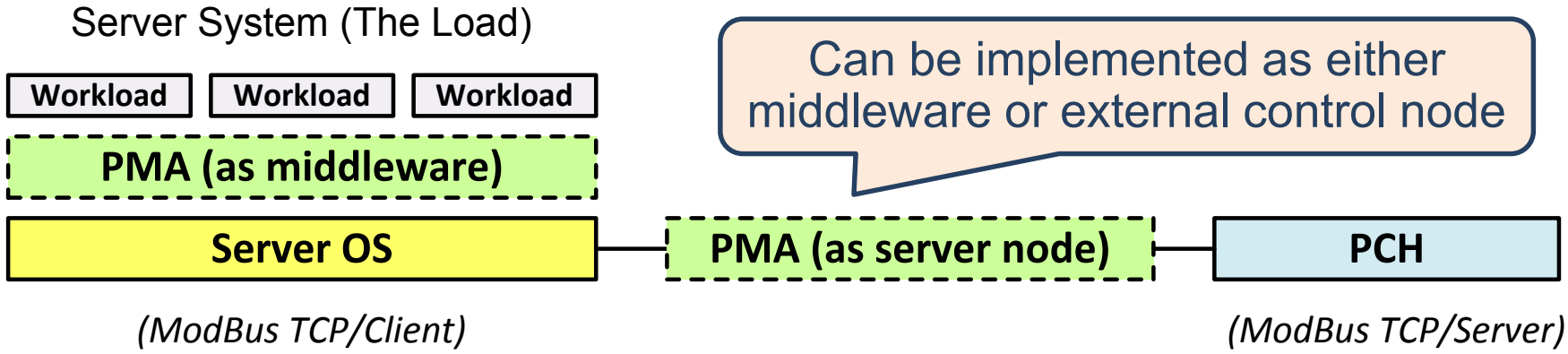


Autonomous Mode

Two Switching Modes!

Coordinated Mode

Power Management Agent (PMA)



- **Adaptive power source switching**
  - Manages utility power usage (affect carbon footprint)
  - Manages solar energy and battery usage
- **Supply-aware server load tuning**
  - Dynamic voltage and frequency scaling (DVFS)
  - Trigger VM migration/checkpointing if necessary

Talk Overview

1. Background and Motivation

2. Oasis: Design and Prototype

3. Optimized Oasis Operation

$P_{Load} > P_{Renewable}$				
1	R	DB > 0	upsC > 60%	Switch = 'N', cpuFreq = 'Highest'
2			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
3		DB = 0	upsC > 60%	Switch = 'N', cpuFreq = 'Low'
4			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
5	U	DB > 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Highest'
6			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'
7		DB = 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Lowest'
8			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'

4. Evaluation and Discussion

Ozone: Optimized Oasis Operation (O_3)

Server load

- Adjust demand level

Power supply

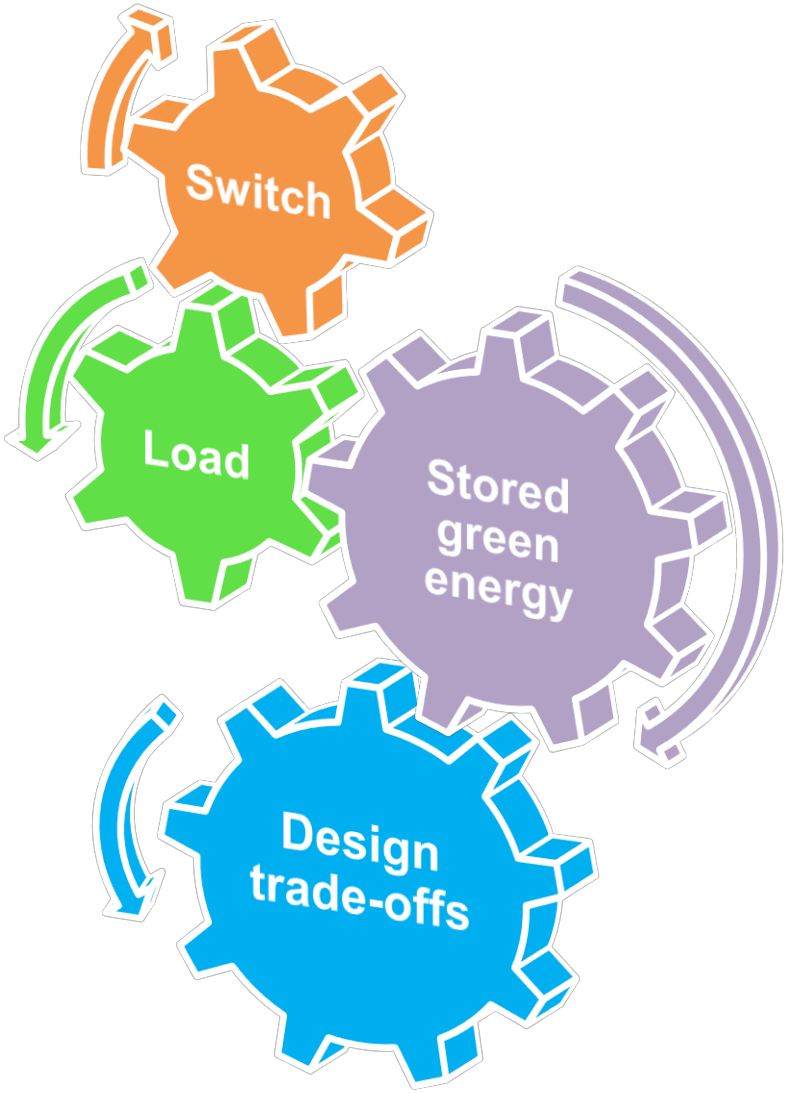
- Switch power source

Battery systems

- Stored green energy usage

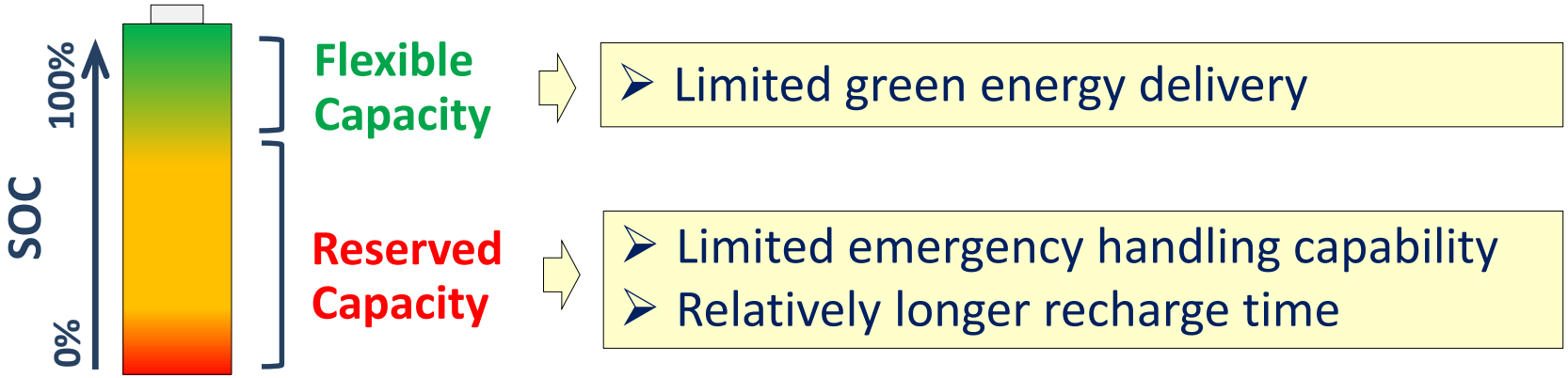
Design trade-offs

- Performance acceptable?
- Carbon-footprint satisfied?
- Battery cost reduced?
- Availability guaranteed?



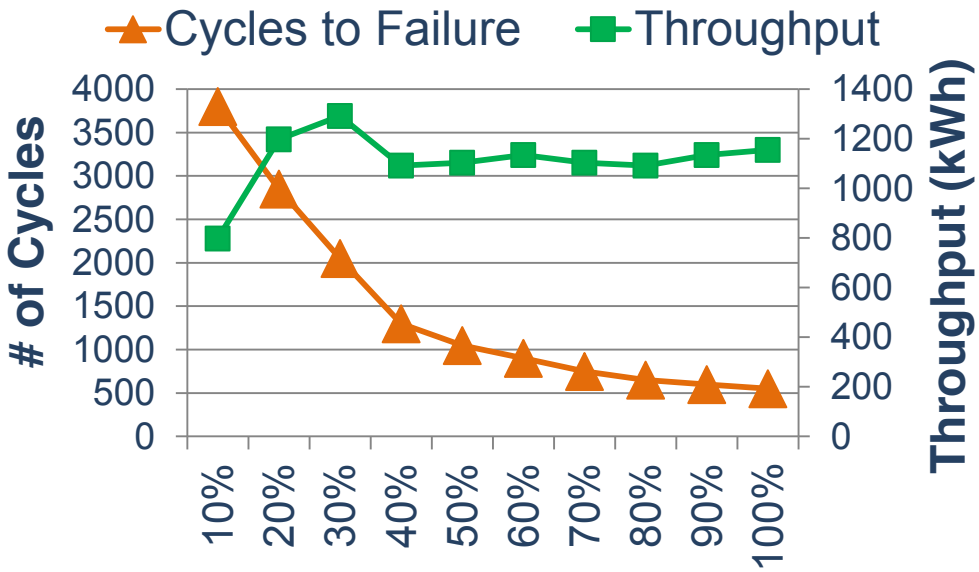
Backup Capacity

- **Capping green energy usage for each discharge cycle**
  - The stored green energy level affects backup time
  - Should avoid low state of charge (SOC)
- **Use different power management schemes at different SOC**
  - Abundant stored energy? (60% ~ 100% SOC)
  - Not enough stored energy? (20% ~ 60% SOC)
  - Should avoid low SOC (i.e., SOC < 20%)



Discharge Budget

- **Discharge throughput model**
  - The total energy that can be cycled through a battery is fixed



$$D_{aggregated} = \sum_0^t D_{Ah}$$

$$D_{budget} = T / Lifetime \times D_{rated}$$

Manage solar energy usage based on $D_{budget} - D_{aggregated}$

- **Capping the aggregated discharge throughput**
  - Predicting lifetime based on the remaining throughput
  - Capping battery discharge to avoid over-use

Supply/Load Control of Ozone

- **Coordinating server load and power supply switch**
  - Based on the capacity level of stored green energy
  - Based on the aggregated stored green energy usage

	Discharge Budget > 0	Discharge Budget = 0
Flexible Capacity > 0	Give Priority to Releasing Stored Solar Energy (Use DVFS if necessary)	Switch to Utility
Flexible Capacity = 0	Give Priority to Server Power Capping (Use battery if necessary)	Switch to Utility

Talk Overview

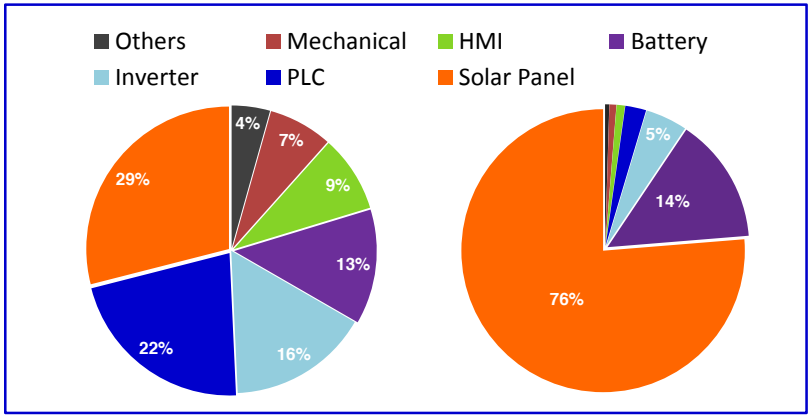
1. Background and Motivation

2. Oasis: Design and Prototype

3. Optimized Oasis Operation

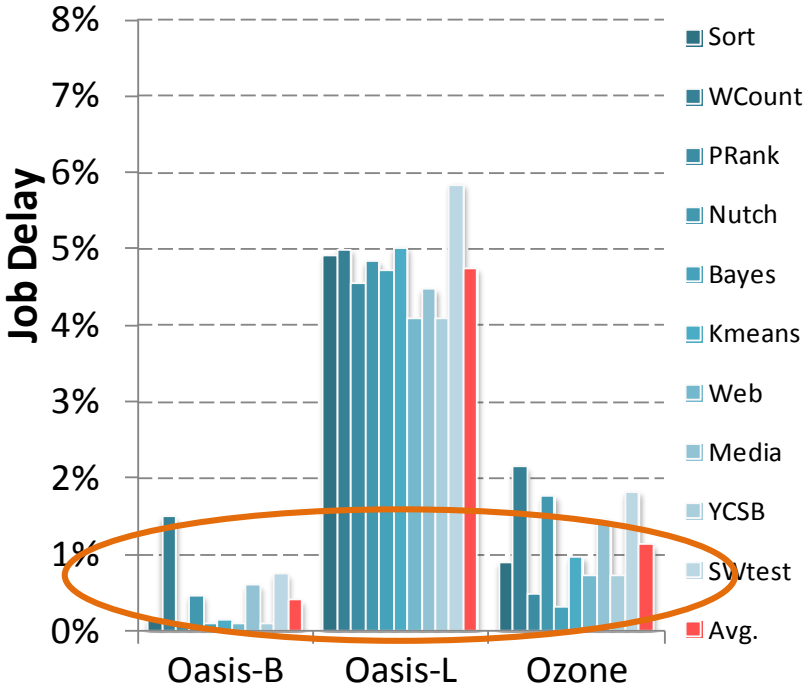
$P_{Load} > P_{Renewable}$				
1	R	DB > 0	upsC > 60%	Switch = 'N', cpuFreq = 'Highest'
2			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
3		DB = 0	upsC > 60%	Switch = 'N', cpuFreq = 'Low'
4			upsC < 60%	Switch = 'Y', cpuFreq = 'TBD'
5	U	DB > 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Highest'
6			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'
7		DB = 0	upsC > 80%	Switch = 'Y', cpuFreq = 'Lowest'
8			upsC < 80%	Switch = 'N', cpuFreq = 'TBD'

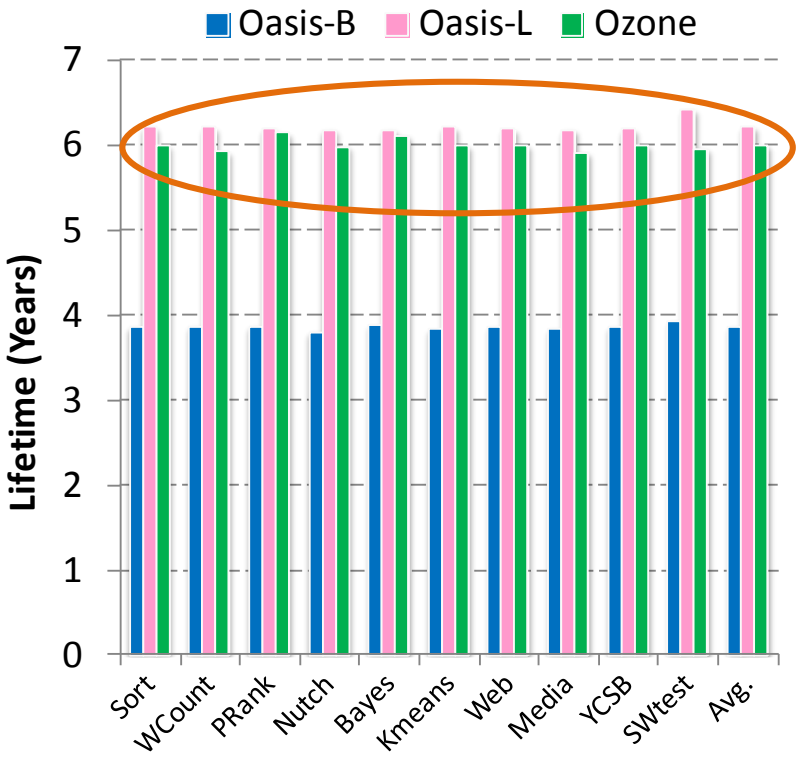
4. Impact of Oasis Design



Job Latency vs. Battery Life

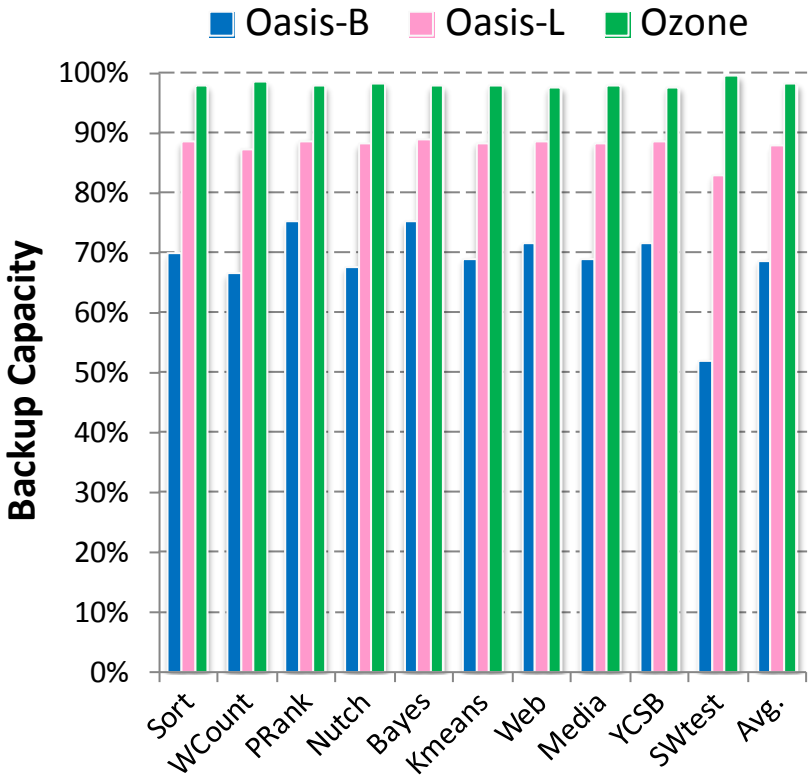
- **Ozone seeks a balance between supply tuning and load tuning**
  - Battery-based design (Oasis-B) emphasis performance
  - Load scaling based design (Oasis-L) emphasis battery lifetime



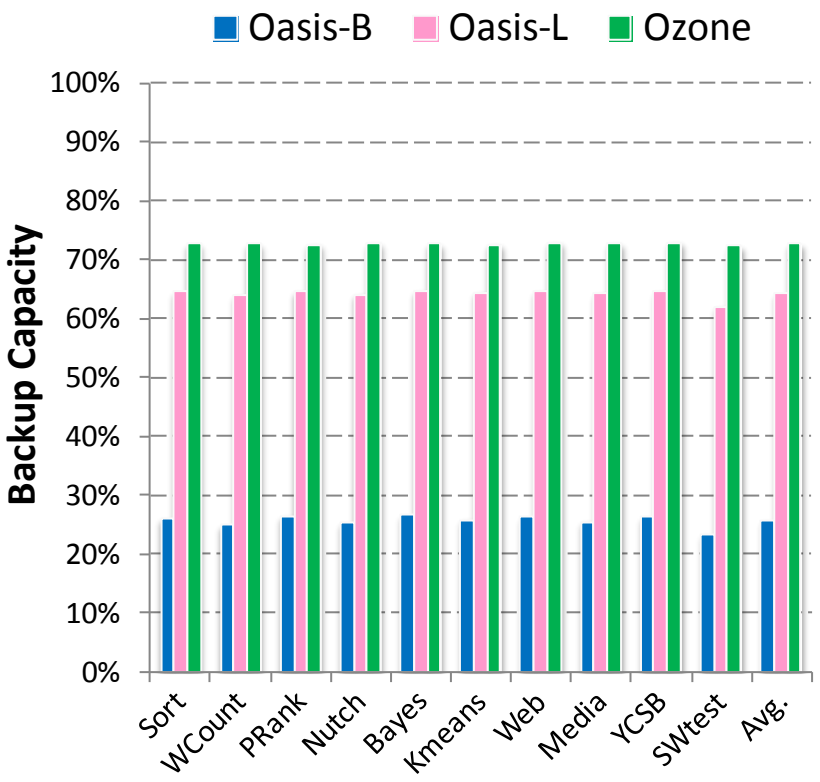


Battery Backup Time

- **Ozone also maintains the best battery backup capacity**
  - Under various renewable power variability



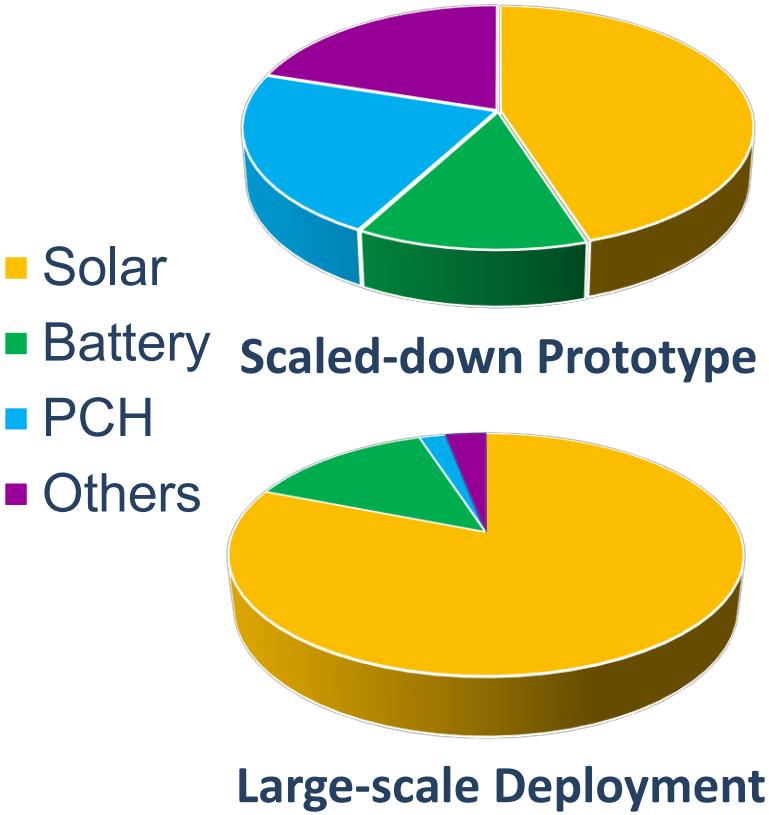
High solar power variability



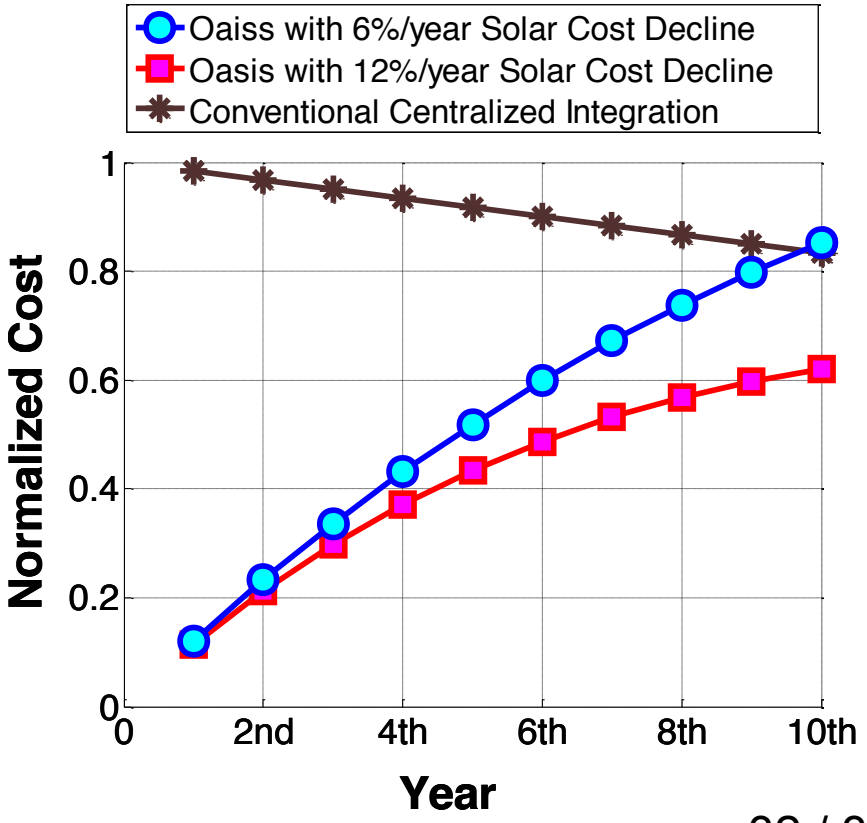
Low solar power variability

Cost Projection

- **Solar systems and batteries are major cost components**
  - PCH: < 4% total cost



- **Oasis could result in 25% less total CapEx**
  - Depending on the hardware cost trend



Conclusions

- Integrating modular green energy sources allows data centers to scale out sustainably
- A distributed, incremental green energy integration method can reduce 25% capital expenditure
- Balancing power supply control and server load control can further improve the design trade-offs
- IT can be the enabler of sustainability: Expanding datacenters using green energy in the big data era!

Welcome!

HPCA-20

Orlando, Florida

February 15-19, 2014

http://hpca20.ece.ufl.edu/

Green Computing