Ph.D. Defense: Resource Allocation Optimization in the Smart Grid and High-performance Computing

Tim Hansen

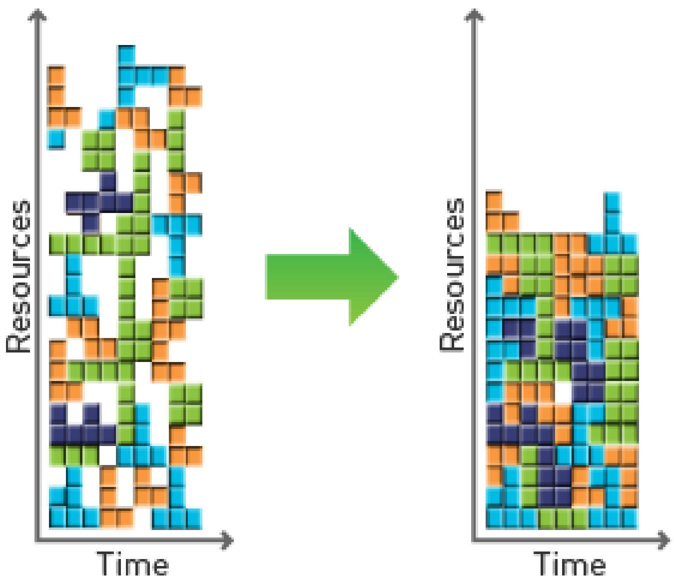
Department of Electrical and Computer Engineering
Colorado State University
Fort Collins, Colorado, USA

- outline
  - introduction to resource allocation
  - resource allocation in Smart Grid
  - resource allocation in high-performance computing
  - conclusions and future directions

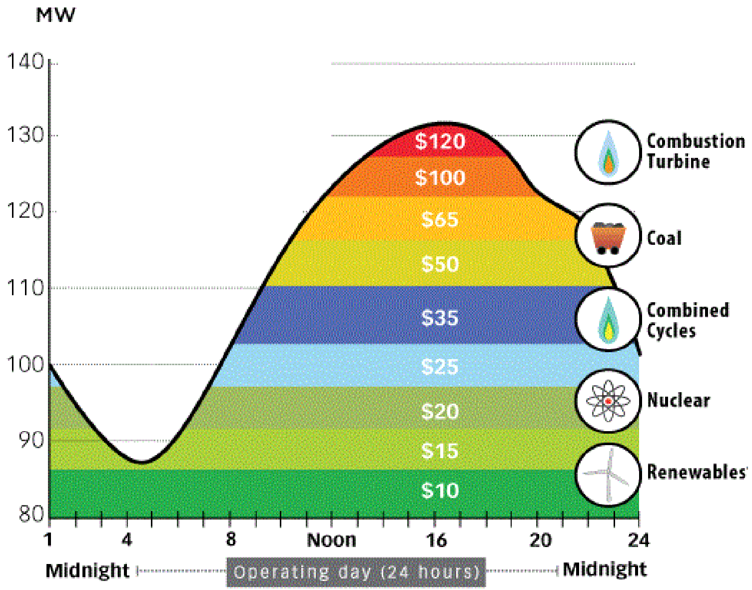
May 8, 2015

Introduction to Resource Allocation

- **resource allocation:** assignment of limited resources to perform useful work
- *optimal* resource allocation problems, in general, are NP-Complete
- in high-performance computing (HPC):
  - allocate HPC resources to parallel applications



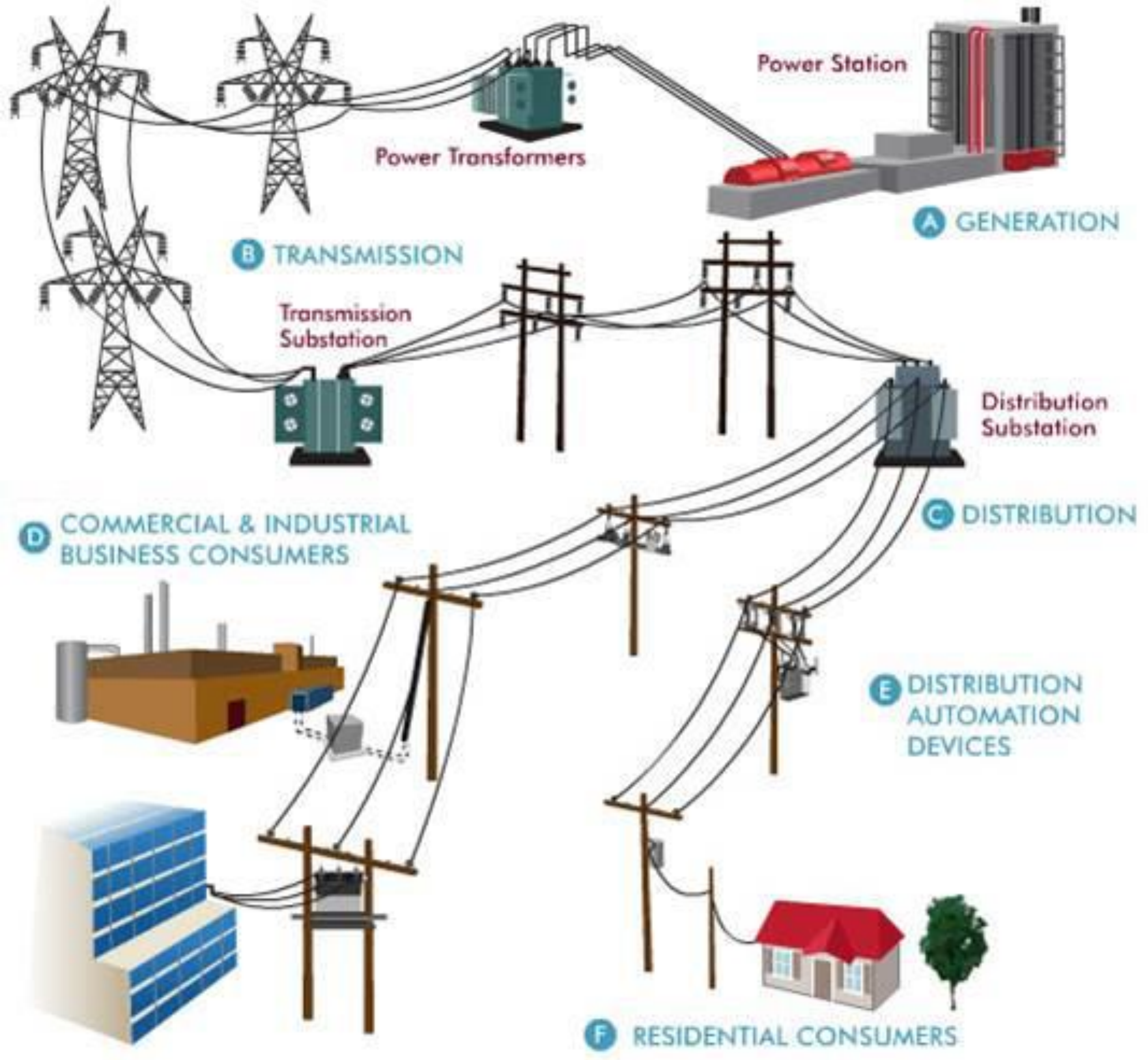
- in electric power systems:
  - allocate generation resources to energy consumers



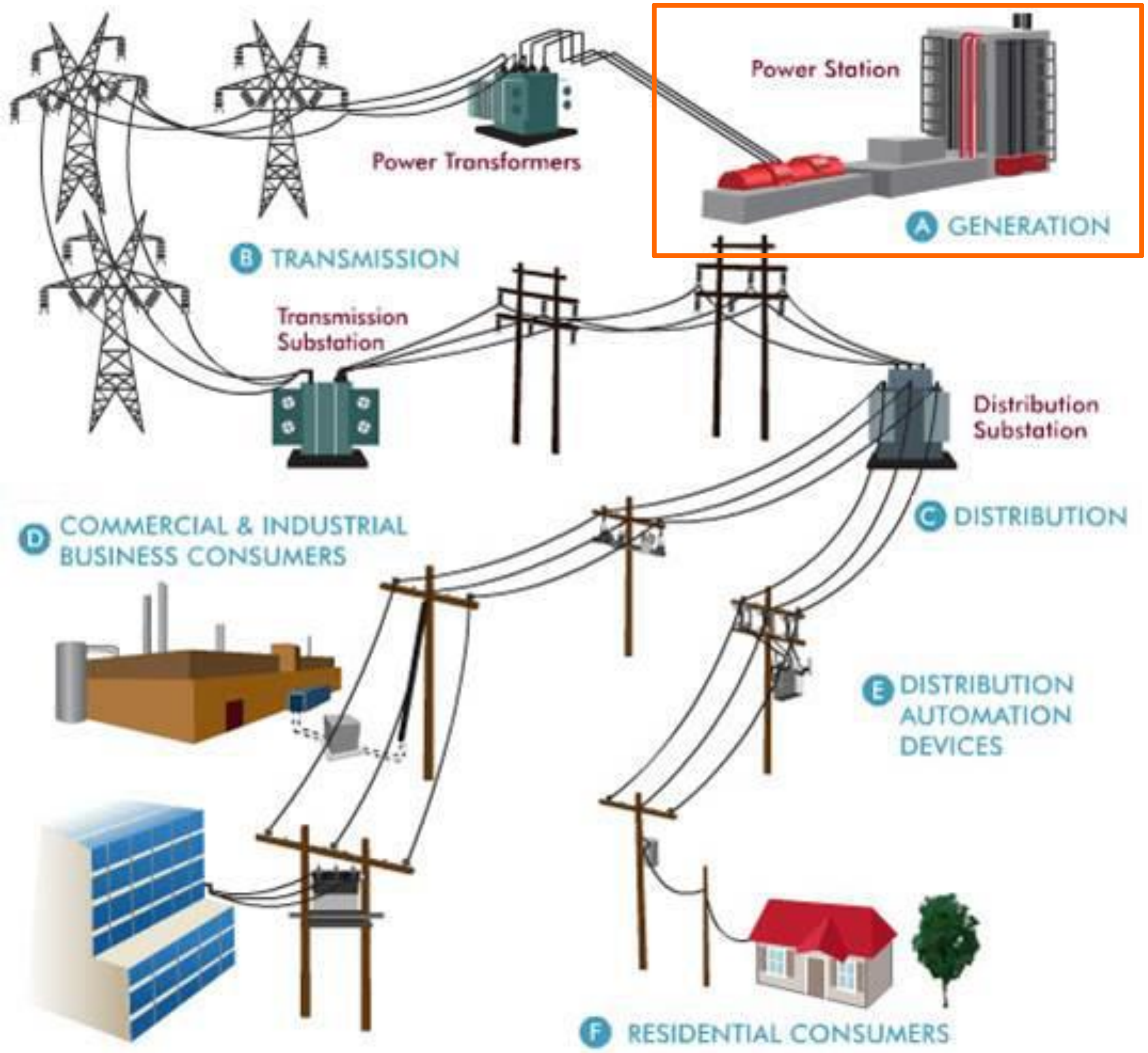
Outline – Resource Allocation in Smart Grid

- introduction to resource allocation
- **resource allocation in Smart Grid**
    - **background and motivation**
    - non-myopic home energy management system
    - aggregator-based residential demand response
    - demand response visualization
    - co-simulation framework
- resource allocation in high-performance computing
- conclusions and future directions

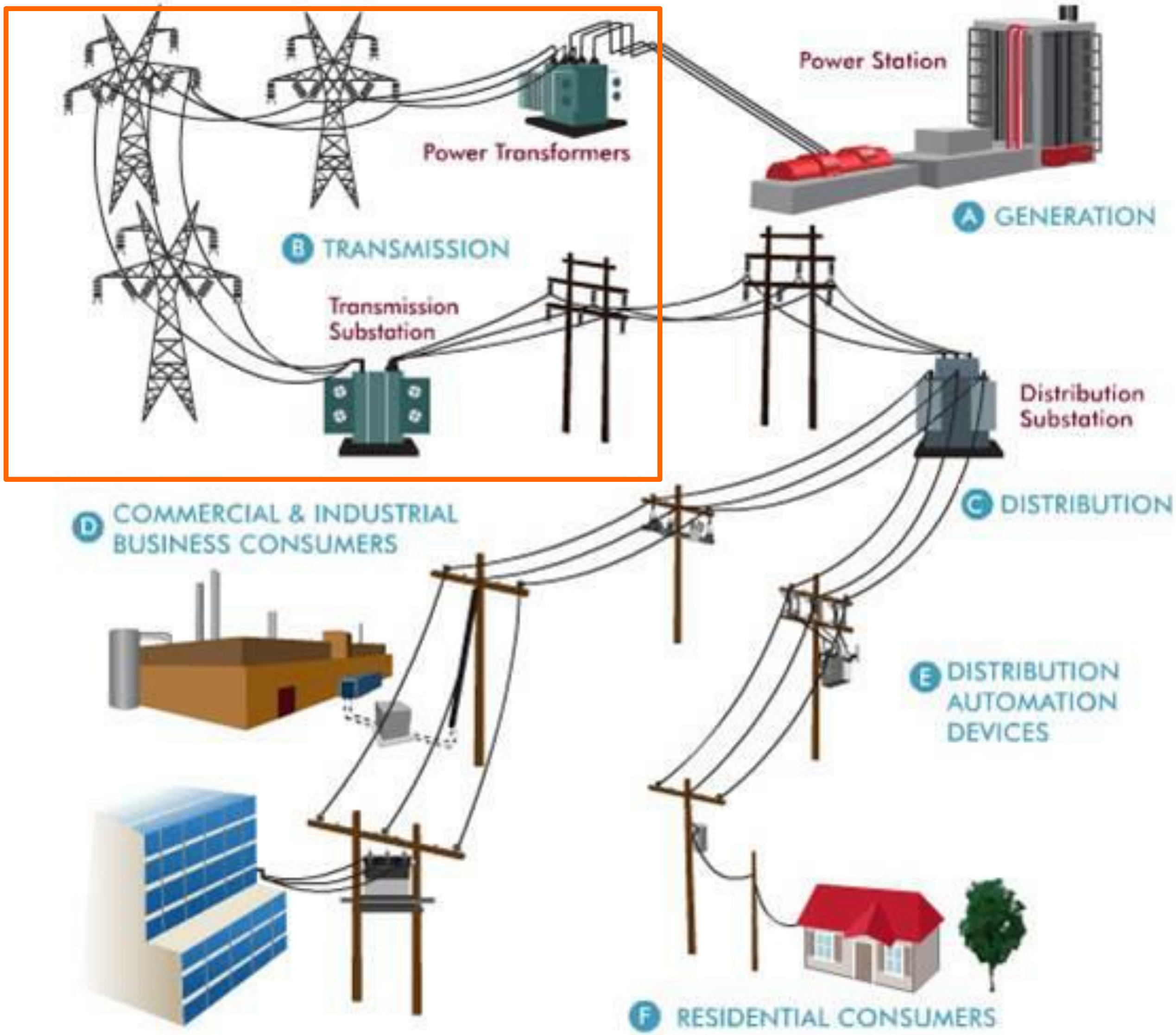
Traditional Bulk Power System



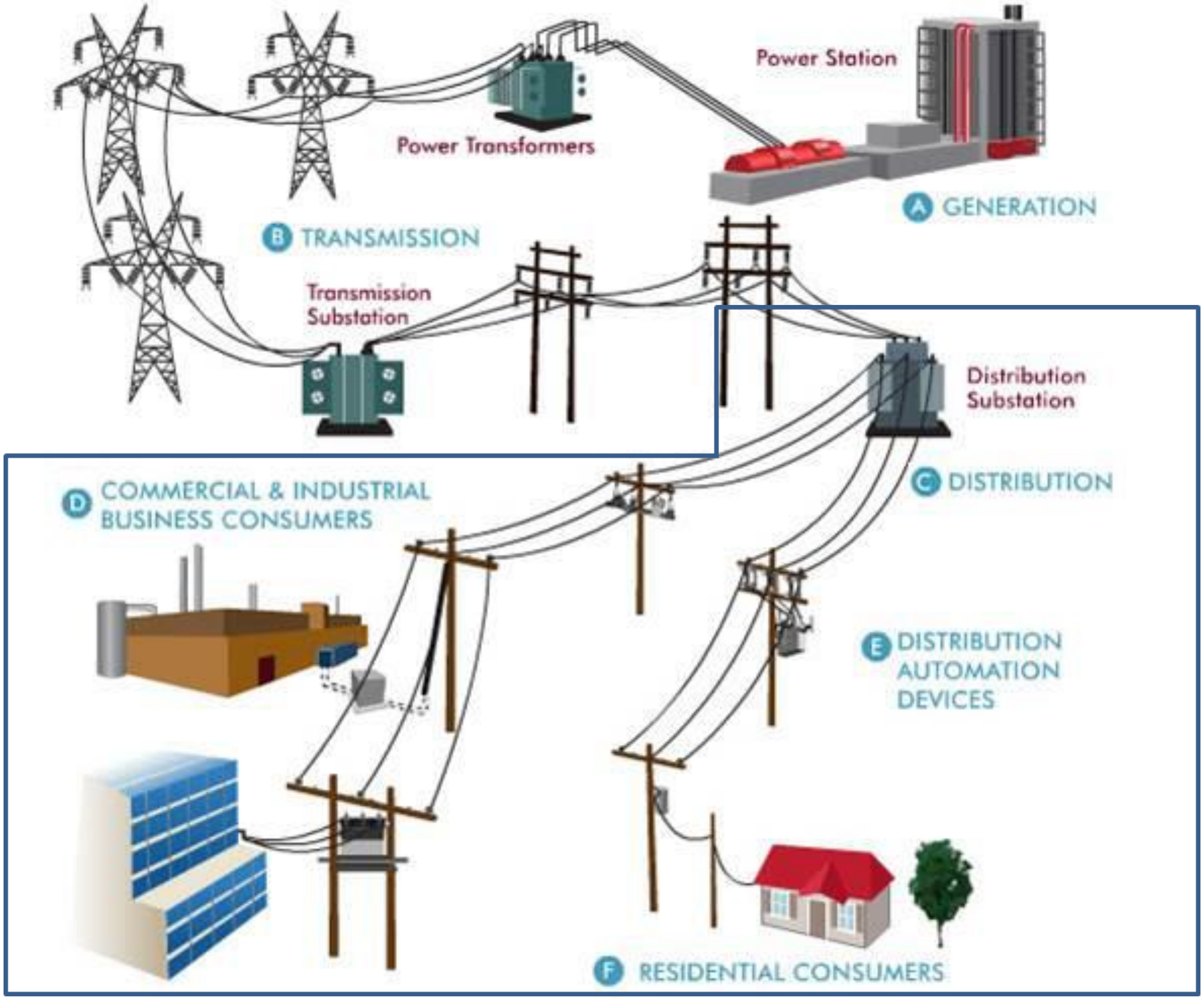
Traditional Bulk Power System



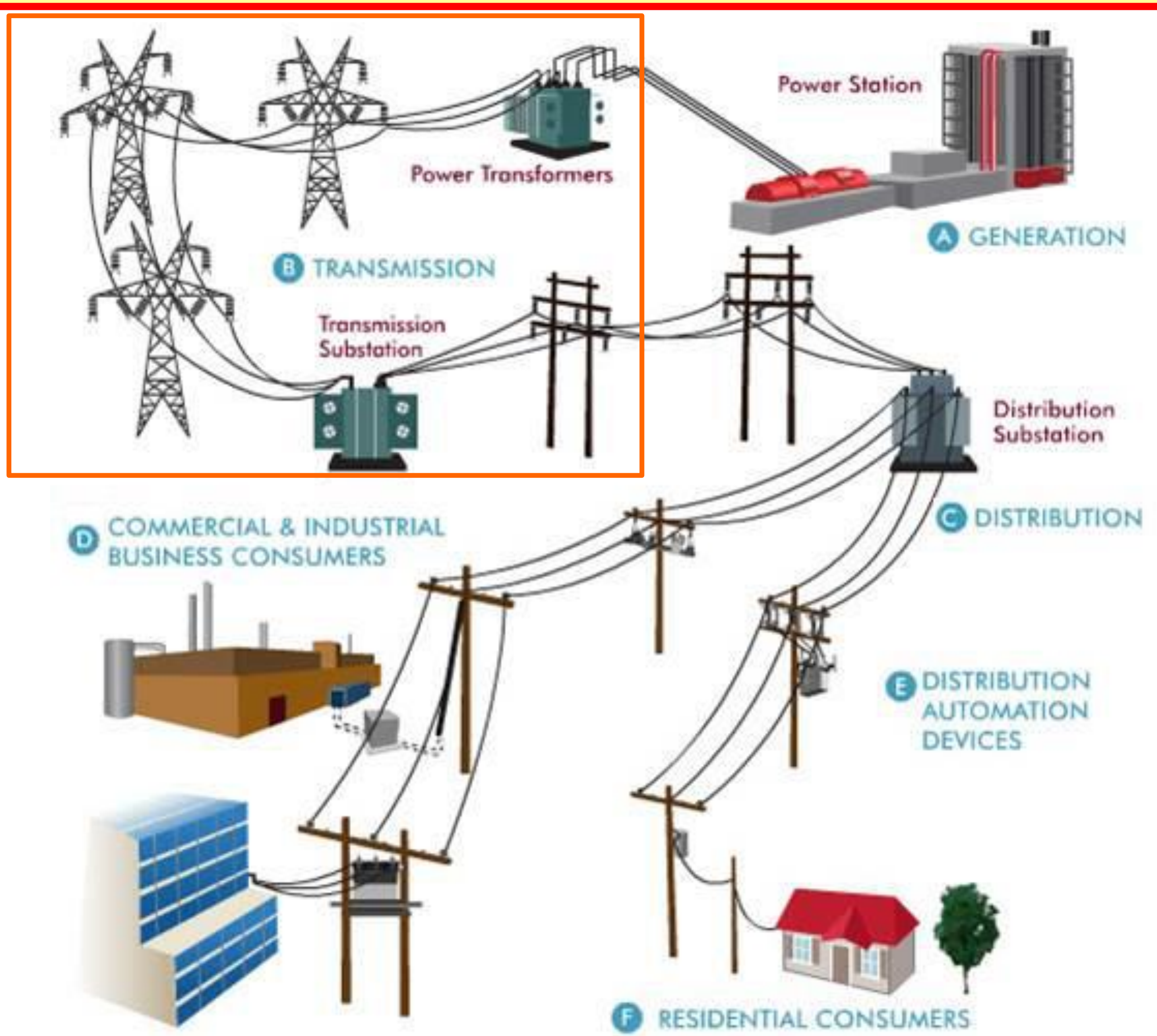
Traditional Bulk Power System



Traditional Bulk Power System

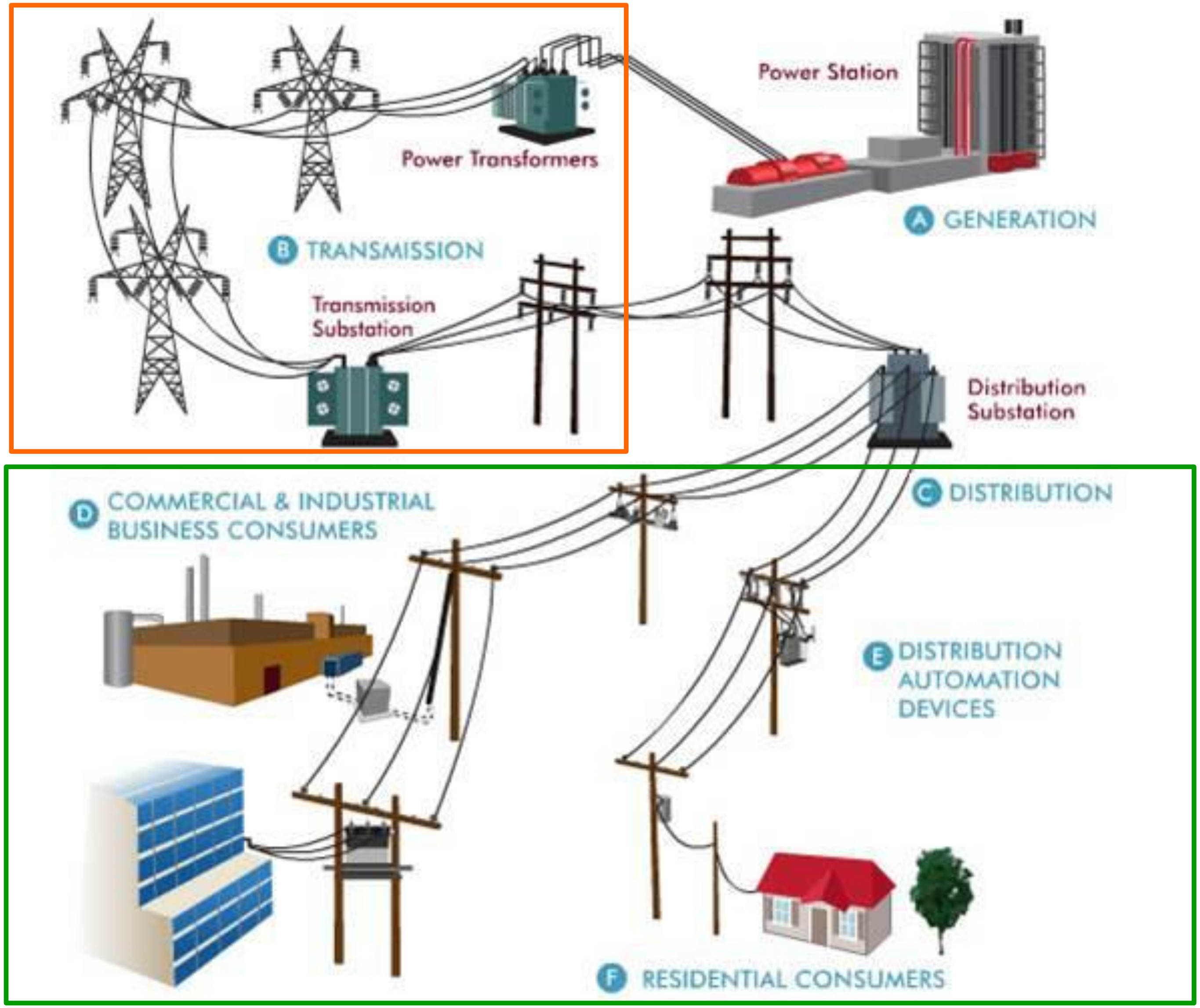


Traditional Bulk Power System



transmission level

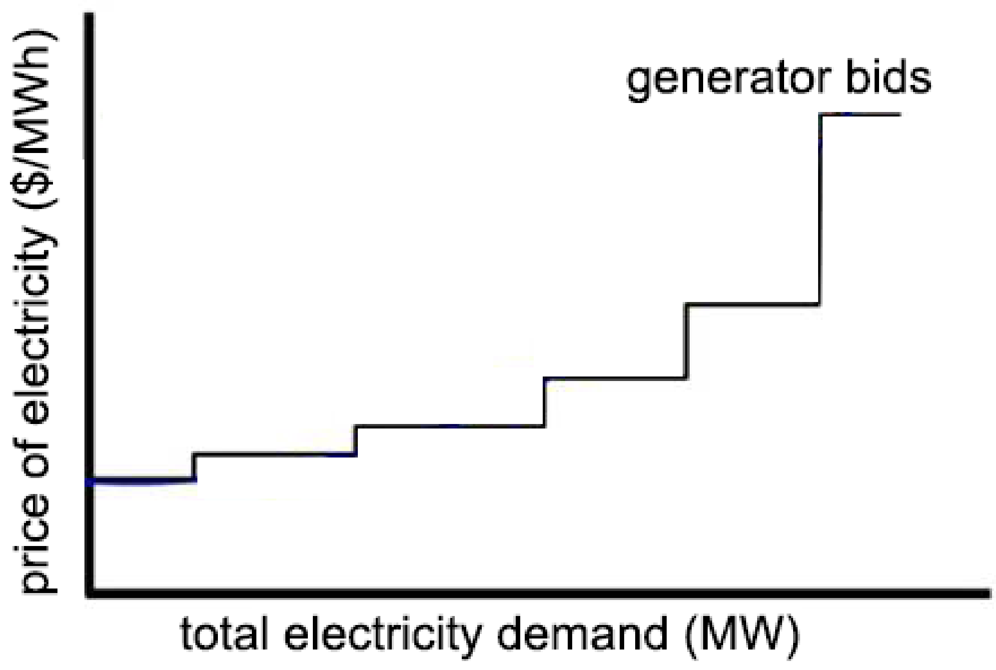
Traditional Bulk Power System



transmission level

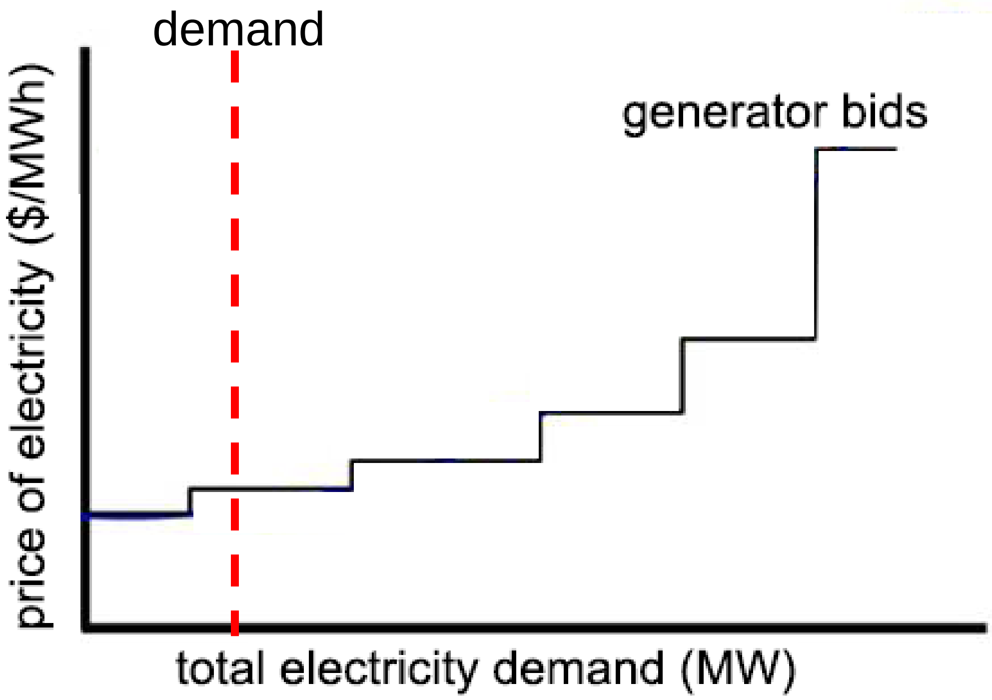
distribution level

Bulk Power Market



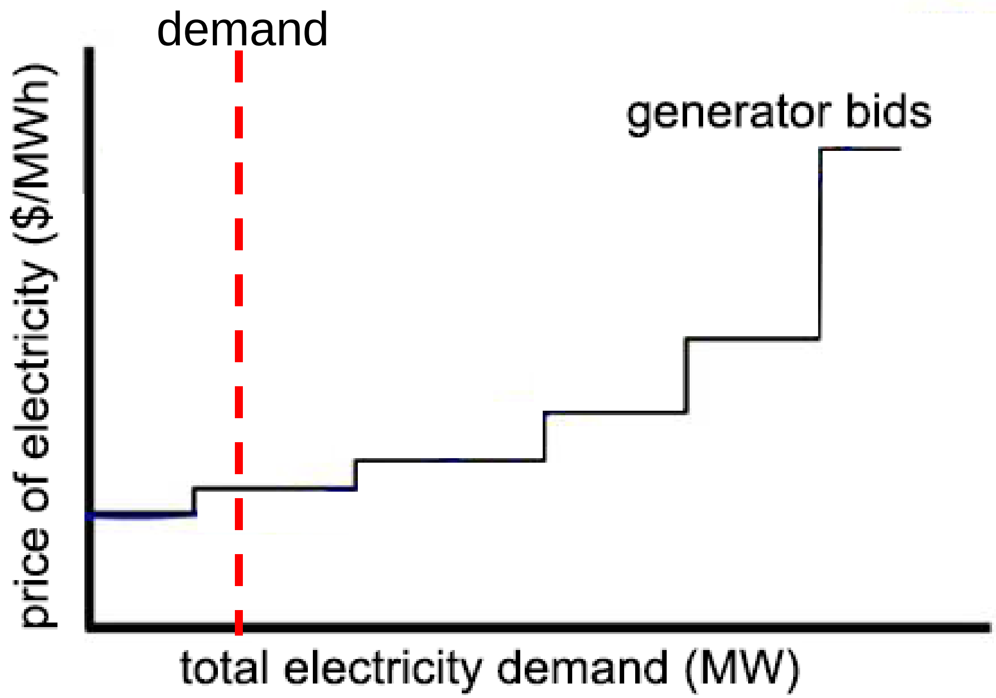
- owners of generators bid into the bulk power market
- as demand increases, more expensive generators are needed to meet the demand

Bulk Power Market



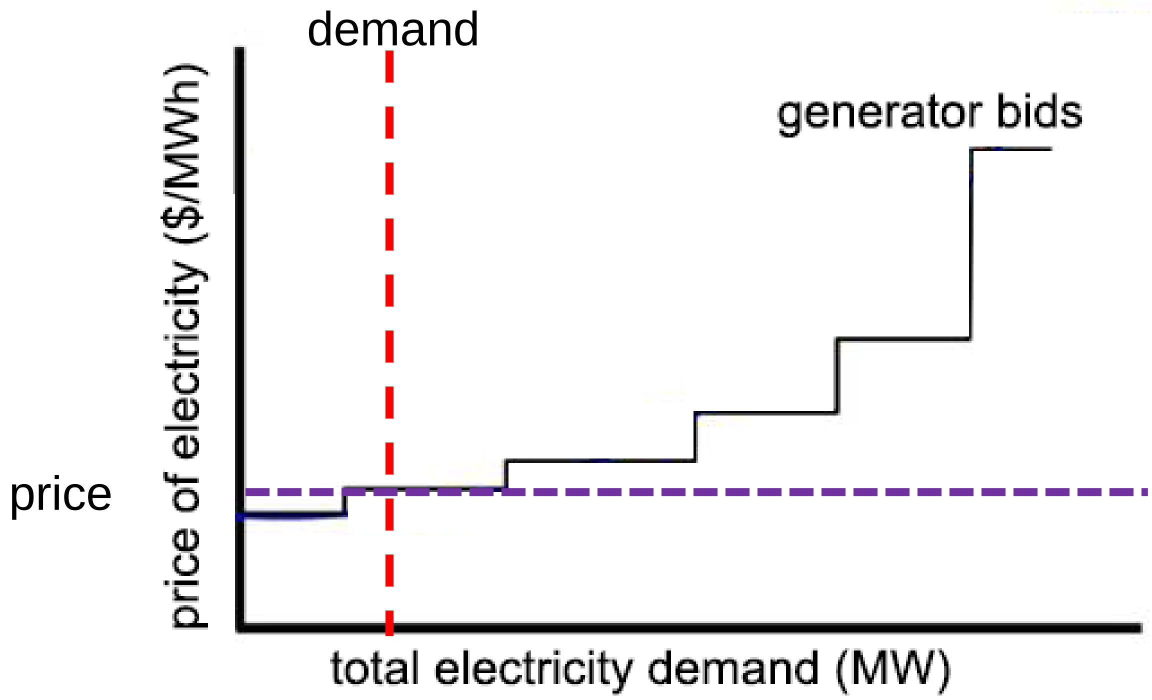
- owners of generators bid into the bulk power market
- as demand increases, more expensive generators are needed to meet the demand

Bulk Power Market



- owners of generators bid into the bulk power market
- as demand increases, more expensive generators are needed to meet the demand

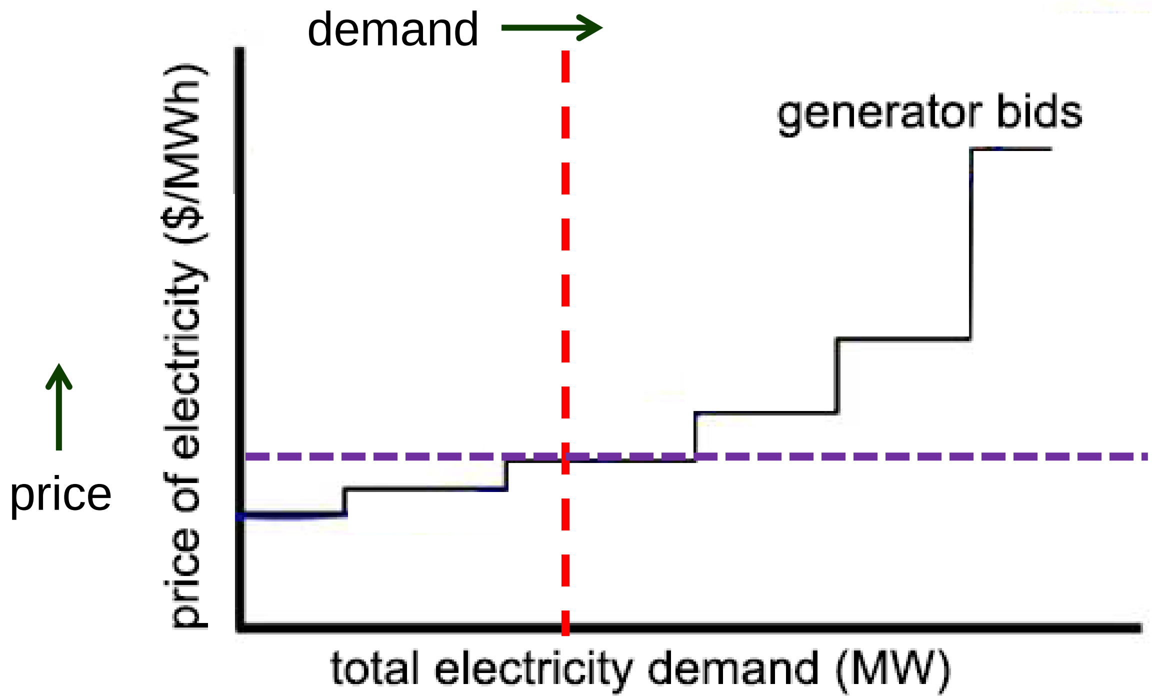
Bulk Power Market

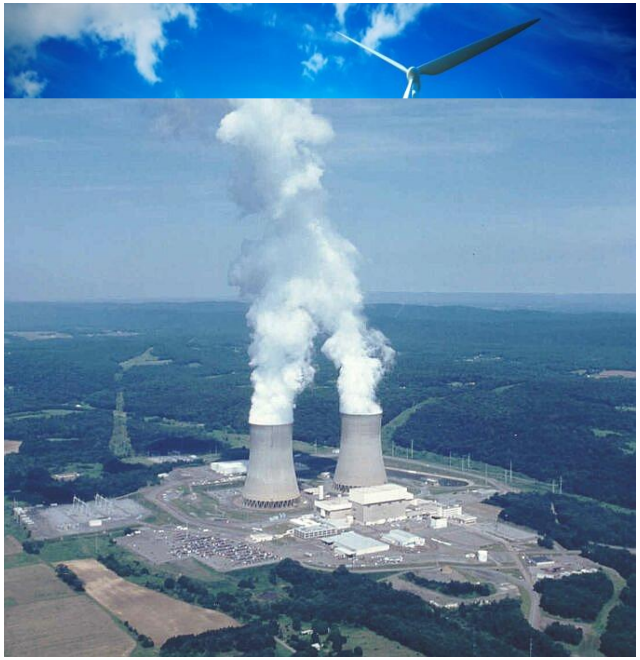


- owners of generators bid into the bulk power market
- as demand increases, more expensive generators are needed to meet the demand

Colorado State University

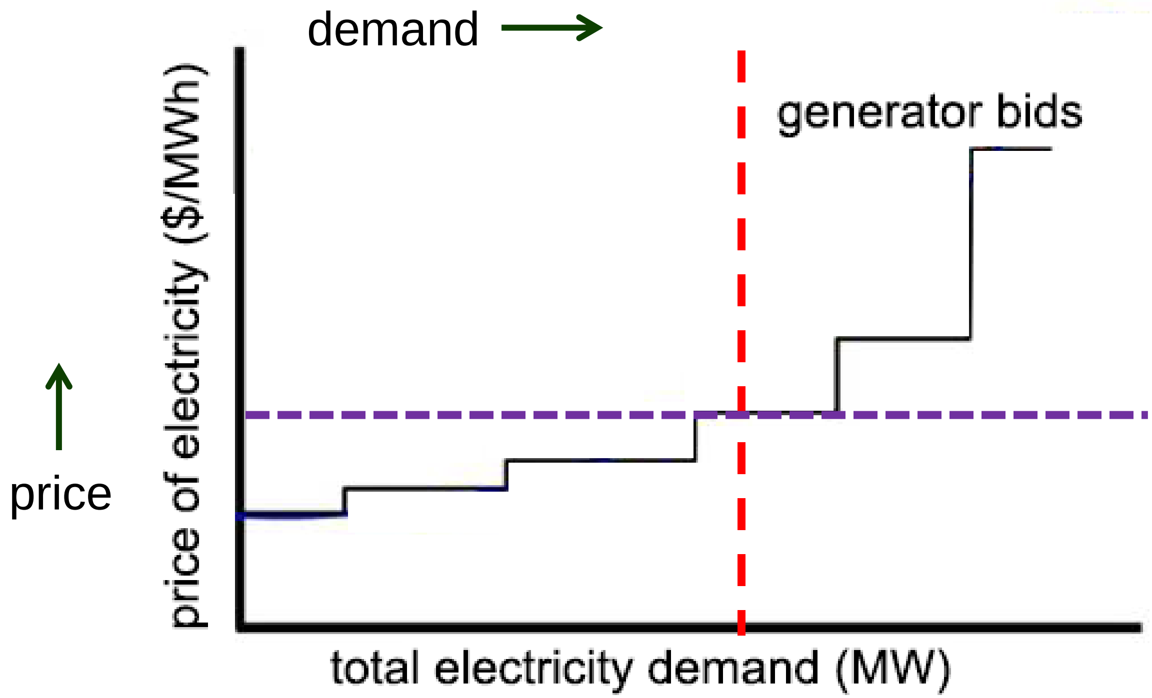
Bulk Power Market





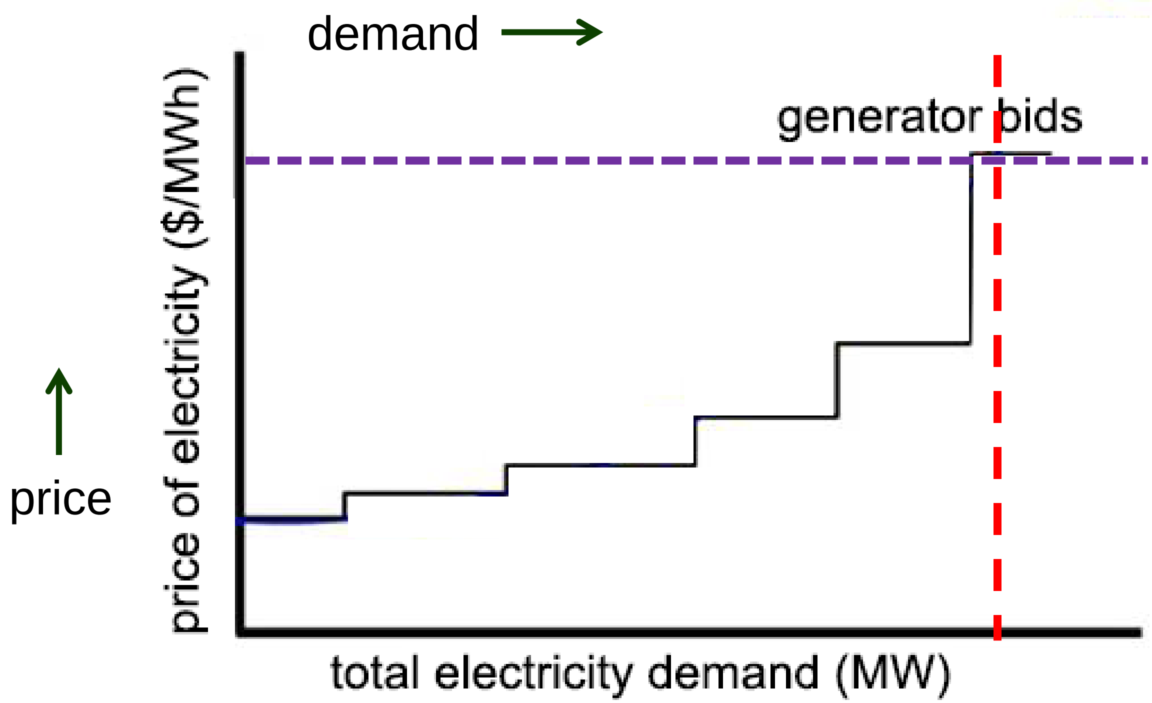
- owners of generators bid into the bulk power market
- as demand increases, more expensive generators are needed to meet the demand

Bulk Power Market



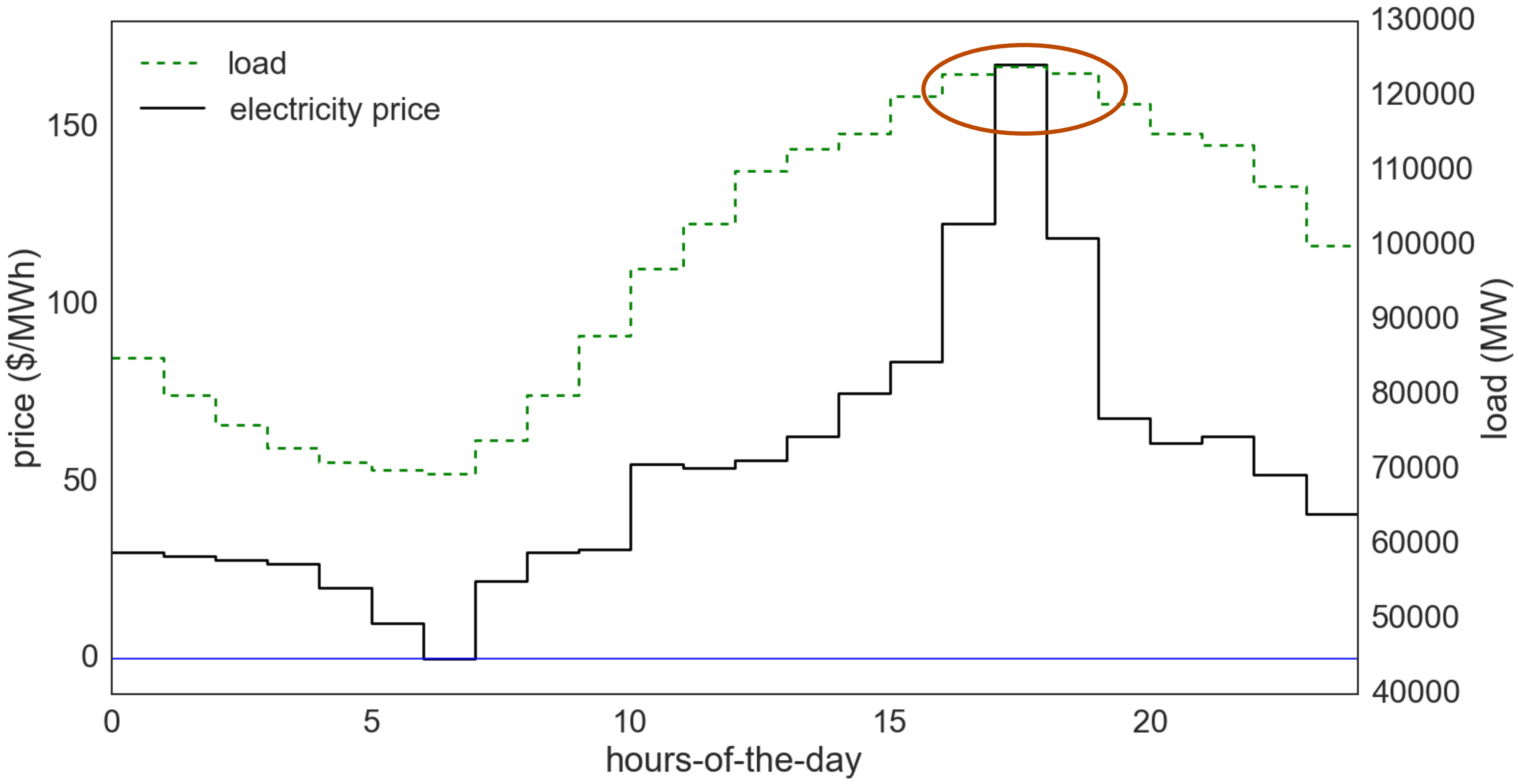
- owners of generators bid into the bulk power market
- as demand increases, more expensive generators are needed to meet the demand

Bulk Power Market



- owners of generators bid into the bulk power market
- as demand increases, more expensive generators are needed to meet the demand

Example Variation of Price

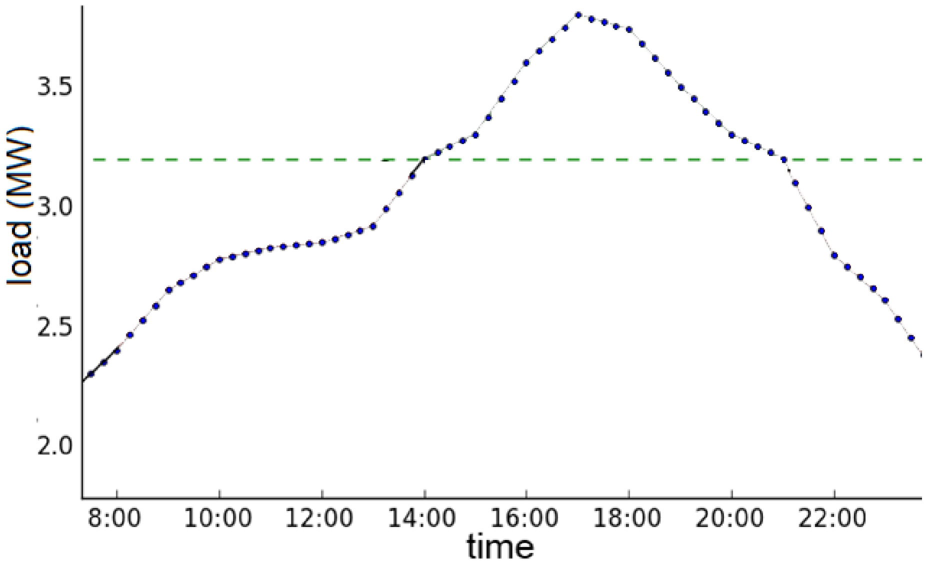


Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions

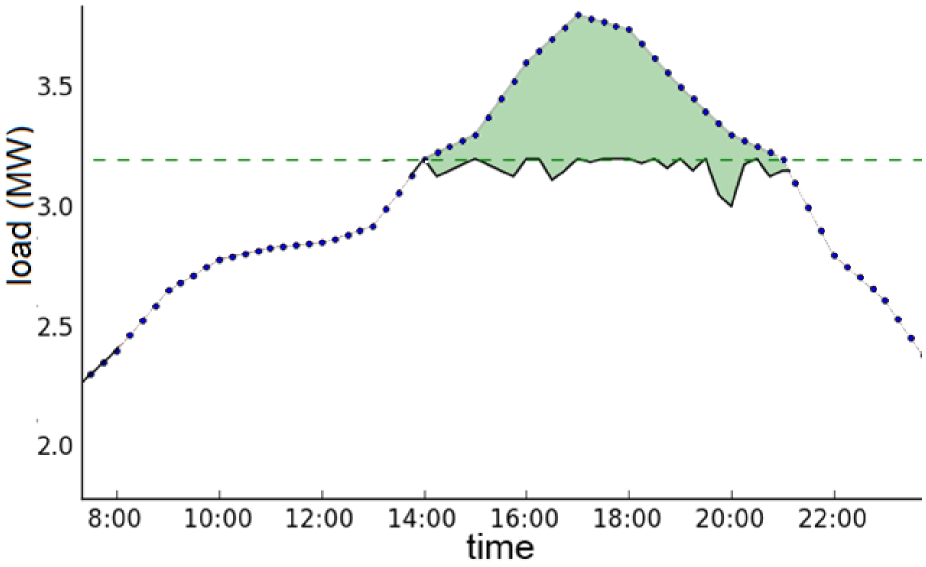
Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- **load shifting**:
  - electric energy consumed by a load (e.g., electric vehicle charging) is moved from one time (e.g., peak) to another
  - total energy consumed is approximately the same



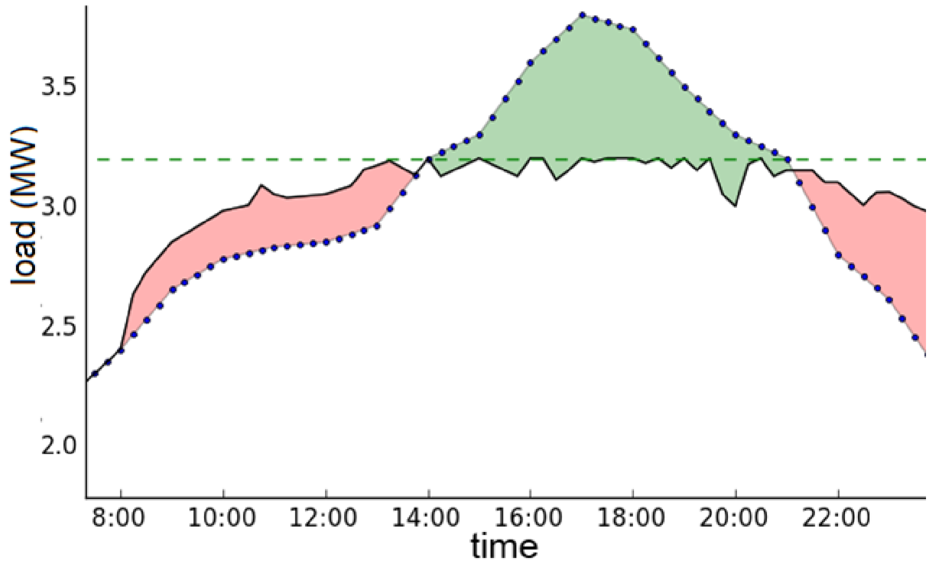
Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- **load shifting**:
  - electric energy consumed by a load (e.g., electric vehicle charging) is moved from one time (e.g., peak) to another
  - total energy consumed is approximately the same



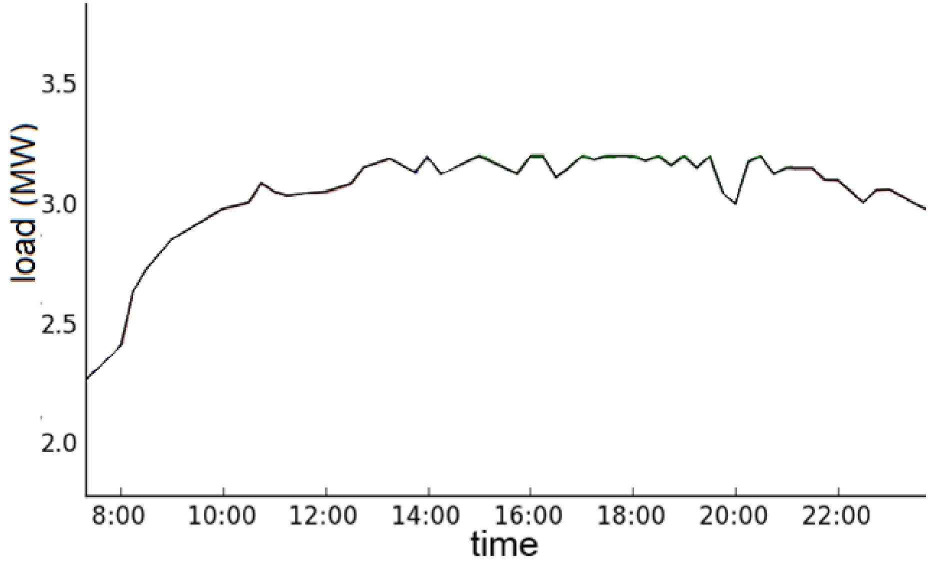
Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- **load shifting**:
  - electric energy consumed by a load (e.g., electric vehicle charging) is moved from one time (e.g., peak) to another
  - total energy consumed is approximately the same



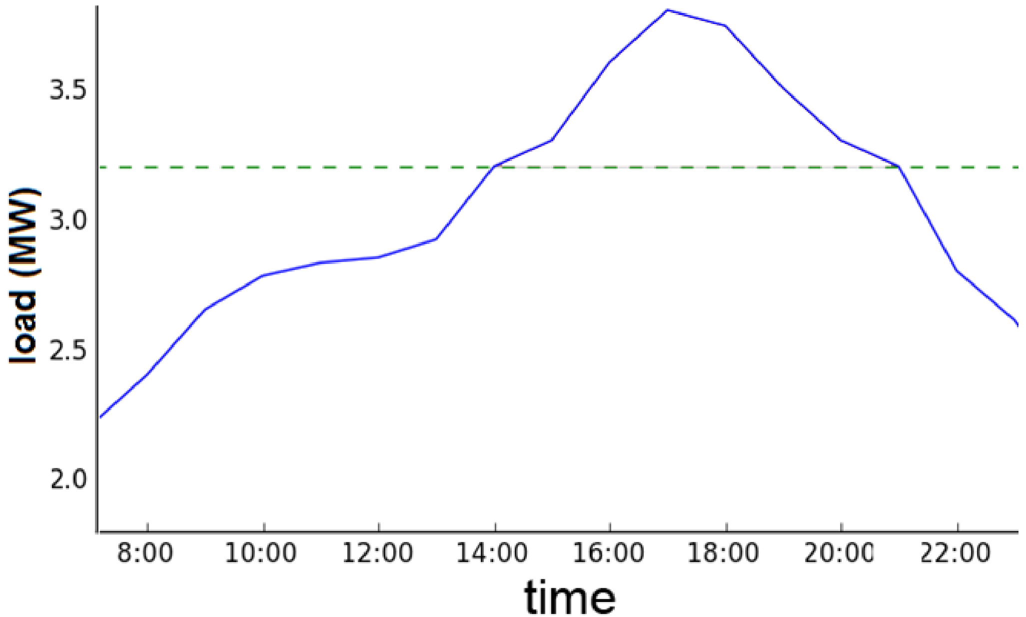
Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- **load shifting**:
  - electric energy consumed by a load (e.g., electric vehicle charging) is moved from one time (e.g., peak) to another
  - total energy consumed is approximately the same



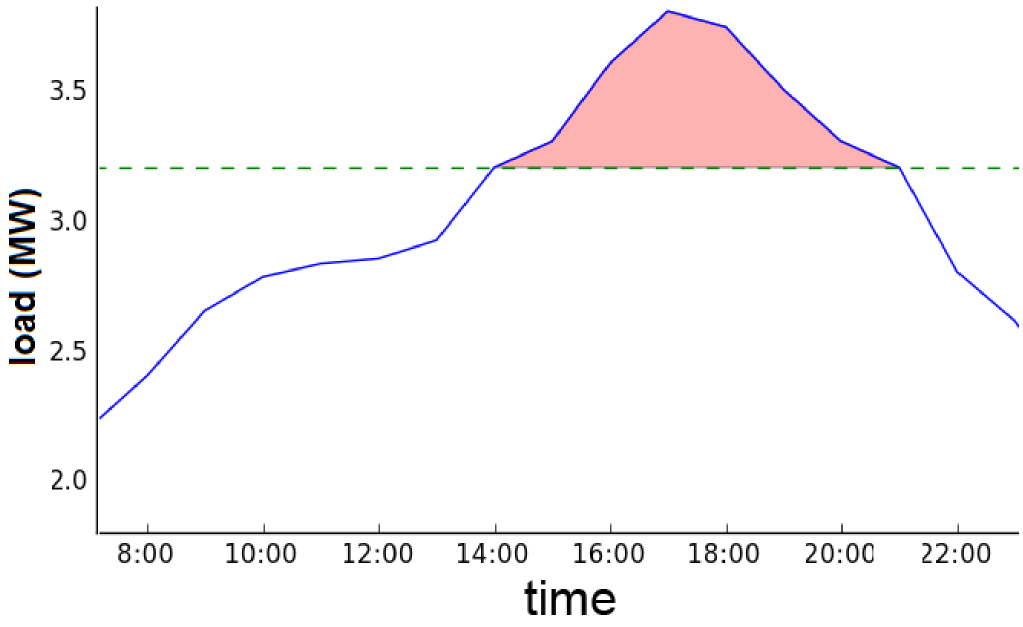
Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- **load shedding**:
  - electric energy consumed by a load is permanently removed from the system
  - total energy consumed is reduced



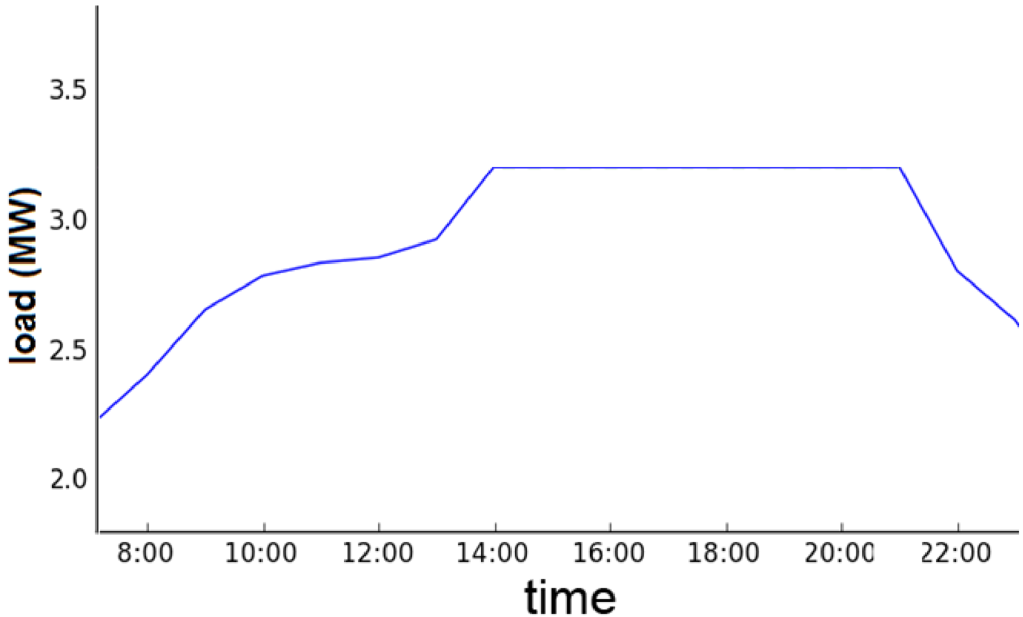
Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- **load shedding**:
  - electric energy consumed by a load is permanently removed from the system
  - total energy consumed is reduced



Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- **load shedding**:
  - electric energy consumed by a load is permanently removed from the system
  - total energy consumed is reduced



Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions

Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions
- sample demand response (DR) programs:
  - direct load control (DLC)
    - air conditioner
    - electric water heater
  - time-varying pricing (indirect)
    - real-time pricing (RTP)
    - time-of-use pricing (ToU)

Colorado State University

Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions

- sample demand response (DR) programs:
  - direct load control (DLC)
    - air conditioner
    - electric water heater
  - time-varying pricing (indirect)
    - real-time pricing (RTP)
    - time-of-use pricing (ToU)

aggregator-based demand response

Demand Response

- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions

- sample demand response (DR) programs:
  - direct load control (DLC)
    - air conditioner
    - electric water heater
  - time-varying pricing (indirect)
    - real-time pricing (RTP)
    - time-of-use pricing (ToU)

aggregator-based demand response

home energy management system

Outline – Home Energy Management

- introduction to resource allocation
- resource allocation in Smart Grid
  - background and motivation
  - **non-myopic home energy management system**
  - aggregator-based residential demand response
  - demand response visualization
  - co-simulation framework
- resource allocation in high-performance computing
- conclusions and future directions

Under Preparation, **Timothy M. Hansen** et al., "A Partially Observable Markov Decision Process Approach to Customer Home Energy Management Systems," *IEEE Transactions on Smart Grid*.

HEMS Research Overview

- **residential home energy management system (HEMS)**
  - **goal**: determine when to use energy within the household to minimize the electricity bill
  - **difficulties**:
    - uncertainty in the time-varying price of electricity
    - as a customer, the benefit of changing energy use must exceed the inconvenience caused

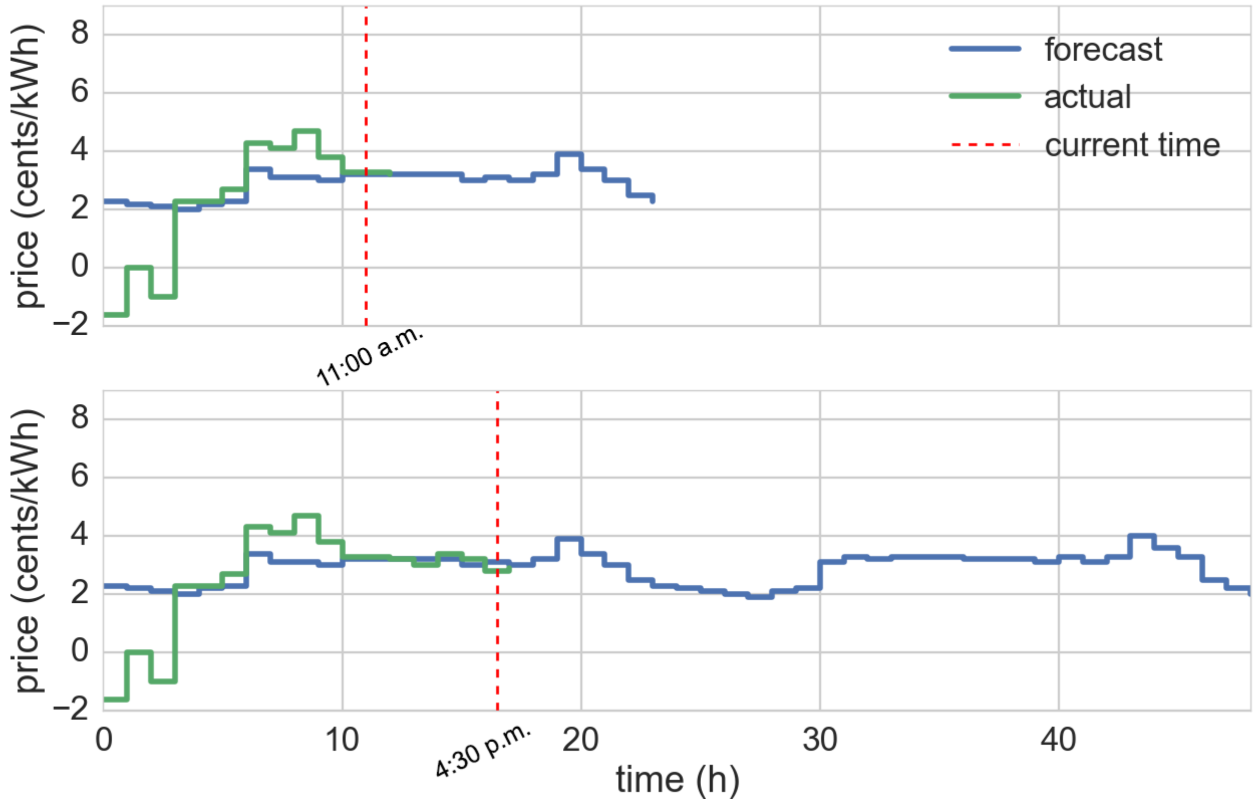
System Model Overview

- single residential household simulation
- household is in a real-time pricing market for electricity
- dynamically “arriving” appliances that need to be used

- **goal:** determine when to run each appliance
  to minimize the total cost of electricity

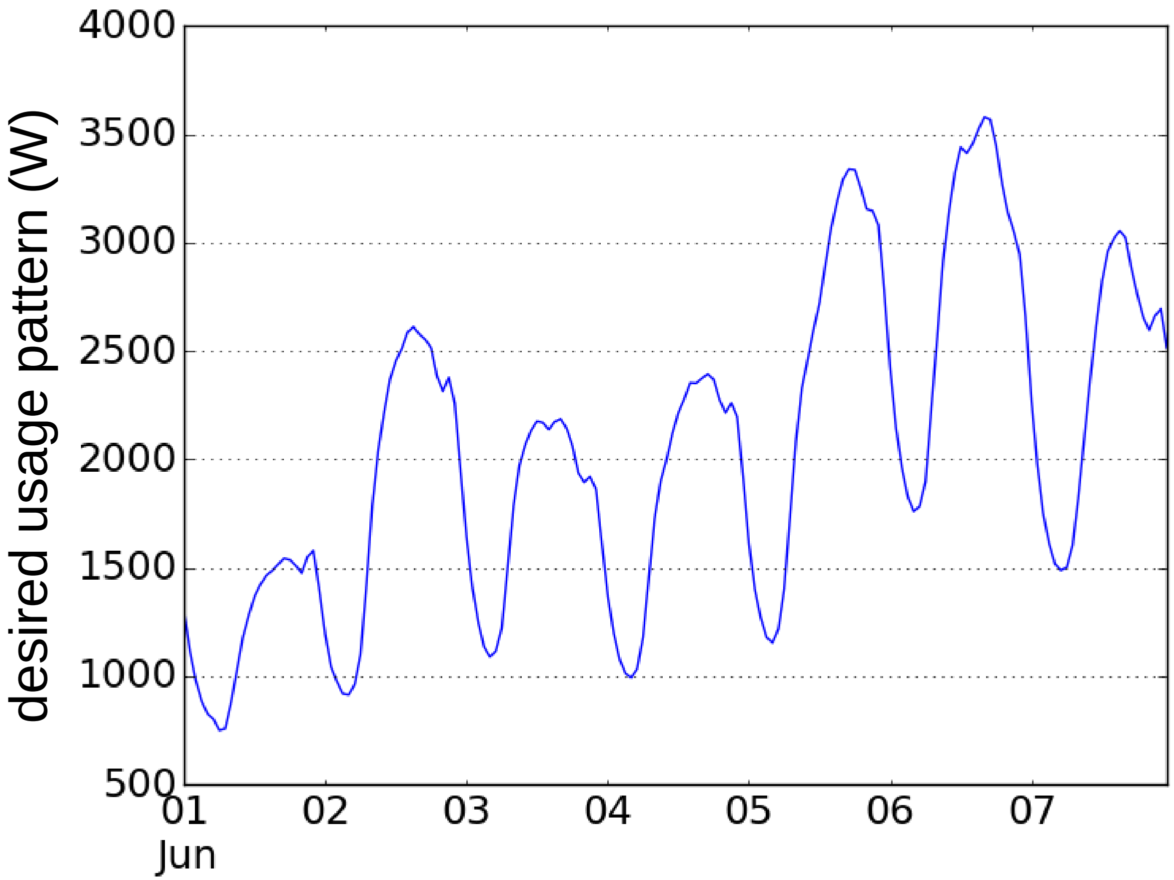
Real-time Pricing Model

- modeled after ComEd real-time pricing (RTP) in Chicago
- price of electricity changes every hour
- provided a forecast for the next day at 4:30 p.m.
- actual price of electricity is known at the start of the hour



Appliance Model

- appliances arrive into the system given a desired usage pattern – models appliance energy usage of a household
  - use queueing theory



Appliance Queue Model

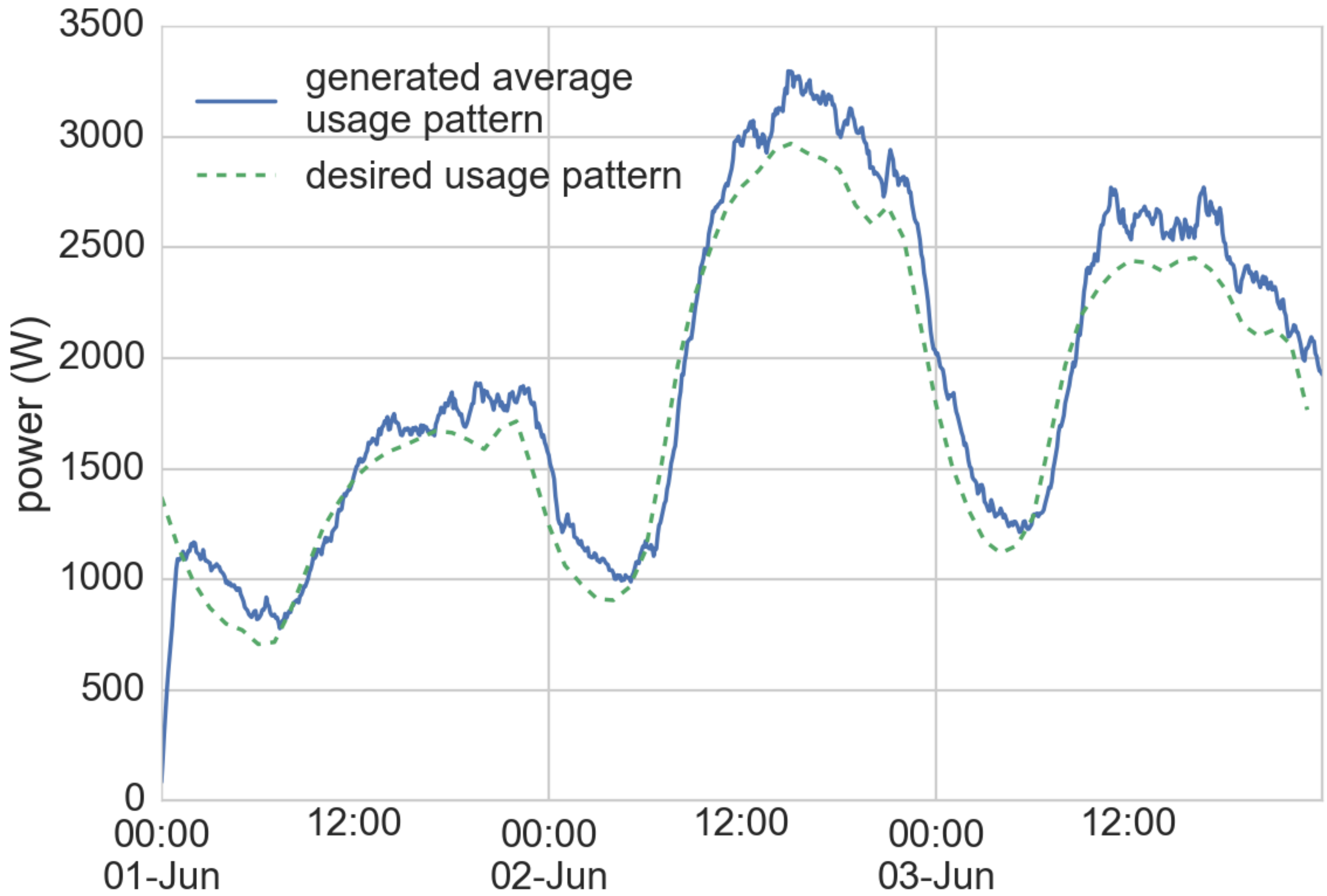
- model the arrival of appliances as an $M_t/G/\infty$ queue
  - arrive with Markovian probability, varying with time
  - appliance run-times are generally distributed
  - household modeled as an infinite capacity compute server
- at time t let:
  - $\lambda(t)$ be the arrival rate of the appliances
  - $l(t)$ be the desired load
  - D be a random variable describing appliance run-times
  - L be a random variable describing appliance loads

$$\lambda(t) = \frac{l(t + \mathbb{E}[D])}{\mathbb{E}[L]\mathbb{E}[D]}$$

Colorado State University

Example Generated Usage Pattern

- appliance arrivals averaged over 500 trials



Problem Statement

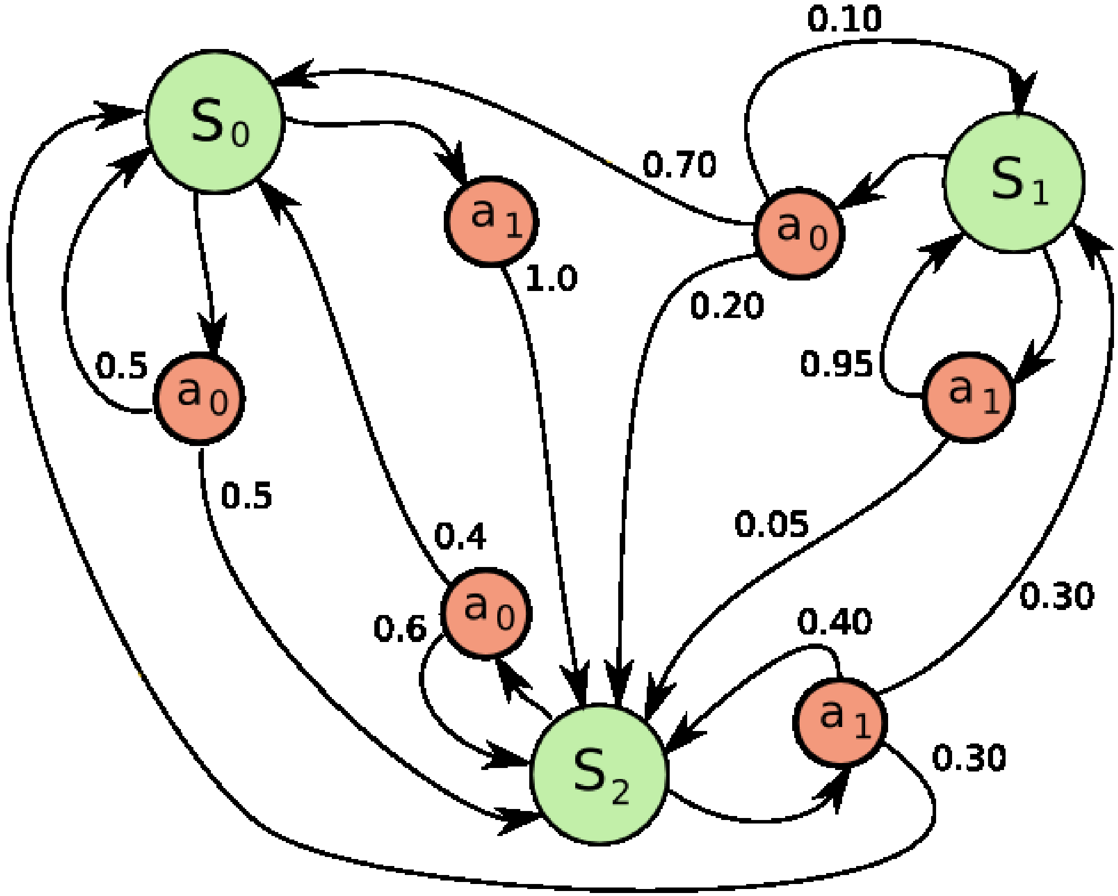
- given:
  - a time-varying RTP modeled after ComEd
  - dynamically starting appliances that statistically follow the provided usage pattern
    - each appliance has a known:
      - power rating
      - duration
      - start time deadline
- constraints:
  - appliances must be started before their start time deadline
- goal:
  - minimize the total cost of electricity of using all appliances

Comparison Methods

- immediate
  - start each appliance at their usage start time (status quo)
- minimum forecast cost
  - schedule appliances to run at the lowest forecast cost
- POMDP
  - Gaussian estimation
  - autoregressive estimation
- lower bound
  - exact minimum cost of running all appliances

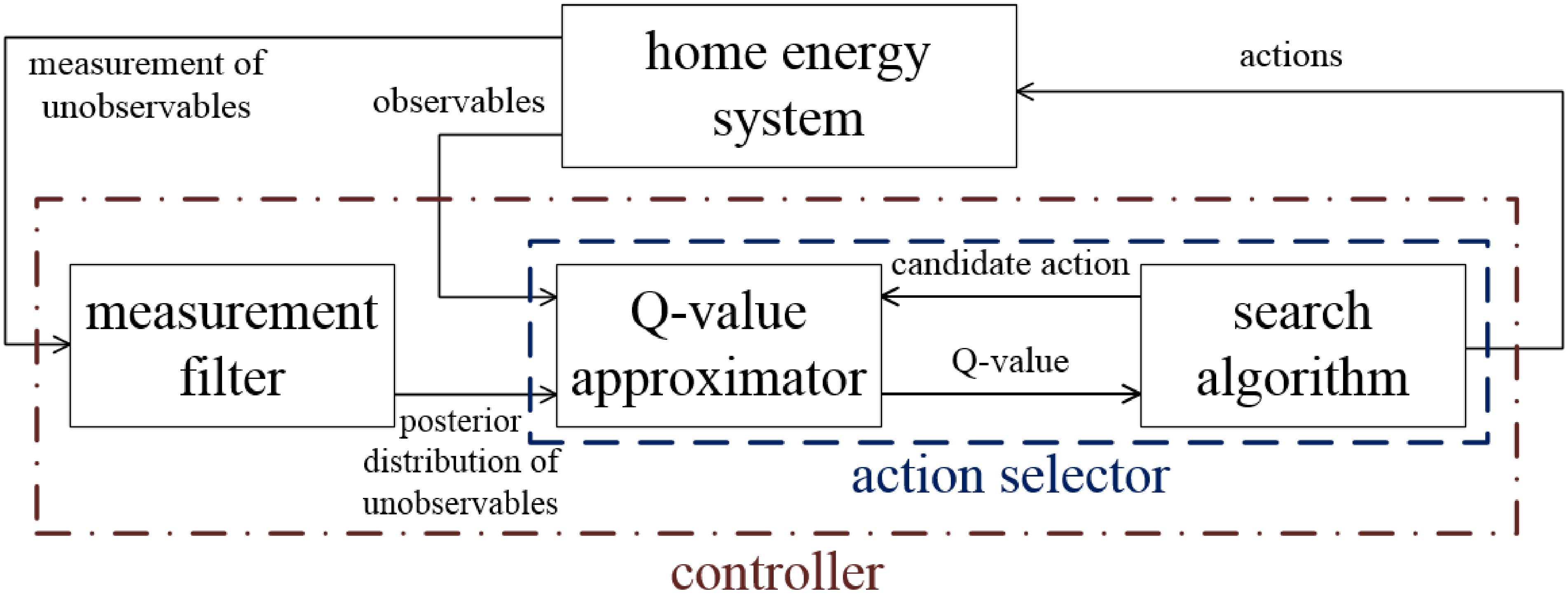
Markov Decision Process (MDP)

- MDP: decision making framework where outcomes have uncertainty
- **goal**: find a policy that maps actions to states to maximize reward



Partially Observable Markov Decision Process

- what if we do not know the underlying state?
- MDP → partially observable MDP (POMDP)
- instead of states, you have **belief** states
  - a probabilistic estimate of the underlying state



POMDP Optimization – Q-value Approximation

- **goal:** maximize total reward over a long time horizon
  - non-myopic, trade-off between immediate and total reward
- use Q-value approximation with Bellman's equation
- for state x and taken actions $\hat{a}$ let:
  - $Q(x,\hat{a})$ be the total expected reward
  - $R(x,\hat{a})$ be the immediate reward
  - x' be the next state
  - $V(x)$ be the optimal reward value following initial state x
- find the policy (action to state mapping) to maximize $Q(x,\hat{a})$

$$Q(x,\hat{a}) = \underbrace{\boxed{R(x,\hat{a})}}_{\text{immediate reward}} + \underbrace{\boxed{E[V(x')|x,\hat{a}]}}_{\text{future reward}}$$

POMDP State

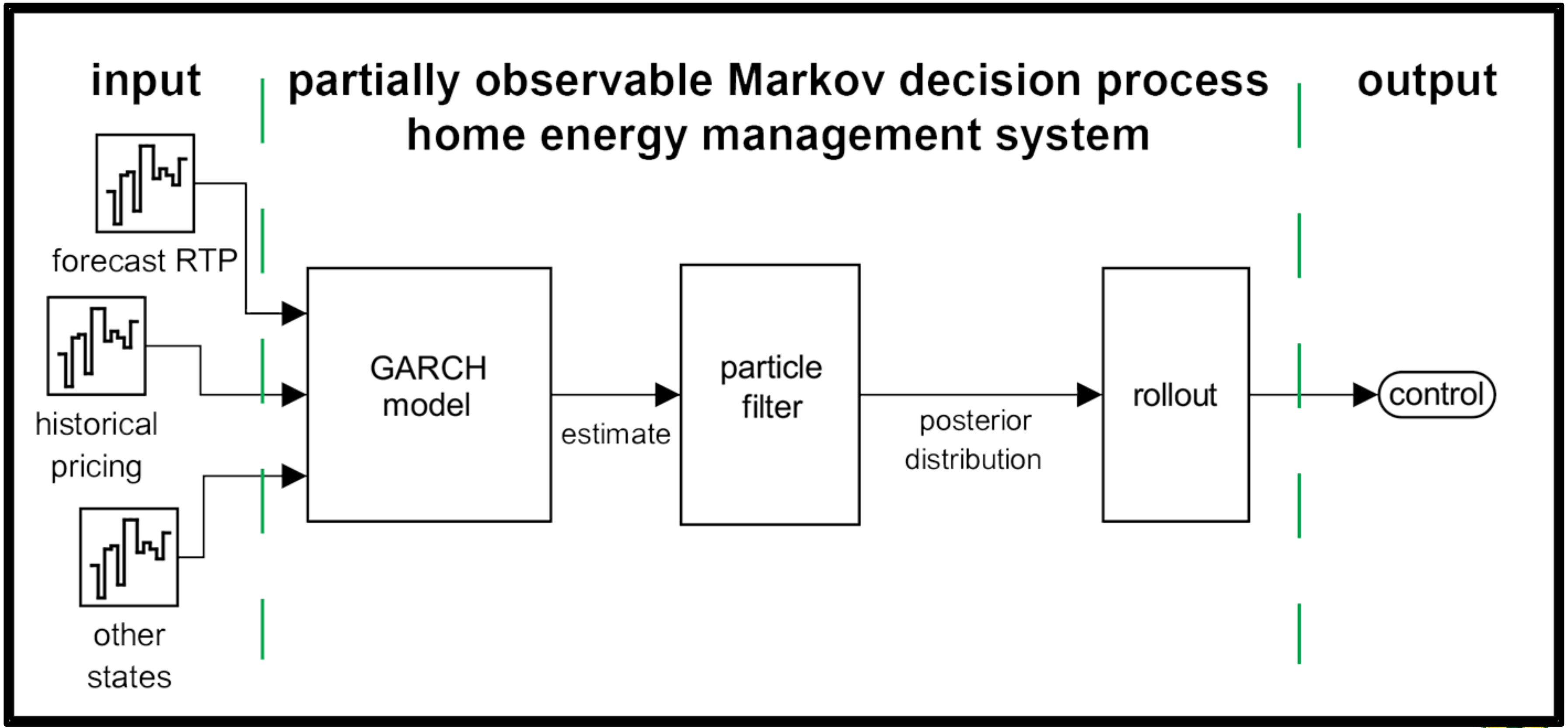
- the underlying state is represented by:
  - the current appliances ready to start (observable)
  - the real-time price (unobservable)
- use a measurement method and filter to determine the belief state of the real-time price

- two measurement methods:
  - Gaussian noise
  - autoregressive time-series model (GARCH)

- use a particle filter to determine conditional probability of the state
  - sequential Monte Carlo sampling method

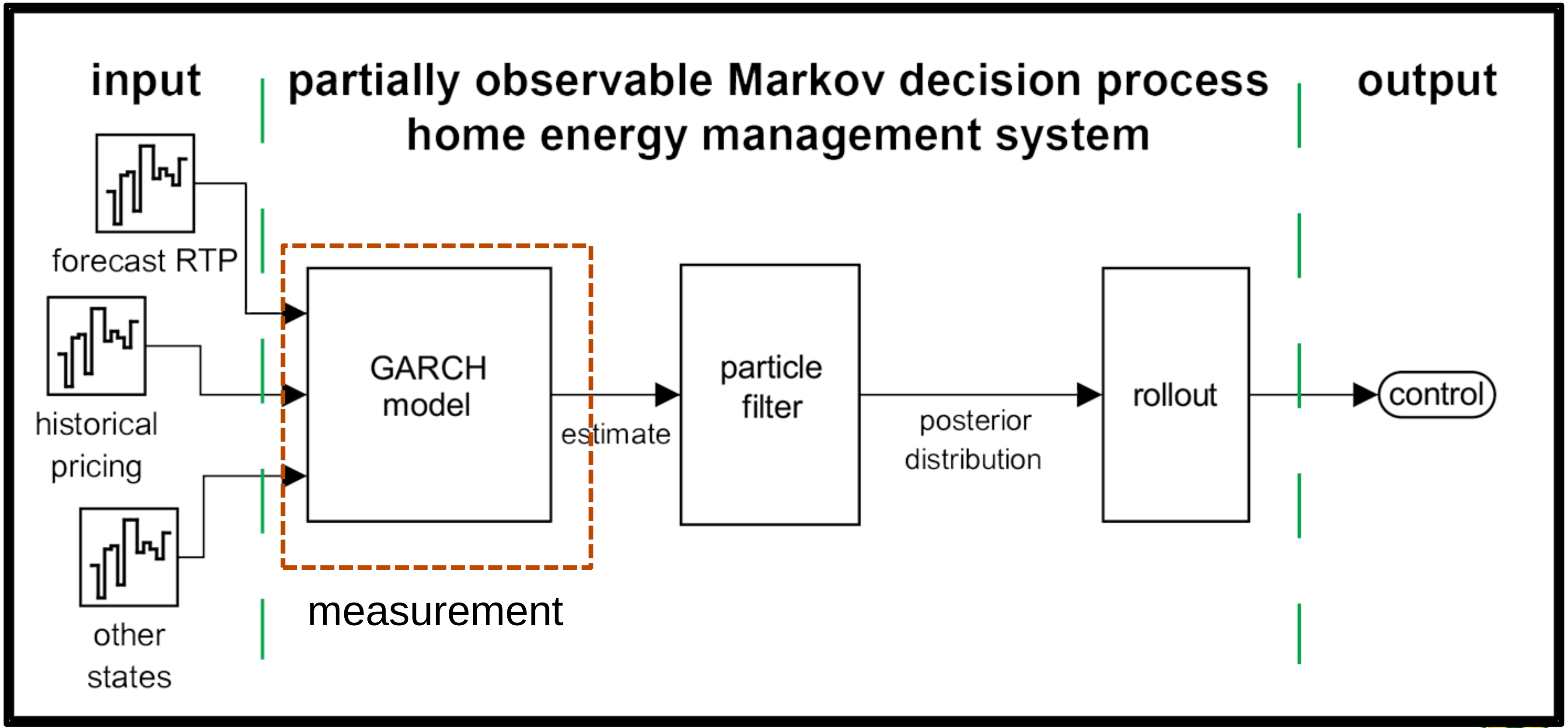
POMDP State

- the underlying state is represented by:
  - the current appliances ready to start (observable)
  - the real-time price (unobservable)



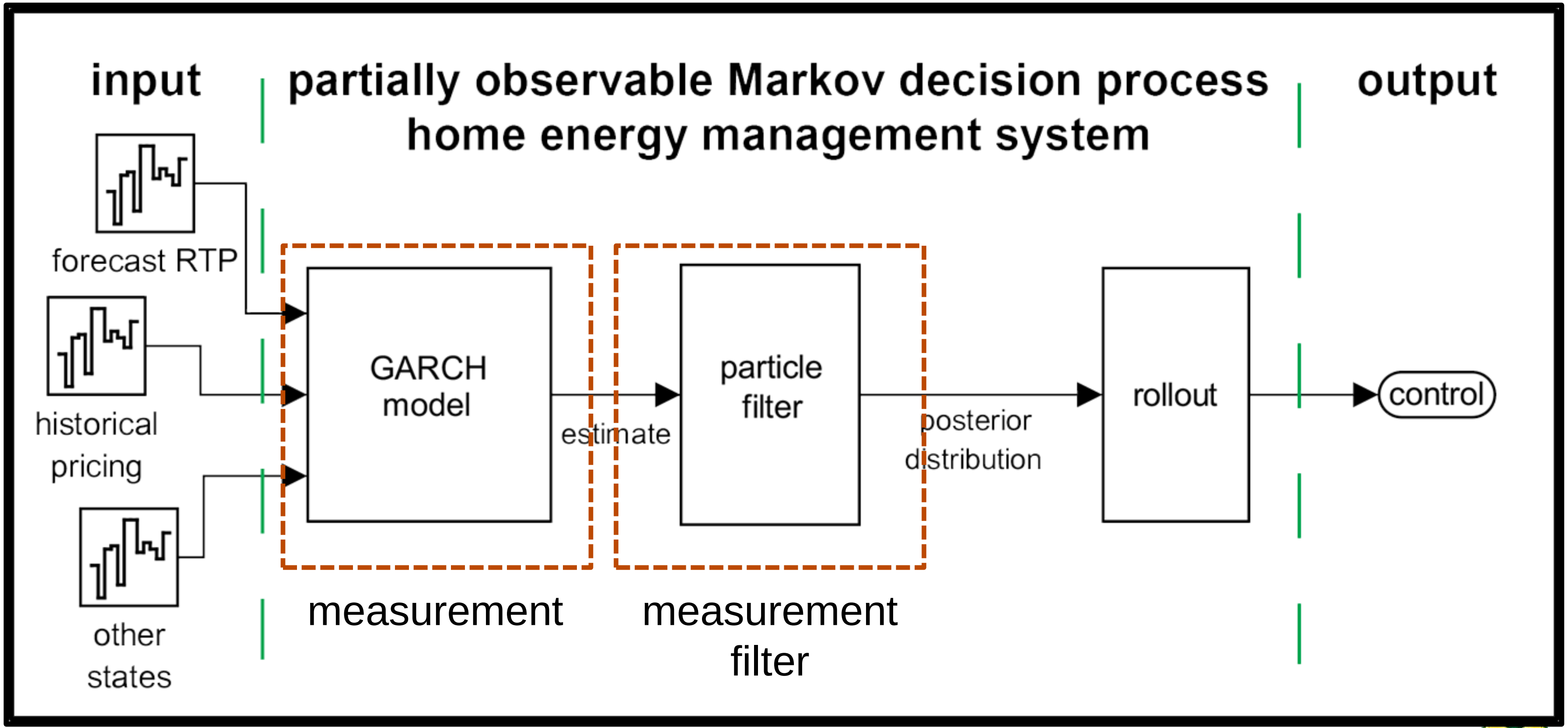
POMDP State

- the underlying state is represented by:
  - the current appliances ready to start (observable)
  - the real-time price (unobservable)



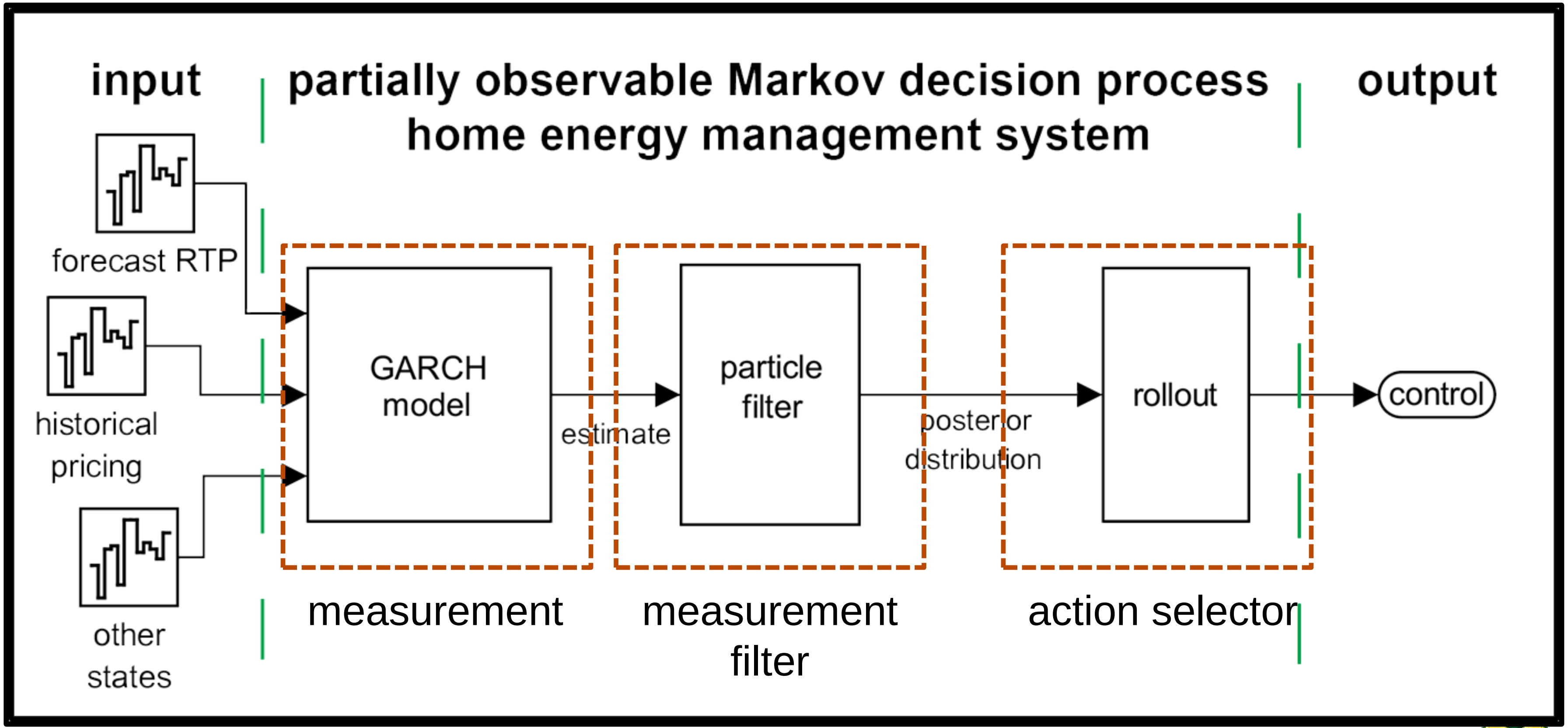
POMDP State

- the underlying state is represented by:
  - the current appliances ready to start (observable)
  - the real-time price (unobservable)



POMDP State

- the underlying state is represented by:
  - the current appliances ready to start (observable)
  - the real-time price (unobservable)



AR Process

- autoregressive (**AR**) process
  - output depends linearly on previous values
- at time t, let:
  - x_t be the output of the AR process
  - c be a constant
  - x_{t-i} be the i^{th} previous output
  - γ_i be the coefficient corresponding to x_{t-i}
  - m be the number of modeled coefficients
  - ϵ_t be the error (usually Gaussian noise)

$$x_t = c + \sum_{i=1}^{m} \gamma_i x_{t-i} + \epsilon_t$$

GARCH Process

- generalized autoregressive conditional heteroskedastic (**GARCH**) process
  - used to model process error (e.g., AR process)
- at time t, let:
  - ϵ_t be the error
  - σ_t be the standard deviation
  - z_t be a Gaussian random variable with mean 0, standard deviation of 1

$$\epsilon_t = \sigma_t z_t$$

- heteroskedastic: the variance of the error (i.e., σ_t^2) varies over time

GARCH – Heteroskedasticity

- at time t, let:
  - k be a constant
  - σ_{t-i}^2 be the i^{th} previous standard deviation
  - p_i be the coefficient corresponding to σ_{t-i}^2
    - GARCH terms
  - P be the number of GARCH terms
  - ϵ_{t-j}^2 be the j^{th} previous square-error
  - q_j be the coefficient corresponding to ϵ_{t-j}^2
    - ARCH terms
  - Q be the number of ARCH terms

$$\sigma_t^2 = k + \sum_{i=1}^{P} p_i \sigma_{t-i}^2 + \sum_{j=1}^{Q} q_j \epsilon_{t-j}^2$$

Colorado State University

AR + GARCH Model for POMDP

- recall AR process:

$$x_t = c + \sum_{i=1}^{m} \gamma_i x_{t-i} + \epsilon_t$$

- recall GARCH process:

$$\epsilon_t = \sigma_t z_t$$

- for POMDP, (number of coefficients from literature):
  - $m = 504$
  - ϵ_t is a GARCH process
    - $P = 1, Q = 3$

Action Selector – Rollout

- recall:

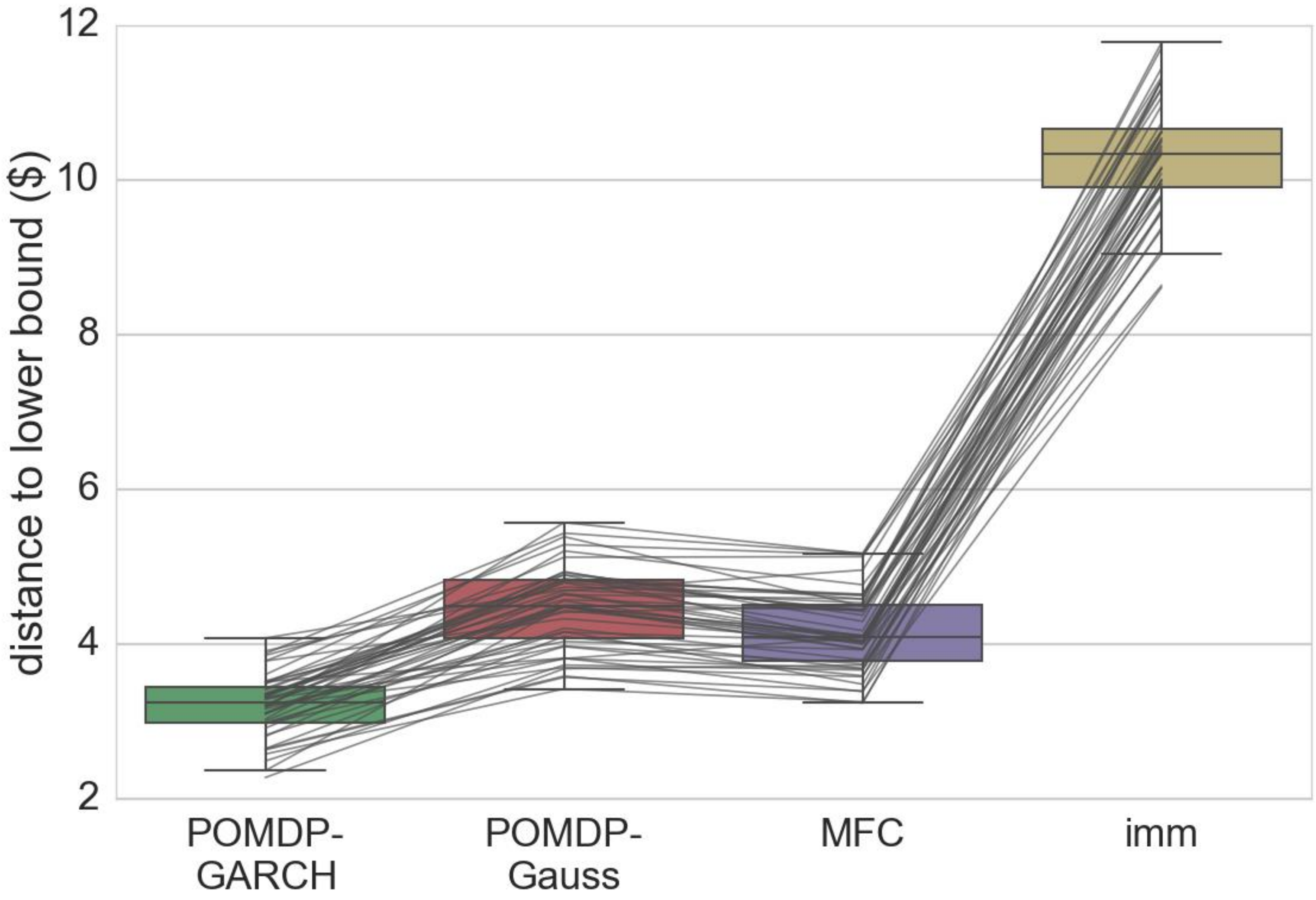
$$Q(x,\hat{a}) = \underbrace{\boxed{R(x,\hat{a})}}_{\text{immediate reward}} + \underbrace{\boxed{E[V(x')|x,\hat{a}]}}_{\text{future reward}}$$

- $\hat{a} = \{now, later\}$
- $R(x,\hat{a})$ – cost of using appliance now
- $E[V(x')|x,\hat{a}]$ – expected cost of waiting
  - determined with a particle filter and the AR + GARCH process

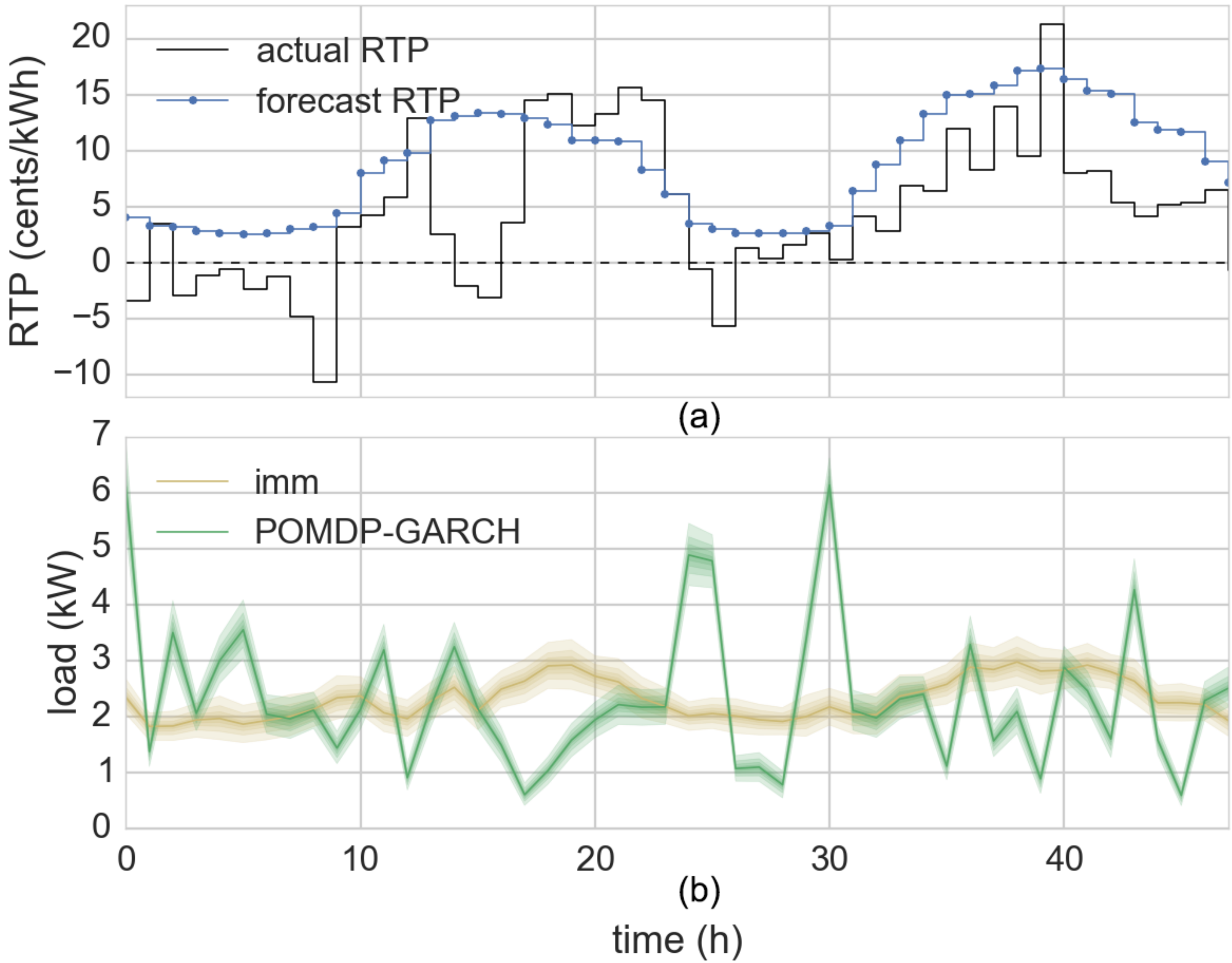
Colorado State University

Example Result – January 2011

- 50 trials of a one-month long simulation of January 2011
- each trial is compared to the lower bound



Change in Energy Usage



Contributions

- design of a non-myopic HEMS using a sequential decision making technique (POMDP)
- creation of a new appliance energy usage pattern based on queueing theory
- comparison of the POMDP HEMS against the lower bound using month-long simulation studies

Outline – Aggregator-based Demand Response

- introduction to resource allocation
- resource allocation in Smart Grid
  - background and motivation
  - non-myopic home energy management system
  - **aggregator-based residential demand response**
  - demand response visualization
  - co-simulation framework
- resource allocation in high-performance computing
- conclusions and future directions

Timothy M. Hansen et al., "Heuristic Optimization for an Aggregator-based Resource Allocation in the Smart Grid," *IEEE Transactions on Smart Grid*, 10 pages, accepted 2015, to appear. DOI: 10.1109/TSG.2015.2399359.

Recall: Demand Response

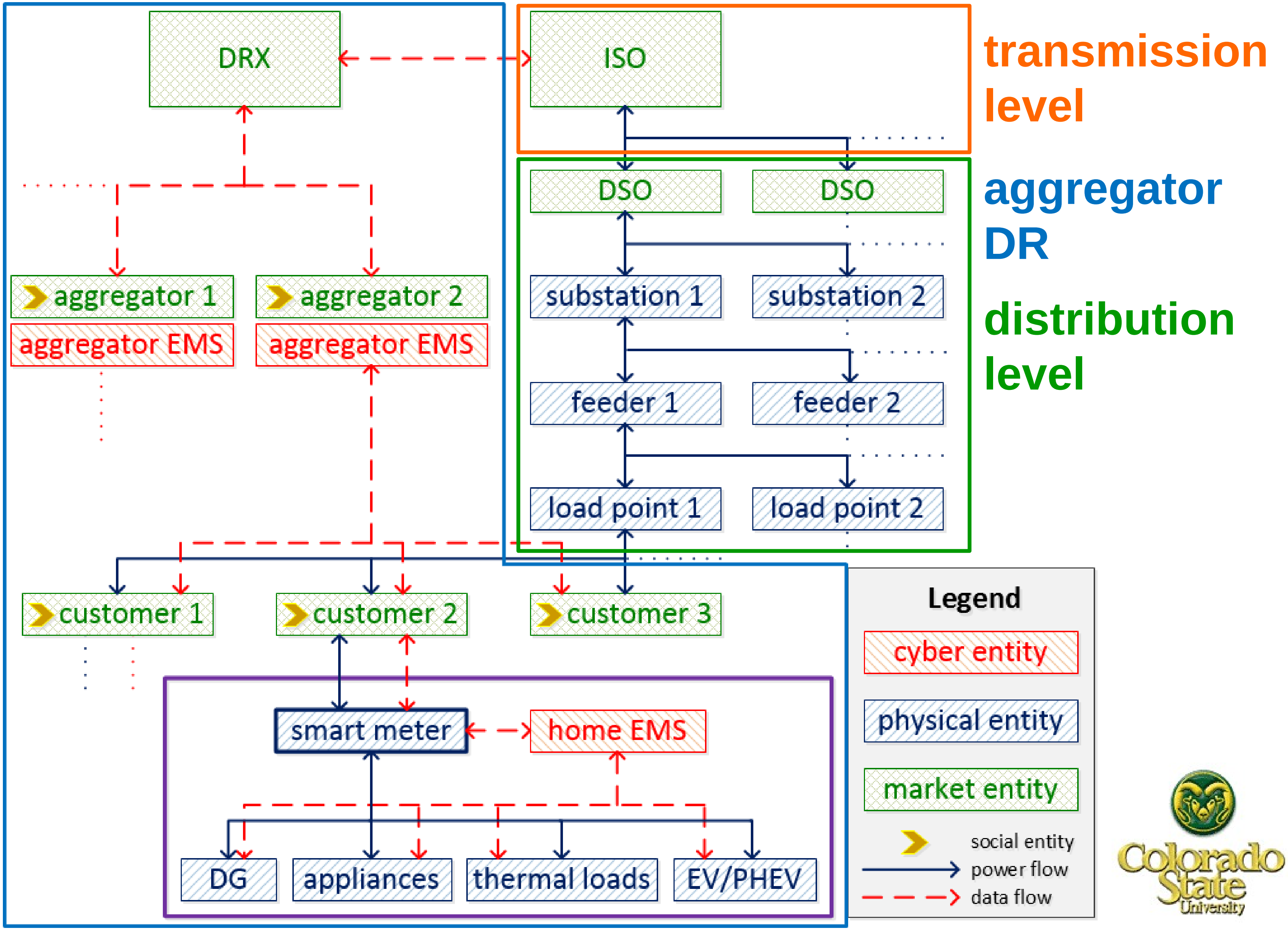
- **demand response:**
  - peak demand reduction by shifting or shedding loads in response to system or economic conditions

- sample demand response (DR) programs:
  - direct load control (DLC)
    - residential assets
  - time-varying pricing (indirect)
    - real-time pricing (RTP)
    - time-of-use pricing (ToU)

aggregator-based demand response

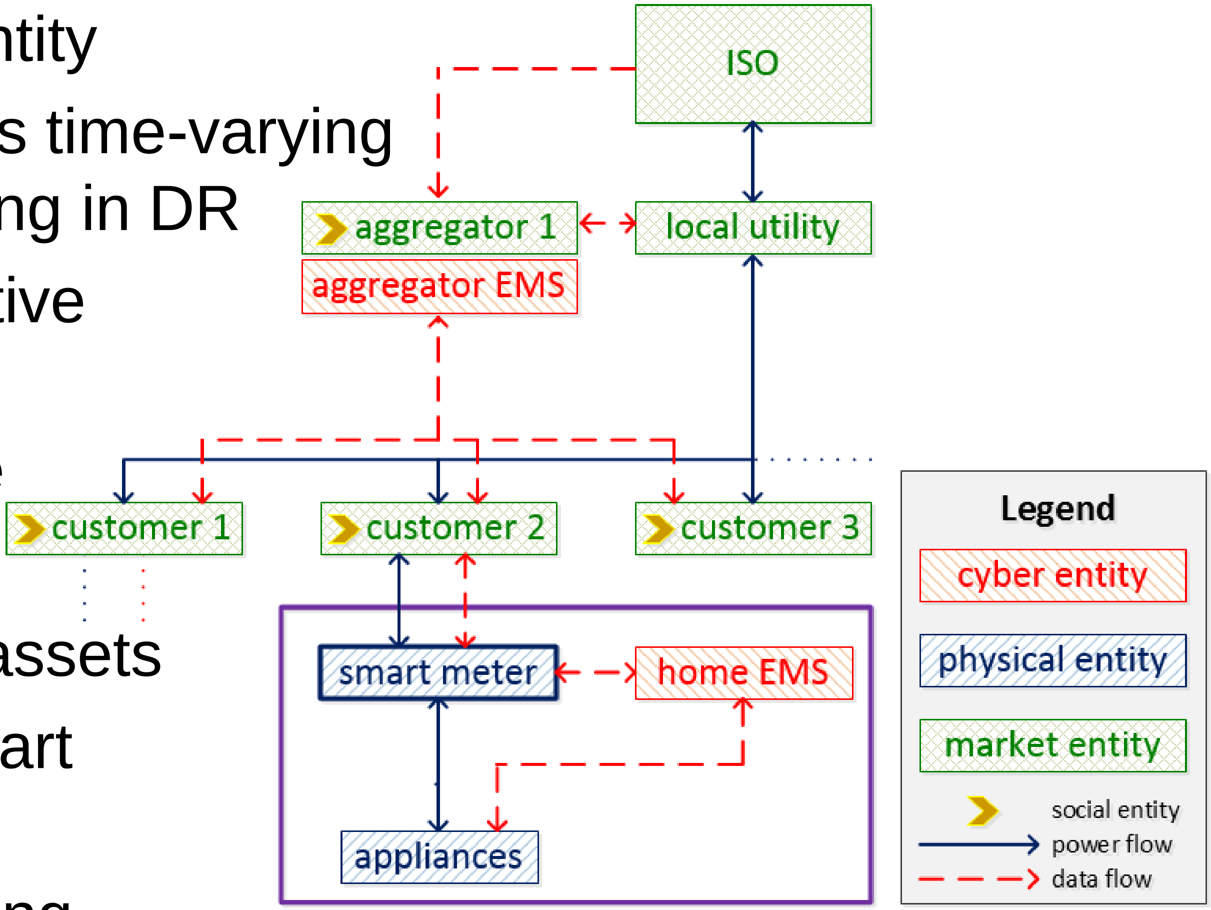
Colorado State University

Cyber-physical Social System (CPSS)



Aggregator-based Residential Demand Response

- for-profit aggregator entity
  - offers all customers time-varying price for participating in DR
    - customer incentive price (CIP)
    - competitive rate
- customer
  - owns a set of DR assets
    - schedulable smart appliances
  - pays CIP for allowing aggregator use of DR assets
    - generally cheaper than utility price



Assumptions

- price is exogenous
  - at the load levels one aggregator changes, bulk price changes marginally
- retail electricity market is fully deregulated
  - allows customer to choose supplier
- control and communication infrastructure
  - exchange of information and control of DR assets
- customer willingness to participate

Smart Grid Resource Allocation

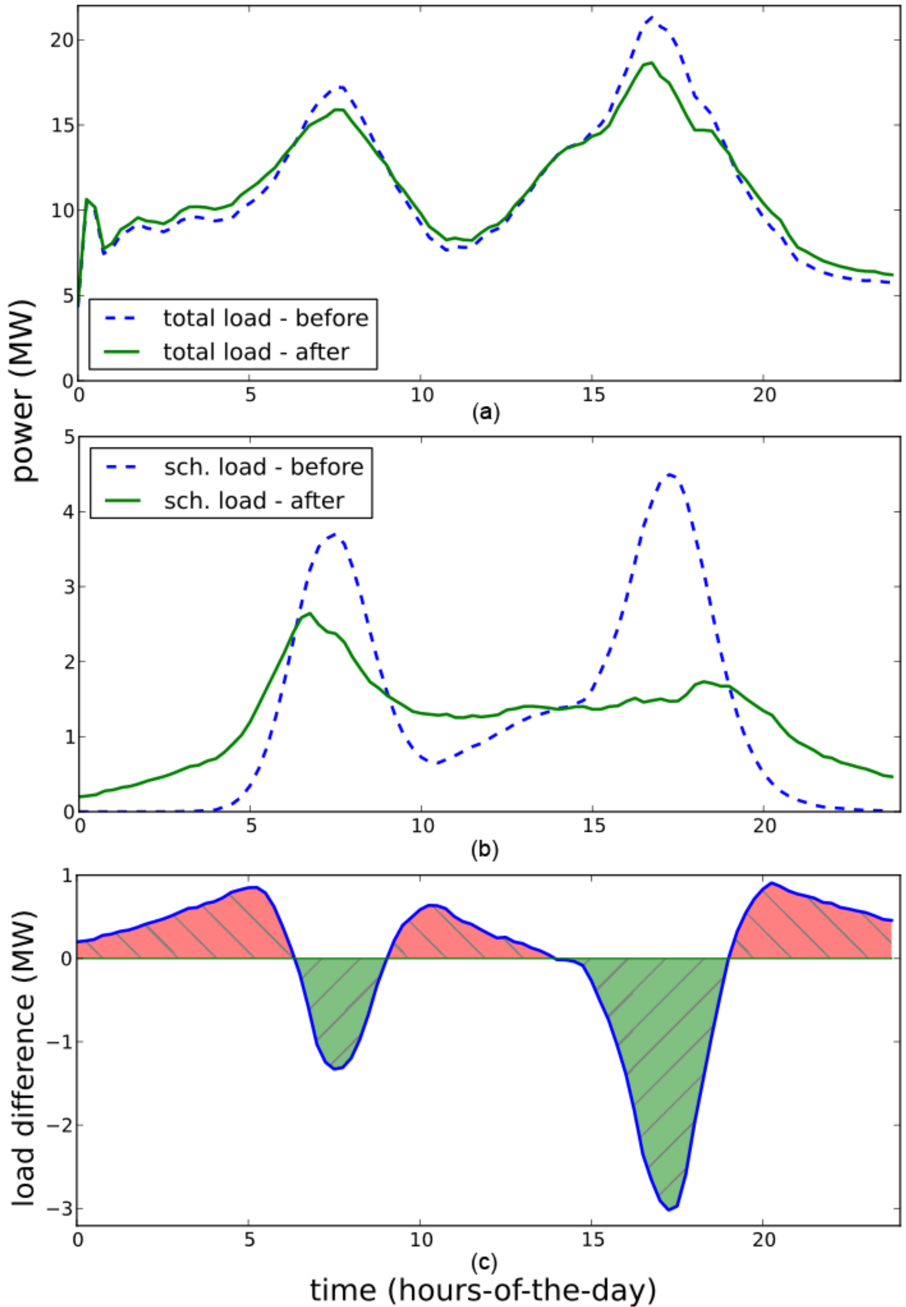
- Smart Grid Resource Allocation (**SGRA**)
- from the point-of-view of the aggregator
- given
  - set of customers
  - information about customer loads
- constraints
  - customer constraints
    - availability of loads to be rescheduled
    - incentive pricing requirements
  - system
- objective
  - aggregator find *customer incentive pricing* and *schedule of loads* to maximize aggregator profit

Simulation Setup

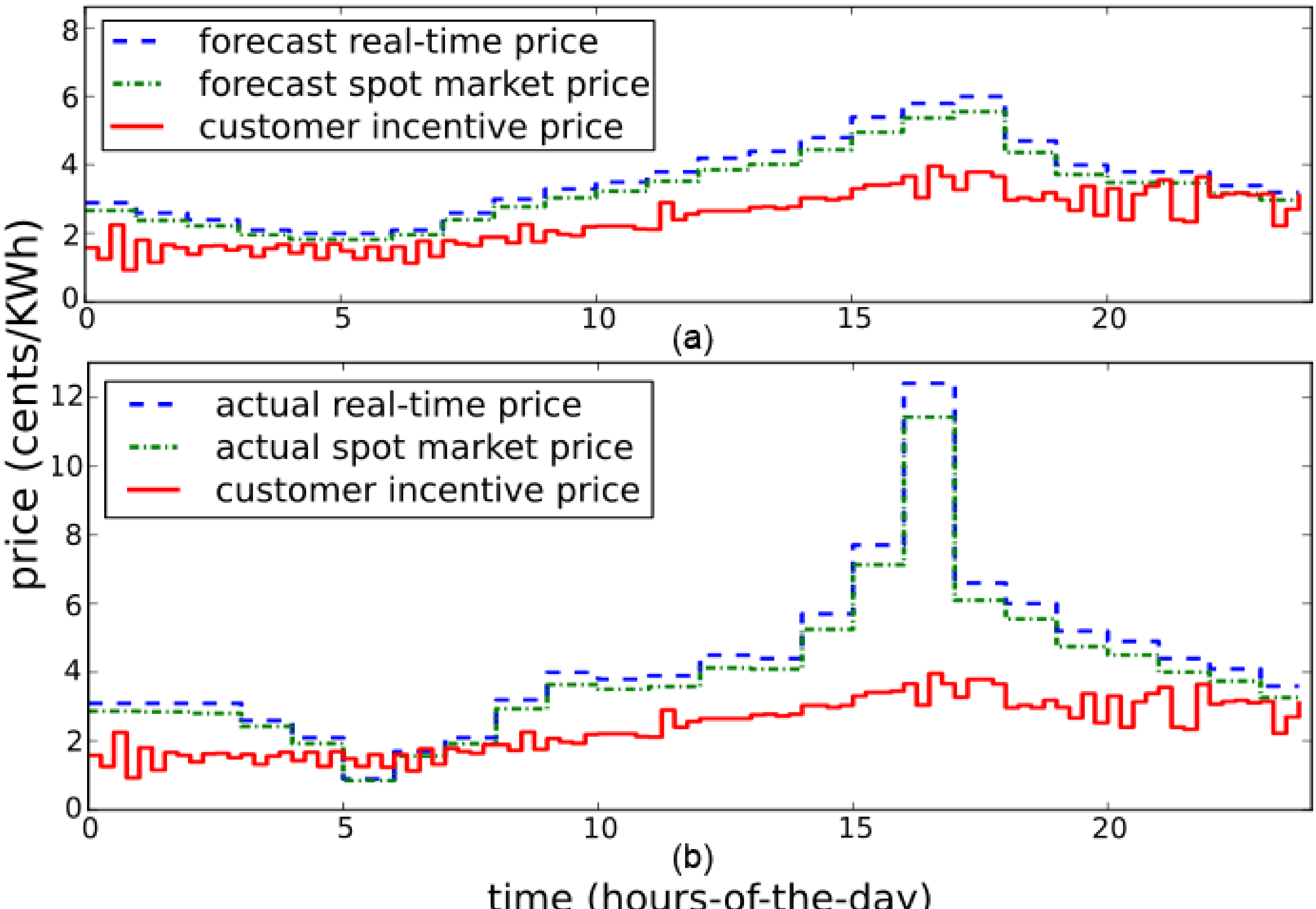
- 5,555 customers
  - each customer has a threshold for determining whether the CIP offers enough discount
  - each customer defines a time period to which each load can be rescheduled
- 56,498 schedulable loads
  - probabilistically generated to simulate use of an average household
- pricing data
  - bulk power spot market price from PJM
  - utility price from ComEd
- genetic algorithm used as optimization method

Results – Demand Response Load Shifting

- peak reduction of 2.66MW (12.6%)
- aggregator profit
  - $947.90
- total customer savings
  - $794.93



Customer Incentive Pricing



Contributions

- alternative customer pricing structure
  - *customer incentive pricing*
- heuristic optimization framework
  - mathematical models for the customer and aggregator entities
- large-scale test simulation consisting of 5,555 customers and ~56,000 schedulable loads
  - used real pricing data from ComEd and PJM
- showed that aggregator optimizing for economic reasons:
  - benefits participating customers
  - benefits aggregator
  - benefits *non-participating* customers
    - system peak reduced as a common good

Outline – Demand Response Visualization

- introduction to resource allocation
- resource allocation in Smart Grid
  - background and motivation
  - non-myopic home energy management system
  - aggregator-based residential demand response
  - **demand response visualization**
  - co-simulation framework
- resource allocation in high-performance computing
- conclusions and future directions

Timothy M. Hansen et al., "A Visualization Aid for Demand Response Studies in the Smart Grid," *The Electricity Journal*, vol. 28, no. 3, pp. 100-111, Apr. 2015.

Colorado State University

Power Systems Visualization

- as new power systems technologies become more pervasive, the amount of data available increases dramatically
- abundance of data makes it difficult to:
  - assess power system states in a fast, user-friendly manner
  - find key information
- application of visualization methods provide unique insight and information to users

- **goal:** design new visualization methods for evaluating demand response programs

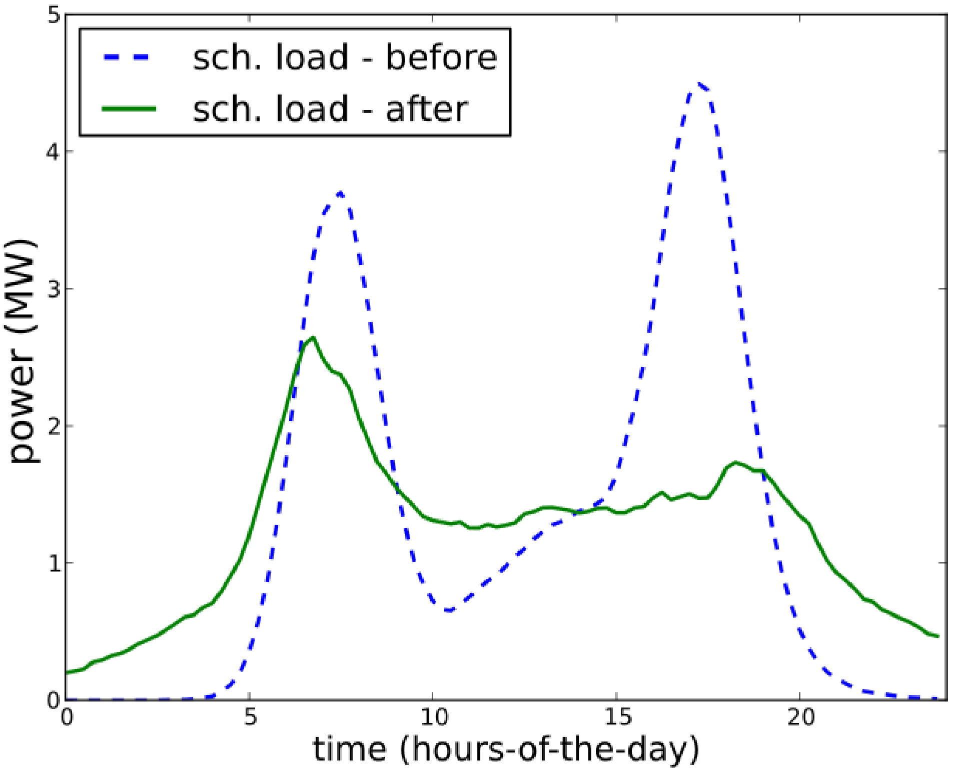
System Model

- use the system model and results from the aggregator-based DR work
- aggregator profit, P, is composed of three pieces:
  - N – *income received* by aggregator for selling negative load to spot market at times load *rescheduled from*
  - S – *income received* by aggregator for selling electricity to customer at times load *scheduled to*
  - B – *cost paid* by the aggregator for buying electricity from spot market at times load *scheduled to*

$$P = N + S - B$$

Visualization Method Overview

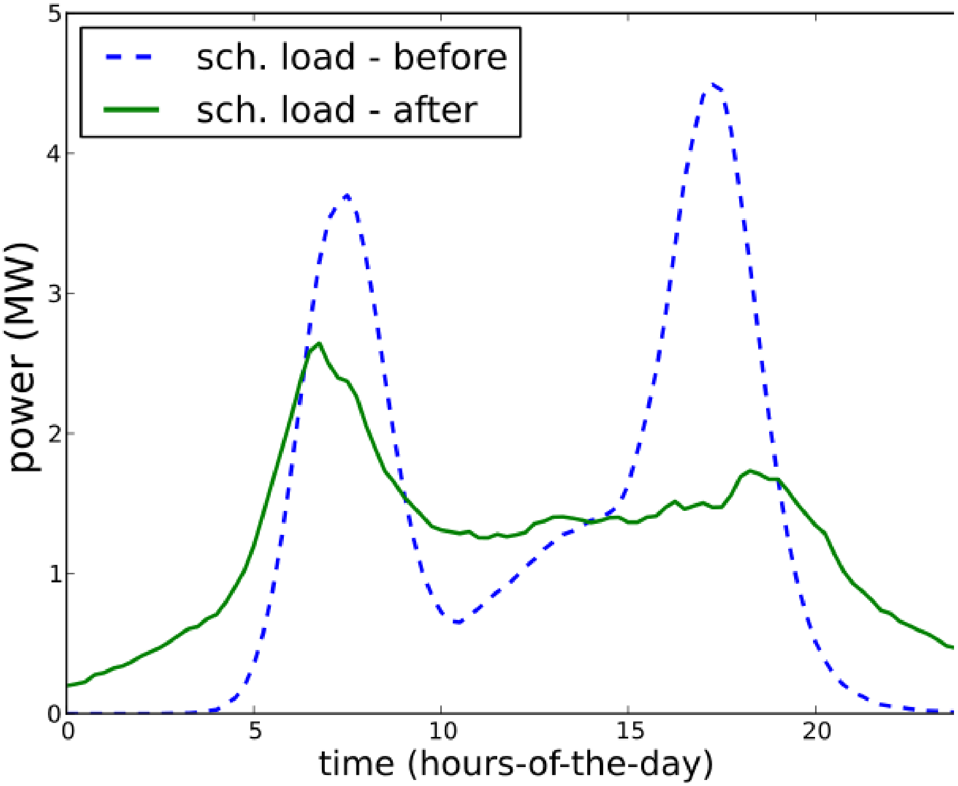
- DR actions normally visualized as 2D load curves

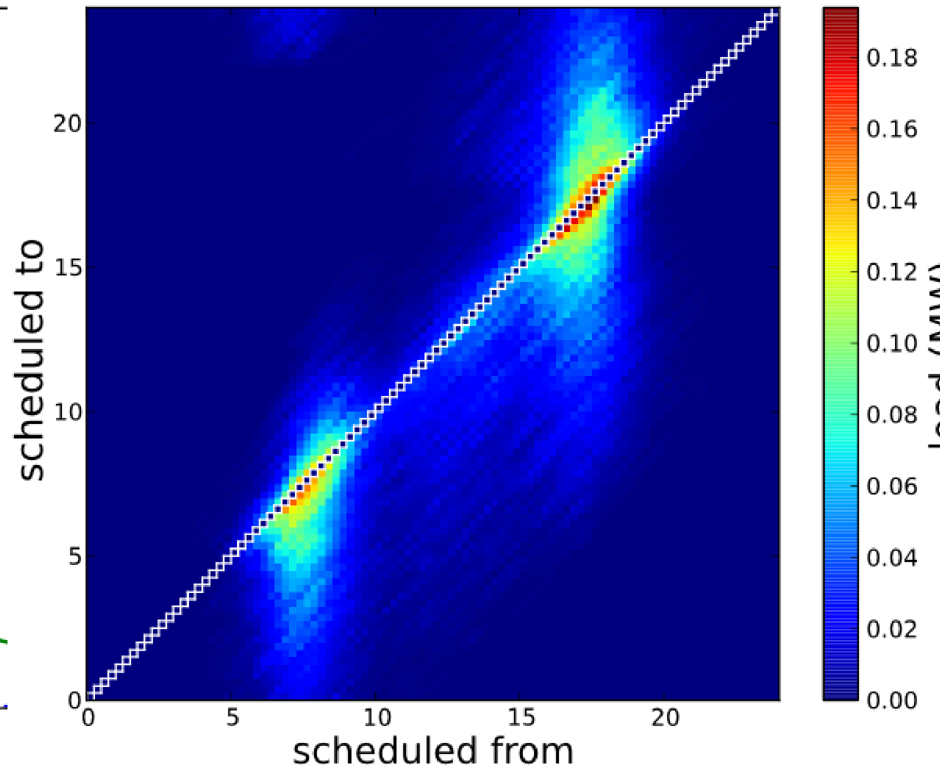


- three new methods for DR visualization:
  - heat map representation
  - 3D load curve
  - spatially distributed DR

Heat Map Representation

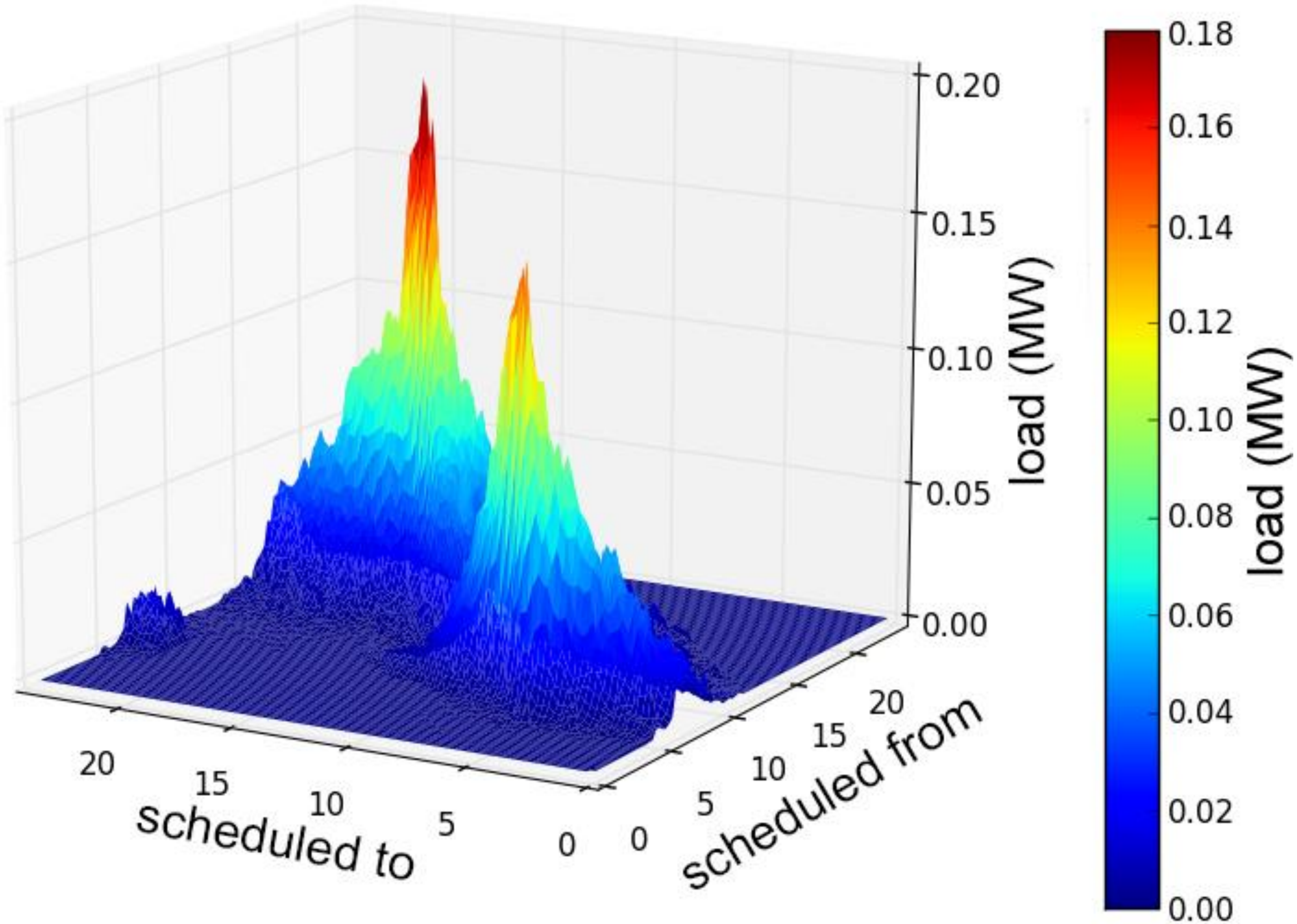
- color at a point (x,y) indicates the load moved from time x to time y (magnitude given by associated color bar)





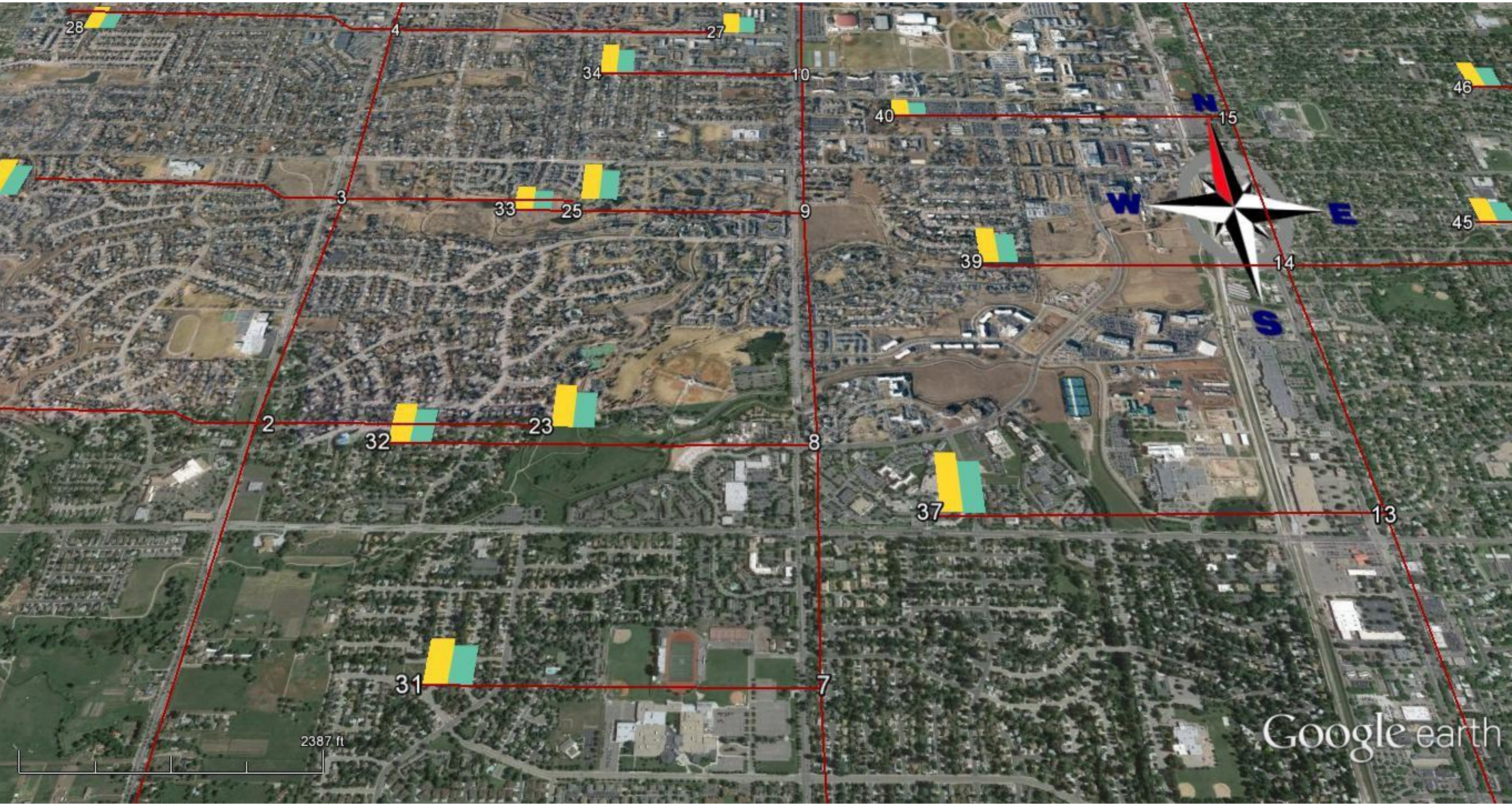
3D Load Curve Representation

- at each surface point (x,y,z), the magnitude z represents the load moved from time x to time y



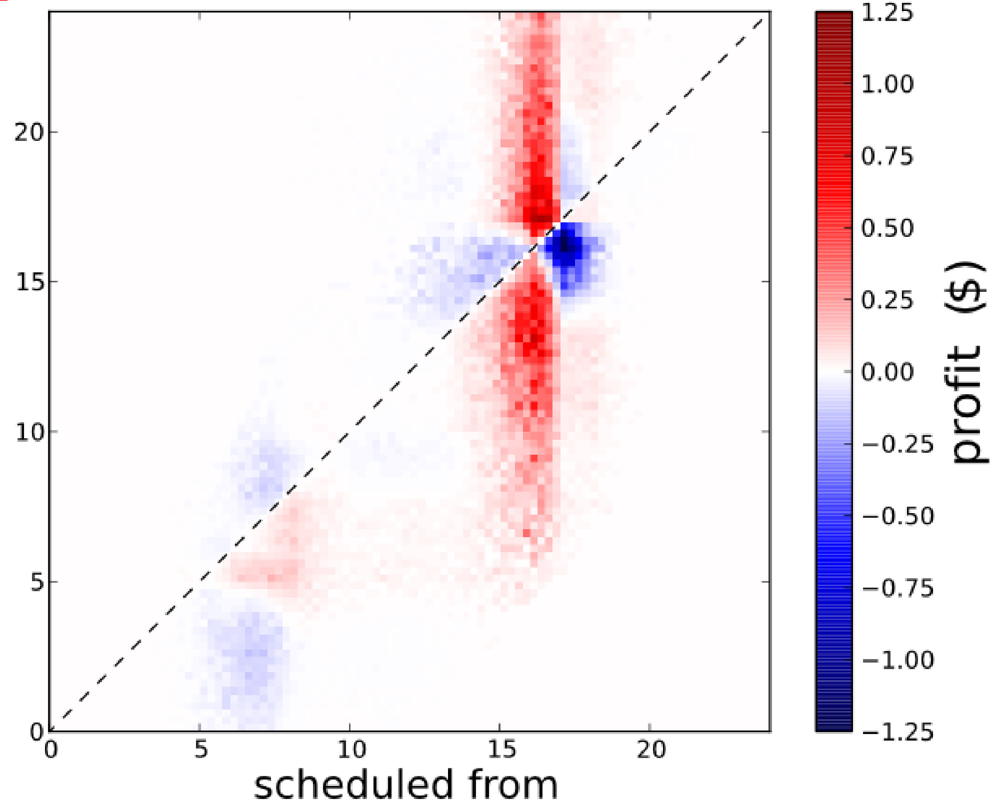
Spatially Distributed GIS Representation

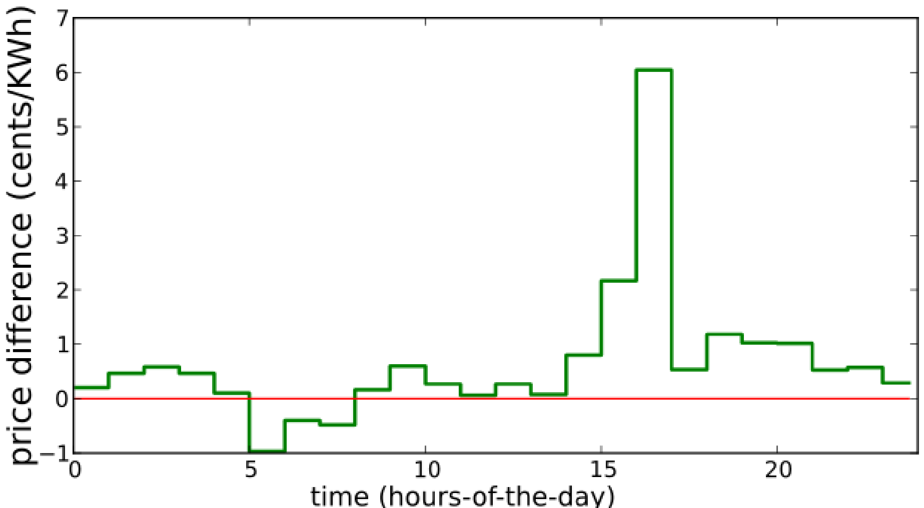
- mapped a distribution system spatially onto Fort Collins
- spread aggregator customers across system



Profit Heat Map

- shows the *difference* in profit between the forecast and actual real-time price
- recall:
  - *N* – income, rescheduled from
  - *B* – cost, scheduled to





Contributions

- design of two new temporal visualization techniques for DR that answer:
  - did the DR plan work effectively?
  - when did the DR entity make a profit or loss?
  - how does differences in the real-time price affect profit margin?
- creation of a spatial visualization technique for DR that answers *where* in the distribution system did the DR affect load
- discussion of how the methods can be used to analyze the effectiveness of DR optimization techniques

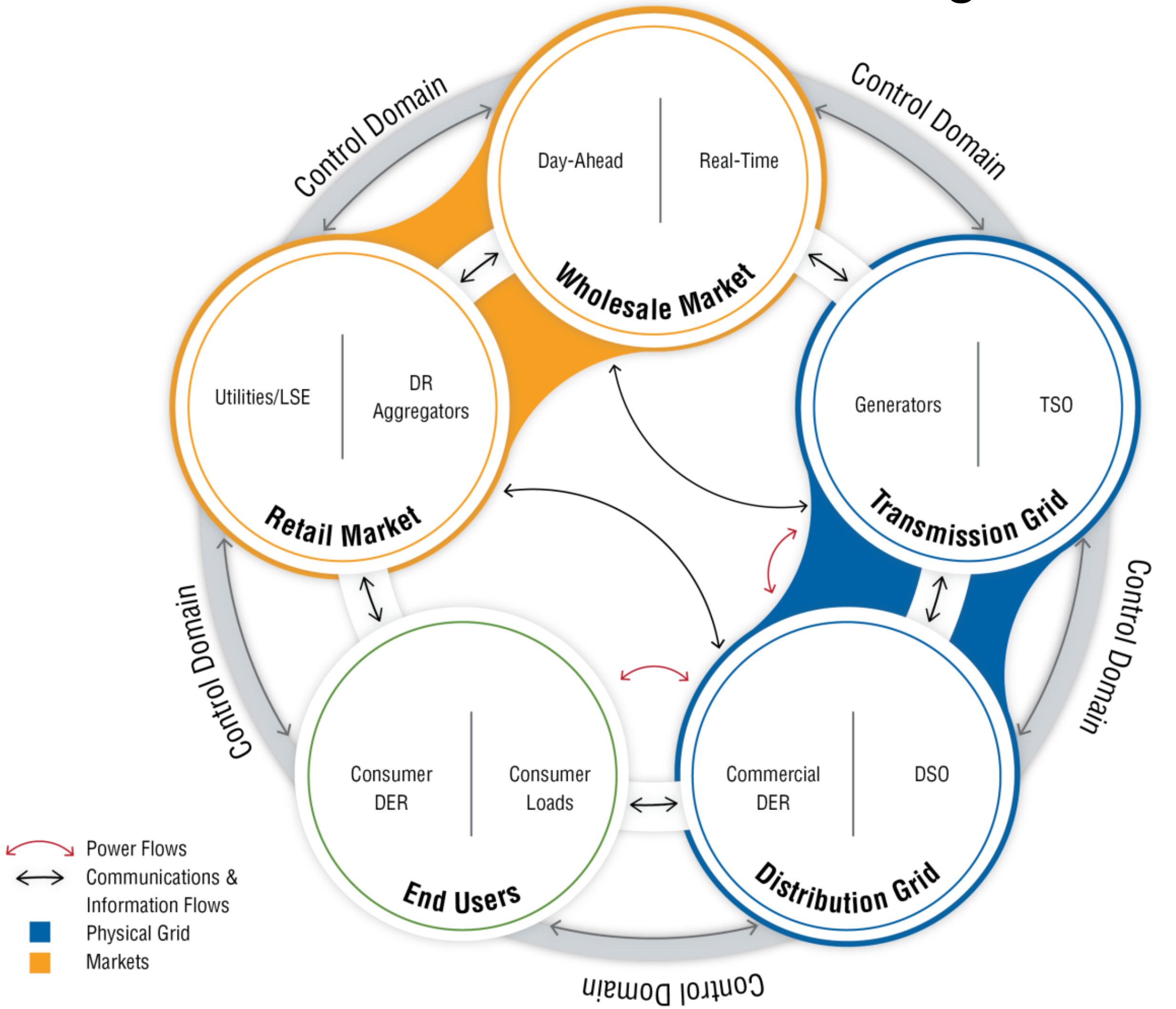
Outline – Co-simulation Framework

- introduction to resource allocation
- resource allocation in Smart Grid
  - background and motivation
  - non-myopic home energy management system
  - aggregator-based residential demand response
  - demand response visualization
  - **co-simulation framework**
- resource allocation in high-performance computing
- conclusions and future directions

Timothy M. Hansen et al., "Bus.py: A GridLAB-D Communication Interface for Smart Distribution Grid Simulations," *IEEE Power and Energy Society General Meeting 2015*, 5 pages, accepted 2015, to appear.

Co-simulation in Power Systems

- **co-simulation:** multiple individual tools, each specializing in a specific domain, interact while running simultaneously

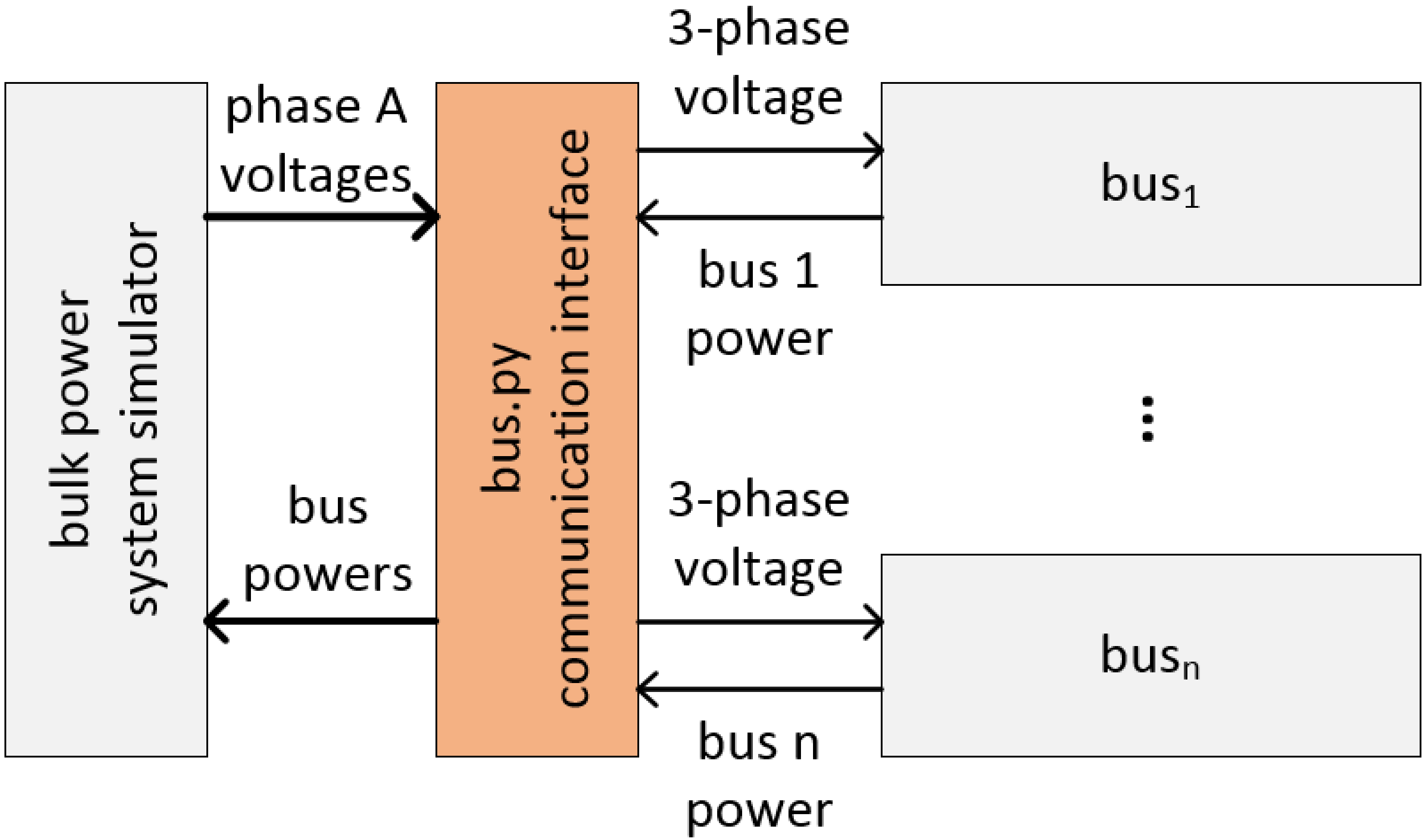


Bus.py

- introduce bus.py – a transmission-level bus simulator and communication interface

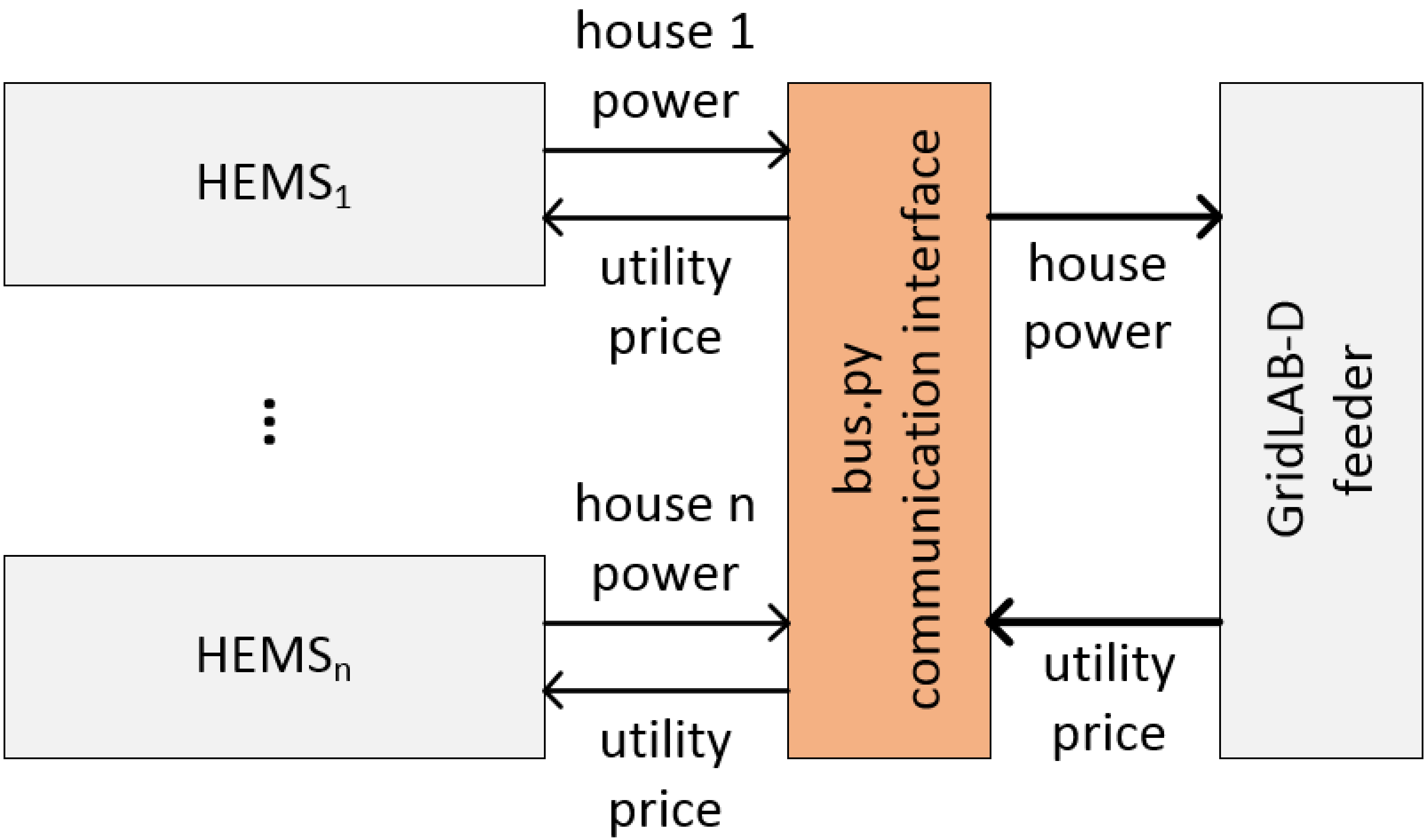
Bus.py

- introduce bus.py – a transmission-level bus simulator and communication interface
- enables co-simulation between:



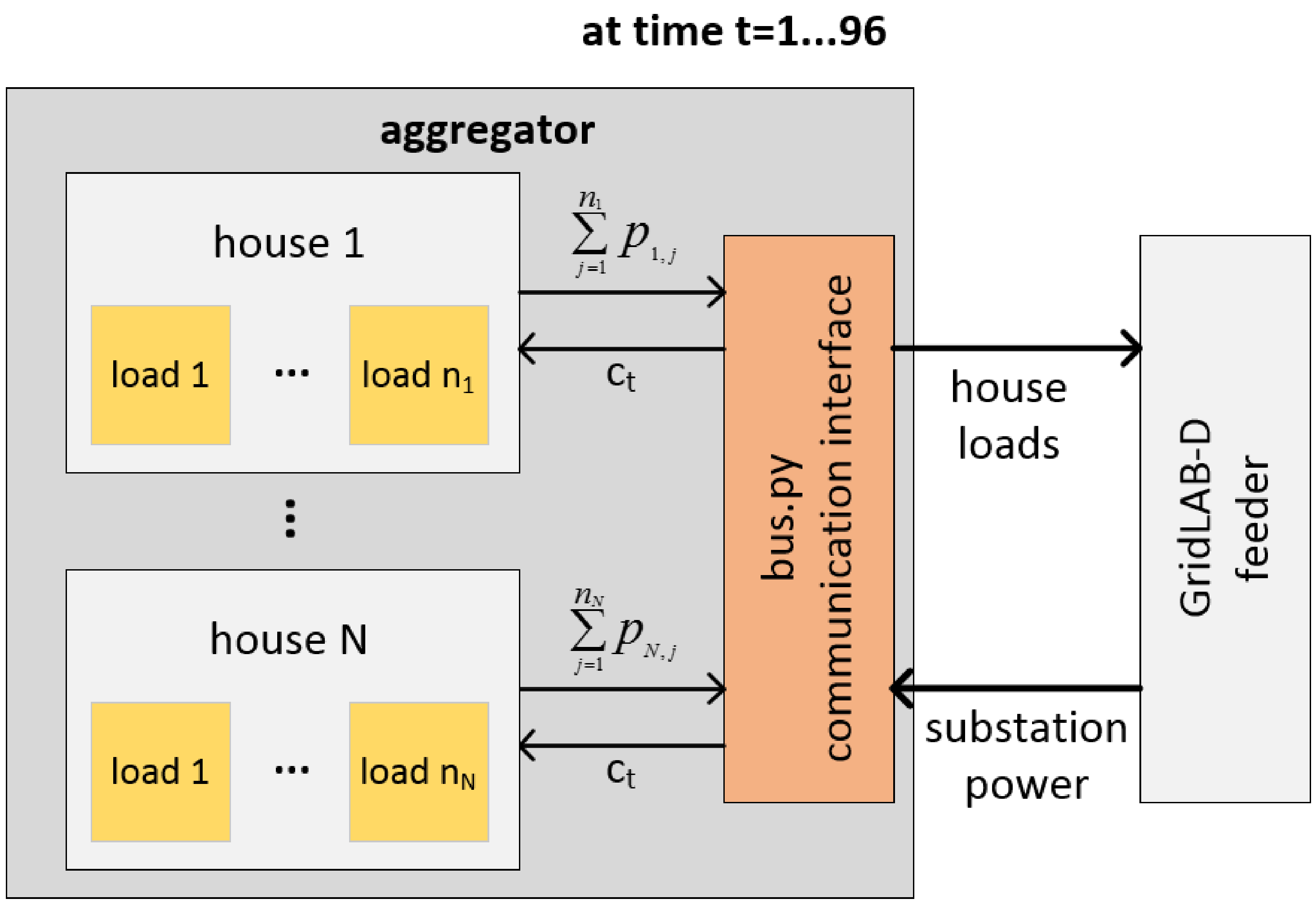
Bus.py

- introduce bus.py – a transmission-level bus simulator and communication interface
- enables co-simulation between:



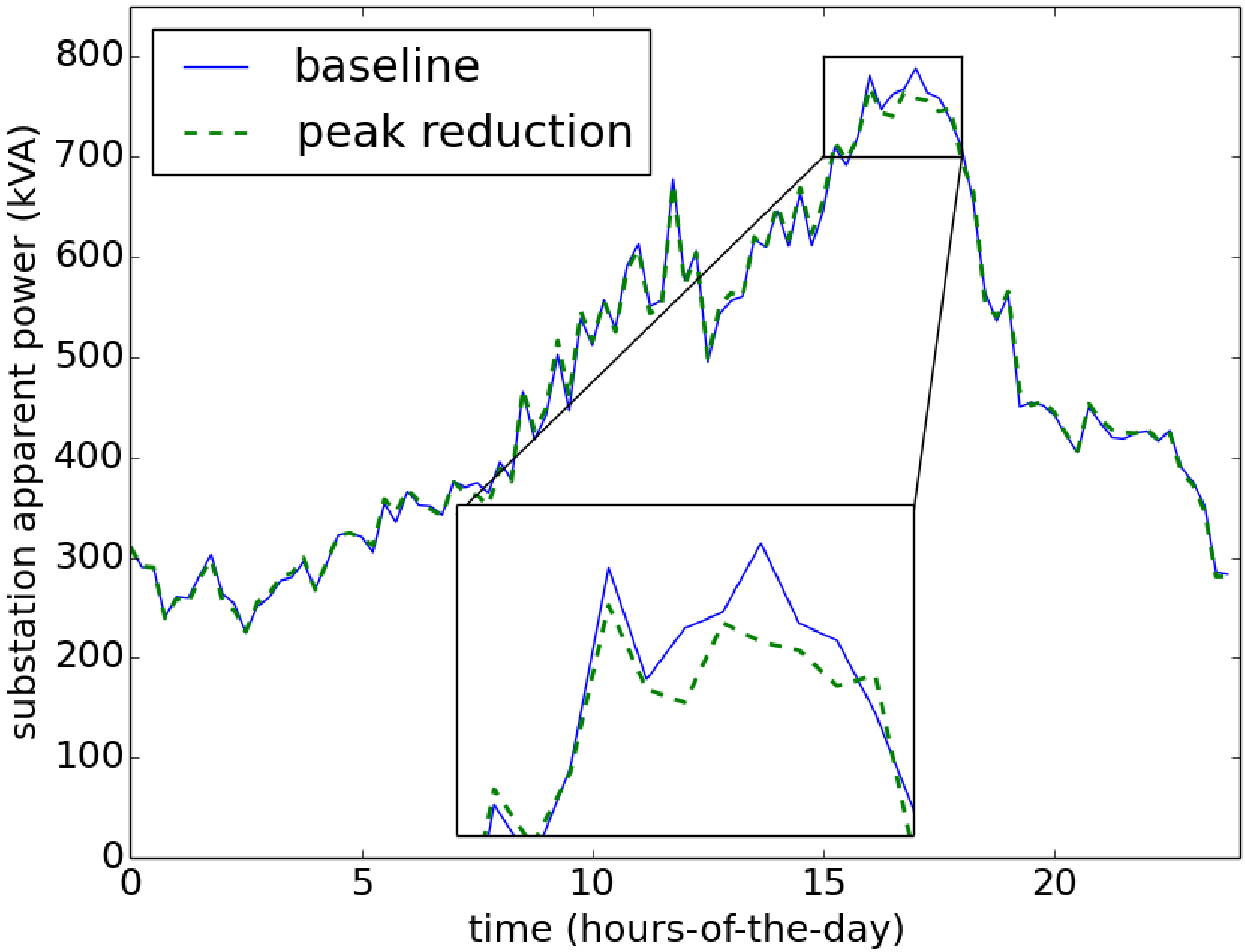
System Model

- aggregator controls the individual loads within a household
- each house is represented on a GridLAB-D distribution feeder
  - GridLAB-D – a PNNL distribution system simulator



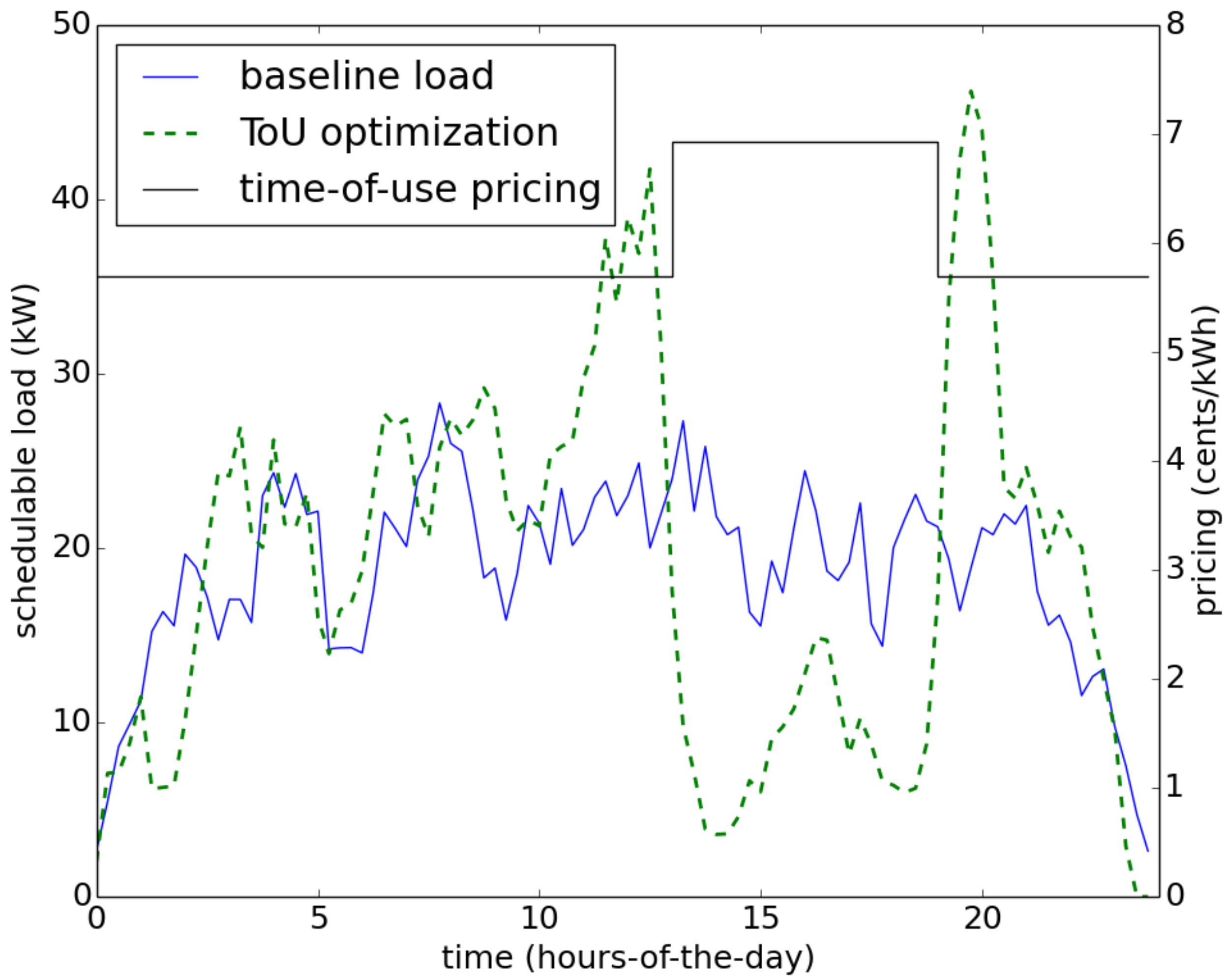
Peak Load Reduction

- reduction in peak load of 19.2 kVA (total available at peak time)



Cost Minimization in Time-of-Use Pricing

- change in schedulable load when minimizing cost in a time-of-use market



Contributions

- design of bus.py, a software transmission bus interface for use in Smart Grid co-simulation studies
- demonstration of bus.py interfacing with GridLAB-D simulating a small set of customers on a distribution feeder and an aggregator entity

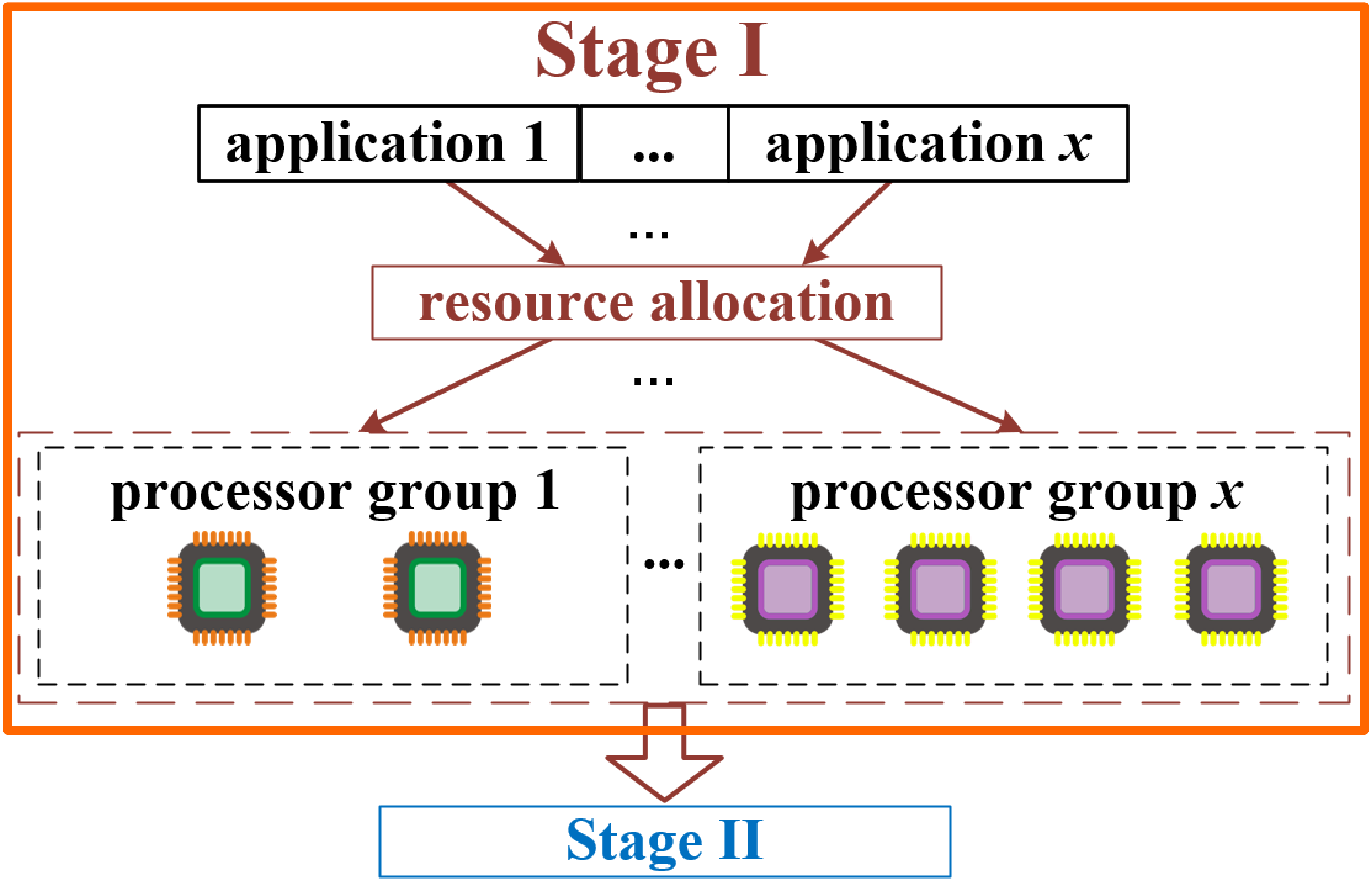
Outline – Resource Allocation in HPC

- introduction to resource allocation
- resource allocation in Smart Grid
- **resource allocation in high-performance computing**
  - **combined dual-stage resource allocation**
- conclusions and future directions

Timothy M. Hansen, Florina M. Ciorba et al., "Heuristics for Robust Allocation of Resources to Parallel Applications with Uncertain Execution Times in Heterogeneous Systems with Uncertain Availability," in *2014 International Conference on Parallel and Distributed Computing*, pp. 536-541, July 2014, **received the best paper award**.

Florina M. Ciorba, **Timothy M. Hansen** et al., "A Combined Dual-Stage Framework for Robust Scheduling of Scientific Applications in Heterogeneous Environments with Uncertain Availability," in *21st Heterogeneity in Computing Workshop*, pp. 187-200, May 2012.

Overview



Stage I
application 1
...
application x
...
resource allocation
...
processor group 1
...
processor group x
Stage II

Motivation

- obtain the most performance from a system
  given a batch of parallel scientific applications
  - i.e., maximize the investment in the computing infrastructure
- computing systems today are often heterogeneous in nature
- scheduling becomes more difficult with:
  - uncertain application execution times
    - e.g., varies based on input data
  - uncertain processor slowdown
    - e.g., system jitter, sharing of resources

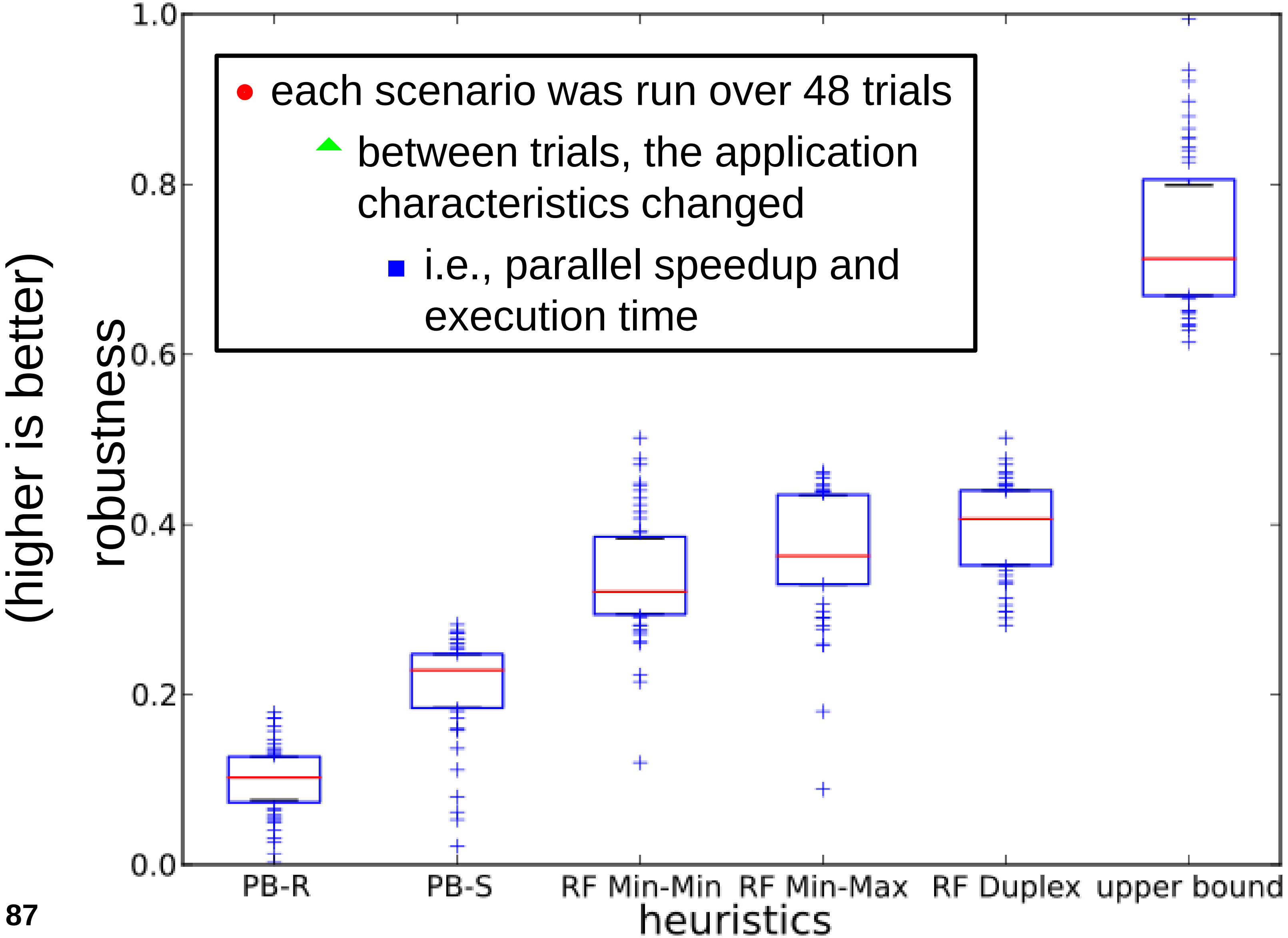
Colorado State University

Optimization Methods

- processor balance
  - random (PB-R)
  - smart (PB-S)
- robustness floor
  - min-min (RF Min-Min)
  - min-max (RF Min-Max)
  - duplex (RF Duplex)
- upper bound on robustness

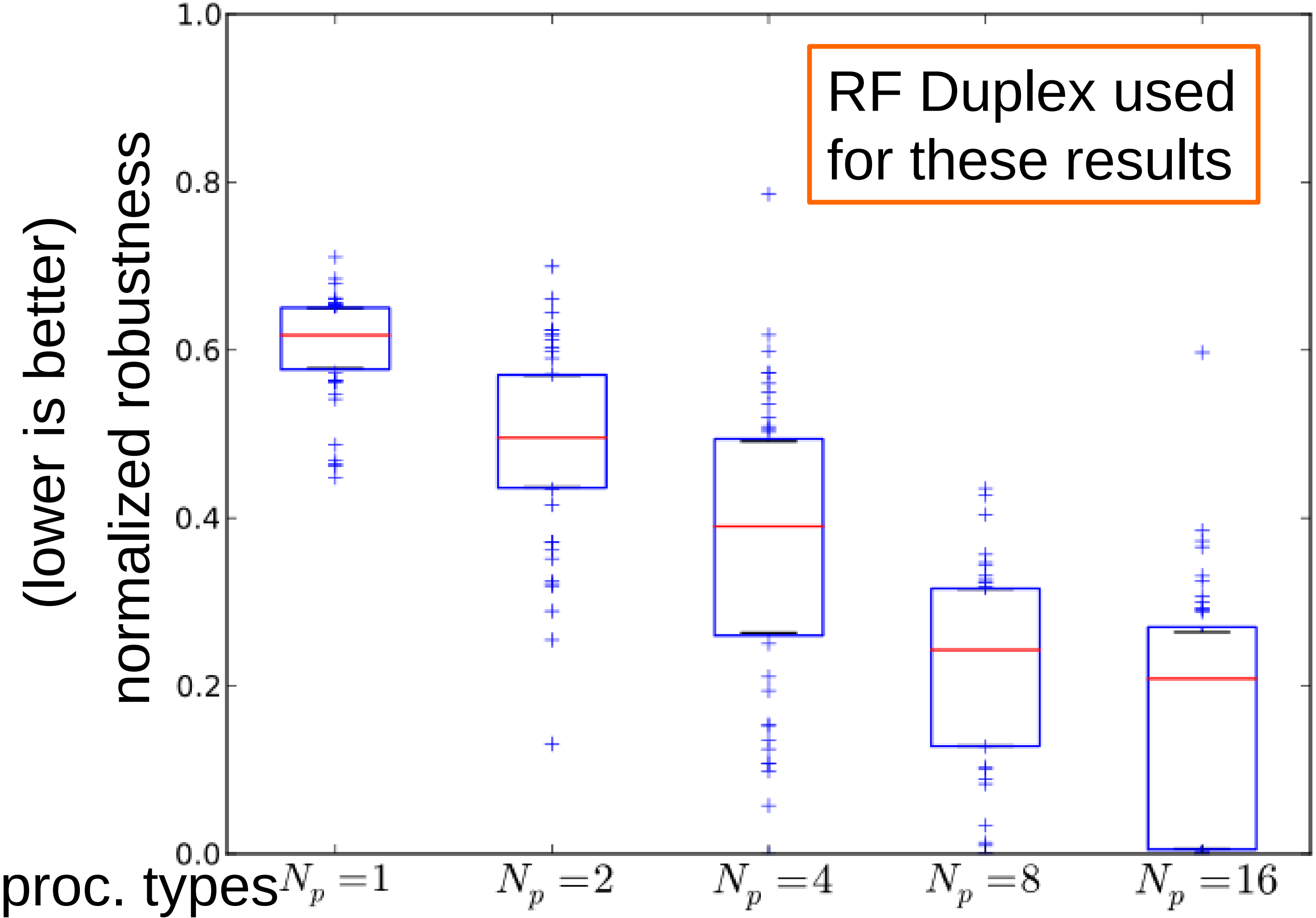
Typical Result

- 32 applications, 4 processor types, high system slowdown



Performance Trend – Number of Processor Types

- normalized robustness = (upper bound – robustness)/upper bound
  - a value of zero is better (i.e., equal to upper bound)
- 32 applications, low system slowdown



\# of proc. types

Contributions

- design of a model for moldable parallel applications with stochastic execution times running in a heterogeneous computing environment
- design of a new robustness metric that models the effect of two uncertainties
- design and analysis of three novel iterative-greedy heuristics that mitigate the effects of the two uncertainties

Outline – Conclusions

- introduction to resource allocation
- resource allocation in Smart Grid
- resource allocation in high-performance computing
- **conclusions and future directions**

Colorado State University

Conclusions: Resource Allocation in Two Domains

- resource allocation in Smart Grid
  - system-view with aggregator-based demand response
  - individual house view with POMDP HEMS
- resource allocation in HPC
  - dual-stage scheduling of applications to HPC systems

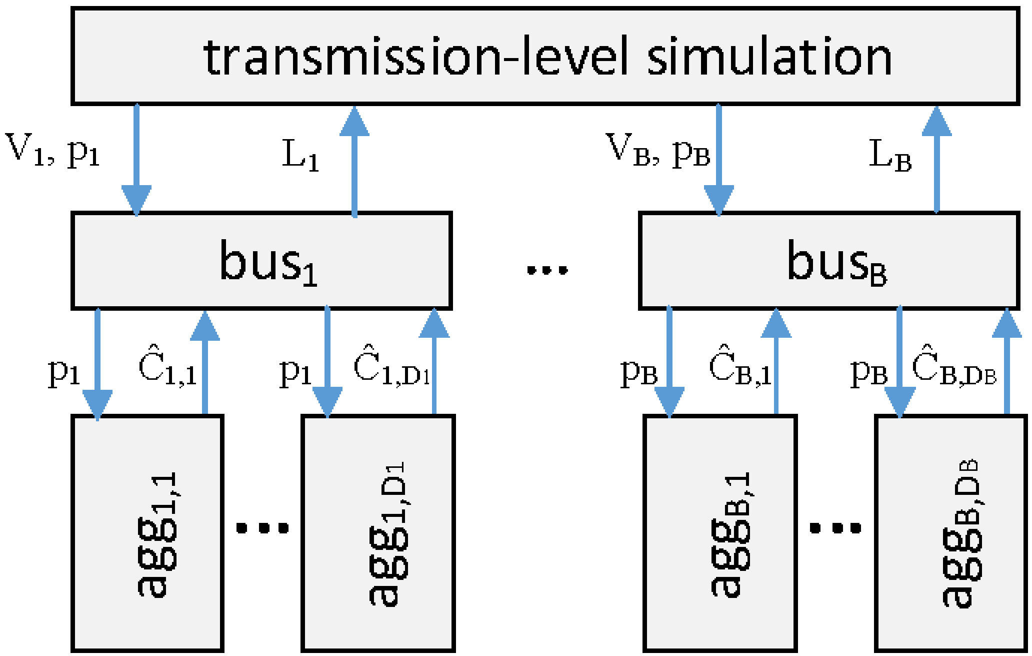
- the demand response methods were shown to reduce peak demand
- reduction in peak demand can:
  - reduce the cost of electricity
  - reduce the output of dirty diesel peaking generators
  - defer building new transmission lines

Future Directions

- combine the system and end-user approaches for a unique multi-level multi-scale end-to-end DR simulation
  - leverage the use of HPC
  - study market responses (as opposed to assuming an exogenous price)

Future Directions

- combine the system and end-user approaches for a unique multi-level multi-scale end-to-end DR simulation
  - leverage the use of HPC
  - study market responses (as opposed to assuming an exogenous price)



Future Directions

- combine the system and end-user approaches for a unique multi-level multi-scale end-to-end DR simulation
  - leverage the use of HPC
  - study market responses (as opposed to assuming an exogenous price)
- model more end-user assets

Future Directions

- combine the system and end-user approaches for a unique multi-level multi-scale end-to-end DR simulation
  - leverage the use of HPC
  - study market responses (as opposed to assuming an exogenous price)
- model more end-user assets
- study sustainability through the use of long-term simulations
  - new metrics such as reduction in capacity factor of dirty generators

Questions and Discussion

- collaborators:

- CV available: www.engr.colostate.edu/~hansentm