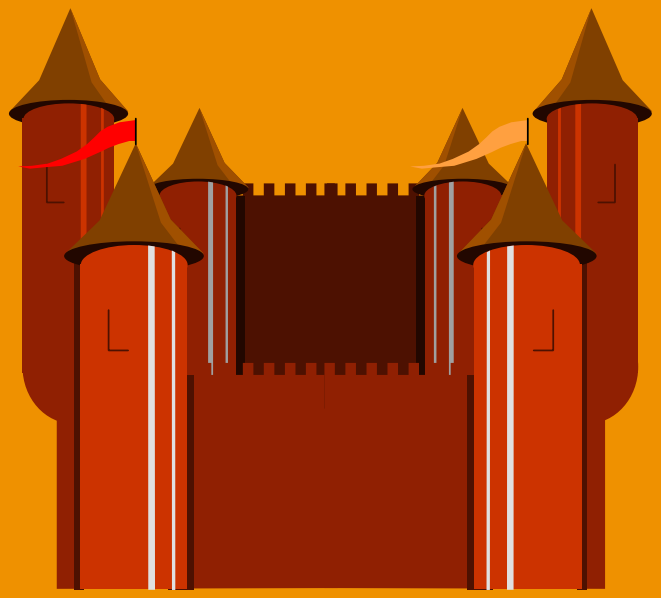
ORF 544

Stochastic Optimization and Learning

Spring, 2019



Warren Powell
Princeton University
http://www.castlelab.princeton.edu

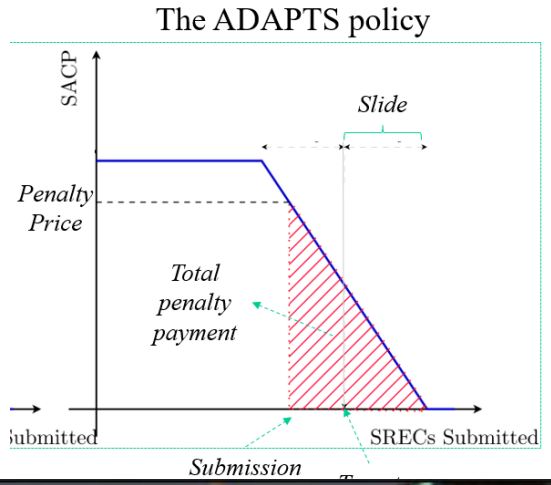
Week 1 - Monday

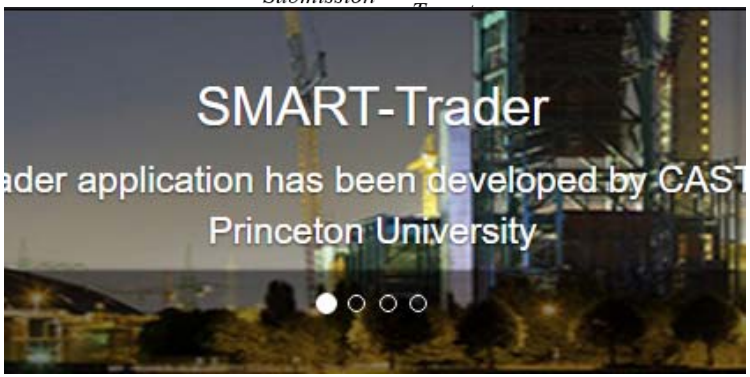
Introduction

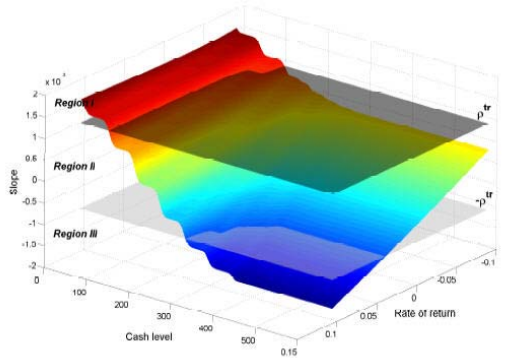
Sample applications

My experience

- Economics
  - » Analysis of SREC certificates for valuing solar energy credits.
  - » System for valuing 10-year forward contracts for electricity.
  - » Optimizing cash for mutual funds.



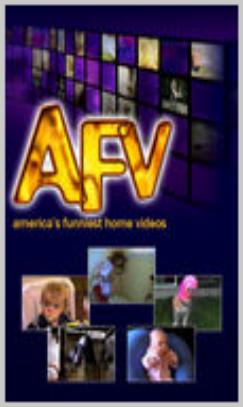




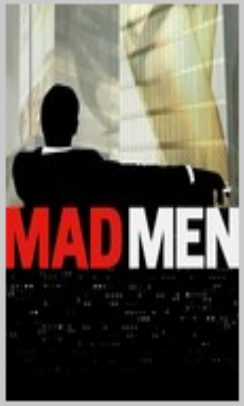
NETFLIX

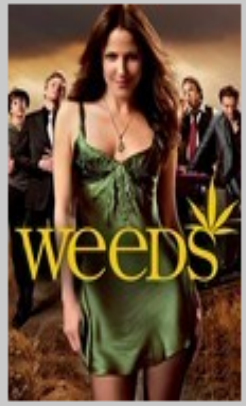
Member Sign In

New Arrivals in TV



TV Drama





TV Comedy

Learning the market in Africa

- How do African consumers respond to energy pricing strategies for recharging cell phones?
  - » Cell phone use is widespread in Africa, but the lack of a reliable power grid complicates recharging cell phone batteries.
  - » A low cost strategy is to encourage an entrepreneurial market to develop which sells energy from small, low cost solar panels.
  - » We do not know the demand curve, and we need to learn it as quickly as possible.

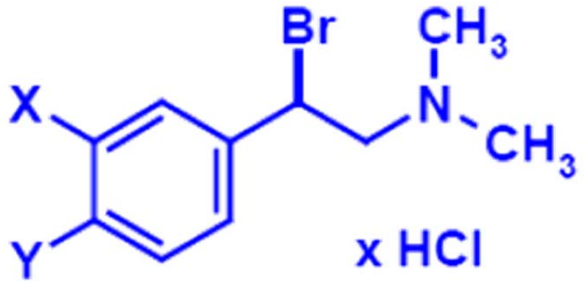
Cell Charging Challenges:
Using Optimal Learning to Determine the Profitability of a Solar Mobile Charging System in Africa

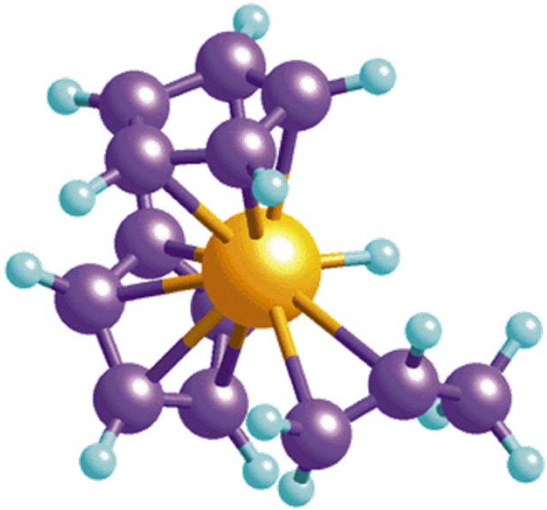
Interim Progress Report

Megan J. Wong
1/31/2011

Drug discovery

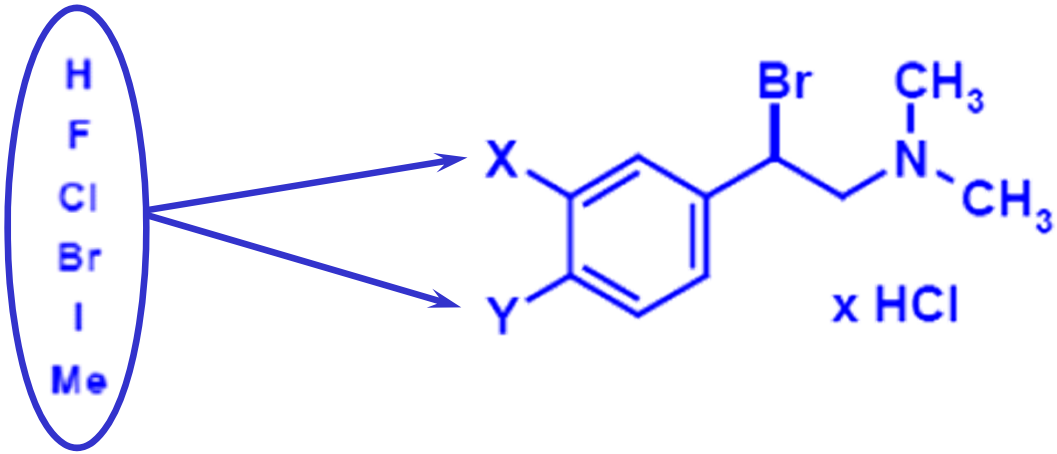
- Designing molecules





disubstituted N,N - Dimethyl-α-bromophenethylamines.

» X and Y are *sites* where we can hang *substituents* to change the behavior of the molecule



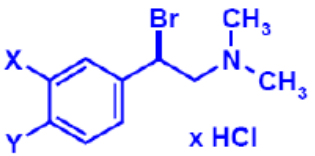
Drug discovery

- We express our belief using a linear, additive QSAR model
  - $X^m = \left(X_{ij}^m\right)_{ij}$ = Indicator variable for molecule *m*.
  - $Y = \theta_0 + \sum_{\text{sites } i} \sum_{\text{substituents } j} \theta_{ij} X_{ij}$

Hugo Kubinyi, www.kubinyi.de

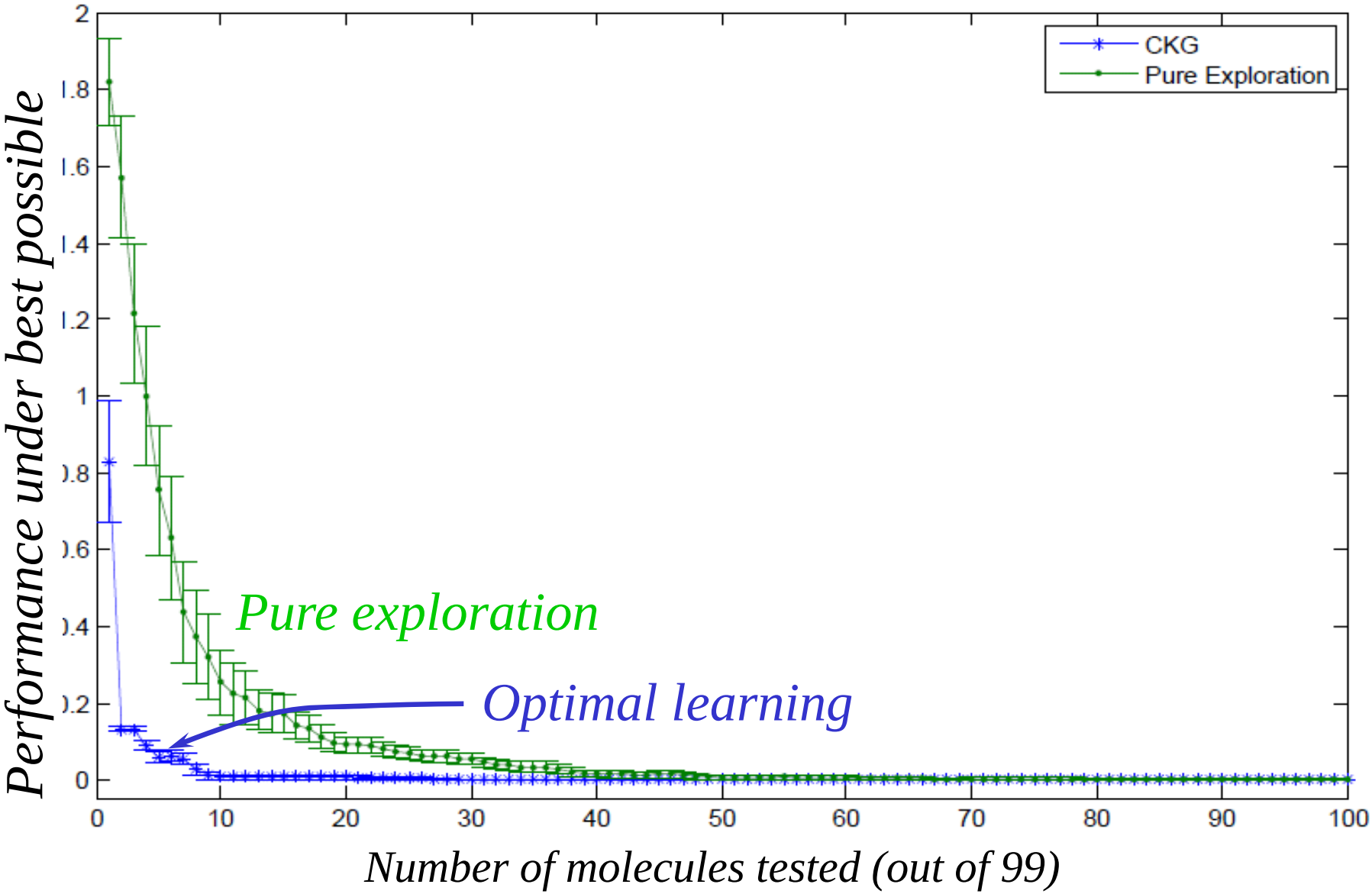
meta (X)	para (Y)	log 1/C obs.	meta- F	Cl	Br	I	Me	para- F	Cl	Br	I	Me	log 1/C calc.
H	H	7.46											7.82
H	F	8.16						1					8.16
H	Cl	8.68							1				8.59
H	Br	8.89								1			8.84
H	I	9.25									1		9.25
H	Me	9.30										1	9.08
F	H	7.52	1										7.52
Cl	H	8.16		1									8.03
Br	H	8.30			1								8.26
I	H	8.40				1							8.40
Me	H	8.46					1						8.28
Cl	F	8.19		1				1					8.37
Br	F	8.57			1			1					8.60
Me	F	8.82					1	1					8.62
Cl	Cl	8.89		1					1				8.80
Br	Cl	8.92			1				1				9.02
Me	Cl	8.96					1		1				9.04
Cl	Br	9.00		1						1			9.05
Br	Br	9.35			1					1			9.28
Me	Br	9.22					1			1			9.30
Me	Me	9.30					1					1	9.53
Br	Me	9.52			1							1	9.51

Matrix for Free Wilson Analysis



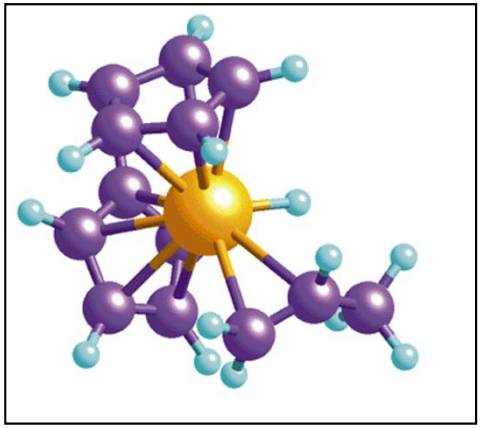
Drug discovery

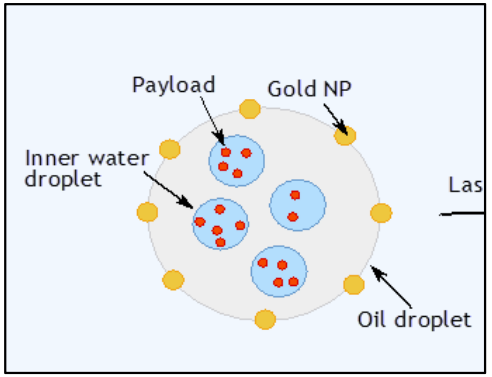
- Knowledge gradient versus pure exploration for 99 compounds

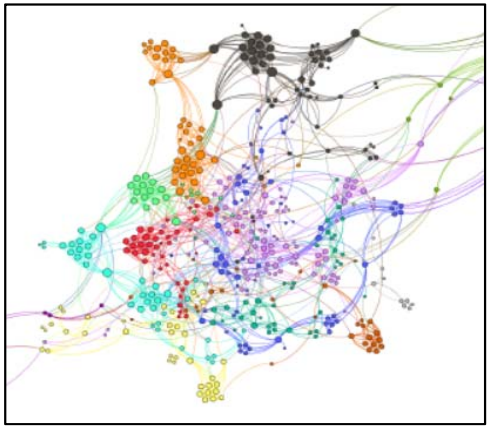


Health applications

- Health sciences
  - » Sequential design of experiments for drug discovery
  - » Drug delivery – Optimizing the design of protective membranes to control drug release
  - » Medical decision making – Optimal learning for medical treatments.

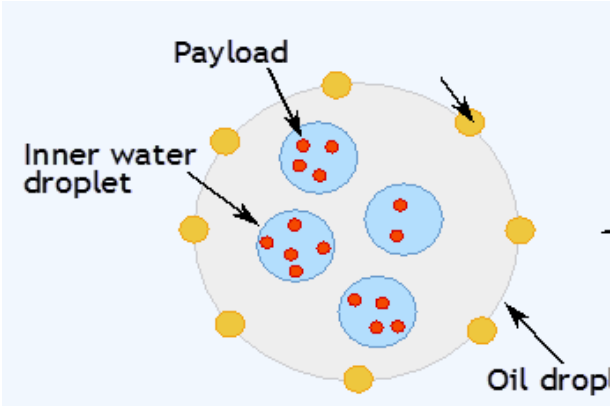






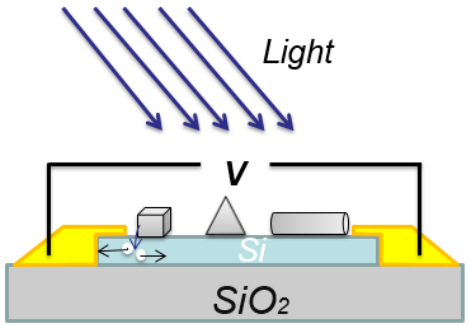
State-dependent applications

- Materials science
  - » Optimizing payloads: reactive species, biomolecules, fluorescent markers, …



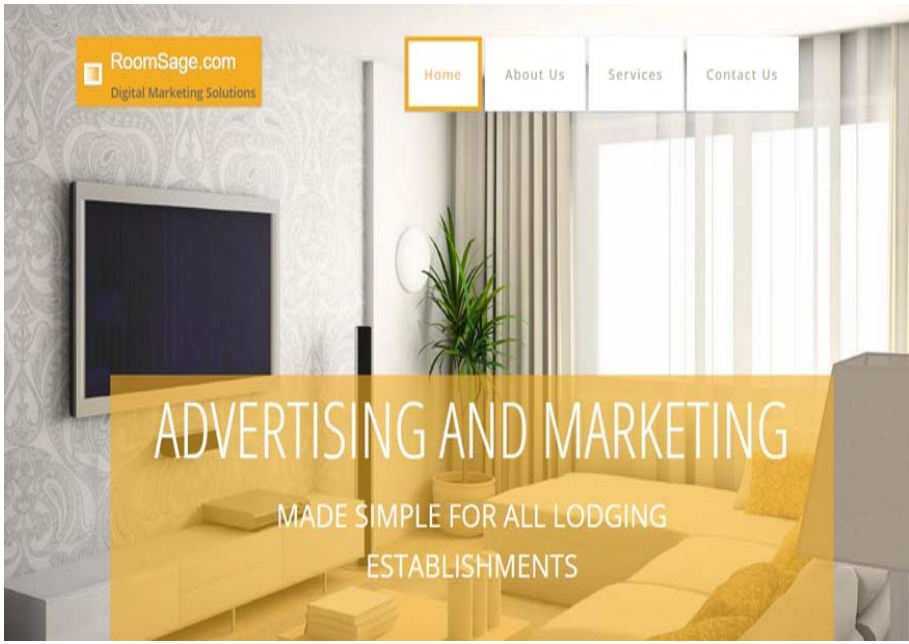
  - » Controllers for robotic scientist for materials science experiments

  - » Optimizing nanoparticles to maximize photoconductivity



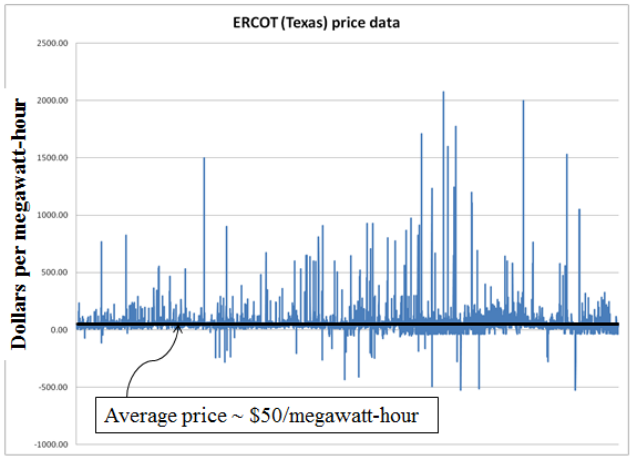
E-commerce

- Revenue management
  - » Optimizing prices to maximize total revenue for a particulate night in a hotel.
- Ad-click optimization
  - » How much to bid for ads on the internet.
- Personalized offer optimization
  - » Designing offers for individual customers.



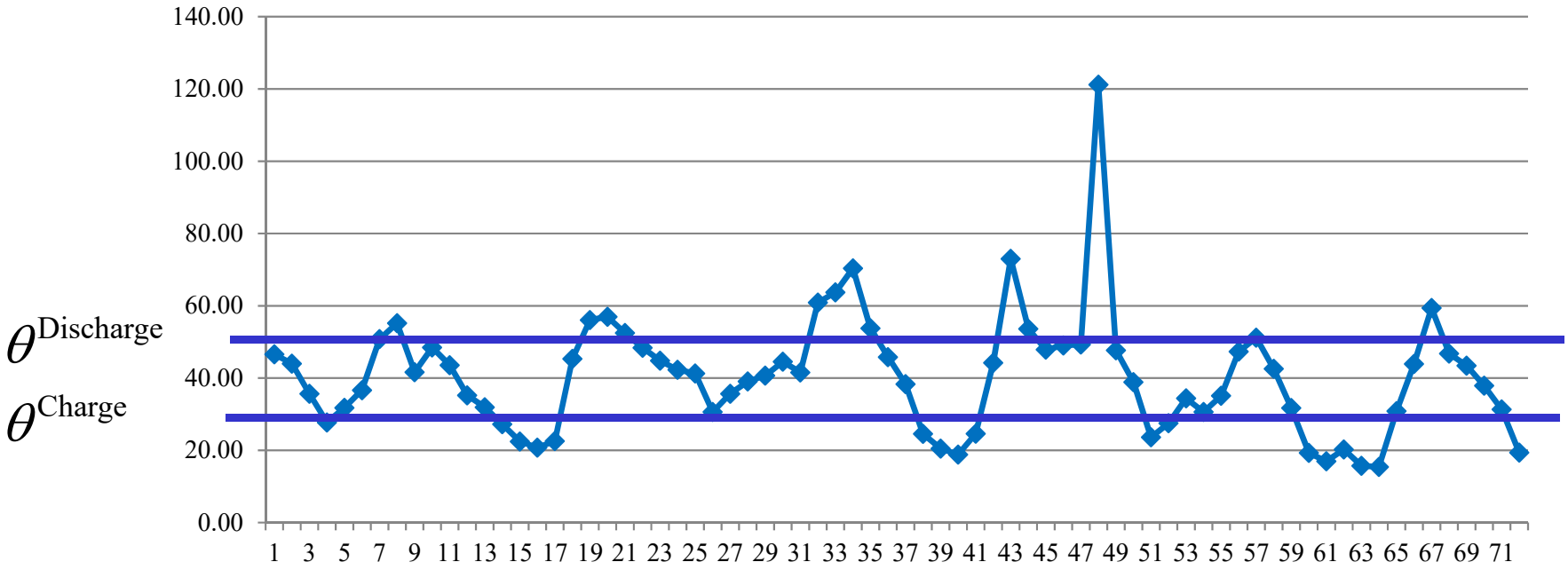
Policy function approximations

- Battery arbitrage – When to charge, when to discharge, given volatile LMPs



Policy function approximations

- Grid operators require that batteries bid charge and discharge prices, an hour in advance.

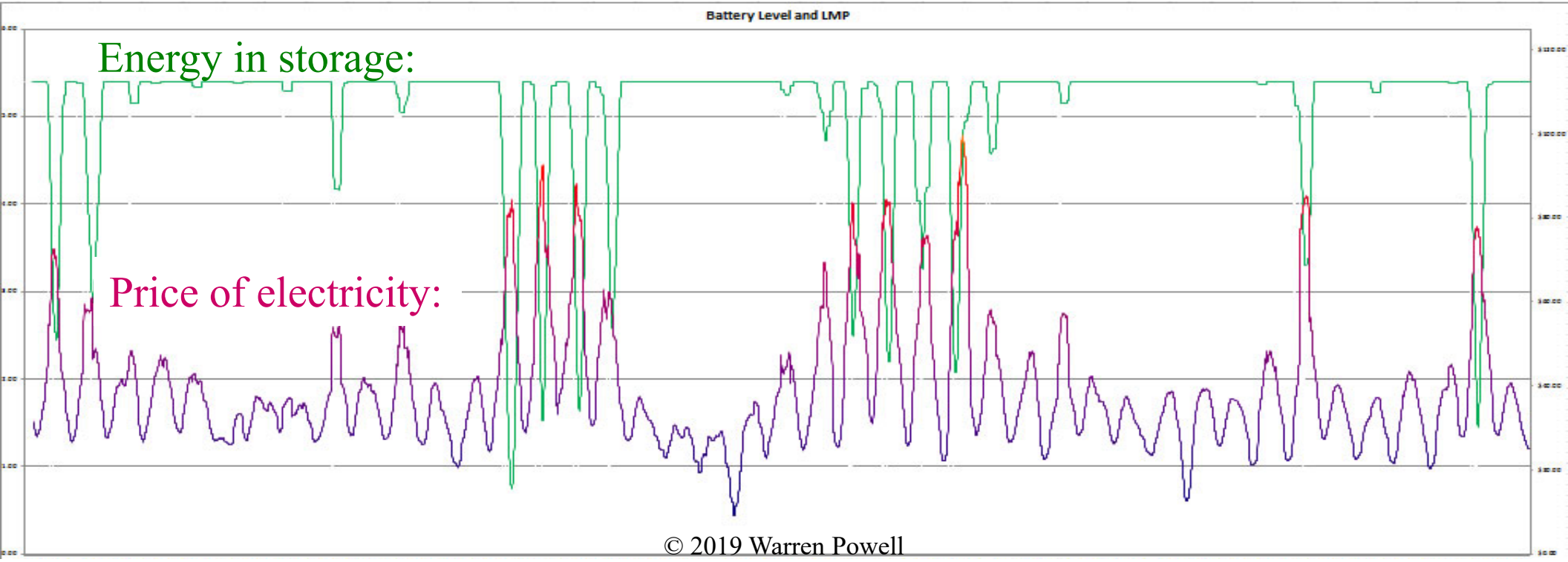


- We have to search for the best values for the policy parameters θ^{Charge} and $\theta^{\text{Discharge}}$.

Policy function approximations

- Our policy function might be the parametric model (this is nonlinear in the parameters):

$$X^{\pi}(S_t \mid \theta) = \begin{cases} +1 & \text{if } p_t < \theta^{\text{charge}} \\ 0 & \text{if } \theta^{\text{charge}} < p_t < \theta^{\text{discharge}} \\ -1 & \text{if } p_t > \theta^{\text{charge}} \end{cases}$$

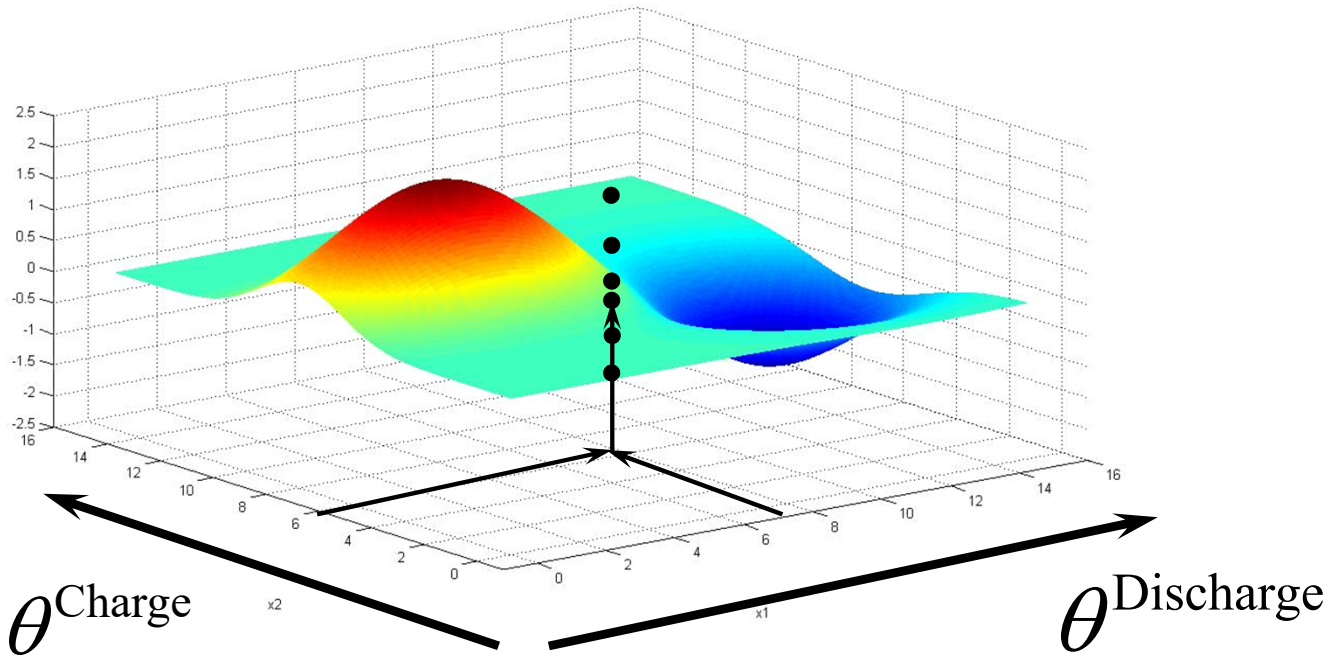


Policy function approximations

- Finding the best policy
  - » We need to maximize

$$\max_\theta F(\theta) = \mathbb{E}\sum_{t=0}^{T}\gamma^t C\left(S_t, X_t^\pi(S_t \mid \theta)\right)$$

  - » We cannot compute the expectation, so we run simulations:



Energy storage

- How much energy to store in a battery to handle the volatility of wind and spot prices to meet demands?

Planning cash reserves

- How much money should we hold in cash given variable market returns and interest rates to meet the needs of a business?

Stock prices

Bonds

The Big Issues Facing Fed Chair Janet Yellen

Janet Yellen faces some knotty questions as she prepares to take the helm of the Federal Reserve. The answers aren't academic

by Peter Coy
petercoy

from BloombergBusinessweek

January 9, 2014 – 6:11 AM EST

On Feb. 1, Janet Yellen will become the chairman of the Federal Reserve. Considering that the Federal Reserve System just celebrated its 100th birthday and employs more Ph.D. economists than any other institution on earth, you'd think it would have monetary policy down pat. It doesn't. Yellen herself acknowledges that setting interest rates sometimes involves groping in the dark. "I consider it essential, in making judgments about the stance of policy, to recognize at the outset the limits of our understanding regarding the dynamics of the economy and the transmission of monetary policy," she said in a 2012 speech to the Money Marketeers of New York University.

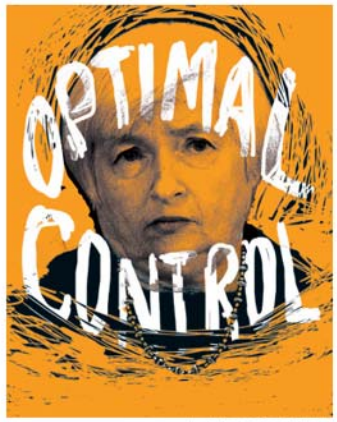


Illustration by 731; AP Photo(2)

Yellen is wise to be humble, as just about every issue she will confront is fraught with uncertainty. Is the U.S. economy stagnating, or threatening to overheat? How quickly should the Fed taper its purchases of long-term Treasury bonds and mortgage-backed securities? When should it begin to raise the federal funds rate, which has been nailed to the floor at zero to 0.25 percent since the end of 2008?

These debates are typically viewed through the familiar hawk-vs.-dove monetary prism, but they are deeper and more interesting. Here's a guide to three of the Big Questions that will keep Madame Chairman occupied in the years ahead:

Optimal Control

Lev Pontryagin, born in Moscow in 1908, was 14 years old when his family's kerosene stove exploded, blinding him, yet he became one of the greatest mathematicians of th 20th century. Among other things, he built the foundation for optimal control theory, which explains how to maximize the effectiveness of a process given constraints on c or time. In 1953 the American Richard Bellman developed a variant called dynamic programming that became a mainstay of science and engineering—it was even used f landing the Apollo Lunar Module.

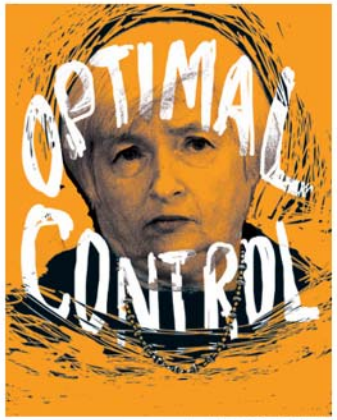


Illustration by 731; AP Photo(2)

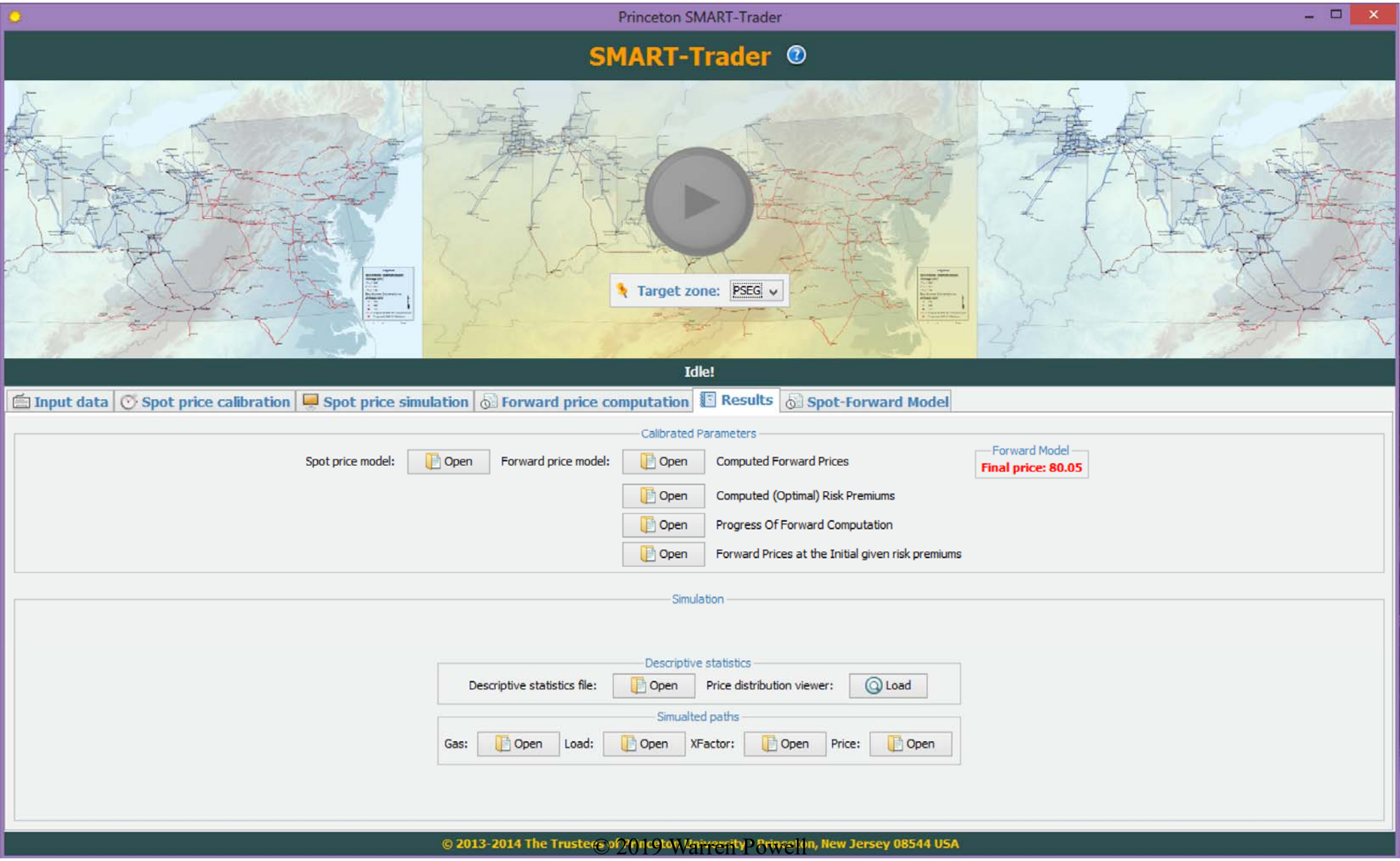
Yellen has repeatedly expressed fascination with the possibility that the Fed could "optimally control" the U.S. economy by raising and lowering interest rates in a more scientific manner. Under the optimal control approach, Fed economists would use a macroeconomic model to calculate the mathematically ideal path of short-term inter rates needed to hit established inflation and unemployment targets.

If Yellen started channeling Pontryagin and got the rest of the Federal Open Market Committee to buy in, short-term interest rates could stay near zero longer than anyon is expecting. In her 2012 talk to the Money Marketeers, she flashed a "purely illustrative" graph of what the economy might look like under optimal control. The Fed would keep the funds rate low longer, and unemployment would fall faster. Inflation would rise slightly above the Fed's target, but only for a few years. Overall, that's much better than the current outlook.

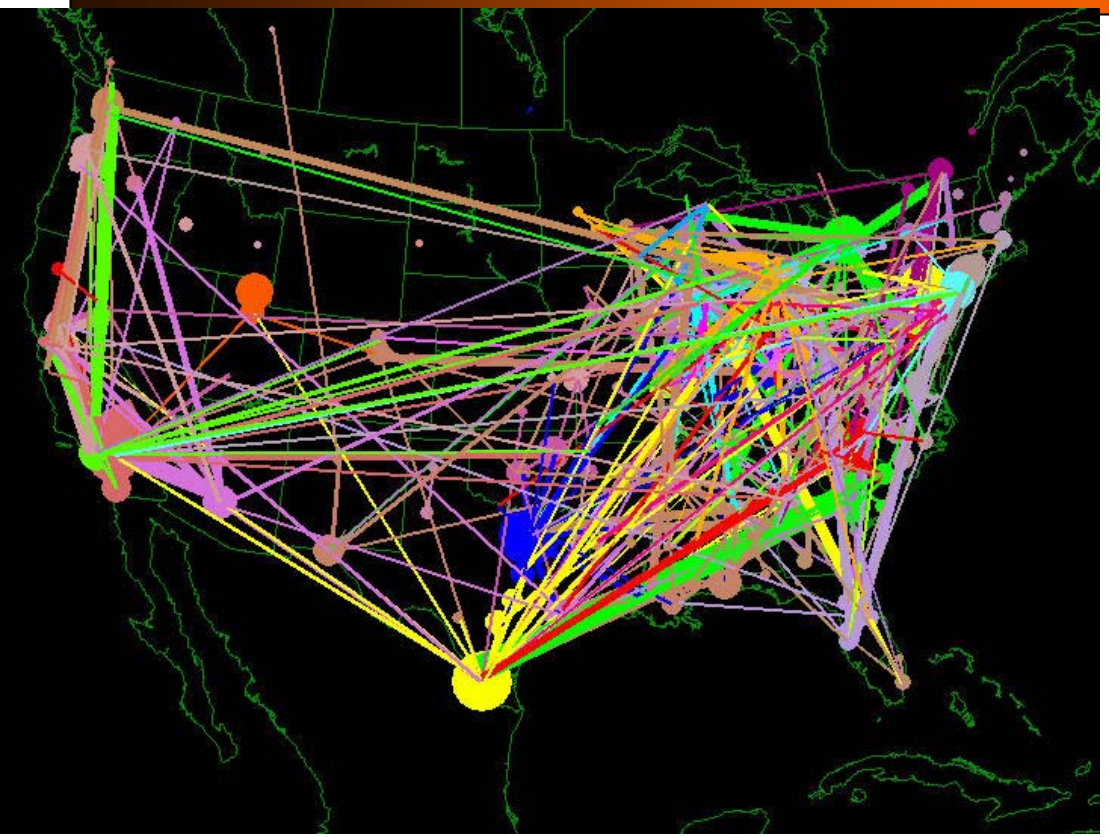
There is a catch, and Yellen is the first to admit it. Optimal control assumes that the Fed has perfect foresight and flawless data about the economy, and also that inflationary psychology never takes hold, because businesspeople and consumers unfailingly trust the Fed to keep prices under control. That's not realistic, of course. So Yellen in her 2012 presentation also considered a "simple rule" that prescribes an interest rate based on nothing more than available data about the divergence of inflation and economic output from their targets. (This is the Taylor Rule, named after conservative economist John Taylor of the Hoover Institution.) The Taylor Rule is more straightforward but has its own drawback, Yellen told the Marketeers: After periods of extreme weakness like the past few years, it calls for rates to rise too much, too soon.

In other words, you can have a rate-setting rule that's very good but unreliable, or one that's reliable but not very good. Some choice. "A dose of good judgment will always be essential as well," Yellen concluded. That's also known as trusting your gut.

Electricity forward contracts



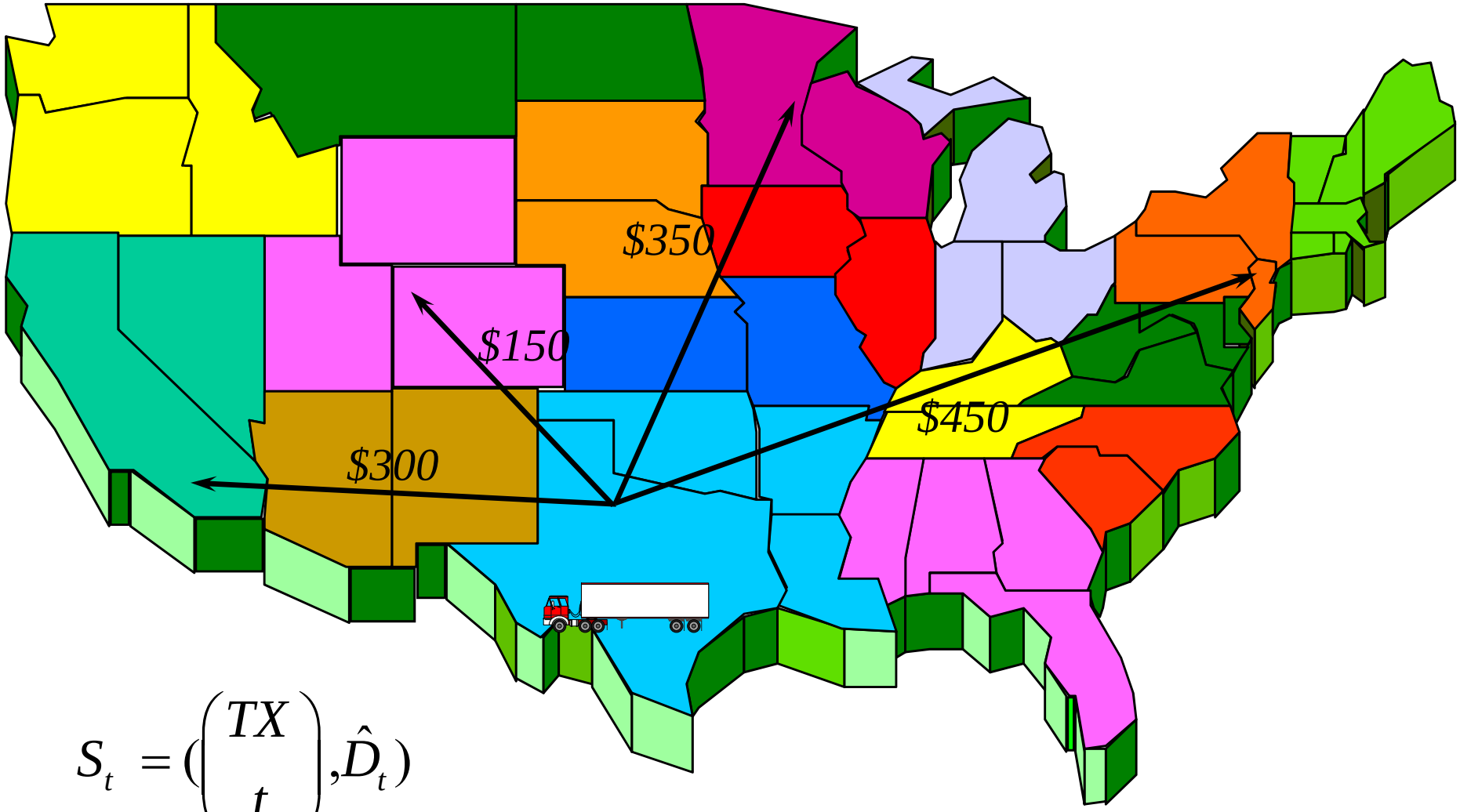
Fleet management



- Fleet management problem
  - » Optimize the assignment of drivers to loads over time.
  - » Tremendous uncertainty in loads being called in

Nomadic trucker illustration

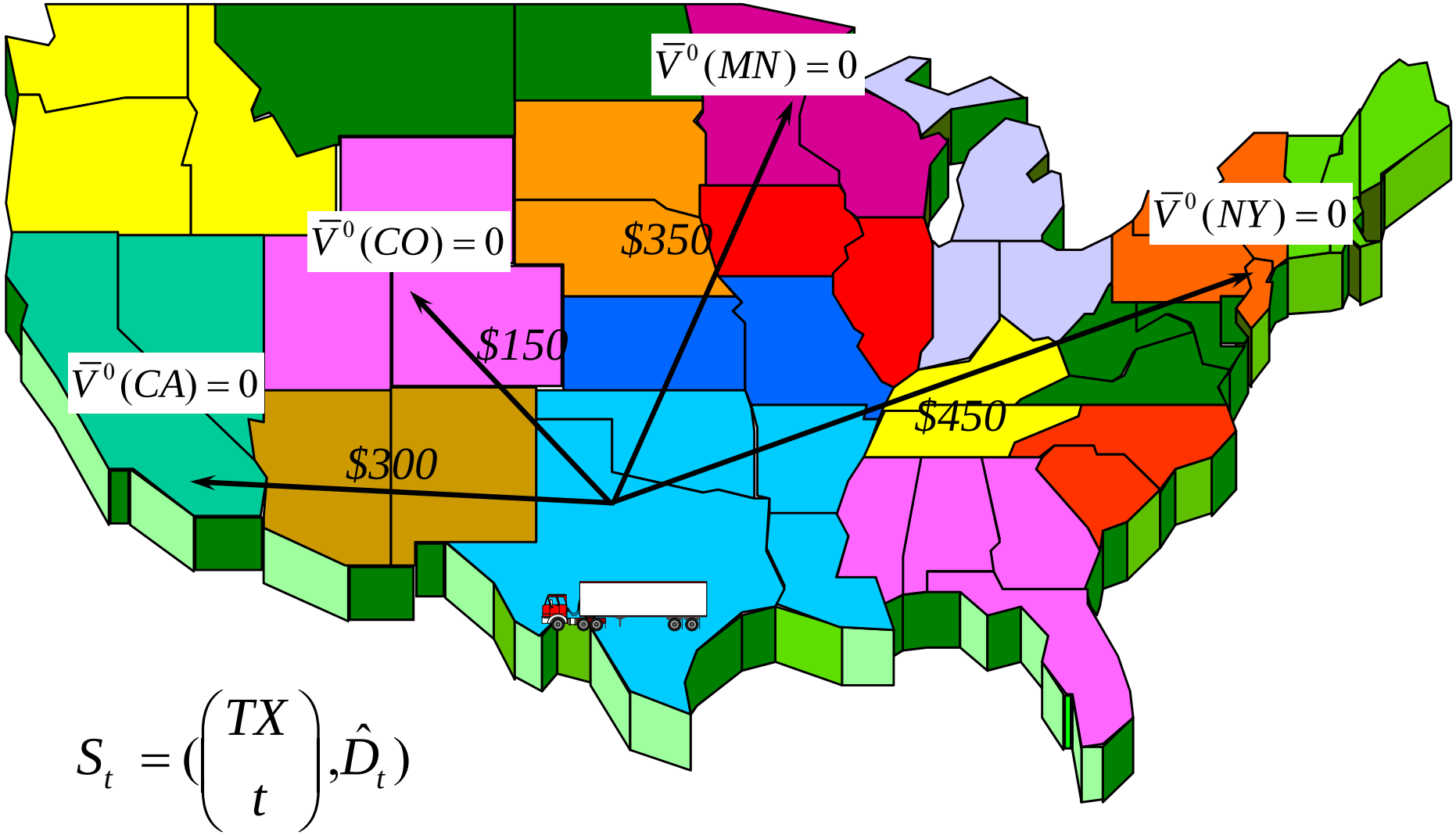
- Pre-decision state: we see the demands



$$S_t = \left(\begin{pmatrix} TX \\ t \end{pmatrix}, \hat{D}_t\right)$$

Nomadic trucker illustration

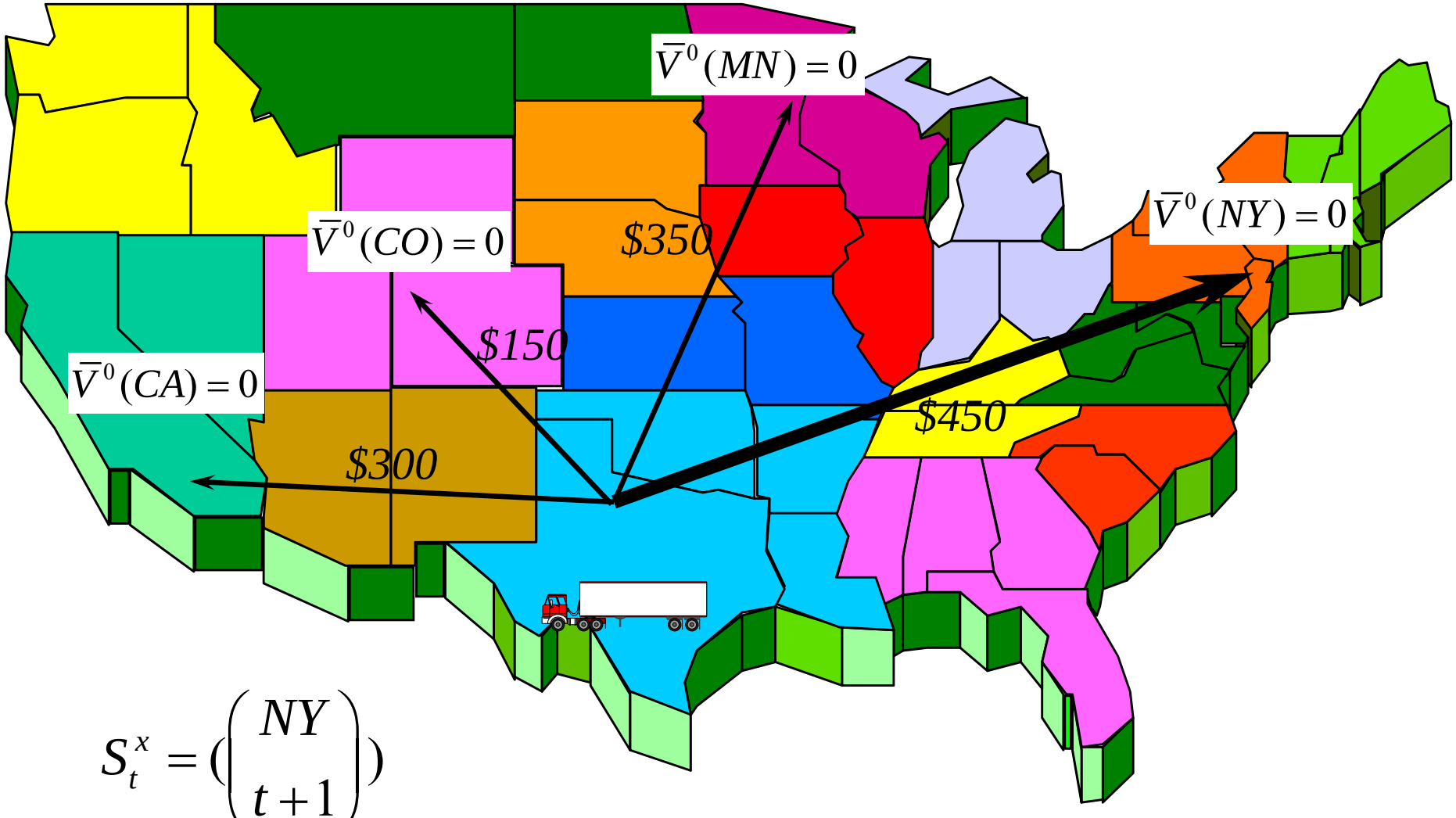
- We use initial value function approximations…



$$S_t = \left(\begin{pmatrix} TX \\ t \end{pmatrix}, \hat{D}_t\right)$$

Nomadic trucker illustration

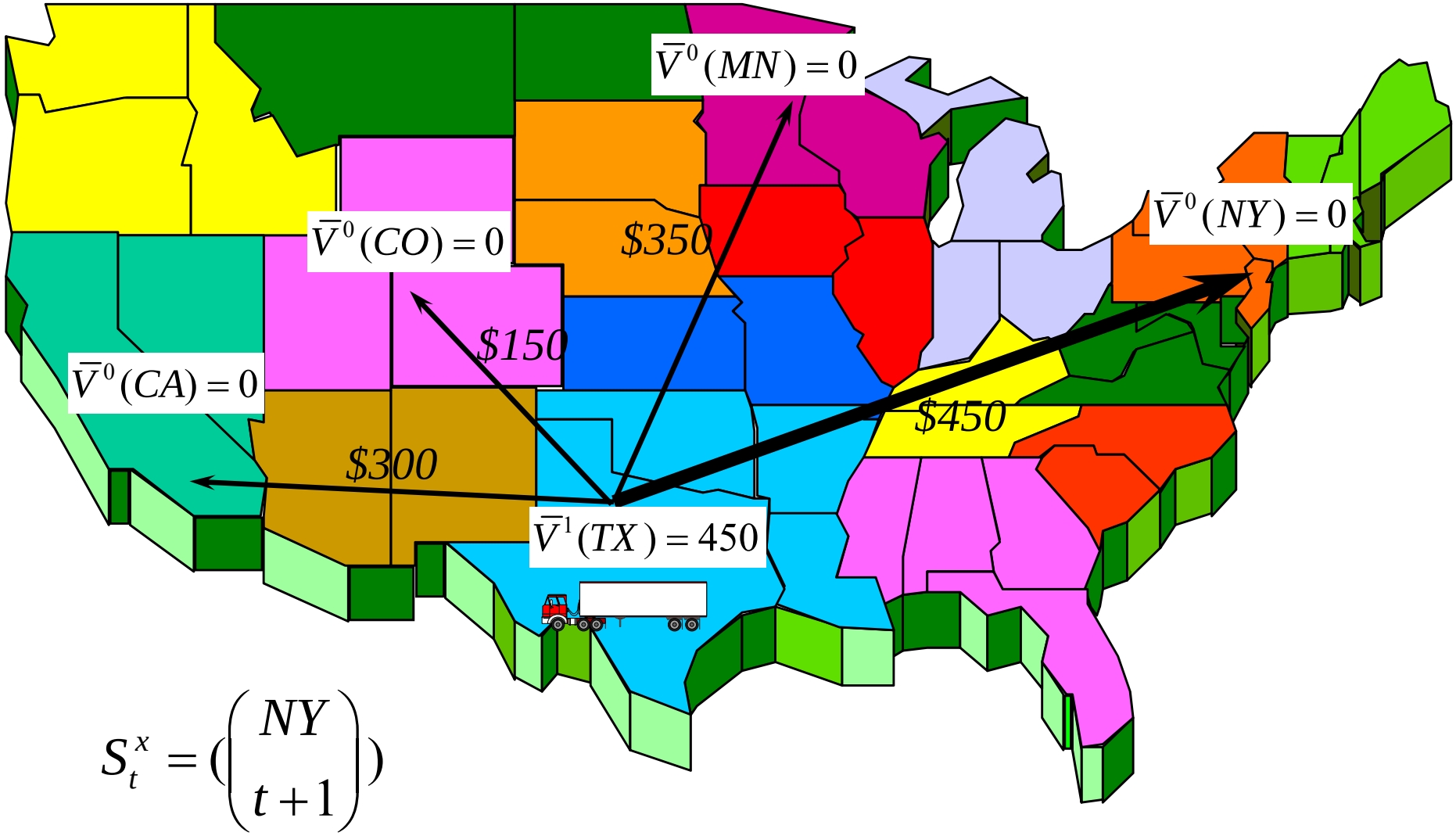
- … and make our first choice: x^1



$$S_t^x = \begin{pmatrix} NY \\ t+1 \end{pmatrix}$$

Nomadic trucker illustration

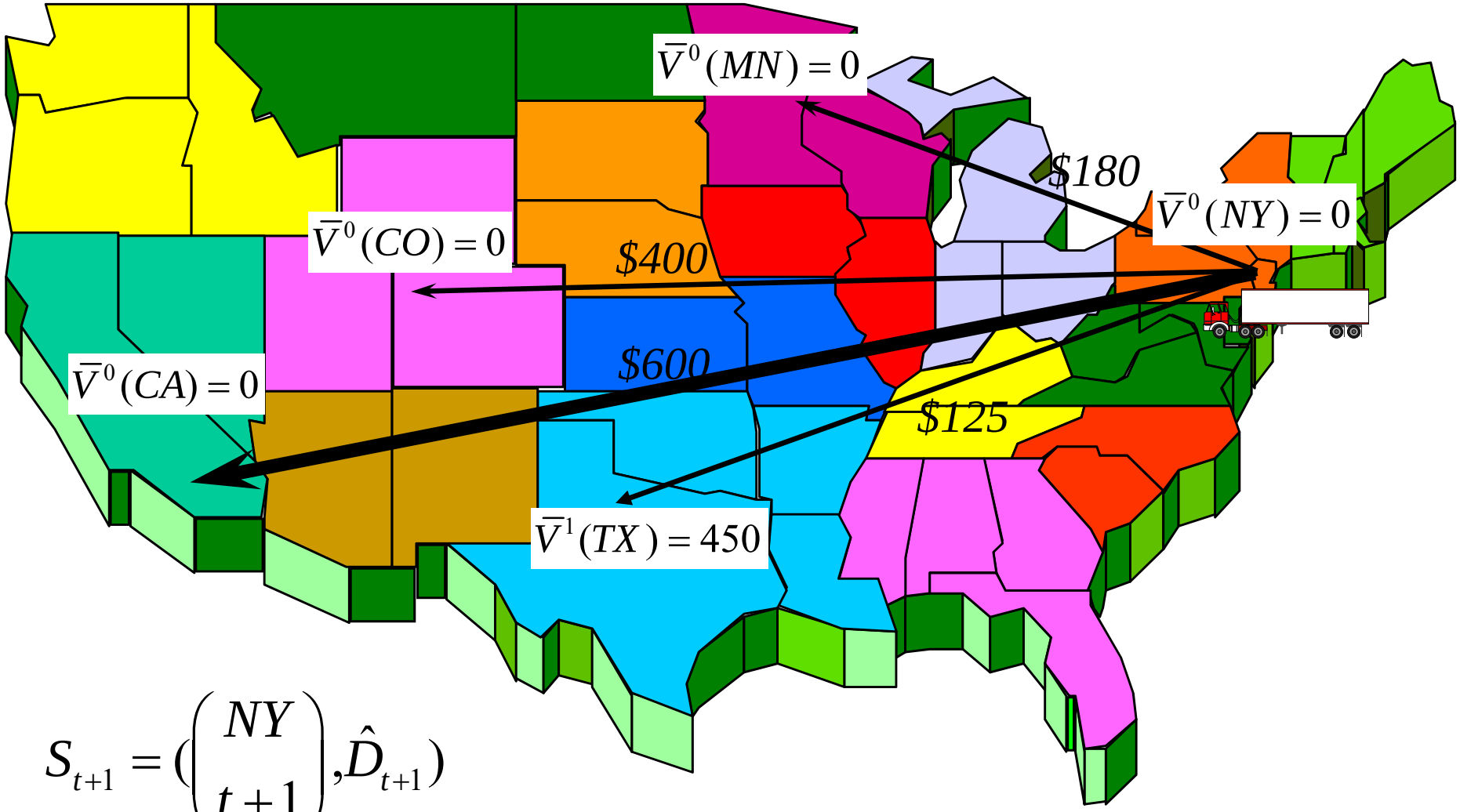
- Update the value of being in Texas.



$$S_t^x = \begin{pmatrix} NY \\ t+1 \end{pmatrix}$$

Nomadic trucker illustration

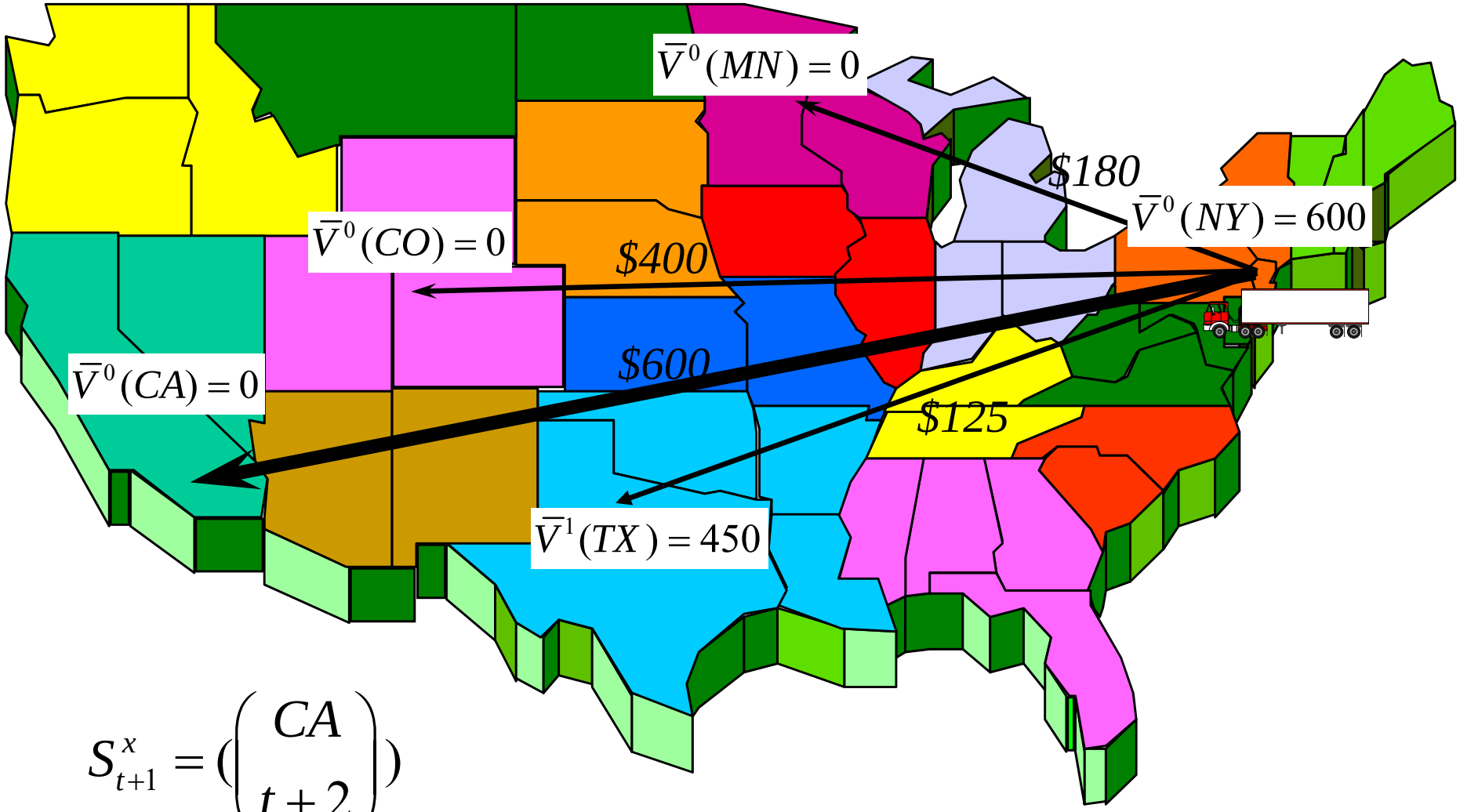
- Now move to the next state, sample new demands and make a new decision



$$S_{t+1} = \left(\begin{pmatrix} NY \\ t+1 \end{pmatrix}, \hat{D}_{t+1}\right)$$

Nomadic trucker illustration

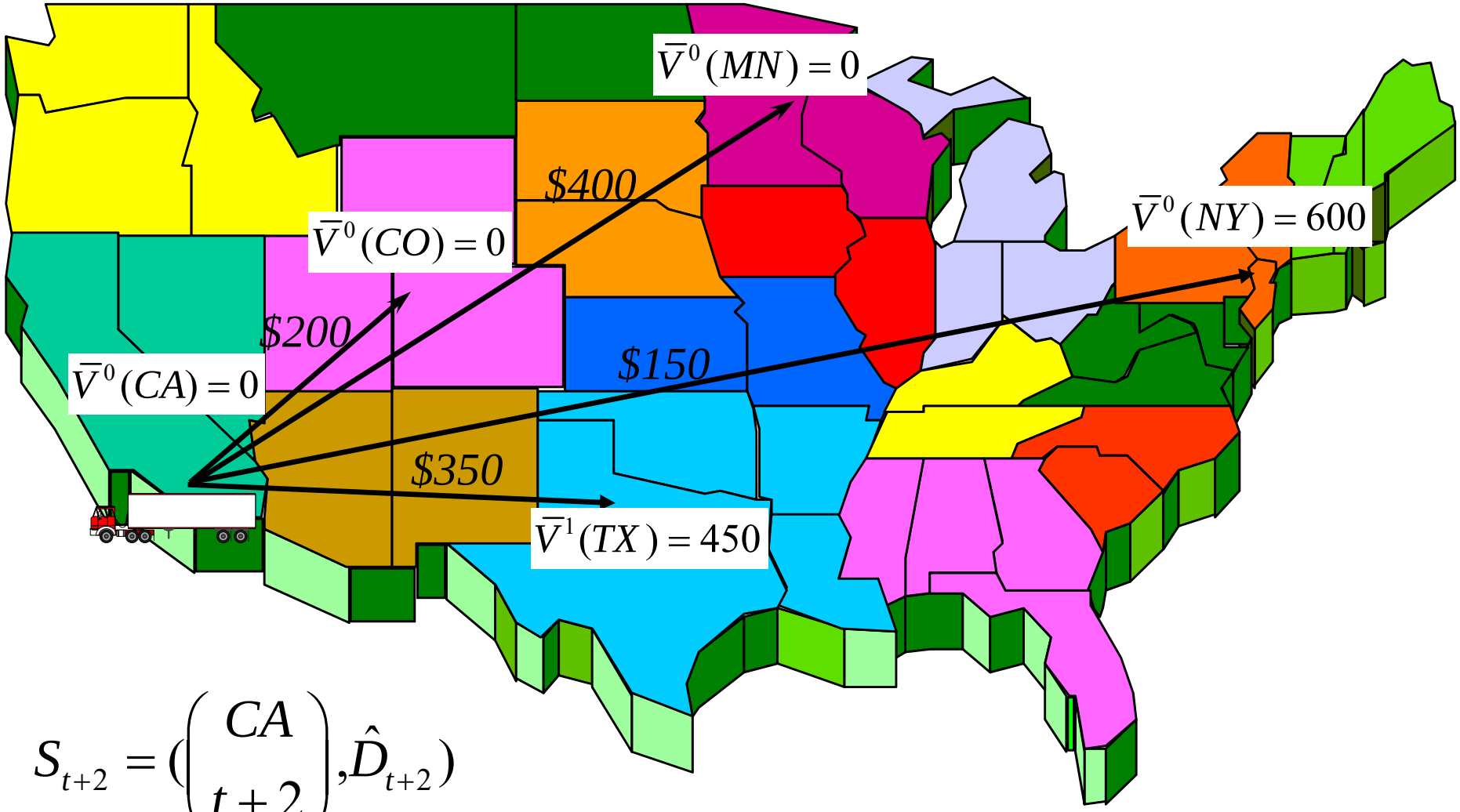
- Update value of being in NY



$$S_{t+1}^x = \begin{pmatrix} CA \\ t+2 \end{pmatrix}$$

Nomadic trucker illustration

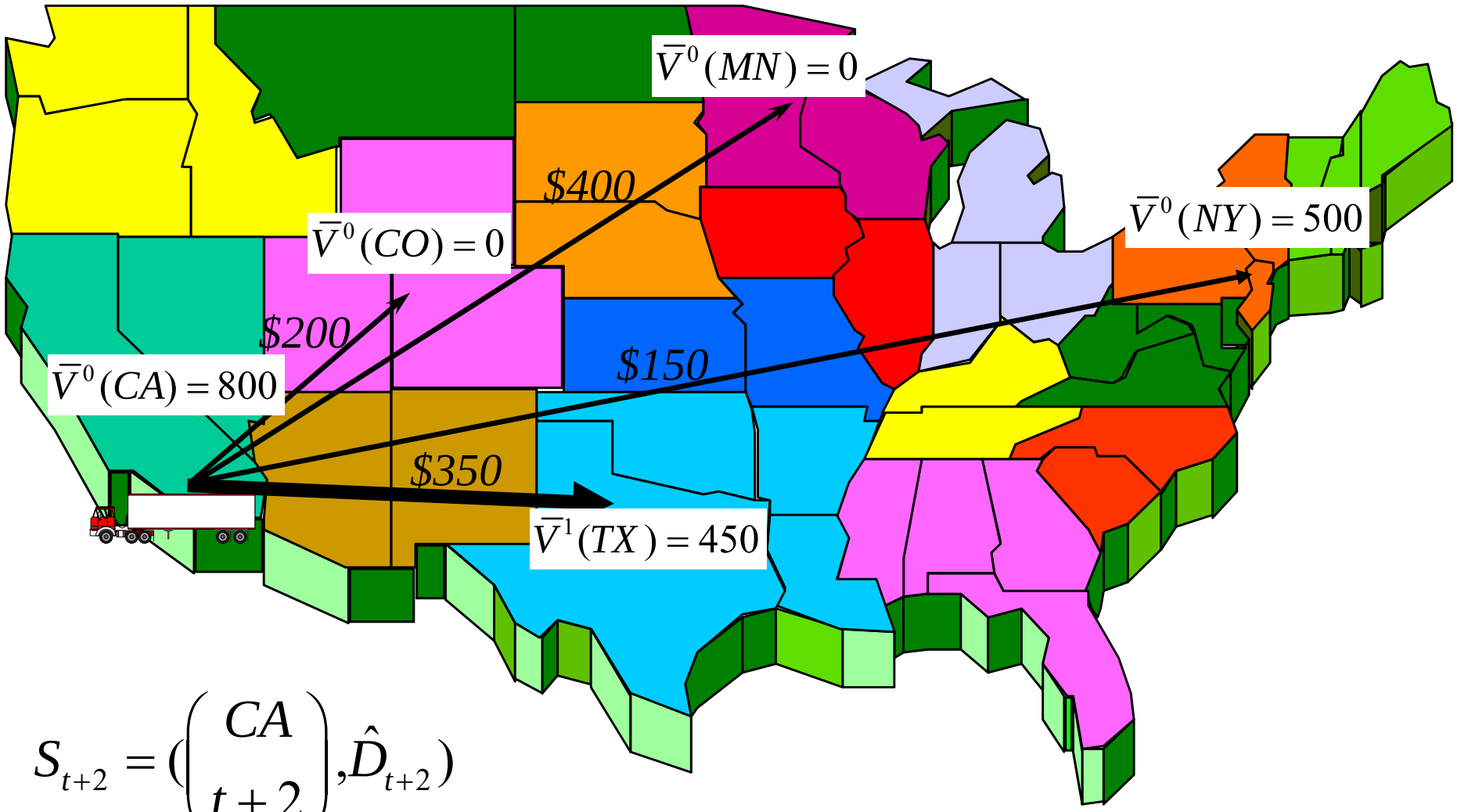
- Move to California.



$$S_{t+2} = \left(\begin{pmatrix} CA \\ t+2 \end{pmatrix}, \hat{D}_{t+2}\right)$$

Nomadic trucker illustration

- Make decision to return to TX and update value of being in CA



$$S_{t+2} = \left(\begin{pmatrix} CA \\ t+2 \end{pmatrix}, \hat{D}_{t+2}\right)$$

Nomadic trucker illustration

- Updating the value function:

Old value:

$$\bar{V}^1(TX) = \$450$$

New estimate:

$$\hat{v}^2(TX) = \$800$$

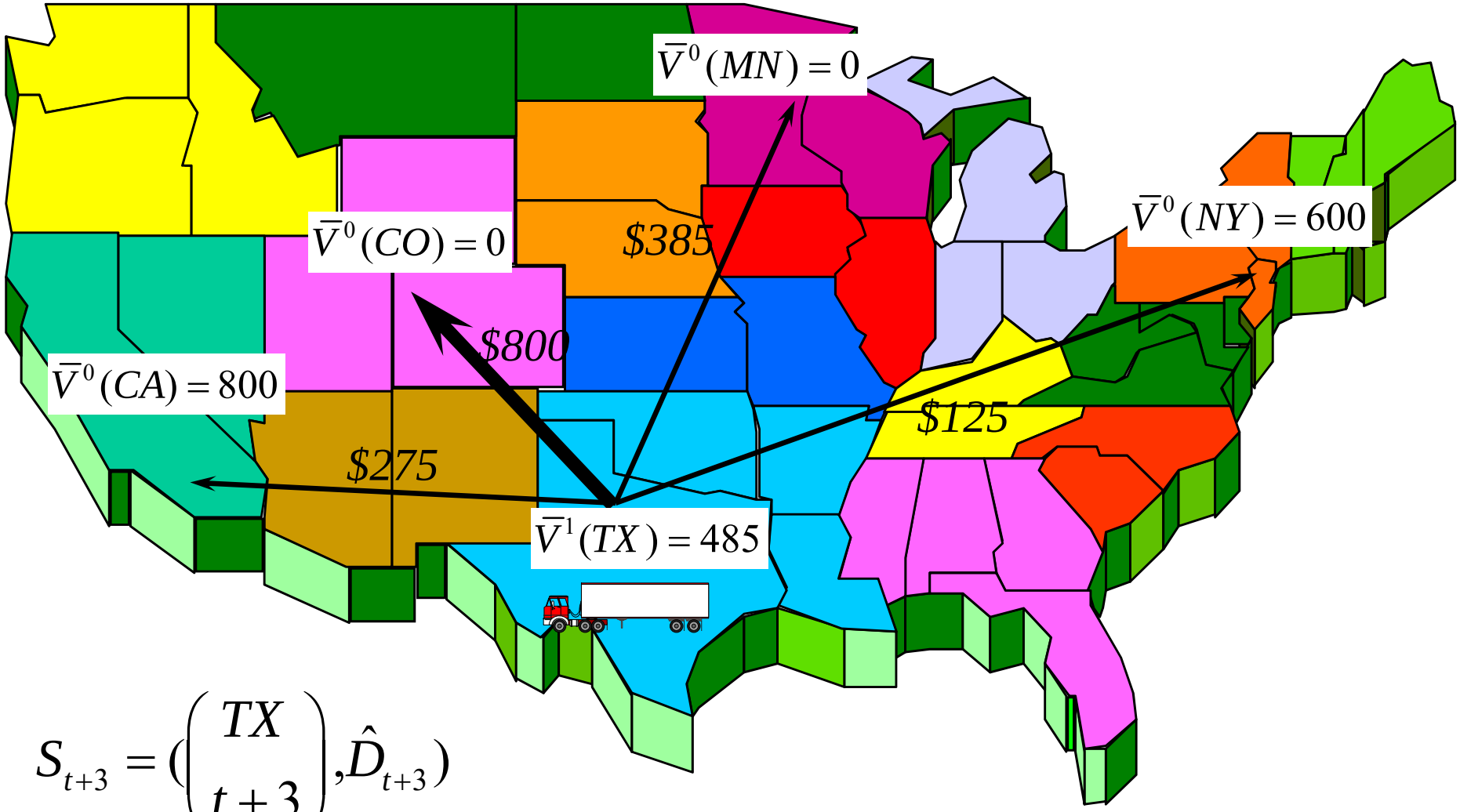
How do we merge old with new?

$$\begin{aligned}\bar{V}^2(TX) &= (1-\alpha)\bar{V}^1(TX) + (\alpha)\hat{v}^2(TX) \\ &= (0.90)\$450 + (0.10)\$800 \\ &= \$485\end{aligned}$$

» We are updating the previous post-decision state (describe).

Nomadic trucker illustration

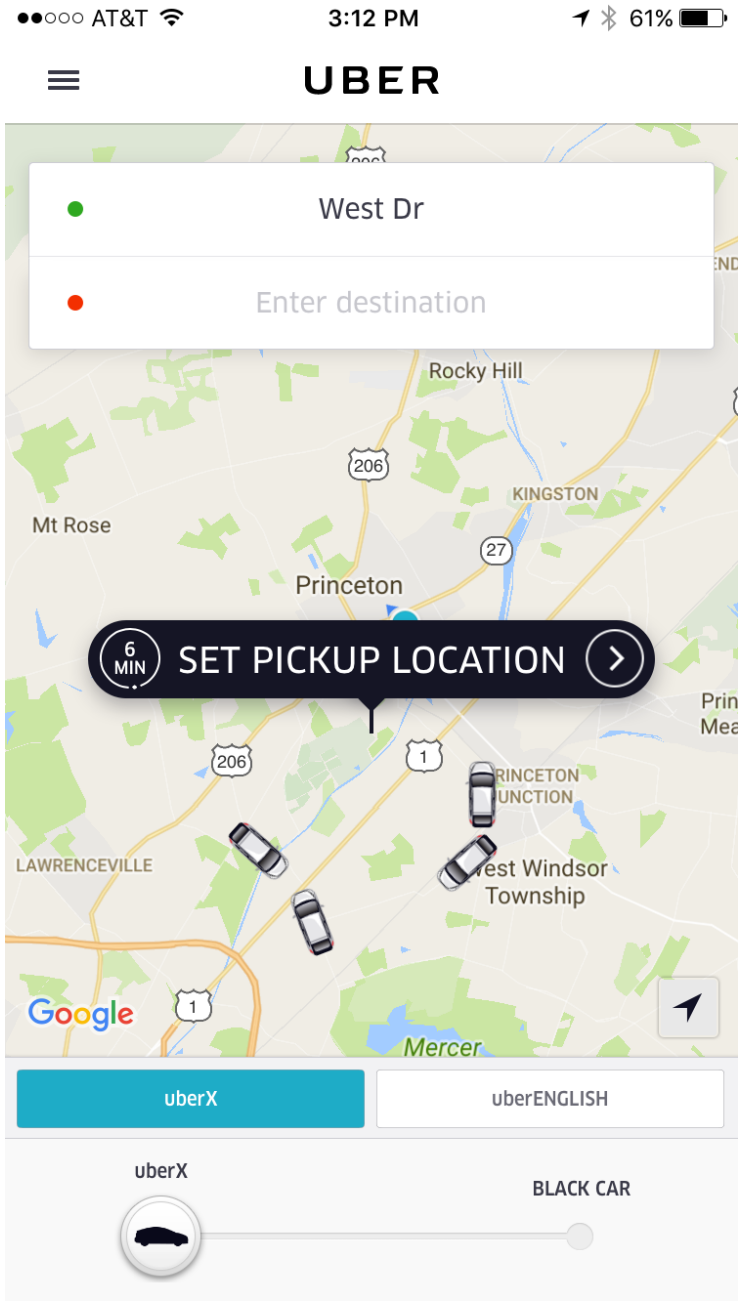
- An updated value of being in TX



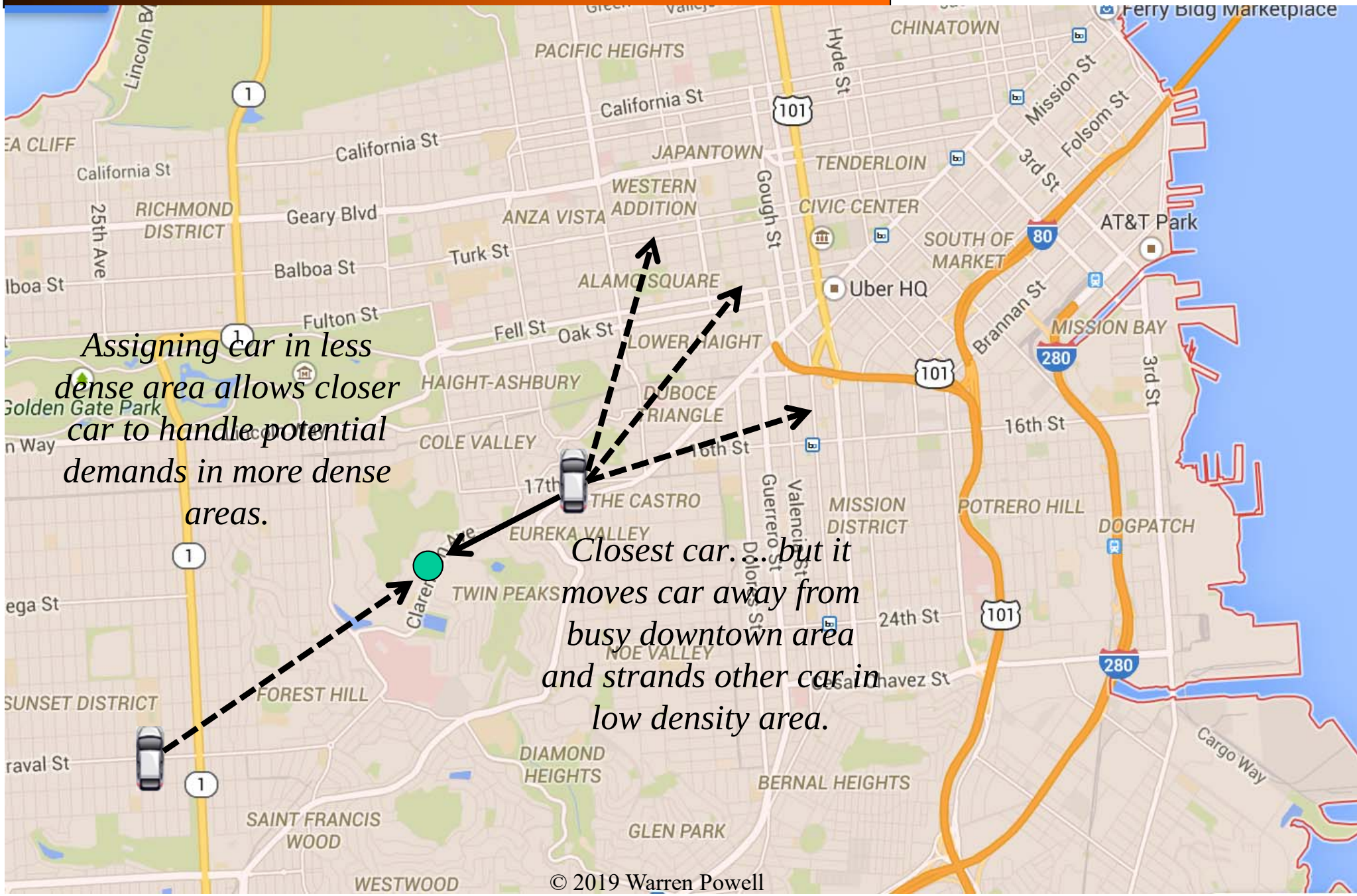
$$S_{t+3} = \left(\begin{pmatrix} TX \\ t+3 \end{pmatrix}, \hat{D}_{t+3}\right)$$

Real-time logistics

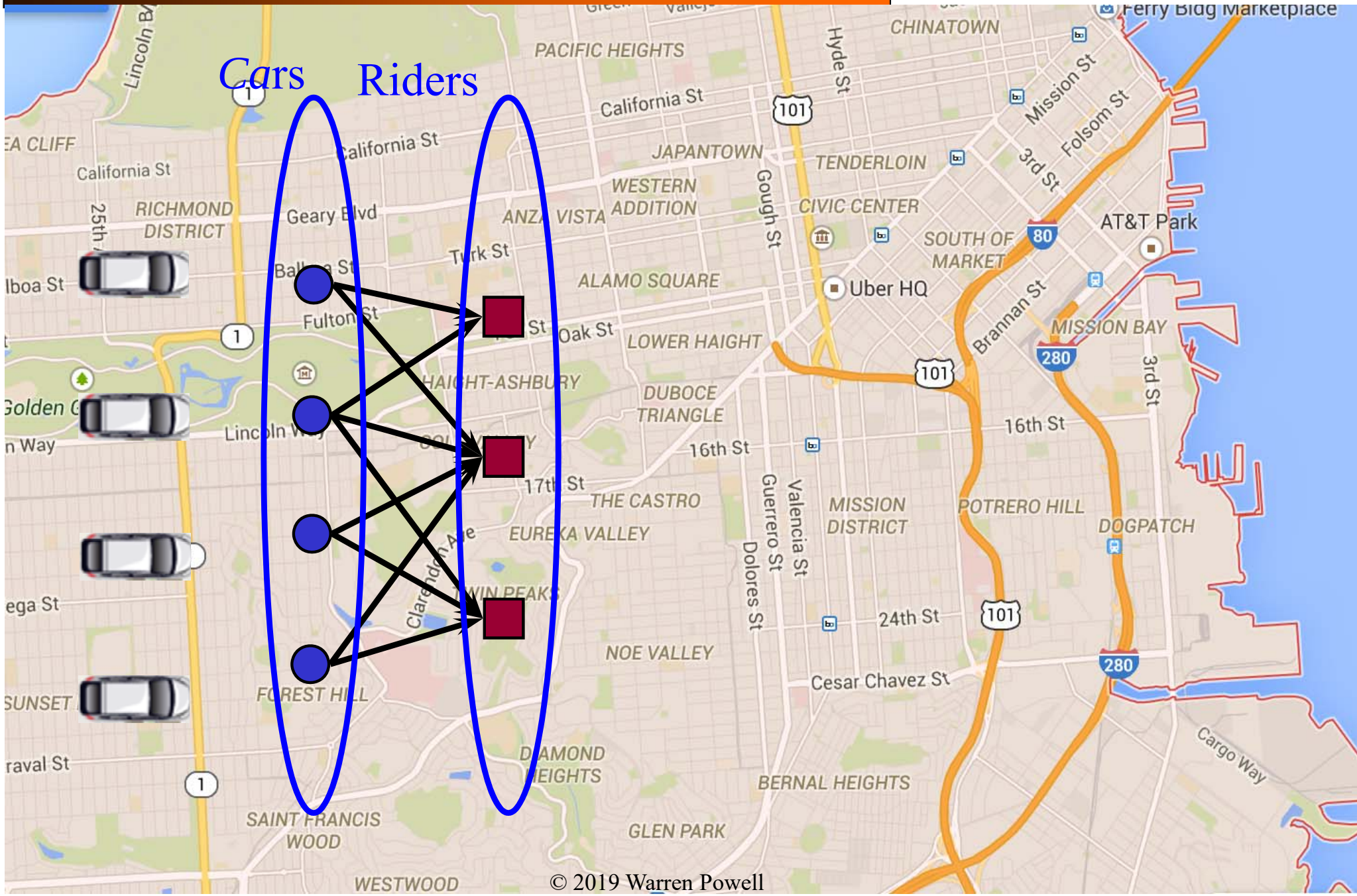
- Uber
  - » Provides real-time, on-demand transportation.
  - » Drivers are encouraged to enter or leave the system using pricing signals and informational guidance.
- Decisions:
  - » How to price to get the right balance of drivers relative to customers.
  - » Assigning and routing drivers to manage Uber-created congestion.
  - » Real-time management of drivers.
  - » Pricing (trips, new services, …)
  - » Policies (rules for managing drivers, customers, …)



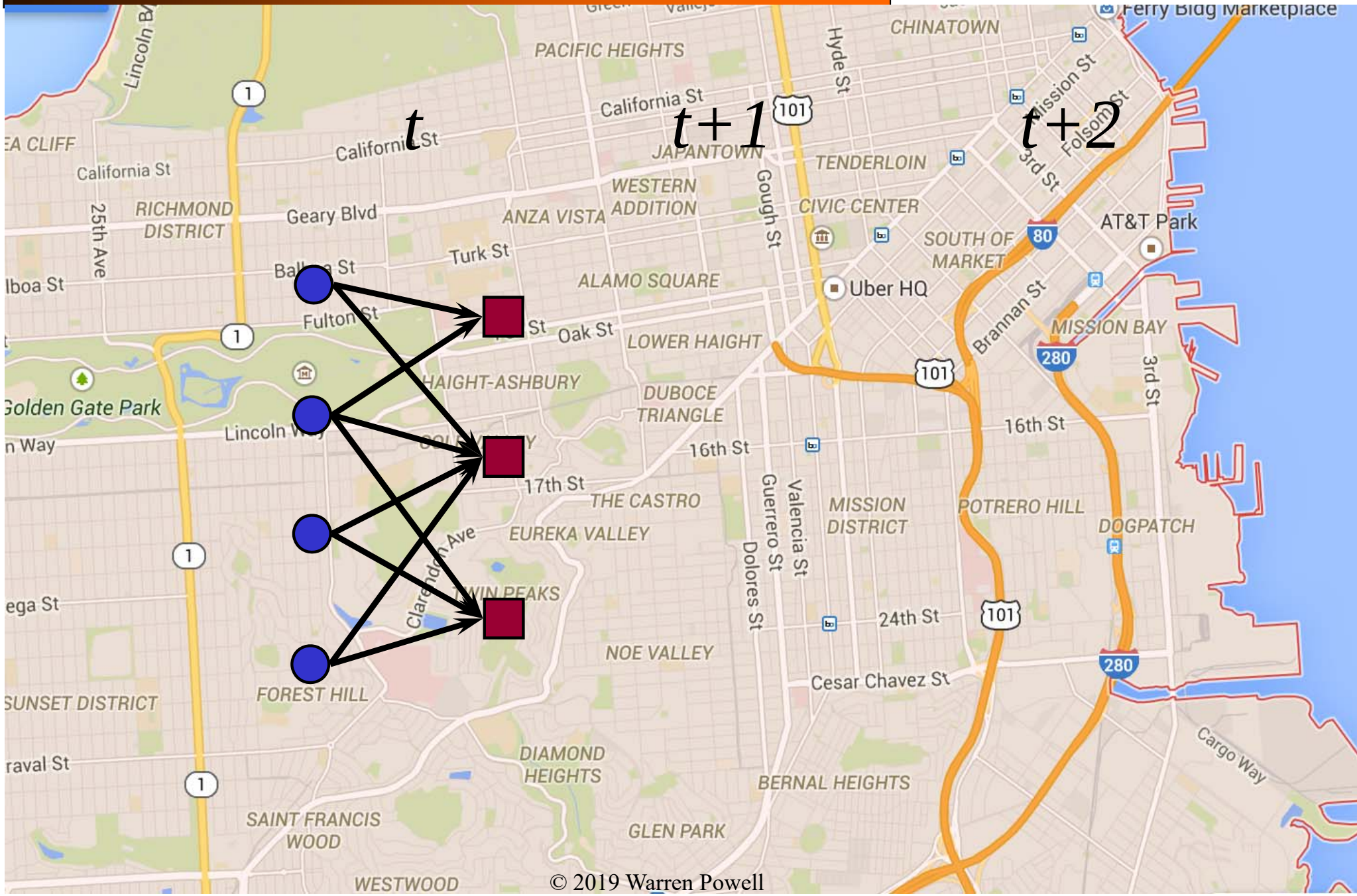
Effect of Current Decision on the Future



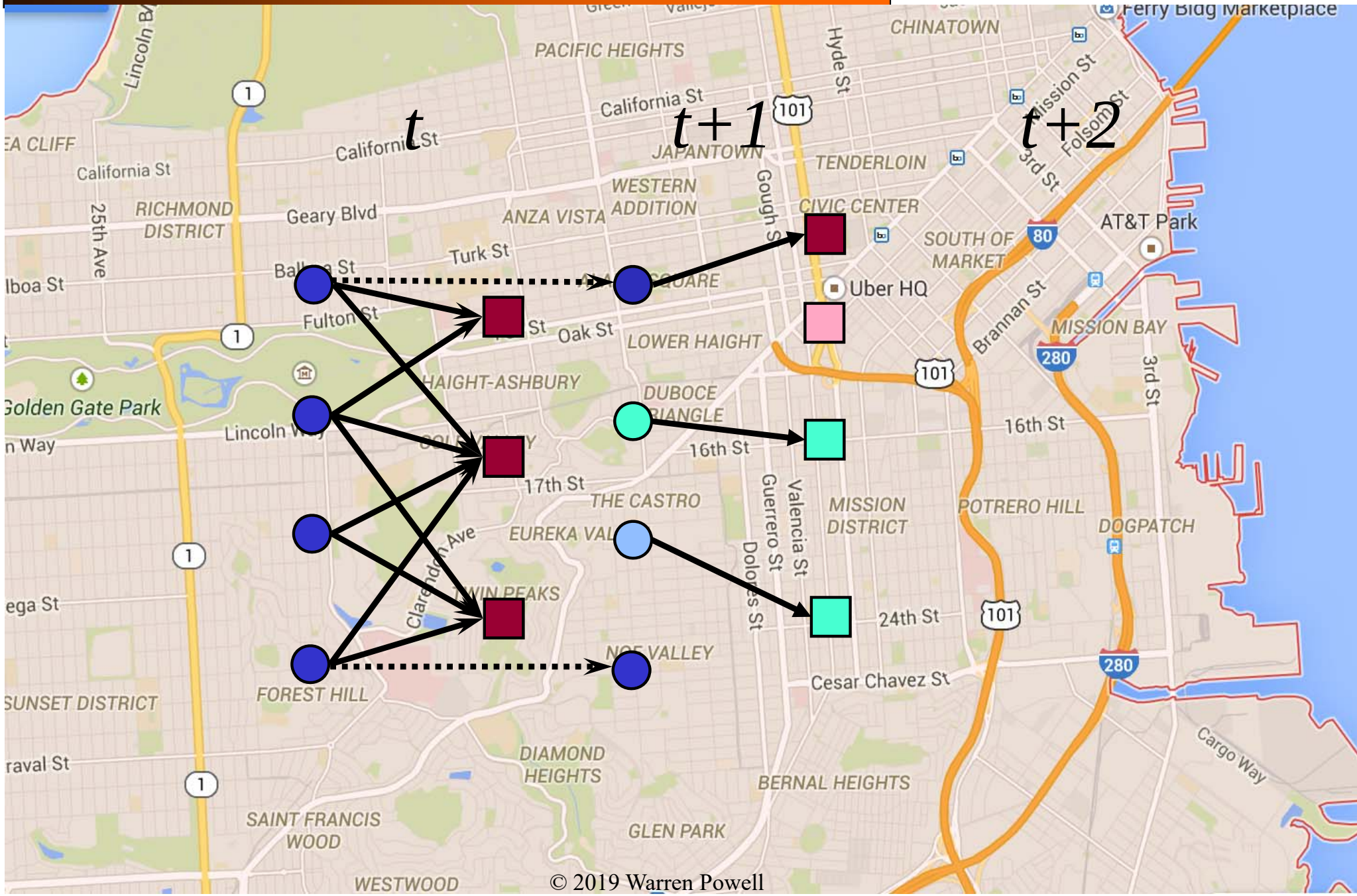
Cost function approximations



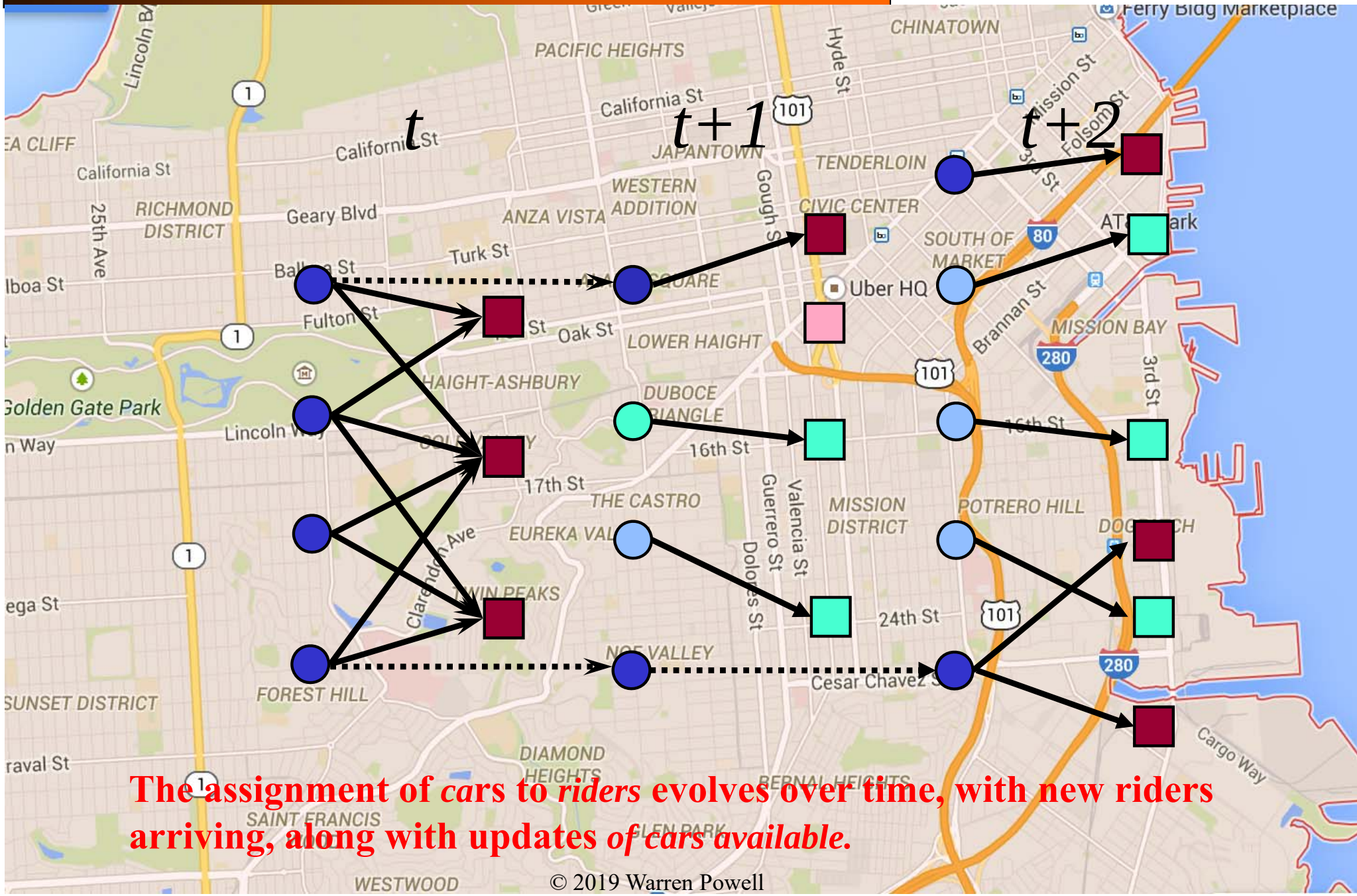
Optimizing over time



Optimizing over time

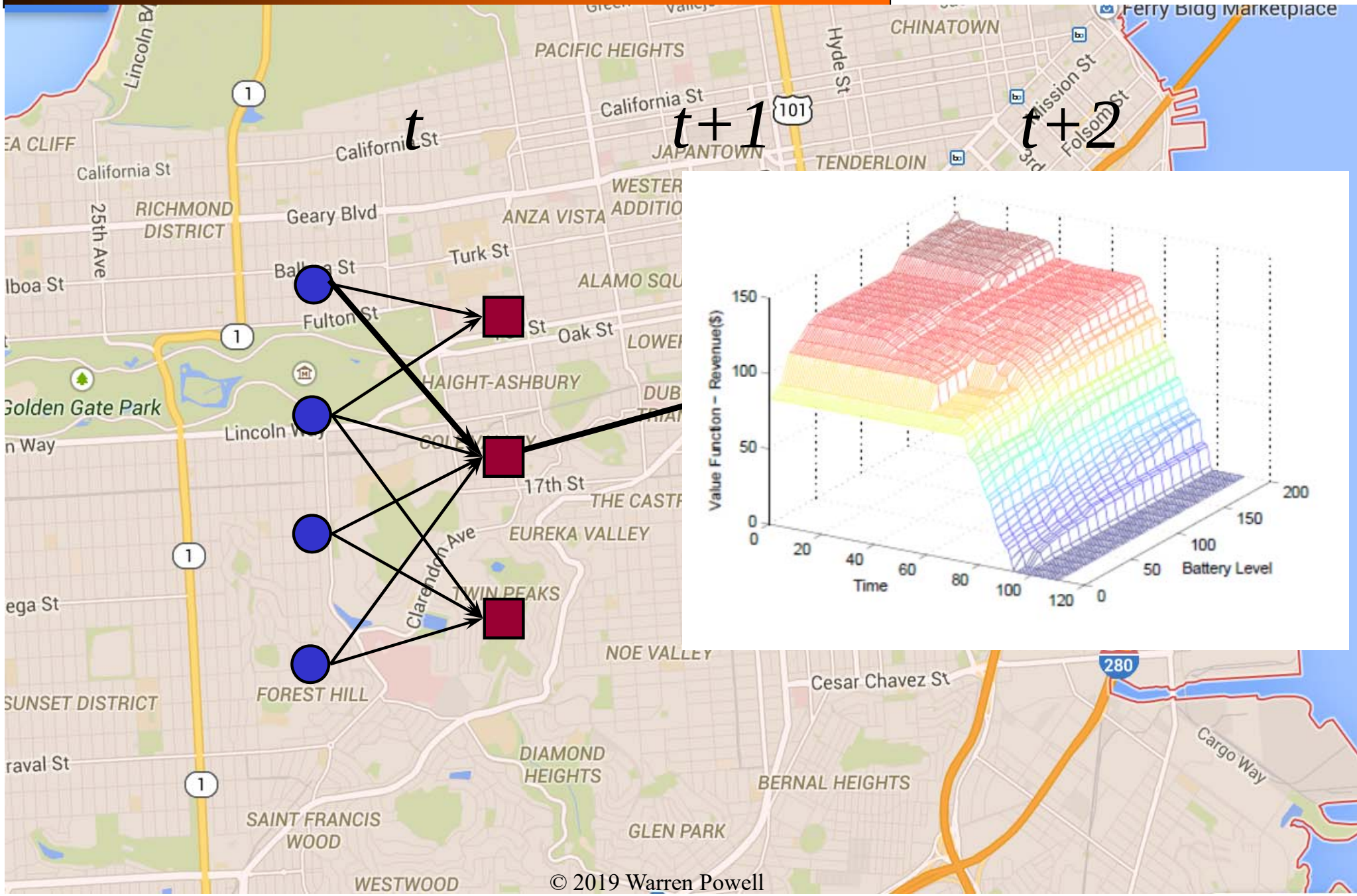


Optimizing over time



The assignment of *cars* to *riders* evolves over time, with new riders arriving, along with updates *of cars available.*

Autonomous EVs



Matching buyers with sellers

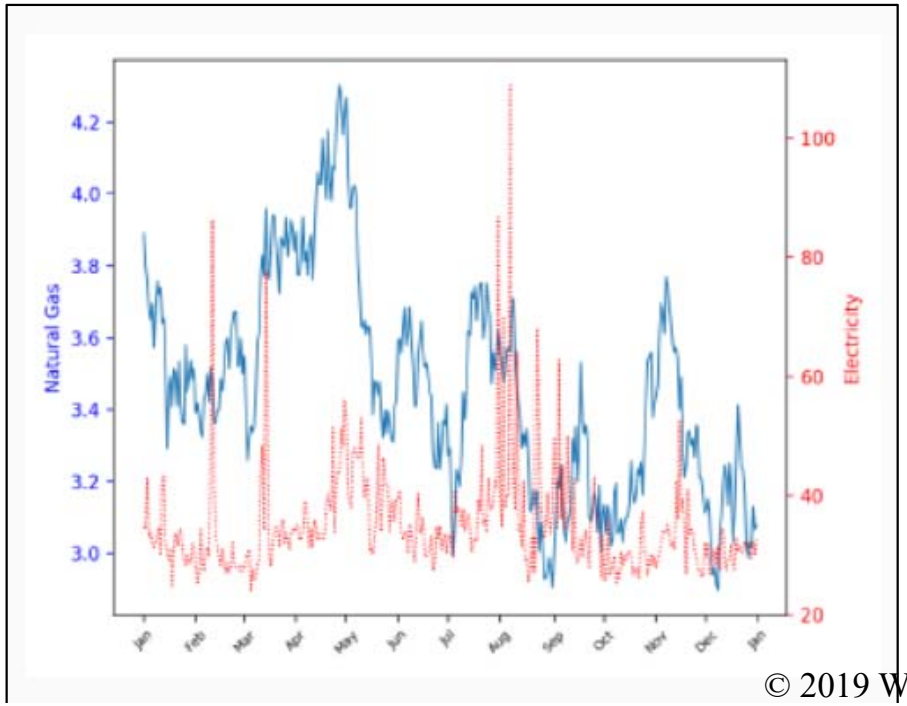
- Now we have a logistic curve for each origin-destination pair (i,j)

$$P^Y(p,a\,|\,\theta) = \frac{e^{\theta^0_{ij}+\theta_{ij}p+\theta^a_{ij}a}}{1+e^{\theta^0_{ij}+\theta_{ij}p+\theta^a_{ij}a}}$$

- Number of offers for each (i,j) pair is relatively small.
- Need to generalize the learning across hundreds to thousands of markets.

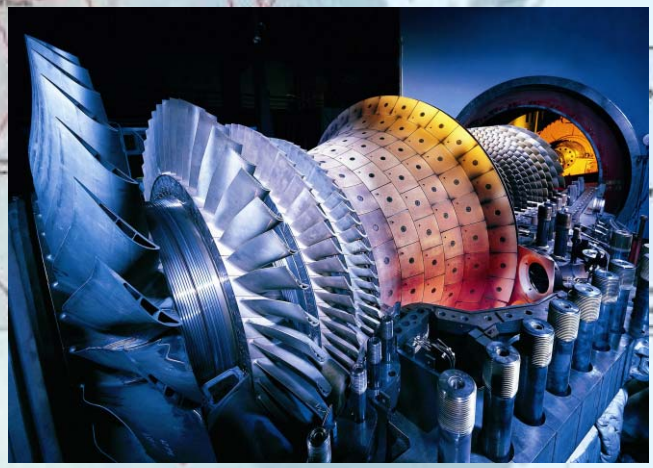
Industrial sponsors

- Air Liquide
  - » Largest industrial gases company with 64,000 employees.
  - » Consumes 0.1 percent of *global* electricity.

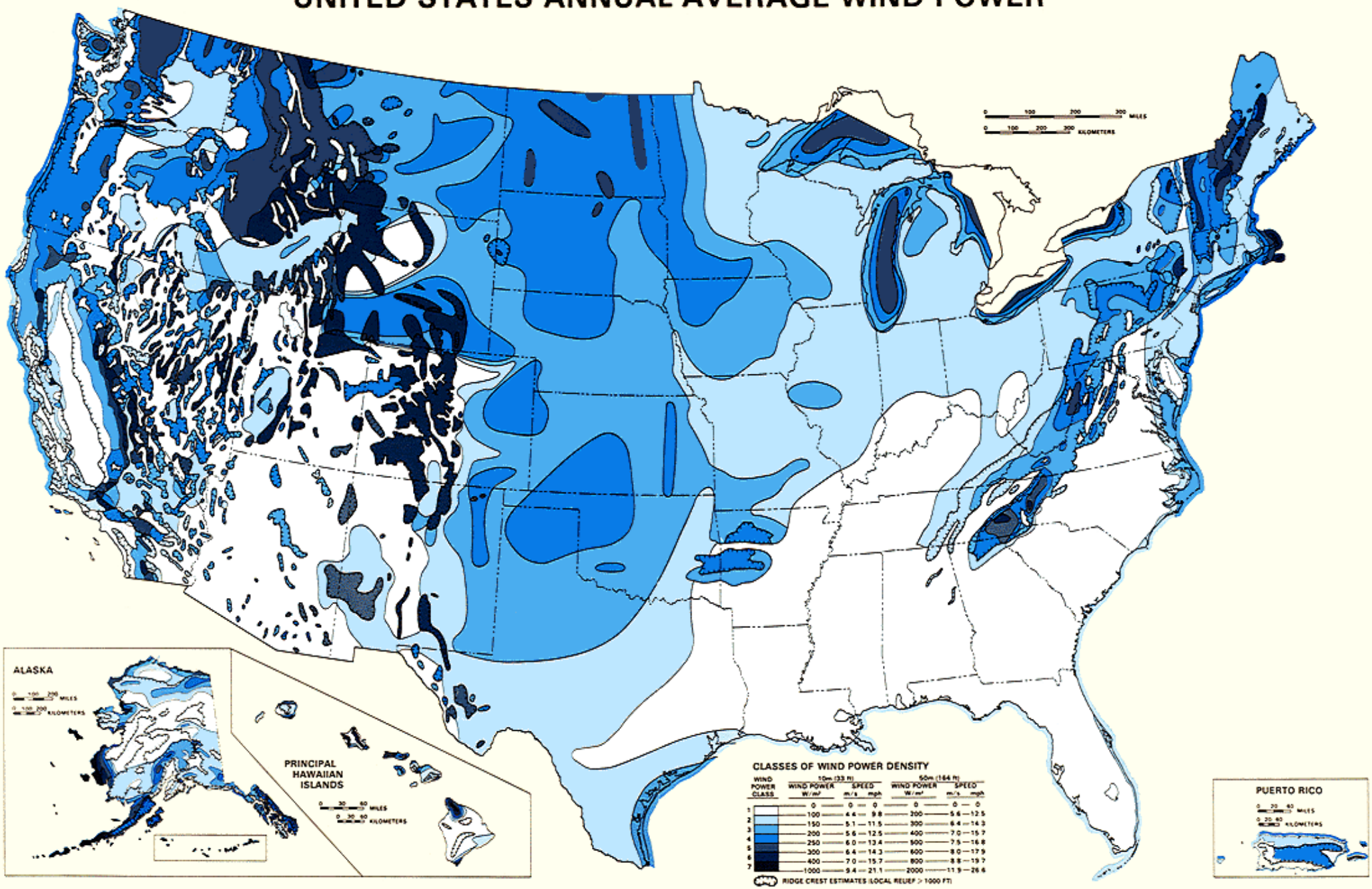


- Challenges
  - » Faces a variety of challenges to manage risk:
  - » Spikes in natural gas prices, electricity prices.
  - » Pipeline outages due to storms.

An energy generation portfolio

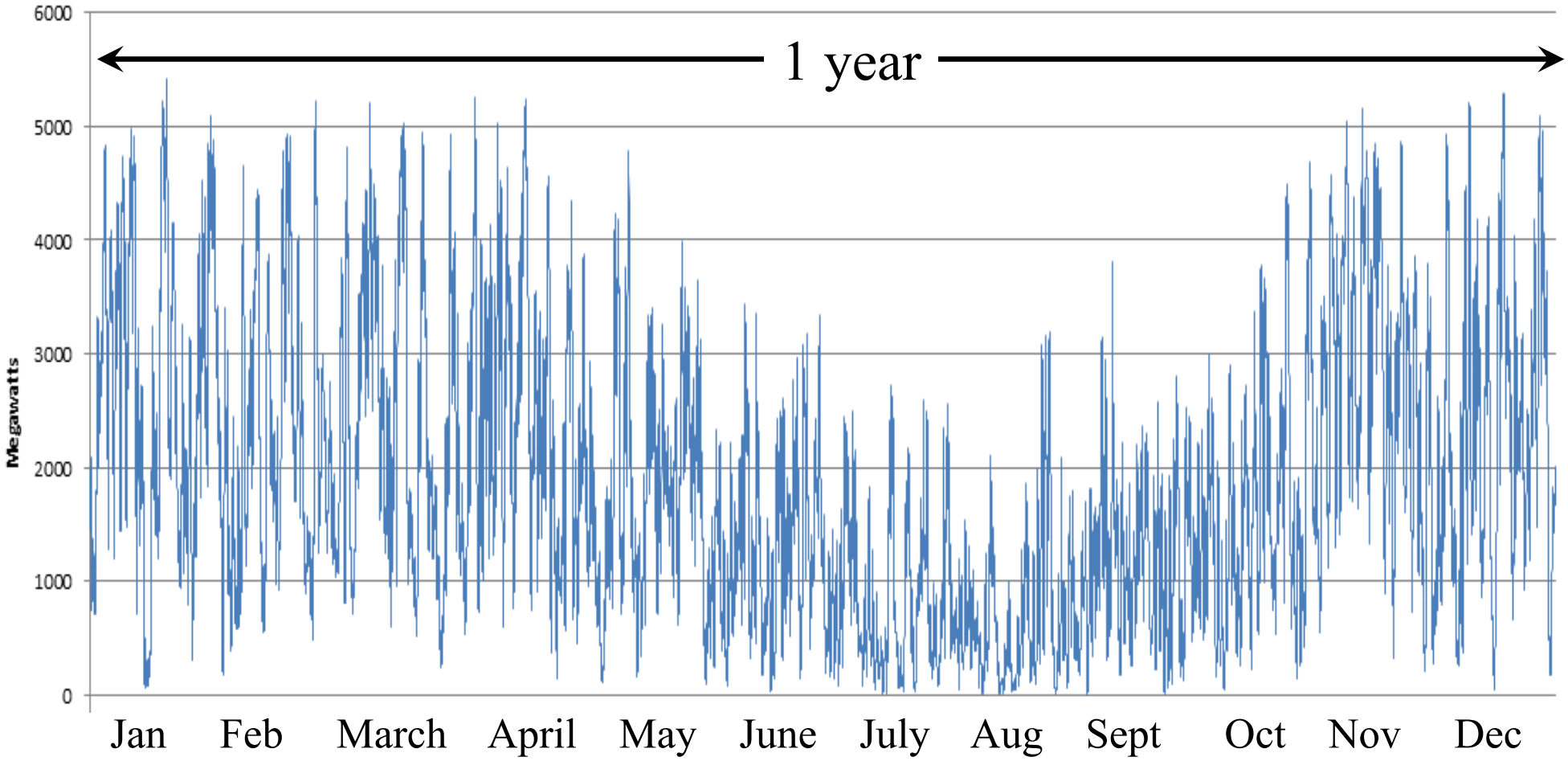


UNITED STATES ANNUAL AVERAGE WIND POWER



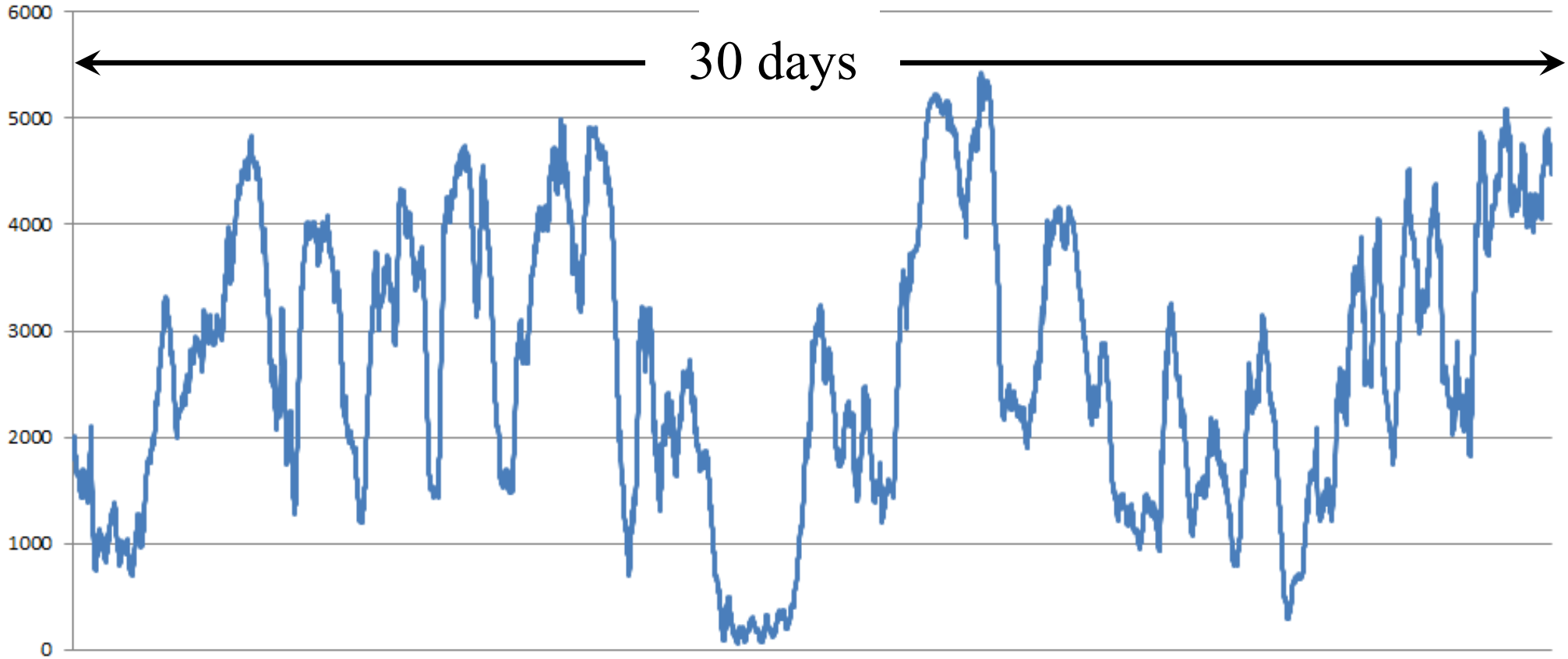
Energy from wind

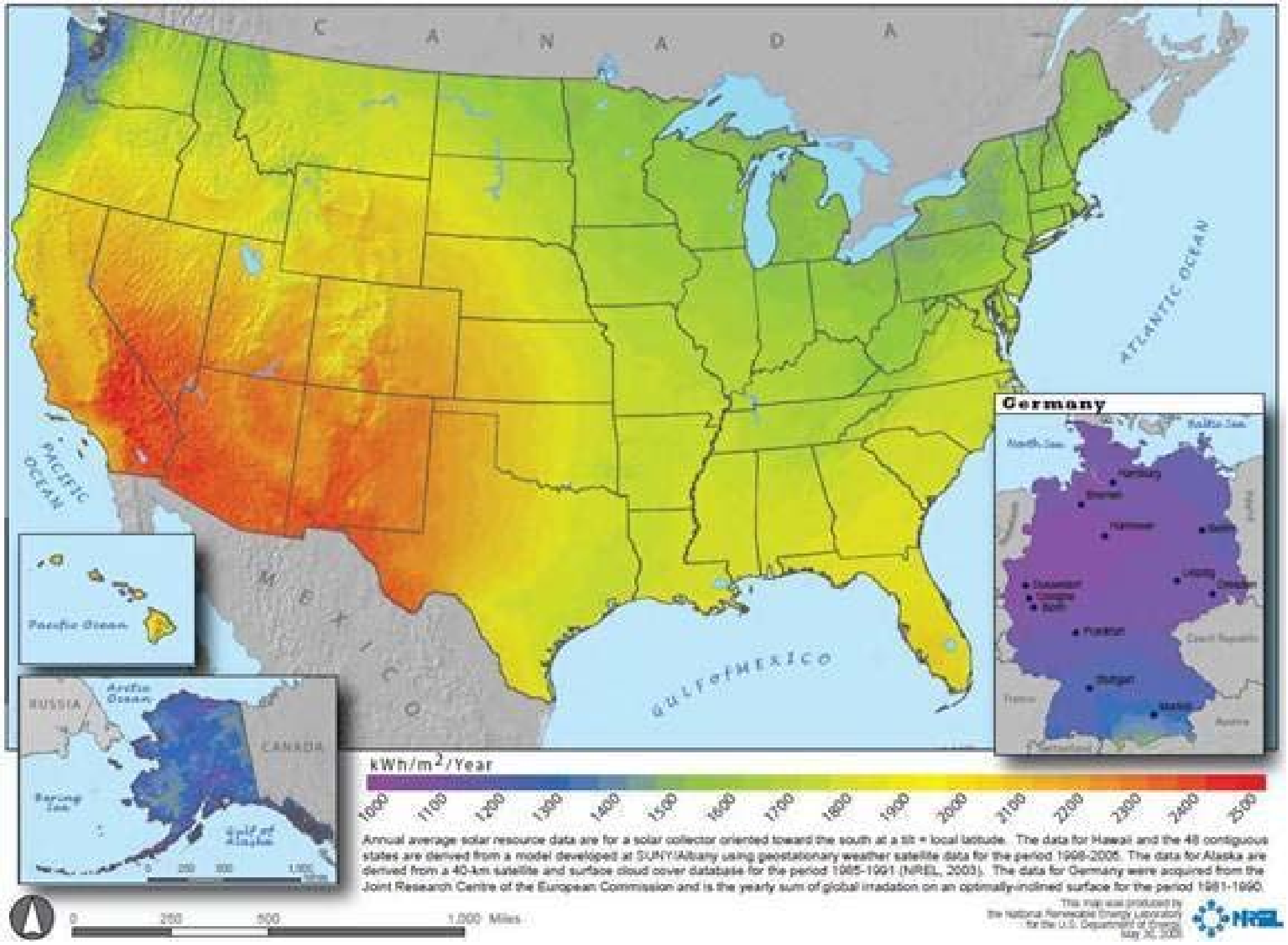
- Wind power from all PJM wind farms



Energy from wind

- Wind from all PJM wind farms

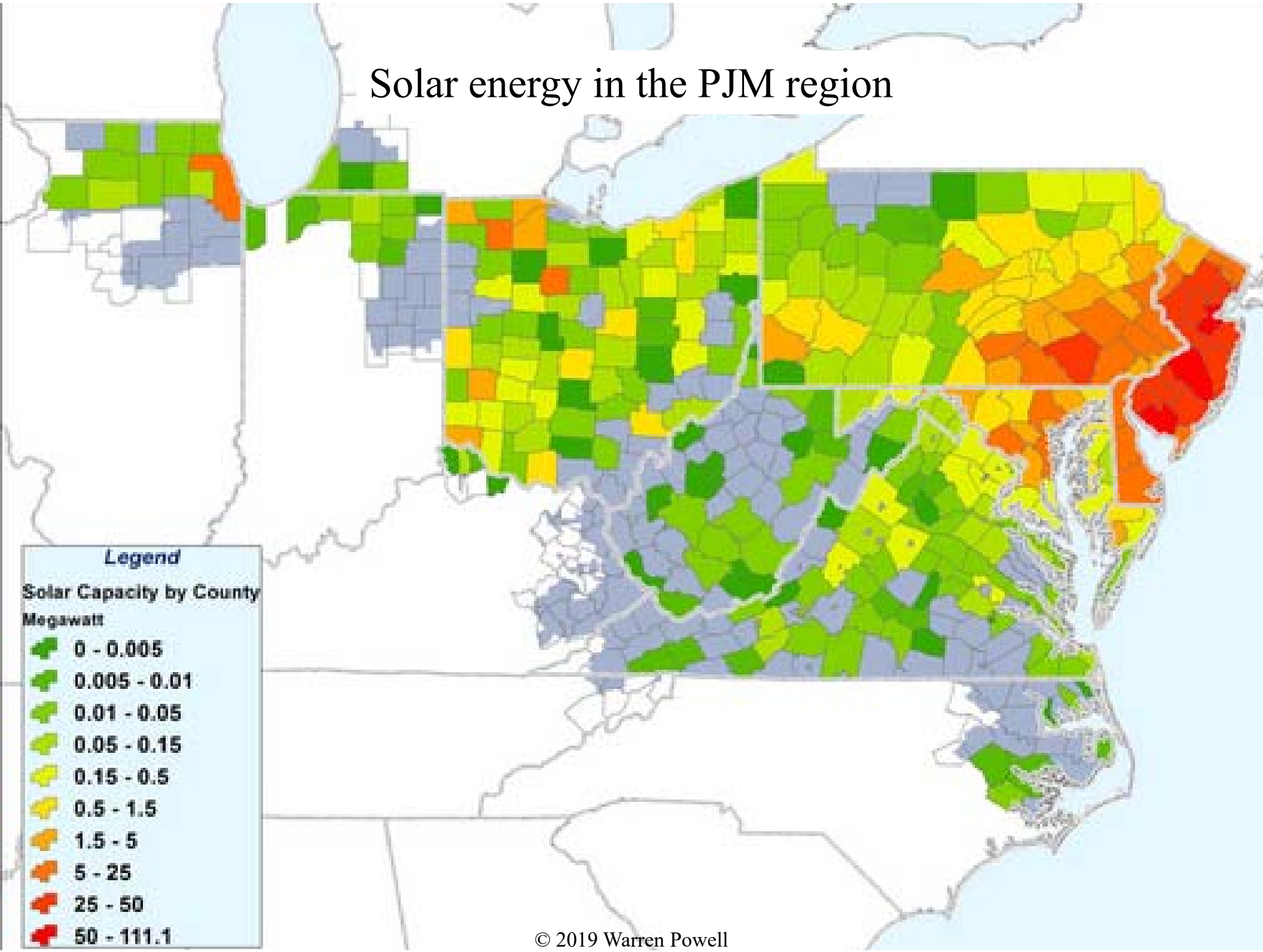




Annual average solar resource data are for a solar collector oriented toward the south at a tilt = local latitude. The data for Hawaii and the 48 contiguous states are derived from a model developed at SUNY/Albany using geostationary weather satellite data for the period 1998-2005. The data for Alaska are derived from a 40-km satellite and surface cloud cover database for the period 1985-1991 (NREL, 2003). The data for Germany were acquired from the Joint Research Centre of the European Commission and is the yearly sum of global irradiation on an optimally-inclined surface for the period 1981-1990.

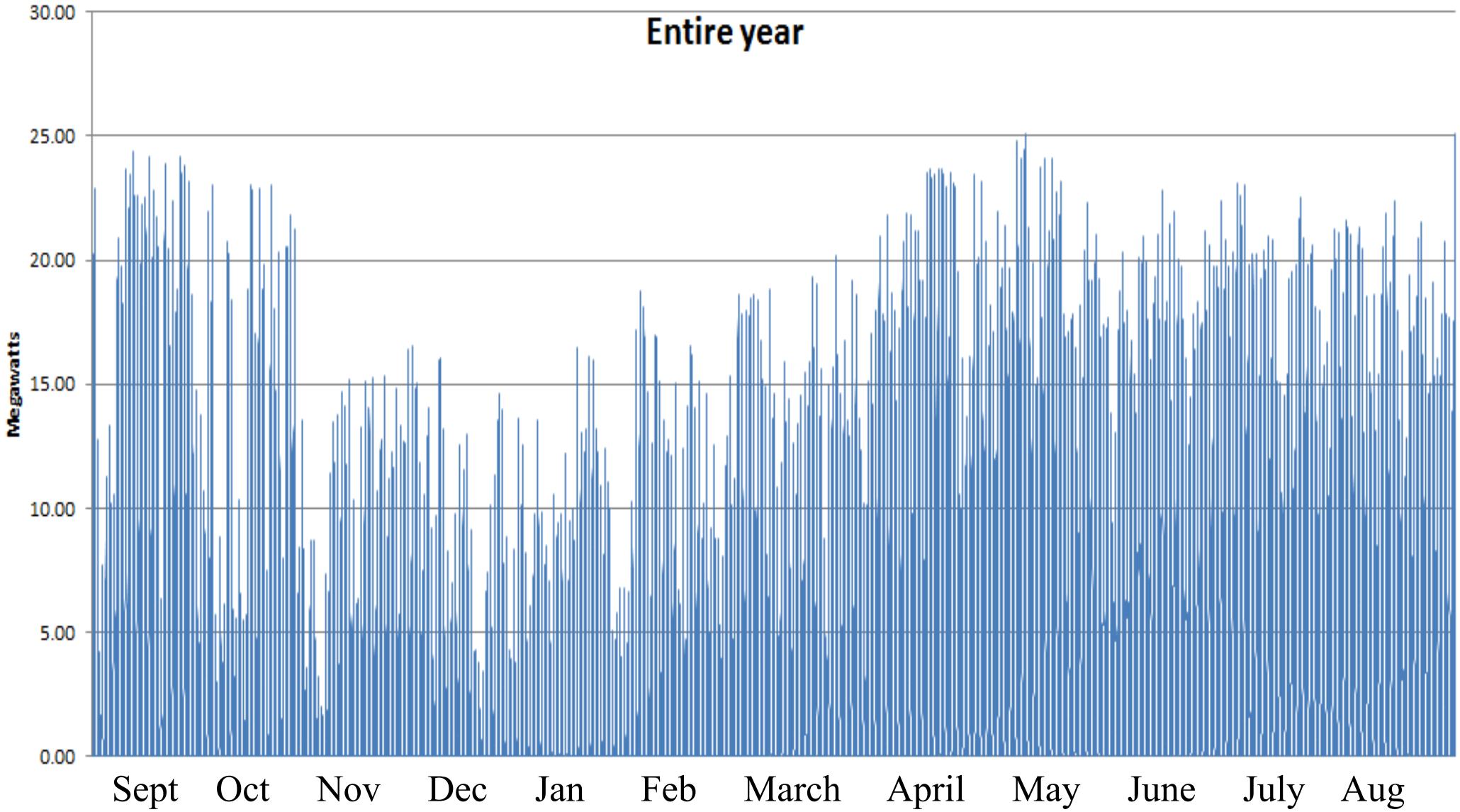
This map was produced by the National Renewable Energy Laboratory for the U.S. Department of Energy May 30, 2003

Solar energy in the PJM region

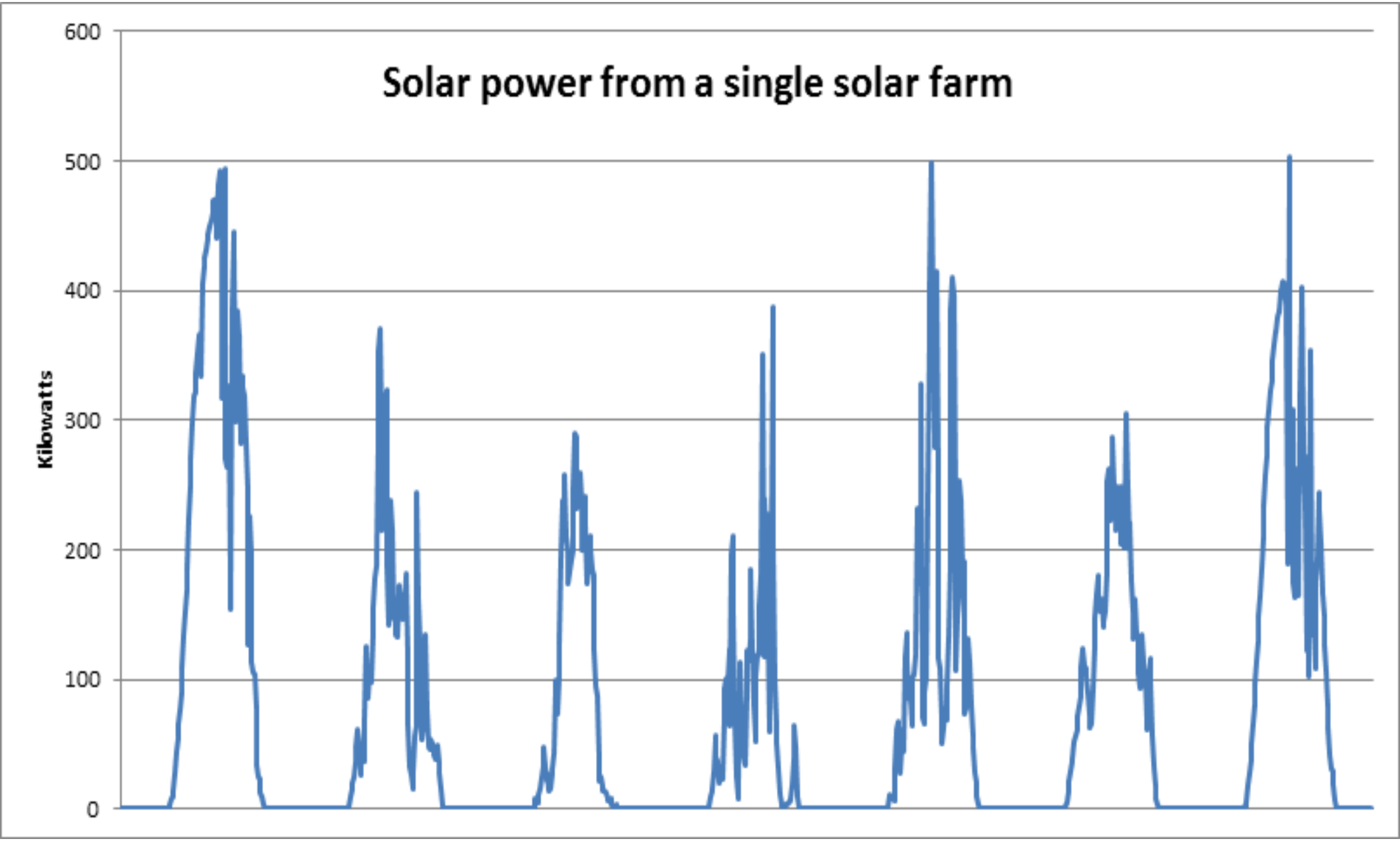


Solar energy

- PSE&G solar farms

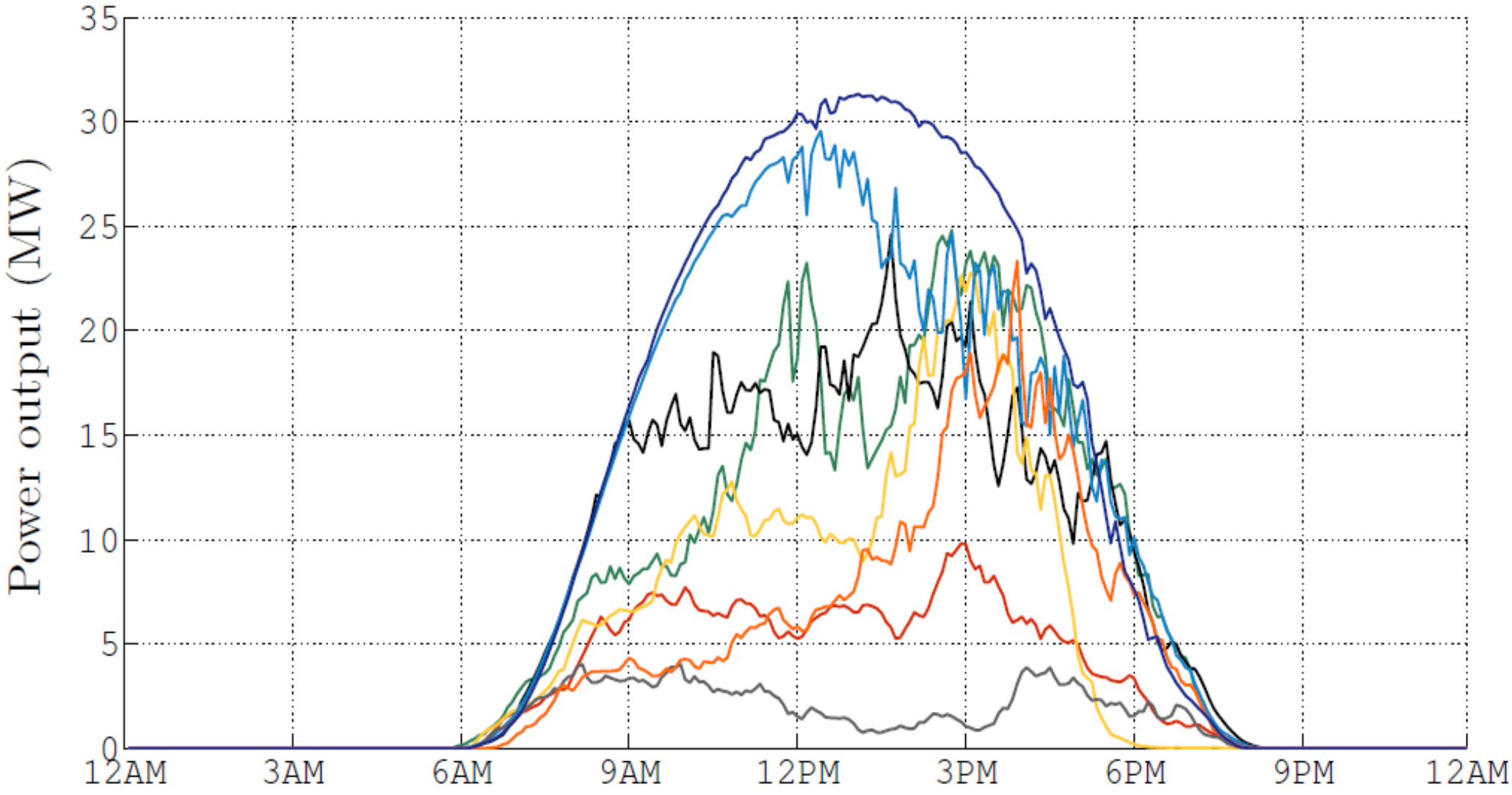


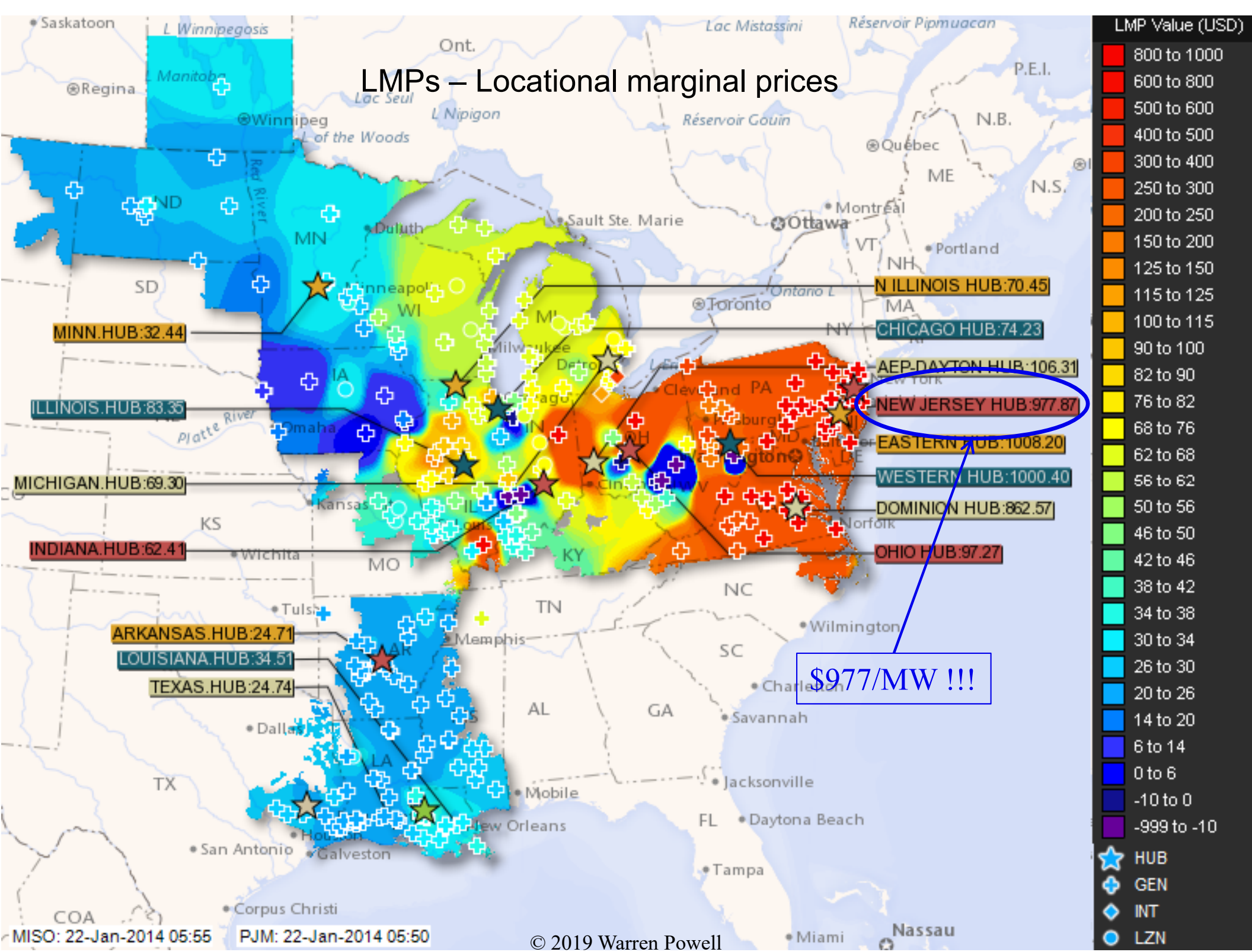
Solar energy



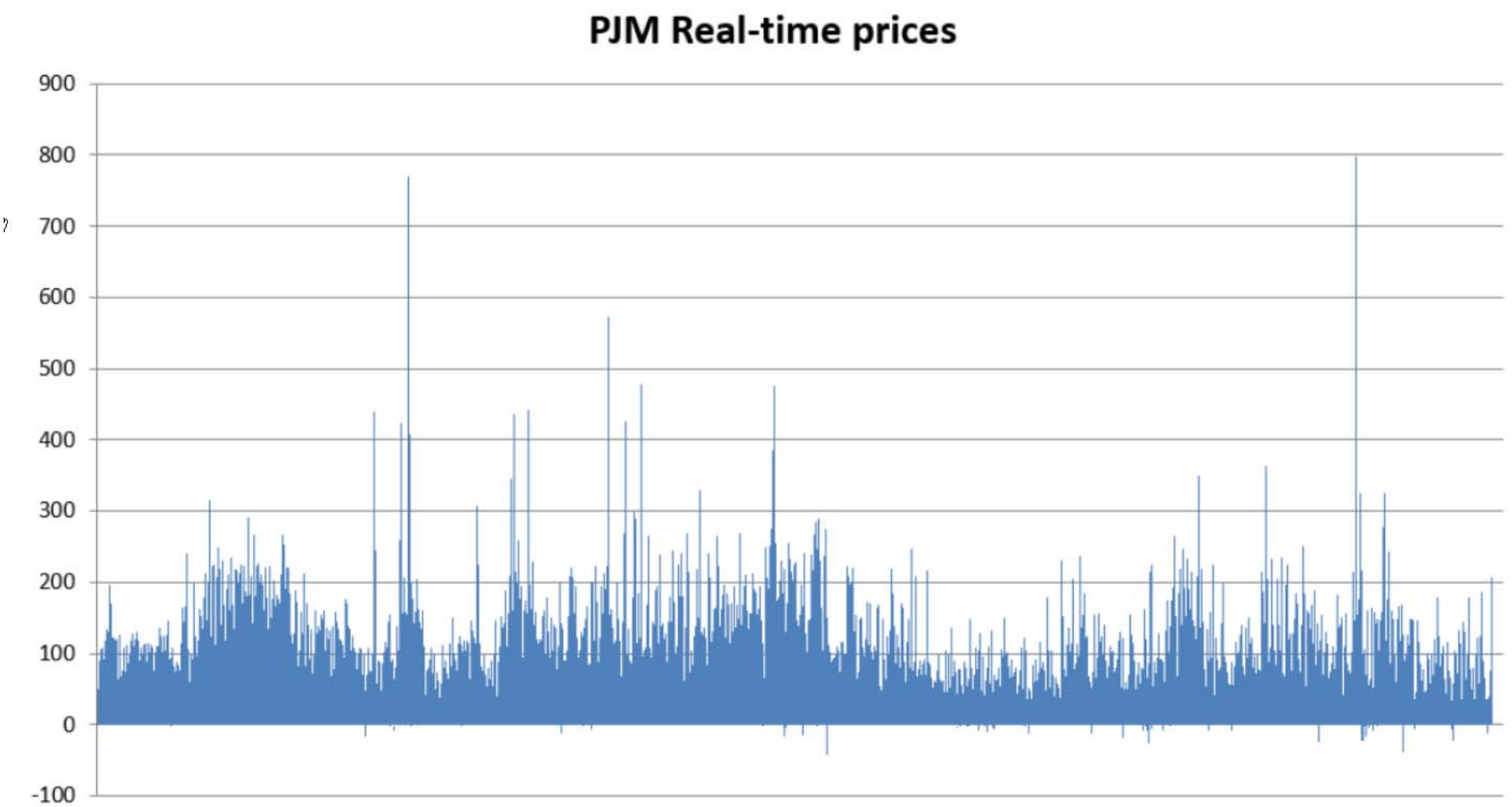
Solar energy

- Within-day sample trajectories



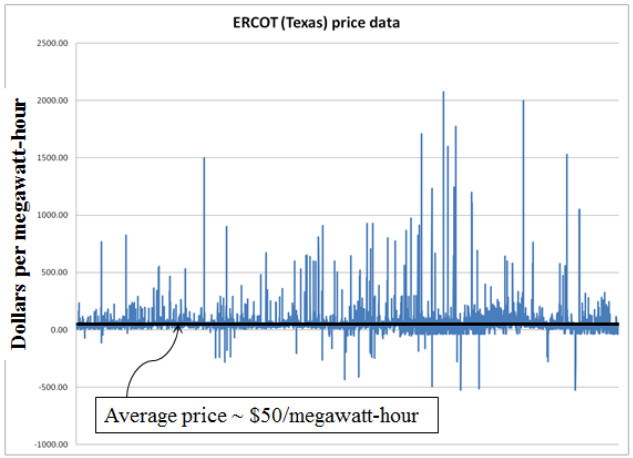


Locational marginal prices on the grid



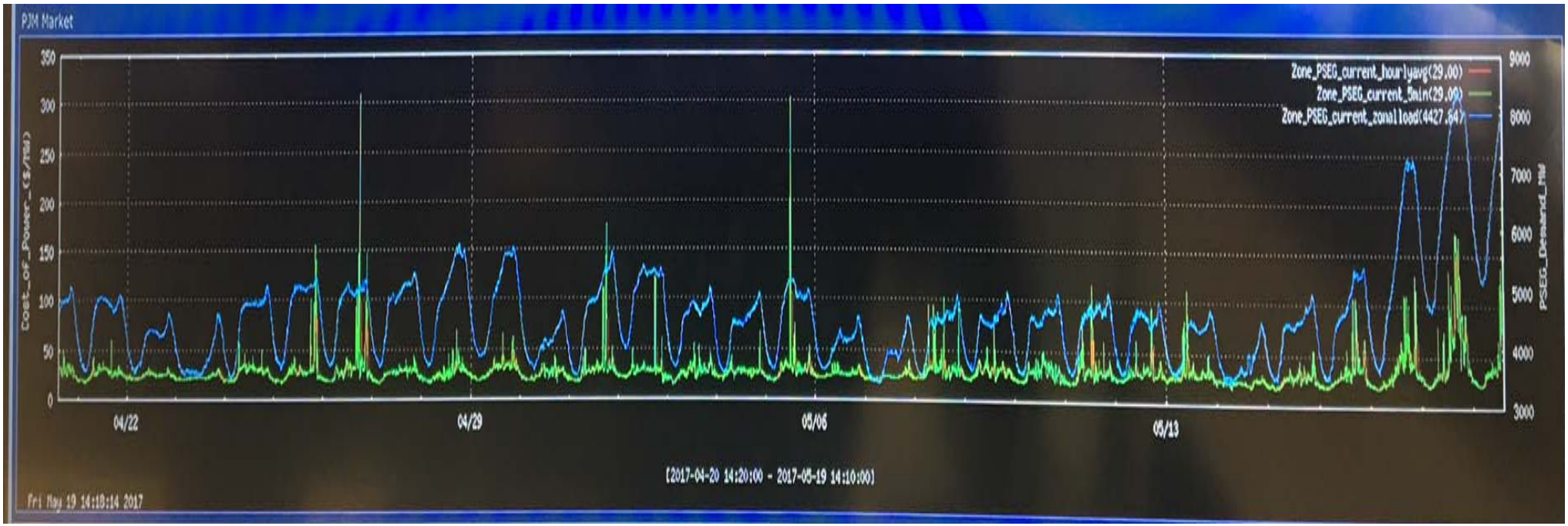
Planning under uncertainty

- Battery arbitrage – When to charge, when to discharge, given volatile LMPs



Planning under uncertainty

- Snapshot of electricity prices for New Jersey:



Planning under uncertainty

ARTICLES
PUBLISHED ONLINE: 13 JUNE 2016 | DOI: 10.1038/NCLIMATE3045
nature climate change

Value of storage technologies for wind and solar energy

$$R_{\text{total}} = \max \left(\sum_{t=0}^{N} P(t) \left(x_{\text{generation}}(t) + x_{\text{discharge}}(t) - x_{\text{charge}}(t)/\eta \right) \right)$$

subject to:

Stochastic prices

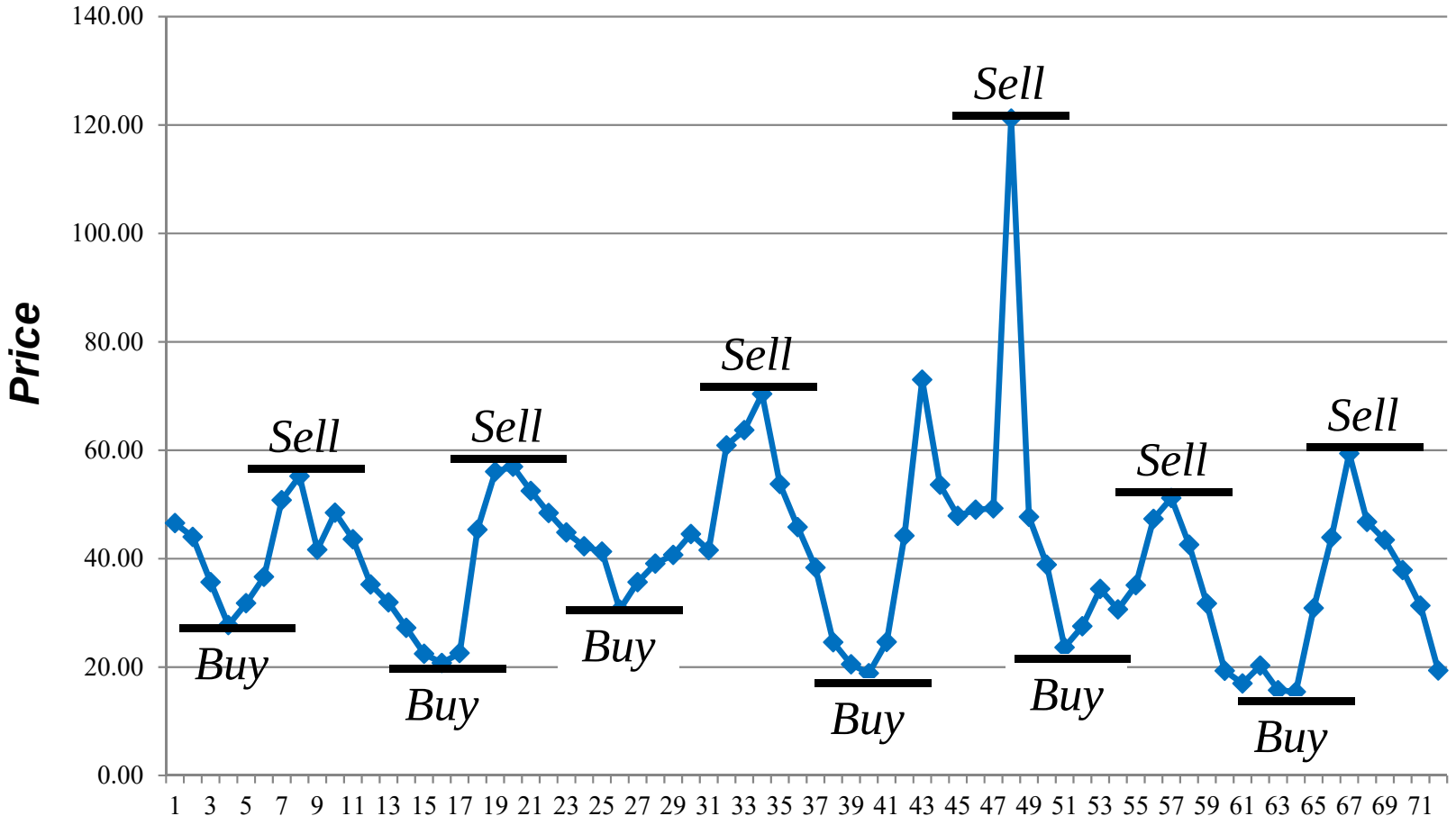
$$0 \le x_{\text{discharge}} \le \dot{E}_{\max}$$

$$0 \le x_{\text{charge}} \le \min(\eta x_{\text{generation}}(t), \eta \dot{E}_{\max})$$

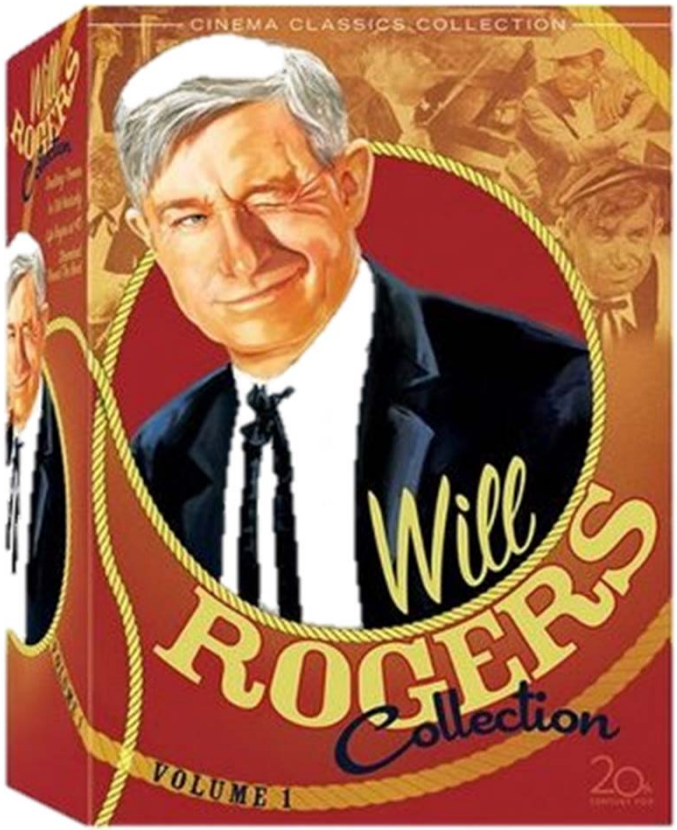
$$0 \le \sum_{t=0}^{N} \left(x_{\text{charge}}(t) - x_{\text{discharge}}(t) \right) \le h \dot{E}_{\max} \qquad (2)$$

Planning under uncertainty

- The models in these papers allow decisions to see into the future:
  - » Strategy posed by the battery manufacturer: “Buy low, sell high”



Planning under uncertainty



Don't gamble; take all your savings and buy some good stock and hold it till it goes up, then sell it. If it don't go up, don't buy it.

Will Rogers

It is not enough to mix "optimization" (intelligent decision making) and uncertainty. You have to be sure that each decision has access to the information available at the time.

From deterministic to stochastic

- Imagine that you would like to solve the time-dependent linear program:

$$\min_{x_0,\ldots,x_T} \sum_{t=0}^{T} c_t x_t$$

$$R_{\text{total}} = \max \left(\sum_{t=0}^{N} P(t)\left(x_{\text{generation}}(t) + x_{\text{discharge}}(t) - x_{\text{charge}}(t)/\eta\right) \right)$$

  - » subject to

$$A_0 x_0 = b_0$$

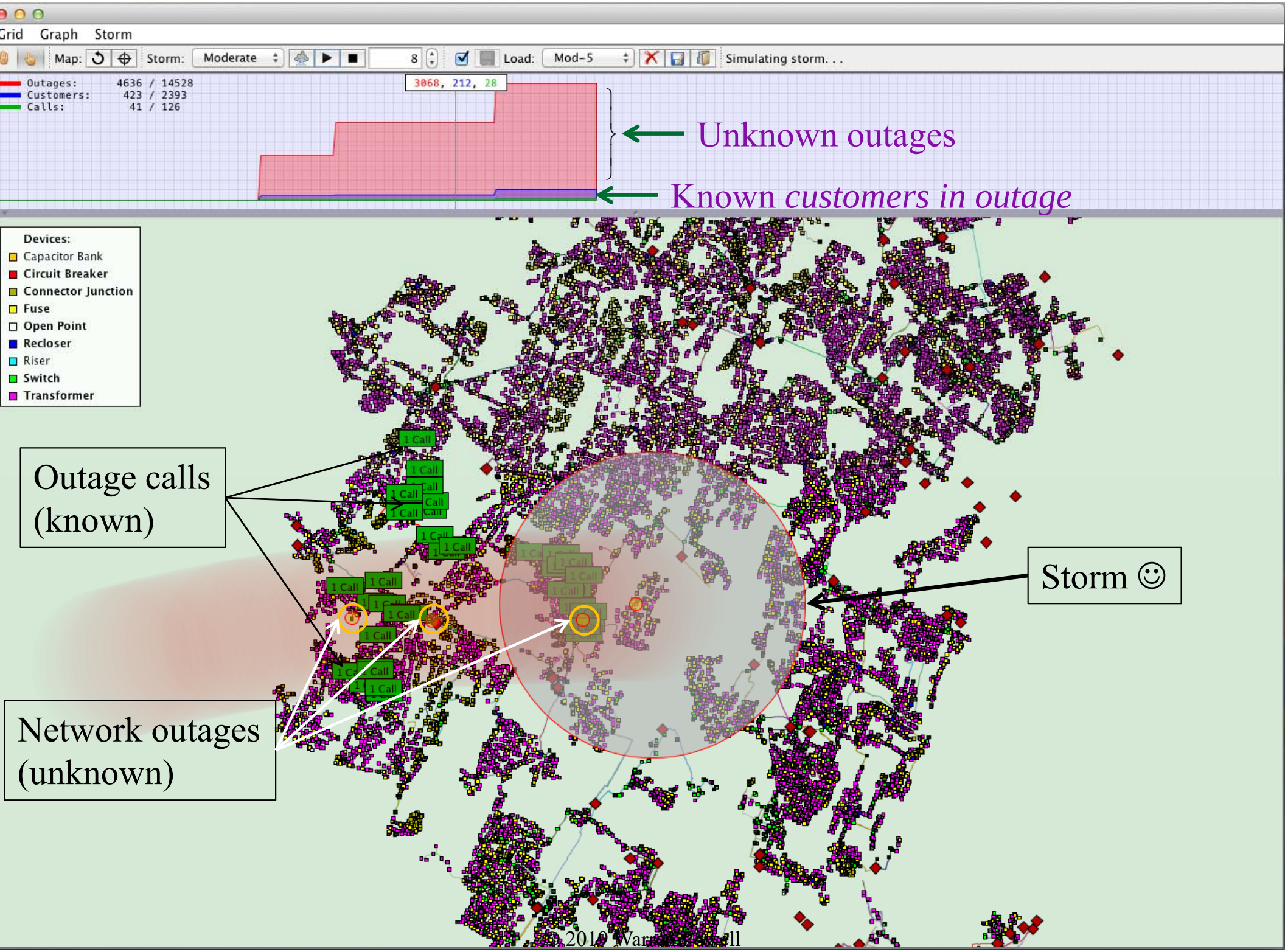
$$A_t x_t - B_{t-1} x_{t-1} = b_t, \quad t \geq 1.$$

- We can convert this to a proper stochastic model by replacing x_t with $X_t^\pi(S_t)$ and taking an expectation:

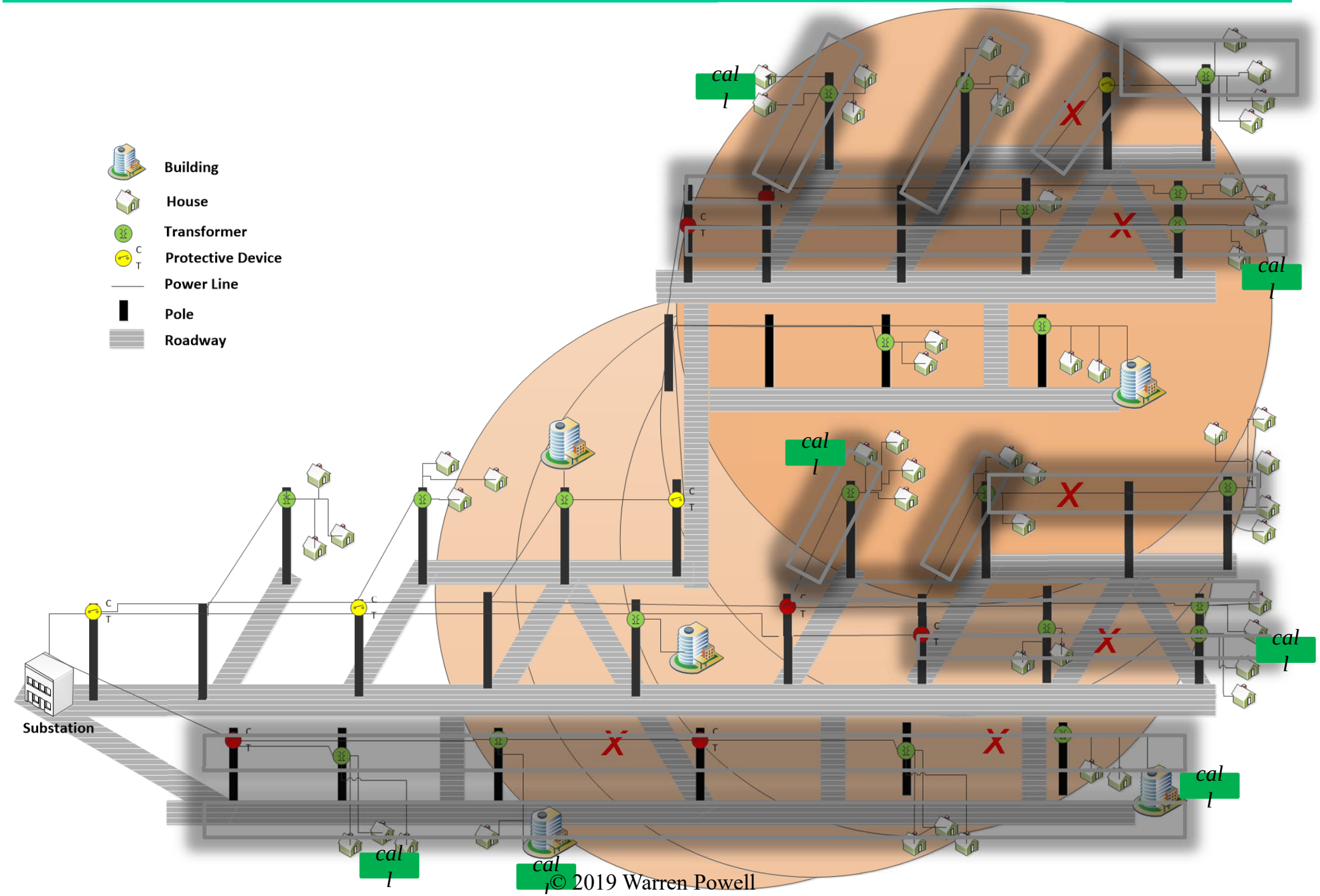
$$\min_\pi \mathbb{E} \sum_{t=0}^{T} c_t X_t^\pi(S_t)$$

The policy $X_t^\pi(S_t)$ has to satisfy $A_t x_t = R_t$ with transition function:

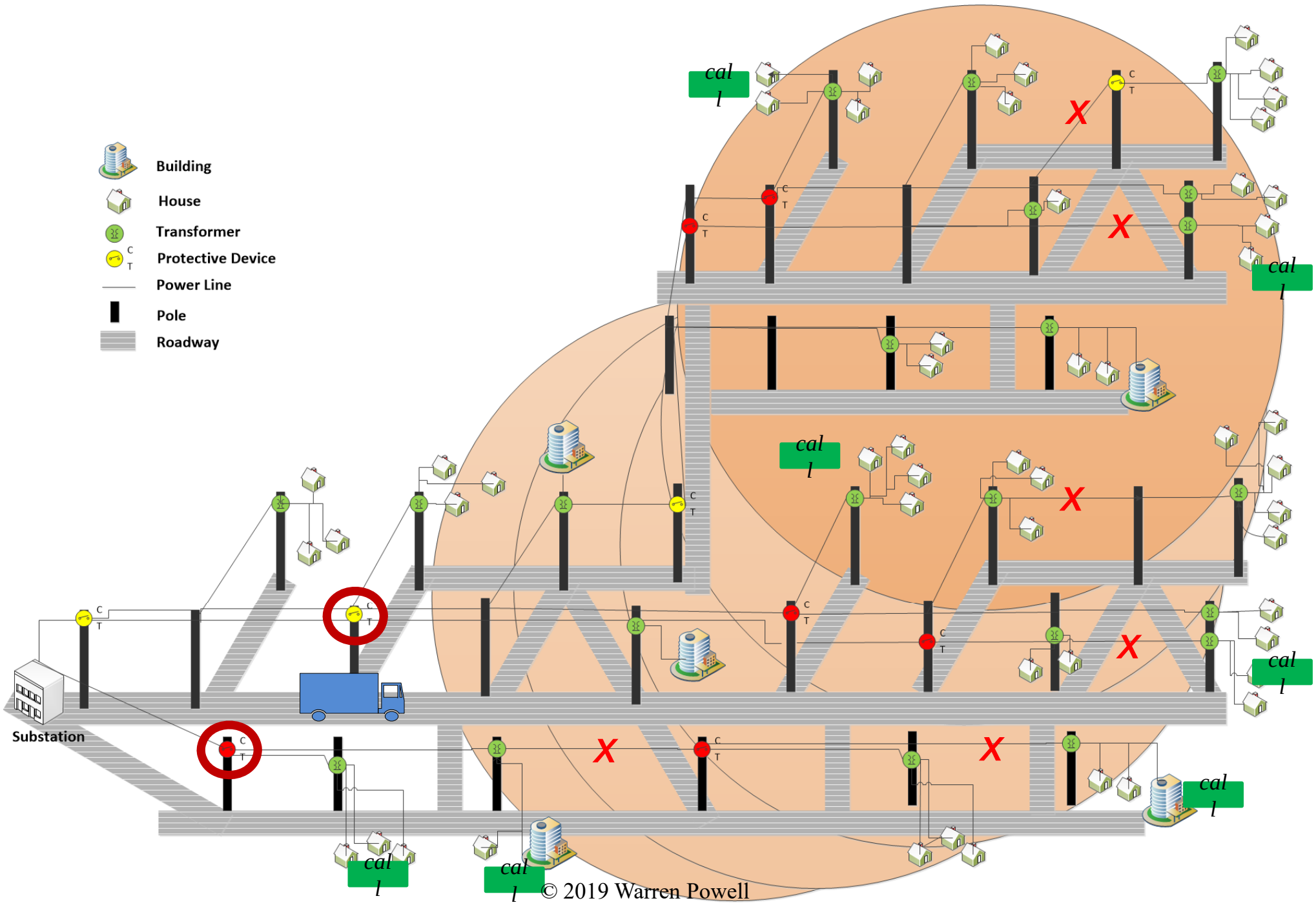
$$S_{t+1} = S^M\left(S_t, x_t, W_{t+1}\right)$$


Grid Graph Storm
Map: Storm: Moderate 8 Load: Mod-5 Simulating storm. . .
Outages: 4636 / 14528
Customers: 423 / 2393
Calls: 41 / 126
3068, 212, 28
Unknown outages
Known customers in outage
Devices:
Capacitor Bank
Circuit Breaker
Connector Junction
Fuse
Open Point
Recloser
Riser
Switch
Transformer
Outage calls (known)
Storm ☺
Network outages (unknown)
1 Call

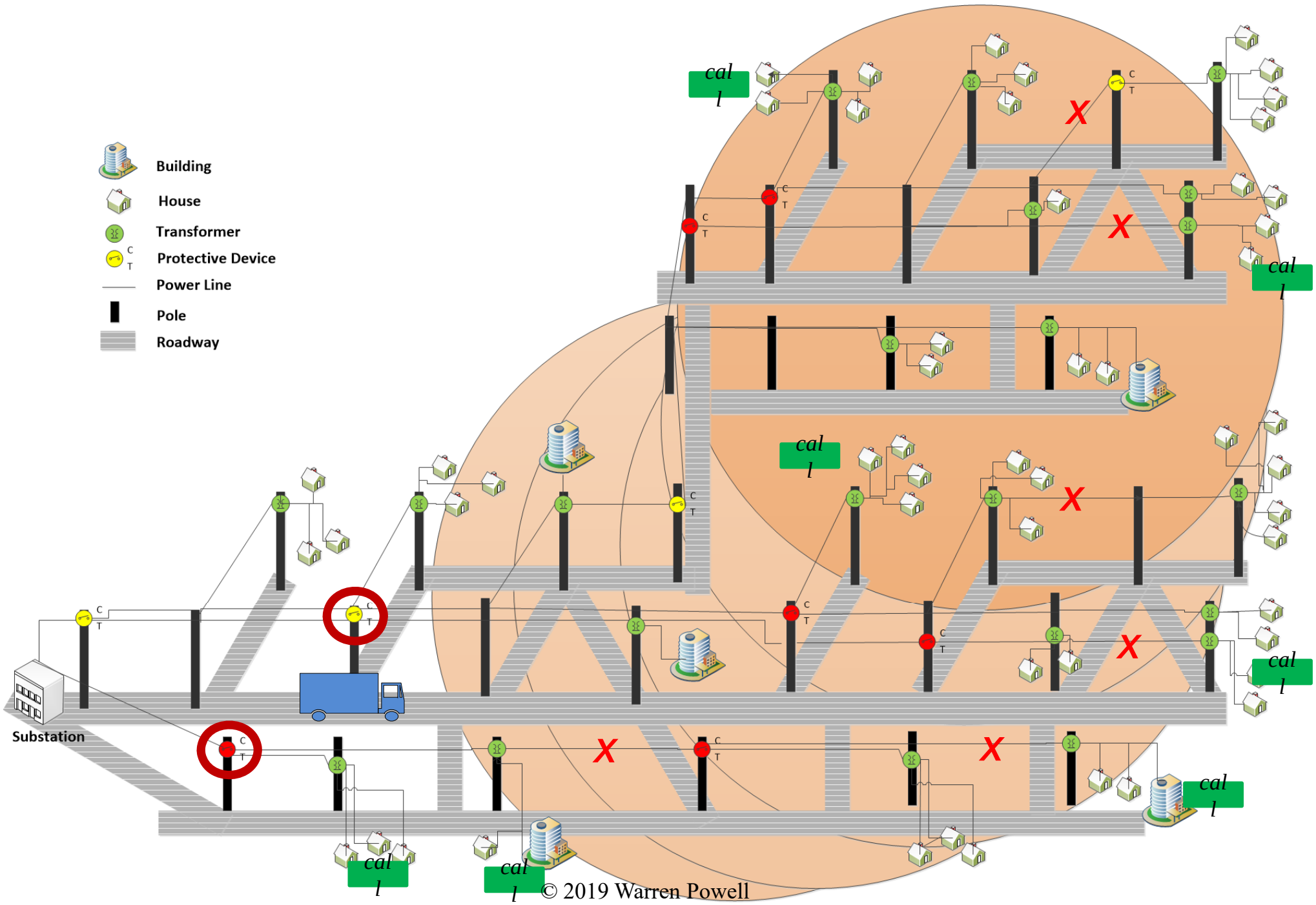
Problem Description - Emergency Storm Response



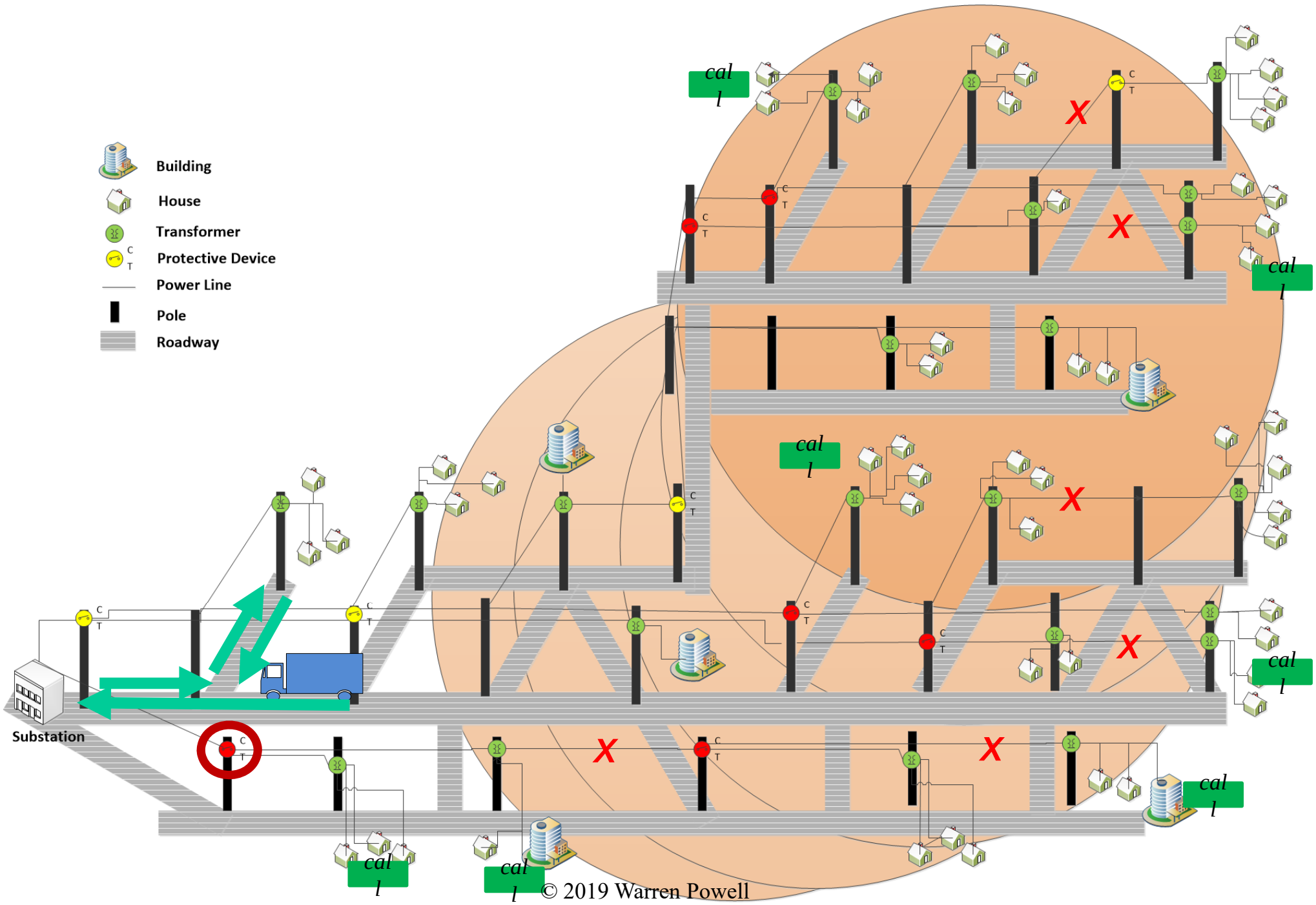
Escalation Algorithm



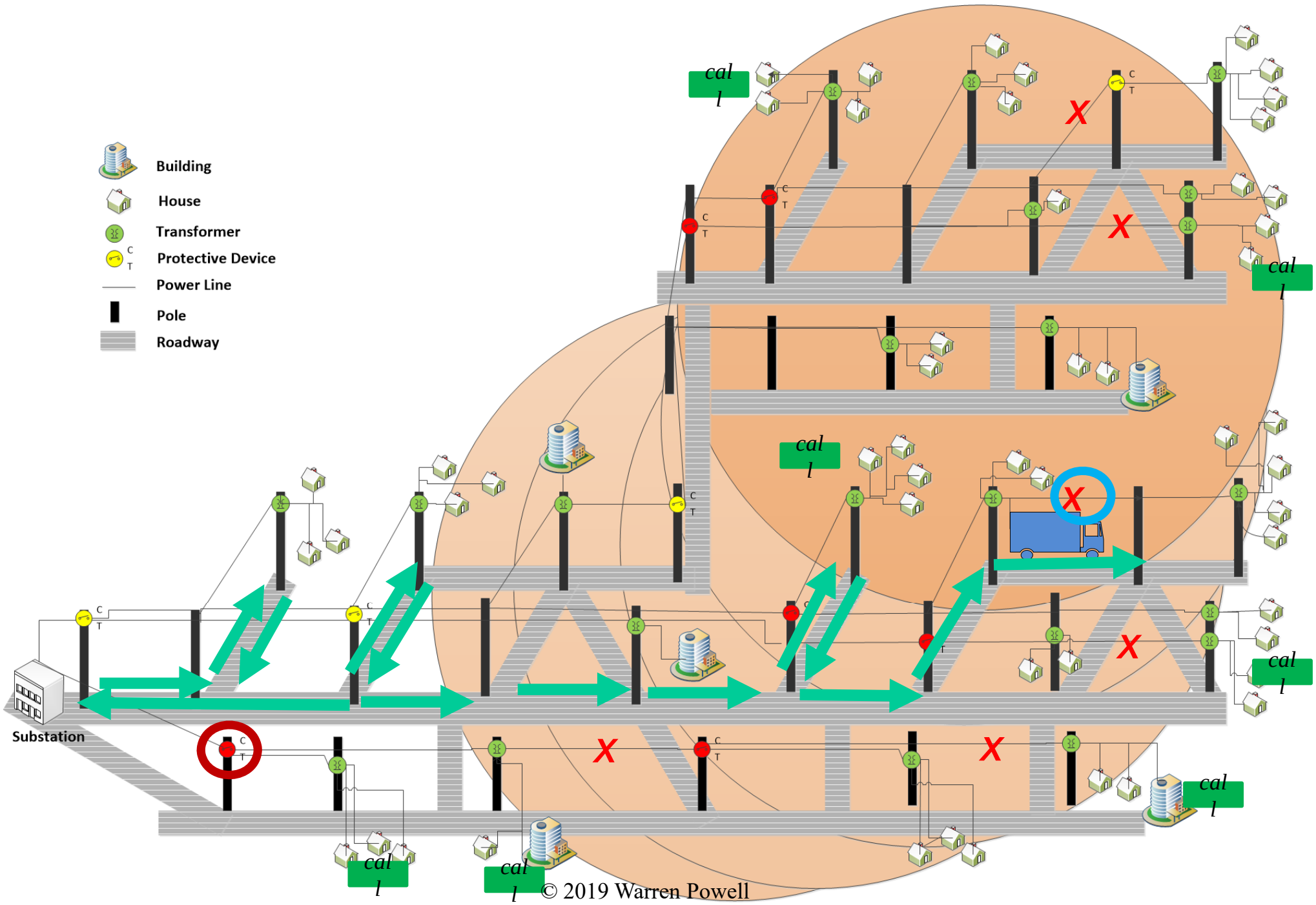
Escalation Algorithm



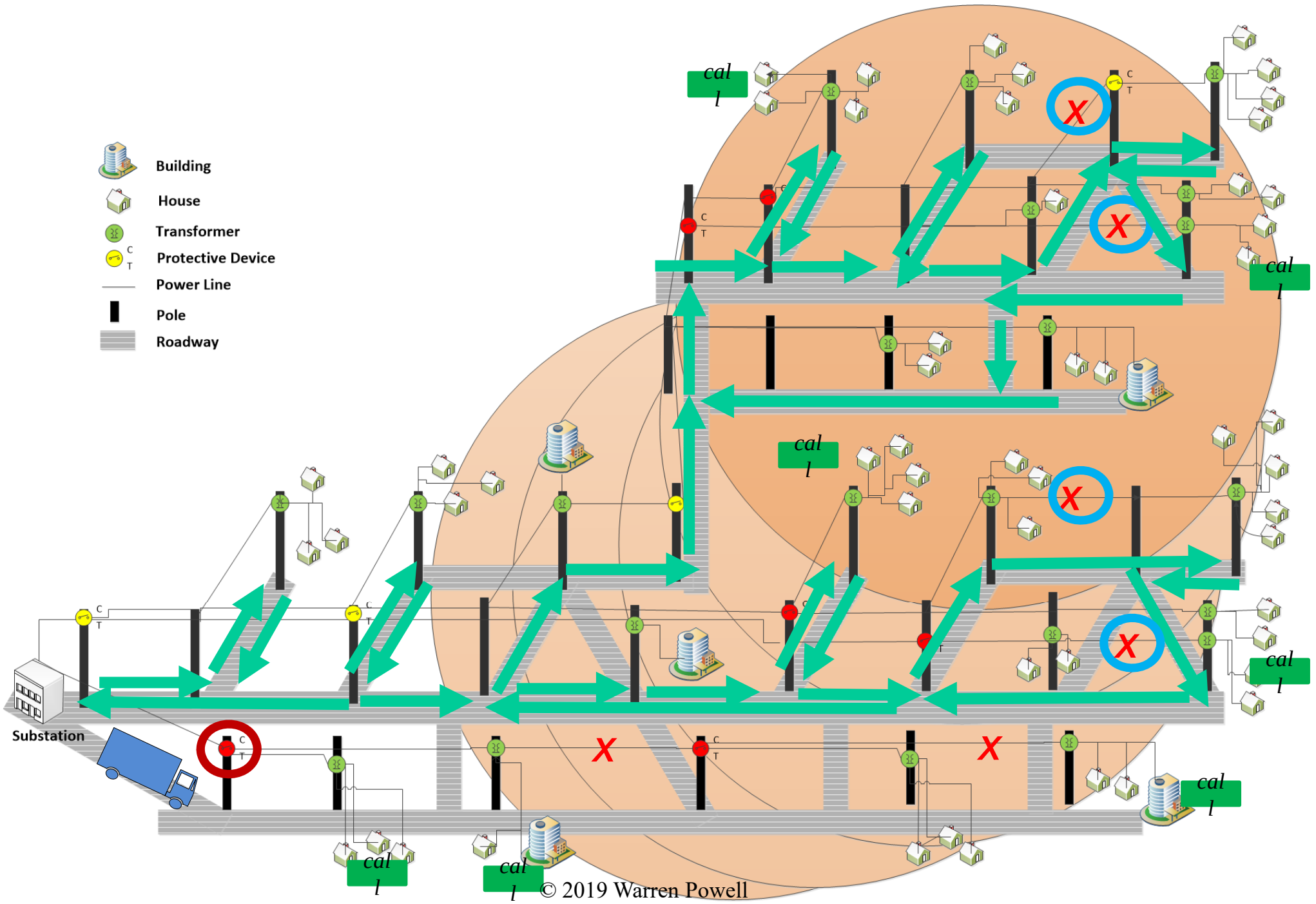
Escalation Algorithm



Escalation Algorithm

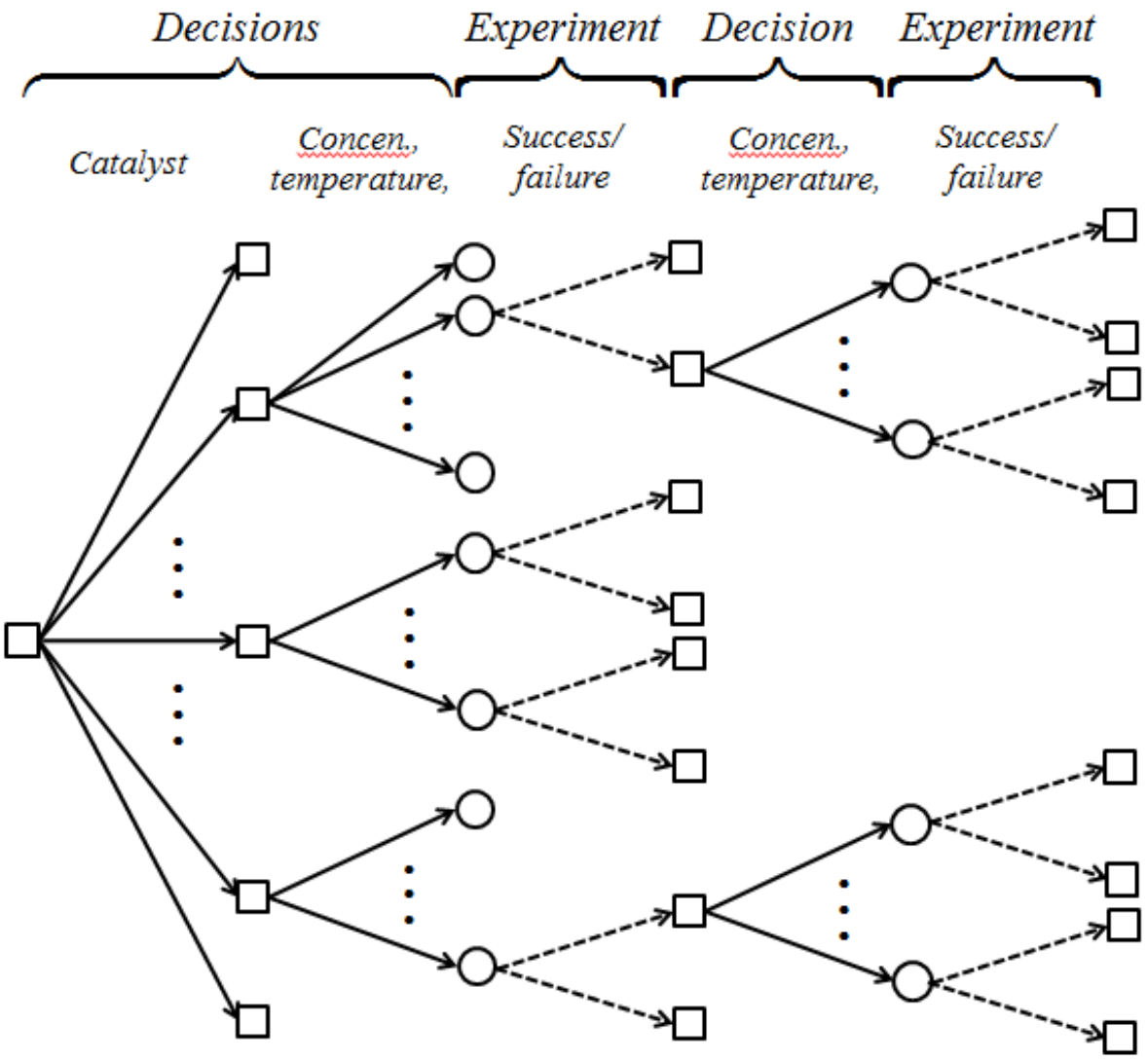


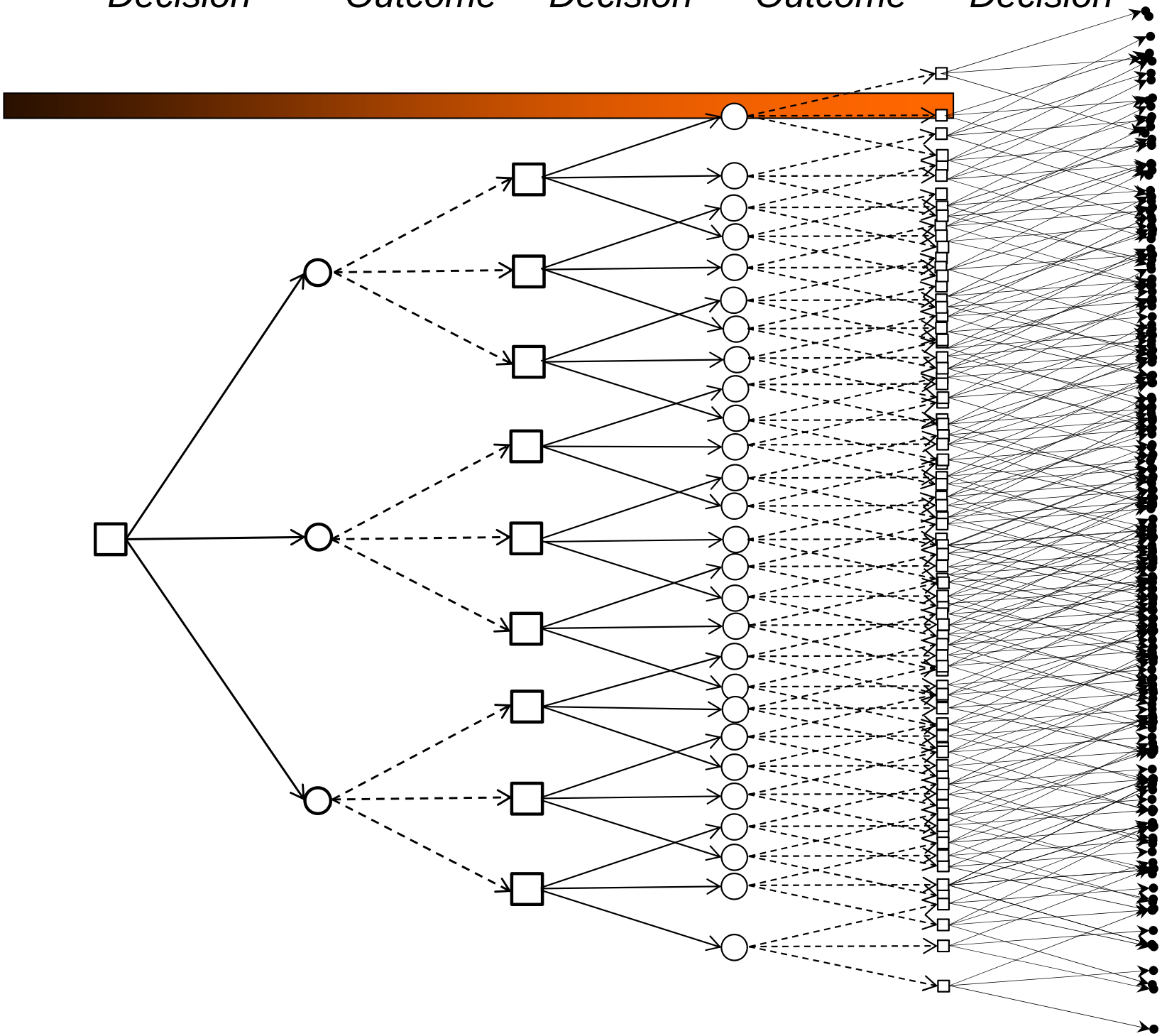
Escalation Algorithm



Lookahead policies

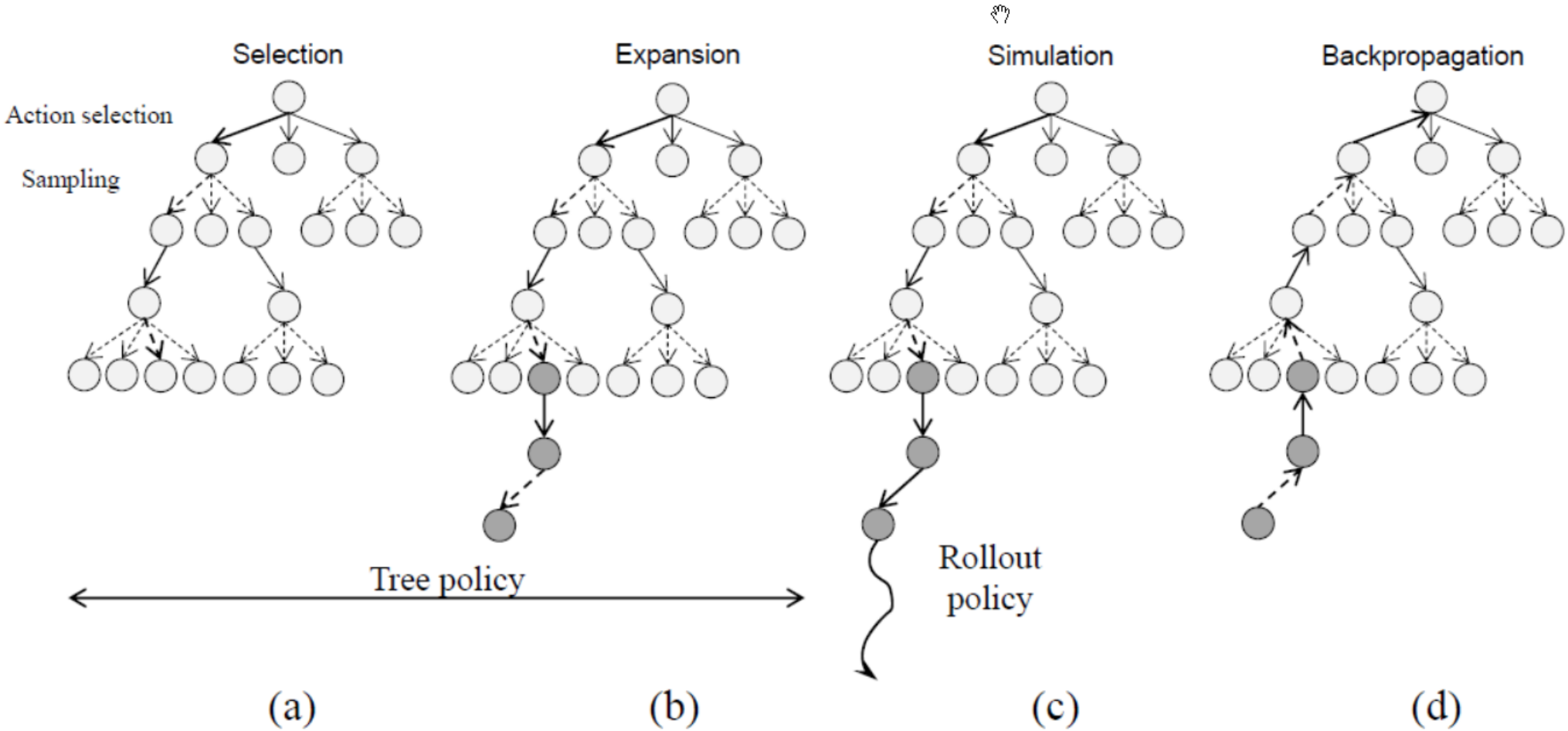
- Decision trees:



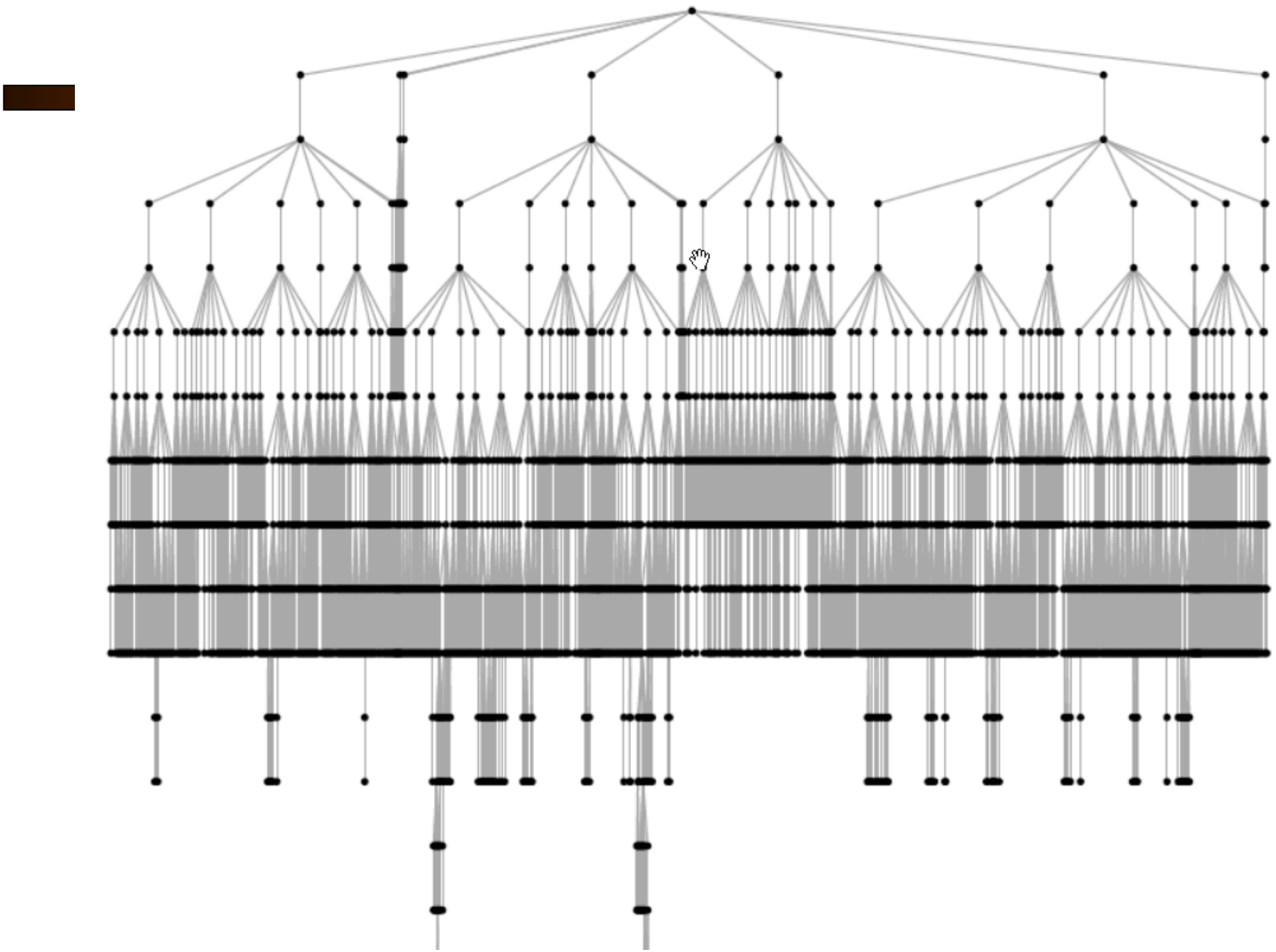


Monte Carlo tree search

- Steps of MCTS:

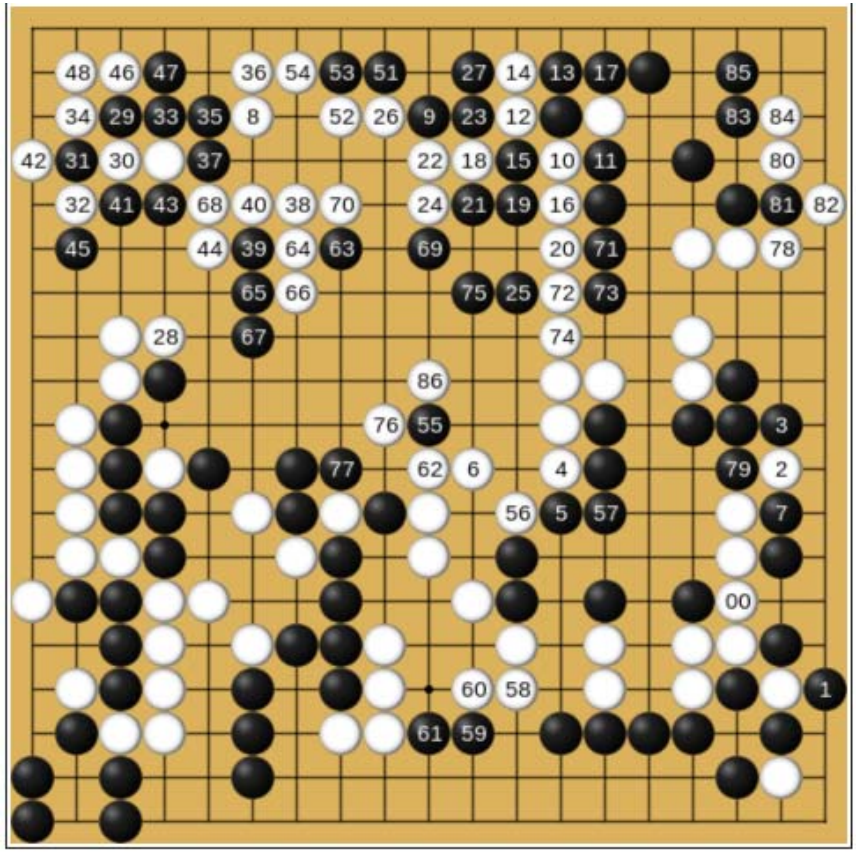


C. Browne, E. Powley, D. Whitehouse, S. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis and S. Colton, "A survey of Monte Carlo tree search methods," IEEE Transactions on Computational Intelligence and AI in Games, vol. 4, no. 1, pp. 1–49, March 2012.



Monte Carlo tree search

- AlphaGo
  - » Much more complex state space.
  - » Uses hybrid of policies:
    - MCTS
    - PFA
    - VFA

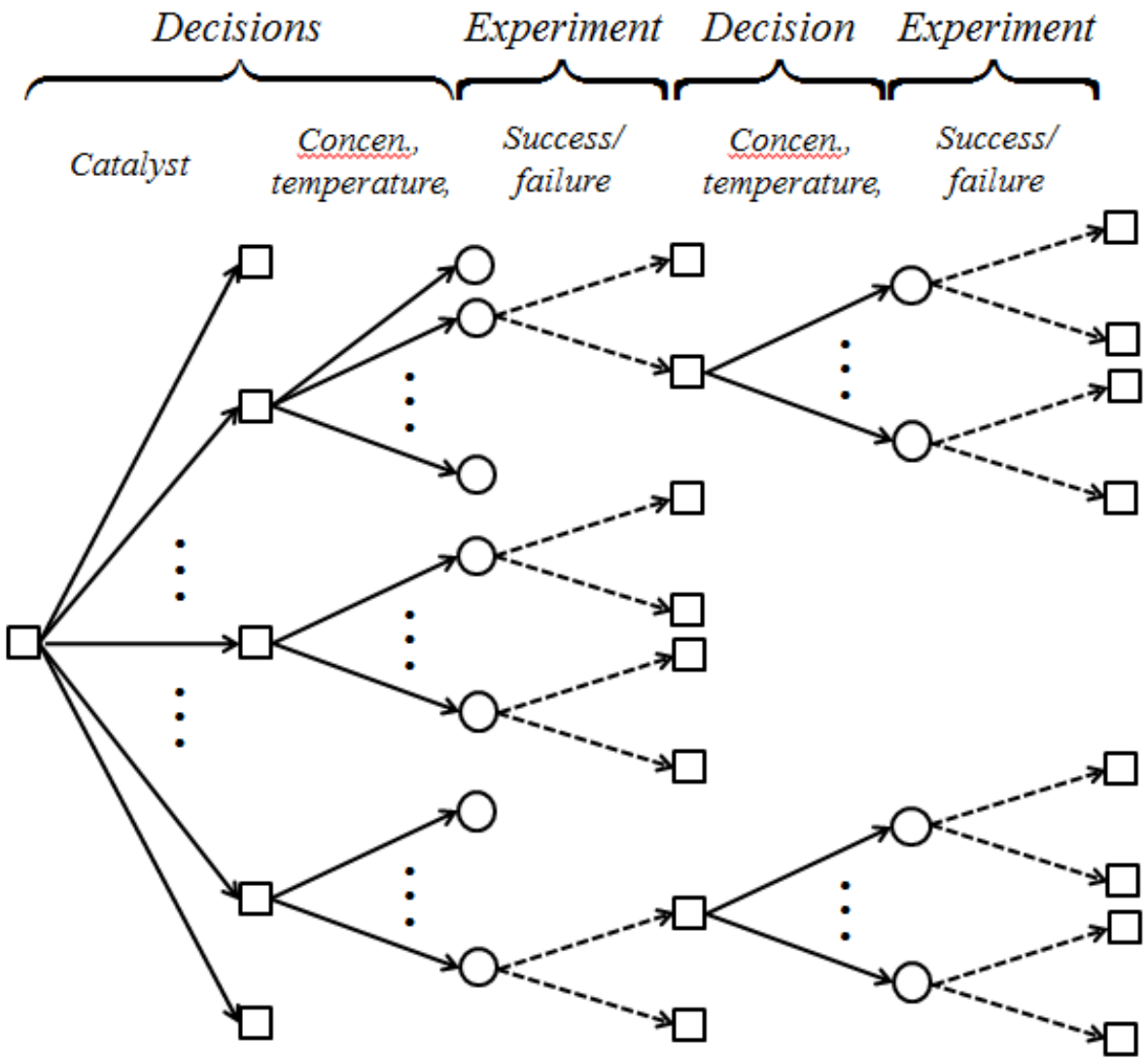


Canonical problems


John R. Birge
François Louveaux
Introduction to Stochastic Programming
Second Edition
Michael C. Fu Editor
Handbook of Simulation Optimization
Princeton Series in Applied Mathematics
Robust Optimization
Introduction to Decision Analysis
A Practitioner's Guide to Improving Decision Quality
SECOND EDITION
Approximate Dynamic Programming
Solving the Curses of Dimensionality
Warren B. Powell
VOLUME 2 • 4th EDITION
Dynamic Programming and Optimal Control
APPROXIMATE DYNAMIC PROGRAMMING
Optimal Learning
SECOND EDITION
Model Predictive Control
OPTIMAL CONTROL
Dimitri P. Bertsekas
Athena Scientific
WILEY
INTRODUCTION TO STOCHASTIC SEARCH AND OPTIMIZATION
Estimation, Simulation, and Control
JAMES C. SPALL
MULTI-ARMED BANDIT ALLOCATION INDICES
SECOND EDITION
John Gittins, Kevin Glazebrook and Richard Weber
Reinforcement Learning
An Introduction
second edition
Richard S. Sutton and Andrew G. Barto
Markov Decision Processes
Discrete Stochastic Dynamic Programming
MARTIN L. PUTERMAN
Online Computation and Competitive Analysis
Allan Borodin
Ran El-Yaniv
STOCHASTIC SIMULATION OPTIMIZATION
An Optimal Computing Budget Allocation
Chun-Hung Chen • Loo Hay Lee
Vol. 1
43
Jiongmin Yong
Xun Yu Zhou
Stochastic Controls
Hamiltonian Systems and HJB Equations

Canonical problems

- Decision trees



Canonical problems

- Stochastic search (derivative based)
  - » Basic problem:

$$\max_x \mathbb{E}F(x,W)$$

Manufacturing network (x=design)
Unit commitment problem (x=day ahead decisions)
Inventory system (x=design, replenishment policy)
Battery system (x=choice of material)
Patient treatment cost (x=drug, treatments)
Trucking company (x=fleet size and mix)

  - » Stochastic gradient

$$x^{n+1} = x^n + \alpha_n \nabla_x F(x^n, W^{n+1})$$

  - » Asymptotic convergence:

$$\lim_{n\to\infty} \mathbb{E}F(x^n,W) \to \mathbb{E}F(x^*,W)$$

  - » Finite time performance

$$\max_\pi \mathbb{E}F(x^{\pi,n},W)$$ where π is an algorithm (or policy)

Canonical problems

- Ranking and selection (derivative free)
  - » Basic problem:

$$\max_{x\in\{x_1,\ldots,x_M\}} \mathbb{E}F(x,W)$$

  - » We need to design a policy $X^{\pi}(S^n)$ that finds a design given by $x^{\pi,N}$

$$\max_{\pi} \mathbb{E}F\left(x^{\pi,N},W\right)$$

  - » We refer to this objective as maximizing the *final reward.*

Canonical problems

- Multi-armed bandit problems
  - » We learn the reward from playing each "arm"
  - » We need to find a policy $X^\pi(S^n)$ for playing machine x that maximizes:

$$\max_\pi \mathbb{E}\sum_{n=0}^{N-1} F(X^\pi(S^n), W^{n+1})$$

where

W^{n+1} = "winnings" — New information

S^n = State of knowledge — What we know about each slot machine

$x^n = X^\pi(S^n)$ — Choose next "arm" to play

We refer to this problem as maximizing *cumulative reward.*

Canonical problems

- (Discrete) Markov decision processes
  - » Bellman's optimality equation

$$V_t(S_t) = \min_{a_t \in \mathcal{A}} \left(C(S_t, a_t) + \gamma \mathbb{E}\left\{ V_{t+1}(S_{t+1}) \,|\, S_t \right\} \right)$$

$$= \min_{a_t \in \mathcal{A}} \left(C(S_t, a_t) + \gamma \sum_{s'} p(S_{t+1} = s' \,|\, S_t, a_t) V_{t+1}(S_{t+1}) \right)$$

  - » This is also the same as solving

$$\min_{\pi} E\left\{ \sum_{t=0}^{T} C\left(S_t, X_t^{\pi}(S_t)\right) \,|\, S_0 \right\}$$

where the optimal policy has the form

$$X^{\pi}(S_t) = \arg\min_{x_t} \left(C(S_t, x_t) + \mathbb{E}\left\{ V_{t+1}(S_{t+1}) \,|\, S_t, x_t \right\} \right)$$

Value function approximations

- Reinforcement learning (computer science)
  - » Q-learning (for discrete actions)

$$\hat{q}^n(s^n,a^n) = r(s^n,a^n) + \gamma \max_{a'} \bar{Q}^{n-1}(s',a')$$

$$\bar{Q}^n(s^n,a^n) = (1-\alpha_{n-1})\bar{Q}^{n-1}(s^n,a^n) + \alpha_{n-1}\hat{q}^n(s^n,a^n)$$

Policy:

$$\pi(s) = \arg\max_a \bar{Q}^n(s,a)$$

  - » The second edition of *Reinforcement Learning* (Sutton and Barto, forthcoming) includes other solution approaches:
    - Policy search (Boltzmann policies)
    - Upper confidence bounding
    - Monte Carlo tree search

Canonical problems

- Optimal stopping I
  - » Model:
    - Exogenous process:

$$\omega = \left(p_1, p_2, ..., p_T \right) = \text{Sequence of stock prices}$$

    - Decision:

$$X_t(\omega) = \begin{cases} 1 & \text{If we stop and sell at time } t \\ 0 & \text{Otherwise} \end{cases}$$

    - Reward:

$$p_t = \text{Price received if we stop at time } t$$

  - » Optimization problem:

$$\max_\tau \mathbb{E} p_\tau X_\tau$$

where τ is a "stopping time" (or "F_t – measurable function")

Canonical problems

- Optimal stopping II
  - » Model:
    - Exogenous process:

$$\omega = \left(p_1, p_2, ..., p_T \right) = \text{Sequence of stock prices}$$

$$\bar{p}_t = (1-\alpha)\bar{p}_{t-1} + \alpha p_t$$

    - State:

$R_t = 1$ if we are holding asset, 0 otherwise.

$$S_t = (R_t, p_t, \bar{p}_t)$$

    - Policy:

$$X_t(S_t \mid \theta) = \begin{cases} 1 & p_t \geq \bar{p}_t + \theta \\ 0 & \text{Otherwise} \end{cases}$$

  - » Optimization problem:

$$\max_\pi \mathbb{E} \sum_{t=0}^{T} p_t X^\pi(S_t \mid \theta) = \max_\theta \mathbb{E} \sum_{t=0}^{T} p_t X^\pi(S_t \mid \theta)$$

Canonical problems

- Linear quadratic regulation (LQR)
  - » A popular optimal control problem in engineering involves solving:

$$\min_{u_0,\ldots,u_T} \mathbb{E}\sum_{t=0}^{T}\left((x_t)^T Q x_t + (u_t)^T R u_t\right)$$

  - » where:

x_t = State at time t

u_t = Control at time t (must be F_t – measurable)

$x_{t+1} = f(x_t, u_t) + w_t$ (w_t is random at time t)

  - » Possible to show that the optimal policy looks like:

$$U_t^*(x_t) = K_t x_t$$

where K_t is a complicated function of Q and R.

Canonical problems

- Stochastic programming
  - » A (two-stage) stochastic programming *problem*

$$\min_{x_0 \in X_0} c_0 x_0 + \mathbb{E} Q(x_0, \xi_1)$$

where

$$Q(x_0, \xi_1(\omega)) = \min_{x_1(\omega) \in X_1(\omega)} c_1(\omega) x_1(\omega)$$

  - » This is the canonical form of stochastic programming, which might also be written over multiple periods:

$$\min c_0 x_0 + \sum_{\omega \in \Omega} p(\omega) \sum_{t=1}^{T} c_t(\omega) x_t(\omega)$$

Canonical problems

- Stochastic programming
  - » A (two-stage) stochastic programming *policy*

$$\min_{x_t \in X_t} c_t x_t + \mathbb{E} Q(x_t, \xi_{t+1})$$

where

$$Q(x_t, \xi_{t+1}(\omega)) = \min_{x_{t+1}(\omega) \in X_{t+1}(\omega)} c_{t+1}(\omega) x_{t+1}(\omega)$$

  - » This is the canonical form of stochastic programming, which might also be written over multiple periods:

$$\min c_t x_t + \sum_{\omega_t \in \Omega_t} p(\omega_t) \sum_{t'=t+1}^{t+H} c_{tt'}(\omega) x_{tt'}(\omega)$$

Canonical problems

- A robust optimization *problem* would be written

$$\min_{x \in X} \max_{w \in \mathcal{W}(\theta)} F(x, w)$$

  - » This means finding the best design x for the worst outcome w in an "uncertainty set" $\mathcal{W}(\theta)$

  - » This has been adapted to multiperiod problems

$$\min_{x_0,\ldots,x_1} \max_{(w_0,\ldots,w_T) \in \mathcal{W}(\theta)} \sum_{t'=0}^{T} c_{t'}(w_{t'}) x_{t'}$$

Canonical problems

- A robust optimization *problem* would be written

$$\min_{x \in X} \max_{w \in \mathcal{W}(\theta)} F(x, w)$$

  - » This means finding the best design x for the worst outcome w in an "uncertainty set" $\mathcal{W}(\theta)$

  - » … but it is often used as a *policy*

$$\min_{x_t, \ldots, x_{t+H}} \max_{(w_t, \ldots, w_{t+H}) \in \mathcal{W}(\theta)} \sum_{t'=t}^{t+H} c_{t'}(w_{t'}) x_{t'}$$

Canonical problems

- Stochastic control (from text by Rene Carmona)

3.1.3 ▪ The Optimization Problem

We assume that we are given an $\mathcal{F}_T$-measurable random variable representing the *terminal cost*. It is assumed to be square integrable. Most often, it will be of the form $g(X_T)$, where $g : \Omega \times \mathbb{R}^d \hookrightarrow \mathbb{R}$ is $\mathcal{F}_T \times \mathcal{B}(\mathbb{R}^d)$-measurable, and of polynomial growth in $x \in \mathbb{R}^d$ uniformly in $\omega \in \Omega$. We also assume that the cost includes a *running cost* given by a function $f : [0,T] \times \Omega \times \mathbb{R}^d \times A \hookrightarrow \mathbb{R}$ satisfying the same assumptions (S1) and (S2) as the drift b. Finally, we define the cost functional J by

$$J(\alpha) = \mathbb{E}\left[\int_0^T f(s, X_s, \alpha_s)ds + g(X_T)\right], \qquad \alpha \in \mathbb{A}. \tag{3.4}$$

As explained earlier, the goal of a stochastic control problem is to find an admissible control $\alpha \in \mathbb{A}$ which minimizes the cost functional $J(\alpha)$. The cost functional J is often called the objective, or objective functional.

» This is mathematically correct, but does not suggest a path to computation, or even how to model a real problem.

Canonical problems

- Why do we need a unified framework?
  - » The classical frameworks and algorithms are *fragile*.
  - » Small changes to problems invalidate optimality conditions, or make algorithmic approaches intractable.
  - » Practitioners need robust approaches that will provide high quality solutions for all problems.

A modeling framework

Modeling

- We propose to model problems along five fundamental dimensions:
  - » State variables
  - » Decision variables
  - » Exogenous information
  - » Transition function
  - » Objective function
  - » This framework draws heavily from Markov decision processes and the control theory communities, but it is not the standard form used anywhere.

Modeling dynamic problems

- The system state:

Controls community

x_t = "Information state"

Operations research/MDP/Computer science

$S_t = (R_t, I_t, K_t)$ = System state, where:

R_t = Resource state (physical state)

Location/status of truck/train/plane

Energy in storage

I_t = Information state

Prices

Weather

K_t = Knowledge state ("belief state")

Belief about traffic delays

Belief about the status of equipment

Bizarrely, only the controls community has a tradition of actually defining state variables. We return to state variables later.

Modeling dynamic problems

- Decisions:

Markov decision processes/Computer science

a_t = Discrete action

Control theory

u_t = Low-dimensional continuous vector

Operations research

x_t = Usually a discrete or continuous but high-dimensional vector of decisions.

At this point, we do not specify *how* to make a decision. Instead, we define the function $X^\pi(s)$ (or $A^\pi(s)$ or $U^\pi(s)$), where π specifies the type of policy. "π" carries information about the type of function f, and any tunable parameters $\theta \in \Theta^f$.

Problem classes

- Types of decisions
  - » Binary
    $$x \in X = \{0,1\}$$
  - » Finite
    $$x \in X = \{1,2,...,M\}$$
  - » Continuous scalar
    $$x \in X = [a,b]$$
  - » Continuous vector
    $$x = (x_1,...,x_K), \quad x_k \in \mathbb{R}$$
  - » Discrete vector
    $$x = (x_1,...,x_K), \quad x_k \in \mathbb{Z}$$
  - » Categorical
    $$x = (a_1,...,a_I), \quad a_i \text{ is a category (e.g. red/green/blue)}$$

Modeling dynamic problems

- Exogenous information:

W_t = New information that first became known at time t

$= \left(\hat{R}_t, \hat{D}_t, \hat{p}_t, \hat{E}_t\right)$

$\hat{R}_t$ = Equipment failures, delays, new arrivals
New drivers being hired to the network

$\hat{D}_t$ = New customer demands

$\hat{p}_t$ = Changes in prices

$\hat{E}_t$ = Information about the environment (temperature, ...)

Note: Any variable indexed by t is known at time t. This convention, which is not standard in control theory, dramatically simplifies the modeling of information.

Below, we let ω represent a sequence of actual observations $W_1, W_2, \ldots$.

$W_t(\omega)$ refers to a sample realization of the random variable W_t.

Modeling dynamic problems

- The transition function

$$S_{t+1} = S^M(S_t, x_t, W_{t+1})$$

$$R_{t+1} = R_t + x_t + \hat{R}_{t+1} \quad \text{Inventories}$$

$$p_{t+1} = p_t + \hat{p}_{t+1} \quad \text{Spot prices}$$

$$D_{t+1} = D_t + \hat{D}_{t+1} \quad \text{Market demands}$$

Also known as the:

- "System model"
- "State transition model"
- "Plant model"
- "Plant equation"
- "Transition law"
- "Transfer function"
- "Transformation function"
- "Law of motion"
- "Model"

For many applications, these equations are unknown. This is known as "model-free" dynamic programming.

Modeling dynamic problems

- The objective function

Dimensions of objective functions

- » Type of performance metric
- » Final cost vs. cumulative cost
- » Expectation or risk measures
- » Mathematical properties (convexity, monotonicity, continuity, unimodularity, …)
- » Time to compute (fractions of seconds to minutes, to hours, to days or months)

Elements of a dynamic model

- Objective functions
  - » Cumulative reward ("online learning")

$$\max_\pi \mathbb{E}\left\{\sum_{t=0}^{T} C_t\left(S_t, X_t^\pi(S_t), W_{t+1}\right) \mid S_0\right\}$$

    - Policies have to work well *over time.*
  - » Final reward ("offline learning")

$$\max_\pi \mathbb{E}\left\{F(x^{\pi,N}, \hat{W}) \mid S_0\right\}$$

    - We only care about how well the final decision $x^{\pi,N}$ works.
  - » Risk

$$\max_\pi \rho\left\{C(S_0, X_0^\pi(S_0)), C(S_1, X_1^\pi(S_1)), \ldots, C(S_T, X_T^\pi(S_T)) \mid S_0\right\}$$

Elements of a dynamic model

- The complete model:
  - » Objective function
    - Cumulative reward ("online learning")

$$\max_{\pi} \mathbb{E}\left\{\sum_{t=0}^{T} C_t\left(S_t, X_t^{\pi}(S_t), W_{t+1}\right) \mid S_0\right\}$$

    - Final reward ("offline learning")

$$\max_{\pi} \mathbb{E}\left\{F(x^{\pi,N}, \hat{W}) \mid S_0\right\}$$

    - Risk:

$$\max_{\pi} \rho\left\{C(S_0, X_0^{\pi}(S_0)), C(S_1, X_1^{\pi}(S_1)), \ldots, C(S_T, X_T^{\pi}(S_T)) \mid S_0\right\}$$

  - » Transition function:

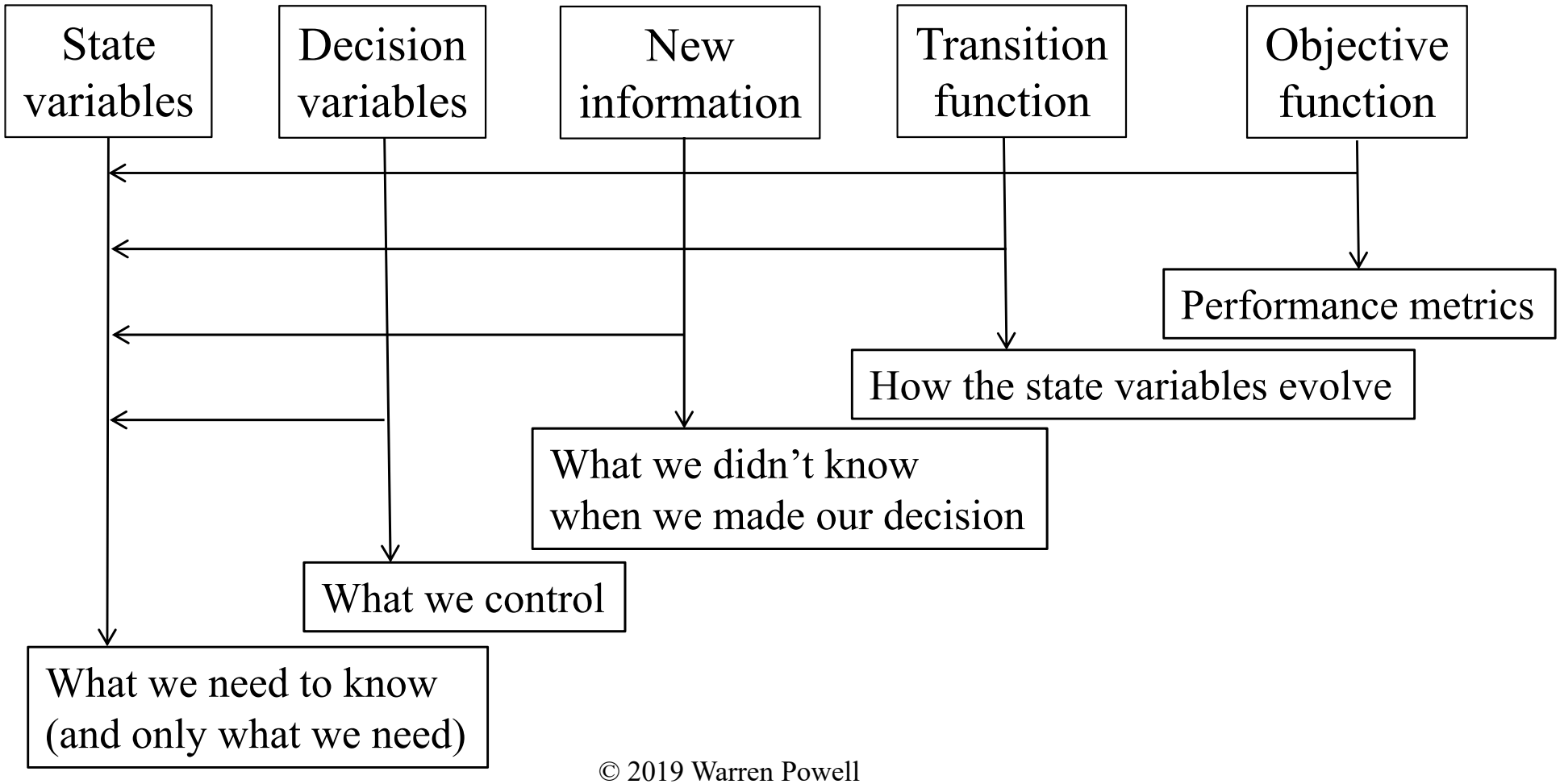
$$S_{t+1} = S^M\left(S_t, x_t, W_{t+1}\right)$$

  - » Exogenous information:

$$\left(S_0, W_1, W_2, \ldots, W_T\right)$$

Elements of a dynamic model

- The modeling process
  - » I conduct a conversation with a domain expert to fill in the elements of a problem:



Course overview

Course overview

- Requirements
  - » Weekly problem sets. These will emphasize modeling and the testing of different algorithmic strategies
  - » Take-home midterm and final
  - » "Diary problem" – You will propose a problem of your choosing, and we will periodically return to this problem to develop a model and suggest solution approaches.
  - » Active participation in class – We have a diverse class with different backgrounds and contexts. I need active discussion to relate the notation to your problems.

Course overview

- Programming

  » Most numerical work will require nothing more than Excel.

  » We have two libraries available for testing more advanced algorithms:
  - Python – This is a newer set of modules.
  - Matlab – We have an older package, MOLTE, that was developed to illustrate a wide range of algorithms.

Course overview

- Prerequisites
  - » ORF 544 primarily needs a basic course in statistics and probability. From this we will need:
    - General understanding of probability distributions and random variables.
    - Basic understanding of conditional probability and Bayes theorem.
    - In this course, we will emphasize *recursive statistics* which is covered in chapter 3.
  - » So why is this a grad course?
    - Translating real problems into notation (mathematical modeling) requires some patience and an interest in solving problems using analytics.
    - Stochastic modeling, which means modeling the flow of decisions and information, can be fairly subtle.

Course overview

- What you have to do.
  - » Most important, *ask questions!!!* If you are not going to ask questions, then just read the book!
  - » The course will follow the structure of the book *Stochastic Optimization and Learning*. I will present selected topics from most chapters, with the goal that you will feel comfortable reading the rest of the chapter on your own.
  - » You will best understand the material when you can relate the notation to problems you are familiar with.
  - » I will do my best to provide a range of applications, but it really helps when you contribute your own problems.