Distributed Power Management under Limited Communication

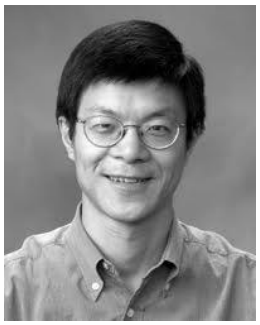
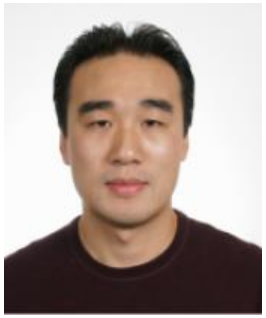
Na Li

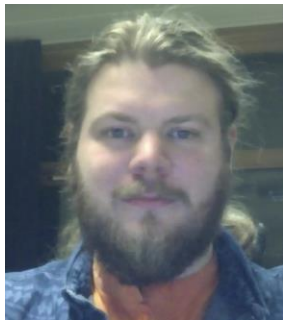
Harvard University

Rutgers, 08/22/2017

Acknowledgment :

Harvard Univ: Guannan Qu, Chinwendu Enyioha, Vahid Tarokh
KTH: Sindri Magnusson, Carlo Fishchione
Caltech: Steven Low
NREL: Changhong Zhao
Univ. of Colorado, Boulder: Lijun Chen

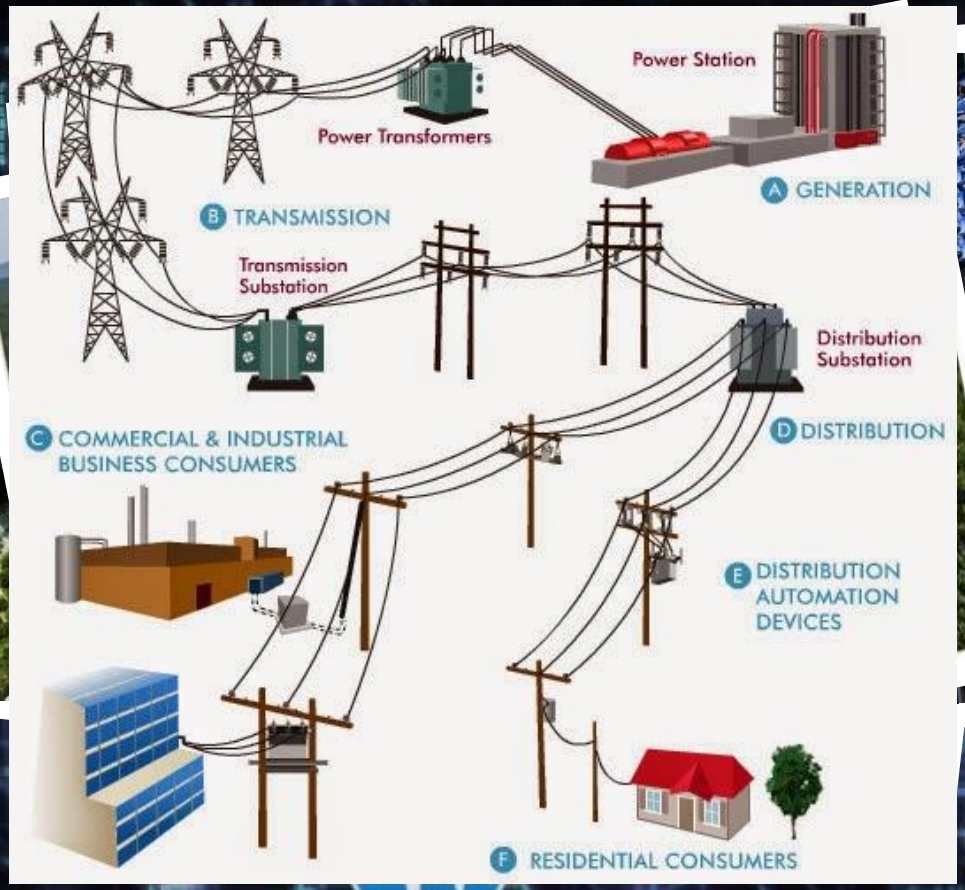
A Vision of Future(IoT)?

All devices are **connected** **and** **coordinated** to

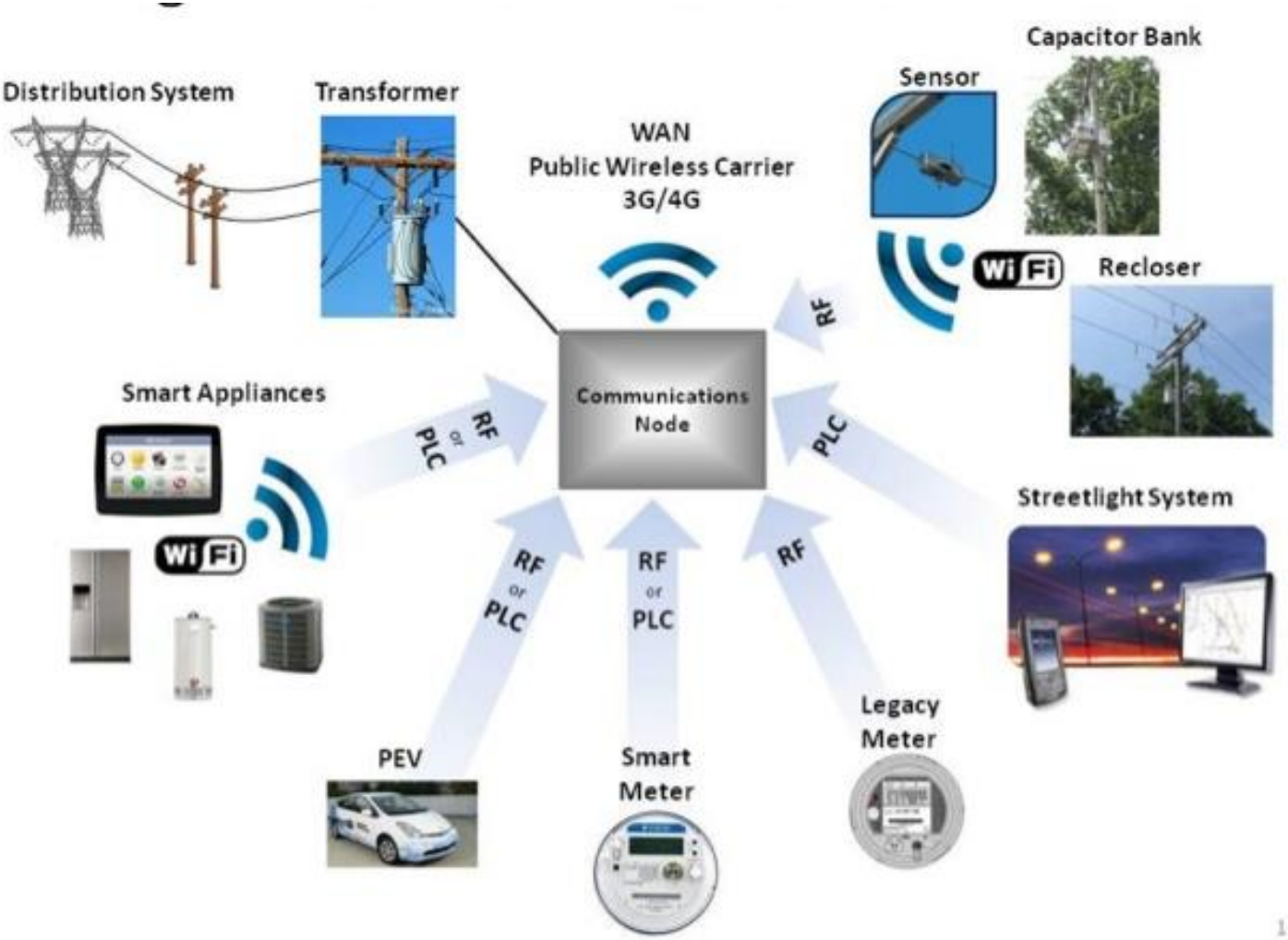
- Maximize social welfare
- Satisfy operation constraints

Distributed Optimization

Devices communicate, compute decisions, & communicate, …
until reach an efficient point (**Iterative**, **two-way** comm)



Sensing, Communication, Computation

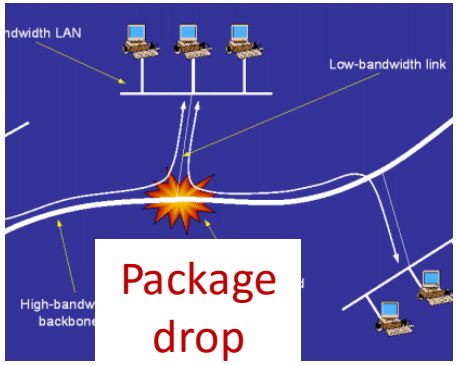


Intelligent Power Systems

Sources: Gigaom

Communication Challenges

- Lack reliability
- Unaccepted delays
- Vulnerable to malicious attacks
- Leak privacy
- Limited bandwidth
  (e.g. Power Line Comm.)
- High deployment cost
- …



How about reducing communication needs?

Reduce communication in power management

- ***Extract* information from *physical measurements (Feedback)***
  - Load frequency control
  - Power allocation in buildings/data centers

- ***Recover* information from *local computation***
  - Quantized dual gradient for power allocation

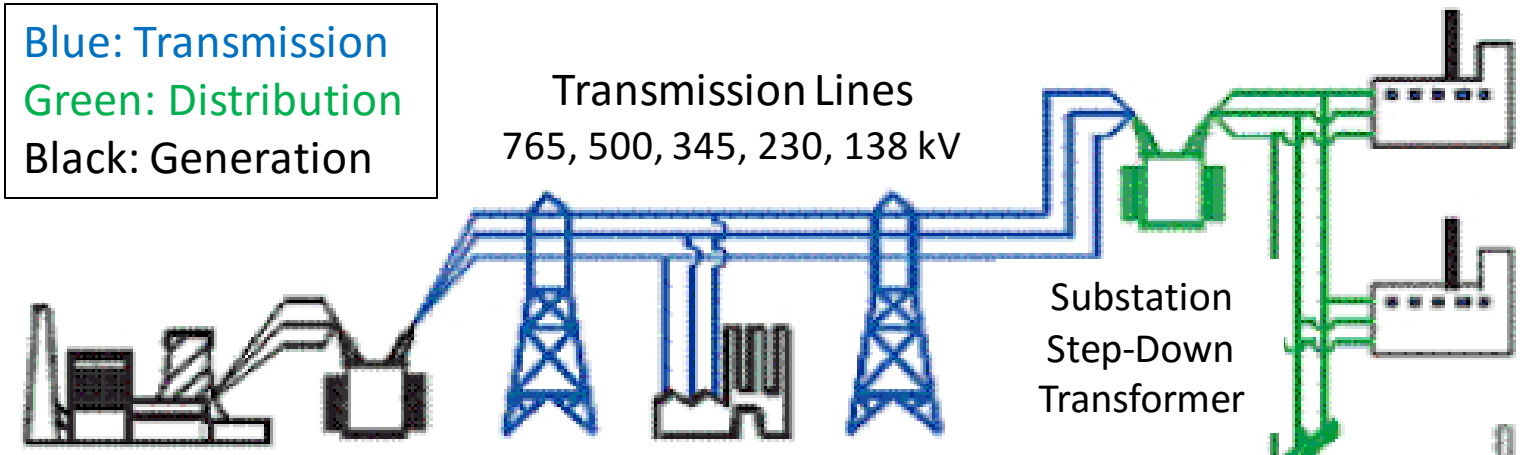
This talk: Limited communication in power systems

- ***Extra* information from *physical measurements (Feedback)***
  - Load frequency control
  - Power allocation in buildings/data centers

- ***Recover* information from *local computation***
  - Quantized dual gradient for power allocation

Power Systems

Blue: Transmission
Green: Distribution
Black: Generation



Supply = Demand

DR appliances

Storage

EV

Source: Graphic courtesy of North American Electric Reliability Corportion (NERC)

Optimal Load Control

- Balance total generation and load
- Keep frequency deviation small
- Minimize aggregate load disutility

$$\min \quad \sum_i \left(\sum_{l \in L(i)} C_l(d_l) + \tfrac{1}{2} D_i \omega_i^2 \right)$$

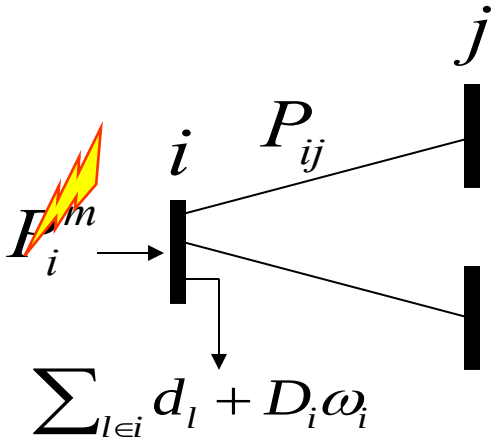
disutility

Freq deviation

$$\text{over} \quad d_l \in \left[\underline{d}_l, \bar{d}_l \right], \omega_i$$

$$\text{s. t.} \quad P_i^m = \sum_{l \in L(i)} d_l + D_i \omega_i + \sum_{j:i \to j} P_{ij} - \sum_{k:k \to i} P_{ki}$$

Power balance at each *i*

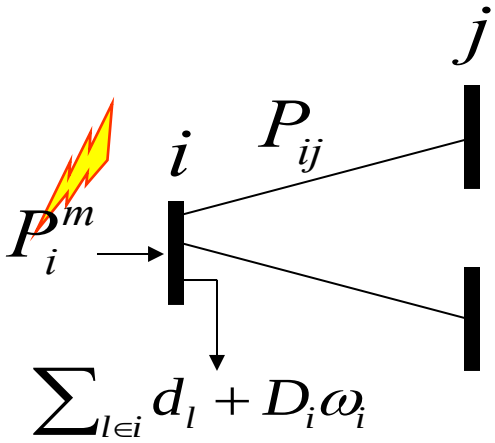


i: control area /aggregated bus

Distributed Optimization (e.g. ADMM) Applies.

But…

But…



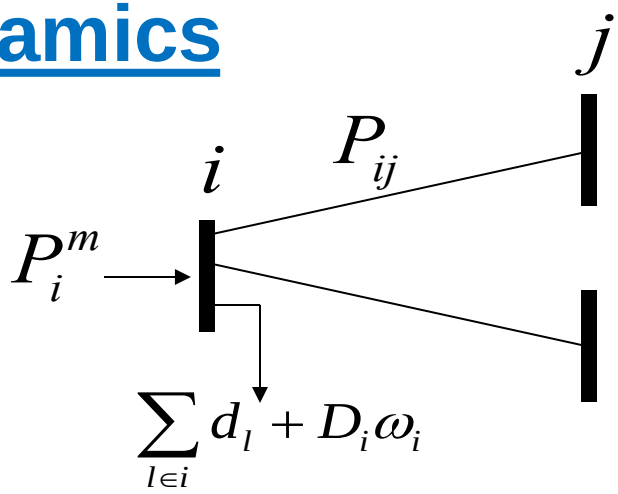
- Hard to get the real-time disturbance information
- Heavily relies on iterative communication

Can loads response in real-time and closed-loop?

➢ ***Network physical dynamics help!***

Physical dynamics: Swing Dynamics

i: Aggregated bus/control area/balance authority



$$M_i\dot{\omega}_i = P_i^m - \sum_{l\in i} d_l - D_i\omega_i - \sum_{j:i\to j} P_{ij} + \sum_{k:k\to i} P_{ki}$$

Inertia

frequency

Mechanical power

freq-insensitive loads

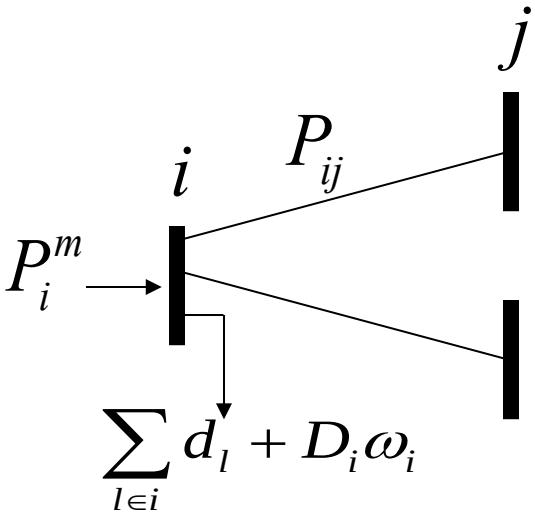
freq-sensitive load

power flow

Variables denote the deviations from their reference (steady state) values

Network dynamics

DC approximation of power flow



$$\dot{P}_{ij} = b_{ij}\left(\omega_i - \omega_j\right)$$

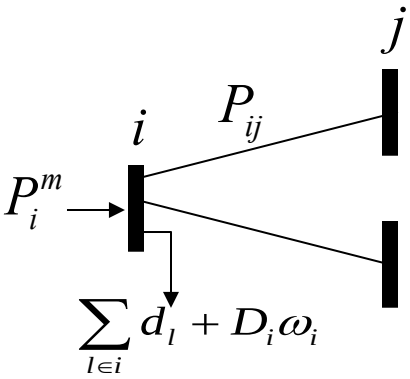
- Lossless (resistance=0)
- Fixed voltage magnitudes $|V_i|$
- Small deviation of angles

System Model Recap

$$M_i\dot{\omega}_i = P_i^m - \sum_{l\in i} d_l - D_i\omega_i - \sum_{j:i\to j} P_{ij} + \sum_{k:k\to i} P_{ki}$$

$$\dot{P}_{ij} = b_{ij}\left(\omega_i - \omega_j\right)$$

Load Control



?

$$\min \quad \sum_i \left(\sum_{l\in L(i)} C_l\left(d_l\right) + \tfrac{1}{2} D_i\omega_i^2 \right)$$

$$\text{over} \quad d_l \in \left[\underline{d}_l, \bar{d}_l\right], \omega_i$$

$$\text{s. t.} \quad P_i^m = \sum_{l\in L(i)} d_l + D_i\omega_i + \sum_{j:i\to j} P_{ij} - \sum_{k:k\to i} P_{ki}$$

Load frequency control

System Dynamics

$$M_i\dot{\omega}_i = P_i^m - \sum_{l\in i} d_l - D_i\omega_i - \sum_{j:i\to j} P_{ij} + \sum_{k:k\to i} P_{ki}$$

$$\dot{P}_{ij} = b_{ij}\left(\omega_i - \omega_j\right)$$

Dual Dynamics

Load Control

$$d_l(t) = \left[C_l'^{-1}\left(\omega_i(t)\right)\right]_{\underline{d}_l}^{\bar{d}_l} \quad \text{for } l \in L(i)$$

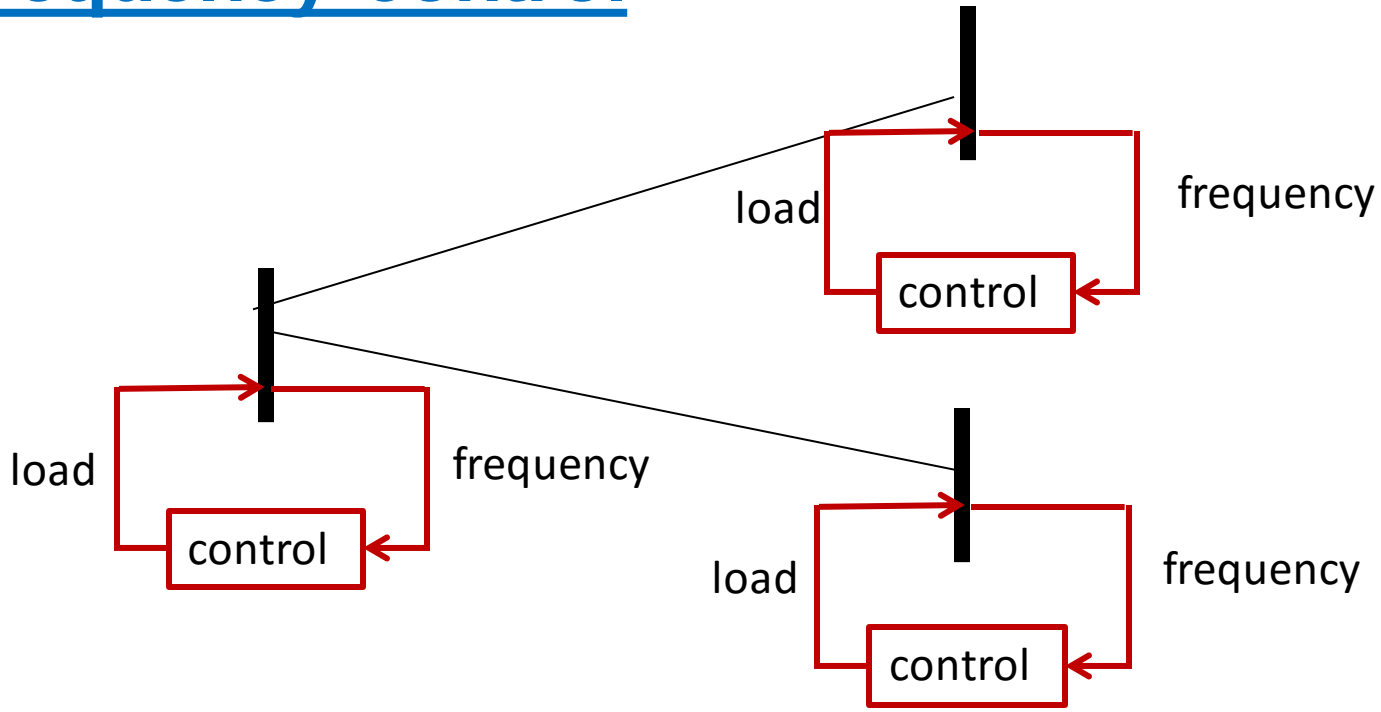
Converge to the optimal solution (Primal-Dual Gradient Flow)

$$\begin{aligned} \min \quad & \sum_i \left(\sum_{l\in L(i)} C_l\left(d_l\right) + \tfrac{1}{2} D_i\omega_i^2 \right) \\ \text{over} \quad & d_l \in \left[\underline{d}_l, \bar{d}_l\right], \omega_i \\ \text{s. t.} \quad & P_i^m = \sum_{l\in L(i)} c_l\left(d_l\right) + D_i\omega_i + \sum_{j:i\to j} P_{ij} - \sum_{k:k\to i} P_{ki} \end{aligned}$$

Optimal Load Control

Primal-Dual Gradient Flow: Arrow etc 1958, Feijer and Paganini 2010, Zhao, Low etc 2013, You, Chen etc 2014, Cherukuri, Mallada, Cortes, 2015, etc

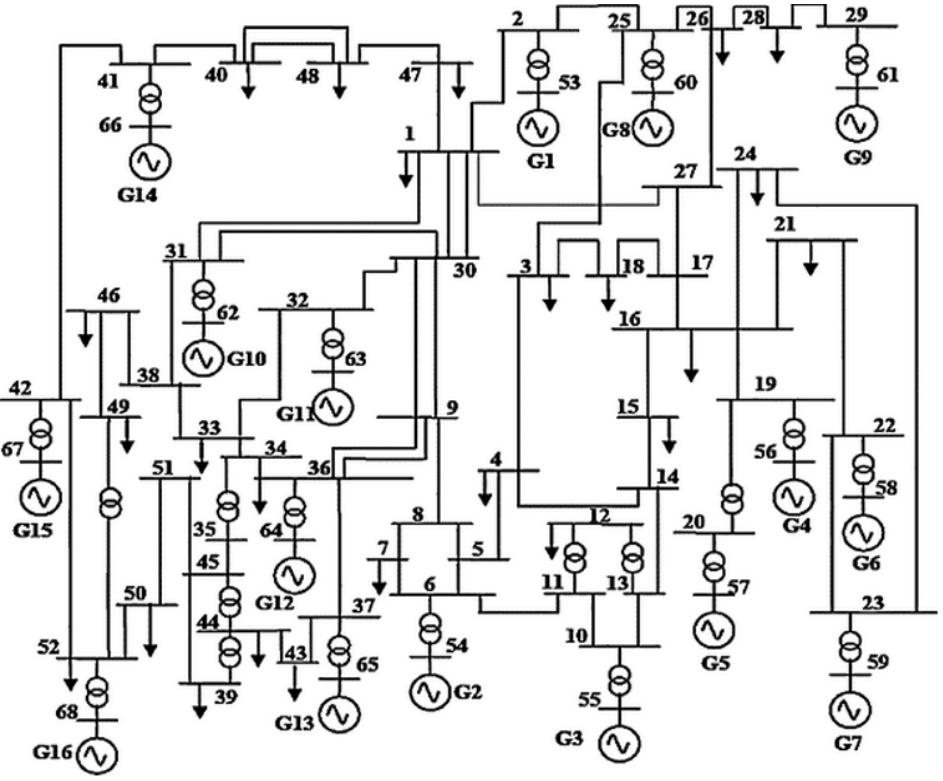
Load frequency control



- ***Frequency: a locally measurable signal ("price" of imbalance)***
- ***Completely decentralized; no explicit communication necessary***

Simulations

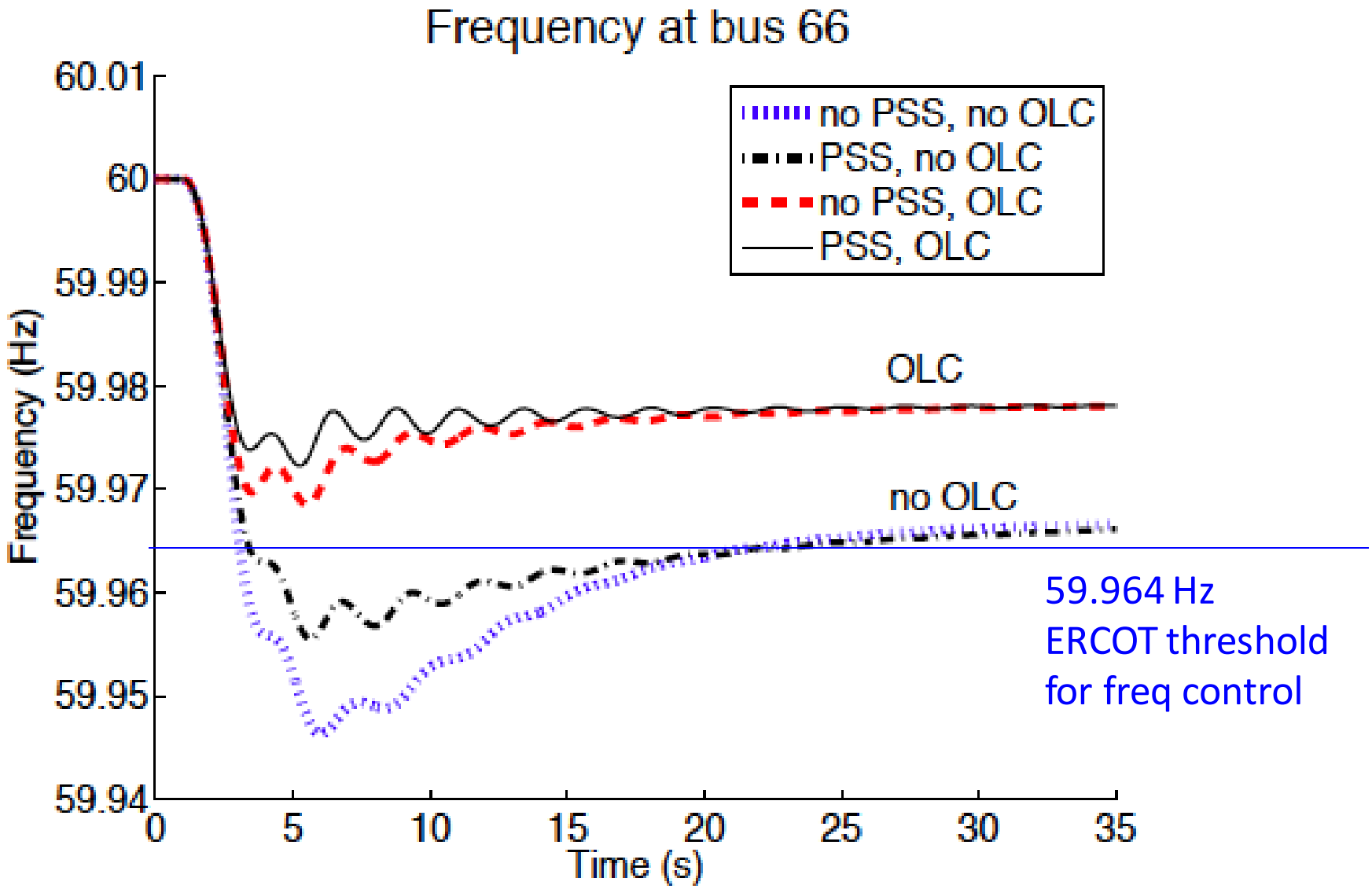
Dynamic simulation of IEEE 68-bus system (New England)



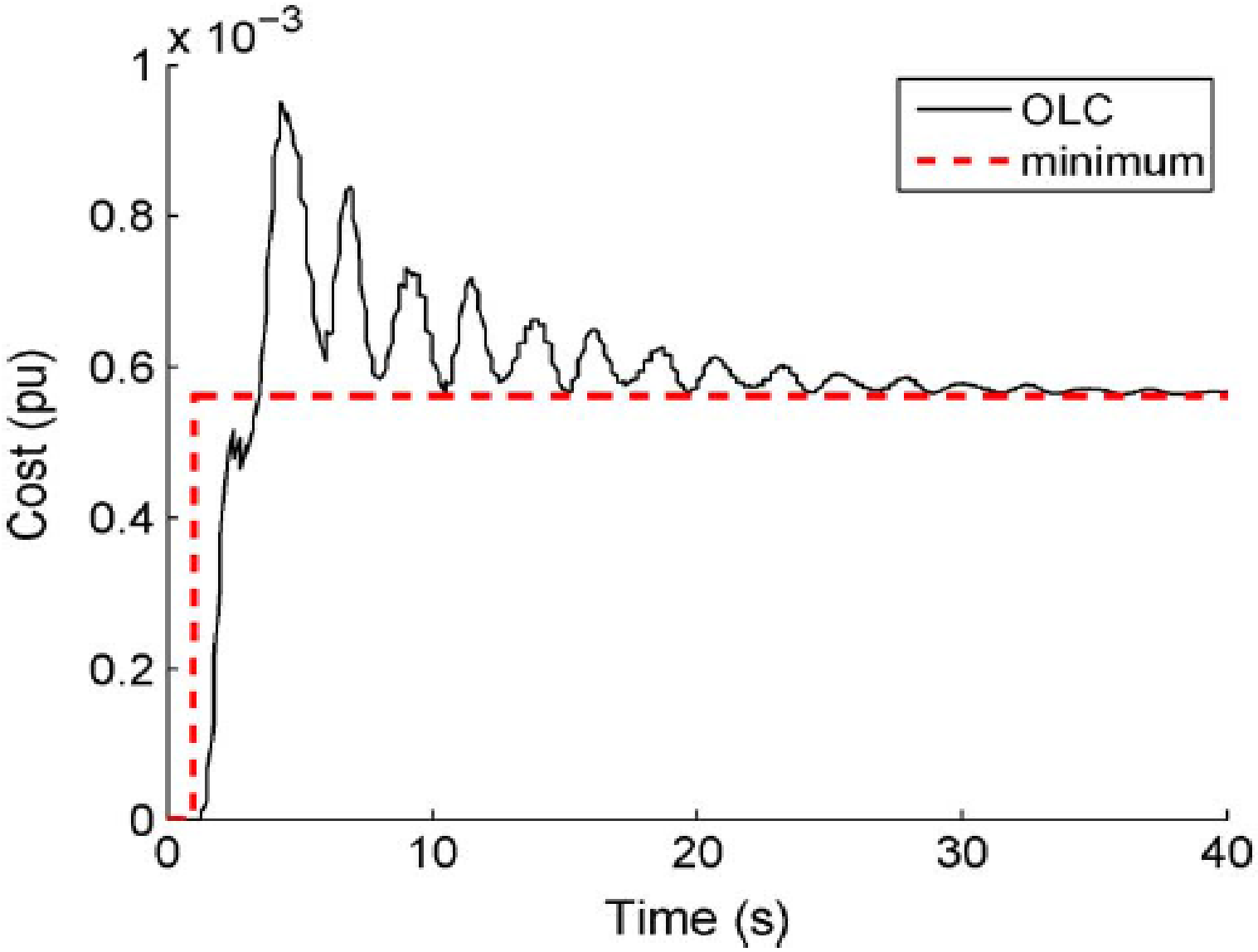
- Power System Toolbox (RPI)
- Detailed generation model
- Exciter model, power system stabilizer model
- Nonzero resistance lines

Sample rate 250ms
Step increase of loads on bus 1, 7, 27

Simulations



Simulations

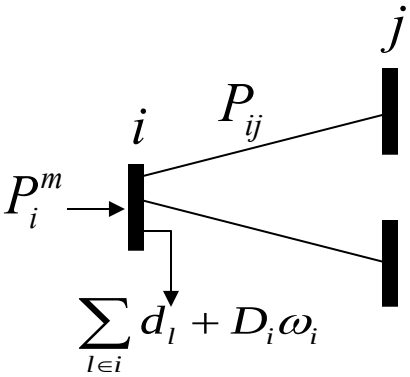


Recap

$$M_i\dot{\omega}_i = P_i^m - \sum_{l\in i} d_l - D_i\omega_i - \sum_{j:i\to j} P_{ij} + \sum_{k:k\to i} P_{ki}$$

$$\dot{P}_{ij} = b_{ij}\left(\omega_i - \omega_j\right)$$

Network Dynamics



Optimization

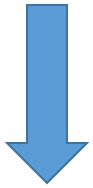
$$\min \quad \sum_i \left(\sum_{l\in L(i)} C_l\left(d_l\right) + \tfrac{1}{2} D_i\omega_i^2 \right)$$

$$\text{over} \quad d_l \in \left[\underline{d}_l, \bar{d}_l\right], \omega_i$$

$$\text{s. t.} \quad P_i^m = \sum_{l\in L(i)} d_l + D_i\omega_i + \sum_{j:i\to j} P_{ij} - \sum_{k:k\to i} P_{ki}$$

This Idea Extends to General Systems

Network Dynamics: $\dot{x} = \sum_{j:i\sim j} A_{ij}x_j + B_i u_i + C_i \omega_i$



How to design distributed, closed-loop controller u?

Optimization:

$$\min_{x_i, u_i} \sum_i f_i(x_i) + \sum_i g_i(u_i)$$

$$\text{s. t.} \quad \sum_{j:i\sim j} A_{ij}x_j + B_i u_i + C_i w_i = 0$$

$$h_i(x_i, u_i) \le 0$$

- [Li, Chen, Zhao, 2015]: Economic Automatic Generation Control
- [Zhang, Antonois, Li, 2016]: Sufficient and Necessary Conditions
- [Zhang, Malkawi, Li, 2016]: Thermal Control for HVAC

This talk: limited communication

- ***Extract* information from *physical measurements (Feedback)***
  - Load frequency control
  - Decentralized voltage control (distribution network) (Qu, Li, Dahleh, 2014)
  - Power allocation in buildings/data center

- ***Recover* information from *local computation***
  - Quantized dual gradient for power allocation

This talk: limited communication

- ***Extract* information from *physical measurements (Feedback)***
  - Load frequency control
  - Decentralized voltage control (distribution network) (Qu, Li, Dahleh, 2014)
  - Power allocation in buildings/data center

- ***Recover* information from *local computation***
  - Quantized dual gradient for power allocation

This talk: limited communication

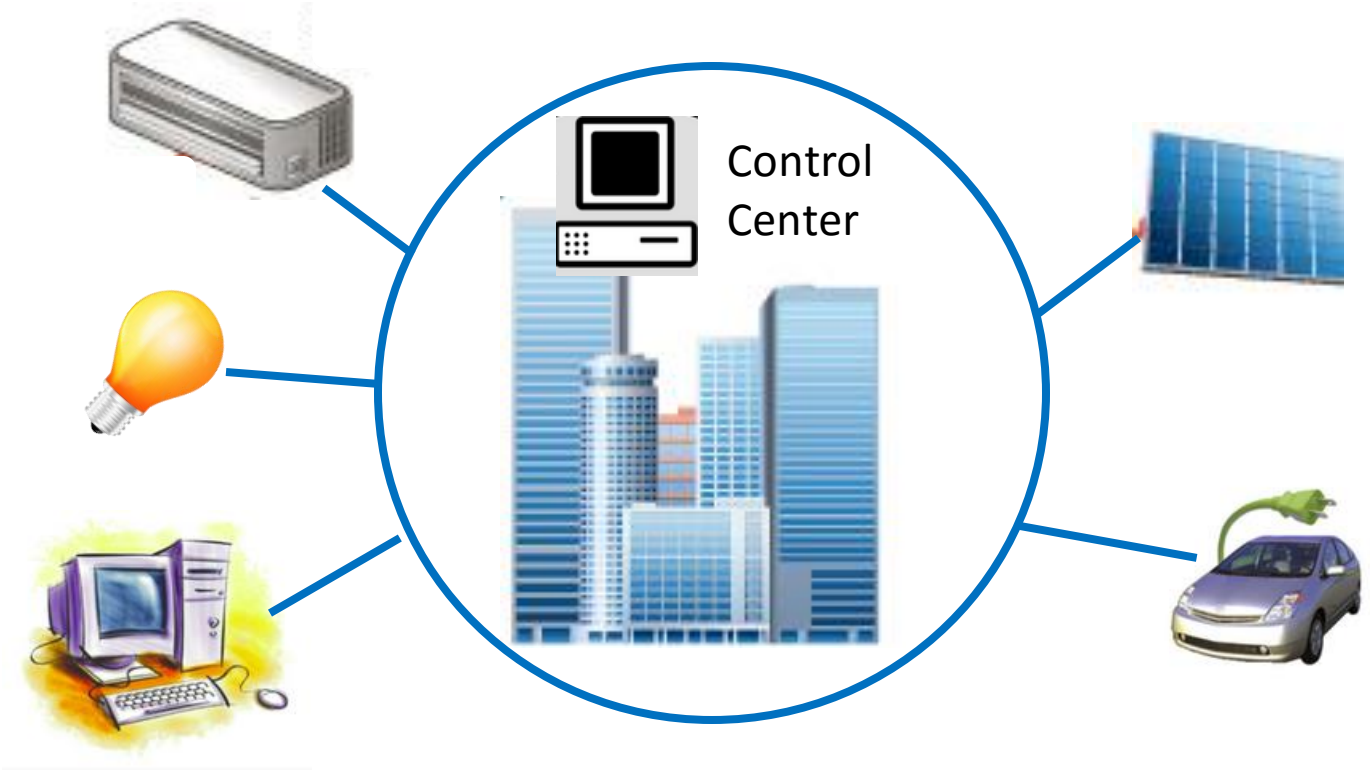
- ***Extract* information from *physical measurements (Feedback)***
  - Load frequency control
  - Decentralized voltage control (distribution network) (Qu, Li, Dahleh, 2014)
  - Power allocation in buildings/data center

- ***Recover* information from *local computation***
  - Quantized dual gradient for power allocation

Power management within buildings

Control center coordinates power consumption of appliances

- Maximize utility, minimize cost
- Satisfy operation constraints, e.g. power capacity constraints



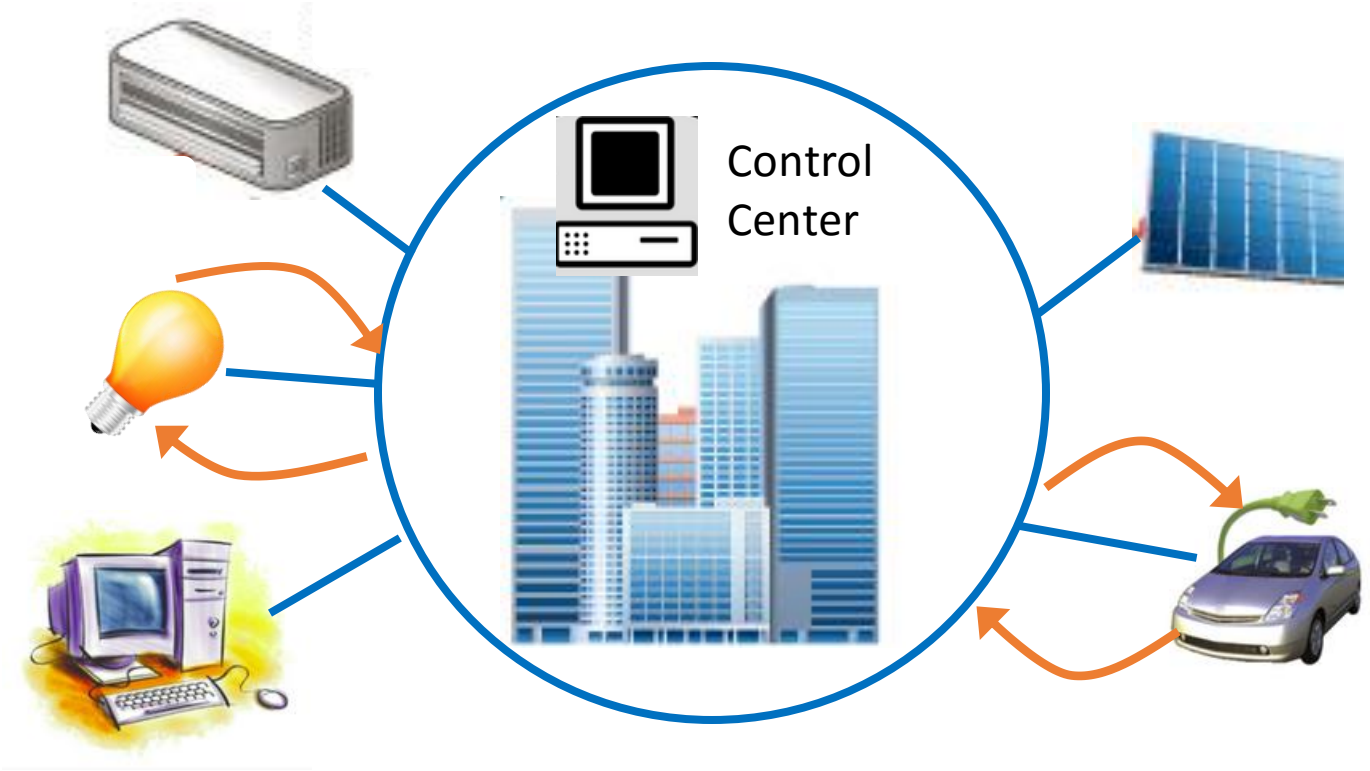
Distributed Coordination under Two-way Comm.

Iterate

Step 1: Appliances to center: Power request

Step 2: Center to appliances: Coordination signal

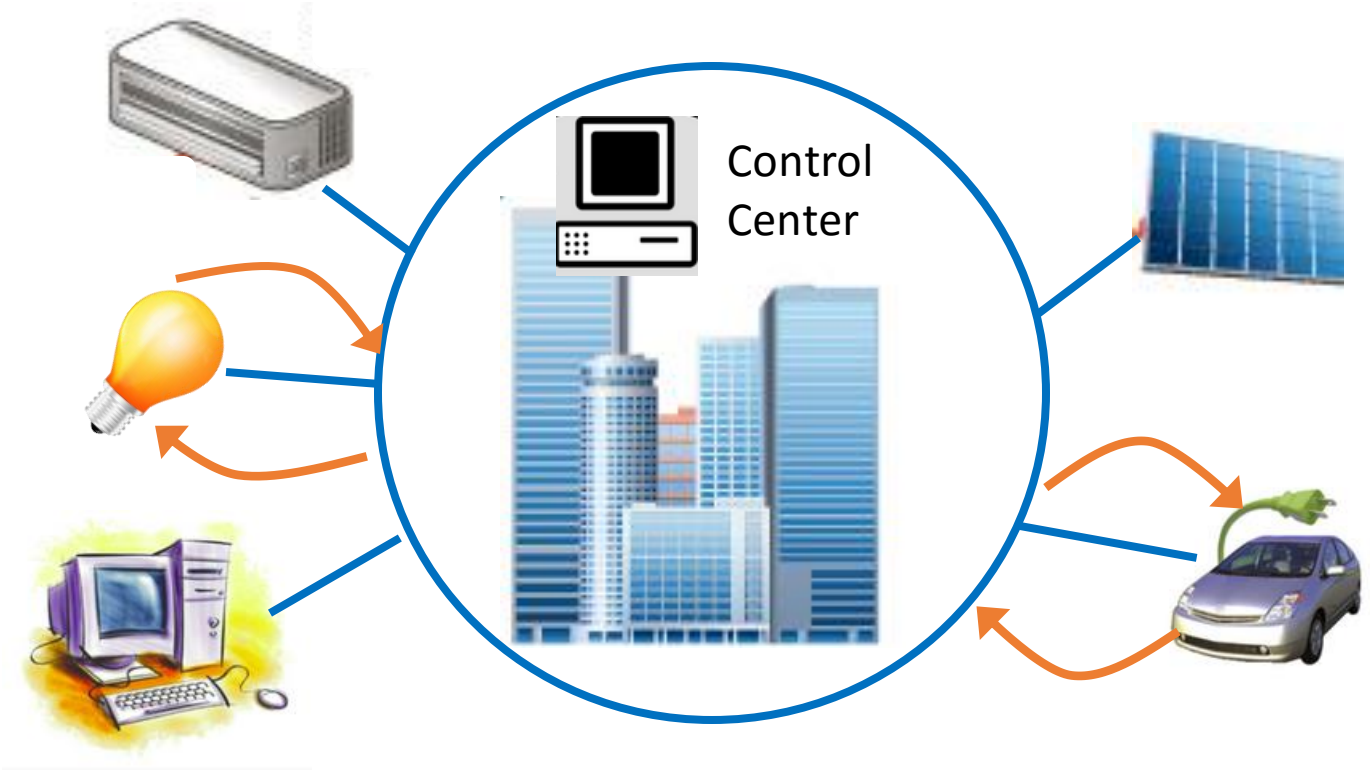
Assume perfect, reliable, and ubiquitous communication resources



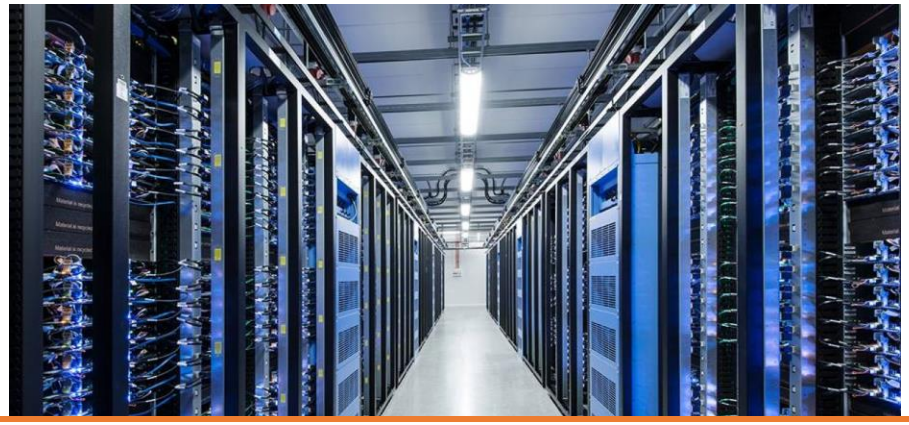
Reduce Communication Needs

Q 1: Is it possible to use only one-way comm.?

Q 2: How many bits are needed?



Not just for the buildings/grids

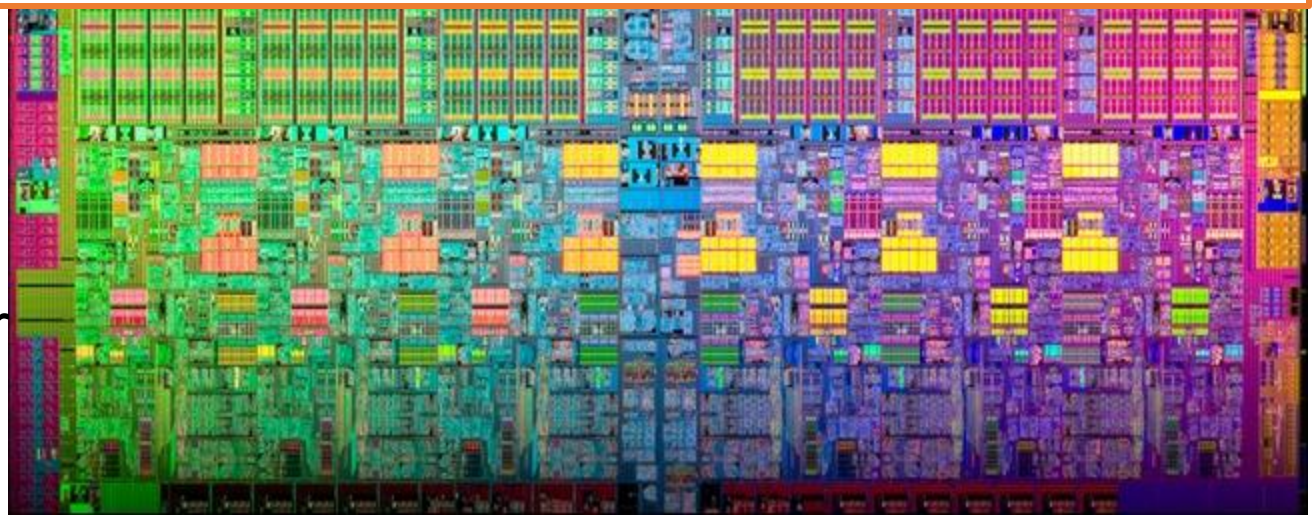


Data Center

Communication cost is much higher than computation

[Bolsens I., 2002]

Multi-core Processor

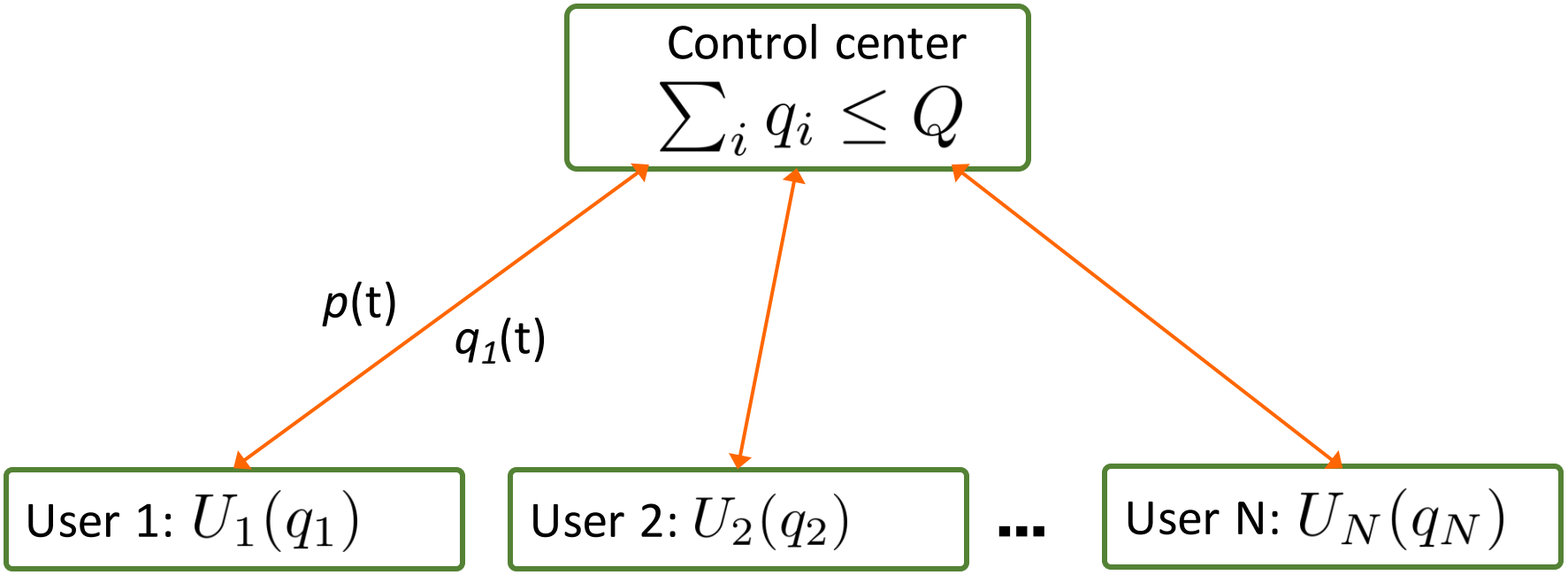


Power allocation problem

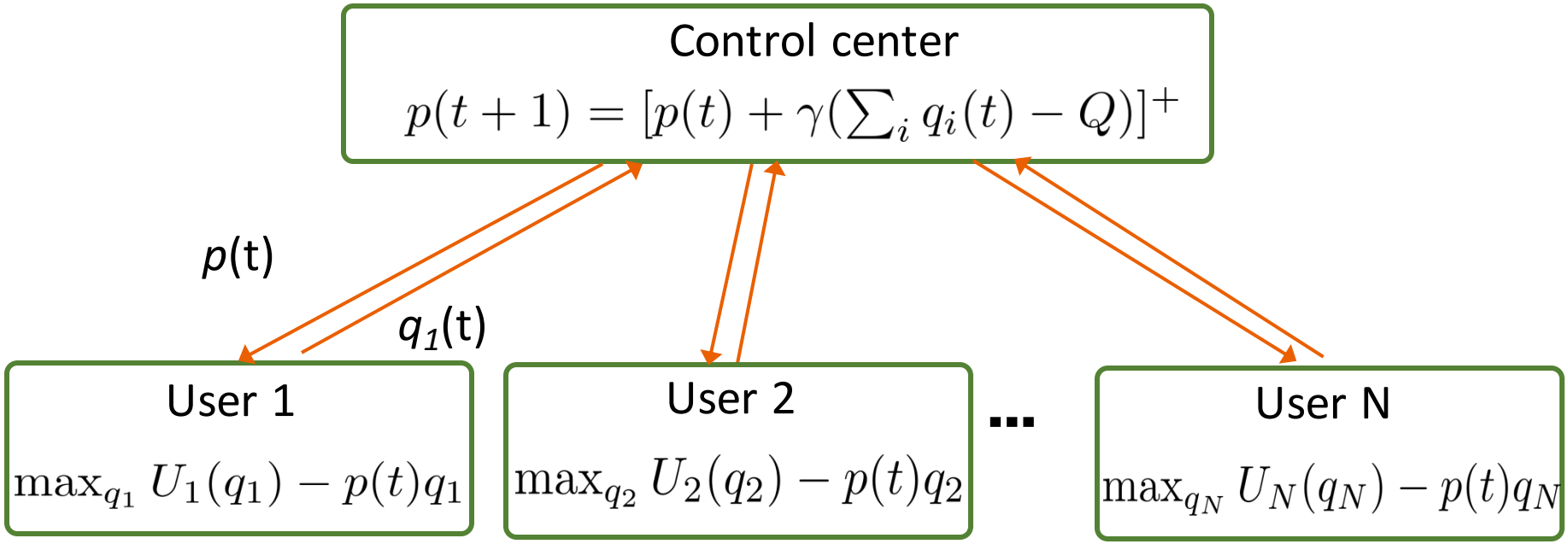
$$\max_{q_i \in [m_i, M_i]} \quad \sum_{i=1}^{N} U_i(q_i)$$

$$s.t. \quad \sum_i q_i \leq Q$$

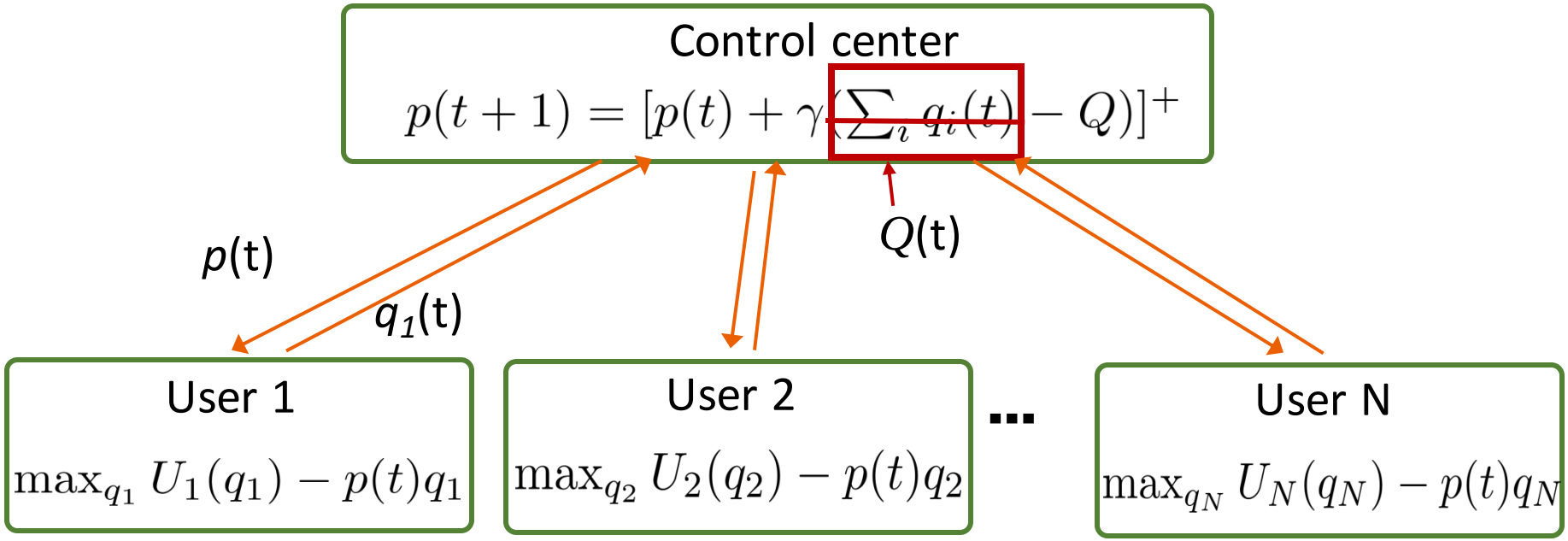
Appliances/Users: $i = 1, \ldots, N$
Power consumption of appliance i : q_i
total consumption: r
Utility function: U_i
Total power capacity: Q



A distributed algorithm: Dual gradient descent

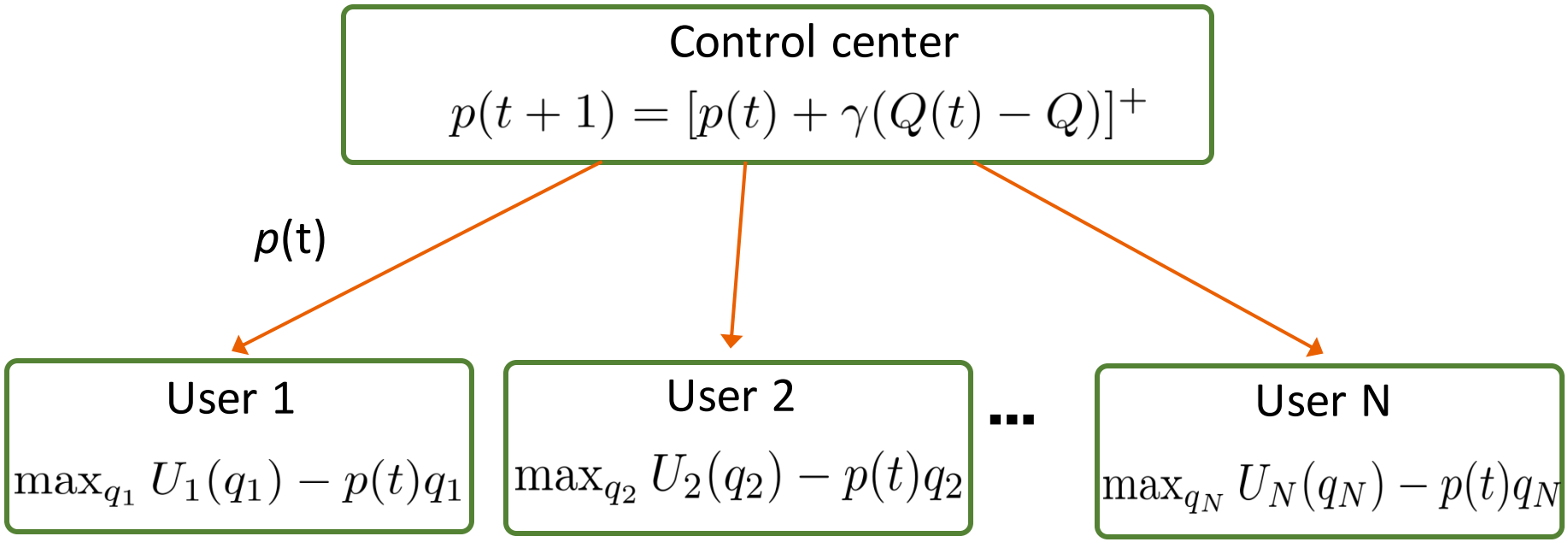


A distributed algorithm: One-way comm.



Replace this with *true measurement* of total power consump. Q(t).

What's the problem here?



It might violates hard physical constraint $\sum q_i \leq Q$

Theorem: If the *step size* and *initial setting* are chosen properly, the constraint will hold all the time.

"Distributed resource allocation using one-way communication", Magnusson, Enyioha, Li, Fischione, Tarokh, 2016

This Talk

- ***Extract* information from *physical measurements* (*Feedback*)**
  - Load frequency control
  - Power allocation in buildings/data center

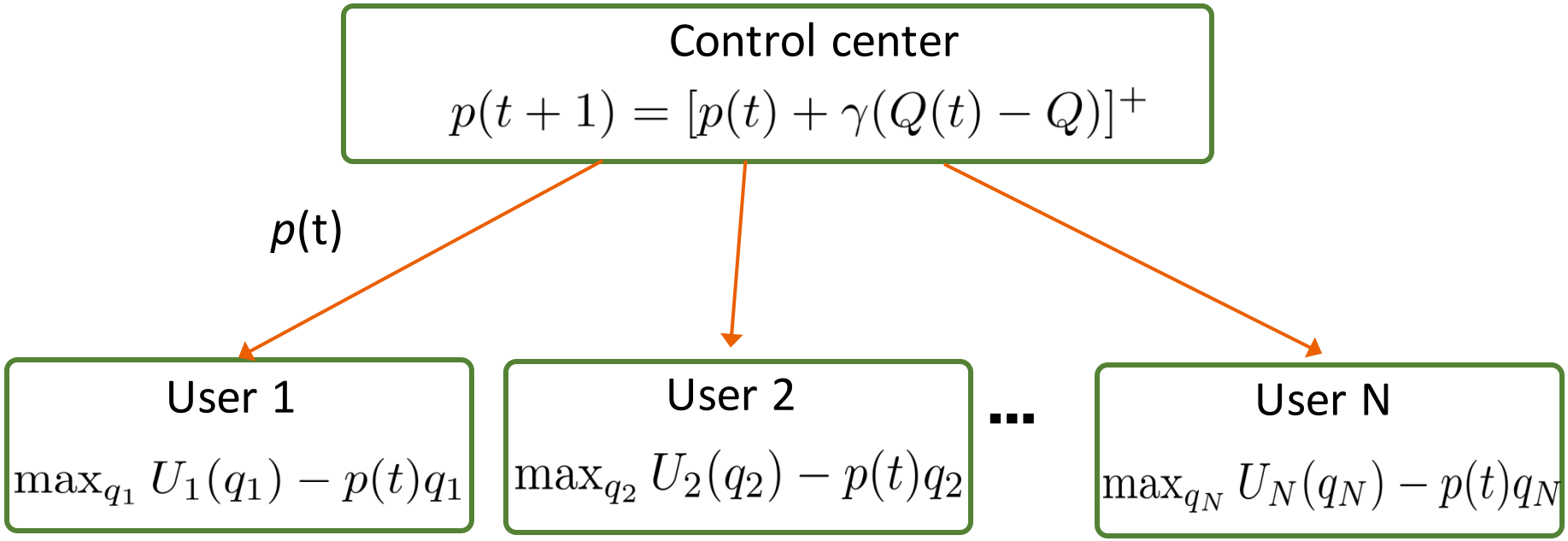
- ***Recover* information from *local computation***
  - Quantized dual gradient for power allocation

This Talk

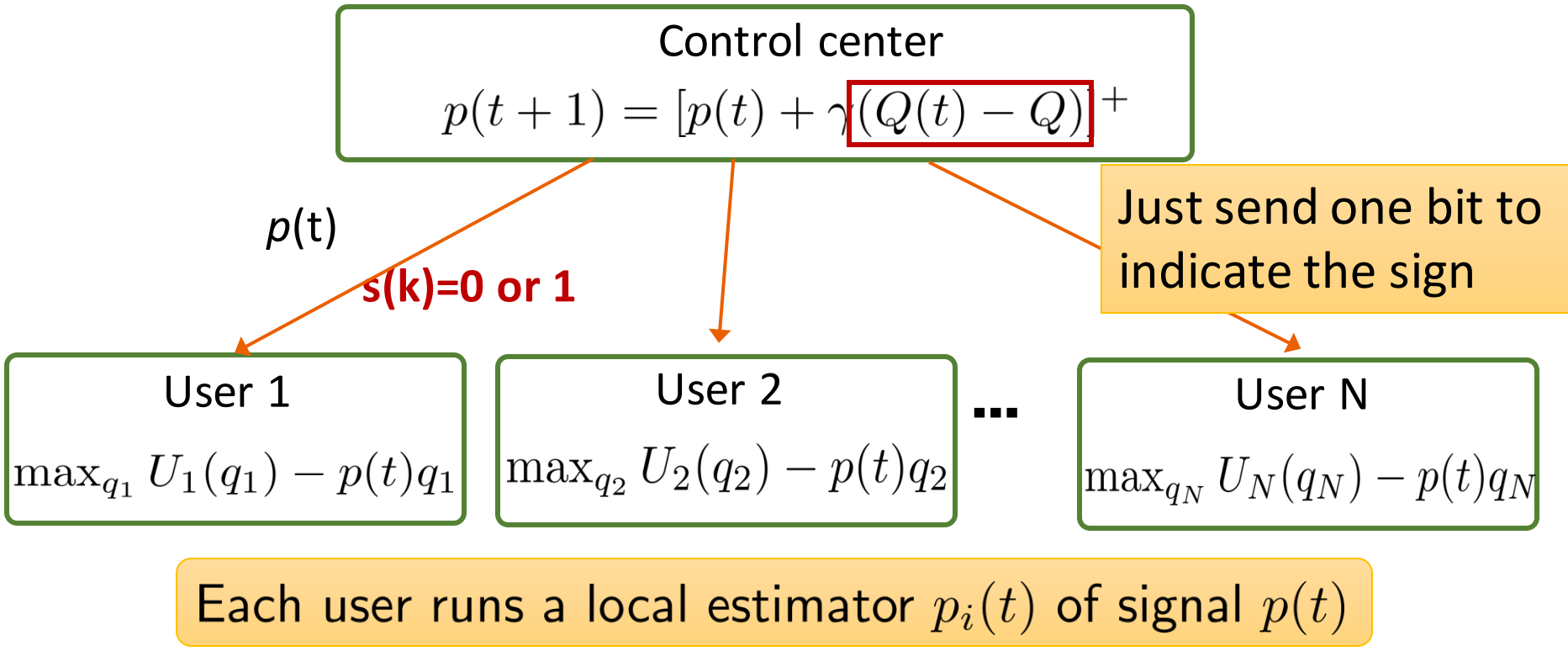
- ***Extract* information from *physical measurements (Feedback)***
  - Load frequency control
  - Power allocation in buildings/data center

- ***Recover* information from *local computation***
  - Quantized dual gradient for power allocation

Recall: Dual Gradient with One-way Comm.

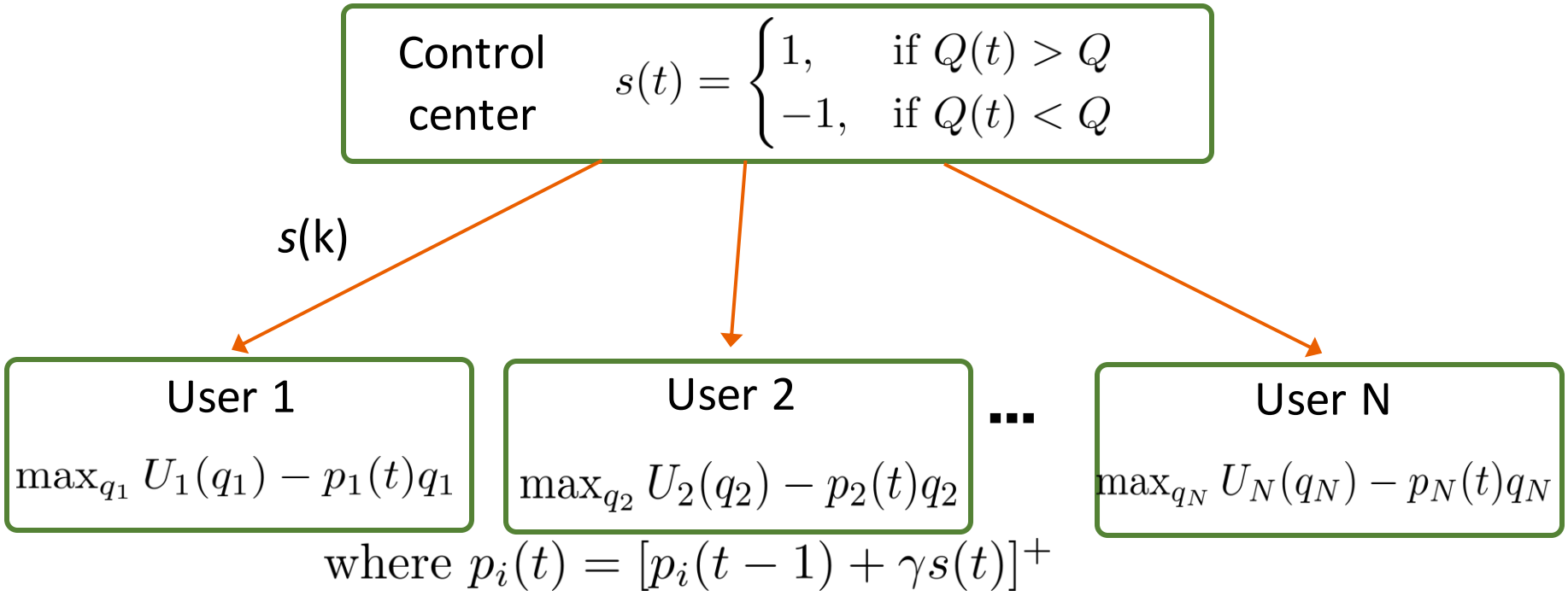


Further reduce comm.



Each user runs a local estimator $p_i(t)$ of signal $p(t)$

Dual Gradient with One-bit One-way Comm.



This is quantized (normalized) gradient descent of dual function

Normalized Gradient Descent [Shor 1985]:

$$\min_p f(p)$$

$$p(t+1) = p(t) - \gamma(t) \frac{f'(p(t))}{||f'(p(t))||}$$

Quantized (Normalized) Gradient Descent (QGD)

Problem: $\min_p f(p)$

QGD: $p(t+1) = p(t) - \gamma(t)d(t), d(t) \in \mathcal{D} \in \mathcal{S}^{K-1}$

Question: When $\mathcal{D}$ is a good quantization?

Definition: A quantization is proper (good) if and only if the algorithm is able to converge to the optimal points for **any well-behaviored** *f, e.g. convex smooth function.*

Quantized (Normalized) Gradient Descent (QGD)

Problem: $\min_p f(p)$

QGD: $p(t+1) = p(t) - \gamma(t)d(t), d(t) \in \mathcal{D} \in \mathcal{S}^{K-1}$

Questions:

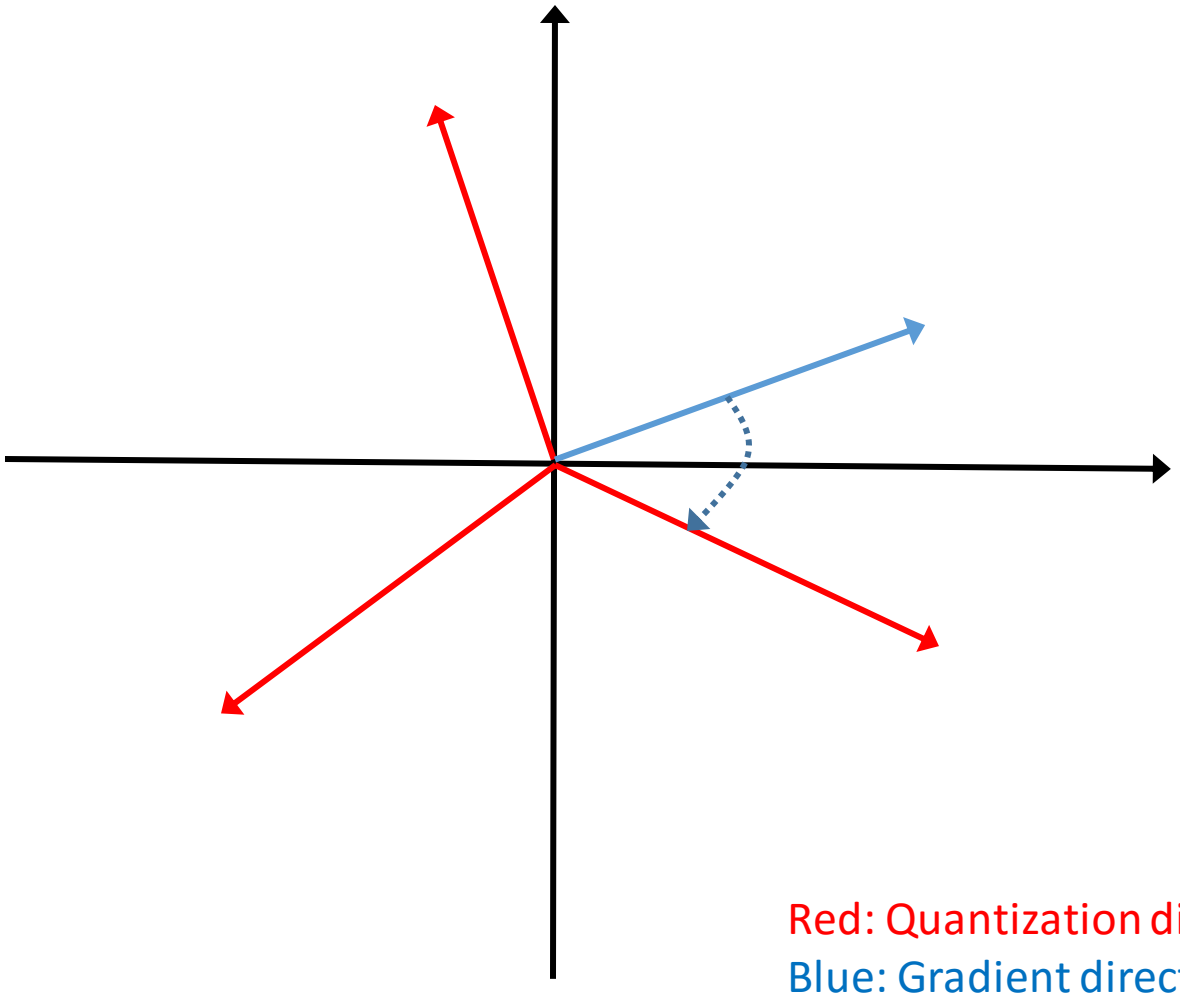
A) How to determine a quantization is proper?

B) What is the *minimal size* of the quantization to be proper?

C) How to choose d(t) and ε(t) , given a good quantization?

D) What are the *connections* between the *fineness* of the quantization to the *convergence* of the algorithm?

Descent direction



Red: Quantization direction
Blue: Gradient direction

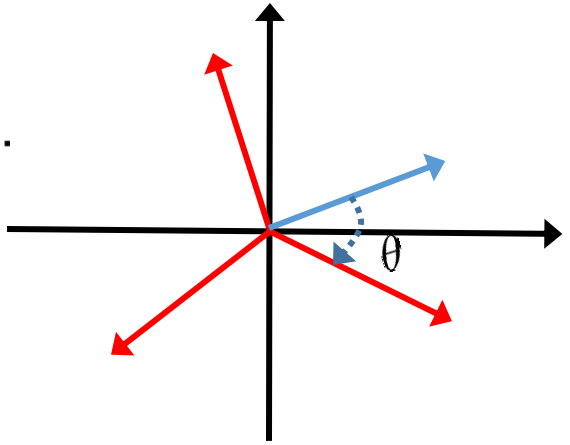
Proper quantization

Proposition. A set $\mathcal{D}$ is a proper quantization if and only if there exists $\theta \in [0, \pi/2)$, for every $g \in \mathcal{S}^{K-1}$ there exists $d \in \mathcal{D}$ such that

$$\text{ang}(g, d) \leq \theta.$$

Such quantization $\mathcal{D}$ is also called as $\theta-$cover.

- Minimal proper quantization, $|\mathcal{D}| = K + 1$.
- Only $\log_2(K+1)$ bits at each iteration.



Red: Quantization direction
Blue: Gradient direction

"Convergence of limited communications gradient methods", Magnusson, Enyioha, Li, Fischione, Tarokh, Transactions on Automatic Control, 2017

Convergence rate

One Stopping Criterion: $\mathcal{P}(\epsilon) := \left\{p \in \mathbb{R}^K : ||\nabla f(p)|| \leq \epsilon\right\}$

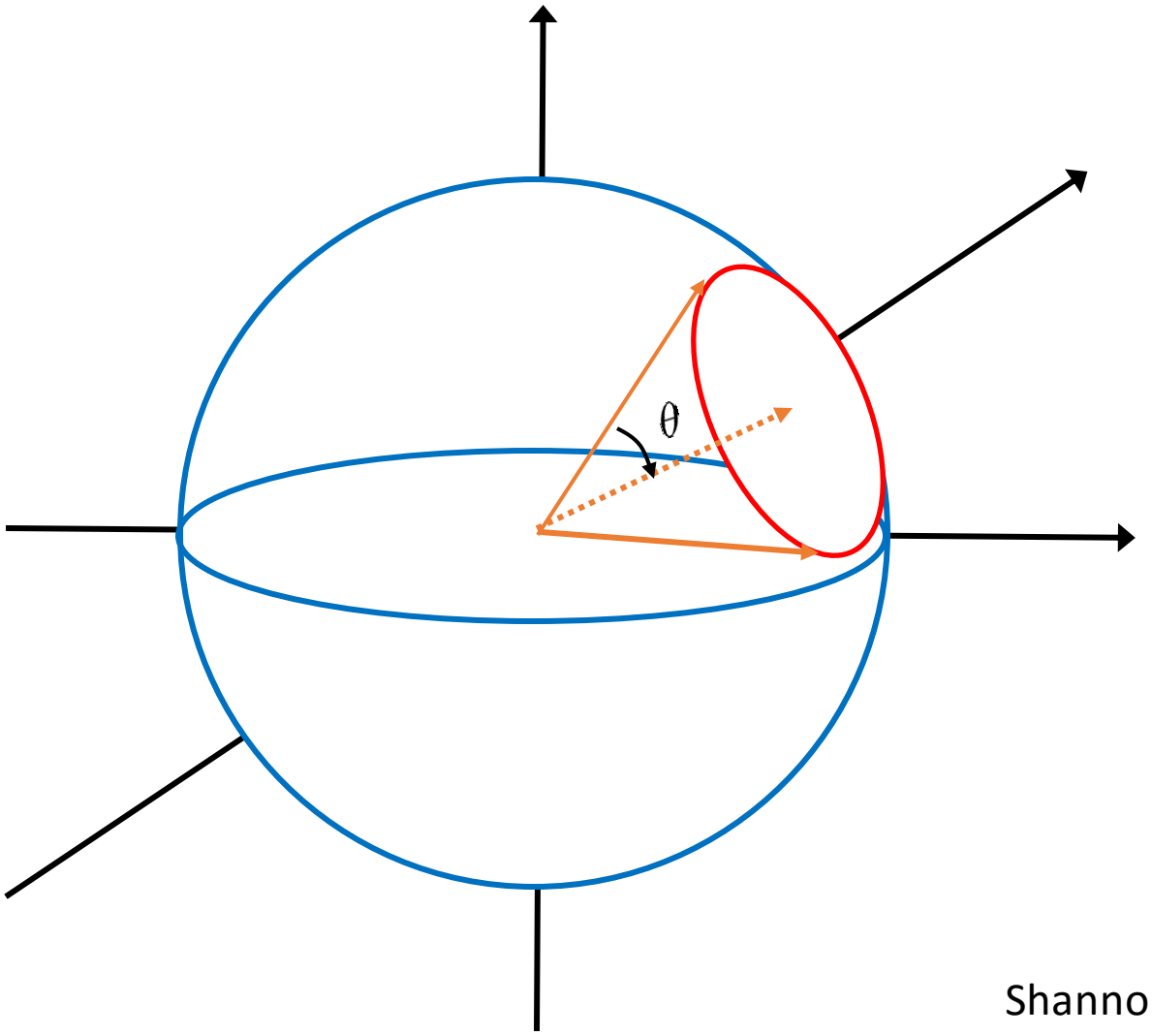
Theorem: For any $\epsilon > 0$, if $\gamma \in (0, 2\cos(\theta)\epsilon/L)$, then there exists $T \in \mathbb{N}$ such that $p(T) \in \mathcal{P}(\epsilon)$, with T bounded by

$$T \leq \frac{2(f(p(0)) - f^*)}{\gamma(2\cos(\theta)\epsilon - L\gamma)}.$$

- Finer quantization, larger stepsize is allowed
- Finer quantization, faster convergence

*More convergence results are available in the paper

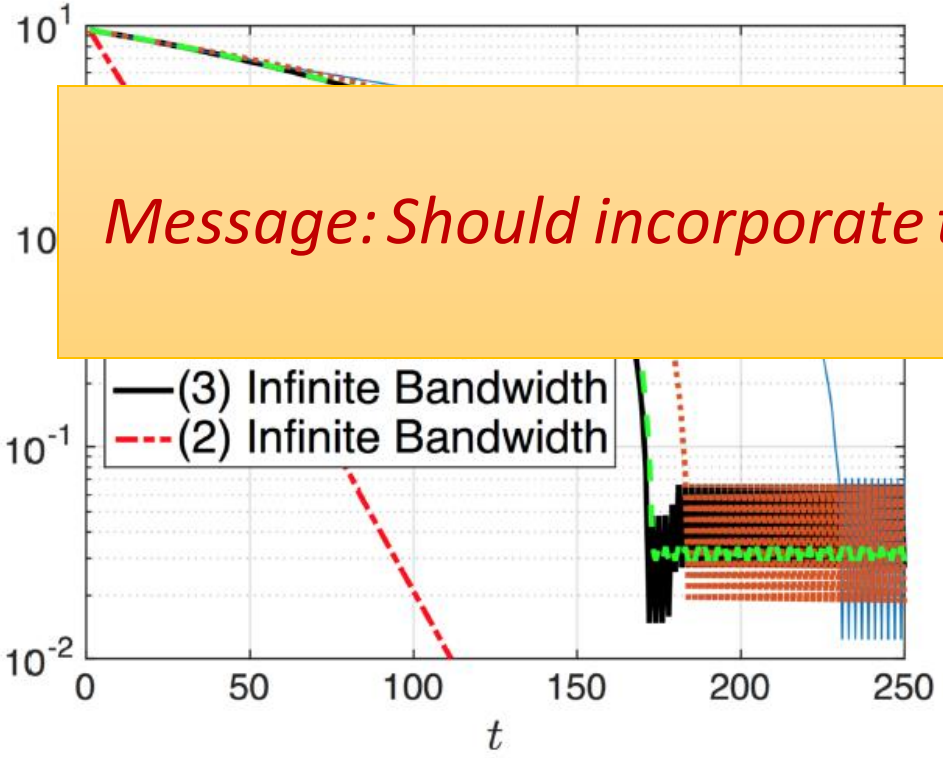
Quantization size



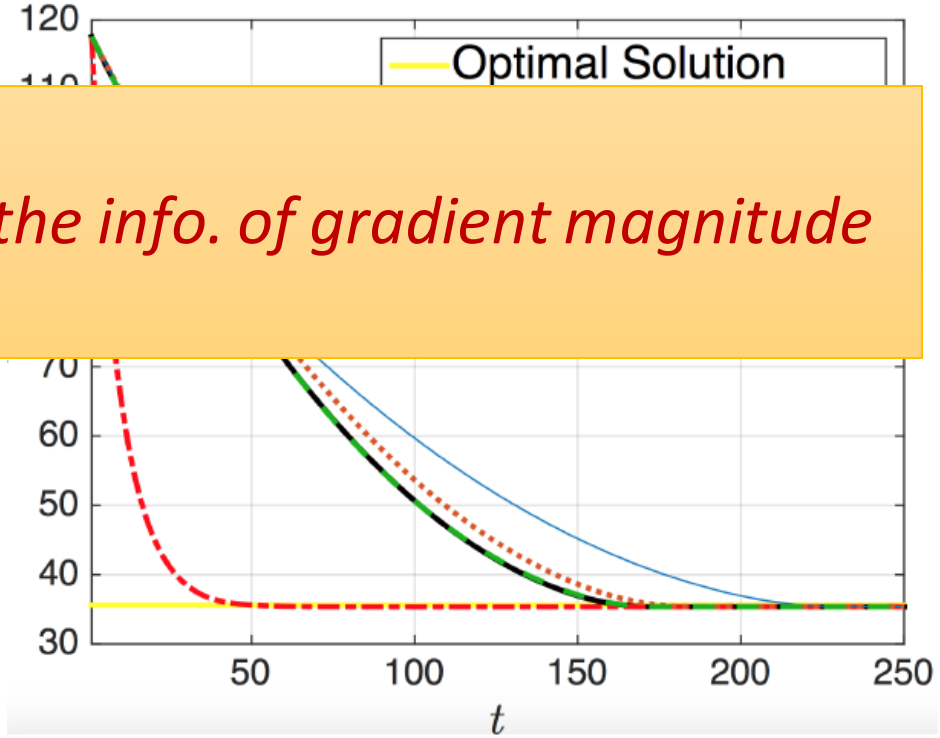
Shannon,1959

Simulation

f is the dual function of a two-dimension resource allocation problem.



$||\nabla f(p(t))||$



$f(p(t))$

Message: Should incorporate the info. of gradient magnitude

(3) Infinite bandwidth: normalized gradient

(2) Infinite bandwidth: gradient

Summary of QGD

Problem: $\min_p f(p)$

QGD: $p(t+1) = p(t) - \gamma(t)d(t), d(t) \in \mathcal{D} \in \mathcal{S}^K$

A) Proper quantization = θ-cover

B) Minimal size of proper quantization is K+1

C) Pick the quantized direction closest to gradient direction

D) θ plays an important role in the convergence

"Convergence of limited communications gradient methods", Magnusson, Enyioha, Li, Fischione, Tarokh, Transactions on Automatic Control, 2017

Extension to Constrained Case

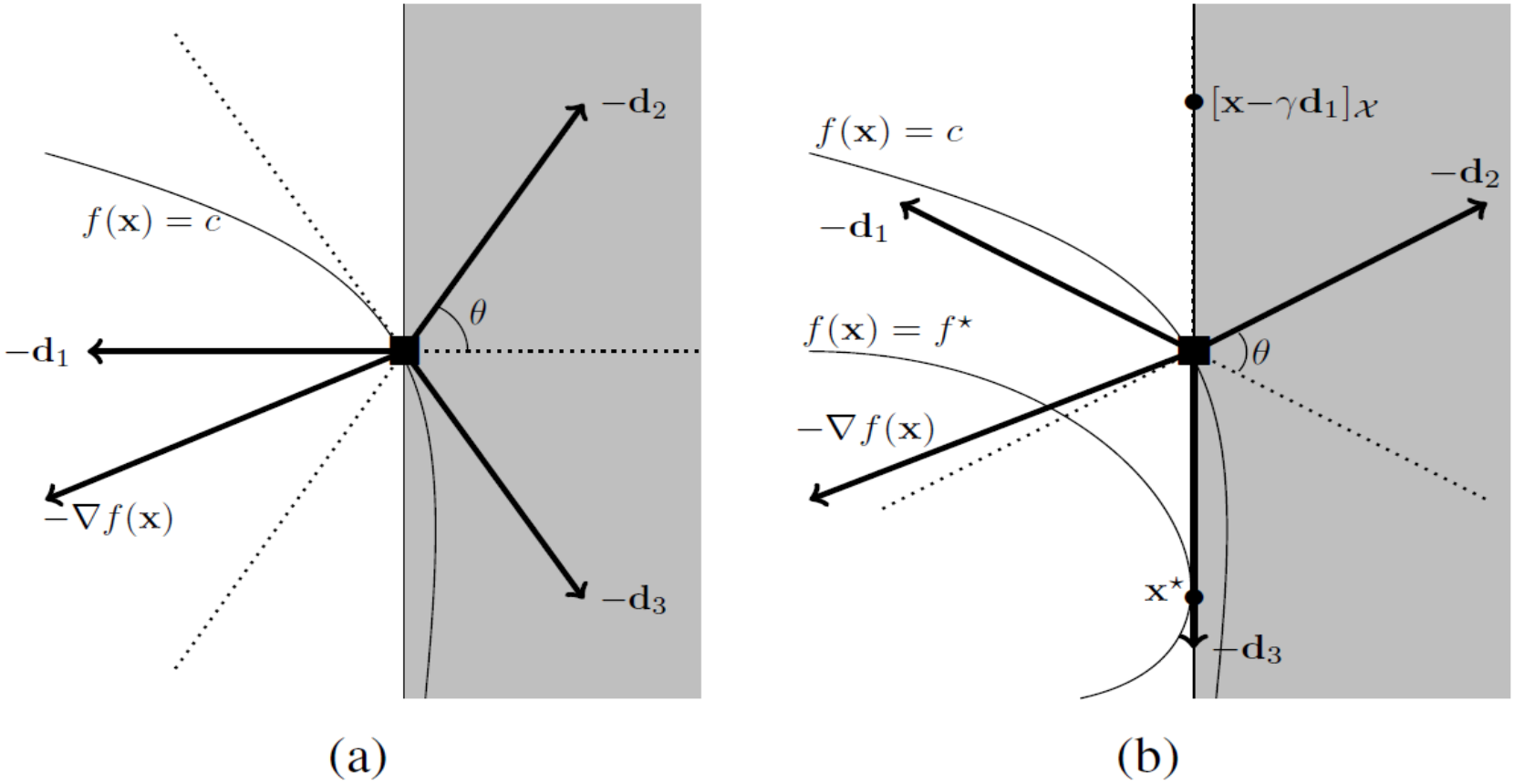
Problem: $\min_{p \in \mathcal{P}} f(p)$

QGD: $p(t+1) = [p(t) - \gamma(t) d(t)]_{\mathcal{P}}, d(t) \in \mathcal{D} \in \mathcal{S}^K$

Can the results of unconstrained case extend?

Θ-cover does not work for constrained case

Grey: Constraints set; ■ : x(t)



Get stuck at non-optimal points

Not necessarily a descent direction

Extension to Constrained Case

Problem: $\min_{p \in \mathbb{R}^n_+} f(p)$

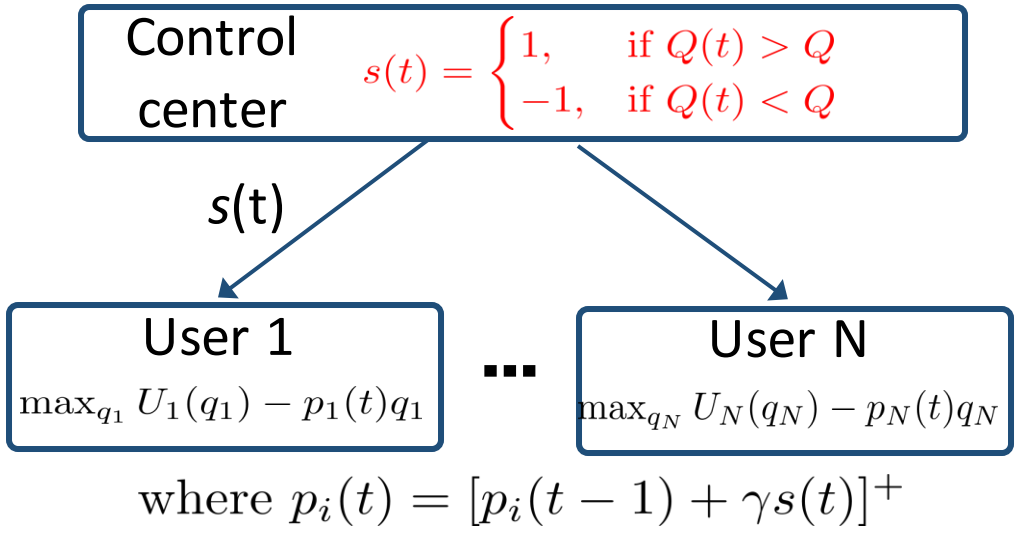
QGD: $p(t+1) = [p(t) - \gamma(t)\text{sign}(\nabla f(x))]_+$

- This is a θ-cover with $\cos(\theta) = \frac{1}{\sqrt{K}}$.
- 1 bit for one dimension (constraint). N bits in total.
- For this special quantization, similar convergence results hold.
- Applications: TCP flow control, Optimal network flow, Power management, Voltage control in distribution networks, etc

Communication Complexity

$$\max_{q_i \in [m_i, M_i]} \quad \sum_{i=1}^{N} U_i(q_i)$$

$$s.t. \quad \sum_i q_i \leq Q$$

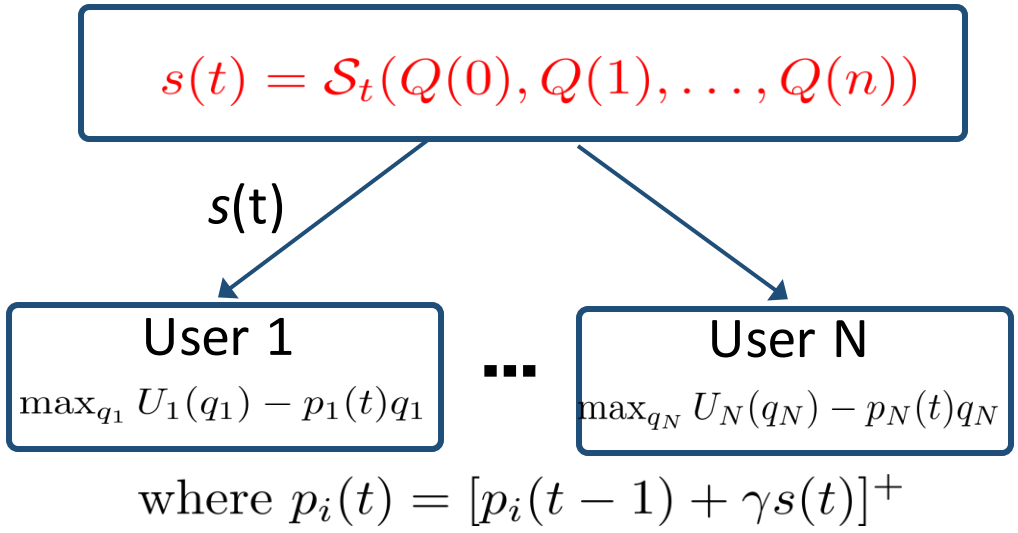


where $p_i(t) = [p_i(t-1) + \gamma s(t)]^+$

Communication Complexity

$$\max_{q_i \in [m_i, M_i]} \quad \sum_{i=1}^{N} U_i(q_i)$$

$$s.t. \quad \sum_{i} q_i \leq Q$$



where $p_i(t) = [p_i(t-1) + \gamma s(t)]^+$

Question:

What ***minimal*** bits (in total) are needed to achieve $\mathcal{E}$-optimal solution?

$\mathcal{E}$-complexity (a min max definition)

What ***optimal*** accuracy is able to be achieved using *b*-bits (in total)?

b-complexity (a min max definition)

Is there a simple coding scheme that reaches the complexity?

Yes.

"Communication Complexity of Distributed Resource Allocation Optimization", Magnusson, Enyioha, Li, Fischione, Tarokh, submitted, 2017

Summary: Limited Communication

- ***Extract* information from *physical measurements (Feedback)***
  - Load frequency control
  - Power allocation in buildings/data center
- ***Recover* information from *local computation***
  - Quantized dual (normalized) gradient for power allocation

Question:

How to choose the right algorithms and integrate them together?

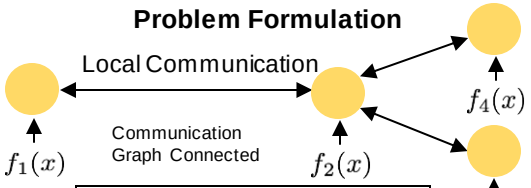
Tradeoff: Efficiency, Robustness, Communication, Sensing, Computation, Convergence speed

Thank you!

Accelerated Distributed Nesterov Gradient Descent

Guannan Qu, Na Li, John A. Paulson School of Engineering and Applied Sciences, Harvard University

Problem Formulation



$$\min f(x) \triangleq \frac{1}{n}\sum_{i=1}^n f_i(x)$$

Agent i keep updating $x_i(t)$ using local info

s.t. $\forall i,\ x_i(t) \to x^* \triangleq \arg\min_x f(x)$

Background

Centralized Gradient Methods for minimizing f

Gradient Descent

(for f convex and L-smooth, or f μ-strongly convex and L-smooth)

$$x(t+1) = x(t) - \frac{1}{L}\nabla f(x(t))$$

Nesterov Gradient Descent (for f convex, L-smooth)

$$x(t+1) = y(t) - \frac{1}{L}\nabla f(y(t))$$
$$v(t+1) = v(t) - \frac{1}{L\alpha_t}\nabla f(y(t))$$
$$y(t+1) = (1-\alpha_{t+1})x(t+1) + \alpha_{t+1}v(t+1)$$

where $\alpha_0 \in (0,1)$, $\alpha_{t+1}^2 = (1-\alpha_{t+1})\alpha_t^2$

Nesterov Gradient Descent (for f μ-strongly convex, L-smooth)

$$x(t+1) = y(t) - \frac{1}{L}\nabla f(y(t))$$
$$v(t+1) = (1-\alpha)v(t) + \alpha y(t) - \frac{1}{L\alpha}\nabla f(y(t))$$
$$y(t+1) = \frac{x(t+1) + \alpha v(t+1)}{1+\alpha} \qquad \text{where } \alpha = \sqrt{\frac{\mu}{L}}$$

Convergence Rates

f type / algo.	GD		Nesterov GD
Convex and L-Smooth	$O(\frac{1}{t})$	faster ➡	$O(\frac{1}{t^2})$
μ-Strongly Convex and L-Smooth	$O((1-\frac{\mu}{L})^t)$	better dependence on condition number $\frac{L}{\mu}$ ➡	$O((1-\sqrt{\frac{\mu}{L}})^t)$

Nesterov GD brings acceleration!

Most distributed gradient methods are based on GD, and the convergence rate is not better than GD.

Can Nesterov momentum be used in *Distributed Gradient Methods* and accelerate the convergence?

Main Results

μ-strongly convex and L-smooth cost functions

Proposed Algorithm

Initialize: $x_i(0) = y_i(0) = v_i(0)$ arbitrary $\quad s_i(0) = \nabla f_i(y_i(0))$

Step size $\eta > 0$, and set $\alpha = \sqrt{\mu\eta}$

$$x_i(t+1) = \sum_j w_{ij}y_j(t) - \eta s_i(t)$$
$$v_i(t+1) = (1-\alpha)\sum_j w_{ij}v_j(t) + \alpha\sum_j w_{ij}y_j(t) - \frac{\eta}{\alpha}s_i(t)$$
$$y_i(t+1) = \frac{x_i(t+1) + \alpha v_i(t+1)}{1+\alpha}$$
$$s_i(t+1) = \sum_j w_{ij}s_j(t) + \nabla f_i(y_i(t+1)) - \nabla f_i(y_i(t))$$

Summary of Results

Theorem. Assume f_i is L-smooth and μ-strongly convex. When

$$0 < \eta < \frac{\sigma^3(1-\sigma)^3}{61909L}(\frac{\mu}{L})^{3/7}$$

where σ is the second largest singular value of the W, we have

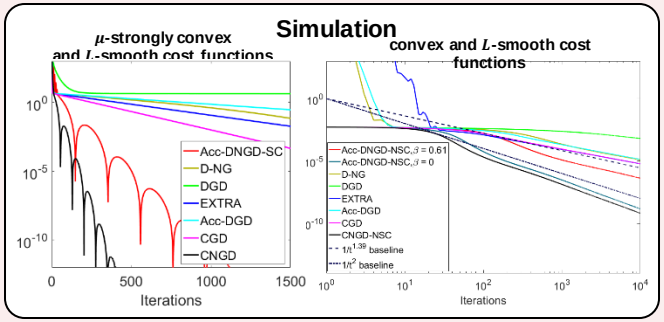
- $f(\bar{x}(t)) - f^* = O((1-\sqrt{\mu\eta})^t)$
- $\|y(t) - \mathbf{1}x^*\| = O((1-\sqrt{\mu\eta})^{t/2})$

Convergence Rate: $\left[1 - \frac{1}{\sqrt{61909}}(\frac{\mu}{L})^{5/7}\sigma^{3/2}(1-\sigma)^{3/2}\right]^t$

Dependence on the condition number $(\frac{\mu}{L})^{5/7}$ strictly better than that of GD $(\frac{\mu}{L})$!

Nesterov does bring acceleration!

Simulation



For more detailed results, see
Guannan Qu and Na Li, "Accelerated Distributed Nesterov Gradient Descent," *arXiv preprint arXiv:1705.07176*(2017).

Main Results

convex and L-smooth cost functions

Proposed Algorithm

Initialize: $x_i(0) = y_i(0) = v_i(0) = 0 \quad s_i(0) = \nabla f(0)$

Step sizes $\eta_t \in (0, \frac{1}{L})$; Define α_t by

$$\alpha_0 = \sqrt{\eta_0 L} \qquad \alpha_{t+1}^2 = \frac{\eta_{t+1}}{\eta_t}(1-\alpha_{t+1})\alpha_t^2$$

$$x_i(t+1) = \sum_j w_{ij}y_j(t) - \eta_t s_i(t)$$
$$v_i(t+1) = \sum_j w_{ij}v_j(t) - \eta_t\frac{1}{\alpha_t}s_i(t)$$
$$y_i(t+1) = (1-\alpha_{t+1})x_i(t+1) + \alpha_{t+1}v_i(t+1)$$
$$s_i(t+1) = \sum_j w_{ij}s_j(t) + \nabla f_i(y_i(t+1)) - \nabla f_i(y_i(t))$$

Summary of Results

(Throughout this section, f_i is convex and L-smooth.)

Theorem. When using vanishing step size $\eta_t = \frac{\eta}{(t+t_0)^{0.6+\epsilon}}$ with small η and large t_0, and if the set of minimizers are compact, then $f(\bar{x}(t)) - f^* = O(\frac{1}{t^{1.4-\epsilon}})$.

Remark. In simulation, $\eta = \frac{1}{2L}$ and $t_0 = 1$ suffice.

Theorem. When using small enough fixed step size $\eta_t = \eta$, and assuming $f_i(x) = h_i(A_i x)$, for some matrix A_i and some h_i that is strongly convex and smooth, we have $f(\bar{x}(t)) - f^* = O(\frac{1}{t^2})$.

Conjecture. When using small enough fixed step size $\eta_t = \eta$, we have $f(\bar{x}(t)) - f^* = O(\frac{1}{t^2})$.

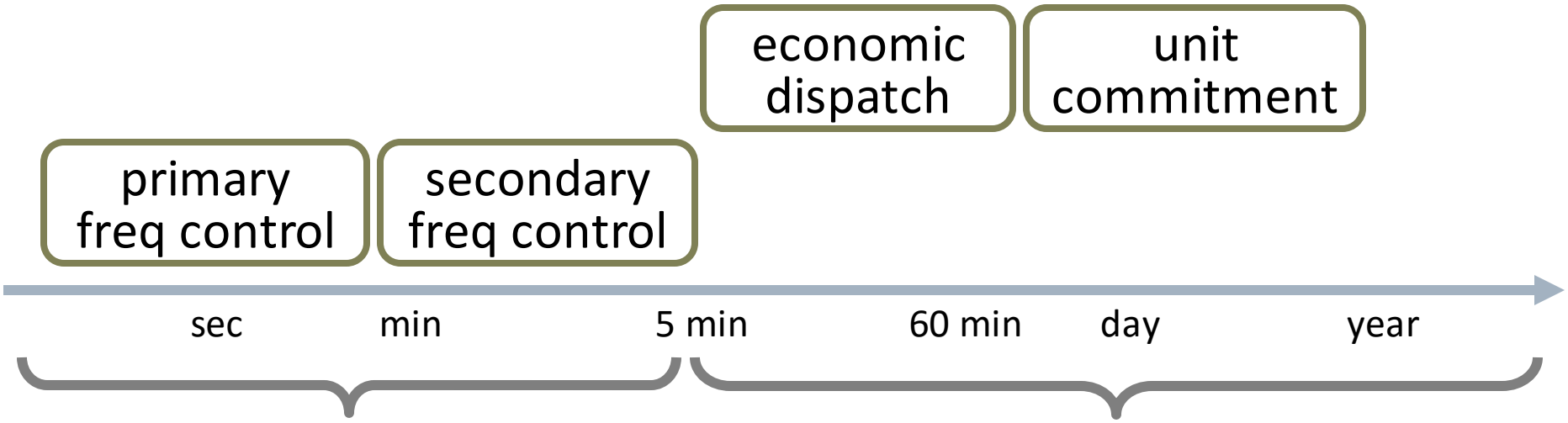
Nesterov does bring acceleration, from $O(\frac{1}{t})$ to $O(\frac{1}{t^{1.4-\epsilon}})$, or even $O(\frac{1}{t^2})$!

Preliminaries

- f_i is L-smooth: $\langle \nabla f_i(x) - \nabla f_i(y), x-y\rangle \le L\|x-y\|^2$
- f_i is μ-strongly convex: $\langle \nabla f_i(x) - \nabla f_i(y), x-y\rangle \ge \mu\|x-y\|^2$
- $W = [w_{ij}]$ is a consensus averaging matrix:
  - Graph is connected
  - $w_{ii} > 0$, $w_{ij} > 0$ if i, j adjacent.
  - W is doubly stochastic
  - $w_{ij} = 0$ if i, j NOT adjacent.

Back up: Reverse and Forward Engineering

Frequency control

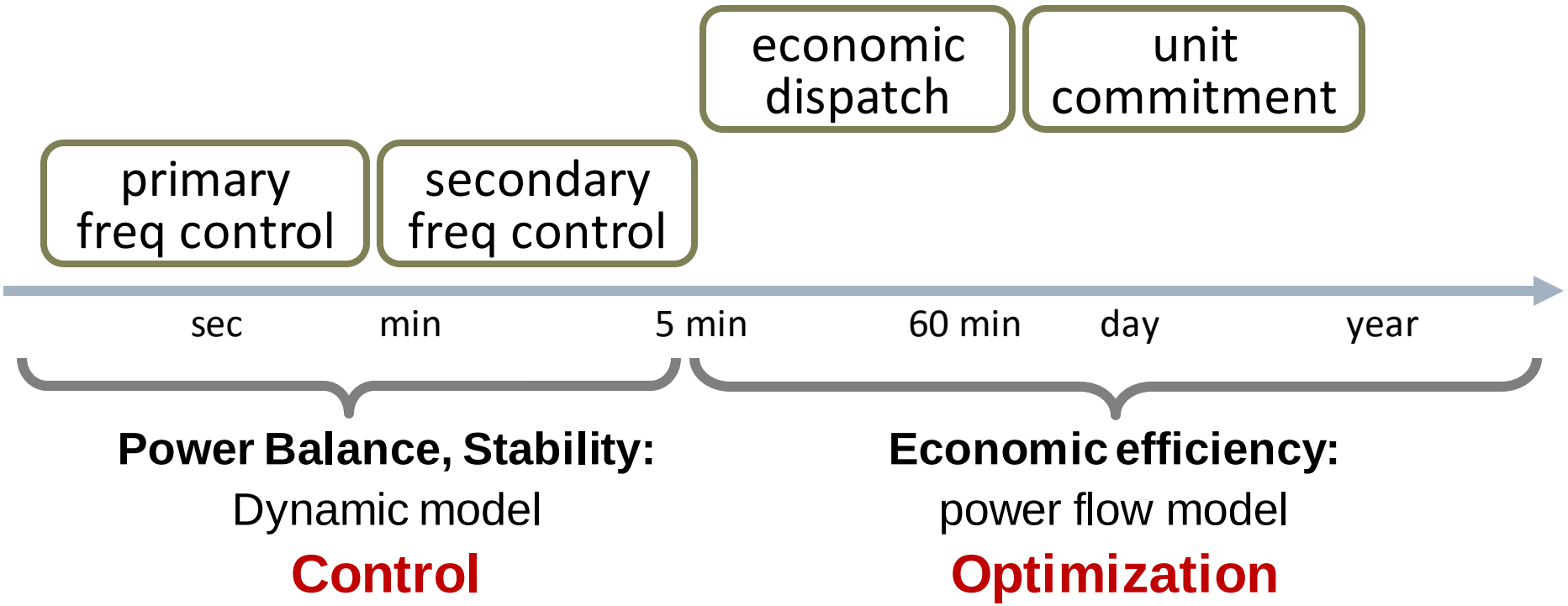


Power Balance, Stability:
Dynamic model

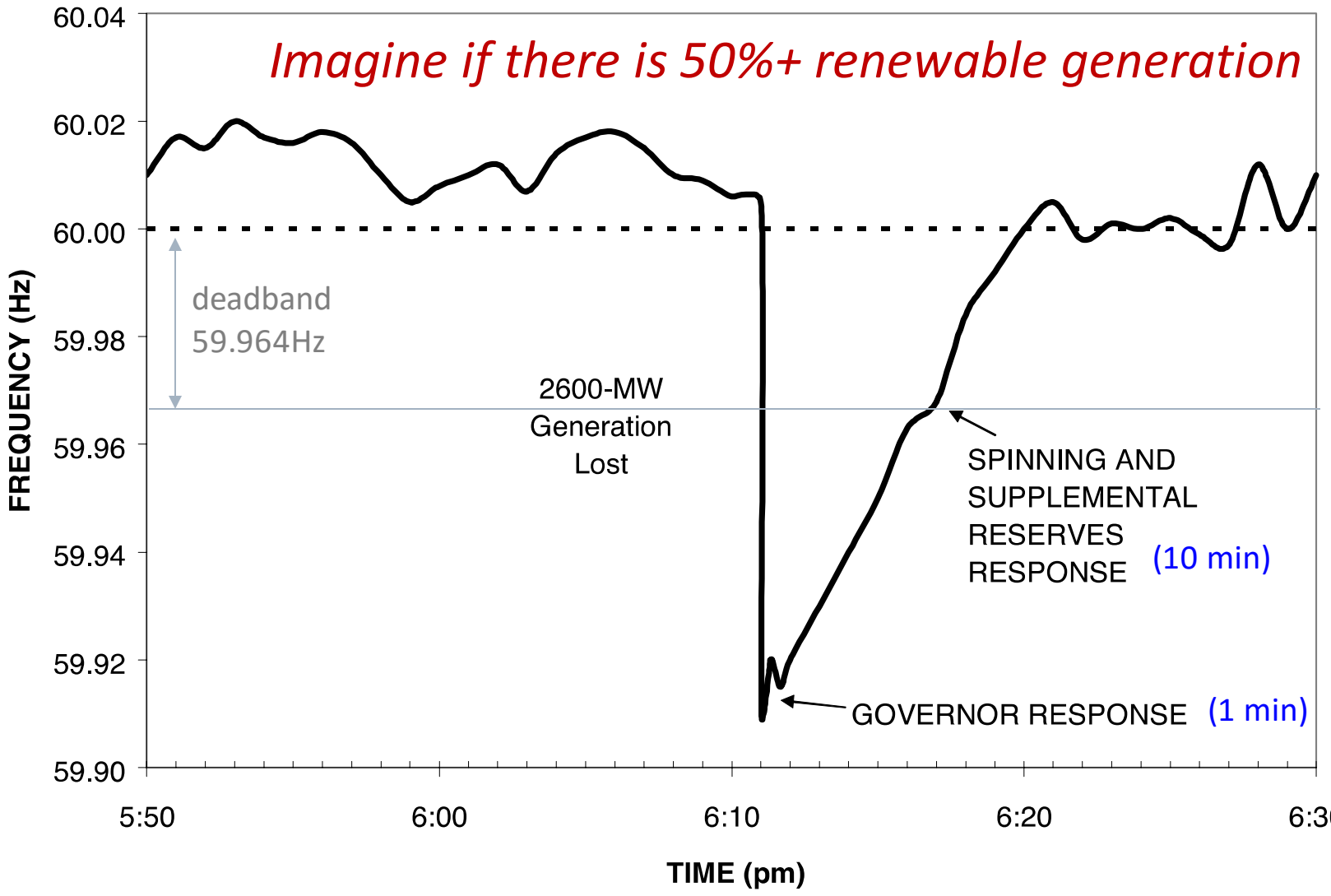
Economic efficiency:
power flow model

Frequency control

- Goal: Balance the grid in an optimal (cost-effective) way
- Traditionally this is done at the generation side.



Frequency response



Loss of 2 nuclear plants in ERCOT

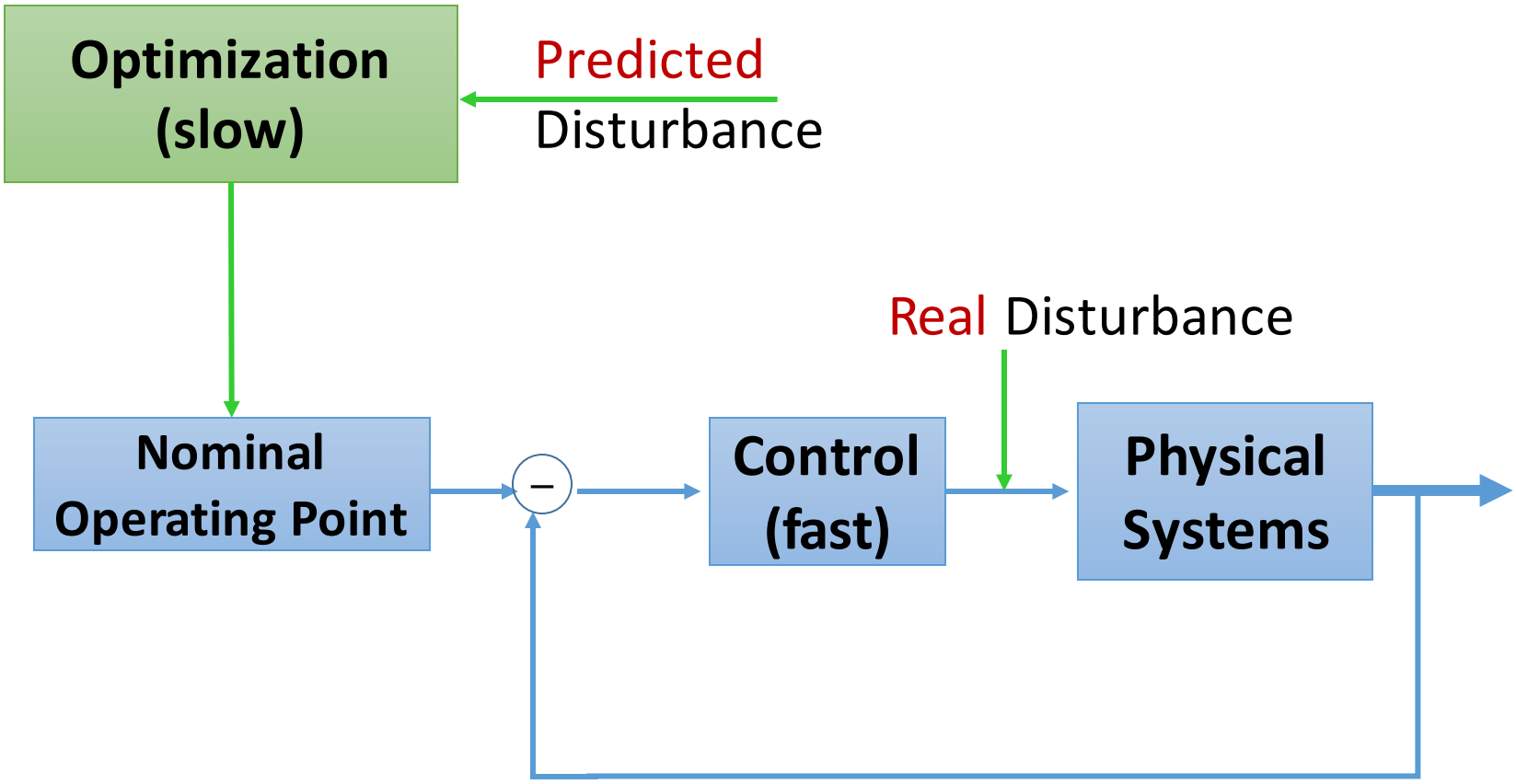
Kirby 2003 [ORNL/TM-2003/19]

Advantages of load-side control

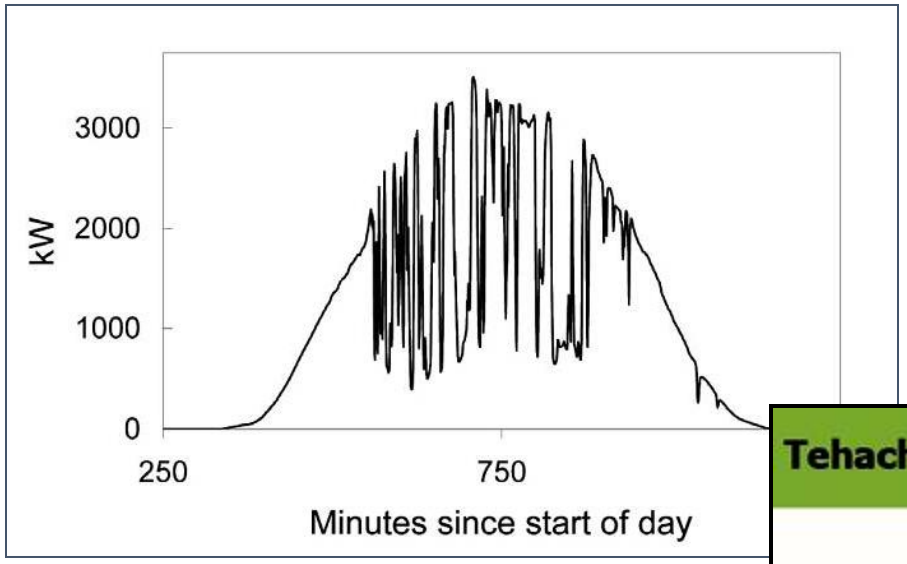
- faster (no/low inertia!)
- no waste or emission
- more resources (large #)
- localize disturbance

Idea dates back to 1970s (Schweppe et al (1979, 1980))

Hierarchical Control at Different Time-scales

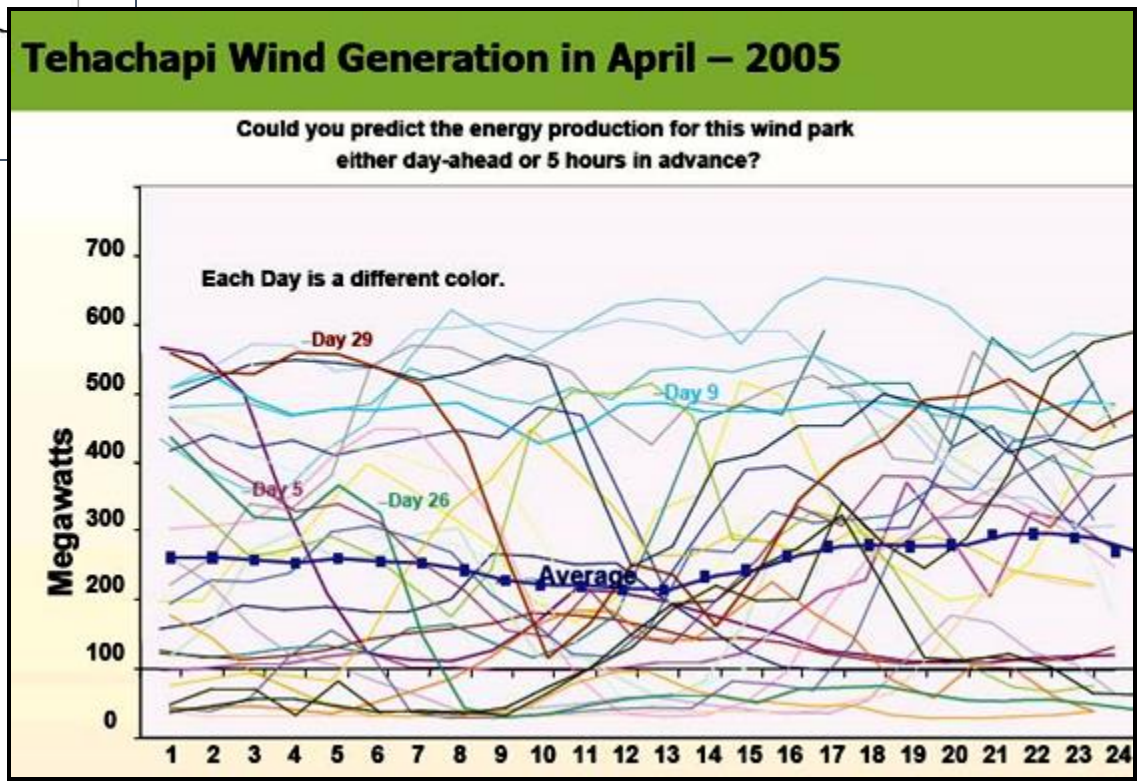


Challenges

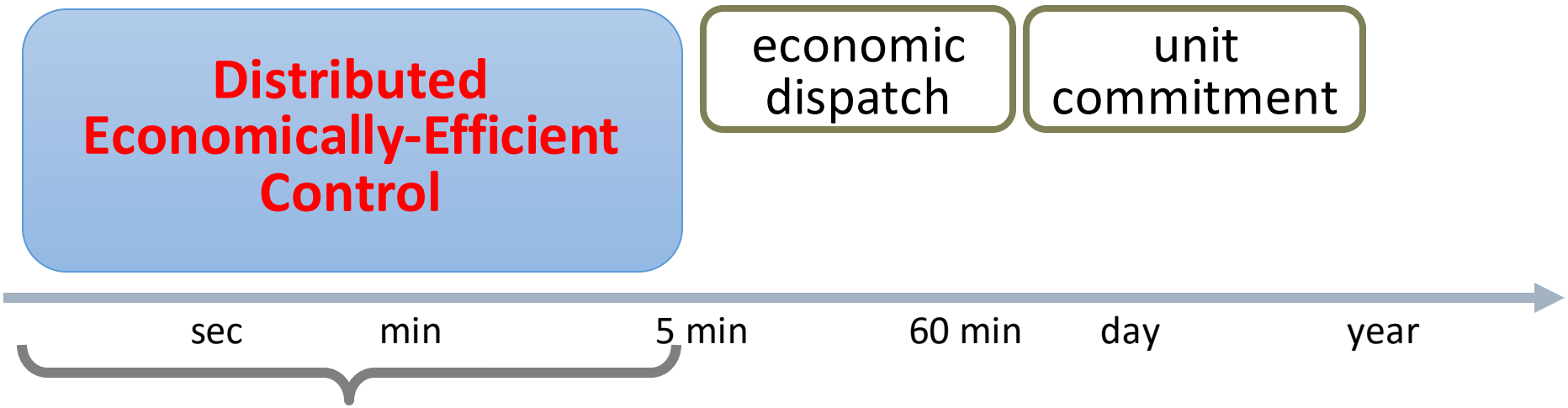


Imagine when we have 33%+ renewable generation …

Can the grid follow its own PV/Wind production faster and more efficiently?



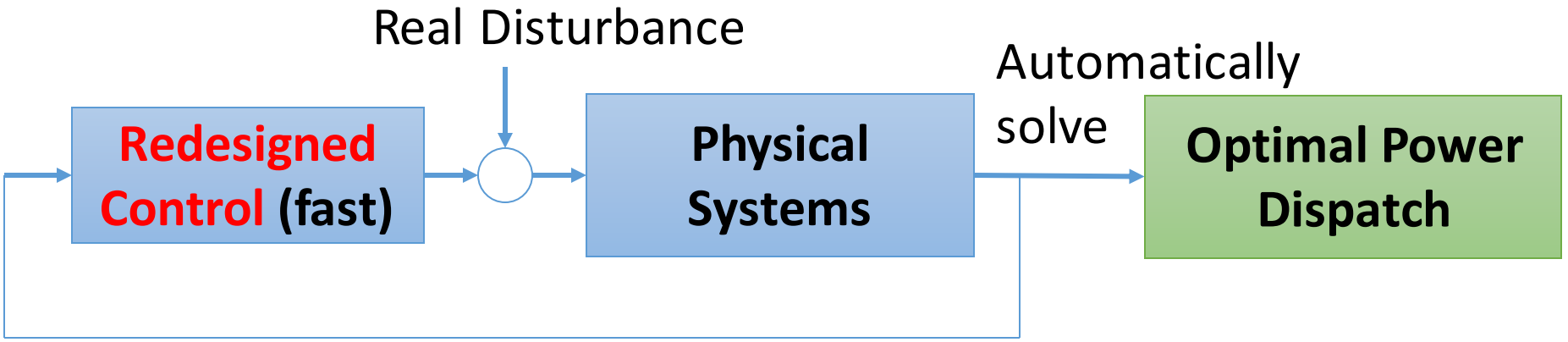
Distributed Economically-Efficient Control



Control Goals:

- Rebalance power
- Stabilize frequency
- Restore nominal frequency
- **Re-dispatch power optimally (min cost/disutility)**

Distributed Economically-Efficient Control



Advantages:

For the ***control***: Stable and more economically-efficient

For the ***optimization***: Save sensing/communication/computation

Problem setup

System Dynamics & Existing Control

$$\dot{x}_i = \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i,\ i = 1, \cdots, N$$

$$\dot{u}_i = \sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i,\ i = 1, \cdots, N$$

Example:

Frequency dynamics, Voltage dynamics

Primary/Secondary frequency/Voltage control

Inverter dynamics/control

(Model limitation: linear approximation)

Problem setup

Optimization Problem

$$\min_{x\in\mathbb{R}^n, u\in\mathbb{R}^m} \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N} g_i(u_i)$$

$$\text{s.t.} \sum_{j\in N(i)} A_{ij}x_j + B_i u_i + C_i w_i = 0$$

$$\sum_{j\in N(i)} D_{ij}x_j + \sum_{j\in N(i)} E_{ij}u_j + F_i w_i = 0$$

$$h_i(x_i, u_i) \le 0,\ i = 1, \cdots, N$$

Example:

Economic Dispatch,

Optimal Load Response.

Problem setup

System Dynamics & Existing Control

$$\dot{x}_i = \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i,\ i = 1, \cdots, N$$

$$\dot{u}_i = \sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i,\ i = 1, \cdots, N$$

Economical Efficient State

$$\min_{x \in \mathbb{R}^n, u \in \mathbb{R}^m} \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N} g_i(u_i)$$

$$\text{s.t.} \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i = 0$$

$$\sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i = 0$$

$$h_i(x_i, u_i) \le 0,\ i = 1, \cdots, N$$

How to (re)design the control *u* to reach the optimal solution?

- Distributed
- Closed-loop (state-feedback)

Tool: reverse/forward engineering

Reverse

System Dynamics & Existing Control

$$\dot{x}_i = \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i,\ i = 1, \cdots, N$$

$$\dot{u}_i = \sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i,\ i = 1, \cdots, N$$

Economically Efficient State

$$\min_{x \in \mathbb{R}^n, u \in \mathbb{R}^m} \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N} g_i(u_i)$$

s.t. $$\sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i = 0$$

$$\sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i = 0$$

$$h_i(x_i, u_i) \le 0,\ i = 1, \cdots, N$$

solve

Analogy ?

Optimization Problem

Forward

System Dynamics & Existing Control

$$\dot{x}_i = \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i,\ i = 1, \cdots, N$$

$$\dot{u}_i = \sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i,\ i = 1, \cdots, N$$

Economically Efficient State

$$\min_{x \in \mathbb{R}^n, u \in \mathbb{R}^m} \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N} g_i(u_i)$$

$$\text{s.t.} \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i = 0$$

$$\sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i = 0$$

$$h_i(x_i, u_i) \le 0,\ i = 1, \cdots, N$$

Equivalent

Modified

Optimization Problem

Forward

System Dynamics & Modified Control

$$\dot{x}_i = \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i,$$

$$\dot{u}_i = \sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i + g(z_i),$$

$$\dot{z}_i = f\left(x_j, u_j, z_j : j \in N_i\right)$$

solve

Economically Efficient State

$$\min_{x \in \mathbb{R}^n, u \in \mathbb{R}^m} \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N} g_i(u_i)$$

s.t. $$\sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i = 0$$

$$\sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i = 0$$

$$h_i(x_i, u_i) \le 0, \; i = 1, \cdots, N$$

Equivalent

Modified

Optimization Problem

Distributed Economically-Efficient Control

System Dynamics & Existing Control

$$\dot{x}_i = \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i,\ i = 1, \cdots, N$$

$$\dot{u}_i = \sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i,\ i = 1, \cdots, N$$

Economical Efficient State

$$\min_{x \in \mathbb{R}^n, u \in \mathbb{R}^m} \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N} g_i(u_i)$$

$$\text{s.t.} \sum_{j \in N(i)} A_{ij} x_j + B_i u_i + C_i w_i = 0$$

$$\sum_{j \in N(i)} D_{ij} x_j + \sum_{j \in N(i)} E_{ij} u_j + F_i w_i = 0$$

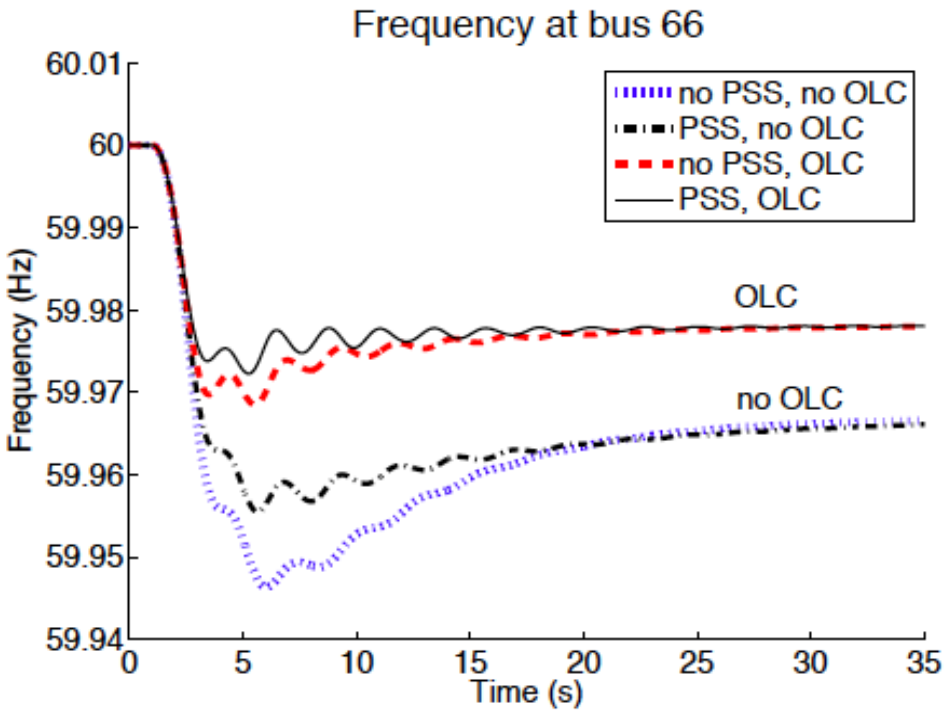
$$h_i(x_i, u_i) \le 0,\ i = 1, \cdots, N$$

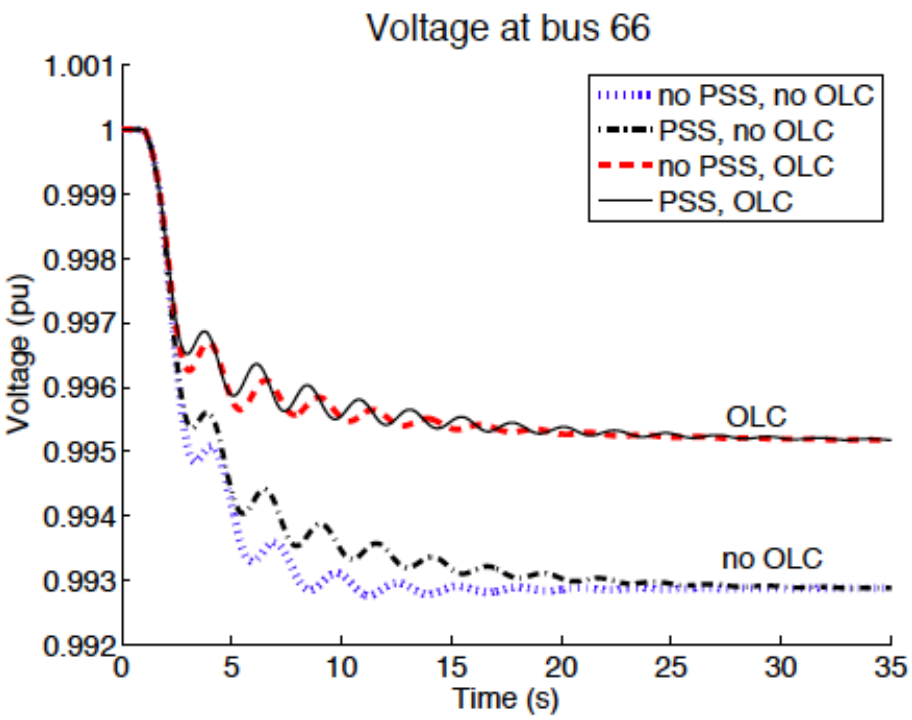
Sufficient and necessary conditions are available at

[Zhang, Antonois, Li, 2015, 2016]

Application to optimal load control for primary freq. control

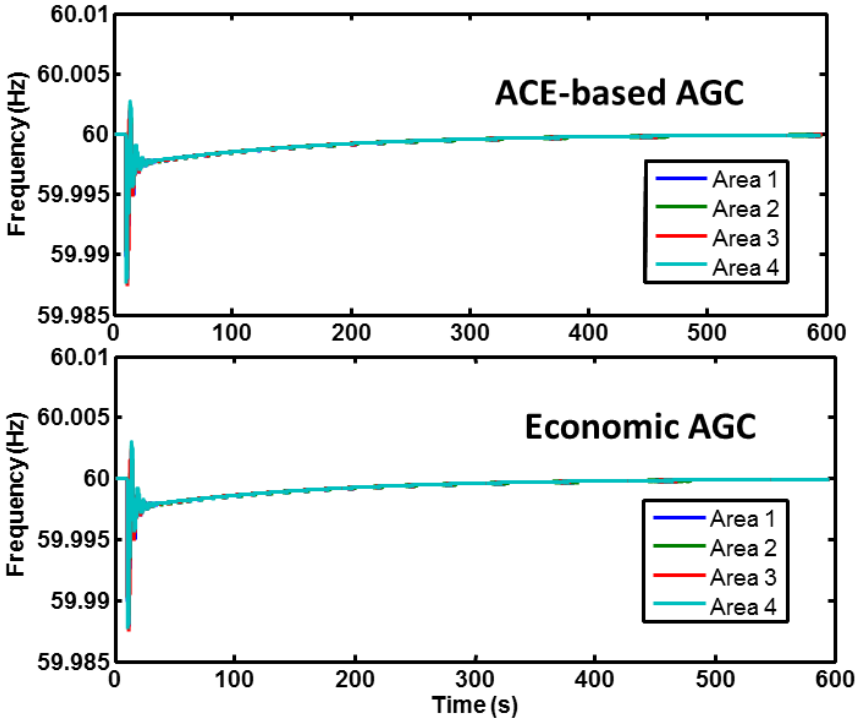
Simulation of IEEE a 68-bus system

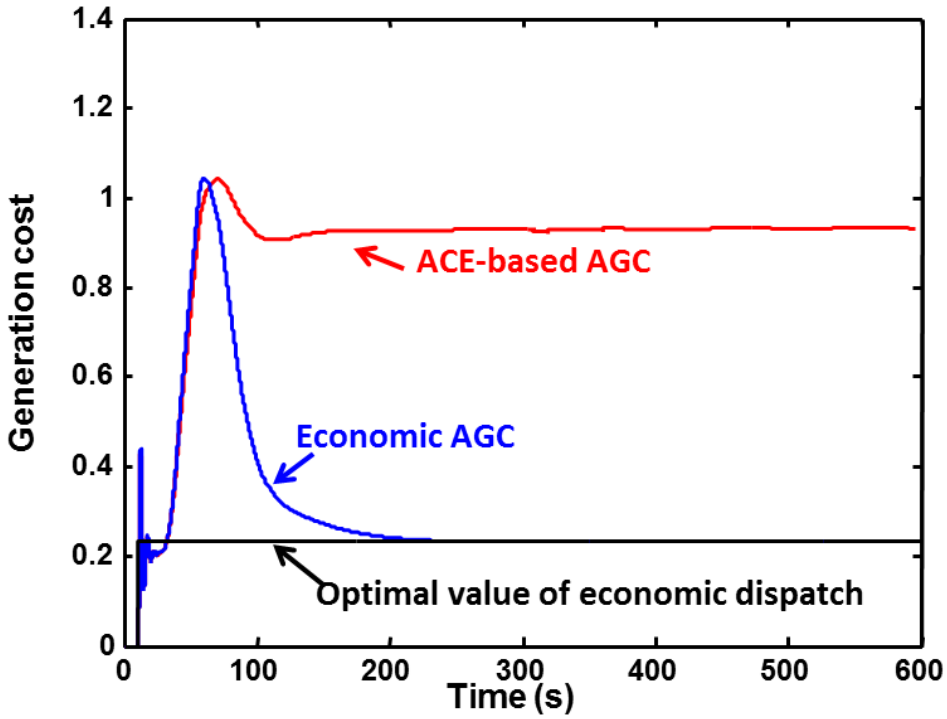




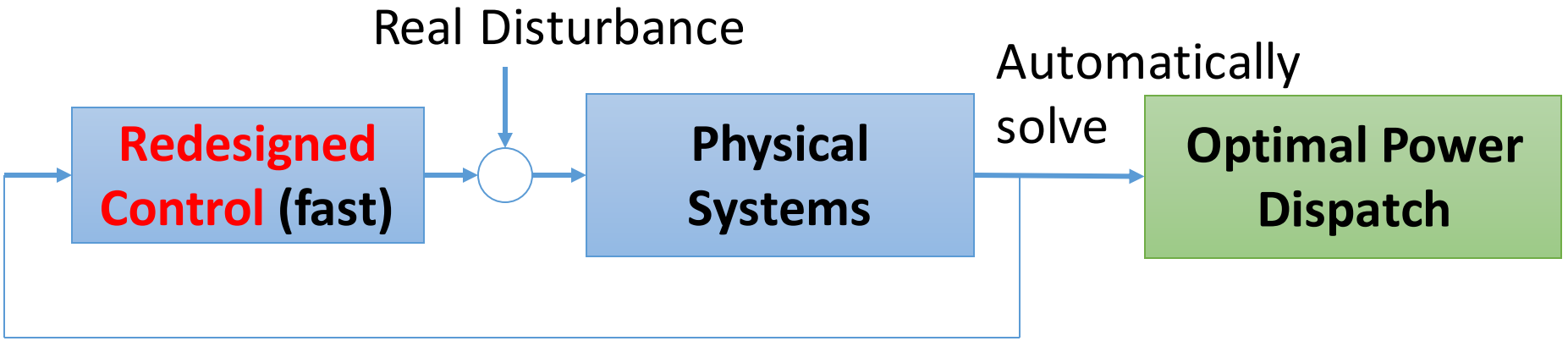
Application to automatic generation control

Simulation of a 4-bus system





Distributed Economically-Efficient Control



Advantages:

For the ***control***: Stable and more economically-efficient

For the ***optimization***:

A large amount of sensing, comm. and comp. is saved

Thank you!

Load frequency control

System Dynamics

$$\dot{\omega}_i = -\frac{1}{M_i}\left(\sum_{l \in L(i)} d_l(t) + D_i \omega_i(t) - P_i^m + \sum_{j:i\to j} P_{ij}(t) - \sum_{k:k\to i} P_{ji}(t) \right)$$

$$\dot{P}_{ij} = b_{ij}\left(\omega_i(t) - \omega_j(t)\right)$$

Load Control

$$d_l(t) = \left[C_l'^{-1}\left(\omega_i(t)\right) \right]_{\underline{d}_l}^{\bar{d}_l} \quad \text{for } l \in L(i)$$

- ***Frequency: a locally measurable signal ("price" of imbalance)***
- ***Completely decentralized; no explicit communication necessary***

Dual Gradient Algorithms

Step 1: Each appliance i updates the power request $q_i(t)$

~~& sends to the control center~~

$$q_i(t) = \arg\max_{q_i} U_i(q_i) - p(t)q_i$$

Step 2: Control center updates the signal $p(t)$

& sends to each appliance

$$p(t+1) = [p(t) + \gamma(\sum_i q_i(t) - Q)]^+$$

Replace this with *true measurement* of total power consump. Q(t).

Normalized Gradient

Problem: $\min_p f(p)$

Gradient Descent: $p(t+1) = p(t) - \gamma(t) f'(p(t))$

Normalized Gradient Descent: $p(t+1) = p(t) - \gamma(t) \frac{f'(p(t))}{||f'(p(t))||}$

Questions: Reduce the communication?

If $p \in \mathbb{R}$, normalized gradient reduces to binary signal, ± 1.

If $p \in \mathbb{R}^K$, $K > 1$, normalized gradient is in $\mathcal{S}^{K-1}$(unit spher

Primal Problem

$$\begin{aligned} \max_{q_i} \quad & \sum_i U_i(q_i) \\ \text{s.t.} \quad & Aq \leq B \end{aligned}$$

e.g. Network constraints;
Multi-resource allocation

Quantized (Normalized) Gradient Descent (QGD)

Problem: $\min_p f(p)$

QGD: $p(t+1) = p(t) - \gamma(t)d(t), d(t) \in \mathcal{D} \in \mathcal{S}^{K-1}$

Question: When $\mathcal{D}$ is a good quantization?

$\mathcal{F}$: set of convex and differential functions whose gradients are $L-$Lipschitz continuous and $\mathcal{P}^*$ is nonempty and bounded.

$\mathcal{P}^*$: set of optimizers

Proper Quantization

Definition [Proper Quantization]
A set $\mathcal{D}$ is a proper quantization if for every $f \in \mathcal{F}$ and every initiation $p(0) \in \mathbb{R}^K$, there are $d(t) \in \mathcal{D}$ and $\epsilon(t) \in \mathbb{R}_+$, where $t \in \mathbb{N}$ such that

$$\lim_{t \to \infty} \text{dist}(p(t), \mathcal{P}^*) = 0.$$

A quantization is proper if and only if the algorithm is able to converge to the optimal points for any well-behaviored f.

Quantized (Normalized) Gradient Descent (QGD)

Problem: $\min_p f(p)$

QGD: $p(t+1) = p(t) - \gamma(t)d(t), d(t) \in \mathcal{D} \in \mathcal{S}^{K-1}$

Questions:

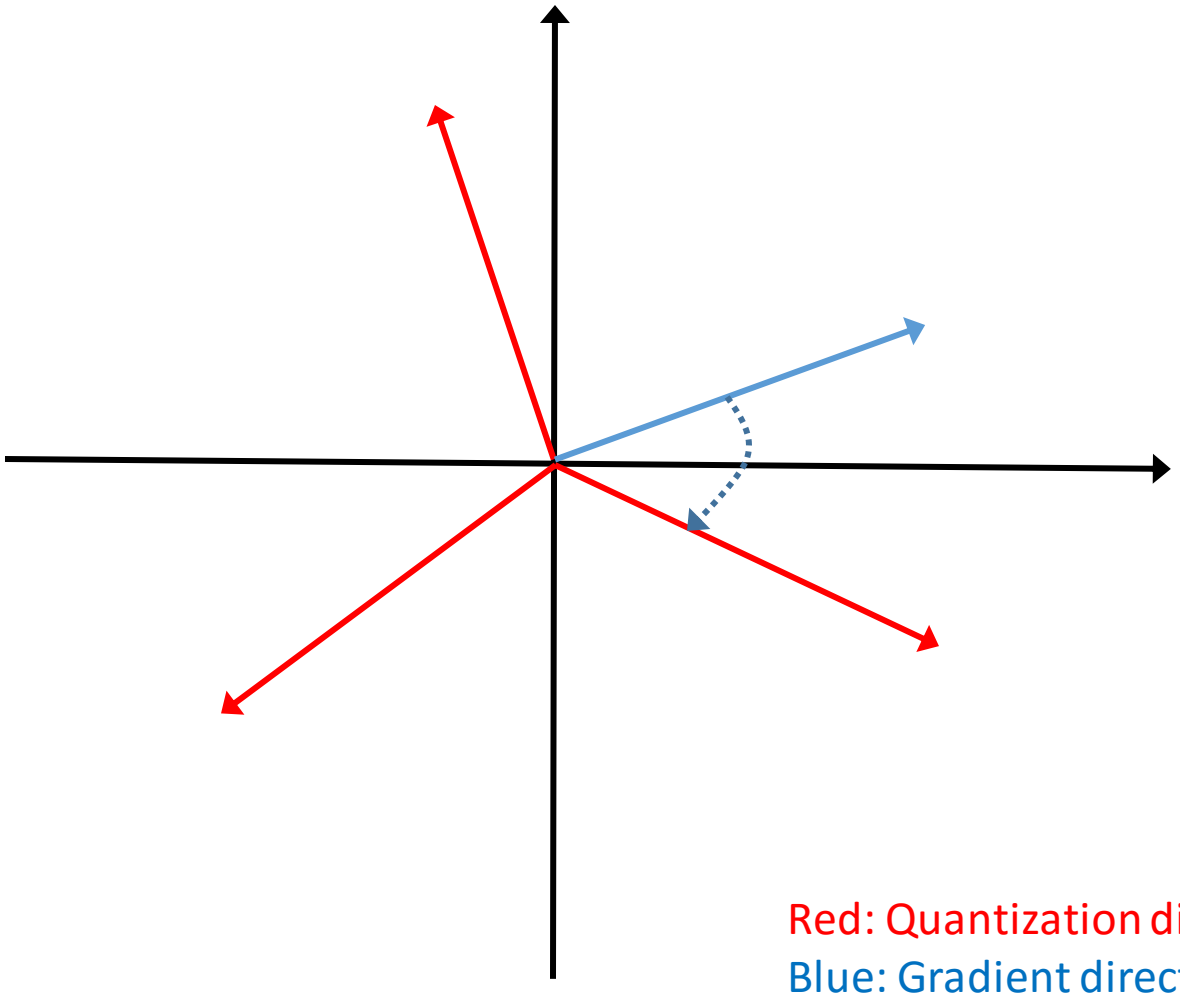
A) How to determine a quantization is proper?

B) What is the *minimal size* of the quantization to be proper?

C) How to choose d(t) and ε(t) , Given a proper quantization?

D) What are the *connections* between the *fineness* of the quantization to the *convergence* of the algorithm?

Descent direction



Red: Quantization direction
Blue: Gradient direction

Proper quantization

Proposition. A set $\mathcal{D}$ is a proper quantization if and only if there exists $\theta > 0$, for every $g \in \mathcal{S}^{K-1}$ there exists $d \in \mathcal{D}$ such that

$$\cos\left(\text{ang}(g, d)\right) \geq \theta.$$

Such quantization $\mathcal{D}$ is also called as $\theta-$cover.

Minimal proper quantization, $|\mathcal{D}| = N + 1$.

Only $\log_2(N + 1)$ bits are needed at each iteration.

"Convergence of limited communications gradient methods", Magnusson, Heal, Enyioha, Li, Fischione, Tarokh, ACC submitted

Convergence rate

Stopping Criterion: $\mathcal{P}(\epsilon) := \left\{p \in \mathbb{R}^K : ||\nabla f(p)|| \leq \epsilon\right\}$

Theorem: For any $\epsilon > 0$, if $\gamma \in (0, 2\theta\epsilon/L)$, then there exists $T \in \mathbb{N}$ such that $p(T) \in \mathcal{P}(\epsilon)$, with T bounded by

$$T \leq \frac{2(f(p(0)) - f^*)}{\gamma(2\theta\epsilon - L\gamma)}.$$

- Finer quantization, larger stepsize is allowed
- Finer quantization, faster convergence

"Convergence of limited communications gradient methods", Magnusson, Heal, Enyioha, Li, Fischione, Tarokh, ACC submitted

Convergence rate

Stopping Criterion: $F(\epsilon) := \sup\{f(p) : p \in \mathcal{P}(\epsilon)\}$

Theorem:

1) For any $\epsilon > 0$, if $\gamma \in (0, 2\theta\epsilon/L)$, and $T \in \mathbb{N}$ such that $p(T) \in \mathcal{P}(\epsilon)$, then for all $t \geq T$

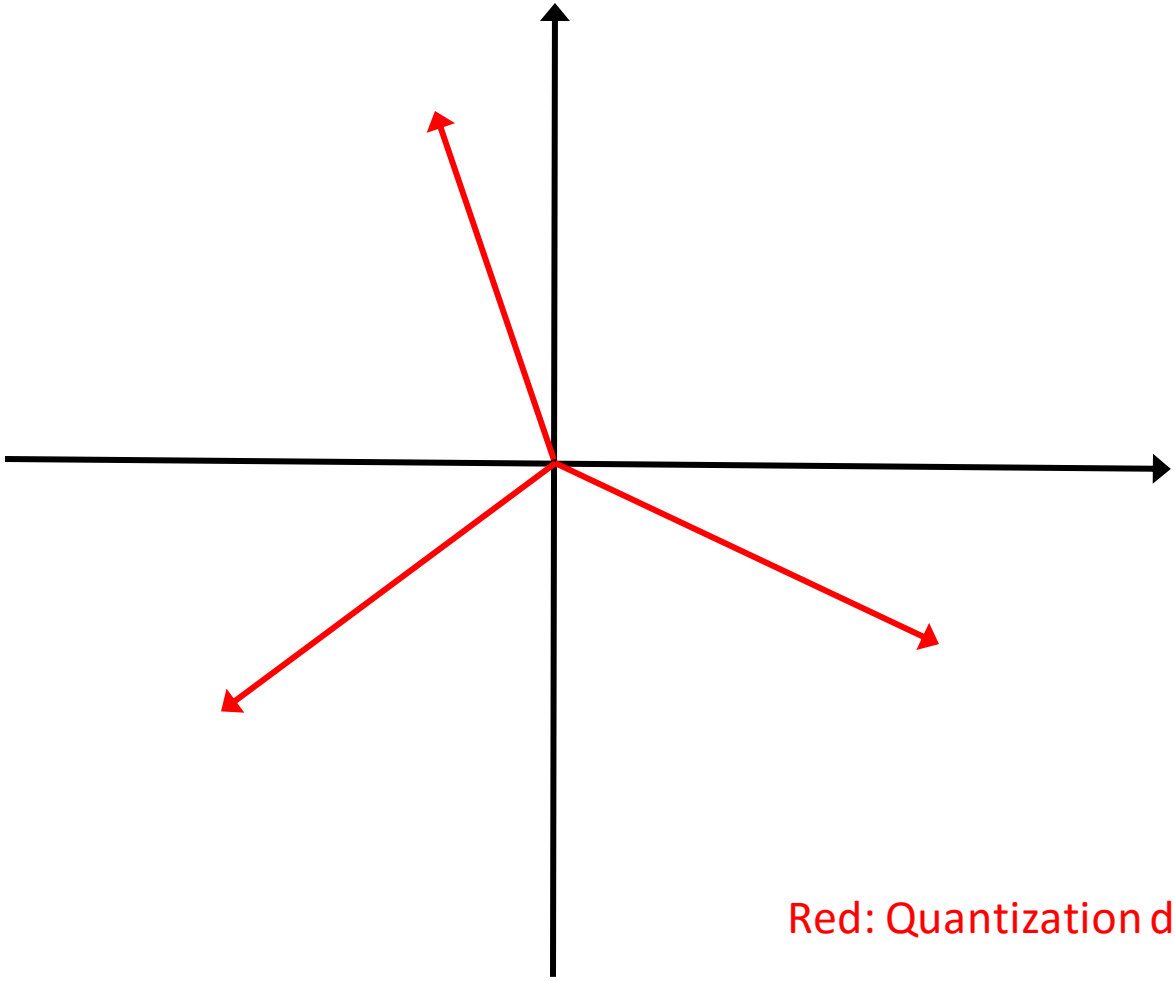
$$f(p(t)) \leq F(\epsilon) + \left(\epsilon + \frac{L}{2}\gamma\right)\gamma.$$

2) If the stepsize $\gamma(t)$ is non-summable and square summable, then $\lim_{t\to\infty} \text{dist}(p(t), \mathcal{P}^*) = 0$.

"Convergence of limited communications gradient methods", Magnusson, Heal, Enyioha, Li, Fischione, Tarokh, ACC submitted

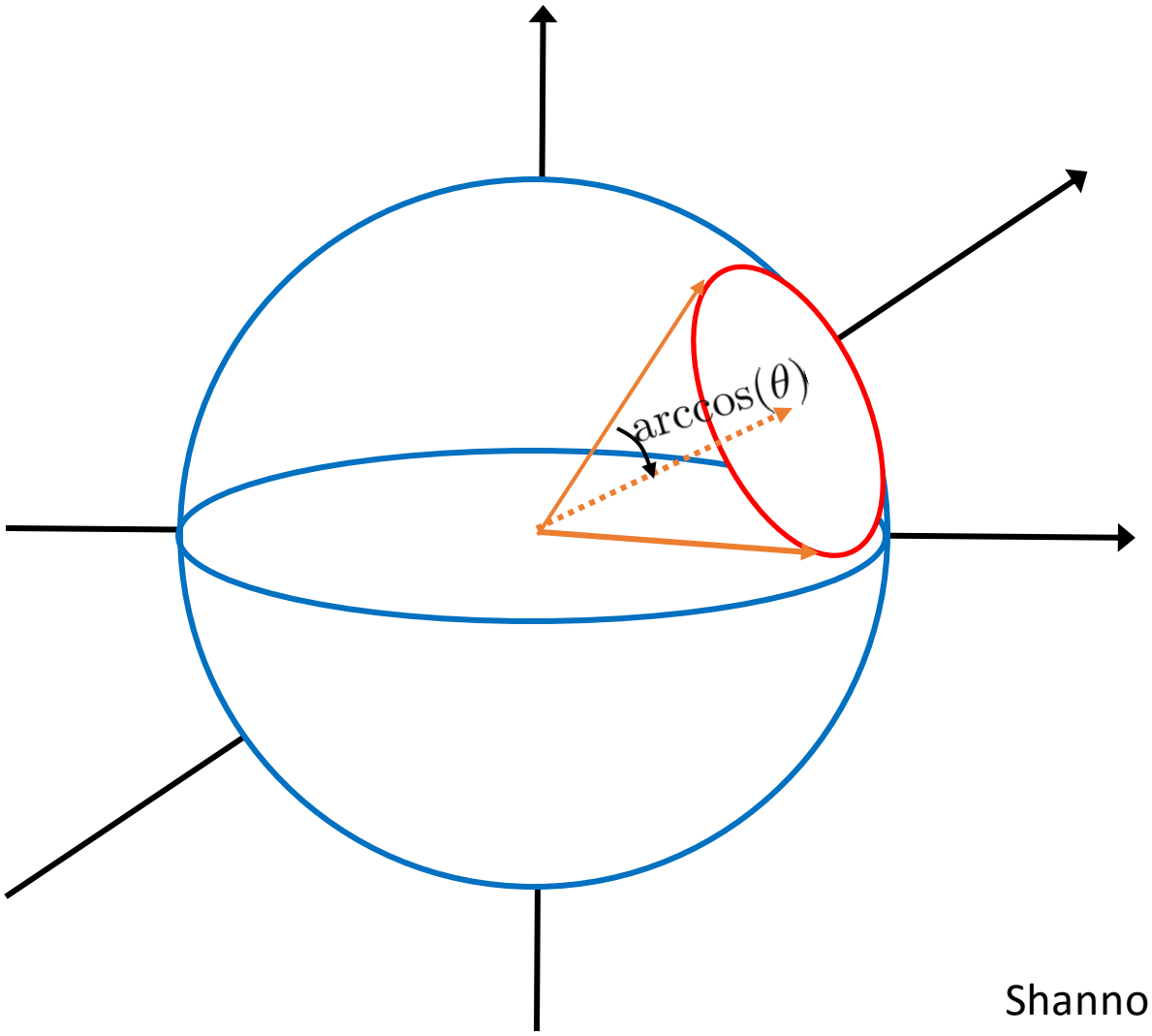
Quantization size

$K = 2$: Given θ, minimal $|\mathcal{D}| = \frac{\pi}{\arccos(\theta)}$



Red: Quantization direction

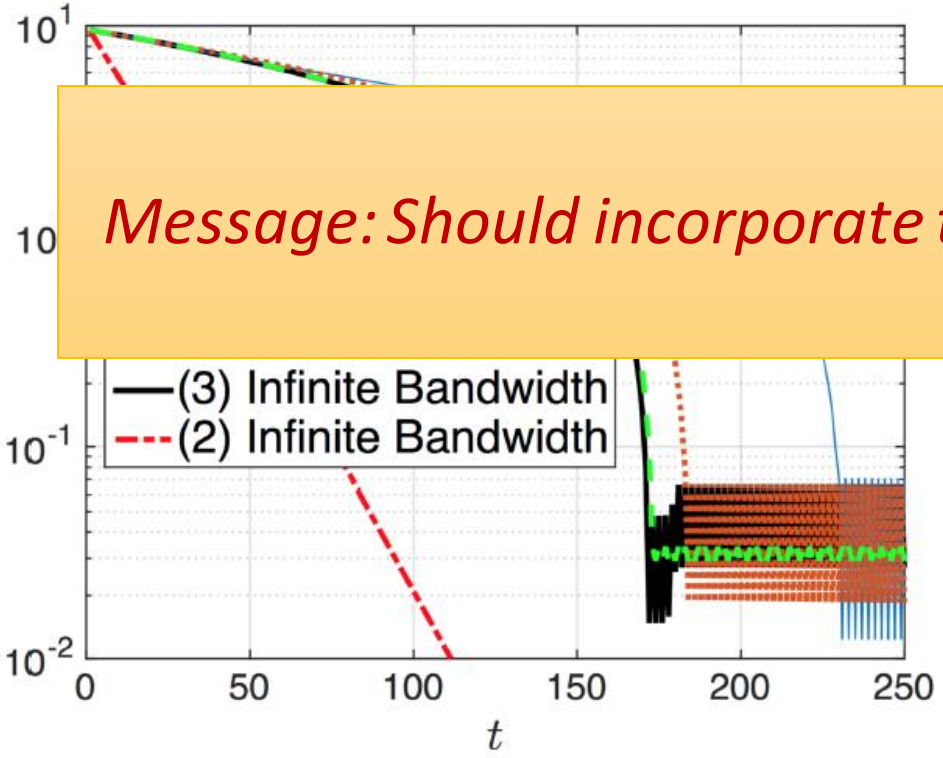
Quantization size



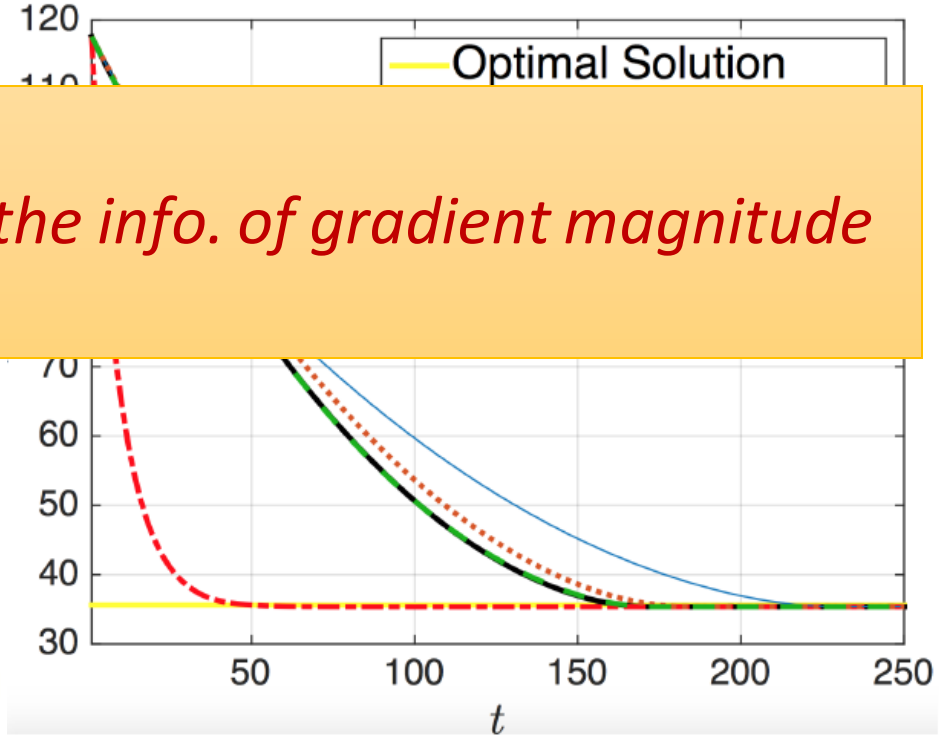
Shannon,1959

Simulation

f is the dual function of a two-dimension resource allocation problem.



$||\nabla f(p(t))||$



$f(p(t))$

Message: Should incorporate the info. of gradient magnitude

(3) Infinite bandwidth: normalized gradient

(2) Infinite bandwidth: gradient

Summary of QGD

Problem: $\min_p f(p)$

QGD: $p(t+1) = p(t) - \gamma(t)d(t), d(t) \in \mathcal{D} \in \mathcal{S}^K$

A) Proper quantization = θ-cover

B) Minimal size of proper quantization is N+1

C) Pick the quantized direction closest to gradient direction

D) θ plays an important role in determines the convergence

Proper quantization

Minimal proper quantization, $|\mathcal{D}| = N + 1$.

Only $\log_2(N+1)$ bits are needed at each iteration.

Finer quantization, larger stepsize is allowed

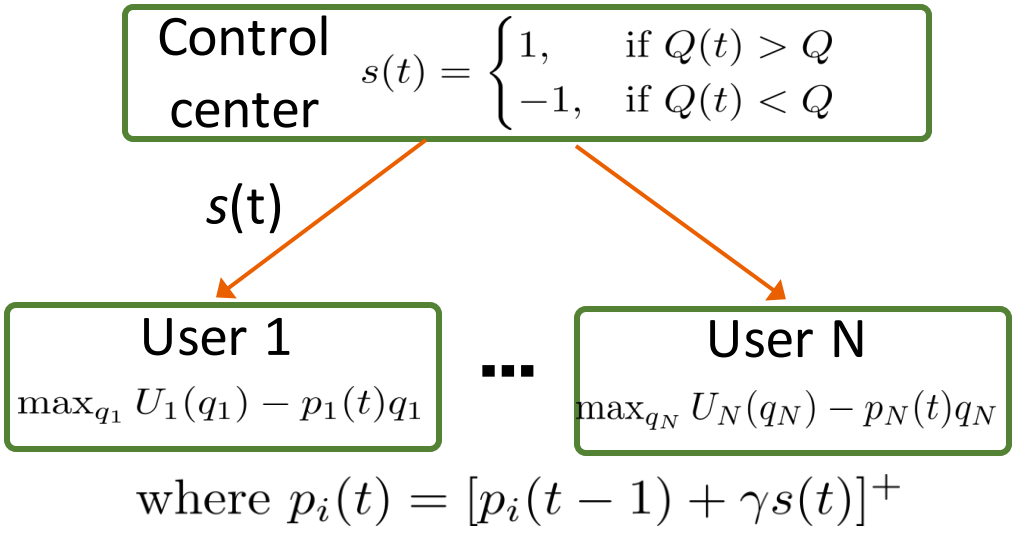
Finer quantization, faster convergence

"Convergence of limited communications gradient methods",
Magnusson, Heal, Enyioha, Li, Fischione, Tarokh, ACC submitted

Communication Complexity

$$\max_{q_i \in [m_i, M_i]} \sum_{i=1}^{N} U_i(q_i)$$

$$s.t. \quad \sum_i q_i \leq Q$$

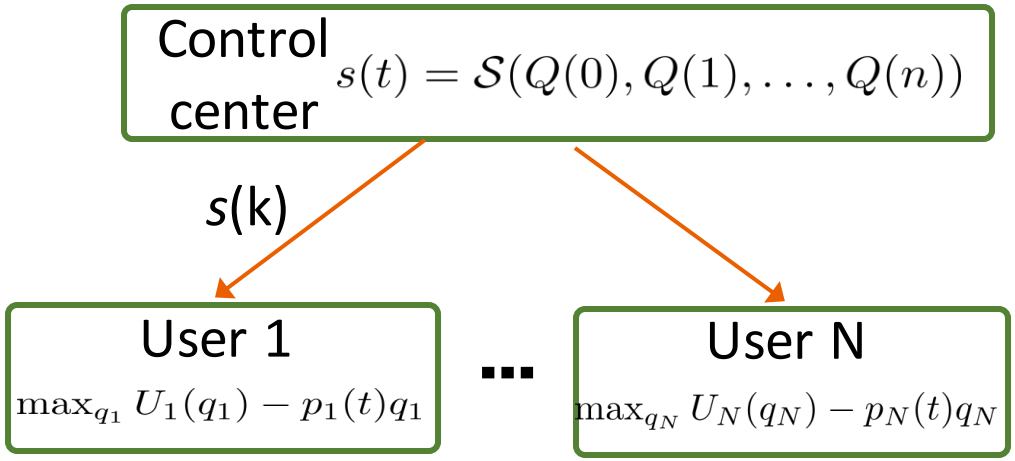


where $p_i(t) = [p_i(t-1) + \gamma s(t)]^+$

$$q_i(t) = \arg\max U_i(q_i) - p(t)q_i$$

$$\Delta_t = \theta_t(f'(p(0)), \ldots, f'(p(t)))$$

$$t+1) = [p(t) - \Delta(t)]_+$$



where $p_i(t) = [p_i(t-1) + \gamma s(t)]^+$