Structure-Exploiting Reinforcement Learning Control with Applications to Power System Dynamics

Our webinar will begin in a few minutes

Introduction

- Welcome
- FREEDM Overview
- Zoom Functionality

freedm.ncsu.edu

FREEDM SYSTEMS CENTER

Structure-Exploiting Reinforcement Learning Control with Applications to Power System Dynamics

Sayak Mukherjee, Aranya Chakrabortty

April 29, 2020

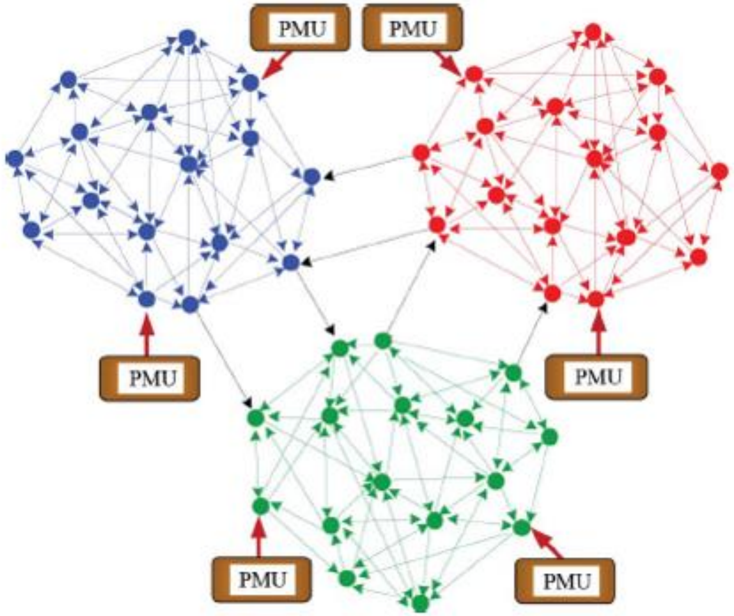
Towards data-driven control

Wide-area damping control of power grid

- The system dynamics of the large-scale network systems may not always be known

US Eastern Interconnection power grid consists of over 70,000 buses and thousands of dynamic elements

- Most of the designs in the existing literature are performed on a relatively small system with known dynamics



Use measurement-driven approach

Controlling inter-area oscillations in power systems

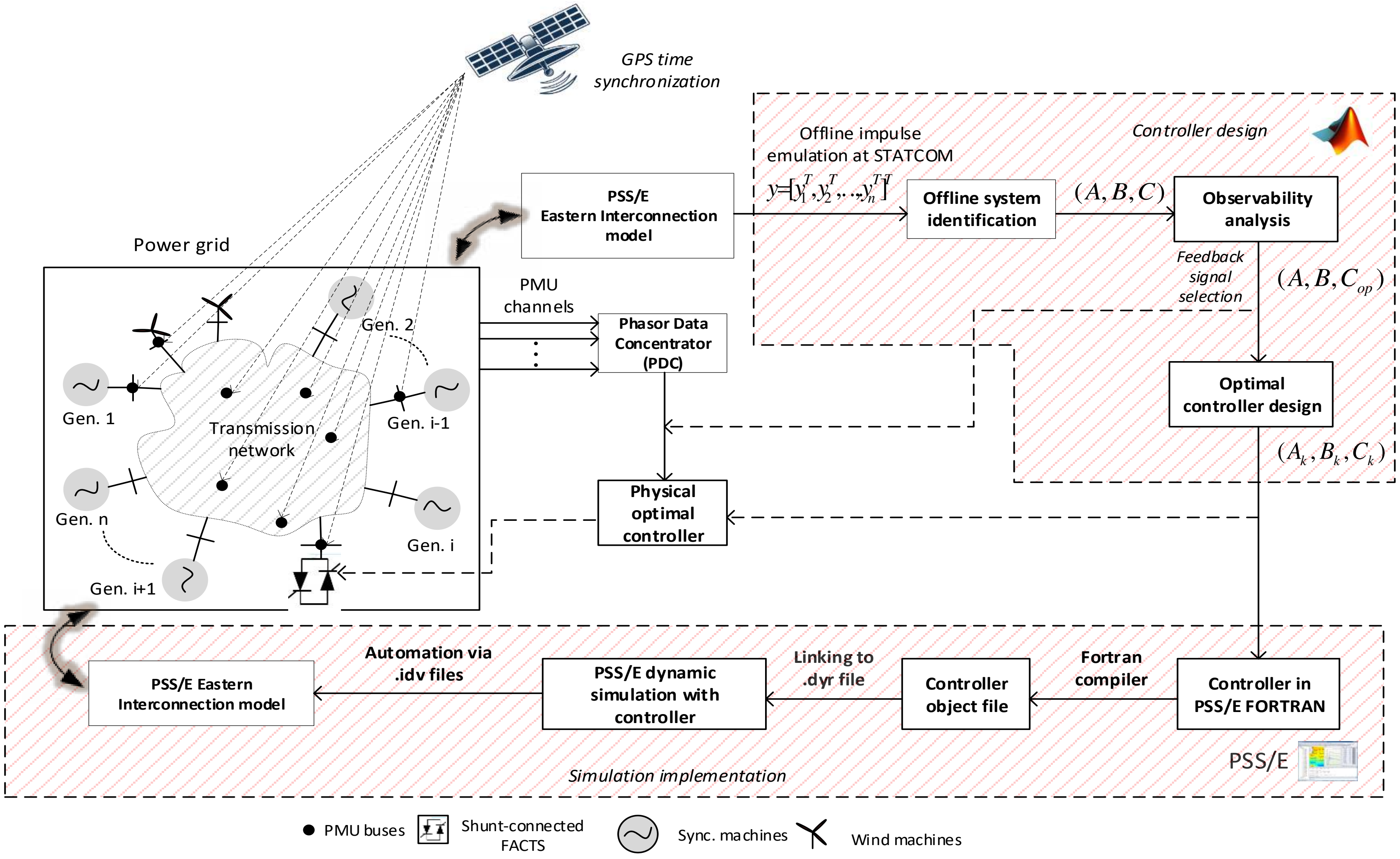
Approach 1: an indirect control design with intermediate identification

Approach 2: a direct control design via Reinforcement Learning (RL)

- Motivating example using approach 1 : Measurement-based optimal controller on a practical full-scale transmission model of New York State grid

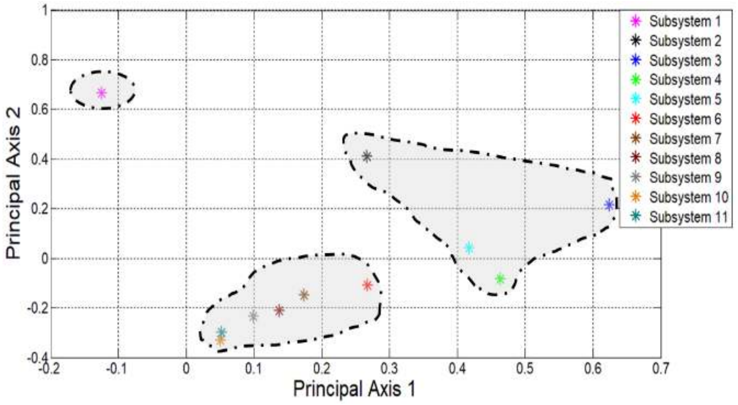
Approach 1: Measurement-driven control of NYS grid

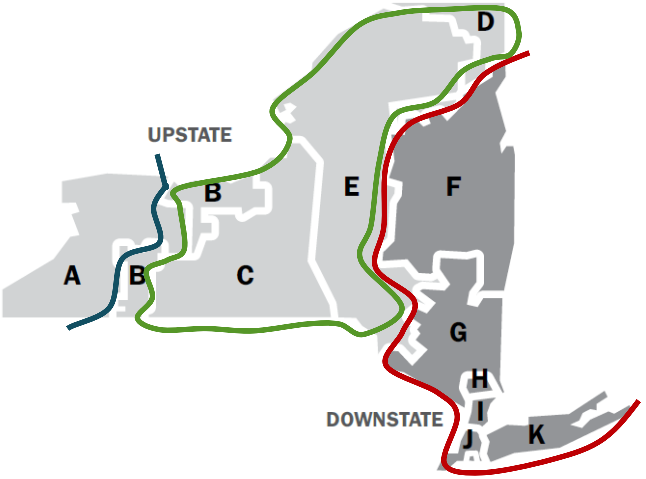
- Objective: Improve the inter-area oscillation performance pertaining to NYS grid



Approach 1: Control design

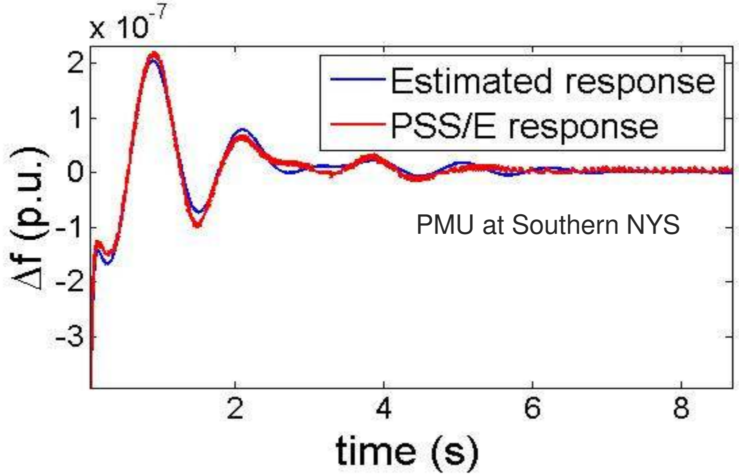
Coherency of NYS grid using PCA



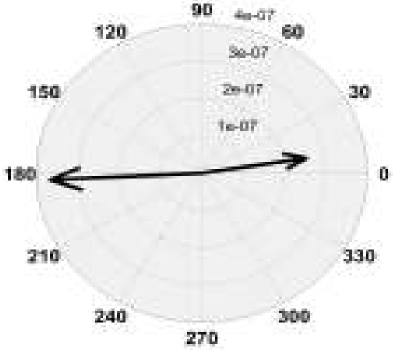


Map source: 2018 Power Trends, NYISO

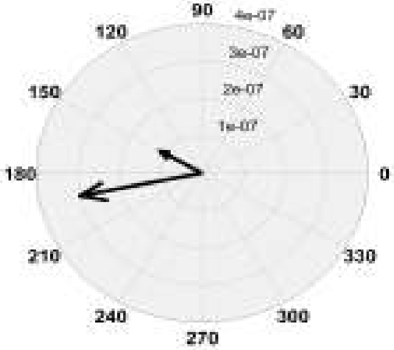
Reduced-order system identification using ERA



Observability analysis



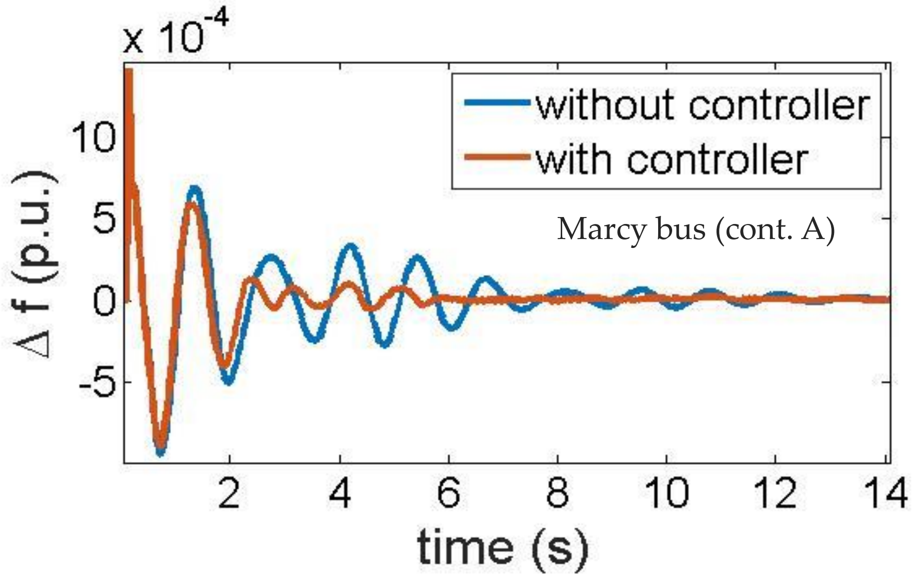
Western PMU

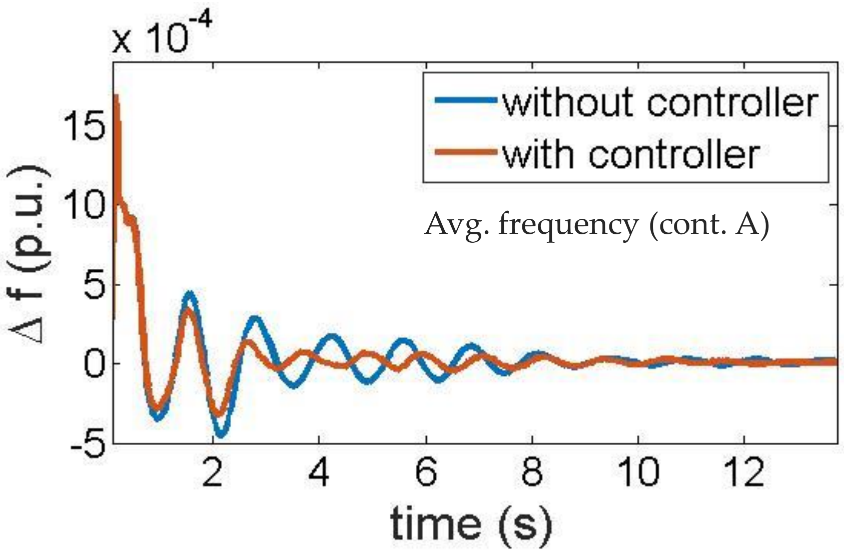


Central PMU

Approach 1: Control performance

- A Linear Quadratic Gaussian type optimal controller is designed. It is then coded in FORTRAN as a user written module and linked with PSS/E EI dynamic suit and tested with NYISO specified contingencies.





Freq. scaling translates to several hundreds of MWs of fluctuations

- We showcase a sequential data-driven damping control design approach on the FACTS facility in NYS using tools from machine learning, linear system identification and optimal control theory.

From indirect to direct approach

Indirect approach

Plant (e.g. power system)

System identification: (A,B) dynamic matrices

Optimal controller solution

$$A^T P + PA - PBR^{-1}B^T P + Q = 0$$

Model dependent

- What if operating conditions keep on changing?
- New components are integrated to the system (e.g. DERs for power grids)

Run approach 1 multiple times

- Can we avoid that?

➢ Yes, perform direct gain computation via Reinforcement Learning

Optimality + Adaptiveness

FREEDM SYSTEMS CENTER

A direct approach via reinforcement learning

NC STATE UNIVERSITY

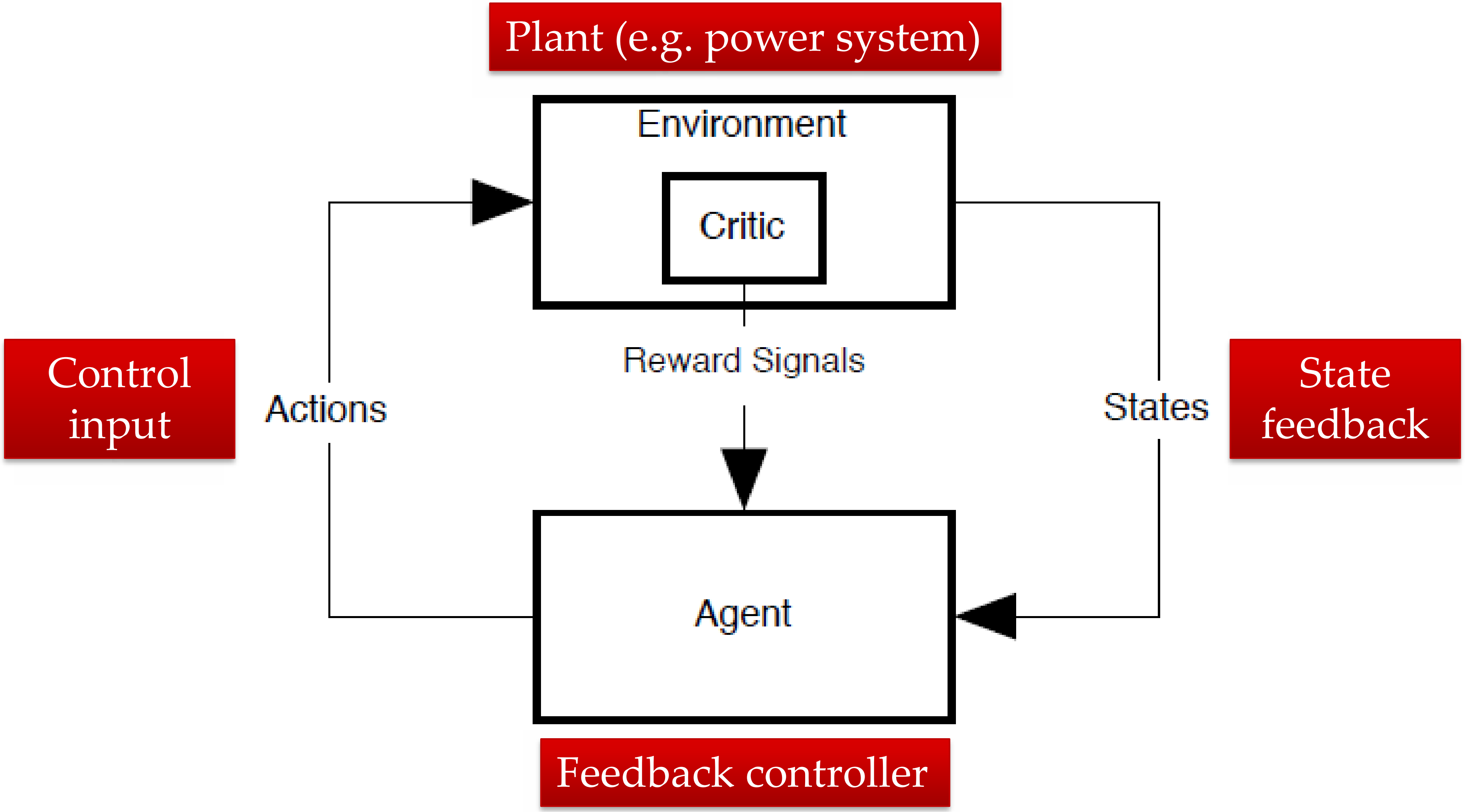
- Reinforcement Learning (RL)/ Adaptive Dynamic programming (ADP) is generally used in finding the feedback gains without explicit state dynamic descriptions
- Classically RL has been used for sequential decision making using *Markov Decision Processes (MDPs)* in AI community
- However, in the last decade, RL has been used to control *dynamical systems*
  (Vrabie et al., Automatica, 2009, Jiang et al., Automatica, 2012, etc.)

Opportunity to design RL controls for power system dynamics

- Major challenges : 1. Learning the feedback gain using full-dimensional system *increases learning time*, and 2. results in *dense feedback control* structure

One solution – Incorporate the ideas of model reduction in conjunction with RL

What is RL?
How is RL related to feedback control?



Barto et al. (2004)

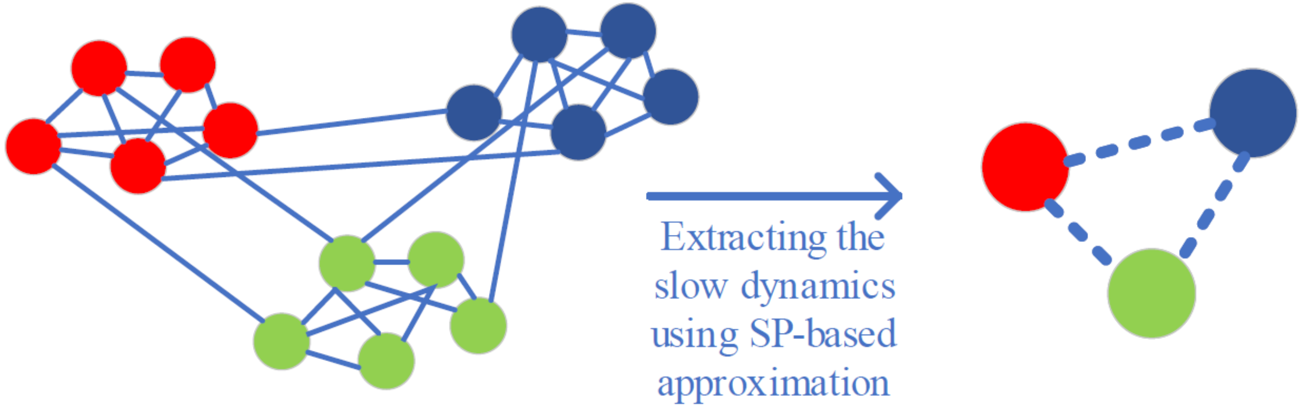
Exploit the structure : reduced-dim. learning

- Large-scale physical networks → specific structures in their dynamics

- Power grid is a complex dynamic network which shows coherency/ clustering, resulting in time-scale separated state dynamics

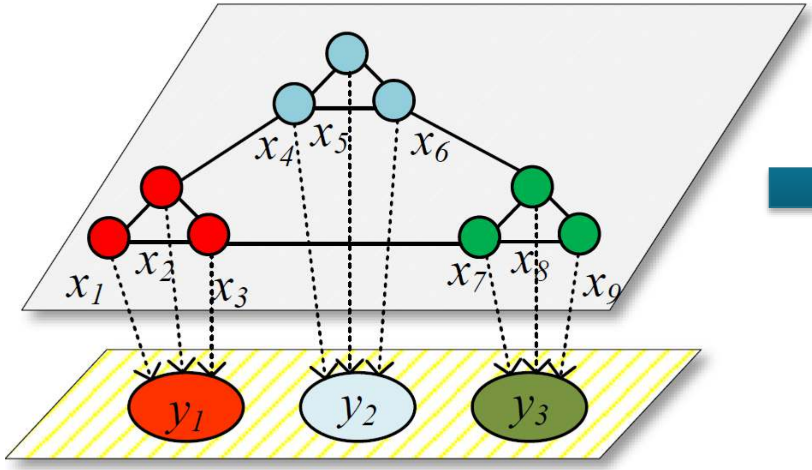
mathematically represented by singular perturbation theory

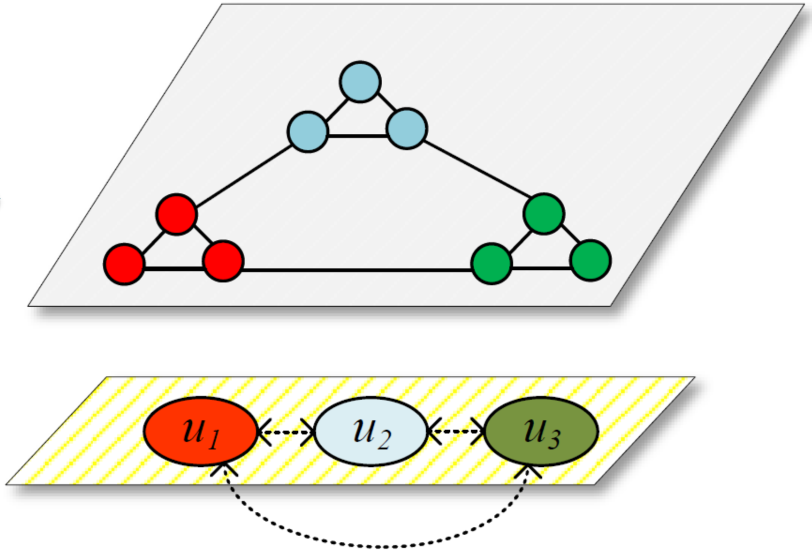
- Use the slow states to control inter-cluster oscillations, known as *inter-area modes* in multi-area power systems



Aggregate-control-inversion RL

Mukherjee, Bai, Chakrabortty (2018)

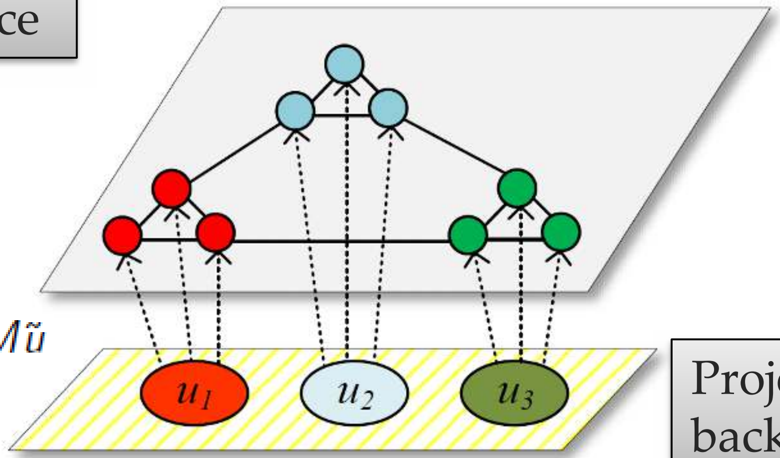




Projection to lower dimensional space

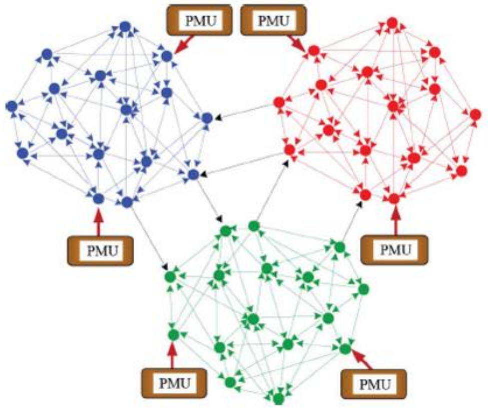
Reduced dimensional learning of feedback gains

$u = M\tilde{u}$



Projecting the gains back to original coordinates

Brief overview of theoretical development



Overall dynamics: $\dot{x} = Ax + Bu, \quad x(0) = x_0, x \in \mathbb{R}^n, u \in \mathbb{R}^m$

Underlying two time-scale dynamics:

$$\dot{y} = A_{11}y + A_{12}z + B_1u, \; y(0) = y_0,$$
$$\epsilon\dot{z} = A_{21}y + A_{22}z + B_2u, \; z(0) = z_0$$

Control: u=-M(Ky) without knowing A and B

$$J(y(0); \tilde{u}) = \int_0^\infty (y^TQy + \tilde{u}^TR\tilde{u})dt,$$
$$\text{s.t.} \; A - BMKC \in \mathbb{RH}_\infty.$$

Performance criterion

Stability

Model-free Linear Quadratic Regulator (LQR)

Ideal slow dynamics $\epsilon = 0$: $\dot{y}_s = A_sy_s + B_su_s, \quad y_s(0) = y(0), u = u_s + u_f,$

$A_s = A_{11} - A_{12}A_{22}^{-1}A_{21}$

$B_s = B_1 - A_{12}A_{22}^{-1}B_2$

Algebraic Riccati Equation:

$$A_s^T\bar{P} + \bar{P}A_s + Q - \bar{P}(B_sM)R^{-1}(B_sM)\bar{P} = 0,$$
$$\bar{K} = R^{-1}M^TB_s^T\bar{P}$$

Still model-based solution

From model-based to model-free formulation

Theorem [Kleinman, TAC, 1968]

Let $\bar{K}_0 \in \mathbb{R}^{r\times r}$ be any stabilizing gain matrix, then for $k = 0, 1, \ldots$

1. Solve for $\bar{P}_k$ starting with stabilizing $\bar{K}_0$(Policy Evaluation) :

$$A_{ck}^T \bar{P}_k + \bar{P}_k A_{ck} + Q + \bar{K}_k^T R \bar{K}_k = 0, A_{ck} = A_s - B_s M \bar{K}_k$$

2. Update feedback matrix (Policy update) :

$$\bar{K}_{k+1} = R^{-1} M^T B_s^T \bar{P}_k$$

Then $A_s - B_s M \bar{K}$ is Hurwitz and $\bar{K}_k$ and $\bar{P}_k$ would converge to optimal $\bar{K}, \bar{P}$.

Model-based iterative solution

Derivative along the trajectory

$$\frac{d}{dt}(y_s^T \bar{P}_k y_s) = y_s^T (A_{ck}^T \bar{P}_k + \bar{P}_k A_{ck}) y_s + 2(\bar{K}_k y_s + u_0)^T M^T B_s^T P_k y$$

$$= -y_s^T \bar{Q}_k y_s + 2(\bar{K}_k y_s + u_0)^T R \bar{K}_{k+1} y_s$$

Eliminate model information using Kleinman's algorithm

Taking integral

Trajectory-based Data-driven solution

$$y_s^T(t+T)\bar{P}_k y_s(t+T) - y_s^T(t)\bar{P}_k y_s(t)$$

$$-2\int_t^{t+T} ((\bar{K}_k y_s + u_0)^T R \bar{K}_{k+1} y_s) d\tau = -\int_t^{t+T} y_s^T \tilde{Q}_k y_s d\tau$$

$y_s(t)$ is ideal, and not implementable, so replace with *$y(t)$*

The SP-RL algorithm

Algorithm steps:

- 1. *Data storage:* *Store* data (y and u_0) during the exploration phase for interval $(t_1, t_2, \cdots, t_l)$, $t_i - t_{i-1} = T$, where T is the learning time step. The learning requires $r^2 + 2r^2(n = r, m = r)$ time samples. Then *construct* the following matrices,

$$\delta_{yy} = \left[y \otimes y|_{t_1}^{t_1+T}, \quad \cdots, \quad y \otimes y|_{t_l}^{t_l+T}\right]^T,$$

$$I_{yy} = \left[\int_{t_1}^{t_1+T}(y \otimes y)d\tau, \quad \cdots, \quad \int_{t_l}^{t_l+T}(y \otimes y)d\tau\right]^T,$$

$$I_{yu_0} = \left[\int_{t_1}^{t_1+T}(y \otimes u_0)d\tau, \quad \cdots, \quad \int_{t_l}^{t_l+T}(y \otimes u_0)d\tau\right]^T.$$

Store measurements

$y_s(t)$ is related to $y(t)$ by *singular perturbation parameter* following Chow, Kokotovic, TAC, 1976, 85

- 2. *Controller update iteration :* Starting with a stabilizing K_0, *Solve* for K iteratively ($k = 0, 1, \cdots$) once matrices δ_{yy}, I_{yy}, I_{yu_0} are constructed and iterative equation can be written for each small learning steps as,

$$\underbrace{\left[\delta_{yy} \quad -2I_{yy}(I_n \otimes K_k^T R) - 2I_{yu_0}(I_n \otimes R)\right]}_{\Theta_k} \begin{bmatrix} vec(P_k) \\ vec(K_{k+1}) \end{bmatrix} = \underbrace{-I_{yy}\, vec(Q_k)}_{\Phi_k}.$$

Run least-square

Hence P_k and K_{k+1} are iteratively solved such that $|P_k - P_{k-1}| < \bar{\epsilon}$, $\bar{\epsilon}$ is a small threshold.

- 3. *Applying K on the system :* Next $\tilde{u} = -Ky$ is applied and u_0 source is removed.

Theoretical results in a nutshell

Sub-optimality theorem [Mukherjee, Bai, Chakrabortty (2018)]

Assuming $||y_s(t)||$ *and* $||u_0(t)||$ *are bounded the sub-optimal solutions are given by*

$$P = \bar{P} + O(\epsilon), K = \bar{K} + O(\epsilon).$$

Here $||.||$ *means the Euclidean norm.*

Corollary

The optimal objective value J with $y(t)$ feedback is related to $\bar{J}$ for the reduced slow sub-system with $y_s(t)$ feedback as,

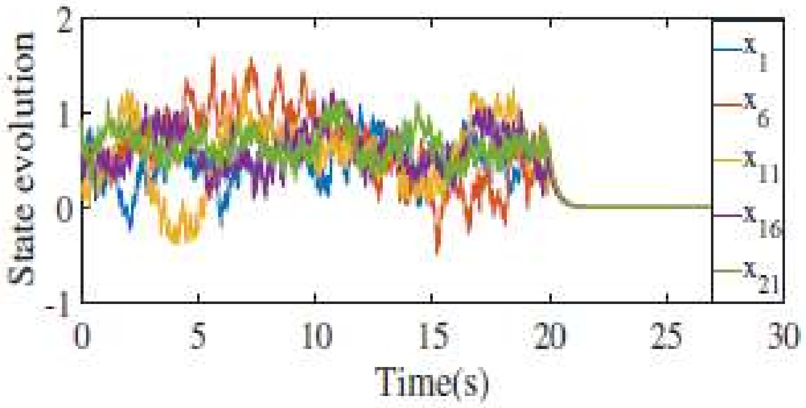
$$J = \bar{J} + O(\epsilon)$$

Stability theorem [Mukherjee, Bai, Chakrabortty (2018)]

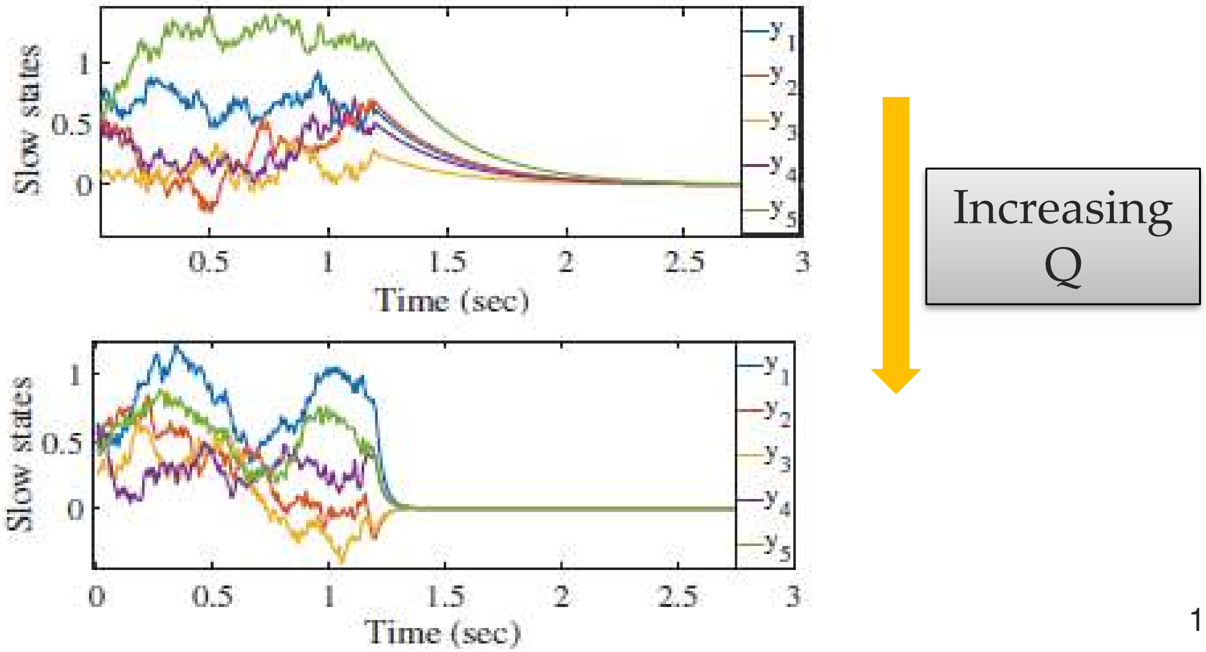
Assume that the control policy $u = -MK_k y$ *at the* k^{th} *iteration is asymptotically stable. Suppose that* $R \succ 0$ *and* $Q \succ 0$ *with* $\lambda_{min}(Q)$ *is sufficiently large. Then the control policy at the* $(k+1)^{th}$ *iteration given by* $u = -MK_{k+1}y$ *is asymptotically stable.*

Simulations on clustered network

- We next consider a simplified LTI clustered multi-agent network with 25 agents divided into 5 clusters
- The network has 4 slow eigen-values, one zero eigen-value and rest are the fast eigen-values. The slow eigen-values are -0.128, -0.195, -0.196, and -0.2638.
- The control is enabled by Aggregation-control-inversion architecture

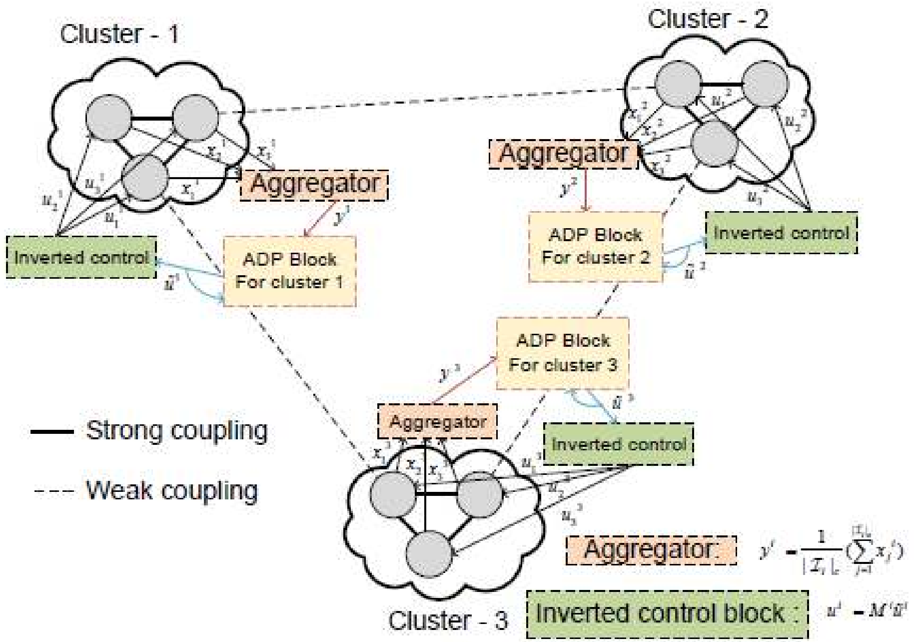


- It takes at least 18.75 seconds to learn the $K \in \mathbb{R}^{25\times 25}$ matrix accurately (atleast $n^2 + 2mn$ samples are needed to learn K; here $n = m = 25$, $T = 0.01s$)
- The reduced-order control requires at least only $r^2 + 2r^2 = 75$ number of samples, and for simulation we consider $1.2s$ for learning with 120 samples



Other design variants

- Output feedback design using Neuro-adaptive observer
- Block-decentralized design ⇨



- Incorporating robustness to the projection-based RL

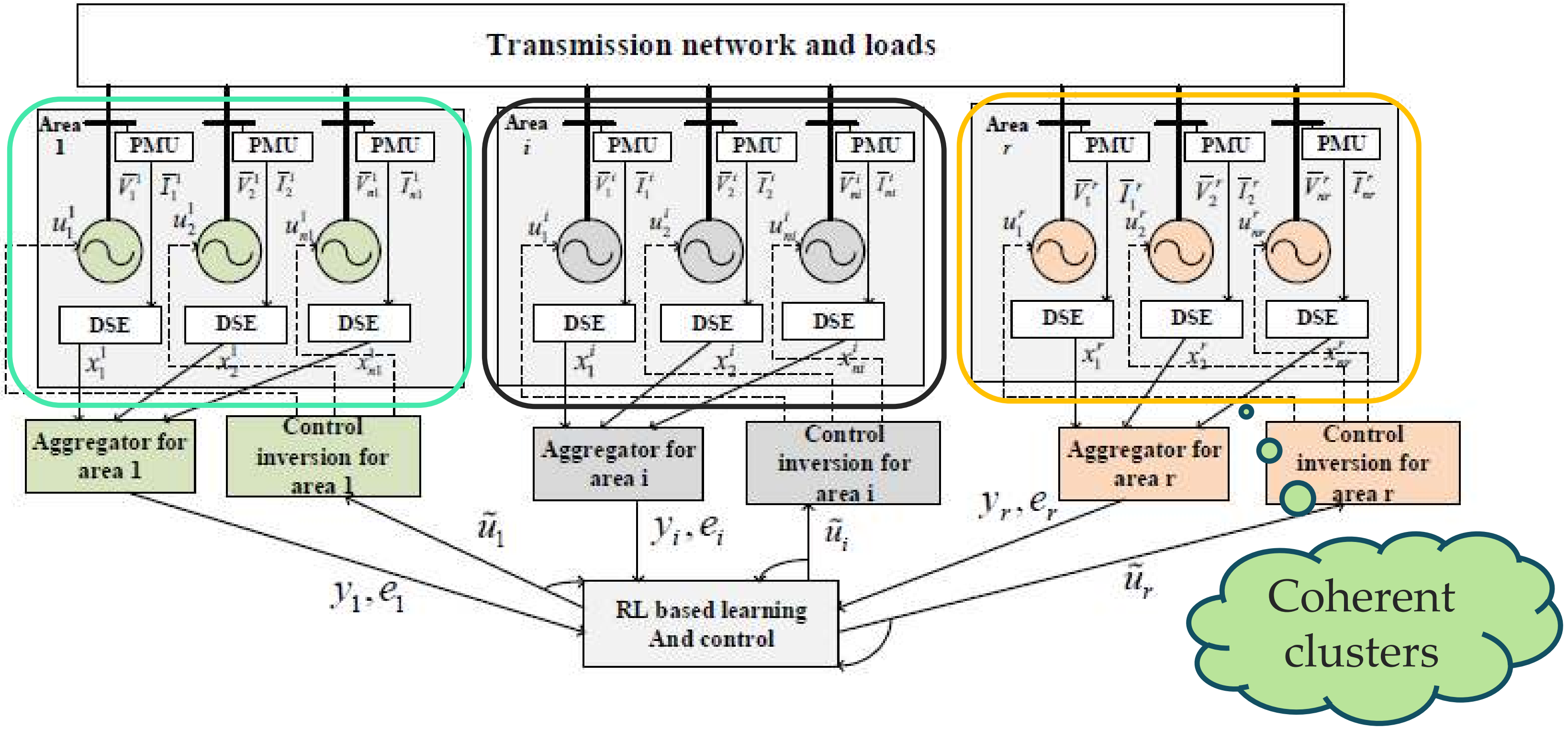
Bringing RL ideas to power systems – Direct SP-RL application

Practical large-scale power grid contains hundreds of dynamic components (say, n components)

Full-scale learning will take long time making it useless

Exploit the clustered nature (say, r clusters) to make considerable faster feedback gain computation

With r<< n, effectiveness increases



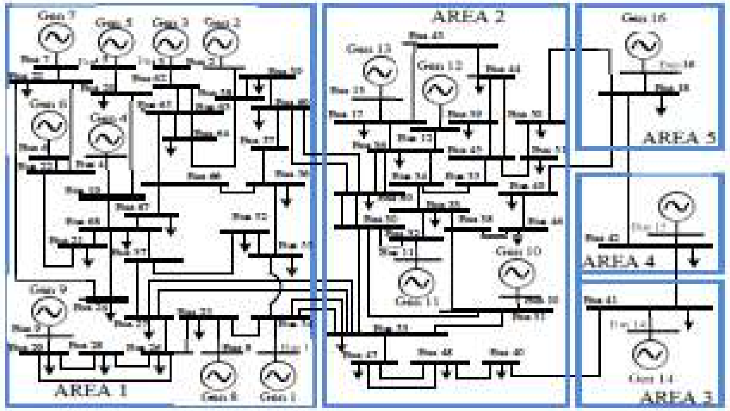
- The synchronous generator states can be estimated by *Decentralized State Estimators (DSEs)* using the voltages and currents of terminal buses obtained from PMUs
- The controls on the generators are actuated via excitation dynamics
- Along with *inertia weighted average of electromechanical states*, practical design will also require to consider *inertia weighted averaged excitation states*

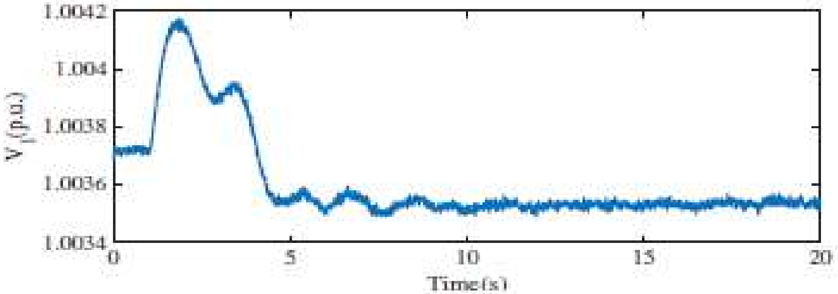
Key Aspects

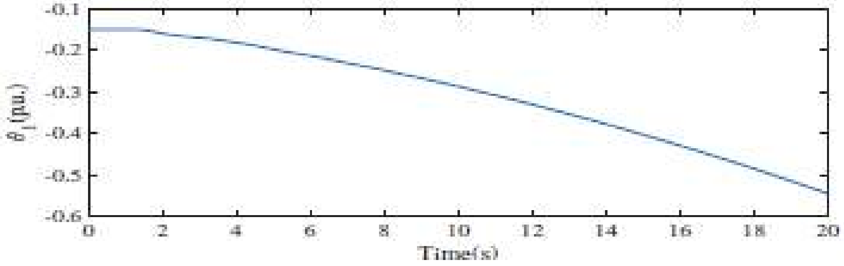
- Transmission model information is not required
- Generator states are estimated locally via DSEs
- A reduced-dimensional RL design that compensates the projection error
- Faster learning with reduced-dimensional feedback than full-dim. design

Simulations

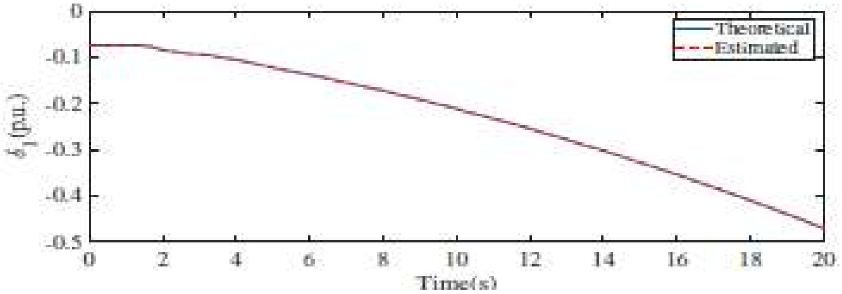
- Voltage and current measurements are generated by adding the white Gaussian noises

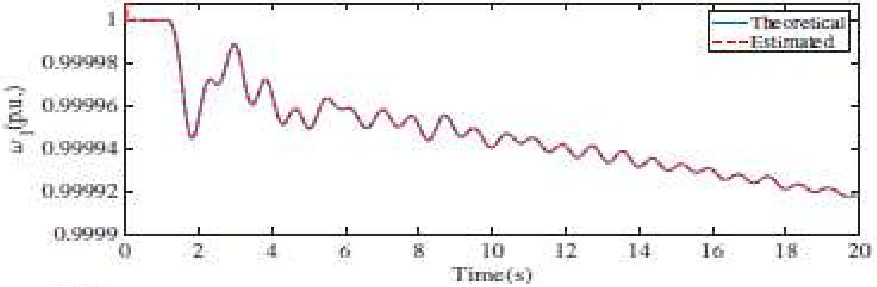


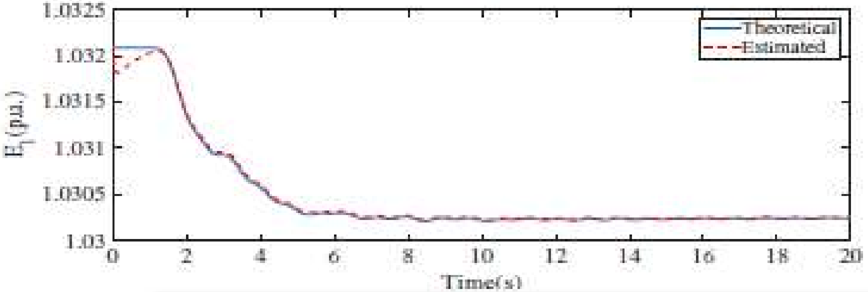




- The unscented Kalman filter is implemented in a decentralized fashion for the individual generating units showing high accuracy



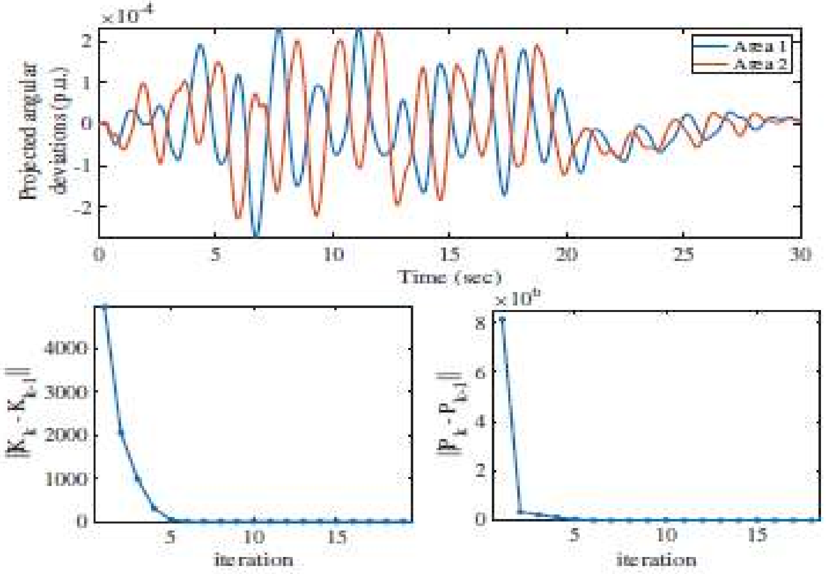




Performance of decentralized state estimation using UKF

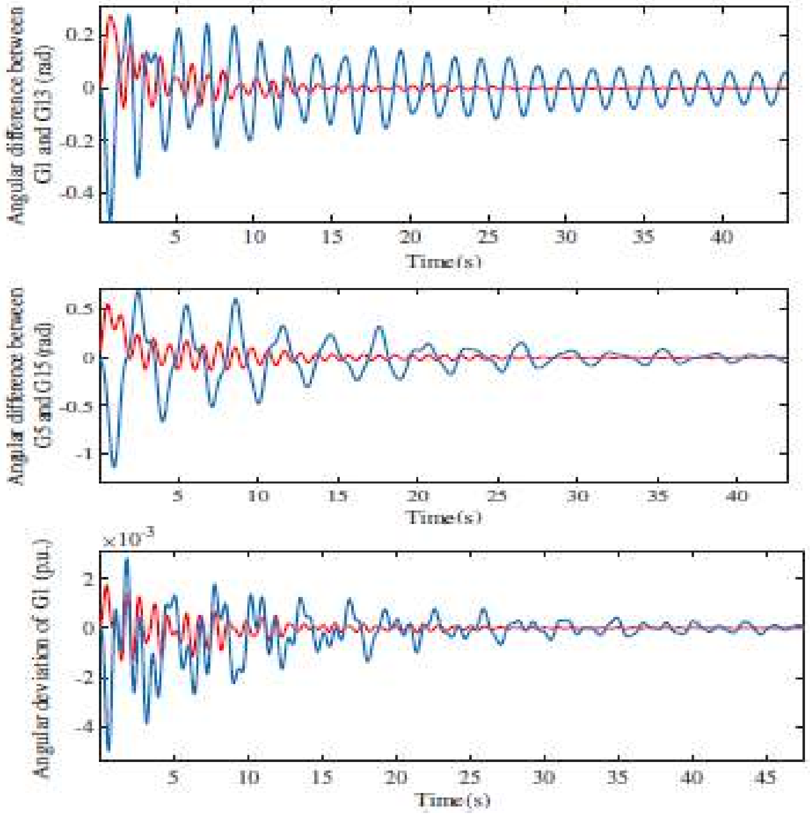
Simulations

- Here, $n = 64, m = 16$, and $r = 5 \rightarrow$ Full dim. design: rank condition of Θ_k would require at-least $n(n+1)/2 + nm = 3104$ time samples $\rightarrow$more than 31 s (T=0.01 s) of exploration
- However, the projected design needs at-least $4r(4r+1)/2 + 4rm + 4rp = 1810$ samples without control back projection (i.e., $m = 16$ and $M = I_{16}$), and 1590 samples with control back projection (i.e., $m = 5$)



Exploration and convergence

- The power grid model is excited with exploration signal with minimal perturbations $\rightarrow$ the supplementary actuation added at V_{ref} $\rightarrow$ improved performances on inter-area oscillations and angular frequencies



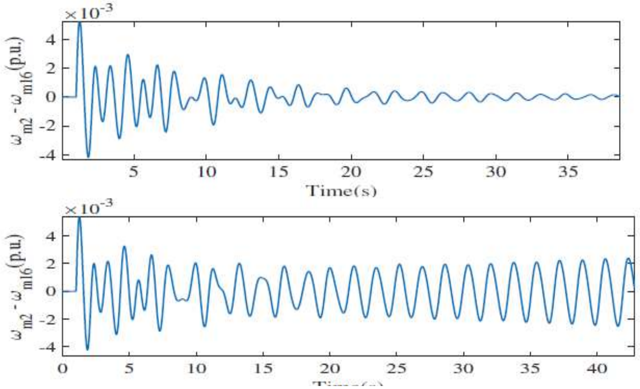
Performance improvement

FREEDM SYSTEMS CENTER

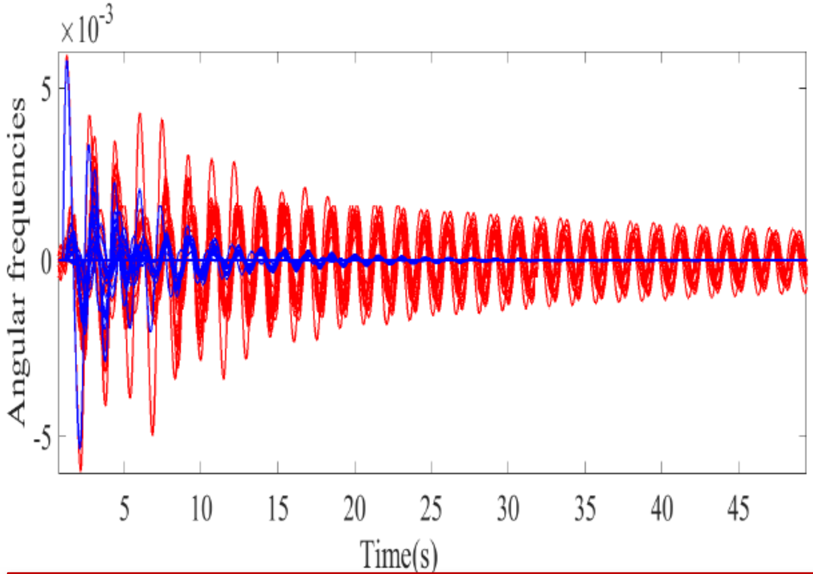
Decentralized RL on DERs

NC STATE UNIVERSITY

- Wind plants (DFIGs) are connected at buses 19 and 62

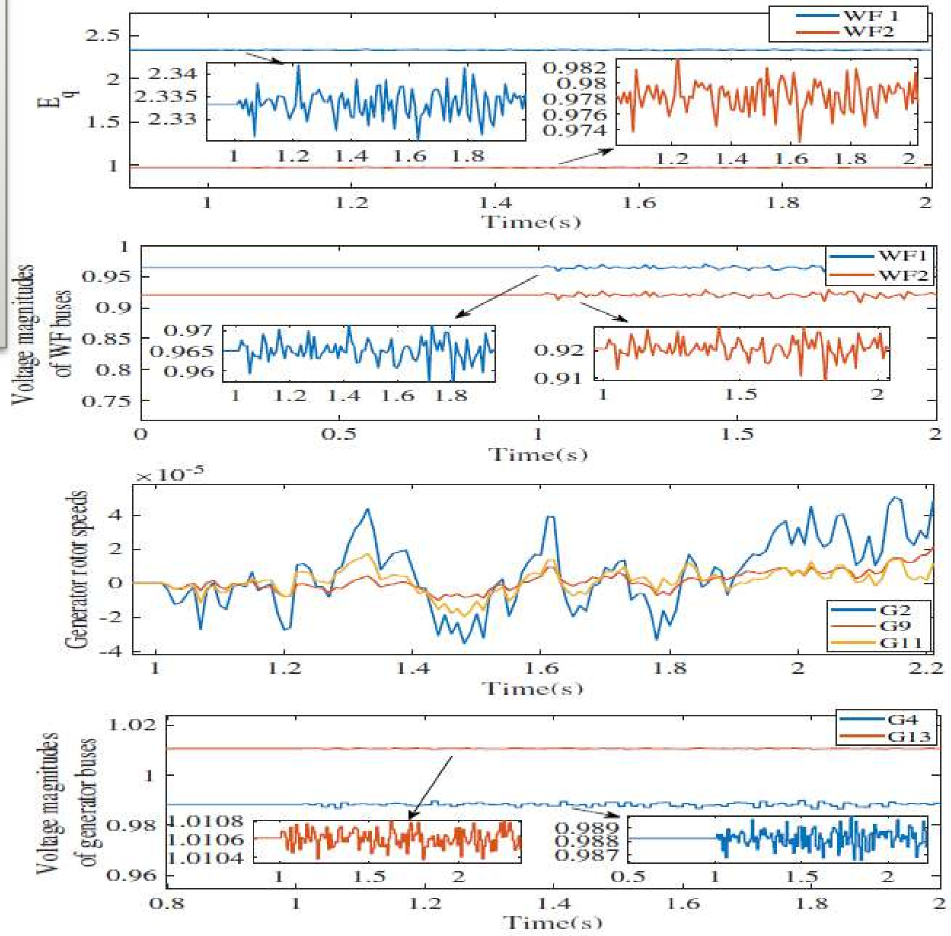


Wind penetration increased from (528,528) MW to (650,650) MW



Performance improvement using learned control

- RL design has been performed with (616, 616) MW penetration levels



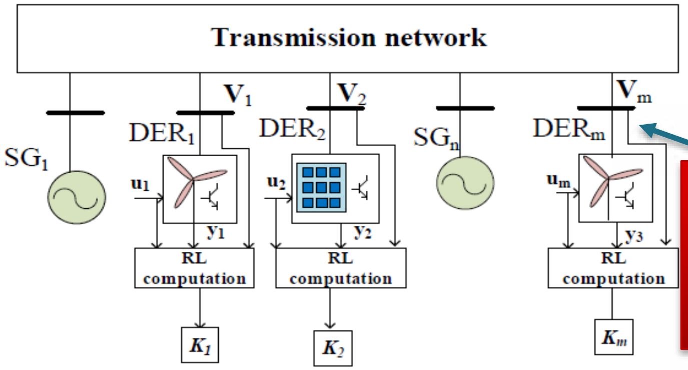
Exploration through DFIGs at (616, 616) MW case

Decentralized RL on DERs

Mukherjee, Bai, Chakrabortty (2020)

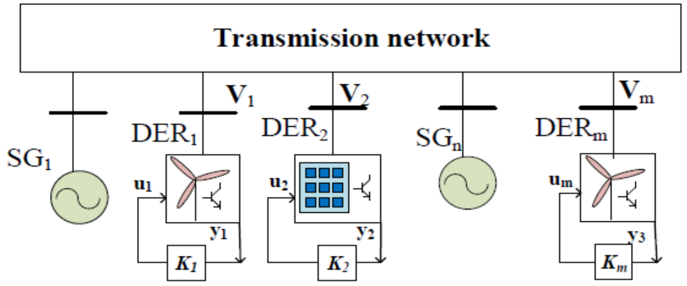
- Next generation power grid is envisioned to be equipped with distributed energy resources (DERs)
- Decentralized control on DERs helps in *scalable* design
- Operator doesn't need to know detailed DER models in RL based designs (*supplementary design with plug-and-play*)
- Localized learning *reduces exploration time and control dimension*

Learning Phase



Decentralization requires DER terminal bus voltages

Implementation Phase



Algorithm 1 Decentralized RL on DERs

For $i = n+1, \ldots, n+m$ **do**

Step 1- *Data storage: Store* data (y_i, V_{wi} and u_{i0}) for time intervals $(t_1, t_2, \cdots, t_l), t_i - t_{i-1} = T$ with time step T. The total exploration time is $T \times l$. There exists a sufficiently large number of sampling intervals such that $\text{rank}(I_{y_i y_i} \;\; I_{y_i u_{i0}} \;\; I_{y_i V_{wi}}) = d(d+1)/2 + dk + 2d$. Then *construct* the following matrices,

$$\delta_{y_i y_i} = \left[y_i \otimes y_i |_{t_1}^{t_1+T}, \quad \cdots \quad , y_i \otimes y_i |_{t_l}^{t_l+T} \right]^T, \tag{17}$$

$$I_{y_i y_i} = \left[\int_{t_1}^{t_1+T} (y_i \otimes y_i) d\tau, \quad \cdots \quad , \int_{t_l}^{t_l+T} (y_i \otimes y_i) d\tau \right]^T, \tag{18}$$

$$I_{y_i u_{i0}} = \left[\int_{t_1}^{t_1+T} (y_i \otimes u_{i0}) d\tau, \quad \cdots \quad , \int_{t_l}^{t_l+T} (y_i \otimes u_{i0}) d\tau \right]^T. \tag{19}$$

$$I_{y_i V_{wi}} = \left[\int_{t_1}^{t_1+T} (y_i \otimes V_{wi}) d\tau, \quad \cdots \quad , \int_{t_l}^{t_l+T} (y_i \otimes V_{wi}) d\tau \right]^T. \tag{20}$$

Step 2- *Controller update:* Starting with a stabilizing K_{i0}, *Solve* for K_i iteratively ($k = 0, 1, \cdots$) once matrices $\delta_{y_i y_i}, I_{y_i y_i}, I_{y_i u_{i0}}, I_{y_i V_{wi}}$ are constructed by the following iterative equation

$$\underbrace{\left[\delta_{y_i y_i} \quad -2I_{y_i y_i}(I_n \otimes \bar{K}_{ik}^T R_i) - 2I_{y_i u_{i0}}(I_n \otimes R_i) \quad -2I_{y_i V_{wi}} \right]}_{\Theta_{ik}} \times$$

$$\begin{bmatrix} vec(\bar{P}_{ik}) \\ vec(\bar{K}_{i(k+1)}) \\ vec(G_{3i}^T \bar{P}_{ik}) \end{bmatrix} = \underbrace{-I_{y_i y_i} vec(Q_{ik})}_{\Phi_{ik}}. \tag{21}$$

$\bar{P}_{ik}$ and $\bar{K}_{i(k+1)}$ are iteratively solved such that $|\bar{P}_{ik} - \bar{P}_{i(k-1)}| < \bar{\epsilon}$, $\bar{\epsilon}$ is a small threshold. This process is performed in a very fast computing platform when the system is under control $u = u_{i0}$.

Step 3- *Applying control on the system :* Next $u_i = -K_i y_i$ is applied and u_{i0} source is removed.

End For

Concluding Remarks

- Measurement-driven designs are discussed using multiple approaches
- RL-based designs are proposed to solve the curse of dimensionality associated with the learning control design using singular perturbation approximations
- The proposed methodology enjoys faster learning and reduction in feedback control dimensionality
- Sub-optimality and stability analyses are performed using the singular perturbation approximations
- Designs can be extended to control oscillations in power systems considering the projection error in the dimensionality reduction
- Decentralized design on DERs enjoys scalability and modularity along with fast learning

Moving forward

- Towards a data-driven decision making for analysis and control of next generation power grid
- Advances in the computing capabilities involving machine learning and data-driven control algorithms and the deployment of high resolution measurement devices in power grid make this union possible
- Inter-disciplinary research areas :

  Core subject area : Power system operations, Power Systems Dynamics, Renewable-integrated power grid etc.

  Computing and analytical tools : Convex Optimization, Control theory, Statistical learning with ML/DL , Reinforcement learning, Cyber-physical system security etc.
- Research towards an autonomous energy grid where mathematical tools from controls, optimization, data analytics, etc. all can work together for solving energy problems of the future

References for more details

- S. Mukherjee, Ph.D. Dissertation, Data-Driven Reinforcement Learning Control using Model Reduction Techniques: Theory and Applications to Power Systems, 2020, available via NC State Libraries, https://www.lib.ncsu.edu/resolver/1840.20/37368.
- S. Mukherjee, H. Bai, and A. Chakrabortty. On Model-Free Reinforcement Learning of Reduced-order Optimal Control for Singularly Perturbed Systems. *59th IEEE Conference on Decision and Control*, FL, Dec. 2018.
- S. Mukherjee, S. Babaei, A. Chakrabortty, and B. Fardanesh, “Measurement-Driven Optimal Control of Utility-Scale Power Systems: A New York State Grid Perspective,” *International Journal of Electrical Power and Energy Systems*, *Elsevier*, 2020.
- S. Mukherjee, H. Bai, A. Chakrabortty, “Reduced-Dimensional Reinforcement Learning Control using Singular Perturbation Approximations”, Submitted to Automatica, 2019, available on arxiv.
- S. Mukherjee, A. Chakrabortty, and H. Bai. Block-Decentralized Model-Free Reinforcement Learning Control of Two Time-Scale Networks. *American Control Conference*, Philadelphia, 2019.
- S. Mukherjee, H. Bai, A. Chakrabortty, Model-Free Decentralized Reinforcement Learning Control of Distributed Energy Resources, accepted in IEEE PES General Meeting, Montreal, Canada, 2020.

Thank you!

Q & A

Contacts: smukher8@ncsu.edu, achakra2@ncsu.edu

More on FREEDM: **https://www.freedm.ncsu.edu/**