Data-Driven Power System Frequency Control

数据驱动的电力系统频率控制

Dr Yan Xu | Nanyang Assistant Professor
School of Electrical & Electronic Engineering
Nanyang Technological University (NTU)
Email: xuyan@ntu.edu.sg
Web: https://eexuyan.github.io/soda/index.html

1. Background

2. Methodology

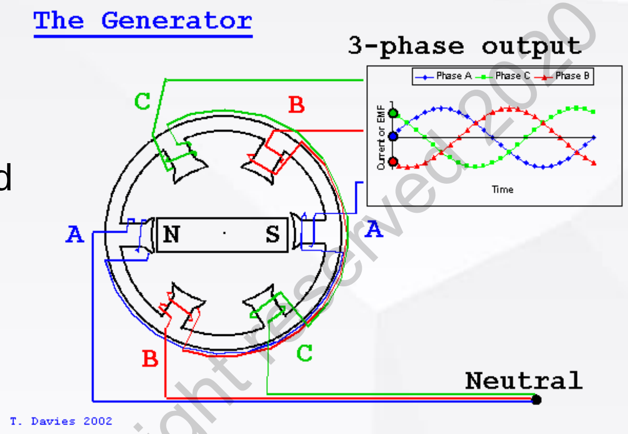
3. Single area

4. Multi-area systems

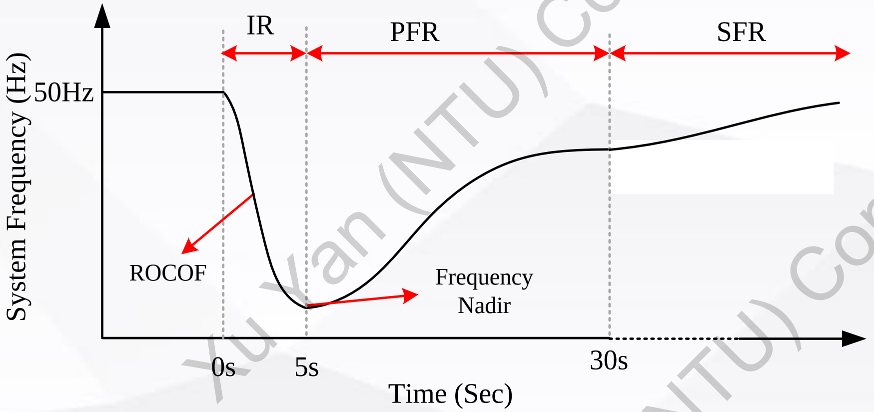
5. Optimal BESS control

Power System Frequency

- **Frequency**
  - AC power system
  - Reflection of rotation speed of synchronous generators
- **Importance**
  - Grid: system stability
  - Consumers: power quality



Source of pictures: website (searched in Google)



- **Inertia Response (IR):** the inherent releasing of energy at the rotor of synchronous machines.
- **Primary control:** mitigate frequency variation (seconds)
- **Secondary control:** eliminate frequency deviation (seconds to minutes)

Country/Region	Australia	Europe	North America	Singapore
Nominal frequencies (Hz)	50	50	60	50
Normal operating frequency bands (Hz)	Interconnected system: ±0.15 Islanded system: ±0.5	±0.2	Targeted frequency band: Eastern Interconnection: ±0.018 Western Interconnection: ±0.0228 Texas Interconnection: ±0.030 Quebec Interconnection: ±0.021	±0.2
Emergency frequency tolerance bands (Hz)	±1 Extreme frequency tolerance band: 47–52	±0.8	Under-frequency load shedding: Eastern Interconnection: 59.5 Western Interconnection: 59.5 Texas Interconnection: 59.3 Quebec Interconnection: 58.5	Under-frequency load shedding: 49.7

1. Background

2. Methodology

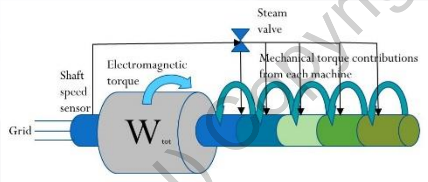
3. Single area

4. Multi-area systems

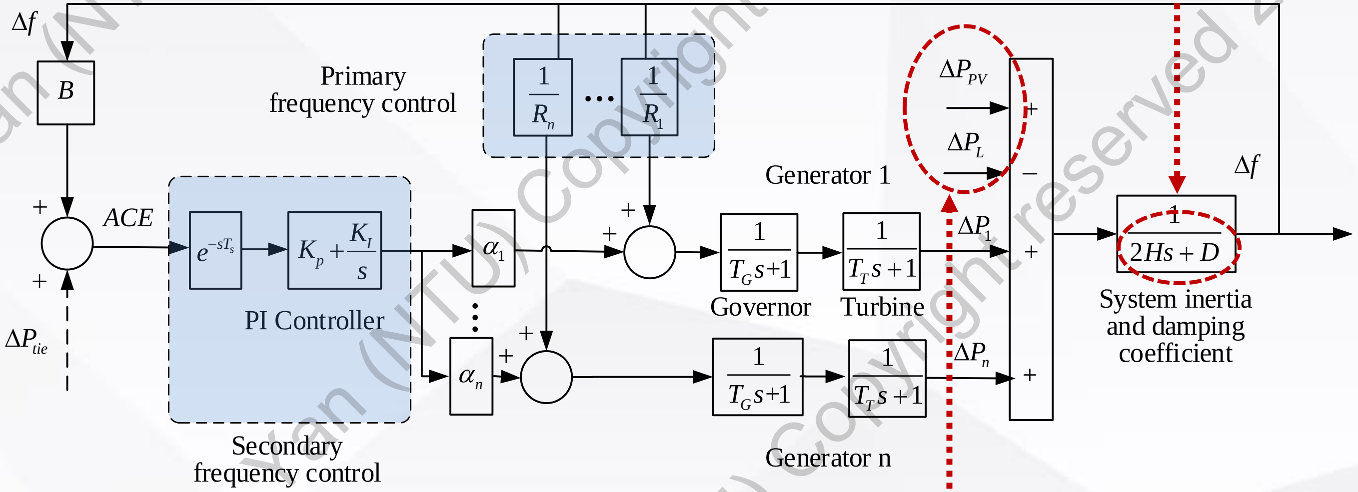
5. Optimal BESS control

Load frequency control (LFC)

➤ **Lower inertia and load damping:**

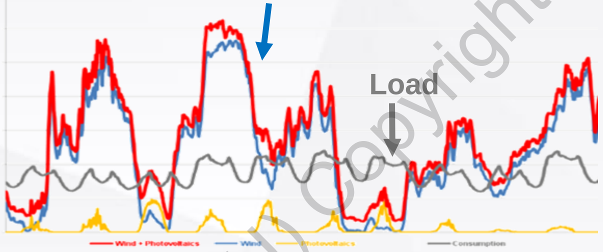


- Generation side: power-converter interfaced generators (wind, solar).
- Transmission side: asynchronous interconnection through HVDC links.
- Load side: inverter-based loads.



➤ **Larger and faster power fluctuations:**

- Generation side: intermittent renewable power generation
- Load side: demand response program, EV charging load, etc.

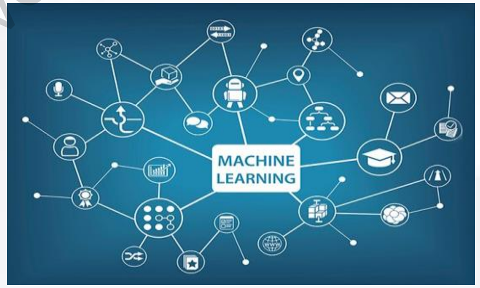


Conventional methods

Model-based:

1. Robust control
Parametric uncertainties.
2. Fuzzy control
Adaptive for unknown system.
3. Variable structure control
Robustness and response speed.
4. Disturbance rejection control
Augmented model to reject effects.
5. Model-predictive control
Predict system's behavior and control.
6. etc.

Data-driven methods



- Stronger modelling capability
- Better control performance
- Higher flexibility and scalability
- etc.

Source of pictures: website (searched in Google)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background

2. Methodology

3. Single area

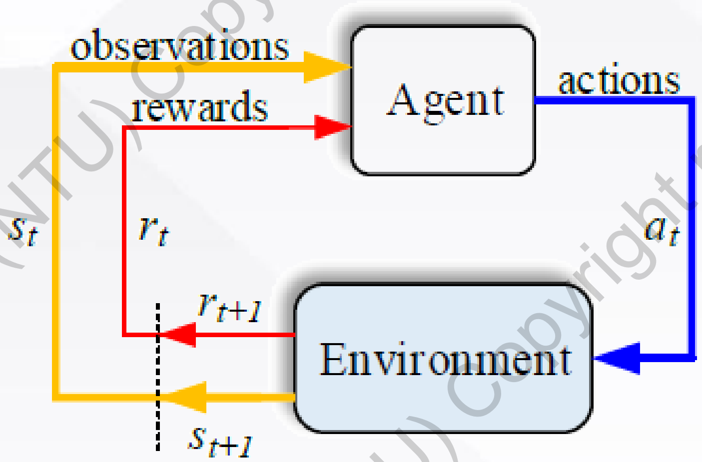
4. Multi-area systems

5. Optimal BESS control

Reinforcement Learning (RL)

Principle & Framework

- **Principle**: training an **agent** via iterative interactions with the **environment**.



- **Agent**: decision-maker → frequency controller
- **Environment**: physical world → power system
- **State (s)**: current situation of the agent → f, ACE, P
- **Action (a)**: agent's decision → generation control signal
- **Reward (r)**: feedback from the environment → power system's frequency performance (at time t)
- **Action value (Q-value)**: total expected reward over a certain time period T

➢ How to **model** the frequency control problem into a RL process?

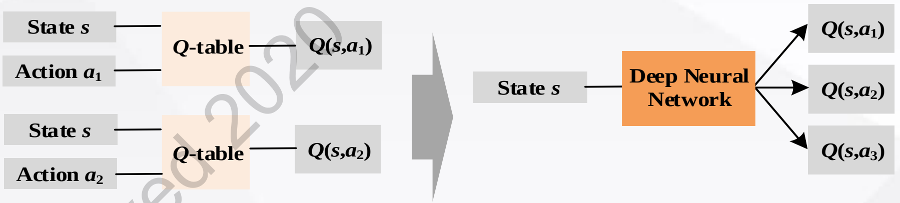
➢ How to **solve** the RL training process considering power system's own characteristics/model?

RL methods

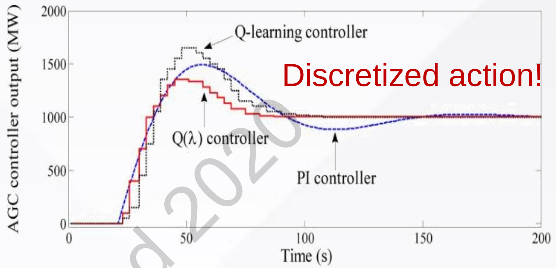
1. Value-based methods – train a Q-value predictor (Q-table)

Given an action, it evaluates the how good the action is.

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha(R_{t+1} + \gamma \max_{\alpha} Q(s_{t+1},a_t) - Q(s_t,a_t))$$



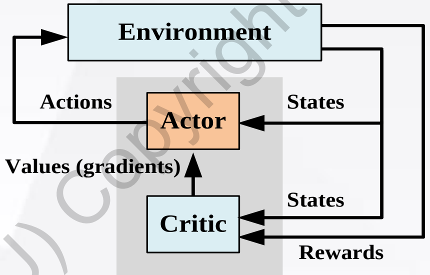
Disadvantages:
- Discretized action.
- Non-satisfactory performance due to discretized action space.



2. Policy-based methods – train an action predictor (actor)

Explicitly learn a mapping policy $\pi: s \rightarrow a$

$$\nabla_{\theta^u} J \approx \frac{1}{N}\sum_i \nabla_a Q(s,a)|_{s=s_i,a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s=s_i}$$



Advantages:
- Continuous action space.
- Better performance in convergence and stability.

1. Background

2. Methodology

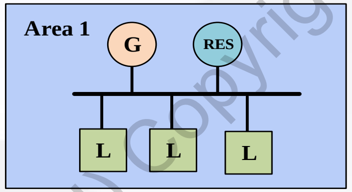
3. Single area

4. Multi-area systems

5. Optimal BESS control

Our research works

Single-area controller [1]

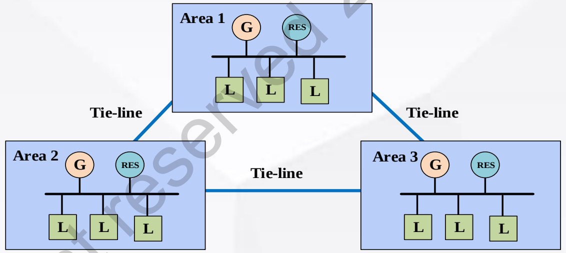


G: generation; L: load;
RES: renewable energy resources;
BESS: battery energy storage system

- Developed a policy-based DRL model for single-area power system frequency control
- Minimize expected frequency deviations
- Model-assisted gradients derivation
- Stacked denoising auto-encoder (SDAE) for feature learning

[1] Z. Yan, Y. Xu, "Data-Driven Load Frequency Control for Stochastic Power Systems: A Deep Reinforcement Learning Method With Continuous Action Search," ***IEEE Trans. Power Systems*, 2019.**

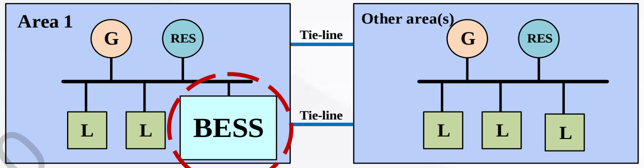
Multi-area controllers [2]



- Developed a set of cooperative DRL models for multi-area power system
- Centralized learning, decentralized implementation
- Optimize global action-value function
- Constraints-aware gradients derivation
- Network initialization to quick start

[2] Z. Yan, Y. Xu, "A Multi-Agent Deep Reinforcement Learning Method for Cooperative Load Frequency Control of Multi-Area Power Systems," ***IEEE Trans. Power Systems*, 2020.**

BESS controller for frequency support [3]



- Optimal control of BESS for f support
- Minimize expected total control cost considering the degradation of battery
- Modelling of BESS lifetime degradation
- Actor-critic framework
- Cost approximation with critic

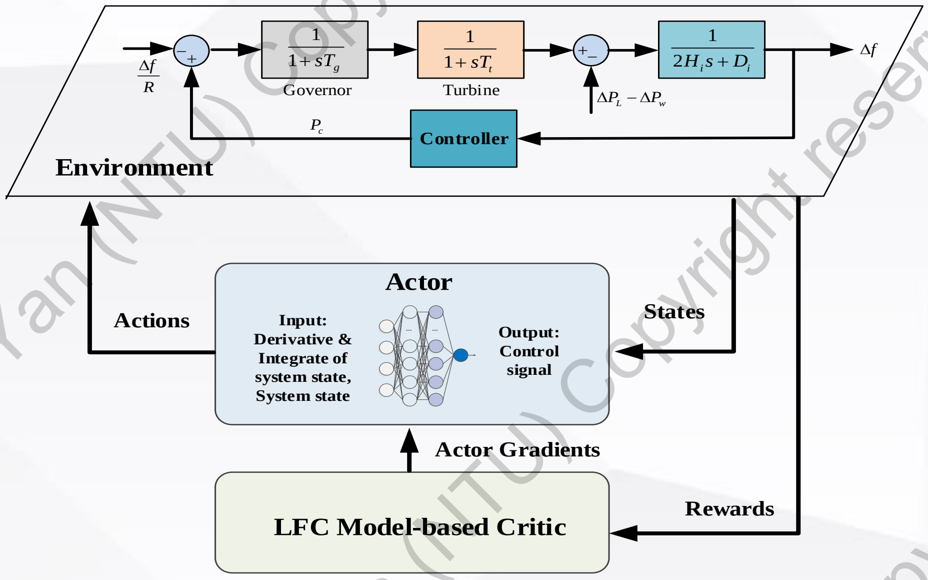
[3] Z. Yan, Y. Xu, et al, "Data-driven Economic Control of Battery Energy Storage System Considering Battery Degradation," ***IET Generation. Transmission & Distribution*, 2020.**

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Single-area LFC controller

- **Principle**

Optimize the parameters $\theta = [\mathrm{W}^{\mathrm{T}}, b]$ of DRL agent based on data, such that the control policy is optimized and expected frequency deviations are minimized.



Deep reinforcement learning process

Agents-Environment Interaction

- Action-value function:

$$\underset{\theta}{\text{Maximize}}\ E_D[Q^{\mu}(s_t, a_t)] = E_D[-(\sum_{i=1}^{N} \Delta t_i \Delta f_i^2)]$$

- Training process

DNN's parameters

$$\theta^{\mu} \leftarrow \theta^{\mu} + \eta \cdot \nabla_{\theta^{\mu}} J$$

$$\nabla_{\theta_i^{(k)}} J \approx E_D[\nabla_{\theta^{\mu}} \mu(s \mid \theta^{\mu}) \nabla_a Q(s,a)]$$

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{N} \sum_i \nabla_a Q(s,a)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^{\mu}} \mu(s \mid \theta^{\mu})|_{s=s_i}$$

Model-assisted gradient derivation

The gradient of expected action-value with respect to control action

$$\nabla_a Q^{\mu}(s_t, a_t) \approx -2\Delta t \Delta f(s_t, a_t)(R - k(\frac{\Delta f(s_{t+\Delta t}, a_{t+\Delta t}) - \Delta f(s_t, a_t)}{\Delta t}))$$

DNN Updating rule

The gradient of action with respect to agent' parameters

$$\nabla_{\theta^{\mu}} \mu(s \mid \theta^{\mu}) = \nabla_{\theta^{\mu}} (f_{\theta}^{(n)}[\ldots f_{\theta}^{(1)}(\boldsymbol{X})])|_{\boldsymbol{X} \text{ is input vector with } s=s_i}$$

[1] Z. Yan, Y. Xu, "Data-Driven Load Frequency Control for Stochastic Power Systems: A Deep Reinforcement Learning Method With Continuous Action Search," ***IEEE Trans. Power Systems*****, 2019.**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Single-area LFC controller

- Model-based gradient derivation process

Model-assisted gradient derivation

1. $$\nabla_a Q^{\mu}(s_t,a_t) = -2\Delta t \Delta f(s_t,a_t)\frac{\partial \Delta f(s_t,a_t)}{\partial a}$$

2. $$\begin{cases} a(t) = b_3\dfrac{d^3 f(t)}{dt^3} + b_2\dfrac{d^2 f(t)}{dt^2} + b_1\dfrac{df(t)}{dt} + b_0\Delta f(t) \\ b_0 = 1/R,\ b_1 = 2HT_gT_t[2H+(T_g+T_t)D]/D, \\ b_2 = 2HT_gT_t[T_gT_tD+2HT_g+2HT_t]/D,\ b_3 = 2HT_gT_t \\ \nabla_a f(t) = \dfrac{1}{b_0}(-b_3\nabla_a\dfrac{d^3 f(t)}{dt^3} - b_2\nabla_a\dfrac{d^2 f(t)}{dt^2} - b_1\nabla_a\dfrac{df(t)}{dt} + 1) \\ \nabla_a f(t) \approx R - k\dfrac{df(t)}{dt} \end{cases}$$

Modifying DDPG

3. $$\nabla_a Q^{\mu}(s_t,a_t) \approx -2\Delta t\Delta f(s_t,a_t)(R - k(\frac{\Delta f(s_{t+\Delta t},a_{t+\Delta t}) - \Delta f(s_t,a_t)}{\Delta t}))$$

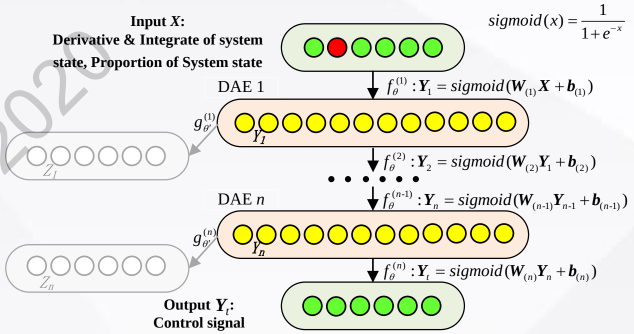
4. $$\nabla_{\theta^{\mu}}\mu(s\,|\,\theta^{\mu}) = \nabla_{\theta^{\mu}}(f_{\theta}^{(n)}[\ldots f_{\theta}^{(1)}(\boldsymbol{X}))])\,|_{\boldsymbol{X}\text{ is input vector with } s=s_i}$$

Improved agent updating rule

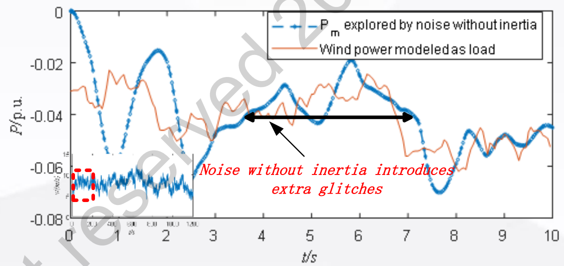
5. $$\begin{cases} W_{ij}^{(l,T+1)} = W_{ij}^{(l,T)} - \eta\dfrac{1}{m}\sum_{k=r}^{r+m}\nabla_a Q^{\mu}(s_t,a_t)\dfrac{\partial}{\partial W_{ij}^{(l,T)}}a(\boldsymbol{W},\boldsymbol{b}) \\ b_i^{(l,T)} = b_i^{(l,T)} - \eta\dfrac{1}{m}\sum_{k=r}^{r+m}\nabla_a Q^{\mu}(s_t,a_t)\dfrac{\partial}{\partial b_i^{(l,T)}}a(\boldsymbol{W},\boldsymbol{b}) \end{cases}$$

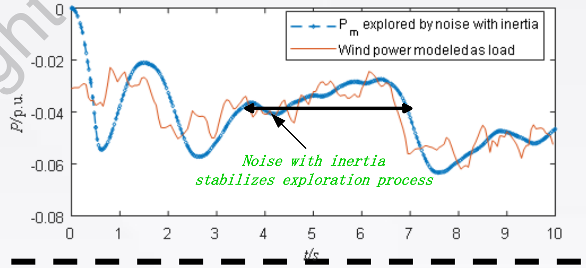
Tricks to improve performance

Stacked denoising auto-encoders:
Initialize the DRL agent by SDAE (supervised learning with data generated by PID controller), a deep learning tool widely used for feature extraction.



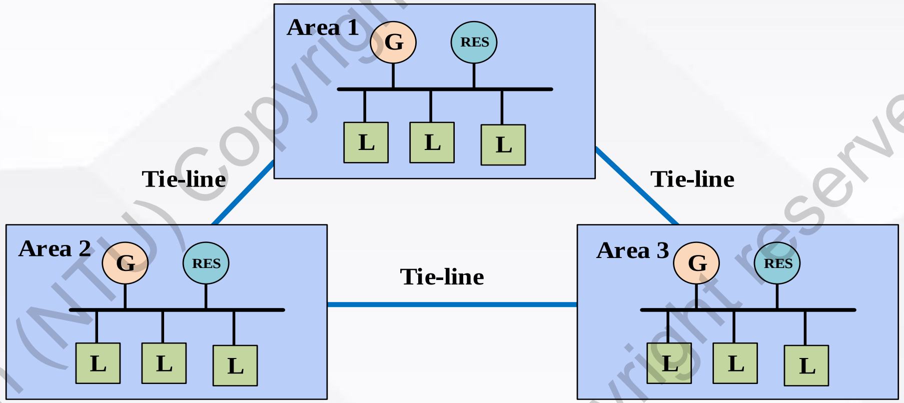
Auto-correlated exploration noise:
Stabilize the exploration process with moving average.





Multi-area LFC controller

Multi-area power system

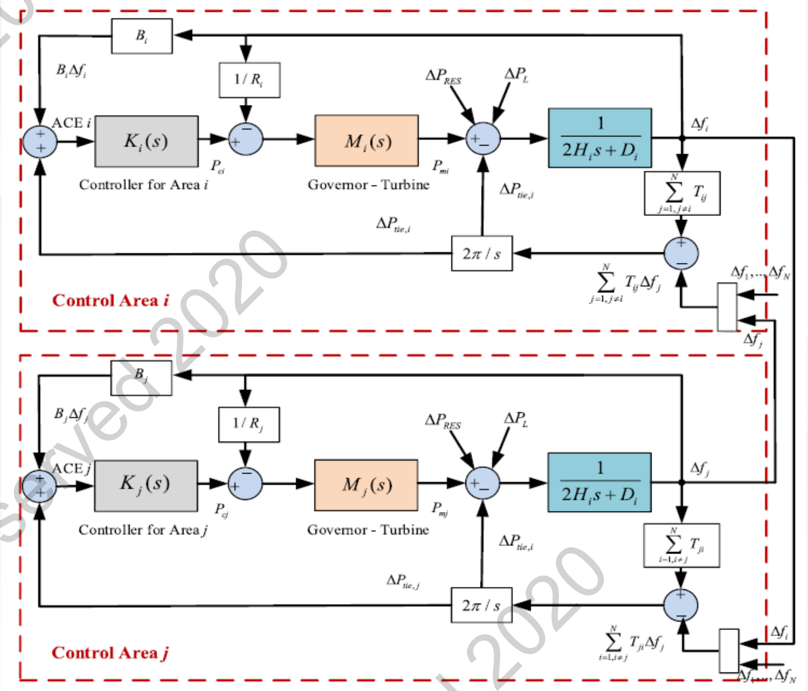


G: generation **L**: load

RES: renewable energy sources

Each area has its own control agent.

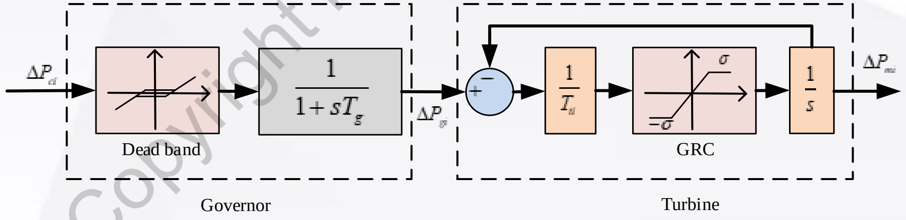
Multi-area LFC block diagram



Problem descriptions

- Intermittent RES: complex **cross-area power balancing** between generation and demand.
- Cooperative control: how to **coordinate** the multiple controllers in all areas.
- Constraints: how to consider nonlinear **physical limits** while optimizing the controllers.

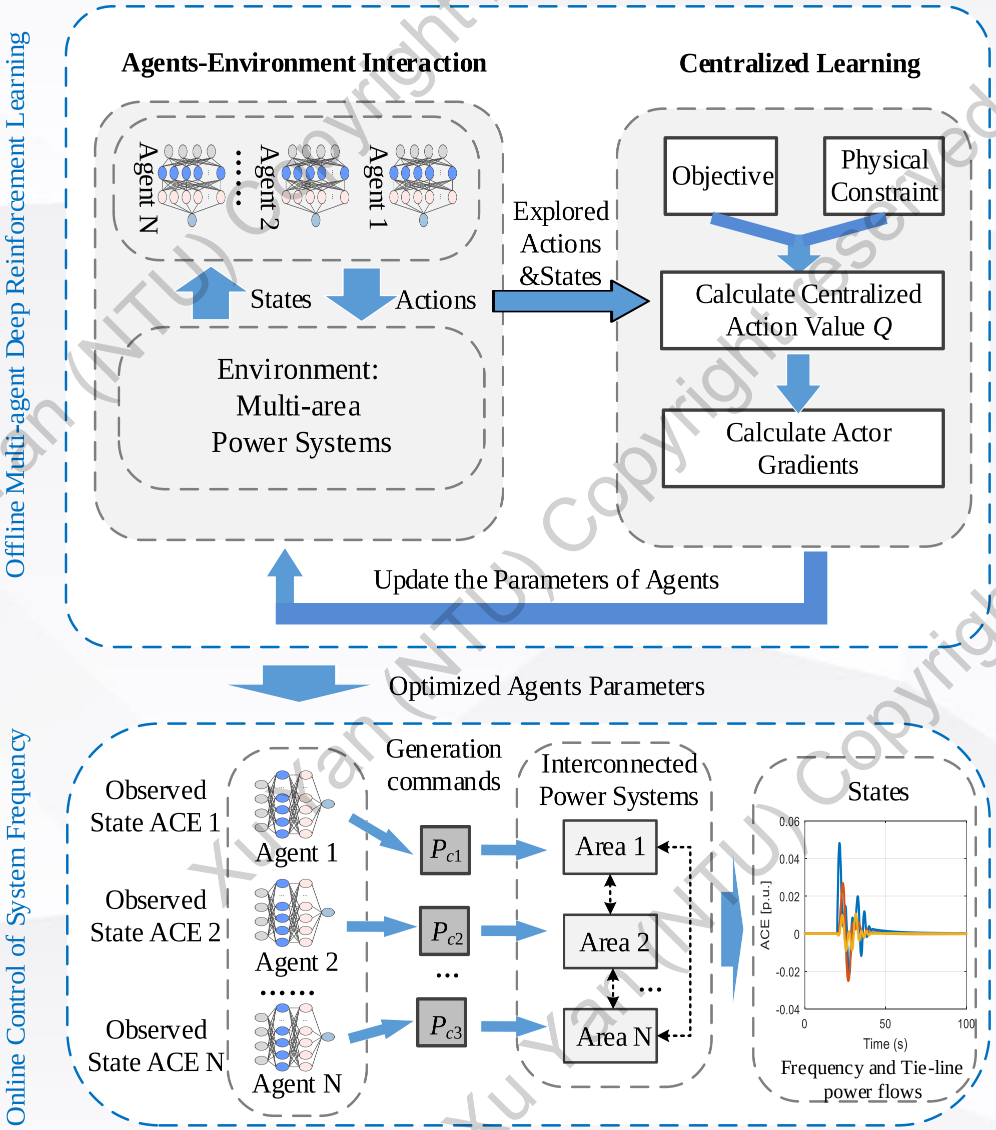
Nonlinear parts



generation dead band (GDB) and generation rate constraints (GRC)

Multi-area LFC controller

Centralized training and decentralized implementation



Agents-Environment Interaction

- Global expected action-value:

$$Q^{\mu}(s,a_1,a_2,...,a_n) = -\sum_{t=1}^{T}[\Delta t\sum_{i=1}^{n}[(B_i\Delta f_i)^2+(\Delta P_{tie,i})^2]]$$

$$\underset{\theta_1,\theta_2,...,\theta_N}{\text{Maximize}}\ E_D[Q^{\mu}(s,a_1,a_2,...,a_n)]$$

- Training process for **each agent**:

$$\theta_i^{(k+1)} = \theta_i^{(k)} + \eta\nabla_{\theta_i^{(k)}} J$$

$$\nabla_{\theta_i^{(k)}} J \approx E_D[\nabla_{\theta_i^{(k)}}\mu_i^{(k)}(o_i)\nabla_{a_i}Q^{\mu_i}(s,a_1,a_2,...,a_n)]$$

$$\nabla_{\theta_i^{(k)}} J \approx \frac{1}{m}\sum_i \nabla_{a_i}Q^{\mu_i}(s,a_1,a_2,...,a_n)\nabla_{\theta_i^{(k)}}\mu_i^{(k)}(o_i)$$

Model-assisted gradient derivation

Gradient of global objective to each action

$$\nabla_{a_i}Q^{\mu_i}(s,a_1,a_2,...,a_n) \approx -2B_i\Delta f_i(R_i-\kappa\frac{d\Delta f_i}{dt}) -4\pi\Delta P_{tie,i}\sum_{j\neq i}^{N}T_{ij}(R_i-\kappa\frac{d\Delta f_i}{dt})$$

DNN Updating rule

Gradient of the action to each agent' parameters

$$\nabla_{\theta^{\mu}}\mu(s|\theta^{\mu}) = \nabla_{\theta^{\mu}}(f_{\theta}^{(n)}[...f_{\theta}^{(1)}(X)])\Big|_{X \text{ is input vector with } s=s_i}$$

[2] Z. Yan, Y. Xu, "A Multi-Agent Deep Reinforcement Learning Method for Cooperative Load Frequency Control of Multi-Area Power Systems," *IEEE Trans. Power Systems*, 2020.

Multi-area LFC controller

Gradients for all actors (MA-DDPG)

1.
$$Q^{\mu}(s,a_1,a_2,...,a_n) = -\sum_{t=1}^{T}[\Delta t\sum_{i=1}^{n}[(B_i\Delta f_i)^2+(\Delta P_{tie,i})^2]]$$
$$\theta_i^{(k+1)} = \theta_i^{(k)} + \eta\nabla_{\theta_i^{(k)}} J$$
$$\nabla_{\theta_i^{(k)}} J \approx \frac{1}{m}\sum_i \nabla_{\theta_i^{(k)}}\mu_i^{(k)}(o_i)\nabla_{a_i}Q^{\mu_i}(s,a_1,a_2,...,a_n)$$

Expand

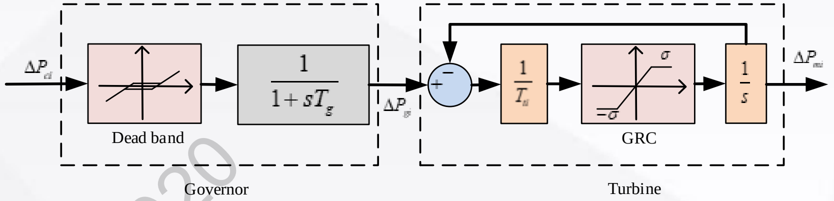
2.
$$\frac{\partial Q^{\mu_i}}{\partial a_i} \approx -2B_i\Delta f_i\frac{\partial\Delta f_i}{\partial a_i} - 2\Delta P_{tie,i}\frac{\partial \Delta P_{tie,i}}{\partial a_i} - \sum_{j\neq i}^{n}[2\Delta P_{tie,j}\frac{\partial\Delta P_{tie,j}}{\partial a_i}]$$
$$\frac{\partial \Delta P_{tie,i}}{\partial a_i}\approx 2\pi[\sum_{j\neq i}^{N}T_{ij}\frac{\partial\Delta f_i}{\partial a_i} - \sum_{j\neq i}^{N}T_{ij}\frac{\partial \Delta f_j}{\partial a_i}]$$
$$\frac{\partial\Delta P_{tie,j}}{\partial a_i}\approx 2\pi[\sum_{k\neq j}^{N}T_{jk}\frac{\partial\Delta f_j}{\partial a_i} - \sum_{k\neq j}^{N} T_{jk}\frac{\partial\Delta f_k}{\partial a_i}]$$

Model-assisted gradient approximation

3.
$$\nabla_a\Delta f(t) = \frac{1}{\beta_0}(1-\beta_1\frac{d\Delta f(t)}{dt}) - \frac{\beta_2}{\beta_0}\nabla_a\frac{d^2\Delta f(t)}{dt^2} - \frac{\beta_3}{\beta_0}\nabla_a\frac{d^3\Delta f(t)}{dt^3}$$
$$\beta_0 = 1/R,\ \beta_1 = 2HT_gT_t[2H+(T_g+T_t)D]/D,$$
$$\beta_2 = 2HT_gT_t[T_gT_tD+2HT_g+2HT_t]/D,\ \beta_3 = 2HT_gT_t$$
$$\frac{\partial Q^{\mu_i}}{\partial a_i}\approx -2B_i\Delta f_i(R_i-\kappa\frac{d\Delta f_i}{dt}) - 4\pi\Delta P_{tie,i}\sum_{j\neq i}^{N}T_{ij}(R_i-\kappa\frac{d\Delta f_i}{dt})$$

Considering generation rate constraints (GRC)

4.
$$|dP'_m(t)/dt| > \sigma$$
$$P_m(t-\Delta t)\pm\sigma\Delta t = 2H\frac{d\Delta f(t)}{dt}+D\Delta f(t)$$
$$\nabla_a(\Delta f(t)) + (2H/D)\nabla_a(d\Delta f(t)/dt) = 0$$
$$\partial Q/\partial a_i \approx 0 \quad \begin{cases} if\ \partial Q/\partial a_i>0,\ da_i/dt>\sigma T_{ti} \\ if\ \partial Q/\partial a_i<0,\ da_i/dt<-\sigma T_{ti} \end{cases}$$

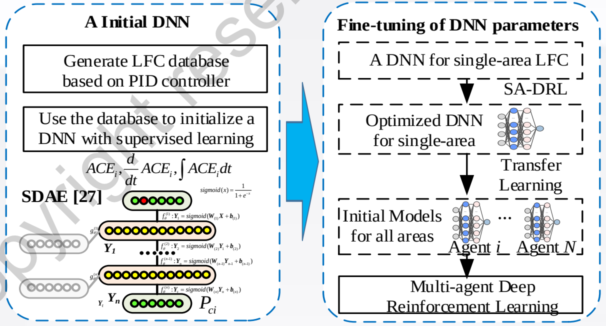


Agent updating rule considering physical limits

5.
$$W_{ij}^{(l,T+1)} = W_{ij}^{(l,T)} - \eta\frac{1}{m}\sum_{k=r}^{r+m}\nabla_a Q^{\mu}(s_t,a_t)\frac{\partial}{\partial W_{ij}^{(l,T)}}a(\boldsymbol{W},\boldsymbol{b})$$
$$b_i^{(l,T)} = b_i^{(l,T)} - \eta\frac{1}{m}\sum_{k=r}^{r+m}\nabla_aQ^{\mu}(s_t,a_t)\frac{\partial}{\partial b_i^{(l,T)}}a(\boldsymbol{W},\boldsymbol{b})$$

Tricks to improve performance

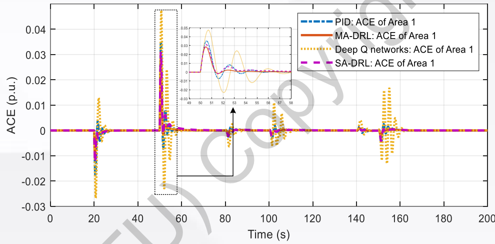
Initialization:
Initialize the DRL agent by supervised learning (data generated by PID controller), then further improved with reinforcement learning.

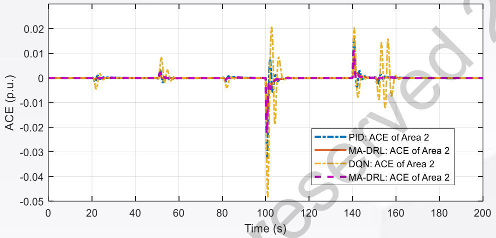


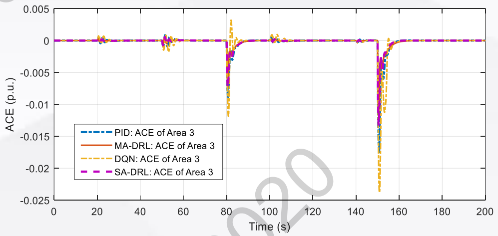
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Testing results (LFC model)

- Linearized LFC model (no physical limits):

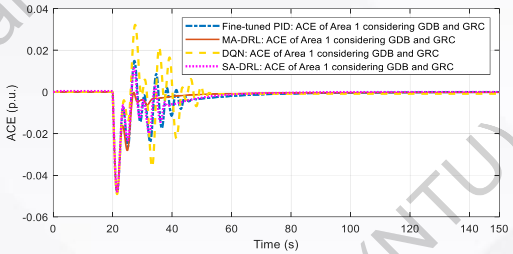


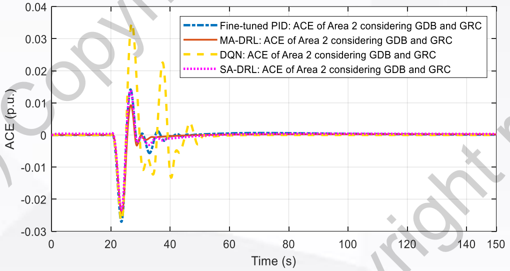


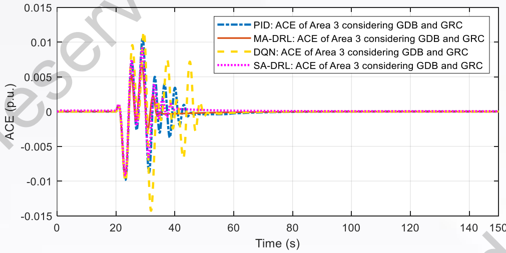


- Less expected frequency deviations: 87.7% better than DQN, 57.5% better than PID.
- Smaller frequency nadir: 39.6% better than DQN, 17.1% better than PID.

- Nonlinearity (GRC&GDB):

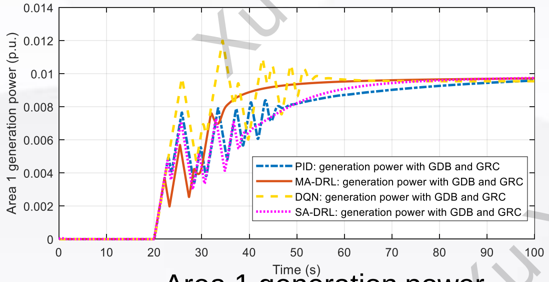




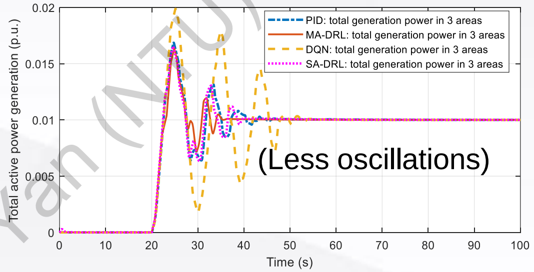


- Less deviations: 62.5% better than DQN, 22.2% better than PID.
- Improves the LFC performance by better coordination among all the areas

- Generation power under GRC&GDB:



Area 1 generation power



Total generation in 3 areas

Method	Q	Mean |ACE| %	Max |ACE| [p.u.]
Fine-tuned PID	-0.0247	0.037	0.035
(Deep) Q-learning	-0.0851	0.093	0.048
Proposed method	-0.0105	0.023	0.029
Fine-tuned PID (GRC and GDB)	-1.8e-3	0.042	0.049
(Deep) Q-learning (GRC and GDB)	-3.2e-3	0.061	0.049
Proposed method (GRC and GDB)	-1.2e-3	0.029	0.048

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background

2. Methodology

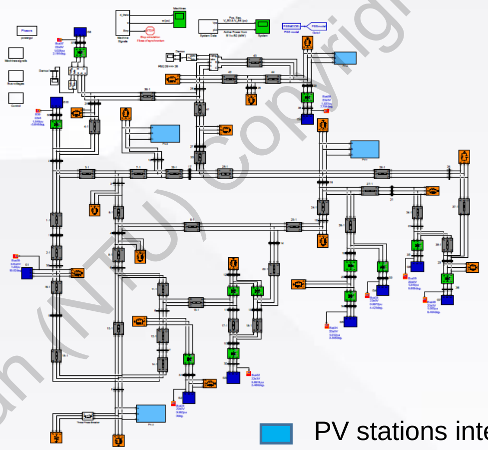
3. Single area

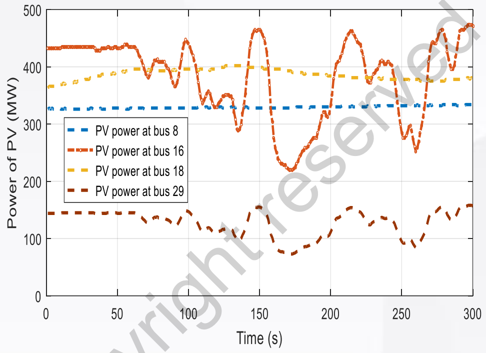
4. Multi-area systems

5. Optimal BESS control

Testing results (time-domain model)

- NE 39-bus system with full dynamic model:





PV stations integrated at load bus 8, 16, 18, 29.

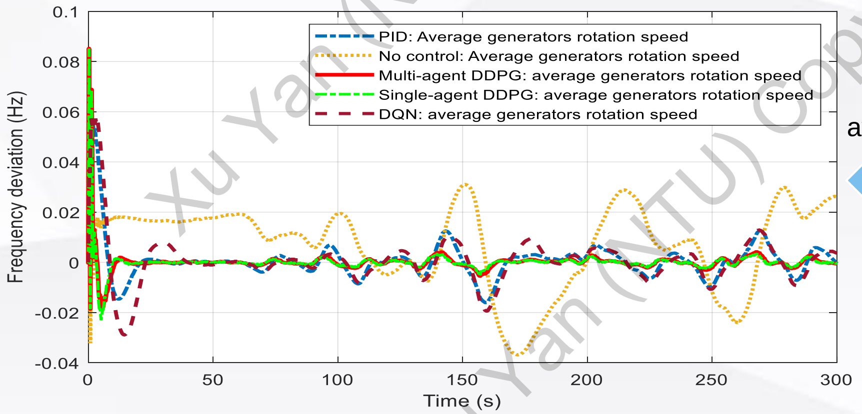
- Numeric comparison

Method	Q	Mean |ACE| %	Max |ACE| [p.u.]
Fine-tuned PID	-7.0e-05	0.0095	0.002
(Deep) Q-learning	-1.35e-4	0.0119	0.002
Single-agent DDPG	-3.4e-05	0.0044	0.002
Proposed MA-DRL	**-3.2e-05**	**0.0047**	0.002
No control	-0.013	0.21	0.002

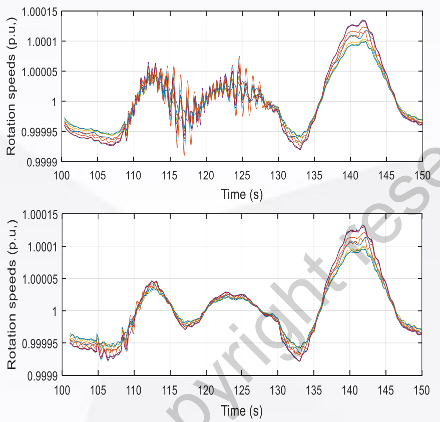
Objective function: less frequency deviations in data-driven methods

More related with system's inertia

- System frequency for different methods



average



Rotation speed of 9 different generators

- **Less frequency deviations**: 76.3% better than DQN, 54.3% better than PID.
- Better **coordination** among all the agents

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background

2. Methodology

3. Single area

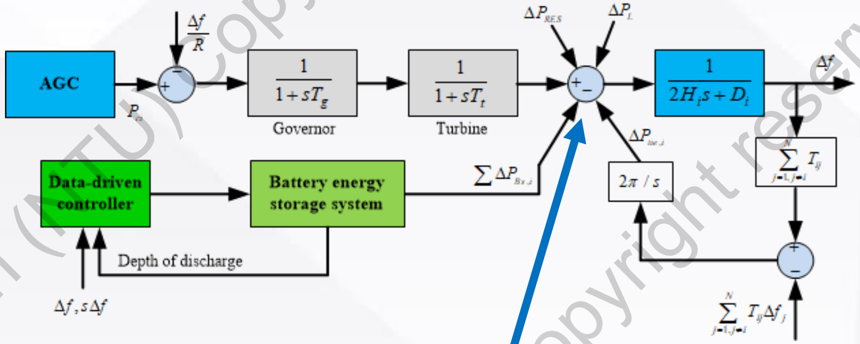
4. Multi-area systems

5. Optimal BESS control

Battery energy storage system control for frequency support

Battery Energy Storage System

- High control flexibility and response speed.
- Intensive usage can cause **battery aging.**



System dynamics with BESS

$$\Delta\dot{f}_i = \frac{1}{2H_i}(\Delta P_{mi} + \Delta P_{BESSi} - \Delta P_{Li} - \Delta P_{tie,i}) - \frac{D}{2H_i}\Delta f_i \quad (1)$$

$$\Delta\dot{P}_{mi} = \frac{1}{T_{ti}}\Delta P_{gi} - \frac{1}{T_{ti}}\Delta P_{mi} \quad (2)$$

$$\Delta\dot{P}_{gi} = \frac{1}{T_{gi}}\Delta P_{ci} - \frac{1}{R_i T_{gi}}\Delta f_i - \frac{1}{T_{gi}}\Delta P_{gi} \quad (3)$$

$$\Delta\dot{P}_{tie,i} = 2\pi\sum_{j=1,j\neq i}^{N} T_{ij}(\Delta f_i - \Delta f_j) \quad (4)$$

System frequency

$$SoC_i(t) = SoC_i(0) - \int_0^t \frac{\eta_i g_i(t)}{3600 E_{rate,i}} dt \quad (5)$$ BESS SoC

$$\Delta P_{ci}(t) = -K_P ACE_i(t) - K_I \int ACE_i(t) \quad (6)$$ AGC

Problem description

Optimize a DRL agent, such that the **expected total control cost** is minimized

$$\underset{\theta^\mu}{\text{Minimize}}\ E_D[\sum_{j=1}^{T}\sum_{i=1}^{J}(c_{u,i} + c_{b,i} + c_{g,i})\Delta t_j]$$

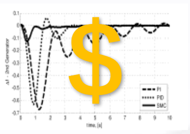
- Modelling of **BESS control cost**

1) Battery Aging Cost

- Cost due to battery marginal degradation.

$$c_b(t) = \text{sgn}[g_t(t)]\frac{g_t(t)R}{\eta EL}\frac{\partial\Phi(\delta)}{\partial\delta}$$

2) Unscheduled interchange



- Cost due to frequency deviations and unscheduled power interchanges.

$$p(f) = \begin{cases} 0 & \text{if } f \in [1.006 f_0, 1.02 f_0) \text{ Hz} \\ \alpha_3 + \beta_3\Delta f & \text{if } f \in [0.99 f_0, 1.006 f_0) \text{ Hz} \\ \alpha_2 + \beta_2\Delta f & \text{if } f \in [0.984 f_0, 0.99 f_0) \text{ Hz} \\ \alpha_1 & \text{if } f \in [0.98_0, 0.984 f_0) \text{ Hz} \end{cases}$$

4) AGC generation cost

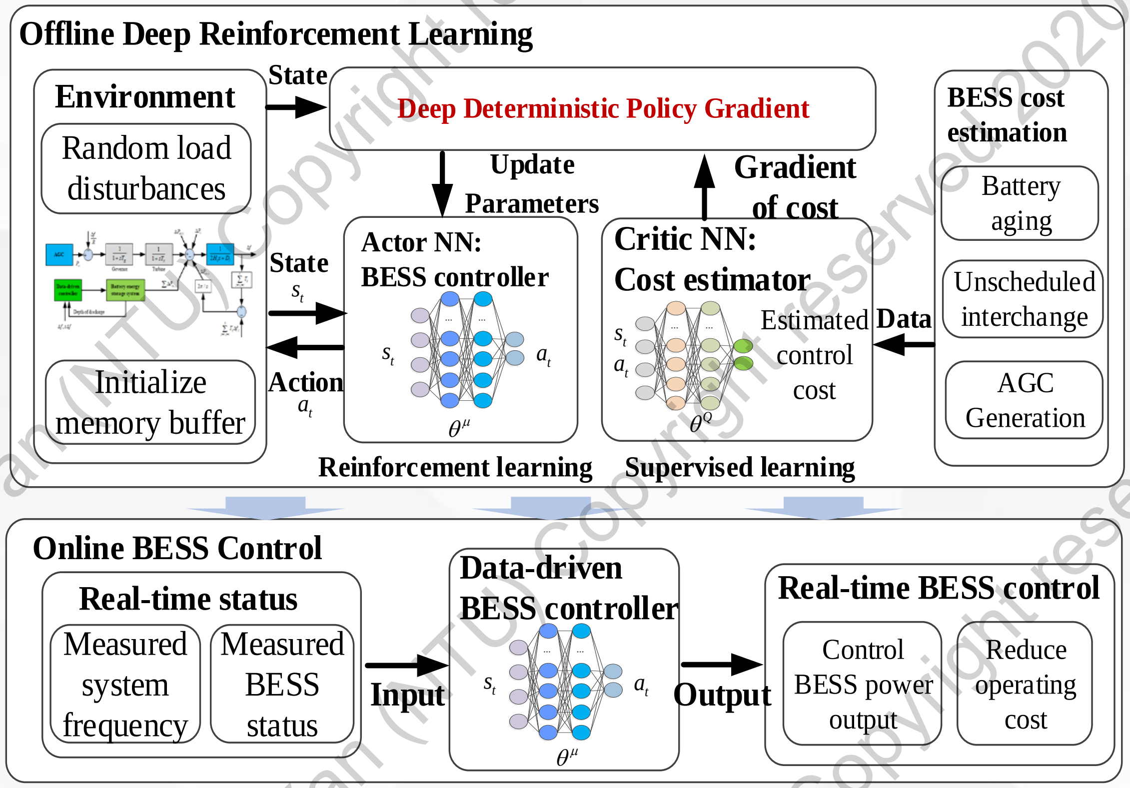
- Additional generations to maintain frequency

$$c_g(t) = \sum_{i=1}^{K}(b_i p_{g,i} + c_i p_{g,i}^2)$$

- Control cost **approximated** by critic network

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

BESS control for frequency support



Agent-Environment Interaction

- Expected action-values:

$$\underset{\theta^{\mu}}{\text{Maximize}}\ E_D[\,Q^{\mu}(s_t,a_t)]$$

- Cost: battery marginal aging, unscheduled interchange, AGC generation
- Cost approximation with critic:

$$Q^{\mu}(s_t,a_t) = -\sum_{T}\,[c_b(t)+c_u(t)+c_g(t)]\Delta t$$

$$\min_{\theta^Q}||Q_R - h^{(n)}_{\theta^Q}[\dots h^{(1)}_{\theta^Q}(\boldsymbol{s},a))]\,||^2$$

- Training process

$$\theta^{\mu'} = \theta^{\mu} + \eta\cdot\nabla_{\theta^{\mu}}J$$

$$\nabla_{\theta^{\mu}}J \approx \frac{1}{N}\sum_i \nabla_a Q(s,a|\theta^{Q})\,\nabla_{\theta^{\mu}}\mu(s|\theta^{\mu})$$

Critic-based gradients

Gradient of objective to BESS action

$$Q_R \approx h^{(n)}_{\theta^Q}[\dots h^{(1)}_{\theta^Q}(\boldsymbol{s},a))$$

$$\nabla_a Q(s,a) \approx \nabla_a h^{(n)}_{\theta^Q}[\dots h^{(1)}_{\theta^Q}(\boldsymbol{s},a))$$

DNN Updating rule

Gradient of action to agent' parameters

$$\nabla_{\theta^{\mu}}\mu(s|\theta^{\mu}) = \nabla_{\theta^{\mu}}(f_{\theta}^{(n)}[\dots f_{\theta}^{(1)}(X))])$$

- Offline Deep Reinforcement learning

The critic NN approximates total control cost and actor gradients. The actor NN (BESS control agent) is optimized with actor gradients.

- Online BESS control

The real-time control action by the optimized DRL agent already considers the control cost.

[3] Z. Yan, Y. Xu, et al, "Data-driven Economic Control of Battery Energy Storage System Considering Battery Degradation," *IET Generation. Transmission & Distribution*, 2020.

1. Background

2. Methodology

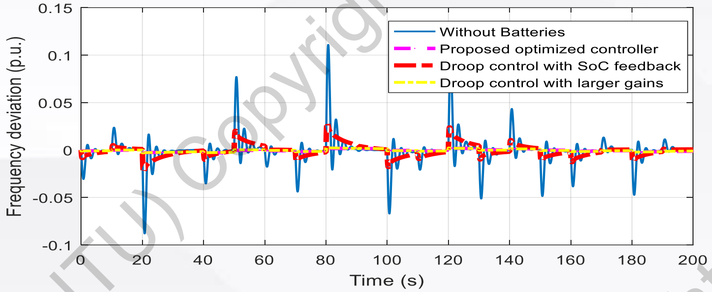
3. Single area

4. Multi-area systems

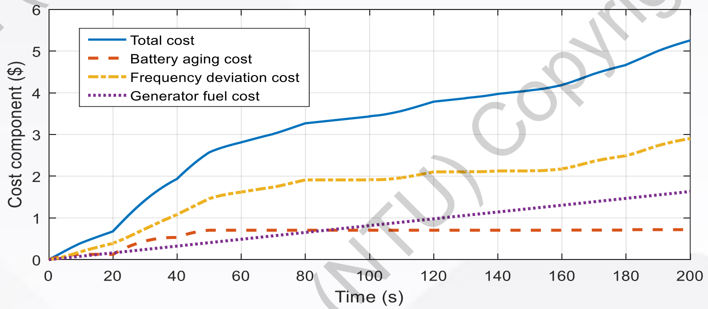
5. Optimal BESS control

Battery energy storage system control for frequency support

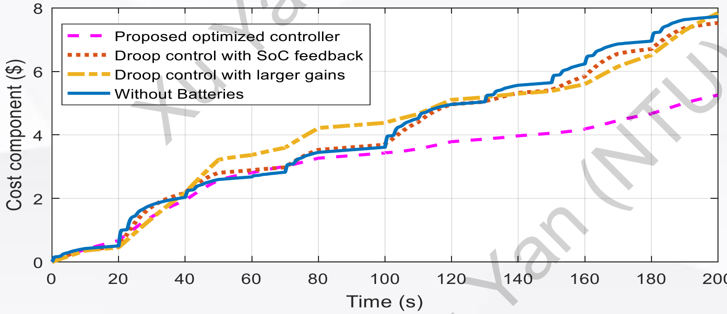
- System frequency in 3 areas



- Accumulative cost (each component)



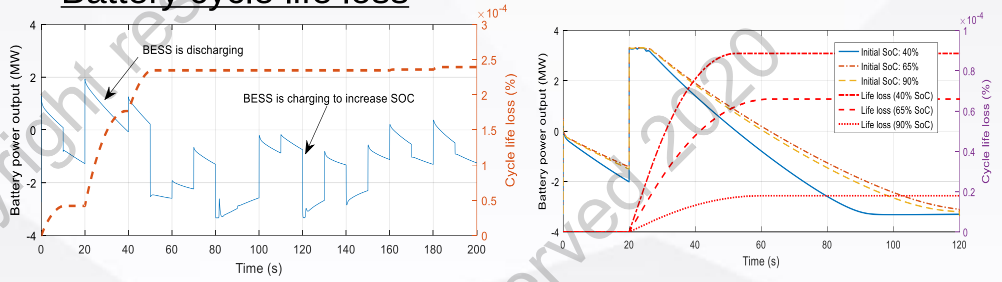
- Accumulative cost (total)



- Numerical results (random load changes)

Method	C ($)	C_b ($)	C_u($)	C_g ($)	Saving (%)
No Batteries	7.73	0.00	6.10	1.63	0.0
Proposed	5.25	0.72	2.90	1.63	32.1
Droop with SoC	7.53	1.43	4.47	1.62	2.6
Droop with larger gains	7.83	4.92	1.29	1.62	-1.3

- Battery cycle life loss



- **Reduced 32.1%** total control cost.
- The BESS control is improved by **avoiding discharging when depth-of-discharge is relatively high**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background

2. Methodology

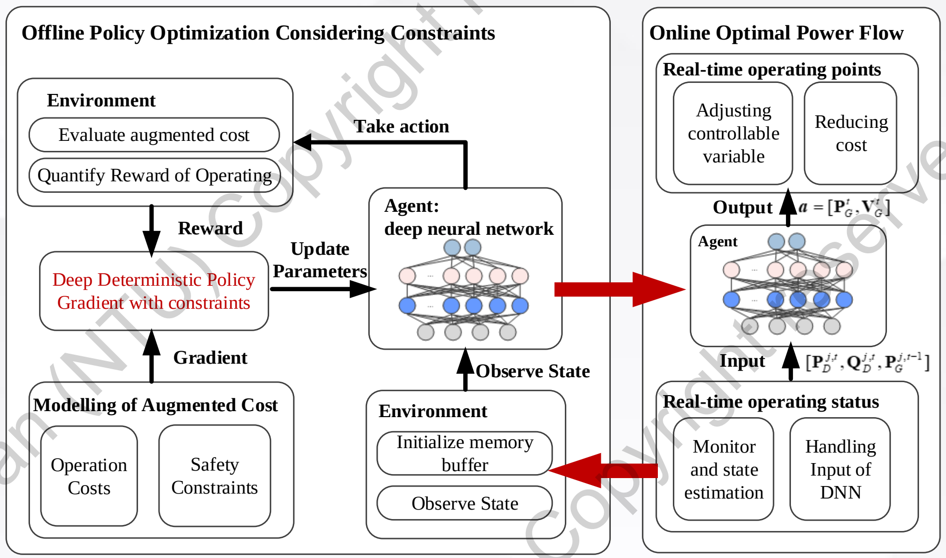
3. Single area

4. Multi-area systems

5. Optimal BESS control

6. Other related works

Related work 1: real-time computation of optimal power flow (RT-OPF)



Train the DRL agent by optimizing augmented action-value function to consider constraints

$$\min_{\theta} \sum_{i}^{N} L_i(\boldsymbol{a}_i, \theta, \lambda, \mu)$$

$$\boldsymbol{L}(\boldsymbol{a}_t, \theta, \lambda, \mu) = \sum_{i=1}^{N_G} \boldsymbol{C}_{Gi}(\boldsymbol{a}_t) + \sum_{j=1}^{N_\lambda} \lambda_j \boldsymbol{g}_j(\boldsymbol{a}_t) + \sum_{k=1}^{N_\lambda} \mu_k \boldsymbol{h}_k(\boldsymbol{a}_t)$$

***Lagrangian* function**

(primal-dual safe reinforcement learning)

Method	Average generation cost (USD$)	Average absolute errors of P_G (MW)	Inequality Constraints	Average time saving (%)
IP method OPF [7] (benchmark)	1.3018×10^5	0.00	All satisfied	0.0%
DC OPF [7]	1.3076×10^5	0.610	Branch flow and nodal voltage not satisfied	90.1%
Supervised learning [3] using a DNN	1.2997×10^5	5.018	Branch flow and generator ramping not satisfied	99.8%
Proposed method	1.3018×10^5	0.186	All satisfied	99.8%

Model-assisted gradient derivation

Expand with mini-batch gradient descent:

$$\nabla_\theta L = \nabla_{\boldsymbol{a}} L \cdot \nabla_\theta \boldsymbol{a}$$

$$\nabla_{\boldsymbol{a}} L = \nabla_{\boldsymbol{a}} (C'_{P_G}(\boldsymbol{a})) + \nabla_{\boldsymbol{a}} (\sum_{k=1}^{N_\lambda} \mu_k h_k(\boldsymbol{a}))$$

$$\nabla_\theta \boldsymbol{a} = \nabla_\theta (f_\theta^{(n)} [\ldots f_\theta^{(1)} ([\mathbf{P}_D^{j,t}, \mathbf{Q}_D^{j,t}, \mathbf{P}_G^{j,t-1}]^T))])$$

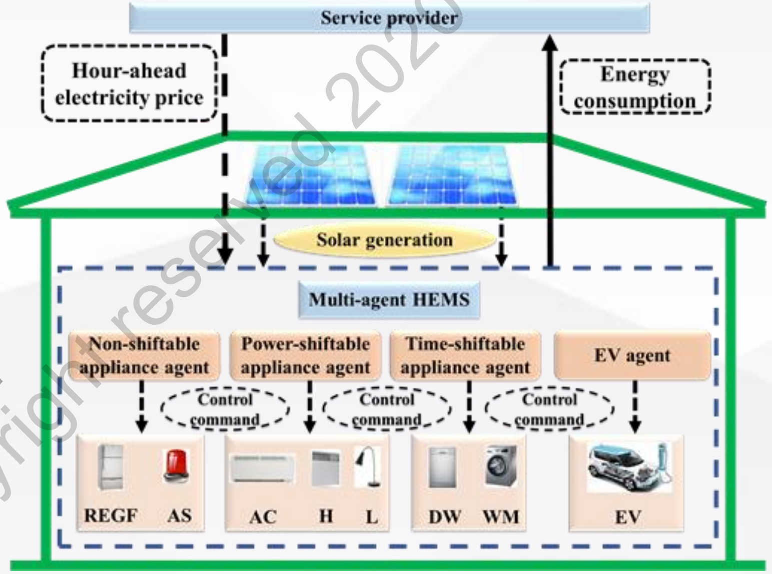
$$\begin{bmatrix} \nabla_{\boldsymbol{a}} L \\ \Delta\lambda \end{bmatrix} \approx \begin{bmatrix} W & G^T \\ G & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\nabla C(\boldsymbol{a}) - H^T \mu \\ -g(\boldsymbol{a}) \end{bmatrix} - \begin{pmatrix} H^T \\ 0 \end{pmatrix} \Delta\mu$$

where, $G = \partial g(\boldsymbol{a}) / \partial \boldsymbol{a}$, W is the Hessian matrix of Lagrangian, $H = \partial h(\boldsymbol{a}) / \partial \boldsymbol{a}$.

Z. Yan and Y. Xu, "Real-Time Optimal Power Flow: A *Lagrangian* based Deep Reinforcement Learning Approach," ***IEEE Trans. Power Systems*, 2020.**

Related work 2: data-driven home energy management system (HEMS)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE + Singtel

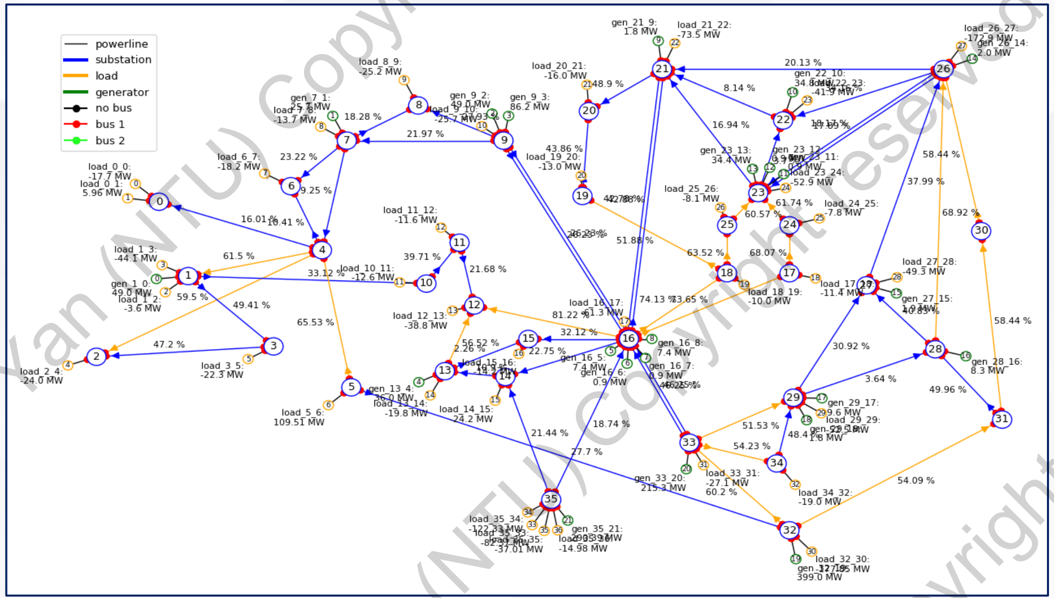


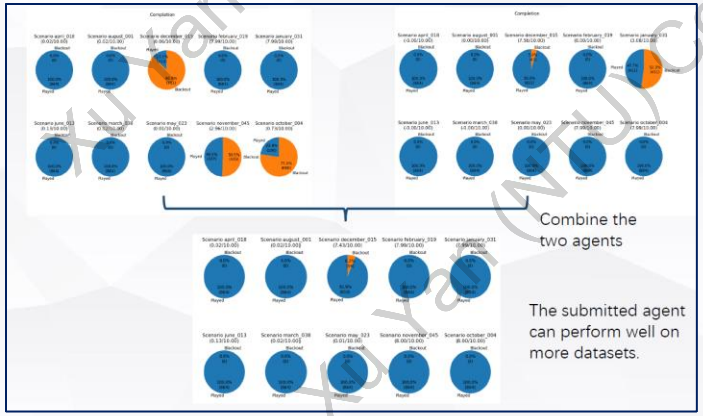
[1] X. Xu, Y. Jia, Y. Xu, Z. Xu, et al, “A Multi-agent Reinforcement Learning based Data-driven Method for Home Energy Management,” ***IEEE Trans. Smart Grid*, 2020.**
[2] X. Xu, Y. Xu, Z. Xu, et al, “Data-driven Game-based Pricing for Sharing Rooftop Photovoltaic Generation and Energy Storage in the Residential Building Cluster under Uncertainties,” ***IEEE Trans. Industrial Informatics*, 2020.**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Related work 3: corrective control optimization

- ✓ Generation rescheduling
- ✓ Network reconfiguration
- ✓ Splitting or coupling busbars at substations
- ✓ Asynchronous Actor-Critic Agents





Detailed introduction of our method can be found at: https://l2rpn.chalearn.org/competitions

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Thank You!