Deep Reinforcement Learning (DRL)-based Data-Driven Control and Operation of Modern Power Systems

Dr Yan Xu
Cham Tao Soon Professor in Engineering
Director, Center for Power Engineering (CPE)
Co-Director, SingaporePower Group – NTU Joint Lab
Nanyang Technological University

1 **Overview**

- Research Background
- Preliminaries of DRL
- Our Research Framework

2 **DRL for Bulk Power Grids**

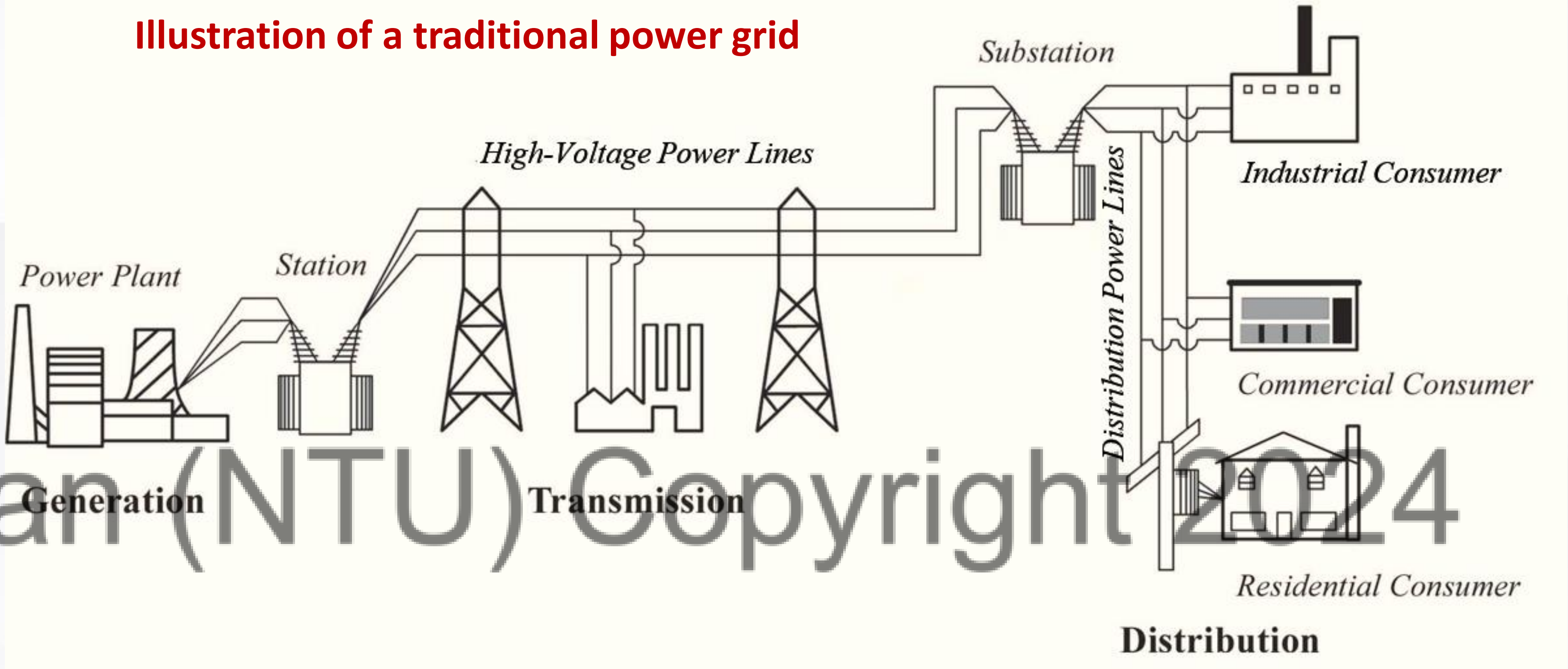
- Load Frequency Control
- Real-Time Optimal power flow
- Topology Optimization

3 **DRL for Microgrids & Active Distribution Grids**

- Frequency Control
- Controller Tuning
- Energy Management
- Volt/Var Control

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Research Background



- High-level renewable energy resources (RES) integration
- Power-converter interfaced RES power plants
- Higher operational uncertainty and lower effective inertia

- Longer transmission distance
- Higher transmission capacity
- Higher voltage
- HVDC & HVAC
- Inter-area (nation) connection

- Distributed energy resources (DER) integration: distributed generators, energy storage systems, flexible loads, new loads (EVs, data centres)
- Microgrids & Active distribution networks
- Smart homes, smart appliances

Various operational and control challenges…

Research Background

Calero *et al.*: Review of Modeling and Applications of Energy Storage Systems in Power Grids

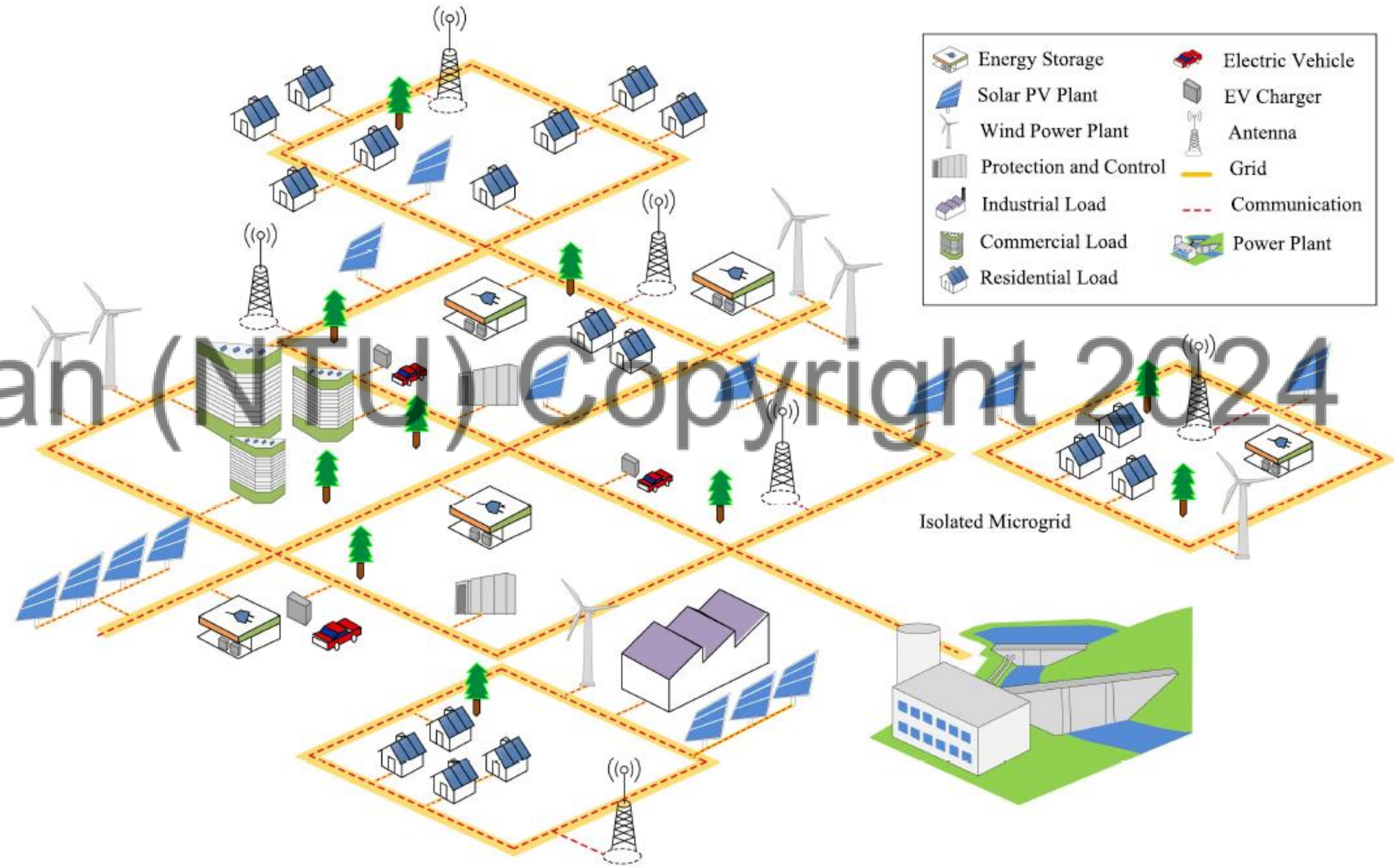
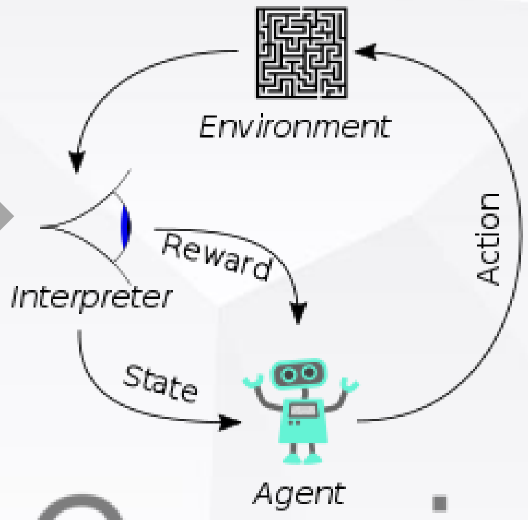


Fig. 1. *Renewable-based smart grids of the future.*

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Preliminaries of Reinforcement Learning (RL)

[Wikipedia]: **Reinforcement learning (RL)** is an interdisciplinary area of machine learning and optimal control concerned with how an intelligent agent ought to take actions in a dynamic environment in order to maximize the cumulative reward.

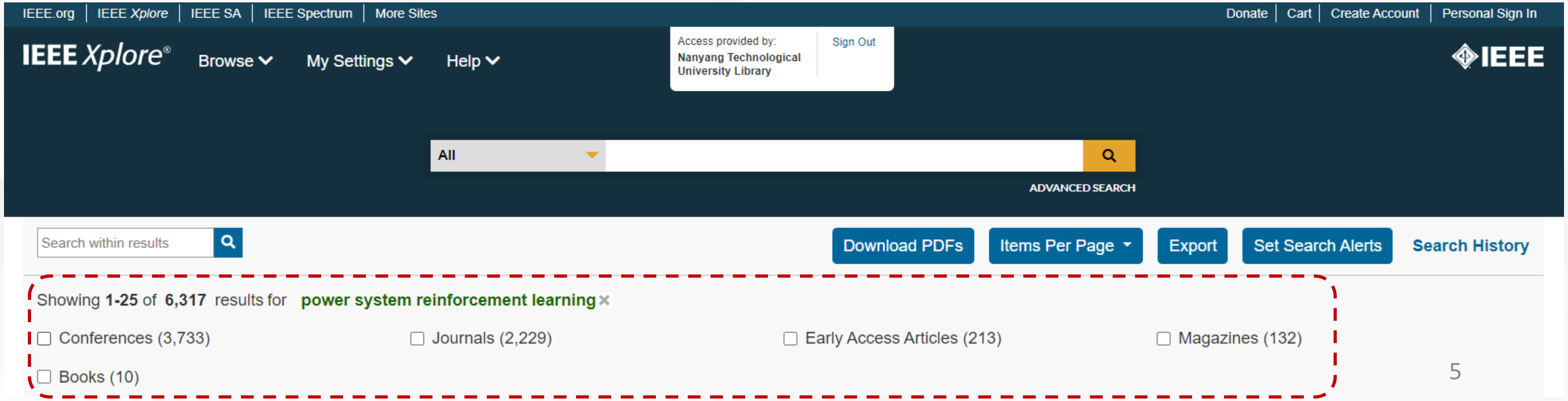


AlphaGo

Google's AlphaGo Beat a World Champion in 2016

Deep RL (DRL): RL + deep learning (e.g., deep neural networks)

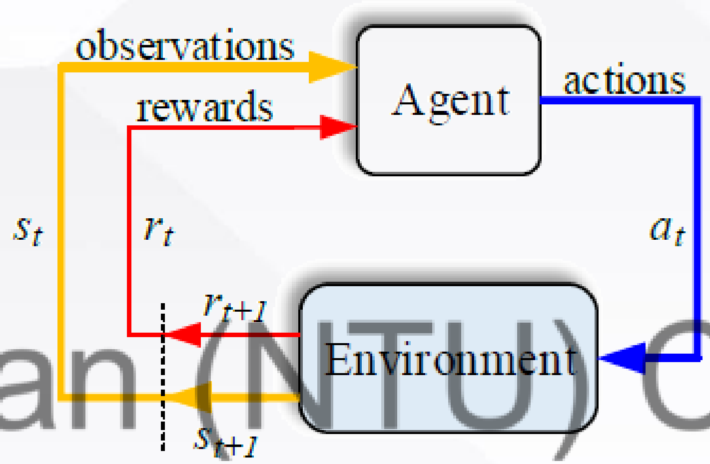
One of the most popular research topics in power systems area, with numerous publications…



Preliminaries of Reinforcement Learning (RL)

Principle & Framework

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



- **Agent**: decision-maker (controller, operator)
- **Environment**: physical system (modeled as a Markov decision process)
- **State (*s*)**: current situation of environment (measurements)
- **Action (*a*)**: agent's decision (control signal, dispatch order)
- **Reward (*r*)**: feedback from the environment (system performance, operation objective)
- **Action value (Q-value)**: total expected reward over *T*

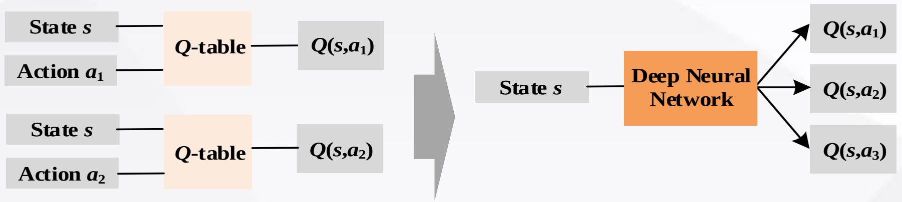
➢ How to **model** power system control and operation problems into a RL process?
➢ How to **solve** the RL training process considering power system's own characteristics/model?

RL methods

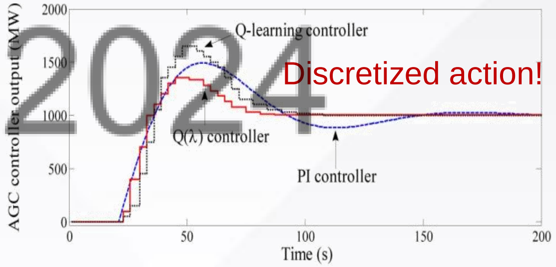
1. Value-based methods – train a Q-value predictor (Q-table)

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

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



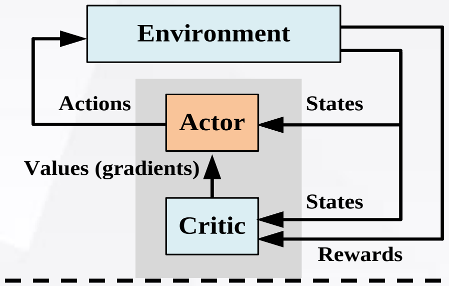
Disadvantages:
- Discretized action.
- Non-satisfactory performance due to discretized action space.



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

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

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



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

What are interesting research problems in this field?

A quick answer: simply running existing DRL algorithms/open-sourced codes for well-modeled problems are NOT interesting..

Our viewpoints and our research efforts (since early 2018):

- **Problem modeling** – target at power system engineering problems that really need DRL
- **Learning framework** – problem-specific design, rather than a universal framework
- **Physics-informed learning** – make use of power system physical models for training
- **Constraints satisfaction** – handle equality and inequality constraints more effectively
- **Safety learning** – maintain (ensure) safety during the learning and decision process
- **Learning efficiency** – improve the learning speed and convergence performance
- **Vulnerability of DRL models** – enhance robustness against adversarial examples
- **Interface with large-language models (LLMs)** – leverage LLMs for modeling difficult problems and solving through interactive learning

SODA group started this research in early 2018 (our first paper was published in early 2019)

Our Research Framework

Deep Reinforcement Learning for Modern Power Systems

Research Problems

Bulk Power Grids

- Power system control
  - Load frequency control
  - Battery storage system
- Power system operation
  - Optimal power flows
  - Network reconfiguration

Microgrids & Active Distribution Networks

- Microgrid frequency control
  - Decentralized Control
  - Distributed Control
  - Controller Tuning
- Microgrid operation
  - Energy Management
  - Volt/Var Control

Funding sources

NATIONAL RESEARCH FOUNDATION
PRIME MINISTER'S OFFICE
SINGAPORE
Research . Innovation . Enterprise

Ministry of Education
SINGAPORE

Singtel

eDF

LITEON

ROLLS ROYCE

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

1 **Overview**

- Research Background
- Preliminaries of DRL
- Our Research Framework

2 **DRL for Bulk Power Grids**

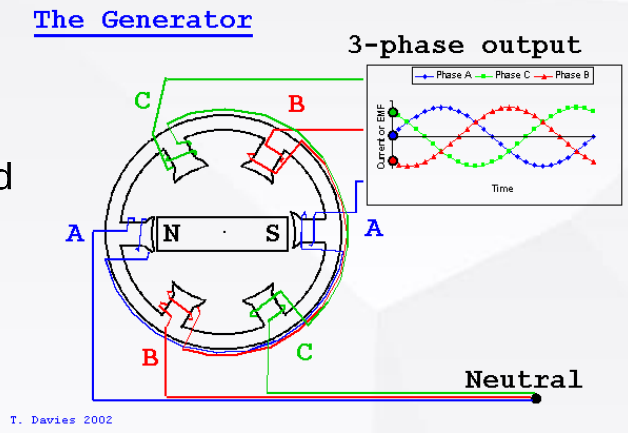
- Load Frequency Control
- Real-Time Optimal power flow
- Network Reconfiguration

3 **DRL for Microgrids & Active Distribution Grids**

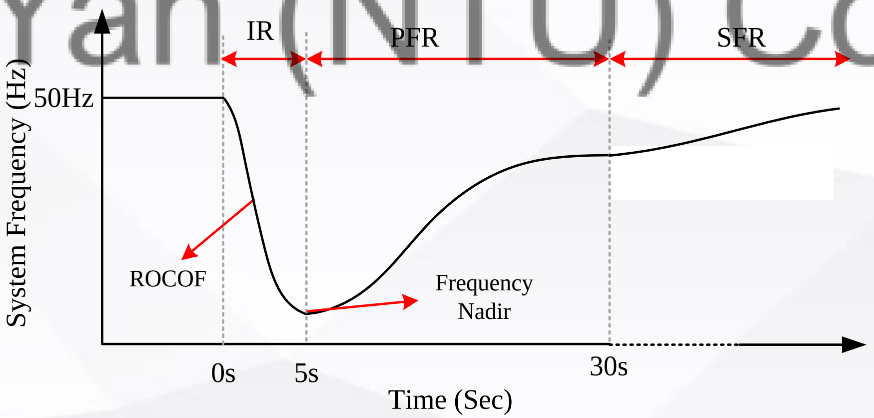
- Frequency Control
- Control Parameter Scheduling
- Energy Management
- Volt/Var Control

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)

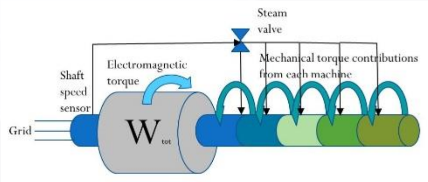
Country/Region	Australia	Europe	North America	Singapore
Nominal frequencies (Hz)	50	50	60	50
Normal operating frequency bands (Hz)	Interconnected system: ±0.15 Islanded system: ±0.5	±0.2	Targeted frequency band: Eastern Interconnection: ±0.018 Western Interconnection: ±0.0228 Texas Interconnection: ±0.030 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

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

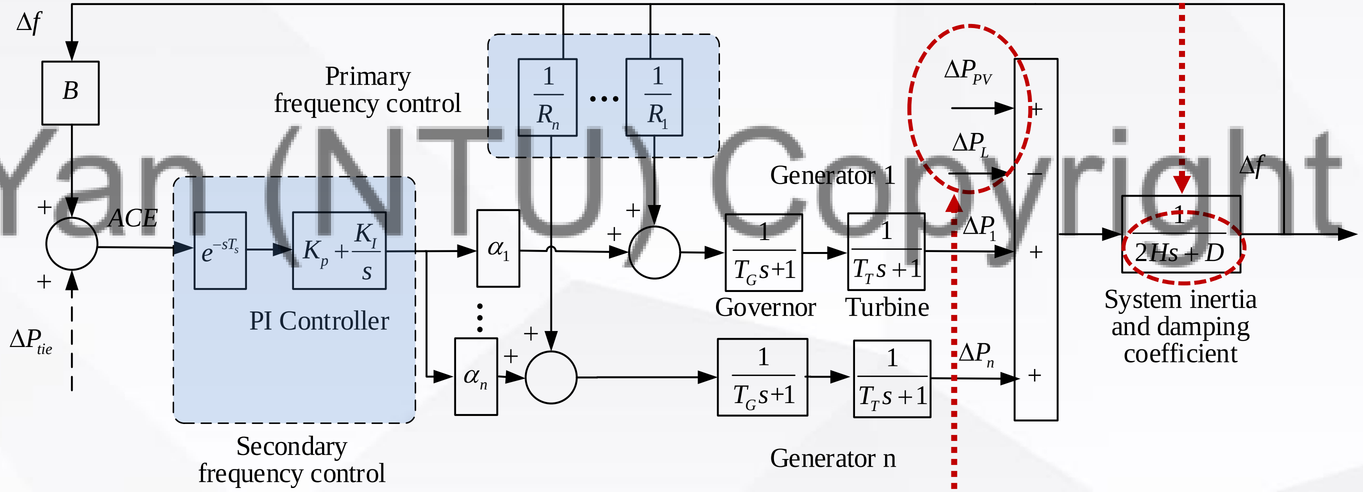
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Load frequency control (LFC)

➢ Lower inertia and load damping:

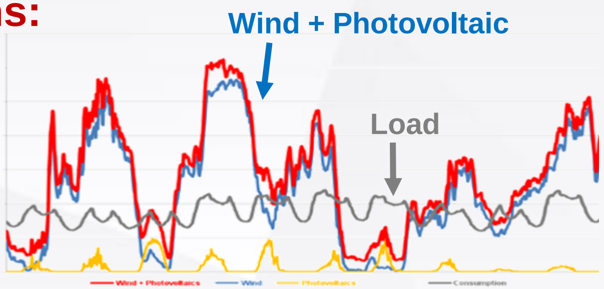


- Generation side: power-converter interfaced generators (wind, solar).
- Transmission side: asynchronous interconnection through HVDC links.
- Load side: inverter-based loads.



➢ Larger and faster power fluctuations:

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

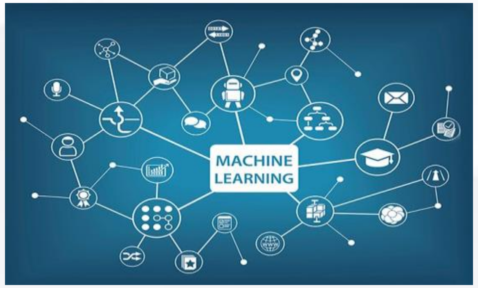


Conventional methods

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

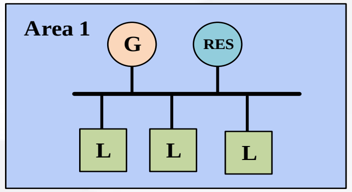


- Stronger modelling capability
- Better control performance
- Higher flexibility and scalability
- etc.

Source of pictures: website (searched in Google)

Our research works in LFC

Single-area controller [1]

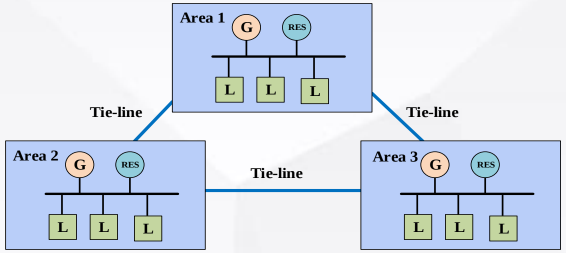


G: generation; L: load;
RES: renewable energy resources;
BESS: battery energy storage system

- 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 – Web of Science highly cited paper**

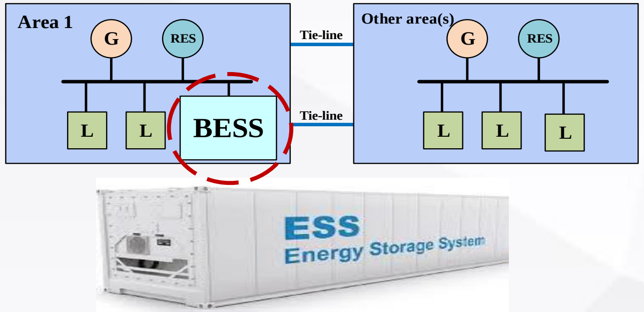
Multi-area controllers [2]



- 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. – Web of Science highly cited paper**

BESS 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

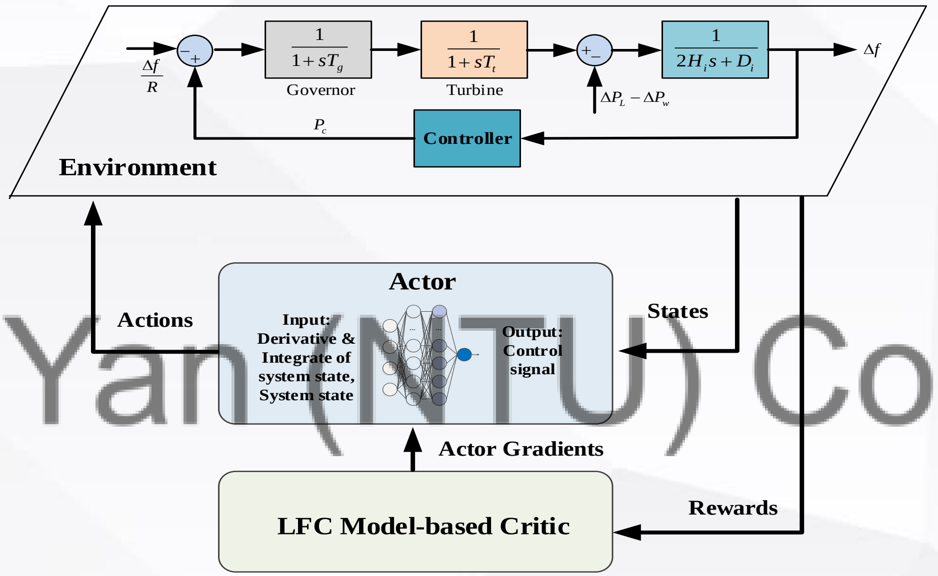
[3] Z. Yan, Y. Xu*, et al, "Deep reinforcement learning-based optimal data-driven control of battery energy storage for power system frequency support," ***IET Gen. Trans. & Dist.*****, 2020.**

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

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

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

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

Physics-informed 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

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

[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** – **Web of Science highly cited paper**

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

Modifying DDPG

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

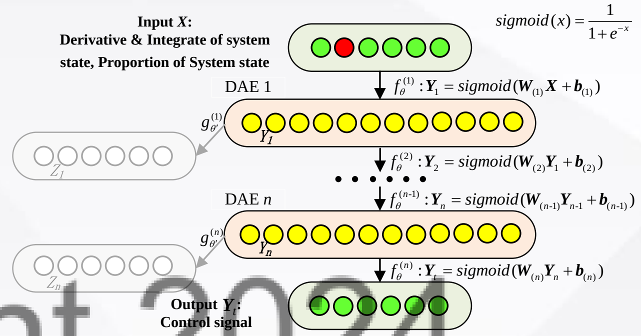
4. $$\nabla_{\theta^{\mu}}\mu(s\,|\,\theta^{\mu}) = \nabla_{\theta^{\mu}}\left(f_{\theta}^{(n)}[\ldots f_{\theta}^{(1)}(\boldsymbol{X})]\right)\Big|_{\boldsymbol{X}\ \text{is input vector with}\ s=s_i}$$

Improved agent updating rule

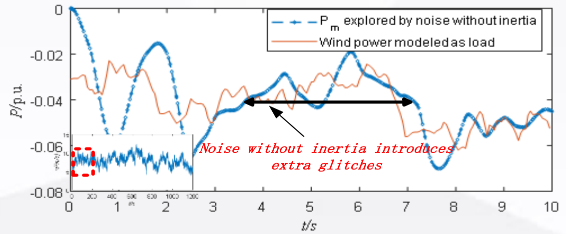
5. $$\begin{cases} W_{ij}^{(l,T+1)} = W_{ij}^{(l,T)} - \eta\dfrac{1}{m}\sum_{k=r}^{r+m}\nabla_a Q^{\mu}(s_t,a_t)\dfrac{\partial}{\partial W_{ij}^{(l,T)}}a(\boldsymbol{W},\boldsymbol{b}) \\ b_i^{(l,T)} = b_i^{(l,T)} - \eta\dfrac{1}{m}\sum_{k=r}^{r+m}\nabla_a Q^{\mu}(s_t,a_t)\dfrac{\partial}{\partial b_i^{(l,T)}}a(\boldsymbol{W},\boldsymbol{b})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

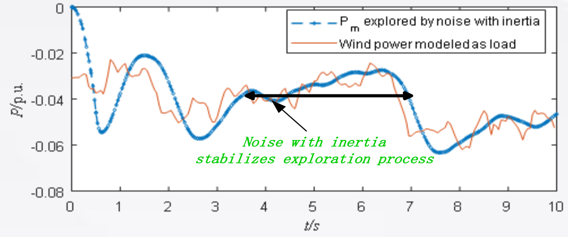
Tricks to improve performance

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.



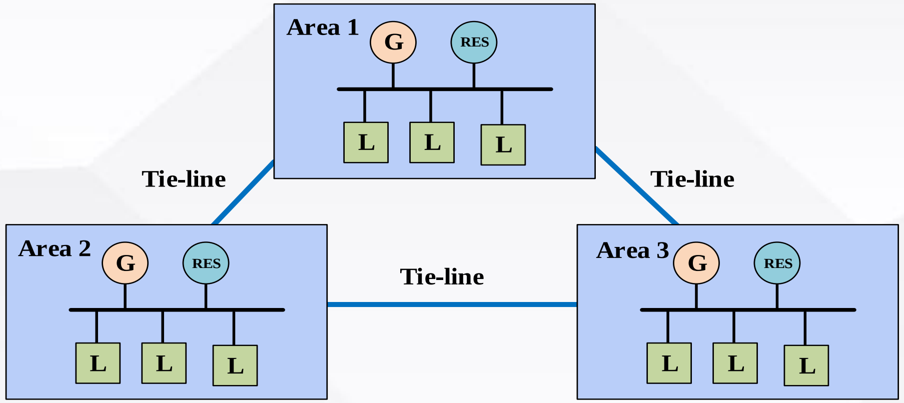
Auto-correlated exploration noise:
Stabilize the exploration process with moving average.





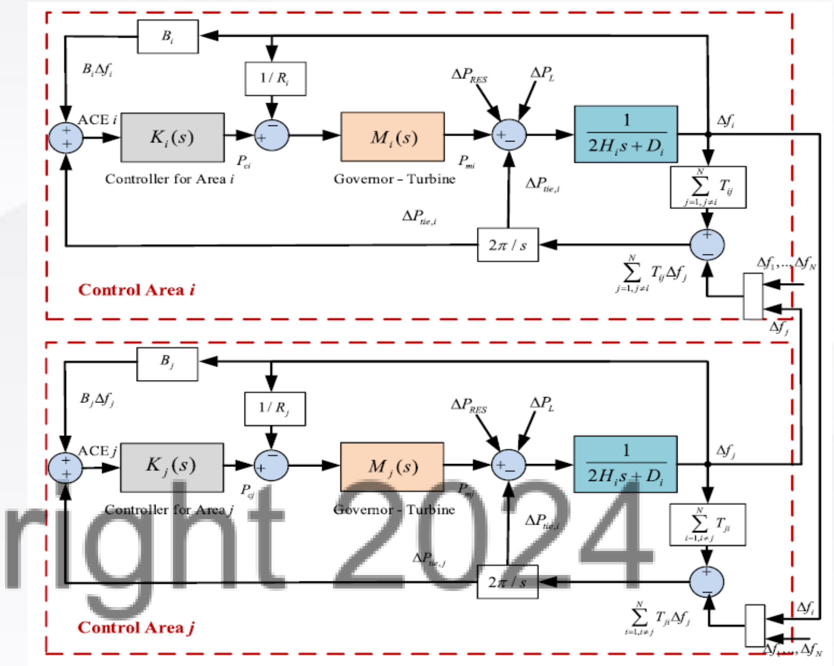
Multi-area LFC controller

Multi-area power system



G: generation **L**: load
RES: renewable energy sources
Each area has its own control agent.

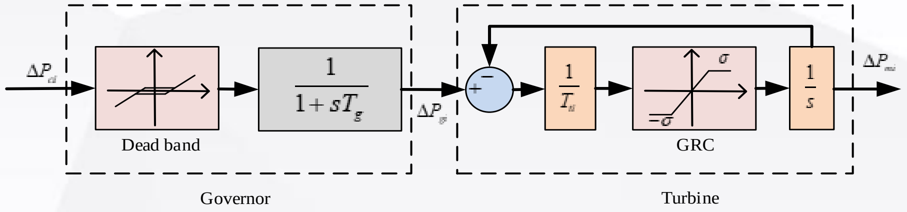
Multi-area LFC block diagram



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

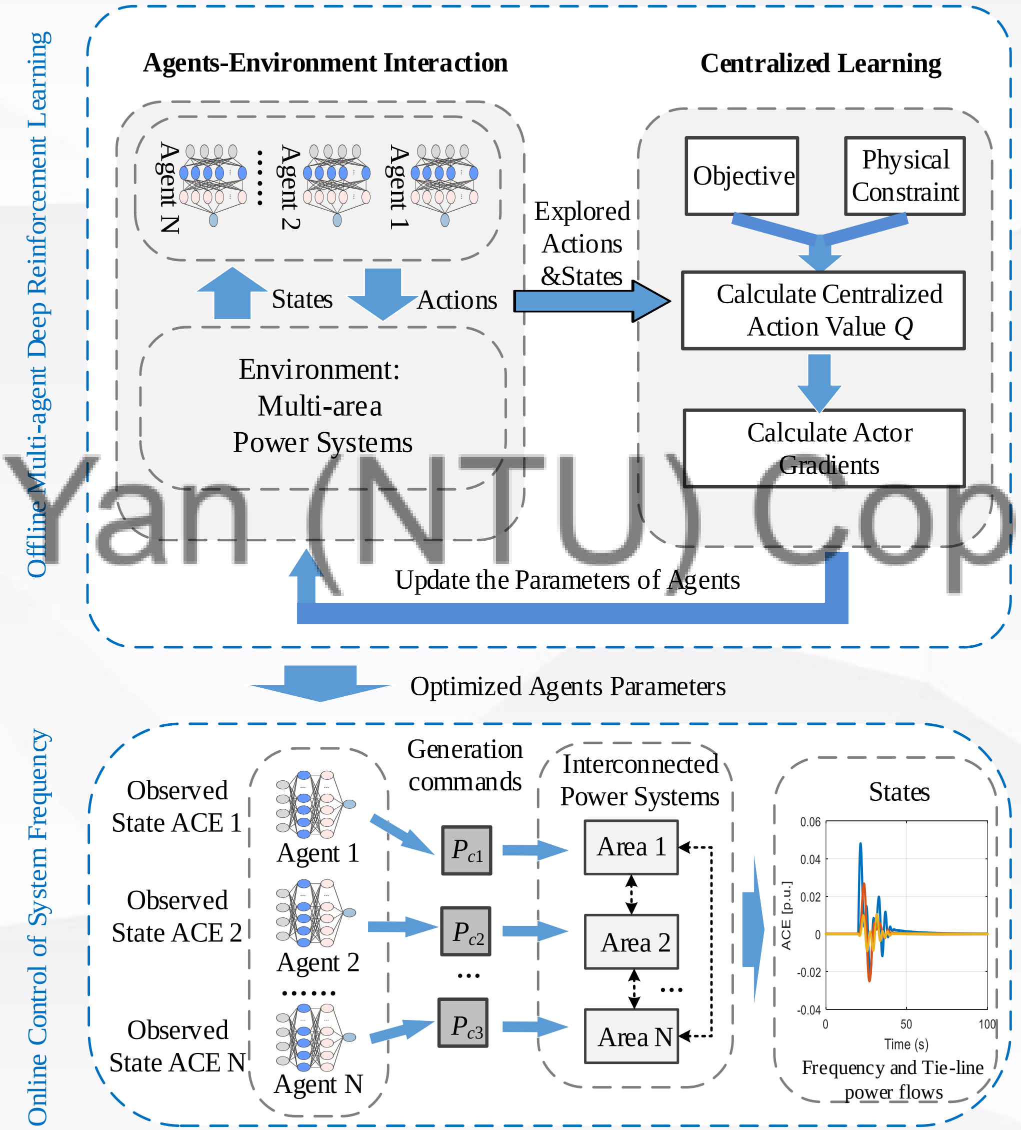


generation dead band (GDB) and generation rate constraints (GRC)

[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

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**:

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

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

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

Physics-informed 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

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

$|_{X \text{ is input vector with } s=s_i}$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Multi-area LFC controller

Gradients for all actors (MA-DDPG)

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

Expand

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

Model-assisted gradient approximation

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

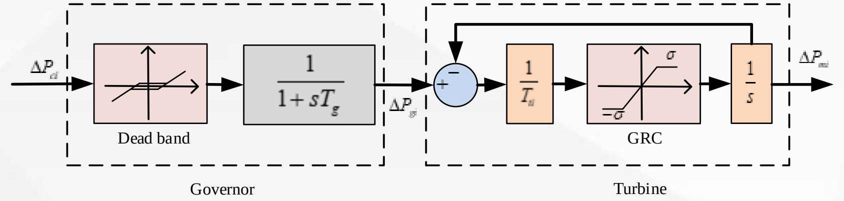
Considering generation rate constraints (GRC)

4.
$$|dP'_m(t)/dt| > \sigma$$

$$P_m(t-\Delta t) \pm \sigma\Delta t = 2H\frac{d\Delta f(t)}{dt} + D\Delta f(t)$$

$$\nabla_a(\Delta f(t)) + (2H/D)\nabla_a(d\Delta f(t)/dt) = 0$$

$$\partial Q/\partial a_i \approx 0 \quad \begin{cases} if\ \partial Q/\partial a_i > 0,\ da_i/dt > \sigma T_{ti} \\ if\ \partial Q/\partial a_i < 0,\ da_i/dt < -\sigma T_{ti} \end{cases}$$

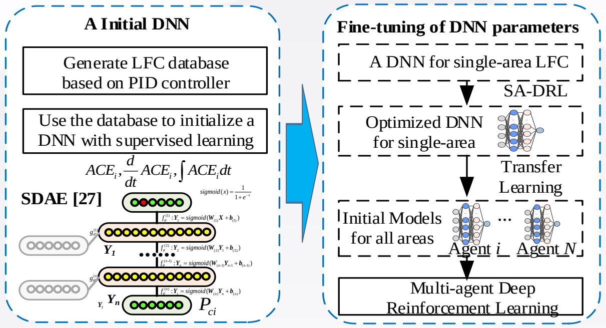


Agent updating rule considering physical limits

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

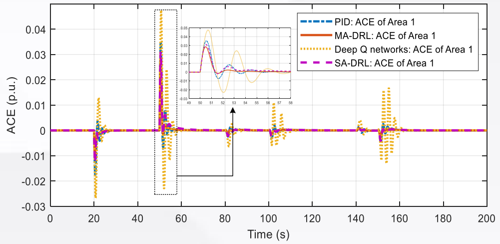
Tricks to improve performance

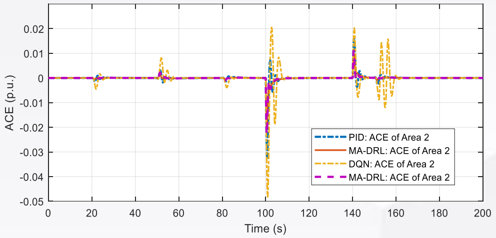
Initialization:
Initialize the DRL agent by supervised learning (data generated by PID controller), then further improved with reinforcement learning.

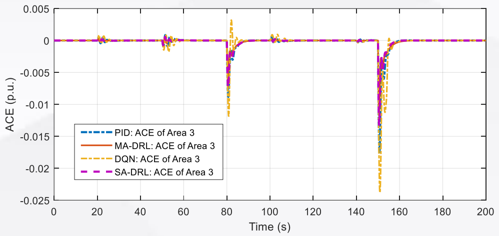


Testing results (on LFC model)

- Linearized LFC model (no physical limits):

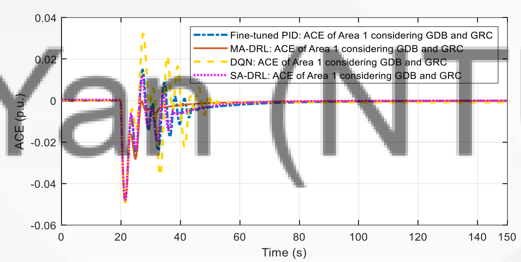


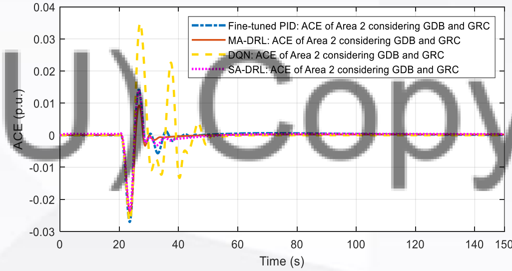


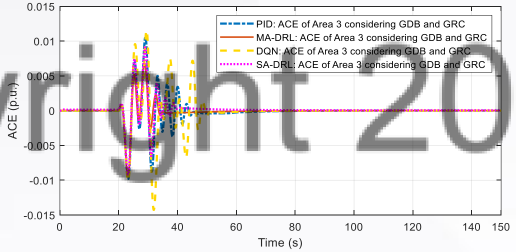


- Less expected frequency deviations: 87.7% better than DQN, 57.5% better than PID.
- Smaller frequency nadir: 39.6% better than DQN, 17.1% better than PID.

- Nonlinearity (GRC&GDB):

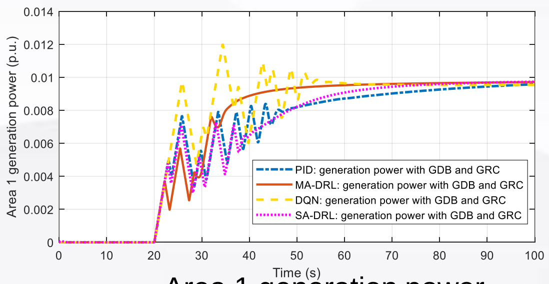




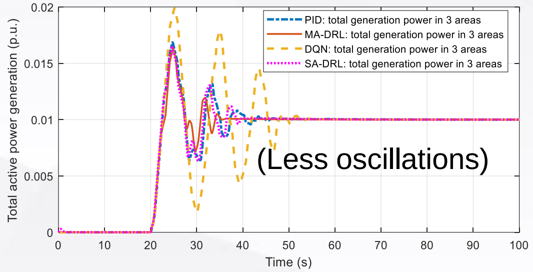


- Less deviations: 62.5% better than DQN, 22.2% better than PID.
- Improves the LFC performance by better coordination among all the areas

- Generation power under GRC&GDB:



Area 1 generation power

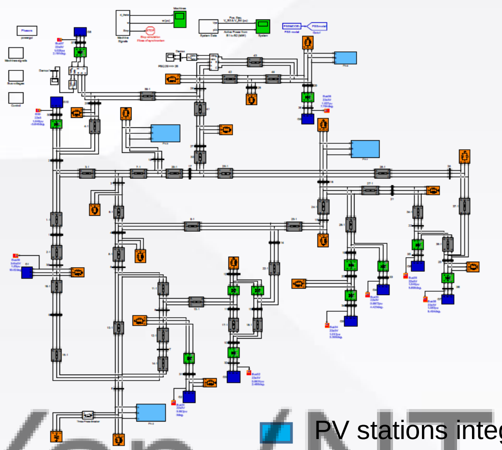


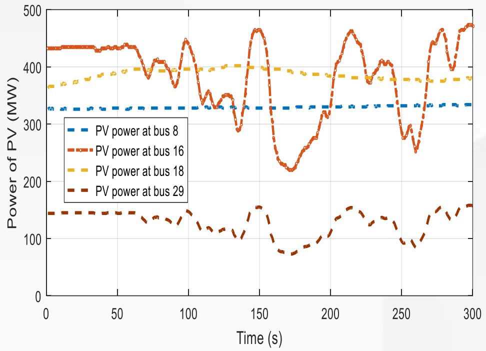
Total generation in 3 areas

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

Testing results (on time-domain model)

- NE 39-bus system with full dynamic model:





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

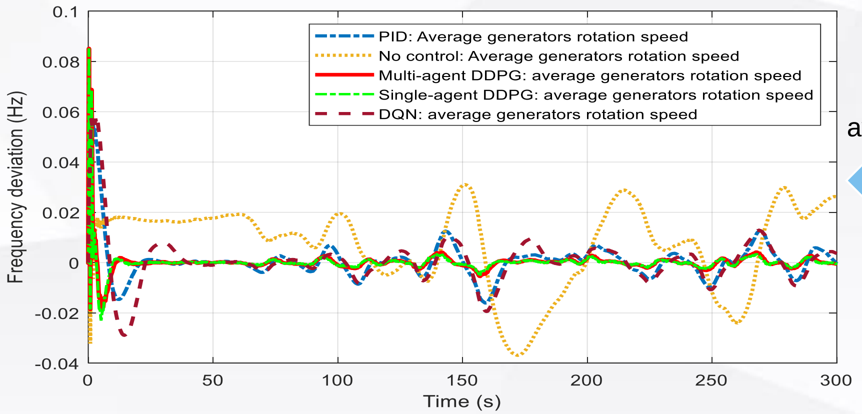
- Numeric comparison

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

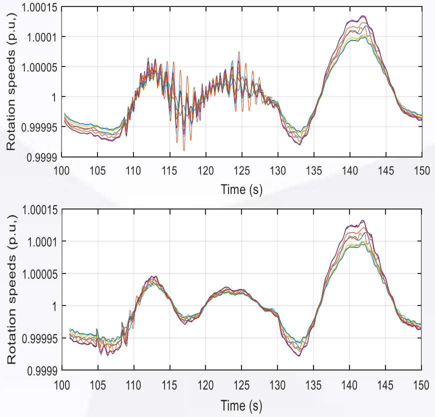
Objective function: less frequency deviations in data-driven methods

More related with system's inertia

- System frequency for different methods



average



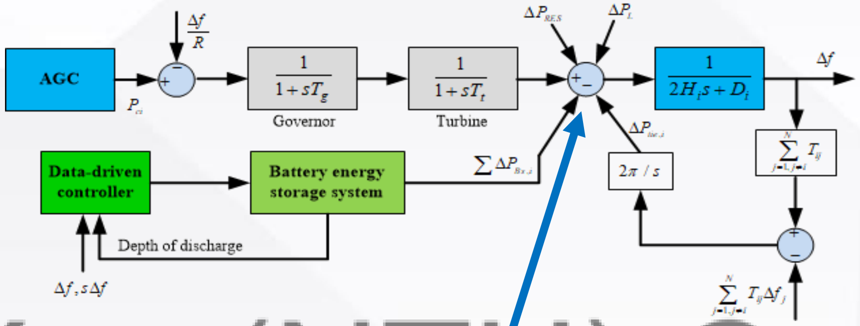
Rotation speed of 9 different generators

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

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

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

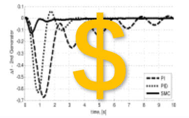
- Modelling of **BESS control cost**

1) Battery Aging Cost

- Cost due to battery marginal degradation.

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

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

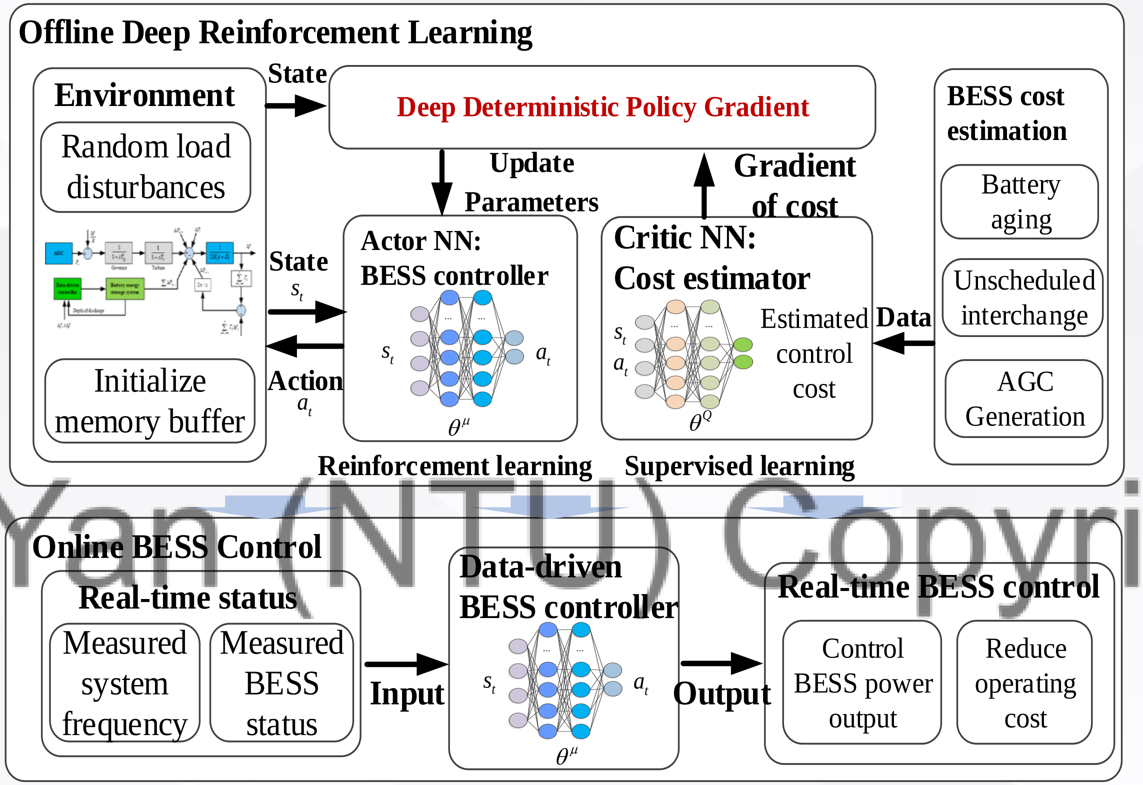
$$c_g(t) = \sum_{i=1}^{K}(b_i p_{g,i} + c_i p_{g,i}^2)$$

- Control cost **approximated** by critic network

[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

Battery energy storage system control for frequency support



Agent-Environment Interaction

- Expected action-values:

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

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

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

$$\min_{\theta^Q} \left|\left| Q_R - h_{\theta^Q}^{(n)}\left[\ldots h_{\theta^Q}^{(1)}(\boldsymbol{s}, a)\right)\right] \right|\right|^2$$

- 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

$$Q_R \approx h_{\theta^Q}^{(n)}\left[\ldots h_{\theta^Q}^{(1)}(\boldsymbol{s}, a)\right)$$

$$\nabla_a Q(s, a) \approx \nabla_a h_{\theta^Q}^{(n)}\left[\ldots h_{\theta^Q}^{(1)}(\boldsymbol{s}, a)\right)$$

DNN Updating rule

Gradient of action to agent' parameters

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

- Offline Deep Reinforcement learning

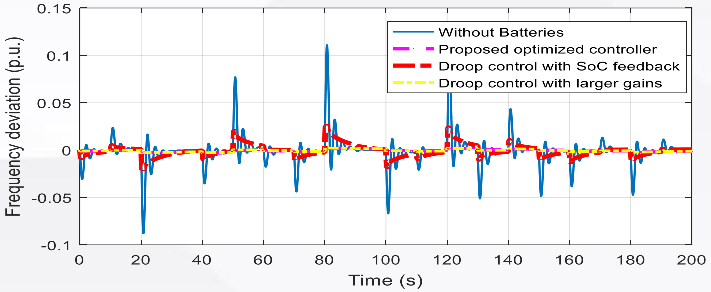
The critic NN approximates total control cost and actor gradients. The actor NN (BESS control agent) is optimized with actor gradients.

- Online BESS control

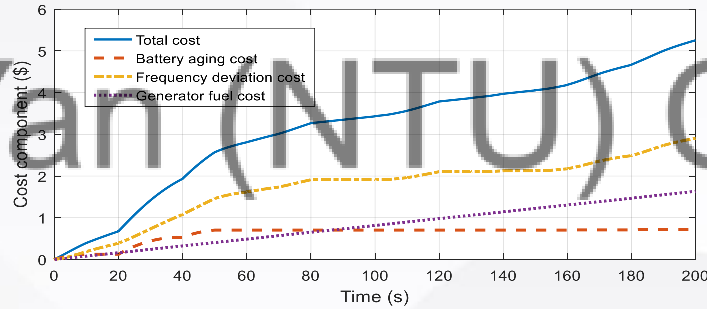
The real-time control action by the optimized DRL agent already considers the control cost.

Battery energy storage system control for frequency support

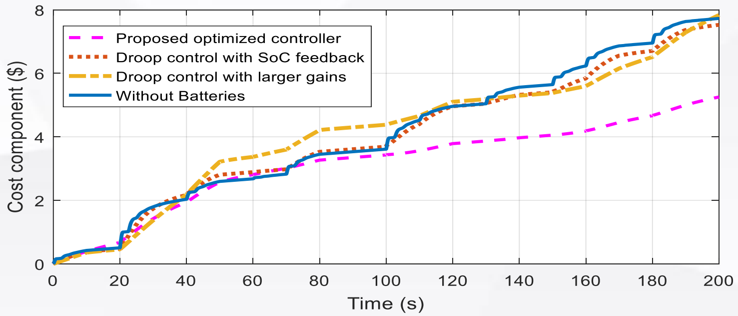
- System frequency in 3 areas



- Accumulative cost (each component)



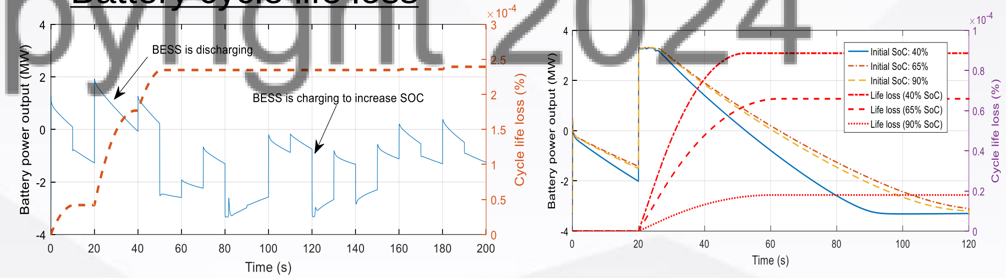
- Accumulative cost (total)



- 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



➢ **Simulation Results**

- **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

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Power System Real-time Operation Challenges

- Renewable energy resources (RES)

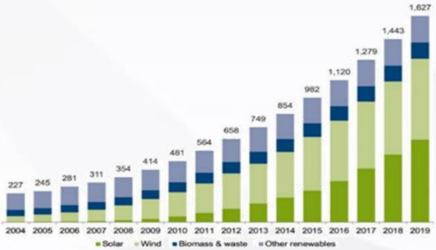
➢ Converter-interfaced RES

- Interfaced into the grid with power electronics **converters**.
- Fast changing power and lower synchronous **inertia**…

➢ Intermittency of RES

- Weather changes and cloud movements.
- **Intermit** RESs power supply that may cause stability issues

- Power systems with high-level RES



Follow the variations of generations and loads under uncertainties

➢ Modeling Difficulties

- The behaviours and dynamics of power systems become more complex to model.

➢ Faster Decision-Makings

- To timely and safely provide operation decision under uncertainties.

Fundamental Changes

➢ High-level RES

- Difficult to have **accurate models**.
- More complexity to handle the **uncertainties** in real-time in faster changing environments.

➢ Power System Operation

- To provide quality power at a reasonable cost across different timescale (**faster**).
- Power balances, power flows, contingencies.

Data-driven methods

Modelling Capabilities

- Modelling capabilities to uncertainties.
- Still complying constraints.

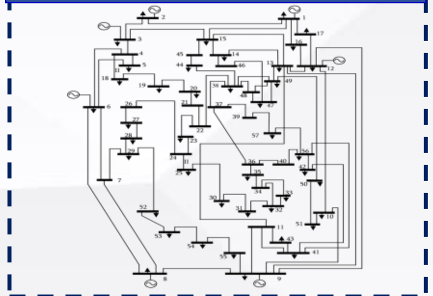
Fast Response Speed

- Inherent faster decision speed.

Source of pictures: website (searched in Google)

Our research works in Power System Operation

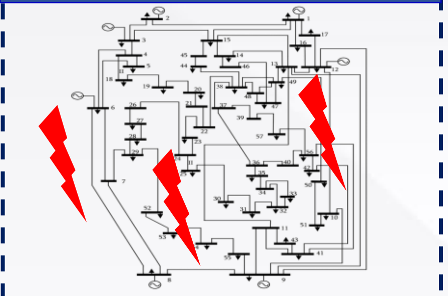
Real-time OPF [4]



- Developed a policy-based DRL model for real-time calculation of optimal power flow.
- Minimize augmented operation cost.
- Model-assisted gradients derivation
- Stacked denoising auto-encoder (SDAE) for feature learning

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

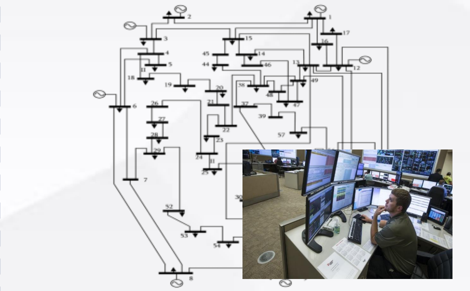
Security Constrain-OPF [5]



- Developed a hybrid data-driven model for fast solution of security-constrained OPF considering credible contingencies.
- Hybridizes primal dual DDPG and SCOPF power flow models.
- Optimize augmented operation cost with model-assisted gradients derivation

[5] Z. Yan and Y. Xu, " A Hybrid Data-driven Method for Fast Solution of Security-Constrained Optimal Power Flow," IEEE Trans. Power Systems, 2022.

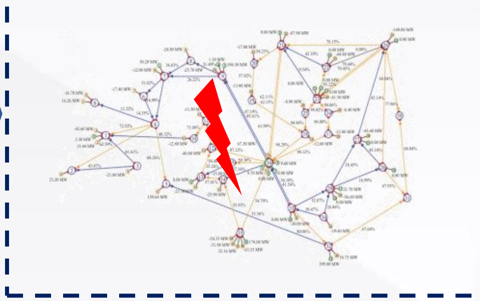
Linguistic OPF [6]



- Leverage GPT-Agent to interpret the language requirements of operation.
- Improving operation performance by considering Grid Code.
- Optimize augmented operation cost with additional GPT-interpreted rewards.
- Constraints with Primal-Dual DRL.

[6] Z. Yan, Y. Xu, "Real-Time Optimal Power Flow with Linguistic Stipulations: Integrating GPT-Agent and Deep Reinforcement Learning," IEEE Trans. Power Systems, 2023.

Topology Optimization [7]



- DRL model for topology optimization.
- Evaluate and mitigate the security risks of DRL models in power systems
- Vulnerability assessment method for DRL models under noisy data and cyber-attack.
- Perturbations to minimize the model's performance. Several vulnerability indices.

[7] Y. Zheng, Z. Yan, K. Chen, J. Sun, Y. Xu and Y. Liu, "Vulnerability Assessment of Deep Reinforcement Learning Models for Power System Topology Optimization," IEEE Trans. Smart Grid, 2021.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Overview

2. Power Systems

2.1 Frequency control

2.2 Optimal power flow

2.3 Topology optimization

3. Microgrids

3.1 Frequency control

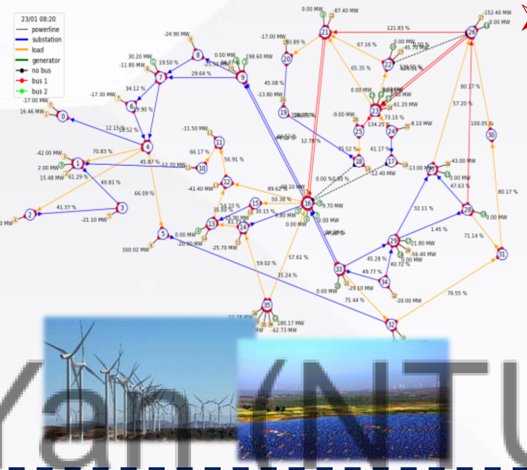
3.2 Controller tuning

3.3 Energy management

3.4 Volt/Var control

Real-time computation of optimal power flow (RT-OPF)

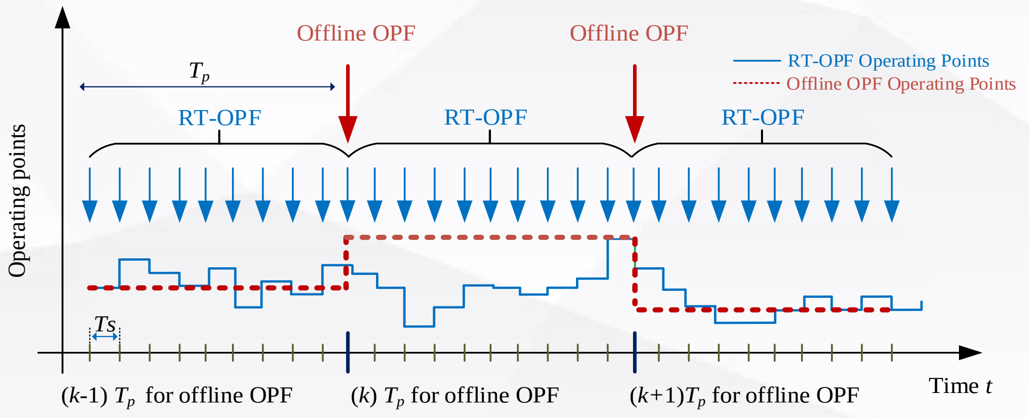
- <u>Real-Time Optimal Power Flow Problem Formulation</u>



➢ **High Uncertainties**

- **Faster changing** environments bring more he uncertainties in real-time
- Maintaining **accurate models** of the continuously changing system becomes challenging.

Real-time solution



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

RT-OPF Formulation

Minimize the operation cost (fuels)

$$\min \sum_{i=1}^{N_G} C_{Gi}(P_{Gi}^t)$$

Satisfying operation constraints

$$P_{Gi} - P_{Di} = V_i \sum_{j=1}^{n} V_j (G_{ij}\cos\delta_{ij} + B_{ij}\sin\delta_{ij})$$

$$Q_{Gi} - Q_{Di} = V_i \sum_{j=1}^{n} V_j (G_{ij}\sin\delta_{ij} - B_{ij}\cos\delta_{ij})$$

$$\max[P_{Gi}^{\min}, P_{Gi}^{t-1} - R_{Gi}^{down}] \le P_{Gi}^t \le \min[P_{Gi}^{\max}, P_{Gi}^{t-1} + R_{Gi}^{up}]$$

$$V_i^{\min} \le V_i \le V_i^{\max}$$

$$|V_i V_j (G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}) - V_i^2 G_{ij}| \le L_{ij}^{\max}$$

Constrained DRL Formulation

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

$$\boldsymbol{L}(\boldsymbol{a_t}, \theta, \lambda, \mu) = \sum_{i=1}^{N_G} \boldsymbol{C_{Gi}}(\boldsymbol{a_t}) + \sum_{j=1}^{N_\lambda} \lambda_j \boldsymbol{g_j}(\boldsymbol{a_t}) + \sum_{k=1}^{N_\lambda} \mu_k \boldsymbol{h_k}(\boldsymbol{a_t})$$

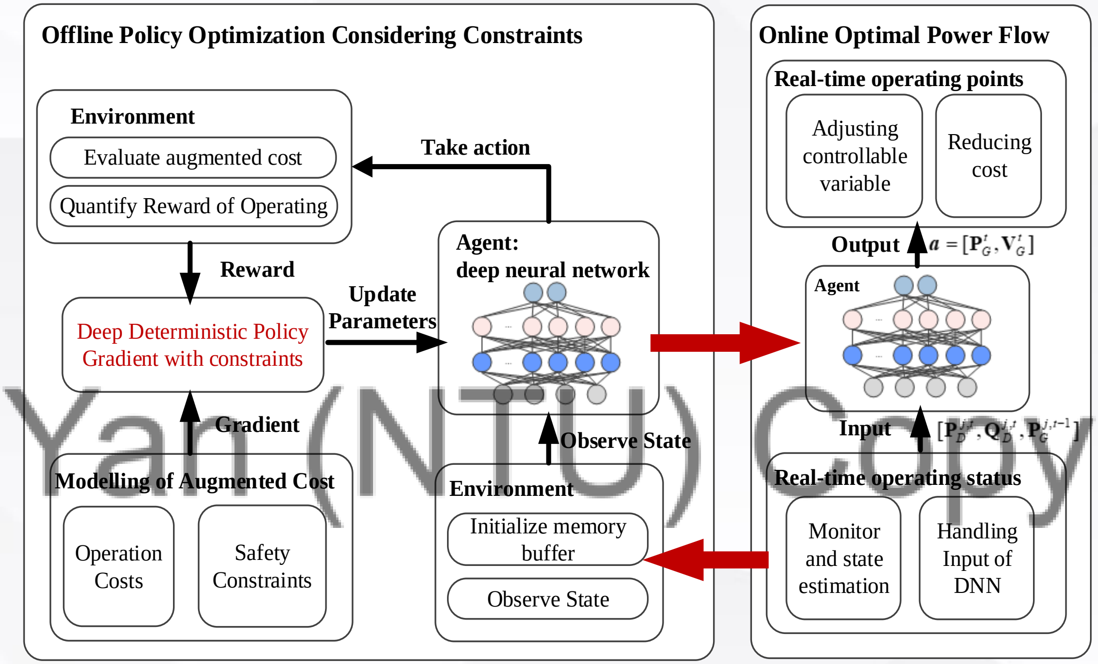
***Lagrangian* function**

(primal-dual safe reinforcement learning)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Real-time computation of optimal power flow (RT-OPF)

- Physics-Informed Policy Training



Formulation of deep reinforcement learning for real-time optimal power flow

Constrained optimization for multiple scenarios → Optimization of deep neural network parameters

Primal-dual safe reinforcement learning

Constrained DRL Formulation

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

$$L(\boldsymbol{a}_t, \theta, \lambda, \mu) = \sum_{i=1}^{N_G} C_{Gi}(\boldsymbol{a}_t) + \sum_{j=1}^{N_\lambda} \lambda_j g_j(\boldsymbol{a}_t) + \sum_{k=1}^{N_\lambda} \mu_k h_k(\boldsymbol{a}_t)$$

Solving the training problem

$$\theta^{k+1} = \theta^k - \eta \nabla_\theta L(\boldsymbol{a}, \theta, \lambda, \mu) \quad \text{(descent)}$$

$$\mu^{k+1} = \mu^k + \eta \nabla_\mu \Big(\sum_{k=1}^{N_\lambda} \mu_k h_k(\boldsymbol{a})\Big) \quad \text{(ascent)}$$

Physics-informed gradient derivation (KKT)

Expand with mini-batch gradient descent:

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

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

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

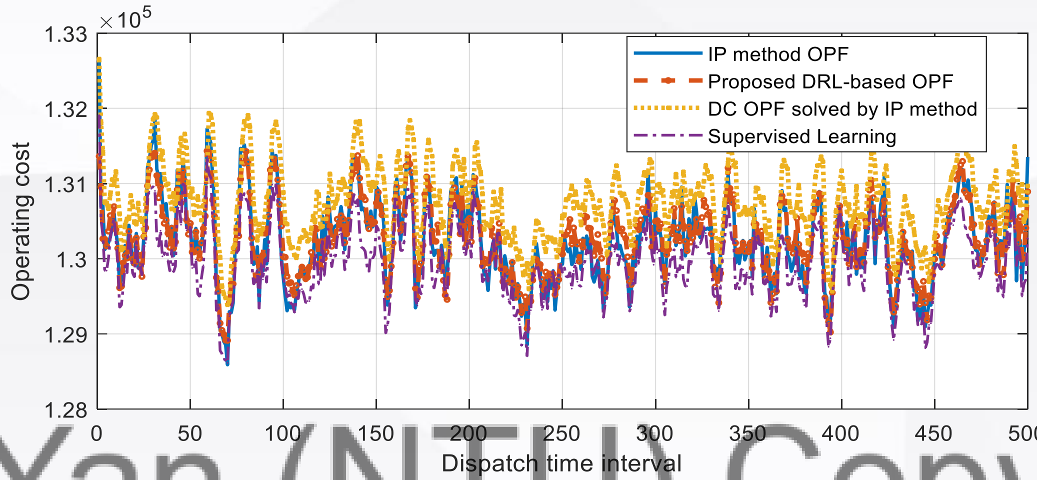
$$\begin{bmatrix} \nabla_{\boldsymbol{a}} L \\ \Delta\lambda \end{bmatrix} \approx \begin{bmatrix} W & G^T \\ G & 0 \end{bmatrix}^{-1} \left[\begin{matrix} -\nabla C(\boldsymbol{a}) - H^T \mu \\ -g(\boldsymbol{a}) \end{matrix} - \begin{pmatrix} H^T \\ 0 \end{pmatrix} \Delta\mu \right]$$

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

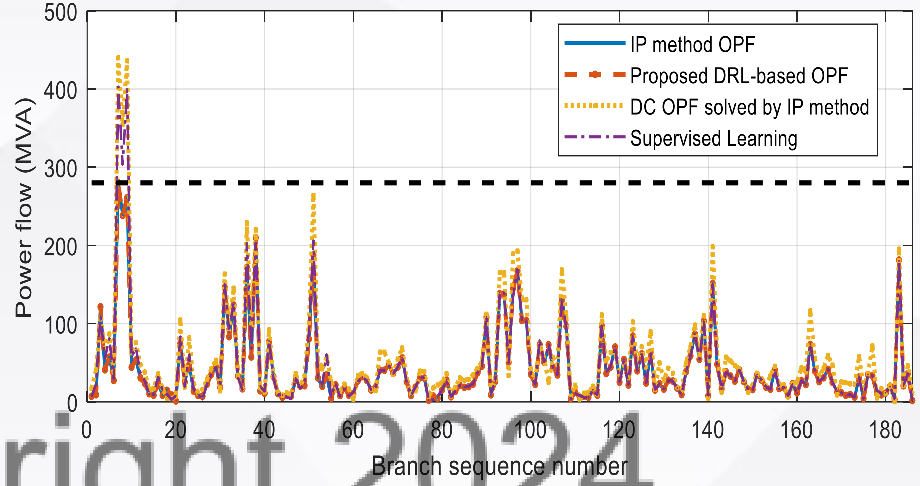
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Real-time computation of optimal power flow (RT-OPF)

- Simulation Results on 118-bus system



Optimality: operating cost comparison of different OPF methods against random load changes



Constraints: branch power flow comparison of different OPF methods

Method	Average generation cost (USD$)	Average absolute errors of P_G (MW)	Inequality Constraints	Average time saving
IP method OPF [73] (benchmark)	1.3018×10^5	0.00	All satisfied	0.0%
DC OPF [74]	1.3076×10^5	0.610	Branch flow and nodal voltage not satisfied	90.1%
Supervised learning [29] 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%

- **Accuracy**
  - Mostly closed to converged IPO;
  - The average cost is similar to IPOPT.
- **Speed**
  - Average 99.8% time saving.
  - 0.000625s. Feasible for real-time applications.
- **Constraints**
  - Satisfied

✓**Best balanced performance**

1. Overview

2. Power Systems

2.1 Frequency control

2.2 Optimal power flow

2.3 Topology optimization

3. Microgrids

3.1 Frequency control

3.2 Controller tuning

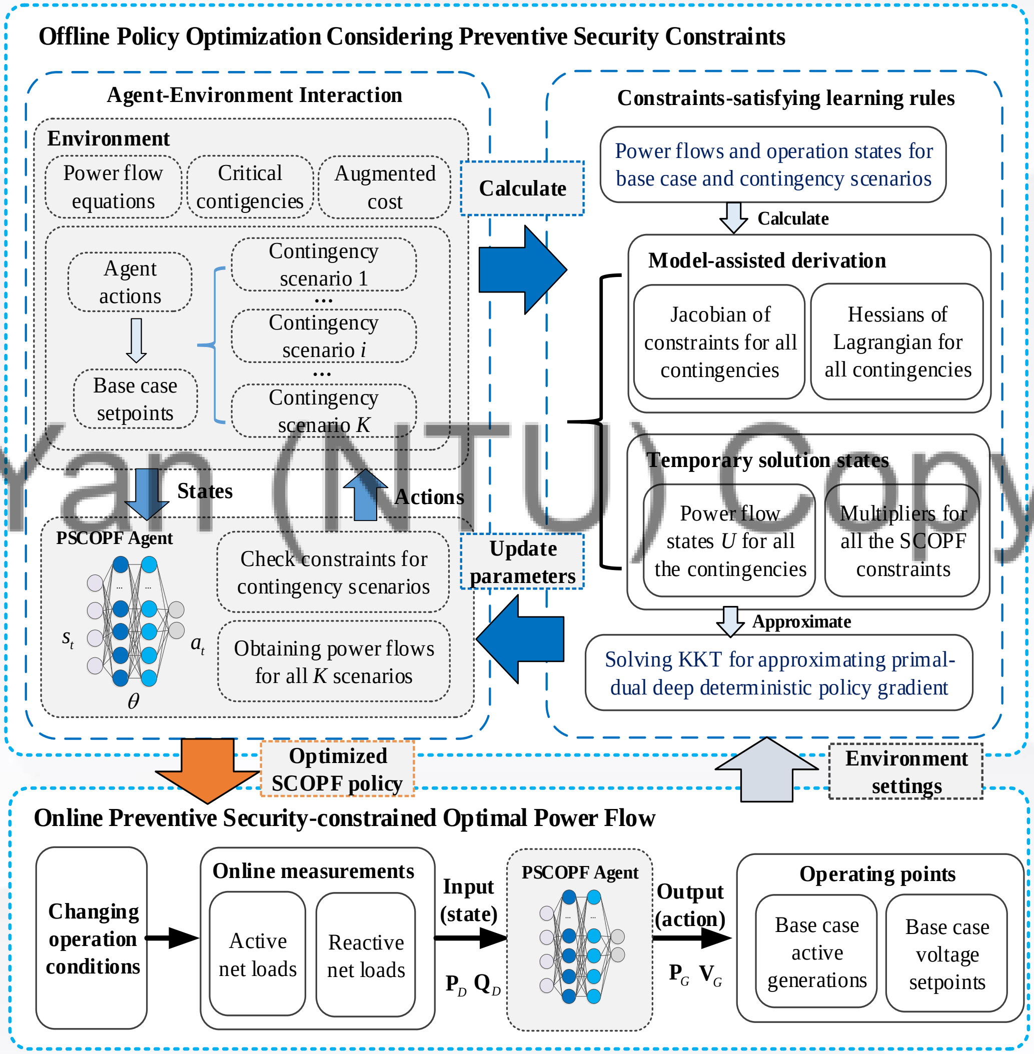
3.3 Energy management

3.4 Volt/Var control

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Hybrid Data-driven Method for Security-Constrained OPF

- Preventive SCOPF through physics-informed safe DRL



Preventive SCOPF Formulation

- Objective considering contingencies:

$$\min_{\mathbf{u}_0} f_0(\mathbf{x}_0, \mathbf{u}_0) \quad \text{Base scenario}$$

$$\mathbf{g}_k(\mathbf{x}_k, \mathbf{u}_0) = \mathbf{0} \quad k = 0, \dots, N_C$$
$$\mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_0) \le \mathbf{0} \quad k = 0, \dots, N_C$$

Contingency scenarios

- Constraints for each contingency scenario:

$$P_{Gi} - P_{Di} = V_i \sum_{j=1}^{n} V_j (G_{ij} \cos\delta_{ij} + B_{ij} \sin\delta_{ij})$$
$$Q_{Gi} - Q_{Di} = V_i \sum_{j=1}^{n} V_j (G_{ij} \sin\delta_{ij} - B_{ij} \cos\delta_{ij})$$

$$P_{Gi} \le P_{Gi}^{\max}$$
$$V_i^{\min} \le V_i \le V_i^{\max}$$
$$|S_{f,i}| \le S_{f,\max}$$

Constrained DRL Formulation

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

$$\begin{cases} U_k = (\boldsymbol{\theta}_k, \mathbf{v}_k, \mathbf{p}_k, \mathbf{q}_k) \\ U = (U_0, U_1, \dots, U_n, \mathbf{v}_g, \mathbf{p}_g) \\ \boldsymbol{L} = \boldsymbol{f}(U) + \lambda^T \boldsymbol{G}(U) + \mu^T[\boldsymbol{H}(U) + \boldsymbol{Z}] - \gamma \sum_{k=1}^{N_\mu} \ln(\mathbf{z}_k) \end{cases}$$

Safe Reinforcement Learning

Z. Yan and Y. Xu, "A Hybrid Data-driven Method for Fast Solution of Security-Constrained Optimal Power Flow," ***IEEE Transactions on Power Systems***, 2022.

Hybrid Data-driven Method for Security-Constrained OPF

- Physics-informed gradient derivation

Constrained DRL Formulation

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

$$\begin{cases} U_k = (\boldsymbol{\theta}_k, \mathbf{v}_k, \mathbf{p}_k, \mathbf{q}_k) \\ U = (U_0, U_1, \ldots, U_n, \mathbf{v}_g, \mathbf{p}_g) \\ \boldsymbol{L} = \boldsymbol{f}(U) + \lambda^T \boldsymbol{G}(U) + \mu^T[\boldsymbol{H}(U) + \boldsymbol{Z}] - \gamma \sum_{\boldsymbol{k}=1}^{N_\mu} \ln(\boldsymbol{z_k}) \end{cases}$$

Constraints w.r.t. variables

Prepare KKT

Formulation of Jacobians and Hessians

- **Formulate sparse Jacobians**

$$J_{local,k} = \begin{bmatrix} \frac{\partial g}{\partial \theta_k} & \frac{\partial g}{\partial v_k} & \frac{\partial g}{\partial p_k} & \frac{\partial g}{\partial q_k} \\ \frac{\partial h}{\partial \theta_{k\ k}} & \frac{\partial h}{\partial v_k} & \frac{\partial h}{\partial p_k} & \frac{\partial h}{\partial q_k} \end{bmatrix} \quad J_{global,k} = \begin{bmatrix} \frac{\partial g}{\partial v_{g,k}} & \frac{\partial g}{\partial p_{g,k}} \\ \frac{\partial h}{\partial v_{g,k}} & \frac{\partial h}{\partial p_{g,k}} \end{bmatrix}$$

- **Spare Jacobians of Constraints**

$$J = \begin{bmatrix} 0_{m\times 0} & J_{local,0} & 0_{m\times(Nc-1)n} & J_{global,0} \\ & & \cdots & \\ 0_{m\times kn} & J_{local,k} & 0_{m\times(Nc-1-k)n} & J_{global,k} \\ & & \cdots & \\ 0_{m\times(N_C-1)n} & J_{local,(N_C-1)} & 0_{m\times 0} & J_{global,(N_C-1)} \end{bmatrix}$$

- **Spare Hessian of Lagrangian**

$$H = \begin{bmatrix} 0_{n\times 0} & H_{l0,l0} & 0_{n\times(Nc-1)n} & H_{l0,global} \\ & & \cdots & \\ 0_{n\times kn} & H_{lk,lk} & 0_{n\times(Nc-1-k)n} & H_{lk,global} \\ & & \cdots & \\ 0_{n\times(Nc-1)n} & H_{l(Nc-1),l(Nc-1)} & 0_{n\times 0} & H_{l(Nc-1),global} \\ H_{global,l0} & \cdots & H_{global,lk} & \cdots \quad H_{global,global} \end{bmatrix}$$

Prepare KKT

Constraints-satisfying updating rule

KKT conditions

$$\nabla_U \boldsymbol{L} = \boldsymbol{0}$$

$$\boldsymbol{G}(\boldsymbol{U}) = \boldsymbol{0}$$

$$\boldsymbol{H}(\boldsymbol{U}) + \boldsymbol{Z} = \boldsymbol{0}$$

$$[\mu]\boldsymbol{Z} - \gamma \boldsymbol{e} = 0$$

Newton's step

$$\Delta \boldsymbol{Z} = -\boldsymbol{H}(\boldsymbol{U}) - \boldsymbol{Z} - \nabla \boldsymbol{H}(\boldsymbol{U})^T \Delta \boldsymbol{U}$$

$$\Delta\mu = -\mu + [\boldsymbol{Z}]^{-1}(\gamma \boldsymbol{e} - [\mu]\Delta \boldsymbol{Z})$$

$$\begin{bmatrix} \Delta U \\ \Delta\lambda \end{bmatrix} \approx \begin{bmatrix} M & G \\ G^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} -N \\ -G \end{bmatrix}$$

$$M = \nabla^2_{UU} L + \nabla_U H[Z]^{-1}[\mu]\nabla_U H^T$$

$$N = \nabla_U L + \nabla_U H[Z]^{-1}([\mu]H + \gamma e)$$

Hybrid Data-driven Method for Security-Constrained OPF

- Physics-informed gradient derivation

Model-assisted gradient derivation

$$\Delta \boldsymbol{Z} = -\boldsymbol{H}(\boldsymbol{U}) - \boldsymbol{Z} - \nabla \boldsymbol{H}(\boldsymbol{U})^T \Delta \boldsymbol{U}$$

$$\Delta \mu = -\mu + [\boldsymbol{Z}]^{-1}(\gamma \boldsymbol{e} - [\mu]\Delta \boldsymbol{Z})$$

$$\begin{bmatrix} \Delta U \\ \Delta \lambda \end{bmatrix} \approx \begin{bmatrix} M & G \\ G^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} -N \\ -G \end{bmatrix}$$

$$M = \nabla_{UU}^2 L + \nabla_U H[Z]^{-1}[\mu]\nabla_U H^T$$

$$N = \nabla_U L + \nabla_U H[Z]^{-1}([\mu]H + \gamma e)$$

$$\nabla_{\boldsymbol{a}} L = \nabla_U L \setminus \nabla_{U \setminus \{\boldsymbol{a}\}} L \approx \Delta a$$

Actor gradients are solved along with the updating of SCOPF optimization variables

Modifying DDPG

Model-assisted gradient derivation

$$\theta^{k+1} = \theta^k - \eta \nabla_{\boldsymbol{a}} L \cdot \nabla_\theta \boldsymbol{a}$$

$$\lambda^{k+1} = \lambda^k + \eta \Delta \lambda$$

$$\mu^{k+1} = \mu^k + \eta \Delta \mu$$

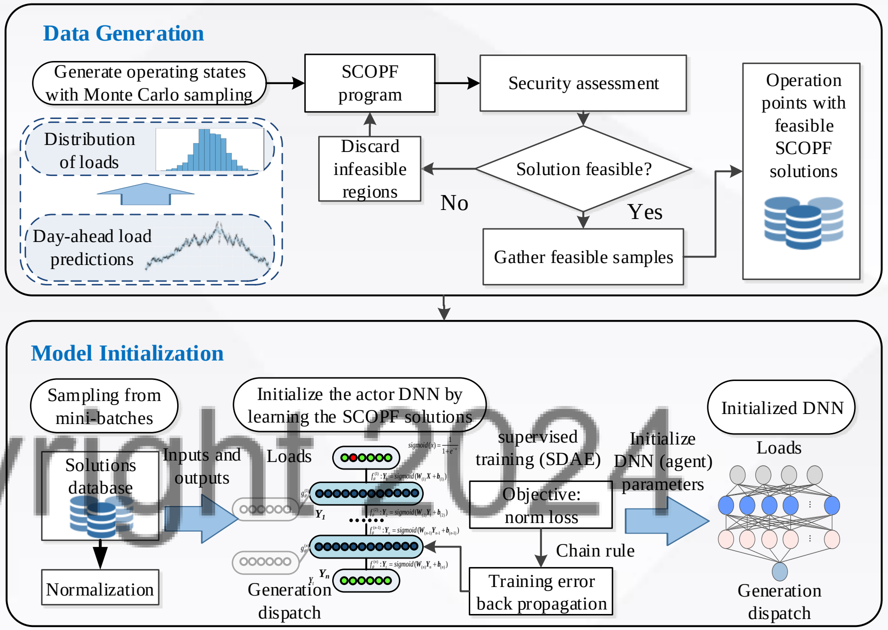
Update neural networks for descending augmented costs

Improved agent updating rule

$$W^{(l,k+1)} = W^{(l,k)} - \eta \frac{1}{m} \sum_{k=r}^{r+m} \nabla_{\boldsymbol{a}} L^\mu(s,\boldsymbol{a}) \frac{\partial}{\partial W^{(l,T)}} \mu^{(k)}(\boldsymbol{W},\boldsymbol{b})$$

$$b^{(l,k)} = b^{(l,k)} - \eta \frac{1}{m} \sum_{k=r}^{r+m} \nabla_{\boldsymbol{a}} L^\mu(s,\boldsymbol{a}) \frac{\partial}{\partial b^{(l,T)}} \mu^{(k)}(\boldsymbol{W},\boldsymbol{b})$$

Tricks to improve performance



Model initialization process

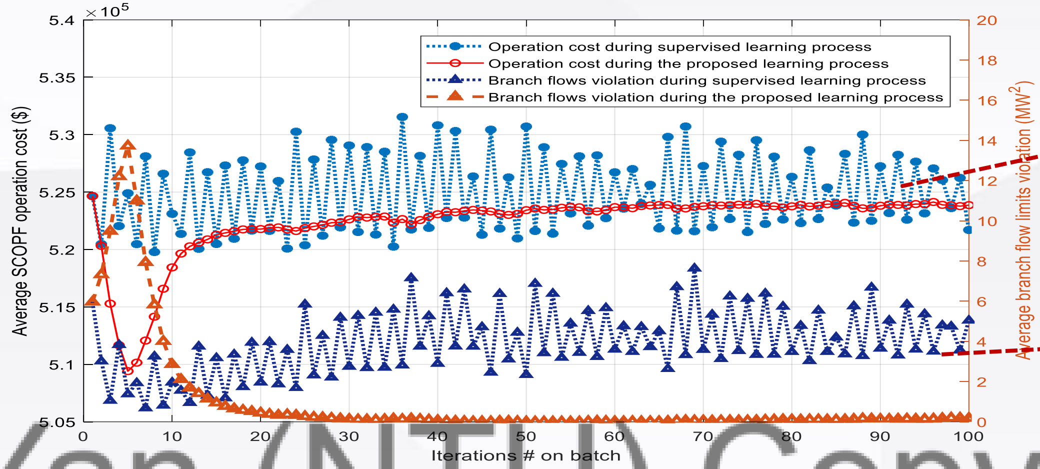
Data generation:
Monte Carlo Simulation to generate day-ahead distribution of loads. The states with infeasible SCOPF solutions are discarded.

Model initialization:
Initialize the DRL agent by supervised learning with SDAE (using the OPF output to train the DNN). Initialize network by minimizing differences of solutions.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Hybrid Data-driven Method for Security-Constrained OPF

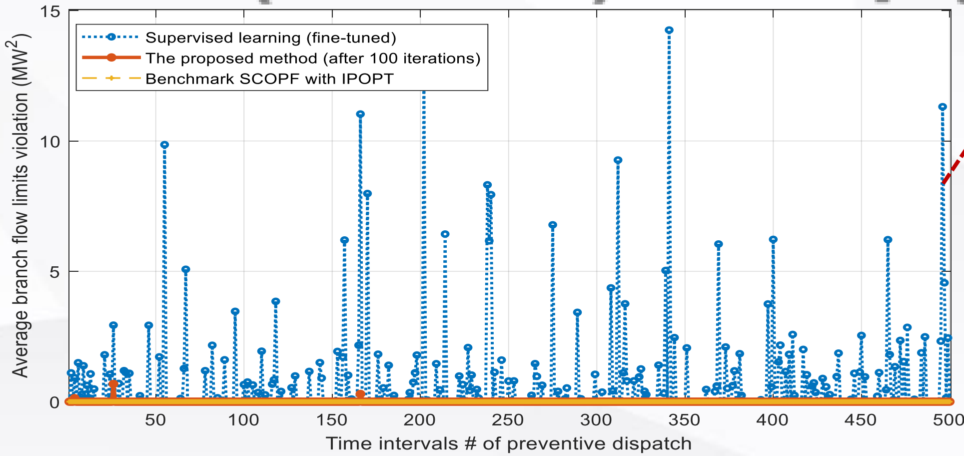
- Testing on 57-bus system



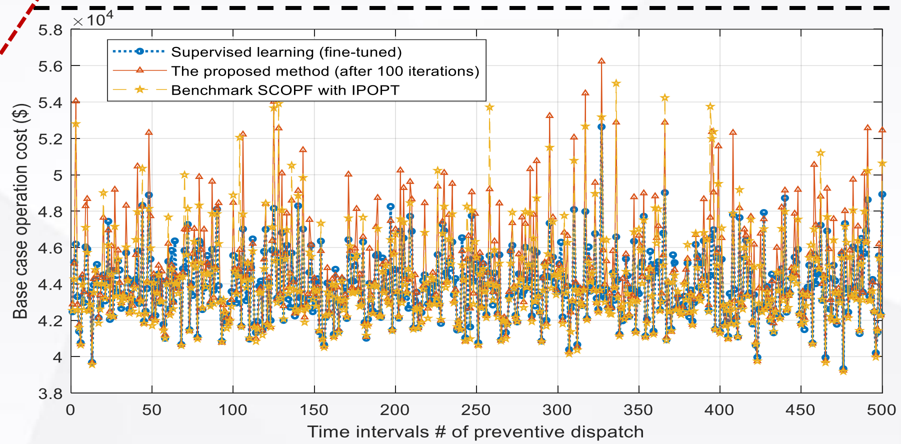
The comparison of convergence characteristics for the 57-bus test system with the same initializations

Discussion of Results

- Supervised learning (OPF output) for training DNN gets **oscillated costs** (after fine-tuned initialization). The physical model-based gradients tend to be more stable.
- Supervised learning (OPF output) for training DNN does **not consistently reduce the constraints violation**. The physical model-based gradients tend to minimize constraints violation.
- The supervised learning sometimes satisfies the branch flow constraints while sometimes violates.



The comparison of branch flow limits violation with respect to different samples obtained by different methods on 57-bus system



The comparison of operation cost with respect to different samples obtained by different methods on 57-bus system

OPF with Linguistic Stipulations – Problem Description

- **Actual Power Systems are 'Human-in-the-loop'**

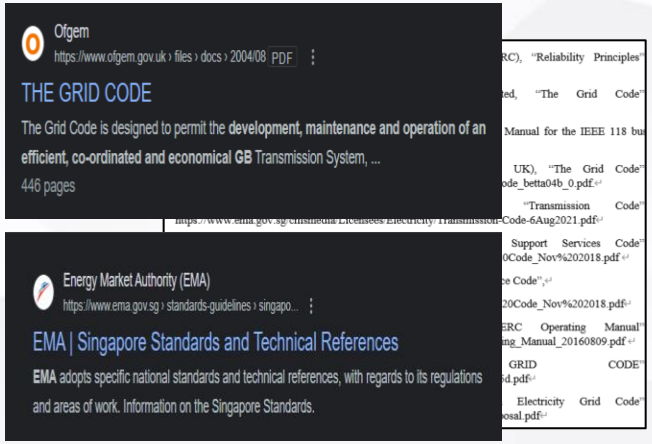
Human Operators in the Decision-Making Loop for Power Systems

➢ **Human > Algorithm**

Human oversight in:

- Interpreting regulations
- Making decisions
- Implementing corrective actions to ensure operational safety.

- **Grid Code and Operation Manual**



"Guide"

- **Linguistic stipulations**
- **Difficult to model.**

Language-based Standard that specify the performance of operation

➢ **Human-in-the-Loop**

- Human expertise interprets power system operation in the Grid Codes.

➢ **Informed-Decisions**

- Ensuring compliance and safety in operational practices following standards

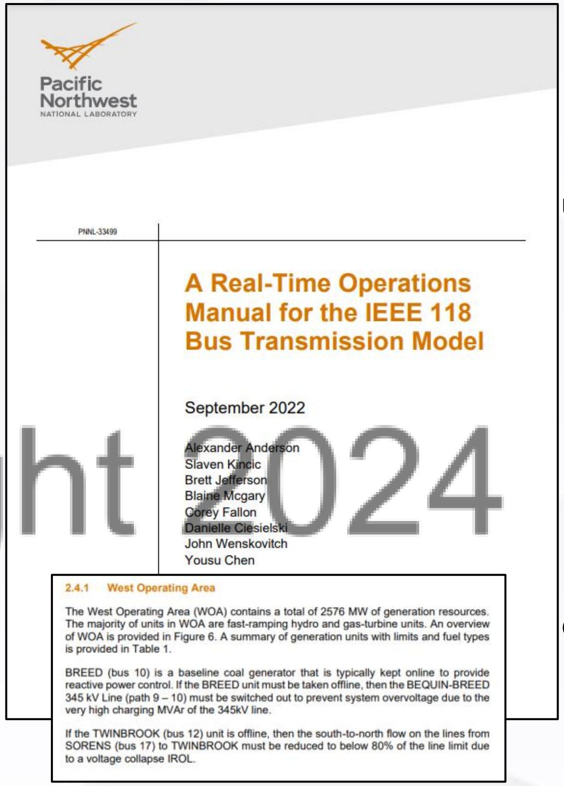
Source of pictures: website (searched in Google)

Z. Yan and Y. Xu, "Real-Time Optimal Power Flow with Linguistic Stipulations: Integrating GPT-Agent and Deep Reinforcement Learning," ***IEEE Transactions on Power Systems***, 2023.

OPF with Linguistic Stipulations – Problem Description

- **Grid Code and Operation Manual**

Country	Grid Code or Operation Manual	Ref.
UK	The Grid Code, documented by National Grid Electricity System Operator Limited and Ofgem	[2] [R1]
Italy	ITALIAN Grid Code, documented by TERNA	[R2]
India	Indian Electricity Grid Code, documented by Central Electricity Regulatory Commission	[R3]
US	Operation Manual for the IEEE 118 bus system documented by the Pacific Northwest National Laboratory	[3]
US	Operating Manual for general operation of interconnected systems documented by the North American Energy Standards Board (NAESB).	[R4]
Singapore	Transmission Code, which outlines the conditions that the Transmission Licensee must meet.	[R5]
Singapore	Market Support Services Code, which defines the standards of service provider performance	[R6]
Singapore	Regulated Supply Service Code, which outlines the requirements for market support services	[R7]


Pacific Northwest NATIONAL LABORATORY

PNNL-33499

A Real-Time Operations Manual for the IEEE 118 Bus Transmission Model

September 2022

Alexander Anderson
Slaven Kincic
Brett Jefferson
Blaine Mcgary
Corey Fallon
Danielle Ciesielski
John Wenskovitch
Yousu Chen

2.4.1 West Operating Area

The West Operating Area (WOA) contains a total of 2576 MW of generation resources. The majority of units in WOA are fast-ramping hydro and gas-turbine units. An overview of WOA is provided in Figure 6. A summary of generation units with limits and fuel types is provided in Table 1.

BREED (bus 10) is a baseline coal generator that is typically kept online to provide reactive power control. If the BREED unit must be taken offline, then the BEQUIN-BREED 345 kV Line (path 9 – 10) must be switched out to prevent system overvoltage due to the very high charging MVAr of the 345kV line.

If the TWINBROOK (bus 12) unit is offline, then the south-to-north flow on the lines from SORENS (bus 17) to TWINBROOK must be reduced to below 80% of the line limit due to a voltage collapse IROL.

Prior experience of operation under different scenarios

↓

Linguistic Stipulations

↓

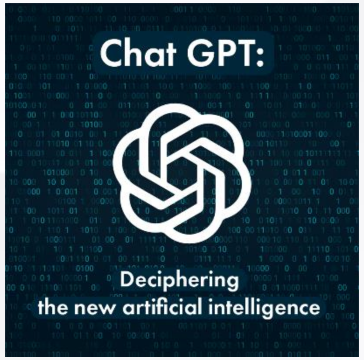
Formulate context (input) for GPT-Agent

[1] The North American Electric Reliability Corporation (NERC), “Reliability Principles”, https://www.nerc.com/pa/Stand/Pages/default.aspx. 2023.
[2] National Grid Electricity System Operator Limited, “The Grid Code”, https://www.nationalgrideso.com/document/162271/download, 2023.
[3] Pacific Northwest National Laboratory, “A Real-Time Operation Manual for the IEEE 118 bus Transmission Model”, PNNL-33499, pp. 28-30, 2022.
[R1] The Office of Gas and Electricity Markets (Ofgem, UK), “The Grid Code”, https://www.ofgem.gov.uk/sites/default/files/docs/2004/08/7885-grid_code_betta04b_0.pdf.
[R2] TERNA, “ITALIAN GRID CODE”, https://download.terna.it/terna/Chapter_1_Section_1B_8db5644575f445d.pdf
[R3] Central Electricity Regulatory Commission, “Indian Electricity Grid Code”, https://cercind.gov.in/2010/ORDER/February2010/IEGC_Review_Proposal.pdf
[R4] Electric Reliability Organization Enterprise, “NERC Operating Manual”, https://www.nerc.com/comm/OC/Operating%20Manual%20DL/Operating_Manual_20160809.pdf
[R5] Energy Market Authority of Singapore, “Transmission Code”, https://www.ema.gov.sg/cmsmedia/Licensees/Electricity/Transmission-Code-6Aug2021.pdf
[R6] Energy Market Authority of Singapore, “Market Support Services Code”, https://www.ema.gov.sg/cmsmedia/Market%20Support%20Services%20Code_Nov%202018.pdf
[R7] Energy Market Authority of Singapore, “Regulated Supply Service Code”,
https://www.ema.gov.sg/cmsmedia/Regulated%20Supply%20Service%20Code_Nov%202018.pdf

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

OPF with Linguistic Stipulations – Proposed Method

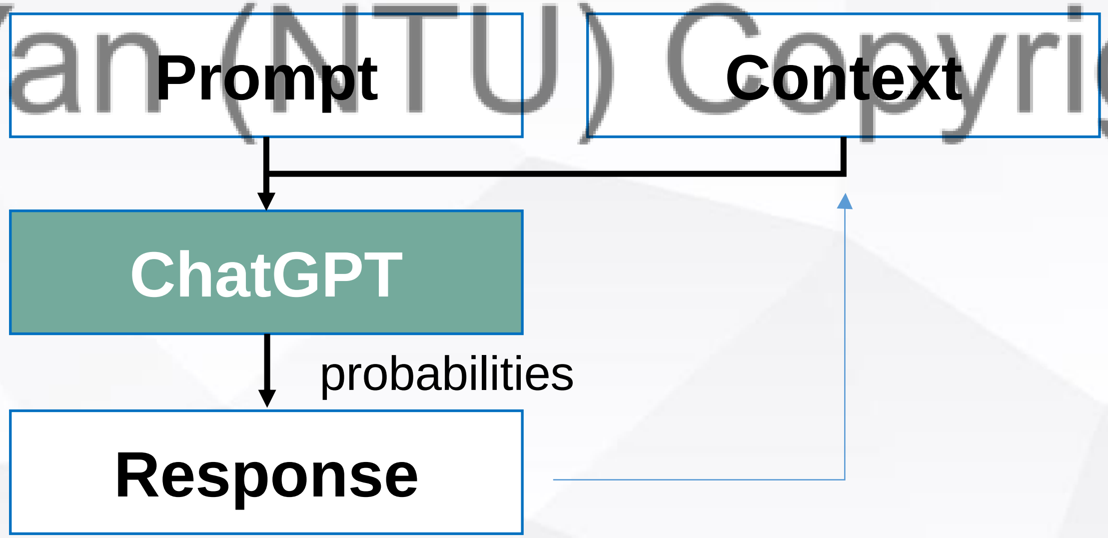
- **Recent breakthrough of Large Language Model**



ChatGPT is a chatbot developed by OpenAI based on a large language model to produce text outputs.

Essence of ChatGPT: Probabilistic model that analyses texts.

- Probabilistic: Answer is generated based on maximum likelihood.
- Text: any questions; any requirements; any text formats.
- Model: given a Context and Question, provide the Answer.



"Sampling via conditional probability"

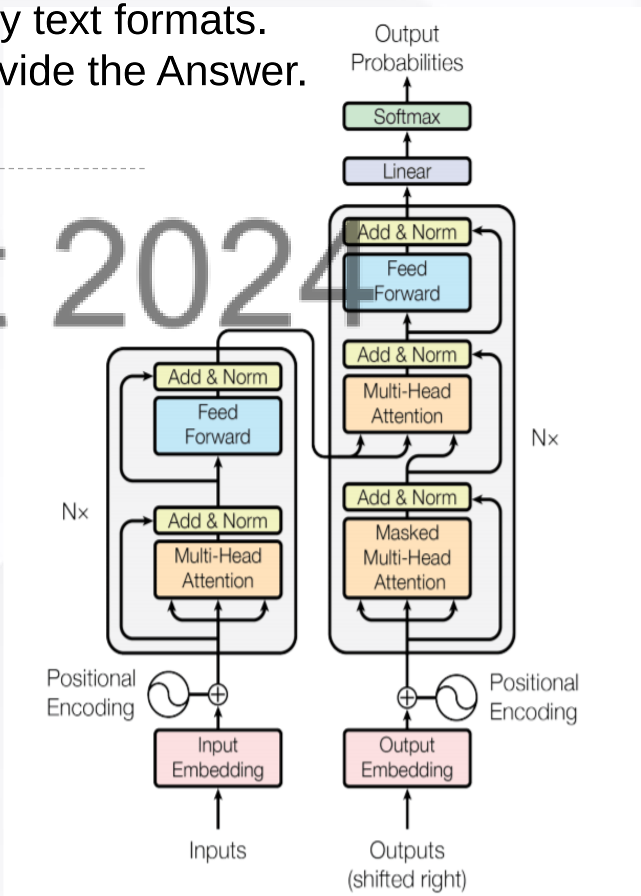
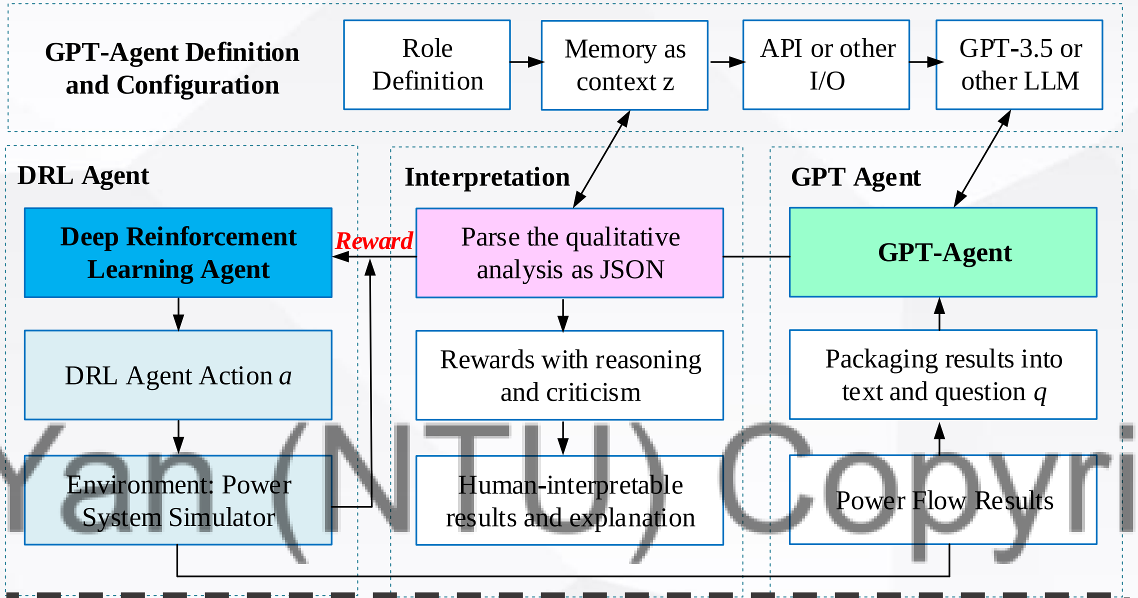


Figure 1: The Transformer - model architecture. **[1]**

[1] https://machinelearningmastery.com/the-transformer-model/

OPF with Linguistic Stipulations – Proposed Method

▪ Mathematical Modeling



GPT Agent is a Generative Model: Sample from a probability distribution under a context (conditional probability).

$$\mathbf{Pr}(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} \mathbf{Pr}(x_i \mid x_1, \dots, x_{i-1})$$

Parse the Results of GPT-Agent: process string as **JSON**, then process JSON as dictionary with multiple keys and values

$$C_{Qj}(P_{Gi}^t, z, q_j) = parse_{CQ}\{x_{i+1}, \dots, x_{i+n} \mid a, s, z, q_j, a_{k,j}\}$$

$$R_{\text{target}} = [-\sum_{i=1}^{N_G} C_{Gi}(P_{Gi}^t) - w_j \sum_{j=1}^{N_Q} C_{Qj}(a, s, z, q_j, a_{k,j})]$$

RT-OPF Formulation

OPF with **linguistic stipulations**

$$\min \sum_{i=1}^{N_G} C_{Gi}(P_{Gi}^t) + \sum_{j=1}^{N_Q} w_j C_{Qj}(P_{Gi}^t, Z, Q_j, A_j)$$

Satisfying operation constraints

$$P_{Gi} - P_{Di} = V_i \sum_{j=1}^{n} V_j (G_{ij}\cos\delta_{ij} + B_{ij}\sin\delta_{ij})$$

$$Q_{Gi} - Q_{Di} = V_i \sum_{j=1}^{n} V_j (G_{ij}\sin\delta_{ij} - B_{ij}\cos\delta_{ij})$$

$$\max[P_{Gi}^{\min}, P_{Gi}^{t-1} - R_{Gi}^{down}] \le P_{Gi}^t \le \min[P_{Gi}^{\max}, P_{Gi}^{t-1} + R_{Gi}^{up}]$$

$$V_i^{\min} \le V_i \le V_i^{\max}$$

$$|V_i V_j (G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}) - V_i^2 G_{ij}| \le L_{ij}^{\max}$$

$$A_{k,i}(a, s, z, q_i) \le A_{k,i,\max}(s, z, q_i)$$

Constrained DRL Formulation

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

Primal-dual safe reinforcement learning

$$\boldsymbol{L} = -\boldsymbol{R}(\boldsymbol{s}_i, \boldsymbol{a}_i, \theta, \boldsymbol{z}, \boldsymbol{q}_i) + \lambda \boldsymbol{C}(\boldsymbol{s}_i, \boldsymbol{a}_i, \theta, \boldsymbol{z}, \boldsymbol{q}_i)$$

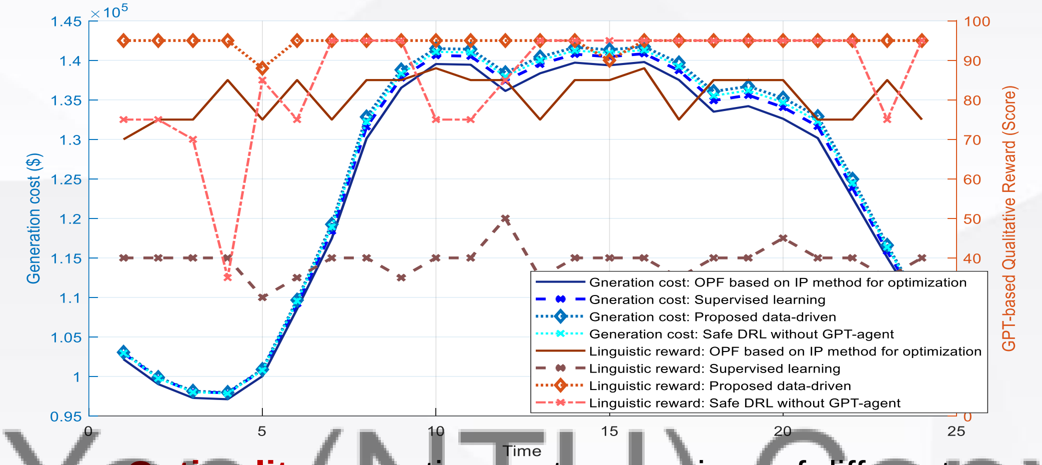
Z. Yan and Y. Xu, "Real-Time Optimal Power Flow with Linguistic Stipulations: Integrating GPT-Agent and Deep Reinforcement Learning," ***IEEE Transactions on Power Systems***, 2023.

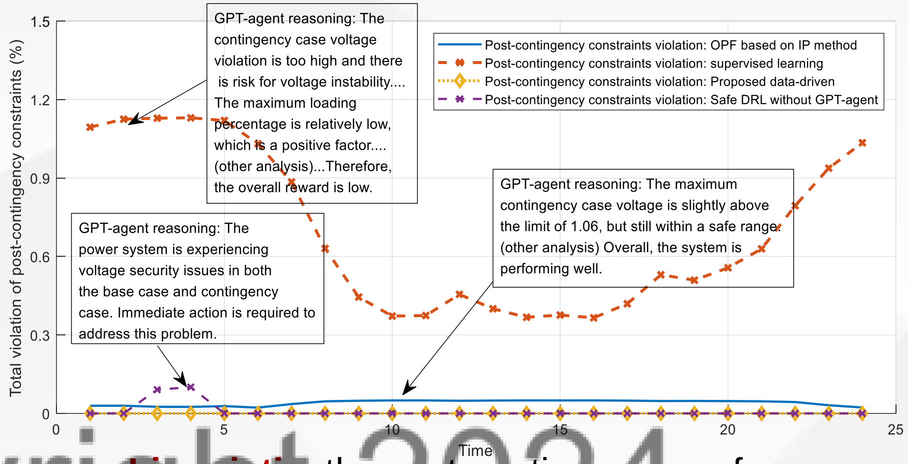
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Overview
2. Power Systems
2.1 Frequency control
2.2 Optimal power flow
2.3 Topology optimization
3. Microgrids
3.1 Frequency control
3.2 Controller tuning
3.3 Energy management
3.4 Volt/Var control

Simulation Results

- **Simulation Results on 118-bus system**





Optimality: operating cost comparison of different OPF methods against random load changes

Linguistic: the post-contingency performance are interpreted by GPT-agent

Method	Average generation costs (USD$)	Average performance score evaluated by GPT-Agent (linguistic reward)	Average contingency constraints violation (%)	Qualitative objectives
OPF based on IP method for optimization (benchmark)	1.2393e5	80.87	0.0404	No
Supervised learning	1.2532e5	39.58	0.6963	No
Proposed method	1.2575e5	94.50	0.0000	Yes
Safe DRL without GPT-agent	1.2540e5	85.63	0.0080	No

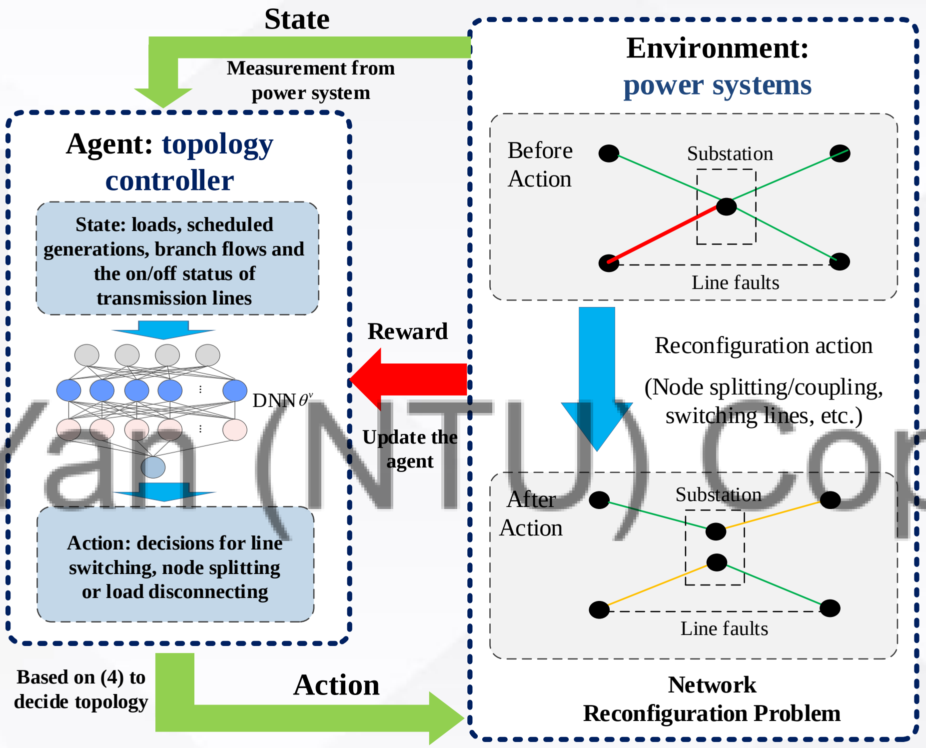
- **Performance**
  - Highest average score considering costs and satisfaction of linguistic stipulations;
  - Slightly higher costs than benchmark optimization.
- **Speed**
  - Average 99.8% time saving.
  - 0.000625s. Feasible for real-time applications.

✓Best balanced performance

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

DRL for Power Grid Topology Optimization

- DRL for Topology Optimization



Environment: Learning to Run a Power Network

Network reconfiguration:
Reconfigure the topology for power network: lines switching, substations busbars splitting and coupling.

Maximize the remaining transfer capabilities for all time (sometimes after contingencies)

$$Maximize \quad \sum_{i}^{N}[1-(S_{Li}/S_{Lm,i})^2] \quad \textbf{(7.1a)}$$

$$\text{s.t. } P_{Gi}-P_{Di}=V_i\sum_{j=1}^{n}V_j(G_{ij}\cos\delta_{ij}+B_{ij}\sin\delta_{ij}) \quad (7.1b)$$

$$Q_{Gi}-Q_{Di}=V_i\sum_{j=1}^{n}V_j(G_{ij}\sin\delta_{ij}-B_{ij}\cos\delta_{ij}) \quad (7.1c)$$

DQN Agent

$$r(s,a)=\begin{cases}\text{Penalty } (-r_e), & \text{power flow diverges}\\ \frac{1}{N}\sum_{i}^{N} max[0,1-(S_{Li}/S_{Lm,i})^2], & \text{otherwise}\end{cases} \quad \textbf{(7.2)}$$

$$Q^{\pi}(s,a)\leftarrow Q^{\pi}(s,a)+\eta*[r+\gamma*\max Q^{\pi}(s',a')-Q^{\pi}(s,a)] \quad \textbf{(7.3)}$$

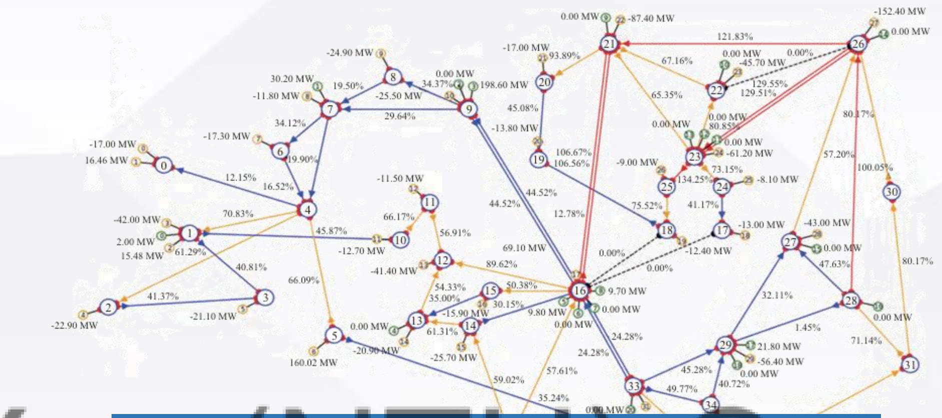
$$\mathcal{L}(\theta)=\sum_{i}[(r_t+\gamma\, max\, Q^{\pi}(s_{t+1},a_{t+1};\theta^{-})-Q^{\pi}(s_t,a_t;\theta))^2] \quad (7.4)$$

Z. Yan, Y. Xu, "Topology Optimization of Power Systems Combining Deep Reinforcement Learning and Domain Knowledge," Automation of Electric Power Systems 46 (1), 60-68, 2022. (In Chinese)

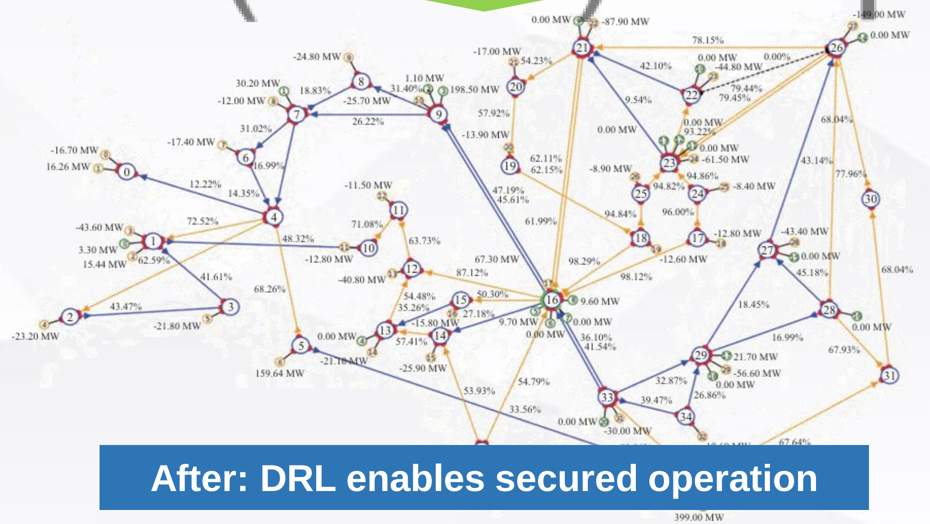
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

DRL for Power Grid Topology Optimization

- Topology Optimization Results: IEEE WCCI Competition



Before: branch flow constraint violations



After: DRL enables secured operation

Much better performance than exhaustion and no control with 45.5% less costs, while saving 92.3% computation time.

Method	Blackout Cost/$	Operation Cost/$	Total Cost/$	Improvement /%
Proposed	0.0	4.738e6	4.738e6	97.17%
No Control	1.650e8	2.212e6	1.672e8	0.0
Exhaustion	8.723e7	3.853e6	9.108e7	45.53%

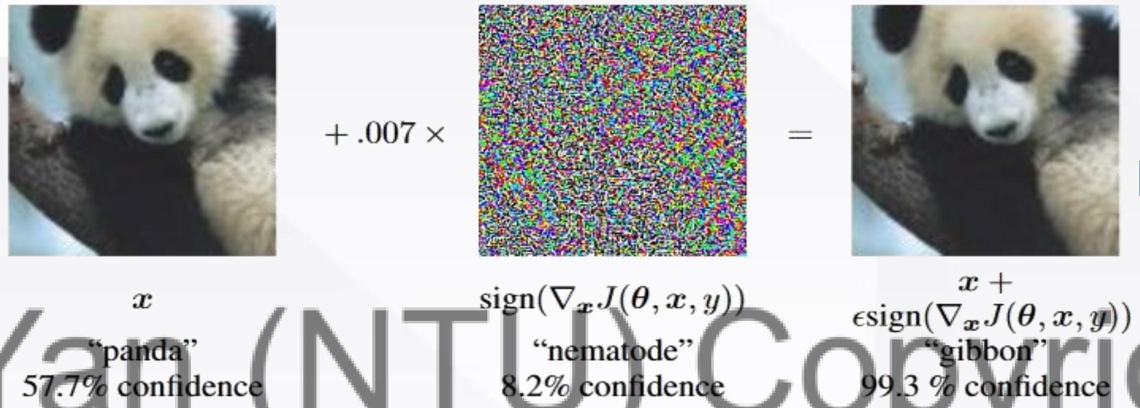
Recognition of methods by Global Competitions

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

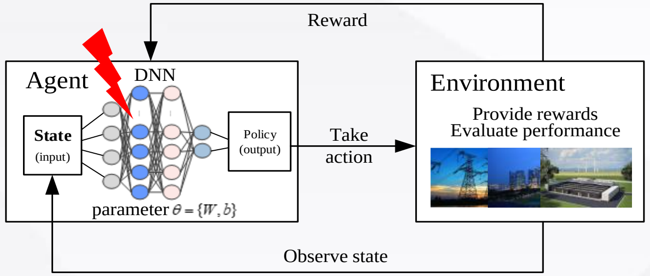
Vulnerability Assessment of DRL Model

- **Motivation**

The previous Chapters are mainly based on deep reinforcement learning methods, whose performance partially inherits from deep neural networks.



Vulnerability of deep neural networks [1]



The DRL agent may be **mislead** by small perturbations (i.e., adversarial attacks) of input data and induce danger

It is important to evaluate and mitigate the security risks of DRL models in power systems before deploying the DRL models in real grids.

- **Reviews**

Power systems community
- Many recent attempts for leveraging DRL into power systems.
- Very limited works have investigated the vulnerability of the DRL models in power system.
- State estimator not sufficient against attacks.

DRL research community
- The performance of DRL agent inherits from deep neural networks (DNN).
- Practically verified that adding small perturbations in the input data may lead to drastically different control actions for DQN.
- DRL model's behaviour can be intriguing and not always predictable

[1] Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572. 2014 Dec 20.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Vulnerability Assessment of DRL Model

- Identify the vulnerabilities

Construct adversarial perturbations

Adversarial perturbation to mislead the DRL model to make a wrong action

$$s'_t = s_t + \lambda \quad (7.6)$$

$$a'_t = argmax\ Q^{\pi}(s'_t|\theta) \quad (7.7)$$

$$\|\lambda\|_p \leq \epsilon \quad \text{s.t.} \quad \pi_\theta(s) \neq \pi_\theta(s+\lambda) \quad (7.8)$$

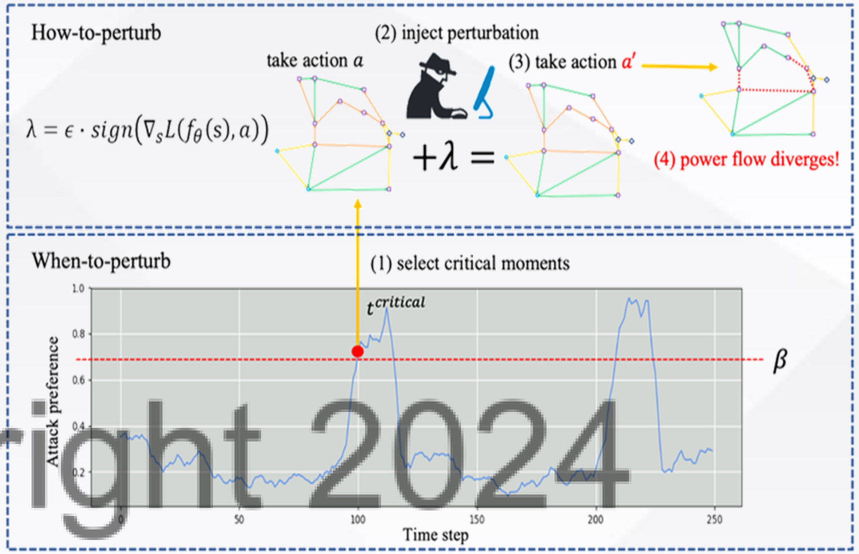
Optimal attack strategy with maximum effects

$$\min_{k_0,\dots,k_{T-1},\lambda_0,\dots,\lambda_1} \sum_{t=0}^{T-1} \mathbb{E}_{a\sim\pi(s_t+k_t\lambda_t)} \gamma^t r_t \quad (7.10a)$$

$$\text{s.t.} \quad \|\lambda_t\|_p \leq \epsilon \quad \text{for all } t \quad (7.10b)$$

$$\sum_{t=0}^{T-1} k_t \leq N \quad (7.10c)$$

When-to-perturb: criticality-based timing



The value of vulnerability indices with respect to time under FGSM attack (actual index value divided by average value for normalization)

The attacker-preference is calculated:

$$p(s_t) = max_{a_t \in A}\pi(s_t|a_t) - max_{a_t \in A\backslash\{a^*\}}\pi(s_t|a_t) \quad (7.11)$$

How-to-perturb: gradient-based perturbation

The fast gradient sign method (FGSM) is leveraged:

$$s^{adv} = s + \epsilon \cdot sign\left(\nabla_s L(\pi_\theta(s), \vec{a})\right) \quad (7.12)$$

Y. Zheng, Z. Yan, K. Chen, J. Sun, Y. Xu and Y. Liu, "Vulnerability Assessment of Deep Reinforcement Learning Models for Power System Topology Optimization," ***IEEE Transactions on Smart Grid***, 2021.

Vulnerability Assessment of DRL Model

- Vulnerability index: to evaluate the risks of DRL models in power systems against the attack and data perturbation.

Probability-based criteria

- Overall vulnerability

failure rate, expected performance decay (EPD), expected performance decay rate (EPDR)

$$EPD = \frac{1}{M}\sum_{i}^{M} \pi_i(s_i')[R_i(s_i|\theta^v) - R_i'(s_i'|\theta^v)] \quad (7.13)$$

$$EPDR = \frac{1}{M}\sum_{i}^{M} \pi_i(s_i')[1 - R_i'(s_i'|\theta^v)/R_i(s_i|\theta^v)] \quad (7.14)$$

where,
Failure rate N_v/N_t: percentage of diverged power flow solutions
$\pi_i(s_i')$: the probability of i-th abnormal state s_i' to happen
$R_i(s_i|\theta^v)$, $R_i'(s_i'|\theta^v)$ the control rewards for environment state before and after perturbations:

The indices quantify **the overall DRL performance** under massively sampled datasets.

Gradient-based criteria

- Operational vulnerability

p-function: risks of being mislead.
Gradient saliency: sensitivity of DNN to perturbations

$$p(s_t) = max_{a_t \in A}\pi(s_t|a_t) - max_{a_t \in A\backslash\{a^*\}}\pi(s_t|a_t) \quad (7.16)$$

$$GS(s_t) = \frac{1}{N}\sum_{i}^{N} \left|\frac{\partial L(f_\theta(x), a)}{\partial x_i}\right| \quad (7.15)$$

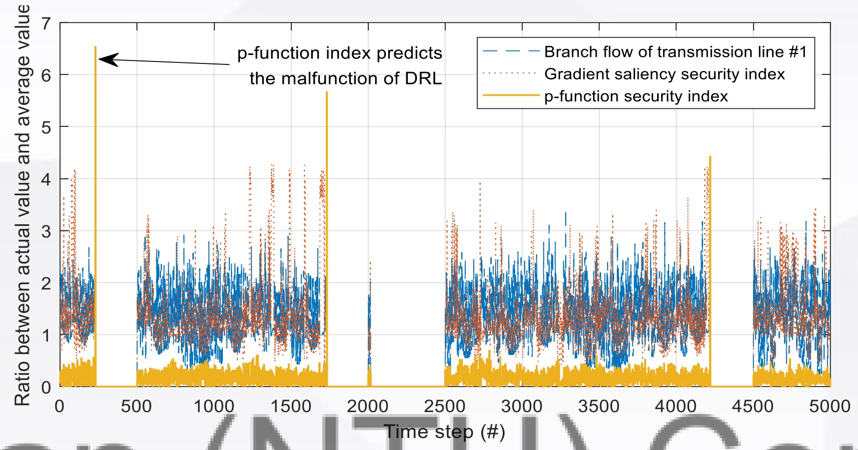
where,
$L(f_\theta(x), a)$: the DRL training objective function.
x_i : the i-th state variable.
A: the action set.
$a^* = max_{a_t \in A}\pi(s_t|a_t)$ is the optimal action with the highest future return

The indices quantify **the importance of control actions** and the **sensitivity of neural networks** to perturbations under certain states

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Vulnerability Assessment of DRL Model

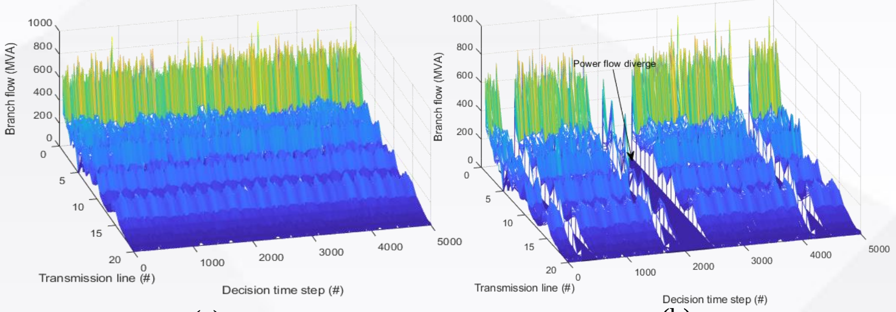
- Assess the vulnerabilities: 14-bus system



The value of **vulnerability indices** with respect to time under FGSM attack (actual index value divided by average value for normalization)

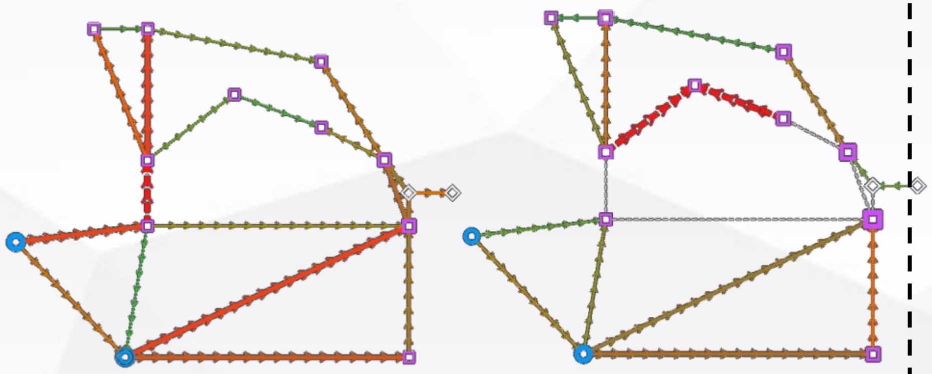
Method	CAR = N_v/N_t	EPDR = EPD / A.S	EPDR (C/Non-C)	A.T
No attack	0% = 0/0	0% = 0 / 3347.1	0 / 0	0
Random Noise	0.021% = 5/23709	4.6% = 823 / 3157.7	4.6% / 0%	0.2ms
FGSM (every step)	0.101% = 20/19817	32.1% = 1073.2 / 2273.9	31.4% / 0.7%	2.2ms
Critical attack (chosen step)	0.386% = 18/4658	32.1% = 1072.6 / 2274.5	31.3% / 0.8%	3.1ms

The performance of DRL with and without attacks



(a) (b)

DRL-based control branch flows results with respect to time in competition 14-bus system against FGSM attack



(a) (b)

Topology decisions based on DRL (a) before and (b) after deploying perturbations (attacks) for IEEE 14-bus system at time 18:55 of 03, Jan

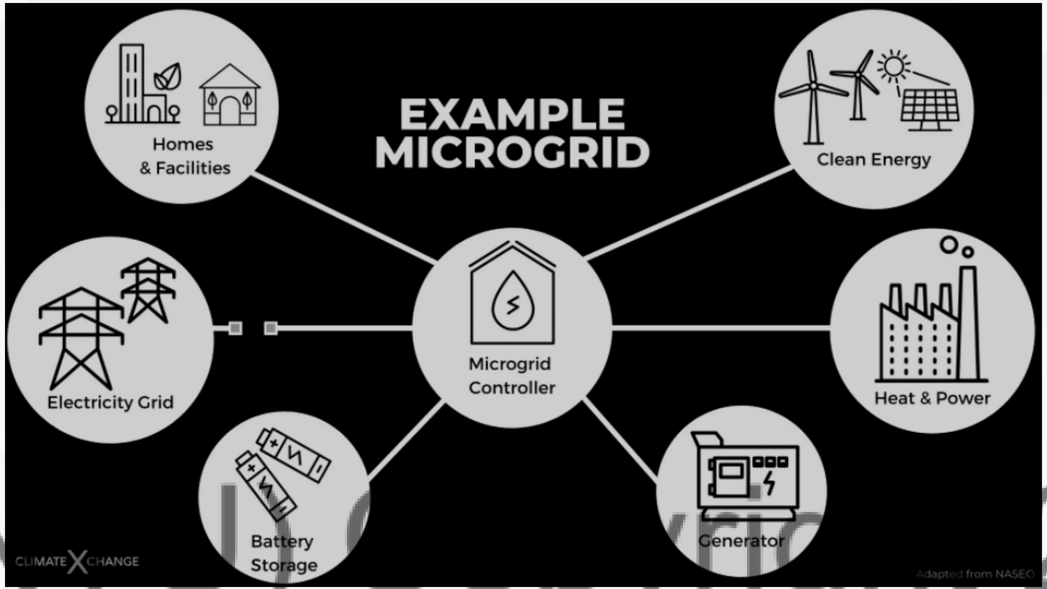
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1 **Overview**
- Research Background
- Preliminaries of DRL
- Our Research Framework

2 **DRL for Bulk Power Grids**
- Load Frequency Control
- Real-Time Optimal power flow
- Network Reconfiguration

3 **DRL for Microgrids & Active Distribution Grids**
- Frequency Control
- Control Parameter Scheduling
- Energy Management
- Volt/Var Control

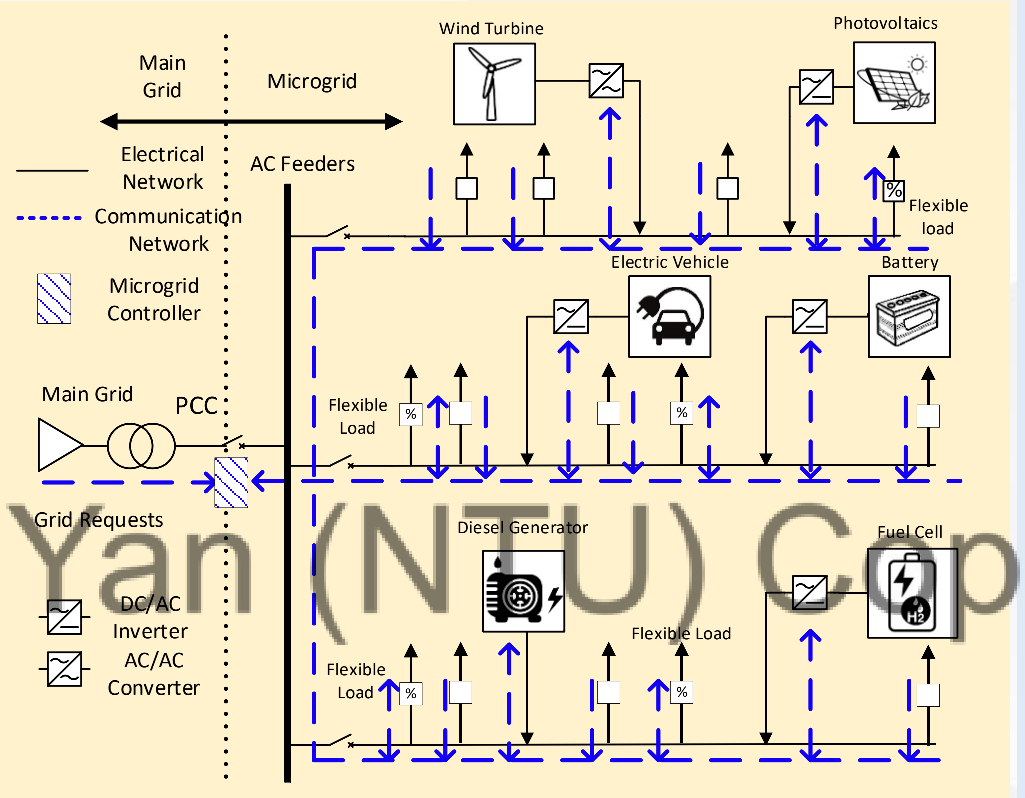
Microgrid Definitions

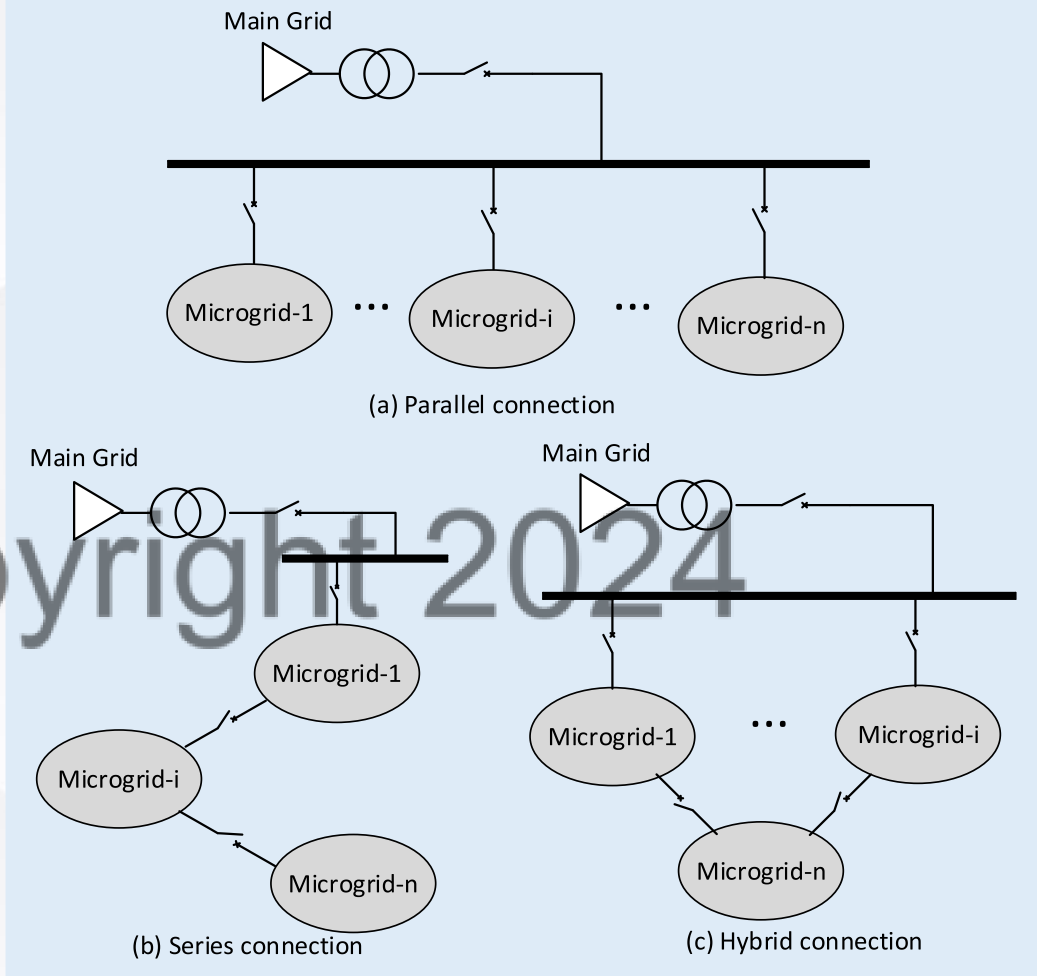


Source of picture: World Economic Forum

1. The U.S. Department of Energy (DOE) defines a microgrid as '*a group of interconnected loads and DERs within clearly defined electrical boundaries that acts as a single controllable entity with respect to the main grid. A microgrid can connect and disconnect from the main grid to enable it to operate in both connected or island-mode*'.
2. The CIGRE C6.22 Working Group defines that '*Microgrids are electricity distribution systems containing loads and DERs, (such as distributed generators, storage devices, or controllable loads) that can be operated in a controlled, coordinated way either while connected to the main power network or while islanded*'.
3. N. Hatziargyriou, Microgrids: Architectures and Control, UK: Wiley-IEEE Press, 2014, ISBN: 978-1-118-72068-4. describes the microgrid as '*comprising low-voltage (LV) distribution systems with DERs. Such systems can operate either connected or disconnected from the main grid. The operation of DERs in the network can provide benefits to the overall system performance, if managed and coordinated efficiently*'.

Microgrid Structure





Single Microgrid

- Islanded mode or grid-connected mode
- Limited generation capability, low system inertia, and geographical boundary

Y. Xu, Y. Wang, C. Zhang, and Z. Li, "*Coordination of Distributed Energy Resources in Microgrids: Optimisation, control, and hardware-in-the-loop validation*," IET Press, 2021, ISBN-13: 978-1-83953-268-9

Networked-Microgrids (NMG)

- Interconnected individual microgrids
- Diverse supply and demand profiles
- More complex network structure and interaction

Smart Home: Nano-grid

Smart Home Appliances

Flexibilities of Smart Appliances

Time-shiftable Appliances: Washing machine, EV charging, etc.

Power-shiftable Appliances: Air Conditioning, Lighting, etc.

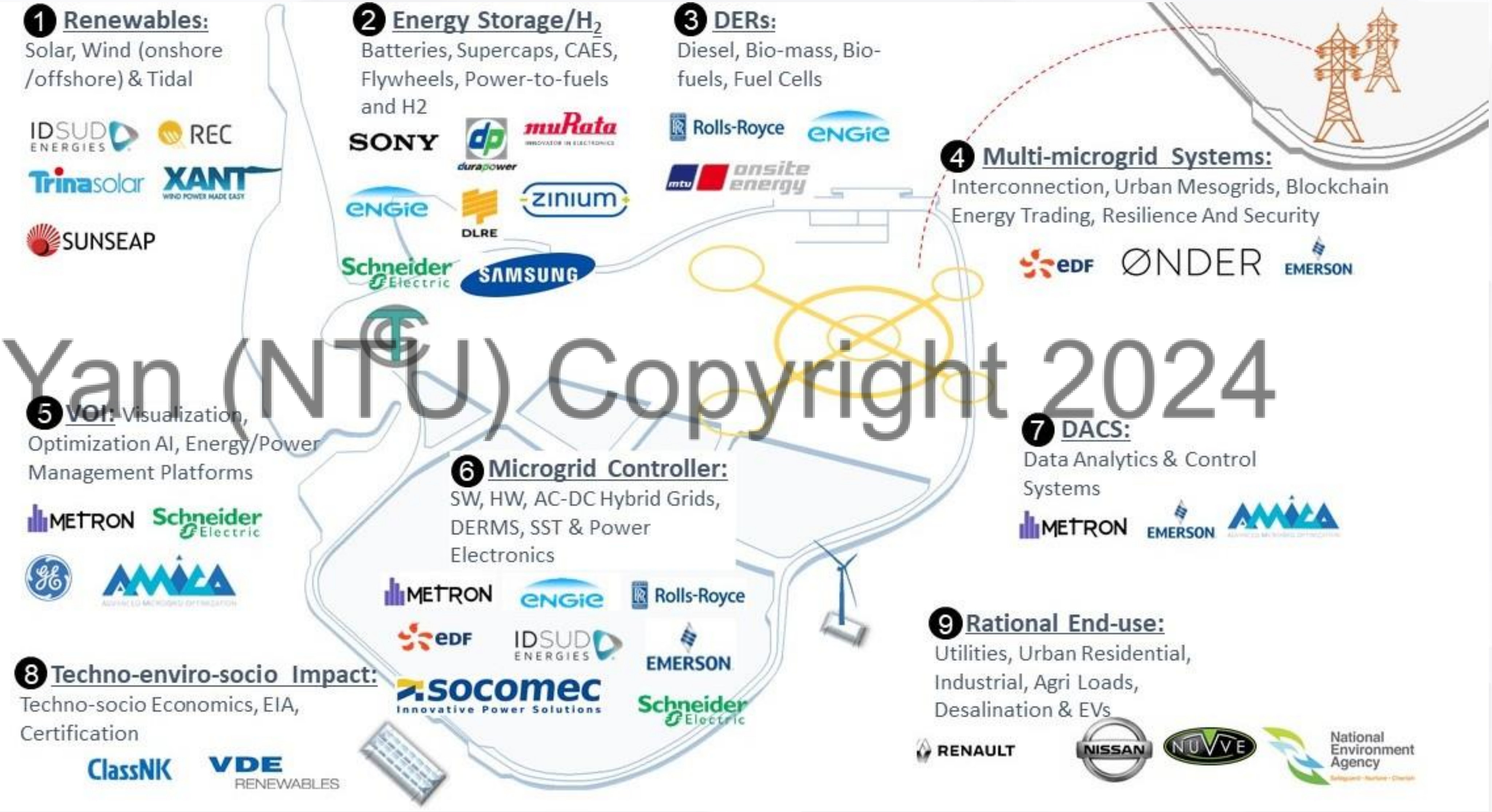
Optimally managing flexible appliances for maximum usage of renewable energies (or minimal user electricity bills)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Renewable Energy Integration Demonstrator – Singapore (REIDS)

http://erian.ntu.edu.sg/REIDS/Pages/AboutREIDS.aspx

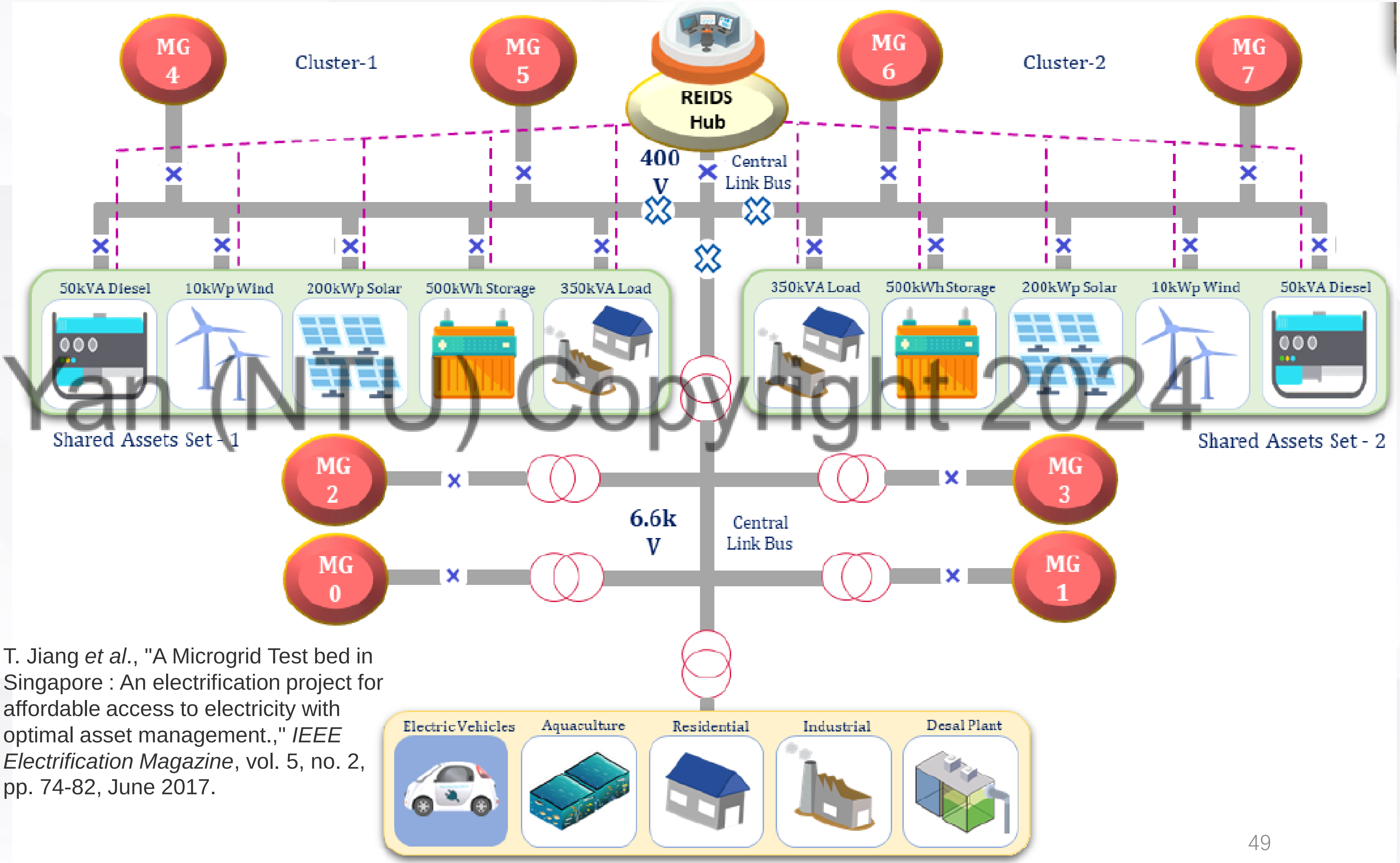
REIDS Industry Collaborators



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

http://erian.ntu.edu.sg/REIDS/Pages/AboutREIDS.aspx

REIDS Electrical Structure



T. Jiang *et al*., "A Microgrid Test bed in Singapore : An electrification project for affordable access to electricity with optimal asset management.," *IEEE Electrification Magazine*, vol. 5, no. 2, pp. 74-82, June 2017.

REIDS Components

1. Overview

2. Power Systems

2.1 Frequency control

2.2 Optimal power flow

2.3 Topology optimization

3. Microgrids

3.1 Frequency control

3.2 Controller tuning

3.3 Energy management

3.4 Volt/Var control

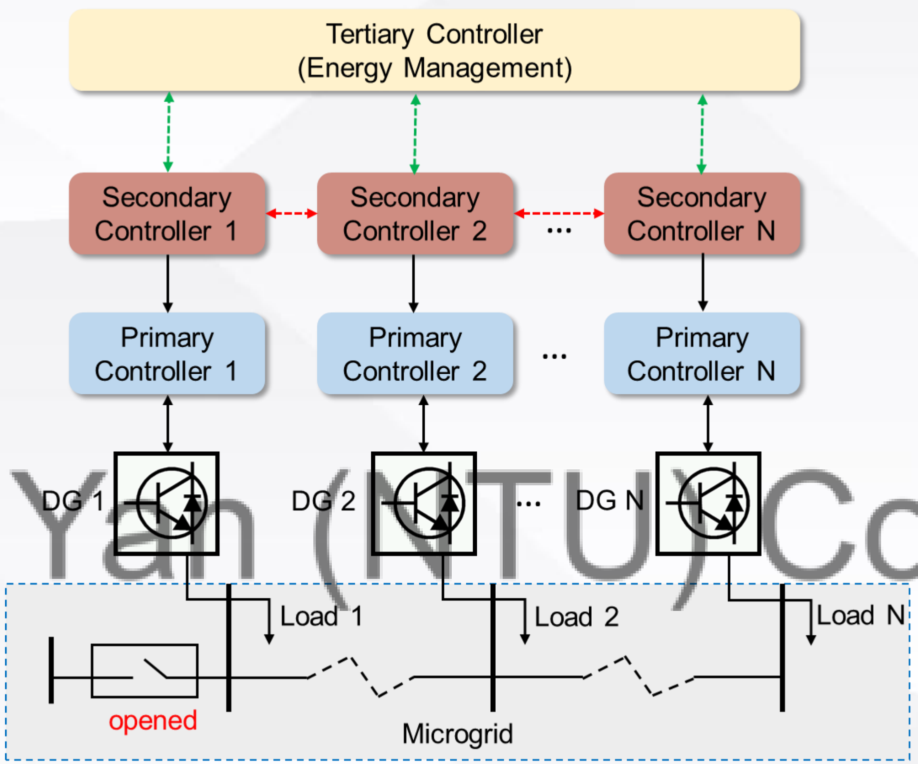
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

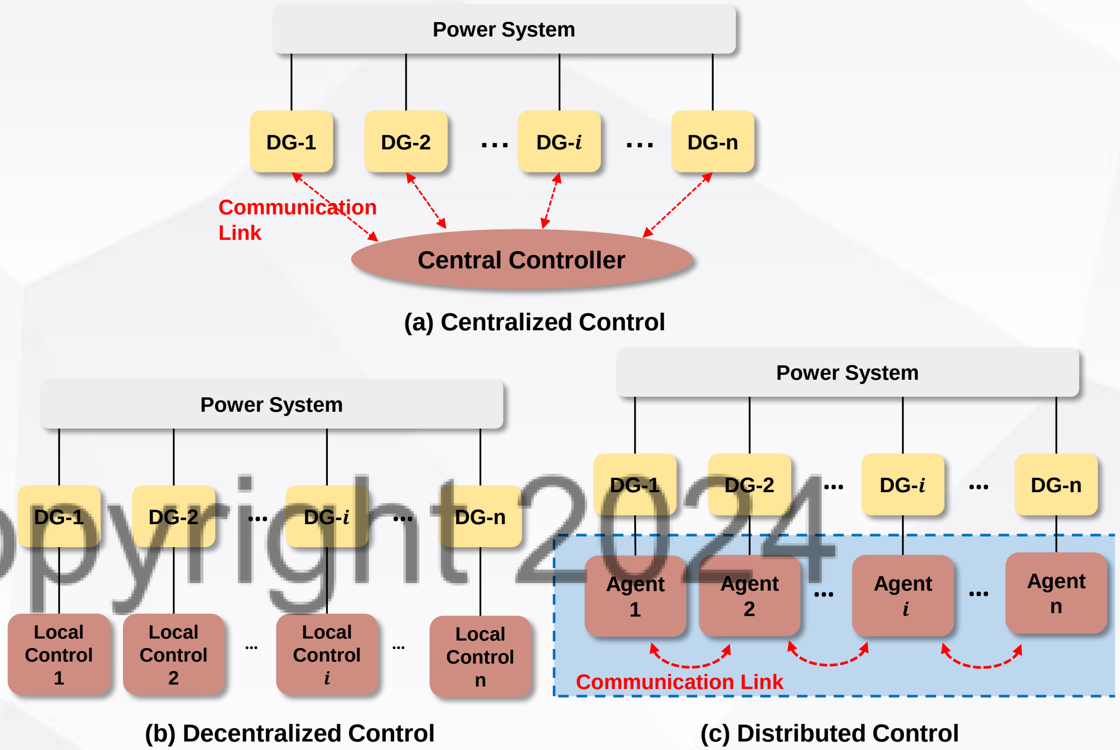
▪ REIDS Components

REIDS Components

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Microgrid Control Hierarchy and Architecture

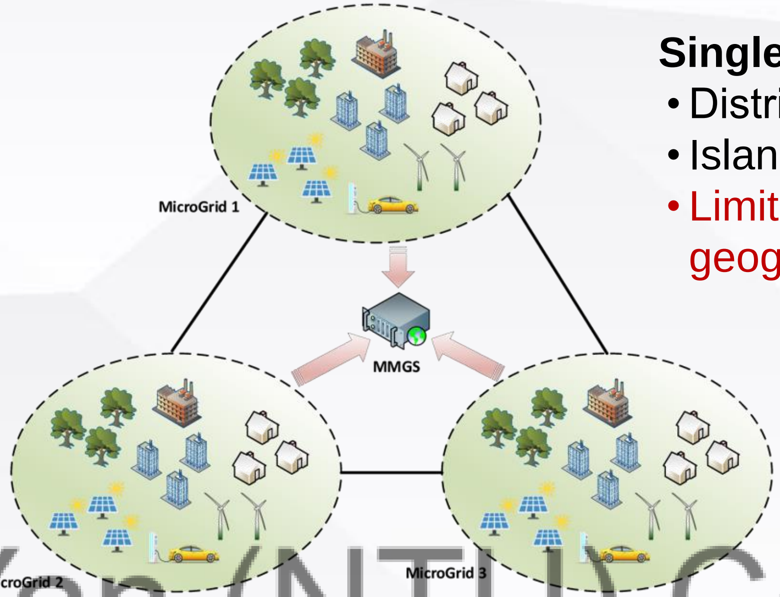




- **Tertiary control (centralized or distributed)**
  - Energy management, volt/var optimization
- **Secondary control (centralized or distributed)**
  - V/f restoration and accurate power balancing
- **Primary control (decentralized)**
  - Inner control loops and droop control
  - Local V/f regulation and power sharing

Control Architecture	Advantages	Disadvantages
Centralized	Global optimization	• Suffer from single-point failure • Unsatisfying reliability • Heavy communication burden
Decentralized	Fast and do not need communication	Lack of global information
Distributed	Good performance with reduced communication burden	Sensitive to time delay

Decentralized Frequency Control of Networked-Microgrids



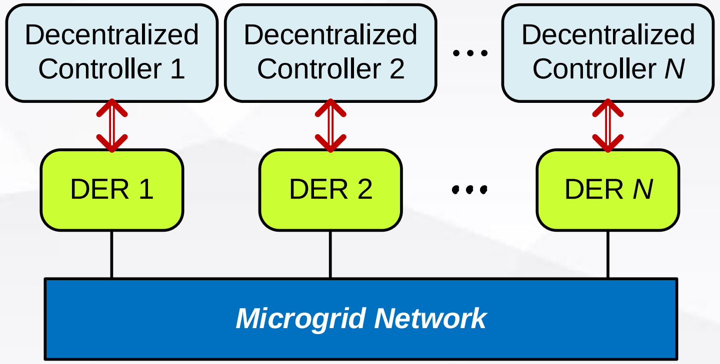
Source of pictures: website (searched in Google)

Single Microgrid
- Distributed energy resources (DERs): DG, ESS, flexible loads…
- Islanded mode or grid-connected mode
- Limited generation capability, low system inertia, and geographical boundary

Networked Microgrid
- Interconnect individual microgrids
- Diverse supply and demand profiles
- More complex network structure and interaction

Decentralized Frequency Control

Mostly model-based: modelling complexity and parameter uncertainty

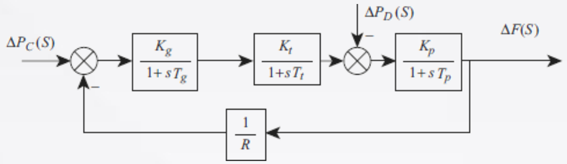


Frequency Control and Economic Issues:

- Separately considered and hierarchically designed (**different time-scales** of secondary and tertiary controls)
- System will deviate from the optimal operation points (**renewable prediction errors** during dispatch intervals)

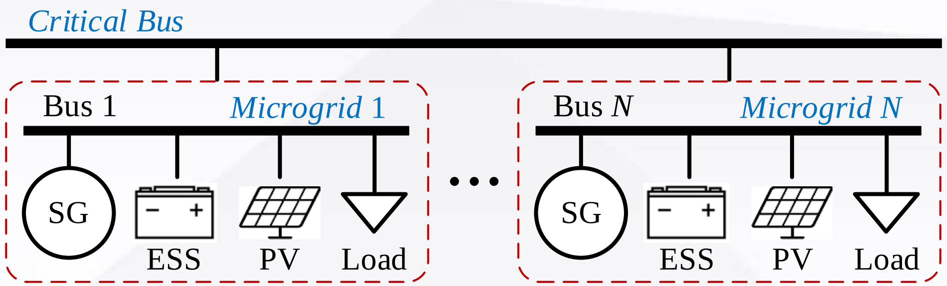
Economic Frequency Control

Control objectives: economic issues + frequency restoration



Y. Xia, **Y. Xu**, Y. Wang, S. Mondal, S. Dasgupta, A. Gupta, and G. Gupta, "A Safe Policy Learning-Based Method for Decentralized and Economic Frequency Control in Isolated Networked-Microgrid Systems," ***IEEE Trans. Sustainable Energy*, 2022.**

Problem Formulation



Schematic of an islanded NMG system

ESS modeling

State of Charge (SOC):

$$E_t^i = E_{t-1}^i - \left(P_t^{i,dis} / \eta_{dis} - P_t^{i,ch} \cdot \eta_{ch}\right) \cdot \Delta t \qquad SOC_t^i = E_t^i / E^{i,rate}$$

$E_i(t)$ is the battery energy at time t in ith microgrid,
Δt is the time interval,
E_i^{rate} is the rated battery capacity,
η_d, η_c indicate the discharging and charging efficiency.

State of Health (SOH):

SOH degradation: $\Delta SOH(t) = SOH(t) - SOH(t-1)$

$$\Delta SOH_t = SOH_t - SOH_{t-1} = h_t \cdot SOH_{t-1}$$

$$h_t = \alpha_1 / \left(\Delta SOC_t\right)^{\alpha_2}$$

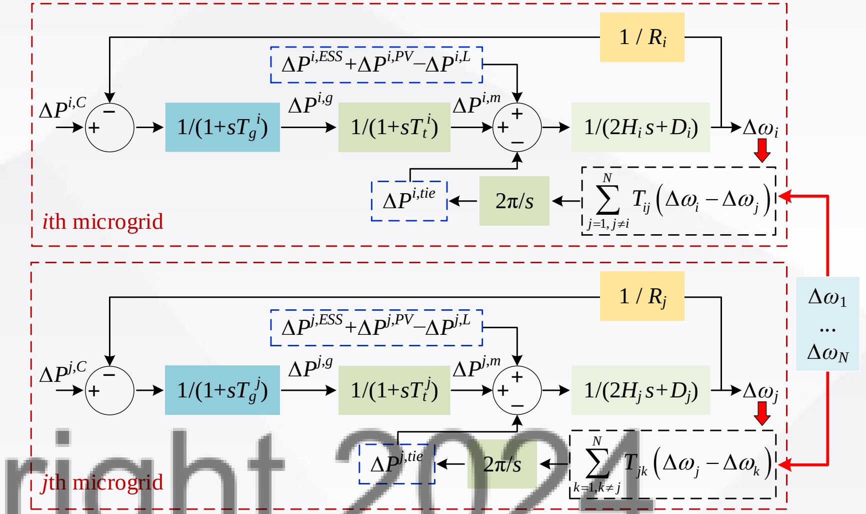
$$\Delta SOC_t = SOC_t - SOC_{t-1}$$

h_t is the degradation factor,
α_1, α_2 are degradation coefficients determined by battery characteristic.

Note that *SOC* and *SOH* values are normalized within the range of [0, 1].

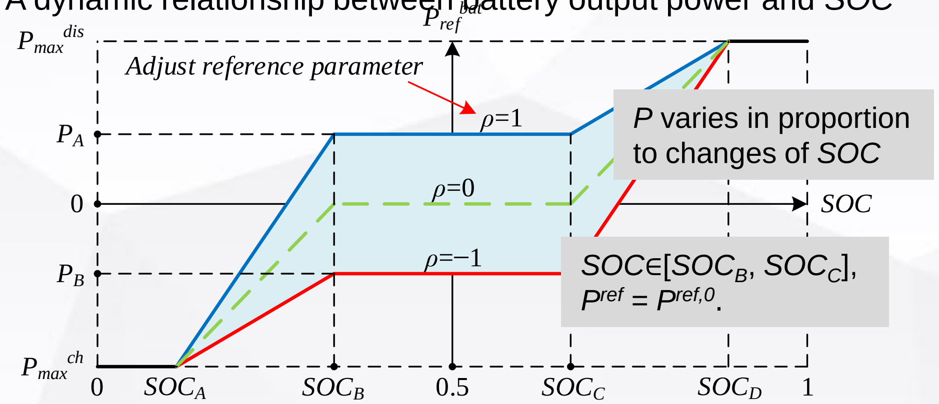
Frequency Response Model of NMG System

A linearized load frequency control (LFC) model with multiple control areas



ESS for frequency support

A dynamic relationship between battery output power and *SOC*

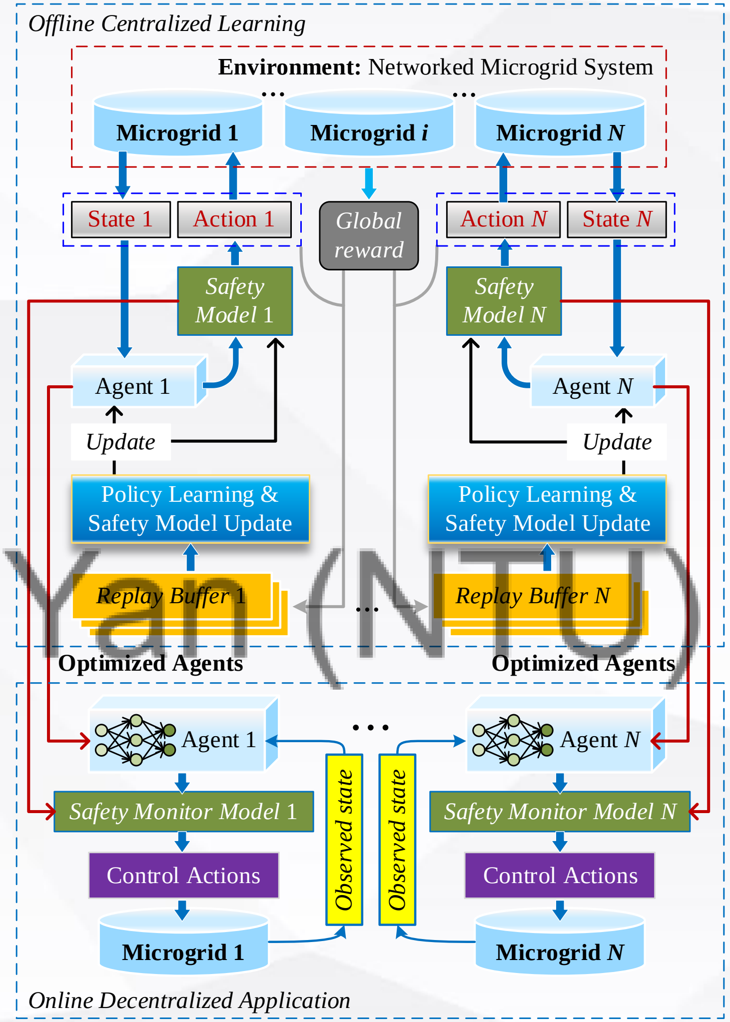


Restore SOC under extremely high/low conditions.
If SOC is within a normal range: ESS output can be adjusted by adjusting the parameter of the curve ρ.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

DRL-based Decentralized Economic Frequency Control



Two stages

- Offline **centralized** learning
- Online **decentralized** application

State (Input of the agent): $s_t^i = \left\{\Delta f_t^i, P_{t-1}^{i,L}, P_{t-2}^{i,L}, ..., P_{t-t_D}^{i,L}, P_{t-1}^{i,PV}, P_{t-2}^{i,PV}, ..., P_{t-t_D}^{i,PV}\right\}$

It implicitly integrates the **forecasting** of PV and load into the proposed method.

Action (Output of the agent): $a_t^i = \left\{\Delta P_t^{i,C},\ \rho_t^i\right\}$

$\Delta P_t^{i,C}$ denotes **the change of control command** in the frequency response model

ρ_t^i denotes **the reference parameter** of ESS characteristic curve at time slot *t*.

Reward: Evaluate the performance for a pair of state and action {*s*(*t*), *a*(*t*)}

$$r_t = \alpha^{\omega}\cdot R_t^{\omega} + \alpha^{DG}\cdot R_t^{DG} \begin{cases} R_t^{\omega} = -\sum_{i\in N}\left|\omega_t^i - \omega^{ref}\right| \\ R_t^{DG} = -\sum_{i\in N}\left|a^i\cdot\left(P_t^{i,DG}\right)^2 + b^i\cdot P_t^{i,DG} + c^i\right| \end{cases}$$

minimize frequency deviation and total generation costs

Constraints: Guarantee the safety of the policy $\omega^- \le \omega_t^i \le \omega^+$

$0 \le P_t^{i,C} \le P_{max}^{i,C}$ $\quad -1 \le \rho_t^i \le 1$ $\quad 0 \le SOC_{min}^i \le SOC_t^i \le SOC_{max}^i \le 1$ $\quad 0 \le SOH_{min}^i \le SOH_t^i \le SOH_{max}^i \le 1$

Soft Actor-Critic (SAC)

- An actor network $\pi(s)$ for action generation
- A **critic network** $Q(s,a)$ for evaluation: approximate the accumulated reward

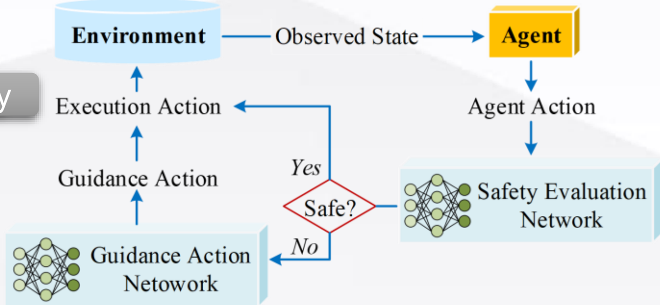
learn the optimal policy towards maximizing the reward while acting for exploration

$$\pi^* = \arg\max_{\pi} \sum_t \mathrm{E}_{(s_t,a_t)}\left[r_t + \tau \boxed{\mathrm{H}\left(\pi\left(s_t\,|\,\theta\right)\right)}\right]$$

H(·) is the function of entropy

Actor network (continuous space): $a_t = \pi(s_t, \xi_t) = \tanh\left(\mu_\theta(s_t) + \sigma_\theta(s_t)\cdot\xi_t\right)$

Safety Model



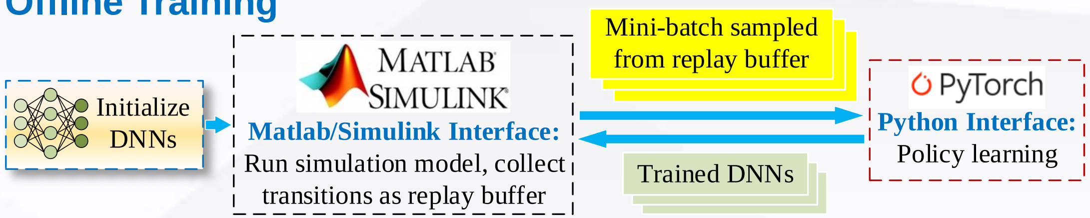
- Avoid constraint violations during offline learning
- Monitor the safety of action during online application

Safety evaluation network: predict the cost value based on (s_t,a_t) before the action is executed

Guidance action network: generate a conservative action to replace the unsafe action

Case Studies

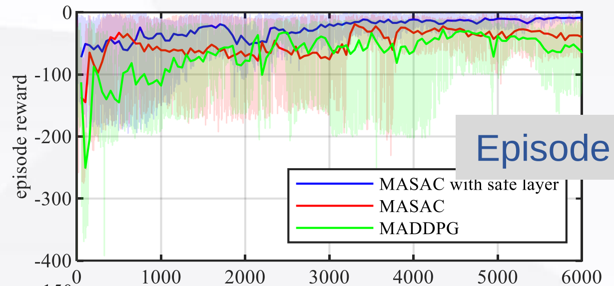
Offline Training



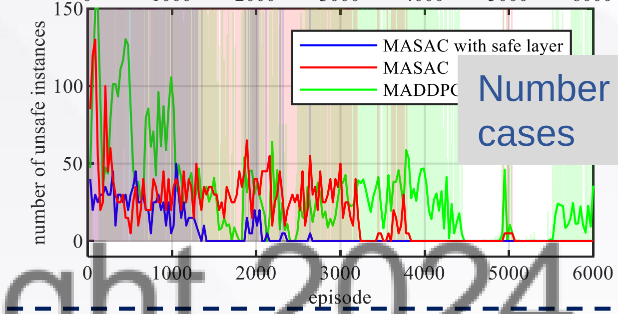
The offline training framework of the proposed method.

TABLE PARAMETERS SETTINGS OF PROPOSED ALGORITHM

Parameters	Value	Parameters	Value
Learning rate λ	1e−3	Discount factor γ	0.99
Optimizer	Adam	Entropy weight τ	0.02
Size of minibatch m	4000	Soft update rate λ_t	0.01
Activation function of hidden layers	relu	Starting episode of safe layers	200

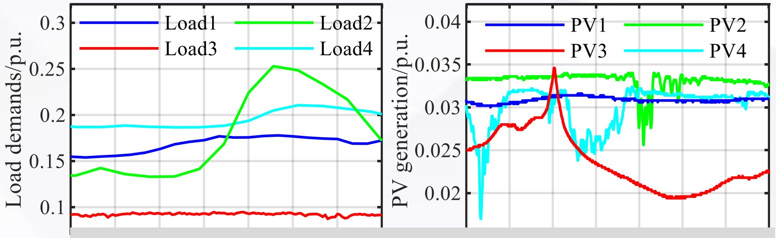


Episode rewards

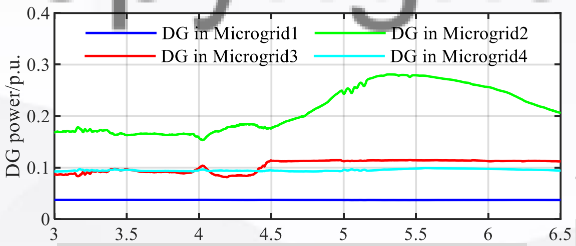


Number of unsafe cases

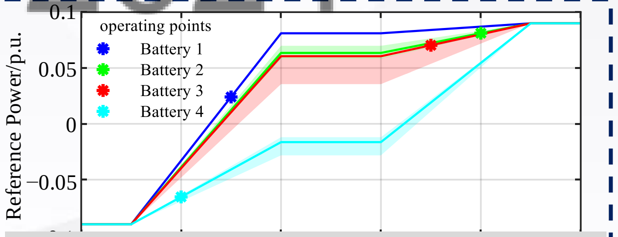
Testing Performance



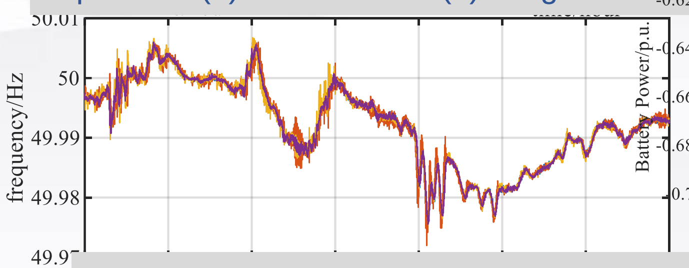
Input data (a) load demand (b) PV generation



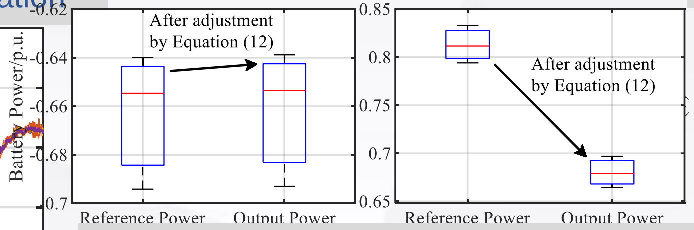
Generated power of each DG



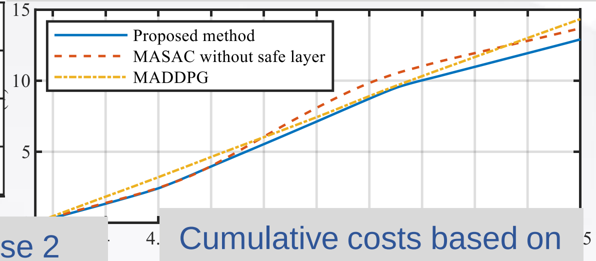
The characteristic curves and operating points of ESSs in the NMG system



Frequency response of the NMG system



Reference power and output power in case 2 (a) ESS 2, SOH2=20% (b) ESS 4, SOH4=70%

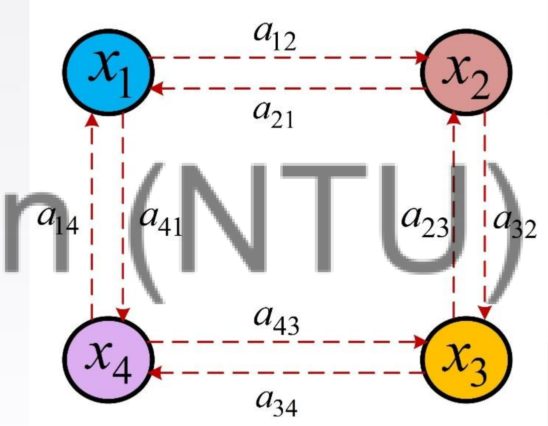


Cumulative costs based on different algorithms

Distributed Control

- ✓ No need for a central controller
- ✓ One node only communicates with neighbouring nodes
- ✓ Share communication and computation burden among nodes
- ✓ Higher resilience, plug-and-play, scalability, data privacy

Example of communication graph



Adjacency matrix of the graph

$$A = \begin{bmatrix} 0 & a_{12} & 0 & a_{14} \\ a_{21} & 0 & a_{23} & 0 \\ 0 & a_{32} & 0 & a_{34} \\ a_{41} & 0 & a_{43} & 0 \end{bmatrix}$$

Two conventional consensus rules:

a) Average consensus control

$$\dot{x}_i(t) = \sum_{j \in N_i} a_{ij}(t)(x_j(t) - x_i(t))$$

$$\lim_{t \to \infty} \left\| x_i(t) - x_j(t) \right\| = 0$$

b) Leader-follower consensus control

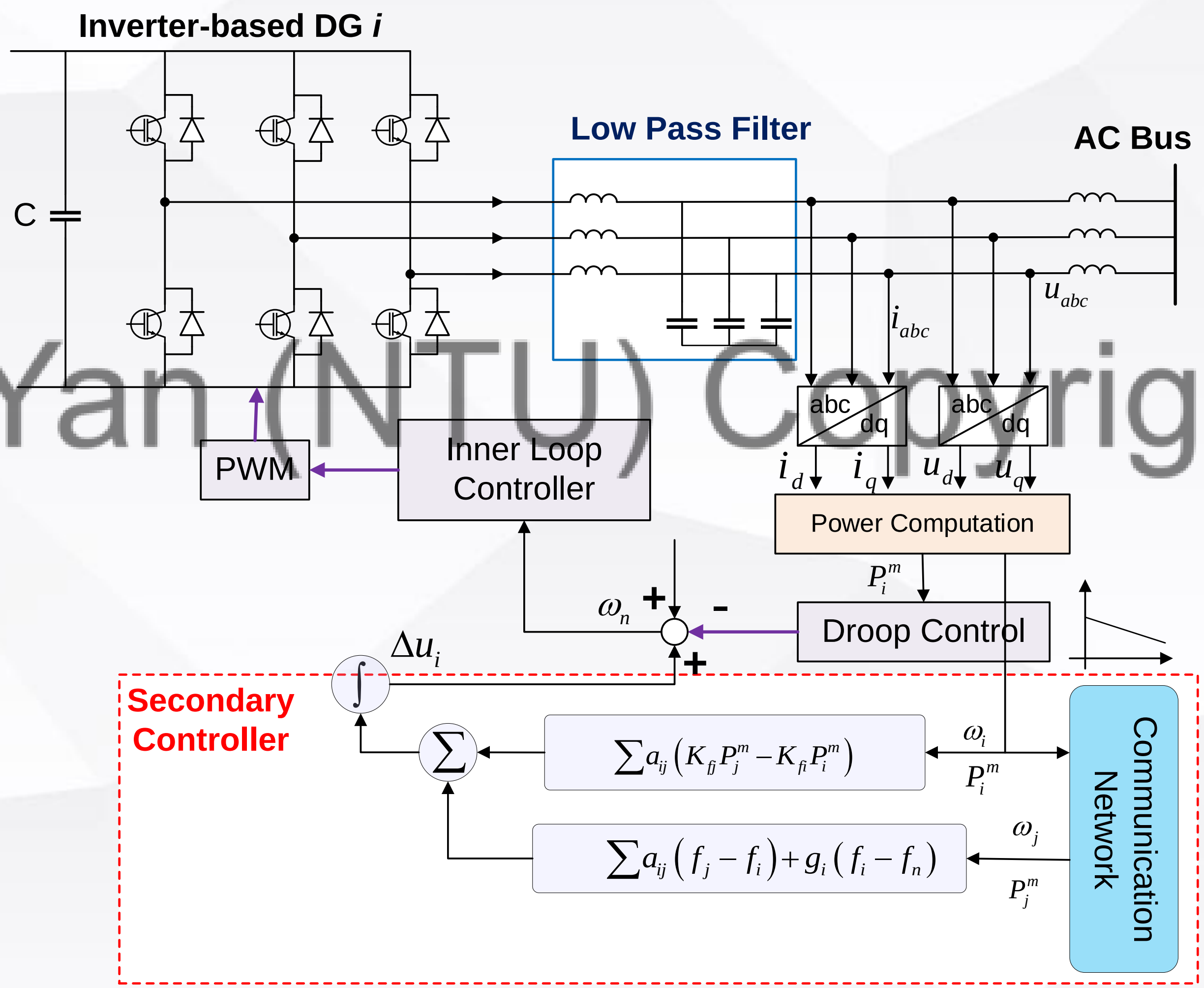
$$\dot{x}_i(t) = \sum_{j=1}^{n} a_{ij}(t)(x_j(t) - x_i(t)) + g_i(x_0(t) - x_i(t)).$$

$$\lim_{t \to \infty} \left\| x_i(t) - x_0(t) \right\| = 0$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Distributed Frequency Control of Islanded Microgrid

❑ Model-Based Method (Leader-Follower Consensus Control)



Primary Control

$$\dot{\xi}_i = \omega_i - \omega_n$$

$$\omega_i = \omega_n - K_{Pi}\left(P_i^m - P_i^r\right), \quad i \in \boldsymbol{I}_{inv}$$

$$\tau_i \dot{P}_i^m = P_i - P_i^m$$

Secondary Control

$$\left(\omega_j - \omega_n\right) + \tau_i \dot{\omega}_i - K_{Pi}\left(P_i^r - P_i\right) + \Delta u_i = 0$$

$$\Delta \dot{u}_k = \sum_{l=1}^{N} a_{kl}\left(f_l - f_k\right) + g_k\left(f_k - f_n\right)$$

$$+\sum_{l=1}^{N} a_{kl}\left(K_{fl}P_l^m - K_{fk}P_k^m\right)$$

Drawbacks

1. Fixed control rules are non-adaptive
2. Hard to realize optimal control
3. Susceptible to time-delay
4. Relies on accurate system model and parameters

Distributed Frequency Control of Islanded Microgrid

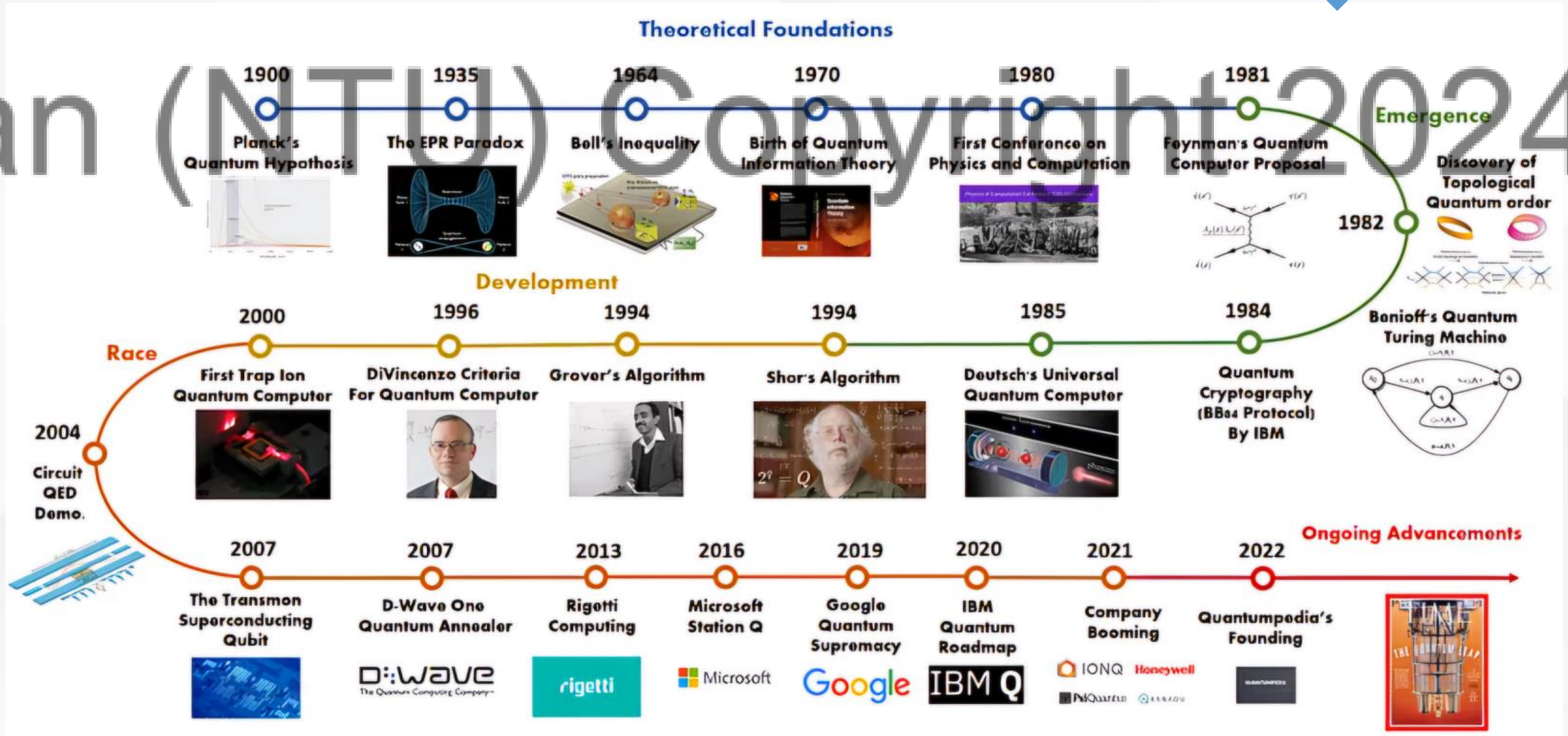
Data-Driven Methods (Deep Reinforcement Learning)

- ✓ Model-free
- ✓ Higher flexibility and scalability
- ✓ Faster solving speed
- ✓ "Trial and Error" interaction with a dynamic system to find an optimal policy.

Problems in the Conventional DRL methods

1. Large amounts of parameters to be trained.
2. Heavy training burden
3. Problem of Scalability: more DGs in the system means more agents

Development of Quantum Computing



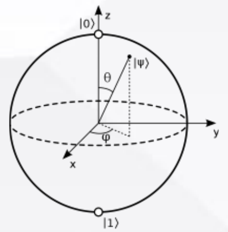
A Brief History of Quantum Computing (Web Source: Quantumpedia)

Quantum Computing and Quantum Machine Learning

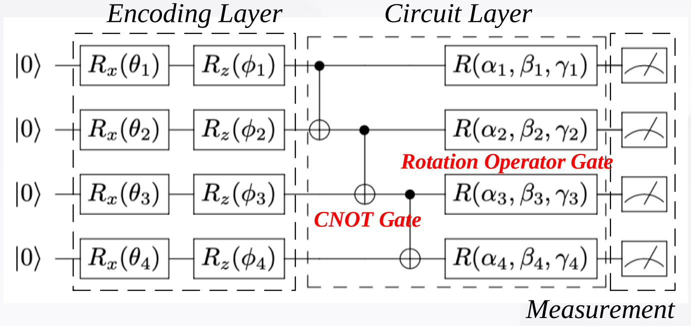
Quantum computing exploits the **principle of state superposition** to achieve **exponential-scale computation space and accelerate computing speed.**

Quantum Bit (Qubit)

$$|\psi\rangle = \rho|0\rangle + \varsigma|1\rangle$$



Variational Quantum Circuit (VQC)



$$R(\alpha,\beta,\gamma) = R_z(\alpha)R_y(\beta)R_x(\gamma)$$

$$R_x(\theta) = \begin{bmatrix} \cos(\theta/2) & -i\sin(\theta/2) \\ -i\sin(\theta/2) & \cos(\theta/2) \end{bmatrix}$$

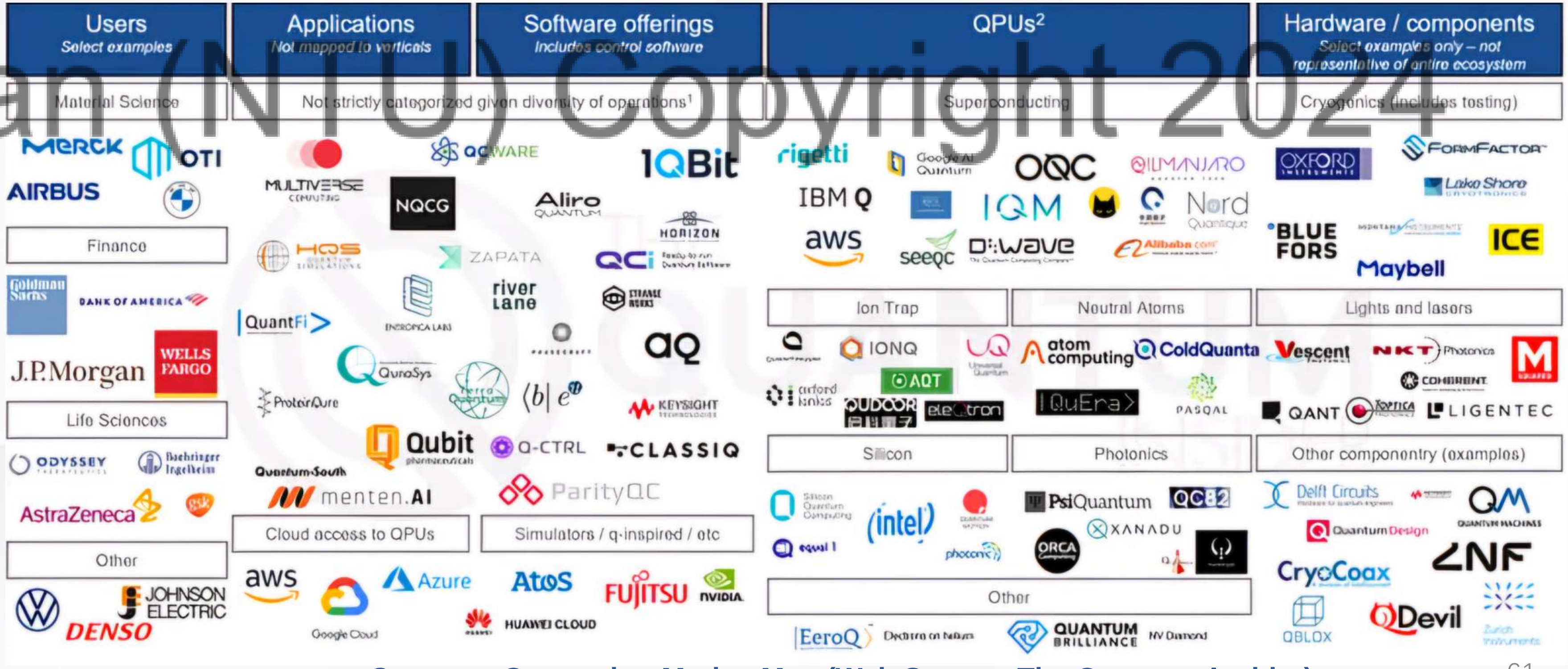
$$R_y(\theta) = \begin{bmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} \exp(-\theta/2) & 0 \\ 0 & \exp(-\theta/2) \end{bmatrix}$$

Unitary Operation

$$\boldsymbol{U^{+}U = UU^{+} = I}$$

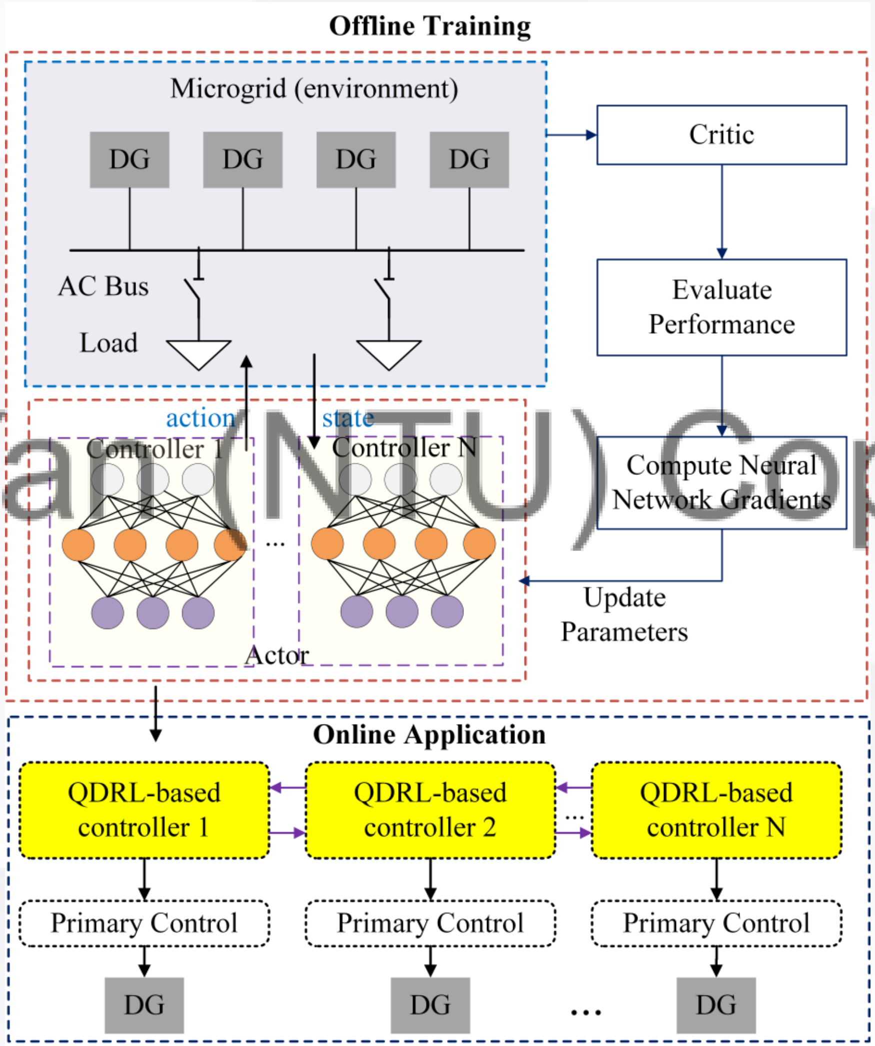
$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$



Quantum Computing Market Map (Web Source: The Quantum Insider)

Distributed Frequency Control of Islanded Microgrid

❑ Proposed Data-Driven Method (Multi-Agent Quantum Deep Reinforcement Learning)



✓ **Critic:** $Q(s,a) = Q_f + Q_p$

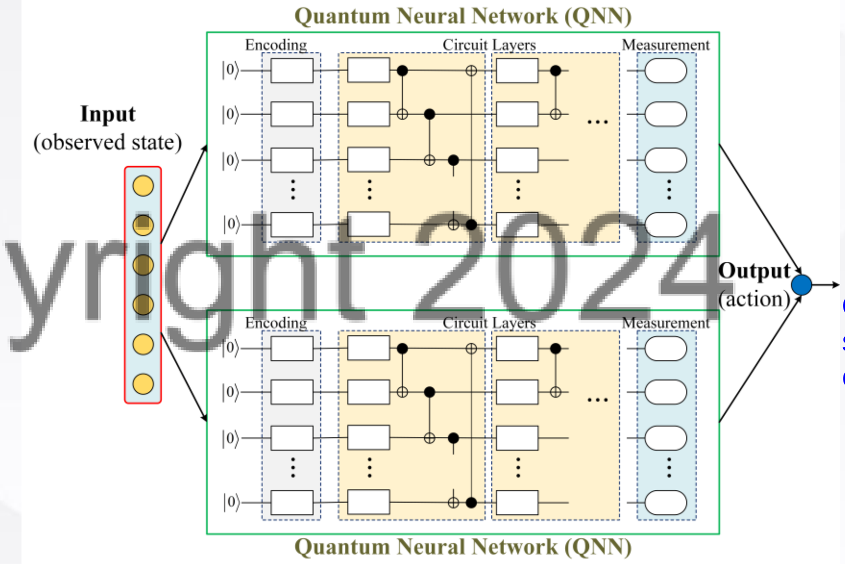
frequency deviations

$$Q_f = -\sum_{t=1}^{T}\sum_{i=1}^{N}\left(f_i(t) - f_n\right)^2$$

power sharing deviations

$$Q_p = -\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(P_i^m(t) - \frac{K_{fj}}{K_{fi}}\cdot P_j^m(t)\right)^2$$

✓ **Actor Network (Hybrid Parallel Structure)**



Output: secondary control signal $a_i = \Delta u_i$

Input: local measurements and neighboring information

$$s_i = \left(f_i, P_i^m, \boldsymbol{F}_i^{nb}, \boldsymbol{P}_i^{nb}\right)$$

$\boldsymbol{F}_i^{nb} = \left\{ f_j \mid j \in \boldsymbol{I}_i \right\}, \boldsymbol{P}_i^{nb} = \left\{ P_j^m \mid j \in \boldsymbol{I}_i \right\}$ Neighboring information

✓ **Reward Function**

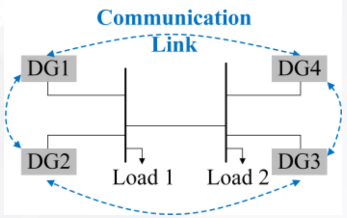
$$r_i = -\sum_{i=1}^{N}\left(f_i - f_n\right)^2 - \sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(P_i^m - \frac{K_{fj}}{K_{fi}}\cdot P_j^m\right)^2 = -\sum_{i=1}^{N}\Delta f_i^2 - \sum_{i=1}^{N}\left(\Delta P_i^m\right)^2$$

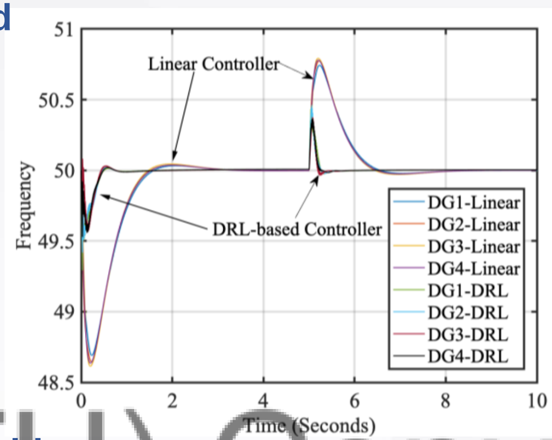
R. Yan, Y. Wang, **Y. Xu***, et. al, "A Multiagent Quantum Deep Reinforcement Learning Method for Distributed Frequency Control of Islanded Microgrids," ***IEEE Transactions on Control of Network Systems*, 2022.**

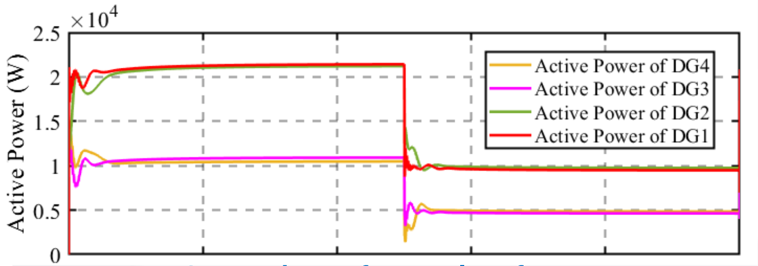
Distributed Frequency Control of Islanded Microgrid

Simulation Results

- Case 1: 4-DG Microgrid



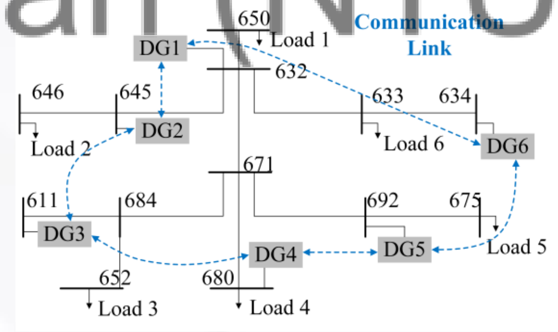




Comparison of control performance

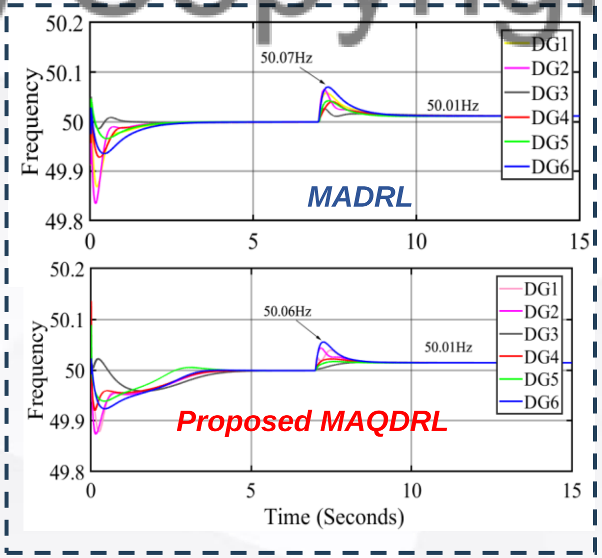
Method	Overshoot	Settling Time	Steady State
Linear	50.74 Hz	2.28s	50 Hz
MADRL	50.44 Hz	0.37s	50 Hz

- Case 2: 13-bus Microgrid

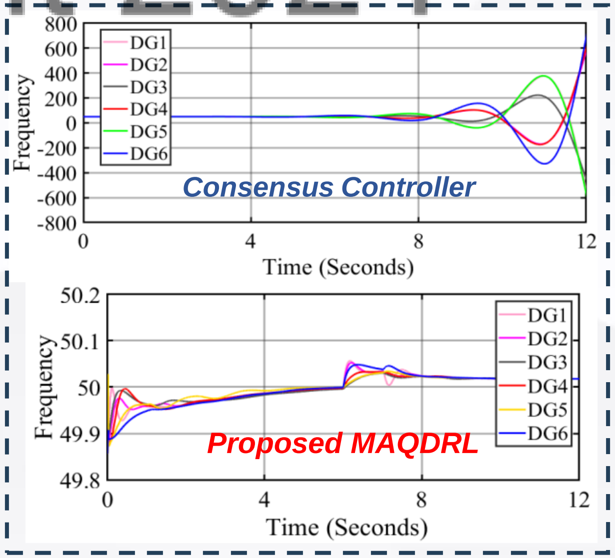


Criteria	MADRL	MAQDRL*
Overshoot	50.07 Hz	50.06 Hz
Settling Time	1.52 s	1.08 s
Steady State	50.01 Hz	50.01 Hz
No. of Parameters	4001	57

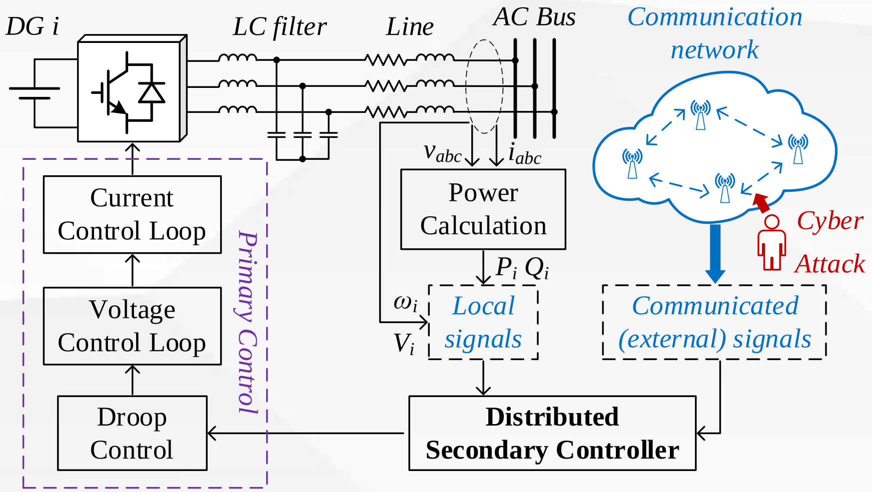
without time delay



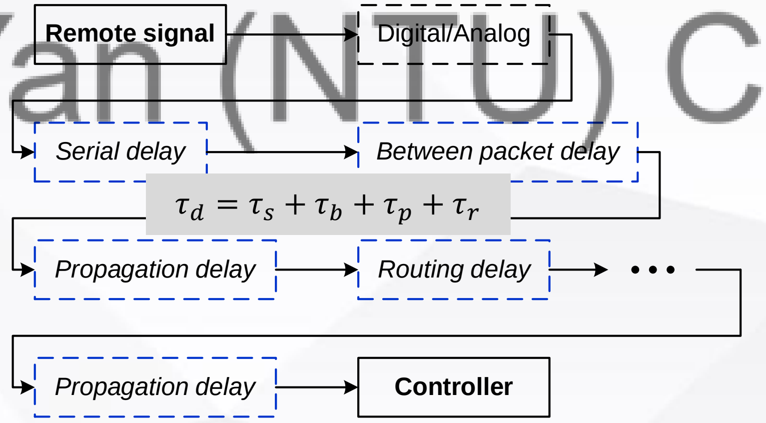
with time delay



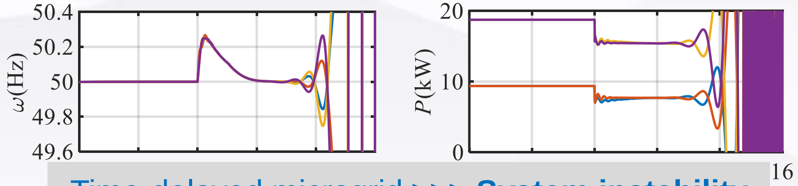
Communication Time-Delay in Distributed Control of Microgrids



Primary and secondary control of inverter-based DGs



Time delays associated with a remote signal to a controller



Time-delayed microgrid >>> **System instability**

❑**Primary control**: fast response for preliminary frequency /voltage regulation (0.1ms~1ms)

❑**Secondary control**: fully restore frequency/voltage and achieve accurate power sharing by global coordination (100ms~1s)

$$\omega_i(t) = \omega^{ref} - K_i^P \cdot P_i(t) + \varepsilon_{\omega,i}(t), V_{di}(t) = V^{ref} - K_i^Q \cdot Q_i(t) + \varepsilon_{V,i}(t)$$

Distributed secondary control: Rely on **communication network**

$$\dot{\varepsilon}_{\omega,i} = \beta_{\omega,i}\left[\theta_i\left(\omega^{ref} - \omega_i(t-\tau_{in})\right) + \sum_{j\in N_i}\alpha_{ij}\left(\omega_j(t-\tau_{co}) - \omega_i(t-\tau_{in})\right)\right] + \beta_{P,i}\left[\sum_{j\in N_i}\alpha_{ij}\left(K_j^P P_j(t-\tau_{co}) - K_i^P P_i(t-\tau_{in})\right)\right]$$

τ_{co}: communication time-delay; τ_{in}: internal time-delay value

$$\dot{\varepsilon}_{V,i} = \beta_{V,i}\left[\theta_i\left(V^{ref} - V_{di}(t-\tau_{in})\right) + \sum_{j\in N_i}\alpha_{ij}\left(V_{dj}(t-\tau_{co}) - V_{di}(t-\tau_{in})\right)\right] + \beta_{P,i}\left[\sum_{j\in N_i}\alpha_{ij}\left(K_j^Q Q_j(t-\tau_{co}) - K_i^Q Q_i(t-\tau_{in})\right)\right]$$

appropriate secondary control gains

>>> a higher time-delay margin (TDM)

Existing gain-scheduling methods:

- Control gains are determined by **offline empirical set:** conservative, ineffective for time-varying delay in practical scenarios
- Online gain-scheduling by solving linear matrix inequality (LMI) equation: rely on an accurate and detailed **system model**, **heavy online computational burden**

Y. Xia, **Y. Xu***, Y. Wang, W. Yao, S. Mondal, S. Dasgupta, A. Gupta, and G. Gupta, "A Data-Driven Method for Online Gain Scheduling of Distributed Secondary Controller in Time-Delayed Microgrids," ***IEEE Trans. Power Syst.*, 2024**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Delay-Dependent Stability Analysis

Small Signal Modeling of Time-Delayed Microgrids

$\Delta\dot{\boldsymbol{X}}_{sys} = \boldsymbol{A}_{sys} \cdot \Delta\boldsymbol{X}_{sys}$ $\boldsymbol{A}_{sys}$: coefficient matrix of the system, $\Delta\boldsymbol{X}_{sys}$: the state variable

$$\Delta\boldsymbol{X}_{sys} = \begin{bmatrix} \Delta\boldsymbol{X}_{inv,1}, \dots, \Delta\boldsymbol{X}_{inv,N} & \varepsilon_{\omega,1}, \dots, \varepsilon_{\omega,N} & \varepsilon_{V,1}, \dots, \varepsilon_{V,N} \\ \Delta i_{dq,load\ 1}, \dots, \Delta i_{dq,load\ K} & \Delta i_{dq,line\ 1}, \dots, \Delta i_{dq,line\ L} & \end{bmatrix}$$

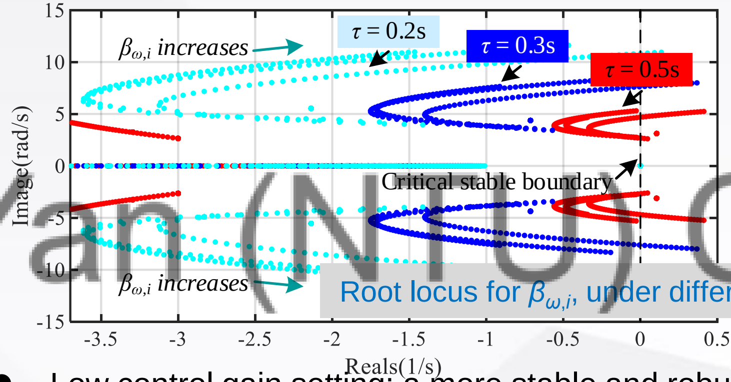
Delayed small-signal model:

$$\Delta\dot{\boldsymbol{X}}_{sys}(t) = \boldsymbol{A}_{sys1} \cdot \Delta\boldsymbol{X}_{sys}(t) + \boldsymbol{A}_{sys2} \cdot \Delta\boldsymbol{X}_{sys}(t-\tau)$$

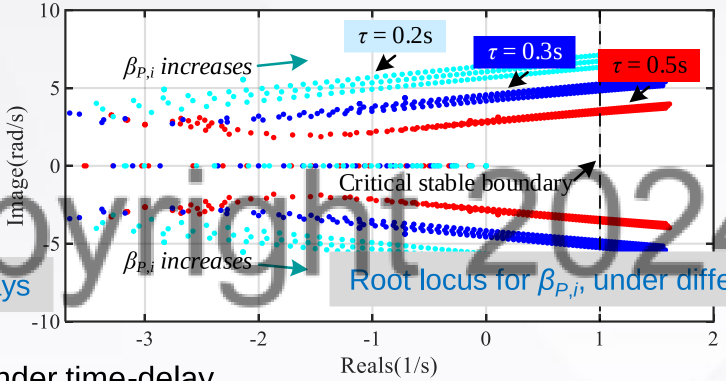
>>> linear delay differential algebraic equation (DDAE): $\Delta\dot{x}(t) = A_0 \cdot \Delta x(t) + \sum_k A_k \cdot x(t-\tau_k)$

>>> characteristic equation of the system: $det(-sI + A_0 + A_1 \cdot e^{-s\tau}) = 0$

Root locus of the eigen pair (under different control gains $\beta_{\omega,i}$, $\beta_{P,i}$, time delays τ = {0.2s, 0.3s, 0.5s})



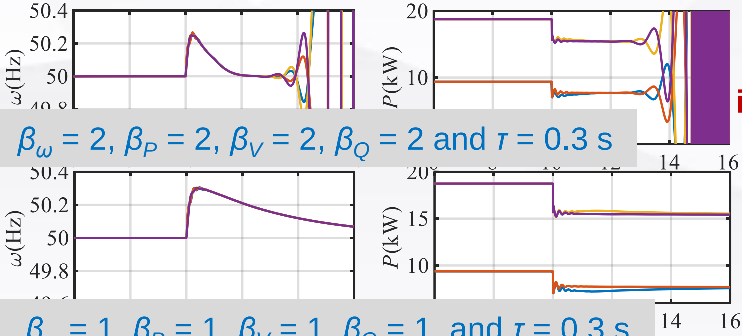
Root locus for $\beta_{\omega,i}$, under different time delays



Root locus for $\beta_{P,i}$, under different time delays

- Low control gain setting: a more stable and robust system under time-delay.
- A slow convergence speed and an ineffective operation (no time-delay)
- An appropriate control gain setting at online stage to improve delay-dependent stability

Small Signal Model Validation



$\beta_\omega = 2$, $\beta_P = 2$, $\beta_V = 2$, $\beta_Q = 2$ and $\tau = 0.3$ s

System instability

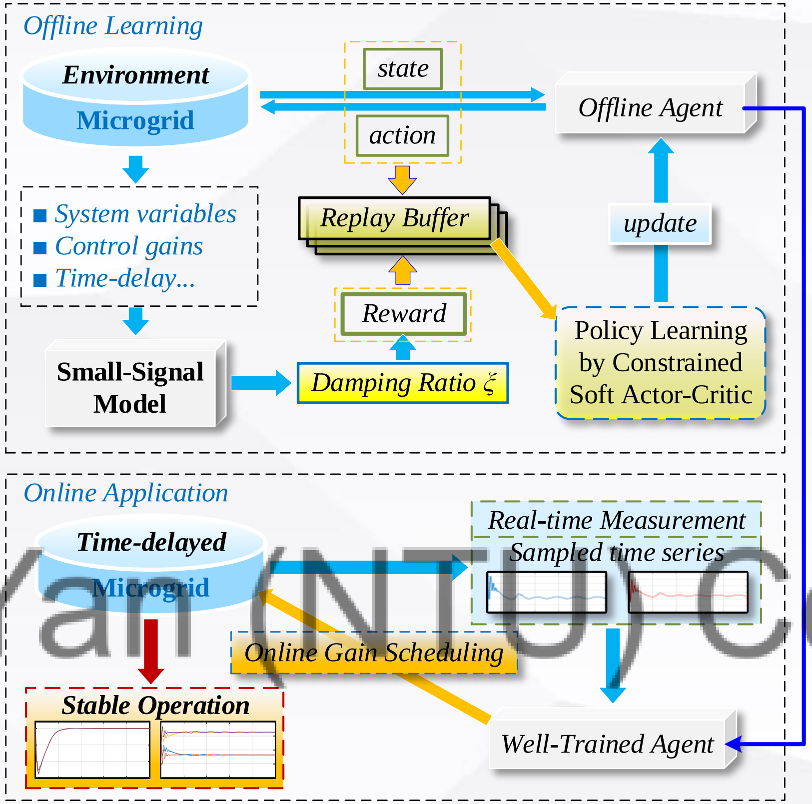
$\beta_\omega = 1$, $\beta_P = 1$, $\beta_V = 1$, $\beta_Q = 1$, and $\tau = 0.3$ s

System stability

Table Test Results of Small-Signal Model

Test	Secondary Control Gains	Time delay	Damp ratio
1	$\beta_\omega = 2, \beta_P = 2, \beta_V = 2, \beta_Q = 2$	$\tau = 0.3$ s	**−9.250 %**
2	$\beta_\omega = 2, \beta_P = 2, \beta_V = 2, \beta_Q = 2$	$\tau = 0.2$ s	**4.543 %**
3	$\beta_\omega = 1, \beta_P = 1, \beta_V = 1.5, \beta_Q = 1.5$	$\tau = 0.3$ s	**0.265 %**
4	$\beta_\omega = 1, \beta_P = 1, \beta_V = 1, \beta_Q = 1$	$\tau = 0.3$ s	**12.982 %**

DRL-based Controller Gain Scheduling



Markov Decision Process Modeling for Gain Scheduling

State/input: A time-series state set of current

$$s_t = [i_{odi}(t-t_d),\dots,i_{odi}(t-1) \quad i_{oqi}(t-t_d),\dots,i_{oqi}(t-1)]$$

Action/output: $a_t = [\beta_\omega(t) \quad \beta_P(t) \quad \beta_V(t) \quad \beta_Q(t)]$

(distributed control gains)

Reward: evaluate the performance of action a_t at state s_t.

Damping ratio ξ of secondary control: $\psi = \psi_{re} + j\psi_{im}$

$$r_t = \begin{cases} 0 & if\ \xi \geq \xi^{ref} \\ -|\xi - \xi^{ref}| & if\ \xi \leq \xi^{ref} \end{cases} \qquad \xi = -\psi_{re}/\sqrt{(\psi_{re})^2 + (\psi_{im})^2}$$

ψ: the critical root of characteristic equation

Constraints:

Frequency/voltage: $\omega_{min} < \omega_i < \omega_{max}$, $V_{min} < V_i < V_{max}$

Active/reactive power: $P_{i,min} < P_i < P_{i,max}$, $Q_{i,min} < Q_i < Q_{i,max}$

Ensure the stability of the system: $\xi > 0$

Offline learning and Online application

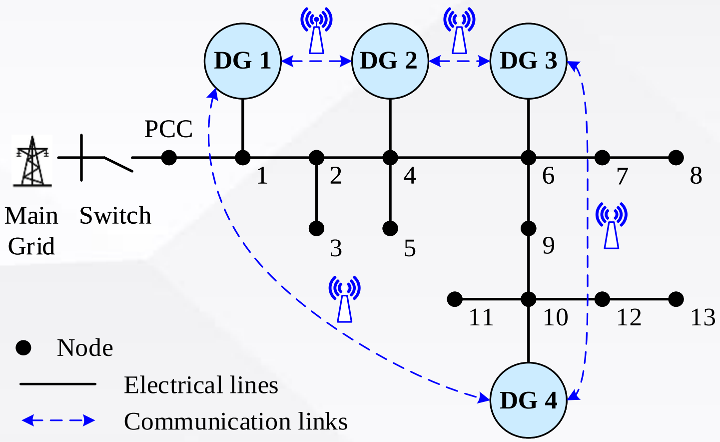
- **Offline learning:**

1. The agent for gain scheduling is initialized as a deep neural network (DNN)
2. Environment >> **state** >> agent
3. Agent >> **new gain set (action)** >> environment
4. Small-signal model >> damping ratio >> **reward**
5. {**state**, **action**, **reward**} into a memory buffer, update the agent (search for optimal policy)

- **DRL training (various scenarios)**
  - Different time-delays
  - Initial system setting
  - Load variation

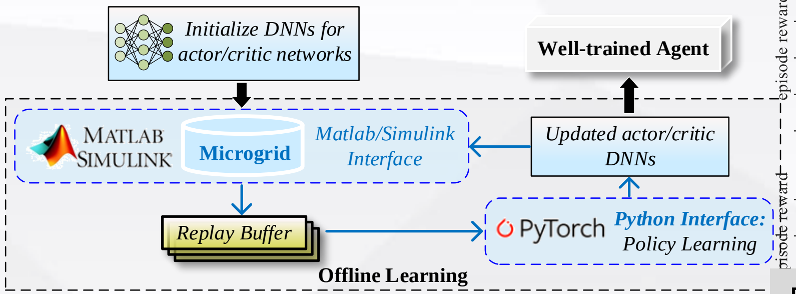
- **Online application:**
  - Work based on the real-time measured current
  - Get rid of heavy computational burden
  - Applicable under time/line-varying time-delays

Case Study



Electrical structure and communication topology of the simulation model

Offline Training



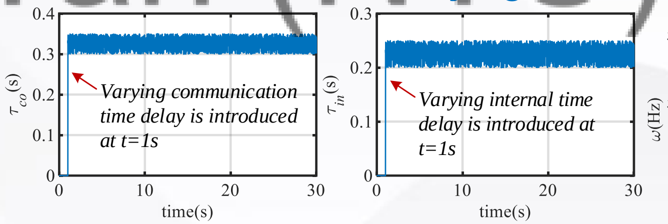
Offline learning framework

Episode reward (a) proposed method (CSAC) (b) SAC

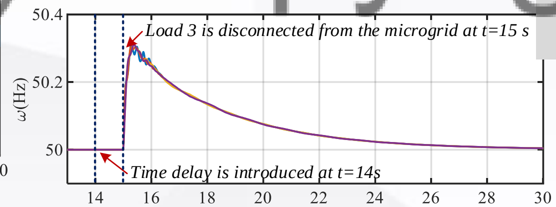
Table Detailed Parameters of The Proposed Algorithm

Parameter	Value	Parameter	Value
Actor learning rate γ_θ	1e−5	Entropy weight α	0.02
Critic learning rate γ_ϕ	1e−4	Soft update rate σ	0.01
Optimizer	Adam	Size of minibatch m	1000

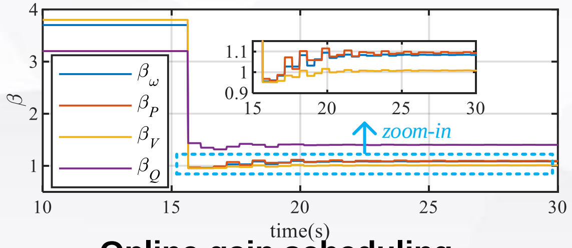
Test case under time-varying time-delay



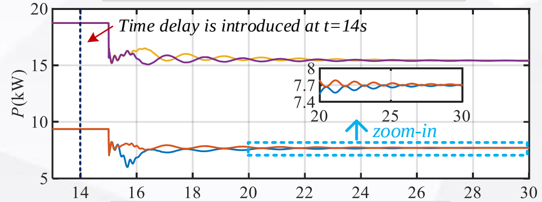
Time-varying time-delays



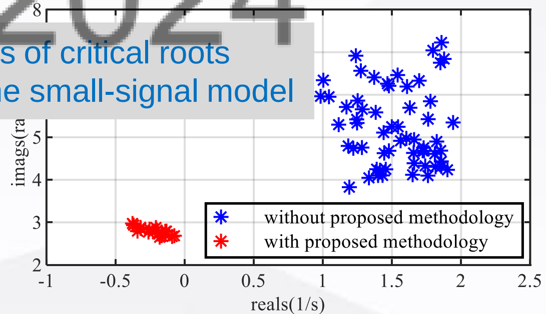
frequency restoration



Online gain scheduling



active power sharing



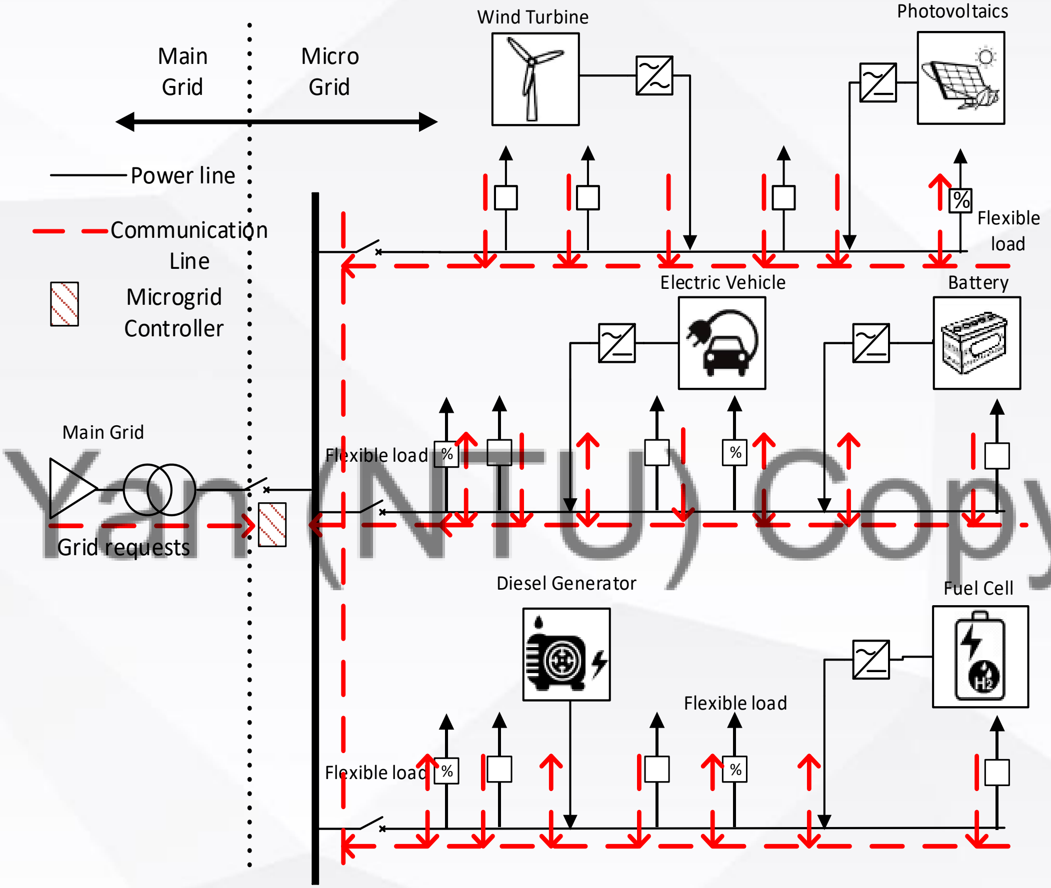
The distributions of critical roots calculated by the small-signal model

Table Eigenvalue and Damping Ratio during the Simulation

Test	Before scheduling		After scheduling	
	Eigenvalue	Ratio	Eigenvalue	Ratio
1	0.69+3.91i	−17.545 %	-0.21+2.60i	8.1999 %
2	0.88+2.58i	−12.119 %	-0.12+1.75i	7.1262 %

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Microgrid Operation: Energy Management & Volt/Var Regulation



- Network model:
1) Linearized Dist-Flow
2) Second-order cone programming (SOCP) model

- Control variables:
1) Micro-turbine
2) Energy storage
3) Demand response

Active power resource

4) Capacitor banks
5) On-load tap changers
6) PV inverters

Reactive power resource

- Parameters:
1) Load demand
2) Wind and PV output
3) Electricity price

Uncertain

4) Network parameters (R,X,B)

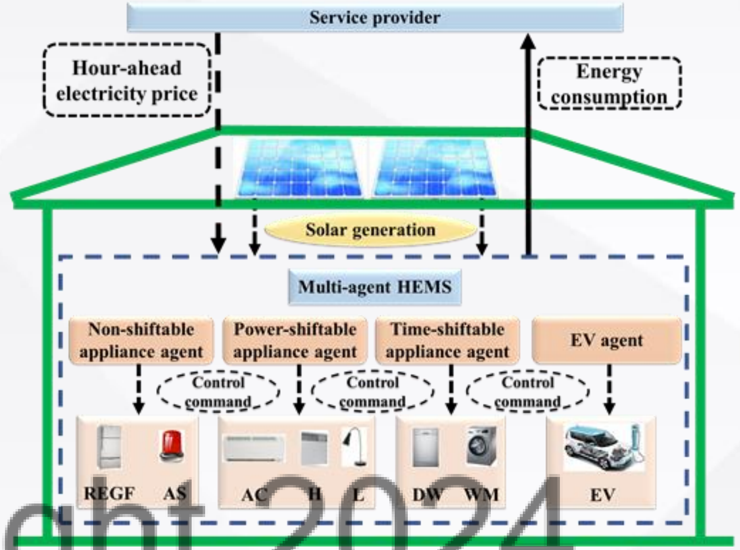
- State variables:
1) Bus voltage
2) Branch power flow
3) Power exchange with main grid

Data-driven Home Energy Management (HEM): Problem Description

Importance of HEM
- Power Grid: local renewable energy consumption
- Consumers: Reduction of electricity bills (demand response)

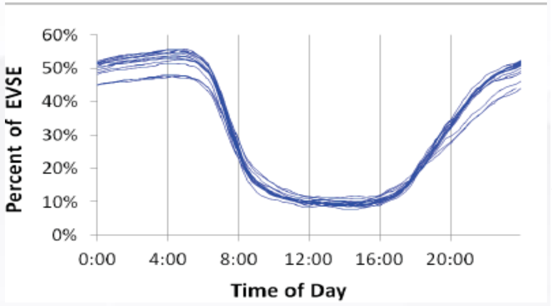
Different load types
- Non-shiftable loads, e.g. refrigerator and alarm system
- Power-shiftable loads, e.g. air conditioner, heating and light
- Time-shiftable loads, e.g. wash machine and dishwasher

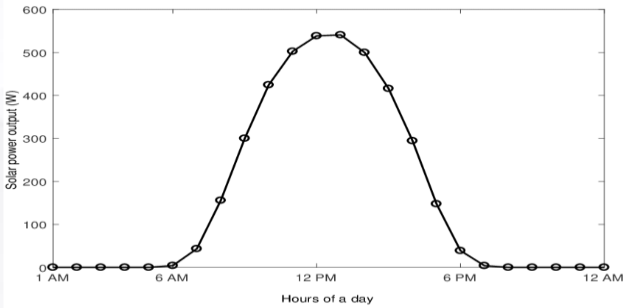
limits of classic optimization methods
- Low computation efficiency
- Non-optimal results for nonlinear and nonconvex models

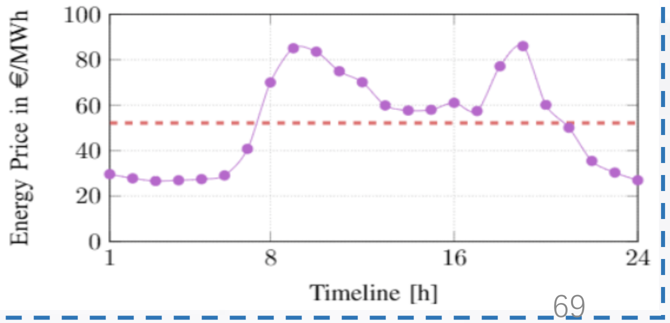


Data-driven based HEM
- Uncertainty prediction
- On-line optimal energy scheduling

Uncertainties
- Electric vehicle (EV) loads
- Rooftop photovoltaic (PV) generation
- Electricity prices

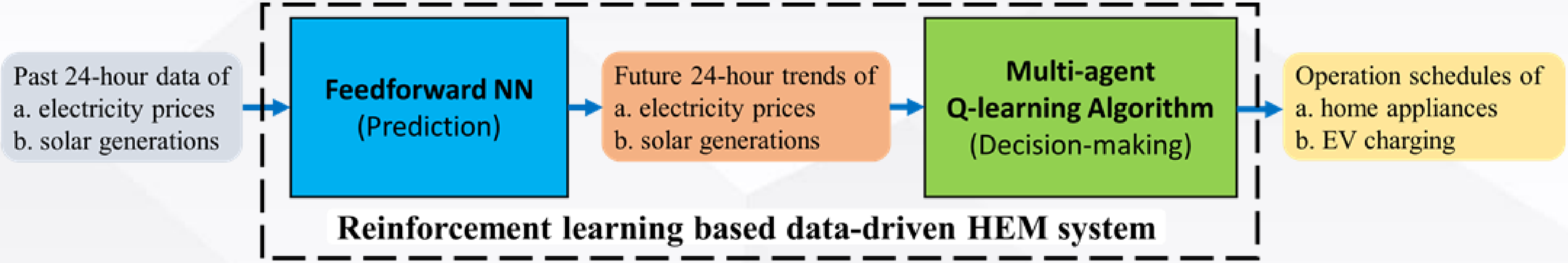




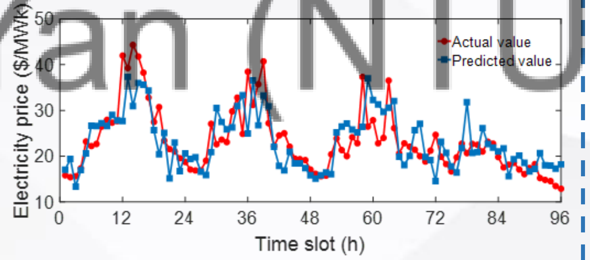


Data-driven Home Energy Management (HEM): Methodology

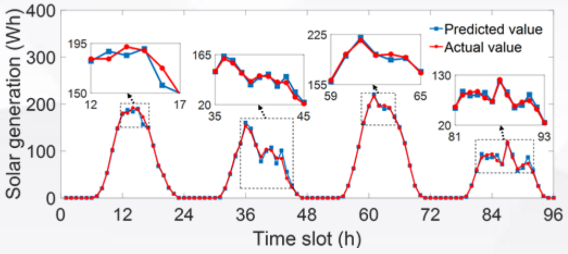
Schematic of the reinforcement learning based data-driven HEM system



Neural Network (NN) based Uncertainty Prediction

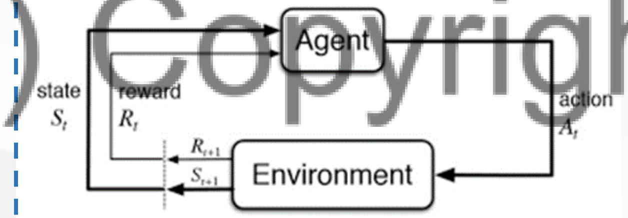


Electricity prices



PV generation

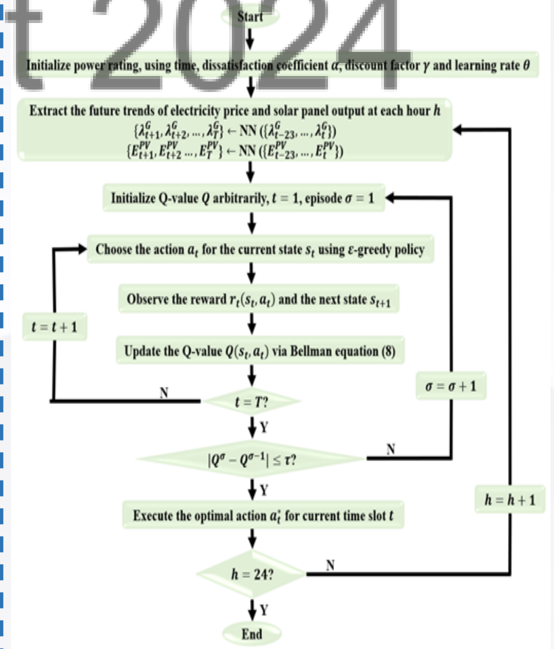
Markov Decision Process (MDP)



- Agent: house owner
- State: predicted information
- Action: energy scheduling
- Reward: (-) utility cost

$$R = -\sum_{t\in T}\left\{\begin{array}{l}\lambda_t^G\left(\begin{array}{l}[P_{it}^{d,NS}-E_{it}^{PVs}]^+ - [P_{jt}^{d,PS}-E_{jt}^{PVs}]^+ \\ -[u_{mt}P_{mt}^{d,TS}-E_{mt}^{PVs}]^+ - P_{nt}^{d,EV}\end{array}\right) \\ -\left(\begin{array}{l}\alpha_j^{PS}(P_{j,\max}^{d,PS}-P_{jt}^{d,PS})^2-\alpha_m^{TS}(t_m^s-t_m^{ini})^2 \\ -\alpha_n^{EV}(P_{n,\max}^{d,EV}-P_{nt}^{d,EV})^2\end{array}\right)\end{array}\right\}$$

Q-learning Algorithm

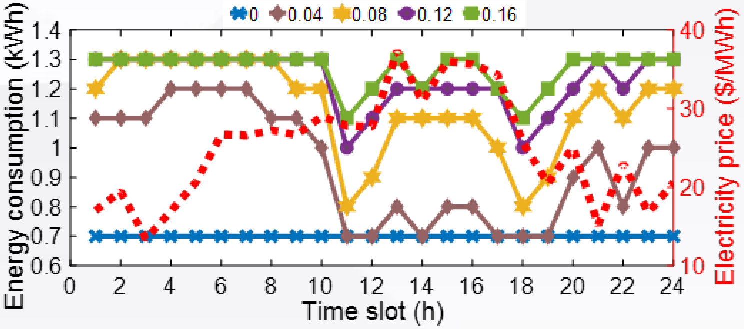


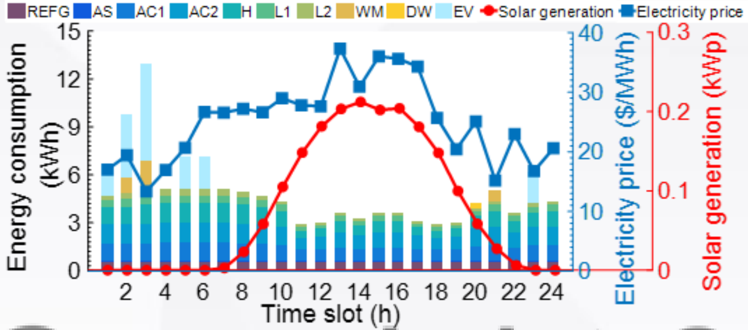
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Data-driven Home Energy Management (HEM): Results

Performance of proposed data-driven model

- Reduce electricity costs (via demand response)

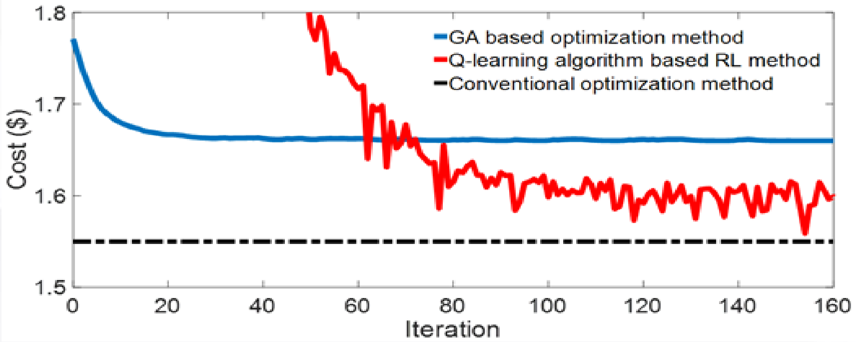


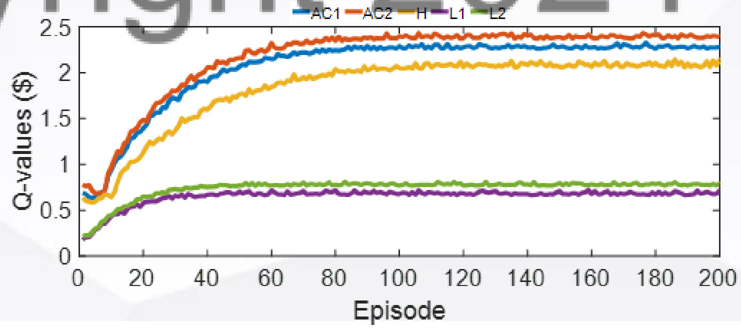


Item ID	Electricity cost ($)	
	With DR	Without DR
REFG	0.492	0.492
AS	0.098	0.098
AC1	0.836	1.378
AC2	0.942	1.378
H	0.731	1.476
L1	0.301	0.591
L2	0.223	0.591
WM	0.023	0.051
DW	0.012	0.012
EV	0.399	1.262
Total	4.057	7.329

Comparison with genetic algorithm

- Higher computation efficiency
- Near-optimal results





	Average computation time of running 1000 times
GA based optimization method	46.296 s
Q-learning algorithm based RL method	1.107 s

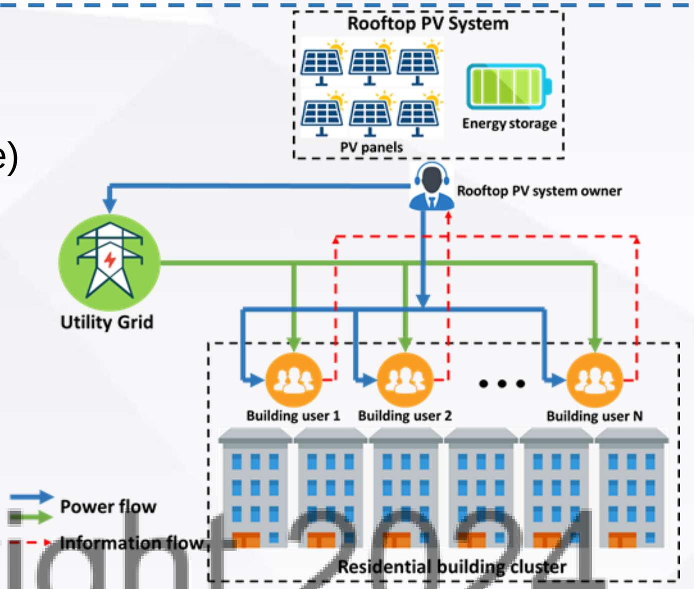
X. Xu, Y. Jia, Y. Xu, Z. Xu, et al, “A Multi-agent Reinforcement Learning based Data-driven Method for Home Energy Management,” **IEEE Trans. Smart Grid, 2020. – Web of Science highly cited paper**

Data-driven Energy Sharing among Buildings: Problem Description

Importance of energy sharing among builidngs
- Power Grid: local renewable energy consumption
- Consumers: reduction of electricity bills (demand response)
- PV system owner: profits

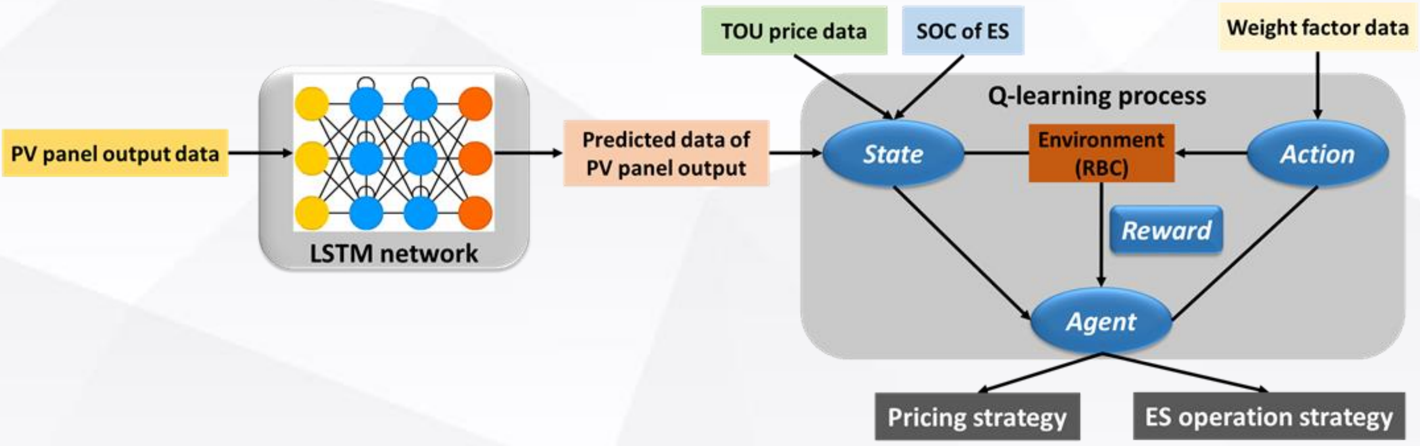
Several deficiencies
- Uncertain renewable generation
- Multiple electricity consumers
- Conflicts of interest

Limits of iterative optimization methods
- Certain assumptions and simplifications for convergence
- Impractical to be used



Data-driven Game-based Energy Sharing

Advantages
- Off-line training and on-line implementation
- Uncertainty consideration
- Near-optimal results



X. Xu, Y. Xu, M. Wang, Z. Xu, J. Li, and S. Chai, "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.**

Data-driven Energy Sharing among Buildings: Framework

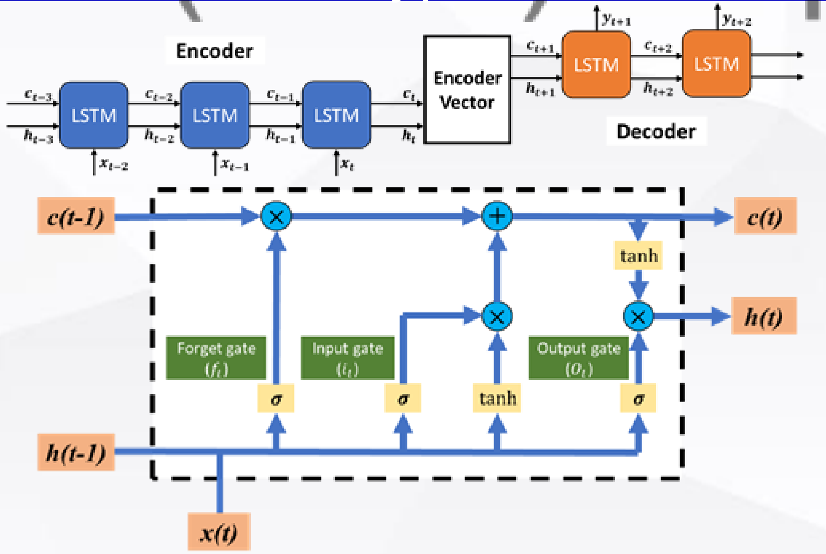
Schematic of the data-driven game-based energy sharing

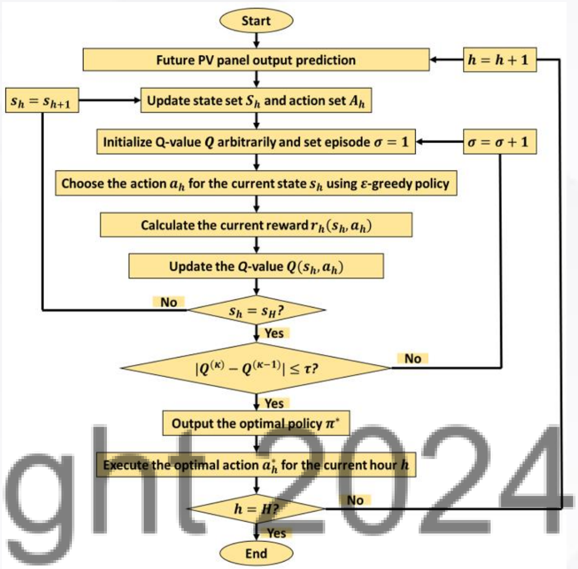
Stackelberg game-based energy sharing

- Leader: Rooftop PV system owner
- Followers: consumers

$$G=\begin{Bmatrix} (Owner \cup Building\ Users) \\ \{\lambda_h^U\},\{P_h^{ES_{in}}\},\{P_h^{ES_{grid}}\} \\ \{P_{ih}^{PV_{user}}\},\{P_h^{ES_{user}}\},\{P_{ih}^{G}\} \\ \{Rev_h^O\},\{U_{ih}^C\} \end{Bmatrix}$$

Long short-term memory (LSTM) based uncertainty prediction





Markov Decision Process (MDP)

- Agent: Rooftop PV system owner
- State: all system information
- Action: pricing strategies
- Reward: revenue

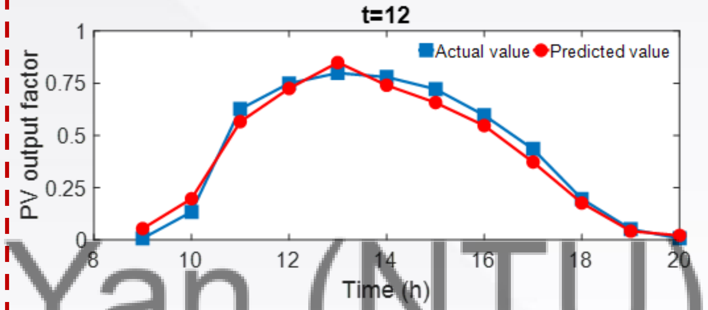
$$Rev^O=\sum_{h\in H}\begin{Bmatrix} \sum_{i\in N^C}\lambda_h^U(P_{ih}^{PV_{user}}+P_{ih}^{ES_{user}}) \\ +\lambda^{FiT}(P_h^{PV_{grid}}+P_h^{ES_{grid}}) \\ -\lambda_h^{TOU}[\sum_i(P_{ih}^{PV_{user}}+P_{ih}^{ES_{user}})-\overline{P}_h^{PV}]^+ \end{Bmatrix}$$

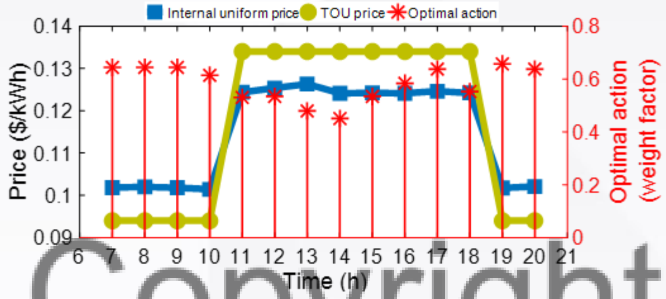
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Data-driven Energy Sharing among Buildings: Results

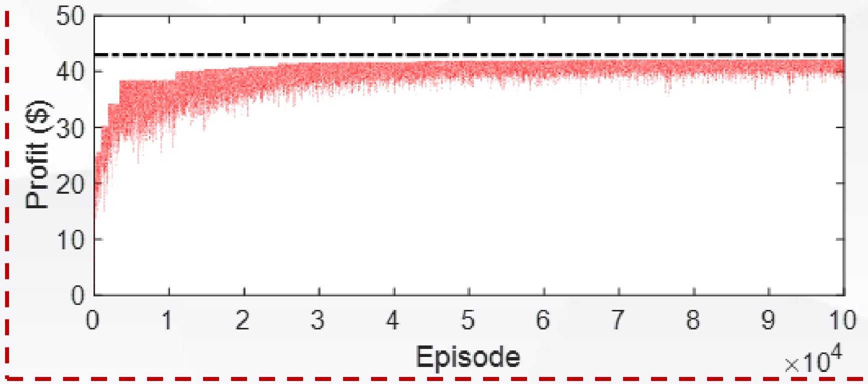
Performance of proposed method
- Accurate PV prediction
- High daily profit
- Well utilization of PV energy

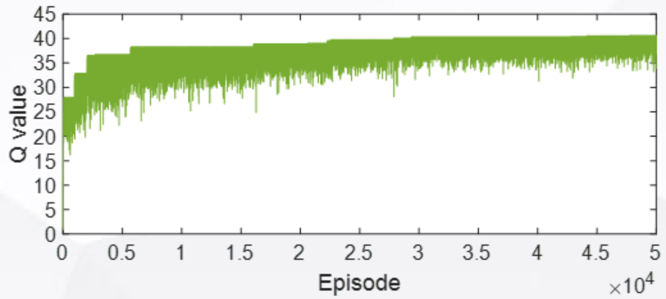
Pricing strategy	Daily profit ($)
Strategy 1: Internal uniform price	41.99
Strategy 2: TOU price	39.71
Strategy 3: Market clearing price	24.47





Comparison with optimization solvers
- High computation efficiency
- Near-optimal results

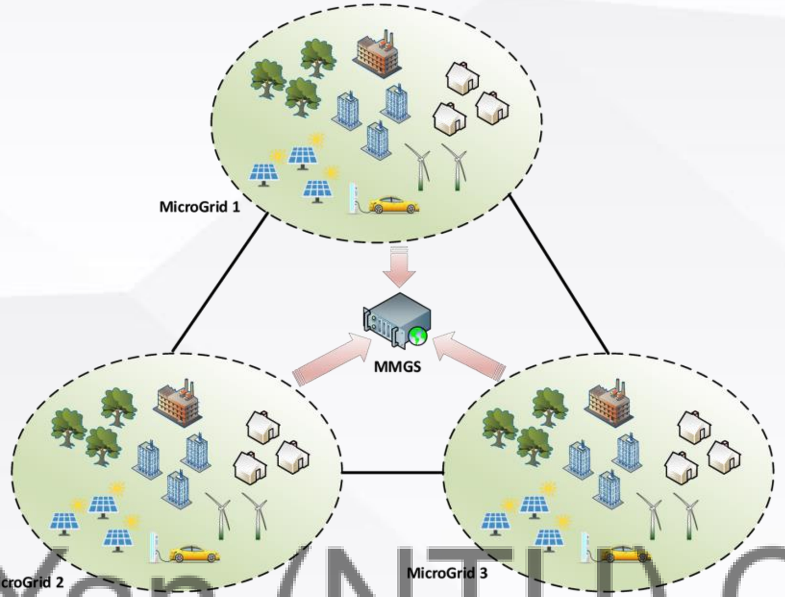




Solution method	Profit ($)	Computation time (s)
Conventional optimization method	43.075	3400.42
Q-learning algorithm	41.994	15.339

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Hierarchical Coordination of Networked-Microgrids



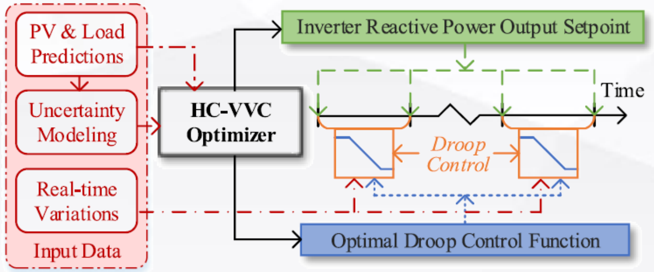
Source of pictures: website (searched in Google)

Uncertainty from renewable energy sources (RESs)

- **Penetration of RESs** (more complexity and uncertainty): flexible and self-adaptive controller is required.
- **Economy and efficiency**: integration of secondary and tertiary layer

P/Q coordination

- **Energy Management System (EMS):** P-management (microturbine, energy storage system, dispatchable load…)
- **Volt-Var Control (VVC)**: voltage regulation, network loss minimization (OLTC, …)



Data-driven (RL-based) methods

- Complexity induced by model-based methods/ uncertainty from RESs
- Resilient and adaptive approach for microgrid operation and control
- Only measurement data are required without parametric identification

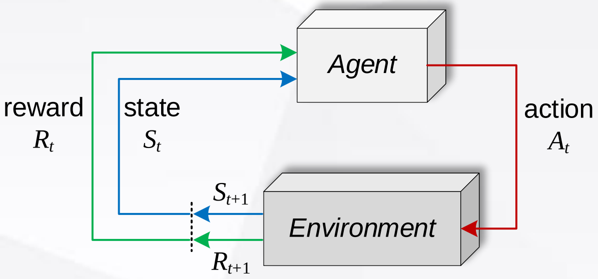
Reinforcement learning:

- ☐ Use “trial and error” interaction with a dynamic system to find **an optimal policy**
- ☐ State => Action => **Reward**

Action Exploration

conflict

Policy’s Safety



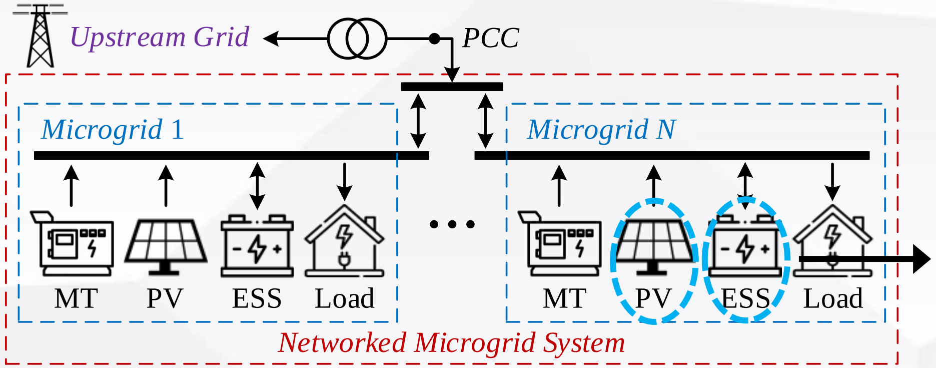
Design of algorithm

- ☐ Action exploration
- ☐ Operational constraints
- ☐ Achieve the **balance**

Y. Xia, **Y. Xu**, and X. Feng, “Hierarchical Coordination of Networked-Microgrids towards Decentralized Operation: A Safe Deep Reinforcement Learning Method,” ***IEEE Trans. Sustainable Energy***, **2022.**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Problem Formulation



Objective: minimize the total operation costs and network loss

P-management:

$$\boxed{C_{MGi}=\sum_{t=1}^{T}\left(C^t_{MTi}+C^t_{ESSi}+C^t_{TRi}\right)}$$ ***Minimize operation cost***

Cost of MT: $C^t_{MTi}=a_i\left(P^t_{MTi}\Delta t\right)^2+b_i\left(P^t_{MTi}\Delta t\right)+c_i$

Cost of ESS: $C^t_{ESSi}=d_i\left(P^t_{ESSi}\Delta t\right)^2$

Cost of energy trading: $C^t_{TRi}=\lambda^t_B P^t_{INi}\Delta t-\lambda^t_S P^t_{OUTi}\Delta t$

Q-management:

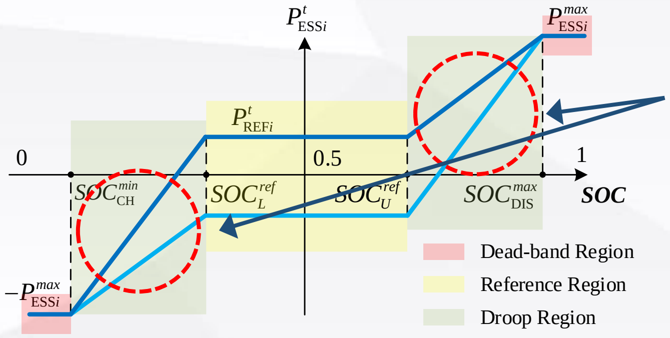
$$\boxed{\min\sum_t P^t_L}$$ ***Minimize network loss & voltage regulation***

$$V_{j+1}{}^2=V_j{}^2-2\left(R_jP_j+X_jQ_j\right)+\left(R_j{}^2+X_j{}^2\right)\frac{P_j{}^2+Q_j{}^2}{V_j^2}$$ *power flow relationship*

$$q^{g,t}_j\approx Q^t_{j,\mathrm{PV}},\ \left|Q^t_{j,\mathrm{PV}}\right|\le\sqrt{S_{j,\mathrm{PV}}{}^2-P^t_{j,\mathrm{PV}}{}^2}$$ *available reactive power capacity*

$$\boxed{V^{min}_j\le V^t_j\le V^{max}_j}$$ *Voltage constraint*

P/SoC droop-based ESS control

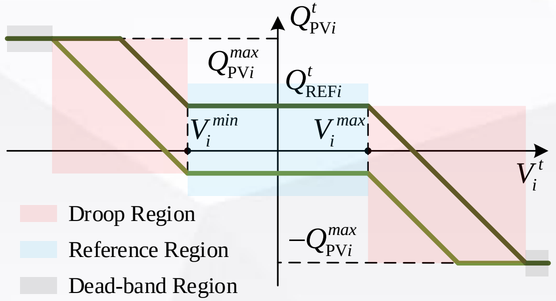


Objective: *restore SoC under extremely high/low conditions.*

$$P^t_{ESSi}=\begin{cases}-P^{max}_{ESSi}, & SoC<SoC^{min}_{CH}\\ P^t_{REFi}+K_{1,i}\left(SoC^{ref}_L-SoC\right), & SoC^{min}_{CH}\le SoC<SoC^{ref}_L\\ P^t_{REFi}, & SoC^{ref}_L\le SoC\le SoC^{ref}_U\\ P^t_{REFi}+K_{2,i}\left(SoC-SoC^{ref}_U\right), & \\ P^{max}_{ESSi}, & \end{cases}$$

Optimize P reference value
(purpose: minimize operation cost)

Q/V Droop-based PV Inverter Control



Objective*:* *restore the voltage if the voltage exceeds limits*

$$Q^t_{PVi}=\begin{cases}\min\left(Q^t_{REFi}+K_{V,i}\left(V^{min}_i-V^t_i\right),Q^{max}_{PVi}\right), & V^t_i<V^{min}_i\\ Q^t_{REFi}, & \\ \max\left(Q^t_{REFi}-K_{V,i}\left(V^t_i-V^{max}_i\right)\right. & \end{cases}$$

Optimize Q reference value
(purpose: minimize network loss)

Proposed Method

Markov Decision Process (MDP) Modeling

Reactive Power Management

state: $s_{Qi}^t = \left[\Delta V_i^t, \Delta p_{h,i}^{t-1}, \ldots, \Delta p_{h,i}^{t-t_D}, Q_{\mathrm{LOAD}i}^{t-1}, \ldots, Q_{\mathrm{LOAD}i}^{t-t_D}\right]$

$$\Delta p_{h,i}^{t-t_D} = P_{\mathrm{MT}i}^{t-t_D} + P_{\mathrm{ESS}i}^{t-t_D} + P_{\mathrm{PV}i}^{t-t_D} - P_{\mathrm{LOAD}i}^{t-t_D}$$

action: $a_{Qi}^t = Q_{\mathrm{REF}i}^t$ *reference Q output of PV inverter*

reward: $r_q^t = -\alpha_1 P_L^t$ *global reward (active power loss)*

Active Power Management

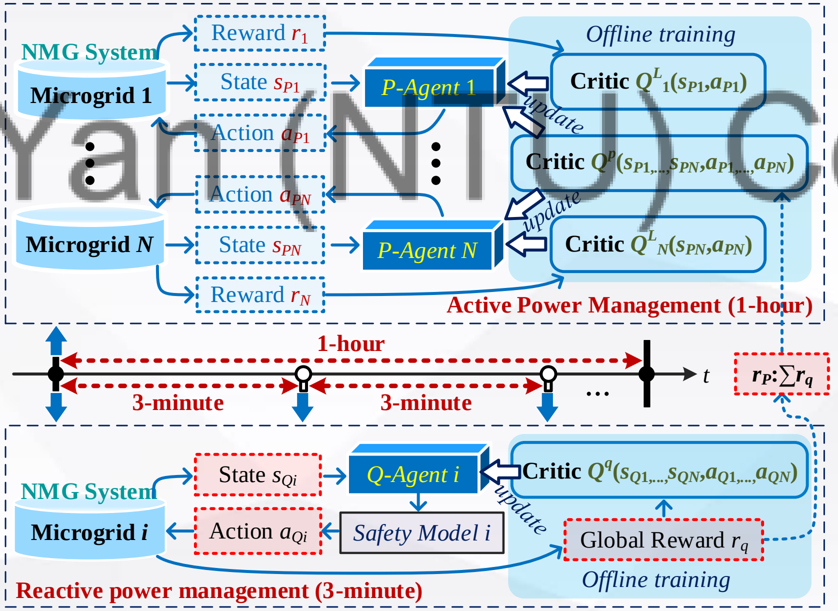
state: $s_{Pi}^T = \left[\lambda_B^T, \lambda_S^T, SoC_i^T, \Delta p_i^{T-1}, \Delta p_i^{T-2}, \ldots, \Delta p_i^{T-T_D}\right]$ $\Delta p_i^{T-T_D} = P_{\mathrm{PV}i}^{T-T_D} - P_{\mathrm{LOAD}i}^{T-T_D}$

action: $a_{Pi}^T = \left[P_{\mathrm{MT}i}^T, P_{\mathrm{REF}i}^T\right]$ *P outputs of MTs, reference P outputs of ESSs*

reward: $r_i^T = -\alpha_2 C_{\mathrm{MG}i}$ *local reward* $r_p^T = -\alpha_3 \lambda_B^T \sum^T r_q^t$ *global reward*

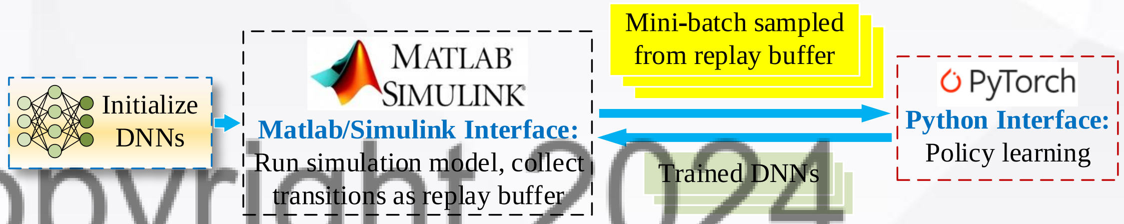
towards minimizing the local generation cost, also coordinate to minimize the total network loss.

Proposed Methodology



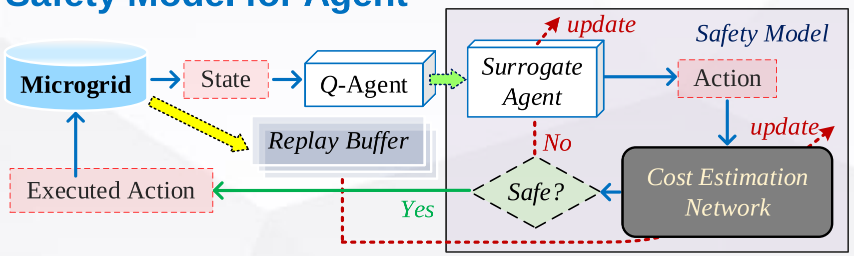
➢ **Centralized** offline training
➢ **Decentralized** online application

P outputs of MTs and ESSs are managed in **a slow time-scale (1-hour)**.
Q outputs of PV inverters are dispatched in **a fast time-scale (3-minute).**



Training Process (based on soft actor critic)

Safety Model for Agent



cost estimation network $\Phi(s_Q^t, a_Q^t|\Omega)$: $\Phi_{Tar}\left(s_Q^t, a_Q^t | \Omega\right) = \left(V^{t+1} - V_{\mathrm{REF}}\right)^2$

predict the condition of voltage violation

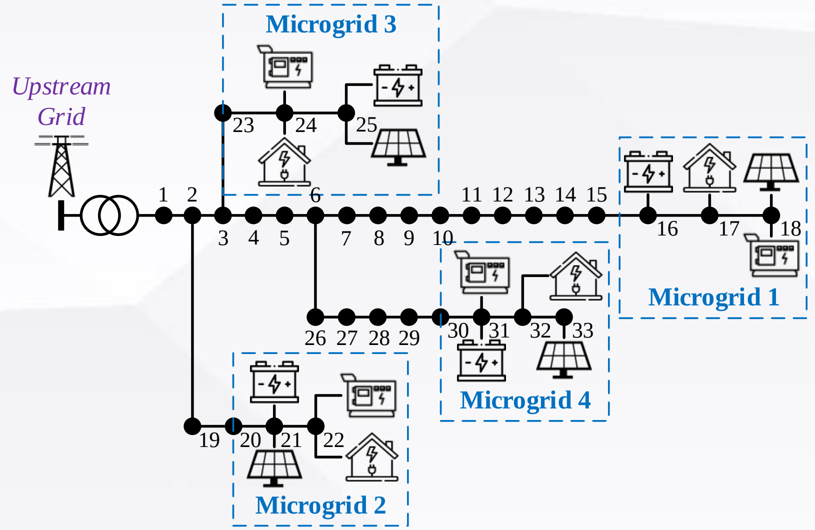
If the predicted value exceeds the threshold Φ_{th} (unsafe case):

A surrogate agent is generated and adjusted by minimizing the cost function value Φ:

$$\Omega_i \leftarrow \Omega_i - \lambda_S \nabla_{\Omega i} J_{\Phi i}(\Omega_i) \qquad J_S\left(\theta_{Qi}\right) = \mathop{\mathrm{E}}_{s_{Qi}\in R}\left[\Phi_i\left(s_{Qi}, \pi_{Qi}\left(s_{Qi}|\theta_{Qi}\right)|\Omega_i\right)\right]$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Test Results

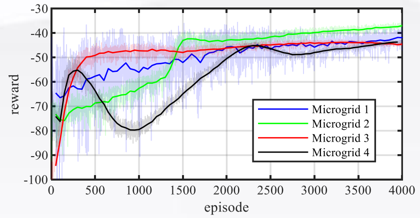


NMG test model based on IEEE 33-bus distribution network

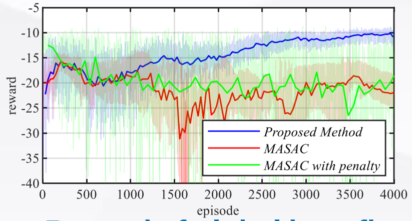
Offline Training Results

TABLE Parameters of Offline Training

Parameters	Value	Parameters	Value
Learning rate λ_Φ	1e−4	$\boldsymbol{R}_p$ size	2×10^4
Learning rate λ_Q	1e−4	$\boldsymbol{R}_q$ size	5×10^4
Learning rate λ_π	1e−5	m_p size	2×10^3
Learning rate λ_S	1e−3	m_q size	5×10^3
Threshold Φ_{th}	0.05	Maximum N_{epi}	5×10^4
Discount factor γ	0.99	Entropy weight τ	0.02

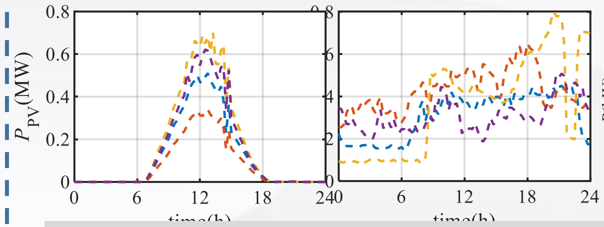


Reward of local benefit

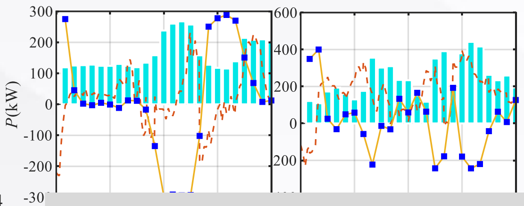


Reward of global benefit

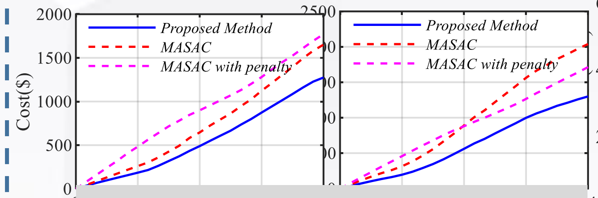
Simulation Results



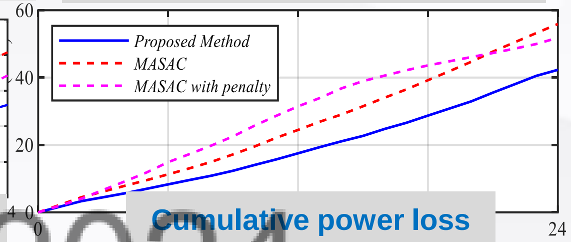
Input profiles of PV, load historic data



Active power outputs
(a) microgrid 1 (b) microgrid 2

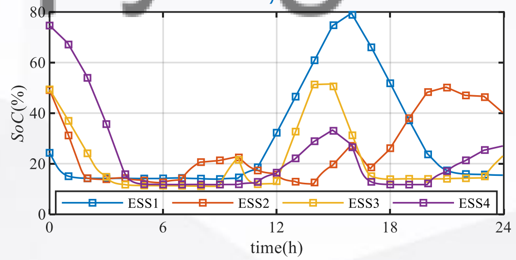


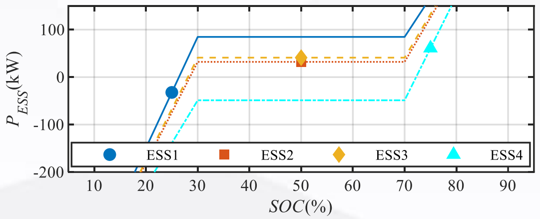
Cumulative cost for microgrids
(a) microgrid 1 (b) microgrid 2



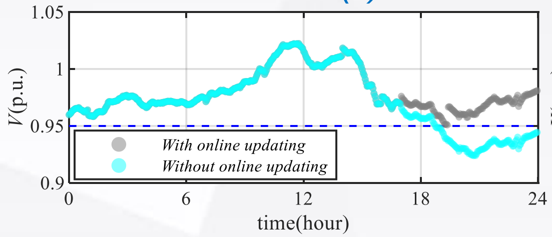
Cumulative power loss
for the NMG system

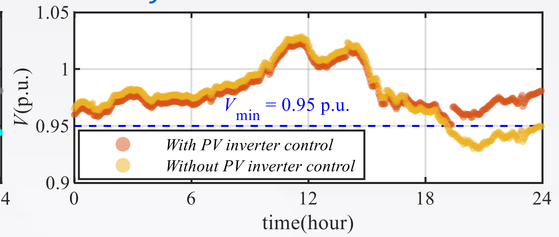
ESS SoC curves, and control curves with corresponding operating points





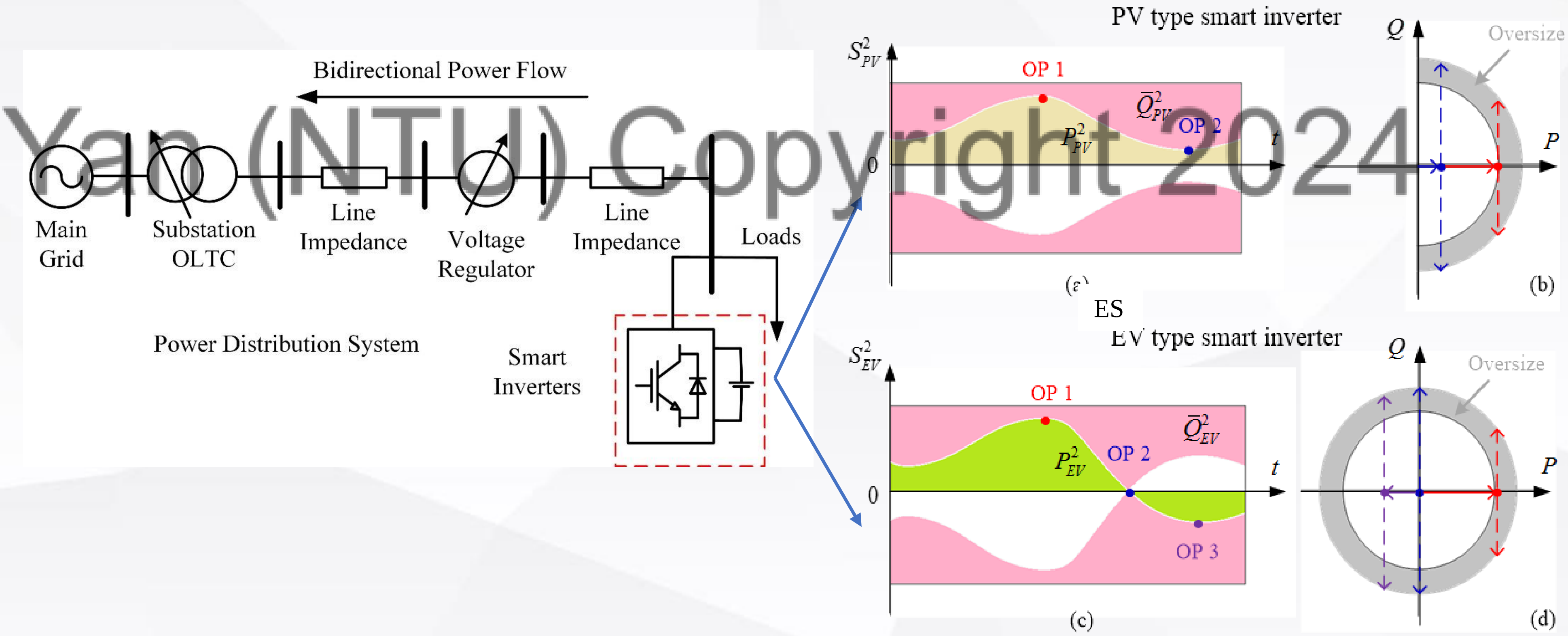
Node voltages (a) with/without PV inverter control
(b) with/without the safety model.





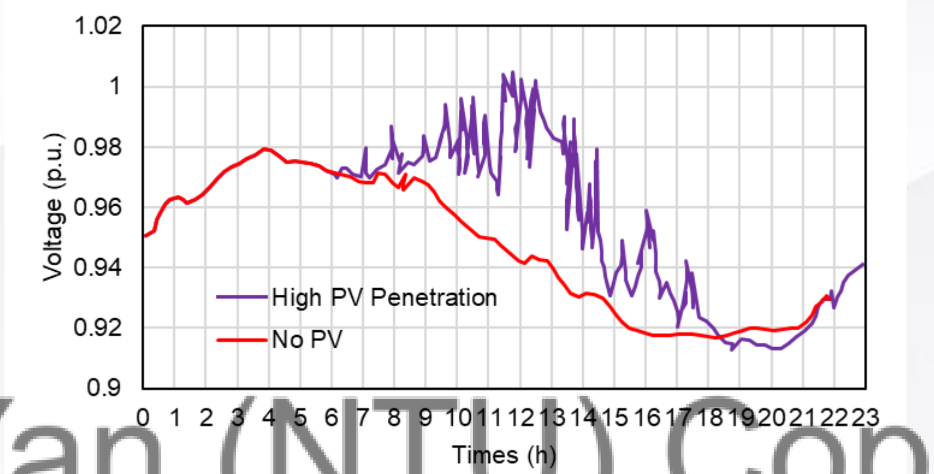
Real-time Voltage/Var Control (VVC) Support from DERs

- **Existing Challenges:** High PV penetration level, massive EV charging.
- **Voltage quality issues:** Voltage rise, drop and fast fluctuations.
- **Potential solutions:** inverter-assisted voltage/var support



Voltage/Var Control (VVC) in Active Distribution Network

Impacts of RES high penetration level on VVC in distribution networks



An example of voltage fluctuations with high PV penetration

High level of RES will lead to:
- Voltage fluctuations and violations
- High network energy loss

Traditional VVC Device
- On-load Tap Changers (OLTC)
- Step Voltage Regulator (SVR)
- Capacitor Bank (CB)

Problems: mechanical apparatuses cannot respond quickly to fast voltage fluctuations caused PV generation.

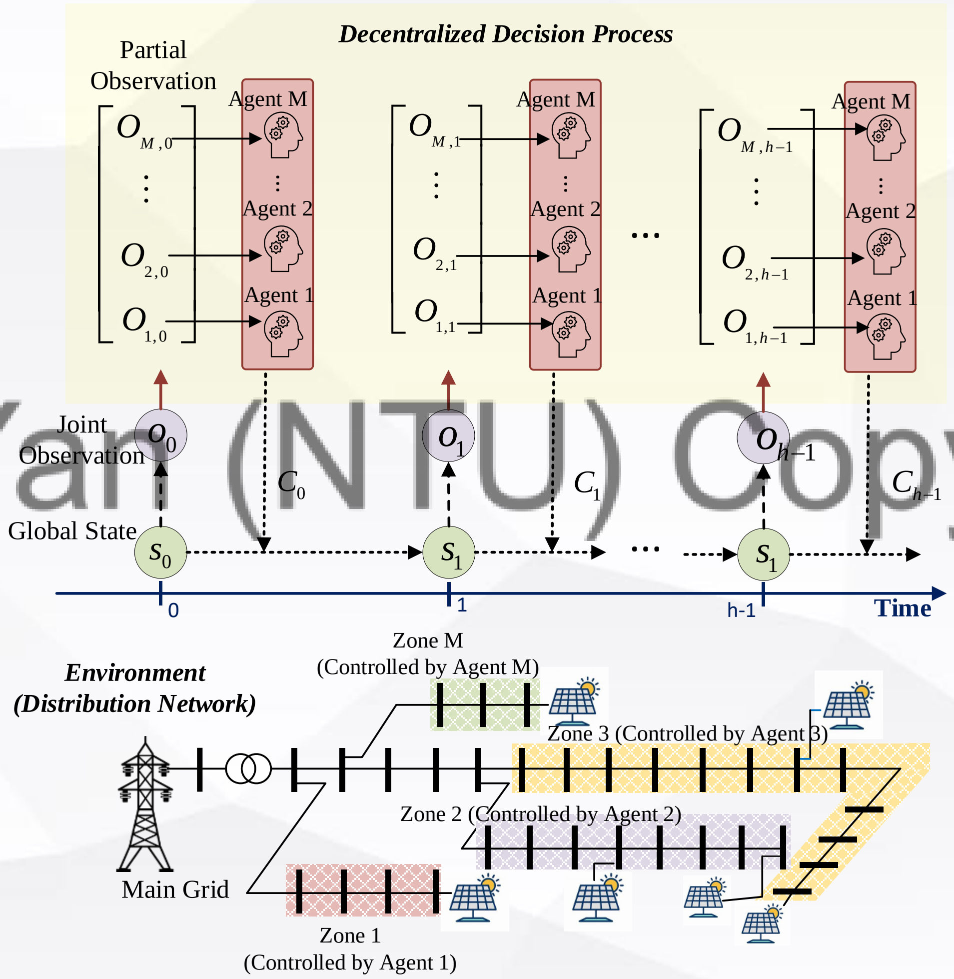
Power Electronics Based Device
- Static Var Compensator (SVC)
- STATCOM
- **PV Inverter (High Flexibility)**

Centralized, Distributed and Decentralized VVC

Centralized	Distributed	Decentralized
▪ Central controller/optimizer	▪ Neighboring communication ▪ Consensus based method, ADMM, accelerated ADMM, etc.	▪ Based on local measurements ▪ Reduced communication requirement ▪ Faster control actions

Voltage/Var Control (VVC) in Active Distribution Network

PV Inverter based Decentralized VVC



- Each Zone has a central control agent/controller.
- Each agent controls the reactive power output of all the inverters in the zone.

❖ **Input System State (Zone Information)**

$$O = \otimes O_m \qquad O_{m,t} = [X_1^m, X_2^m, \dots, X_{N_m^B}^m]^T$$

$$X_{i_m}^m = [V_{i_m,m}, P^{PV}_{i_m,m}, P^{L}_{i_m,m}, Q^{inv}_{i_m,m}, Q^{L}_{i_m,m}]$$

❖ **Output Dispatch Command (Zonal Var dispatch)**

Scaling Factor

$$C = \otimes C_m$$

$$C_m = \left[c_{1,m}, c_{2,m}, \dots, c_{N_m^{PV},m}\right]^T \qquad \boxed{c_{q,m} \in [-1,1]}$$

$$Q^{inv}_{q,m} = c_{q,m} \boxed{Q^{inv,\max}_{q,m}} = c_{q,m} \boxed{\sqrt{(S^{cap}_{q,m})^2 - (P^{PV}_{q,m})^2}}$$

Maximum available reactive power output of the q[th] inverter in m[th] zone

Formulation of constrained Markov Decision Process

❖ **Problem Formulation**

$$\max_{\pi=\{\pi_1,\pi_2,\dots,\pi_M\}} \sum_{m=1}^{M} E\left[\sum_{t=0}^{h-1} \gamma^t r_t^m\right]$$

$$s.t.\quad F(\pi_m) \le d \qquad \forall m = 1,2,\dots,M$$

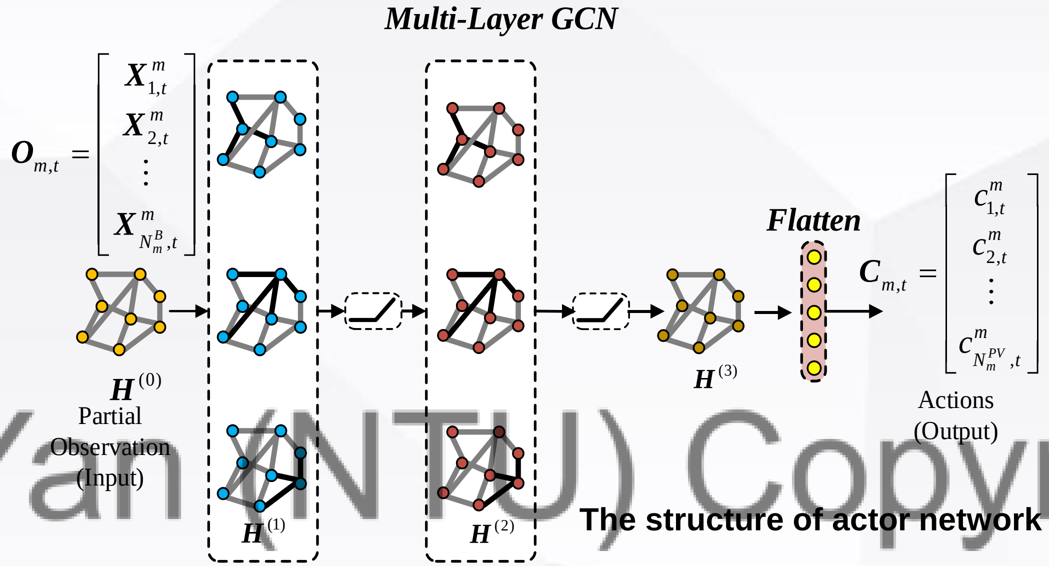
❖ **Lagrangian Relaxation**

$$L(\pi,\lambda) = \sum_{m=1}^{M}\left[R(\pi_m) - \lambda_m\left(F(\pi_m) - d\right)\right]$$

$$\lambda = [\lambda_1, \lambda_2, \dots, \lambda_M],\ \pi = \{\pi_1, \pi_2, \dots, \pi_M\}$$

Voltage/Var Control (VVC) in Active Distribution Network

❑ Data-Driven Method (Multi-Agent Primal Dual Graph Reinforcement Learning)



The structure of actor network

- Actor Network: Graph Convolutional Network (GCN) Embedded

Input: System Operation State

(node features organized as graph-structured data)

$$O_{m,t} = [X^m_{1,t}, X^m_{2,t}, \ldots, X^m_{N^B_m,t}]^T$$

$$X^m_{i_m,t} = [V_{i_m,m,t}, P^{PV}_{i_m,m,t}, P^L_{i_m,m,t}, Q^{inv}_{i_m,m,t}, Q^L_{i_m,m,t}]$$

Output: Var Dispatch Command

$$C_m = \left[c_{1,m}, c_{2,m}, \ldots, c_{N^{PV}_m,m}\right]^T, c_{q,m} \in [-1,1]$$

❖ Feedforward Calculation Process

Message-Passing Scheme: Use neighboring data to fill missing value

Graph Fourier Transform: Low-Pass Filter

$$\begin{cases} H^{(0)}_{m,t} = O_{m,t} \\ H^{(l+1)}_{m,t} = \sigma\left(\boxed{\hat{D}_m^{-\frac{1}{2}} \hat{A}_m \hat{D}_m^{-\frac{1}{2}}} H^{(l)}_{m,t} \boxed{W^{(l)}_{G,m}}\right), \ l = 0,1,\ldots,L-1 \\ h_{m,t} = flatten(H^{(L)}_{m,t}), \ h_{m,t} \in R^{N^B_m}, H^{(L)}_{m,t} \in R^{N^B_m} \times R \\ C_{m,t} = W_{F,m} h_{m,t}, \ W_{F,m} \in R^{N^{PV}_m} \times R^{N^B_m} \end{cases}$$

A_m is the adjacency matrix, D_m is diagonal matrix of node degrees

❖ Loss Function & Updating Process

$$L^P(\phi_m) = \frac{1}{K}\sum_{k=1}^{K} Q^R_{\varphi^R_m}(O_{m,k}, C_{m,k}) - \lambda_m Q^C_{\varphi^C_m}(O_{m,k}, C_{m,k})$$

$$= \frac{1}{K}\sum_{k=1}^{K} Q^R_{\varphi^R_m}(O_{m,k}, \mu_{\phi_m}(O_{m,k})) - \lambda_m Q^C_{\varphi^C_m}(O_{m,k}, \mu_{\phi_m}(O_{m,k}))$$

Main actor network: $\phi_m \leftarrow \phi_m + \alpha \nabla_{\phi_m} L^P(\phi_m)$

Target actor network: $\phi'_m \leftarrow \tau\phi_m + (1-\tau)\phi'_m$

R. Yan, Q. Xing and **Y. Xu***, "Multi Agent Safe Graph Reinforcement Learning for PV Inverters Based Real-Time Decentralized Volt/Var Control in Zoned Distribution Networks," ***IEEE Transactions on Smart Grid***, **May 2023.**

Voltage/Var Control (VVC) in Active Distribution Network

Data-Driven Method (Multi-Agent Primal Dual Graph Reinforcement Learning)

- Critic Networks: Reward Critic Network and Cost Critic Network

Description	Reward Critic Network	Cost Critic Network
Neural Network Parameters	φ_m^R	φ_m^C
Input	system state $\boldsymbol{s}$, action $\boldsymbol{a}$	system state $\boldsymbol{s}$, action $\boldsymbol{a}$
Output	reward Q value $Q^R_{\varphi_m^R}(\boldsymbol{s},\boldsymbol{a})$	cost Q value $Q^C_{\varphi_m^C}(\boldsymbol{s},\boldsymbol{a})$
Loss Function	reward function $L^R(\varphi_m^R)=\frac{1}{K}\sum_{k=1}^{K}[r_k+\gamma Q^R_{\varphi'^R_m}-Q^R_{\varphi_m^R}]^2$	cost function $L^C(\varphi_m^C)=\frac{1}{K}\sum_{k=1}^{K}[F_k+\gamma Q^C_{\varphi'^C_m}-Q^C_{\varphi_m^C}]^2$
Updating Rules	$\varphi_m^R\leftarrow\varphi_m^R-\alpha\nabla_{\varphi_m}L^R(\varphi_m^R)$	$\varphi_m^C\leftarrow\varphi_m^C-\alpha\nabla_{\varphi_m^C}L^C(\varphi_m^C)$

- Primal-Dual Policy Optimization Algorithm

❖ Primal Space: policy π updating

$$\varphi_m^R\leftarrow\varphi_m^R-\alpha\nabla_{\varphi_m}L^R(\varphi_m^R)$$

$$\phi_m\leftarrow\phi_m+\alpha\nabla_{\phi_m}L^P(\phi_m)$$

$$\phi_m\leftarrow\phi_m+\alpha\nabla_{\phi_m}L^P(\phi_m)$$

Update primal policy and dual variable in turn

❖ Dual Space: dual variable λ updating

$$\lambda_m\leftarrow\left[\lambda_m+\beta\frac{1}{K}\sum_{k=1}^{K}\left(Q^C_{\varphi_m^C}(\boldsymbol{O}_{m,k},\mu_{\phi_m}(\boldsymbol{O}_{m,k}))-d\right)\right]^+$$

$$\lambda_m\geq 0$$

1. Overview

2. Power Systems

2.1 Frequency control

2.2 Optimal power flow

2.3 Topology optimization

3. Microgrids

3.1 Frequency control

3.2 Controller tuning

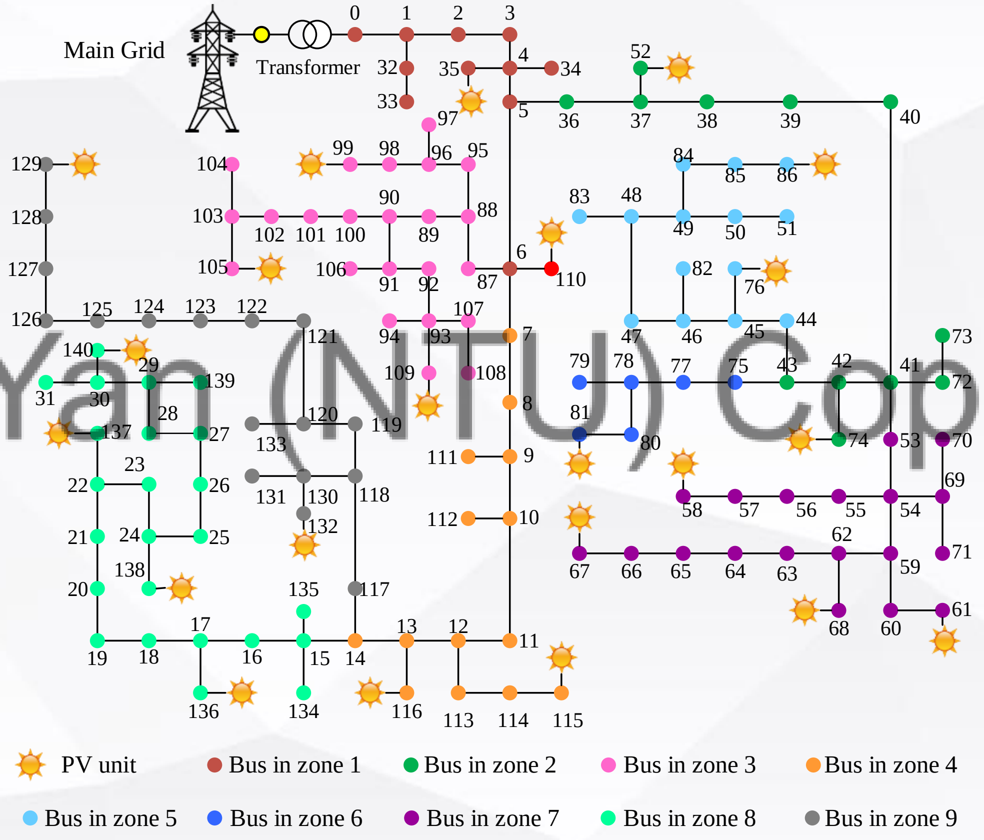
3.3 Energy management

3.4 Volt/Var control

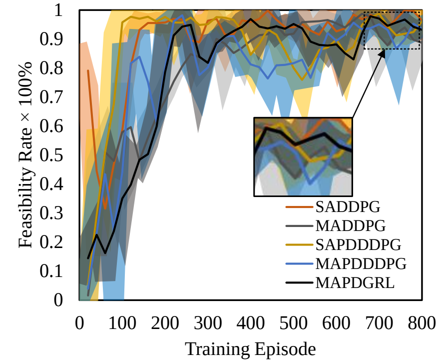
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

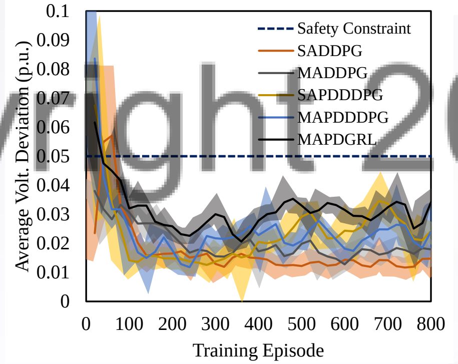
Voltage/Var Control (VVC) in Active Distribution Network

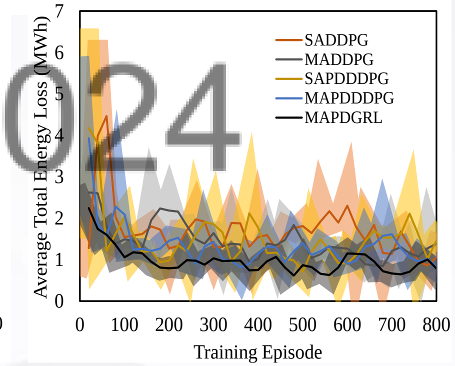
Case Study



141-bus distribution network with 9 zones and 22 PV units.





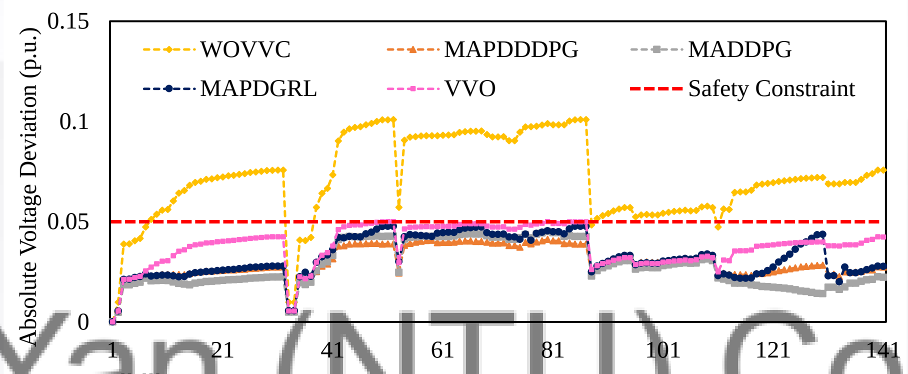


Method	Training Time
Single Agent DDPG (SADDPG)	1h 49min 54s
Multi-Agent DDPG (MADDPG)	2h 54min 5s
Single Agent PDDDPG (SAPDDDPG)	4h 34min 35s
Multi-Agent PDDDPG (MAPDDDPG)	5h 22min 33s
Multi-Agent PDGRL (MAPDGRL)	**6h 32min 50s**

Voltage/Var Control (VVC) in Active Distribution Network

Simulation Results on 141-bus system

Testing Performance in Ideal Scenario (no noise and missing data)



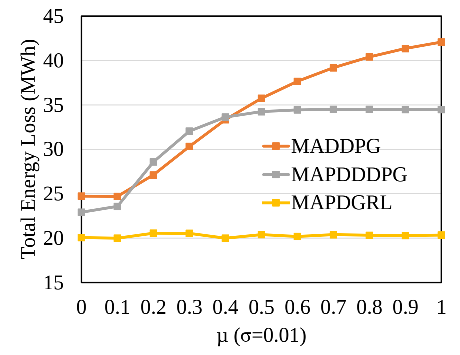
FR: Feasibility Rate

Method	Without VVC	MADDPG	MAPDDDPG	MAPDGRL
FR	36.09 %	45.24 %	97.48 %	95.95 %

† theoretical best result, * proposed method.

Methods	Network Energy Loss	Max Voltage Deviation	Voltage Violation Rate	Computation Time (Each Step)
WO VVC	**16.22 MWh**	**0.101 p.u.**	**46.57 %**	**--**
VVO †	**10.11 MWh**	**0.05 p.u.**	**0%**	**2.6 mins**
SADDPG	16.51 MWh	0.0403 p.u.	0%	25 ms
MADDPG	15.89 MWh	0.0452 p.u.	0%	17 ms
SAPDDDPG	19.94 MWh	0.05 p.u.	0%	27 ms
MAPDDDPG	15.22 MWh	0.05 p.u.	0%	18 ms
MAPDGRL*	**12.61 MWh**	**0.05 p.u.**	**0%**	**81 ms**

Robustness against Noise $\tilde{\boldsymbol{O}}_m = \boldsymbol{O}_m + \boldsymbol{\mathcal{N}}(\mu, \sigma)$



♦ Disturbed by noise

Methods	Network Energy Loss
MAPDDDPG	15.22 MWh
♦MAPDDDPG	35.49 MWh
MAPDGRL	**12.61 MWh**
♦MAPDGRL	**16.218 MWh**

Robustness against Missing Data

♦ Disturbed by noise

Methods	Network Energy Loss	Maximum Voltage	Minimum Voltage
MAPDDDPG	16.96 MWh	1.028 p.u.	0.979 p.u.
♦MAPDDDPG	19.51 MWh	1.019 p.u.	0.979 p.u.
MAPDGRL	**12.14 MWh**	**1.028 p.u.**	**0.978 p.u.**
♦MAPDGRL	**13.16 MWh**	**1.029 p.u.**	**0.973 p.u.**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

▪ Publication List – Bulk Power Grids

Frequency control

1. Z. Yan and Y. Xu*, “Data-driven Load Frequency Control for Stochastic Power Systems: A Deep Reinforcement Learning Method with Continuous Action Search,” *IEEE Trans. Power Syst*., vol. 34, no. 2, pp. 1653-1656, Mar. 2019. **– Web of Science highly cited paper, started in 2018**
2. Z. Yan and Y. Xu*, “A Multi-Agent Deep Reinforcement Learning Method for Cooperative Load Frequency Control of Multi-Area Power Systems,” *IEEE Trans. Power Syst*., vol. 35, no. 6, pp. 4599-4608, Nov. 2020.
3. Z. Yan, Y. Xu*, Y. Wang, and X. Feng, “Deep reinforcement learning-based optimal data-driven control of battery energy storage for power system frequency support,” *IET Gen. Trans. & Dist*., no.14, pp. 6071-6078, 2020.

Real-Time Operation

1. Z. Yan and Y. Xu*, “Real-Time Optimal Power Flow: A Lagrangian based Deep Reinforcement Learning Approach,” *IEEE Trans. Power Syst*., vol.35, no.4, pp.3270-3273, Jul. 2020.
2. Z. Yan and Y. Xu*, “A Hybrid Data-driven Method for Fast Solution of Security-Constrained Optimal Power Flow,” *IEEE Trans. Power Syst*., vol. 37, no. 6, pp. 4365-4374, Nov. 2022.
3. Z. Yan and Y. Xu*, “Real-Time Optimal Power Flow with Linguistic Stipulations: Integrating GPT-Agent and Deep Reinforcement Learning,” *IEEE Trans. Power Syst*., vol. 39, no. 2, pp. 4747-4750, March 2024.

Topology optimization

1. Y. Zheng, Z. Yan, K. Chen, J. Sun, Y. Xu, and Y. Liu, “Vulnerability Assessment of Deep Reinforcement Learning Models for Power System Topology Optimization”, *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3613-3623, July 2021.
2. Z. Yan and Y. Xu*, “结合深度强化学习与领域知识的电力系统拓扑结构优化,” *《电力系统自动化》*, 2021.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Publication List – Microgrids & Active Distribution Networks

Microgrid control

1. Y. Xia, Y. Xu*, Y. Wang, S. Mondal, S. Dasgupta, A. Gupta, and G. Gupta, “A Safe Policy Learning-Based Method for Decentralized and Economic Frequency Control in Isolated Networked-Microgrid Systems,” *IEEE Trans. Sustainable Energy*, vol. 13, no. 4, pp. 1982-1993, Oct. 2022.
2. R. Yan, Y. Wang, Y. Xu*, and J. Dai, “A Multi-Agent Quantum Deep Reinforcement Learning Method for Distributed Frequency Control of Islanded Microgrids,” *IEEE Trans. Control of Network Systems*, vol. 9, no. 4, pp. 1622-1632, Dec. 2022.
3. Y. Xia, Y. Xu*, Y. Wang, W. Yao, S. Mondal, S. Dasgupta, A. Gupta, and G. Gupta, “A Data-Driven Method for Online Gain Scheduling of Distributed Secondary Controller in Time-Delayed Microgrids,” *IEEE Trans. Power Syst*., vol. 39, no. 3, pp. 5036-5049, May 2024.

Microgrid Energy Management & Volt/Var Control

1. X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C.S. Lai, “A Multi-agent Reinforcement Learning based Data-driven Method for Home Energy Management,” *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201-3211, July 2020. **– Web of Science highly cited paper**
2. X. Xu, Y. Xu, M. Wang, Z. Xu, J. Li, and S. Chai, “Data-driven Game-based Pricing for Sharing Rooftop Photovoltaic Generation and Energy Storage in the Residential Building Cluster under Uncertainties,” *IEEE Trans. Industrial Informatics*, vol. 17, no. 7, pp. 4480-4491, July 2021.
3. Y. Xia, Y. Xu*, “Hierarchical Coordination of Networked-Microgrids towards Decentralized Operation: A Safe Deep Reinforcement Learning Method,” *IEEE Trans. Sustainable Energy*, 2024.
4. R. Yan, Q. Xing, and Y. Xu*, “Multi-Agent Safe Graph Reinforcement Learning for PV Inverters Based Real-Time Decentralized Volt/Var Control in Zoned Distribution Networks,” *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 299-311, Jan. 2024.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SODA

Stability & Optimization & Data-Analytics For MORDERN POWER SYSTEMS

Thank You!