Scalable Chance-Constrained Optimisation for Reliable Electricity Markets

Dr Yihong Zhou
University of Oxford
20 Jan 2026

Content

- Background
- Optimisation with Uncertainty in Constraints
- Speeding-up Chance-Constrained Optimisation through Valid Cuts
- Speeding-up Decision-Making through AI
- Summary

Background

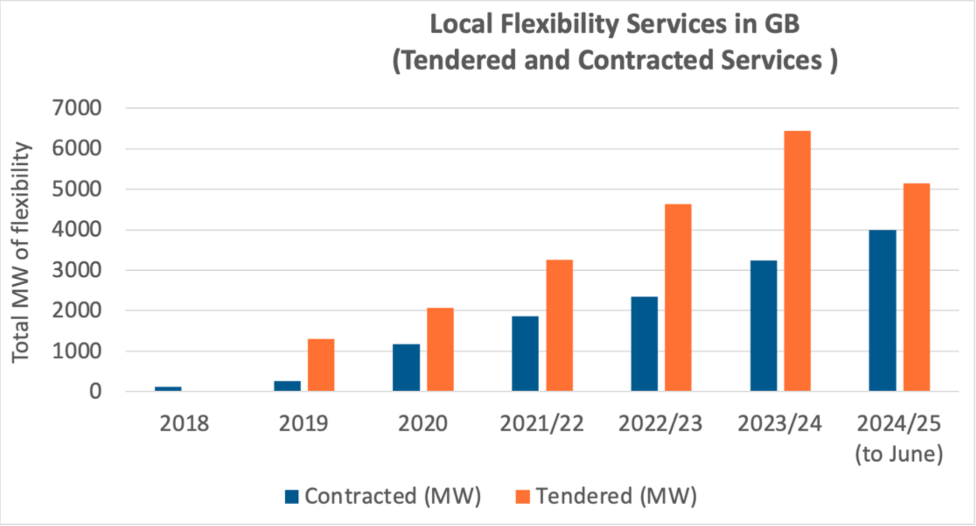
Electricity Markets and Optimisation

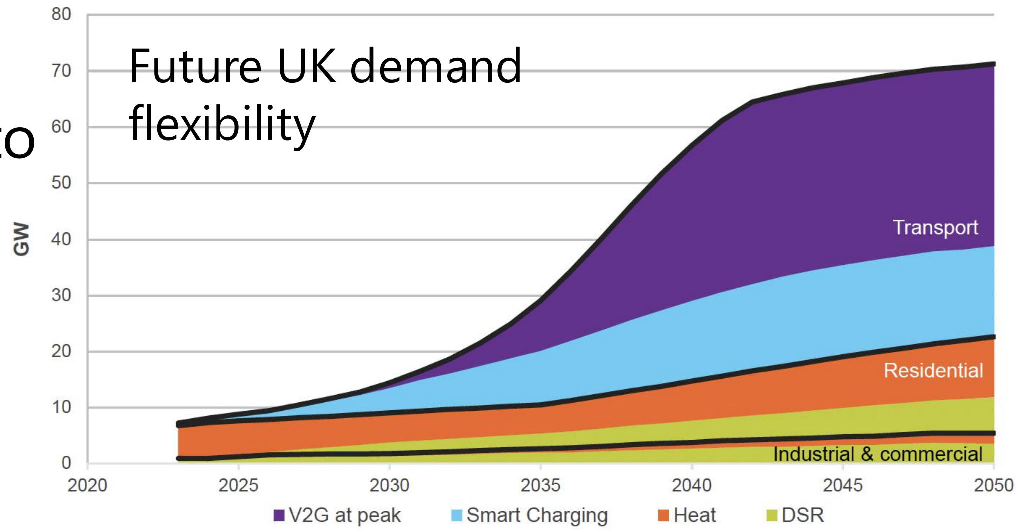
- Mathematical optimisation is usually used for market clearing.
- The objective is to maximise the social welfare.
- Constraints include market rules and **physical network and reliability constraints**.
- Market prices are derived from the optimisation outcome
  - e.g., from the optimal dual variables under pay-as-clear designs.

What Will Change in Future Electricity Markets

1. Increasing Number of Participants

- Millions of flexible grid-edge devices will be added:
  - 15m electric vehicles by 2030 (UK alone)
  - 600k heat pumps per year from 2028 (UK alone)
  - Increasing home solar paired with batteries
- These devices can provide **valuable flexibility** to the grid (USD $270bn globally based on IEA).
- However, the large number of participants also **dramatically increases the scale** of the market optimisation problems.



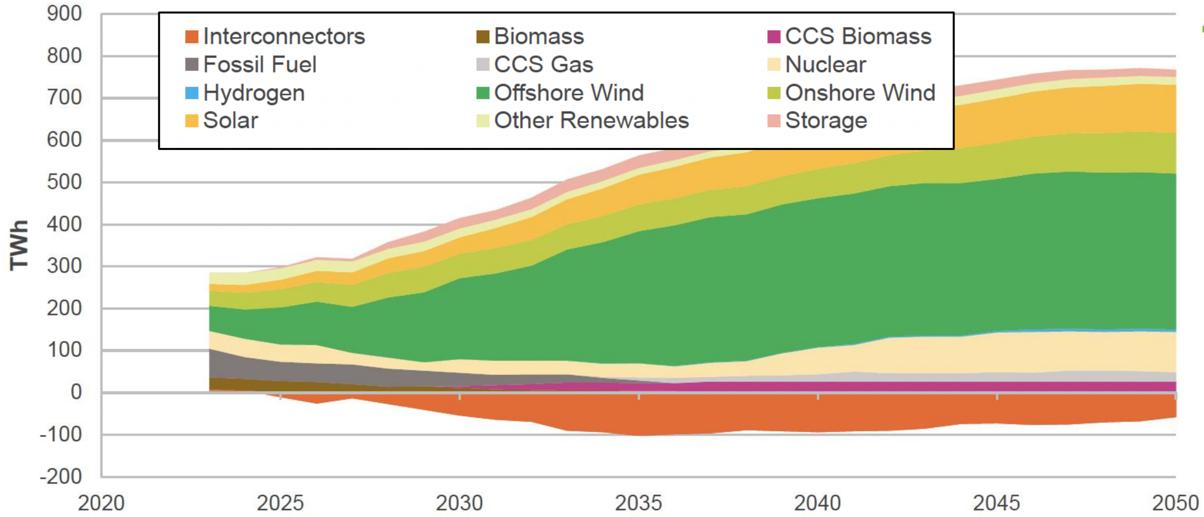


UK Future Energy Scenarios (NESO, 2024)

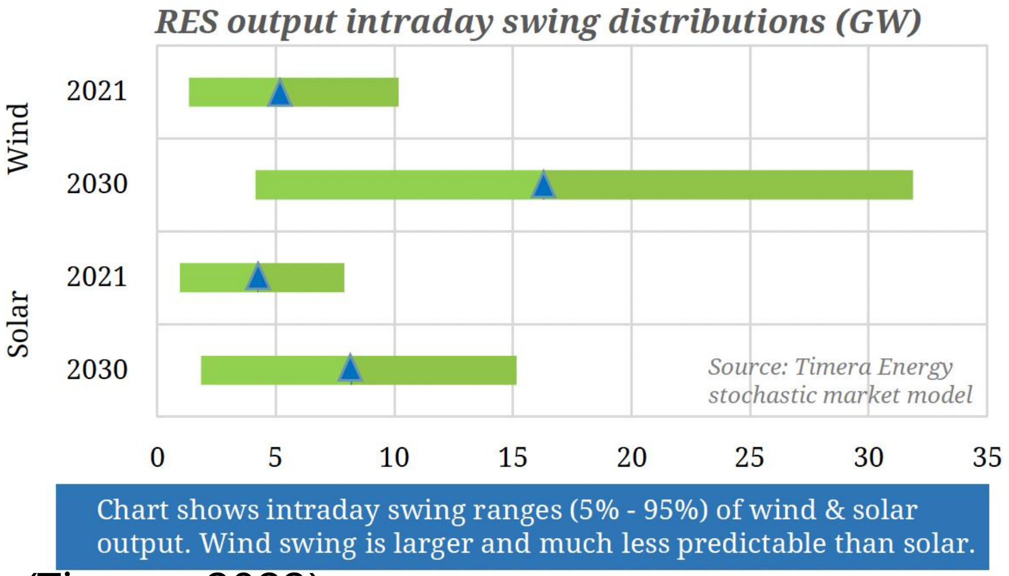
What Will Change in Future Electricity Markets

2. Increasing Renewables with High Uncertainty

- There is also increasing renewable energy sources (RES) in the grid
- However, this creates challenges to reliable power grid operation
  - Increased generation variability and uncertainty (which can be **three times** larger than load uncertainty, based on NREL).



UK Future Energy Scenarios (NESO, 2024)



(Timera, 2022)

Challenges and Needs in Future Electricity Markets

- The two concurrent trends are making optimisation problems in electricity markets:
  - have much larger problem scale;
  - have more uncertain parameters with greater magnitude.
- Therefore, future electricity markets **require computationally scalable and uncertainty-aware optimisation tools**!

For system operator, particularly ***scalable (no need for optimal) uncertainty management in these safety-critical hard constraints.***

Since optimisation is widely involved in electricity markets, when thinking about how markets can deal with uncertainty (*especially uncertainty that affects safety constraints*), **we can start from a more fundamental optimisation perspective**.

Optimisation with Uncertainty in Constraints

Optimisation with Uncertainty in Constraints

To deal with uncertain parameters $\boldsymbol{\xi}$ in **hard** constraints, one can either

1. Ensure no-violation for all possible scenarios, i.e., robust optimisation (RO);
2. Ensure violation below small probability ϵ, i.e., chance constraints (CC):

$$\min_{\boldsymbol{x}\in\mathcal{X}} \quad c(\boldsymbol{x})$$
$$\text{s.t.} \quad \mathbb{P}[\boldsymbol{\xi} \notin \mathcal{S}(\boldsymbol{x})] \leq \epsilon.$$

- $\mathbb{P}$ is a probability measure for $\boldsymbol{\xi}$. $\mathcal{S}(\boldsymbol{x})$ is a set of constraints that couple $\boldsymbol{x}$ and $\boldsymbol{\xi}$.
- Individual CC (ICC) when $|\mathcal{S}(\boldsymbol{x})| = 1$, joint CC (JCC) when $|\mathcal{S}(\boldsymbol{x})| > 1$.
- **JCC is often desired due to the greater safety (also TODAY's main focus).** **Unfortunately JCC is often intractable and nonconvex.**

Chance-Constrained Optimisation

- Why JCC instead of robust optimisation (RO)?
  - More intuitive parameter setting. Directly control the failure probability.
  - Can be more **computationally efficient** under **a small amount of data**.
    - RO needs to introduce many ancillary variables (polytope uncertainty set) or conic constraints (ellipsoid).
- Why JCC instead of multi-stage stochastic programming (MSSP)?
  - MSSP may need to add slack variables to ensure constraints satisfaction for the whole scenario tree (ensure hard satisfaction **less strictly/explicitly**)
  - JCC-based market clearing (with decision rule approx) is more interpretable (Dvorkin, Y., 2019)

Dvorkin, Y., 2019. A chance-constrained stochastic electricity market. *IEEE Transactions on Power Systems*, *35*(4), pp.2993-3003.

Terminology: RHS and LHS

JCC can be divided into those with right-hand-side (RHS) uncertainty and left-hand-side (LHS) uncertainty.

1. RHS means the safety set $\mathcal{S}(\boldsymbol{x})$ can be expressed as a **separable** form:

$$\mathcal{S}(\boldsymbol{x}) := \left\{ \boldsymbol{\xi} \mid \boldsymbol{b}_p^\top \boldsymbol{\xi} + d_p - \boldsymbol{a}_p^\top \boldsymbol{x} \geq 0, \ p \in [P] \right\},$$

   - $\boldsymbol{b}_p$, d_p, and $\boldsymbol{a}_p$ are parameters. $\boldsymbol{\xi}$ is the random vector, and $\boldsymbol{x}$ is the decision vector.
   - There are P constraints indexed by $[P] = \{1, \cdots, P\}$ and thus a JCC.
   - The $\boldsymbol{x}$ and $\boldsymbol{\xi}$ appear in separate terms and thus RHS.

2. LHS means the decision variable $\boldsymbol{x}$ is multiplied with the random vector $\boldsymbol{\xi}$:

$$\mathcal{S}(\boldsymbol{x}) := \left\{ \boldsymbol{\xi} \mid (\boldsymbol{b}_p - \boldsymbol{A}_p^\top \boldsymbol{x})^\top \boldsymbol{\xi} + d_p - \boldsymbol{a}_p^\top \boldsymbol{x} \geq 0, \ p \in [P] \right\},$$

*Other terms in $\mathcal{S}(\boldsymbol{x})$ are known parameters.

Wasserstein Ambiguity

Another question naturally arises: *How can we get the distribution* $\mathbb{P}$*?*

- Collect historical data? This will require A LOT OF data to have a confident estimate of $\mathbb{P}$.
- And usually we only have an insufficient amount of data. In this case there is **ambiguity** towards the true $\mathbb{P}$!

A Wasserstein joint chance constraint (WJCC) can be expressed as:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{X}} \quad & c(\boldsymbol{x}) \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{F}_N(\theta)} \mathbb{P}[\boldsymbol{\xi} \notin \mathcal{S}(\boldsymbol{x})] \le \epsilon, \end{aligned}$$

where $\mathcal{F}_N(\theta)$ is a collection of probability distributions that are within θ-Wasserstein distance to the historical data we have.

Wasserstein Ambiguity

- Why Wasserstein instead of other ambiguity set?
  - **Avoid over-conservativeness**. Moment-based ambiguity set can include distributions with implausible shapes.
  - **Stronger out-of-sample performance**. A better metric compared to KL divergence.

Speeding-up Chance-Constrained Optimisation through Valid Cuts

Proposed SFLA and FICA

- Exact WJCC formulation is mixed-integer (NP-hard), but we don’t really need an exact reformulation.
- We propose SFLA (for RHS) and FICA (for LHS with partial special structure), two convex (linear) inner approximations.
    - *Proved*: Less ancillary constraints than standard CVaR approximation.
    - *Proved*: Less or equally conservative compared to CVaR.
    - *Proved*: Approaching exact reformulation under **a small amount of data**.
- The link to CVaR provides a better interpretation: compared to JCC, CVaR additionally controls **the tail expectation of the violation magnitude**.

[1] Yihong Zhou, Hanbin Yang, and Thomas Morstyn. "FICA: Faster Inner Convex Approximation of Chance Constrained Grid Dispatch with Decision-Coupled Uncertainty." *arXiv preprint arXiv:2506.18806* (2025).
[2] Yihong Zhou, Yuxin Xia, Hanbin Yang, and Thomas Morstyn. "Strengthened and faster linear approximation to joint chance constraints with wasserstein ambiguity." *arXiv preprint arXiv:2412.12992* (2024).

Proposed SFLA and FICA

Existing linear approximation (LA) or CVaR

$$s \geq 0, \boldsymbol{r} \geq \boldsymbol{0},$$

$$\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N,$$

$$\kappa_i \left(\frac{\boldsymbol{b}_p^\top \boldsymbol{\xi}_i + d_p - \boldsymbol{a}_p^\top \boldsymbol{x}}{\|\boldsymbol{b}_p\|_*} \right) \geq s - r_i, \forall i \in [N], p \in [P].$$

Proposed SFLA

$$s \geq 0, \boldsymbol{r} \geq \boldsymbol{0},$$

$$\epsilon N s - \sum_{i \in [N]} r_i \geq \theta N,$$

$$\kappa_i \left(\frac{\boldsymbol{b}_p^\top \boldsymbol{\xi}_i + d_p - \boldsymbol{a}_p^\top \boldsymbol{x}}{\|\boldsymbol{b}_p\|_*} \right) \geq s - r_i, \forall i \in [N]_p, p \in [P],$$

$$\frac{q_p + d_p - \boldsymbol{a}_p^\top \boldsymbol{x}}{\|\boldsymbol{b}_p\|_*} \geq s, \forall p \in [P].$$

where we have $q_p :=$ the $(\lfloor \epsilon N \rfloor + 1)$-th smallest element of the set $\{\boldsymbol{b}_p^\top \boldsymbol{\xi}_i\}_{i \in [N]}$, and $[N]_p := \{i \in [N] \mid \boldsymbol{b}_p^\top \boldsymbol{\xi}_i < q_p\}$. This form is also known as quantile strengthening.

[1] Yihong Zhou, Hanbin Yang, and Thomas Morstyn. "FICA: Faster Inner Convex Approximation of Chance Constrained Grid Dispatch with Decision-Coupled Uncertainty." *arXiv preprint arXiv:2506.18806* (2025).

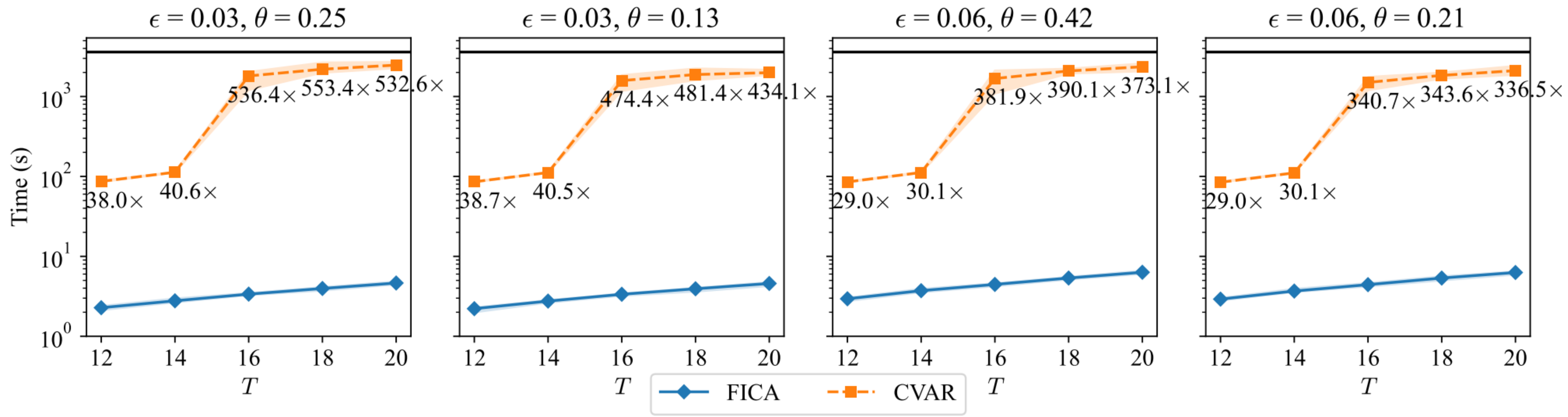
[2] Yihong Zhou, Yuxin Xia, Hanbin Yang, and Thomas Morstyn. "Strengthened and faster linear approximation to joint chance constraints with wasserstein ambiguity." *arXiv preprint arXiv:2412.12992* (2024).

SFLA and FICA Case Study Setting

- Wind-uncertainty-aware economic dispatch.
  - This closely resembles market-clearing formulations, using generation costs rather than bid-based representations.
  - Automatic generation factor (AGC, or participation factors) α are decision variables.

$$
\begin{aligned}
\min_{\boldsymbol{p},\boldsymbol{\alpha}} \quad & \mathcal{C}(\boldsymbol{p},\boldsymbol{\alpha}) && && \text{(1a)}\\
\text{s.t.} \quad & \boldsymbol{p},\ \boldsymbol{\alpha} \in \mathcal{X}, && && \text{(1b)}\\
& \mathbf{1}^\top(\boldsymbol{p}_t - \boldsymbol{d}_t) = 0, && \forall t \in [T], && \text{(1c)}\\
& \mathbf{1}^\top \boldsymbol{\alpha}_t = 1, && \forall t \in [T], && \text{(1d)}\\
& -\mathbf{1} \le \boldsymbol{\alpha}_t \le \mathbf{1}, && \forall t \in [T], && \text{(1e)}\\
& \underline{\boldsymbol{p}}_t \le \boldsymbol{p}_t \le \overline{\boldsymbol{p}}_t, && \forall t \in [T], && \text{(1f)}\\
& \inf_{\mathbb{P}\in\mathcal{F}(\theta)} \mathbb{P}\left\{ \begin{aligned} & \underline{\boldsymbol{p}}_t \le \boldsymbol{p}_t - \sum_{w\in\mathcal{W}} e_{t,w}\boldsymbol{\alpha}_t \le \overline{\boldsymbol{p}}_t, \quad \forall t \in [T], \\ & \underline{\boldsymbol{f}} \le \boldsymbol{S}^G \tilde{\boldsymbol{p}}_t + \boldsymbol{S}^W(\boldsymbol{\omega}_t + \boldsymbol{e}_t) - \boldsymbol{S}^D \boldsymbol{d}_t \le \overline{\boldsymbol{f}}, \forall t \in [T] \end{aligned} \right\} \ge 1-\epsilon, && && \text{(1g)}
\end{aligned}
$$

SFLA and FICA Case Study Results



- When the optimisation horizon T exceeds 16, FICA achieves 500$\times$ speedup compared to CVaR in solving the LHS-WJCC economic dispatch problem.
- Besides, we have observed FICA being super memory-saving, due to the $[N]_p$ set that reduces the N constraints in $[N]$ to $\lfloor \epsilon N \rfloor$.

* The proportion of the one-dimensional constraints is only 50%.

Speeding-up Decision-Making through AI

AI for Greater Computational Efficiency

- Although SFLA and FICA provide speedup, it remains hard to scale to millions+ grid-edge devices.
- AI represents a promising complement as it enables possibilities to train a **neural network surrogate for the chance constrained optimisation solver**.
  - This may need intense offline training but fast online decision-making.
  - A typical example that aligns with our context is safe reinforcement learning (SRL).

$$
\begin{aligned}
\text{Maximize} \quad & V^{\pi}(s_0) = \mathbb{E}_{\pi}\left[\textstyle\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)\right] \\
\text{Subject to} \quad & \Pr\left[\textstyle\sum_{t=0}^{\infty} \gamma^t c(S_t, A_t) > d_{th}\right] \le \epsilon_0 \\
& \text{for } S_0 = s_0, A_t \sim \pi(\cdot|S_t), S_{t+1} \sim M(\cdot|S_t, A_t).
\end{aligned}
$$

Jung, W., Cho, M., Park, J. and Sung, Y., 2022. Quantile constrained reinforcement learning: A reinforcement learning framewo rk constraining outage probability. *Advances in Neural Information Processing Systems*, *35*, pp.6437-6449.

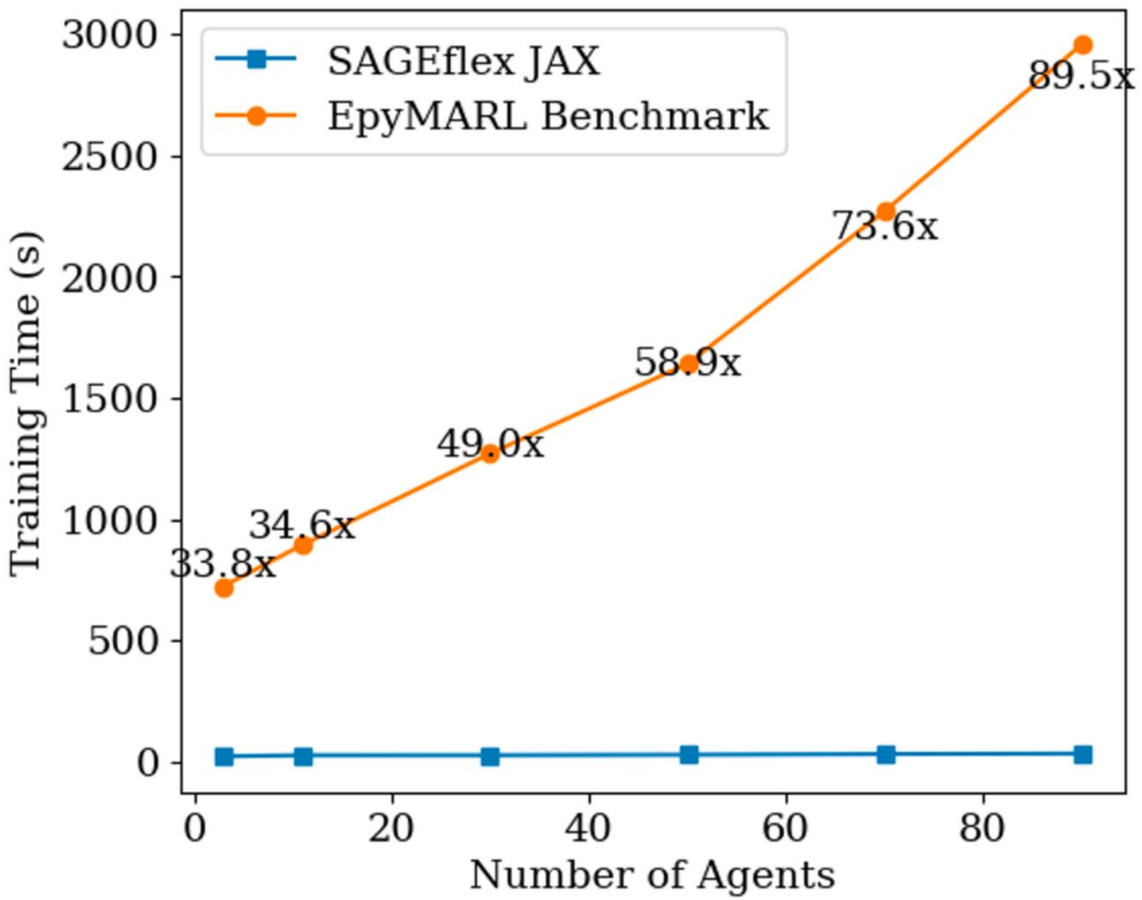
Computationally Efficient Training/Testing Platform

- The development of AI (including RL) requires a large amount of experiments.
- We have a "SAGEflex: Safeguarded AI Agents for Grid-Edge Flexibility" project to speed up this process.
- One of our targets is to develop a **software platform** for training, testing, and benchmarking, with **JAX for end-to-end GPU acceleration.**

SAGEflex is funded by the Advanced Research + Invention Agency (ARIA)

Computationally Efficient Training/Testing Platform

- Comparison with another software platform “EpyMARL” on MARL training.
  - The MARL algorithm is IPPO.
  - EpyMARL only uses CPU (faster than using GPU for their implementation)
  - SAGEflex JAX uses GPU both for the env and RL



Summary

Main Takeaways

- The large number of grid-edge devices and the increasing uncertainty motivate computationally scalable optimisation tools for future electricity markets.
- Chance constraints provide a principled way to incorporate reliability under uncertainty.
  - We propose SFLA and FICA that achieve 100x speedup over CVaR, without being more conservative.
- AI offers new opportunities to scale these ideas to future markets.
  - Our SAGEflex project is developing a platform to support its development.

Thank you!
Questions?