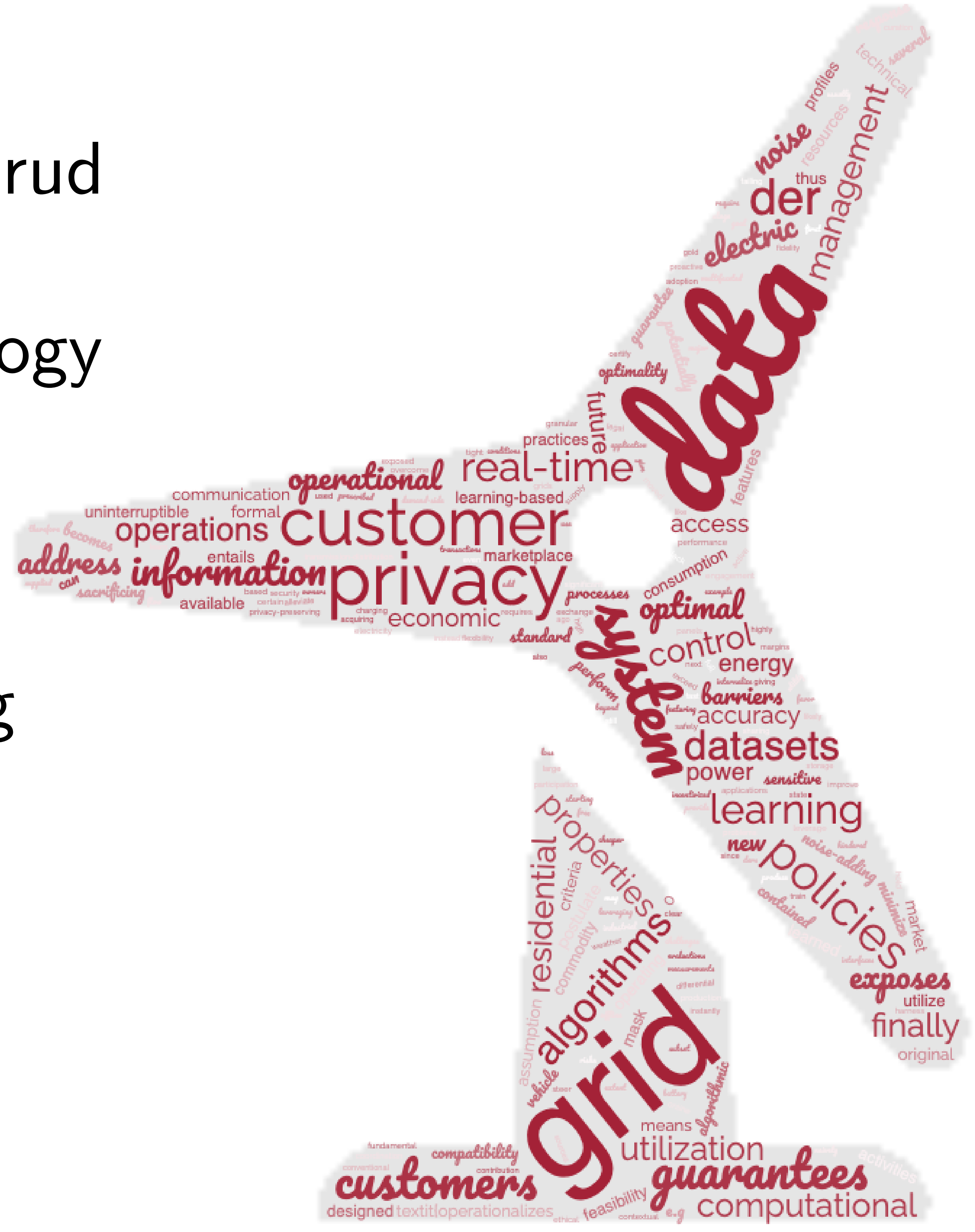
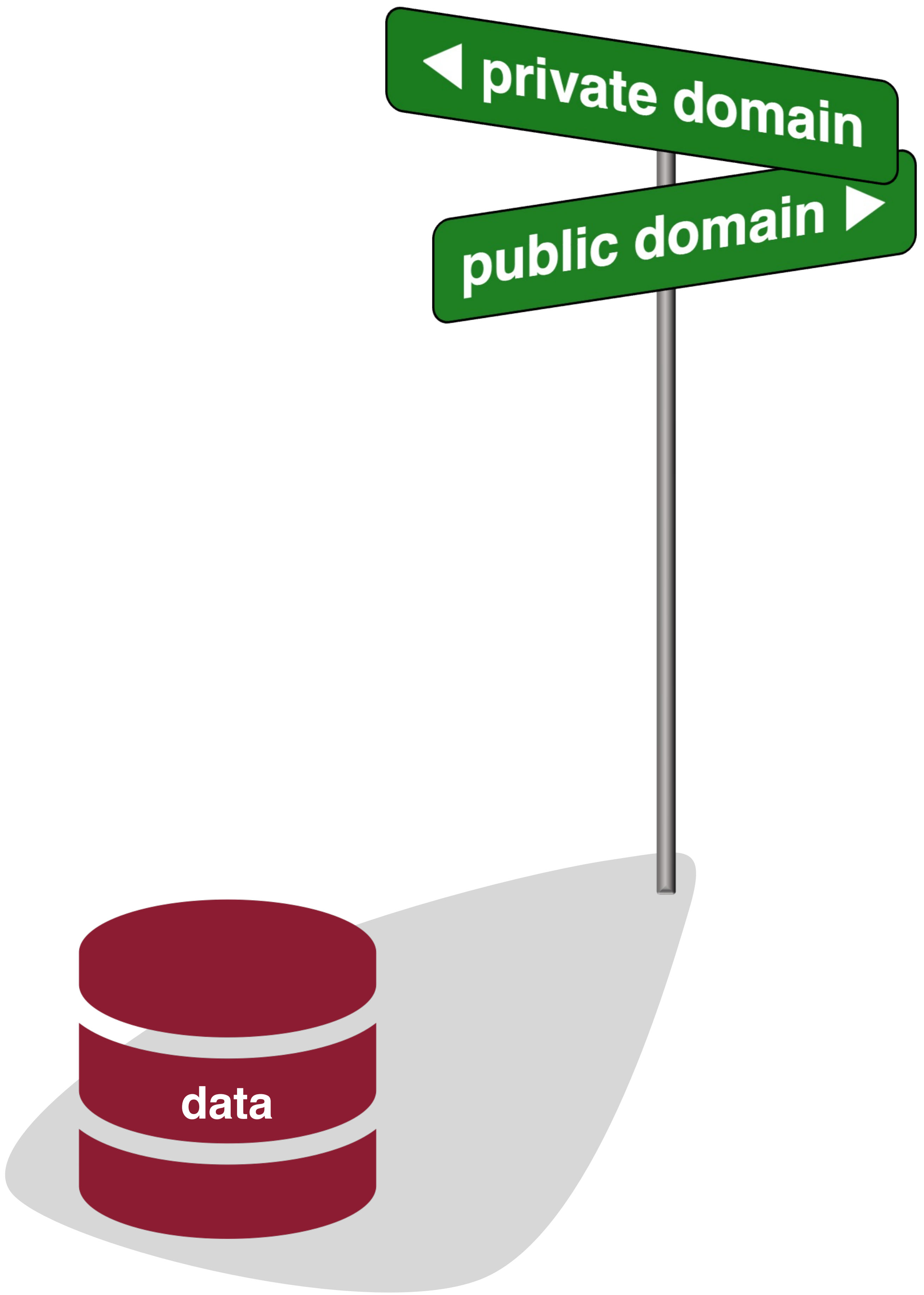
Differentially Private Algorithms for Synthetic Power System Datasets

Vladimir Dvorkin and Audun Botterud
Energy Initiative & LIDS
Massachusetts Institute of Technology

2023 INFORMS Annual Meeting
October 17, 2023



Where data should go?



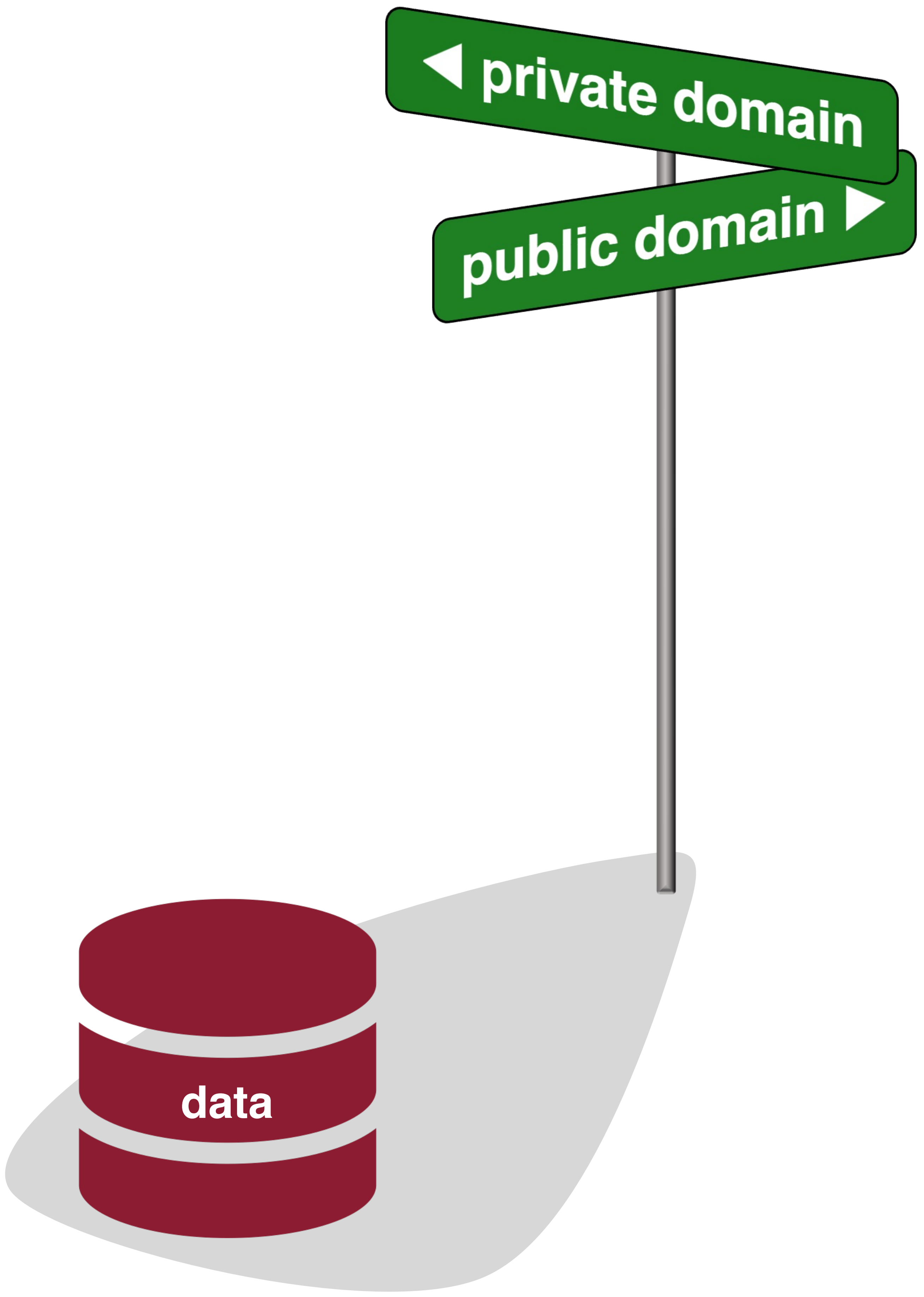
Arguments in favor of **private** data:

- Privacy and security
- Regulatory compliance
- Competitive advantage

Arguments in favor of **public** data:

- Improved decision-making
- Less barriers for entry
- Innovation, research

Where data should go?



Arguments in favor of **private** data:

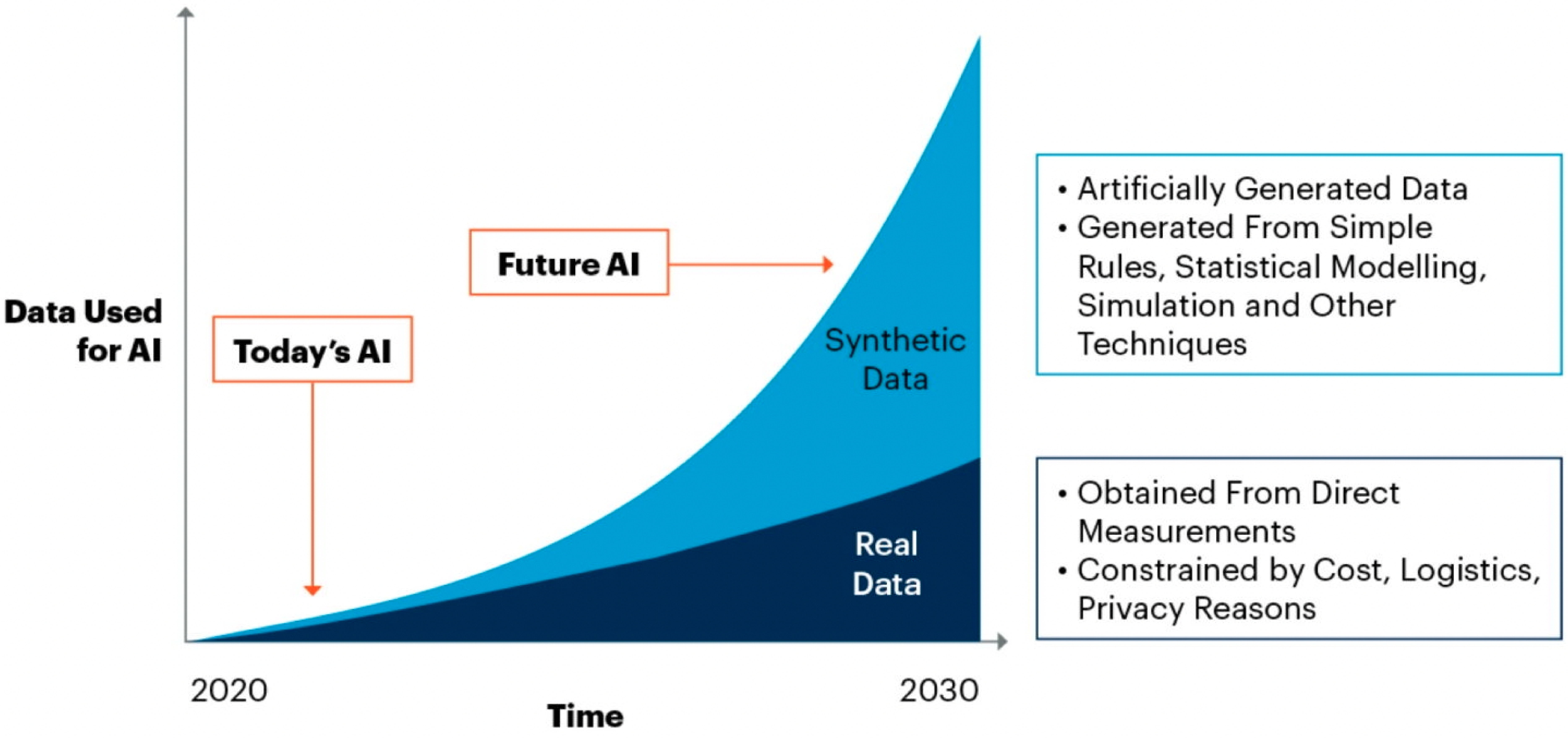
- Privacy and security
- Regulatory compliance
- Competitive advantage

Arguments in favor of **public** data:

- Improved decision-making
- Less barriers for entry
- Innovation, research

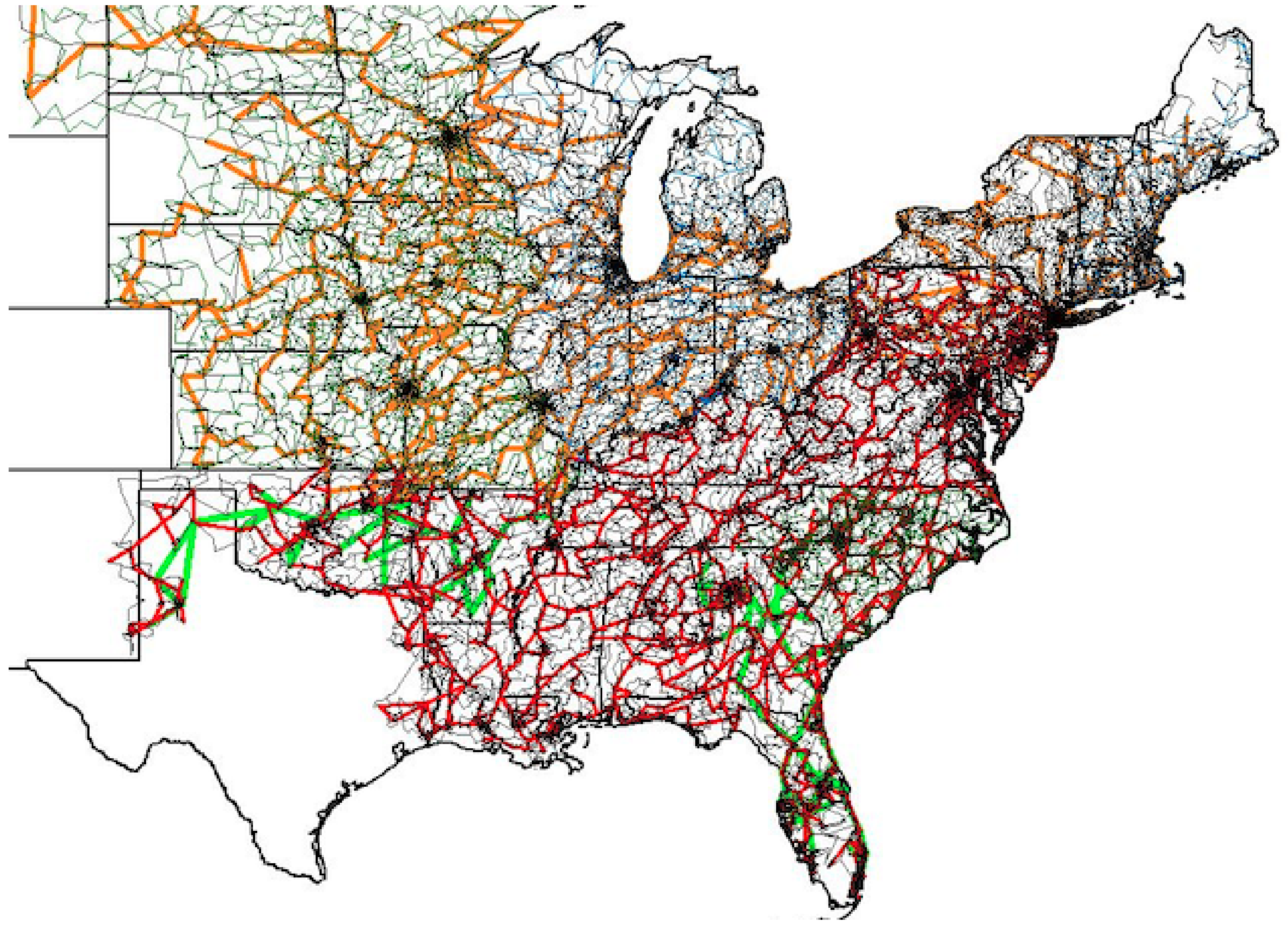
Synthetic data serves as a middle ground !

By 2030 synthetic data will completely overshadow real data in AI models

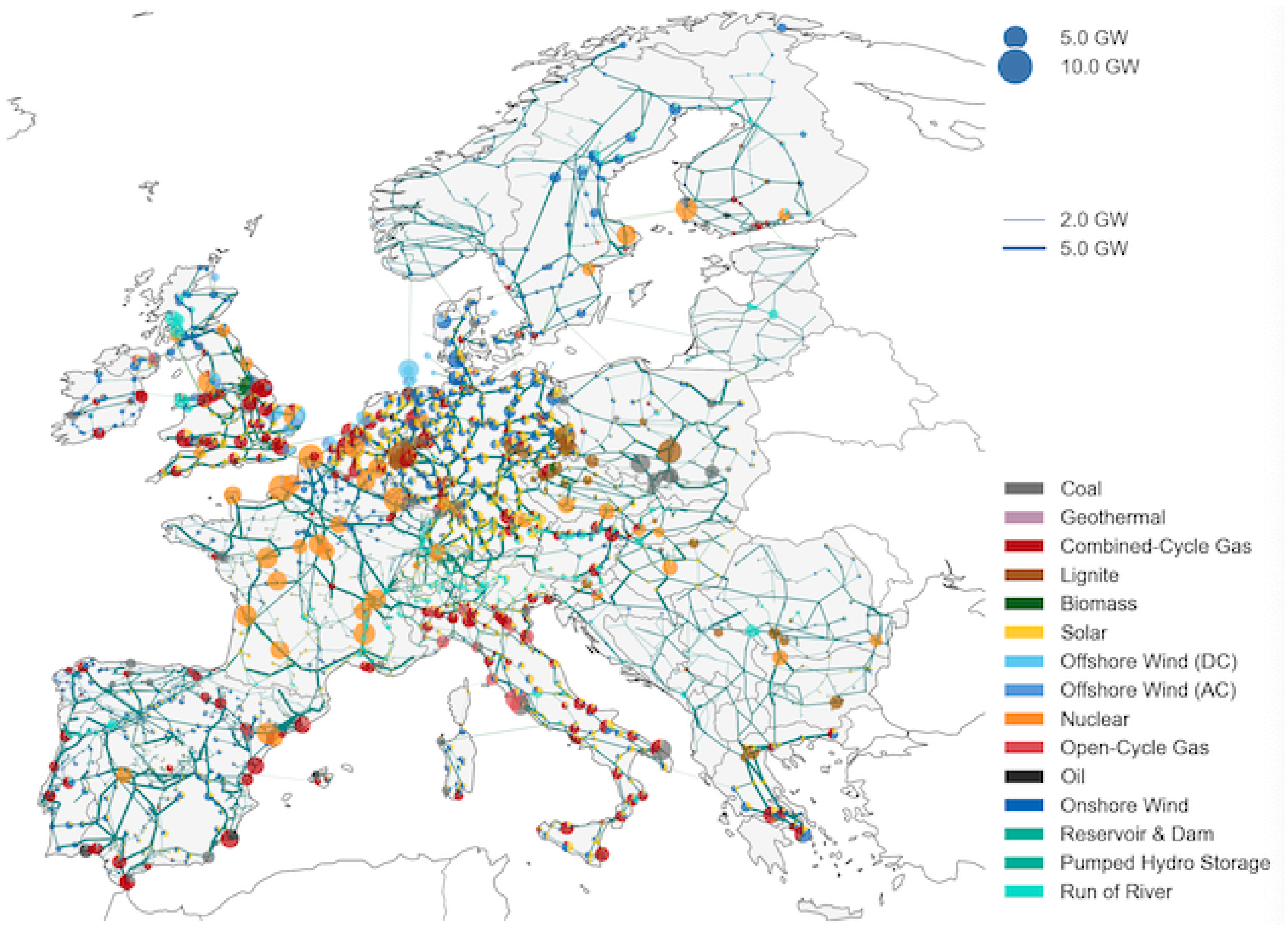


Source: Gartner

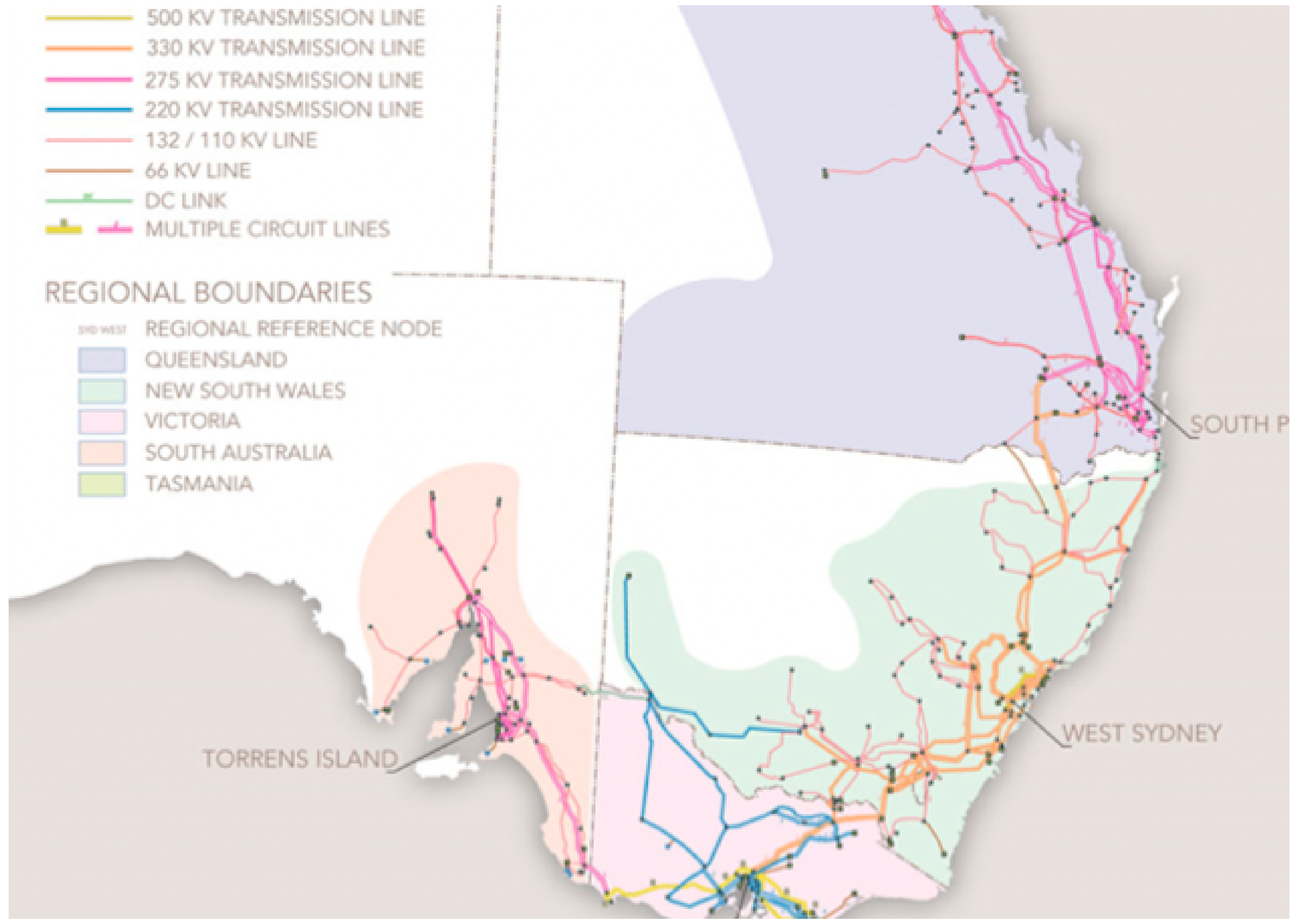
Synthetic power systems datasets



Texas A&M University Grid Datasets (from 37 to 80k+ bus networks)



PyPSA-Eur: synthetic dataset of Europe covering the full ENTSO-E area

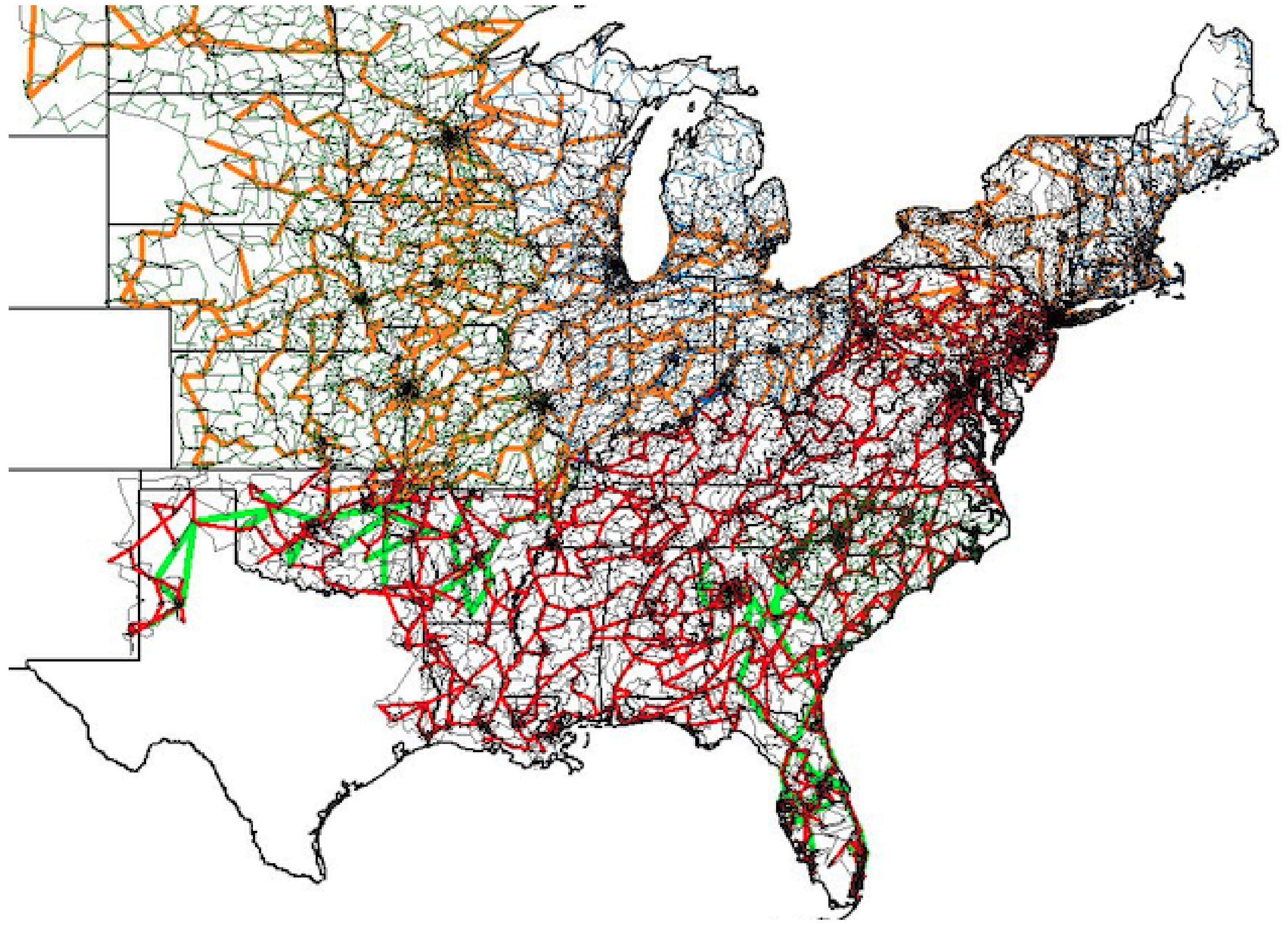


Synthetic Data of the National Electricity Market (Australia)

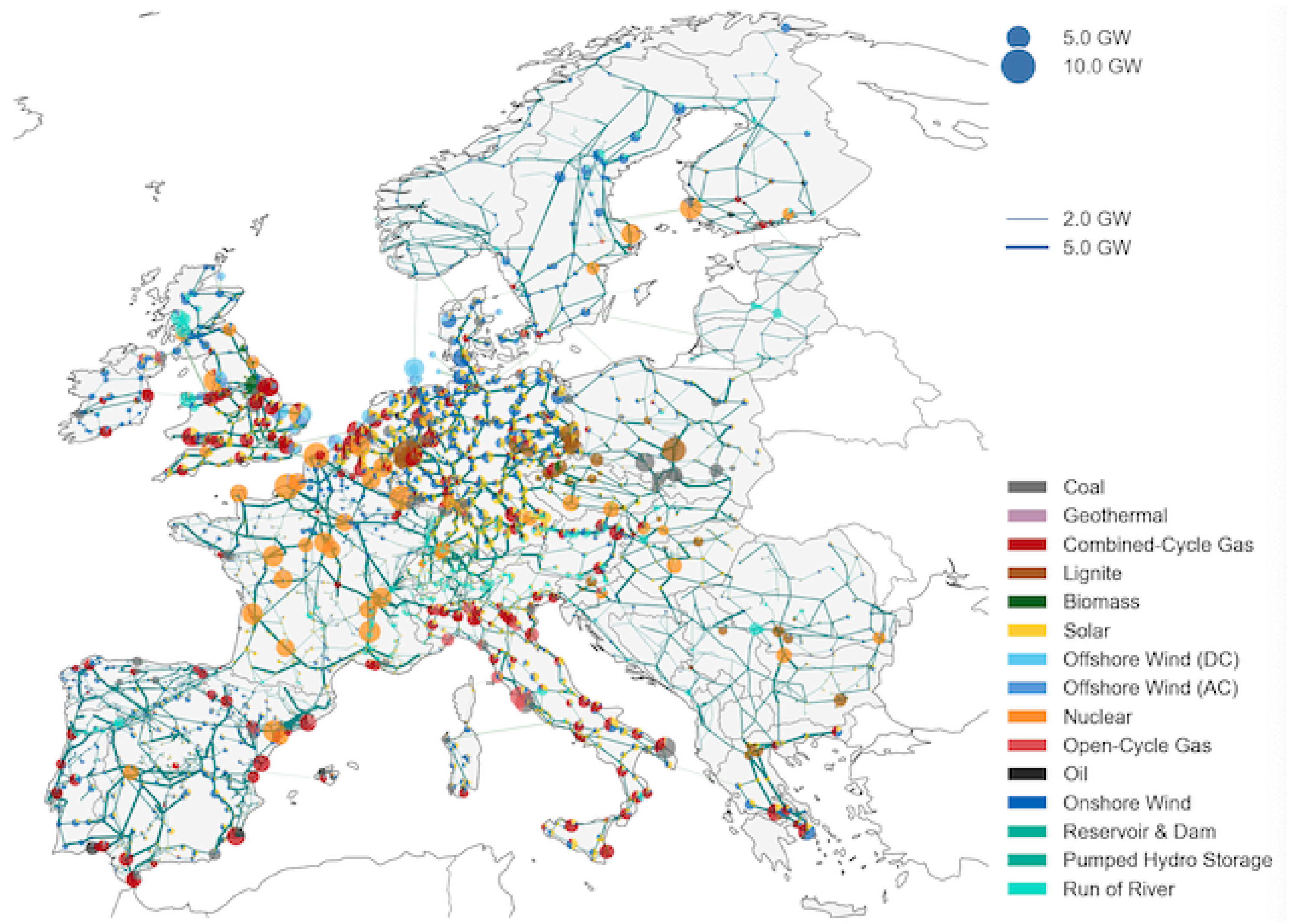
Why these datasets may not satisfy our needs?

- "[...] data bears **no relation** to the actual grid [...] except that generation and load profiles are similar, based on public data"
- "This test case represents a synthetic (**fictitious**) transmission"
- "This case is synthetic and **does not** model the actual grid"

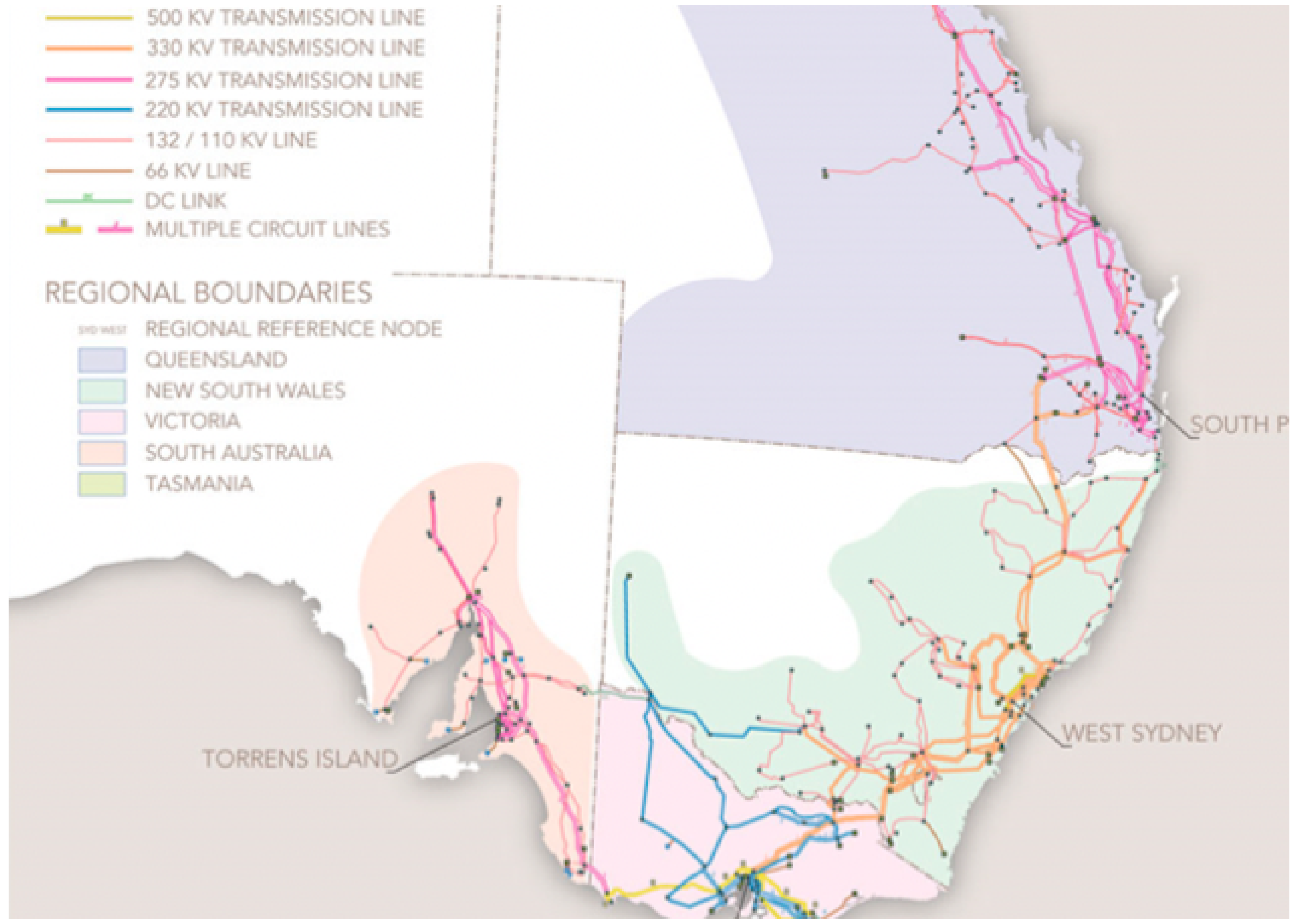
Synthetic power systems datasets



Texas A&M University Grid Datasets (from 37 to 80k+ bus networks)



PyPSA-Eur: synthetic dataset of Europe covering the full ENTSO-E area

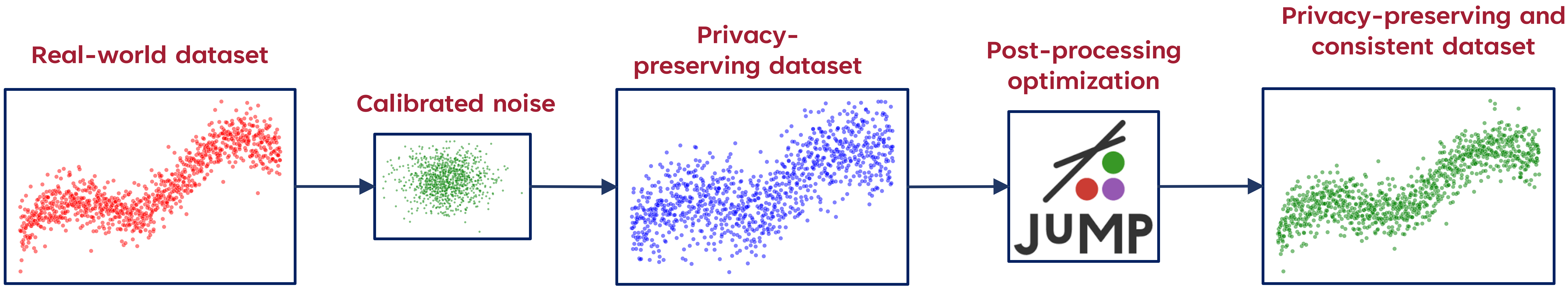


Synthetic Data of the National Electricity Market (Australia)

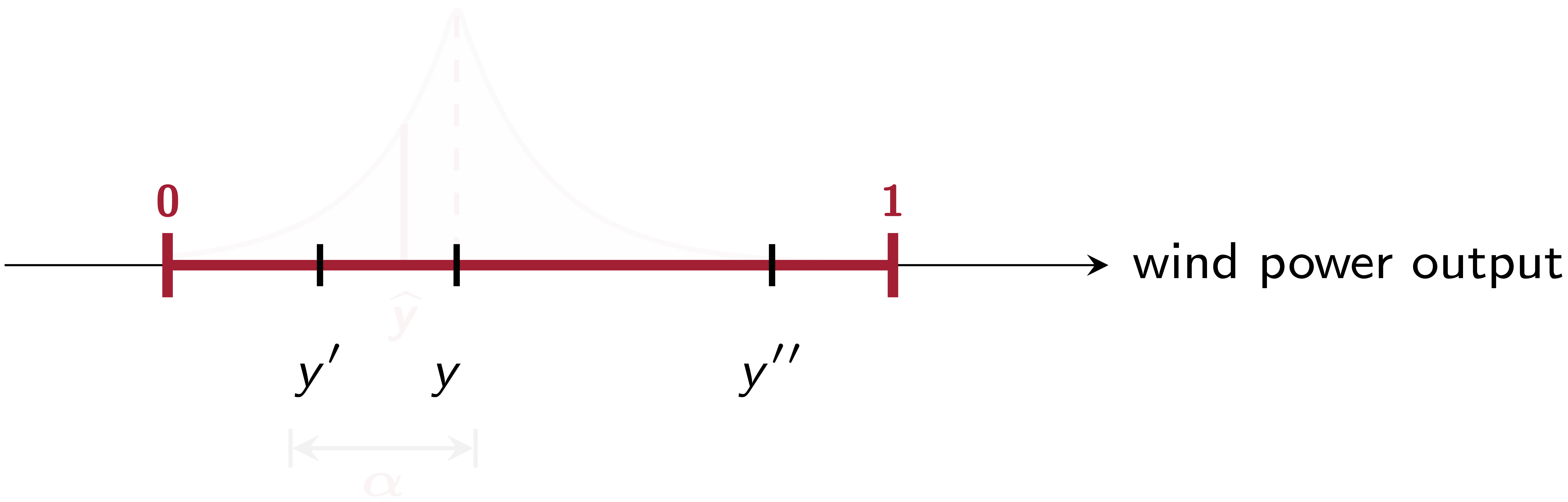
Why these datasets may not satisfy our needs?

- "[...] data bears **no relation** to the actual grid [...] except that generation and load profiles are similar, based on public data"
- "This test case represents a synthetic (**fictitious**) transmission"
- "This case is synthetic and **does not** model the actual grid"

Differential privacy & optimization for synthetic power systems data

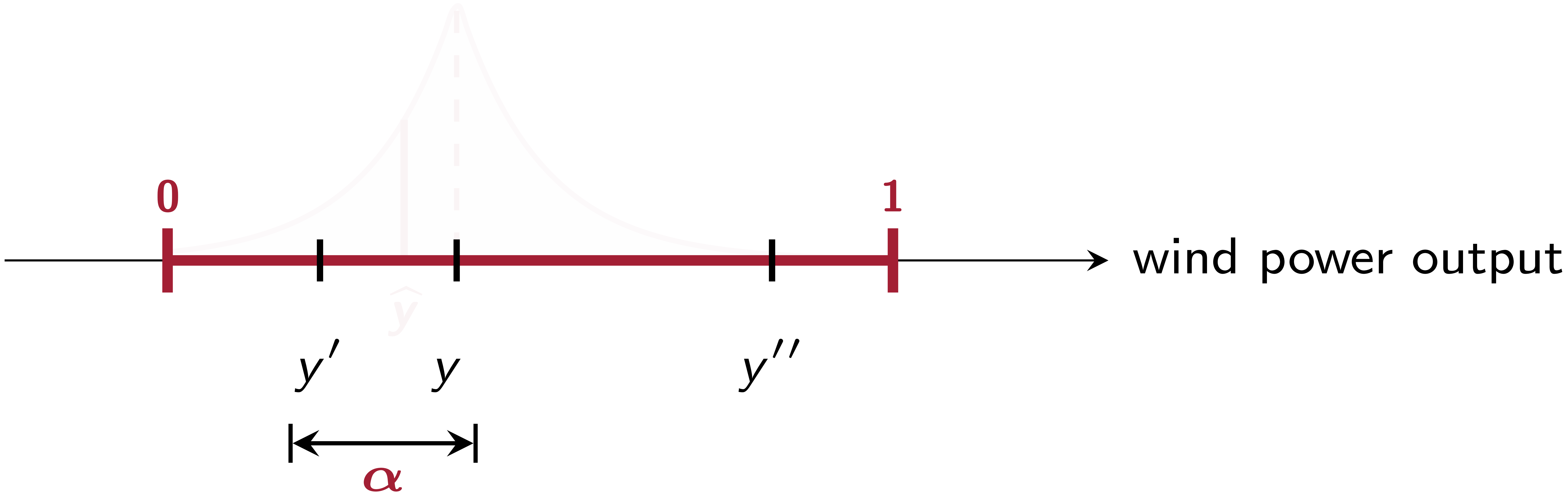


Formalizing differential privacy (DP)



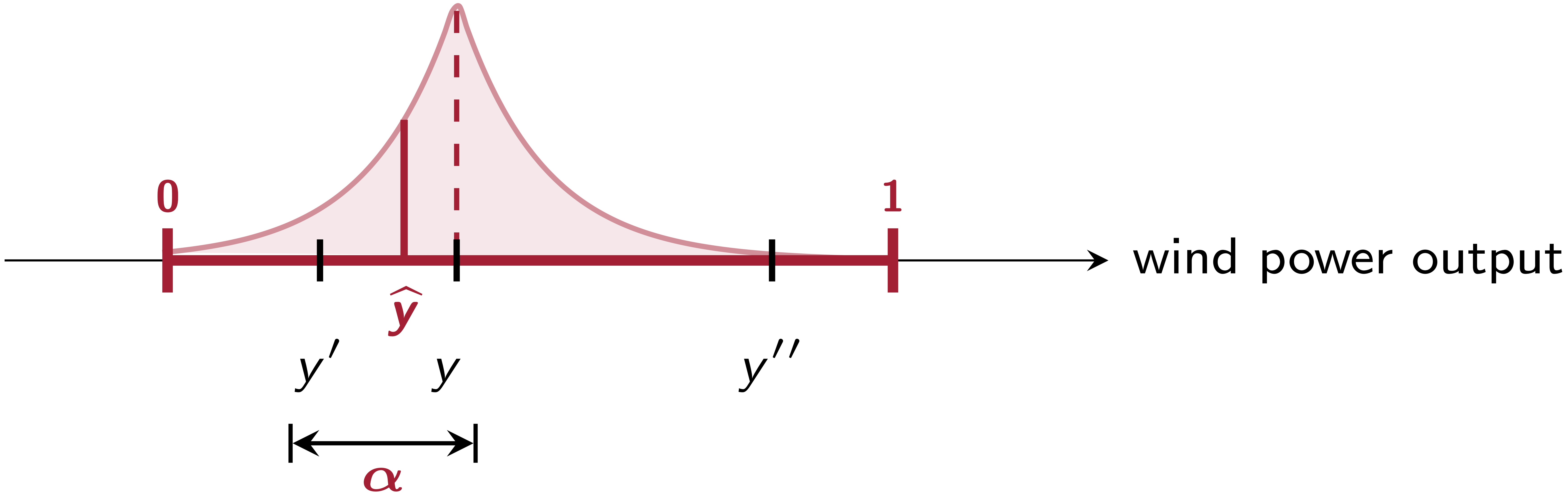
- Wind power records $y, y', y'', \ldots \in [0, 1]$
- For given $\alpha > 0$, records y and y' are $\alpha-$adjacent if $\|y - y'\| \leqslant \alpha$
- The goal is to obfuscate differences in records up to α

Formalizing differential privacy (DP)



- Wind power records $y, y', y'', \ldots \in [0,1]$
- For given $\alpha > 0$, records y and y' are $\alpha-$adjacent if $\|y - y'\| \leqslant \alpha$
- The goal is to obfuscate differences in records up to α

Formalizing differential privacy (DP)

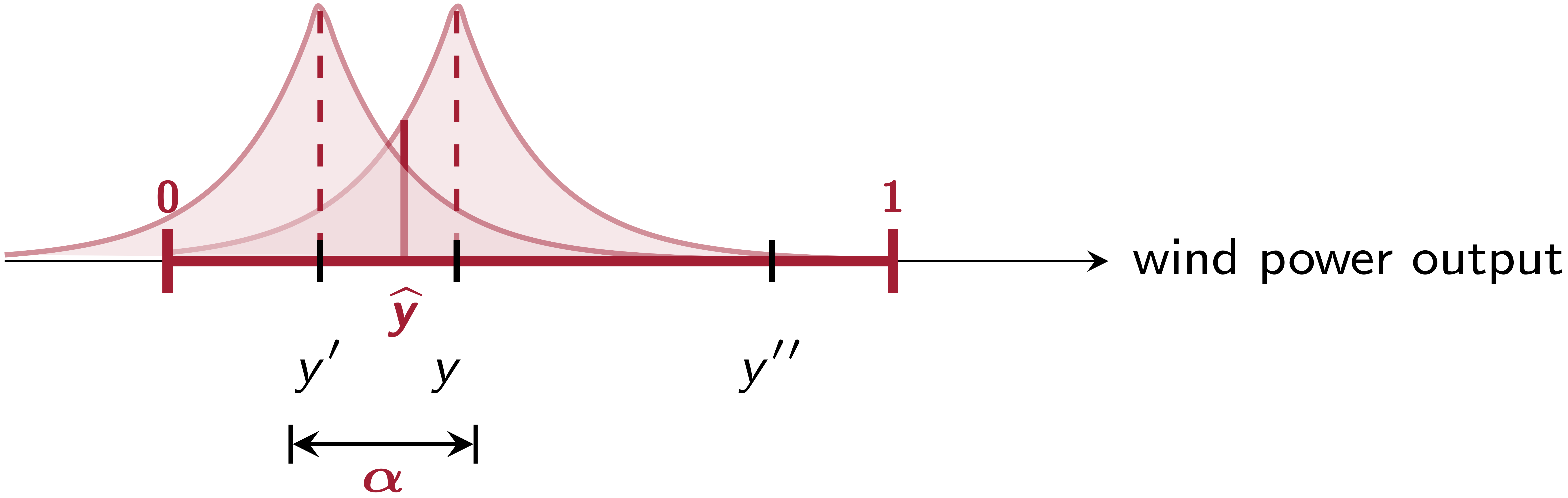


- Wind power records $y, y', y'', \dots \in [0, 1]$
- For given $\alpha > 0$, records y and y' are $\alpha-$adjacent if $\|y - y'\| \leqslant \alpha$
- The goal is to obfuscate differences in records up to α

- Let $\zeta \sim \text{Lap}(\alpha/\varepsilon)$ be a zero-mean random Laplacian noise
- For some small parameter $\varepsilon > 0$, the release is $\varepsilon-$DP if

$$\frac{\Pr[\, y \;+ \zeta \in \widehat{y}\,]}{\Pr[\, y' + \zeta \in \widehat{y}\,]} \leqslant \exp(\varepsilon)$$

Formalizing differential privacy (DP)

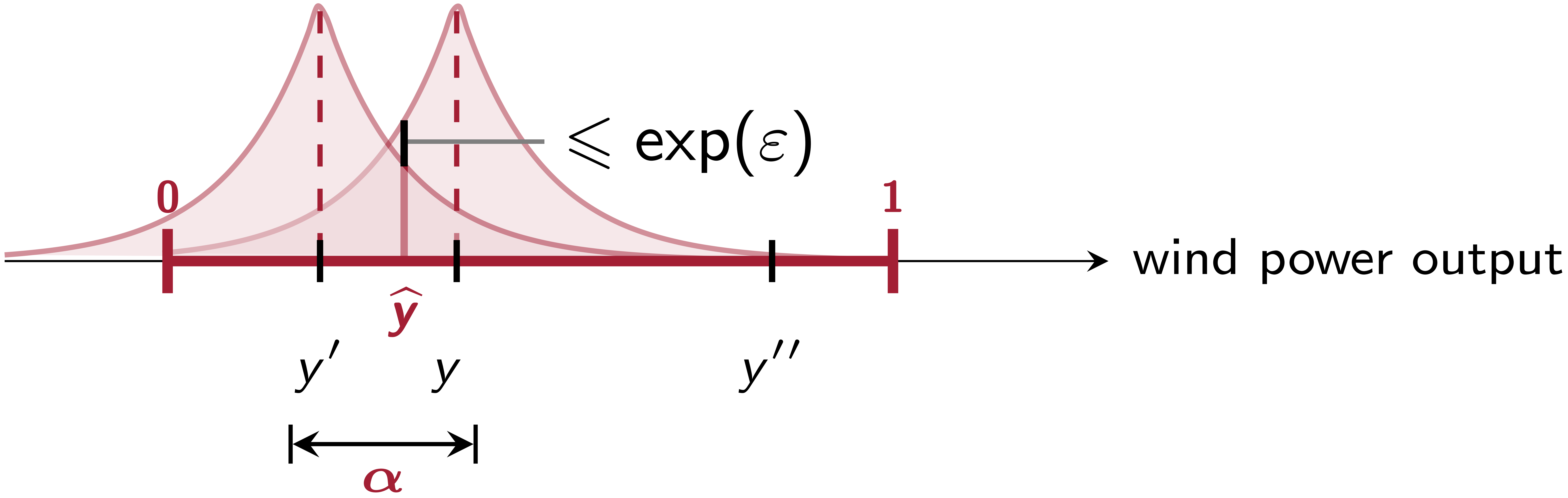


- Wind power records $y, y', y'', \ldots \in [0, 1]$
- For given $\alpha > 0$, records y and y' are $\alpha-$adjacent if $\|y - y'\| \leqslant \alpha$
- The goal is to obfuscate differences in records up to α

- Let $\zeta \sim \text{Lap}(\alpha/\varepsilon)$ be a zero-mean random Laplacian noise
- For some small parameter $\varepsilon > 0$, the release is $\varepsilon-$DP if

$$\frac{\Pr[\, y \;+ \zeta \in \widehat{y}\,]}{\Pr[\, y' + \zeta \in \widehat{y}\,]} \leqslant \exp(\varepsilon)$$

Formalizing differential privacy (DP)

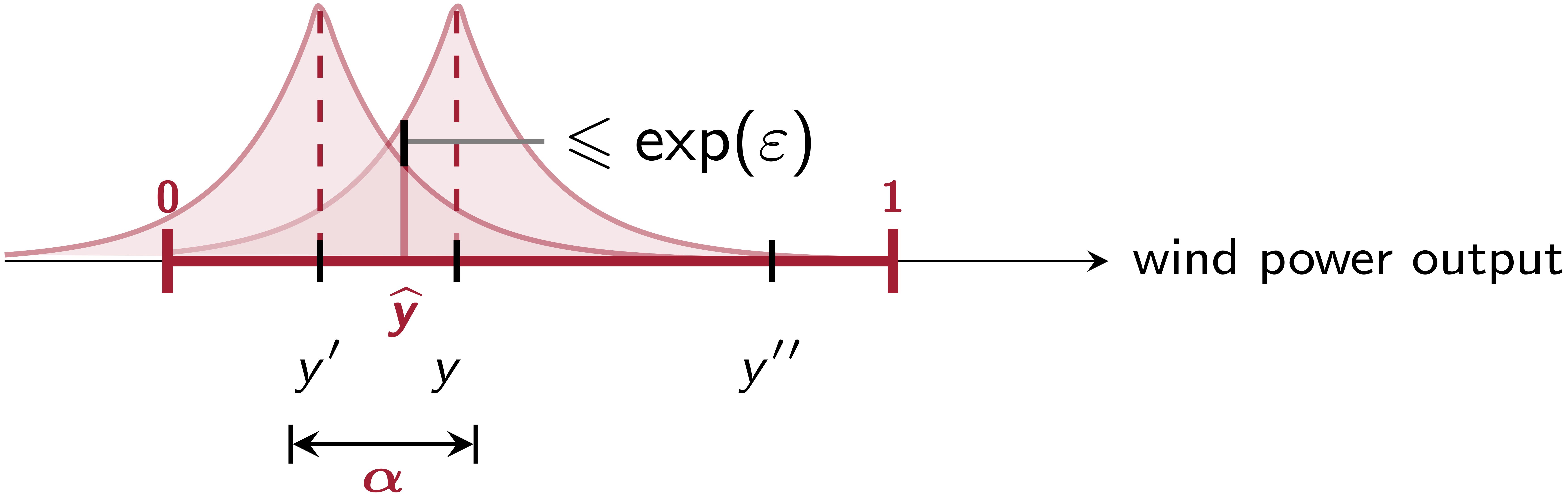


- Wind power records $y, y', y'', \ldots \in [0, 1]$
- For given $\alpha > 0$, records y and y' are $\alpha-$adjacent if $\|y - y'\| \leqslant \alpha$
- The goal is to obfuscate differences in records up to α

- Let $\zeta \sim \text{Lap}(\alpha/\varepsilon)$ be a zero-mean random Laplacian noise
- For some small parameter $\varepsilon > 0$, the release is $\varepsilon-$DP if

$$\frac{\Pr[\, y + \zeta \in \widehat{y}\,]}{\Pr[\, y' + \zeta \in \widehat{y}\,]} \leqslant \exp(\varepsilon)$$

Formalizing differential privacy (DP)



- Wind power records $y, y', y'', \ldots \in [0, 1]$
- For given $\alpha > 0$, records y and y' are $\alpha-$adjacent if $\|y - y'\| \leqslant \alpha$
- The goal is to obfuscate differences in records up to α

Strong theoretical properties

- Rigorous, quantifiable privacy guarantees
- Immunity to post-processing! Arbitrary transformations of noisy data preserve privacy

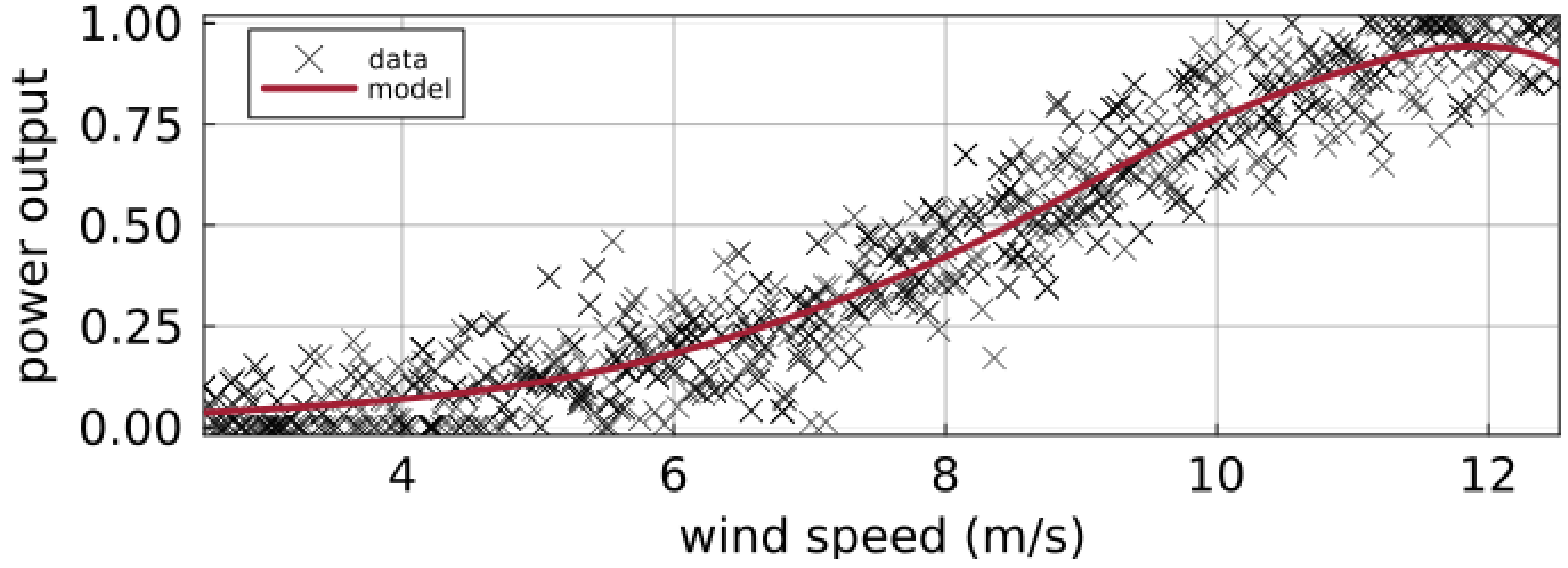
Wind power obfuscation (WPO) algorithm (Part I)

$$\underset{\beta}{\text{minimum}} \;\; \|X\beta - y\| + \lambda\|\beta\|$$

real dataset: $\mathcal{D} = \{(y_1, x_1), \dots, (y_n, x_n)\}$

Power measurement (private data) → y; Wind speed (public data) → x

synthetic dataset: $\tilde{\mathcal{D}} = \{(\tilde{y}_1, x_1), \dots, (\tilde{y}_n, x_n)\}$



- Regression on synthetic data $\tilde{y}$ must match the regression on real data y
- We use regression loss and weights as a measure of accuracy
- Private estimation of regression parameters:

$$\textbf{loss}: \;\; \overline{\ell} = \ell(y) + \text{Lap}\left(\frac{\delta_\ell}{\varepsilon}\right), \qquad \textbf{weights}: \;\; \overline{\beta} = \beta(y) + \text{Lap}\left(\frac{\delta_\beta}{\varepsilon}\right)$$

where $\delta_{(\cdot)}$ is the sensitivity of $(\cdot)$ to data $\alpha-$adjacent datasets

- Lemma (**global sensitivity bounds**):

$$\delta_\ell \leqslant \underset{i=1,\dots,n}{\text{maximum}} \left\|(X(X^\top X + \lambda I)^{-1}X^\top - I)(e_i \circ \alpha)\right\| \qquad \delta_\beta \leqslant \left\|(X^\top X + \lambda I)^{-1}X^\top\right\|_1 \alpha$$

Wind power obfuscation (WPO) algorithm (Part I)

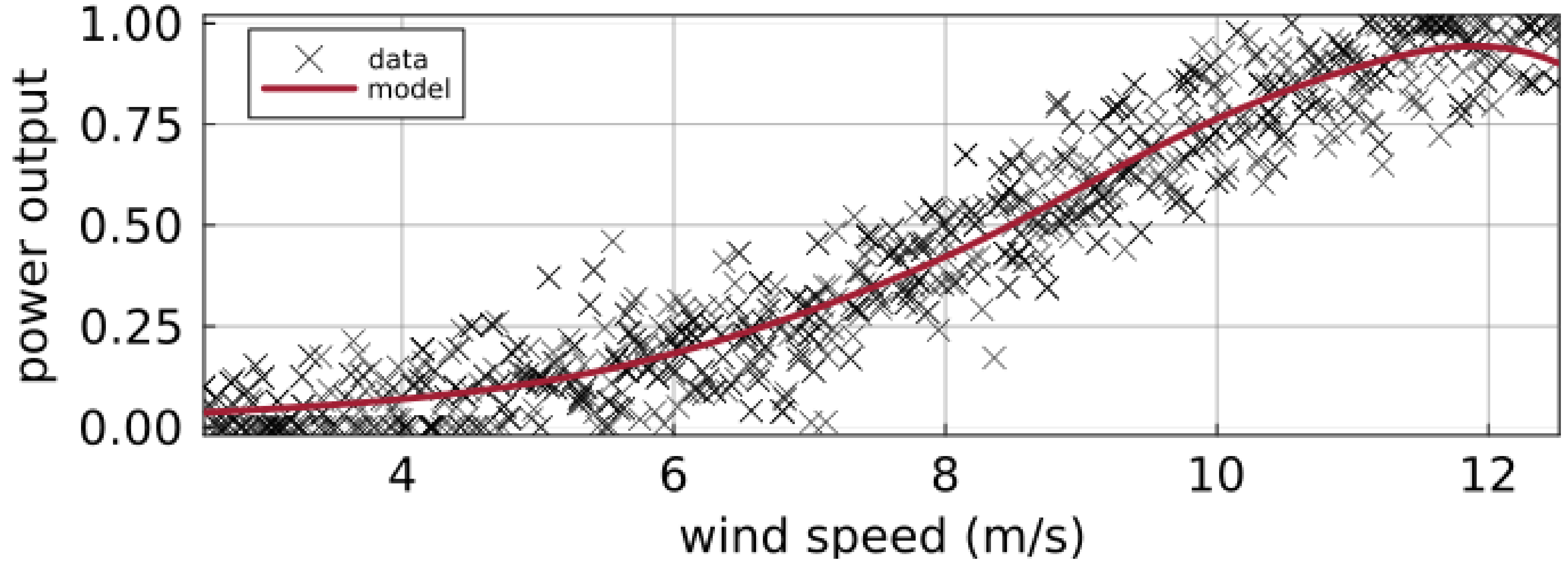
real dataset: $\mathcal{D} = \{(y_1, x_1), \dots, (y_n, x_n)\}$

Power measurement (private data) → y_1

Wind speed (public data) → x_1

synthetic dataset: $\tilde{\mathcal{D}} = \{(\tilde{y}_1, x_1), \dots, (\tilde{y}_n, x_n)\}$

$$\underset{\beta}{\text{minimum}} \; \|X\beta - y\| + \lambda \|\beta\|$$



- Regression on synthetic data $\tilde{y}$ must match the regression on real data y
- We use regression loss and weights as a measure of accuracy
- Private estimation of regression parameters:

$$\textbf{loss}: \quad \overline{\ell} = \ell(y) + \text{Lap}\left(\frac{\delta_\ell}{\varepsilon}\right), \quad \textbf{weights}: \quad \overline{\beta} = \beta(y) + \text{Lap}\left(\frac{\delta_\beta}{\varepsilon}\right)$$

where $\delta_{(\cdot)}$ is the sensitivity of $(\cdot)$ to data $\alpha-$adjacent datasets

- Lemma (global sensitivity bounds):

$$\delta_\ell \leqslant \underset{i=1,\dots,n}{\text{maximum}} \left\| (X(X^\top X + \lambda I)^{-1} X^\top - I)(e_i \circ \alpha) \right\| \qquad \delta_\beta \leqslant \left\| (X^\top X + \lambda I)^{-1} X^\top \right\|_1 \alpha$$

Wind power obfuscation (WPO) algorithm (Part I)

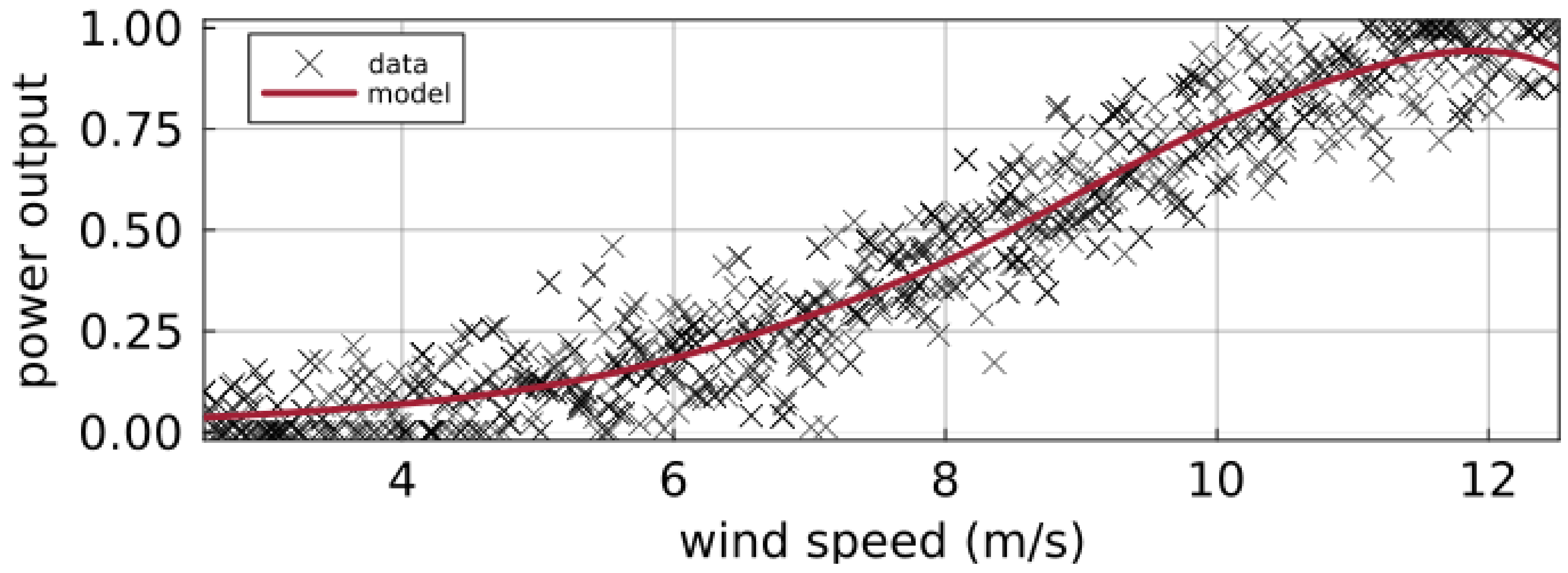
real dataset: $\mathcal{D} = \{(y_1, x_1), \dots, (y_n, x_n)\}$

Power measurement (private data)

Wind speed (public data)

synthetic dataset: $\tilde{\mathcal{D}} = \{(\tilde{y}_1, x_1), \dots, (\tilde{y}_n, x_n)\}$

$$\underset{\beta}{\text{minimum}} \;\; \|X\beta - y\| + \lambda\|\beta\|$$



- Regression on synthetic data $\tilde{y}$ must match the regression on real data y
- We use regression loss and weights as a measure of accuracy
- Private estimation of regression parameters:

$$\textbf{loss}: \quad \bar{\ell} = \ell(y) + \text{Lap}\left(\frac{\delta_\ell}{\varepsilon}\right), \quad \textbf{weights}: \quad \bar{\beta} = \beta(y) + \text{Lap}\left(\frac{\delta_\beta}{\varepsilon}\right)$$

where $\delta_{(\cdot)}$ is the sensitivity of $(\cdot)$ to data $\alpha-$adjacent datasets

- Lemma (**global sensitivity bounds**):

$$\delta_\ell \leqslant \underset{i=1,\dots,n}{\text{maximum}} \left\|(X(X^\top X + \lambda I)^{-1}X^\top - I)(e_i \circ \alpha)\right\| \qquad \delta_\beta \leqslant \left\|(X^\top X + \lambda I)^{-1}X^\top\right\|_1 \alpha$$

Wind power obfuscation (WPO) algorithm (Part II)

Step 1 Synthetic wind power measurements:

$$\tilde{y}^0 = y + \text{Lap}\left(\alpha/\varepsilon_1\right)$$

Step 2 Private regression parameters estimation:

$$\overline{\ell} = \ell(y) + \text{Lap}\left(\delta_\ell/\varepsilon_2\right) \quad \overline{\beta} = \beta(y) + \text{Lap}\left(\delta_\beta/\varepsilon_2\right)$$

Step 3 Synthetic dataset post-processing:

$$\tilde{y} \in \underset{\tilde{y}}{\text{argmin}} \quad \underbrace{\left\|\overline{\ell} - \ell(\tilde{y})\right\|}_{\text{loss accuracy}} + \gamma_\beta \underbrace{\left\|\overline{\beta} - \beta(\tilde{y})\right\|}_{\text{weight accuracy}} + \gamma_y \underbrace{\left\|\tilde{y}^0 - \tilde{y}\right\|}_{\text{regularization}}$$

$$\text{s.t.} \quad \mathbb{0} \leqslant \tilde{y} \leqslant \mathbb{1}$$

$$\beta(\tilde{y}), \ell(\tilde{y}) \in \underset{\beta}{\text{argmin}} \ \underbrace{\left\|X\beta - \tilde{y}\right\|}_{\ell} + \lambda \left\|\beta\right\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/4$ renders WPO $\varepsilon-$DP for $\alpha-$adjacent wind power datasets.

Wind power obfuscation (WPO) algorithm (Part II)

Step 1 Synthetic wind power measurements:

$$\tilde{y}^0 = y + \text{Lap}\left(\alpha/\varepsilon_1\right)$$

Step 2 Private regression parameters estimation:

$$\overline{\ell} = \ell(y) + \text{Lap}\left(\delta_\ell/\varepsilon_2\right) \quad \overline{\beta} = \beta(y) + \text{Lap}\left(\delta_\beta/\varepsilon_2\right)$$

Step 3 Synthetic dataset post-processing:

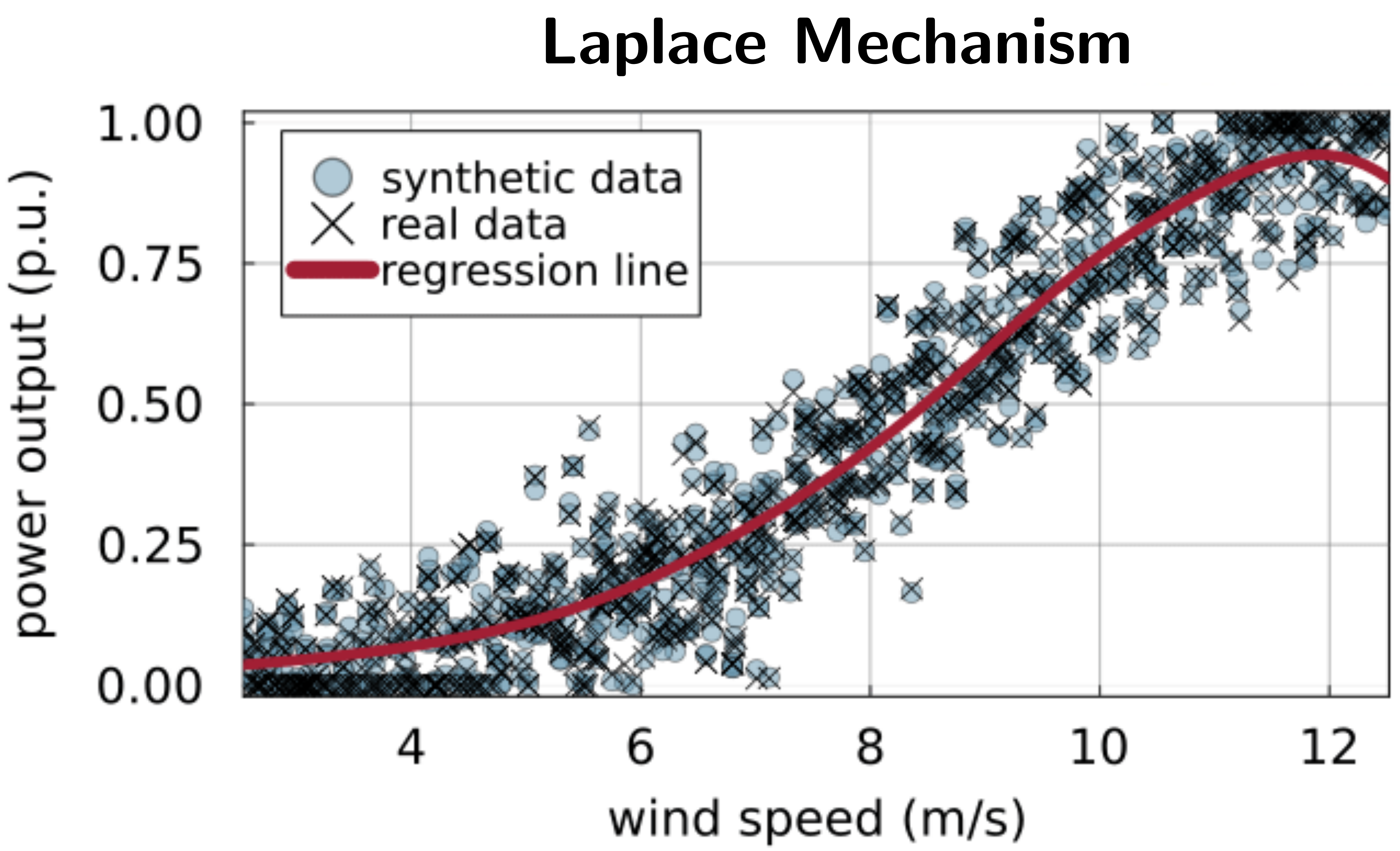
$$\tilde{y} \in \underset{\tilde{y}}{\text{argmin}} \quad \underbrace{\left\|\overline{\ell} - \ell(\tilde{y})\right\|}_{\text{loss accuracy}} + \gamma_\beta \underbrace{\left\|\overline{\beta} - \beta(\tilde{y})\right\|}_{\text{weight accuracy}} + \gamma_y \underbrace{\left\|\tilde{y}^0 - \tilde{y}\right\|}_{\text{regularization}}$$

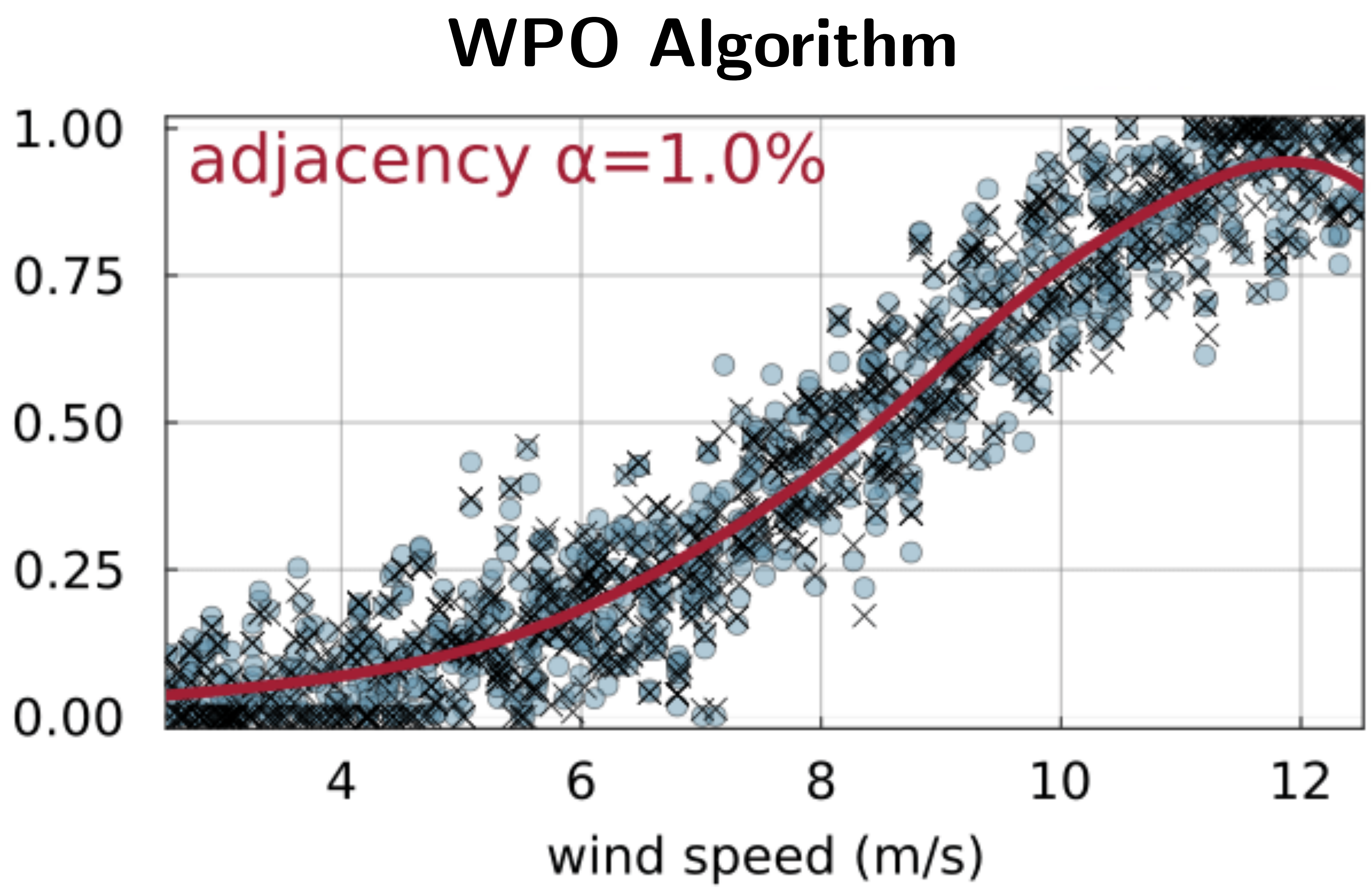
$$\text{s.t.} \quad \mathbb{0} \leqslant \tilde{y} \leqslant \mathbb{1}$$

$$\beta(\tilde{y}), \ell(\tilde{y}) \in \underset{\beta}{\text{argmin}} \; \underbrace{\left\|X\beta - \tilde{y}\right\|}_{\ell} + \lambda \left\|\beta\right\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/4$ renders WPO $\varepsilon-$DP for $\alpha-$adjacent wind power datasets.

WPO algorithm: Application to Alstom Eco 80 wind turbine





Accuracy of the WPO Algorithm remains high with a growing privacy requirement α

Differentially private release of network parameters

Optimal Power Flow (OPF) problem

$$\begin{aligned} \mathcal{C}(\overline{f}) = \min_{p\in\mathcal{P}} \quad & c^\top p && \textit{dispatch costs} \\ \text{s.t.} \quad & \mathbb{1}^\top(p-d)=0 && \textit{power balance} \\ & |F(p-d)| \leqslant \overline{f} && \textit{power flow limit} \end{aligned}$$

How to release vector of transmission capacities $\overline{f}$ privately?

Laplace mechanism:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon)$$

Almost never feasible

Laplace + Bilevel optimization:

$$\begin{aligned} \min_{\hat{\varphi}} \quad & \left\|\overline{\varphi}^0 - \hat{\varphi}\right\| \\ \text{s.t.} \quad & |\mathcal{C}(\hat{\varphi}) - \mathcal{C}^\star| \leqslant \beta\mathcal{C}^\star \end{aligned}$$

Embedded OPF

Feasible and cost-consistent with respect to a **single** OPF model

Laplace & Exponential mechanisms + Bilevel optimization:

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a **population** of OPF models

Differentially private release of network parameters

Optimal Power Flow (OPF) problem

$$\begin{aligned} \mathcal{C}(\overline{f}) = \min_{p \in \mathcal{P}} \quad & c^\top p && \textit{dispatch costs} \\ \text{s.t.} \quad & \mathbb{1}^\top (p - d) = 0 && \textit{power balance} \\ & |F(p-d)| \leqslant \overline{f} && \textit{power flow limit} \end{aligned}$$

How to release vector of transmission capacities $\overline{f}$ privately?

Laplace mechanism:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon)$$

Almost never feasible

Laplace + Bilevel optimization:

$$\begin{aligned} \min_{\hat{\varphi}} \quad & \left\| \overline{\varphi}^0 - \hat{\varphi} \right\| \\ \text{s.t.} \quad & |\mathcal{C}(\hat{\varphi}) - \mathcal{C}^\star| \leqslant \beta \mathcal{C}^\star \end{aligned}$$

Embedded OPF

Feasible and cost-consistent with respect to a **single** OPF model

Laplace & Exponential mechanisms + Bilevel optimization:

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a **population** of OPF models

Differentially private release of network parameters

Optimal Power Flow (OPF) problem

$$\begin{aligned} \mathcal{C}(\overline{f}) = \min_{p\in\mathcal{P}} \quad & c^\top p && \textit{dispatch costs} \\ \text{s.t.} \quad & \mathbb{1}^\top(p-d)=0 && \textit{power balance} \\ & |F(p-d)| \leqslant \overline{f} && \textit{power flow limit} \end{aligned}$$

How to release vector of transmission capacities $\overline{f}$ privately?

Laplace mechanism:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon)$$

Almost never feasible

Laplace + Bilevel optimization:

$$\begin{aligned} \min_{\hat{\varphi}} \quad & \left\|\overline{\varphi}^0 - \hat{\varphi}\right\| \\ \text{s.t.} \quad & |\mathcal{C}(\hat{\varphi}) - \mathcal{C}^\star| \leqslant \beta\mathcal{C}^\star \end{aligned}$$

Embedded OPF

Feasible and cost-consistent with respect to a **single** OPF model

Laplace & Exponential mechanisms + Bilevel optimization:

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a **population** of OPF models

Differentially private release of network parameters

Optimal Power Flow (OPF) problem

$$\begin{aligned} \mathcal{C}(\overline{f}) = \min_{p \in \mathcal{P}} \quad & c^\top p && \textit{dispatch costs} \\ \text{s.t.} \quad & \mathbb{1}^\top (p - d) = 0 && \textit{power balance} \\ & |F(p-d)| \leqslant \overline{f} && \textit{power flow limit} \end{aligned}$$

How to release vector of transmission capacities $\overline{f}$ privately?

Laplace mechanism:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon)$$

Almost never feasible

Laplace + Bilevel optimization:

$$\begin{aligned} \min_{\hat{\varphi}} \quad & \left\| \overline{\varphi}^0 - \hat{\varphi} \right\| \\ \text{s.t.} \quad & |\mathcal{C}(\hat{\varphi}) - \mathcal{C}^\star| \leqslant \beta \mathcal{C}^\star \end{aligned}$$

Embedded OPF

Feasible and cost-consistent with respect to a **single** OPF model

Laplace & Exponential mechanisms + Bilevel optimization:

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a **population** of OPF models

Differentially private transmission capacity obfuscation (TCO) Algorithm

Step 1 Initialize synthetic data using LM:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)$$

Step 2 Find the worst-case OPF model using EM:

$$\Delta\mathcal{C}_i = \left\|\mathcal{C}_i(\overline{f}) - \mathcal{C}_i^R(\overline{\varphi}^0)\right\|_1 + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right), \forall i = 1, \dots, m$$

return index k of the worst-case model

Step 3 Compute the worst-case cost using LM:

$$\overline{\mathcal{C}} = \mathcal{C}_k(\overline{f}) + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right)$$

Step 4 Post-processing bilevel optimization:

$$\overline{\varphi}^t \in \underset{\overline{\varphi}}{\text{argmin}} \; \left\|\overline{\mathcal{C}} - \mathcal{C}_k(\overline{\varphi})\right\| + \left\|\overline{\varphi} - \overline{\varphi}^0\right\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/(4T)$ achieve $\varepsilon-$differential privacy

Differentially private transmission capacity obfuscation (TCO) Algorithm

Step 1 Initialize synthetic data using LM:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)$$

Step 2 Find the worst-case OPF model using EM:

$$\Delta\mathcal{C}_i = \left\| \mathcal{C}_i(\overline{f}) - \mathcal{C}_i^R(\overline{\varphi}^0) \right\|_1 + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right), \forall i = 1, \dots, m$$

return index k of the worst-case model

Step 3 Compute the worst-case cost using LM:

$$\overline{\mathcal{C}} = \mathcal{C}_k(\overline{f}) + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right)$$

Step 4 Post-processing bilevel optimization:

$$\overline{\varphi}^t \in \underset{\overline{\varphi}}{\text{argmin}} \ \left\|\overline{\mathcal{C}} - \mathcal{C}_k(\overline{\varphi})\right\| + \left\|\overline{\varphi} - \overline{\varphi}^0\right\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/(4T)$ achieve $\varepsilon-$differential privacy

Differentially private transmission capacity obfuscation (TCO) Algorithm

Step 1 Initialize synthetic data using LM:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)$$

Step 2 Find the worst-case OPF model using EM:

$$\Delta\mathcal{C}_i = \left\|\mathcal{C}_i(\overline{f}) - \mathcal{C}_i^R(\overline{\varphi}^0)\right\|_1 + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right), \forall i = 1, \dots, m$$

return index k of the worst-case model

Step 3 Compute the worst-case cost using LM:

$$\overline{\mathcal{C}} = \mathcal{C}_k(\overline{f}) + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right)$$

Step 4 Post-processing bilevel optimization:

$$\overline{\varphi}^t \in \underset{\overline{\varphi}}{\text{argmin}} \ \left\|\overline{\mathcal{C}} - \mathcal{C}_k(\overline{\varphi})\right\| + \left\|\overline{\varphi} - \overline{\varphi}^0\right\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/(4T)$ achieve $\varepsilon-$differential privacy

Differentially private transmission capacity obfuscation (TCO) Algorithm

Step 1 Initialize synthetic data using LM:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)$$

Step 2 Find the worst-case OPF model using EM:

$$\Delta\mathcal{C}_i = \left\|\mathcal{C}_i(\overline{f}) - \mathcal{C}_i^R(\overline{\varphi}^0)\right\|_1 + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right), \forall i = 1, \ldots, m$$

return index k of the worst-case model

Step 3 Compute the worst-case cost using LM:

$$\overline{\mathcal{C}} = \mathcal{C}_k(\overline{f}) + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right)$$

Step 4 Post-processing bilevel optimization:

$$\overline{\varphi}^t \in \underset{\overline{\varphi}}{\text{argmin}} \; \left\|\overline{\mathcal{C}} - \mathcal{C}_k(\overline{\varphi})\right\| + \left\|\overline{\varphi} - \overline{\varphi}^0\right\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/(4T)$ achieve $\varepsilon-$differential privacy

Differentially private transmission capacity obfuscation (TCO) Algorithm

repeat T times

Step 1 Initialize synthetic data using LM:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)$$

Step 2 Find the worst-case OPF model using EM:

$$\Delta\mathcal{C}_i = \left\|\mathcal{C}_i(\overline{f}) - \mathcal{C}_i^R(\overline{\varphi}^{t-1})\right\|_1 + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right), \forall i = 1, \dots, m$$

return index k^t of the worst-case model

Step 3 Compute the worst-case cost using LM:

$$\overline{\mathcal{C}}_t = \mathcal{C}_{k^t}(\overline{f}) + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right)$$

Step 4 Post-processing bilevel optimization:

$$\overline{\varphi}^t \in \underset{\overline{\varphi}}{\text{argmin}} \sum_{\tau=1}^{t} \|\overline{\mathcal{C}}_\tau - \mathcal{C}_{k^\tau}(\overline{\varphi})\| + \|\overline{\varphi} - \overline{\varphi}^{t-1}\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/(4T)$ achieve $\varepsilon-$differential privacy

Differentially private transmission capacity obfuscation (TCO) Algorithm

repeat $\boldsymbol{T}$ times

Step 1 Initialize synthetic data using LM:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)$$

Step 2 Find the worst-case OPF model using EM:

$$\Delta\mathcal{C}_i = \left\|\mathcal{C}_i(\overline{f}) - \mathcal{C}_i^R(\overline{\varphi}^{\boldsymbol{t-1}})\right\|_1 + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right), \forall i = 1, \dots, m$$

return index $k^{\boldsymbol{t}}$ of the worst-case model

Step 3 Compute the worst-case cost using LM:

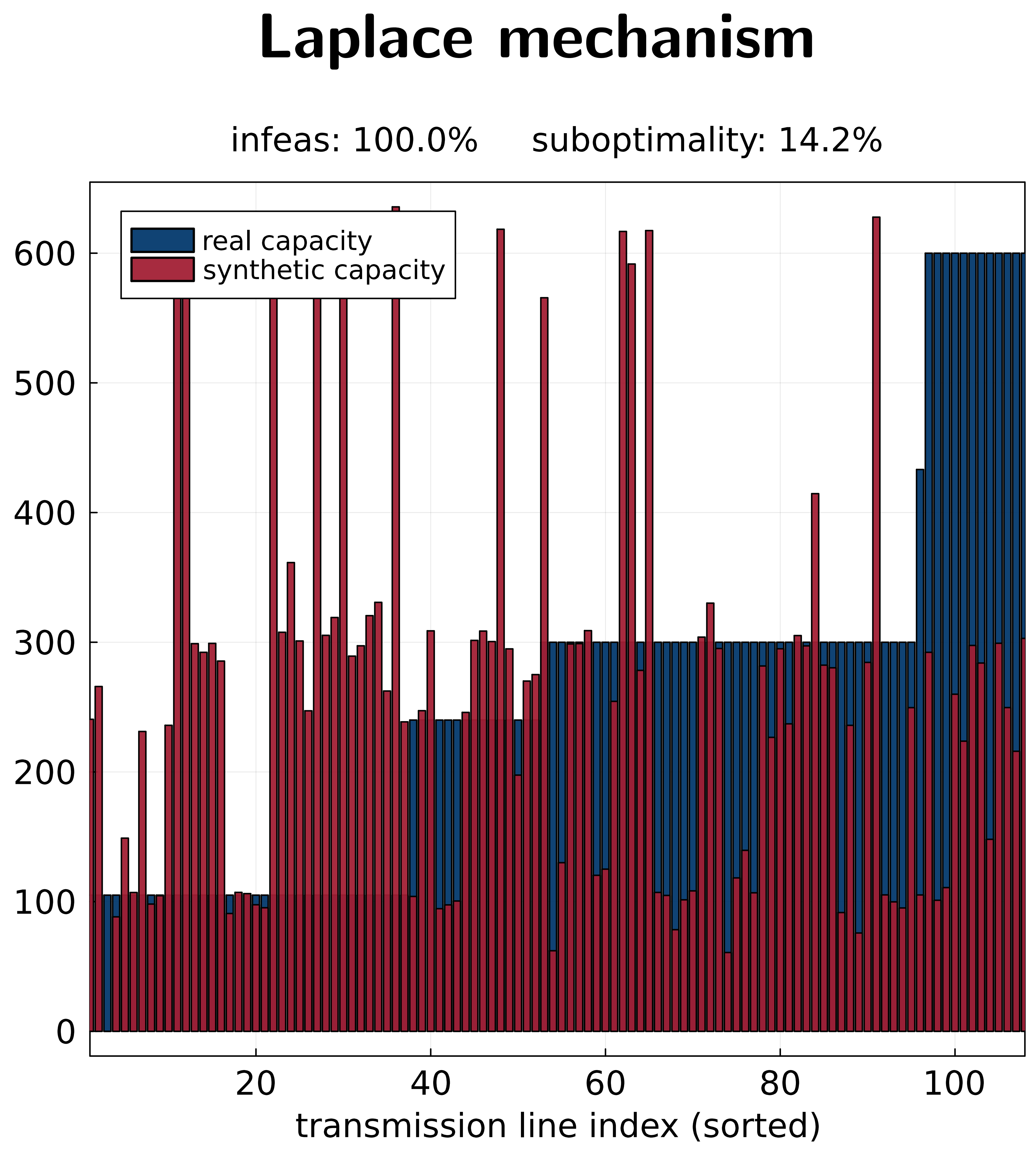
$$\overline{\mathcal{C}}_{\boldsymbol{t}} = \mathcal{C}_{k^{\boldsymbol{t}}}(\overline{f}) + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right)$$

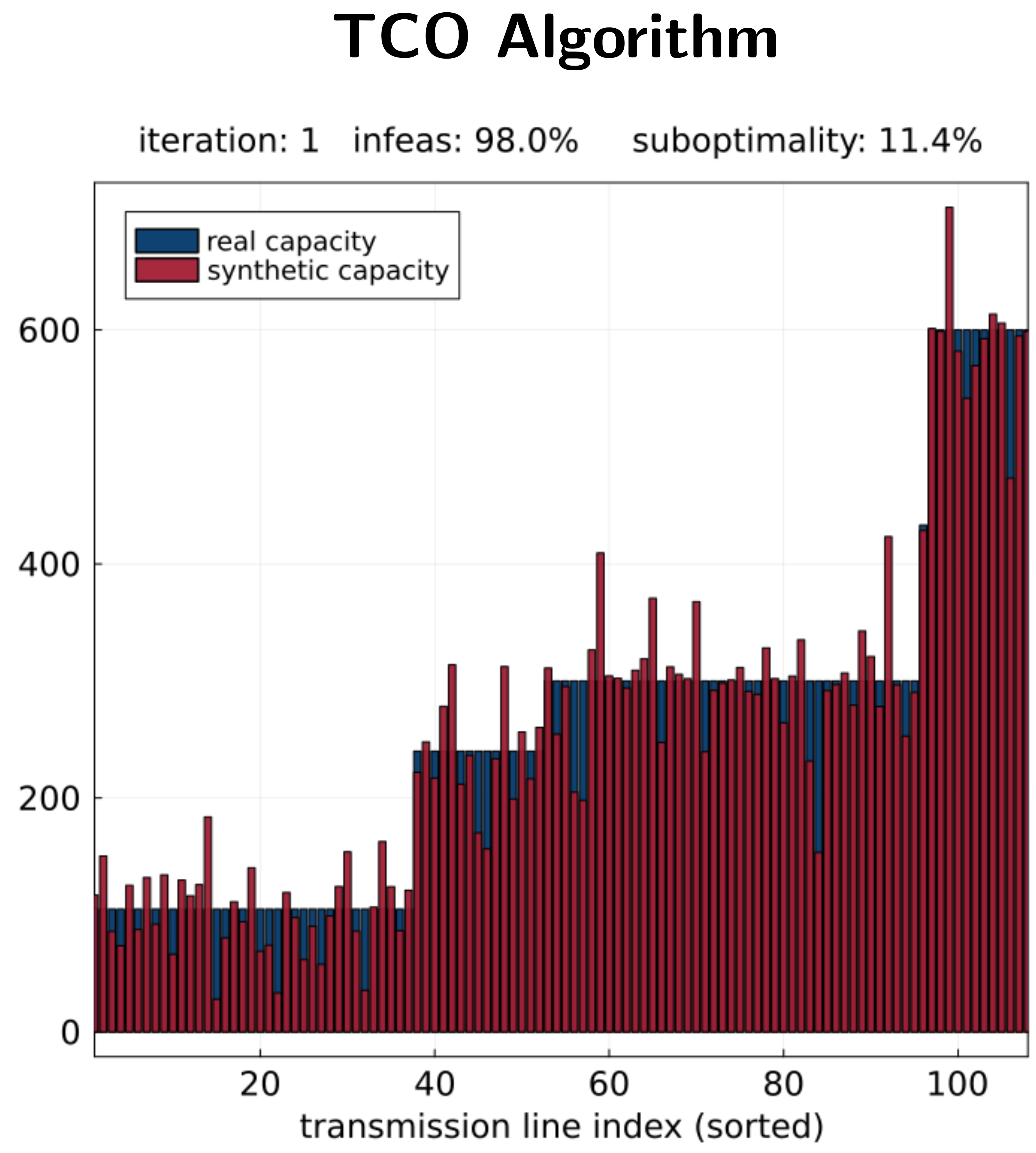
Step 4 Post-processing bilevel optimization:

$$\overline{\varphi}^t \in \underset{\overline{\varphi}}{\text{argmin}} \sum_{\boldsymbol{\tau=1}}^{\boldsymbol{t}} \left\|\overline{\mathcal{C}}_{\boldsymbol{\tau}} - \mathcal{C}_{k^{\boldsymbol{\tau}}}(\overline{\varphi})\right\| + \left\|\overline{\varphi} - \overline{\varphi}^{\boldsymbol{t-1}}\right\|$$

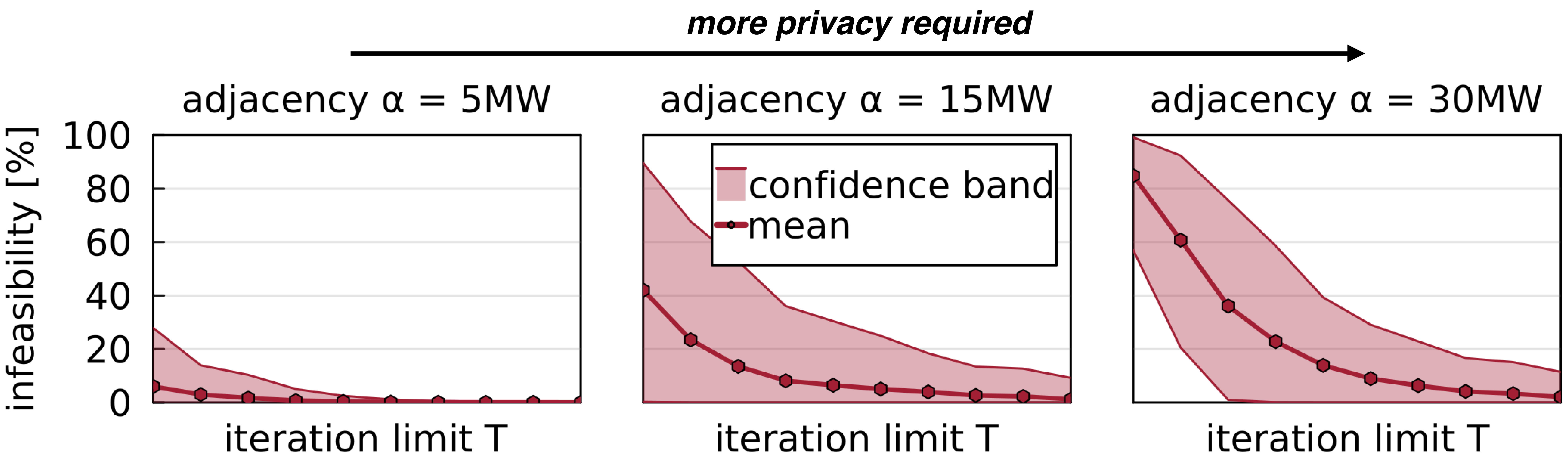
Theorem: $\boldsymbol{\varepsilon_1 = \varepsilon/2}$ and $\boldsymbol{\varepsilon_2 = \varepsilon/(4T)}$ achieve $\varepsilon-$differential privacy

IEEE 73-RTS benchmark: Laplace versus TCO Algorithm

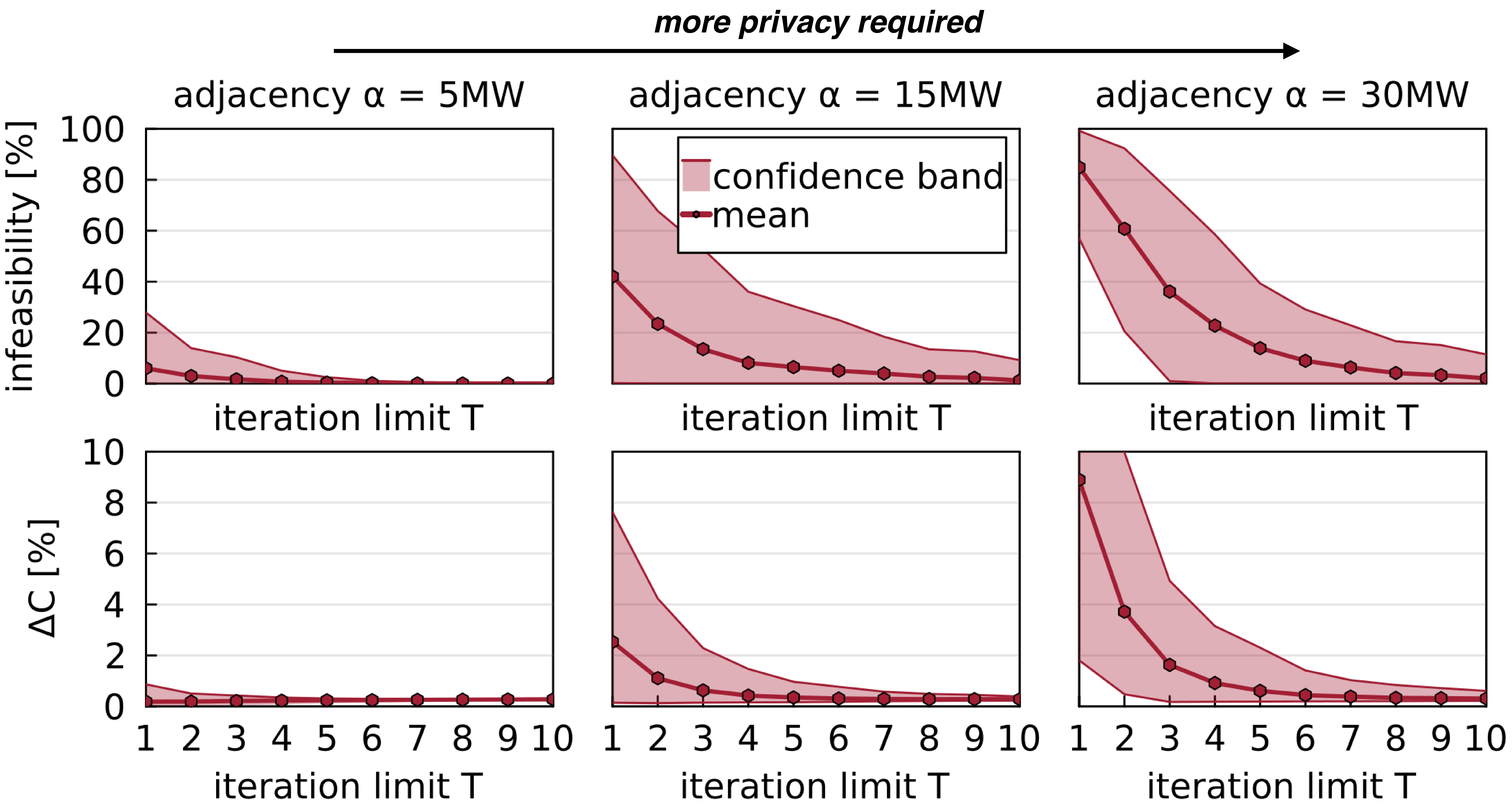




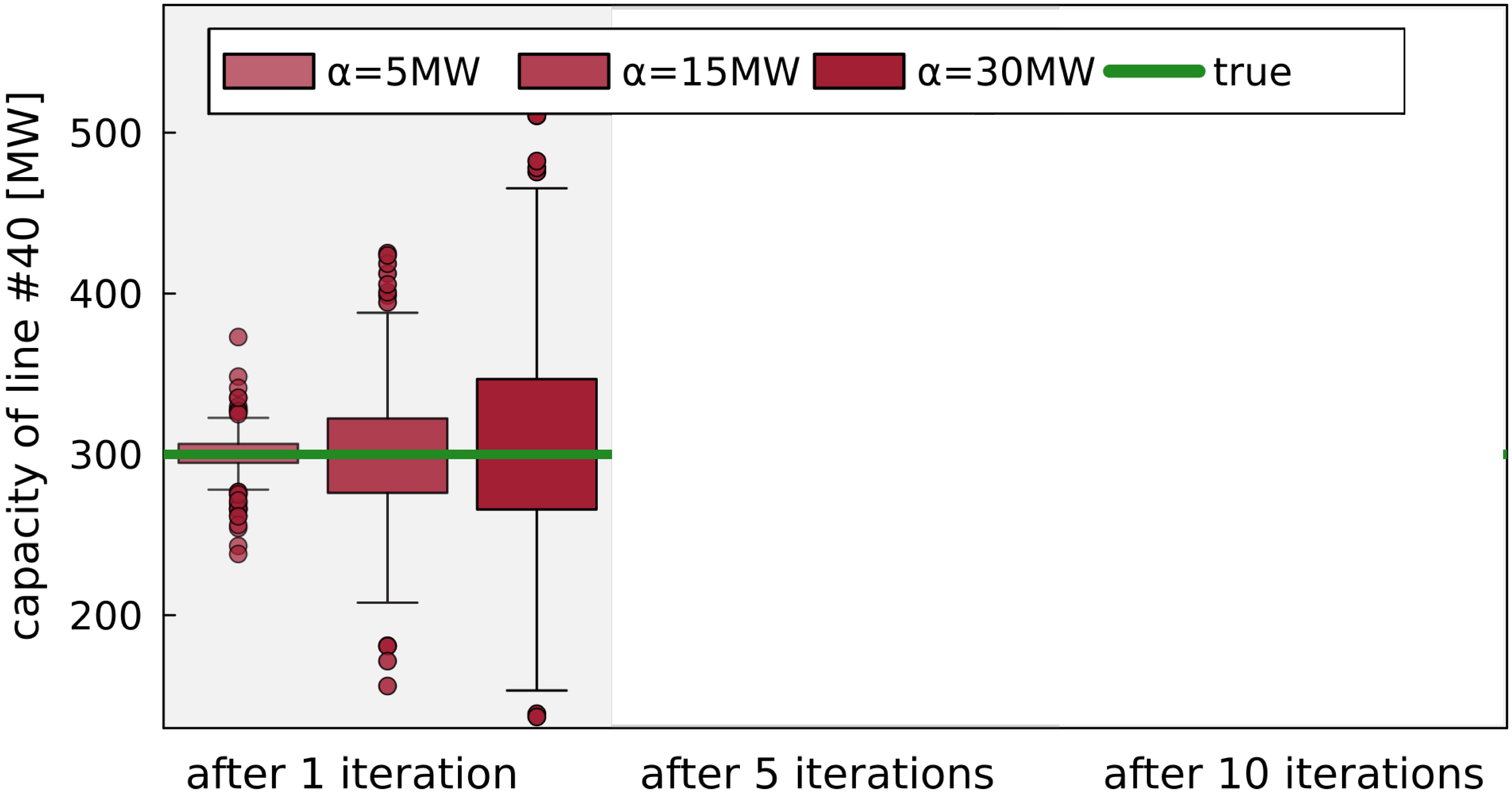
IEEE 73-RTS benchmark: TCO feasibility and sup-optimality



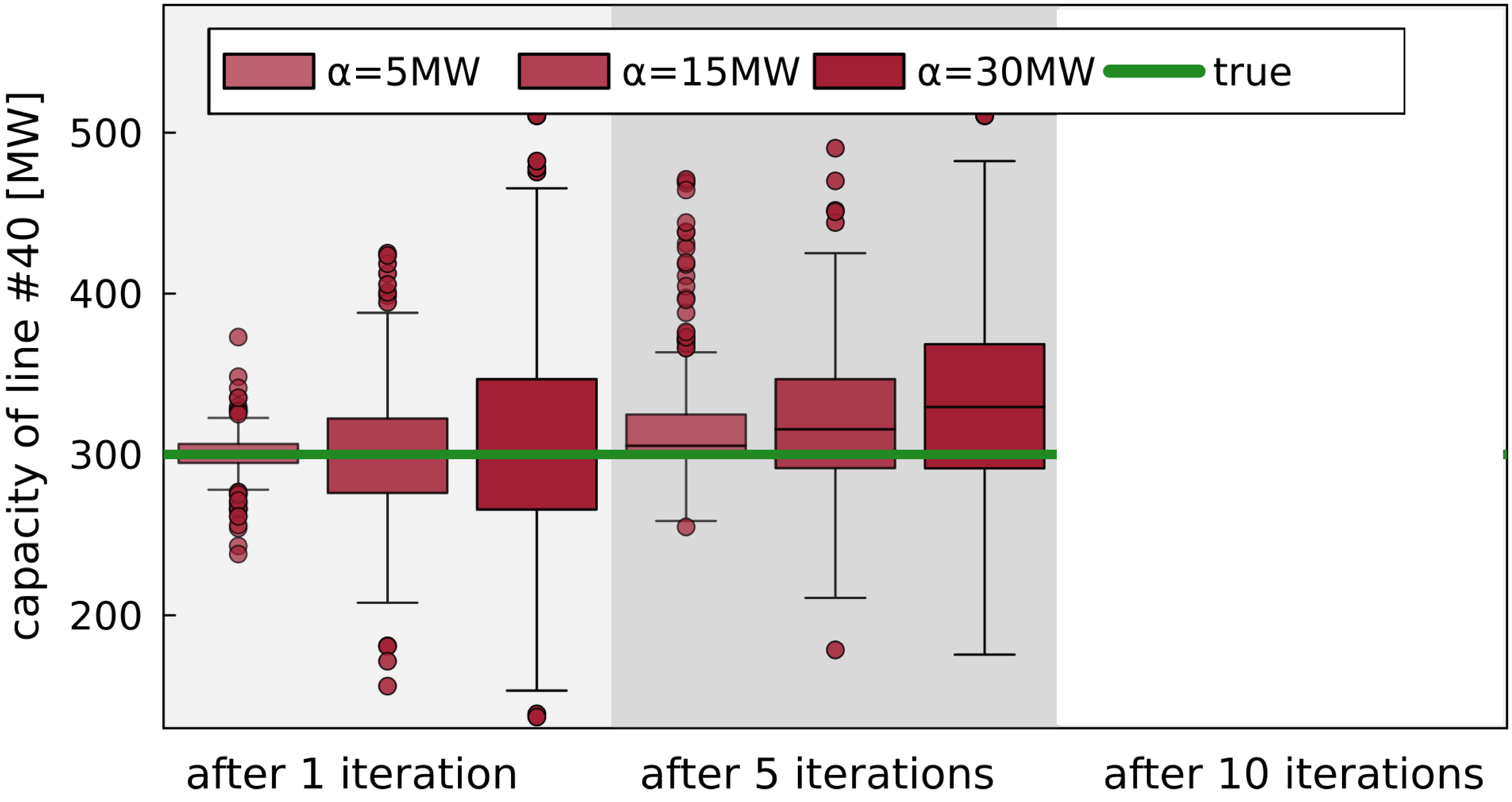
IEEE 73-RTS benchmark: TCO feasibility and sup-optimality



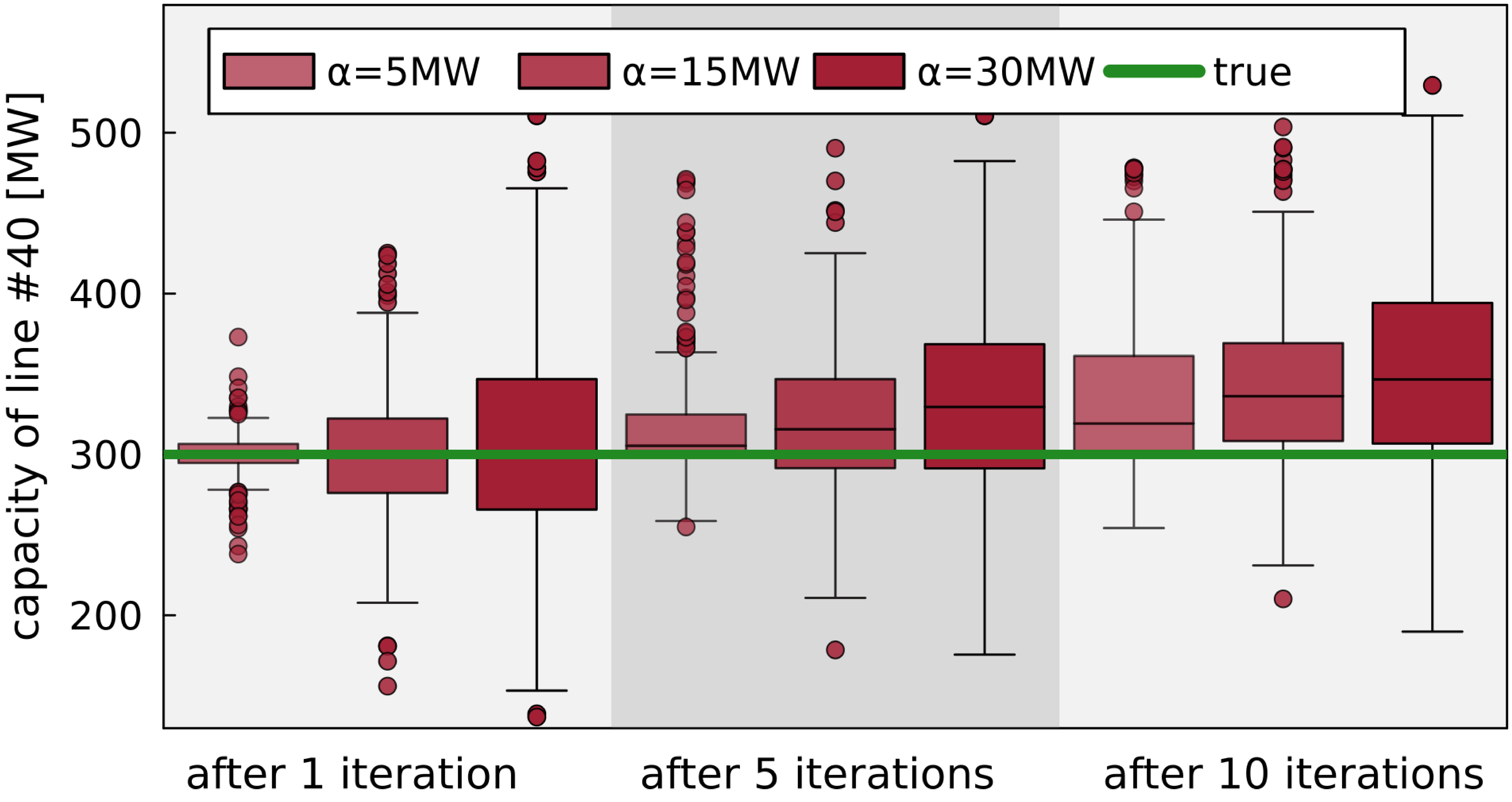
IEEE 73-RTS benchmark: TCO robustness bias



IEEE 73-RTS benchmark: TCO robustness bias



IEEE 73-RTS benchmark: TCO robustness bias



Future of synthetic power system datasets

What we used to say about synthetic datasets:

- "[...] data bears **no relation** to the actual grid [...]"
- "This test case represents [...] **fictitious** transmission"
- "This case is synthetic and **does not** model the actual grid"

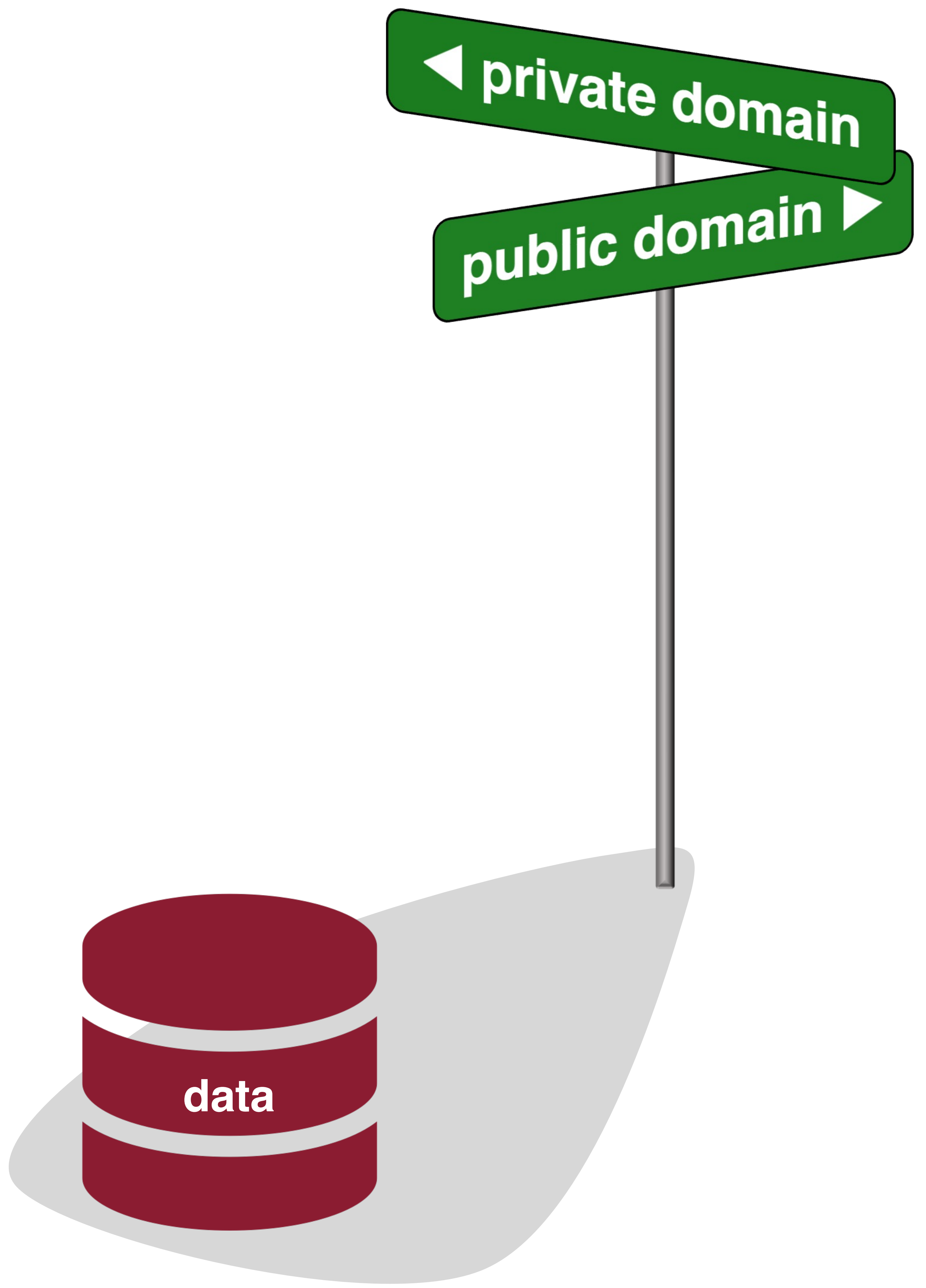
What we will say about synthetic datasets:

- "This synthetic dataset is produced based on the data from a real-world power grid"
- "It is not possible to infer the real data from this synthetic dataset"
- "Computational results on this data are consistent with the real data"

What does it mean for electricity market operators?

- New algorithms for **controllable** market transparency:
  - infrastructure data (grid topology, network parameters, generation, loads, etc.)
  - market participation data (bidding quantities, prices, etc.)
- No need for aggregation:
  - system cost/load $\Longrightarrow$ nodal cost/load
  - aggregated generation $\Longrightarrow$ highly granular generation records
- Rigorous privacy quantification $\Longrightarrow$ legal compliance (e.g., US Census Bureau)

Where data should go?



Our $\varepsilon-$differentially private algorithms provide a **non-discrete** answer to this question!

Thank you for your attention!

From this talk:

1 Dvorkin, V., Botterud A.
Differentially private algorithms for synthetic power system datasets
IEEE Control Systems Letters, 2023

Other references:

2 Dvorkin, V., Fioretto, F., Van Hentenryck, P., Kazempour, J. and Pinson, P.
Privacy-preserving convex optimization: When differential privacy meets stochastic programming
Priprint, arXiv preprint arXiv:2006.12338, 2022

3 Dvorkin, V., Fioretto, F., Van Hentenryck, P., Pinson, P. and Kazempour J.
Differentially private optimal power flow for distribution grids
IEEE Transactions on Power Systems, 2021
Best 2019–2021 Paper Award

4 Dvorkin, V., Van Hentenryck, P., Kazempour, J. and Pinson P.
Differentially private distributed optimal power flow
2020 Conference on Decision and Control

Let's stay in touch:

DvorkinVladimir   Vladimir-Dvorkin   dvorkin@mit.edu