Differentially Private Federated Bayesian Optimization with Distributed Exploration

Zhongxiang Dai [1]   Bryan Kian Hsiang Low [1]   Patrick Jaillet [2]

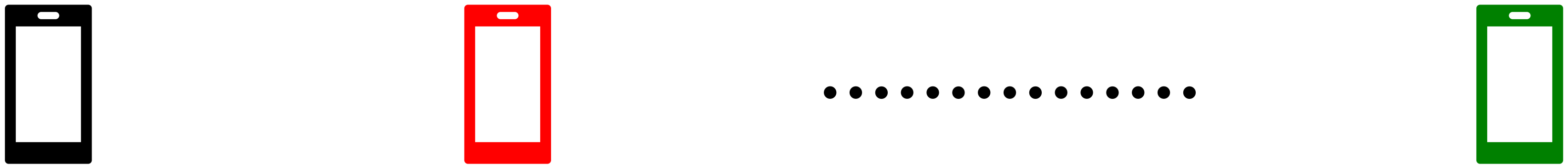
[1] Department of Computer Science, National University of Singapore

[2] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Differentially Private Federated Bayesian Optimization

- **Bayesian optimization** (BO) has been extended to the federated setting, yielding the **federated Thompson sampling** (**FTS**) algorithm (Dai et. al., 2020)

- **FTS** facilitates collaborative black-box optimization **without sharing raw data**:
  - Multiple mobile phone users can collaborate to optimize the hyperparameters of their deep neural networks for a smart keyboard
  - Multiple hospitals can collaborate to select patients for performing a medical test



Differentially Private Federated Bayesian Optimization

- **Rigorous privacy preservation** has been an important consideration for both federated learning (FL) and BO.
- However, **the FTS algorithm** (Dai et. al., 2020) is **not** equipped with a rigorous preservation of the privacy of the users/agents.

Differentially Private Federated Bayesian Optimization

Differential Privacy

- **Differential Privacy (DP)** has been widely used in privacy-preserving ML
  - DP-SGD: adding DP to the training of DNN
  - **DP-FedAvg:** adding DP to FL to preserve the **user-level privacy**

An adversary cannot infer whether a user has participated in the algorithm

Differentially Private Federated Bayesian Optimization

Differential Privacy

- **Differential Privacy (DP)** has been widely used in privacy-preserving ML
  - DP-SGD: adding DP to the training of DNN
  - **DP-FedAvg:** adding DP to FL to preserve the **user-level privacy**

An adversary cannot infer whether a user has participated in the algorithm

- An algorithm satisfying **user-level (ϵ, δ)-DP** ensures that adding/removing any single user has an imperceptible impact on its output.

- Smaller $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ indicate a better privacy guarantee

Differentially Private Federated Bayesian Optimization

Differential Privacy

- Both DP-SGD and DP-FedAvg follow a general framework for **adding DP to generic iterative algorithms**
  - Apply a **subsampled Gaussian mechanism** in every iteration

preserves user-level privacy

Differentially Private Federated Bayesian Optimization

Further Improve Utility via Distributed Exploration

Differentially Private Federated Bayesian Optimization

Further Improve Utility via Distributed Exploration

- The general DP framework is able to handle **different parameter vectors**
  - E.g., parameters from different layers of a DNN

Differentially Private Federated Bayesian Optimization

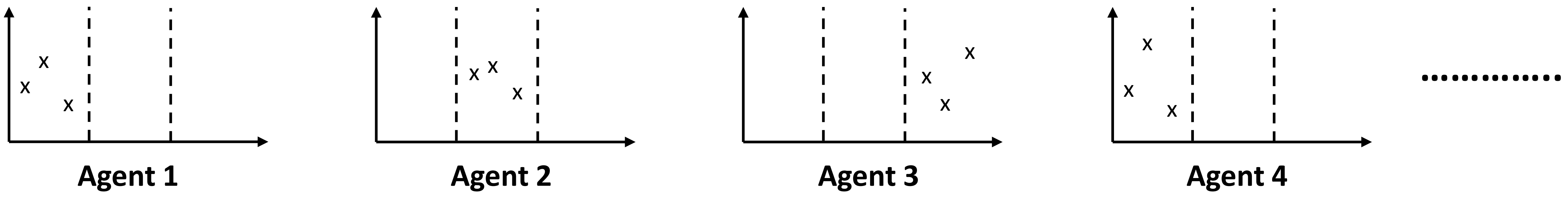
Further Improve Utility via Distributed Exploration

- The general DP framework is able to handle **different parameter vectors**
  - E.g., parameters from different layers of a DNN

↓ **+ local modelling for BO**

Distributed Exploration (DE)
(at initialization)

smaller local sub-regions
↓
improves modelling of GP surrogate
↓
better performances



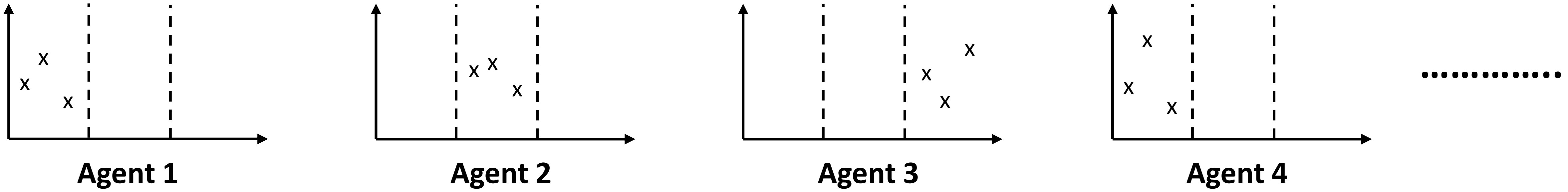
Differentially Private Federated Bayesian Optimization

Further Improve Utility via Distributed Exploration

- The general DP framework is able to handle **different parameter vectors**
  - E.g., parameters from different layers of a DNN

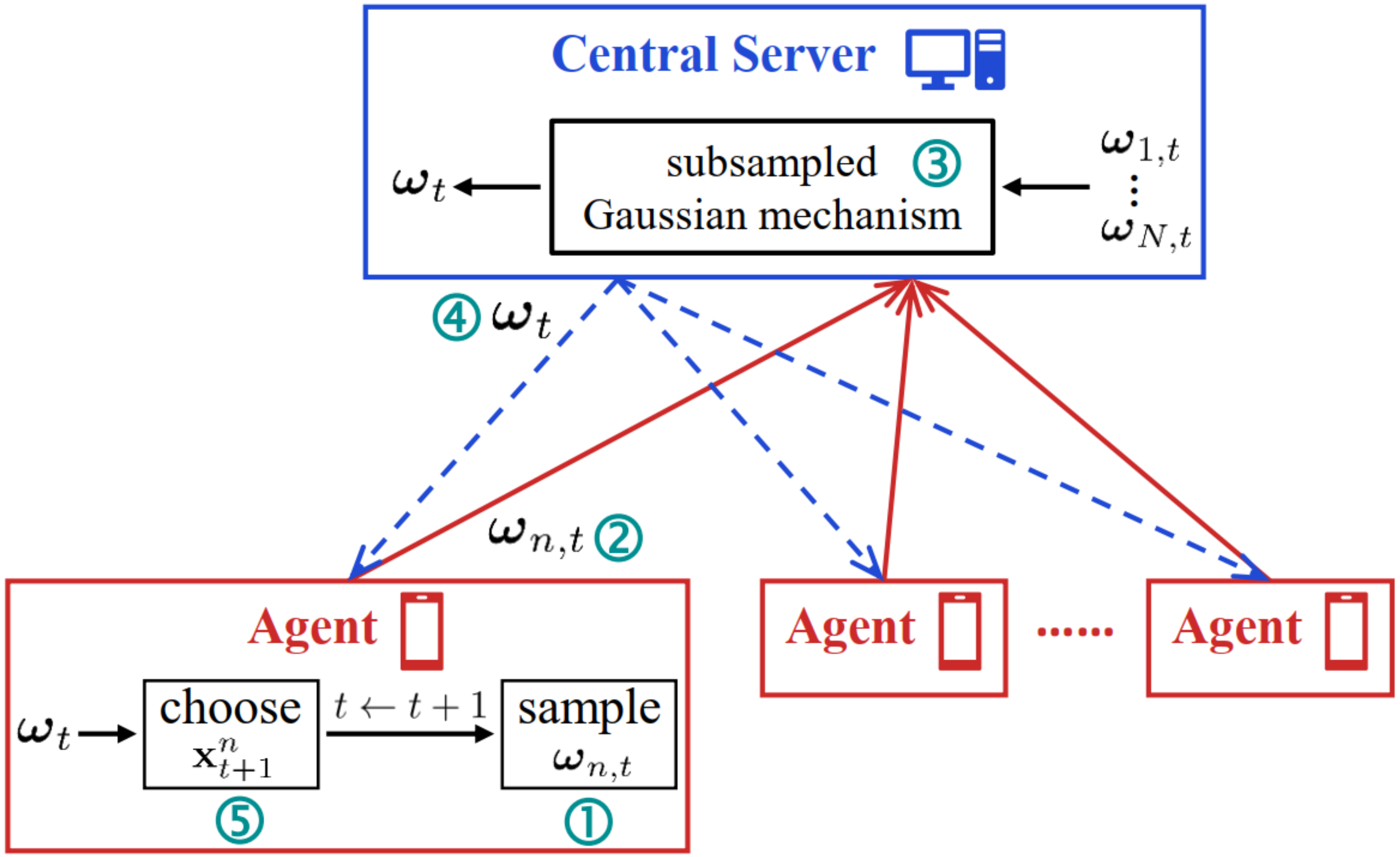
↓ **+ local modelling for BO**

DP-FTS + **Distributed Exploration (DE)** (at initialization) → **DP-FTS-DE**



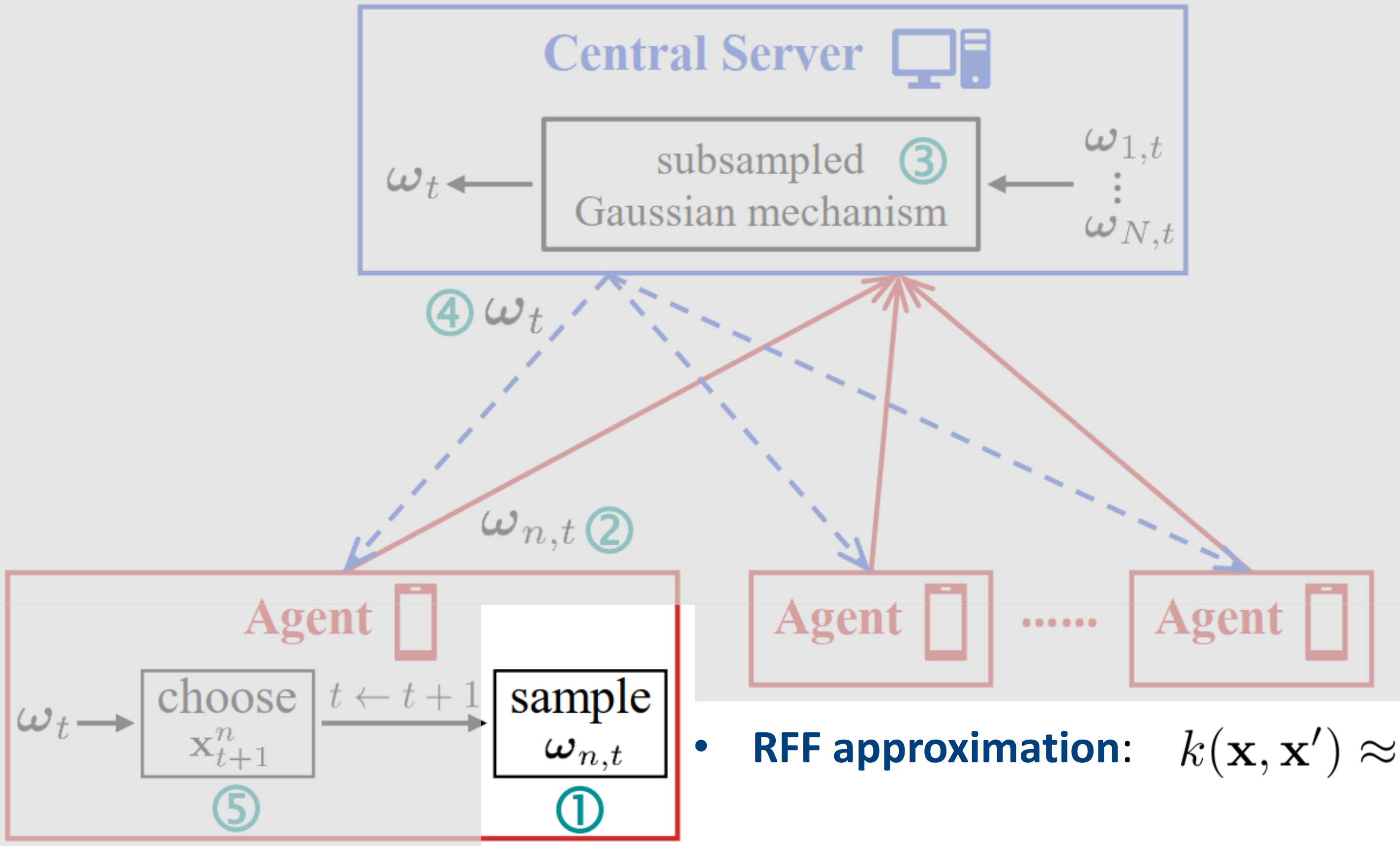
Differentially Private Federated Bayesian Optimization

DP-FTS (without DE)



Differentially Private Federated Bayesian Optimization

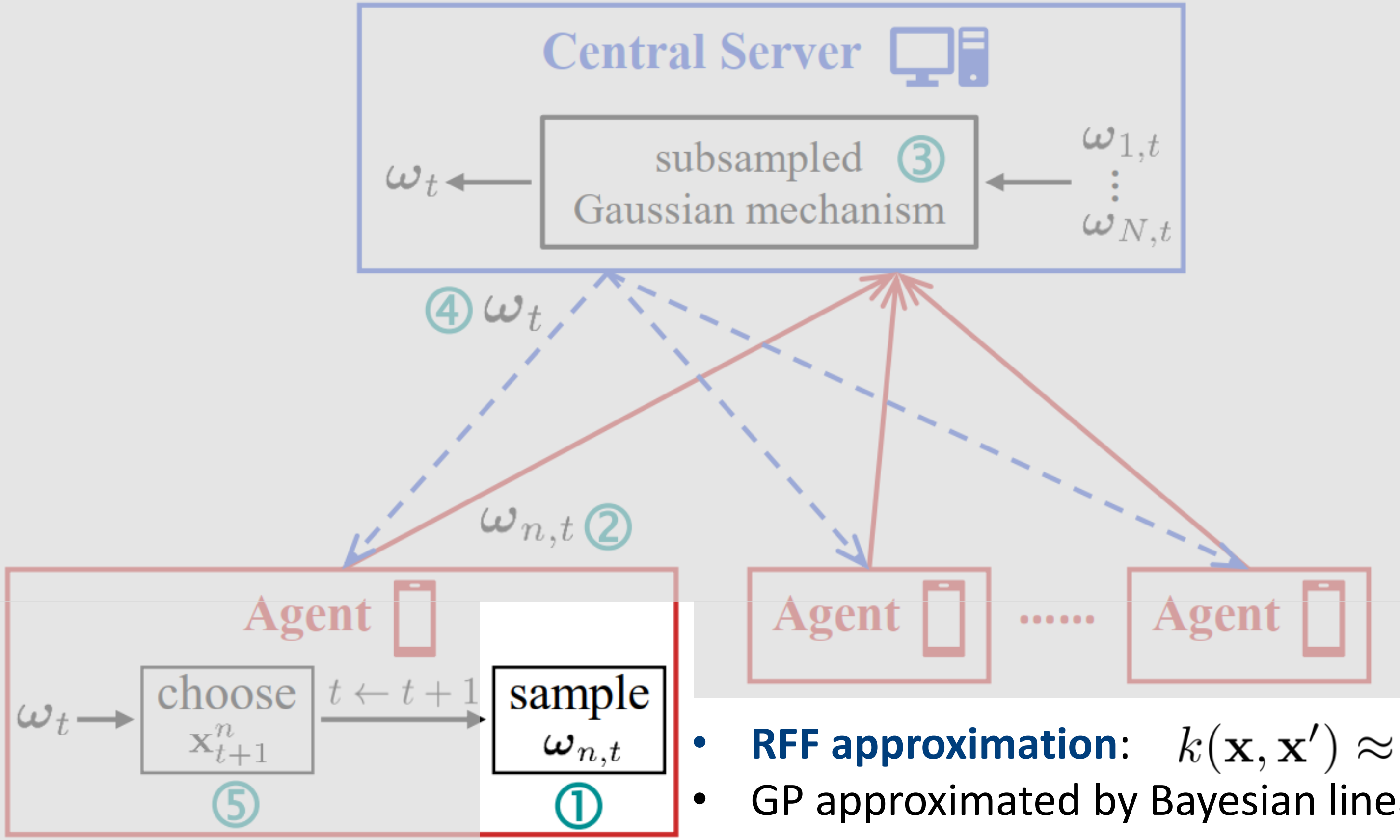
DP-FTS (without DE)



- **RFF approximation:** $k(\mathbf{x}, \mathbf{x}') \approx \phi(\mathbf{x})^\top \phi(\mathbf{x}')$

Differentially Private Federated Bayesian Optimization

DP-FTS (without DE)

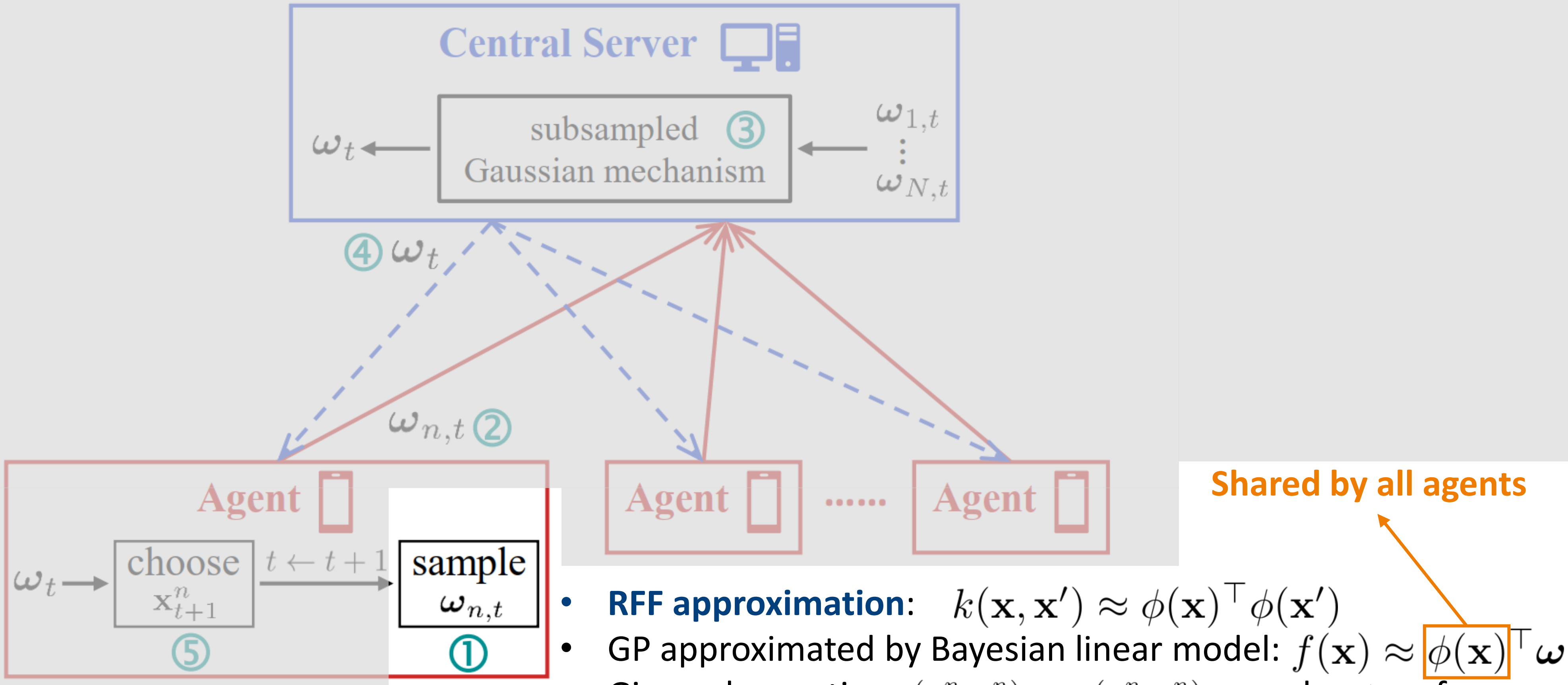


Shared by all agents

- **RFF approximation**: $k(\mathbf{x}, \mathbf{x}') \approx \phi(\mathbf{x})^\top \phi(\mathbf{x}')$
- GP approximated by Bayesian linear model: $f(\mathbf{x}) \approx \phi(\mathbf{x})^\top \omega$

Differentially Private Federated Bayesian Optimization

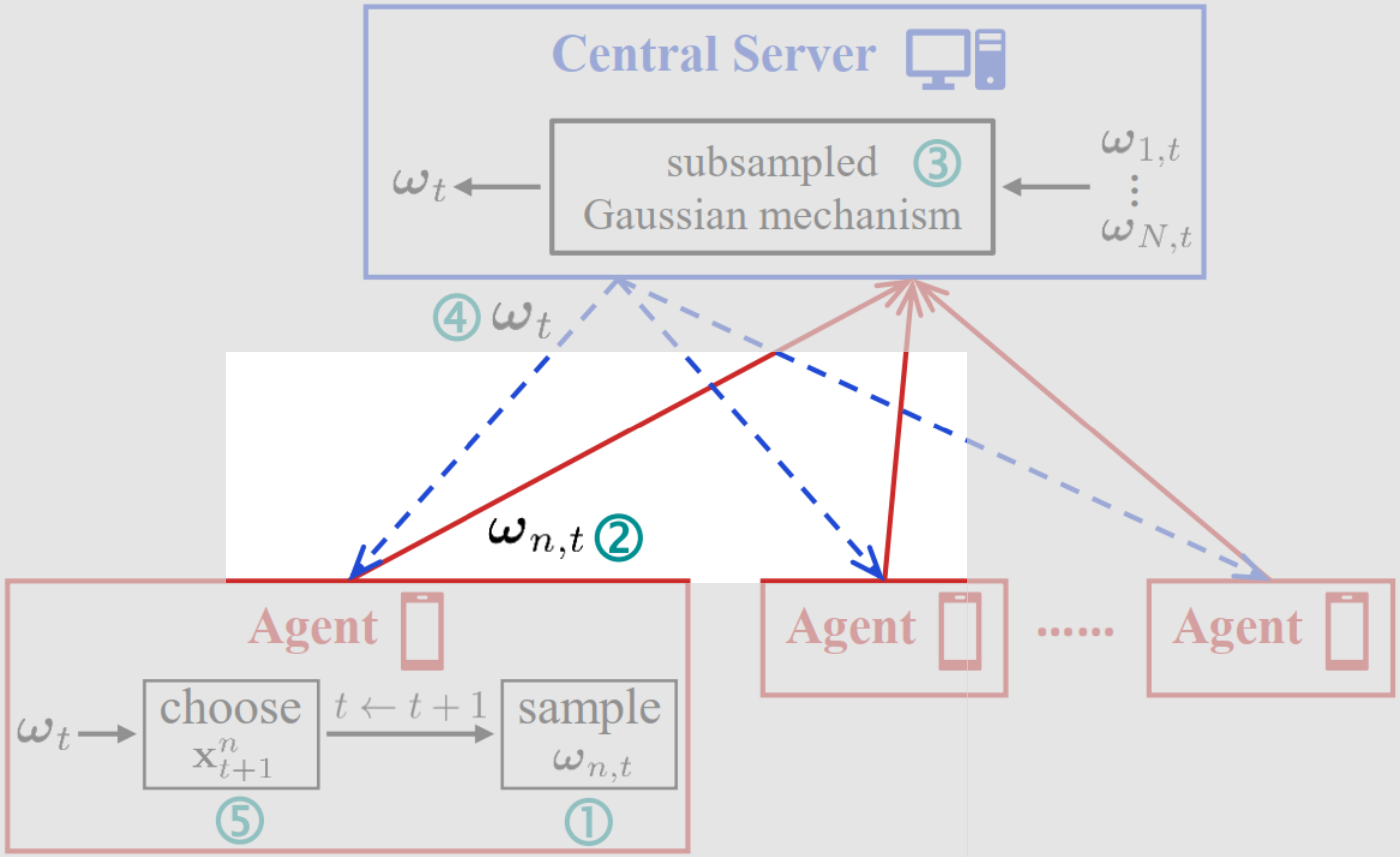
DP-FTS (without DE)



- **RFF approximation**: $k(\mathbf{x}, \mathbf{x}') \approx \phi(\mathbf{x})^\top \phi(\mathbf{x}')$
- GP approximated by Bayesian linear model: $f(\mathbf{x}) \approx \phi(\mathbf{x})^\top \boldsymbol{\omega}$
- Given observations $(\mathbf{x}^n_1, y^n_1), \ldots, (\mathbf{x}^n_t, y^n_t)$, sample $\boldsymbol{\omega}_{n,t}$ from **posterior of** $\boldsymbol{\omega}$ (sampled function: $f_{n,t}(\mathbf{x}) = \phi(\mathbf{x})^\top \boldsymbol{\omega}_{n,t}$)

Differentially Private Federated Bayesian Optimization

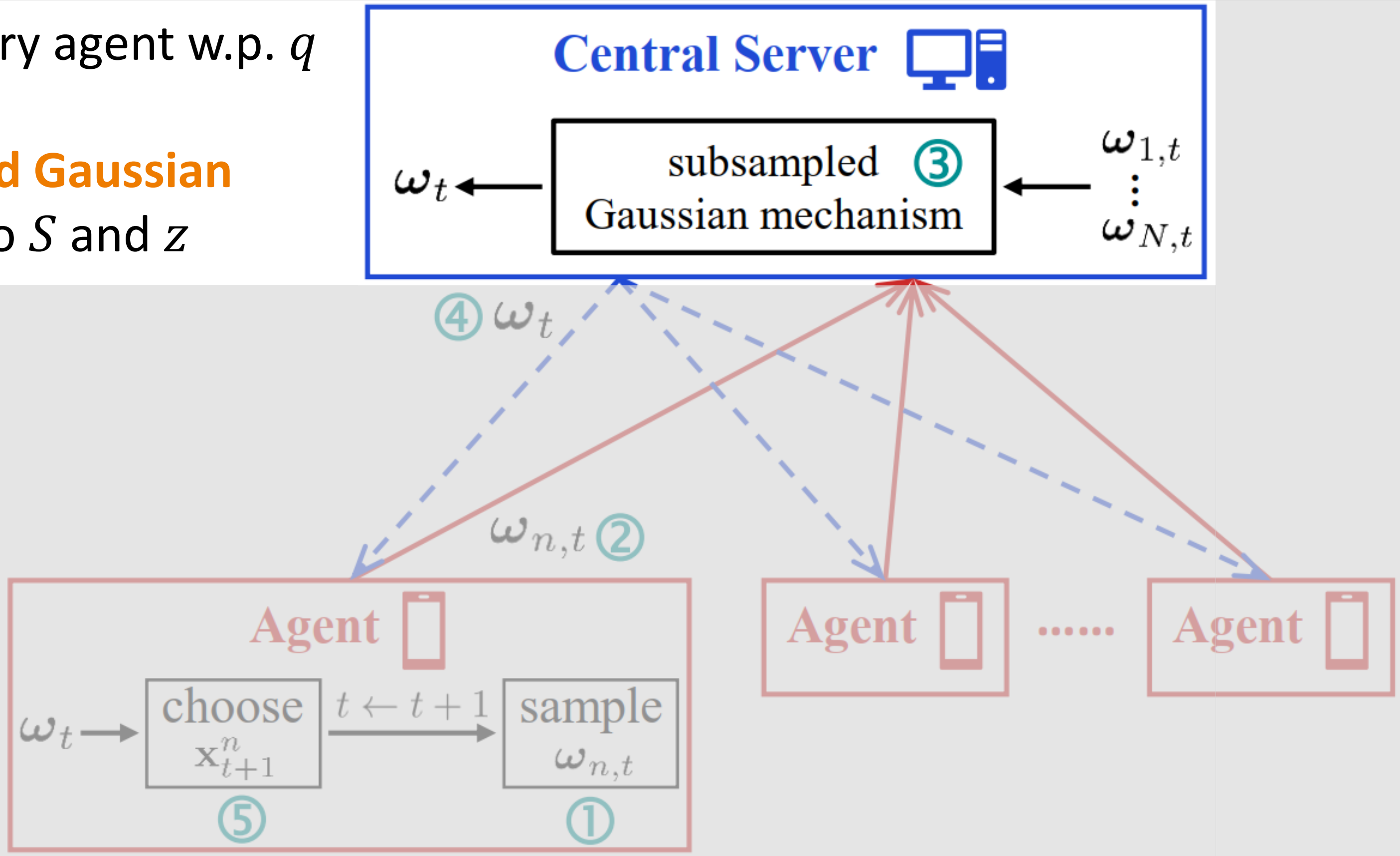
DP-FTS (without DE)



Differentially Private Federated Bayesian Optimization

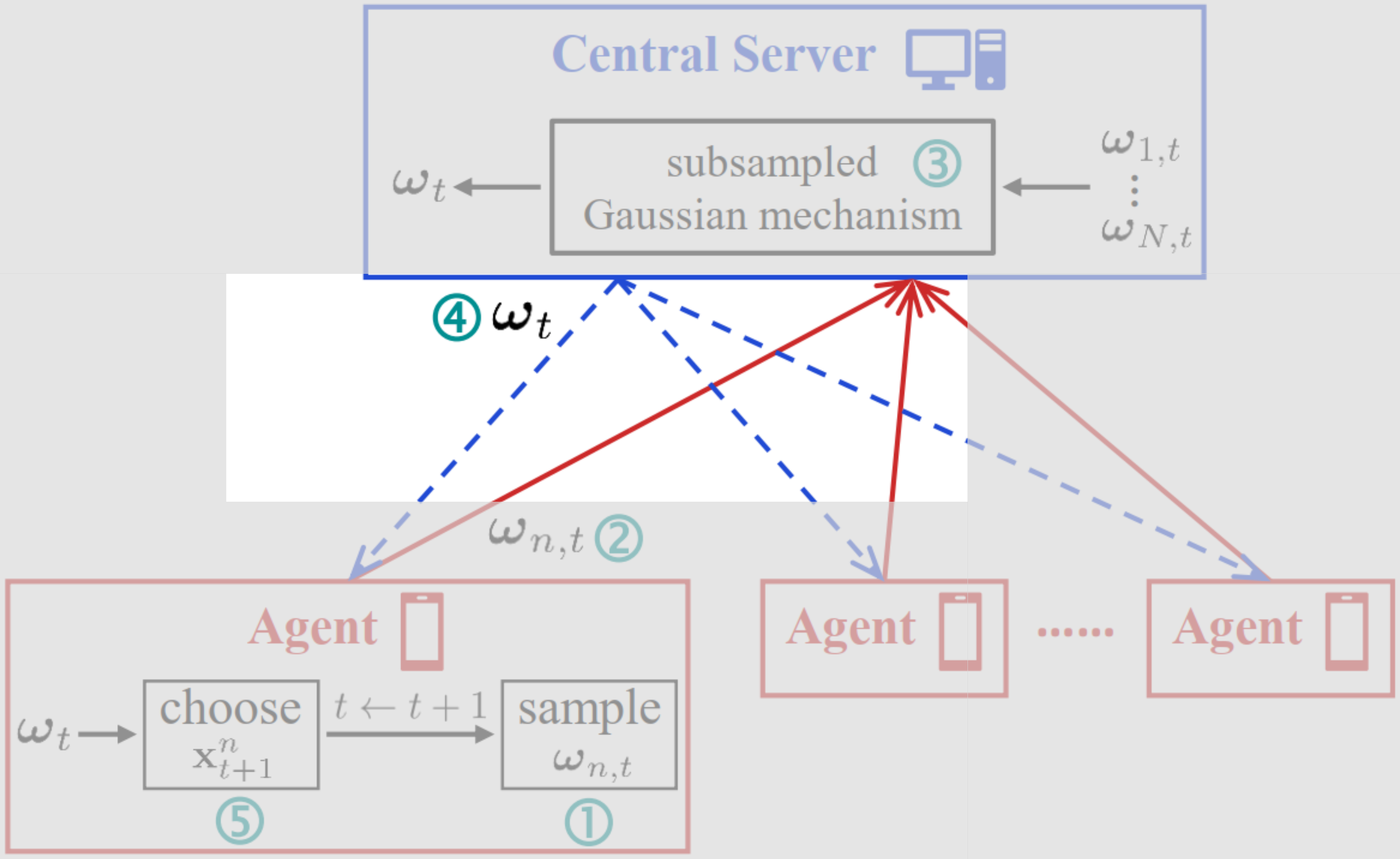
DP-FTS (without DE)

1. **Subsample:** select every agent w.p. q
2. **Clip:** $||\omega_{1,n}||_2 \leq S$
3. **Weighted average, add Gaussian noise** with std. prop. to S and z



Differentially Private Federated Bayesian Optimization

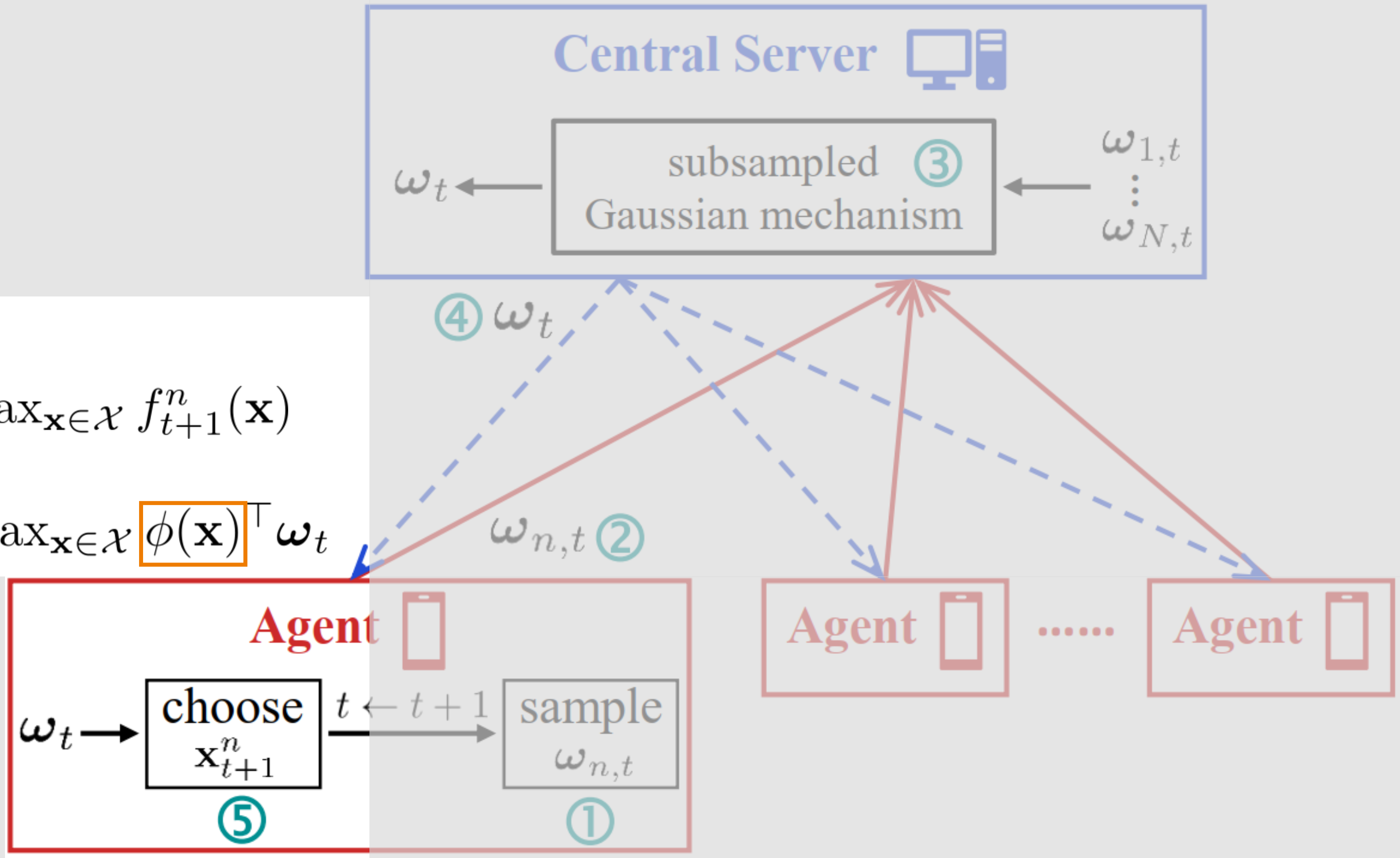
DP-FTS (without DE)



Differentially Private Federated Bayesian Optimization

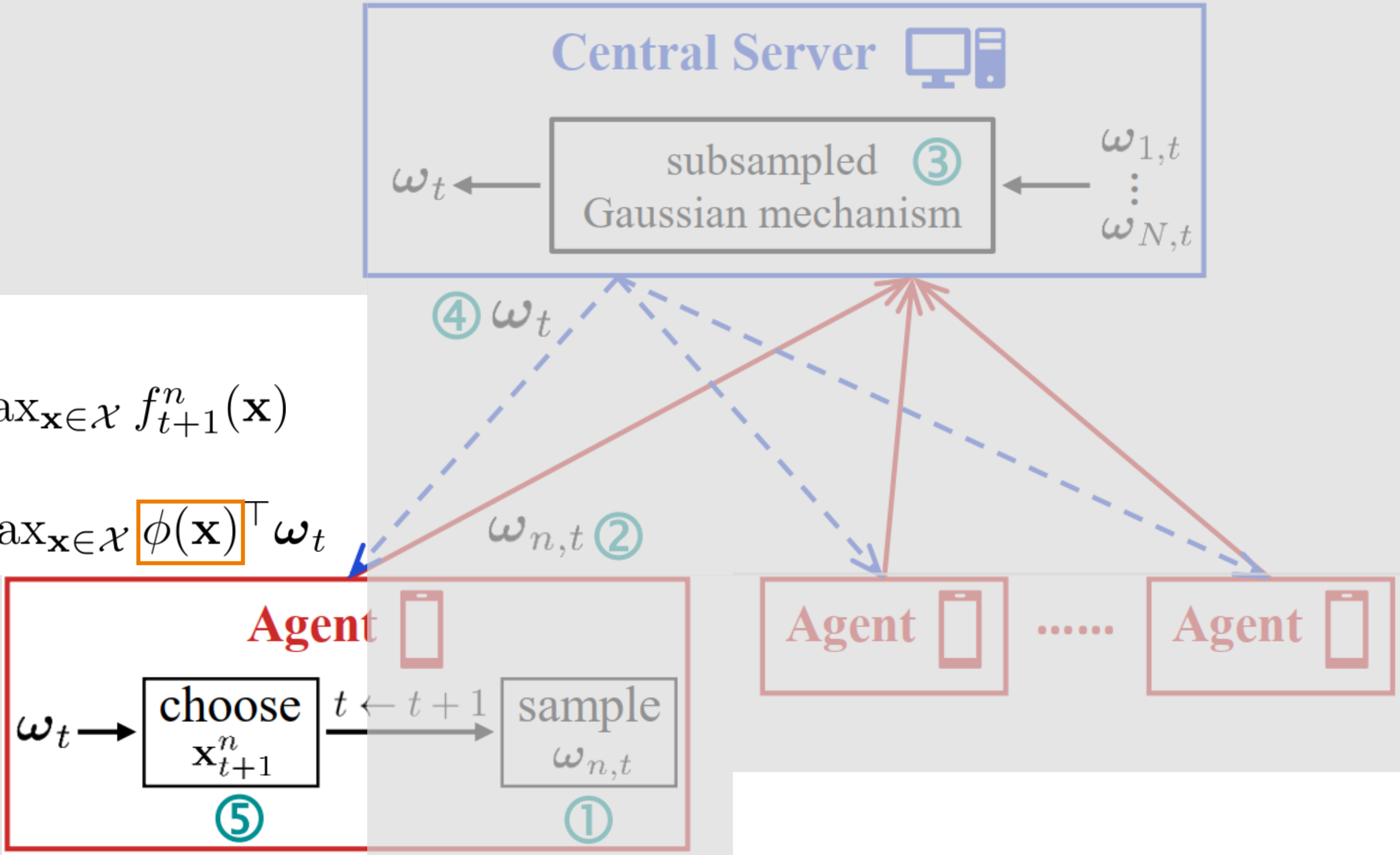
DP-FTS (without DE)

- With probability p_t:
  - Choose $\mathbf{x}_{t+1}^n = \arg\max_{\mathbf{x}\in\mathcal{X}} f_{t+1}^n(\mathbf{x})$
- With probability $1 - p_t$:
  - Choose $\mathbf{x}_{t+1}^n = \arg\max_{\mathbf{x}\in\mathcal{X}} \phi(\mathbf{x})^\top \boldsymbol{\omega}_t$



Differentially Private Federated Bayesian Optimization

DP-FTS (without DE)



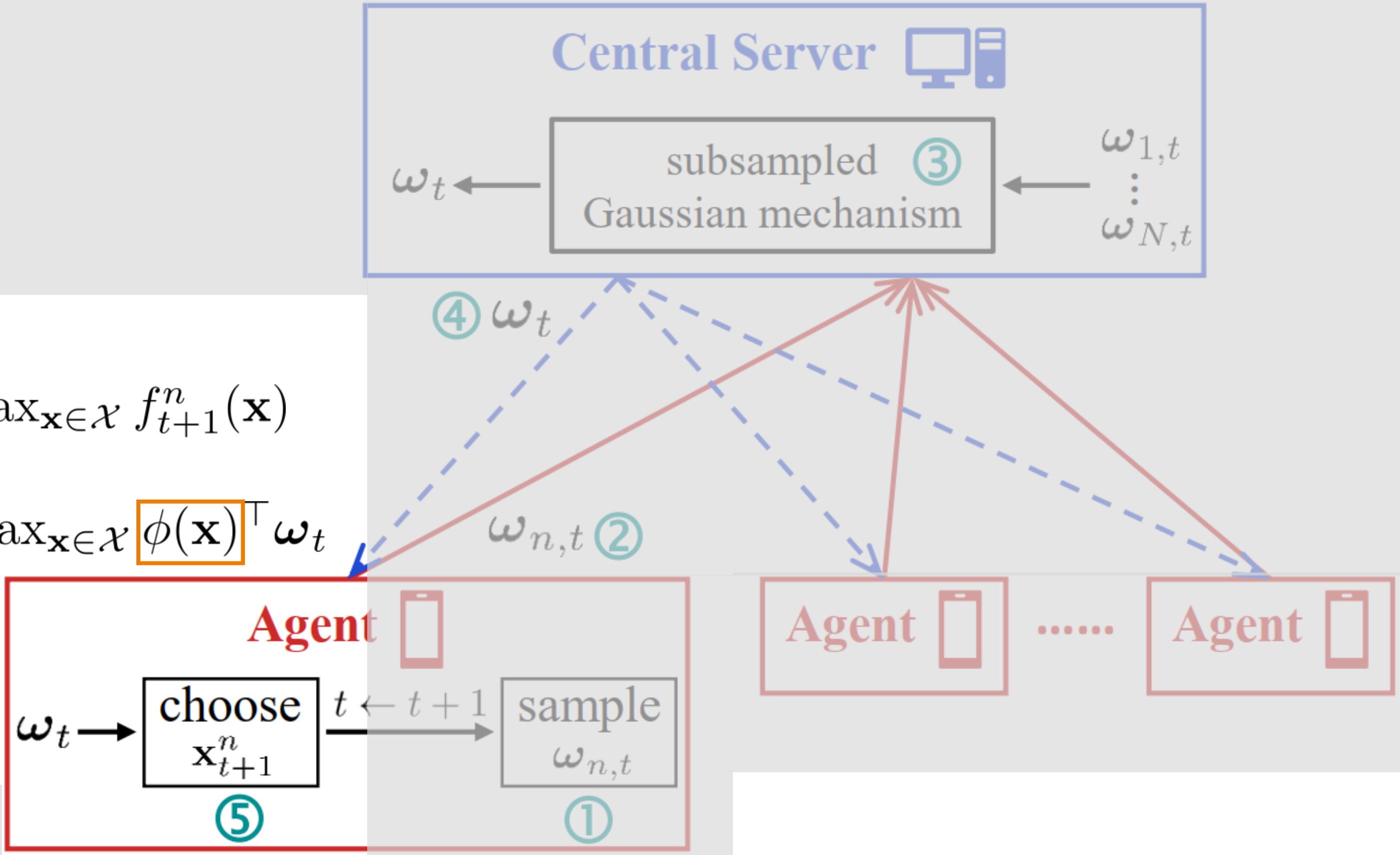
- With probability p_t:
  - Choose $\mathbf{x}^n_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{X}} f^n_{t+1}(\mathbf{x})$
- With probability $1 - p_t$:
  - Choose $\mathbf{x}^n_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{X}} \phi(\mathbf{x})^\top \boldsymbol{\omega}_t$

Shared by all agents

- Given observations $(\mathbf{x}^n_1, y^n_1), \ldots, (\mathbf{x}^n_t, y^n_t)$, sample $\boldsymbol{\omega}_{n,t}$ from **posterior of $\boldsymbol{\omega}$** (sampled function: $f_{n,t}(\mathbf{x}) = \phi(\mathbf{x})^\top \boldsymbol{\omega}_{n,t}$)

Differentially Private Federated Bayesian Optimization

DP-FTS (without DE)



- With probability p_t:
  - Choose $\mathbf{x}_{t+1}^n = \arg\max_{\mathbf{x}\in\mathcal{X}} f_{t+1}^n(\mathbf{x})$
- With probability $1 - p_t$:
  - Choose $\mathbf{x}_{t+1}^n = \arg\max_{\mathbf{x}\in\mathcal{X}} \phi(\mathbf{x})^\top \boldsymbol{\omega}_t$

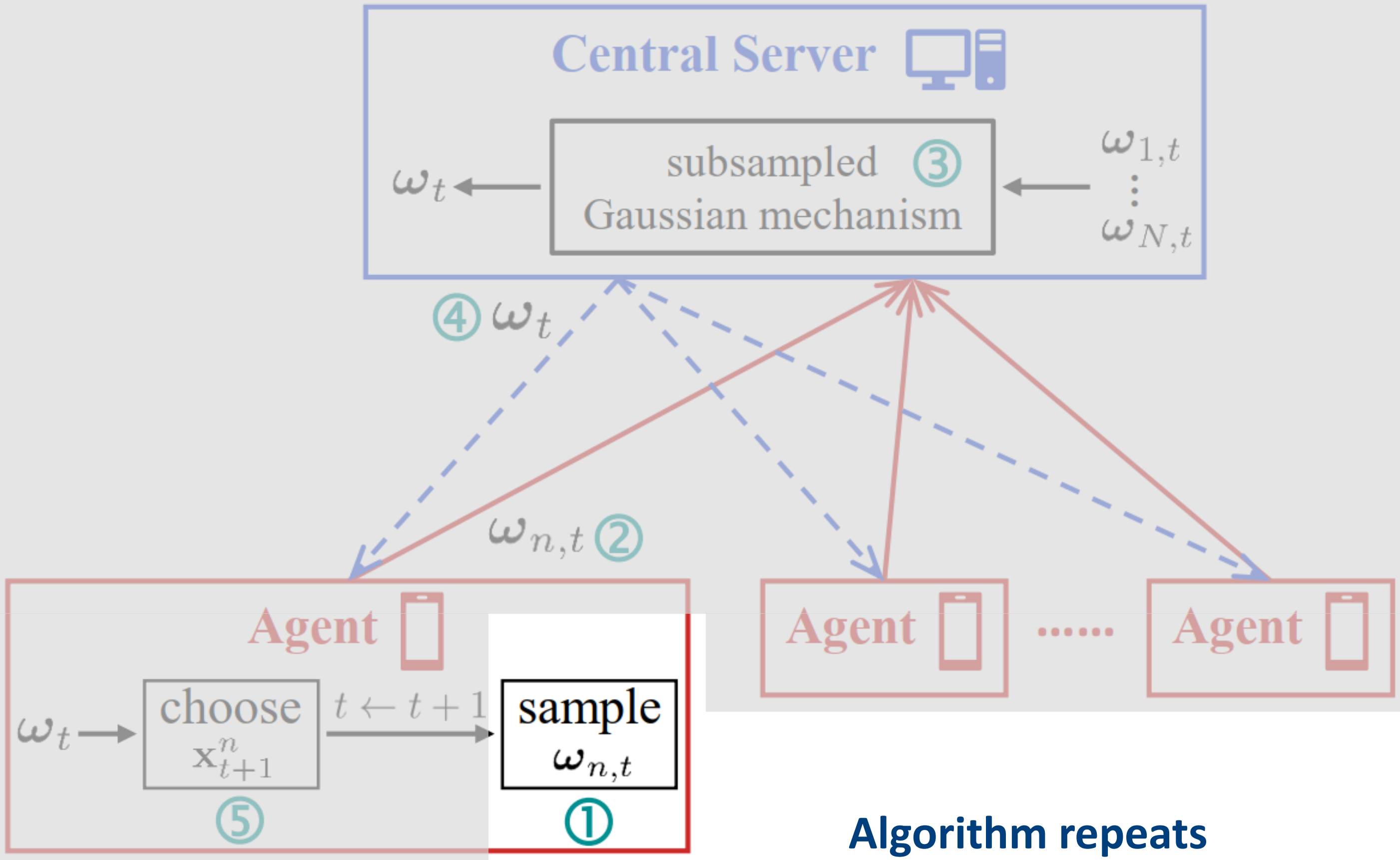
p_t monotonically increasing
$1 - p_t = \mathcal{O}(1/t^2)$

Shared by all agents

- Given observations $(\mathbf{x}_1^n, y_1^n), \ldots, (\mathbf{x}_t^n, y_t^n)$, sample $\boldsymbol{\omega}_{n,t}$ from **posterior of $\boldsymbol{\omega}$** (sampled function: $f_{n,t}(\mathbf{x}) = \phi(\mathbf{x})^\top \boldsymbol{\omega}_{n,t}$)

Differentially Private Federated Bayesian Optimization

DP-FTS (without DE)



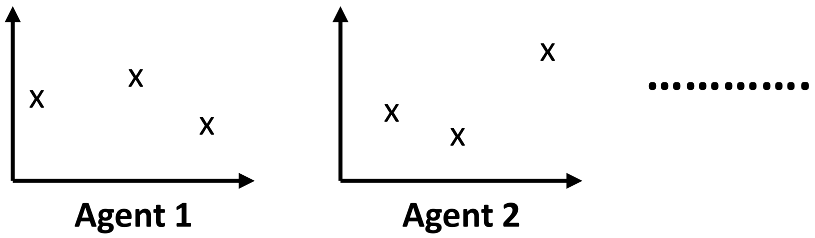
Algorithm repeats

Differentially Private Federated Bayesian Optimization

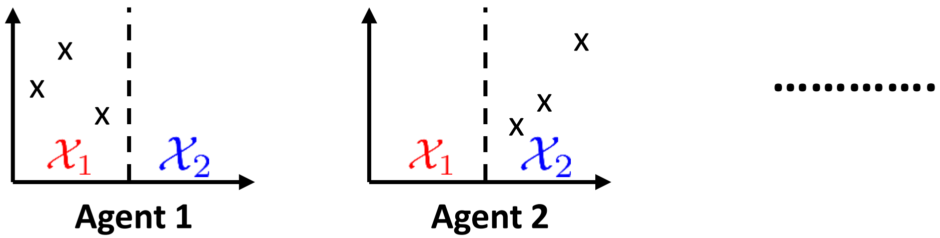
Distributed Exploration (DE)

DP-FTS

Initialization



DP-FTS-DE (P=2)

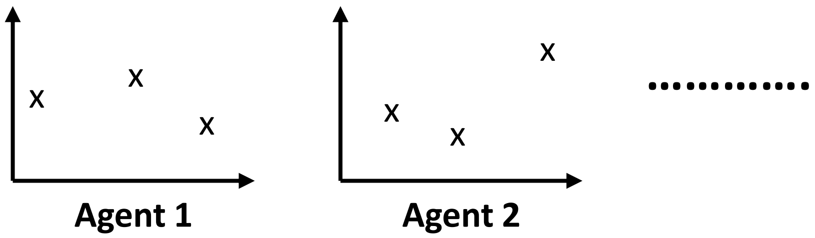


Differentially Private Federated Bayesian Optimization

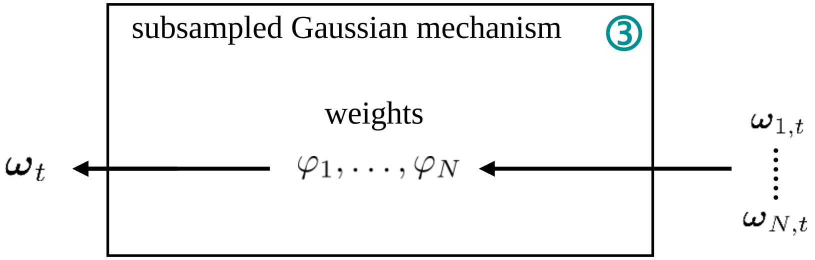
Distributed Exploration (DE)

DP-FTS

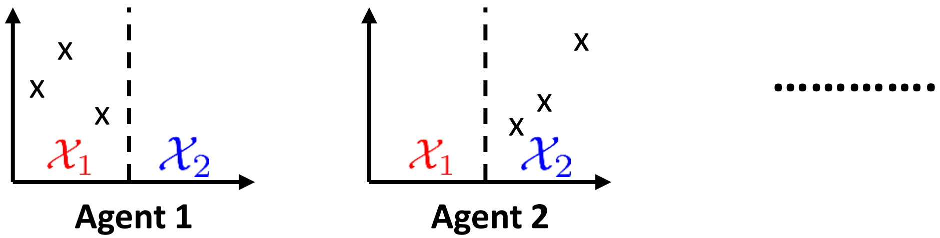
Initialization

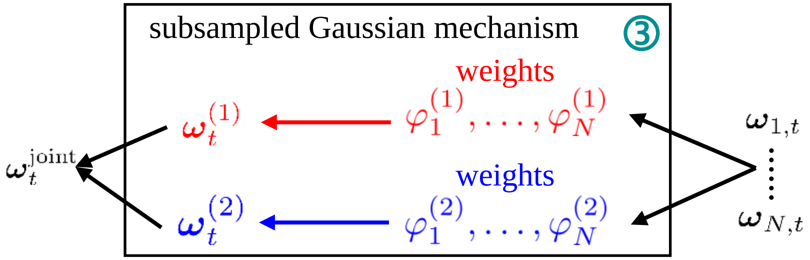


DP Transformations



DP-FTS-DE (P=2)



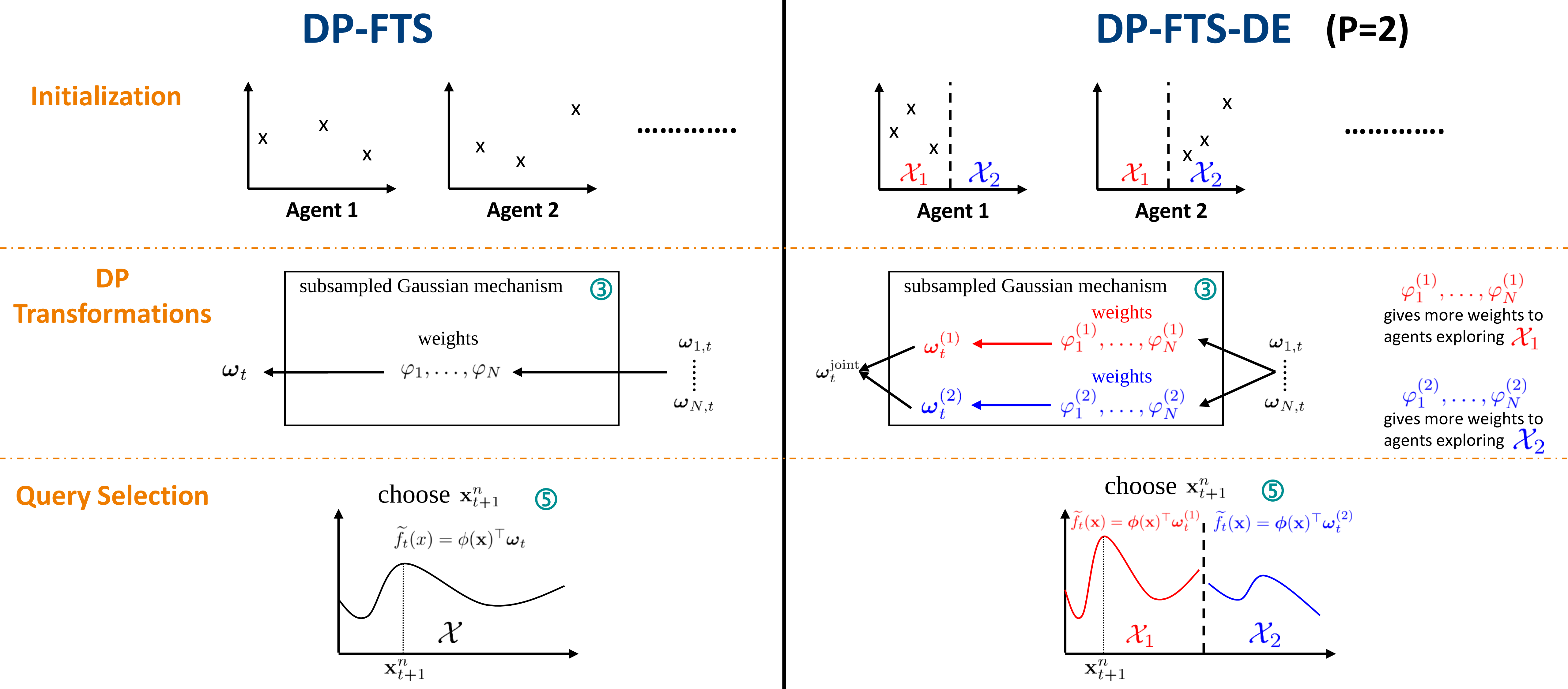


$\varphi_1^{(1)}, \ldots, \varphi_N^{(1)}$ gives more weights to agents exploring $\mathcal{X}_1$

$\varphi_1^{(2)}, \ldots, \varphi_N^{(2)}$ gives more weights to agents exploring $\mathcal{X}_2$

Differentially Private Federated Bayesian Optimization

Distributed Exploration (DE)



Differentially Private Federated Bayesian Optimization

Theoretical Analysis

Proposition 1 (Privacy Guarantee). *There exist constants c_1 and c_2 such that for fixed q and T and any $\epsilon < c_1 q^2 T, \delta > 0$, DP-FTS-DE (Algo. 1) is (ϵ, δ)-DP if $z \geq c_2 q\sqrt{T \log(1/\delta)}/\epsilon$.*

Theorem 1 (Utility Guarantee). *Define $\mathcal{C}_t \triangleq \{n \in [N] \big| ||\boldsymbol{\omega}_{n,t}||_2 > S/\sqrt{P}\}$. W.p. $\geq 1-\delta$,*

$$R_T^1 = \tilde{\mathcal{O}}\Big((B + 1/p_1)\, \gamma_T \sqrt{T} + \textstyle\sum_{t=1}^{T} \psi_t + B \sum_{t=1}^{T} \vartheta_t \Big)$$

where $\psi_t \triangleq \tilde{\mathcal{O}}((1-p_t) P \varphi_{\max} q^{-1} (\Delta_t + zS\sqrt{M}))$, $\Delta_t \triangleq \sum_{n=1}^{N} \Delta_{n,t}$, $\Delta_{n,t} \triangleq \tilde{\mathcal{O}}(\varepsilon B t^2 + B + \sqrt{M} + d_n + \sqrt{\gamma_t})$, and $\vartheta_t \triangleq (1-p_t) \sum_{i=1}^{P} \sum_{n \in \mathcal{C}_t} \varphi_n^{(i)}$.

- **Privacy-utility trade-off**

Differentially Private Federated Bayesian Optimization

Theoretical Analysis

Proposition 1 (Privacy Guarantee). *There exist constants c_1 and c_2 such that for fixed q and T and any $\epsilon < c_1 q^2 T, \delta > 0$, DP-FTS-DE (Algo. 1) is (ϵ, δ)-DP if $z \geq c_2 q \sqrt{T \log(1/\delta)}/\epsilon$.*

Theorem 1 (Utility Guarantee). *Define $\mathcal{C}_t \triangleq \{n \in [N] \,|\, ||\boldsymbol{\omega}_{n,t}||_2 > S/\sqrt{P}\}$. W.p. $\geq 1 - \delta$,*

$$R_T^1 = \tilde{\mathcal{O}}\Big(\left(B + 1/p_1\right) \gamma_T \sqrt{T} + \textstyle\sum_{t=1}^{T} \psi_t + B \sum_{t=1}^{T} \vartheta_t \Big)$$

where $\psi_t \triangleq \tilde{\mathcal{O}}((1 - p_t) P \varphi_{\max} q^{-1}(\Delta_t + z S \sqrt{M}))$, $\Delta_t \triangleq \sum_{n=1}^{N} \Delta_{n,t}$, $\Delta_{n,t} \triangleq \tilde{\mathcal{O}}(\varepsilon B t^2 + B + \sqrt{M} + d_n + \sqrt{\gamma_t})$, and $\vartheta_t \triangleq (1 - p_t) \sum_{i=1}^{P} \sum_{n \in \mathcal{C}_t} \varphi_n^{(i)}$.

- **Privacy-utility trade-off**
  - **Larger z** (larger noise variance) -> **better privacy** (Prop. 1) & **worse utility** (Theorem 1)

Differentially Private Federated Bayesian Optimization

Theoretical Analysis

Proposition 1 (Privacy Guarantee). *There exist constants c_1 and c_2 such that for fixed q and T and any $\epsilon < c_1 q^2 T, \delta > 0$, DP-FTS-DE (Algo. 1) is (ϵ, δ)-DP if $z \geq c_2 q \sqrt{T \log(1/\delta)}/\epsilon$.*

Theorem 1 (Utility Guarantee). *Define $\mathcal{C}_t \triangleq \{n \in [N] \,|\, ||\boldsymbol{\omega}_{n,t}||_2 > S/\sqrt{P}\}$. W.p. $\geq 1 - \delta$,*

$$R_T^1 = \tilde{O}\Big((B + 1/p_1)\, \gamma_T \sqrt{T} + \textstyle\sum_{t=1}^{T} \psi_t + B \sum_{t=1}^{T} \vartheta_t \Big)$$

where $\psi_t \triangleq \tilde{O}((1 - p_t) P \varphi_{\max} q^{-1} (\Delta_t + zS\sqrt{M}))$, $\Delta_t \triangleq \sum_{n=1}^{N} \Delta_{n,t}$, $\Delta_{n,t} \triangleq \tilde{O}(\varepsilon B t^2 + B + \sqrt{M} + d_n + \sqrt{\gamma_t})$, and $\vartheta_t \triangleq (1 - p_t) \sum_{i=1}^{P} \sum_{n \in \mathcal{C}_t} \varphi_n^{(i)}$.

- **Privacy-utility trade-off**
  - Larger z (larger noise variance) -> **better privacy** (Prop. 1) & **worse utility** (Theorem 1)
  - Larger q (more selected agents in an iteration) -> **worse privacy** (Prop. 1) & **better utility** (Theorem 1)

Differentially Private Federated Bayesian Optimization

Theoretical Analysis

Proposition 1 (Privacy Guarantee). *There exist constants c_1 and c_2 such that for fixed q and T and any $\epsilon < c_1 q^2 T, \delta > 0$, DP-FTS-DE (Algo. 1) is (ϵ, δ)-DP if $z \geq c_2 q \sqrt{T \log(1/\delta)}/\epsilon$.*

Theorem 1 (Utility Guarantee). *Define $\mathcal{C}_t \triangleq \{n \in [N] \,|\, ||\boldsymbol{\omega}_{n,t}||_2 > S/\sqrt{P}\}$. W.p. $\geq 1 - \delta$,*

$$R_T^1 = \tilde{O}\Big((B + 1/p_1)\, \gamma_T \sqrt{T} + \sum_{t=1}^{T} \psi_t + B \sum_{t=1}^{T} \vartheta_t \Big)$$

where $\psi_t \triangleq \tilde{O}((1 - p_t) P \varphi_{\max} q^{-1} (\Delta_t + zS\sqrt{M}))$, $\Delta_t \triangleq \sum_{n=1}^{N} \Delta_{n,t}$, $\Delta_{n,t} \triangleq \tilde{O}(\varepsilon B t^2 + B + \sqrt{M} + d_n + \sqrt{\gamma_t})$, and $\vartheta_t \triangleq (1 - p_t) \sum_{i=1}^{P} \sum_{n \in \mathcal{C}_t} \varphi_n^{(i)}$.

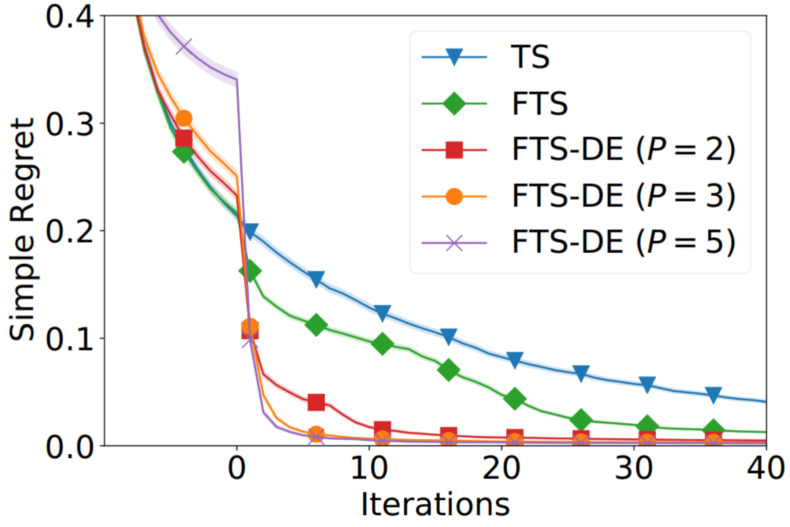
- **Two conflicting impacts of S** (clipping threshold)
  - A smaller S reduces the value of ψ_t -> better regret (due to smaller noise variance)
  - A smaller S increases the cardinality of the set $\mathcal{C}_t$ -> worse regret (due to clipping more vectors)

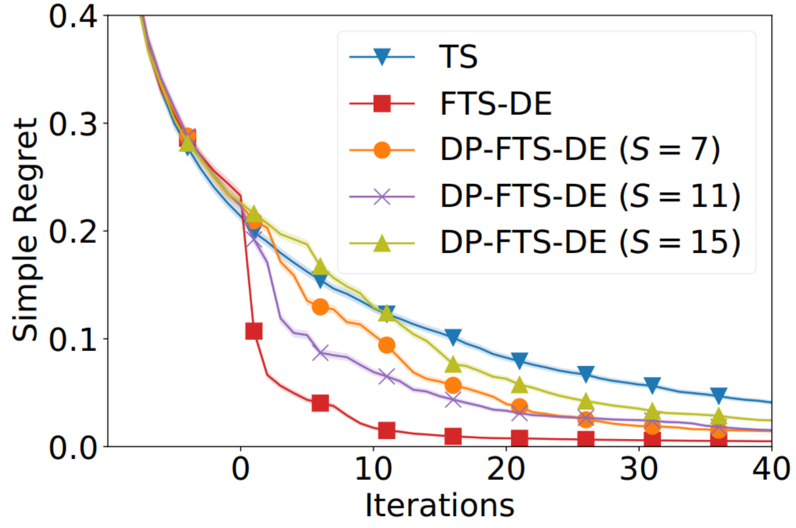
practical guideline

Choose **a small S**, while ensuring **a small number of vectors are clipped**

Differentially Private Federated Bayesian Optimization

Synthetic Experiments



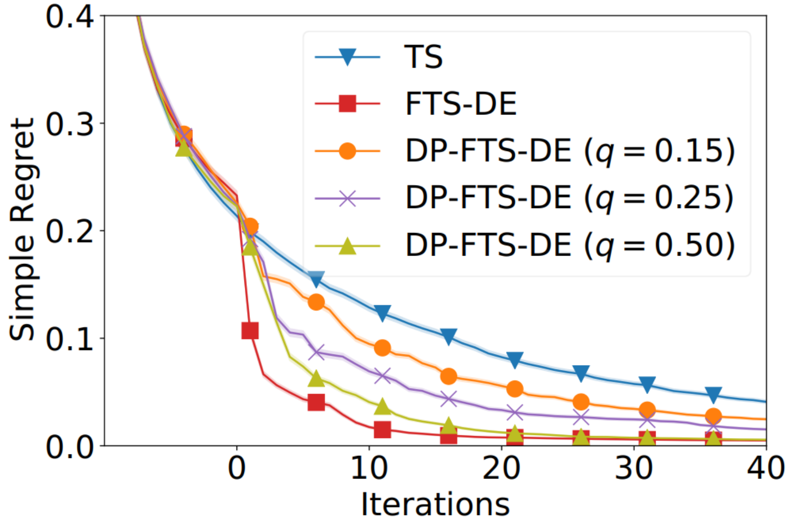


- Impact of P **(number of sub-regions in DE)** on FTS
  - Larger P improves the performance

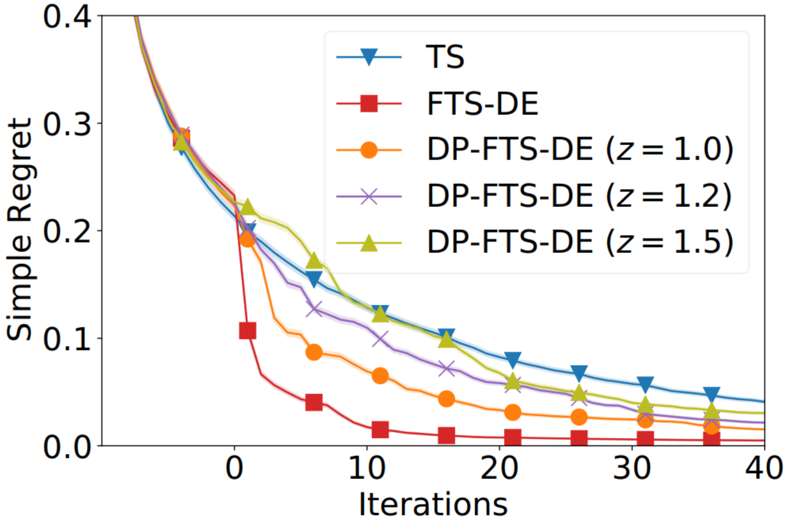
- Impact of S **(the clipping threshold)**
  - Overly small S -> more vectors clipped
  - Overly large S -> more noises added

Differentially Private Federated Bayesian Optimization

Synthetic Experiments



Privacy losses (top to bottom):
5.93, 9.91, 20.12



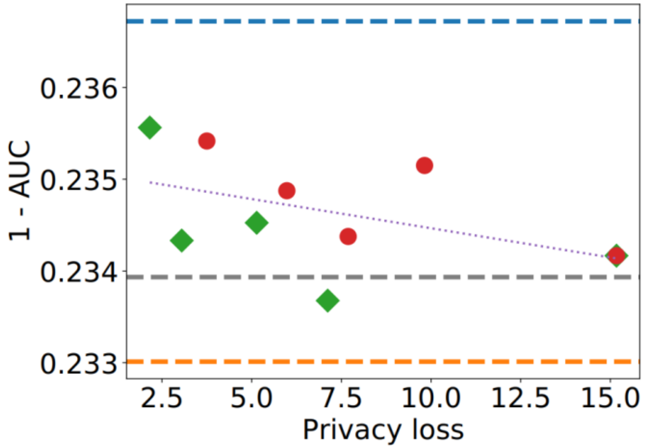
Privacy losses (top to bottom):
9.91, 7.39, 5.22

- Impact of q **(prob. of selecting an agent)**
  - Larger q improves utility & deteriorates privacy
- Impact of z **(prop. to noise variance)**
  - Larger z deteriorates utility & improves privacy

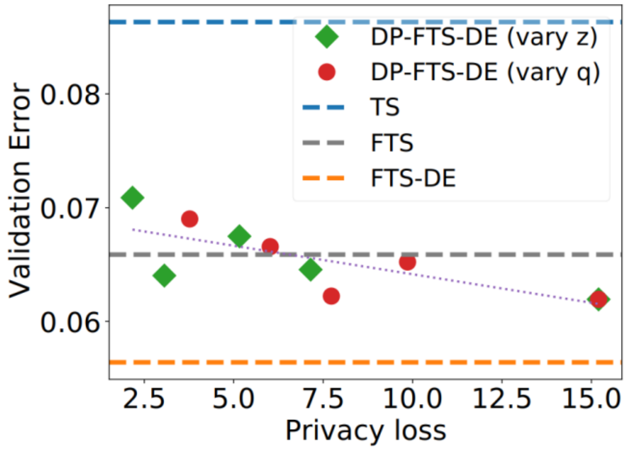
Differentially Private Federated Bayesian Optimization

Real-world Experiments

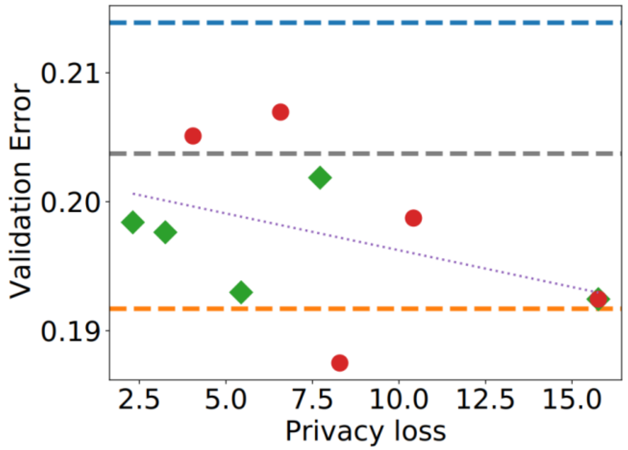
- **Privacy-utility trade-off:**
  - More to the left: better privacy
  - More to the bottom: better utility



Landmine detection (N=29),
hyper tuning for SVM



Activity recognition using mobile phone (N=30),
hyper tuning for logistic regression

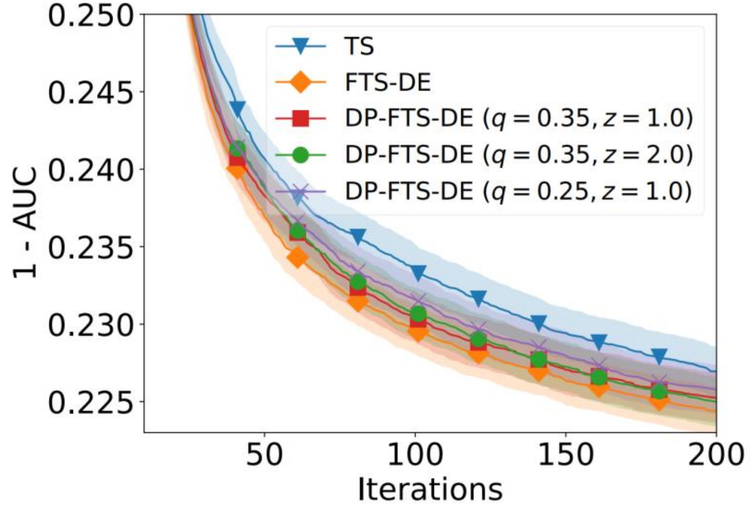


EMNIST (N=50),
hyper tuning for CNN

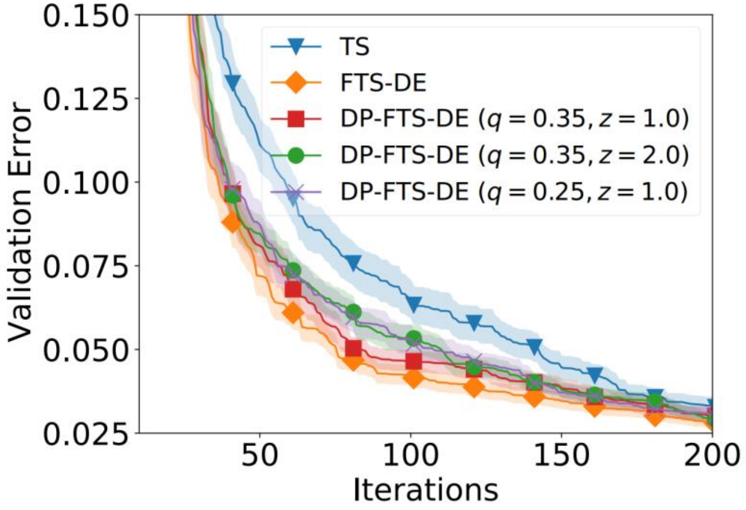
Differentially Private Federated Bayesian Optimization

Real-world Experiments

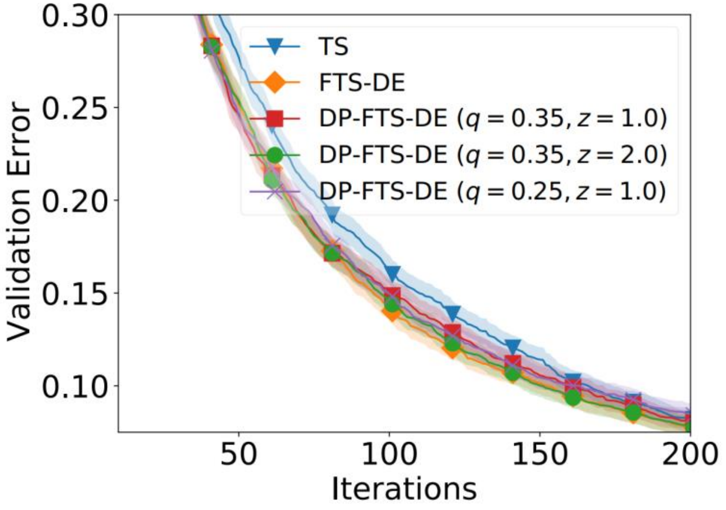
- **Privacy-utility trade-off:**
  - Convergence



Landmine detection



Activity recognition using mobile phone



EMNIST

Thank you!