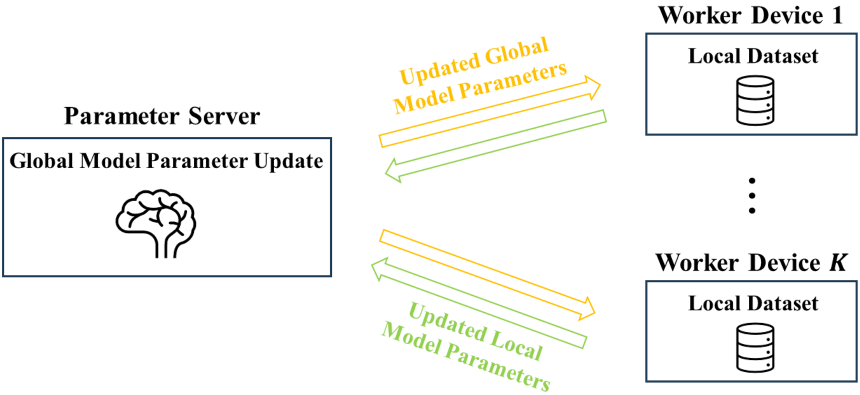
Joint Downlink-Uplink Beamforming for Wireless Multi-Antenna Federated Learning

Chong Zhang*, Min Dong†, Ben Liang*, Ali Afana‡, Yahia Ahmed‡

*Department of Electrical and Computer Engineering, University of Toronto, Canada
†Department of Electrical, Computer and Software Engineering, Ontario Tech University, Canada
‡Ericsson Canada

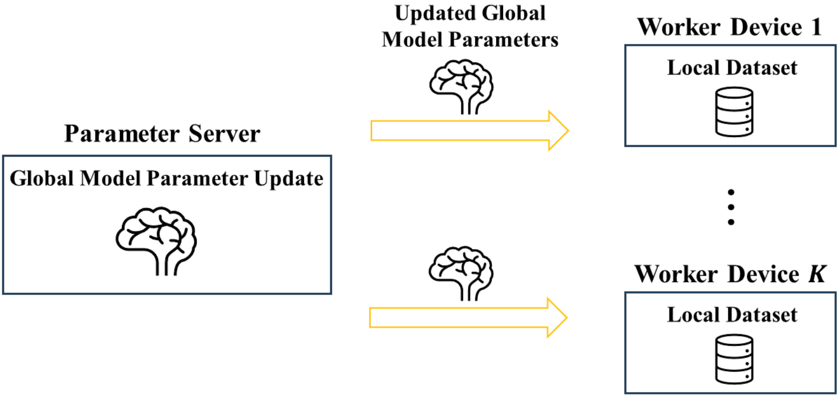
WiOpt 2023

Federated Learning



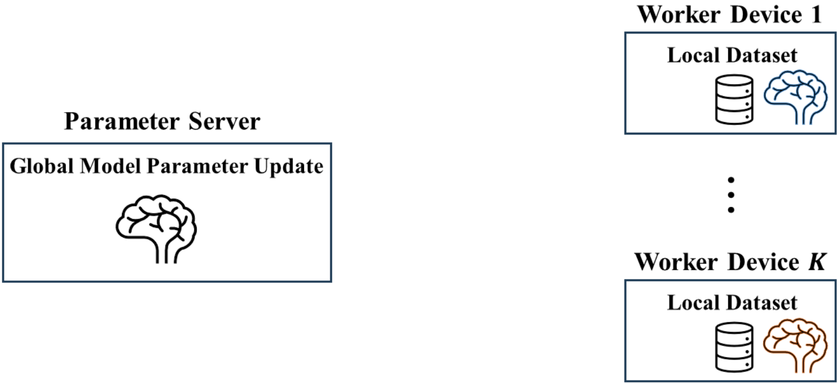
- Federated Learning (FL): collaborative model training using local datasets.
  - Protects data privacy of local worker devices.

Federated Learning Algorithm



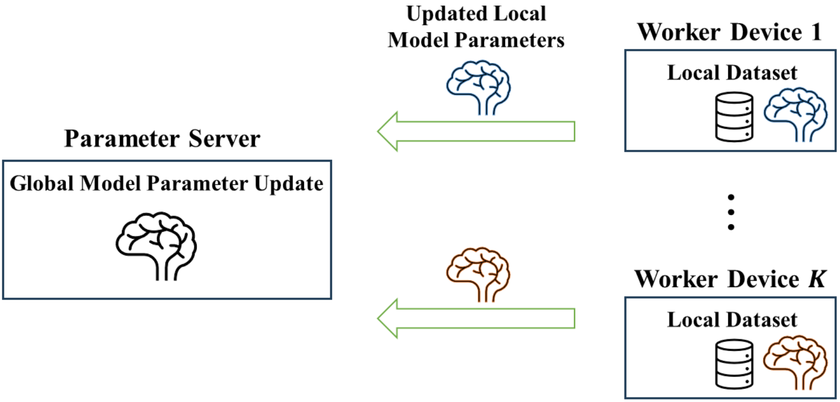
- Step 1: Worker devices download current global model parameters.

Federated Learning Algorithm



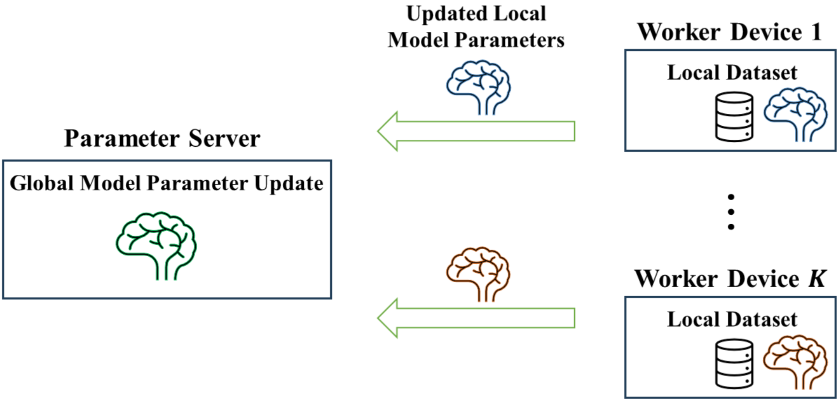
- Step 2: Worker devices generate updated local parameters using local datasets.

Federated Learning Algorithm



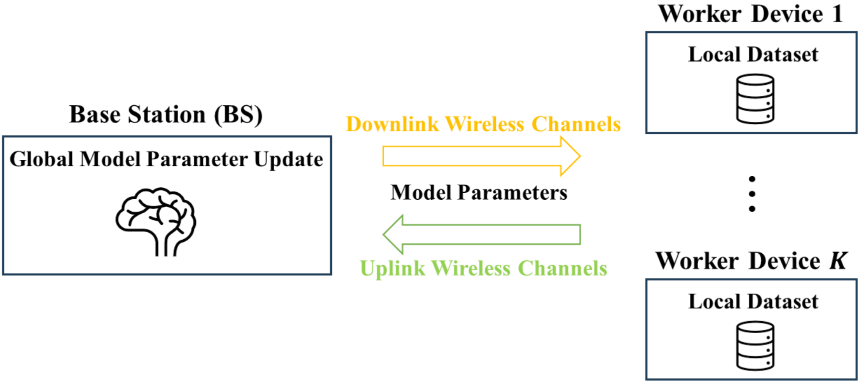
- Step 3: Worker devices upload their locally updated model parameters.

Federated Learning Algorithm



- Step 4: Central server aggregates received local models to update global model parameters.

Wireless Federated Learning



- Parameter server: hosted by base station (BS).
- Model parameter exchange: over downlink/uplink wireless channels.

Existing Works

- Transmission design to improve FL communication efficiency

Existing Works

- Transmission design to improve FL communication efficiency
  - Uplink design assuming error-free downlink (Yang&etal'20, Zhu&etal'20, Zhang&Tao'21).

Existing Works

- Transmission design to improve FL communication efficiency
  - Uplink design assuming error-free downlink (Yang&etal'20, Zhu&etal'20, Zhang&Tao'21).
  - Downlink design assuming error-free uplink (Amiri&etal'22).

Existing Works

- Transmission design to improve FL communication efficiency
  - Uplink design assuming error-free downlink (Yang&etal'20, Zhu&etal'20, Zhang&Tao'21).
  - Downlink design assuming error-free uplink (Amiri&etal'22).
  - Analog schemes outperform digital schemes in both downlink and uplink.

Existing Works

- Transmission design to improve FL communication efficiency
  - Uplink design assuming error-free downlink (Yang&etal'20, Zhu&etal'20, Zhang&Tao'21).
  - Downlink design assuming error-free uplink (Amiri&etal'22).
  - Analog schemes outperform digital schemes in both downlink and uplink.
- Joint Downlink-Uplink Transmission

Existing Works

- Transmission design to improve FL communication efficiency
  - Uplink design assuming error-free downlink (Yang&etal'20, Zhu&etal'20, Zhang&Tao'21).
  - Downlink design assuming error-free uplink (Amiri&etal'22).
  - Analog schemes outperform digital schemes in both downlink and uplink.
- Joint Downlink-Uplink Transmission
  - Training convergence analysis using a generic receiver noise model (Wei&Shen'22).

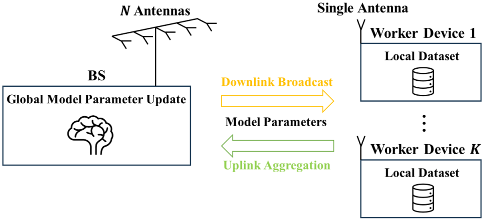
Existing Works

- Transmission design to improve FL communication efficiency
  - Uplink design assuming error-free downlink (Yang&etal'20, Zhu&etal'20, Zhang&Tao'21).
  - Downlink design assuming error-free uplink (Amiri&etal'22).
  - Analog schemes outperform digital schemes in both downlink and uplink.
- Joint Downlink-Uplink Transmission
  - Training convergence analysis using a generic receiver noise model (Wei&Shen'22).
  - Design assuming single-antenna BSs in single-cell (Guo&etal'22) or multi-cell (Wang&etal'22).

Existing Works

- Transmission design to improve FL communication efficiency
  - Uplink design assuming error-free downlink (Yang&etal'20, Zhu&etal'20, Zhang&Tao'21).
  - Downlink design assuming error-free uplink (Amiri&etal'22).
  - Analog schemes outperform digital schemes in both downlink and uplink.
- Joint Downlink-Uplink Transmission
  - Training convergence analysis using a generic receiver noise model (Wei&Shen'22).
  - Design assuming single-antenna BSs in single-cell (Guo&etal'22) or multi-cell (Wang&etal'22).
- Goal of this work: joint downlink-uplink beamforming with a multi-antenna BS to improve wireless FL performance.

FL System Model



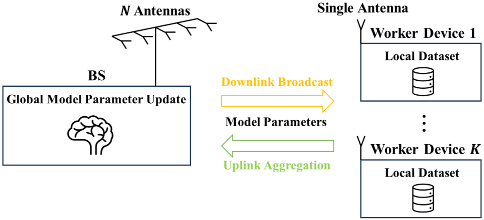
At device k

- **Data sample set $\mathcal{S}_k$: sample $\mathbf{s}_{k,i}$, label $v_{k,i}$**
- **Number of samples: S_k**
- **Training loss function: $L(\cdot)$**
- **Local loss function**

$$F_k(\boldsymbol{\theta}) = \frac{1}{S_k} \sum_{i=1}^{S_k} L(\boldsymbol{\theta};\ \mathbf{s}_{k,i}, v_{k,i})$$

- Model parameter vector: $\boldsymbol{\theta} \in \mathbb{R}^D$
  — D: number of model parameters.

FL System Model



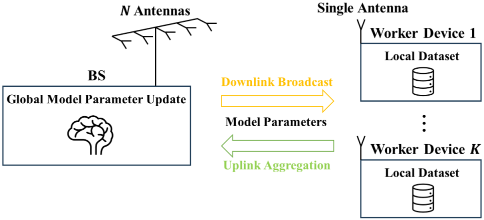
At device k

- **Data sample set $\mathcal{S}_k$: sample $\mathbf{s}_{k,i}$, label $v_{k,i}$**
- **Number of samples: S_k**
- **Training loss function: $L(\cdot)$**
- **Local loss function**

$$F_k(\boldsymbol{\theta}) = \frac{1}{S_k} \sum_{i=1}^{S_k} L(\boldsymbol{\theta};\ \mathbf{s}_{k,i}, v_{k,i})$$

- Model parameter vector: $\boldsymbol{\theta} \in \mathbb{R}^D$
  — D: number of model parameters.
- Global loss function: $F(\boldsymbol{\theta}) = \sum_{k=1}^{K} \frac{S_k}{S} F_k(\boldsymbol{\theta})$.

FL System Model



At device k

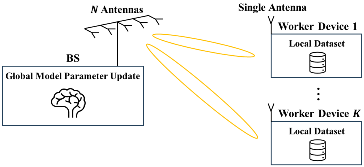
- **Data sample set $\mathcal{S}_k$: sample $\mathbf{s}_{k,i}$, label $v_{k,i}$**
- **Number of samples: S_k**
- **Training loss function: $L(\cdot)$**
- **Local loss function**

$$F_k(\boldsymbol{\theta}) = \frac{1}{S_k} \sum_{i=1}^{S_k} L(\boldsymbol{\theta};\ \mathbf{s}_{k,i}, v_{k,i})$$

- Model parameter vector: $\boldsymbol{\theta} \in \mathbb{R}^D$
  — D: number of model parameters.
- Global loss function: $F(\boldsymbol{\theta}) = \sum_{k=1}^{K} \frac{S_k}{S} F_k(\boldsymbol{\theta})$.
- Goal: find optimal global model $\boldsymbol{\theta}^\star$ that minimizes global loss $F(\boldsymbol{\theta})$.
  - Iteratively update $\boldsymbol{\theta}_t \in \mathbb{R}^D$
    — t: FL round index.

Downlink Broadcast at Communication Round t

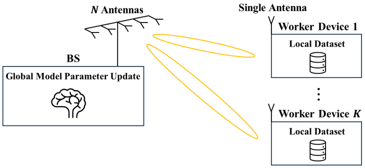
- BS broadcasts current global model to K devices via multicast beamforming.



- For efficient transmission: convert real $\boldsymbol{\theta}_t \in \mathbb{R}^D \Rightarrow$ complex $\tilde{\boldsymbol{\theta}}_t \in \mathbb{C}^{\frac{D}{2}}$.
  - $\boldsymbol{\theta}_t = [(\tilde{\boldsymbol{\theta}}_t^{\text{re}})^T, (\tilde{\boldsymbol{\theta}}_t^{\text{im}})^T]^T \quad \Leftrightarrow \quad \tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t^{\text{re}} + j\tilde{\boldsymbol{\theta}}_t^{\text{im}}$.

Downlink Broadcast at Communication Round t

- BS broadcasts current global model to K devices via multicast beamforming.



- For efficient transmission: convert real $\boldsymbol{\theta}_t \in \mathbb{R}^D \Rightarrow$ complex $\tilde{\boldsymbol{\theta}}_t \in \mathbb{C}^{\frac{D}{2}}$.
  - $\boldsymbol{\theta}_t = [(\tilde{\boldsymbol{\theta}}_t^{\text{re}})^T, (\tilde{\boldsymbol{\theta}}_t^{\text{im}})^T]^T \quad \Leftrightarrow \quad \tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t^{\text{re}} + j\tilde{\boldsymbol{\theta}}_t^{\text{im}}$.
- Channel between BS and device k: $\mathbf{h}_{k,t}$
  — unchanged during round t.
- Downlink multicast beamformer: $\mathbf{w}_t^{\text{dl}}$.

Downlink Broadcast at Communication Round t

- Transmitted complex signal vector at BS: $\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t^{\text{re}} + j\tilde{\boldsymbol{\theta}}_t^{\text{im}} \in \mathbb{C}^{\frac{D}{2}}$.

Downlink Broadcast at Communication Round t

- Transmitted complex signal vector at BS: $\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t^{\text{re}} + j\tilde{\boldsymbol{\theta}}_t^{\text{im}} \in \mathbb{C}^{\frac{D}{2}}$.
- Received signal at device k

$$\mathbf{u}_{k,t} = (\mathbf{w}_t^{\text{dl}})^H \mathbf{h}_{k,t} \tilde{\boldsymbol{\theta}}_t + \underbrace{\mathbf{n}_{k,t}^{\text{dl}}}_{\textit{noise vector}} .$$

Downlink Broadcast at Communication Round t

- Transmitted complex signal vector at BS: $\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t^{\text{re}} + j\tilde{\boldsymbol{\theta}}_t^{\text{im}} \in \mathbb{C}^{\frac{D}{2}}$.
- Received signal at device k

$$\mathbf{u}_{k,t} = (\mathbf{w}_t^{\text{dl}})^H \mathbf{h}_{k,t} \tilde{\boldsymbol{\theta}}_t + \underbrace{\mathbf{n}_{k,t}^{\text{dl}}}_{\textit{noise vector}}.$$

- Post-processed received signal at device k

$$\hat{\tilde{\boldsymbol{\theta}}}_{k,t} = \frac{\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}|^2} \mathbf{u}_{k,t} = \tilde{\boldsymbol{\theta}}_t + \tilde{\mathbf{n}}_{k,t}^{\text{dl}}.$$

where $\tilde{\mathbf{n}}_{k,t}^{\text{dl}} \triangleq \frac{\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}|^2} \mathbf{n}_{k,t}^{\text{dl}}$.

Downlink Broadcast at Communication Round t

- Transmitted complex signal vector at BS: $\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t^{\text{re}} + j\tilde{\boldsymbol{\theta}}_t^{\text{im}} \in \mathbb{C}^{\frac{D}{2}}$.
- Received signal at device k

$$\mathbf{u}_{k,t} = (\mathbf{w}_t^{\text{dl}})^H \mathbf{h}_{k,t} \tilde{\boldsymbol{\theta}}_t + \underbrace{\mathbf{n}_{k,t}^{\text{dl}}}_{\textit{noise vector}} .$$

- Post-processed received signal at device k

$$\hat{\tilde{\boldsymbol{\theta}}}_{k,t} = \frac{\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}|^2} \mathbf{u}_{k,t} = \tilde{\boldsymbol{\theta}}_t + \tilde{\mathbf{n}}_{k,t}^{\text{dl}}.$$

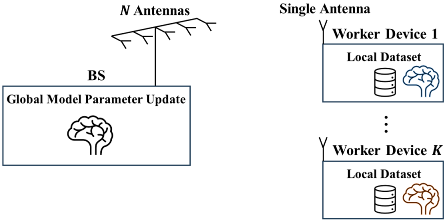
where $\tilde{\mathbf{n}}_{k,t}^{\text{dl}} \triangleq \frac{\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{dl}}|^2} \mathbf{n}_{k,t}^{\text{dl}}$.

- Convert to real-valued estimate of global model $\boldsymbol{\theta}_t$

$$\hat{\boldsymbol{\theta}}_{k,t} = \left[\mathfrak{Re}\{\hat{\tilde{\boldsymbol{\theta}}}_{k,t}\}^T, \mathfrak{Im}\{\hat{\tilde{\boldsymbol{\theta}}}_{k,t}\}^T \right]^T = \boldsymbol{\theta}_t + \hat{\mathbf{n}}_{k,t}^{\text{dl}}.$$

where $\hat{\mathbf{n}}_{k,t}^{\text{dl}} \triangleq [\mathfrak{Re}\{\tilde{\mathbf{n}}_{k,t}^{\text{dl}}\}^T, \mathfrak{Im}\{\tilde{\mathbf{n}}_{k,t}^{\text{dl}}\}^T]^T$.

Local Model Update at Communication Round t



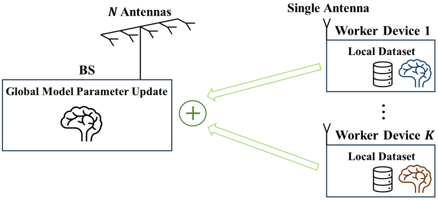
- J mini-batch stochastic gradient descent (SGD) iterations at device k

$$\begin{aligned}\boldsymbol{\theta}_{k,t}^{\tau+1} &= \boldsymbol{\theta}_{k,t}^{\tau} - \eta_t \nabla F_k(\boldsymbol{\theta}_{k,t}^{\tau}; \mathcal{B}_{k,t}^{\tau}) \\ &= \boldsymbol{\theta}_{k,t}^{\tau} - \frac{\eta_t}{|\mathcal{B}_{k,t}^{\tau}|} \sum_{(\mathbf{s},v)\in\mathcal{B}_{k,t}^{\tau}} \nabla L(\boldsymbol{\theta}_{k,t}^{\tau}; \mathbf{s}, v).\end{aligned}$$

  - SGD iteration index: τ.
  - Initial point: $\boldsymbol{\theta}_{k,t}^{0} = \hat{\boldsymbol{\theta}}_{k,t}$.
  - Mini-batch: $\mathcal{B}_{k,t}^{\tau}$.
  - Learning rate: η_t.

Uplink Aggregation at Communication Round t

- Over-the-air aggregation: BS aggregates local models via receive beamforming.



- **Convert real $\boldsymbol{\theta}_{k,t}^{J} \in \mathbb{R}^{D} \rightarrow$ complex $\widetilde{\boldsymbol{\theta}}_{k,t}^{J} \in \mathbb{C}^{\frac{D}{2}}$**
- **At device k: transmit beamforming weight $a_{k,t}$**
  - **Form distributed transmit beamforming among K devices**

- BS receive beamformer: $\mathbf{w}_t^{\text{ul}}$.
- Post-processed received aggregated signal:

$$\mathbf{z}_t = \sum_{k=1}^{K} (\mathbf{w}_t^{\text{ul}})^H \mathbf{h}_{k,t} a_{k,t} \tilde{\theta}_{k,t}^{J} + \underbrace{\mathbf{n}_t^{\text{ul}}}_{\textit{noise vector}} .$$

Over-the-Air Aggregation: Transmit Phase Alignment

- Post-processed received signal: $\mathbf{z}_t = \sum_{k=1}^{K} (\mathbf{w}_t^{\text{ul}})^H \mathbf{h}_{k,t} a_{k,t} \tilde{\theta}_{k,t}^J + \mathbf{n}_t^{\text{ul}}$.

Over-the-Air Aggregation: Transmit Phase Alignment

- Post-processed received signal: $\mathbf{z}_t = \sum_{k=1}^{K} (\mathbf{w}_t^{\text{ul}})^H \mathbf{h}_{k,t} a_{k,t} \tilde{\theta}_{k,t}^{J} + \mathbf{n}_t^{\text{ul}}$.
- At devices: transmit phase alignment for uplink distributed transmit beamforming
  - Transmit weight at device k

$$a_{k,t} = \sqrt{p_{k,t}} \frac{\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}|}. \qquad \Rightarrow \qquad (\mathbf{w}_t^{\text{ul}})^H \mathbf{h}_{k,t} a_{k,t} = \sqrt{p_{k,t}} |\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}|.$$

— $p_{k,t}$: transmit power scaling factor at device k.

Over-the-Air Aggregation: Transmit Phase Alignment

- Post-processed received signal: $\mathbf{z}_t = \sum_{k=1}^{K} (\mathbf{w}_t^{\text{ul}})^H \mathbf{h}_{k,t} a_{k,t} \tilde{\boldsymbol{\theta}}_{k,t}^J + \mathbf{n}_t^{\text{ul}}$.
- At devices: transmit phase alignment for uplink distributed transmit beamforming
  - Transmit weight at device k

$$a_{k,t} = \sqrt{p_{k,t}} \frac{\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}|}. \qquad \Rightarrow \qquad (\mathbf{w}_t^{\text{ul}})^H \mathbf{h}_{k,t} a_{k,t} = \sqrt{p_{k,t}} |\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}|.$$

— $p_{k,t}$: transmit power scaling factor at device k.

- At BS: scale $\mathbf{z}_t$ to obtain complex equivalent global model update

$$\tilde{\boldsymbol{\theta}}_{t+1} = \frac{\mathbf{z}_t}{\sum_{k=1}^{K} \sqrt{p_{k,t}} |\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}|} = \sum_{k=1}^{K} \rho_{k,t} \tilde{\boldsymbol{\theta}}_{k,t}^J + \tilde{\mathbf{n}}_t^{\text{ul}}.$$

— $\rho_{k,t} \triangleq \frac{\sqrt{p_{k,t}} |\mathbf{h}_{k,t}^H \mathbf{w}_t^{\text{ul}}|}{\sum_{j=1}^{K} \sqrt{p_{j,t}} |\mathbf{h}_{j,t}^H \mathbf{w}_t^{\text{ul}}|}$, $\tilde{\mathbf{n}}_t^{\text{ul}} \triangleq \frac{\mathbf{n}_t^{\text{ul}}}{\sum_{j=1}^{K} \sqrt{p_{j,t}} |\mathbf{h}_{j,t}^H \mathbf{w}_t^{\text{ul}}|}$.

Global Model Updating Equation

- Round-trip model update at communication round t

$$\tilde{\boldsymbol{\theta}}_{t+1} = \tilde{\boldsymbol{\theta}}_t + \sum_{k=1}^{K} \rho_{k,t} \Delta \tilde{\boldsymbol{\theta}}_{k,t} + \sum_{k=1}^{K} \rho_{k,t} \tilde{\mathbf{n}}_{k,t}^{\mathrm{dl}} + \tilde{\mathbf{n}}_t^{\mathrm{ul}}.$$

- $\Delta \tilde{\boldsymbol{\theta}}_{k,t}$: Equivalent (complex) local model change at device k.
- $\tilde{\mathbf{n}}_{k,t}^{\mathrm{dl}}$: Post-processed downlink receiver noise at device k.
- $\tilde{\mathbf{n}}_t^{\mathrm{ul}}$: Post-processed uplink receiver noise at BS.

- Recovered real-valued global model update

$$\boldsymbol{\theta}_{t+1} = [\mathfrak{Re}\{\tilde{\boldsymbol{\theta}}_{t+1}\}^T, \mathfrak{Im}\{\tilde{\boldsymbol{\theta}}_{t+1}\}^T]^T.$$

Global Model Updating Equation

- Round-trip model update at communication round t

$$\tilde{\boldsymbol{\theta}}_{t+1} = \tilde{\boldsymbol{\theta}}_t + \sum_{k=1}^{K} \rho_{k,t} \Delta \tilde{\boldsymbol{\theta}}_{k,t} + \sum_{k=1}^{K} \rho_{k,t} \tilde{\mathbf{n}}_{k,t}^{\text{dl}} + \tilde{\mathbf{n}}_t^{\text{ul}}.$$

- $\Delta \tilde{\boldsymbol{\theta}}_{k,t}$: Equivalent complex local model change at device k.
- $\tilde{\mathbf{n}}_{k,t}^{\text{dl}}$: Post-processed downlink receiver noise at device k.
- $\tilde{\mathbf{n}}_t^{\text{ul}}$: Post-processed uplink receiver noise at BS.

- Obtained from round-trip wireless FL procedure
  - Downlink-uplink transmission.
  - Local device model update.
  - Reflects noisy communication and transmitter-receiver processing effect.

Joint Downlink-Uplink Beamforming Design

- Objective: minimize expected global loss function after T rounds through joint downlink-uplink beamforming design.

$$
\begin{aligned}
\mathcal{P}_o : \min_{\{\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t\}_{t\in\mathcal{T}}} \quad & \mathbb{E}[F(\boldsymbol{\theta}_T)] \\
\text{s.t.} \quad & \|\mathbf{w}_t^{\text{dl}}\|^2 \|\boldsymbol{\theta}_t\|^2 \le DP^{\text{dl}}, \quad t \in \mathcal{T}, \quad \text{(DL transmit power constraint)} \\
& p_{k,t} \|\boldsymbol{\theta}_{k,t}^J\|^2 \le DP_k^{\text{ul}}, \quad k \in \mathcal{K}, t \in \mathcal{T}, \quad \text{(UL transmit power constraint)} \\
& \|\mathbf{w}_t^{\text{ul}}\|^2 = 1, \quad t \in \mathcal{T}.
\end{aligned}
$$

- $\mathbf{p}_t \triangleq [p_{1,t}, \ldots, p_{K,t}]^T$.
- $\mathbb{E}[\cdot]$: over receiver noise and mini-batch sampling in local training.
- P^{dl}: maximum downlink transmit power limit.
- P_k^{ul}: maximum uplink transmit power limit of device k.

- DL/UL power constraints: power budgets for sending $\tilde{\boldsymbol{\theta}}_t$ (DL) or $\tilde{\boldsymbol{\theta}}_{k,t}^J$ (UL) in D channel uses.

Joint Downlink-Uplink Beamforming Design

$$
\begin{aligned}
\mathcal{P}_o : \min_{\{\mathbf{w}_t^{\mathrm{dl}}, \mathbf{w}_t^{\mathrm{ul}}, \mathbf{p}_t\}_{t\in\mathcal{T}}} \quad & \mathbb{E}[F(\boldsymbol{\theta}_T)] \\
\text{s.t.} \quad & \|\mathbf{w}_t^{\mathrm{dl}}\|^2 \|\boldsymbol{\theta}_t\|^2 \le DP^{\mathrm{dl}}, \quad t \in \mathcal{T}, \quad \text{(DL transmit power constraint)} \\
& p_{k,t} \|\boldsymbol{\theta}_{k,t}^J\|^2 \le DP_k^{\mathrm{ul}}, \quad k \in \mathcal{K}, t \in \mathcal{T}, \quad \text{(UL transmit power constraint)} \\
& \|\mathbf{w}_t^{\mathrm{ul}}\|^2 = 1, \quad t \in \mathcal{T}.
\end{aligned}
$$

- $\mathbf{p}_t \triangleq [p_{1,t}, \ldots, p_{K,t}]^T$.
- $\mathbb{E}[\cdot]$: over receiver noise and mini-batch sampling in local training.
- P^{dl}: maximum downlink transmit power limit.
- P_k^{ul}: maximum uplink transmit power limit of device k.

- Finite-horizon stochastic optimization problem: challenging to solve.

Joint Downlink-Uplink Beamforming Design

$$\mathcal{P}_o: \min_{\{\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t\}_{t\in\mathcal{T}}} \mathbb{E}[F(\boldsymbol{\theta}_T)]$$

$$\text{s.t.} \quad \|\mathbf{w}_t^{\text{dl}}\|^2\|\boldsymbol{\theta}_t\|^2 \leq DP^{\text{dl}}, \quad t \in \mathcal{T}, \quad \text{(DL transmit power constraint)}$$

$$p_{k,t}\|\boldsymbol{\theta}_{k,t}^J\|^2 \leq DP_k^{\text{ul}}, \quad k \in \mathcal{K}, t \in \mathcal{T}, \quad \text{(UL transmit power constraint)}$$

$$\|\mathbf{w}_t^{\text{ul}}\|^2 = 1, \quad t \in \mathcal{T}.$$

- $\mathbf{p}_t \triangleq [p_{1,t}, \ldots, p_{K,t}]^T$.
- $\mathbb{E}[\cdot]$: over receiver noise and mini-batch sampling in local training.
- P^{dl}: maximum downlink transmit power limit.
- P_k^{ul}: maximum uplink transmit power limit of device k.

- Finite-horizon stochastic optimization problem: challenging to solve.
- Solution: minimize an upper bound on $\mathbb{E}[F(\boldsymbol{\theta}_T)]$.

Upper Bound of Expected Optimality Gap

- Common assumptions for convergence analysis of FL model training
  - Local loss function $F_k(\boldsymbol{\theta})$ is differentiable, L-smooth, and strongly convex.
  - Unbiasedness and bounded gradient variance of mini-batch SGD.
  - Bounded gradient difference between global and weighted average of local loss functions.

Upper Bound of Expected Optimality Gap

- Common assumptions for convergence analysis of FL model training
  - Local loss function $F_k(\boldsymbol{\theta})$ is differentiable, L-smooth, and strongly convex.
  - Unbiasedness and bounded gradient variance of mini-batch SGD.
  - Bounded gradient difference between global and weighted average of local loss functions.
- Bound of expected change of $F(\boldsymbol{\theta}_t)$

$$\mathbb{E}[F(\boldsymbol{\theta}_{t+1}) - F(\boldsymbol{\theta}_t)] \leq \underbrace{\Re\{\mathbb{E}[(\sum_{k=1}^{K} \rho_{k,t} \Delta\tilde{\boldsymbol{\theta}}_{k,t} + \sum_{k=1}^{K} \rho_{k,t} \tilde{\mathbf{n}}_{k,t}^{\text{dl}} + \tilde{\mathbf{n}}_t^{\text{ul}})^H \nabla \tilde{F}(\boldsymbol{\theta}_t)]\}}_{\triangleq A_{1,t}}$$

$$+ \frac{L}{2} \underbrace{\mathbb{E}[\| \sum_{k=1}^{K} \rho_{k,t} \Delta\tilde{\boldsymbol{\theta}}_{k,t} + \sum_{k=1}^{K} \rho_{k,t} \tilde{\mathbf{n}}_{k,t}^{\text{dl}} + \tilde{\mathbf{n}}_t^{\text{ul}} \|^2]}_{\triangleq A_{2,t}}.$$

Upper Bound of Expected Optimality Gap

- Lemma 1: Under Assumptions 1–3, $A_{1,t}$ is upper bounded as

$$A_{1,t} \leq \eta_t J \left(\frac{2}{Q_t} - \frac{5}{2} \right) \mathbb{E}\left[\|\nabla F(\boldsymbol{\theta}_t)\|^2 \right] + \frac{D(1-Q_t)}{4\eta_t J Q_t} \sum_{k=1}^{K} \frac{\rho_{k,t}\sigma_{\mathrm{d}}^2}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\mathrm{dl}}|^2} + \frac{\eta_t J}{2} \left(\frac{\delta+\mu}{Q_t} + \frac{\delta-\mu}{2} \right)$$

— $Q_t \triangleq 1 - 4\eta_t^2 J^2 L^2$ and assume $\eta_t J < \frac{1}{2L}$.

Upper Bound of Expected Optimality Gap

- Lemma 1: Under Assumptions 1–3, $A_{1,t}$ is upper bounded as

$$A_{1,t} \leq \eta_t J \left(\frac{2}{Q_t} - \frac{5}{2} \right) \mathbb{E}\left[\|\nabla F(\boldsymbol{\theta}_t)\|^2 \right] + \frac{D(1-Q_t)}{4\eta_t J Q_t} \sum_{k=1}^{K} \frac{\rho_{k,t}\sigma_{\mathrm{d}}^2}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\mathrm{dl}}|^2} + \frac{\eta_t J}{2} \left(\frac{\delta+\mu}{Q_t} + \frac{\delta-\mu}{2} \right)$$

— $Q_t \triangleq 1 - 4\eta_t^2 J^2 L^2$ and assume $\eta_t J < \frac{1}{2L}$.

- Lemma 2: Under Assumptions 1–3, $A_{2,t}$ is upper bounded as

$$A_{2,t} \leq \frac{2}{L^2} \left(\frac{1-Q_t}{Q_t} \right) \mathbb{E}\left[\|\nabla F(\boldsymbol{\theta}_t)\|^2 \right] + D \left(\frac{1-Q_t}{Q_t} \sum_{k=1}^{K} \frac{\rho_{k,t}}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\mathrm{dl}}|^2} + \sum_{k=1}^{K} \frac{\rho_{k,t}^2 \sigma_{\mathrm{d}}^2}{|\mathbf{h}_{k,t}^H \mathbf{w}_t^{\mathrm{dl}}|^2} \right)$$
$$+ \frac{D\sigma_{\mathrm{u}}^2}{2(\sum_{k=1}^{K} \alpha_{k,t}^{\mathrm{ul}})^2} + \frac{1-Q_t}{2L^2 Q_t} \left(\left(1 - Q_t + \frac{Q_t}{J} \right) \mu + 4\delta \right)$$

where $\sigma_{\mathrm{d}}^2/\sigma_{\mathrm{u}}^2$ is device/BS receiver noise variance.

Upper Bound of Expected Optimality Gap

Proposition 1

For the FL system described above, under the assumptions and for $\frac{1}{10L} \leq \eta_t J < \frac{1}{2L}$, $\forall\, t \in \mathcal{T}$, the expected gap $\mathbb{E}[F(\boldsymbol{\theta}_T)] - F^\star$ after T communication rounds is upper bounded by

$$\mathbb{E}[F(\boldsymbol{\theta}_T)] - F^\star \leq \Gamma \prod_{t=0}^{T-1} G_t + \Lambda + \sum_{t=0}^{T-2} H(\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t) \prod_{s=t+1}^{T-1} G_s + H(\mathbf{w}_{T-1}^{\text{dl}}, \mathbf{w}_{T-1}^{\text{ul}}, \mathbf{p}_{T-1})$$

where $\Gamma \triangleq \mathbb{E}[F(\boldsymbol{\theta}_0)] - F^\star$, $\Lambda \triangleq \sum_{t=0}^{T-2} C_t \big(\prod_{s=t+1}^{T-1} G_s \big) + C_{T-1}$ with

$$G_t \triangleq \frac{1 - Q_t}{4\eta_t J \lambda Q_t} \left(5(1 - Q_t) + 4\sqrt{1 - Q_t} - 1\right) + 1,$$

$$C_t \triangleq \frac{\eta_t J}{2}\left(\frac{\delta + \mu}{Q_t} + \frac{\delta - \mu}{2}\right) + \frac{1 - Q_t}{2L^2 Q_t}\left(\left(1 - Q_t + \frac{Q_t}{J}\right)\mu + 4\delta\right).$$

- $H(\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t)$: function of joint DL-UL beamforming design.

Upper Bound of Expected Optimality Gap

- $H(\mathbf{w}_t^{\mathrm{dl}}, \mathbf{w}_t^{\mathrm{ul}}, \mathbf{p}_t)$ is given by

$$H(\mathbf{w}_t^{\mathrm{dl}}, \mathbf{w}_t^{\mathrm{ul}}, \mathbf{p}_t) \triangleq \frac{LD}{2}\left(\frac{1-Q_t+\sqrt{1-Q_t}}{Q_t}\right)\frac{\sigma_{\mathrm{d}}^2\left(\sum_{k=1}^K \frac{\sqrt{p_{k,t}}|\mathbf{h}_{k,t}^H\mathbf{w}_t^{\mathrm{ul}}|}{|\mathbf{h}_{k,t}^H\mathbf{w}_t^{\mathrm{dl}}|^2}\right)}{\sum_{k=1}^K \sqrt{p_{k,t}}|\mathbf{h}_{k,t}^H\mathbf{w}_t^{\mathrm{ul}}|} + \frac{LD}{2}\frac{\sigma_{\mathrm{d}}^2\left(\sum_{k=1}^K \frac{p_{k,t}|\mathbf{h}_{k,t}^H\mathbf{w}_t^{\mathrm{ul}}|^2}{|\mathbf{h}_{k,t}^H\mathbf{w}_t^{\mathrm{dl}}|^2}\right)+\frac{\sigma_{\mathrm{u}}^2}{2}}{\left(\sum_{k=1}^K \sqrt{p_{k,t}}|\mathbf{h}_{k,t}^H\mathbf{w}_t^{\mathrm{ul}}|\right)^2}$$

- A weighted sum of the inverse of two types of SNRs.
  - Post-processing SNR at BS receiver due to downlink noise.
  - Post-processing SNR at BS receiver due to uplink noise.

Joint Downlink-Uplink Beamforming Design

$$
\begin{aligned}
\mathcal{P}_o : \min_{\{\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t\}_{t \in \mathcal{T}}} \quad & \mathbb{E}[F(\boldsymbol{\theta}_T)] \\
\text{s.t.} \quad & \|\mathbf{w}_t^{\text{dl}}\|^2 \|\boldsymbol{\theta}_t\|^2 \le D P^{\text{dl}}, \quad t \in \mathcal{T}, \quad \text{(DL transmit power constraint)} \\
& p_{k,t} \|\boldsymbol{\theta}_{k,t}^J\|^2 \le D P_k^{\text{ul}}, \quad k \in \mathcal{K}, t \in \mathcal{T}, \quad \text{(UL transmit power constraint)} \\
& \|\mathbf{w}_t^{\text{ul}}\|^2 = 1, \quad t \in \mathcal{T}.
\end{aligned}
$$

Joint Downlink-Uplink Beamforming Design

$$
\begin{aligned}
\mathcal{P}_o: \min_{\{\mathbf{w}_t^{\text{dl}},\mathbf{w}_t^{\text{ul}},\mathbf{p}_t\}_{t\in\mathcal{T}}} \quad & \mathbb{E}[F(\boldsymbol{\theta}_T)] \\
\text{s.t.} \quad & \|\mathbf{w}_t^{\text{dl}}\|^2\|\boldsymbol{\theta}_t\|^2 \le DP^{\text{dl}}, \quad t\in\mathcal{T}, \quad \text{(DL transmit power constraint)} \\
& p_{k,t}\|\boldsymbol{\theta}_{k,t}^J\|^2 \le DP_k^{\text{ul}}, \quad k\in\mathcal{K}, t\in\mathcal{T}, \quad \text{(UL transmit power constraint)} \\
& \|\mathbf{w}_t^{\text{ul}}\|^2 = 1, \quad t\in\mathcal{T}.
\end{aligned}
$$

- Minimizing the upper bound of optimality gap $\mathbb{E}[F(\boldsymbol{\theta}_T)] - F^\star$, subject to transmit power constraints:

$$
\begin{aligned}
\mathcal{P}_1: \min_{\{\mathbf{w}_t^{\text{dl}},\mathbf{w}_t^{\text{ul}},\mathbf{p}_t\}_{t\in\mathcal{T}}} \quad & \sum_{t=0}^{T-2} H(\mathbf{w}_t^{\text{dl}},\mathbf{w}_t^{\text{ul}},\mathbf{p}_t)\prod_{s=t+1}^{T-1} G_s + H(\mathbf{w}_{T-1}^{\text{dl}},\mathbf{w}_{T-1}^{\text{ul}},\mathbf{p}_{T-1}) \\
\text{s.t.} \quad & \|\mathbf{w}_t^{\text{dl}}\|^2\|\boldsymbol{\theta}_t\|^2 \le DP^{\text{dl}}, \quad t\in\mathcal{T}, \\
& p_{k,t}\|\boldsymbol{\theta}_{k,t}^J\|^2 \le DP_k^{\text{ul}}, \quad k\in\mathcal{K}, t\in\mathcal{T}, \\
& \|\mathbf{w}_t^{\text{ul}}\|^2 = 1, \quad t\in\mathcal{T}.
\end{aligned}
$$

Joint Downlink-Uplink Beamforming Design

- Minimizing the upper bound of optimality gap $\mathbb{E}[F(\boldsymbol{\theta}_T)] - F^\star$, subject to transmit power constraints:

$$
\begin{aligned}
\mathcal{P}_1: \min_{\{\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t\}_{t\in\mathcal{T}}} \quad & \sum_{t=0}^{T-2} H(\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t) \prod_{s=t+1}^{T-1} G_s + H(\mathbf{w}_{T-1}^{\text{dl}}, \mathbf{w}_{T-1}^{\text{ul}}, \mathbf{p}_{T-1}) \\
\text{s.t.} \quad & \|\mathbf{w}_t^{\text{dl}}\|^2 \|\boldsymbol{\theta}_t\|^2 \le DP^{\text{dl}}, \quad t \in \mathcal{T}, \\
& p_{k,t} \|\boldsymbol{\theta}_{k,t}^J\|^2 \le DP_k^{\text{ul}}, \quad k \in \mathcal{K}, t \in \mathcal{T}, \\
& \|\mathbf{w}_t^{\text{ul}}\|^2 = 1, \quad t \in \mathcal{T}.
\end{aligned}
$$

Joint Downlink-Uplink Beamforming Design

- Minimizing the upper bound of optimality gap $\mathbb{E}[F(\boldsymbol{\theta}_T)] - F^\star$, subject to transmit power constraints:

$$\begin{aligned} \mathcal{P}_1: \min_{\{\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t\}_{t\in\mathcal{T}}} \quad & \sum_{t=0}^{T-2} H(\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t) \prod_{s=t+1}^{T-1} G_s + H(\mathbf{w}_{T-1}^{\text{dl}}, \mathbf{w}_{T-1}^{\text{ul}}, \mathbf{p}_{T-1}) \\ \text{s.t.} \quad & \|\mathbf{w}_t^{\text{dl}}\|^2 \|\boldsymbol{\theta}_t\|^2 \leq DP^{\text{dl}}, \quad t \in \mathcal{T}, \\ & p_{k,t} \|\boldsymbol{\theta}_{k,t}^J\|^2 \leq DP_k^{\text{ul}}, \quad k \in \mathcal{K}, t \in \mathcal{T}, \\ & \|\mathbf{w}_t^{\text{ul}}\|^2 = 1, \quad t \in \mathcal{T}. \end{aligned}$$

- T-horizon joint optimization.
- By Proposition 1, $G_t > 0, \forall\, t \in \mathcal{T}$.
  - $\mathcal{P}_1$ can be decomposed into T subproblems, one for each communication round t.

Per-Round Beamforming Optimization

- Joint downlink-uplink beamforming optimization at round t:

$$\begin{aligned}
\mathcal{P}_2^t: \min_{\mathbf{w}_t^{\mathrm{dl}}, \mathbf{w}_t^{\mathrm{ul}}, \mathbf{p}_t} \; & H(\mathbf{w}_t^{\mathrm{dl}}, \mathbf{w}_t^{\mathrm{ul}}, \mathbf{p}_t) \quad \text{(weighted sum of the inverse of SNRs)} \\
\text{s.t.} \; & \|\mathbf{w}_t^{\mathrm{dl}}\|^2 \|\boldsymbol{\theta}_t\|^2 \leq DP^{\mathrm{dl}}, \\
& p_{k,t} \|\boldsymbol{\theta}_{k,t}^J\|^2 \leq DP_k^{\mathrm{ul}}, \; k \in \mathcal{K}, \\
& \|\mathbf{w}_t^{\mathrm{ul}}\|^2 = 1.
\end{aligned}$$

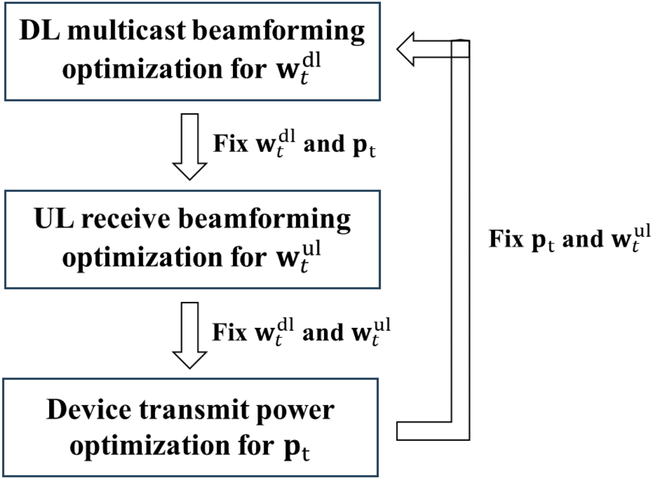
Per-Round Beamforming Optimization

- Joint downlink-uplink beamforming optimization at round t:

$$\begin{aligned}
\mathcal{P}_2^t : \min_{\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t} \quad & H(\mathbf{w}_t^{\text{dl}}, \mathbf{w}_t^{\text{ul}}, \mathbf{p}_t) \quad \text{(weighted sum of the inverse of SNRs)} \\
\text{s.t.} \quad & \|\mathbf{w}_t^{\text{dl}}\|^2 \|\boldsymbol{\theta}_t\|^2 \le DP^{\text{dl}}, \\
& p_{k,t} \|\boldsymbol{\theta}_{k,t}^J\|^2 \le DP_k^{\text{ul}}, \; k \in \mathcal{K}, \\
& \|\mathbf{w}_t^{\text{ul}}\|^2 = 1.
\end{aligned}$$

- Proposed Algorithm: alternating optimization (AO) approach.
  - DL multicast beamforming subproblem.
  - UL beamforming and power optimization subproblem.
  - Each subproblem is solved by projected gradient descent (PGD).

Proposed Algorithm for Per-Round Optimization $\mathcal{P}_2^t$



Simulation Settings

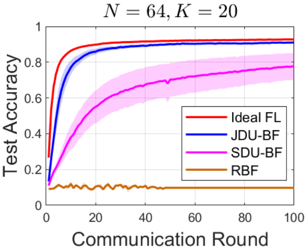
- Typical LTE wireless system settings
  - Bandwidth: 10 MHz.
  - Max BS transmit power: 47 dBm.
  - Max device transmit power 23 dBm.
  - Randomly located devices with pathloss channel.
- Image classification using a CNN based on MNIST dataset.

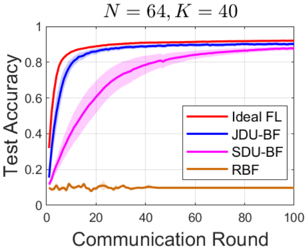
No. parameters	No. training samples at each device	Batch size	Learning rate
1.361×10^4	$\frac{6\times10^4}{K}$	$\frac{2\times10^3}{K}$	$\frac{1}{10JL}$

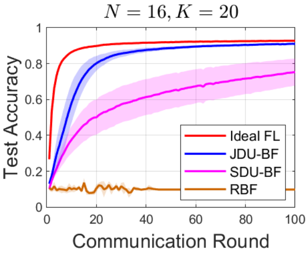
Simulation Settings

- Proposed method:
  - **JDU-BF**: joint DL-UL beamforming design by minimizing the upper bound on optimality gap after T rounds.
- Benchmark comparison methods:
  - **Ideal FL**: perform FL, assuming error-free DL/UL and perfect recovery of model parameters at the receivers.
  - **SDU-BF**: separate DL/UL beamforming design by maximizing received SNR at the receiver of each link.
  - **RBF**: perform FL with random DL/UL beamforming.

Test Accuracy vs. Communication Round T







- JDU-BF outperforms SDUBF and nearly attains ideal FL performance.

Conclusions

- Wireless multi-antenna FL

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.
  - Derive impact of imperfect communication/processing on the global model update.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.
  - Derive impact of imperfect communication/processing on the global model update.
- Joint DL-UL beamforming design

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.
  - Derive impact of imperfect communication/processing on the global model update.
- Joint DL-UL beamforming design
  - Minimizing expected global loss after T rounds.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.
  - Derive impact of imperfect communication/processing on the global model update.
- Joint DL-UL beamforming design
  - Minimizing expected global loss after T rounds.
  - Obtain an upper bound on the expected optimality gap.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.
  - Derive impact of imperfect communication/processing on the global model update.
- Joint DL-UL beamforming design
  - Minimizing expected global loss after T rounds.
  - Obtain an upper bound on the expected optimality gap.
  - AO-based fast first-order algorithm to solve joint optimization problem.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.
  - Derive impact of imperfect communication/processing on the global model update.
- Joint DL-UL beamforming design
  - Minimizing expected global loss after T rounds.
  - Obtain an upper bound on the expected optimality gap.
  - AO-based fast first-order algorithm to solve joint optimization problem.
  - Substantially outperforms the separate-link design approach.

Conclusions

- Wireless multi-antenna FL
  - Noisy DL/UL wireless channels.
  - Transmit/receive beamforming for model parameter transmission/reception.
  - Obtain round-trip closed-form global model update under the noisy channel and transmitter/receiver processing.
  - Derive impact of imperfect communication/processing on the global model update.
- Joint DL-UL beamforming design
  - Minimizing expected global loss after T rounds.
  - Obtain an upper bound on the expected optimality gap.
  - AO-based fast first-order algorithm to solve joint optimization problem.
  - Substantially outperforms the separate-link design approach.
  - Provides near-optimal learning performance for wireless FL.