ROYAL INSTITUTE
OF TECHNOLOGY

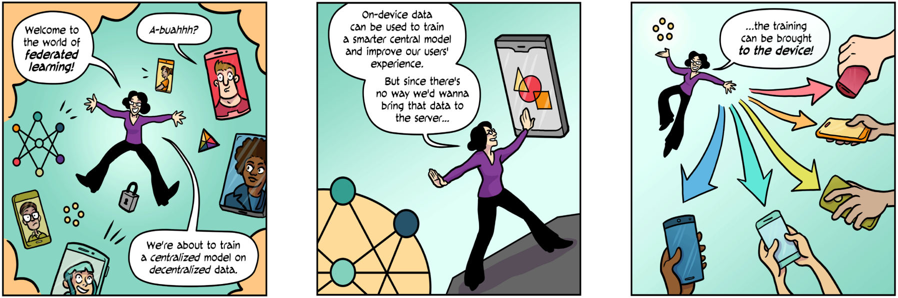
Simultaneous Wireless Information and Power Transfer for Federated Learning

José Mairton B. da Silva Jr[⋆], Konstantinos Ntougias[‡], Ioannis Krikidis[‡], Gábor Fodor[⋆†], Carlo Fischione[⋆]

IEEE SPAWC 2021 - SS3 Wireless for Machine Learning
27-30 September, 2021

[⋆]KTH Royal Institute of Technology, Stockholm, Sweden
[‡]University of Cyprus, Nicosia, Cyprus
[†]Ericsson Research, Stockholm, Sweden
https://people.kth.se/~jmbdsj/index.html
jmbdsj@kth.se

Federated Learning meets Wireless Internet of Things



Source: Federated Learning – an online comic from Google AI

Attempt to overcome learning challenges

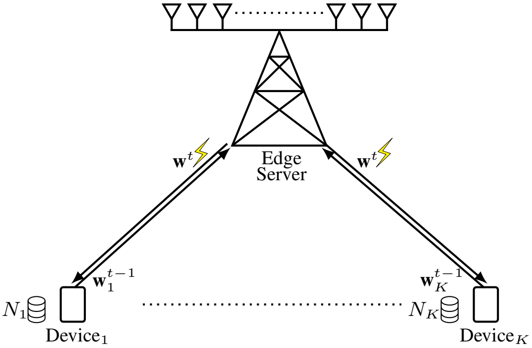
- Communication efficiency, heterogeneous data and devices, privacy
- What about the wireless IoT challenges, especially the energy consumption and latency?

Outline

1. Overview & Main Contributions
2. System Model
3. Minimization of Communication Rounds and Round Time
4. Numerical Results and Discussions
5. Concluding Remarks

Outline

Federated Learning meets Energy Harvesting



Harvesting for Learning

- Use simultaneous wireless information and power transfer for learning tasks
- Harvest energy in the downlink while receiving model updates
- Harvest energy from RF signals to enable federated learning

Challenges and Research Gap

Challenges

- Number of local iterations
- Number of global communication rounds
- Transmit power at devices and beamforming at the edge server
- Local training without depleting devices battery

Research Gap

- Lack of time- and energy-efficient resource allocation in federated learning over wireless methods [Zeng21]

[Zeng21] Q. Zeng et al., "Wirelessly Powered Federated Edge Learning: Optimal Tradeoffs Between Convergence and Power Transfer," arXiv, 2021.

Research Questions and Contributions

Research Questions

- Q1: What is the learning impact on IoT scenarios?
- Q2: How much of the energy used can we compensate?
- Q3: What is trade-off between the number of communication rounds and latency per round?

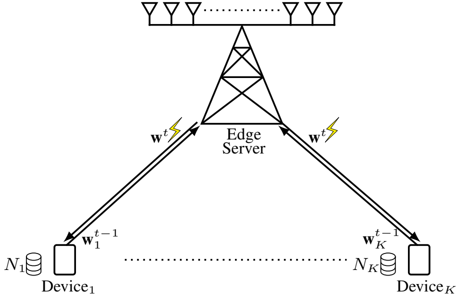
Contributions

- Joint minimization of the latency and communication rounds
  - Convex optimization problem with learning, time, and energy objective function and constraints
- A1: $82\,\%$ vs $69\,\%$ accuracy with MRT and ZF compared to a learning-centric system
- A2: $100\,\%$ with MRT and ZF
- A3: MRT has much lower latency than ZF while showing similar accuracy

Outline

1. Overview & Main Contributions
2. System Model
3. Minimization of Communication Rounds and Round Time
4. Numerical Results and Discussions
5. Concluding Remarks

Communication Model (1/2)



- M antennas at edge server serving K single-antenna devices
- Fixed uplink power p_k and beamforming $\mathbf{v}_k$ (MRT or ZF)
- UL and DL received signals

$$\overline{\mathbf{y}}^u = \underbrace{\mathbf{h}_k \sqrt{p_k} s_k^u}_{\text{Interest signal}} + \underbrace{\sum_{j \neq k} \mathbf{h}_j \sqrt{p_j} s_j^u}_{\text{Interf. signal}} + \underbrace{\boldsymbol{\eta}^u}_{\text{Noise}},$$

$$\overline{y}_k^d = \underbrace{\mathbf{h}_k^{\mathrm{H}} \mathbf{v}_k s_k^d}_{\text{Interest signal}} + \underbrace{\sum_{j \neq k} \mathbf{h}_k^{\mathrm{H}} \mathbf{v}_j s_j^d}_{\text{Interf. signal}} + \underbrace{\eta^d}_{\text{Noise}}.$$

Communication Model (2/2)

- The total received power

$$P_k^r = \sum_{j=1}^{K} \left|\mathbf{h}_k^{\mathrm{H}} \mathbf{v}_j\right|^2 + \sigma^2.$$

- The power splitting constant $\delta_k \in (0,1) \rightarrow \; \delta_k P_k^r$ to data decoding and $(1-\delta_k)P_k^r$ to energy harvesting
- The uplink and downlink rates

$$R_k^u = B_c \log_2\left(1 + \frac{p_k \left|\mathbf{u}_k^{\mathrm{H}} \mathbf{h}_k\right|^2}{\sum_{j\neq k} p_j \left|\mathbf{u}_k^{\mathrm{H}} \mathbf{h}_j\right|^2 + \sigma^2}\right),$$

$$R_k^d = B_c \log_2\left(1 + \frac{\delta_k \left|\mathbf{h}_k^{\mathrm{H}} \mathbf{v}_k\right|^2}{\delta_k \left(\sum_{j\neq k} \left|\mathbf{h}_k^{\mathrm{H}} \mathbf{v}_j\right|^2 + \sigma^2\right) + \sigma_c^2}\right).$$

Time and Energy Models (1/2)

- Time to transmit the model $\rightarrow t_k^u$
- Time to receive the model $\rightarrow t_k^d$
- Time to compute the model [Yang21]

$$t_k^c = \underbrace{C_k}_{\text{CPU cycles/bit}} \times \underbrace{A_k}_{\text{dataset size in bits}} \times \underbrace{I_k}_{\text{\#local iter.}} \; / \; \underbrace{f_k}_{\text{CPU freq.}}$$

- Total time for one communication round (latency per round)

$$t^r = \max_k \left(t_k^u + t_k^c\right) + \max_k \left(t_k^d\right)$$

- Uplink time constraint $\rightarrow \; t_k^u R_k^u \geq D_k$
- Downlink time constraint $\rightarrow \; t_k^d R_k^d \geq D_k$

[Yang21] Z. Yang et al., "Energy Efficient Federated Learning Over Wireless Communication Networks," IEEE TWC, March 2021.

Time and Energy Models (2/2)

- The energy to compute the model [Yang21]

$$E_k^c = \underbrace{\kappa}_{\text{capacitance}} \times C_k A_k I_k f_k^2.$$

- Energy to transmit the model $\rightarrow$ $E_k^t = t_k^u p_k$
- Harvested power [Xu17]

$$P_k^h = \alpha_1 \Big((1-\delta_k) P_k^r \Big)^2 + \alpha_2 \Big((1-\delta_k) P_k^r \Big) + \alpha_3.$$

- Energy harvested at device k $\rightarrow$ $E_k^h = t_k^d P_k^h$
- Energy harvesting constraint $E_k^h \geq \zeta(E_k^t + E_k^c)$, with $\zeta \in (0, 1]$

[Yang21] Z. Yang et al., "Energy Efficient Federated Learning Over Wireless Communication Networks," IEEE TWC, March 2021.

[Xu17] X. Xu et al., "Simultaneous Information and Power Transfer under a Non-Linear RF Energy Harvesting Model," IEEE ICC, 2017.

Learning Model (1/3)

- FedProx method [Li20]
  - Convergence guarantees for heterogeneous devices and non-convex learning objectives
- Solve local surrogate objective function

$$\underset{\mathbf{w}_k}{\text{minimize}}\ h_k(\mathbf{w}_k; \mathbf{w}^t) = F_k(\mathbf{w}_k) + \frac{\mu}{2} \left\| \mathbf{w}_k - \mathbf{w}^t \right\|^2 .$$

- Solve inexactly local problem with $\gamma_k \in [0, 1]$ inexactness

$$\left\| \nabla h_k(\mathbf{w}_k^{(n)}; \mathbf{w}^t) \right\| \leq \gamma_k^t \left\| \nabla h_k(\mathbf{w}^t; \mathbf{w}^t) \right\| .$$

[Li20] T. Li et al., "Federated Optimization in Heterogeneous Networks," PMLR, 2020.

Learning Model (2/3)

Convergence Analysis [Theorem 4 and Corollary 9, Li20]

Assume F_k's are non-convex and L-Lipschitz smooth. Assume B is a global measure of dissimilarity between the gradients of the devices, and suppose that $\mathbf{w}^t$ is not a stationary solution and the local functions F_k are B-dissimilar. If μ, K, and γ_k^t are chosen such that

$$\rho = \left(\frac{(1 - \gamma^t B)}{\mu} - (1 + \gamma^t)(a_1 + a_2(1 + \gamma^t)) \right) > 0,$$

then we have the following expected decrease ρ in the global objective

$$\mathbb{E}_{\mathcal{S}_t}[f(\mathbf{w}^{t+1})] \leq f(\mathbf{w}^t) - \rho \left\| \nabla f(\mathbf{w}^t) \right\|^2,$$

where $\mathcal{S}_t$ is the set of K devices selected, $\gamma^t = \max_{k \in \mathcal{S}_t} \ \gamma_k^t$, and $a_1, \ a_2$ are constants.

Learning Model (3/3)

- For $\rho > 0 \rightarrow \gamma B < 1, \quad B < \sqrt{K}$
- The total number of communication rounds $\rightarrow T = O(\frac{\Delta}{\rho\epsilon})$, where $\Delta = f(\mathbf{w}^1) - f^*$

Number of Local Iterations [Silva21]

Consider that the local problem at device k is solved via gradient descent with step size $\alpha < 2/(L+\mu)$. Consider that the initial iteration for device k is given by $\mathbf{w}_k^0 = \mathbf{w}^t$, and that $\beta = 2/(\alpha\bar{\mu}\,(2 - \alpha(L+\mu)))$. Then, the number of local iterations I_k is lower-bounded by

$$I_k \geq 2\beta \log\left(\frac{L+\mu}{\gamma_k \bar{\mu}}\right).$$

[Silva21] J. M. B. da Silva Jr. et al., "Simultaneous Wireless Information and Power Transfer for Federated Learning," arXiv, 2021

Outline

Problem Formulation

- Minimization of communication rounds and round time (latency round)

$$
\begin{aligned}
\underset{\{t_k^u\},\{t_k^d\},\{\gamma_k\},\gamma}{\text{minimize}} \quad & t^r - \rho && \text{(Objective)} \\
\text{subject to} \quad & t_k^u R_k^u \geq D_k, \forall k, && \text{(min. UL time)} \\
& t_k^d R_k^d \geq D_k, \forall k, && \text{(min. DL time)} \\
& E_k^h \geq \zeta \left(E_k^t + E_k^c \right), \forall k, && \text{(min. harvested energy)} \\
& \gamma_k \leq \frac{1-\xi}{B}, \forall k, && \text{(learning bound)} \\
& \gamma \geq \gamma_k, \forall k, && \text{(max. const.)} \\
& t_k^u, t_k^d, \gamma_k \geq 0, \forall k. && \text{(nonzero)}
\end{aligned}
$$

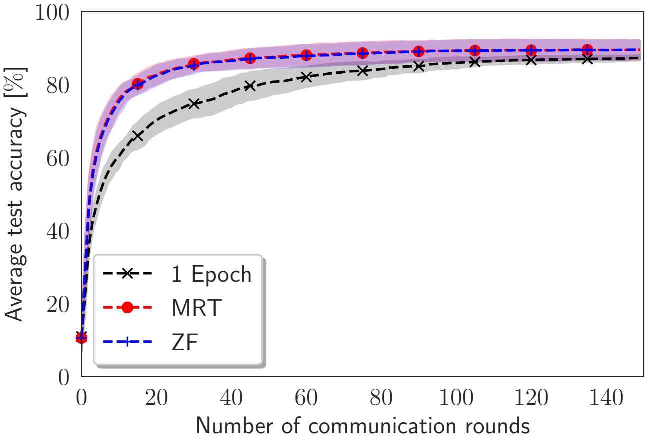
- Convex problem whose solution has computational complexity of $\mathcal{O}(K^4)$

Outline

Simulation Paramters

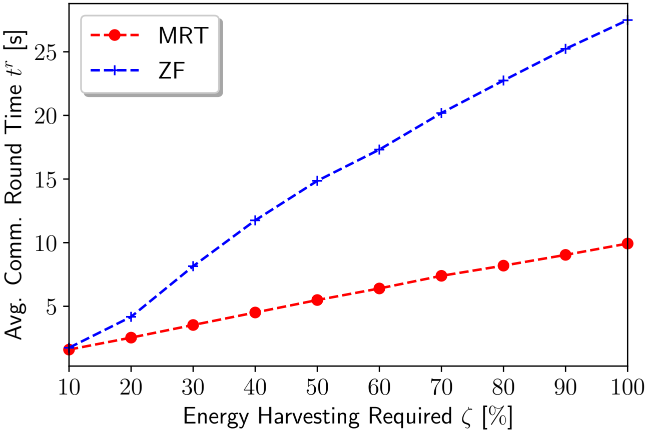
- Small cell with $40\,\mathrm{m}$ radius, with $M = 16$ and $K = 10$
- Image classification using the MNIST dataset with logistic regression
- Pathloss models according to 3GPP urban-micro
- 200 different channel realizations
- Comparisons
  - Test accuracy with $\zeta = 1.0 \rightarrow$ proposed solution using MRT/ZF vs learning-centric FedProx using 1 local iteration (epoch)
  - Trade-off between round time and energy harvesting: proposed solution using MRT and ZF

Test Accuracy vs Number of Communication Rounds



- $82\,\%$ accuracy for MRT and ZF at 20 communication rounds
- $69\,\%$ accuracy for learning-centric FedProx

Communication Round Time vs Energy Harvested



- Energy harvesting constraint impacts heavily ZF
- MRT has the best trade-off on number of communication rounds and round time

Outline

Some takeaways & Future works

Takeaway message

- Federated Learning meets SWIPT
  - Communication, energy, time, and learning models
  - Non-trivial optimization problem to minimize the number of communication rounds and round time
- Harvesting for learning is possible
  - A1: $82\,\%$ accuracy with MRT and ZF
  - A2-A3: $100\,\%$ with MRT requiring small number of communication rounds and round time

Future works

- Optimization of beamformers and splitting parameters
- Scheduling of users to harvest and learn
- More realistic learning tasks on IoT (water monitoring)

ROYAL INSTITUTE
OF TECHNOLOGY

Simultaneous Wireless Information and Power Transfer for Federated Learning

José Mairton B. da Silva Jr[⋆], Konstantinos Ntougias[‡],
Ioannis Krikidis[‡], Gábor Fodor[⋆†], Carlo Fischione[⋆]

IEEE SPAWC 2021 - SS3 Wireless for Machine Learning
27-30 September, 2021

[⋆]KTH Royal Institute of Technology, Stockholm, Sweden
[‡]University of Cyprus, Nicosia, Cyprus
[†]Ericsson Research, Stockholm, Sweden
https://people.kth.se/~jmbdsj/index.html
jmbdsj@kth.se