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

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

1. Overview & Main Contributions

2. System Model

3. Minimization of Communication Rounds and Round Time

4. Numerical Results and Discussions

5. Concluding Remarks
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 the 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
• $M$ antennas at edge server serving $K$ single-antenna devices
• Fixed uplink power $p_k$ and beamforming $v_k$ (MRT or ZF)
• UL and DL received signals

\[
\bar{y}^u = \underbrace{h_k \sqrt{p_k} s^u_k}_{\text{Interest signal}} + \underbrace{\sum_{j \neq k} h_j \sqrt{p_j} s^u_j}_{\text{Interf. signal}} + \underbrace{\eta^u}_{\text{Noise}},
\]

\[
\bar{y}^d_k = \underbrace{h_k^H v_k s^d_k}_{\text{Interest signal}} + \underbrace{\sum_{j \neq k} h_k^H v_j s^d_j}_{\text{Interf. signal}} + \underbrace{\eta^d}_{\text{Noise}}.
\]
The total received power

\[ P_k^r = \sum_{j=1}^{K} |h_k^H v_j|^2 + \sigma^2. \]

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

\[ R_k^u = B_c \log_2 \left( 1 + \frac{p_k |u_k^H h_k|^2}{\sum_{j \neq k} p_j |u_k^H h_j|^2 + \sigma^2} \right), \]

\[ R_k^d = B_c \log_2 \left( 1 + \frac{\delta_k |h_k^H v_k|^2}{\delta_k \left( \sum_{j \neq k} |h_k^H v_j|^2 + \sigma^2 \right) + \sigma_c^2} \right). \]
Time and Energy Models (1/2)

- Time to transmit the model $\rightarrow t^u_k$
- Time to receive the model $\rightarrow t^d_k$
- Time to compute the model [Yang21]

\[
t^c_k = C_k \times A_k \times I_k / f_k
\]

CPU cycles/bit dataset size in bits #local iter. CPU freq.

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

\[
t^r = \max_k (t^u_k + t^c_k) + \max_k (t^d_k)
\]

- Uplink time constraint $\rightarrow t^u_k R^u_k \geq D_k$
- Downlink time constraint $\rightarrow t^d_k R^d_k \geq D_k$

[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]
  \[ E^c_k = \kappa \times C_k A_k I_k f_k^2. \]

- Energy to transmit the model \( \rightarrow E^t_k = t^u_k p_k \)

- Harvested power [Xu17]
  \[ P^h_k = \alpha_1 ((1 - \delta_k)P^r_k)^2 + \alpha_2 ((1 - \delta_k)P^r_k) + \alpha_3. \]

- Energy harvested at device \( k \) \( \rightarrow E^h_k = t^d_k P^h_k \)

- Energy harvesting constraint \( E^h_k \geq \zeta (E^t_k + E^c_k) \), 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.
FedProx method [Li20]

Convergence guarantees for heterogeneous devices and non-convex learning objectives

Solve local surrogate objective function

\[
\minimize_{w_k} h_k(w_k; w^t) = F_k(w_k) + \frac{\mu}{2} \|w_k - w^t\|^2.
\]

Solve inexactly local problem with \(\gamma_k \in [0, 1]\) inexactness

\[
\|\nabla h_k(w_k^n; w^t)\| \leq \gamma_k^t \|\nabla h_k(w^t; w^t)\|.
\]

[Li20] T. Li et al., “Federated Optimization in Heterogeneous Networks,” PMLR, 2020.
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 $w^t$ is not a stationary solution and the local functions $F_k$ are $B$-dissimilar. If $\mu$, $K$, and $\gamma^t_k$ 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 $\rho$ in the global objective

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

where $S_t$ is the set of $K$ devices selected, $\gamma^t = \max_{k \in S_t} \gamma^t_k$, and $a_1, a_2$ are constants.
For $\rho > 0 \rightarrow \gamma B < 1$, $B < \sqrt{K}$

- The total number of communication rounds $\rightarrow T = O\left(\frac{\Delta}{\rho \epsilon}\right)$, where $\Delta = f(w^1) - f^*$

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 $w^0_k = w^t$, 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

1. Overview & Main Contributions

2. System Model

3. Minimization of Communication Rounds and Round Time

4. Numerical Results and Discussions

5. Concluding Remarks
Problem Formulation

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

\[
\text{minimize } \begin{cases} t^r - \rho \\ \{t^u_k, t^d_k, \gamma_k, \gamma\} \end{cases}
\]

subject to

- Min. UL time
\[
t^u_k R^u_k \geq D_k, \forall k,
\]

- Min. DL time
\[
t^d_k R^d_k \geq D_k, \forall k,
\]

- Min. harvested energy
\[
E^h_k \geq \zeta \left( E^t_k + E^c_k \right), \forall k,
\]

- Learning bound
\[
\gamma_k \leq \frac{1 - \xi}{B}, \forall k,
\]

- Max. const.
\[
\gamma \geq \gamma_k, \forall k,
\]

- Nonzero
\[
t^u_k, t^d_k, \gamma_k \geq 0, \forall k.
\]

- Convex problem whose solution has computational complexity of \( O(K^4) \)
Outline

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

- Small cell with 40 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
82% accuracy for MRT and ZF at 20 communication rounds

69% accuracy for learning-centric FedProx
Energy harvesting constraint impacts heavily ZF
MRT has the best trade-off on number of communication rounds and round time
Outline

1. Overview & Main Contributions

2. System Model

3. Minimization of Communication Rounds and Round Time

4. Numerical Results and Discussions

5. Concluding Remarks
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)
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