A Demonstration of Over-the-Air Computation for Federated Edge Learning

Alphan Şahin

December 4, 2022

Email: asahin@mailbox.sc.edu

Overview

- Motivation
- Preliminaries on over-the-air computation (OAC)
- Proposed synchronization method
- Experiment and results
- Final remarks

Motivation

OAC is investigated theoretically for the implementation of federated learning (FL) over a wireless network, i.e., federated edge learning (FEEL) [1,2]

- Only few demonstrations exist in the literature

Motivation of this study: A plausible demonstration of an OAC scheme for FEEL in practice

Motivation

OAC is investigated theoretically for the implementation of federated learning (FL) over a wireless network, i.e., federated edge learning (FEEL) [1,2]

- Only few demonstrations exist in the literature

Motivation of this study: A plausible demonstration of an OAC scheme for FEEL in practice

Motivation

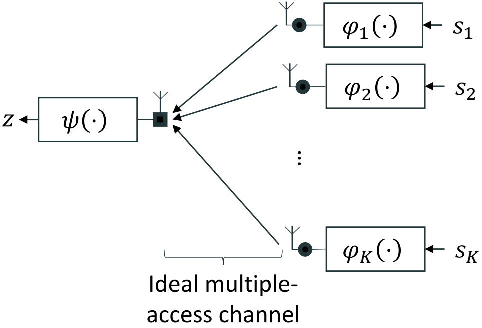
OAC is investigated theoretically for the implementation of federated learning (FL) over a wireless network, i.e., federated edge learning (FEEL) [1, 2]

- Only few demonstrations exist in the literature

Motivation of this study: A plausible demonstration of an OAC scheme for FEEL in practice

What is OAC?

- OAC leverages the signal-superposition property of wireless multiple-access channels to compute a nomographic function [3]



$$f(s_1, s_2, \ldots, s_K) = \psi\left(\sum_{k=1}^{K} \phi_k(s_k)\right)$$

$$f : \mathbb{R}^K \to \mathbb{R}$$

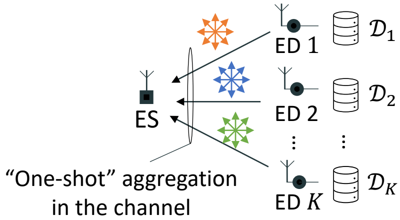
$\phi_k(\cdot)$: Pre-processing function

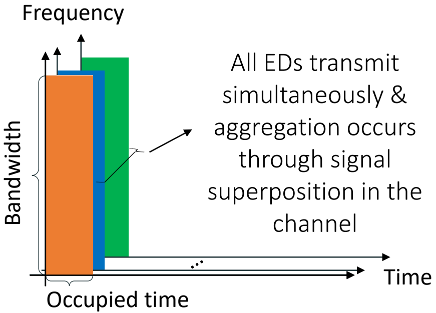
$\psi(\cdot)$: Post-processing function

Several "air-computable" functions: Mean, weighted sum, polynomials, norm

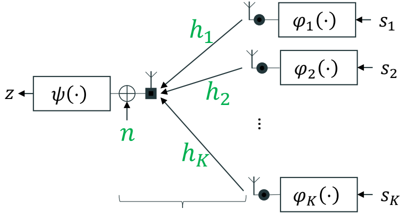
Merit OAC for FEEL

- The communication latency does not scale with the number of edge devices (EDs) if OAC is employed for FEEL
- Edge server (ES) does not need the local information, e.g., local gradients, but the aggregated one, e.g., average of the local gradients





Challenges: Simultaneous Transmissions & Wireless Channel

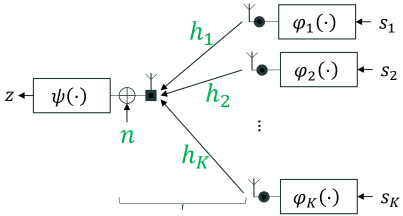


Multi-path channel & noise?

$$z = f(s_1, s_2, \dots, s_K) = \psi\left(\underbrace{\sum_{k=1}^{K} h_k \phi_k(s_k)}_{\text{incoherent additions}} + \underbrace{n}_{\text{noise}}\right)$$

- Challenge 1: Not trivial to achieve a reliable computation in practice
  - Wireless channel **does not allow coherent addition** as in ideal channel
- Challenge 2: EDs need to start their transmissions synchronously
  - Can impose **stringent requirements**, depending on the OAC scheme
  - Hard to implement **without losing the flexibility of software-defined radios (SDRs)**

Challenges: Simultaneous Transmissions & Wireless Channel



Multi-path channel & noise?

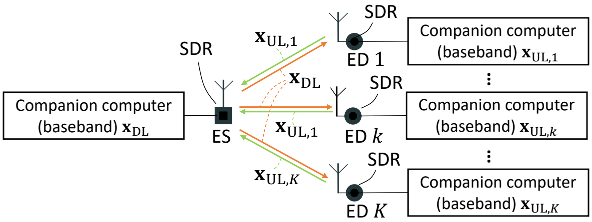
$$z = f(s_1, s_2, \dots, s_K) = \psi\left(\underbrace{\sum_{k=1}^{K} h_k \phi_k(s_k)}_{\text{incoherent additions}} + n\right)$$

noise

- Challenge 1: Not trivial to achieve a reliable computation in practice
  - Wireless channel **does not allow coherent addition** as in ideal channel
- Challenge 2: EDs need to start their transmissions synchronously
  - Can impose **stringent requirements**, depending on the OAC scheme
  - Hard to implement **without losing the flexibility of software-defined radios (SDRs)**

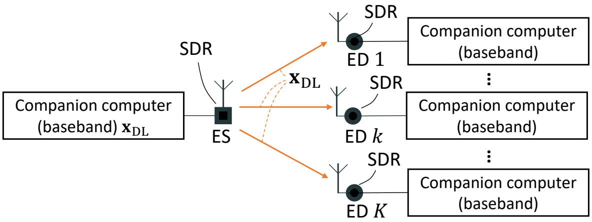
Proposed Synchronization Method

Scenario



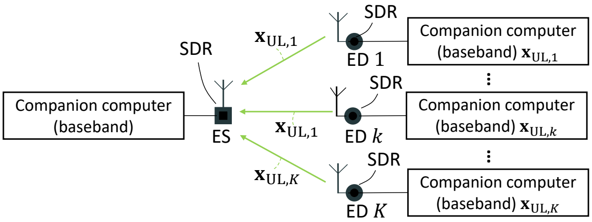
- Scenario: K EDs transmit a set of vectors denoted by $\{\mathbf{x}_{\mathrm{UL},k} \in \mathbb{C}^{1\times N_{\mathrm{UL}}}, \forall k\}$ simultaneously to an ES in response to $\mathbf{x}_{\mathrm{DL}} \in \mathbb{C}^{1\times N_{\mathrm{DL}}}$ transmitted from the ES
- For OAC, we need to ensure the followings:
  - The reception of the vector $\mathbf{x}_{\mathrm{DL}}$ at the companion computer (CC) of each ED
  - The reception of the superposed vector $\sum_{k=1} \mathbf{x}_{\mathrm{UL},k}$ at the ES
- Challenge: Maintaining the baseband processing for OAC at the CCs (i.e., the flexibility of host-based processing) while ensuring synchronous transmissions

Scenario



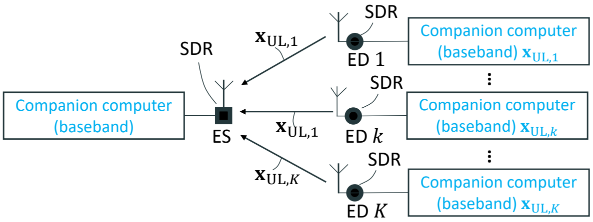
- Scenario: K EDs transmit a set of vectors denoted by $\{\mathbf{x}_{\mathrm{UL},k} \in \mathbb{C}^{1\times N_{\mathrm{UL}}}, \forall k\}$ simultaneously to an ES in response to $\mathbf{x}_{\mathrm{DL}} \in \mathbb{C}^{1\times N_{\mathrm{DL}}}$ transmitted from the ES
- For OAC, we need to ensure the followings:
  - The reception of the vector $\mathbf{x}_{\mathrm{DL}}$ at the companion computer (CC) of each ED
  - The reception of the superposed vector $\sum_{k=1} \mathbf{x}_{\mathrm{UL},k}$ at the ES
- Challenge: Maintaining the baseband processing for OAC at the CCs (i.e., the flexibility of host-based processing) while ensuring synchronous transmissions

Scenario



- Scenario: K EDs transmit a set of vectors denoted by $\{\mathbf{x}_{\mathrm{UL},k} \in \mathbb{C}^{1\times N_{\mathrm{UL}}}, \forall k\}$ simultaneously to an ES in response to $\mathbf{x}_{\mathrm{DL}} \in \mathbb{C}^{1\times N_{\mathrm{DL}}}$ transmitted from the ES
- For OAC, we need to ensure the followings:
  - The reception of the vector $\mathbf{x}_{\mathrm{DL}}$ at the companion computer (CC) of each ED
  - The reception of the superposed vector $\sum_{k=1} \mathbf{x}_{\mathrm{UL},k}$ at the ES
- Challenge: Maintaining the baseband processing for OAC at the CCs (i.e., the flexibility of host-based processing) while ensuring synchronous transmissions

Scenario



- Scenario: K EDs transmit a set of vectors denoted by $\{\mathbf{x}_{\mathrm{UL},k} \in \mathbb{C}^{1\times N_{\mathrm{UL}}}, \forall k\}$ simultaneously to an ES in response to $\mathbf{x}_{\mathrm{DL}} \in \mathbb{C}^{1\times N_{\mathrm{DL}}}$ transmitted from the ES
- For OAC, we need to ensure the followings:
  - The reception of the vector $\mathbf{x}_{\mathrm{DL}}$ at the companion computer (CC) of each ED
  - The reception of the superposed vector $\sum_{k=1} \mathbf{x}_{\mathrm{UL},k}$ at the ES
- Challenge: Maintaining the baseband processing for OAC at the CCs (i.e., the flexibility of host-based processing) while ensuring synchronous transmissions

How to Maintain Baseband Processing in CCs while Ensuring Synchronization?

Key strategy

Separate any signal processing blocks that maintain the synchronization from the ones that do not need to be implemented under strict timing requirements

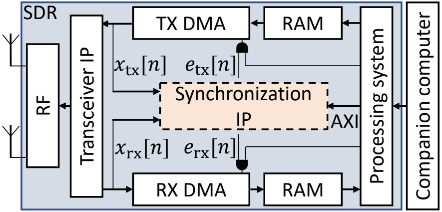
- We propose a synchronization IP that is solely responsible for time synchronization
  - Agnostic to the in-phase/quadrature (IQ) data desired to be communicated
  - Checks if the synch waveform $\mathbf{x}_{\mathrm{SYNC}}$ exists on the IQ samples in TX or RX directions, i.e., $x_{\mathrm{tx}}[n]$ and $x_{\mathrm{rx}}[n]$
  - Controls TX direct-memory access (DMA) and the RX DMA via $e_{\mathrm{tx}}[n]$ and $e_{\mathrm{rx}}[n]$

How to Maintain Baseband Processing in CCs while Ensuring Synchronization?

Key strategy

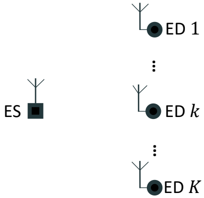
Separate any signal processing blocks that maintain the synchronization from the ones that do not need to be implemented under strict timing requirements

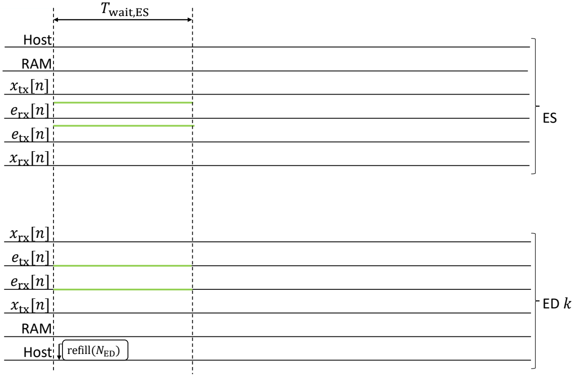
- We propose a synchronization IP that is solely responsible for time synchronization
  - Agnostic to the in-phase/quadrature (IQ) data desired to be communicated
  - Checks if the synch waveform $\mathbf{x}_{\mathrm{SYNC}}$ exists on the IQ samples in TX or RX directions, i.e., $x_{\mathrm{tx}}[n]$ and $x_{\mathrm{rx}}[n]$
  - Controls TX direct-memory access (DMA) and the RX DMA via $e_{\mathrm{tx}}[n]$ and $e_{\mathrm{rx}}[n]$



Behavioral Description of Synchronization IP and Proposed Procedure

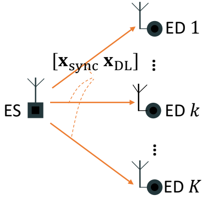
- Step 1/6: The EDs request to receive $N_{\mathrm{ED}} \geq N_{\mathrm{DL}}$ IQ samples
  - The synch. IP @ ED does not allow the SDR to receive since $\mathbf{x}_{\mathrm{SYNC}}$ has not been detected yet





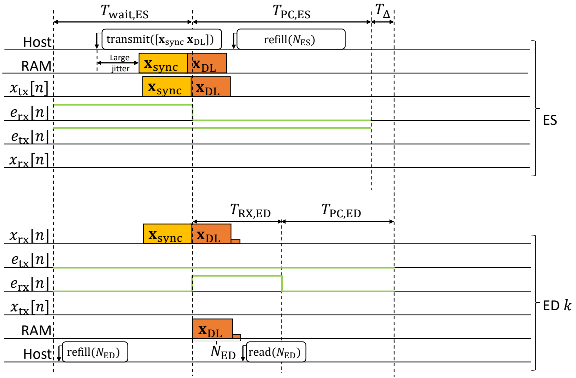
Behavioral Description of Synchronization IP and Proposed Procedure

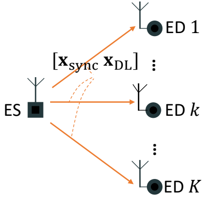
- Step 2/6: The ES transmits $[\mathbf{x}_{\mathrm{SYNC}}\ \mathbf{x}_{\mathrm{DL}}]$ and requests for N_{ES} IQ samples to be acquired
  - The synch. IP @ ES detects $\mathbf{x}_{\mathrm{SYNC}}$ in the TX direction, but it disables RX for $T_{\mathrm{PC,ES}}$ seconds



Behavioral Description of Synchronization IP and Proposed Procedure

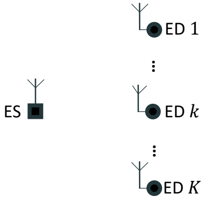
- Step 3/6: EDs detect $\mathbf{x}_{\mathrm{SYNC}}$ in the receive direction
  - The synch. IP @ ED writes N_{ED} samples to the RAM
  - It also starts a counter for $T_{\mathrm{RX,ED}} + T_{\mathrm{PC,ED}}$ seconds to respond to the DL signal

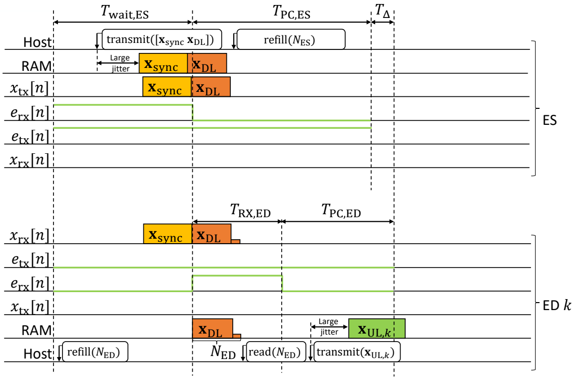




Behavioral Description of Synchronization IP and Proposed Procedure

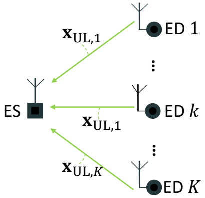
- Step 4/6: The ED processes $\mathbf{x}_{\mathrm{DL}}$ and requests to transmit $\mathbf{x}_{\mathrm{UL},k}$ as a response within $T_{\mathrm{RX,ED}} + T_{\mathrm{PC,ED}}$ seconds

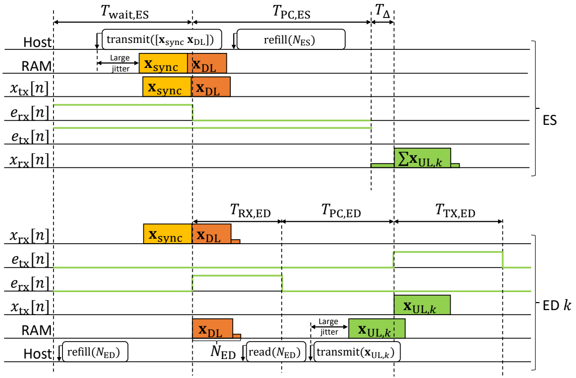




Behavioral Description of Synchronization IP and Proposed Procedure

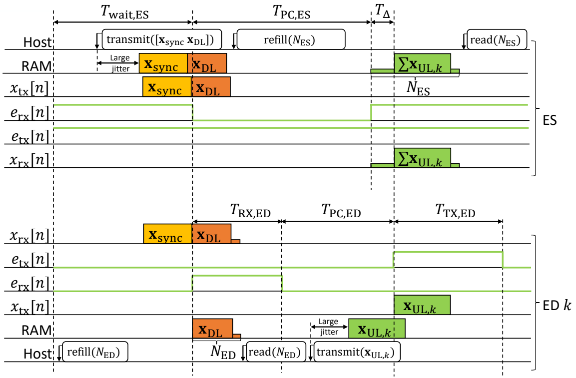
- Step 5/6: The synch. IP @ ED enables the TX-DMA for $T_{\mathrm{TX,ED}}$ seconds
  - At this point, the EDs start their transmissions simultaneously

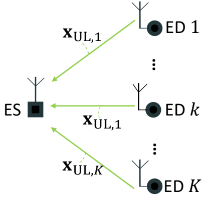




Behavioral Description of Synchronization IP and Proposed Procedure

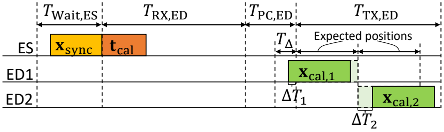
- Step 6/6: Assuming that $T_{\mathrm{PC,ES}}$ is set properly based on $T_{\mathrm{PC,ED}}$, the RX DMA at the ES starts to transfer N_{ES} IQ samples (due to the request in Step 2)
  - At this point, the ES receives the signal $\sum_{k=1} \mathbf{x}_{\mathrm{UL},k}$



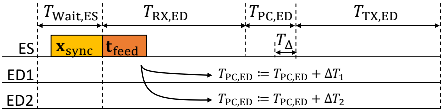


How to Use the Proposed Synchronization Method?

- The proposed synchronization IP along with corresponding procedure is agnostic to the IQ data samples in UL and DL
  - **Calibration**: We utilize it for coarse time offset, carrier frequency offset (CFO), and average signal power calibrations
  - **Computation**: We use it to implement frequency-shift keying (FSK)-based majority vote (MV) (FSK-MV) with absentees [4] with following triggers



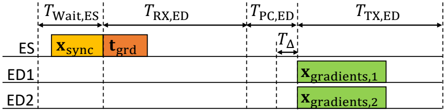
(a) Calibration trigger



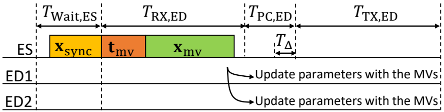
(b) Calibration feedback

How to Use the Proposed Synchronization Method?

- The proposed synchronization IP along with corresponding procedure is agnostic to the IQ data samples in UL and DL
  - **Calibration**: We utilize it for coarse time offset, CFO, and average signal power calibrations
  - **Computation**: We use it to implement FSK-MV with absentees [4] with following triggers



(c) Gradient trigger



(d) MV feedback

Experiment

Experiment

Scenario: A learning task of handwritten-digit recognition by using MNIST dataset

- SDRs: Adalm Pluto, Rev.C, v34, Xilinx Zynq XC7Z010
- Sample rate: 20 Msps
- IDFT size: 256
- # of EDs: 5
- # of active subcarriers: 192
- A CNN with 29034 parameters
- In-door environment
- Baseband language: Python
- Synchronization IP: MATLAB HDL Coder

(a) The ES with an Jetson Nano

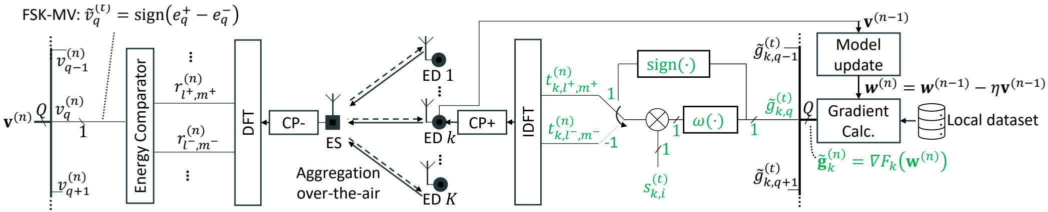
(b) The EDs with Surface Pro 4. An independent thread runs for each SDR.

$T_{\mathrm{PC,ED}} = 750$ ms, $T_{\mathrm{RX,ED}} = 50$ ms, $T_{\mathrm{TX,ED}} = 50$ ms

Data Distribution

- To prepare the datasets, we first choose $|\mathcal{D}| = 25000$ training images from the database, where each digit has distinct 2500 images
  - For homogeneous data distribution, each ED has 500 distinct images for each digit
  - For heterogeneous data distribution, kth ED has the data samples with the labels $\{k-1, k, 1+k, 2+k, 3+k, 4+k\}$
  - For both distributions, the EDs do not have common training images

OAC Scheme: FSK-based Majority Vote (FSK-MV) with Absentees [4]

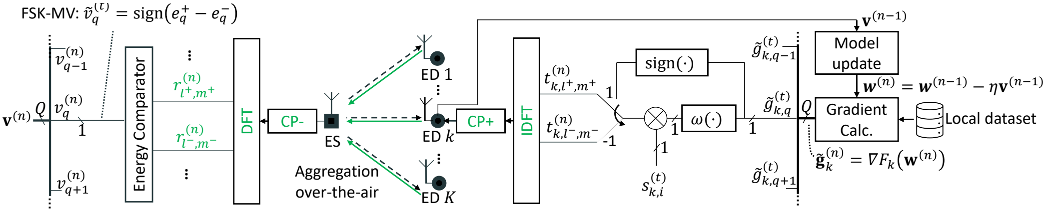


- Step 1/4: The kth ED calculates the local gradients and activates the l^-th or l^+th subcarriers of m^-th or m^+th OFDM symbols based on the sign of qth gradient

$$t_{k,l^+,m^+}^{(n)} = \sqrt{E_\text{s}}s_{k,q}^{(n)}\omega\left(\tilde{g}_{k,q}^{(n)}\right)\mathbb{I}\left[\text{sign}(\tilde{g}_{k,q}^{(n)}) = 1\right]$$
$$t_{k,l^-,m^-}^{(n)} = \sqrt{E_\text{s}}s_{k,q}^{(n)}\omega\left(\tilde{g}_{k,q}^{(n)}\right)\mathbb{I}\left[\text{sign}(\tilde{g}_{k,q}^{(n)}) = -1\right]$$

$E_\text{s} = 2$, $\omega(x) = \mathbb{I}\left[|x| \geq t\right]$ where $t \in \{0, 0.005\}$ is a threshold for absentees, $s_{k,q}^{(n)}$: Randomization symbol

OAC Scheme: FSK-based Majority Vote (FSK-MV) with Absentees [4]



- Step 2/4: The signals are superposed over the air

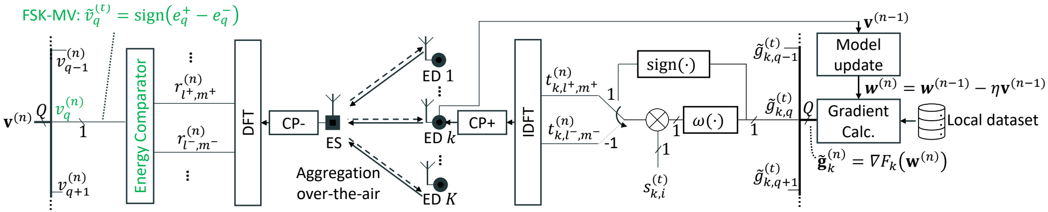
$$r_{l^+,m^+}^{(n)} = \sum_{\forall k} h_{k,l^+,m^+}^{(n)} t_{k,l^+,m^+}^{(n)} + n_{l^+,m^+}^{(n)}$$

$$r_{l^-,m^-}^{(n)} = \sum_{\forall k} h_{k,l^-,m^-}^{(n)} t_{k,l^-,m^-}^{(n)} + n_{l^-,m^-}^{(n)}$$

Channel coefficients: $h_{k,l^+,m^+}^{(n)}, h_{k,l^-,m^-}^{(n)}$:

Noise: $n_{l^+,m^+}^{(n)}, n_{l^-,m^-}^{(n)}$

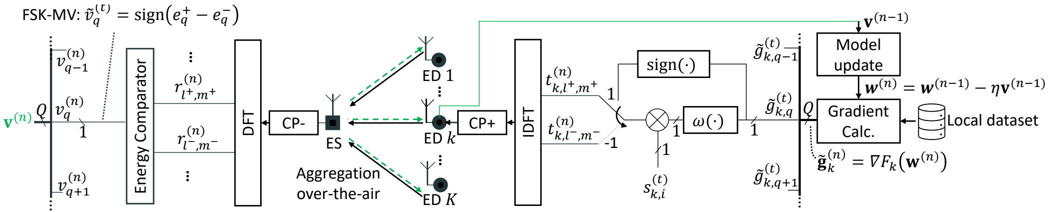
OAC Scheme: FSK-based Majority Vote (FSK-MV) with Absentees [4]



- Step 3/4: For $e_q^+ \triangleq |r_{l^+,m^+}^{(n)}|_2^2$ and $e_q^- \triangleq |r_{l^+,m^+}^{(n)}|_2^2$, the ES determines the gradient direction based on MV with an energy detector

$$v_q^{(n)} = \text{sign}\left(e_q^+ - e_q^-\right)$$

OAC Scheme: FSK-based Majority Vote (FSK-MV) with Absentees [4]



- Step 4/4: The ES broadcasts $\mathbf{v}^{(n)} = [v_1^{(n)}, \ldots, v_Q^{(n)}]^{\mathrm{T}}$ and the models at the EDs are updated as

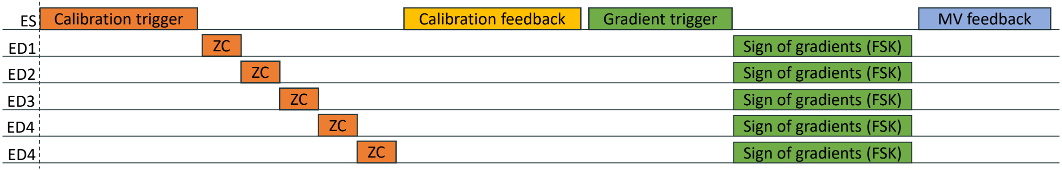
$$\mathbf{w}^{(n+1)} = \mathbf{w}^{(n)} - \eta\mathbf{v}^{(n)}$$

Key Benefit

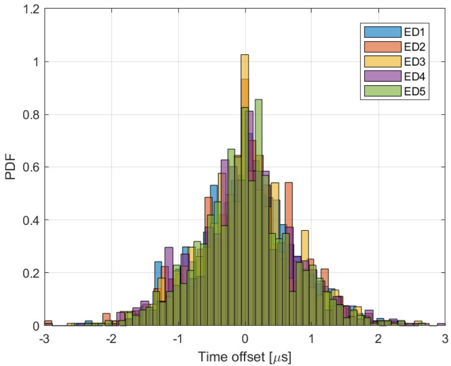
FSK-MV is robust against time synchronization errors as 1) it does not need CSI at the EDs and ES and 2) it does not rely on phase synchronization

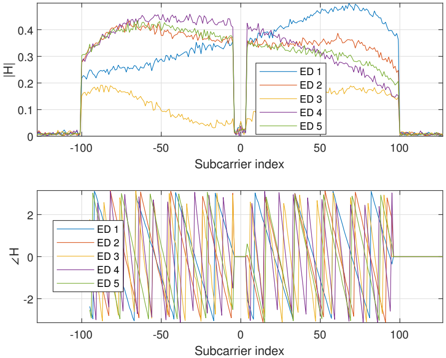
Complete Procedure for Each Communication Round

- The procedure along with calibration for each communication round is given below
  - With the calibration feedback, average power offset, CFO, and time offset are fed back to each ED for calibration



Results: Time Offset Distribution and channel frequency response (CFR)





- The jitter (standard deviation $\sim 1\ \mu$s) is large due to the imperfect clock in the SDR
- The magnitudes of the channel frequency coefficients do not change significantly, but their phases change in an intractable manner due to the random time offsets

Results: Without Absentee Votes under Homogeneous Data Distribution

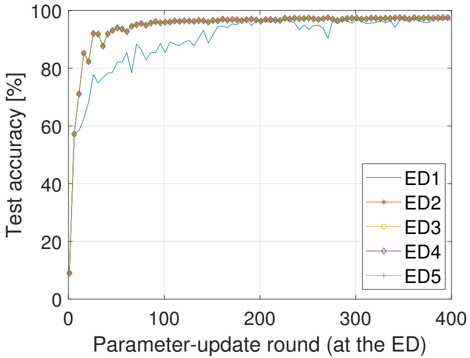


Figure 1: Test accuracy versus rounds

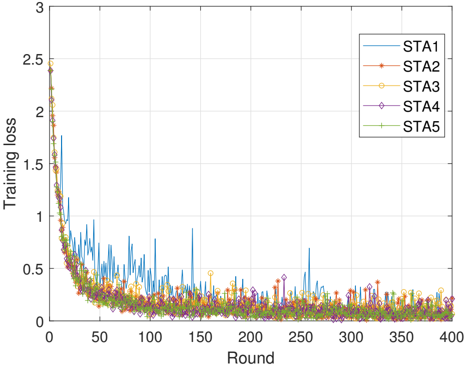


Figure 2: Training loss versus rounds

- For homogeneous data distribution without absentee votes, the test accuracy for each ED quickly reaches 97.5% and the training losses decrease gradually

Results: Without Absentee Votes under Heterogeneous Data Distribution

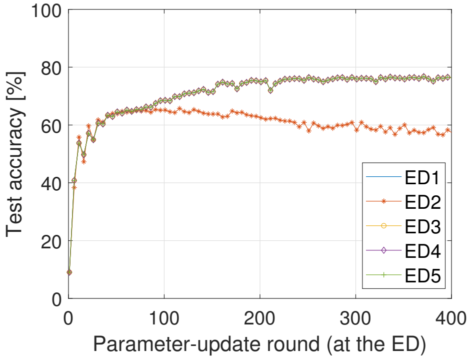


Figure 3: Test accuracy versus rounds

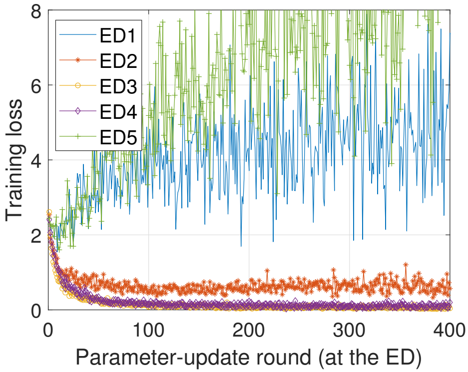


Figure 4: Training loss versus rounds

- For heterogeneous data distribution without absentee votes, the test accuracy drops below 80% and ED1 and ED5 suffer from high training loss

Results: With Absentee Votes under Heterogeneous Data Distribution

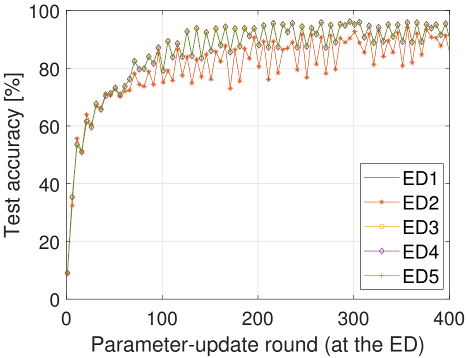


Figure 5: Test accuracy versus rounds

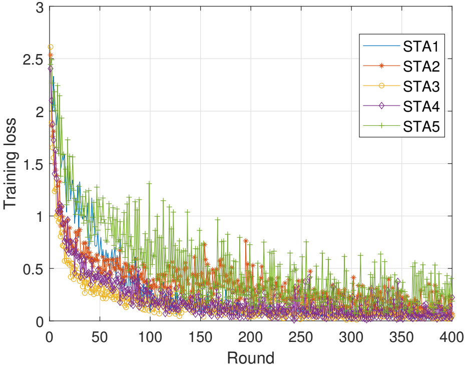


Figure 6: Training loss versus rounds

- For heterogeneous data distribution scenario, eliminating converging ED by using a larger threshold improves the test accuracy considerably

Final Remarks

- We propose a method and the corresponding procedures that can maintain the synchronization in an SDR-based network without implementing the baseband as a hard-coded block
- By implementing the proposed concept with Adalm Pluto SDRs, for the first time, we demonstrate the performance of an OAC, i.e., FSK-MV, for FEEL
- Our experiment shows that FSK-MV provides robustness against time synchronization errors and can result in a high test accuracy in practice

References i

[1] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. Vincent Poor, "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, pp. 1–26, 2021.

[2] T. Gafni, N. Shlezinger, K. Cohen, Y. C. Eldar, and H. V. Poor, "Federated learning: A signal processing perspective," 2021. [Online]. Available: arXiv:2103.17150

[3] M. Goldenbaum, H. Boche, and S. Stańczak, "Nomographic functions: Efficient computation in clustered gaussian sensor networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2093–2105, 2015.

[4] A. Şahin, "Distributed learning over a wireless network with non-coherent majority vote computation," 2022. [Online]. Available: https://arxiv.org/abs/2209.04692

Questions?