Communications for Distributed Computations

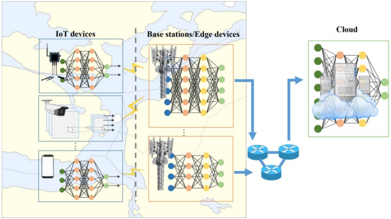
I Brazilian Signal Processing Forum:
cooperating for a connected world

Carlo Fischione

KTH Royal Institute of Technology
carlofi@kth.se

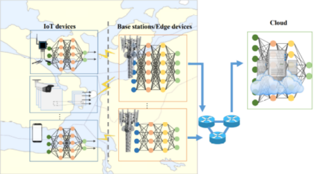
June 2024

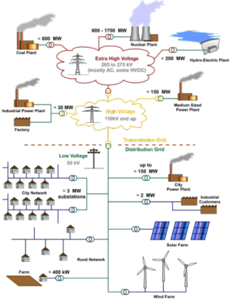
ML and Wireless: Challenges



- In wireless networks and devices, it is difficult to make ML training and inference in real time.
- The networks and devices are distributed, and heterogeneous, even using different communication protocols.
- Inference on a device/access network needs data from other devices and network locations as a collaborative effort.
- A major concern is energy efficiency, bandwidth limitations, privacy, and security.

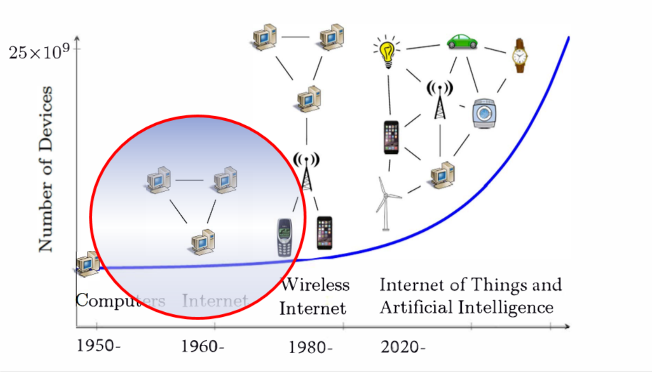
Many Use-cases of ML in Wireless Networks





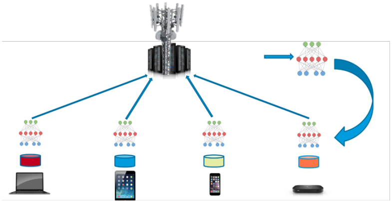
- Smart Cities, Smart Grids, Autonomous Vehicles
- Personal Health Monitoring, Communication Infrastructure

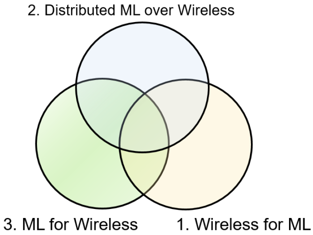
But ML is Still Conceived for Past Technological Revolutions!



- ML is still conceived for centrally collected data or private powerful networks of processors having clean, easy to access, statistically rich data, without communication delays or bandwidth limitations
- Traditional ML is challenged by wireless networks
- Current wireless networks are inefficient for ML services

ML and Wireless Research





- ML over Wireless Networks is concerned with
  - **Distributed model training**
  - **Distributed inference**
- We can use ML in wireless networks for
  1. redesign or adaptation of wireless access protocols to support ML/AI services;
  2. ML services over wireless networks;
  3. data-driven redesign and management of the network (e.g., in difficult channels, handover predictions, resource allocations).

Do We Need Communication Protocols for ML Computations?




- "The Americans have need of the telephone, but we do not. We have plenty of messenger boys". Sir William Preece, Chief Engineer of the British Post Office, 1876.

- "Cellular phones will absolutely not replace local wire systems". Marty Cooper, the father of the cell phone, 1974

Analog Over-the-Air Computation: OAC

State-of-the-art

OAC Federated Learning with Retransmissions

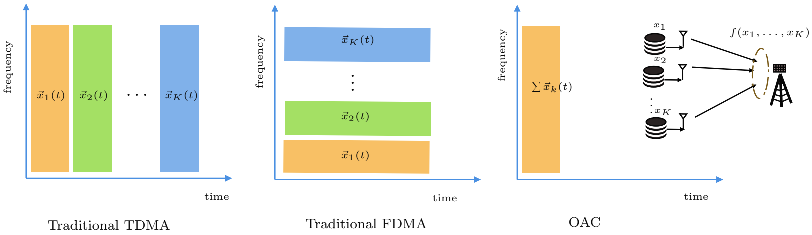
Static Retransmissions

Fast-fading Retransmissions

Digital Over-the-Air Computation: ChannelComp

Conclusions

Over-the-Air Computation (OAC)

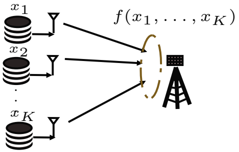


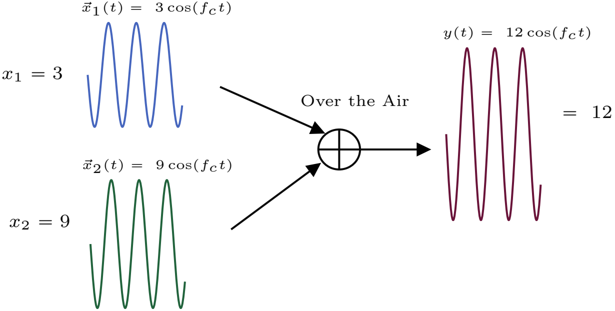
- In Federated Learning, the model/gradient sum at the central server can be "automatically computed" by wireless interference.
- The devices transmit simultaneously over the same channels, which leads to a natural sum:

$$\vec{y}(t) = \sum_k \vec{x}_k(t), \quad t = 1, 2, \quad \ldots \tag{1}$$

- Potentially, tremendous energy, frequency, privacy, security, and efficiency benefits!

OAC Uses Analog Modulations





The OAC state-of-the-art assumes Amplitude Analog Modulations.

Analog Over-the-Air Computation: OAC

State-of-the-art

OAC Federated Learning with Retransmissions

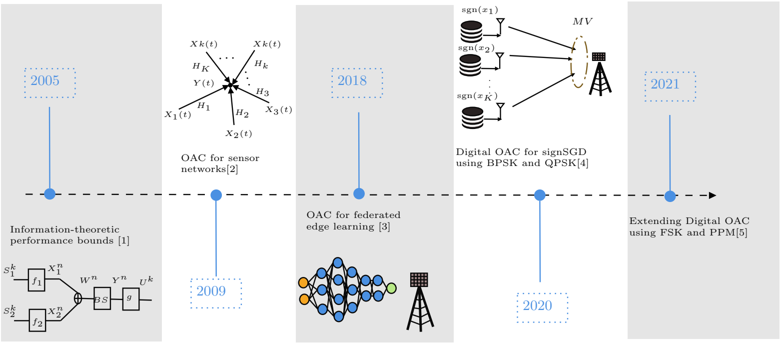
Static Retransmissions

Fast-fading Retransmissions

Digital Over-the-Air Computation: ChannelComp

Conclusions

History of OAC



[1] B. Nazer *et al.*, "Reliable computation over multiple-access channels," in *Allerton Conf. on Commun., Control, and Computing*, 2005

[2] M. Goldenbaum *et al.*, "On function computation via wireless sensor multiple-access channels," in *IEEE Wire. Commun. and Net. Conf.*, 2009

[3] G. Zhu *et al.*, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. on Wire.Commun.*, 2019

[4] G. Zhu *et al.*, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Wireless Commun.*, 2020

[5] A. Şahin *et al.*, "Distributed learning over a wireless network with FSK-based majority vote," in *IEEE CommNet*, 2021

State-of-the-art (1/2)

Topic	**Ref.**	**Summary**
Broadband Analog Aggregation	[3]	FL using AirComp over broadband channels with truncated channel inversion to handle fading.
Gradient Sparsification	[6]	Sparsification of gradients combined with error accumulation for compression.
	[7]	Extension of [6] to consider fading channels, uses truncated channel inversion.
	[8]	Performance comparison of [7] scheme, sequential digital transmission, and BAA.
	[9]	Utilization of temporal structures in the gradient updates to form a Bayesian prior in the gradient estimation step.
Federated Distillation	[10]	Trains by communicating model outputs instead of model parameters. Over-the-air computation is used to combine model output vectors for each class.

[6] M. M. Amiri *et al.*, "Machine Learning at the Wireless Edge: Distributed Stochastic Gradient Descent Over-the-Air," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2155–2169, Mar. 2020

[7] M. M. Amiri *et al.*, "Over-the-Air Machine Learning at the Wireless Edge," in *SPAWC*, IEEE, Aug. 2019, pp. 1–5

[8] M. M. Amiri *et al.*, "Federated Learning over Wireless Fading Channels," *IEEE Trans. on Wire. Commun.*, vol. 19, no. 5, pp. 3546–3557, Feb. 2020

[9] D. Fan *et al.*, "Temporal-Structure-Assisted Gradient Aggregation for Over-the-Air Federated Edge Learning," *arXiv*, vol. abs/2103.02270, Mar. 2021

[10] J.-H. Ahn *et al.*, "Wireless Federated Distillation for Distributed Edge Learning with Heterogeneous Data," in *PIMRC*, IEEE, Jul. 2019, pp. 1–6

State-of-the-art (2/2)

Topic	Ref.	Summary
Training with Noisy Gradients	[11]	Proposal of gradient-based multiple-access scheme that does not cancel the fading effect but operates directly with noisy gradients.
	[12]	Convergence rate analysis for gradient-based multiple-access.
Data Sharing	[13]	DSGD training using combined gradients. Introduces data redundancy to combat non-IID. data.
Analog Federated ADMM	[14]	Second-order training algorithm with CoMAC communication.

- For a detailed exposition of the literature, see [15].

[11] T. Sery *et al.*, "A Sequential Gradient-Based Multiple Access for Distributed Learning over Fading Channels," in *Allerton*, IEEE, Dec. 2019, pp. 303–307

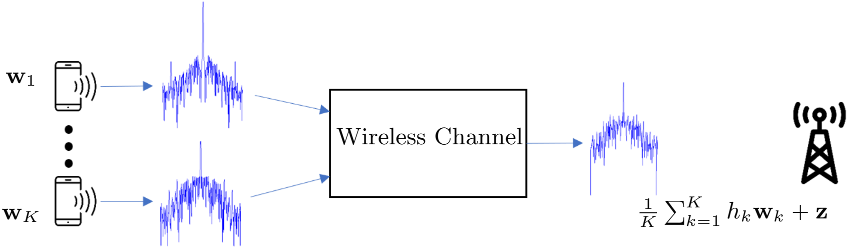
[12] T. Sery *et al.*, "On Analog Gradient Descent Learning over Multiple Access Fading Channels," *IEEE Trans. on Sig. Proc.*, vol. 68, pp. 2897–2911, Apr. 2020

[13] Y. Sun *et al.*, "Energy-Aware Analog Aggregation for Federated Learning with Redundant Data," in *ICC*, IEEE, Jul. 2020, pp. 1–7

[14] A. Elgabli *et al.*, "Harnessing Wireless Channels for Scalable and Privacy-Preserving Federated Learning," *IEEE Trans. on Commun.*, May 2021

[15] H. Hellström *et al.*, "Wireless for machine learning: A survey," *Foundations and Trends® in Sig. Proc.*, vol. 15, no. 4, pp. 290–399, 2022

OAC Introduces Estimation Errors



- OAC deliberately generates interference over the wireless channel
  - The desired function is estimated using the superimposed received signal
  - The individual model parameter vectors $\mathbf{w}_k$ are never recreated at the receiver
- Due to the analog modulations, channel attenuation and additive noise, there are inevitable estimation errors

Estimation Errors

- With heterogeneous fading and additive noise, the received signal is a noisy and distorted sum of the transmitted messages
  - $y[t] = \sum_{k=1}^{K} \frac{h_k[t] b_k[t] \mathbf{w}_k}{\sqrt{\eta}} + \frac{z[t]}{\sqrt{\eta}}$
- Given independent Gaussian sources and global channel knowledge, the minimum mean-squared error estimator (MMSE) is biased [16]
  - $\eta* = \min_k \left(\frac{\sigma_z^2 + \sum_{i=1}^{k} P_{\max} |h_i|^2}{\sum_{i=1}^{k} P_{\max} |h_i|} \right)^2$
  - $b_k^* = \frac{h_k[t]^H}{|h_k[t]|^2} \min \left(P_{\max}, \frac{\eta*}{|h_k|^2} \right)$
- Even with optimal estimation, significant estimation errors, due to bias, remain
- How do we reduce them?

---

[16] X. Cao *et al.*, "Optimized power control for over-the-air computation in fading channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7498–7513, 2020

Retransmissions over Static Channels

- With $M-1$ retransmissions over block-fading static channels, the received signal becomes
  - $y[t] = \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{k=1}^{K}\frac{h_k[t]b_k[t]\mathbf{w}_k}{\sqrt{\eta}} + \frac{z[t,m]}{\sqrt{\eta}}\right)$
- Signal-part interferes constructively, while ergodic noise leads to destructive interference
- Federated Learning algorithm with retransmissions:
  1. Random model initialization
  2. Broadcast model in downlink
  3. Local training at User Devices
  4. for $m = 1 : M$
     4.1 Uplink OAC aggregation of model updates
  5. Compute mean at Access Point
  6. Repeat 2-5 until convergence

Retransmissions over Static Channels

- With standard tools from convex optimization theory, we can prove upper bounds on over-the-air federated learning convergence with retransmissions [17]
- Let

$$c_2 := 1 - 2\beta \frac{\mu L}{\mu + L}, \tag{2}$$

and

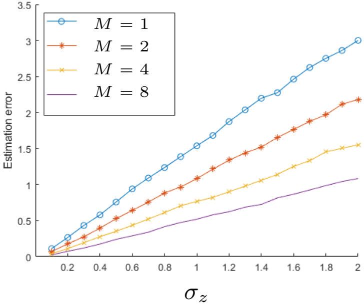
$$c_3 := \beta^2 ||\sigma||^2 K \sum_{k=1}^{K} p_k |h_k|^2 + \frac{d\sigma_z^2}{M}. \tag{3}$$

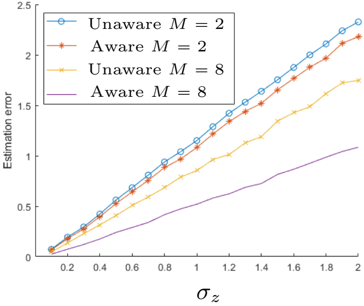
Then,

$$\mathbb{E}\left[F(\mathbf{w}_n)\right] - F(\mathbf{w}^*) \leqslant \frac{L}{2} c_2^n \mathbb{E}[r_0^2] + \frac{L c_3}{2\left(\sum_{k=1}^{M} \sqrt{p_k}|h_k|\right)^2 (1 - c_2)}, \tag{4}$$

[17] H. Hellström *et al.*, "Federated learning over-the-air by retransmissions," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9143–9156, 2023

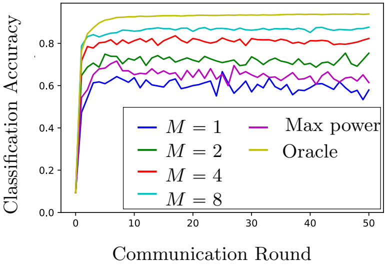
Retransmissions over Static Channels

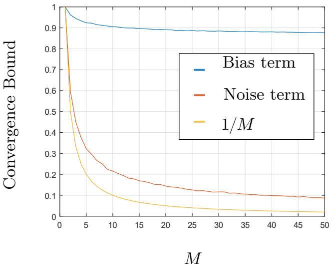




- MSE-minimizing power control is dependent on the number of retransmissions, i.e., the devices should be aware of M when selecting their transmission powers
- MSE reductions are expensive compared to channel codes, but offer a first step toward enabling an estimation-communication tradeoff

Retransmissions over Static Channels

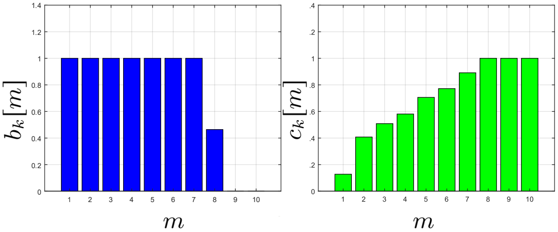




- Retransmissions improve post-convergence accuracy
- More expensive in terms of communication

- Noise-related term falls of at approximately $1/M$
- Slight decline in bias-related term

Optimal Precoder with Retransmissions over Fast-fading

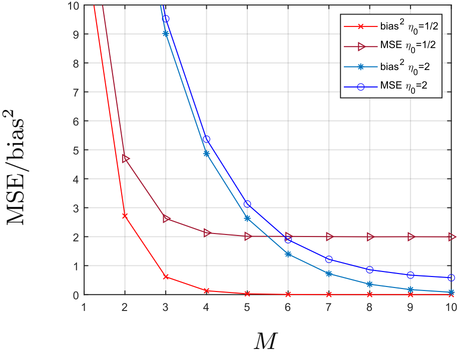


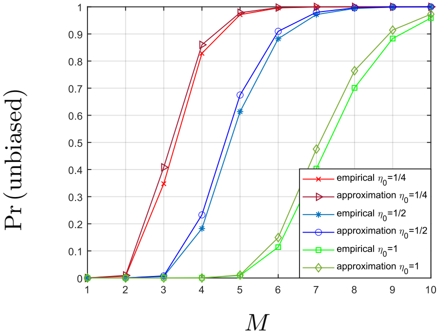
- OAC function estimation is plagued by bias
- We can exploit the time-diversity of fast fading to reduce the bias
- With tools from Algorithm Analysis, we prove that the optimal causal power control scheme follows a greedy approach [18]

$$b_k[m]^* = \min\left(\sqrt{p_{\max}}, \left(1 - \sum_{i=1}^{m-1} \frac{|h_k[i]| b_k[i]}{M\sqrt{\eta}}\right) \frac{M\sqrt{\eta}}{|h_k[t]|}\right), \tag{5}$$

[18] H. Hellstrom *et al.*, "Unbiased over-the-air computation via retransmissions," in *IEEE Global Communications Conference*, 2022, pp. 782–787

Numerical Results over Fast-fading Channels





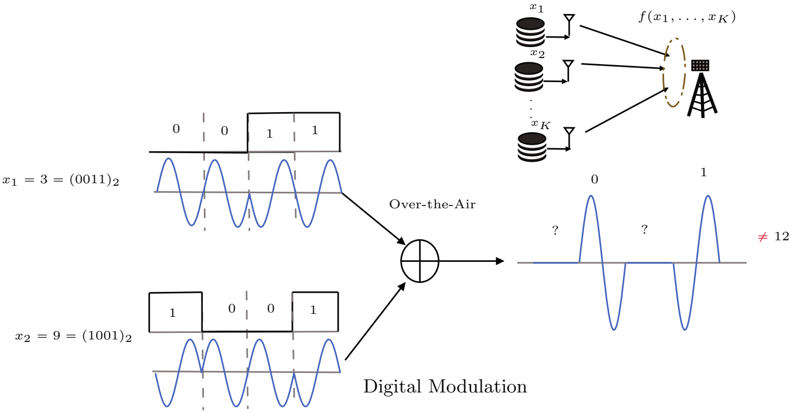
- Rapid decrease in bias
- MSE floor that depends on choice of post-transmission scalar η
- More efficient tradeoff than for static channels

- Without retransmissions ($M = 1$), estimator bias is inevitable
- Unbiased probability increases toward 100% within finite number of retransmissions

Take-home Message

- The quality of the estimation provided by OAC can be significantly improved by retransmissions.
- Over-the-air computation introduces a bias.
- Power control and retransmissions can help to significantly reduce or eliminate the bias, especially for fast-fading channels

OAC Does not Work with Digital Modulations



- OAC with digital modulations seems unfeasible, because the overlapping of digital waveforms returns incomprehensible signals.

Research Gap in OAC and Digital Modulations

Methods / Subjects	Analog Modulation		Digital Modulation			
Papers	[2], [3], [19]	[20]	[4]	[5], [21]	[22]	Ours
Spectral Efficiency	✓	✗	✓	✗	✓	✓
Low Latency	✓	✗	✓	✗	✓	✓
BPSK and QPSK	✗	✗	✓	✗	✓	✓
QAM 16, 32, . . .	✗	✗	✗	✗	✗	✓
Analog Modulation	✓	✓	✗	✗	✗	✓
Sign Function	✗	✗	✓	✓	✓	✓
Nomographic Functions	✓	✓	✗	✗	✗	✓
General Functions	✗	✗	✗	✗	✗	✓
Ubiquitous implementation	✗	✗	✓	✓	✓	✓

✓: Performance is very good! ✗: It is not studied at all.

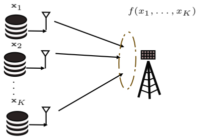
[19] M. Goldenbaum *et al.*, "Robust analog function computation via wireless multiple-access channels," *IEEE Trans. on Commun.*, 2013

[20] A. Şahin *et al.*, "Distributed learning over a wireless network with FSK-based majority vote," in *IEEE CommNet*, 2021

[21] A. Şahin, "A demonstration of over-the-air computation for federated edge learning," in *IEEE Globecom Workshops*, 2022

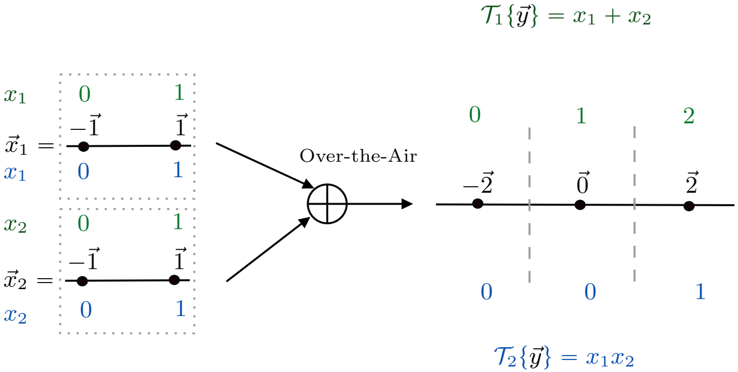
[22] A. Şahin *et al.*, "Over-the-air computation over balanced numerals," in *IEEE Globecom Workshops*, 2022

Our Goal with ChannelComp



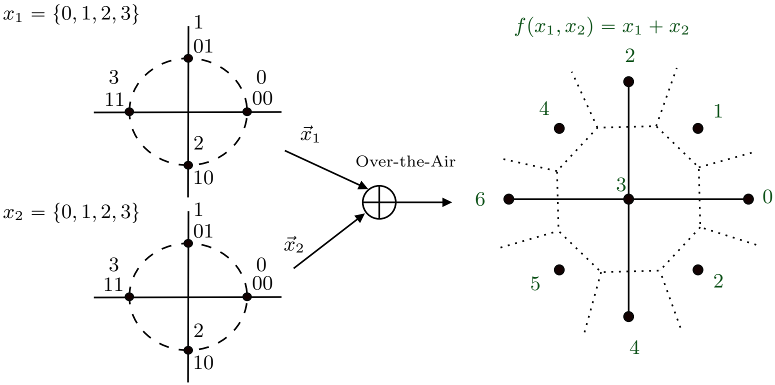
- Create OAC methods that are inherently built for **digital communications**.
- The methods should be able to perform the computation of any function $f(x_1, \dots, x_K)$ over-the-air, where the inputs belong to different units/nodes.
- The key idea: rethink how the receiver works!

Key Idea: a GOOD Example for BPSK Modulation



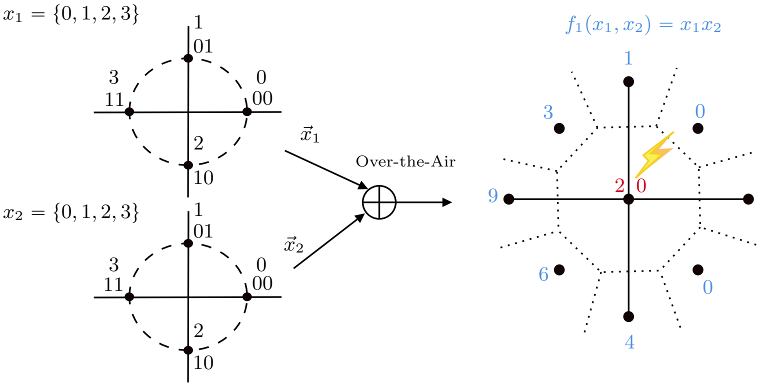
We can attach mechanically the value of the computation to a received constellation point.

Key Idea: a GOOD Example for QPSK Modulation (1/2)



By assigning specific values to the reshaped constellation points, QPSK modulation enables the computation of the summation function.

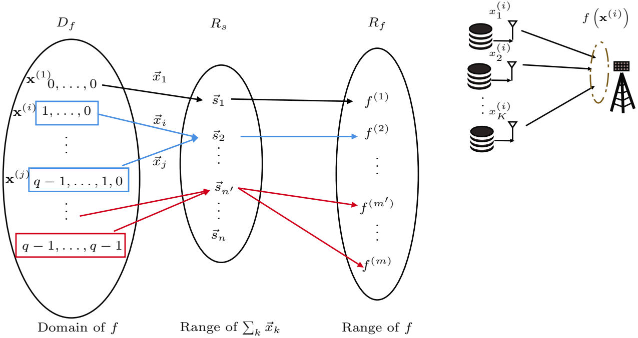
Key Idea: A BAD Example for QPSK Modulation (2/2)



The overlaps of the reshaped constellation points of QPSK modulation do NOT allow us to compute the product function.

Key Idea: the General Case of ChannelComp

$f^{(i)} := f(\mathbf{x}^{(i)})$ $\qquad$ $\mathbf{x}^{(i)} := (x_1^{(i)}, \ldots, x_K^{(i)})$



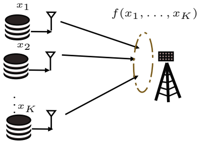
- Because of the input x is digital, the domain of f is over a finite set.
- In a noise-free channel, the constellation points at the receiver are finite.
- Consider two inputs $x^{(i)}$ and $x^{(j)}$ that generate two different function's output $f^{(i)}$ and $f^{(j)}$:
  - if $\vec{s}_i \neq \vec{s}_j$, we can associate the value $f^{(i)}$ to $\vec{s}_i$, and $f^{(j)}$ to $\vec{s}_j$.
  - if $\vec{s}_i = \vec{s}_j$, we cannot make the correct association, unless we enforce a splitting of $\vec{s}_i$ and $\vec{s}_j$ by a proper encoding.

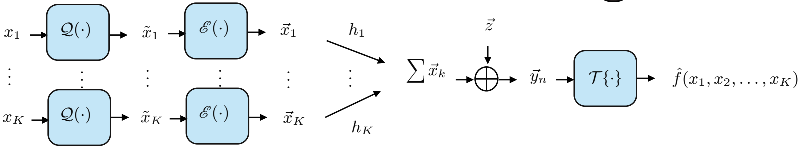
An Old Idea? The Hydraulic Telegraph, 4-th Century BC

- Attributed to Aeneas Tacticus, 4th century BC.
- Used to send messages between Sicily and Carthage (modern Tunisia) [23].
- The water levels were associated with (possibly complex) messages.
- The water levels do not mean anything by themselves, it is their association/mapping to messages that is meaningful.

[23] *Polybius: The Histories. Loeb Classical Library (in Ancient Greek, English, and Latin).* Translated by Paton, W.R. Chicago; University of Chicag, 2012

ChannelComp's Problem Formulation



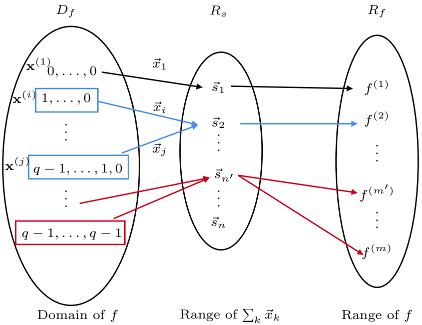


Goal: Find the constellation encoder $\mathscr{E}(\cdot)$ and the mapping $\mathcal{T}\{\cdot\}$ to do the computation for a given quantisation $\mathcal{Q}(\cdot)$:

$$\mathcal{T}^*, \mathscr{E}(\cdot)^* = \underset{\mathscr{E}}{\operatorname{argmin}} \sum_{x_1,\ldots,x_K \in \mathcal{D}_f} \Big| f(x_1,\ldots,x_K) - \underbrace{\mathcal{T}\{\vec{y}_n\}}_{\hat{f}} \Big|^2$$

Constellation Design

$f^{(i)} := f(\mathbf{x}^{(i)})$



To find the encoder, we pose the following feasibility optimization

$$\mathcal{P}_1 = \underset{\mathbf{x}}{\text{find}} \quad \mathbf{x}$$

$$\text{s.t.} \quad f^{(i)} \neq f^{(j)} \Rightarrow \vec{s}_i \neq \vec{s}_j, \ \forall (i,j) \in [M]^2, \tag{6a}$$

$$\|\mathbf{x}\|_2^2 \leqslant P. \tag{6b}$$

ChannelComp 1: Any Modulation

Proposition (Necessary condition [24])

Let the K multivariate function $f(x_1, x_2, \ldots, x_K)$ with domain $\mathcal{D}_f$, where $x_k \in \mathcal{D}_f$ for $k \in [K]$ be a symmetric function, i.e.,

$$f(x_1, \ldots, x_K) = f(\pi(x_1), \ldots, \pi(x_K)), \tag{7}$$

for all possible permutations by $\pi : \{1, \ldots, K\} \mapsto \{1, \ldots, K\}$. Let each node use the identical modulation $\mathscr{E}$. Then, function f can be computed by the constellation diagram of $\sum_{k=1}^{K} \mathscr{E}(x_k)$.

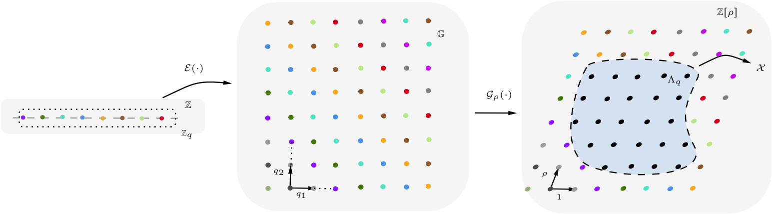
---

[24] S. Razavikia *et al.*, "ChannelComp: A general method for computation by communications," *IEEE Transactions on Communications*, 2023. DOI: 10.1109/TCOMM.2023.3324999

Proposition ([24])

Let $\epsilon^{-1} \geqslant \max_{(i,j)\in[M]^2} |f^{(i)} - f^{(j)}|^2$. Then, Problem $\mathcal{P}_2$ is feasible, and thus there exists a modulation vector $\mathbf{x}$ satisfying the constraints.

ChannelComp 2: q-QAM by the Ring of Integers (SumComp)



Theorem (MAE Analysis [25])

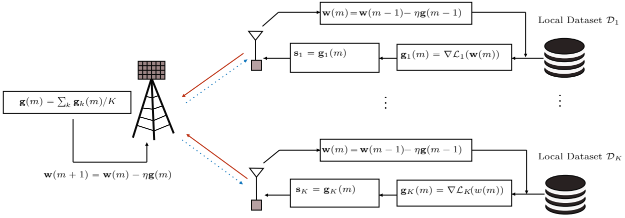
Under the same conditions given in [25, Theorem 1], except for $f := \psi\Big(\sum_{k=1}^{K} \varphi_k(s_k)\Big)$, *we have*

$$\mathrm{MAE}(\hat{f}) := \mathbb{E}\{|f - \hat{f}|\}, \leqslant w_\psi(\sqrt{q_1^2 e_1 + q_2^2 e_2}), \tag{8}$$

where w_ψ *signifies the modulus of continuity of* ψ *(an extension of Lipschitz continuity).*

[25] S. Razavikia *et al.*, *SumComp: Coding for digital over-the-air computation via the ring of integers*, 2023. arXiv: 2310.20504 [cs.IT]

ChannelComp 3: SumComp for Blind Federated Learning



Proposition (Number of Antennas [26])

Let $\mathbf{g} \in \mathbb{C}^N$ be the global gradient averaged by the ES, and $\hat{\mathbf{g}}$ be the estimated gradient of $\mathbf{g}$. Then, with probability no less than $1 - \delta$, the error of the estimated gradient, $\hat{\mathbf{g}}$, as well as the MSE of $\hat{\mathbf{g}}$, are bounded by scalar σ^2_{fad}, i.e.,

$$\|\tilde{\mathbf{g}} - \hat{\mathbf{g}}\| \leqslant \epsilon_{\text{fad}}, \qquad \mathbb{E}[\|\mathbf{g} - \hat{\mathbf{g}}\|^2] \leqslant \sigma^2_{\text{fad}} + \sigma^2_q, \tag{9}$$

where $\sigma^2_{\text{fad}} := 16\frac{N\gamma^2_{\max}q}{N_r c^2_{\min}}(\pi + 2\ln(6K)^2)$, if the number of antennas, N_r, is greater than $\frac{16\gamma^2_{\max}Nq}{\epsilon^2_{\text{fad}}c^2_{\min}}\ln\left(\frac{6K}{\delta}\right)$, where $\gamma_{\max} := \max_n \gamma_n$, $c_{\min} := \min_n c_n$, and q is the order of modulations.

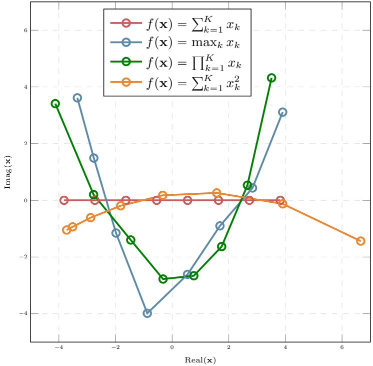
[26] S. Razavikia *et al.*, *Blind federated learning via over-the-air q-QAM*, 2023. arXiv: 2311.04253 [eess.SP]

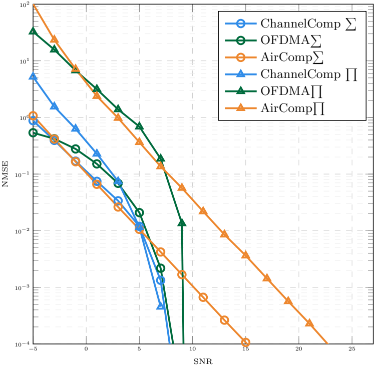
Simulations Setup

- ChannelComp performance is compared to
  - OFDMA.
  - OAC, which uses analog modulation.
- Functions tested with $K = 4$ nodes:
  - $f_1 = \sum_{k=1}^{4} x_k$
  - $f_2 = \prod_{k=1}^{4} x_k$
  - $f_3 = \sum_{k=1}^{4} x_k^2$
  - $f_4 = \max_k x_k$

  for $x_k \in \{0, 1, 2, \ldots, 7\}$
- Input signals transmitted over an AWGN channel.
- NMSE used to characterize computation error over $N_s = 100$ Monte Carlo trials for different SNRs.

Performance Comparison





- Thanks to constructive overlaps of the reshaped modulation, ChannelComp outperforms AirComp and OFDMA with more than **10** dB improvement for the product function.

Analog Over-the-Air Computation: OAC

Digital Over-the-Air Computation: ChannelComp

Conclusions

Conclusions

- In ML over networks, we will often encounter the problem of computing functions from distributed units connected by wireess multiple access network (MAC)
- Our work is *the first attempt* to propose general digital modulations for function computation over the MAC.
- The proposed ChannelComp properties:
  - Ultra-low-latency
  - General functions computation
  - Any digital modulations
  - Simple communication architecture
  - Integration of both the encoder and modulation
  - Extension of OAC (it works for analog as well)
- Generalization to MIMO, fading channels, asynchronous, etc.
- Applications of ChannelComp for, e.g., federated edge learning, or distributed sensing problems.

Acknowledgements

This presentation is based on the following papers:

- L. Turchet, C. Fischione, G. Essl, D. Keller, M. Barthet, "Internet of Musical Things: Vision and Challenges", *IEEE Access*, 2018
- H Hellström, J. M. Barros da Silva Jr., M. M. Amiri, M. Chen, V. Fodor, V. Poor, C. Fischione, "Wireless for Machine Learning: A Survey", *NOW Foundations and Trends in SP*, 2022.
- S. Razavikia, J. M. Barros da Silva Jr., C. Fischione, "ChannelComp: A General Method for Computation by Communications", *IEEE TCOM*, 2023.
- S. Razavikia, J. M. Barros da Silva Jr., C. Fischione, "SumComp: Coding for Digital Over-the-Air Computation via the Ring of Integers", Submitted to *IEEE TCOM*, 2023.
- S. Razavikia, J. M. Barros da Silva Jr., C. Fischione, "Blind federated learning via over-the-air q-QAM ", Submitted to *IEEE TWC*, 2023.

Q&A

Thanks for your attention! Any question?

References I

[1] B. Nazer *et al.*, "Reliable computation over multiple-access channels," in *Allerton Conf. on Commun., Control, and Computing*, 2005.

[2] M. Goldenbaum *et al.*, "On function computation via wireless sensor multiple-access channels," in *IEEE Wire. Commun. and Net. Conf.*, 2009.

[3] G. Zhu *et al.*, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. on Wire.Commun.*, 2019.

[4] G. Zhu *et al.*, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Wireless Commun.*, 2020.

[5] A. Şahin *et al.*, "Distributed learning over a wireless network with FSK-based majority vote," in *IEEE CommNet*, 2021.

[6] M. M. Amiri *et al.*, "Machine Learning at the Wireless Edge: Distributed Stochastic Gradient Descent Over-the-Air," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2155–2169, Mar. 2020.

[7] M. M. Amiri *et al.*, "Over-the-Air Machine Learning at the Wireless Edge," in *SPAWC*, IEEE, Aug. 2019, pp. 1–5.

References II

[8] M. M. Amiri *et al.*, "Federated Learning over Wireless Fading Channels," *IEEE Trans. on Wire. Commun.*, vol. 19, no. 5, pp. 3546–3557, Feb. 2020.

[9] D. Fan *et al.*, "Temporal-Structure-Assisted Gradient Aggregation for Over-the-Air Federated Edge Learning," *arXiv*, vol. abs/2103.02270, Mar. 2021.

[10] J.-H. Ahn *et al.*, "Wireless Federated Distillation for Distributed Edge Learning with Heterogeneous Data," in *PIMRC*, IEEE, Jul. 2019, pp. 1–6.

[11] T. Sery *et al.*, "A Sequential Gradient-Based Multiple Access for Distributed Learning over Fading Channels," in *Allerton*, IEEE, Dec. 2019, pp. 303–307.

[12] T. Sery *et al.*, "On Analog Gradient Descent Learning over Multiple Access Fading Channels," *IEEE Trans. on Sig. Proc.*, vol. 68, pp. 2897–2911, Apr. 2020.

[13] Y. Sun *et al.*, "Energy-Aware Analog Aggregation for Federated Learning with Redundant Data," in *ICC*, IEEE, Jul. 2020, pp. 1–7.

References III

[14] A. Elgabli *et al.*, "Harnessing Wireless Channels for Scalable and Privacy-Preserving Federated Learning," *IEEE Trans. on Commun.*, May 2021.

[15] H. Hellström *et al.*, "Wireless for machine learning: A survey," *Foundations and Trends® in Sig. Proc.*, vol. 15, no. 4, pp. 290–399, 2022.

[16] X. Cao *et al.*, "Optimized power control for over-the-air computation in fading channels," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7498–7513, 2020.

[17] H. Hellström *et al.*, "Federated learning over-the-air by retransmissions," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9143–9156, 2023.

[18] H. Hellstrom *et al.*, "Unbiased over-the-air computation via retransmissions," in *IEEE Global Communications Conference*, 2022, pp. 782–787.

[19] G. Zhu *et al.*, "Over-the-air computing for wireless data aggregation in massive IoT," *IEEE Wireless Commun.*, 2021.

References IV

[20] M. Goldenbaum *et al.*, "Robust analog function computation via wireless multiple-access channels," *IEEE Trans. on Commun.*, 2013.

[21] A. Şahin, "A demonstration of over-the-air computation for federated edge learning," in *IEEE Globecom Workshops*, 2022.

[22] A. Şahin *et al.*, "Over-the-air computation over balanced numerals," in *IEEE Globecom Workshops*, 2022.

[23] *Polybius: The Histories. Loeb Classical Library (in Ancient Greek, English, and Latin)*. Translated by Paton, W.R. Chicago; University of Chicag, 2012.

[24] S. Razavikia *et al.*, "ChannelComp: A general method for computation by communications," *IEEE Transactions on Communications*, 2023. DOI: 10.1109/TCOMM.2023.3324999.

[25] S. Razavikia *et al.*, *SumComp: Coding for digital over-the-air computation via the ring of integers*, 2023. arXiv: 2310.20504 [cs.IT].

[26] S. Razavikia *et al.*, *Blind federated learning via over-the-air q-QAM*, 2023. arXiv: 2311.04253 [eess.SP].