Over-the-Air Computation for Distributed Systems: Something Old and Something New

Zheng Chen, Erik G. Larsson, Carlo Fischione, Mikael Johansson, and Yura Malitsky

Abstract

Facing the upcoming era of Internet-of-Things and connected intelligence, efficient information processing, computation, and communication design becomes a key challenge in large-scale intelligent systems. Recently, Over-the-Air (OTA) computation has been proposed for data aggregation and distributed computation of functions over a large set of network nodes. Theoretical foundations for this concept exist for a long time, but we mainly investigate within the context of wireless sensor networks. There are still many open questions when applying OTA computation in different types of distributed systems where modern wireless communication technology is applied. In this article, we provide a comprehensive overview of the OTA computation principle and its applications in distributed learning, control, and inference systems, for both server-coordinated and fully decentralized architectures. Particularly, we highlight the importance of the statistical heterogeneity of data and wireless channels, the temporal evolution of model updates, and the choice of performance metrics, for the communication design in OTA federated learning (FL) systems. Several key challenges in privacy, security, and robustness aspects of OTA FL are also identified for further investigation.

Introduction

Wireless communications networks, designed with communication among humans in mind during the last fifty years, are now becoming overwhelmed with data traffic that stems from emerging artificial intelligence (AI) and machine learning (ML) applications. One distinction of the data traffic required by AI/ML applications is that it is subject to entirely different performance requirements than traditional human-perceived data (e.g., text messages, voice calls or multimedia content). More explicitly, the traffic generated by AI/ML applications often contains computation results, which may be useful even if they are only received approximately correctly. Traditional quality-of-service performance metrics in communications such as error probability and rate may not be relevant for AI/ML applications. Rather, entirely different criteria will matter more, such as convergence speed of learning models, and statistical inference performance on average, in expectation over an ensemble of random data. This will require a re-design of the networks, spanning from the physical layer to the application layer. It will also require going beyond classical Shannon information theory, which has underpinned all generations of networks so far.

Within the AI/ML field, several classes of distributed algorithms have been developed with which autonomous agents (clients), that could be mobile users or sensors, collaboratively solve large-scale inference and learning tasks. The progress on collaborative and distributed intelligent systems has motivated the concept of wireless edge intelligence, which means that the computation tasks that were traditionally performed in centralized cloud servers are moved towards the wireless network edge. Implementing distributed algorithms in wireless networks often requires iterative exchange and aggregation of information among the agents over resource-constrained communication links. One main challenge is the resource allocation among the agents, where the orthogonal division of frequency/time resources can be inefficient when the number of agents is very large. If we consider that the goal of communication for machine-perceived data differs from human-perceived data, joint computation and communication design without error-free bits has the potential to achieve better performance than classical digital transmission designs.

Over-the-Air (OTA) computation, also referred to as AirComp, has emerged recently as a promising solution for computing functions of data from distributed nodes over wireless links by exploiting direct signal superposition in the analog domain [1]. In this article, we provide an overview of the basics, applications and research questions of OTA aggregation in distributed systems. As compared to existing overviews in the literature, we summarize previously overlooked or under-explored aspects of OTA FL, such as statistical heterogeneity and temporal evolution of model updates, choices of performance metrics, and the communication design for fully decentralized systems.

The organization of this article is summarized as follows. In Sec. II, we briefly explain the concept of OTA aggregation for distributed computation of nomographic functions. In Sec. III, we present applications of OTA computation in distributed learning, control and inference systems. Then, we focus on the communication and signal...
processing aspects of OTA FL in Sec. IV, and highlight the challenges in privacy and security aspects in Sec. V. At last, in Sec. VI, we conclude the article and summarize important research directions.

**Over-the-Air (OTA) Aggregation as a Computational Tool**

With OTA computation, by virtue of the superposition principle of the wave equation that governs the wireless multiple-access channel, information sent by the agents naturally adds up “in the air.” Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is a nomographic function [2], such that

$$f(x_1, \ldots, x_N) = \phi \left( \sum_{i=1}^{N} \psi(x_i) \right)$$

for some scalar functions $\phi$ and $\psi$. Then every agent $i$ can perform $\psi(\cdot)$ prior to transmission, the medium (wave superposition) performs the actual aggregation ($\phi\cdot$), and the receiver performs $\phi(\cdot)$. Any nomographic function can be computed this way through the OTA mechanism. Commonly used nomographic functions include arithmetic mean, geometric mean, and Euclidean norm.

OTA aggregation makes a clean break with the notion that modern data transmission should be digital. Its principal benefit is that rather than giving the agents orthogonal resources (as in traditional digital communications), they send their data simultaneously, saving a factor of $N$ in spectrum resources.

Though the core idea of OTA aggregation seems straightforward, in practice the picture is more complicated. First, interference and noise (say $n$) will enter at the receiver, resulting in received post-processed signal as $\phi\left( \sum_{i=1}^{N} \psi(x_i) + n \right)$.

If the inversion carried out by $\phi$ is ill-conditioned, then the presence of $n$ may be detrimental. Second, the basic OTA abstraction assumes that agents send their updates synchronously with perfect phase alignment. But in practice, some phase mismatches are unavoidable due to imperfectly known propagation delays and hardware impairments. The effect of imperfect synchronization needs to be accounted for in the construction of aggregated model updates, and such aggregation errors are structurally different from other sources of “noise” in traditional ML systems. While in theory the synchronization issue can be resolved using calibration protocols, these are difficult to implement in emerging Internet-of-Things devices that must be low-cost and highly energy-efficient, and may operate at $>30$ GHz carrier frequencies where phase drift is a major issue.

**Applications of OTA Computation in Distributed Systems**

Theoretically, OTA computation can be useful in any distributed system that requires combining and aggregating data from multiple sources at one common receiver without the need of receiving each independent stream correctly. In this section, we provide some application examples and highlight the relation between the accuracy of data aggregation and the underlying tasks/applications performed by the system.

**Distributed ML with Collaborative Training**

Currently, most research activities on the application of OTA computation focus on distributed ML with multiple agents collaborating in training a common ML model, as shown in Fig. 1. The most outstanding example is server-coordinated FL with a master-worker architecture, where the computation objective at the server is to obtain the weighted sum of the model update vectors from the agents after each round of local training. Collaborative ML can also be implemented in fully decentralized systems, where information exchange and fusion take place between locally connected agents without communication with a central server. This part will be elaborated with more details in Sec. IV-E.

In server-coordinated FL systems, to effectively compute the weighted average of the model updates at the server, each agent pre-processes its data to compensate for the amplitude degradation and phase rotation in the wireless channel, subject to some power constraints. At the server side, the received superimposed signal is post-processed to obtain an estimate of the function to be computed. This procedure is illustrated in Fig. 2.

**Distributed Control Systems**

Distributed control systems rely on multiple geographically distributed sensors and actuators to control a larger system, such as a factory or an infrastructure network. Increasingly often, the control loops are closed over wireless communication links. OTA computation could potentially improve the efficiency of such systems. For example, Park et al. [3] considers a simple scenario where a single control loop with a single actuator uses measurements from multiple sensors. OTA aggregation is applied such that all sensors simultaneously send scaled signals containing the measured plant state to the actuator. The actuator receives the superimposed control signals in the presence of channel noise. The scaling factors at the sensor and the actuator sides can be designed in a way to minimize the distortion of the received control signal, which eliminates the need of a control unit. Thus the sensor-to-actuator feedback occurs without a separate unit. The study in [3] is arguably just the initial application of OTA to distributed control, and thus several questions remain to be answered in this setting, such as the relation between system stability and the distortion of aggregated signal, and the impact of output feedback and transmit power limitations on the system performance.

**Distributed Estimation and Inference**

One example in this application is to estimate some random field with distributed sensors when the dimension of the observation at each sensor is much smaller than the dimension of the field.
This can be done by applying distributed linear estimation with consensus + innovations algorithms for the information exchange and fusion among cooperative sensors following an iterative process. In every iteration, the exchanged information contains the current estimate of the random field, where O TA can be applied for data aggregation. The impact of communication constraints and costs on the reconstruction of the random field is an interesting direction to explore.

There are numerous potential O TA applications that fall within the general framework of federated analytics. The most important characteristics of these applications are: 1) the communication goal is to compute functions of distributed data; 2) the computation result only needs to be approximately correct. In other words, not all distributed systems can potentially benefit from O TA computation for data aggregation. Currently, most articles on the application of O TA computation focus on server-coordinated FL. Therefore, in the following sections, we focus on communication design and signal processing in O TA FL. Rather than summarizing existing work on this topic, the purpose of this article is to point out what is missing in the current literature and which aspects of O TA FL should be further investigated.

**Communication and Signal Processing Perspective of O TA FL**

Even though the main advantage of O TA computation for wireless data aggregation is the efficient usage of communication resources, there are still many factors that can affect the computation and communication efficiency in O TA systems. In the existing literature on O TA FL, the following aspects have been either ignored or under-explored:

- the impact of non-IID data on the statistical heterogeneity of the local model updates;
- the disparity between the channel conditions of different agents and its impact on the statistics of the reconstruction error;
- the temporal evolution/correlation of the model updates (or gradient vectors) along the communication rounds;
- other performance metrics than the mean-square error (MSE) of the aggregated model updates.

One example is that many existing works assume that the model updates from different agents follow the same distribution with unit variance. This inherently suggests that all the agents in the system are statistically identical. As illustrated in Fig. 3, this assumption is generally not true with non-IID training data across different agents.

Another observation from Fig. 3 is that the gradient norms evolve over time. Only a few existing papers have considered the temporal structure of model updates and its impact on the data compression design [5], [6]. The joint consideration of these aspects mentioned above brings new challenges and questions in O TA FL that differ from a classical statistical estimation setting.

**Computation, Communication and Convergence**

An important aspect of FL over wireless links is the interplay between computation and communication costs and the impact of radio resource management on the convergence performance. The design space of resource allocation in wireless FL is vast, but the fundamental question is: under given constraints on frequency/time/power, how can we allocate these resources to different agents in every communication round such that the FL algorithm converges as fast and as accurately as possible? Computational complexity is another factor that can affect the convergence performance and energy consumption. Consequently,
resource limitations pose constraints on the number of agents that can participate in each communication round and the power allocation between computation (local training) and communication (sending model updates).

In the digital domain, work on communication-efficient ML has centered around compression techniques such as quantization and sparsification of gradient vectors. For FL over rate-limited wireless networks, scheduling and resource (frequency, time, power) allocation can be optimized to improve the system convergence performance under the limitation of communication resources. In the OtA domain, since all users share the same frequency/time resources, the communication efficiency is mostly affected by the data compression and scheduling design. Particularly, sparsification plays an important role in reducing the dimension of the transmitted data, thus reducing the consumed radio resources. For this to work, the locations of the sparsified elements need to be sent as side information to the receiver through some reliable channels. Furthermore, we can explore linear compression techniques inspired by compressed sensing method to recover the aggregated model parameter vector at the server side.

In the digital domain, the impact of scheduling lies in the fact that with more devices participating in model uploading, less frequency/time resources are allocated to each user, which results in higher compression loss that may deteriorate the accuracy of the aggregated model. With OtA aggregation, scheduling is also important because all the agents apply a common scaling factor in the pre-processing step such that the transmitted signals satisfy the power constraints at all agents. This scaling factor affects the variance of the effective noise after post-processing at the receiver side. From the perspective of minimizing the reconstruction error, it might be beneficial to drop several bottleneck users with large gradient norms and bad channel conditions. Furthermore, we can explore linear compression techniques inspired by compressed sensing method to recover the aggregated model parameter vector at the server side.

In the digital domain, the impact of scheduling lies in the fact that with more devices participating in model uploading, less frequency/time resources are allocated to each user, which results in higher compression loss that may deteriorate the accuracy of the aggregated model. With OtA aggregation, scheduling is also important because all the agents apply a common scaling factor in the pre-processing step such that the transmitted signals satisfy the power constraints at all agents. This scaling factor affects the variance of the effective noise after post-processing at the receiver side. From the perspective of minimizing the reconstruction error, it might be beneficial to drop several bottleneck users with large gradient norms and bad channel conditions. Furthermore, we can explore linear compression techniques inspired by compressed sensing method to recover the aggregated model parameter vector at the server side.

In the digital domain, the impact of scheduling lies in the fact that with more devices participating in model uploading, less frequency/time resources are allocated to each user, which results in higher compression loss that may deteriorate the accuracy of the aggregated model. With OtA aggregation, scheduling is also important because all the agents apply a common scaling factor in the pre-processing step such that the transmitted signals satisfy the power constraints at all agents. This scaling factor affects the variance of the effective noise after post-processing at the receiver side. From the perspective of minimizing the reconstruction error, it might be beneficial to drop several bottleneck users with large gradient norms and bad channel conditions. Furthermore, we can explore linear compression techniques inspired by compressed sensing method to recover the aggregated model parameter vector at the server side.

**Temporal Statistics and Adaptive Learning**

For distributed ML over wireless networks, the temporal evolution of the generated model updates (gradient vectors) and the communication link conditions is another fresh aspect that makes the OtA problem more interesting.

For gradient vectors in FL, there is some evidence of a predictable structure [6], in that the amplitude of the different gradient elements changes slowly over time. This inherently suggests that the impact of communication resources on the system performance is not equal over time. Additionally, some recent work on FL has shown that the convergence in later iterations is more sensitive to large compression loss than the earlier ones. This observation has motivated the consideration of adapting communication resources along the iterations such that the aggregated model is received with higher
accuracy during the later iterations. This fact was exploited in [7] to derive an adaptive gradient compression mechanism that maximizes the optimization progress (reduction in expected loss function value) per communication resource (e.g., channel or power) used in every iteration. Similar ideas can be explored for O TA FL, where energy allocation and scheduling decisions can be adapted along the iterations to achieve faster convergence under a total energy budget for the entire training process. In this case, the channel statistics and data statistics must be jointly accounted for when making transmission decisions. The development of adaptive scheduling and resource allocation policies for iterative learning algorithms with O TA computation is a challenging open problem.

**OTA Computation With Multiple-Antenna Technology**

Most published research on O TA FL considers the case with single-antenna agents and servers. Given the enormous success of multiple-antenna technology in cellular communications, especially massive MIMO, it is conceivable that the use of multiple antennas would greatly benefit also O TA FL setups. A crucial aspect of O TA FL is that the agents need to know, on a per-antenna basis, the physical channel response in order to pre-compensate for the channel phase-rotation and amplitude attenuation, before transmission. This is required such that the signals from all agents can be added up constructively, inphase at the server. But with multiple antennas at the server, there will be no universal phase-rotation per agent antenna that works for all server antennas.

Two basic cases can be distinguished. Either the agents can obtain knowledge of the per-antenna channel responses (including their phase), or they cannot. If these channel responses are available, then the agents can optimize their transmission, which includes selecting beam patterns (in case of multiple agent antennas) and phase-rotations, such that the weighted sum of the transmitted gradient updates can be inferred from the signals collectively received at the server antennas. In contrast, if the channel responses are unknown, then the agents cannot meaningfully pre-compensate, or even beamform, their transmissions to the server and one has to either resort to “blind” O TA aggregation techniques [8], or to spatial multiplexing of the gradients using a conventional multiple-antenna decoding algorithm such as zero-forcing. The latter option allows for individual sparsification and compression of the gradients, and it was shown in [9] that this approach is superior as long as the number of antennas across all agents is smaller than at the server and the channel coherence allows for the transmission of individual uplink reference signals (which are required to separate the spatially multiplexed streams).

The design of optimal transmission schemes for multi-antenna agents and servers remains an open research problem, whose solution entails an accurate modeling of the effects of channel coherence, channel estimation errors, gradient compression, and calibration - which eventually determines to what extent the channel response phases can be obtained by the agents. It also requires the use of appropriate performance metrics, that go beyond the metrics conventionally used in the design of multiple-antenna communication systems.

**Other Metrics Than the MSE of Aggregated Data**

Many existing research on O TA FL uses the MSE of the aggregated model updates as the performance metric to optimize the communication (scheduling and power control) design. From this perspective, the formulation of the optimization problem becomes independent of the ML setting. There is no clear evidence that minimizing the MSE (without optimal rescaling) is equivalent to optimizing the learning performance. From the learning perspective, the direction of the aggregated gradient vector matters more than the magnitude. The impact of O TA aggregation error on the system performance cannot be simply measured by the squared norm of the difference between transmitted and received signal vectors.

Additionally, it is shown in [10] that the channel noise does not have a significant impact on the convergence performance if the channel noise does not dominate the noise in stochastic gradient vectors caused by random data sampling. Based on this remark, we expect that the effect of O TA estimation error on the learning performance will depend on the level of noise in the computation of stochastic gradients. Another concern with using the MSE as the optimization objective is the impact of scheduling. As briefly explained in Sec. IV-A, minimizing the MSE of aggregated data will motivate us to drop many bottleneck agents, which in turn increases the bias of the aggregated model, especially under the non-IID training data setting. In Table 1, we show an example of the accuracy-MSE tradeoff in O TA FL. This result suggests that the “plain” MSE (without proper re-scaling) of the aggregated signals from the set of scheduled users does not fully reflect how close the aggregated gradient vector is to the true gradient. An efficient scheduling design needs to balance the bias caused by scheduling users with heterogeneous data and the estimation error of O TA aggregation under the power constraints.

The same consideration can be extended to many other applications of O TA in distributed systems, such as the ones mentioned in Section III. Finding meaningful performance metrics for O TA computation in different distributed systems remains an important open question.

**OTA Computation for Fully Decentralized Systems**

With fully decentralized ML, the system only relies on local computations and information exchange between locally connected nodes to solve a decentralized optimization problem. Although several conceptual algorithms have been proposed for such systems, such as distributed sub-gradient method and distributed CoCoA, using O TA to perform such computations has rarely been investigated. The convergence speed of such algorithms is affected by the connectivity of the network nodes over time [11], which in turn is affected by how the scheduling of O TA nodes is performed in every communication round. This raises new challenges in terms of joint computation-communication methods, stability, and optimality of the algorithms.

The first challenge is the communication scheduling problem. Here, as opposed to the usual
In general, distributed learning systems can be vulnerable to different types of attacks, depending on which level of information is shared among the agents in the system [13]. In this article, we focus on OTa FL with transmissions of shared training model parameters over wireless links. The specific features of aggregating model updates through OTa computation generate new perspectives of potential threats and possible defense methods that are unique in OTa systems.

### Privacy and Security

One of the main advantages of FL is to protect data privacy by only sharing model parameter updates instead of the raw training data with the central server. However, FL can still be vulnerable to privacy and security attacks, such as membership inference attack and model poisoning attack. As compared to FL with digital transmission of model updates, one distinct feature of OTa FL is that channel noise can be exploited as a natural random perturbation function for improving differential privacy [14]. Also, the server only receives the aggregated signals from the agents without being able to decode each of the data stream, which makes it difficult to infer information about the local data samples at each agent.

Nonetheless, existing methods for improving differential privacy are mostly built on the idea of gradient obfuscation (e.g., adding random perturbations to the gradient vectors), which leads to an accuracy-privacy tradeoff since the added perturbations in the aggregated gradient can compromise the learning performance. One possibility is to apply spatially correlated perturbations across different agents such that the aggregated perturbations add up to zero at the server side [15]. In this direction, many other ideas can be developed, such as adaptive privacy budget allocation and long-term optimization of privacy-accuracy tradeoffs.

### Robustness and Reliability

Distributed SGD algorithms are generally vulnerable to Byzantine attacks, where the goal of attack is to prevent the system from converging. Conventional gradient aggregation techniques that are Byzantine fault-tolerant, typically rely on distance-based clustering and truncation schemes to eliminate the influence of outliers that may be Byzantine attackers. With OTa aggregation, in contrast to with digital transmission schemes, the different data streams are inseparable. Because of the inherent anonymity of OTa, it is in principle easy for an adversary to send malicious messages (e.g., false gradients) and harm the computations. Any random perturbation signal can affect directly possible for OTa.

In addition to model poisoning, it is well-known that carefully designed adversarial examples can degrade the trained model. These are inputs that

| Scheduled users | 50  | 40  | 30  | 20  | 10  |
|-----------------|-----|-----|-----|-----|-----|
| Test accuracy   | 93% | 71.8% | 56.6% | 42.8% | 27.2% |
| MSE             | 275e-14 | 3.72e-16 | 754e-17 | 786e-18 | 1.6e-18 |

**TABLE I.** Test accuracy (after 100 communication rounds) vs. MSE of aggregated model updates for different numbers of scheduled users. The total number of users is 30. The entire training dataset (MNIST) is equally divided among the users but each user only contains data samples with at most 2 digits (non-IID data setting). The scheduling decision is made by minimizing the MSE of the aggregated model updates from the set of scheduled users.

“many-to-one” topology, fully decentralized learning requires “many-to-many” communication over unreliable wireless links with limited range, where each node needs to send and aggregate information at the same time. Thus, the communication protocol for information broadcasting, reception and fusion will play a major role in the performance (optimality and convergence) and scalability of ML algorithms. Should the protocol be synchronous where model updating occurs after all agents receive information from their neighbors, or asynchronous as in gossip algorithms? There are many open questions when considering the broadcast nature and superposition of signals in OTa channels. Randomized communication protocols that combine features of distributed medium access, spatial reuse of communication resources and compressed communication could be particularly attractive to develop.

Another challenge is the “precoding problem.” Normally, an agent adapts the phase and amplitude of its message before transmission so that the OTa channel returns the desired computation at a specific receiver. But in a fully decentralized system, the OTa channel must simultaneously return many desired computation results at different nodes despite the fact that we can control only one precoder at each user. How to select one precoder that can return several OTa computations at the same time? One possible solution is to combine the degrees of freedom in space (joint precoding and decoding at different nodes) and in time (coding over multiple slots) [12]. Other designs can be developed by following the methods of distributed estimation theory. In such an estimation setup, the nodes of a wireless network cooperate to perform a distributed estimation process, where each node broadcasts its local estimate to the neighbors, which in their turn, use such received estimate to update their local estimates. The main challenge in this case is that the channel quality and dynamics in OTa channels can threaten the stability and convergence of the ML algorithms if they would merely follow the distributed estimation algorithms.

**Privacy, Security and Robustness of OTa FL**

In general, distributed learning systems can be vulnerable to different types of attacks, depending on which level of information is shared among the agents in the system [13]. In this article, we focus on OTa FL with transmissions of shared training model parameters over wireless links. The specific features of aggregating model updates through OTa computation generate new perspectives of potential threats and possible defense methods that are unique in OTa systems.
are virtually indistinguishable from the original input to a human, but on which the trained model shows drastically different results. To circumvent that, the prevailing approach is adversarial training, which is often formulated as a minimax optimization problem, where the adversarial examples and the model weights are alternatively updated. Such learning typically requires many more epochs, and hence, more communication rounds. Because of the latter, the advantages of applying OTA computation instead of digital communication design are obvious. However, the robustness of adversarial training with OTA computation remains to be investigated.

In general, very little research has been performed on robust model update aggregation in OTA FL. How to achieve robust and reliable computation will be a main challenge and open question for OTA-assisted distributed systems.

CONCLUSION
As we enter the 6G era, future wireless systems need to support efficient information processing and data aggregation in large-scale distributed intelligent systems. OTA computation shows a clear advantage in resource efficiency as compared to traditional digital communication with separated source and channel coding schemes. Though the concept of distributed computation of functions over multiple access channels is not new, there are still many new challenges and open questions that are worth investigating. In this article, we provided an overview of the theory and applications of OTA computation, and highlighted the importance of the following aspects:

- The statistical heterogeneity of users (both in data distribution and channel condition) and temporal evolution of model updates will have a strong impact on communication efficiency. This is often overlooked in the existing literature on OTA FL.
- Finding appropriate performance metrics for different types of distributed systems with OTA computation is crucial for optimizing the resource efficiency.
- Many open questions remain in the communication and computation design for OTA in fully decentralized systems with local information exchange.
- OTA FL is vulnerable to model poisoning attacks since any perturbation signal can directly affect the aggregated model update. The robustness and reliability of OTA systems will be a great challenge for its application in real-world networks.

ACKNOWLEDGMENT
The authors would like to give acknowledgment to Chung-Hsuan Hu, doctoral student at Linköping University (Sweden), and Jeremy Weill, master student at EPFL (Switzerland), for their contributions with the simulation results presented in the article.

REFERENCES
[1] G. Zhu et al., “Over-the-air computing for wireless data aggregation in massive IoT,” IEEE Wireless Commun., vol. 28, no. 4, pp. 57–65, Aug. 2021.
[2] M. Goldenbaum, H. Boche, and S. Stanicak, “Nomographic functions: Efficient computation in clustered Gaussian sensor networks,” IEEE Trans. Wireless Commun., vol. 14, no. 4, pp. 2093–2105, Apr. 2015.
[3] P. Park, P. D. Marco, and C. Fischione, “Optimized over-the-air computation for wireless control systems,” IEEE Commun. Lett., vol. 26, no. 2, pp. 424–428, Feb. 2022.
[4] S. Kar and J. M. Moura, “Consensus + innovations distributed inference over networks: Cooperation and sensing in networks,” IEEE Signal Process. Mag., vol. 30, no. 3, pp. 99–109, May 2013.
[5] E. Ozfatura, K. Ozfatura, and D. Gunduz, “Time-correlated sparsification for communication-efficient federated learning,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2021, pp. 461–465.
[6] D. Fan, X. Yuan, and Y.-J. A. Zhang, “Temporal-structure-assisted gradient aggregation for over-the-air federated edge learning,” IEEE J. Sel. Areas Commun., vol. 39, no. 12, pp. 3757–3771, Dec. 2021.
[7] S. Khiriat et al., “A flexible framework for communication-efficient machine learning,” in Proc. AAAI Conf. Artif. Intell., 2021, vol. 35, no. 9, pp. 8101–8109.
[8] M. M. Amiri et al., “Blind federated edge learning,” IEEE Trans. Wireless Commun., vol. 20, no. 8, pp. 5129–5143, Aug. 2021.
[9] E. Becirovic, Z. Chen, and E. G. Larsson, “Optimal MIMO combining for blind federated edge learning with gradient sparsification,” in Proc. IEEE 23rd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Jul. 2022, pp. 1–5.
[10] X. Wei and C. Shen, “Federated learning over noisy channels: Convergence analysis and design examples,” IEEE Trans. Cogn. Commun. Netw., vol. 8, no. 2, pp. 1253–1268, Jun. 2022.
[11] A. Nedić, A. Olshhevsky, and M. G. Rabbat, “Network topology and communication-computation tradeoffs in decentralized optimization,” Proc. IEEE, vol. 106, no. 5, pp. 953–976, May 2018.
[12] Z. Chen and Y. Malitsky, “Over-the-air computation with multiple receivers: A space-time approach,” in IEEE Wireless Commun. Lett., vol. 12, no. 8, pp. 1399–1403, Aug. 2023.
[13] C. Ma et al., “Trusted AI in multi-agent systems: An overview of privacy and security for distributed learning,” 2022, arXiv:2202.09027.
[14] D. Liu and O. Simeone, “Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control,” IEEE J. Sel. Areas Commun., vol. 39, no. 1, pp. 170–185, Jan. 2021.
[15] J. Liao, Z. Chen, and E. G. Larsson, “Over-the-air federated learning with privacy protection via correlated additive perturbations,” in Proc. 58th Annu. Allerton Conf. Commun., Control, Comput. (Allerton), Sep. 2020, pp. 1–8.

BIографIES
ZHENG CHEN (zheng.chen@liu.se) is an Associate Professor at Linköping University, Sweden. Her research focuses on distributed information processing and machine learning over wireless networks. She was the recipient of the 2020 IEEE Communications Society Young Author Best Paper Award.

ERIK G. LARSSON (Fellow, IEEE) (erik.g.larsson@liu.se) is a Professor at Linköping University, Sweden. He is highly cited according to ISI Web of Science, and received many awards for his work.

CARLO FISCHIONE (carlo@kth.se) is a Professor at the KTH Royal Institute of Technology. He is the Chair of IEEE Communication Society Emerging Technology Initiative on Machine Learning for Communications and is a Distinguished Lecturer of IEEE Communications Society. He received a number of awards, including the IEEE “S. O. Rice” Award of 2018.

MIKAIL JOHANSSON (mikael@kth.se) is a Professor at KTH, Stockholm, Sweden. He has coauthored two books and over 200 papers, several of which are highly cited and have received recognition in terms of best paper awards. His research interests include large-scale and distributed optimization, autonomous decision making and control, and machine learning.

YURA MALITSKY (Yuriy Malitskiy)(yuriy.malitsky@liu.se) is an Assistant Professor in optimization at the University of Vienna, Austria. His research focuses on algorithms for continuous optimization and machine learning.