FEDREC: FEDERATED LEARNING OF UNIVERSAL RECEIVERS OVER FADING CHANNELS

Mahdi Boloursaz Mashhadi, Nir Shlezinger, Yonina C. Eldar and Deniz Gündüz

ABSTRACT

Wireless communications is often subject to channel fading. Various statistical models have been proposed to capture the inherent randomness in fading, and conventional model-based receiver designs rely on accurate knowledge of this underlying distribution, which, in practice, may be complex and intractable. In this work, we propose a neural network-based symbol detection technique for downlink fading channels, which is based on the maximum a-posteriori probability (MAP) detector. To enable training on a diverse ensemble of fading realizations, we propose a federated training scheme, in which multiple users collaborate to jointly learn a universal data-driven detector, hence the name FedRec. The performance of the resulting receiver is shown to approach the MAP performance in diverse channel conditions without requiring knowledge of the fading statistics, while inducing a substantially reduced communication overhead in its training procedure compared to centralized training.

1. INTRODUCTION

Fading in wireless communications encapsulates the fact that the relationship between the transmitted signal and the received one is determined by the propagation of electromagnetic waves, which is typically dynamic and subject to different forms of randomness induced by the environment. Various distributions have been proposed to represent the statistical behavior of fading channels, including the Rayleigh, Rice, and Nakagami-m models [1], where each approximates the propagation profile in different settings [2].

The inherent randomness of fading channels makes symbol detection a challenging task. The common strategy is to periodically transmit a-priori known pilot signals for the receiver to estimate the channel, which in turn is utilized for detection [3]. The main drawback of this approach is that pilots must be transmitted anew each time the channel changes, i.e., on each coherence duration, inducing notable overhead in rapidly-changing channels. Alternatively in fast fading conditions, one can utilize a single detection rule for all channel conditions, which accounts for its statistical model [4]. However, such model-based detection relies on the knowledge of the fading distribution, and tends to be inaccurate when the assumed distribution does not faithfully capture the real statistical propagation profile, or in the presence of a model mismatch.

An alternative strategy, which does not require the knowledge of the underlying statistical model, is to learn the detection mapping from data. In particular, neural networks (NNs) have demonstrated unprecedented success over recent years in learning complex mappings in a data-driven fashion [5]. Consequently, a multitude of NN-based receiver architectures that can operate without the prior knowledge of the underlying statistical model have been proposed, see, e.g., [6–11]. NN-based receivers require labeled data to learn their mapping. If one has prior knowledge of the channel conditions and can generate such data artificially, the NN can be trained offline. However, this may not be the case in many practical scenarios, where labeled data must be obtained from pilot transmissions.

When the channel conditions change rapidly, NN-based receivers must be frequently retrained with new pilot signals, inducing significant computation and communication overhead. NN-based receivers can track dynamic channel conditions by periodic online training combined with methods to reduce the training complexity as in [12–16]. Alternatively, one can train a single NN that would work for a broad range of channel conditions, by learning a universal rule based on the fading distribution rather than its specific realization [17, 18]. Nonetheless, for a NN to learn the subtleties of a fading distribution, which may be complex and intractable, the training data must contain a sufficiently large number of channel realizations. This may be difficult to achieve even with long pilot sequences, motivating the design of universal NN-based detectors for fading channels with limited pilots.

In this work, we propose FedRec, which is a data-driven universal symbol detection scheme for multi-user downlink fading channels, designed to learn its mappings from a limited amount of pilots. FedRec is comprised of two algorithmic components: The first is the NN-based symbol detection architecture, which uses sufficient statistics from the maximum a-posteriori probability (MAP) symbol detection rule as input features. This allows FedRec to utilize a relatively compact NN, which can be trained with a smaller number of samples. The second component is the training mechanism, which exploits the fact that, in a wireless network of many users, while each user observes only a limited number of channel realizations, the realizations observed by the overall network are expected to be sufficiently diverse. FedRec builds upon this insight to have the users collaborate for training via federated learning (FL) [19–21], allowing to train a single NN over a diverse dataset without additional pilots, at the cost of several iterations of parameter exchanges with the base station (BS). Our numerical results show that FedRec yields an accurate symbol detector, with a performance approaching that of a MAP detector, and outperforms the model-based approach in the presence of inaccurate knowledge of the fading distribution. Moreover, FedRec induces substantially less communication overhead compared to learning a NN-based symbol detector in a centralized fashion.

The rest of this paper is organized as follows: Section 2 presents the system model and the problem formulation. Section 3 details the proposed FedRec receiver. Numerical examples are presented in Section 4. Finally, Section 5 concludes the paper. FedRec codes are available at: https://github.com/MahdiBoloursazMashhadi/FedRec

D. Gündüz received funding from the European Research Council (ERC) through project BEACON under grant No. 677854. Y. C. Eldar received funding from the European Union's Horizon 2020 research and innovation program under grant No. 646804-ERC-COG-BNYQ, and from the Israel Science Foundation under grant No. 0100101. M. B. Mashhadi and D. Gündüz are with the Dept. of EE, Imperial College, London, UK (email: {m.boloursaz-mashhadi, d.gunduz}@imperial.ac.uk). N. Shlezinger is with the School of ECE, Ben-Gurion University of the Negev, Be'er-Sheva, Israel (e-mail: nirshl@bgu.ac.il). Y. C. Eldar is with the Faculty of Math and CS, Weizmann Institute, Rehovot, Israel (e-mail: yonina@weizmann.ac.il).

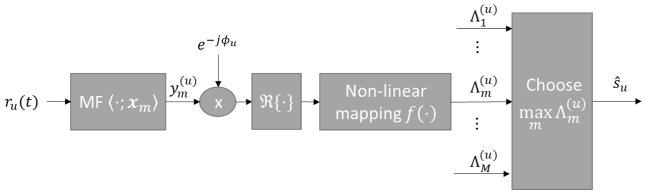

Fig. 1: MAP symbol detector block diagram.

2. SYSTEM MODEL

We consider a downlink communication scenario, where a BS serves U users indexed by $u \in \mathcal{U} \triangleq \{1,\ldots,U\}$. Although we focus on single-antenna terminals in this paper, our approach can be extended to multiple-antenna terminals. Letting $x(t) \in \mathcal{C}$ be the baseband continuous-time (CT) channel input transmitted by the BS at time instance t, the corresponding channel output at the uth user is

$$r_u(t) = h_u(t)x(t) + w_u(t), \quad u \in \mathcal{U}, \tag{1}$$

where $\{w_u(t)\}$ are independent identically distributed (i.i.d.) Gaussian noise signals with unit power spectral density (PSD), while $\{h_u(t)\}$ are i.i.d. flat fading coefficients, following a common distribution $p_h(\cdot)$. Various different models exist for $p_h(\cdot)$, which approximate the statistical behavior under different channel conditions [22, Ch. 2], but we do not assume prior knowledge of $p_h(\cdot)$.

Downlink communication is typically comprised of pilot and data transmission. During downlink data transmission to user u commencing at time instance t_d, the BS encodes the message s_u, uniformly distributed over $\mathcal{M} \triangleq \{1,\ldots,M\}$, into a signal of temporal duration of T_s seconds, denoted by $x_{s_u}(t)$, $t \in t_\mathrm{d} + [0, T_\mathrm{s})$. Each user $u \in \mathcal{U}$ uses its channel output $r_u(t)$, obtained via (1), to recover s_u. During pilot transmission commencing at $t = 0$, the BS broadcasts a sequence of N_T a-priori known pilot symbols, denoted by $\{s_i^p\}_{i=1}^{N_T}$, over a duration of $N_T \cdot T_\mathrm{s}$ seconds.

We focus on regimes in which the number of pilot symbols N_T is relatively small; and hence, it is unlikely to span a sufficient amount of different realizations of the fading coefficient at a single user. While the number of pilots is limited, we allow the users to collaborate and share their detection mappings, in order to jointly learn a unified symbol detector, i.e., one that is universally applicable not only to the users participating in the training, but also to any arbitrary user, by learning a broad fading distribution from the pilots received at all the users and the corresponding messages, i.e., $\{\mathcal{D}_u\}_{u=1}^U$, where $\mathcal{D}_u = \{s_i^p, \mathcal{R}_i^u\}_{i=1}^{N_T}$, with $\mathcal{R}_i^{(u)} \triangleq \{r_u(t)|t \in [(i-1)T_\mathrm{s}, iT_\mathrm{s})\}$.

The resulting symbol detector at an arbitrary user u' recovers the message $s_{u'}$ from the received $r_{u'}(t)$, $t \in t_\mathrm{d} + [0, T_\mathrm{s})$, assuming only the knowledge of the channel input-output relationship (1), but not the fading distribution $p_h(\cdot)$. Our proposed NN-based symbol detection scheme, which combines model knowledge of (1) with data-driven tools to utilize $\{\mathcal{D}_u\}_{u=1}^U$ in a collaborative fashion to train such a universal receiver, is detailed in the following section.

3. FEDERATED RECEIVER

The need for a universal symbol detector applicable to a diverse range of channel fading statistics without the exact knowledge of the underlying distribution motivates using NNs, which were empirically shown to operate reliably in complex unknown statistical environments [5]. However, since NNs require large volumes of diverse training data, having each user train a local NN based on its limited pilot observations is expected to yield a less reliable model with limited generalization performance. To overcome this, we design FedRec based on the following guidelines: (i) The fact that the channel is modeled via (1) is exploited as partial domain knowledge for feature extraction, inspired by the MAP rule for such channels. This approach facilitates utilizing compact NNs (avoiding feature extraction layers), which are trainable using relatively small datasets. (ii) While each local training dataset $\mathcal{D}_u$ encapsulates a relatively small number of fading channel realizations, the diversity among these sets at different users is exploited to obtain a unified model for all the users by training in a federated manner.

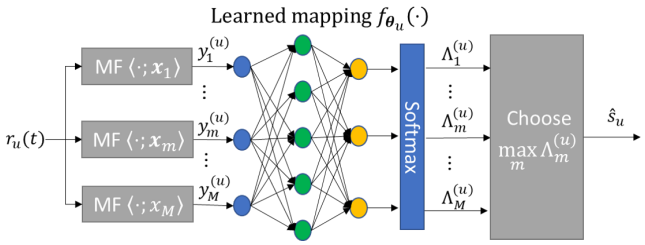

Fig. 2: The proposed symbol detector architecture.

3.1. Model-Based Symbol Detection for Fading Channels

Here, we briefly recall the model-based MAP symbol detector, which requires knowledge of $p_h(\cdot)$. We focus on scenarios in which the fading coefficient $h_u(t)$ takes a single realization during the transmission of each message, and that its phase, denoted by ϕ_u, is known. The following derivation is based on [22, Ch. 7.2].

Let $\mathcal{R}^{(u)} \triangleq \{r_u(t)\}_{t=t_\mathrm{d}}^{t_\mathrm{d}+T_\mathrm{s}}$. Since the message s_u is uniformly distributed, the MAP rule at the uth user is given by

$$\hat{s}_u^{\mathrm{map}} = \arg\max_{m\in\mathcal{M}} \Pr\left(\mathcal{R}^{(u)}|s_u = m\right). \tag{2}$$

By defining $y_m^{(u)} \triangleq \int_{t=t_\mathrm{d}}^{t_\mathrm{d}+T_\mathrm{s}} r_u(t)x_m(t)dt \triangleq \langle r_u; x_m\rangle$, and similarly, $e_m \triangleq \langle x_m; x_m\rangle$, the conditional distribution in (2) becomes

$$\Pr\left(\mathcal{R}^{(u)}|s_u=m\right) = c\int_0^\infty e^{\alpha\Re\{e^{-j\phi_u}y_m^{(u)}\}-\alpha^2 e_m} p_{|h|}(\alpha)d\alpha, \tag{3}$$

where $c > 0$ is a constant that does not depend on m and $\mathcal{R}^{(u)}$.

The complex structure of the conditional distribution in (3) for general $p_{|h|}(\cdot)$ makes evaluating (3) a challenging task. If $p_{|h|}(\cdot)$ follows a simple Rayleigh distribution with scale parameter σ_u, the MAP decision criteria (3) reduces to maximizing $\Lambda_m^{(u)}$ given by

$$\Lambda_m^{(u)} = \ln\{1+\sqrt{\pi}\mu_m e^{\frac{\mu_m^2}{4}}[1-Q(\mu_m/\sqrt{2})]\} - \ln(1+\gamma_m), \tag{4}$$

where $\gamma_m = 2\sigma_u^2 e_m$, $\mu_m = \sqrt{\gamma_m/[e_m(1+\gamma_m)]}\cdot\Re\{e^{-j\phi_u}y_m^{(u)}\}$ and Q denotes the Gaussian Q-function [22, Ch. 7.2.1].

Despite its complex form, (3) reveals that, regardless of the fading distribution, MAP detection over any fading channel conforming to (1), is comprised of the following steps: First, the observations $\mathcal{R}^{(u)}$ are processed by a set of matched filters (MFs) to produce $\{y_m^{(u)}\}_{m=1}^M$; then, when the phase information is given, the MF outputs are phase-shifted accordingly and their real part is taken; this quantity is processed by a non-linear mapping $f(\cdot)$ that depends on the fading distribution, encapsulating the integral in (3), to produce $\Lambda_m^{(u)} \propto \Pr\left(\mathcal{R}^{(u)}|s=m\right)$. This procedure is illustrated in Fig. 1.

3.2. Data-Driven Symbol Detector Architecture

We next present a NN architecture which learns to recover s_u from $\mathcal{R}^{(u)}$. Recall that NNs process inputs that can be represented as vectors. Here, the input is a CT signal $r_u(t)$, defined over the uncountable set $t \in [t_\mathrm{d}, t_\mathrm{d} + T_\mathrm{s})$. The intuitive approach to design the network is therefore to uniformly sample $r_u(t)$ via, e.g., Nyquist rate sampling. Taking a large number of samples is expected to facilitate handling the presence of noise and unknown channel [23], while resulting in processing high dimensional inputs, which in turn typically involves using highly-parametrized NNs.

Nonetheless, the model-based MAP detector in Fig. 1 reveals that the MF outputs $\{y_m^{(u)}\}_{m=1}^M$ constitute a sufficient statistics for identifying the message s_u. As a result, we exploit this domain knowledge, and design a classification network whose inputs are the matched filter features $\{y_m^{(u)}\}_{m=1}^M$. This significantly simplifies the NN architecture in comparison with feeding the Nyquist samples of the received signal as input. The proposed NN architecture avoids additional layers that would be required for feature extraction from Nyquist samples of the received signal and directly uses the matched filter outputs as features for classification.

As matched filter outputs take complex values, the NN input is a $2M \times 1$ dimensional vector, which feeds a fully connected network. As s_u can take one of M different values, we use a softmax output layer with M possible labels. Letting $\boldsymbol{\theta}_u$ be the NN parameters of the uth user, and $f_{\boldsymbol{\theta}_u} : \mathcal{C}^M \mapsto [0,1]^M$ be its mapping, then the NN prediction can be written as

$$\hat{s}_u = \underset{m \in \mathcal{M}}{\arg\max}\, f_{\boldsymbol{\theta}_u, m}\left(\{y_m^{(u)}\}\right), \tag{5}$$

where $f_{\boldsymbol{\theta}_u, m}(\cdot)$ is the mth output. The NN is illustrated in Fig. 2.

Note that while the NN architecture in Fig. 2 is comprised of M MF components, in many transmission schemes the outputs of some of the MFs can be obtained as linear combinations of the remaining features. In particular, in modulation schemes where the modulation order is less than M, (e.g., 2 for QAM), we use fewer MF features (determined by the modulation order) to simplify the NN.

3.3. FedRec Training

Here, we describe how the NN is trained. In particular, we consider a training procedure in which the users collaborate in a federated manner, exploiting the diversity in the observed channels. The resulting learned symbol detector is referred to as *FedRec*.

As FedRec is comprised of a neural classifier, we use the empirical cross entropy loss for training. The resulting loss is

$$\mathcal{L}_u(\boldsymbol{\theta}, \mathcal{D}) = -\frac{1}{|\mathcal{D}|} \sum_{\{s_i, \mathcal{R}_i\} \in \mathcal{D}} \log f_{\boldsymbol{\theta}, s_i}\left(\{\langle \mathcal{R}_i; x_m \rangle\}_{m=1}^M\right). \tag{6}$$

When the training set $\mathcal{D}_u$ captures a sufficiently large number of realizations of the fading channel h_u, then each user should be able to tune its local NN parameters to carry out accurate detection. This can be achieved via conventional training mechanisms, e.g., stochastic gradient descent (SGD), for which $\boldsymbol{\theta}_u$ is iteratively updated via

$$\boldsymbol{\theta}_u^{(n)} = \boldsymbol{\theta}_u^{(n-1)} - \eta_n \nabla \mathcal{L}_u(\boldsymbol{\theta}_u^{(n-1)}, \{s_{i_n}, \mathcal{R}_{i_n}\}), \tag{7}$$

where n is the iteration index, $\eta_n > 0$ is the step-size, and i_n is an index drawn uniformly in an i.i.d. fashion from $\{1, \ldots, N_T\}$.

Nonetheless, when the channel coefficient takes a limited number of realizations in each interval of $N_T T_\mathrm{s}$ seconds, the local dataset may not suffice to capture the subtleties of a universal fading distribution. In such cases, the trained NN is likely to be highly biased towards its observed channel conditions, and may not operate reliably for future realizations of the channel in case the fading statistics change over time. To tackle this challenge, we exploit the fact that while each user may observe a limited amount of channel realizations, the overall set of different realizations observed by all the users in a cell is likely to be sufficiently diverse to train the NN accurately.

Based on the above insight, we propose to jointly train a single instance of the NN shared by all the users via FL [19]. Such distributed learning orchestrated by the BS requires the users to periodically exchange and synchronize their local parameters, possibly by utilizing low-rate transmissions, to train a reliable universal symbol detector. In particular, the training mechanism, based on the local SGD algorithm [24], consists of τ SGD iterations as in (7), carried out by each user locally, after which the BS averages the trained updates into a global parameter vector $\boldsymbol{\theta}$, which is distributed to the users. This training mechanism is summarized as Algorithm 1.

Algorithm 1: FedRec Training

Init: Initial parameters $\boldsymbol{\theta}_u^{(0)} = \boldsymbol{\theta}^{(0)}, \forall u \in \mathcal{U}$.
for *each* $n = 1, 2, \ldots$ **do**
  Each user u sets $\boldsymbol{\theta}_u^{(n)}$ via (7) ;
  if n *is an integer multiple of* τ **then**
    Each user u sends $\boldsymbol{g}_u^{(n)} = \boldsymbol{\theta}_u^{(n)} - \boldsymbol{\theta}_u^{(n-\tau)}$ to BS;
    BS computes $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta}^{(n-\tau)} + \frac{1}{U}\sum \boldsymbol{g}_u^{(n)}$;
    BS distributes $\boldsymbol{\theta}^{(n)}$ s.t. $\boldsymbol{\theta}_u^{(n)} = \boldsymbol{\theta}^{(n)}, \forall u \in \mathcal{U}$;
  end
end
Output: Trained FedRec $\boldsymbol{\theta}^{(n)}$ shared among all users.

3.4. Three-Phase Operation

FedRec requires the users to exchange messages with the BS during training, i.e., before the NN is tuned. Consequently, its application involves a three phase scheme: (i) data collection, (ii) federated training, and (iii) pilot-free communication using FedRec.

During phase (i), wireless users in the coverage area of the BS utilize a conventional pilot-based scheme to communicate with the BS. However, they keep collecting the noisy received pilot signals $\mathcal{R}_i^u$ and the corresponding labels s_i^p to form local datasets $\mathcal{D}_u = \{s_i^p, \mathcal{R}_i^u\}_{i=1}^{N_T}$. To communicate with the BS during this phase, the users perform coherent symbol detection utilizing channel estimates from the received pilots.

During phase (ii), the wireless users utilize their local datasets $\{\mathcal{D}_u\}_{u=1}^U$ to collaboratively train FedRec in a federated fashion. They continue to utilize pilot-aided coherent communications to exchange model updates with the BS for federated training.

Finally, during phase (iii), users utilize the trained FedRec receiver for pilot-free communication. As FedRec is trained with data from many users with diverse channel conditions, it does not rely on the knowledge of the exact channel realization; and hence, eliminates the need for periodic pilot transmissions. During this phase, the BS can use a low rate control channel to transmit the trained FedRec receiver to any new user entering its coverage area.

4. NUMERICAL EVALUATIONS

In this section we compare FedRec with model-based detection and learning-based schemes for Rayleigh fading channels. We consider two cases with i.i.d. and non-i.i.d. fading channels across the users. In the i.i.d. case, the coefficients $h_u(t)$ are generated from a Rayleigh distribution with unit scale parameter, and a new

Table 1: BER comparison for various symbol detectors, $U = 5$.

ρ (dB)		5	7.5	10	12.5
non-iid	NL	0.0677	0.0574	0.0490	0.0438
	CL	0.0661	0.0544	**0.0439**	**0.0382**
	FedRec	**0.0649**	**0.0530**	**0.0439**	**0.0382**
	MD	0.0663	0.0544	0.0445	0.0392
	MAP	**0.0647**	**0.0526**	**0.0437**	**0.0382**
iid	NL	0.0577	0.0458	0.0377	0.0325
	CL	0.0573	0.0453	0.0369	**0.0307**
	FedRec	**0.0561**	**0.0444**	**0.0361**	0.0308
	MD	0.0558	0.0439	0.0357	0.0303
	MAP	**0.0557**	**0.0439**	**0.0357**	**0.0302**

realization is generated on each T_s time instances. We also consider a non-i.i.d. case, where the scale parameter for different users is not identical, e.g., due to different statistics of the local environment for each user, such that for each user $u \in \mathcal{U}$, the scale parameter σ_u is randomized from a uniform distribution over the range $[0.5, 1.5]$.

We use 16QAM modulations, i.e., $M = 16$, with average transmit power per bit ρ, representing the signal-to-noise ratio (SNR). For the learning-based methods, the training and test datasets are comprised of $|\mathcal{D}_{train}| = 2 \cdot 10^4$ and $|\mathcal{D}_{test}| = 10^7$ symbols, respectively. Training data is collected by the users during phase (i) of network operation, with each user recording $N_T = |\mathcal{D}_u| = 2 \cdot 10^4/U$ 16QAM pilot symbols, which are faded according to its local fading statistics. The learning-based methods train a NN with an input layer of size 2 corresponding to the in-phase and quadrature signal components, followed by a hidden layer with 16 neurons and softmax, thus $|\boldsymbol{\theta}| = 48$. This NN architecture is trained using the Adam optimizer [25] over 25 epochs with batch-size of 20, and tested for each SNR value. We compare the following symbol detection schemes:

-Non-collaborative learning (NL): Each user trains the NN solely on its local dataset. The bit error rate (BER) is averaged over all user trained NNs using the test dataset.

-Centralized learning (CL): The users transmit their datasets to the BS over a noise-free link, and the NN is trained centrally.

-FedRec: The users follow Algorithm 1 with 5 local epochs and over 5 rounds of aggregation to train the NN collaboratively.

-Model-based detection (MD): The model-based detector uses the decision rule (4), with an estimate of the Rayleigh scale parameter, denoted by $\hat{\sigma}$, obtained from the training data. Here, each user obtains a maximum likelihood estimate of its scale parameter denoted by $\hat{\sigma}_u$ utilizing its local dataset. The overall estimate is then obtained by $\hat{\sigma}^2 = 1/U \sum_{u=1}^{U} \hat{\sigma}_u^2$ inserted in (4).

-MAP bound: The MAP receiver follows the closed form (4) for the i.i.d. case. Unlike the model-based approach which estimates the Rayleigh scale parameter from the user data, we here assume exact knowledge of the scale parameter $\sigma = 1$ and insert it into (4). For the non-i.i.d. case, however, derivation of the MAP receiver is cumbersome and we use a numerical approach to calculate the integrals in (3) and to evaluate the BER.

In Table 1, we evaluate the BER performance of these symbol detectors for various SNRs with $U = 5$ users. Among the data-driven detectors, FedRec consistently outperforms the non-collaborative scheme, while achieving similar or improved BER compared to centralized learning for both the i.i.d. and non-i.i.d cases. It is observed that the BER is significantly improved over non-collaborative learning, most notably in high SNRs when users collaborate either through FL or by exchanging their local datasets with the BS for centralized training. For the i.i.d. case, this gain is due to the increased amount of data made available for training through collaboration. For the non-i.i.d. case, the improvement is more significant as it is not only due to increased number of data samples, but also due to the diversity captured by these samples.

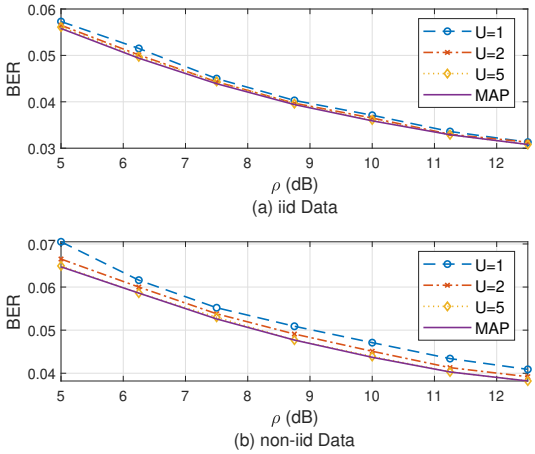

Fig. 3: FedRec BER curves versus SNR.

Table 2: Communication overhead in float32 words.

	CL		FedRec	
	UL	DL	UL	DL
U=1	40000	48	240	240
U=2	40000	48	480	240
U=5	40000	48	1200	240

For the i.i.d. case, FedRec approaches the performance of the model-based and MAP detectors. For the more realistic non-i.i.d. case, FedRec generally outperforms the model-based approach computed with the estimated scale parameter. This indicates that while the model-based approach is sensitive to uncertainty in the scale parameter, FedRec learns from data how to cope with such heterogeneous fading and performs closer to the MAP lower bound.

In Fig. 3, we compare the BER versus SNR of FedRec for different number of users $U = 1, 2, 5$. We have also added the MAP lower bound for comparison. For both i.i.d and non-i.i.d cases, FedRec BER rapidly decreases as the number of users participating in federated training grows, approaching the MAP lower bound with merely $U = 5$ users. Hence, increasing the number of users participating in federated training of FedRec not only decreases N_T, hence reducing the duration of the data collection phase (i), but also improves the BER performance.

Finally, we evaluate the overhead induced on both uplink (UL) and downlink (DL) communications in the training procedure of the collaborative data-driven schemes of FedRec and centralized learning. The number of float32 words conveyed over the UL and DL channels for FedRec and centralized learning are summarized in Table 2. For centralized training, the overhead does not depend on the number of users, and is comprised of $2 \cdot |\mathcal{D}_{train}|$ and $|\boldsymbol{\theta}|$ words on the UL and DL, respectively. For federated training, the overhead is comprised of 5 parameter exchange rounds of $U|\boldsymbol{\theta}|$ words and $|\boldsymbol{\theta}|$ words on UL and DL, respectively, both are much smaller compared to having the users transmit their received pilots to the BS. Hence, FedRec notably reduces the communication load.

5. CONCLUSION

We proposed FedRec, a data-driven symbol detector for downlink fading channels. FedRec is comprised of a NN designed based on the MAP rule for fading channels, combined with a collaborative training mechanism, which exploits the channel diversity across multiple users through federated learning. Our numerical results demonstrate that FedRec approaches the MAP performance, while achieving improved robustness to uncertainty, and significantly reduces the communication overhead compared to centralized training.

6. REFERENCES

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[2] M. Patzold, *Mobile fading channels: Modelling, analysis and simulation*. John Wiley & Sons, Inc., 2001.

[3] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[4] M. K. Simon and M.-S. Alouini, "A unified approach to the performance analysis of digital communication over generalized fading channels," *Proc. IEEE*, vol. 86, no. 9, pp. 1860–1877, 1998.

[5] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[6] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, 2017.

[7] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 648–664, 2018.

[8] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 2018.

[9] D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, "Machine learning in the air," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2184–2199, 2019.

[10] M. B. Mashhadi and D. Gündüz, "Pruning the pilots: Deep learning-based pilot design and channel estimation for MIMO-OFDM systems," *arXiv preprint arXiv:2006.11796v2*, 2020.

[11] A. Balatsoukas-Stimming and C. Studer, "Deep unfolding for communications systems: A survey and some new directions," *arXiv preprint arXiv:1906.05774*, 2019.

[12] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "ViterbiNet: A deep learning based Viterbi algorithm for symbol detection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3319–3331, 2020.

[13] N. Shlezinger, R. Fu, and Y. C. Eldar, "DeepSIC: Deep soft interference cancellation for multiuser MIMO detection," *IEEE Trans. Wireless Commun.*, 2020.

[14] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Data-driven factor graphs for deep symbol detection," in *Proc. IEEE ISIT*, 2020.

[15] S. Park, H. Jang, O. Simeone, and J. Kang, "Learning to demodulate from few pilots via offline and online meta-learning," *IEEE Trans. Signal Process.*, 2020.

[16] C.-F. Teng and Y.-L. Chen, "Syndrome enabled unsupervised learning for neural network based polar decoder and jointly optimized blind equalizer," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, 2020.

[17] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5663–5678, 2018.

[18] Y. Liao, N. Farsad, N. Shlezinger, Y. C. Eldar, and A. J. Goldsmith, "Deep neural network symbol detection for millimeter wave communications," *arXiv preprint arXiv:1907.11294*, 2019.

[19] H. B. McMahan, E. Moore, D. Ramage, and S. Hampson, "Communication-efficient learning of deep networks from decentralized data," *arXiv preprint arXiv:1602.05629*, 2016.

[20] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.

[21] D. Gunduz, D. B. Kurka, M. Jankowski, M. M. Amiri, E. Ozfatura, and S. Sreekumar, "Communicate to learn at the edge," *arXiv preprint arXiv:2009.13269*, 2020.

[22] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels*. John Wiley & Sons, 2005.

[23] N. Shlezinger, R. J. G. van Sloun, I. A. M. Hujiben, G. Tsintsadze, and Y. C. Eldar, "Learning task-based analog-to-digital conversion for MIMO receivers," in *Proc. IEEE ICASSP*, 2020.

[24] S. U. Stich, "Local SGD converges fast and communicates little," *arXiv preprint arXiv:1805.09767*, 2018.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.