Robust Symbol-Level Precoding Beyond CSI Models: A Probabilistic-Learning Based Approach

Jianjun Zhang, Christos Masouros

Department of Electronic & Electrical Engineering, University College London, London WC1E7JE, U.K.
Email: {jianjun.zhang,c.masouros}@ucl.ac.uk

***Abstract*—The use of large antenna arrays poses great difficulties in obtaining perfect channel state information (CSI) in a multi-antenna communication system, which is essential for precoding optimization. To tackle this issue, in this paper we propose a probabilistic-learning based approach (PLA), aiming at alleviating the requirement of perfect CSI. The rationale is that the existing precoding algorithms that output a single precoder are often overconfident in their abilities and estimated CSI. To avoid overconfidence, we incorporate the idea of regularization in machine learning into precoding models, so as to limit the representative abilities of the precoder models. Compared to the state-of-the-art robust precoding designs, an important advantage of PLA is that CSI uncertainty models are not required. As a specific application of PLA, we design an efficient symbol-level hybrid precoding algorithm for the millimeter wave (mmwave) communication system. Comprehensive simulation results confirm that PLA can achieve a better performance.**

***Index Terms*—Probabilistic precoding, probabilistic-learning, robust symbol-level precoding, millimeter wave communication.**

I. Introduction

Operating in the band of 30-300 GHz, mmwave communications have attracted considerable attention and stimulated surging studies [1]. Since mmwave channels suffer severe path-loss and are more likely to be sparse in the beam/angle domain, it brings new technical challenges for mmwave communication designs. To combat the large path-loss, large-scale antenna array becomes an indispensable component to obtain a large array gain. As a result, CSI is often acquired via beam training or tracking. However, the use of pencil-like beams raises the difficulty of obtaining high-precision CSI [2]–[4]. Moreover, the high cost and energy consumption of mmwave RF chains necessitate the use of the hybrid analog-digital architecture with much reduced number of RF chains, rather than the fully-digital counterpart, which motivates new approaches to design mmwave hybrid precoding [5]–[7].

Precoding, as an effective means of mitigating or exploiting the interference, is an active research area in mmwave communications. Under the assumption that perfect CSI is available, a variety of mmwave hybrid precoding algorithms have been proposed in the past ten years, so as to mitigate interferences and improve system performances [5]–[9]. In classical approaches, interferences are often regarded as a limitation and should be suppressed as much as possible. Such approaches ignore the fact that interference, seen from an instantaneous point of view, can be constructive and can be exploited, e.g., on a symbol-level. In [10]–[12], the concept of constructive interference (CI) was exploited to improve the system performance. In particular, a low-complexity vector precoding scheme that incorporates CI was proposed in [13] for downlink multi-user MISO systems. Instead of suppressing interference, CI is also exploited as green signal power to improve energy efficiency in [14]. The significant gains in performance and energy efficiency, achieved by CI via SL precoding, is very appealing for mmwave communications, due to the poor power efficiency of mmwave hardware devices.

Note that in most precoding schemes, perfect CSI is often required, which is, however, never available in practice. In fact, it is particularly difficult to accurately estimate mmwave CSI, due to the large dimension of mmwave channels and the large path-loss. Moreover, limited feedback operations, quantization errors, and other non-idealities (e.g., hardware constraints) further exacerbate the problem of acquiring perfect CSI. Therefore, precoding algorithms that are robust to CSI errors become crucial. To tackle this issue, robust precoding algorithms have been proposed in [14]–[16], particularly, the robust precoding with CI exploitation in [14]. However, these robust precoding algorithms are based on specific CSI uncertainty models (e.g., the spherical bounded or Gaussian distribution model), which limits their applications, because the CSI uncertainty models in practice change with time and are often unavailable.

To overcome the shortcomings that CSI patterns (typically, the change or uncertainty models of CSI) have to be defined in advance, machine learning (ML) theory and algorithms have been introduced into wireless communications [17]–[19]. For example, a deep learning based hybrid precoding design was proposed in [19] with the aim of maximizing the sum-rate. A coordinated beamforming algorithm was developed in [20] for highly-mobile mmwave systems. Recently, a SL precoding design has been proposed in [18], which, however, focuses on the fully-digital precoding architecture and thus is unfriendly for mmwave hardware implementation. As for general precoding problems, robust solutions are still unavailable.

To address the issue of general robust precoding designs, in this paper we propose PLA based on Bayesian philosophy, by designing a probabilistic neural network and designing a novel loss function, where a probabilistic regularization term (PRT) is incorporated. Thanks to the PRT, in contrast to most robust precoding algorithms that only generate a single precoder, a probability distribution of the desired precoder can be learned. To exploit the interferences and facilitate mmwave hardware implementation, in this paper we formulate the problem of mmwave multi-user SL hybrid precoding as an optimization problem, where CSI acquisition is also taken into account. We further employ PLA to design an efficient robust precoding algorithm for the formulated optimization problem. Finally, comprehensive simulation results are provided to demonstrate the effectiveness and superiority of the proposed algorithm. It is shown that the proposed algorithm can effectively mitigate the impact due to CSI uncertainty.

II. System Model and Problem Statement

Consider a mmwave multi-user communication system, which consists of one base station (BS) and U single-antenna mobile users (MUs) collected into set $\mathcal{U} = \{1, 2, \cdots, U\}$. The BS is equipped with N_{T} transmit antennas controlled by $N_{\mathrm{F}}\,(N_{\mathrm{F}} < N_{\mathrm{T}})$ RF chains. To facilitate practical system implementation, the hybrid analog and digital precoding is considered in this paper [4]. In particular, all analog beams are chosen within a predefined analog codebook of size N_{C}, i.e., $\mathcal{C} = \{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_{N_{\mathrm{C}}}\}$.

To model the sparsity of mmwave channels, an extended Saleh-Valenzuela channel model is adopted in this paper. The channel vector between the BS and MU u is given by

$$\bar{\mathbf{h}}_u = \sqrt{N/\beta} \sum_{l=1}^{L_u} g_{u,l} \mathbf{a}(\phi_{u,l}), \tag{1}$$

where $\mathbf{a}(\cdot)$ is the array response vector, β is the average path-loss, L_u is the number of channel paths of MU u, and $g_{u,l}$ is the complex small-scale fading of the l-th path of MU u. Note that in Eq.(1), $\phi_{u,l} = \cos(\theta_{u,l})$, where $\theta_{u,l}$ is the physical angle of departure (AoD) of the l-th path of MU u.

Without loss of generality, the PSK modulation (with constellation $\mathcal{A}_u$ of size K_u for MU u) is considered in this paper. Nevertheless the developed approach can also be applicable to other modulations [21]. Let $s_u = e^{j\xi_u} \in \mathcal{A}_u$ be the intended information symbol for MU u and $\mathbf{x}$ be the transmitted signal. Then, the signal received at each MU u can be written as

$$y_u = \bar{\mathbf{h}}_u^{\mathrm{H}} \mathbf{x} + n_u,$$

where $n_u \sim \mathcal{CN}(0, \sigma_{\mathrm{N}}^2)$ denotes the random noise. The idea of CI is exploited, so as to improve the energy efficiency. For the PSK modulation, the key of the CI design principle can be captured by the following constraint $(\forall\, u \in \mathcal{U})$ [14]

$$\left|\mathrm{Im}(\bar{\mathbf{h}}_u^{\mathrm{H}} \mathbf{x} e^{-j\xi_u})\right| \le \left(\mathrm{Re}(\bar{\mathbf{h}}_u^{\mathrm{H}} \mathbf{x} e^{-j\xi_u}) - \gamma_u\right) \tan(\pi/K_u), \tag{2}$$

where γ_u measures the quality of received signal of MU u. The above design constraint enforces that the CI pushes the received signal away from the decision boundaries of the PSK constellation, therefore improving the received SNR without the need to increase the transmitted power [14].

Let $\mathbf{s} = [s_1, \cdots, s_U]^{\mathrm{T}}$. For the hybrid analog and digital precoding, the transmitted signal $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{D}\mathbf{s},$$

where $\mathbf{A} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{F}}}$ and $\mathbf{D} \in \mathbb{C}^{N_{\mathrm{F}} \times U}$ represent the analog and digital precoding matrices, respectively. For the codebook-based precoding, each column of $\mathbf{A}$ is chosen within $\mathcal{C}$, i.e., $\mathbf{A}(:, j) \in \mathcal{C}\,(\forall\, j = 1, \cdots, N_{\mathrm{F}})$, with $\mathcal{C}$ defined above. As an example, the design goal is to maximize the worst performance among the U MUs. Then, by letting $\mathbf{d} = \mathbf{D}\mathbf{s}$, the SL hybrid precoding problem can be formulated as

$$\begin{aligned} \max_{\mathbf{A}, \mathbf{d}, \{\gamma_u\}} \quad & \min_{u \in \mathcal{U}} \{\gamma_u\} \\ \text{s.t.} \quad & \left|\mathrm{Im}(\bar{\mathbf{h}}_u^{\mathrm{H}} \mathbf{A}\mathbf{d} e^{-j\xi_u})\right| \le \\ & \qquad C_u\left(\mathrm{Re}(\bar{\mathbf{h}}_u^{\mathrm{H}} \mathbf{A}\mathbf{d} e^{-j\xi_u}) - \gamma_u\right), \, (\forall\, u \in \mathcal{U}) \\ & \mathbf{A}(:, j) \in \mathcal{C}, \, (\forall\, j = 1, \cdots, N_{\mathrm{F}}) \\ & \|\mathbf{A}\mathbf{d}\|^2 \le p_0, \end{aligned} \tag{3}$$

where p_0 represents the transmit power budget and $\{C_u = \tan(\pi/K_u)\}$ are introduced for notation simplicity.

Note that the problem in (3) is a combinatorial optimization problem with non-convex equality constraints, which is difficult to tackle. Moreover, the coupling between the analog precoding matrix $\mathbf{A}$ and the digital precoding vector $\mathbf{d}$ further exacerbates the problem. Therefore, problem (3) is difficult to solve even if accurate CSI (i.e., $\{\bar{\mathbf{h}}_u\}$) is available. However, what is more challenging is that accurate CSI is almost never available in practice, which implies that even if the optimal solution could be obtained, the resulting performance may be still far from satisfactory. In fact, due to the large dimension of the channels, it is difficult (and even impossible) to accurately estimate the CSI. Therefore, a robust precoding approach that takes CSI acquisition and uncertainties into account is desired, which is the focus of the remaining part of this paper.

III. A Probabilistic Learning Framework for Robust Precoding

For practical wireless communication systems (not limited to mmwave communication systems), only imperfect CSI is available. To address this issue, we propose a general robust precoding approach in this section. As an example, in the next section we will design an efficient robust precoding algorithm for the considered SL hybrid precoding problem.

Let $\mathcal{H}_t$ represent CSI at time-slot t and $\hat{\mathcal{H}}_t$ be the (inaccurate) estimation. Given $\hat{\mathcal{H}}_t$, the general system design goal or optimization problem can be formulated as

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x} \,|\, \hat{\mathcal{H}}_t) \\ \text{s.t.} \quad & h_i(\boldsymbol{x} \,|\, \hat{\mathcal{H}}_t) = 0, \, (i = 1, \cdots, I) \\ & g_j(\boldsymbol{x} \,|\, \hat{\mathcal{H}}_t) \le 0, \, (j = 1, \cdots, J), \end{aligned} \tag{4}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the system design (or optimization) variable, and $\{h_i(\cdot \,|\, \cdot) = 0\}$ and $\{g_j(\cdot \,|\, \cdot) \le 0\}$ are equality constraints and inequality constraints, respectively. Without loss of generality, $\boldsymbol{x}$ is referred to as a precoder.

To alleviate the impact due to imperfect CSI, we propose the PLA. In contrast to most existing precoding algorithms that generate only a single precoder, PLA generates a probability distribution of the promising precoder. A specific precoder (e.g., when used for data transmission) is obtained by sampling the distribution. The motivation or rationale behind the design philosophy is that the algorithms that generate a single precoder are overconfident in their abilities or obtained CSI. To avoid overconfidence, from the perspective of ML, the idea of regularization should be incorporated to limit the representative abilities of the precoder models or algorithms.

Note that the probability distribution of a precoding model is often very complex and changes over time. Hence, the form of the probability distribution has to be sufficiently flexible, so as to efficiently represent a complex distribution. The NN models, whose representation abilities are very powerful, are adopted in this paper to tackle this issue. However, the output of a conventional NN is a deterministic function of its input. To tackle this issue, as shown in Fig. 1, all weights of the NN, denoted by $\mathbf{w}$, are represented by probability distributions over possible values [22], [1] rather than having a single fixed value as usual. As a result, instead of training a single network, an ensemble of networks are trained, where each network has its weights drawn from a shared learnt probability distribution.

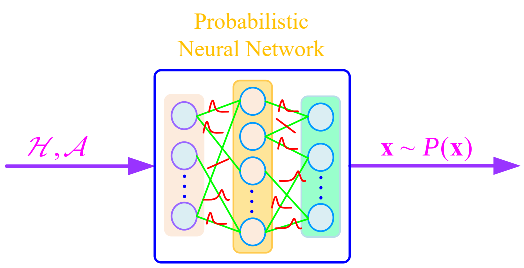

Fig. 1. All weights of the NN are assigned with probability distributions. $\mathcal{A}$ represents other possible input information (e.g., transmit symbol in SL precoding). For simplicity, $\mathcal{A}$ is absorbed in $\mathcal{H}$ and is omitted in the text.

PLA implemented via a NN can be viewed as a probabilistic model $\mathbb{P}(\mathbf{x}|\hat{\mathcal{H}}, \mathbf{w})$, i.e., given an input $\hat{\mathcal{H}}$ the NN assigns a probability to each possible precoder $\mathbf{x}$, using the parameters $\mathbf{w}$ (i.e., weights including biases). To incorporate possible prior knowledge, Bayesian modeling approach is considered and a prior $P(\mathbf{w})$ is placed on $\mathbf{w}$. Let $\mathcal{D} = \{(\hat{\mathcal{H}}_t, \mathbf{x}_t)\}$ denote the training dataset. Then, the prior $P(\mathbf{w})$ can be updated into a posterior $P(\mathbf{w}|\mathcal{D})$. However, an exact inference of the posterior is intractable, due to the large number of parameters and the functional form of the NN.

Fortunately, in practice it is sufficient to find an approximate solution to $P(\mathbf{w}|\mathcal{D})$ via variational inference (VI). Let $\{Q(\mathbf{w}|\boldsymbol{\lambda})\}$ be a family of variational distributions parameterized by $\boldsymbol{\lambda}$. VI finds the optimal parameters $\boldsymbol{\lambda}^\star$ of a variational distribution $Q(\mathbf{w}|\boldsymbol{\lambda})$ which minimizes the Kullback-Leibler (KL) divergence with the true Bayesian posterior on $\mathbf{w}$, i.e.,

$$\begin{aligned}\boldsymbol{\lambda}^\star &= \arg\min_{\boldsymbol{\lambda}} \mathrm{KL}(Q(\mathbf{w}|\boldsymbol{\lambda})\|P(\mathbf{w}|\mathcal{D}))\\ &= \arg\min_{\boldsymbol{\lambda}} \mathrm{KL}(Q(\mathbf{w}|\boldsymbol{\lambda})\|P(\mathbf{w})) - \mathbb{E}_{Q(\mathbf{w}|\boldsymbol{\lambda})}[\log P(\mathcal{D}|\mathbf{w})].\end{aligned}$$

[1]Note however that the amount of perturbation each weight exhibits is learnt in a way that coherently explains variability in the training data.

The remaining problem is to attach the optimization/update of the parameters $\boldsymbol{\lambda}$ and $\mathbf{w}$ to the original design goal, i.e., $\min\ f(\boldsymbol{x}\,|\,\hat{\mathcal{H}}_t)$. This can be achieved by designing a novel loss function, which is defined as

$$\begin{aligned}L = f(\mathbf{x}\,|\,\hat{\mathcal{H}}) + \rho_0 L_0(\mathbf{w}, \boldsymbol{\lambda}) + \rho_1 \sum_{i=1}^{I} \mathcal{E}(h_i(\mathbf{x}\,|\,\hat{\mathcal{H}}))\\ + \rho_2 \sum_{j=1}^{J} \mathcal{E}(g_j(\boldsymbol{x}\,|\,\hat{\mathcal{H}})),\end{aligned} \tag{5}$$

where $L_0(\mathbf{w}, \boldsymbol{\lambda})$ is given by

$$L_0(\mathbf{w}, \boldsymbol{\lambda}) = \mathrm{KL}(Q(\mathbf{w}|\boldsymbol{\lambda})\|P(\mathbf{w})) - \mathbb{E}_{Q(\mathbf{w}|\boldsymbol{\lambda})}[\log P(\mathcal{D}|\mathbf{w})].$$

Note that $\mathcal{E}(\cdot)$ in (5) penalizes possible constraint violations, and ρ_1 and ρ_2 represent the penalty parameters of the corresponding constraints. Typically, for the equality constraints, $\mathcal{E}(\cdot)$ takes the form of quadratic penalty, i.e., $\mathcal{E}(h_i(\boldsymbol{x}\,|\,\hat{\mathcal{H}}_t)) = |h_i(\boldsymbol{x}\,|\,\hat{\mathcal{H}}_t)|^2$, while for the inequality constraints, $\mathcal{E}$ can take the form of exponential function, i.e.,

$$\mathcal{E}(g_j(\boldsymbol{x}\,|\,\hat{\mathcal{H}}_t)) = \exp\left(\eta_j(\boldsymbol{x}\,|\,\hat{\mathcal{H}}_t)\right),$$

where constants$\{\eta_j > 0\}$ control the degree of penalty.

Remark 3.1 The term $L_0(\mathbf{w}, \boldsymbol{\lambda})$, which is referred to as PRT, plays a key role in PLA. From the view of optimization, $L_0(\mathbf{w}, \boldsymbol{\lambda})$ enables the designed algorithm to generate a probabilistic robust precoder, rather than a single precoder. From the view of ML, $L_0(\mathbf{w}, \boldsymbol{\lambda})$, along with $\rho_0 > 0$, plays the role of regularization. As a regularization term, $L_0(\mathbf{w}, \boldsymbol{\lambda})$ and $\rho_0 > 0$ balance the original design goal (i.e., to minimize f) and uncertainty measure or robustness. The rationale behind PLA is that the well-designed PRT limits the representative ability and DoF of the underlying precoder models, which, from the view of machine learning, avoids overconfident presentation.

The prior $P(\mathbf{w})$ plays an important role in designing a probabilistic precoding algorithm (e.g., it affects the number of required iterations or convergence rate), which should be carefully designed. However, due to space limitation, $P(\mathbf{w})$ is simply set to $\mathcal{N}(\mathbf{0}, \mathbf{I})$ in this paper. To obtain a more efficient prior (e.g., an environment-dependent prior), in [**] an efficient algorithm is proposed. Next, we take the SL hybrid precoding problem in (3) as an example to show how to use the PLA to develop a robust precoding algorithm.

IV. PLA Based Symbol-Level Hybrid Precoding for Millimeter Wave Communications

Building on the above framework, in this section we design an efficient robust SL precoding algorithm by employing PLA, along with the technique of beam training.

A. CSI Acquisition and Hybrid Precoding

Before proceeding to details of the hybrid precoding design, we first outline its basic principle. As shown in Fig. 2, the BS first finds out the main beam directions $\{\hat{\phi}_{u,l}\}$ of strong channel paths, based on which the analog precoding matrix $\mathbf{A}$ can be determined at the same time. Then, given the transmit symbols and the analog precoding matrix $\mathbf{A}$, the BS further designs a robust digital precoding vector $\mathbf{d}$. With both $\mathbf{A}$ and $\mathbf{d}$ available, data transmission can be performed. Finally, each MU recovers the intended information symbol based on the received signal via simple decision operations.

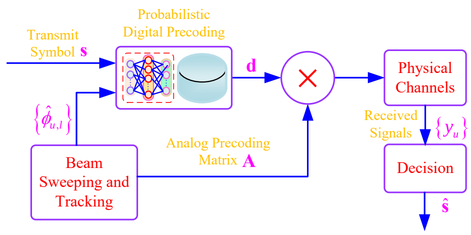

Fig. 2. The principle of the symbol-level hybrid precoding design.

To tackle the challenge of CSI acquisition, beam training (i.e., beam sweeping along with beam tracking) is utilized to circumvent the need for CSI. Let $\mathbf{h}_u = \mathbf{A}^{\mathrm{H}}\bar{\mathbf{h}}_u$, which is often referred to as equivalent channel vector [1]. The key is that, instead of estimating $\{\bar{\mathbf{h}}_u\}$, an estimation of $\{\mathbf{A}^{\mathrm{H}}\bar{\mathbf{h}}_u\}$, denoted by $\{\hat{\mathbf{h}}_u\}$, can be obtained. A simple method is as follows. The BS sends training signal $s = 1$ from each direction defined in $\mathcal{C}$. For each $\mathbf{f}_i \in \mathcal{C}$, the signal received by MU u can be written as $y_{u,i} = \bar{\mathbf{h}}_u^{\mathrm{H}}\mathbf{f}_i + n_{u,i}$ with $n_{u,i}$ denoting noise variable. Then, each MU u simply processes $\{y_{u,i} \,|\, i = 1, \cdots, N_{\mathrm{C}}\}$ (e.g., quantization) and feeds back the processed signals, denoted by $\{\hat{y}_{u,i} \,|\, i = 1, \cdots, N_{\mathrm{C}}\}$, to the BS. Since mmwave channels are sparse, there is no need to sweep the entire beam space and feed back all $\{\hat{y}_{u,i}\}$ to the BS [23].

Based on $\{\hat{y}_{u,i}\}$, the BS can determine the analog precoder $\mathbf{A}$ as follows. The number of RF chains allocated for MU u is assumed to N_u, which should satisfy $L_u \geq N_u \, (\forall\, u = 1, \cdots, U)$ and $\sum_{u=1}^{U} N_u = N_{\mathrm{F}}$. The analog beams of each MU u are chosen as the beams that align the most strongest N_u channel paths. Specifically, the N_u analog beams of MU u, denoted by $\mathcal{C}'_u = \{\mathbf{f}_{i_1}, \mathbf{f}_{i_2}, \cdots, \mathbf{f}_{i_{N_u}}\} \subset \mathcal{C}$, should satisfy

$$|\hat{y}_{u,i}| \approx |\bar{\mathbf{h}}_u^{\mathrm{H}}\mathbf{f}_i| > |\bar{\mathbf{h}}_u^{\mathrm{H}}\mathbf{f}_j| \approx |\hat{y}_{u,j}|, \; (\forall\, \mathbf{f}_i \in \mathcal{C}'_u \text{ and } \mathbf{f}_j \in \mathcal{C}\backslash\mathcal{C}'_u).$$

The analog precoding matrix $\mathbf{A}$ can be obtained by collecting all beams in $\mathcal{C}'_u$ of all U MUs. Then, the SL hybrid precoding problem in (3) can be rewritten as

$$\begin{aligned} \max_{\mathbf{d},\{\gamma_u\}} \quad & \min_{u\in\mathcal{U}} \{\gamma_u\} \\ \text{s.t.} \quad & \big|\mathrm{Im}(\hat{\mathbf{h}}_u^{\mathrm{H}}\mathbf{d}e^{-j\xi_u})\big| \leq C_u\big(\mathrm{Re}(\hat{\mathbf{h}}_u^{\mathrm{H}}\mathbf{d}e^{-j\xi_u}) - \gamma_u\big), \\ & (\forall\, u \in \mathcal{U}), \quad \|\mathbf{A}\mathbf{d}\|^2 \leq p_0. \end{aligned} \tag{6}$$

Problem (6) is convex and can be solved efficiently. The SL hybrid precoding algorithm is summarized in Algorithm 2.

Note that although a good performance can be achieved by Algorithm 1 in some cases (e.g., the CSI is relatively accurate), the use of inaccurate CSI inevitably causes a large performance degradation. In fact, the CSI is not only contaminated by the received noise, the use of an analog codebook also leads to extra quantization error. To further improve the performance, we next propose a more robust algorithm based on PLA.

Algorithm 1: Hybrid Precoding with estimated CSI

1: **input:** analog codebook $\mathcal{C} = \{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_{N_{\mathrm{C}}}\}$

2: **repeat** for each time-slot

(a) find out N_u most strongest channel paths via beam training

(b) determine analog precoding matrix $\mathbf{A}$

(c) solve problem (6) $\Longrightarrow$ digital precoder $\mathbf{d}$

(d) perform data transmission with $\mathbf{A}$ and $\mathbf{d}$

end

B. Learning-Based Symbol-Level Precoding

The estimation of equivalent channel vectors and the design of analog precoder are similar to Algorithm 1, which is mainly based on beam training. The remaining task is to use PLA to design a robust digital precoder, which essentially constructs a mapping based on NN that can generate a distribution of the digital precoder when given inaccurate CSI and transmitted symbols. To facilitate the use of existing deep learning libraries (e.g., Tensorflow or Pytorch) which only support real-value operations, all complex-value input and output are transformed into their real-value counterparts. The network structure, input, output and loss function are designed as follows:

1) Network Structure: The commonly used fully-connected NN is adopted as an example. Moreover, to maximize system design goal $\min_{u\in\mathcal{U}}\{\gamma_u\}$, the inequality constraint $\|\mathbf{A}\mathbf{d}\|^2 \leq p_0$ in (3) is, in fact, active, i.e., $\|\mathbf{A}\mathbf{d}\|^2 = p_0$. To simplify the loss function, this constraint can be absorbed into the NN [18]. Specifically, an extra layer (i.e., the normalization layer) is appended to the NN with activation function given by

$$\sigma(\mathbf{z}) = \min\big(\sqrt{p_0}, \|\mathbf{z}\|\big)\mathbf{z}/\|\mathbf{z}\|.$$

2) Input: The input of the NN consists of transmitted symbols $\mathbf{s}$, estimated beam directions $\{\hat{\phi}_{u,l}\}$ (via beam training) and estimated equivalent channel vectors $\{\hat{\mathbf{h}}_u\}$. Note that both $\{\hat{\mathbf{h}}_u\}$ and $\mathbf{s}$ should be first transformed into their real-value counterparts and then fed into the NN.

3) Output: Similarly, the real part and imaginary part of the digital precoder $\mathbf{d}$ are tackled separately, i.e., $\bar{\mathbf{d}} = [\mathbf{d}_{\mathrm{R}}^{\mathrm{T}}, \mathbf{d}_{\mathrm{I}}^{\mathrm{T}}]^{\mathrm{T}}$ with $\mathbf{d}_{\mathrm{R}} = \mathrm{Re}(\mathbf{d})$ and $\mathbf{d}_{\mathrm{I}} = \mathrm{Im}(\mathbf{d})$. The output of the NN is $\bar{\mathbf{d}}$, which is further transformed into the complex-value counterpart for the subsequent data transmission.

4) Loss Function: According to the principle of PLA, the loss function is given in (7) (the top of the next page),

$$\begin{aligned} L = &-\min_{u\in\mathcal{U}}\{\gamma_u\} + \rho_0 L_0(\mathbf{w}, \boldsymbol{\lambda}) + \\ &\rho_1 \sum_{u=1}^{U}\Bigg[\exp\Big(\mathrm{Im}(\hat{\mathbf{h}}_u^{\mathrm{H}}\mathbf{d}e^{-j\xi_u}) - C_u\big(\mathrm{Re}(\hat{\mathbf{h}}_u^{\mathrm{H}}\mathbf{d}e^{-j\xi_u}) - \gamma_u\big)\Big) \\ &+ \exp\Big(C_u\big(\mathrm{Re}(\hat{\mathbf{h}}_u^{\mathrm{H}}\mathbf{d}e^{-j\xi_u}) - \gamma_u\big) - \mathrm{Im}(\hat{\mathbf{h}}_u^{\mathrm{H}}\mathbf{d}e^{-j\xi_u})\Big)\Bigg]. \end{aligned} \tag{7}$$

where $\exp(\cdot)$ is the exponential function. Note that the penalty term corresponding to $\|\mathbf{A}\mathbf{d}\|^2 \leq p_0$ in (3) is absent since it has been absorbed into the normalization layer of the NN.

Before proceeding to details of the robust precoding algorithm, we shall point out how to compute the loss in (5). The difficulty of computing the loss is mainly caused by the PRT $L_0(\mathbf{w}, \boldsymbol{\lambda})$. The first term in the PRT, i.e., $\mathrm{KL}\big(Q(\boldsymbol{\Phi}|\boldsymbol{\lambda})\|P(\boldsymbol{\Phi})\big)$, can often be analytically calculated. As an example, still the diagonal Gaussian distribution is chosen as the variational posterior, i.e., $Q(\boldsymbol{\Phi}|\boldsymbol{\lambda}) = \mathcal{N}(\boldsymbol{\zeta}, \boldsymbol{\Lambda})$, where $\boldsymbol{\zeta}$ is the mean vector and $\boldsymbol{\Lambda}$ is the covariance matrix (a diagonal matrix). Then, $\mathrm{KL}\big(Q(\boldsymbol{\Phi}|\boldsymbol{\lambda})\|P(\boldsymbol{\Phi})\big)$ can be calculated as

$$\begin{aligned} &\mathrm{KL}\big(Q(\boldsymbol{\Phi}|\boldsymbol{\lambda})\|P(\boldsymbol{\Phi})\big) \\ =&0.5 \cdot \big(\mathrm{tr}(\boldsymbol{\Lambda}) + \boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\zeta} - \log\det(\boldsymbol{\Lambda}) - d\big), \end{aligned} \tag{8}$$

where d denotes the dimension of $\boldsymbol{\Phi}$. As for the second term in the PRT, i.e., $\mathbb{E}_{Q(\boldsymbol{\Phi}|\boldsymbol{\lambda})}[\log P(\mathcal{D}|\boldsymbol{\Phi})]$, it is estimated via the Monte-Carlo sampling.

Algorithm 2: Robust Hybrid Precoding based on PLA

1: **input:** analog codebook $\mathcal{C} = \{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_{N_{\mathrm{C}}}\}$;
K - frequency that each sample is used;
M - sampling frequency of weights $\mathbf{w}$;
Q - update frequency of digital precoder

2: **repeat** for each time-slot
(a) determine analog precoder $\mathbf{A}$ and equivalent channel vectors $\{\hat{\mathbf{h}}_u\}$ via beam training
(b) initialize variational distribution $Q(\mathbf{w}|\boldsymbol{\lambda})$
(c) **for** $k = 1, \cdots, K$
(1) sample $\mathbf{w}$ M times $\Longrightarrow \mathcal{W} = \{\mathbf{w}_1, \cdots, \mathbf{w}_M\}$
(2) forward propagation $\Longrightarrow \{\bar{\mathbf{d}}_1 \cdots, \bar{\mathbf{d}}_M\}$
(3) compute average loss according to (7)
(4) update prior into posterior via BP
end
(d) **as** per frequency Q (or with period $1/Q$):
(1) sample $\mathbf{w}$ and perform forward propagation to obtain precoder $\mathbf{x}_0$
(2) perform data transmission with $\mathbf{x}_0$
end

For clarity, the designed robust SL precoding algorithm is summarized in Algorithm 2. Before starting the algorithm, the hyper-parameters ρ_0 and ρ_1 should be provided. In each time-slot, beam training is invoked to determine the analog precoder $\mathbf{A}$ and equivalent channel vectors. Then, PLA is employed to design a robust digital precoder, which consists of 4 main steps. In step (1), we sample the prior to generate the weights $\mathbf{w}$. [2] In step (2), we perform forward propagation to obtain the digital precoder $\bar{\mathbf{d}}$ (or equivalently $\mathbf{d}$). The average loss can be calculated according to (7) in step (3), and the prior is updated into a posterior via back-propagation (BP) in step (4). Via K iterations, a better distribution is available, and a specific precoder can be obtained by sampling the distribution. Data transmission can be performed between the BS and MUs with the obtained analog and digital precoders.

[2]For the diagonal Gaussian distribution, samples of $\mathbf{w}$ are obtained by first sampling a unit Gaussian distribution, then scaling it by a standard deviation $\boldsymbol{\Lambda}^{1/2}$ and shifting it by a mean $\boldsymbol{\mu}$, i.e., a sample of $\mathbf{w}$ is represented by

$$\mathbf{w} = \boldsymbol{\Lambda}^{1/2} \circ \boldsymbol{\varepsilon} + \boldsymbol{\mu} \quad \text{with} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{9}$$

V. Numerical Results

In this section, simulation results are provided to demonstrate the performance of the proposed algorithms. For comparison, the fully-digital SL precoding (FD-SLP) solution of [14] (with some modifications) is chosen as the benchmark. Note that FD-SLP requires physical CSI (pCSI), i.e., $\bar{\mathbf{h}}_u = \sqrt{N/\beta}\sum_{l=1}^{L_u} g_{u,l}\mathbf{a}(\phi_{u,l})$, to design a precoder. For convenience, the optimization-based hybrid SL precoding without learning (in Algorithm 1) and the hybrid precoding based on PLA (in Algorithm 2) are named as HSLP-Opt and HSLP-PLA, respectively. The symbol error rate (SER) is chosen as performance metric to evaluate different algorithms.

Without loss of generality, the uniform linear array with $N_{\mathrm{F}} = 3$ RF chains is adopted. For all simulation experiments, the channel model in (1) consists of $L_{\mathrm{S}} = 2$ channel paths whose path gains are strong (i.e., strong channel paths) and three channel paths whose path gains are weak (i.e., weak channel paths). The AoDs of all weak paths are uniformly distributed in $[-1, 1]$. The average power ratio between the gain of strong channel paths and the gain of weak channel paths is 10dB. The path gain of each weak path is distributed as $\mathcal{CN}(0, \sigma_{\mathrm{W}}^2)$ with σ_{W}^2 calculated as per the power ratio.

For the analog precoding codebook design, the widely used DFT codebook is adopted, which can be obtained by uniformly sampling $[-1, 1]$. Accordingly, the range of quantization error is $(-1/N, 1/N)$. For simplicity, the quantization error is assumed to be uniformly distributed in $(-1/N, 1/N)$. Let $\hat{\mathbf{h}}_u$ denote the estimation of accurate equivalent channel vector $\mathbf{h}_u$, which, similar to [14], is assumed to satisfy

$$\mathbf{h}_u = \hat{\mathbf{h}}_u + \Delta\mathbf{h}_u, \tag{10}$$

where $\Delta\mathbf{h}_u$ is distributed as $\Delta\mathbf{h}_u \sim \mathcal{CN}(0, \sigma_h^2\mathbf{I})$.

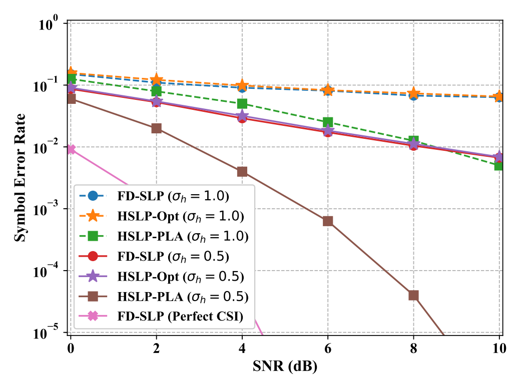

Fig. 3. The SER performance of different precoding algorithms: $N = 64$, $U = 2$ and QPSK modulation.

Firstly, we confirm the effectiveness and superiority of the proposed algorithms. Fig. 3 shows the SER performance of the three precoding algorithms. It is not surprising that FD-SLP with perfect pCSI achieves the best performance. However, HSLP-PLA performs better than FD-SLP or HSLP-Opt for inaccurate CSI. The reason for this is that FD-SLP and HSLP-Opt are overconfident to their abilities and obtained CSI. In contrast, HSLP-PLA generates a probability distribution, and then multiple digital precoders can be obtained via sampling, which makes HSLP-PLA not act too confidently.

It is amazing that even if "accurate" equivalent CSI (i.e., $\sigma_h = 0$) is fed to HSLP-Opt, HSLP-PLA can still achieve a much better performance than HSLP-Opt in some cases (e.g., when $\mathrm{SNR} \geq 0\mathrm{dB}$ and $\sigma_h = 0.5$). The reason for this is as follows. For the codebook-training-based analog beamforming, even if $\sigma_h = 0$ is assumed, the beam directions cannot be aligned accurately (because of the quantization error of the codebook) and the array gains cannot be obtained completely (due to multi-path interferences), which implies that accurate CSI is always unavailable. As a result, performance loss due to inaccurate CSI is inevitable. Nevertheless, since HSLP-PLA is designed based on PLA, randomly generated precoders can compensate for the performance loss due to inaccurate CSI.

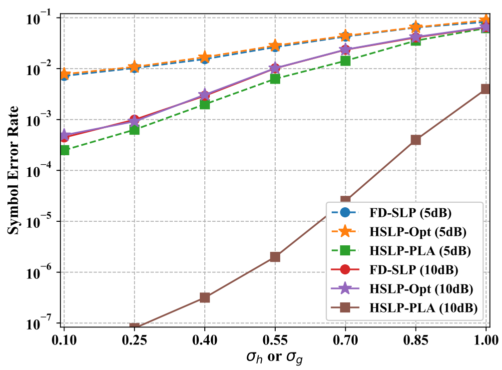

Fig. 4. The SER performance of different precoding algorithms under various degrees of CSI uncertainties: $N = 64$, $U = 2$ and QPSK modulation.

Fig. 4 shows the SER performance of different algorithms under various degrees of CSI uncertainties. Not surprisingly, HSLP-PLA achieves the best performance among the three algorithms. It is observed from Fig. 4 that as the degree of CSI uncertainties decreases, the SER performance of HSLP-Opt and HSLP-PLA both become better and better. However, HSLP-PLA achieves a better perfermance than HSLP-Opt in the entire interval (from $\sigma_h = 0.1$ to $\sigma_h = 1.0$), and the gap in terms of SER performance (in the logarithmic scale) between the two precoding algorithms becomes larger and larger as the degree of CSI uncertainties decreases. It is observed from Fig. 3 and Fig. 4 that HSLP-PLA and FD-SLP achieve a similar SER performance. However, HSLP-PLA is training- and feedback-efficient.

VI. Conclusion

To exploit interferences and facilitate hardware implementation, in this paper we formulated the problem of mmwave multi-user SL hybrid precoding as an optimization problem. To tackle imperfect CSI, we proposed a probabilistic precoding methodology for a general robust precoding problem by designing a NN with a novel loss function. Based on the robust precoding methodology, we proposed a learning-based robust precoding algorithm for the formulated SL precoding problem. Simulations confirm the superiority and effectiveness.

References

[1] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. L. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sept 2017.

[2] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmwave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[3] Z. Xiao, T. He, P. Xia, and X. G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[4] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289–5304, Oct 2017.

[5] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[6] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, May 2015.

[7] A. Alkhateeb, G. Leus, and R. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov 2015.

[8] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, 2016.

[9] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave mimo systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, 2016.

[10] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in mimo broadcast systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1396–1404, 2009.

[11] C. Masouros, "Correlation rotation linear precoding for MIMO broadcast communications," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 252–262, 2011.

[12] C. Masouros and T. Ratnarajah, "Interference as a source of green signal power in cognitive relay assisted co-existing mimo wireless transmissions," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 525–536, 2012.

[13] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Vector perturbation based on symbol scaling for limited feedback miso downlinks," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 562–571, 2014.

[14] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, 2015.

[15] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Robust SINR-constrained symbol-level multiuser precoding with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 68, pp. 1837–1852, 2020.

[16] A. Morsali and B. Champagne, "Robust hybrid analog/digital beamforming for uplink massive-mimo with imperfect csi," in *2019 IEEE WCNC*, 2019, pp. 1–6.

[17] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang, and X. Shen, "Fast mmwave beam alignment via correlated bandit learning," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5894–5908, 2019.

[18] F. Sohrabi, H. V. Cheng, and W. Yu, "Robust symbol-level precoding via autoencoder-based deep learning," in *2020 IEEE ICASSP*, 2020, pp. 8951–8955.

[19] A. M. Elbir and A. K. Papazafeiropoulos, "Hybrid precoding for multiuser millimeter wave massive mimo systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 552–563, 2020.

[20] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37 328–37 348, 2018.

[21] C. Masouros, T. Ratnarajah, M. Sellathurai, C. B. Papadias, and A. K. Shukla, "Known interference in the cellular downlink: a performance limiting factor or a source of green signal power?" *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 162–171, 2013.

[22] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," in *International Conference on Machine (ICML)*, 2015, pp. 1613–1622.

[23] J. Zhang, Y. Huang, Y. Zhou, and X. You, "Beam alignment and tracking for millimeter wave communications via bandit learning," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5519–5533, 2020.