IEEE ComSoc IEEE Communications Society | IEEE COMPUTER SOCIETY | IEEE Signal Processing Society | VTS | IEEE Transactions on Machine Learning in Communications and Networking

Received 23 November, 2023; revised 20 March, 2024; accepted 30 April, 2024; Date of publication XX Month, XXXX; date of current version 7 May, 2024.

# Hierarchical ML Codebook Design for Extreme MIMO Beam Management

**Ryan M. Dreifuerst[1], Graduate Student Member, IEEE, and Robert W. Heath Jr.[2], Fellow, IEEE**

[1]Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA
[2]Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093 USA

Corresponding author: Ryan M. Dreifuerst (email: rmdreifu@ncsu.edu).

This material is based upon work supported by the National Science Foundation under grant nos. NSF-ECCS-2153698, NSF-CCF-2225555, NSF-CNS-2147955 and is supported in part by funds from federal agency and industry partners as specified in the Resilient & Intelligent NextG Systems (RINGS) program.

**ABSTRACT** Beam management is a strategy to unify beamforming and channel state information (CSI) acquisition with large antenna arrays in 5G. Codebooks serve multiple uses in beam management including beamforming reference signals, CSI reporting, and analog beam training. In this paper, we propose and evaluate a machine learning-refined codebook design process for extremely large multiple-input multiple-output (X-MIMO) systems. We propose a neural network and beam selection strategy to design the initial access and refinement codebooks using end-to-end learning from beamspace representations. The algorithm, called Extreme-Beam Management (X-BM), can significantly improve the performance of extremely large arrays as envisioned for 6G and capture realistic wireless and physical layer aspects. Our results show an 8dB improvement in initial access and overall effective spectral efficiency improvements compared to traditional codebook methods.

**INDEX TERMS** Beam Management, Codebook Design, Extreme MIMO, Machine Learning

## I. Introduction

Large-scale or Massive MIMO (M-MIMO) has been an important technology in mobile broadband since the first investigations on 3D beamforming in 3GPP release 13 [1]. Since then, larger arrays have been envisioned with extremely large arrays ($\gg 64$ antennas) for 6G [2]. Effectively employing M-MIMO or X-MIMO, however, requires obtaining accurate CSI. With the inception of 5G, M-MIMO was supported through a process called beam management which integrated beamforming for control signaling, hierarchical codebooks for beam training, and enabled multiple configurations of feedback [1]. Notably, beam management supports hybrid arrays where the number of antenna ports visible to digital communication systems and the number of antennas in the array might be vastly different, as found in the so-called hybrid MIMO architecture [3]. While the beam management framework is designed to be flexible, there has been limited research on optimizing beam management codebooks for this flexible framework, leading to expensive overhead and underperformance [4], [5].

Beam management is a multi-step process that includes transmitting beamformed reference signals and obtaining multiple forms of CSI to balance the overhead, latency, and accuracy [1], [4], [6]. In 5G, the first step is synchronization which requires transmitting one or more synchronization signal blocks (SSB) with a beamformer from a codebook (denoted as the *SSB codebook*). These are typically understood as wide-area coverage beams because synchronization does not require a high signal-to-noise ratio (SNR), but wide-area beamforming improves cell coverage and aids the subsequent beam search. From the set of SSBs, user equipment (UE) provides a small feedback message identifying the strongest SSB and the power received from the transmission [6]. Based on the SSB feedback, the base station (BS) selects a refined beamforming codebook (identified as the *CSI-RS codebook*) for transmitting CSI reference signals (CSI-RS), which contain pilot signals for channel estimation. These two beamforming codebooks support different functions (initial access and synchronization compared to beam refinement and channel estimation) and operate over different time-frequency resource allocations to support these roles. Following CSI-RS reception, UEs estimate the effective channel

and quantize it according to another codebook (the *feedback codebook*) based on the BS array geometry. Finally, the CSI feedback is provided to the BS and data transmission can begin with an appropriately designed precoder to serve one or more users.

Optimizing the various codebooks is challenging due to the difficulty in mathematically capturing the relationships between the channel, mobile users, and the codebook inter-relationships. In fact, there is no standardized technique for optimizing codebooks because of the difficulty in accurately capturing and modeling the distributions and roles served by codebooks in 5G. One possible solution is applying artificial intelligence or machine learning (AI/ML) to learn these underlying characteristics directly from data. While AI/ML has become an integral part of domains like computer vision and language modeling, mobile broadband has not seen as many benefits due to the well-known structure and strict timing requirements of wireless communications [7], [8]. Recently, 3GPP has announced investigations for the integration of AI/ML targeting 3GPP Release 18 and beyond [9]. In particular, beam management and CSI feedback are identified as key areas where learning strategies are envisioned because of the lack of performance and potential for realistic applications [9].

With the initial proposals for 6G underway, new research on AI/ML for extremely large arrays, equipped with hundreds of antennas, is a critical topic, especially operating in new communication bands [2]. 5G initially began supporting massive MIMO in a hybrid format, where analog beam training was performed using the SSB and CSI-RS beamformers, and digital precoding was based on the CSI feedback. Given the support and success of these formats, it is valuable to design and optimize strategies for X-MIMO within this formulation. Successfully accomplishing this task paves the way for beam management with AI/ML in 6G.

There are three directions of work related to AI/ML beam management: 1) beam training, 2) codebook design, and 3) CSI feedback. First, we review a set of relevant prior work incorporating machine learning and beam training. Beam training has received the most attention of the three directions when it comes to machine learning applications. Classical data-driven techniques were explored in [10], [11] using historical beam training data. While the codebooks and algorithms rely on more traditional techniques, the concept of "learning" a reduced codebook from site-specific data is a foundational concept for more recent deep learning techniques [12]–[15]. These papers ( [12]–[15]) focus on millimeter-wave (mmWave) channels and typically exploit the sparsity to aid the beam training process. Furthermore, the codebooks within these papers are standard codebooks like DFT which are unlikely to perform as well in rich scattering environments as seen in upper-mid bands (7-20GHz) envisioned for 6G [2]. In addition, that work [12]–[15] does not consider a MIMO format with multiple UE antennas and streams of data, instead limiting the setting to analog beamforming. In this paper, we consider a scenario with multiple users, multiple data streams, and frequency selective, OFDM channels in rich multipath environments.

Codebook and beamformer design using machine learning has received some focus in recent years [13], [16]–[19]. Hierarchical codebooks, which are a core concept in 5G, were proposed in [20]–[22] to separate the beam search process into multiple resolutions to reduce the overhead of beam search. Deep reinforcement learning for broadcast beam pattern transmission was shown to be effective in small-scale MIMO [16]. Supervised learning for SSB codebook design in narrowband MIMO was also shown to improve over standard codebooks [19]. Unsupervised learning was applied to beamformer design from narrowband channel measurements [17] but required perfect channel state information. Broadcast probing beam patterns were learned using channel state information in [18] which is one of the only papers to consider a hierarchical strategy similar to 5G based on SSB and CSI-RS codebooks. This work [18] provides motivation for researching realistic deep learning codebook methods, but it remains unclear how the integration of ML impacts the overall system performance as opposed to simply improving the beam alignment. Our work addresses this question while further increasing the role that data-driven algorithms play in codebook operations.

While codebook design is typically focused on improving the received signal power, codebooks are also employed for quantizing channel state information into discrete feedback. ML has also been applied to improve feedback quantization in recent work [23]–[26]. In general, ML methods propose encoder-decoder structures [23]–[26], that employ jointly-trained models at the UE and BS. This type of application involves significant overhead, however, to ensure both terminals have synchronized neural network weights or parameters. Furthermore, the system effects in MU-MIMO settings are not considered, even though high resolution CSI feedback is primarily designed for improving MU-MIMO performance [27]. Throughout these investigations, CSI optimization is treated as a separate task beam training and channel estimation, thereby neglecting the relationships throughout the system. In this paper, we incorporate beamformed channel feedback as a result of beam management to study the effects of ML codebooks on the overall system performance.

In our prior work [28], we focused on a 5G, fully digital sub-6GHz system. From that investigation, we found that notable gains can be achieved with improvements to the beam management framework using deep learning. While the strategy was effective for SSB transmission, we now extend the consideration to a new system configuration that supports hybrid, upper-mid band systems with jointly-learned SSB and CSI-RS codebooks. Throughout the process, the overall goal is to balance the computational complexity, cost, and

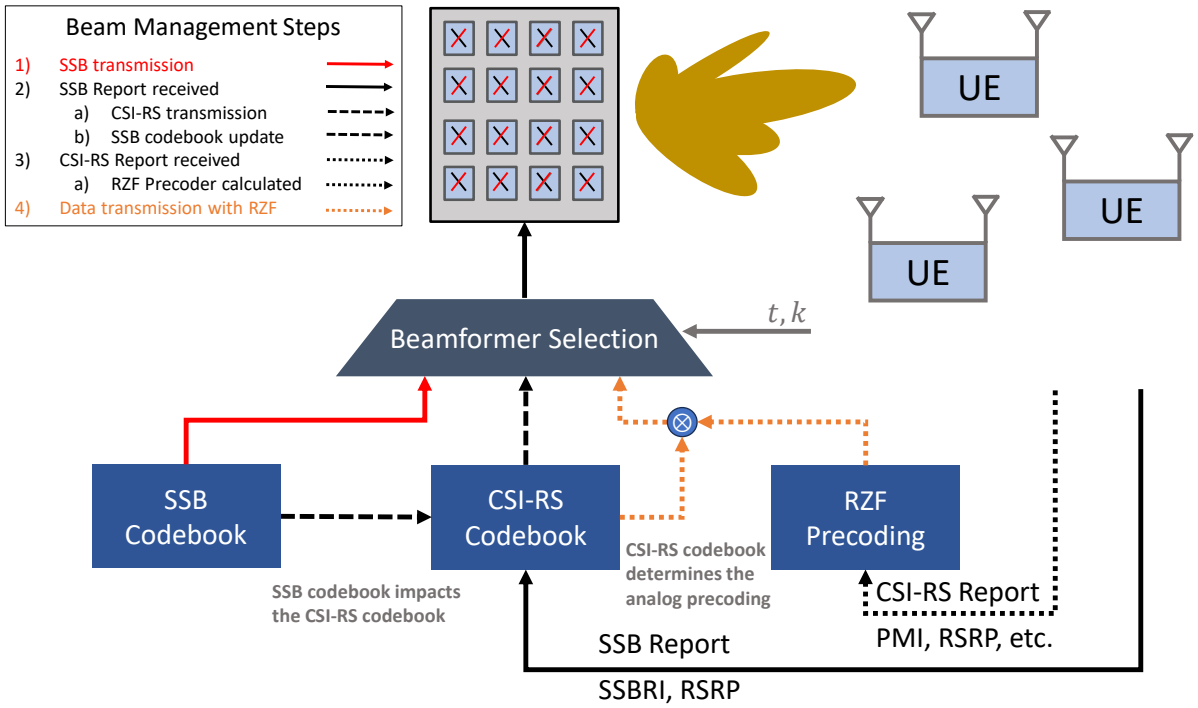



**FIGURE 1.** **A depiction of the timing and interaction between the codebooks and data transmission. The SSB and CSI-RS codebooks are used to determine an effective analog precoding that is combined with the precoding determined by the CSI-RS report.**

performance of CSI acquisition to maximize system-level performance. Our contributions in this manuscript are as follows:

- First, we precisely define the beam management process for an envisioned hybrid, X-MIMO system. The process unifies traditional hybrid array formulations with standardized beam management to support large-scale MIMO arrays. The system includes all relevant steps from initial access up through achievable network spectral efficiency calculation. This formulation enables a holistic evaluation of codebooks that captures the interdependencies of a realistic beam management strategy for 6G and is a natural evolution of that supported in 5G.
- Second, we design a multi-task neural network architecture and processing pipeline to design SSB and CSI-RS codebooks. The pipeline starts by converting the current SSB codebook and SSB feedback into the beamspace. Then the input is fed into two neural networks that are jointly trained in an end-to-end fashion to output an SSB and a set of candidate CSI-RS codebooks. End-to-end training allows for incorporating the hierarchical beam search from the system model into the gradient calculations to enable robust and efficient learning. The joint architecture is formulated to fit the strict timing requirements of beam training. The joint neural network formulation is able to learn codebooks that achieve beam training results similar to if perfect CSI were known a priori.
- Third, we rigorously evaluate the capabilities of the beam management framework in X-MIMO using various codebooks, physically array geometries, and feedback strategies. The most important result we find is that the proposed system, with limited feedback, can achieve comparable performance to DFT-codebook systems with full feedback in MU-MIMO settings. Furthermore, we find that the X-BM design transfers efficiently between new environments, antenna arrays, and frequency ranges.

**Notation**: $\mathbf{A}$ is a matrix, $\mathbf{a}$ and $\{a[i]\}_{i=1}^{N}$ are column vectors and $a, A$ denote scalars. $\mathbf{A}^T$, $\overline{\mathbf{A}}$, $\mathbf{A}^*$, and $\mathbf{A}^\dagger$ represent the transpose, conjugate, conjugate transpose, and psuedo-inverse of $\mathbf{A}$. $\mathbf{A}[k,\ell]$ denotes the entry of $\mathbf{A}$ in the $k^{\text{th}}$ row and the $\ell^{\text{th}}$ column. The same meaning is also associated with $\mathbf{A}_{k,\ell}$. Similarly, $\mathbf{A}[:,k]$ refers to the $k^{\text{th}}$ column of $\mathbf{A}$. Unspecified norm equations are $\|\mathbf{a}\|_2 = \mathbf{a}^*\mathbf{a}$ for vectors and the Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}^*)}$ for matrices. We use the notation $\lfloor\cdot\rceil$ for rounding to the nearest integer. We define $\mathrm{j} = \sqrt{-1}$. The operator $\mathbb{E}[\cdot]$ is used for the expectation of a random variable. Due to the notational complexity of MU-MIMO with OFDM, we will always use $u$ to refer to a specific UE, $t$ as a specific time, $k$ as a specific frequency resource, and $n_{\text{t}}/n_{\text{r}}$ to refer to a specific transmit or receive antenna. We use the superscript $(\cdot)^{\text{RF}}$ to refer to the analog component of hybrid variables within the MIMO system.

The remainder of our paper is organized as follows. First, we introduce the system model and briefly highlight the SSB and CSI-RS beam management processes. We conclude this section with an overview of PMI and define the data transmission stage. We pay particular attention to the beam training as we integrate the beam training into our learning algorithm as an end-to-end model. We introduce the proposed algorithm, X-BM, that includes beamspace preprocessing, a multi-stage neural network, and end-to-end learning. In the final sections, we present the simulation setup and a rigorous evaluation of the various codebooks, feedback formats, and overall performance achieved by our proposed method.

## II. System model

We consider one region served by a single X-MIMO base station. The BS is equipped with an $N_{\mathrm{T}}^{\mathrm{RF}} = N_{\mathrm{X}}^{\mathrm{RF}} \times N_{\mathrm{Y}}^{\mathrm{RF}}$ planar array in a hybrid architecture such that antenna elements are connected with phase shifters to radio frequency chains to form a digital dimension of size $N_{\mathrm{T}} = N_{\mathrm{X}} \times N_{\mathrm{Y}}$. The phase shifters are limited to $b_{\text{phase}}$ bits resolution but we assume a fully-connected structure such that every digital port is connected to each physical antenna with a separate phase shifter [29], [30]. On the other side of the radio link, there are $U$ UEs, where $U$ is a random variable, each equipped with an $N_{\mathrm{R}}$ uniform linear array which we assume to be fully digital for this investigation. This reduces the notational complexity and allows us to focus on the BS operations, whereas UE-side optimization is left to future work. All antennas are half-wavelength spaced in the YZ planes. It remains to be seen what bands and configurations will be standardized in 6G, so we will assume channels are generated according to a model but rely on machine learning to capture channel structures rather than incorporate a specific channel model into the system. The OFDM MU-MIMO channel is defined for $U$ users, over $T$ time slots, and $K$ subcarriers between each of the terminals as

$$\mathbf{H} \in \mathbb{C}^{U \times T \times K \times N_{\mathrm{R}} \times N_{\mathrm{T}}^{\mathrm{RF}}}. \tag{1}$$

Note that we can generate this channel model using a clustered channel model as typically done by Sionna [31], 3GPP, and other raytracing or statistical models and array response vectors $\mathbf{a}_N(\theta,\phi) = \frac{1}{\sqrt{N}}[1, e^{\mathrm{j}\pi\cos\theta\sin\phi}, e^{\mathrm{j}2\pi\cos\theta\sin\phi}, ..., e^{\mathrm{j}(N-1)\pi\cos\theta\sin\phi}]^T$ for the path direction $(\theta,\phi)$.

With this setup in mind, we work with the following generic received signal model for signals $\mathbf{s}$ transmitted with digital and analog precoders $\mathbf{F}_{u,t,k}$, $\mathbf{F}_t^{\mathrm{RF}}$ and received with combiners $\mathbf{W}_{u,t,k}$

$$\mathbf{y}_{u,t,k} = \frac{1}{\sqrt{N_{\mathrm{T}}}}\mathbf{W}_{u,t,k}^{*}\left(\mathbf{H}_{u,t,k}\mathbf{F}_t^{\mathrm{RF}}\sum_{\mathrm{u}=0}^{U-1}\mathbf{F}_{\mathrm{u},t,k}\mathbf{s}_{\mathrm{u},t,k} + \mathbf{N}_{t,k}\right). \tag{2}$$

The noise $\mathbf{N}$ is modeled as independent complex Gaussian random values to account for thermal noise and the noise figure of the receiver. Our model assumes that timing and frequency synchronization have been performed, which is supported through the embedded primary synchronization and secondary synchronization sequences contained within the SSB signals. During the next subsections, we will specify how $\mathbf{W}, \mathbf{F}$, and $\mathbf{s}$ are determined during the SSB, CSI-RS, and data transmission stages according to an SSB codebook $\mathbf{B}^{\mathrm{SSB}} \in \mathbb{C}$, a CSI-RS codebook $\mathbf{B}^{\mathrm{CSI\text{-}RS}} \in \mathbb{C}$, and a predefined feedback codebook. We will assume throughout the paper that all digital beamformers are normalized according to a per-symbol power constraint, i.e. $\|\mathbf{F}\|^2 = N_{\mathrm{T}}$ and all analog beamformers are constant modulus $|\mathbf{F}_{i,j}^{\mathrm{RF}}|^2 = 1/N_{\mathrm{T}}^{\mathrm{RF}}$ for all elements $(i,j)$ to ensure comparable results across array sizes and to reflect the per-antenna power constraint that is common in practice. In addition, the analog phase shifters have a limited resolution such that each can only apply a quantized phase shift with resolution $2^{b_{\text{phase}}}$.

We will assume a working knowledge of the beam management stages and will only briefly highlight the key equations. Readers are encouraged to review [1], [6], [28] for an in-depth review of these processes in 5G. Given the identified focus on beam management for 6G [9], we will assume the same three phases (SSB, CSI-RS, and feedback) as in 5G. We also assume UEs operate similarly to sub-6GHz 5G implementations, although such simplifications are only assumed to as a simplification to focus on the BS operation. Much of the beam training and hybrid operations are not specified in the standards, so we will assume a formulation similar to classical hybrid implementations found in literature i.e. [3], [32], [33].

### A. Synchronization Signal Burst

The SSB process is an initial access procedure where a limited number of broadcast beams are transmitted periodically to allow UEs to connect to the network [1], [6]. Although the first SSB codebooks were designed for wide-area coverage, the concepts of "narrow" and "wide" beams do not make sense in multipath environments, especially with extremely large arrays. Instead, the SSB and CSI-RS codebooks should be designed in a hierarchical format to find more precise beamformers that improve metrics such as the received signal power or data rate. With this in mind, we overview the SSB signal model.

The SSB signals are transmitted over a consistent band $K^{\mathrm{SSB}}$ and predefined timeslots $T_i^{\mathrm{SSB}}$ for each SSB $i$ up to the total SSB size $L_{\max}$. The received signal during SSB reception before combining can be simplified from (2) to a single-stream representation of the demodulation reference signals $s_{t,k}^{\mathrm{DMRS}} \in \mathbb{C}$ as

$$\mathbf{y}_{u,t,k}^{\mathrm{SSB}_i} = \frac{1}{\sqrt{K N_{\mathrm{T}}}}\mathbf{H}_{u,t,k}\mathbf{f}_i s_{t,k}^{\mathrm{DMRS}} + \mathbf{N}_{u,t,k} \tag{3}$$

where $\mathbf{f}_i \in \mathbf{B}^{\mathrm{SSB}}$ is the analog beamformer controlled from a single digital port. Note that a digital beamformer can

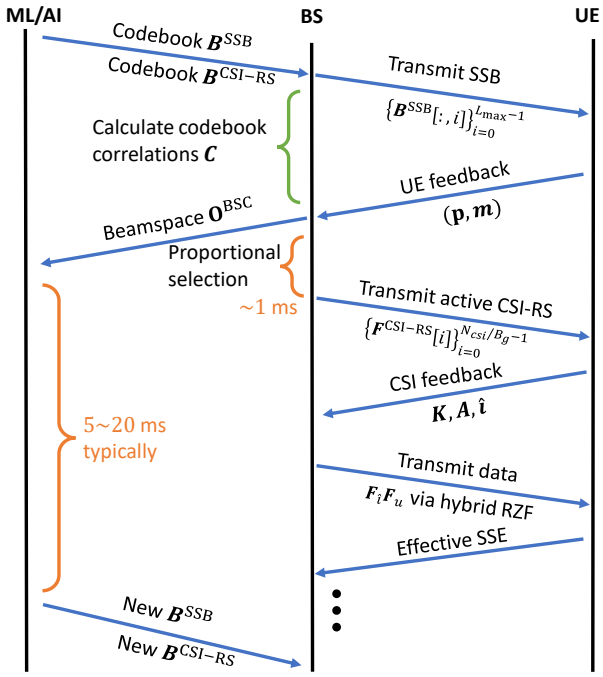



**FIGURE 2. A timing diagram indicating the connection between an ML/AI codebook algorithm, the BS operation, and the UE role in a downlink configuration. Deep neural networks are capable of being deployed in near real-time roles like codebook generation with time spans on the order of 5-20ms. Real-time operations like precoder determination and feedback quantization are challenging due to the strict $\leq$ 1ms latency requirements.**

also be used but it is most reasonably achieved through an analog rank-1 beamformer. While the limited resolution of the phase shifters $b_{\text{phase}}$ in the hybrid array has a noticeable effect during multi-stream precoder design, the effects are almost imperceptible and nearly neglectable during rank-1 beamforming, even with low-resolution phase shifters [29]. Because the focus is on the BS operation, we will assume the UE performs digital combining to maximize the SSB reference signal received power (RSRP) as

$$\text{RSRP}_{i,u} = \sum_{k\in K^{\text{SSB}}} \sum_{t\in T_i^{\text{SSB}}} \left\| \mathbf{y}_{u,t,k}^{\text{SSB}_i} \right\|^2. \tag{4}$$

It is common at high frequencies to employ beam training at the UE instead of digital combining, but we leave such a scenario to future work as it is unclear whether upper-mid band 6G will operate more like sub-6GHz 5G or millimeter-wave 5G. Similarly, it is unknown if additional SSB feedback will be added in 6G, but the minimum SSB feedback corresponds to the best beam index and the RSRP of that beam. This can be aggregated at the base station as

$$\mathbf{p} = \left\{ \max_i \text{RSRP}_{i,u} \right\}_{u=1}^{U} \tag{5}$$

$$\mathbf{m} = \left\{ \arg\max_i \text{RSRP}_{i,u} \right\}_{u=1}^{U}. \tag{6}$$

The vectors $\mathbf{p}$ and $\mathbf{m}$ correspond to the best beam index (6) and the corresponding RSRP (5) for all $U$ users. After the SSB feedback, the BS can transmit CSI-RS to obtain more accurate beam alignment and channel estimates in the form of PMI.

### B. CSI-RS

CSI-RS is the second beamforming stage where the previous SSB feedback is used to refine the beamformers to achieve a high SNR for analog beam training and channel estimation. In the current specifications (3GPP R17), the CSI-RS codebook is large, and a subset of the candidate beamformers are selected based on the SSB feedback [28]. We define $\mathbf{B}^{\text{sub}}$ as the active CSI-RS codebook and the selection algorithm, $\text{Sel}(\cdot)$, that determines the $N_{\text{CSI}}$ active beamformers from $\mathbf{B}^{\text{CSI-RS}}$ based on the SSB codebook and SSBRI feedback $\mathbf{m}$. The subset selection can be formalized as

$$\mathbf{B}^{\text{sub}} = \text{Sel}(\mathbf{m}, \mathbf{B}^{\text{SSB}}, \mathbf{B}^{\text{CSI-RS}}, N_{\text{CSI}}) \tag{7}$$

**CSI-RS subset selection:** An important challenge during CSI-RS is that the selection operates on a short timescale–on the order of 1ms or less. Because of this limitation, neural networks cannot easily be employed for selection; typical neural network execution times with modern hardware are on the order of 10us to 100ms [34], [35]. Historically, the CSI-RS selection simply divides the angular space of the SSB into a proportional number of beams compared to the total allocation of CSI-RS beams, although the exact selection algorithm is left up to private implementations. Given that the learned beams do not represent simple angular spaces, we propose a novel proportional selection process using the cross-correlations of the SSB and CSI-RS codebooks. In the case of DFT beams, the proposed proportional selection produces the same results as angular division but it also extends to arbitrary, learned codebooks as well. The proportional selection algorithm begins by calculating the cross-correlations $\mathbf{C}$ between each of codebook entries, which can be performed while awaiting SSB feedback

$$\mathbf{C} = \left\langle \mathbf{B}_i^{\text{SSB}}, \mathbf{B}_j^{\text{CSI-RS}} \right\rangle \quad \forall i,j. \tag{8}$$

Then the number of CSI-RS beams to be allocated to each SSB reference beam is determined as a proportion of the total number of CSI-RS based on how many users reported a given SSB beam selection $\mathbf{m}$

$$\mathbf{m}^{\text{sel}} = \left\{ \left\lfloor \frac{N_{\text{CSI}}}{U} \sum \mathbf{1}_{\mathbf{m}=i} \right\rfloor \right\}_{i=0}^{i=L_{\max}-1}. \tag{9}$$

Finally, for each of the SSB beams, the $\mathbf{m}_i^{\text{sel}}$ most correlated beams are assigned to the active subset. In other words, the selection algorithm first determines how many of the total CSI-RS beams should "correspond" to each of the SSB beams, and then the most correlated CSI-RS beams for each SSB beam are selected. This can be done based on a column-wise argsort operation, the cumulative sum operation

(cumsum) and vectorized selection

$$\mathbf{A}^{\text{sorted}} = \text{argsort}(\mathbf{C}) \tag{10}$$

$$\mathbf{v}_i = \text{cumsum}(\mathbf{m}_{:i}^{\text{sel}}) \quad \forall i \in \{0, 1, ..., L_{\max} - 1\} \tag{11}$$

$$\mathbf{B}^{\text{sub}}_{\mathbf{v}_{i-1}:\mathbf{v}_i} = \mathbf{B}^{\text{CSI-RS}}[\mathbf{A}^{\text{sorted}}_{0:\mathbf{m}_i^{\text{sel}}}]. \tag{12}$$

The result is a subset of $N_{\text{CSI}}$ CSI-RS beams that are proportionally selected to be highly correlated with the active SSB beams reported. Although calculating the cross-correlations involves significant computation, these operations can be performed as soon as the codebooks are available rather than waiting for SSB feedback. The timing critical operations only require at most $UL_{\max} + 1$ FLOPs for proportioning (9) and $L_{\max}$ FLOPs for the cumulative sum with negligible operations for indexing in (12). Therefore, even resource-constrained systems are capable of performing the proportional selection in real time.

Before discussing the received CSI-RS, it is important to understand the role of logical ports when it comes to hybrid arrays and CSI-RS. Logical ports simply correspond to a set of ports that transmit a signal over a consistent channel. It does not correspond to a specific or unique subset of analog or digital ports in a hybrid system, but rather, allows for a separation between the hardware and a representation of the channel. Although this logical-physical port separation is flexible, this formulation leaves many implementation details out of the specifications. In order to follow traditional hybrid systems, while still keeping the flexibility, we associate three parameters with the CSI-RS: $N_{\text{P}}$ is the total number of ports that can be configured which is typically the digital dimension of the hybrid array, $N_{\text{CSI}}$ is the number of CSI-RS beamformers to be transmitted, and an additional parameter $B_{\text{g}}$ which is the number of ports and beams assigned to one CSI-RS allocation. The connection between the parameters is that $B_{\text{g}}$ CSI-RS beams are transmitted within one resource so that $\lceil N_{\text{CSI}}/B_{\text{g}} \rceil$ total CSI-RS resources must be allocated and up to $\lfloor N_{\text{P}}/B_{\text{g}} \rfloor$ users can be multiplexed together during the subsequent data transmission. The separation here allows for "assigning" portions of the logical array of size $B_{\text{g}}$ to each user and users can select from $\lceil N_{\text{CSI}}/B_{\text{g}} \rceil$ resources to feedback beamformed channel estimates from.

With the active CSI-RS codebook determined, the CSI-RS process involves transmitting training symbols $\mathbf{s}^{\text{tr}}_{t,k} \in \mathbb{C}^{B_{\text{g}} \times 1}$ for channel estimation using each selected beamforming codeword, grouped into $B_{\text{g}}$ beams i.e. $\mathbf{F}_i \subseteq \mathbf{B}^{\text{sub}}$ for each CSI-RS allocation $i$. Note that the symbols and beamformers may be one-hot or code-division multiplexed across the time-frequency resources of the CSI-RS depending on the BS-selected settings. The received signal for each UE is then

$$\mathbf{y}^{\text{CSI-RS}_i}_{u,t,k} = \frac{1}{\sqrt{KN_{\text{T}}}} \mathbf{H}_{u,t,k} \mathbf{F}_i \mathbf{s}^{\text{tr}}_{t,k} + \mathbf{N}_{u,t,k}. \tag{13}$$

From the received signal, the UE can determine the RSRP, estimated SNR, and select the CSI-RS indicator $\hat{\mathbf{i}}$ as

$$\text{RSRP}^{\text{CSI-RS}_i}_{i,u} = \max_{n_{\text{r}}} \sum_{k \in K^{\text{CSI-RS}_i}} \sum_{t \in T^{\text{CSI-RS}_i}} \left\| \mathbf{y}^{\text{CSI-RS}_i}_{u,t,k} \right\|^2 \tag{14}$$

$$\text{SNR}_{i,u} = \max_{n_{\text{r}}} \frac{1}{KN_{\text{T}}} \sum_{k \in K^{\text{CSI-RS}_i}} \sum_{t \in T^{\text{CSI-RS}_i}} \frac{\left\| \widehat{\mathbf{H}_{u,t,k} \mathbf{F}_i} \right\|^2}{\sigma^2} \tag{15}$$

$$\mathbf{F}_{\hat{i}_u} = \arg\max_i \text{SNR}_{i,u} \tag{16}$$

The noise power, $\sigma^2 = \mathbb{E}[\|\mathbf{N}_{u,t,k}\|^2]$ can be estimated from the system parameters and zero-power CSI-RS. Compared to the SSB model (3), the CSI-RS received model is multi-layer due to the need for sounding each of the $B_{\text{g}}$ ports. The CSI-RS therefore must have at least $N_{\text{R}}$ beams for a full-rank beamformed channel, assuming $N_{\text{R}} \leq N_{\text{T}}$. There is a significant benefit to using groupings of a small number of beams to sound the channel as the overhead for both CSI-RS transmission and feedback quantization are reduced, and users can be multiplexed using different analog beams and well-designed digital precoders.

The SNR (15) impacts the channel estimation efficacy and is useful for determining the initial modulation and coding scheme (MCS) for data transmission. The beamformed channel estimate is defined for a channel estimation algorithm $w(\cdot)$ as

$$\widehat{\mathbf{H}_u \mathbf{F}_{\hat{\mathbf{i}}_u}} = w(\mathbf{y}_u^{\text{CSI-RS}_{\hat{\mathbf{i}}_u}}, \mathbf{s}^{\text{tr}}). \tag{17}$$

There is no standardized channel estimation algorithm, but, given the limited time and power budget of mobile UEs, a least squares method is assumed herein. The resulting channel estimate at the UE is frequency selective over a set of frequency subbands $b \in \{1, ...\text{S}_{\text{B}}\}$. In the final step, the channel estimates are quantized according to a feedback codebook and transmitted to the BS [1], [36].

### C. Feedback

CSI feedback is a highly configurable process to achieve a careful balance of overhead and CSI accuracy. In 5G there were two categories of feedback: type-I and type-II [1], [36]. In 3GPP Release 16 an additional format of type-II called "enhanced type-II" was standardized that included additional, frequency-domain compression. In general, the feedback formats are methods for quantizing the channel estimate according to a shared feedback codebook $\mathbf{B}^{\text{FB}}$. Type-II in particular is designed for high-resolution, MU-MIMO settings where the channel information is quantized as a set of $L_{\text{CSI}}$ oversampled DFT components, with over-sampling factors $O_{\text{H}} \times O_{\text{V}}$. We refer readers to [36] for a general description of the feedback formats and [5] for a fast implementation of type-II quantization that is important for large arrays and high resolution feedback considered in X-MIMO.

The BS receives a set of codebook indices $\mathbf{A} \in \mathbb{W}^{\mathrm{S_B} \times N_\mathrm{R} \times L_\mathrm{CSI}-1}$ and cophasing factors $\mathbf{K} \in \mathbb{C}^{\mathrm{S_B} \times N_\mathrm{R} \times L_\mathrm{CSI}-1}$ that enable a reconstruction of the beamformed channel. Channels are reconstructed in the same way as [28] i.e.

$$\widehat{\mathbf{H}_{u,b,n_\mathrm{r}}\mathbf{F}_{\hat{\mathbf{i}}_u}} = \sum_{\ell=1}^{L_\mathrm{CSI}} \mathbf{K}_{b,n_\mathrm{r},\ell} \mathbf{B}^{\mathrm{FB}}_{\mathbf{A}_{b,n_\mathrm{r},\ell}} \quad \forall b,\ n_\mathrm{r}. \tag{18}$$

The quantization error, therefore, can be reduced with large $L_\mathrm{CSI}$, but this introduces additional overhead. In addition, the frequency domain quantization introduces overhead that increases with a larger number of subband elements $\mathrm{S_B}$. The actual overhead is compressed so that the frequency-domain scaling is minimal, although the overhead of this signaling is a significant part of the precoder matrix indicator (PMI) section of the feedback report. In addition, the feedback will include information describing the rank indicator $R$ (RI), the best CSI-RS beam index $\hat{i}_u$ (CRI), and other subfields not critical to this work. The base station should then use the information from both the beam training and channel feedback to design precoders to best serve the users.

### D. Data transmission

In the final steps, the BS uses the CSI acquired through the UE feedback to determine the precoders to serve a set of users. The exact process is not standardized; we take a standard approach here inspired by MIMO fundamentals [37]. The analog beamforming for the hybrid array is controlled through the CRI from the feedback packet while the digital precoder is used to mitigate interference between users and streams. A typical precoder formulation is a regularized zero-forcing precoder (RZF) determined to maximize the signal-to-leakage noise ratio [37, section 9.9]. The RZF precoder with perfect channel state information is

$$\mathbf{F}_{u,t,k} = \frac{(\sum_{i=0}^{U-1} \mathbf{H}^*_{i,t,k}\mathbf{H}_{i,t,k} + UN_\mathrm{T}\mathbb{E}[\mathbf{N}^2_{u,t,k}])^{-1}\mathbf{H}^*_{u,t,k}}{\left\|(\mathbf{H}^*_{i,t,k}\mathbf{H}_{i,t,k} + UN_\mathrm{T}\mathbb{E}[\mathbf{N}^2_{u,t,k}])^{-1}\mathbf{H}^*_{u,t,k}\right\|}. \tag{19}$$

In the case of realistic, hybrid arrays the perfect channel state information $\mathbf{H}$ is replaced with the selected beamformed channels, redefined for ease of notation as $\widehat{(\mathbf{HF})} \equiv \{\widehat{\mathbf{H}_{u,b,n_\mathrm{r}}\mathbf{F}_{\hat{i}_u}}\}_{u,b,n_\mathrm{r}} \in \mathbb{C}^{U \times \mathrm{S_B} \times N_\mathrm{R} \times B_\mathrm{g}}$. The beamformed channels arise from the beam grouping and port assignments from Section B. Without loss of generality, we assume the first $U_a \leq N_\mathrm{Td}/B_\mathrm{g}$ users are allocated for data transmission so that the analog $\{\mathbf{F}_{\hat{i}_u}\}$ and digital precoders $\{\mathbf{F}_{u,t,k}\}$ are designed in a subarray-formulation [3] as

$$\mathbf{F}^{\mathrm{RF}} = [\mathbf{F}_{\hat{i}_0}, \mathbf{F}_{\hat{i}_1}, ..., \mathbf{F}_{\hat{i}_{U_a}}] \tag{20}$$

$$\mathbf{F}_{t,k} = \begin{bmatrix} \mathbf{F}_{0,t,k} & 0 & ... & 0 \\ 0 & \mathbf{F}_{1,t,k} & 0 & ... \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \mathbf{F}_{U_a,t,k} \end{bmatrix}. \tag{21}$$

This formulation neatly unifies hybrid arrays, traditional precoding, and beam management techniques with the received signal model from (2). Because data transmission is not standardized, we propose this formulation as an intersection of traditional hybrid operations and the information available during beam management.

On the receiver side, the UEs also perform a combining strategy based on the embedded DMRS within the signal allowing for decoding each data stream without knowing the exact precoder. In this work, we will assume an LMMSE receiver is used to maximize the signal-to-interference noise ratio (SINR) for any precoder, although the actual SINR and subsequent data rate calculations are not actually performed by the UE but are necessary to evaluate the achievable spectral efficiency in simulation. The choice of LMMSE receiver here makes sense to ensure allocated users have high performance for receiving the signal while still balancing the computational cost. The SINR at resource element $(t, k)$ for a given user $u$ and data stream $r$ with equal power allocation per user is

$$\mathbf{D}_{u,t,k} \triangleq \mathbf{H}_{u,t,k}\mathbf{F}_{\hat{\mathbf{i}}_u}\mathbf{F}_{i,t,k} \tag{22}$$

$$\mathbf{I}_{u,t,k} = \left(\sum_{i=0}^{U-1} \mathbf{D}_{u,t,k}\mathbf{D}^*_{u,t,k} + UN_\mathrm{T}\mathbb{E}[\mathbf{N}^2_{u,t,k}]\right)^{-1} \tag{23}$$

$$\mathrm{SINR}_{u,t,k,r} = \frac{1}{K}\frac{\mathbf{D}^*_{u,t,k,r}\mathbf{I}_{u,t,k}\mathbf{D}_{u,t,k,r}}{1 - \mathbf{D}^*_{u,t,k,r}\mathbf{I}_{u,t,k}\mathbf{D}_{u,t,k,r}}. \tag{24}$$

The SINR expression (24) is a ratio between the signal power of the desired data stream $(u, r)$ compared to the noise and interference power of all other streams in (23).

We evaluate the performance in the network by the sum spectral efficiency (SSE), which is reasonable when there is not more complicated scheduling, fairness constraints, and resource allocation. The sum spectral efficiency can be multiplied by the bandwidth to get the sum rate. Assuming Gaussian signaling and treating interference as Gaussian noise, the achievable spectral efficiency, $\mathrm{SE}_{u,t}$, is

$$\mathrm{SE}_{u,t} = \sum_{k=0}^{K-1}\sum_{r=0}^{R-1} \log_2(1 + \mathrm{SINR}_{u,t,k,r}). \tag{25}$$

The SSE is a sum over the active users, but it does not account for the overhead of beam management or CSI acquisition. In the final step, we consider the effective SSE (ESSE), which accounts for the overhead due to beam training by removing the corresponding time/frequency resources due to training and feedback of the entire beam management system $(T_\mathrm{BM}, K_\mathrm{BM})$ from the spectral efficiency calculation

$$\mathrm{ESSE} = \sum_{u=0}^{U-1}\sum_{\substack{t\notin T_\mathrm{BM} \\ k\notin K_\mathrm{BM}}}\sum_{r=0}^{R-1} \log_2(1 + \mathrm{SINR}_{u,t,k,r}). \tag{26}$$

Although presented in a linear fashion, the SSB, CSI-RS, feedback, and data transmission typically occur over separate resources so that there are no periods of waiting in a full-buffer scenario. With the system model defined, we now

present the core machine learning architecture, X-BM that integrates ML within the codebook design process. The proposed strategy integrates in a realistic fashion such that the system does not require additional side information and the overall architecture is designed to meet the timing constraints implicit within beam management.

## III. Neural codebook design

We propose a novel approach for jointly designing the SSB and CSI-RS codebooks using AI/ML. The primary challenges with integrating machine learning into wireless domains ultimately come down to the tight timing constraints (see Figure 2) and generalizing across varying numbers of users and system configurations. Timing constraints are not particularly concerning for SSB codebook generation, as the periodicity is typically 20ms with a minimum of 5ms. In comparison, the CSI-RS codebook needs to be available within 1ms following the SSB feedback, however, the CSI-RS selection depends on the SSB feedback. It is extremely challenging for modern hardware to employ deep neural networks with such tight, sequential timing constraints [34], [35]. Therefore, we propose the neural architecture, shown in Figure 3, to output the SSB codebook and a set of candidate CSI-RS codebooks in parallel at each SSB period. Then we employ the proportional selection technique from (8)-(12) to determine the CSI-RS codebook from the candidate options, allowing the joint algorithm to overcome the timing limitation and still incorporate the SSB feedback. In preparation for 6G, this operation is similar to current 5G CSI-RS codebooks but extends to arbitrary and time-varying SSB and CSI-RS codebooks. The entire structure is trained in an end-to-end fashion with the goal of minimizing the power loss with respect to singular value decomposition beamforming, which is the power-maximizing strategy for a single user with perfect CSI and rank-1 beamforming. This ensures that we do not have to define a specific beamforming pattern train against, but rather have the network learn patterns that are close to power-maximizing.

The input to X-BM is prepared by converting the prior SSB codebook and SSB feedback to the beamspace. The input only contains previously known users, so the learning algorithm is partially regularized by new users who join the network in the upcoming SSB cycle and therefore are not included in the prior observation. The beamspace is a simple and highly extensible conversion where beamformers are converted to the angular domain with fixed resolution. This format is beneficial for translating beamformer information over time, antenna geometries, and UE numbers because the beamspace is largely unaffected by these parameters [23], [28], [38]. This ensures that the input from a set of previous beamformers represents a consistent two-dimensional grid of projections. This step was originally developed in our prior work [28].

First, we define an angular transformation matrix for antenna size $N_{\mathrm{s}}$ and $N_{\mathrm{m}}$ angular directions as a series of array responses

$$\boldsymbol{\theta}_{N_{\mathrm{m}}} = \frac{1}{\pi}[0, 1, ..., N_{\mathrm{m}} - 1]^T \tag{27}$$

$$\mathbf{U}_{N_{\mathrm{s}},N_{\mathrm{m}}} \triangleq [\mathbf{a}_{N_{\mathrm{s}}}(\boldsymbol{\theta}_0), ...\mathbf{a}_{N_{\mathrm{s}}}(\boldsymbol{\theta}_{N_{\mathrm{m}}-1})]. \tag{28}$$

The beamspace conversion for $N_{x0}$ azimuth directions and $N_{y0}$ elevation direction is calculated as

$$\mathbf{F}_i^{\mathrm{SSB}} = [\mathbf{B}^{\mathrm{SSB}}_{0:N_{\mathrm{X}}^{\mathrm{RF}},i}, ..., \mathbf{B}^{\mathrm{SSB}}_{(N_{\mathrm{X}}^{\mathrm{RF}}-1)N_{\mathrm{Y}}^{\mathrm{RF}}:N_{\mathrm{X}}^{\mathrm{RF}}N_{\mathrm{Y}}^{\mathrm{RF}},i}] \tag{29}$$

$$\mathbf{O}_i^{\mathrm{BSC}} = \mathbf{U}^*_{N_{\mathrm{X}}^{\mathrm{RF}},N_{x0}} \mathbf{F}_i^{\mathrm{SSB}} \mathbf{U}_{N_{\mathrm{Y}}^{\mathrm{RF}},N_{y0}} \quad \forall i. \tag{30}$$

The codebook must be reshaped from a vector of size $N_{\mathrm{T}}^{\mathrm{RF}}$ to the planar dimensions $N_{\mathrm{X}}^{\mathrm{RF}} \times N_{\mathrm{Y}}^{\mathrm{RF}}$ in (29) before the beamspace conversion in (30) that produces the beamspace observation, $\mathbf{O}^{\mathrm{BSC}}$. Note that the beamspace conversion is a reversible operation via the pseudo inverse so long as the number of samples in each direction is equal to or greater than the number of antennas, so we also train the network to predict the beamspace of the desired output, rather than direct beamforming coefficients to aid the learning process. The predicted beamformers are obtained within the training loop, and the computational complexity can be controlled by changing the observation sizes $N_{x0}$ and $N_{y0}$. In addition to the angular representation, the input is also concatenated with the feedback corresponding to the number of users reporting each beam and the sum RSRP. We can express this using the notation $\mathbf{1}_{(\cdot)}$ to refer to a vector with all 0 entries except the indices where the subscript is true contain a 1 and resulting in the input

$$\mathbf{O}^{\mathrm{BSC}} = \left\{ \left[ \mathbf{O}^{\mathrm{BSC}}[i], \sum \mathbf{1}_{\mathbf{m}=i}, \mathbf{1}^T_{\mathbf{m}=i}\mathbf{p} \right] \right\}_{i=0}^{L_{\max}-1}. \tag{31}$$

This input is generated during each SSB period and fed into the X-BM algorithm as shown in Figure 3. The output of the algorithm is an SSB codebook and a CSI-RS codebook that can be subset using the proportional selection algorithm.

The overall system is designed and trained in an end-to-end fashion with respect to maximizing the RSRP. End-to-end training changes the learning goal with data-driven methods to maximize or minimize a system-level metric, rather than training toward a specific output [39]. In particular, we do not identify the exact beamformers we want the neural network to learn, instead relying on the network to learn the beamformers that achieve the highest RSRP. It is difficult to exactly define an "optimal" beamformer formulation in general for the SSB and CSI-RS steps. This arises from the fact that even with perfect CSI, the broadcast beamformers that achieve the highest SNR with a limited number of beams depend heavily on the system parameters, user channels, and number of beamformers used. By employing end-to-end learning, we instead train the neural network to learn beamformers that achieve the highest performance relative to single-user singular value decomposition (SVD) beamforming with perfect CSI. This formulation is even more

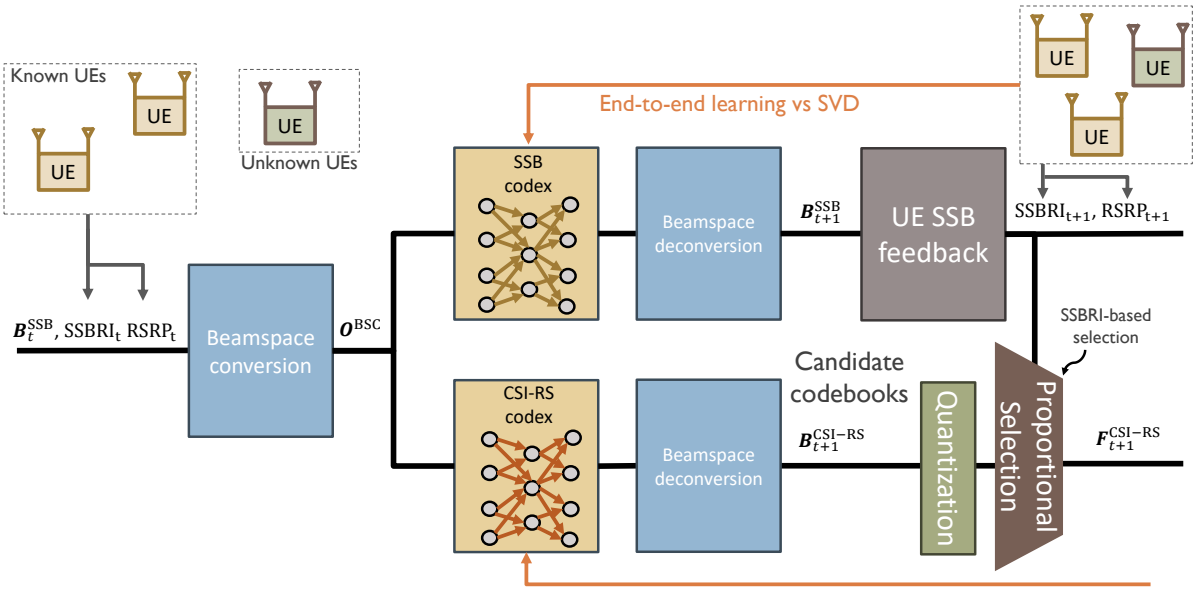



**FIGURE 3.** **A depiction of the X-BM algorithm and gradient propagation. The beamspace is prepared using the previous SSB codebook and known UE feedback and fed into two parallel networks that generate the new SSB codebook and CSI-RS codebook. The CSI-RS codebook is then quantized and a proportional selection according to the new SSB feedback is used to determine the active CSI-RS codebook. Gradients are calculated according to the difference between the UE-received power and the achievable SVD received power for all users.**

advantageous in the initial access scenario where not all users are known by the system prior to codebook generation. The general learning framework is shown in Figure 3 with a loss function comparing the SVD power and the reported RSRP for each user as

$$\mathbf{U}_u, \mathbf{S}_u, \mathbf{V}_u = \mathbf{H}_u \quad \forall u \in \{1, ...U\} \tag{32}$$

$$\mathcal{L}(\mathbf{S}, \mathbf{p}) = \frac{1}{U} \sum_u (10 \log_{10}(\mathbf{S}_{u,0}^2) - 10 \log_{10}(\mathbf{p}_u^2)) \tag{33}$$

where $\mathbf{p}$ is the RSRP from (5) or (14) for the SSB or CSI-RS codebooks. Gradients are calculated from the loss function (33) throughout the network and parameters are updated using an Adam optimizer [40] with cyclic learning rate scheduling between $10^{-3}$ and $10^{-5}$.

One challenge of codebook design for X-MIMO is the massive physical array size, which results in a significant number of parameters for generating codebooks. We mitigate this with efficient neural structures like convolutional layers (See Table 1) that share parameters and efficiently operate over the beamspace representation (30). In fact, this formulation allows the neural network to learn from a universal pattern that allows for training and testing on any array geometries, so long as the beamspace dimensions are equal or greater to the antenna dimensions to prevent aliasing [28]. We examine how the network generalizes across geometries in Figure 8. It is noteworthy that further parameter reduction can be achieved using convolution transpose layers instead of a fully connected layer within the network, although we found this caused a slight performance drop in our previous work [28]. Refining, distilling, or pruning the network are techniques that can be employed to minimize the hardware resource needed during deployment [41].

In addition to hardware considerations for deployment, algorithm periodicity is also a potential source of operator modification. The current formulation is set up to enable a fast, supervised learning strategy where each training step can be performed independently, rather than treated as a sequential time series [28]. This means that the codebooks can be generated every SSB period, or after a series of periods if the channel time dynamics are slow compared to the SSB period. For example, fixed wireless is an application where the codebooks would not need to be updated. Even running the algorithm at every period, as it's used here, would produce the same codebook given that the inputs would be constant over time.

## IV. Data generation framework

The system-level simulation requires two stages: 1) spatially and consistent channel generation for multiple users and 2) beam management integration of the iterative steps for initial access, beam refinement, feedback, and data transmission. We start by generating raytraced channels using Sionna [31]. In our simulations, we consider two environments (labeled A and B) with 10000 potential users scattered over a BS sector site. The BS is equipped with an $N_{\mathrm{X}}^{\mathrm{RF}} = 16$, $N_{\mathrm{Y}}^{\mathrm{RF}} = 16$ planar array using a 3GPP 3D antenna model with half-wavelength spacing and $N_{\mathrm{X}} = 8$ and $N_{\mathrm{Y}} = 4$ with $b_{\mathrm{phase}} = 2$ bit phase shifters facing the serving sector region. Each user has $N_{\mathrm{R}} = 4$ antennas tuned for a 10GHz carrier frequency in environment A and tuned for a 20GHz carrier

**TABLE 1. XBM neural architectures**

| Layer | Primary Parameter | Activation | Output Dimension |
|---|---|---|---|
| Conv+Max Pool | 128 Filters | ReLU | $\lceil (N_{x0}+2)/2 \rceil \times \lceil (N_{y0}+2)/2 \rceil \times 128$ |
| Conv+Max Pool | 96 Filters | ReLU | $\lceil (N_{x0}+2)/4 \rceil \times \lceil (N_{y0}+2)/4 \rceil \times 96$ |
| Dropout | 0.3 Rate | | $\lceil (N_{x0}+2)/4 \rceil \times \lceil (N_{y0}+2)/4 \rceil \times 96$ |
| Conv | 320 Filters | ReLU | $\lceil (N_{x0}+2)/4 \rceil \times \lceil (N_{y0}+2)/4 \rceil \times 320$ |
| Flatten | | | $(\lceil (N_{x0}+2)/4 \rceil)(\lceil (N_{y0}+2)/4 \rceil)(320)$ |
| Dropout | 0.1 Rate | | $(\lceil (N_{x0}+2)/4 \rceil)(\lceil (N_{y0}+2)/4 \rceil)(320)$ |
| Fully Conn. | $2N_{x0}N_{y0}N$ Neurons | | $2N_{x0}N_{y0}N$ |
| Reshape | | | $N_{x0} \times N_{y0} \times 2N$ |

frequency in environment B to consider low-band and high-band frequencies for upper-mid bands [2]. Users are assumed to have random mobility patterns with an average speed of 3m/s in environment A and 10m/s in environment B and arrays oriented vertically. Environment A models the area around the Frauenkirche in Munich with a BS placed on a building 40m tall and Environment B models the area around the Arc de Triomphe in Paris with the BS at a height of 27m [31] (See Figure 4. The user channels are sampled at 1ms intervals over 100MHz with 270 resource blocks assuming 30kHz subcarrier spacing. After a database of channels is generated, the neural network is integrated into a processing pipeline following the system model. The neural network is trained on $40,000$ training samples with a random number of users and varied locations throughout the scene. The validation dataset is drawn from the same distribution and used to determine when the network training is complete. All test results come from newly generated channels according to the test environment, which may be the same or a different environment from the training set.

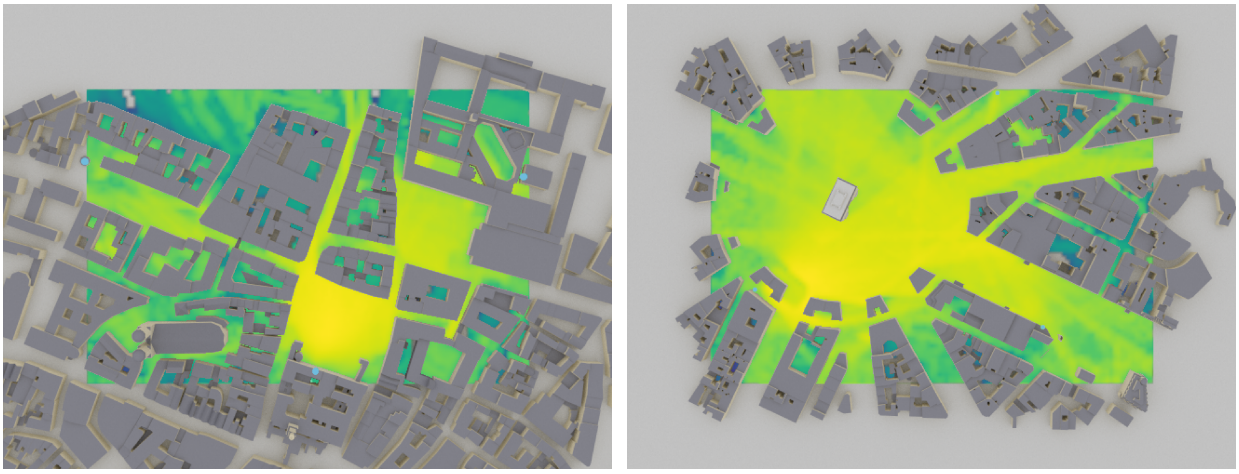
The data processing begins by first selecting a random number of active users $U \sim \mathcal{U}[4, 16]$ for the subsequent timestep and drawing that many users' channels randomly from the channel database. Then, the beamspace is calculated using the prior SSB codebook (initially set as random DFT codebook beams) with $80\%$ of the active users reporting. The remaining $20\%$ are assumed to be new users that are not known prior to the current SSB cycle and helps prevent the neural network from overfitting to known users. The new SSB and CSI-RS codebooks are generated and the current iteration begins with the SSB transmission and feedback $(\mathbf{p}, \mathbf{m})$ according to Section A. Then, the active CSI-RS codebook $\mathbf{B}^{\text{sub}}$ is selected according to the proportional correlation selection, beams are grouped into sets of length $P_{\text{CSI}}$ to form beamformed channels and transmitted. During data transmission, we assume the network calculates the RZF precoders based on the feedback and estimates the expected SSE for each combination of UEs, and selects the highest SSE user set. This information is used as a simple, realistic scheduling algorithm using the CSI available. The resulting SSB RSRP, CSI-RS SNR, and ESSE metrics are used to evaluate the proposed X-BM from a modular and system-level context.

## V. Simulation results

Given the capabilities for overfitting with neural networks, we carefully evaluate the proposed system across different scenarios, array geometries, and out-of-distribution settings. From these evaluations, we seek to better understand the advantages and limitations of a neural codebook algorithm. Unless specified, the default parameters for the results are $L_{\text{max}} = 16$ SSB beams, $N_{\text{CSI}} = 16$ CSI-RS beams, $B_{\text{g}} = 4$ logical ports or CSI-RS beams per CSI-RS resource, $N_{\text{RB}} = 24$ resource blocks occupied by a CSI-RS resource, and $\text{S}_{\text{B}} = 8$ subbands for PMI.

### A. SSB Performance

The first evaluation focuses on the RSRP performance of the proposed algorithm. The RSRP is evaluated using both the SSB and CSI-RS codebooks, although the CSI-RS relies on the SSB feedback and should in general outperform the SSB beamforming given the additional information available to the system. In Figure 5 we show the empirical probability distribution function (PDF) of the reported RSRP using the proposed X-BM, DFT codebooks, and the neural architecture from our prior work [28]. To compare the results, we show the relative power-loss of the RSRP achieved with the codebook method compared to SVD beamforming with perfect CSI (CSI-SVD). In the first results, we show the RSRP performance with $L_{\text{max}} = 16$, which is more than supported in sub-6GHz 5G, but less than the $L_{\text{max}} = 64$ supported by mmWave 5G. It can be seen that the performance of X-BM is almost always within 3dB of SVD beamforming and achieves a gain of over 6dB on average compared to DFT codebooks. Especially interesting, we see that there is very little gain from the CSI-RS with $N_{\text{CSI}} = 32$ for the majority of users, with only the bottom $20\%$ of users seeing noticeable improvements. We also note that the performance of the beamspace codex (BSC) is modestly more performant than DFT codebooks, but ultimately suffers from the difference in learning objective. In particular, learning SVD beamformers is more challenging so reducing the training data to the



(a) Environment A (b) Environment B

**FIGURE 4. A comparison of the powermaps in two environments with different base station heights and carrier frequencies. Environment A corresponds to Munich, Germany operating at $10$GHz and with the BS mounted at $40$m high. The alternate setting in Environment B is Paris, France operating at $20$ GHz, with a $27$m high BS. A low resolution power map is rendered on each scene to highlight possible user locations.**

current $40,000$ samples limits performance while showing the X-BM sample efficiency.

In Fig. 6, the performance with respect to different SSB codebook sizes $L_{\max}$ is plotted. An interesting result of these simulations (Figures 5-6) is that the X-BM SSB performance often outperforms even the CSI-RS performance of DFT codebooks. From a beam training perspective, there appears limited need for using CSI-RS beams when employing X-BM codebooks, because the SSB codebooks are already well-performing. Although a hierarchical search provides significant gains with DFT codebooks, it can be seen that the X-BM SSB codebook typically achieve a sufficient beam alignment. In particular, X-BM achieves better beam training RSRP with $L_{\max} = 8$ beamformers than the CSI-RS beams achieve from a $4$ times larger DFT codebook, although these results can still be improved with a larger X-BM codebook as shown in Figure 6.

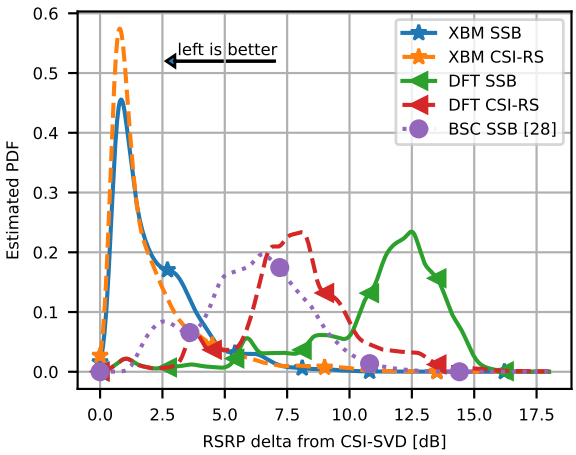



**FIGURE 5. Comparison of the beamforming codebook performance relative to perfect CSI beamforming with $L_{\max} = 16$ and $N_{\text{CSI}} = 16$ beams. $5000$ sets of active users are drawn and each codebook method is used to select the SSB (solid lines) and CSI-RS (dashed lines) codebooks. The RSRP is then calculated and the difference from SVD beamforming is shown as an estimated density function. The X-BM codebooks are extremely effective, with more than $75\%$ of users within $3$dB of perfect CSI and outperforming the prior work on ML codebook algorithms [28]. Even compared to much larger DFT codebooks for CSI-RS, there is a substantial performance improvement with the proposed X-BM method.**

### *B. CSI-RS SNR*

Following the SSB transmission, the candidate CSI-RS codebook is selected based on the SSBRI and transmitted. The first goal is to achieve a high channel estimation SNR, while also ensuring the beamformer can be employed for hybrid beamforming with digital precoding. An inherent challenge of the design and operation of CSI-RS codebooks is the dependence on the SSB codebook and the extensive overhead associated with larger active codebooks. It is often advantageous to employ a larger number of SSB beams, which require relatively low overhead, in combination with a limited number of CSI-RS beams that consume significant wireless resources. At the same time, the SSB periodicity is higher than CSI-RS–typically about $4$ times faster–so there is additional overhead to consider. Furthermore, the design challenge is not simply to maximize SNR, but also to obtain an analog precoder that performs well for the hybrid data transmission.

We begin by addressing the first question of achieving a high-gain CSI-RS codebook. In Figure 7 we compare the CSI-RS SNR obtained from classic DFT codebooks with varying sizes compared to the number of active users. We can see that when the number of users is much larger

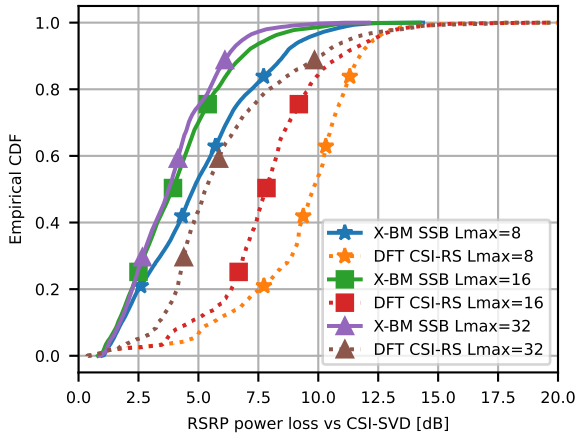



**FIGURE 6.** **Comparison of the X-BM SSB RSRP with different sizes of codebooks ($L_{\max}$) versus CSI-RS DFT codebooks. The DFT CSI-RS codebooks always include $N_{\text{CSI}} = 32$ active beams selected based on the proportional selection algorithm with the reported SSBRI. The proposed X-BM SSB codebooks outperform $4$ times larger SSB+CSI-RS DFT codebooks.**

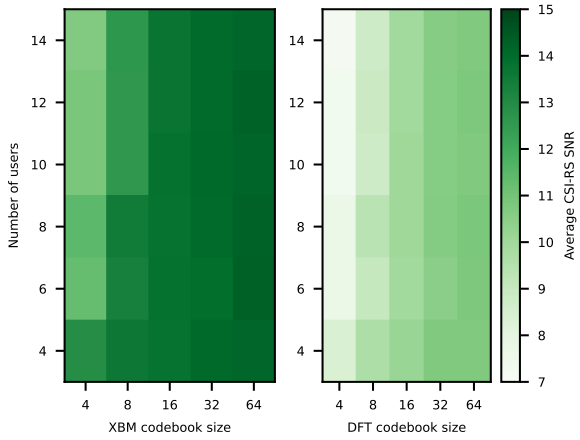



**FIGURE 7.** **A heatmap of the average SNR during CSI-RS reception with different numbers of active CSI-RS beams $N_{\text{CSI}}$ and active users $U$. While increasing the DFT codebook size provides a significant gain for more than $4$ users, the rich multipath environment enables better performance with X-BM codebooks, even with $1/16$ as many CSI-RS beams. This is a limitation of DFT beamformers in low- and mid-band environments.**

than the number of beams, the X-BM method performs significantly better than DFT codebooks. We can see that with just $1$ beam per active user, the X-BM codebooks reach a consistent performance gain of about $4$dB over DFT codebooks of any size. This can be attributed to the multipath propagation environment that does not match DFT beamformers in general.

### *C. Geometry translation*

One of the beneficial aspects of the X-BM structure is a natural translation across array geometries. In particular, the neural network size and parameters do not depend on the antenna array, so the same neural network can be employed

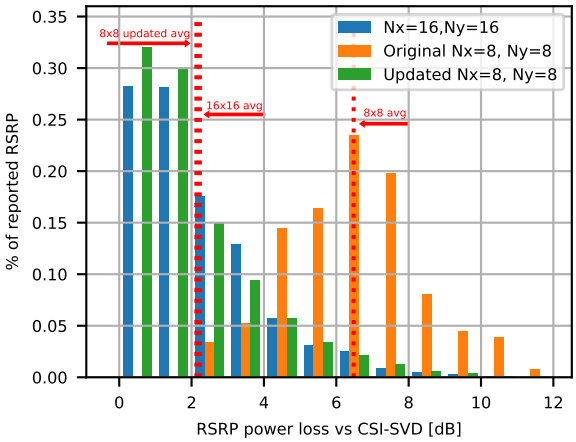



**FIGURE 8.** **A comparison of two X-BM networks for geometry translation. The X-BM models are trained using data from arrays equipped with $N_{\text{X}} = \{8, 16\}$ and $N_{\text{Y}} = \{8, 16\}$ and beamspaces of size $N_{x0} = 16, N_{y0} = 16$, but both are evaluated with $N_{\text{X}} = 16$, $N_{\text{Y}} = 16$ and $L_{\max} = 16$. Initially, there is a performance drop of about $4$dB, but fine-tuning restores the performance to that of a model trained specifically for the array geometry. This highlights the effectiveness and generalization capability of the beamspace conversion.**

or trained across geometries. Figure 8 shows a comparison of the performance of geometry translation, assuming the arrays are placed in the same scene and location. We focus on the more difficult setting of training a neural network on data from a smaller array than it is tested on. It can be seen that there is a drop in performance when testing models trained on $8 \times 8$ arrays in $16 \times 16$ settings, although the performance is still better than the DFT codebooks in Figure 5. In general, this is a limitation of deep learning in that it does not automatically generalize well, even with carefully designed formulations. At the same time, there is a distinct advantage over traditional codebooks, showing that the beamspace formulation is capturing consistent patterns in the beamspace. By applying domain knowledge to convert codebooks to the beamspace and back, the neural architecture can naturally capture patterns through the lens of the MIMO arrays.

While the performance loss is noticeable, the benefit to the X-BM architecture is that the same neural network can be trained with data from different array sizes. This allows network operators to only need to collect data once and then fine-tune after selecting the hardware to maximize performance. To show the fine-tuning performance, we include the original $8 \times 8$ model that has been retrained on the $16 \times 16$ data for 30s, highlighting how effective fine-tuning is for resolving the geometry translation loss. In this setting, fine-tuning is performed by simply training on new data with the same training steps, however, we will consider fine-tuning on imperfect CSI in Section E.

### D. Feedback

The previous results have shown that ML codebooks are extremely effective for beam training, but beam alignment is not the ultimate goal of wireless networks. Effective beam training naturally ensures that a strong rank-1 wireless link exists, however, the actual MU-MIMO performance is less clear from the beam alignment results. In this subsection, we incorporate type-II feedback and multi-stream data transmission to determine the system-level impact of ML codebooks. We first address the general impact of the CSI-RS parameters for hybrid arrays, then look at a comparison between traditional codebooks and X-BM codebooks. Throughout these results, the most important metric is always the effective sum spectral efficiency, which describes the actual performance after accounting for the overhead.

First, we investigate the feedback parameters using traditional DFT codebooks and hybrid arrays in Fig. 9. These results show an interesting result: using higher rank CSI-RS requires more feedback and typically reduces the performance compared to lower rank CSI-RS. This means there is little benefit to configuring the system with many CSI-RS ports. In fact, MU-MIMO with at least one user receiving $4$ data layers is unlikely to maximize performance so even smaller beam groupings can be used. These results are specific to the simulated scenario, which includes many UEs to multiplex between and encourages scheduling more users with lower rank channels and less overhead. The total overhead depends on the number of CSI-RS resources allocated $N_{\text{CSI}}/B_{\text{g}}$, the beams per group $B_{\text{g}}$, and the feedback and overhead associated with each CSI-RS.

In the next investigation, we compare how the SSB and CSI-RS codebooks impact the overall system ESSE. The results in Figure 10 are compared again using DFT and X-BM codebooks with three feedback settings from the previous results, $(L_{\text{CSI}} = 2, B_{\text{g}} = 2)$, $(L_{\text{CSI}} = 2, B_{\text{g}} = 4)$, and $(L_{\text{CSI}} = 4, B_{\text{g}} = 4)$. It can be seen that the X-BM codebooks result in roughly a $10\%$ effective spectral efficiency improvement over DFT codebooks using the standard feedback formats. Furthermore, the gain from using limited feedback is increased with X-BM codebooks due to the more effective beam alignment.

### E. Model updating

For the final evaluation, we evaluate how the model performs as a result of distribution shift. We consider the case that the model is initially trained in the first environment, but then the scene changes significantly causing the channel distributions to be out-of-distribution [42] or arising from a new context. We evaluate the model in the new environment (environment B) and then again after retraining the model using only the available CSI at the BS. In particular, we retrain on potentially erroneous CSI to characterize how effective live-model updating is using imperfect CSI. Because of the lack of perfect CSI, we employ an unsupervised learning strategy

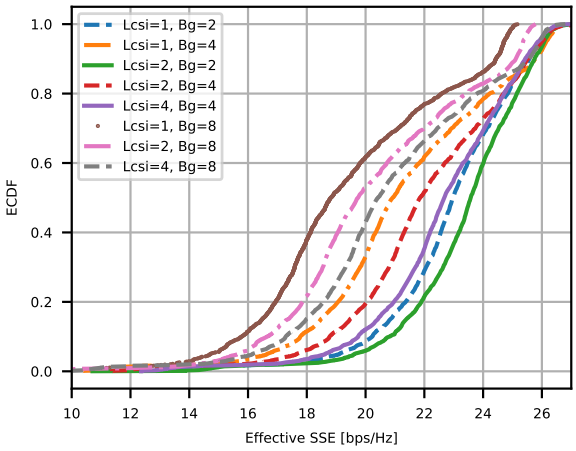



**FIGURE 9. An empirical CDF chart of the ESSE performance with PMI type-II using different ratios of feedback resolution $L_{\text{CSI}}$ and CSI rank $B_{\text{g}}$. Each ratio is plotted with different line formats. It can be seen that there is no gain from using larger beam groupings which require more overhead, more feedback, and reduces the number of UEs that can be multiplexed.**

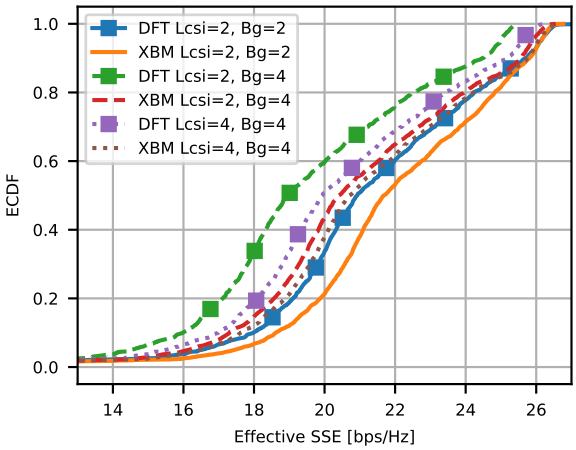



**FIGURE 10. ESSE performance with CSI parameters $(L_{\text{CSI}} = 2, B_{\text{g}} = 2)$, $(L_{\text{CSI}} = 2, B_{\text{g}} = 4)$, and $(L_{\text{CSI}} = 4, B_{\text{g}} = 4)$. The X-BM codebooks provide a noticeable improvement over DFT codebooks, especially for the lowest $20\%$ of users. While the significant gain in beam training does result in improved ESSE, it is not expected to see gains on the order of $4$ times larger ESSE, especially with greedy scheduling.**

to maximize the RSRP (or minimize the negative RSRP)

$$\mathcal{L}(\mathbf{p}) = -\frac{1}{U}\sum_{u} 10\log_{10}(\mathbf{p}_u)^2. \tag{34}$$

This formulation is less stable as the training only seeks to maximize the RSRP without a reference point so a low RSRP is more penalizing, even if that UE cannot achieve a much higher RSRP due to the location/channel. Additionally, the gradients are calculated with respect to the quantized, estimated beamformed CSI which can include estimation and quantization error. At the same time, the training considers a very practical scenario where only the available information obtained from a live network is used for fine-tuning.

The results of the fine-tuning are shown in Figure 11 which characterizes the difference between a site-specific model, the site-agnostic model, and the fine-tuned model performance between the scenarios. It can be seen that there is a significant drop in performance when deploying the initial model (site-agnostic) in the new environment, but that using the imperfect CSI and retraining with 20000 samples for 30s can provide an improvement, especially for the bottom 50% of users. The imperfect CSI and unsupervised fine-tuning strategy, however, is not able to recover the results that could be achieved with perfect CSI shown by the site-specific results. The flexibility of deploying pre-trained models and then quickly fine-tuning is a valuable benefit for network operators, even during cyclic changes in distributions i.e. rush hour or during sporting events.

## VI. Conclusion

In this paper, we proposed a machine learning-assisted beam management strategy (X-BM) for extreme MIMO. The proposed solution integrates machine learning into all stages of beam management while maintaining realistic timing and computation constraints. ML codebooks provide significant gains during beam training, especially for hybrid arrays, with an average improvement of 6dB in received power compared to traditional methods. In fact, the learned initial access codebooks outperform two-step hierarchical search DFT codebooks in all scenarios.

The proposed X-BM algorithm is built upon a beamspace conversion that translates beamformers into a consistent grid representation. We integrate the beamspace conversion, quantization, transmission, and feedback into the model learning to enable an end-to-end learning framework. The proposed algorithm naturally generalizes across array geometries, user distributions, and entirely new environments. Furthermore, X-BM does not require changes to UE operation while increasing the achievable spectral efficiency by 10%. We also show that it is also capable of efficiently updating using only the partial CSI available to the network.

Synthesizing the results of our study, we draw the following conclusions about the evolution of the upper-mid band (so-called FR3) in the next releases of 5G and into 6G. First, we find that the beam training framework can be extended to the upper midband and X-MIMO arrays. Second, DFT beams have worse performance as their narrower beamwidths (compared to sub-6GHz bands) become even more suboptimal in multipath environments. Finally, AI/ML approaches are able to design both SSB and CSI-RS codebooks that can substantially improve beam training and sum spectral efficiency within the network. Future investigations on multi-cell and interference-aware codebook design are necessary for further characterizing the benefits of AI/ML in realistic, network-level beam management. We expect future evaluations using different hardware architectures and hybrid formulations will also be important as 6G refines the physical layer support

## REFERENCES


[1] R. M. Dreifuerst and R. W. Heath, "Massive MIMO in 5G: How beamforming, codebooks, and feedback enable larger arrays," *accepted for publication in IEEE Commun. Mag*, 2023.

[2] H. Holma, H. Viswanathan, and P. Mogensen, "Extreme massive MIMO for macro cell capacity boost in 5G-Advanced and 6G," Nokia Bell Labs, Espoo Finland, Tech. Rep., 2021.

[3] R. W. Heath *et al.*, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[4] Y. Heng *et al.*, "Six key challenges for beam management in 5.5G and 6G systems," *IEEE Commun. Mag.*, vol. 59, no. 7, pp. 74–79, Jul. 2021.

[5] R. M. Dreifuerst and R. W. J. Heath, "CSI type-II codebook of codebooks," in *in Proc. of IEEE 24th Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Sep. 2023, pp. 116–120.

[6] M. Giordani *et al.*, "A tutorial on beam management for 3GPP NR at mmwave frequencies," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 1, pp. 173–196, Jan. 2019.

[7] T. Wang *et al.*, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, 2017.

[8] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[9] X. Lin, "An overview of the 3GPP study on artificial intelligence for 5G New Radio," aug 2023. [Online]. Available: https://arxiv.org/abs/2308.05315v1http://arxiv.org/abs/2308.05315

[10] B. Li *et al.*, "On the efficient beam-forming training for 60GHz wireless personal area networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 504–515, 2013.

[11] V. Va *et al.*, "Inverse multipath fingerprinting for millimeter wave V2I beam alignment," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4042–4058, May 2018.

[12] Y. Wang, N. J. Myers, N. González-Prelcic, and R. W. Heath, "Site-specific online compressive beam codebook learning in mmWave vehicular communication," *IEEE Trans. Wirel. Commun.*, pp. 1–14, 2021.

[13] Q. Xue *et al.*, "Beam management in ultra-dense mmWave network via federated reinforcement learning: An intelligent and secure approach," *IEEE Trans. Cogn. Commun. Netw.*, 2022.

[14] J. Yang, W. Zhu, and M. Tao, "Deep learning for hierarchical beam alignment in mmWave communication systems," in *Proc. of the IEEE Global Communications Conference*, 2022.

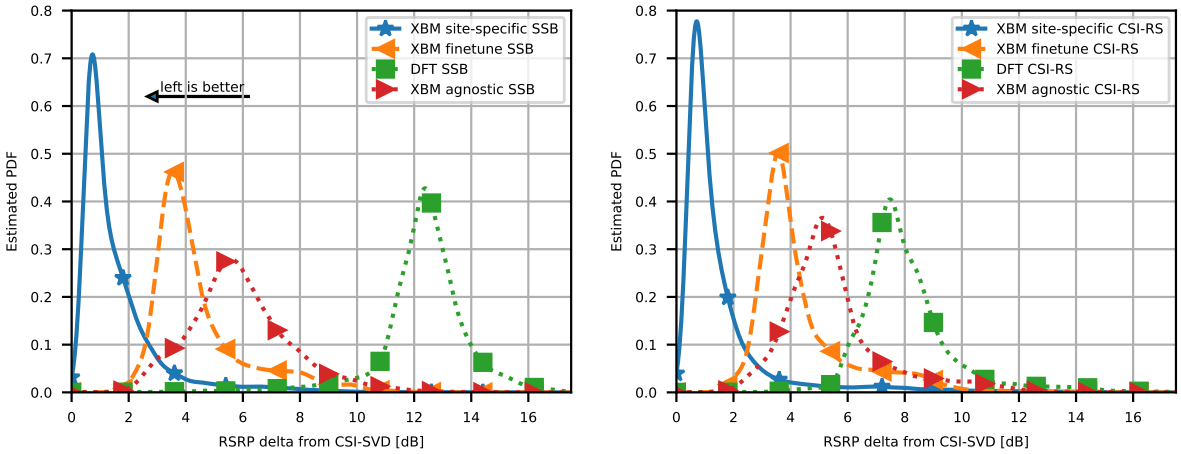
[15] J. Yang, W. Zhu, M. Tao, and S. Sun, "Hierarchical Beam Alignment for Millimeter-Wave Communication Systems: A Deep Learning Approach," *IEEE Trans. Wirel. Commun.*, pp. 1–1, sep 2023.

[16] R. Shafin *et al.*, "Self-tuning sectorization: Deep reinforcement learning meets broadcast beam optimization," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 6, pp. 4038–4053, Jun. 2020.

[17] W. Xia *et al.*, "A deep learning framework for optimization of MISO downlink beamforming," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1866–1880, Mar. 2020.

[18] Y. Heng, J. Mo, and J. G. Andrews, "Learning site-specific probing beams for fast mmWave beam alignment," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 8, pp. 5785–5800, 2022.

[19] R. M. Dreifuerst, R. W. Heath, and A. Yazdan, "Massive MIMO beam management in sub-6 GHz 5G NR," in *Proc. of the IEEE Veh. Technol. Conf.*, 2022, pp. 1–5.




(a) SSB

(b) CSI-RS

**FIGURE 11. Empirical PDF comparison of the SSB RSRP (a) and CSI-RS RSRP (b) for the X-BM model deployed in environment B and training on data from environment A at a different carrier frequency and user distribution. The site-specific model (—), fine-tuned (– –), and site-agnostic model (··) correspond to a model trained directly for environment B, a model trained on environment A and retrained on environment B, and the original model from environment A without retraining. Even trained on a different environment, the proposed X-BM method outperforms DFT codebooks and transfers to new environments efficiently.**


[20] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, 2009.

[21] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[22] 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, 2016.

[23] C. K. Wen, W. T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wirel. Commun. Lett.*, vol. 7, no. 5, pp. 748–751, 2018.

[24] J. Kim, H. Lee, and S.-H. Park, "Learning robust beamforming for MISO downlink systems," *IEEE Communications Letters*, vol. 25, no. 6, pp. 1916–1920, 2021.

[25] M. Chen *et al.*, "Deep learning-based implicit CSI feedback in massive MIMO," *IEEE Trans. Commun.*, vol. 70, pp. 935–950, Dec. 2022.

[26] H. Xiao *et al.*, "Knowledge-driven Meta-learning for CSI Feedback," oct 2023. [Online]. Available: https://arxiv.org/abs/2310.15548v2

[27] G. Morozov, A. Davydov, and V. Sergeev, "Enhanced CSI feedback for FD-MIMO with beamformed CSI-RS in LTE-A Pro systems," in *Proc. of the IEEE Veh. Technol. Conf.*, Jul. 2016, pp. 1–5.

[28] R. M. Dreifuerst and R. W. Heath, "Machine learning codebook design for initial access and CSI type-II feedback in sub-6GHz 5G NR," *Accepted for publication in IEEE Trans. Wirel. Comm.*, 2023. [Online]. Available: https://arxiv.org/abs/2303.02850

[29] M. Alouzi, F. Al-Kamali, C. D'amours, and F. Chan, "Direct conversion of hybrid precoding and combining from full array architecture to subarray architecture for mmWave MIMO systems," *IEEE Access*, vol. 11, pp. 35 457–35 468, 2023.

[30] Q. Xue *et al.*, "A survey of beam management for mmWave and THz communications towards 6G," aug 2023. [Online]. Available: https://arxiv.org/abs/2308.02135v1http://arxiv.org/abs/2308.02135

[31] J. Hoydis *et al.*, "Sionna: An open-source library for next-generation physical layer research," *arXiv preprint*, Mar. 2022.

[32] A. Singh and S. Joshi, "A survey on hybrid beamforming in mmWave massive MIMO system," *J. Sci. Res.*, vol. 65, no. 01, pp. 201–213, 2021.

[33] X. Wu, D. Liu, and F. Yin, "Hybrid Beamforming for Multi-User Massive MIMO Systems," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3879–3891, Sep. 2018.

[34] Y. Ren, S. Yoo, and A. Hoisie, "Performance analysis of deep learning workloads on leading-edge systems," in *2019 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS)*, 2019, pp. 103–113.

[35] V. J. Reddi *et al.*, "MLPerf inference benchmark," ML Commons, 2020, v3.1 updated Aug. 2023.

[36] Z. Qin and H. Yin, "A review of codebooks for CSI feedback in 5G New Radio and beyond," 2023. [Online]. Available: http://arxiv.org/abs/2302.09222

[37] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, 2018.

[38] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.

[39] A. Felix *et al.*, "Ofdm-autoencoder for end-to-end learning of communications systems," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[40] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. of International Conference on Learning Representations*, Dec. 2014, pp. 1–15.

[41] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in *Proc. of NeurIPS Deep Learning and Representation Learning Workshop*, 2015, pp. 1–9.

[42] S. Ben-David *et al.*, "A theory of learning from different domains," *Machine Learning*, vol. 79, pp. 151–175, 2010. [Online]. Available: http://www.springerlink.com/content/q6qk230685577n52/

**Ryan M. Dreifuerst**graduated with a B.S. in Electrical Engineering from Milwaukee School of Engineering and a B.S. in Electrical and Communications Engineering from Technische Hochschule Lübeck. He received a M.S. from the University of Texas at Austin where he researched low-resolution multi-sinusoidal decomposition and estimation methods. Ryan completed his PhD on machine learning for codebook design in beam management at North Carolina State University. Ryan's research interests lie at the intersection of wireless communications, signal processing, and machine learning. Currently, his work is focused on deep learning for massive MIMO systems, including codebook design for beam management and feedback. Ryan has spent the previous summers with Meta, Qualcomm, Samsung Research America, and Nvidia researching novel algorithms for wireless systems. Following his graduate studies, Ryan is working for MITRE on signal processing and advanced wireless techniques. Ryan is an active member and reviewer for IEEE Signal Processing Society, Communications Society, and Vehicular Technology Society.

**Robert W. Heath Jr.**(S'96 - M'01 - SM'06 - F'11) is the Charles Lee Powell Chair in Wireless Communications in the Department of Electrical and Computer Engineering at the University of California, San Diego. He is also President and CEO of MIMO Wireless Inc. From 2002-2020 he was with The University of Texas at Austin, most recently as Cockrell Family Regents Chair in Engineering and Director of UT SAVES. From 2020-2023 he was the Lampe Distinguished Professor at North Carolina State University and co-founder of 6GNC. He authored "Introduction to Wireless Digital Communication" (Prentice Hall, 2017) and "Digital Wireless Communication: Physical Layer Exploration Lab Using the NI USRP" (National Technology and Science Press, 2012), and co-authored "Millimeter Wave Wireless Communications" (Prentice Hall, 2014) and "Foundations of MIMO Communication" (Cambridge University Press, 2018). Dr. Heath has been a co-author of a number award winning conference and journal papers including recently the 2017 Marconi Prize Paper Award, the 2019 IEEE Communications Society Stephen O. Rice Prize, the 2020 IEEE Signal Processing Society Donald G. Fink Overview Paper Award, the 2021 IEEE Vehicular Technology Society Neal Shepherd Memorial Best Propagation Paper Award, and the 2022 IEEE Vehicular Technology Society Best Vehicular Electronics Paper Award. Other notable awards include the 2017 EURASIP Technical Achievement award, the 2019 IEEE Kiyo Tomiyasu Award and the 2021 IEEE Vehicular Technology Society James Evans Avant Garde Award. In 2017, he was selected as a Fellow of the National Academy of Inventors. He was a member-at-large on the IEEE Communications Society Board-of-Governors (2020-2022) and the IEEE Signal Processing Society Board-of-Governors (2016-2018). He was Editor-in-Chief of IEEE Signal Processing Magazine from 2018-2020. He is also a licensed Amateur Radio Operator, a Private Pilot, a registered Professional Engineer in Texas.