Received 20 March 2025; accepted 7 April 2025. Date of publication 10 April 2025; date of current version 29 April 2025.

*Digital Object Identifier 10.1109/OJCOMS.2025.3559737*

# Unsupervised Learning-Based Low-Complexity Integrated Sensing and Communication Precoder Design

**MURAT TEMIZ (Member, IEEE), AND CHRISTOS MASOUROS (Fellow, IEEE)**

Department of Electronic and Electrical Engineering, University College London, WC1E 7JE London, U.K.

CORRESPONDING AUTHOR: M. TEMIZ (e-mail: m.temiz@ucl.ac.uk)

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the Deep Learning-Based Solutions for the Physical Layer of Machine Type Communications Project under Grant EP/S028455/1.

**ABSTRACT** This study proposes an unsupervised deep learning-based (DL-based) approach to precoding design for integrated sensing and communication (ISAC) systems. Designing a dynamic precoder that can adjust the trade-off between the sensing performance and communication capacity for ISAC systems is typically highly compute-intensive owing to requiring solving non-convex problems. Such complex precoders cannot be efficiently implemented on hardware to operate in highly dynamic wireless environments where channel conditions rapidly vary. Accordingly, we propose an unsupervised DL-based precoder design strategy that does not require a data set of the optimum precoders for training. The proposed DL-based precoder can also adapt the trade-off between the communication sum rate and sensing accuracy depending on the required communication and/or sensing performance. It offers a low-complexity precoder design compared to conventional precoder design approaches that require iterative algorithms and computationally intensive matrix operations. To further reduce the memory usage and computational complexity of the proposed precoding solution, we have also explored weight quantization and pruning techniques. The results have shown that a quantized and pruned deep neural network (DNN) can achieve 96% of the sum rate achieved by the full DNN while its memory and computational requirements are less than 17% of the full DNN.

**INDEX TERMS** 6G wireless networks, beamforming design, integrated sensing and communication, unsupervised deep learning.

## I. INTRODUCTION

INTEGRATED sensing and communication (ISAC) is seen as one of the game-changer technologies for the sixth generation (6G) and beyond communication networks [1], [2], [3]. Future communication networks are expected to support a wide range of applications, such as smart vehicles or robotics that, in addition to reliable communication links, also need accurate sensing information of the environment to seamlessly perform their functions [4]. Moreover, sensing information can also assist the communication systems in enabling highly accurate channel estimation, rapid initial connectivity, fast beamforming, and more secure links in communication networks [5]. Accordingly, ISAC is expected to be an essential part of future wireless networks.

One of the major challenges in ISAC systems is to generate the optimum dual-function waveform that provides the desired trade-off between communication capacity and sensing accuracy. This is because the computation of the optimum waveforms typically requires solving non-convex and NP-hard optimization problems, which may not be feasible within the limited computational, memory, and time resources of the base station (BS) and user devices (UEs). Therefore, some algorithms have been recently proposed for designing sub-optimal ISAC waveforms [6], [7], [8], [9], [10], [11], [12], [13], [14]. For instance, [6] proposes

low-complexity algorithms to generate waveforms for maximizing the sum rate and energy efficiency by utilizing the interference in massive multiple-input and multiple-output (MIMO) ISAC systems. The semidefinite relaxation (SDR) technique is widely utilized to solve non-convex optimization problems at the expense of sub-optimal transmit beamformers [7], [8]. Moreover, various algorithms for joint transmit and receive beamformer design under self-interference were proposed in [9], [10]. Another study [12] proposes a hybrid beamforming design based on iteratively optimizing the transmit beam and the phase vector for millimeter-wave MIMO ISAC. Joint communication and sensing with full-duplex operation are also considered in [15], [16], and joint beamforming and power optimization for full-duplex are also studied in [16]. Liu et al. proposed an approach to design ISAC waveforms by optimizing the CRLB of sensing under certain communication sum rate constraints [14]. The studies mentioned above primarily propose iterative algorithms to design sub-optimal ISAC waveforms to avoid solving non-convex optimization problems.

However, even sub-optimum solutions of ISAC waveform design require high computational complexity and memory usage due to employing iterative methods and involving multiple compute-intensive matrix operations such as matrix inversion and multiplication [17]. These operations may need to be concurrently executed in orthogonal frequency division multiplexing (OFDM) systems that employ distinct sub-band precoders in a single communication-sensing waveform frame. Moreover, practical ISAC systems will need to consider the trade-off between communication and sensing since sensing accuracy and communication data rate performance requirements can vary depending on the scenario and channel conditions. Introducing a trade-off parameter between the sensing and communication performances will increase computational complexity such that the optimum waveform needs to be computed dynamically as the trade-off parameter is changed.

Deep neural networks (DNNs) have been shown to satisfactorily approximate linear or nonlinear continuous functions [18], [19]. Thus, DNNs have been recently considered to solve computationally expensive optimization problems encountered in many fields, including wireless communication systems [17], [20], [21], [22]. Deep learning solutions are especially appealing since efficient neural network accelerators are already included in recent central processing units (CPUs), even in the CPUs of resource-constrained edge devices and mobile phones [23]. Deep learning techniques are generally organized under two main categories: Supervised and unsupervised learning techniques. While both can be used for solving physical-layer problems in wireless systems, unsupervised deep learning is desirable for waveform and precoder design problems as it does not require the optimum solution data set (labeled data) for training [24], [25]. On the other hand, supervised deep learning methods require a large training data set containing the optimum solutions. Acquiring such a large data set containing optimum ISAC precoders may not always be possible due to the need to collect and classify measurement or simulation data in a wide variety of ISAC setups.

Unsupervised learning has the potential to achieve higher sum rate performance compared to supervised learning for beamformer design tasks. This has been shown in [26], where a deep neural network (DNN) was first trained via supervised learning, and then the same DNN was retrained via unsupervised learning. Another study in [27] proposed an unsupervised DL-based beamformer design for an analog-digital hybrid massive MIMO transmitter architecture. Their proposed beamformers achieved a near-optimum performance with low complexity without having access to full CSI. Unsupervised deep learning has also been utilized to design decentralized beamformers for cell-free massive MIMO networks [28], delivering a near-optimum capacity (i.e., slightly lower than that of centralized ZF). Unsupervised machine learning has also employed for CSI feedback in massive MIMO systems [29], for joint antenna selection and beamforming [30], for power control in cell-free massive MIMO systems [31], or for efficient and low-complexity symbol level precoding in MIMO systems [32], among others.

DL-based techniques have also been considered for integrated sensing and communications. A recent study proposed a distributed unsupervised learning method for power allocation and interference management in integrated sensing and communication networks [33]. Another study utilized unsupervised deep learning and long short-term memory (LSTM) networks for predictive beamforming for vehicular integrated sensing and communication networks [34]. Researchers in [35] propose an end-to-end DL-based design that jointly optimizes the transmitter and receiver with semi-supervised learning. Moreover, Sankar et al. propose an unsupervised learning-based precoder design that aims to maximize the worst-case target illumination while guaranteeing a desired SNR for the users [36].

Unlike the previous ISAC precoder design studies, where non-convex optimization libraries or iterative numerical methods were used, we propose a lightweight ISAC waveform design method based on unsupervised deep learning in this study. Moreover, depending on the system requirements, the proposed solution enables a swift adjustment of the trade-off between sensing accuracy and communication capacity. The proposed method can be efficiently implemented on a neural network hardware accelerator. Hardware implementation of ISAC techniques and performing experimental measurements are essential to demonstrate the real-time performance of ISAC techniques and trade-offs between communication sum rate and sensing accuracy [37], [38], [39], [40].

Against the above background, the contributions of this paper are:

- We propose an unsupervised DL-based ISAC precoder design for multi-user, multi-target ISAC networks where

the BS communicates with multiple UEs while performing sensing of targets with multiple beams. The proposed unsupervised DL-based model can swiftly adjust the ISAC precoders depending on the desired sensing and communication requirements.

- We introduce a tunable trade-off parameter to adjust the trade-off between the communication sum rate and sensing accuracy. A set of DNNs, each of them corresponding to a specific trade-off weight, are trained to achieve the near optimum precoders for different scenarios. Our results show that the proposed precoder can outperform the weighted minimum mean square error (WMMSE) precoder in terms of communication sum rate when the communication sum rate is aimed to be maximized.
- We investigate the trade-off between the complexity of the DNNs and the performance of communication and sensing. By utilizing weight quantization and pruning, the DNN size is substantially reduced while still providing the desired trade-off between the communication and sensing performances.

*Notation:* Throughout the manuscript, the following mathematical notations are used. Bold uppercase letters (e.g., $\mathbf{H}$) indicate matrices, while bold lowercase letters (e.g., $\mathbf{h}$) indicate vectors. Superscripts $^*$ and $^H$ indicate the conjugate and Hermitian transpose, respectively. The expectation, absolute value, and Euclidean norm operators are denoted by $\mathbb{E}[\cdot]$, $|\cdot|$, and $\|\cdot\|$, respectively.

## II. SYSTEM MODEL

A single sub-6 GHz time division duplex (TDD) cell is considered, where the ISAC BS is equipped with a uniform linear array (ULA) consisting of $M$ antennas to transmit ISAC signals and receive communication signals in addition to $R$ radar receive antennas. ULA antennas are placed with $\lambda/2$ separation, with $\lambda$ being the wavelength of the carrier frequency. $M$ transmit antennas alternately transmit the ISAC signal and receive communication data in TDD mode while $R$ receive antennas are always in the receiving mode to receive the target echo signals. As shown in Fig. 1, the BS communicates with $K$ downlink user equipment (UEs) while performing beamforming on $T$ possible targets in the area of interest.

OFDM is considered in this study such that the precoder matrix is constructed for each subcarrier or a block of subcarriers [41], [42]. The proposed DL-based approach significantly reduces the computational complexity of the precoder design, hence, a large number of beamformer matrices, each of them corresponding to a subcarrier or a block of subcarriers, can be swiftly constructed by the trained DNNs. Accordingly, in the rest of this article, we consider a single subcarrier and explain the system model and the proposed method accordingly, without the loss of generality.

Since TDD is considered, the $M$ antennas transmit ISAC waveform and receive communication data in turns within the coherence time, utilizing the estimated channel for downlink and uplink. However, the signals reflected from the targets or environment must be received and processed continuously for sensing. Accordingly, the BS has $R$ separate receive antennas that are used to receive echoes for sensing. The BS simultaneously intends to transmit the multi-user MIMO communication data $\mathbf{Y} \in \mathbb{C}^{K\times L}$ consisting of $L$ symbols for $K$ UEs. Each symbol is modulated via quadrature amplitude modulation (QAM), i.e., 256-QAM.[1] The same transmitted signals are used to form beams on the targets.

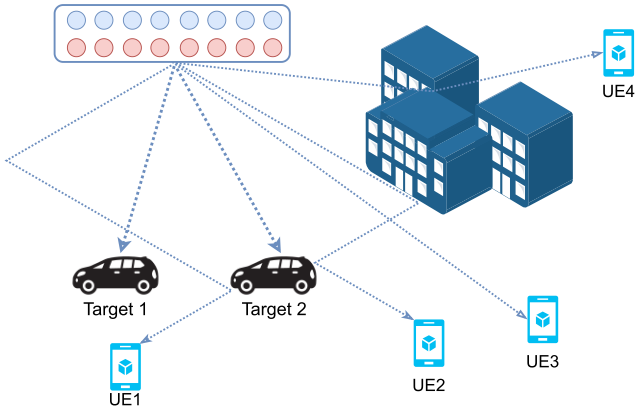


**FIGURE 1.** A typical ISAC cell where the BS performs beamforming on several LOS targets while communicating with multiple downlink users (LOS or NLOS). (Blue circles: ISAC transmit antennas, red circles: radar receive antennas).

### A. COMMUNICATION CHANNEL MODEL

The channel matrix between the BS and UEs, $\mathbf{H} \in \mathbb{C}^{K\times M}$, is modeled as a Rician fading channel model, while radar channels are considered to be line-of-sight (LOS) two-way channels. The Rician fading channel model considered for communication channels is given by [43],

$$\mathbf{H} = \sqrt{\frac{\mathcal{K}}{1+\mathcal{K}}}\mathbf{H}_L + \sqrt{\frac{1}{1+\mathcal{K}}}\mathbf{H}_F, \tag{1}$$

where $\mathbf{H}_L$ denotes the segment of the channel consisting of LOS or strong reflection links while $\mathbf{H}_F$ denotes the non-line-of-sight (NLOS) component of the channel, whose entries follow a complex-valued Gaussian distribution with zero means and $\sigma_h^2$ variance, i.e., $h_F \sim \mathcal{CN}(0, \sigma_h^2)$. In the above, $\mathcal{K}$ denotes the Rician K factor. We further assume that the channels of subsequent time instances are temporally correlated since a small number of the scatters move in most environments [44]. Hence, the channel matrix, given by (1), slowly changes over time, leading to a high correlation between sequential channel instances, i.e., $\mathbf{H}(t)$ and $\mathbf{H}(t+1)$.

The complex-valued entries of the LOS channel segment, $\mathbf{H}_L$, is given by

$$h_{L(k,m)} = \beta_k \exp\left(j2\frac{\pi}{\lambda}d_{k,m}\right), \tag{2}$$

[1]Note that phase-shift keying (PSK) or QAM can be used to modulate each symbol since the proposed ISAC precoder is independent of the modulation scheme.

where $\lambda$ and $d_{k,m}$ denote the wavelength of the carrier and the distance between the $k$th UE and $m$th antenna. Moreover, the linear large-scale fading of the $k$th UE is a function of the path loss given by

$$\beta_k = 10^{-PL_k/10}. \tag{3}$$

The path-loss of the $k$th UE in dB is modeled as [6],

$$PL_k = 10\log_{10}\left(\frac{4\pi f_c d_0}{c_0}\right)^2 + 10\log_{10}\left(\frac{d_k}{d_0}\right)^{\varphi} + \zeta_{sh}, \tag{4}$$

where $f_c$, $d_0$, $c_0$ and $\varphi$ denote the frequency of the carrier signal, the reference distance, the speed of light, and the path-loss exponent, respectively. Moreover, $d_k$ denotes the distance of the $k$th UE to the BS, (i.e., $30 \leq d_k \leq 400$ m), and $\zeta_{sh}$ denotes the log-normal shadow fading, which is a zero-mean Gaussian random variable with standard deviation $\sigma_{sh}$.

Based on the non-line-of-sight (NLOS) urban macrocell measurements in [45], we use $f_c = 5$ GHz, $d_0 = 1$ m, $\varphi = 2.9$, and $\sigma_{sh} = 5.7$ dB. Each UE is randomly located in the cell and may experience significantly different channel gains in relation to the path loss [46]. Moreover, the complex-valued entries of the NLOS channel, $\mathbf{H}_F$, between the $k$th UE and the $m$th BS antenna is modeled as $h_{F,k,m} \sim \mathcal{CN}(0,1)$ which pertains to a Rayleigh channel model. The UEs are randomly located in the x- and y-coordinates in the cell, and their path loss and channel matrix are calculated based on their locations and using the channel model explained above in each network instance.

### B. RADAR CHANNEL MODEL

For radar sensing, as typically considered in most radar scenarios, the targets are assumed to be in line-of-sight (LOS) with the BS. The BS can perform beam-scanning to search the possible targets in the area of interest by scanning various beamforming angles or can focus on specific angles for beamforming. In this study, we have considered that the BS can form two beams for two independent target angles at the same time while communicating with UEs.

The power of the received echo signal from a single target, without considering the transmit and receive beamforming gains, is given by [47],

$$P_r = \frac{P_t \lambda^2 G_A^2 \sigma_T}{(4\pi)^3 R_T^4}, \tag{6}$$

where $\lambda$, $G_A$, $\sigma_T$, $R_T$ denote the wavelength of the signal, antenna gain, radar-cross section (RCS) of the target, and distance of the target from the radar, respectively.

The Cramér-Rao lower bound (CRLB) $\psi(\theta_t)$ for the estimation of target angle $\theta_t$ for the $t$th target is given by (5), shown at bottom of the page [14], where $\eta_t = P_r/P_n$ with $P_n$ being the noise power denotes the signal-to-noise ratio (SNR) of the signals reflected from the $t$th target and received at the radar receive antennas. Moreover, $\mathbf{R}_X$ denotes the covariance matrix of the transmitted waveform $\mathbf{X}$, i.e.,

$$\mathbf{R}_X = \frac{1}{L}\mathbf{X}\mathbf{X}^H, \tag{7}$$

Moreover, $\mathbf{A}(\theta_t)$ and $\dot{\mathbf{A}}(\theta_t)$ in (5) are respectively defined as

$$\mathbf{A}(\theta_t) = \mathbf{b}(\theta_t)\mathbf{a}^H(\theta_t), \tag{8}$$

$$\dot{\mathbf{A}}(\theta_t) = \frac{\partial \mathbf{A}(\theta_t)}{\partial \theta_t}, \tag{9}$$

where $\mathbf{a}(\theta_t) \in \mathbb{C}^{M\times 1}$ and $\mathbf{b}(\theta_t) \in \mathbb{C}^{R\times 1}$ denote the steering vectors for the transmit and receive antenna arrays, respectively. By choosing the center of the ULA array as the reference, the steering vector for the transmit array $\mathbf{a}(\theta_t)$ is given by [14]

$$\mathbf{a}(\theta_t) = \left[e^{-j\frac{M-1}{2}\pi \sin\theta_t}, e^{-j\frac{M-3}{2}\pi \sin\theta_t}, \ldots, e^{j\frac{M-1}{2}\pi \sin\theta_t}\right]^T, \tag{10}$$

and the steering vector $\mathbf{b}(\theta_t)$ for the receive array consisting of $R$ antennas can be written in the same way.

### C. CHANNEL ESTIMATION AND TRANSMISSION

During channel estimation, the UEs transmit pilot symbols, and the ISAC BS estimates the channel state information (CSI) by processing these pilot symbols. CSI estimation can be performed by linear methods such as the least-squares (LS) estimator, linear minimum mean-square error (LMMSE) estimator, or iterative methods [48]. The CSI estimated by the ISAC BS may contain estimation errors due to noise, hardware impairments, frequency, or phase offsets. Since the focus of this study is precoder design rather than channel estimation, an imperfect CSI estimation model is utilized [49], [50]. Accordingly, the estimated CSI matrix consists of estimation errors; $\hat{\mathbf{H}}$ is modeled as [51],

$$\hat{\mathbf{H}} = \mu\mathbf{E} + (1-\mu)\mathbf{H}, \tag{11}$$

where $\mathbf{E}$ denotes the complex-valued CSI error matrix, whose entries follow a complex-valued Gaussian distribution with zero mean and variance $\sigma_h^2$ as $e_k \sim \mathcal{CN}(0, \sigma_h^2)$, where $\sigma_h^2 = 1$ assuming that the channel matrix is normalized. Moreover, $\mu$ denotes the ratio of the mean amplitude of CSI estimation errors to the mean amplitude of the ground-truth channels. For instance, $\mu = 0$ represents perfect CSI estimation, and $\mu = 0.1$ represents the case with 10% average channel estimation amplitude errors.

The transmitted signal matrix $\mathbf{X} \in \mathbb{C}^{M\times L}$ used for both sensing and communications is given by

$$\mathbf{X} = \mathbf{W}\mathbf{Y}, \tag{12}$$

$$\psi(\theta_t) = \frac{\mathrm{tr}\left(\mathbf{A}^H(\theta_t)\mathbf{A}(\theta_t)\mathbf{R}_X\right)}{2\eta_t L\left(\mathrm{tr}\left(\dot{\mathbf{A}}^H(\theta_t)\dot{\mathbf{A}}(\theta_t)\mathbf{R}_X\right)\mathrm{tr}\left(\mathbf{A}^H(\theta_t)\mathbf{A}(\theta_t)\mathbf{R}_X\right) - \left|\mathrm{tr}\left(\dot{\mathbf{A}}^H(\theta_t)\mathbf{A}(\theta_t)\mathbf{R}_X\right)\right|^2\right)} \tag{5}$$

where $\mathbf{Y} \in \mathbb{C}^{K\times L}$ denotes the symbols transmitted to $K$ UEs in each transmission frame. Moreover, $\mathbf{W} \in \mathbb{C}^{M\times K}$ denotes the precoder, which is a function of the estimated channel matrix $\hat{\mathbf{H}}$, the trade-off parameter between the communication and sensing $\rho$, and corresponding angles of the $T$ targets and the BS, $\Theta = [\theta_1, \theta_2, \ldots, \theta_T]$, i.e.,

$$\mathbf{W} = f\left(\hat{\mathbf{H}}, \rho, \Theta\right). \tag{13}$$

In the following, we illustrate how the optimum design of $\mathbf{W}$ requires solving a non-convex optimization problem.

### D. COMMUNICATION SUM RATE

The signal received by the $k$th UE for the $l$th transmitted symbol is given by

$$\hat{y}_{k,l} = \mathbf{h}_k \mathbf{w}_k y_{k,l} + \sum_{i=1, i\neq k}^{K} \mathbf{h}_k \mathbf{w}_i y_{i,l} + \sigma_k^2, \tag{14}$$

where $\mathbf{h}_k$ denotes the $k$th row vector of $\mathbf{H}$, $\mathbf{w}_k$ denotes the $k$th column vector of $\mathbf{W}$, and $y_{k,l}$ denotes intended data of the $k$th UE, and $\sigma_k^2$ denotes the average power (variance) of the additive white Gaussian noise (AWGN) at the $k$th UE. Considering that the communication data symbols of UEs are uncorrelated and have a mean unit power, the capacity of the $k$th UE is given by

$$C_k = \log_2\left(1 + \frac{\mathbb{E}\left[|\mathbf{h}_k \mathbf{w}_k|^2\right]}{\mathbb{E}\left[\sum_{i=1, i\neq k}^{K} |\mathbf{h}_k \mathbf{w}_i|^2\right] + \sigma_k^2}\right), \tag{15}$$

where $\mathbb{E}[.]$ denotes the expectation operator. The communication sum rate of the network is then calculated by

$$C_{sum} = \sum_{k=1}^{K} C_k. \tag{16}$$

Because BS is beamforming on the targets for sensing, there will be some degree of radar interference on the UEs when $\rho < 1$.

### E. ISAC WAVEFORM DESIGN

The trade-off between the radar sensing and communication performance is considered to be determined by a weighting parameter, $0 \leq \rho \leq 1$. Concretely, $\rho = 0$ indicates that the ISAC system aims to deliver the maximum radar sensing performance, while $\rho = 1$ targets a maximization of the communication sum rate. Other values of $\rho$ correspond to a proportional trade-off between communication and sensing.

The BS may derive the ISAC precoder based on the communication channel, target locations, and trade-off parameter $\rho$ to maximize both communication sum capacity and radar sensing performance by minimizing the sum CRLB, i.e.,

$$\begin{aligned} \max_{\mathbf{W}} \quad & \rho\beta \sum_{k=1}^{K} C_k + \frac{(1-\rho)\gamma}{\sum_{t=1}^{T} \psi(\theta_t)} \\ \text{s.t.} \quad & ||\mathbf{W}||_F^2 \leq P_t, \end{aligned} \tag{17}$$

where $P_t$ is the maximum BS transmit power, and the parameters $\beta$ and $\gamma$ normalize the sum rate and sum CRLBs, hence balancing the weights of the sum rate and sum CRLB in the objective function (17). Accordingly, $\beta$ and $\gamma$ are given by

$$\beta = \frac{1}{\left|\sum_{k=1}^{K} C_k\right|_{\rho=1}}, \tag{18}$$

and $\gamma$ is the minimum sum CRLB for the network achieved when $\rho = 0$, i.e.,

$$\gamma = \left|\sum_{t=1}^{T} \psi(\theta_t)\right|_{\rho=0}. \tag{19}$$

The optimization problem given by (17) is non-convex and NP-hard due to the fractional structure of the sum rate and CRLB equations. Indeed, computing the precoder that only maximizes the sum capacity for MIMO systems is also a non-convex and NP-hard problem for which iterative and DL-based solutions have been recently proposed [52], [53]. Thus, in the following section, we explore an approach based on unsupervised deep learning to approximate the optimum ISAC precoders.

## III. DEEP LEARNING MODEL

In the case of supervised learning, one needs to solve (17) in advance many times to create a data set for the input and output of the DNN. Then, the DNN can be trained through this dataset, and the aim is to minimize the loss between the labeled data and the output of the learning. Unsupervised learning eliminates the need to solve (17) since it does not require a labeled dataset. Accordingly, we have employed an unsupervised learning approach to compute the ISAC precoders for a desired trade-off. Moreover, unsupervised learning also facilitates quick adaptation, fine-tuning, online learning, and developing site-specific models.

### A. UNSUPERVISED LEARNING LOSS FUNCTION

The optimization problem (17) aims to maximize the sum capacity while minimizing the sum CRLB for the desired trade-off $\rho$. This optimization problem can be written as a loss function of unsupervised learning, where constraints are enforced by a penalty term. Accordingly, the loss function of the unsupervised learning for the optimization problem (17) to design a precoder $\widehat{\mathbf{W}}$ can be written as,

$$\mathcal{L}_1 = -\rho\hat{\beta} \sum_{k=1}^{K} C_k - \frac{(1-\rho)\hat{\gamma}}{\sum_{t=1}^{T} \widetilde{\psi}(\theta_t)} + \left|\left(P_t - ||\widehat{\mathbf{W}}||_F^2\right)\right|, \tag{20}$$

where the penalty term $|(P_t - ||\widehat{\mathbf{W}}||_F^2)|$ is introduced so that not satisfying the power constraint $||\widehat{\mathbf{W}}||_F^2 \leq P_t$ is penalized. Moreover, the parameters $\hat{\beta} = \beta/\alpha$, $\hat{\gamma} = \gamma/\alpha$, with $\alpha$ being the scaling factor with regard to the power constraint, weigh the objectives and constraints to achieve a suitable balance between them as described in Section V. Note that the value of $\alpha$ can be loosely chosen within a reasonable

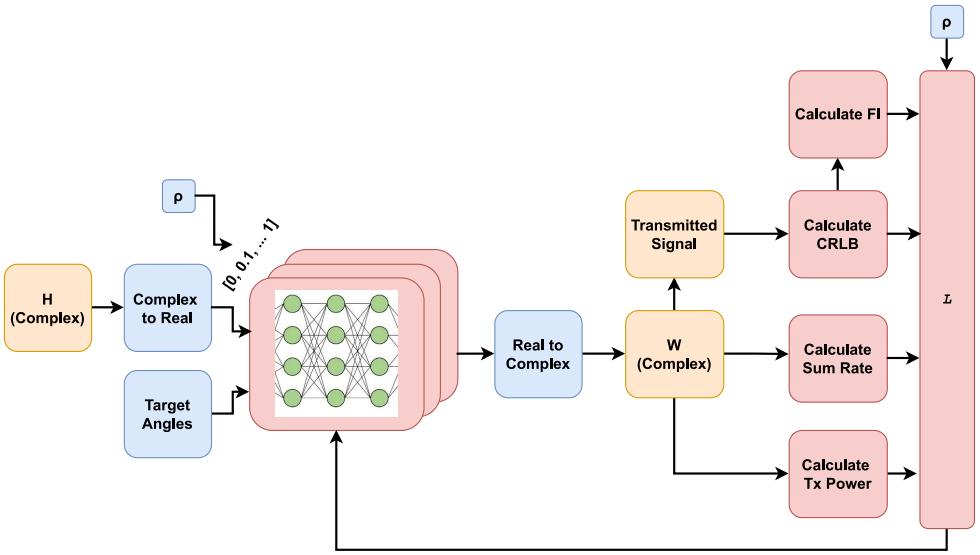


**FIGURE 2.** The unsupervised deep learning model employed during training, where ground truth channels are used.

range, e.g., $5 < \alpha < 10$, since the final power output will be adjusted such that the power constraint $||\widehat{\mathbf{W}}||_F^2 \leq P_t$ is strictly satisfied, as given by

$$\mathbf{W} = \sqrt{P_t}\frac{\widehat{\mathbf{W}}}{\sqrt{||\widehat{\mathbf{W}}||_F^2}}, \tag{21}$$

hence, the final precoder $\mathbf{W}$ is obtained.

The CRLB defined by (5) is not a continuous function since the denominator includes a subtraction, which can also cause the function to be undefined when the subtraction result is zero. Thus, (20) uses an approximation of the CRLB, $\widetilde{\psi}(\theta_t)$, given by

$$\widetilde{\psi}(\theta_t) = \frac{1}{2\eta_t L\left|\mathbf{a}^H(\theta_t)\mathbf{R}_X\mathbf{a}(\theta_t)\right|} \tag{22}$$

$$= \frac{1}{2\eta_t L\left|\mathbf{a}^H(\theta_t)\widehat{\mathbf{W}}\widehat{\mathbf{W}}^H\mathbf{a}(\theta_t)\right|}. \tag{23}$$

This approximation primarily aims to design the covariance matrix of the transmit signals to maximize the signal power on the target, hence does not consider receive beamforming and receiver array steering vector $\mathbf{b}(\theta_t)$. Because the radar receiver beamforming can be formed independently for sensing and does not need to take communication signaling into account.

Notably, the loss function in (20) does not consider the CRLBs of individual targets, which might cause unbalanced beamforming on the targets. Hence, this requires the introduction of a penalty in the loss function to enforce fairness between the targets' CRLBs. There exist various candidate fairness index (FI) metrics available for wireless communication and sensing systems [54]. For this problem, we utilized the Jain's FI metric given by [55],[2]

[2]Note that it is also possible to use any other continuous fairness metric instead of the Jain's FI metric.

$$F_T = \frac{\left(\sum_{t=1}^{T}\widetilde{\psi}(\theta_t)\right)^2}{T\sum_{t=1}^{T}\widetilde{\psi}(\theta_t)^2}, \tag{24}$$

where $0 < F_T \leq 1$. The FI index is equal to 1 when all targets have the same CRLBs whereas it approaches 0 when there are significant differences between the CRLBs of the targets. The loss function accounting for the introduced fairness-related penalty function, $(1 - F_T)$, can be therefore expressed as

$$\mathcal{L}_2 = -\rho\hat{\beta}\sum_{k=1}^{K}C_k - \frac{(1-\rho)\hat{\gamma}}{\sum_{t=1}^{T}\widetilde{\psi}(\theta_t)} + |\left(P_t - ||\widehat{\mathbf{W}}||_F^2\right)| + \delta(1-\rho)(1-F_T), \tag{25}$$

where $\delta$ adjusts the weight of the FI in the loss function. Not that the fairness-related loss function is also adjusted by the trade-off factor, as $(1 - \rho)$, since it needs to only enforce fairness between the targets' CRLBs.

### B. DEEP NEURAL NETWORK ARCHITECTURE

The employed deep learning architecture is given in Fig. 2, where the input and output of the DNN are real-valued vectors. The loss function of the DNN is (25), hence, its training aims to minimize the loss function, resulting in the maximization of the sum rate and minimization of the CRLB for a given trade-off $\rho$ while enforcing the constraints as penalty terms. The real-valued input and output vectors of the DNN are obtained as follows:

- The input vector of the DNN $\mathbf{u} \in \mathbb{R}^{(2MK+T)\times 1}$ is obtained by concatenating i) the real and imaginary parts of the flattened channel matrix $\mathbf{H}$, and ii) the vector of target angles.
- The output vector $\mathbf{z} \in \mathbb{R}^{2MK\times 1}$ consists of the flattened real-valued precoder vector, which is then converted into the complex-valued precoder matrix, $\mathbf{W} \in \mathbb{C}^{M\times K}$ as shown in Fig. 2.

Instead of training the same DNN for all $\rho$ values, distinct DNNs are separately trained for each step of $\rho \in [0, 0.1, \ldots, 1]$. This simplifies the training since each training can be performed for a specific $\rho$ and maximizes the accuracy of the precoding for a given $\rho$.[3] Accordingly, 11 different DNN weight sets are obtained for the same DNN architecture, and the trade-off parameter $\rho$ will determine the specific DNN weight set that will be loaded into the inference hardware. The training of the proposed DL-based precoder design approach is summarized in Fig. 2. After training is completed, the trained DNN can generate the near-optimum precoder matrix for a given input channel matrix and target angles.

## IV. LIGHTWEIGHT DEEP LEARNING ARCHITECTURE

The practical implementation of DNNs will be required to operate with the constrained memory and computational resources available on the BS hardware. Moreover, lightweight architectures also reduce the energy consumption of computational hardware. In this section, we systematically investigate the impact of weight quantization and pruning to reduce the complexity, power consumption, and memory usage of the considered deep learning model [56].

### A. WEIGHT QUANTIZATION

Widely used deep learning libraries such as Pytorch and TensorFlow support 32-bit floating point precision during training, and they also support 32-bit, 16-bit, or 8-bit quantized weights for inference [57], [58]. However, these libraries do not support a lower number of bits quantization, i.e., 1-bit, 2-bit, hence we have developed a dynamic quantization function explained below. Depending on the application, deep learning models can be run with lower precision while providing a similar performance. Thus, we have investigated training the model at full precision (32-bit) and then quantizing it via the dynamic quantization function. Let us denote the vector consisting of all weights in the $i$th layer as $\mathbf{s}_i = \{s_{i,1}, s_{i,2}, \ldots, s_{i,U}\}$, where $U$ denotes the number weights in the $i$th layer. The $Q$-bit quantized weights vector of the $i$th layer are then given by

$$\mathbf{s}_i^Q = \left\lfloor \frac{\mathbf{s}_i}{f(Q, \mathbf{s}_i)} \right\rceil f(Q, \mathbf{s}_i), \tag{26}$$

where

$$f(Q, \mathbf{s}_i) = \frac{\max(\mathbf{s}_i) - \min(\mathbf{s}_i)}{2^Q}. \tag{27}$$

Here, $\lfloor . \rceil$, max(.), min(.) express the element-wise rounding function to the nearest integer, the maximum value, and the minimum value in the vector, respectively. This post-training quantization function dynamically adjusts the range of the quantization to the range of the weights, therefore striving to maximize the utilization of available quantization bits. Furthermore, values that are close to 0 will be automatically pruned after quantization since they will be set to 0 while rounding.

[3]This choice does not introduce additional inference complexity or inference memory requirements since pre-trained DNNs can be loaded for inference for a selected $\rho$. The training duration is also reduced as several DNNs can be concurrently and independently trained. However, this may increase storage requirements since a larger number of neural network weights need to be stored.

### B. PRUNING

Pruning is performed by removing the weights or nodes expected to have the smallest impact on the DNN outcome. Thus, we perform a global pruning by removing the weakest weights (i.e., the weights are closest to 0) after quantization [59]. Additionally, neurons with all input or output weights pruned are also removed. The pruning ratio is denoted by $J$, e.g., $J = 0.1$ indicates that the weakest 10% of all weights across the entire DNN model are pruned.

## V. NUMERICAL RESULTS

In this section, we have examined the proposed ISAC precoder design method via simulations. The trade-off between the communication sum rate, sensing accuracy, and computational complexity is demonstrated. Moreover, the trade-offs between the DNN complexity and communication and sensing performances are also investigated.

### A. DATA GENERATION AND TRAINING

The ISAC system described in Section II is modeled in Python, and PyTorch is used as the deep learning library [57]. The channel dataset is generated for $10^6$ network instances, where each network consists of randomly located $K = 10$ UEs. The channel of each network is generated according to (1), where the path-loss of each UE in each instance is calculated by (4). Two different channel datasets are generated for $M = 10$ and $M = 20$ antennas. Moreover, $10^6$ random target angles between $-90$ and 90 degrees with $1^\circ$ resolution are generated and added to the training data set as described in Section III-B. An additional and independent test dataset consisting of $10^5$ network instances is generated in the same way to test the proposed precoder performance.

The training of the model is concurrently performed for all $\rho = \{0, 0.1, \ldots, 1\}$ values on Nvidia Tesla V100 graphics processing units (GPUs). The training data set consists of $10^6$ channel instances and target angles, as explained above. The DNN is trained via an unsupervised learning approach using 95% of the training dataset to minimize the loss function given by (25) that aims to maximize the sum rate and minimize the CRLB for a given trade-off $\rho$. The training is performed until the loss function converges to its minimum value in both training and validation that is performed via the remaining 5% of the training data. During training, Adam optimizer with 0.001 learning rate and 10% dropout is employed to avoid overfitting. Table 2 summarizes the training parameters used in this study.

While the DNNs are trained using perfect channel information, the performance of the proposed method is also evaluated with imperfect CSI that are obtained via (11) to

**TABLE 1.** Sum rate of the DL-based precoder with various numbers of hidden layers and nodes in each layer ($\rho = 0.5$ $M = 20$, $K = 10$, $T = 2$).

| Number of Nodes | Number of Hidden Layers | Sum Rate (Relatively) |
|---|---|---|
| 512 | 5 | 100% |
| 256 | 4 | 100% |
| 128 | 4 | 99.9% |
| 128 | 3 | 90.4% |
| 128 | 2 | 73.8% |
| 64 | 5 | 74.5% |
| 64 | 4 | 68.3% |
| 64 | 3 | 56.2% |

**TABLE 2.** DNN training parameters.

| Parameter | Value/Description |
|---|---|
| Training Approach | Unsupervised Learning |
| Training Dataset | $10^6$ instances |
| Test Dataset | $10^5$ instances |
| Batch Size | 500 |
| Optimizer | Adam |
| Learning Rate | 0.001 |
| Dropout Rate | 10% (to prevent overfitting) |
| Activation Function | Sigmoid Linear Unit (SiLU) |

reveal its performance under imperfect CSI. After training is completed, the performance of the proposed approach is tested using the test dataset ($10^5$ instances), and the results presented below in this section are the average of many network instances.

First, the impact of the DNN size on the learning of communication and sensing performances is examined to determine a reasonable DNN size. For this, we consider that the BS incorporates $M = 20$ transmit and $R = 10$ receive antennas, two targets located at the angles $\theta_1 = -10°$ and $\theta_2 = 25°$, and $K = 10$ UEs. Moreover, UEs are randomly located, the BS transmit power is $P_t = 1$ W, and the noise variance at UE receivers is $\sigma_k^2 = 10^{-3}$ W. The trade-off parameter is chosen as $\rho = 0.5$ to have a balance between communication sum rate and sensing accuracy. The number of neural network (NN) layers, including input and output layers, is chosen as 1, 2, 3, 4, 5, and 6 hidden layers, and the number of nodes in each layer is chosen as 16, 32, 64, 128, 256, 512, and 1024 nodes. Hence, a total of 42 different DNN configurations were trained. The average results of $10^5$ inference instances of the trained DNN above configurations were examined as shown in Table 1, and we observed that

- a minimal DNN configuration (e.g., 2 hidden layers with 32 nodes in each layer) could not learn the ISAC precoding to achieve a satisfactory sum rate, and
- an overly complex DNN configuration was not necessary (e.g., 5 or more hidden layers with 256 nodes or more in each layer).

Accordingly, the DNN configuration consisting of 4 hidden layers and 128 nodes in each layer was chosen as the configuration providing the maximum learning performance with the lowest computational complexity for the ISAC precoder design.

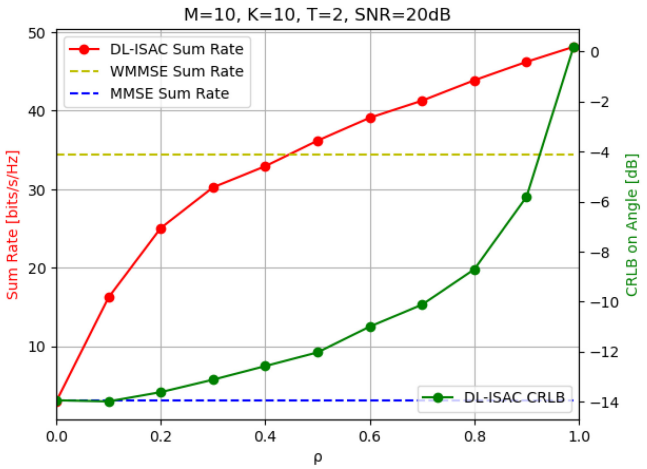


**FIGURE 3.** Sensing and communication performance $M = 10$ and $K = 10$.

### B. SENSING ACCURACY AND COMMUNICATION PERFORMANCE

Sensing and communication performances are examined for two different transmit array configurations, namely $M = 10$ and $M = 20$ antennas, while the number of radar receive antennas, the number of UEs, the number of targets, and SNR are $R = 10$, $K = 10$, $T = 2$, and $SNR = 20$dB, respectively. The targets are located at $\theta_1 = -10°$ and $\theta_1 = 25°$. The MMSE precoder and the iterative WMMSE precoder are utilized as the communication sum rate benchmarks [60] since it is highly complicated to derive an iterative ISAC precoder for the given problem. Note that the sum rate performance of the iterative WMMSE algorithm depends on the initial parameters, and it cannot, therefore, reach the optimum communication performance.

Note that the UEs can encounter significantly different path losses due to their random location in the network. This introduces an additional challenge to linear precoders, e.g., zero-forcing (ZF) or minimum-mean-squared-error (MMSE), especially when $M$ is only slightly larger than $K$.

Fig. 3 and Fig. 4 illustrate the communication sum rate and sensing sum CRLB in terms of dB, i.e., $10\log_{10}(\sum_{t=1}^{T}\psi(\theta_t))$, during the inference for $M = 10$ and $M = 20$, respectively. In the fully-loaded MIMO case of Fig. 3, where $M = 10$, $K = 10$, and $T = 2$, the proposed technique outperforms the WMMSE precoder when $\rho \geq 0.5$, and in this configuration, the MMSE precoder cannot perform precoding due to fully-loaded antennas. For $M = 20$, Fig. 4 shows that the proposed DNN-based precoder can still outperform the sum rate of the WMMSE when $\rho \geq 0.6$. These figures also demonstrate that varying $\rho$ provides a desired trade-off between the communication sum rate and sensing sum CRLB.

**TABLE 3.** Sum rate and complexity of the DNN after quantization and pruning ($\rho = 0.5$, $M = 10$, $K = 10$).

| Quantization<br>Pruning | Q=32-bit<br>J=0% | Q=7-bit<br>J=10% | Q=7-bit<br>J=20% | Q=6-bit<br>J=10% | Q=6-bit<br>J=20% | Q=5-bit<br>J=10% | Q=5-bit<br>J=20% | Q=5-bit<br>J=30% | Q=4-bit<br>J=20% |
|---|---|---|---|---|---|---|---|---|---|
| Complexity<br>(Relatively) | 100% | 19.68% | 17.5% | 16.88% | 15% | 14.06% | 12.5% | 10.93% | 10% |
| Sum Rate<br>[bits/s/Hz] | 36.1 | 35.3 | 33.5 | 34.7 | 32.5 | 32.7 | 30.75 | 27.44 | 27.66 |
| Sum Rate<br>(Relatively) | 100% | 97.8% | 92.8% | 96.1% | 90% | 90.6% | 85.2% | 76% | 76.6% |

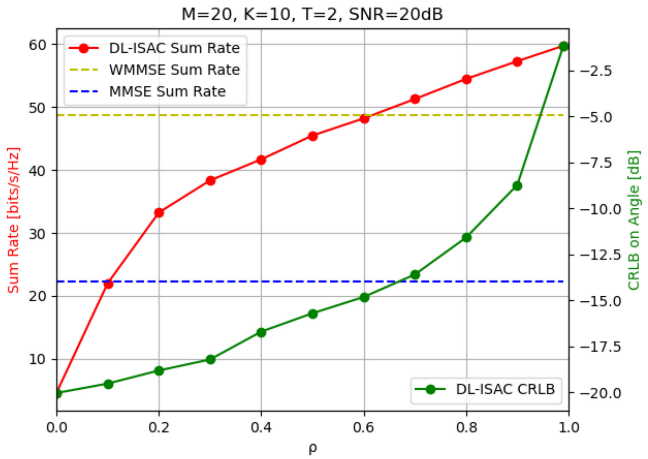


**FIGURE 4.** Sensing and communication performance $M = 20$ and $K = 10$.

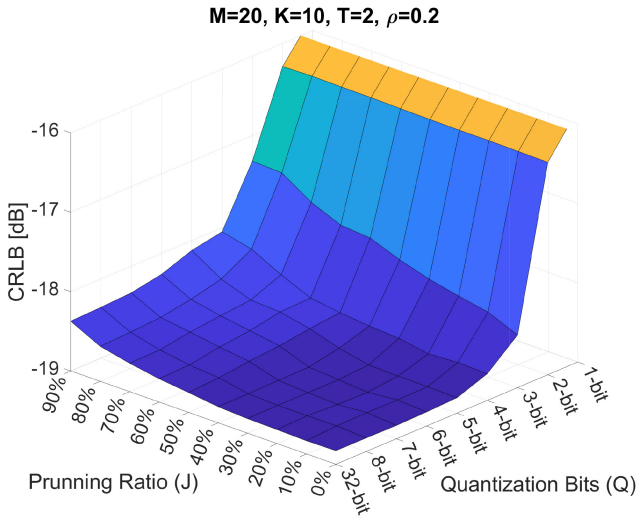


**FIGURE 6.** Sum CRLB of the network with weight quantization and pruning.

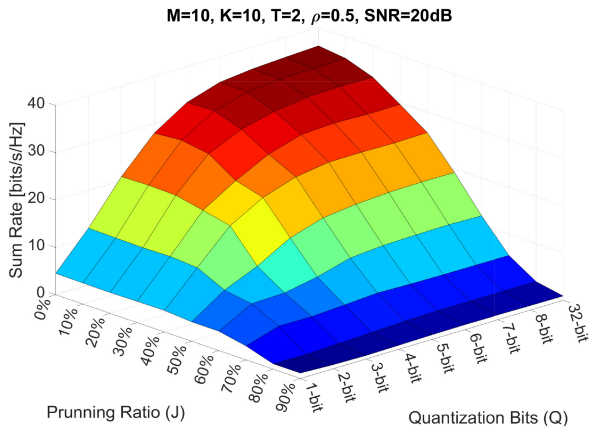


**FIGURE 5.** Sum rate of the network with weight quantization and pruning, M=10.

### C. QUANTIZATION AND PRUNING

Fig. 5 shows the sum rate of the ISAC system with $\rho = 0.5$ when the DNN weights are systematically quantized with $Q$ bits and pruned with various ratios $J$. It can be seen that the DNN with 8-bit and 7-bit weights can approximately achieve the sum rate of the DNN with 32-bit weights. Some significant points with more details on this figure are given in Table 3, where it is shown that a DNN quantized with $Q = 7$ bits and 10% pruning achieves 97.8% of the sum rate attained by full DNN. The number of bits that need to be stored and computed is only 19.68% of the full model, resulting in substantially reduced computational complexity and memory usage by quantization and pruning.

A more compact DNN model with $Q = 4$ bits and $J = 20\%$ achieves 76.6% of the sum rate of the full model with only 10% of its computational complexity. Fig. 6 presents the sum CRLB of the two targets as a function of $Q$ and $J$ when $\rho = 0.2$. This figure demonstrates that beamforming on the targets can be handled with a less sophisticated DNN since beamforming on the targets considers only LOS links. In contrast, precoding for communication considers both LOS and NLOS links, thus making it a more challenging problem.

The trade-off between the communication sum rate and sensing performance is presented in Fig. 7. Having a larger number of antennas enables the ISAC BS to achieve a higher sum rate and lower CRLB due to increasing degrees of freedom. It is also seen that the impact of the DNN weight pruning and quantization on the sensing performance is limited, as also seen in Fig. 6.

Both the communication sum rate and target angle estimation are affected by noise. Fig. 8 illustrates the sum rate and CRLB as functions of SNR with different weight quantization and pruning ratios, where the trade-off factor between communication and sensing is $\rho = 0.5$ and $M = 10$, $K = 10$, $T = 2$. It is observed that when the SNR is low, even the significantly quantized and pruned DNN can perform

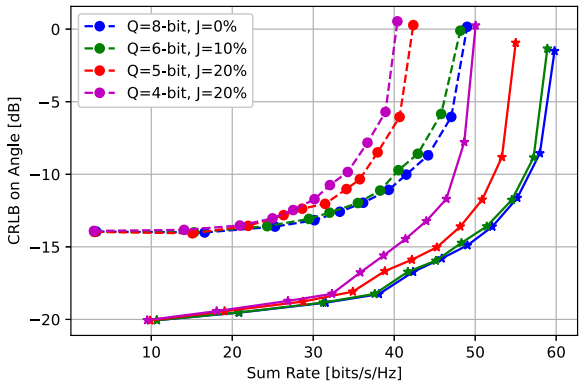


**FIGURE 7.** Communication (sum rate) and sensing (sum CRLB) trade-off with regard to weight quantization (Q) and pruning ratio (J). $M = 20$ is indicated by solid lines with $*$ while $M = 10$ is indicated by dashed lines with $\circ$.

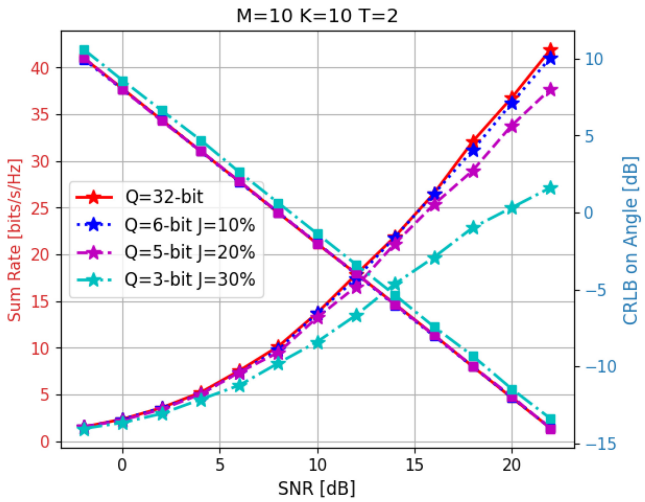


**FIGURE 8.** Communication sum rate and sum CRLB on target angle estimation obtained by DL-based precoders with various quantization bits and pruning ratios as functions of SNR. Lines with star markers indicate the sum rate, while lines with square markers indicate sum CRLB. $M = 10$, $K = 10$, $T = 2$, $\rho = 0.5$.

as well as the full DNN model in terms of communication performance since the sum rate is more restricted by noise rather than interference in low SNR regions. Moreover, 6-bit quantized and 10% pruned DNN can still perform satisfactorily and nearly reach the performance of full DNN even in high SNR regions, as also shown in Table 3.

Target angle estimation is also affected by the noisy received echoes from the target, as seen in this figure. For instance, when the SNR is low, the received echoes from the target are noise-dominated, hence, the CRLB is higher, leading to significant angle estimation errors. Therefore, increasing SNR improves angle estimation accuracy. Fig. 8 shows that the theoretical error bound on the target angle estimation can be as low as $-14$ dB when the SNR is sufficiently high. This figure also shows that target angle estimation is less affected than communication sum rate by quantization and pruning, as is also seen in the beam pattern figures presented in the next section.

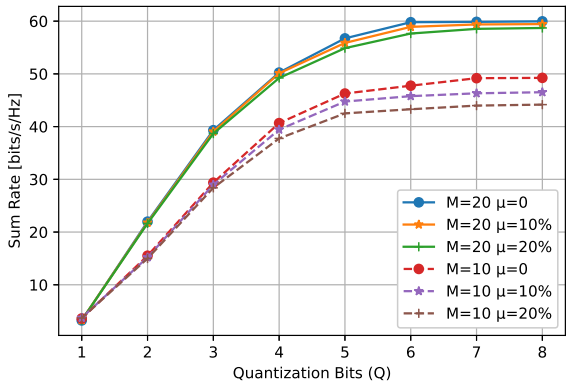


**FIGURE 9.** Impact of channel estimation errors on the sum rate, $\rho = 1$.

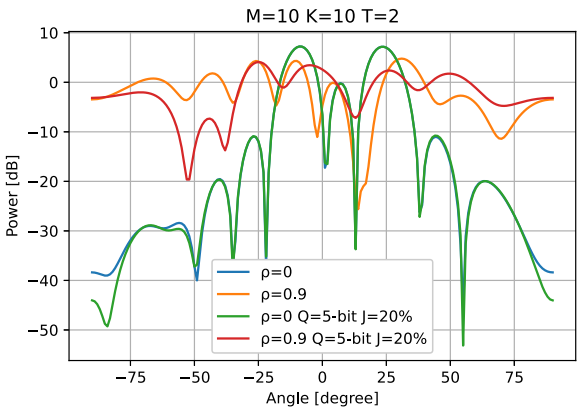


**FIGURE 10.** Beampattern of the ISAC BS when $M = 10$ $K = 10$ with quantization and pruning.

Fig. 9 depicts the sum rate for $M = 20$ and $M = 10$ when channel estimation errors are considered as (11). Here, the sum rate is maximized, i.e., $\rho = 1$. This figure shows that the proposed DL-based precoder is robust to channel estimation errors. It can also be observed that the precoder is more robust to errors in systems with a larger number of antennas since the sum rate loss is minimal even with $\mu = 20\%$ when $M = 20$.

### D. BEAMPATTERNS

Fig. 11 and Fig. 10 illustrate the beam patterns of the ISAC system with $M = 20$ and $M = 10$ antennas. Although the proposed technique relies on an unsupervised learning framework, these figures show excellent beamforming patterns for target sensing when $\rho = 0$. The minimum beamwidth of the beams is limited by the number of antennas, such that when the number of transmit antennas is $M = 20$, the ISAC system produces narrower beams compared to the case that has $M = 10$ antennas, as seen in the comparison of Fig. 10 and Fig. 11. This, in turn, improves the target angle estimation and communication sum rate as the number of antennas increases, as previously shown in Fig. 7.

It can be seen that the beam patterns vary depending on the $\rho$ to adjust the trade-off between the sensing and communication performances in Fig. 11 and Fig. 10.

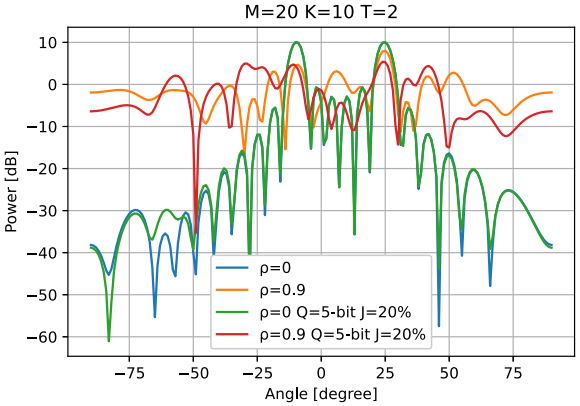


**FIGURE 11.** Beampattern of the ISAC BS when $M = 20$ $K = 10$ with quantization and pruning.

Furthermore, these figures also depict the impact of the selected weight quantization, $Q = 5 - bit$, and pruning, $J = 20\%$ on the beam patterns. Quantization and pruning especially affect the beam patterns when the ISAC system aims to maximize the communication performance, i.e., $\rho = 0.9$, following the trend seen in Fig. 7. Because beamforming for targets considers mainly LOS signals, while beamforming for communication considers both LOS and NLOS signals, it requires more sophisticated DNN architecture.

The proposed unsupervised learning framework can be applied to different problems in ISAC precoder design; for instance, by introducing a minimum capacity penalty term in the loss function for each UE, the sum capacity maximization problem under the quality of service (QoS) constraint can be solved. Moreover, hybrid beamforming techniques for ISAC can also be developed by following a similar approach, where the loss function would aim to design a suitable beamforming strategy based on designing both digital beamforming and analog beamforming matrices to maximize both communication sum rate and sensing accuracy for a given trade-off. It can also be applied to energy efficiency maximization problems or multicell networks. The proposed DL-based ISAC waveform design will be implemented in a future study on a hardware platform, such as on the UCL ARESTOR, which has a flexible modular architecture and supports multiple RF chain and antenna experiments and measurements and ARM cores that can run deep learning models [39], [40].

## VI. CONCLUSION

This work has proposed an unsupervised learning DL-based lightweight precoding method for ISAC systems that allows adjusting the trade-off between communication and sensing depending on the instantaneous sensing and communication requirements. The proposed DL-based method eliminates the need to solve non-convex optimization problems for the precoder design. Moreover, its computational complexity and memory requirements are further reduced by weight quantization and pruning. It is shown that a quantized and pruned model that has only 17% complexity of the full DNN model can achieve 96% of the performance of the full DNN model in terms of sum rate. Moreover, the proposed method has been shown to be robust under operating CSI estimation errors. In future work, we will work on distributed deep learning architectures to reduce the computational complexity of precoder design in multicell ISAC systems.

## ACKNOWLEDGMENT

The authors would like to thank Adrian Garcia-Rodriguez, AI Systems & Research Department, Ericsson, France, for his valuable feedback and contributions to the manuscript.

## REFERENCES

[1] C.-X. Wang et al., "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, 2nd Quart., 2023.
[2] Z. Wei, F. Liu, C. Masouros, N. Su, and A. P. Petropulu, "Toward multi-functional 6G wireless networks: Integrating sensing, communication, and security," *IEEE Commun. Mag.*, vol. 60, no. 4, pp. 65–71, Apr. 2022.
[3] Z. Wei et al., "Integrated sensing and communication signals toward 5G-A and 6G: A survey," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11068–11092, Jul. 2023.
[4] M. Temiz, E. Alsusa, and M. W. Baidas, "A dual-functional massive MIMO OFDM communication and radar transmitter architecture," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14974–14988, Dec. 2020.
[5] F. Liu et al., "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.
[6] M. Temiz, E. Alsusa, and M. W. Baidas, "Optimized precoders for massive MIMO OFDM dual radar-communication systems," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4781–4794, Jul. 2021.
[7] H. Hua, J. Xu, and T. X. Han, "Optimal transmit beamforming for integrated sensing and communication," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10588–10603, Aug. 2023.
[8] X. Li et al., "Integrated sensing, communication, and computation over-the-air: MIMO beamforming design," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5383–5398, Aug. 2023.
[9] N. Zhao, Y. Wang, Z. Zhang, Q. Chang, and Y. Shen, "Joint transmit and receive beamforming design for integrated sensing and communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 662–666, Mar. 2022.
[10] Z. Liu, S. Aditya, H. Li, and B. Clerckx, "Joint transmit and receive beamforming design in full-duplex integrated sensing and communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2907–2919, Sep. 2023.
[11] Z. Lyu, G. Zhu, and J. Xu, "Joint maneuver and beamforming design for UAV-enabled integrated sensing and communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2424–2440, Apr. 2023.
[12] C. Qi, W. Ci, J. Zhang, and X. You, "Hybrid beamforming for millimeter wave MIMO integrated sensing and communications," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1136–1140, May 2022.
[13] J. Chu, R. Liu, M. Li, Y. Liu, and Q. Liu, "Joint secure transmit beamforming designs for integrated sensing and communication systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4778–4791, Apr. 2023.
[14] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér-Rao bound optimization for joint radar-communication beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 240–253, 2022.
[15] Z. Xiao and Y. Zeng, "Waveform design and performance analysis for full-duplex integrated sensing and communication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1823–1837, Jun. 2022.
[16] Z. He, W. Xu, H. Shen, D. W. K. Ng, Y. C. Eldar, and X. You, "Full-duplex communication for ISAC: Joint beamforming and power optimization," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2920–2936, Sep. 2023.

[17] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.

[18] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei, "Deep neural network approximation theory," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2581–2623, May 2021.

[19] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised learning for parametric optimization," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 678–681, Mar. 2021.

[20] Y. Tian, C. Lu, X. Zhang, K. C. Tan, and Y. Jin, "Solving large-scale multiobjective optimization problems with sparse optimal solutions via unsupervised neural networks," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3115–3128, Jun. 2021.

[21] H. Dahrouj et al., "An overview of machine learning-based techniques for solving optimization problems in communications and signal processing," *IEEE Access*, vol. 9, pp. 74908–74938, 2021.

[22] K. M. Attiah, F. Sohrabi, and W. Yu, "Deep learning for channel sensing and hybrid precoding in TDD massive MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10839–10853, Dec. 2022.

[23] M. M. H. Shuvo, S. K. Islam, J. Cheng, and B. I. Morshed, "Efficient acceleration of deep learning inference on resource-constrained edge devices: A review," *Proc. IEEE*, vol. 111, no. 1, pp. 42–91, Jan. 2023.

[24] A. Agrawal, S. Barratt, and S. Boyd, "Learning convex optimization models," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 8, pp. 1355–1364, Aug. 2021.

[25] D. Liu, C. Sun, C. Yang, and L. Hanzo, "Optimizing wireless systems using unsupervised and reinforced-unsupervised deep learning," *IEEE Netw.*, vol. 34, no. 4, pp. 270–277, Jul./Aug. 2020.

[26] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1065–1069, Jan. 2020.

[27] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Unsupervised deep learning for massive MIMO hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7086–7099, Nov. 2021.

[28] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Decentralized beamforming for cell-free massive MIMO with unsupervised learning," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1042–1046, May 2022.

[29] Y. Cui, J. Guo, C.-K. Wen, S. Jin, and S. Han, "Unsupervised online learning in deep learning-based massive MIMO CSI feedback," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2086–2090, Sep. 2022.

[30] Z. Liu, Y. Yang, F. Gao, T. Zhou, and H. Ma, "Deep unsupervised learning for joint antenna selection and hybrid beamforming," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1697–1710, Mar. 2022.

[31] Y. Zhang, J. Zhang, S. Buzzi, H. Xiao, and B. Ai, "Unsupervised deep learning for power control of cell-free massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9585–9590, Jul. 2023.

[32] A. Mohammad, C. Masouros, and Y. Andreopoulos, "An unsupervised deep unfolding framework for robust symbol-level precoding," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1075–1090, 2023.

[33] X. Liu, H. Zhang, K. Long, A. Nallanathan, and V. C. M. Leung, "Distributed unsupervised learning for interference management in integrated sensing and communication systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9301–9312, Dec. 2023.

[34] C. Liu et al., "Learning-based predictive beamforming for integrated sensing and communication in vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2317–2334, Aug. 2022.

[35] J. M. Mateos-Ramos, B. Chatelier, C. Häger, M. F. Keskin, L. Le Magoarou, and H. Wymeersch, "Semi-supervised end-to-end learning for integrated sensing and communications," in *Proc. IEEE Int. Conf. Mach. Learn. Commun. Netw. (ICMLCN)*, May 2024, pp. 132–138.

[36] R. P. Sankar, S. S. Nair, S. Doshi, and S. P. Chepuri, "Learning to precode for integrated sensing and communication systems," in *Proc. 31st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2023, pp. 695–699.

[37] M. Temiz, C. Horne, N. J. Peters, M. A. Ritchie, and C. Masouros, "An experimental study of radar-centric transmission for integrated sensing and communications," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 7, pp. 3203–3216, Jul. 2023.

[38] T. Xu, F. Liu, C. Masouros, and I. Darwazeh, "An experimental proof of concept for integrated sensing and communications waveform design," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1643–1655, 2022.

[39] C. Horne, N. J. Peters, and M. A. Ritchie, "Classification of LoRa signals with real-time validation using the Xilinx radio frequency system-on-chip," *IEEE Access*, vol. 11, pp. 26211–26223, 2023.

[40] N. J. Peters, C. P. Horne, A. D. Amiri, P. Beasley, and M. A. Ritchie, "Modular multi-channel RFSoC system expansion and array design," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2023, pp. 1–4.

[41] C. Shen and M. P. Fitz, "MIMO-OFDM beamforming for improved channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 948–959, Aug. 2008.

[42] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, "Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3403–3418, Jul. 2018.

[43] J. Zhang, L. Dai, X. Zhang, E. Björnson, and Z. Wang, "Achievable rate of Rician large-scale MIMO channels with transceiver hardware impairments," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8800–8806, Oct. 2016.

[44] C. Chen and M. A. Jensen, "Secret key establishment using temporally and spatially correlated wireless channel coefficients," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 205–215, Feb. 2011.

[45] S. Sun et al., "Propagation path loss models for 5G urban micro- and macro-cellular scenarios," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, 2016, pp. 1–6.

[46] M. Temiz, E. Alsusa, L. Danoon, and Y. Zhang, "On the impact of antenna array geometry on indoor wideband massive MIMO networks," *IEEE Trans. Antennas Propag.*, vol. 69, no. 1, pp. 406–416, Jan. 2021.

[47] D. K. Barton, *Radar Equations for Modern Radar*. Boston, MA, USA: Artech House, 2013.

[48] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, "Channel estimation for OFDM," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1891–1908, 4th Quart., 2014.

[49] X. Yang, Z. Wei, J. Xu, Y. Fang, H. Wu, and Z. Feng, "Coordinated transmit beamforming for networked ISAC with imperfect CSI and time synchronization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 18019–18035, Dec. 2024.

[50] J. Dai, J. Ye, K. Wang, C. Pan, and H. Fan, "Joint radar-communication beamforming considering both transceiver hardware impairments and imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 13, no. 7, pp. 1898–1902, Jul. 2024.

[51] M. Temiz, E. Alsusa, and M. W. Baidas, "A dual-function massive MIMO uplink OFDM communication and radar architecture," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 750–762, Jun. 2022.

[52] J. Shi, A.-A. Lu, W. Zhong, X. Gao, and G. Y. Li, "Robust WMMSE precoder with deep learning design for massive MIMO," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 3963–3976, Jul. 2023.

[53] L. Pellaco and J. Jaldén, "A matrix-inverse-free implementation of the MU-MIMO WMMSE beamforming algorithm," *IEEE Trans. Signal Process.*, vol. 70, pp. 6360–6375, 2022.

[54] H. Shi, R. V. Prasad, E. Onur, and I. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 5–24, 1st Quart., 2014.

[55] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," Eastern Res. Lab., Digit. Equip. Corporation, Hudson, MA, USA, Rep. DEC Research-TR-301, 1984.

[56] T. Hoefler, D. Alistarh, T. Ben-Nun, N. Dryden, and A. Peste, "Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 10882–11005, 2021.

[57] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," 2019, *arXiv:1912.01703*. [Online]. Available: http://arxiv.org/abs/1912.01703

[58] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org/

[59] D. Blalock, J. J. Gonzalez Ortiz, J. Frankle, and J. Guttag, "What is the state of neural network pruning?" in *Proc. Mach. Learn. Syst.*, vol. 2, 2020, pp. 129–146.

[60] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility Maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.