Heterogeneous Graph Neural Network for Cooperative ISAC Beamforming in Cell-Free MIMO Systems

Zihuan Wang
zihuanwang@ece.ubc.ca
The University of British Columbia
Vancouver, Canada

Vincent W.S. Wong
vincentw@ece.ubc.ca
The University of British Columbia
Vancouver, Canada

ABSTRACT

Integrated sensing and communication (ISAC) is one of the usage scenarios for the sixth generation (6G) wireless networks. In this paper, we study cooperative ISAC in cell-free multiple-input multiple-output (MIMO) systems, where multiple MIMO access points (APs) collaboratively provide communication services and perform multi-static sensing. We formulate an optimization problem for the ISAC beamforming design, which maximizes the achievable sum-rate while guaranteeing the sensing signal-to-noise ratio (SNR) requirement and total power constraint. Learning-based techniques are regarded as a promising approach for addressing such a nonconvex optimization problem. By taking the topology of cell-free MIMO systems into consideration, we propose a heterogeneous graph neural network (GNN), namely SACGNN, for ISAC beamforming design. The proposed SACGNN framework models the cell-free MIMO system for cooperative ISAC as a heterogeneous graph and employs a transformer-based heterogeneous message passing scheme to capture the important information of sensing and communication channels and propagate the information through the graph network. Simulation results demonstrate the performance gain of the proposed SACGNN framework over a conventional null-space projection based scheme and a deep neural network (DNN)-based baseline scheme.

CCS CONCEPTS

• **Networks** → **Network architectures**; **Network simulations**;
• **Hardware** → *Emerging technologies*.

KEYWORDS

Beamforming, cell-free multiple-input multiple-output (MIMO), integrated sensing and communication (ISAC), heterogeneous graph neural network (GNN).

1 INTRODUCTION

Integrated sensing and communication (ISAC) is envisioned to play an indispensable role in the sixth generation (6G) wireless networks, supporting immersive communications and pervasive sensing, and facilitating paradigm shifts towards the metaverse [4, 8, 11, 14]. Benefiting from the terahertz frequency bands and large antenna arrays, ISAC systems can provide sensing and communication functionalities simultaneously through multiple-input multiple-output (MIMO) beamforming, where the sensing and communication modules share the same infrastructure and spectrum resources. Thus, ISAC can reduce hardware cost and improve the spectral and energy efficiencies for wireless communication networks.

Extensive efforts have been devoted to theoretical analysis and testbed development for ISAC beamforming design [3, 8, 15–17]. These works focus on single base station scenarios. In order to further enhance the spatial degrees of freedom (DoFs), cooperative communication and multi-static sensing are crucial. This motivates the investigation of cooperative ISAC in cell-free MIMO systems. In such cell-free MIMO systems, multiple access points (APs) collaboratively provide seamless communication service and receive uncorrelated sensing observations. The distributed APs are connected to a central processing unit (CPU), which facilitates joint processing between the APs [1, 2, 6, 9]. In [1, 6], power allocation is investigated to maximize the sensing performance while ensuring that the communication performance meets its requirements. Cooperative ISAC beamforming design in cell-free MIMO networks is investigated in [2]. A semidefinite relaxation (SDR) based algorithm is proposed for transmit beamforming design which aims to maximize the sensing signal-to-noise ratio (SNR) while satisfying the communication signal-to-interference-plus-noise ratio (SINR) requirement. In [9], the mode selection of APs is studied for joint communication and multi-static sensing, where each AP can operate either as a transmitter or receiver.

The aforementioned works rely on alternating optimization algorithms for power allocation and beamforming design, which can incur a high computational complexity. Recently, deep learning techniques, which train the deep neural networks (DNNs) offline and then deploy the trained models for online optimization, have proven to reduce the computational complexity of solving optimization problems in wireless networks. Among the various types of DNNs, graph neural networks (GNNs) [5, 10, 13] have demonstrated significant potential for modeling structured data. This makes them particularly well-suited for solving optimization problems in wireless networks with specific architectures, such as cell-free systems. Existing GNNs for wireless networks typically account only for communication functionality, and they primarily focus on extracting and passing the message carried by communication channels.

ISAC systems, on the other hand, also involve sensing information, which should be incorporated into the graph modeling and message passing procedures.

In this paper, we investigate cooperative ISAC beamforming in cell-free MIMO systems. We develop a heterogeneous GNN, which is called sensing and communication GNN (SACGNN), for ISAC beamforming design to maximize the achievable sum-rate while guaranteeing the sensing SNR and total power constraints. The proposed SACGNN framework models the cell-free MIMO system as a heterogeneous graph, with AP antennas and users being different types of nodes. Then, we apply a transformer-based message passing scheme, which extracts important information from sensing and communication channels with attention mechanisms. We conduct simulations to evaluate the performance of our proposed SACGNN framework. The results demonstrate the effectiveness of the proposed approach over a conventional null-space projection based scheme and a DNN-based baseline scheme.

2 SYSTEM MODEL

We consider a cell-free MIMO system for joint communication and multi-static sensing. There are N_T transmit APs and N_R receive APs, which are synchronized and connected to a CPU via fronthaul links, as shown in Fig. 1. Each AP is equipped with a uniform planar array (UPA) with $M = M_v M_h$ antennas, where M_v and M_h denote the number of antennas in the vertical and horizontal dimensions, respectively. For a half-wavelength-spaced UPA, the vertical and horizontal beam steering vectors are respectively given by

$$\mathbf{b}_v(\theta) = \frac{1}{\sqrt{M_v}} \left[1 \ e^{-j\pi \cos\theta} \ \cdots \ e^{-j(M_v-1)\pi\cos\theta} \right]^T, \qquad (1)$$

$$\mathbf{b}_h(\phi,\theta) = \frac{1}{\sqrt{M_h}} \left[1 \ e^{-j\pi\sin\theta\cos\phi} \ \cdots \ e^{-j(M_h-1)\pi\sin\theta\cos\phi} \right]^T, (2)$$

where ϕ and θ are the azimuth and elevation angles, respectively. The beam steering vector can be expressed as $\mathbf{a}(\phi,\theta) \in \mathbb{C}^M$:

$$\mathbf{a}(\phi,\theta) = \mathbf{b}_h(\phi,\theta) \otimes \mathbf{b}_v(\theta). \qquad (3)$$

The N_T transmit APs jointly communicate with K users through beamforming and steer another beamformer toward a specific location for target detection. Note that the beamformers for communication and sensing share the same waveform and time-frequency resources. On the other hand, the N_R receive APs obtain the reflected echo signals, which are used to determine whether there is a target at that particular location or not.

2.1 Signal Model

We consider ISAC downlink transmission, where the transmit APs collaboratively send K communication data streams $\{s_k\}_{k=1}^K$ to K users and transmit another data stream s_0 for target detection. The total number of streams is equal to $K + 1$. Each transmitted symbol has unit power, i.e., $\mathbb{E}\{|s_k|^2\} = 1$, for $k = 0, \ldots, K$. We define $\mathbf{s} = [s_0 \ \cdots \ s_K]^T \in \mathbb{C}^{K+1}$ as the transmitted streams. Similar to [2], we assume the streams are statistically independent, i.e., $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_{K+1}$. Let $\mathbf{x}_i \in \mathbb{C}^M$ denote the transmitted signal at the
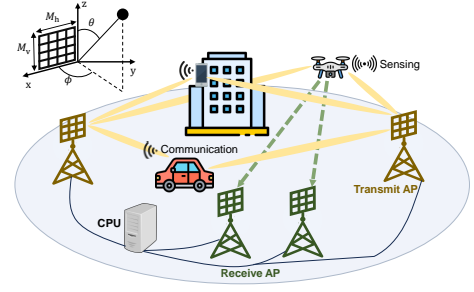

Figure 1: The system model of cooperative ISAC transmission in a cell-free MIMO system. The transmit APs jointly communicate with multiple users and sense a target through beamforming. The receive APs obtain the echo signal reflected by the target. The transmit and receive APs are connected to a CPU through fronthaul links.

i-th transmit AP. It can be expressed as

$$\mathbf{x}_i = \underbrace{\sum_{k=1}^K \mathbf{f}_{i,k} s_k}_{\text{Communication}} + \underbrace{\mathbf{f}_{i,0} s_0}_{\text{Sensing}} = \sum_{k=0}^K \mathbf{f}_{i,k} s_k = \mathbf{F}_i \mathbf{s}, \qquad (4)$$

where $\mathbf{f}_{i,k} \in \mathbb{C}^M$ is the beamforming vector of the i-th transmit AP for the k-th data stream. We define $\mathbf{F}_i = [\mathbf{f}_{i,0} \ \cdots \ \mathbf{f}_{i,K}] \in \mathbb{C}^{M\times(K+1)}$ as the beamforming matrix at the i-th transmit AP. The transmit power at the i-th transmit AP is given by $P_i = \|\mathbf{F}_i\|_F^2$, $i = 1, \ldots, N_T$. Each transmit AP has a constraint on the total power consumption, i.e., $P_i \leq P_{\max}$, for $i = 1, \ldots, N_T$, where $P_{\max}$ is the maximum transmit power.

2.2 Communication Model

Let $\mathbf{h}_{i,k} \in \mathbb{C}^M$ denote the channel vector between the i-th transmit AP and the k-th user, $i = 1, \ldots, N_T$, $k = 1, \ldots, K$. Assuming there is a line-of-sight (LoS) link between each transmit AP and user, the channel vector $\mathbf{h}_{i,k}$ can be expressed as

$$\mathbf{h}_{i,k} = \beta_{i,k} \mathbf{a}(\phi_{i,k}, \theta_{i,k}), \ i = 1, \ldots, N_T, \ k = 1, \ldots, K, \qquad (5)$$

where $\beta_{i,k}$ denotes the complex channel gain, and it follows a Gaussian distribution with zero mean and variance $\zeta_{i,k}^2$. $\phi_{i,k}$ and $\theta_{i,k}$ are the angles of departure (AoDs) in the azimuth and elevation domains of the k-th user observed at the i-th transmit AP, respectively. By stacking the channel vectors between the k-th user and all the transmit APs, we define $\mathbf{h}_k = [\mathbf{h}_{1,k}^T \ \cdots \ \mathbf{h}_{N_T,k}^T]^T \in \mathbb{C}^{MN_T}$. By stacking the beamforming vectors of all the transmit APs for the k-th user, we define $\mathbf{f}_k = [\mathbf{f}_{1,k}^T \ \cdots \ \mathbf{f}_{N_T,k}^T]^T \in \mathbb{C}^{MN_T}$. Then, the received downlink signal at the k-th user can be expressed as follows:

$$y_k^{(c)} = \sum_{i=1}^{N_T} \mathbf{h}_{i,k}^H \mathbf{x}_i + n_k \qquad (6)$$

$$= \underbrace{\mathbf{h}_k^H \mathbf{f}_k s_k}_{\text{Desired signal}} + \underbrace{\sum_{m=1, m\neq k}^K \mathbf{h}_k^H \mathbf{f}_m s_m}_{\text{Multi-user interference}} + \underbrace{\mathbf{h}_k^H \mathbf{f}_0 s_0}_{\text{Sensing interference}} + \underbrace{n_k}_{\text{Noise}}, (7)$$

where n_k denotes white Gaussian noise of the k-th user with zero mean and variance σ_k^2. Then, the SINR of the k-th user, i.e., $\gamma_k^{(c)}$, can be obtained as follows:

$$\gamma_k^{(c)} = \frac{|\mathbf{h}_k^H \mathbf{f}_k|^2}{\sum_{m=1,m\neq k}^{K} |\mathbf{h}_k^H \mathbf{f}_m|^2 + |\mathbf{h}_k^H \mathbf{f}_0|^2 + \sigma_k^2}. \tag{8}$$

The achievable sum-rate of all K users is given by

$$R = \sum_{k=1}^{K} \log_2\left(1 + \gamma_k^{(c)}\right). \tag{9}$$

2.3 Sensing Model

If a target exists at the point of interest, then the transmitted downlink signals will be reflected by the target and the reflected echo signal will be collected by the receive APs. Assuming a LoS path exists between each AP and the target, the echo signal obtained at the j-th receive AP, where $j = 1, \ldots, N_R$, is given by[1]

$$\mathbf{y}_j^{(s)} = \sum_{i=1}^{N_T} \lambda_{i,j} \mathbf{a}(\varphi_j^{(rx)}, \vartheta_j^{(rx)}) \mathbf{a}^H(\varphi_i^{(tx)}, \vartheta_i^{(tx)}) \mathbf{x}_i + \mathbf{z}_j, \tag{10}$$

where $\lambda_{i,j}$ is an unknown complex sensing channel gain. $\lambda_{i,j}$ includes the effects due to path loss and radar cross section of the target. It follows the Gaussian distribution with zero mean and variance $\chi_{i,j}^2$. $\varphi_i^{(tx)}$ and $\vartheta_i^{(tx)}$ correspond to the azimuth and elevation AoDs of the target observed at the i-th transmit AP, respectively. $\varphi_j^{(rx)}$ and $\vartheta_j^{(rx)}$ correspond to the azimuth and elevation angles of arrival (AoAs) of the target observed at the j-th receive AP, respectively. $\mathbf{z}_j \in \mathbb{C}^M$ is the noise at the j-th receive AP, which follows the complex Gaussian distribution with zero mean and variance of $\xi_j^2 \mathbf{I}_M$.

We define matrix $\mathbf{A}_{i,j} \triangleq \lambda_{i,j} \mathbf{a}(\varphi_j^{(rx)}, \vartheta_j^{(rx)}) \mathbf{a}^H(\varphi_i^{(tx)}, \vartheta_i^{(tx)}) \in \mathbb{C}^{M \times M}$. A compact form of the obtained echo signal (10) at the j-th receive AP can be expressed as

$$\mathbf{y}_j^{(s)} = \mathbf{A}_j \mathbf{x} + \mathbf{z}_j, \tag{11}$$

where matrix $\mathbf{A}_j = [\mathbf{A}_{1,j} \cdots \mathbf{A}_{N_T,j}] \in \mathbb{C}^{M \times N_T M}$ and vector $\mathbf{x} = [\mathbf{x}_1^T \cdots \mathbf{x}_{N_T}^T]^T \in \mathbb{C}^{N_T M}$. Then, each receive AP forwards the echo signal (11) to the CPU. The concatenated echo signal obtained from all the receive APs is given by $\mathbf{y}^{(s)} = [(\mathbf{y}_1^{(s)})^T \cdots (\mathbf{y}_{N_R}^{(s)})^T]^T \in \mathbb{C}^{N_R M}$. It can be expressed as

$$\mathbf{y}^{(s)} = \mathbf{A} \mathbf{x} + \mathbf{z}, \tag{12}$$

where $\mathbf{A} = [\mathbf{A}_1^T \cdots \mathbf{A}_{N_R}^T]^T \in \mathbb{C}^{N_R M \times N_T M}$ and $\mathbf{z} = [\mathbf{z}_1^T \cdots \mathbf{z}_{N_R}^T]^T \in \mathbb{C}^{N_R M}$. The sensing SNR for target detection is given by

$$\gamma^{(s)} = \frac{\mathbb{E}\{\|\mathbf{A}\mathbf{x}\|^2\}}{\mathbb{E}\{\|\mathbf{z}\|^2\}} = \frac{\sum_{j=1}^{N_R} \sum_{i=1}^{N_T} \chi_{i,j}^2 \|\widetilde{\mathbf{A}}_{i,j} \mathbf{F}_i\|_F^2}{\sum_{j=1}^{N_R} \xi_j^2}, \tag{13}$$

where we define $\widetilde{\mathbf{A}}_{i,j} \triangleq \mathbf{a}(\varphi_j^{(rx)}, \vartheta_j^{(rx)}) \mathbf{a}^H(\varphi_i^{(tx)}, \vartheta_i^{(tx)}) \in \mathbb{C}^{M \times M}$.

---

[1]There are also direct communication links between the transmit and receive APs. We assume that these direct links are known prior to sensing, and their effects can be removed from the received echo signal at each receive AP [1]. Moreover, there may also exist multipath components. Similar to [1, 2], we assume the contribution of the multipath components is small and can be ignored for simplicity.

2.4 Problem Formulation

In this paper, we consider cooperative ISAC beamforming design to maximize the achievable sum-rate while guaranteeing the sensing SNR requirement and the total power constraint at each transmit AP. The optimization problem can be formulated as follows:

$$\underset{\{\mathbf{F}_i\}_{i=1}^{N_T}}{\text{maximize}} \quad R \tag{14a}$$

$$\text{subject to} \quad \gamma^{(s)} \geq \gamma_{\min}, \tag{14b}$$

$$\|\mathbf{F}_i\|_F^2 \leq P_{\max}, \ i = 1, \ldots, N_T, \tag{14c}$$

where $\gamma_{\min}$ is the minimum SNR requirement for target sensing. Note that the optimal solution to problem (14) is intractable due to the nonconvexity of the objective function. Motivated by the recent success of applying learning-based techniques to solve nonconvex optimization problems, we propose a SACGNN framework in the next section, which leverages graph learning to effectively capture the structural information of cell-free MIMO systems and determine the beamformers $\{\mathbf{F}_i\}_{i=1}^{N_T}$.

3 PROPOSED SACGNN FRAMEWORK

In this section, we introduce the SACGNN framework for cooperative ISAC beamforming design. The proposed SACGNN framework models the cell-free MIMO system as a heterogeneous graph and uses a transformer-based heterogeneous message passing scheme to extract important sensing and communication features.

3.1 Heterogeneous ISAC Graph Modeling

The considered cell-free MIMO system for cooperative ISAC can be modeled as a heterogeneous graph, denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which consists of a set of nodes $\mathcal{V}$ and the corresponding set of edges $\mathcal{E}$. The graph model consists of three types of nodes. The first type is the transmit AP antenna (tAP) which sends ISAC signals. The second type is the receive AP antenna (rAP) which obtains the reflected echo signals for sensing. The third type is the user equipment (UE) which receives communication service. We define the set of node types as $\mathcal{A} = \{\text{tAP, rAP, UE}\}$, and function $\psi(V) : \mathcal{V} \to \mathcal{A}$ which maps a node $V \in \mathcal{V}$ to its node type in $\mathcal{A}$. Accordingly, the set of nodes can be expressed as $\mathcal{V} = \mathcal{V}_{\text{tAP}} \cup \mathcal{V}_{\text{rAP}} \cup \mathcal{V}_{\text{UE}}$, which consists of tAP node set $\mathcal{V}_{\text{tAP}} = \{V_{\text{tAP},1}, \ldots, V_{\text{tAP},N_T M}\}$, rAP node set $\mathcal{V}_{\text{rAP}} = \{V_{\text{rAP},1}, \ldots, V_{\text{rAP},N_R M}\}$, and UE node set $\mathcal{V}_{\text{UE}} = \{V_{\text{UE},1}, \ldots, V_{\text{UE},K}\}$. Each pair of tAP and UE nodes is connected by an undirected edge, with the channel coefficient being their edge feature. In particular, as shown in Fig. 2, $\mathbf{h}_{i,k}(m)$ serves as the edge feature between node $V_{\text{tAP},(i-1)M+m}$ and node $V_{\text{UE},k}$. Note that $\mathbf{h}_{i,k}(m)$ represents the m-th element of vector $\mathbf{h}_{i,k}$, where $i = 1, \ldots, N_T, k = 1, \ldots, K, m = 1, \ldots, M$. The tAP nodes are also connected with the rAP nodes. $\widetilde{\mathbf{A}}_{i,j}(m, n)$ serves as the edge feature between node $V_{\text{tAP},(i-1)M+n}$ and node $V_{\text{rAP},(j-1)M+m}$. Note that $\widetilde{\mathbf{A}}_{i,j}(m, n)$ denotes the element in the m-th row and n-th column of matrix $\widetilde{\mathbf{A}}_{i,j}$, for $i = 1, \ldots, N_T, j = 1, \ldots, N_R, m = 1, \ldots, M, n = 1, \ldots, M$.

3.2 Heterogeneous Message Passing

Consider a SACGNN framework with L hidden layers and denote $\mathbf{g}^{(l)}[V]$ as the hidden state of the l-th layer of node $V \in \mathcal{V}$. Denote
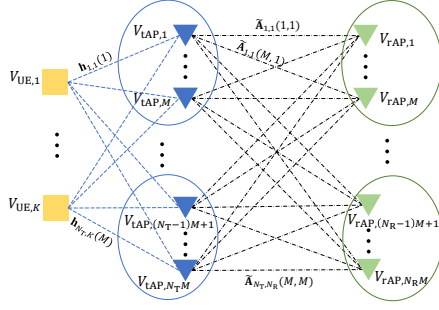
Figure 2: The heterogeneous graph model of a cell-free MIMO system for cooperative ISAC.

$\mathcal{N}(V)$ as the set of neighboring nodes that have edges connected to node $V \in \mathcal{V}$. In each layer, each node updates its own hidden state based on the aggregated message passed from its neighboring nodes. Due to node heterogeneity, the features related to various types of nodes may fall into different feature spaces. It is crucial to handle the diverse feature information and capture the important features. In this work, we apply a transformer [12] based scheme to determine the attention scores for the set of neighbors, which takes the importance of node heterogeneity into consideration.

In particular, we first initialize the hidden state for each node based on the edge features shown in Fig. 2. For a tAP node $V_{\text{tAP},i} \in \mathcal{V}_{\text{tAP}}$, where $i = 1, \ldots, N_T M$, the initial hidden state $\mathbf{g}^{(1)}[V_{\text{tAP},i}] \in \mathbb{C}^{K+N_R M}$ for the first layer combines the edge features associated with the i-th tAP node. It is given by

$$\mathbf{g}^{(1)}[V_{\text{tAP},i}] = \Big[\mathbf{h}_{\lfloor \frac{i}{M} \rfloor,1}(i - M\lfloor (i-1)/M \rfloor), \cdots,$$

$$\mathbf{h}_{\lfloor \frac{i}{M} \rfloor,K}(i - M\lfloor (i-1)/M \rfloor), \widetilde{\mathbf{A}}_{\lfloor \frac{i}{M} \rfloor,1}(i - M\lfloor (i-1)/M \rfloor, 1),$$

$$\cdots, \widetilde{\mathbf{A}}_{\lfloor \frac{i}{M} \rfloor,1}(i - M\lfloor (i-1)/M \rfloor, M), \widetilde{\mathbf{A}}_{\lfloor \frac{i}{M} \rfloor,2}(i - M\lfloor (i-1)/M \rfloor, 1),$$

$$\cdots, \widetilde{\mathbf{A}}_{\lfloor \frac{i}{M} \rfloor,N_R}(i - M\lfloor (i-1)/M \rfloor, M)\Big]^T, \quad (15)$$

where $\lfloor \cdot \rfloor$ represents the floor function. Similarly, for a rAP node $V_{\text{rAP},j} \in \mathcal{V}_{\text{rAP}}$, where $j = 1, \ldots, N_R M$, the initial value $\mathbf{g}^{(1)}[V_{\text{rAP},j}] \in \mathbb{C}^{N_T M}$ is characterized by the edge features associated with the j-th rAP node, which can be written as

$$\mathbf{g}^{(1)}[V_{\text{rAP},j}] = \Big[\widetilde{\mathbf{A}}_{1,\lfloor \frac{j}{M} \rfloor}(1, j - M\lfloor (j-1)/M \rfloor), \cdots,$$

$$\widetilde{\mathbf{A}}_{1,\lfloor \frac{j}{M} \rfloor}(M, j - M\lfloor (j-1)M \rfloor), \widetilde{\mathbf{A}}_{2,\lfloor \frac{j}{M} \rfloor}(1, j - M\lfloor (j-1)/M \rfloor),$$

$$\cdots, \widetilde{\mathbf{A}}_{N_T,\lfloor \frac{j}{M} \rfloor}(M, j - M\lfloor (j-1)/M \rfloor)\Big]^T. \quad (16)$$

For a UE node $V_{\text{UE},k} \in \mathcal{V}_{\text{UE}}$, where $k = 1, \ldots, K$, the initial value $\mathbf{g}^{(1)}[V_{\text{UE},k}] \in \mathbb{C}^{N_T M}$ is given by

$$\mathbf{g}^{(1)}[V_{\text{UE},k}] = \big[\mathbf{h}_{1,k}(1), \cdots, \mathbf{h}_{1,k}(M), \cdots, \mathbf{h}_{N_T,k}(M)\big]^T. \quad (17)$$

After initialization, for a node $V_t \in \mathcal{V}$ with its hidden state vector being $\mathbf{g}^{(l)}[V_t]$ in the l-th ($l > 1$) layer, our goal is to extract the important features from its neighbors $V_s \in \mathcal{N}(V_t)$ and update its hidden state for the next layer, i.e., $\mathbf{g}^{(l+1)}[V_t]$, through a transformer-based heterogeneous message passing scheme. By applying a residual connection, the updated hidden state $\mathbf{g}^{(l+1)}[V_t]$

can be expressed as follows:

$$\mathbf{g}^{(l+1)}[V_t] = f_{\text{agg}}\Big(f_{\text{att}}\big(V_t, \mathbf{g}^{(l)}[V_t]\big) \cdot \underset{V_s \in \mathcal{N}(V_t)}{f_{\text{msg}}}\big(\mathbf{g}^{(l)}[V_s]\big)\Big) + \mathbf{g}^{(l)}[V_t], \quad (18)$$

where $f_{\text{agg}}(\cdot)$, $f_{\text{att}}(\cdot)$, and $f_{\text{msg}}(\cdot)$ represent the functions for aggregation, attention calculation, and message calculation, respectively. The function $f_{\text{msg}}(\cdot)$ is responsible for extracting the message carried by the neighbors of node V_t. The importance of each neighboring node is evaluated through function $f_{\text{att}}(\cdot)$. Finally, the weighted message of the set of neighbors is aggregated through $f_{\text{agg}}(\cdot)$ to obtain the updated hidden state for node V_t.

We use a multi-head self-attention mechanism in function $f_{\text{att}}(\cdot)$. Considering N_{h} self-attention heads, for the h-th attention head, we map the hidden state $\mathbf{g}^{(l)}[V_t]$ of node $V_t \in \mathcal{V}$ into the h-th query vector $\mathbf{q}_h^{(l)}[V_t]$ through a linear projector $\mu_{\text{q}}(\cdot; \mathbf{W}_h^{\text{q}})$ with a weight matrix $\mathbf{W}_h^{\text{q}}$. Similarly, the hidden state $\mathbf{g}^{(l)}[V_s]$ of a neighboring node $V_s \in \mathcal{N}(V_t)$ with type $\psi(V_s)$ is projected into the h-th key vector $\mathbf{k}_h^{(l)}[V_s]$ with a linear projector $\mu_{\text{k},\psi(V_s)}(\cdot; \mathbf{W}_{h,\psi(V_s)}^{\text{k}})$. Note that $\mathbf{W}_{h,\psi(V_s)}^{\text{k}}$ is the weight matrix and different types of nodes have their own unique projectors with different weight matrices. Then, we calculate the similarity between the query and the key as the product $(\mathbf{k}_h^{(l)}[V_s])^T \mathbf{W}_{\psi(V_s)}^{\text{att}} \mathbf{q}_h^{(l)}[V_t]$, where $\mathbf{W}_{\psi(V_s)}^{\text{att}}$ is the weight matrix. By concatenating all the N_{h} self-attention heads together, we can obtain the overall attention vector for node V_t. The function $f_{\text{att}}(\cdot)$ can be expressed as follows:

$$f_{\text{att}}(V_t, \mathbf{g}^{(l)}[V_t]) = \underset{V_s \in \mathcal{N}(V_t)}{\text{softmax}}\Big(\big\|_h \big\{(\mathbf{k}_h^{(l)}[V_s])^T \mathbf{W}_{\psi(V_s)}^{\text{att}} \mathbf{q}_h^{(l)}[V_t]\big\}\Big), (19)$$

$$\mathbf{q}_h^{(l)}[V_t] = \mu_{\text{q}}(\mathbf{g}^{(l)}[V_t]; \mathbf{W}_h^{\text{q}}), \quad (20)$$

$$\mathbf{k}_h^{(l)}[V_s] = \mu_{\text{k},\psi(V_s)}(\mathbf{g}^{(l)}[V_s]; \mathbf{W}_{h,\psi(V_s)}^{\text{k}}), \quad (21)$$

where $\|_h \{\cdot\}$ represents the concatenation operation.

The information from the neighboring nodes is extracted through function $f_{\text{msg}}(\cdot)$. Similar to $f_{\text{att}}(\cdot)$, for the h-th attention head, the hidden state $\mathbf{g}^{(l)}[V_s]$ of a neighboring node $V_s \in \mathcal{N}(V_t)$ is mapped into a value vector $\mathbf{v}_h^{(l)}[V_s]$ through a linear projector $\mu_{\text{v},\psi(V_s)}(\cdot; \mathbf{W}_{h,\psi(V_s)}^{\text{v}})$. The multi-head message is calculated by:

$$f_{\text{msg}}(\mathbf{g}^{(l)}[V_s]) = \big\|_h \big\{\mathbf{v}_h^{(l)}[V_s] \mathbf{W}_{\psi(V_s)}^{\text{msg}}\big\}$$

$$= \big\|_h \big\{\mu_{\text{v},\psi(V_s)}(\mathbf{g}^{(l)}[V_s]; \mathbf{W}_{h,\psi(V_s)}^{\text{v}}) \mathbf{W}_{\psi(V_s)}^{\text{msg}}\big\}, \quad (22)$$

where $\mathbf{W}_{\psi(V_s)}^{\text{msg}}$ is the weight matrix. Finally, the calculated multi-head attention scores and messages from all the neighboring nodes are aggregated followed by an activation operation. We use the rectified linear unit (ReLU) as the activation function. The hidden state of node V_t in layer $(l + 1)$ is then updated as follows:

$$\mathbf{g}^{(l+1)}[V_t] = \mathbf{W}_{\psi(V_t)}^{\text{agg}} \text{ReLU}\Big(\text{sum}\big(f_{\text{att}}(V_t, \mathbf{g}^{(l)}[V_t])$$

$$\cdot \underset{V_s \in \mathcal{N}(V_t)}{f_{\text{msg}}}(\mathbf{g}^{(l)}[V_s])\big)\Big) + \mathbf{g}^{(l)}[V_t], \quad (23)$$

where $\mathbf{W}_{\psi(V_t)}^{\text{agg}}$ denotes the weight matrix to update the hidden state of the node V_t. We show the heterogeneous message passing
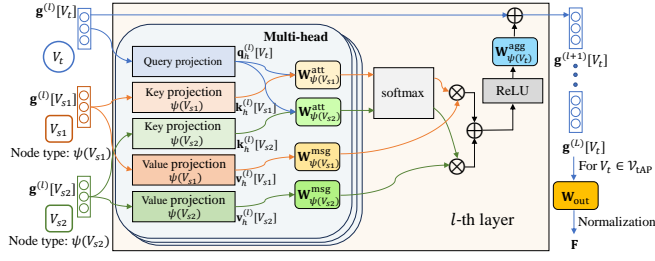
Figure 3: Transformer-based heterogeneous message passing. Given node V_t and its hidden state $\mathbf{g}^{(l)}[V_t]$, as well as the neighboring nodes $V_{s1}, V_{s2} \in \mathcal{N}(V_t)$ and their hidden states $\mathbf{g}^{(l)}[V_{S1}]$ and $\mathbf{g}^{(l)}[V_{S2}]$, the hidden state $\mathbf{g}^{(l+1)}[V_t]$ for node V_t is updated through a transformer-based heterogeneous message passing scheme.

scheme in Fig. 3. By stacking all the L hidden layers, the constructed SACGNN can generate a representation $\mathbf{g}^{(L)}[V_t]$ for each node $V_t \in \mathcal{V}$. The hidden states of the final layer for the tAP nodes, i.e., $\mathbf{g}^{(L)}[V_t]$ for $V_t \in \mathcal{V}_{tAP}$, are then used for the transmit beamforming design. We apply a linear transformation of weight $\mathbf{W}_{out}$ to the hidden states $\mathbf{g}^{(L)}[V_t]$ with $V_t \in \mathcal{V}_{tAP}$. The outputs are normalized to satisfy the power constraint (14c), and we obtain the designed beamforming matrix for all the transmit APs, i.e., $\mathbf{F} \in \mathbb{C}^{N_T M \times (K+1)}$, where the beamforming matrix for the i-th transmit AP is given by $\mathbf{F}_i = \mathbf{F}((i-1)M+1 : iM) \in \mathbb{C}^{M \times (K+1)}$, $i = 1, \ldots, N_T$.

3.3 Loss Function

We construct a training data sample as $(\mathbf{H}, \widetilde{\mathbf{A}})$, where $\mathbf{H} = \{\mathbf{h}_{i,k}\}_{i,k}^{N_T, K}$ and $\widetilde{\mathbf{A}} = \{\widetilde{\mathbf{A}}_{i,j}\}_{i,j}^{N_T, N_R}$. The training dataset is given by $\mathcal{D} = \{(\mathbf{H}^{(1)}, \widetilde{\mathbf{A}}^{(1)}), \ldots, (\mathbf{H}^{(N_{tr})}, \widetilde{\mathbf{A}}^{(N_{tr})})\}$, where N_{tr} is the total number of training samples. The developed SACGNN is trained in an unsupervised manner. In particular, we reformulate the objective function in (14) and define the loss function as follows:

$$\mathcal{L} = -R + \rho \text{ReLU}(\gamma_{min} - \gamma^{(s)}), \tag{24}$$

where the sensing SNR requirement (14b) is moved to the objective with a penalizing coefficient ρ. The developed SACGNN is trained to minimize the loss function (24) using Adam optimizer [7] in an unsupervised manner. After training, the trained model can be applied for online beamforming design. When given the communication channel vectors[2] and AoDs/AoAs of the target to be detected, the network can generate the beamforming vectors for the transmit APs to achieve a high sum-rate while guaranteeing the sensing SNR requirement.

4 PERFORMANCE EVALUATION

In this section, we evaluate the performance of the developed SACGNN framework for cooperative ISAC beamforming. The simulation settings are as follows. We consider there are $N_T = 2$ transmit

---

[2]We note that channel estimation is an important topic and there exist various estimation methods that can be applied to obtain the channel vectors. For simplicity, we assume perfect channel state information is available at the CPU which is widely adopted in the existing works (e.g., [1, 2, 9]).
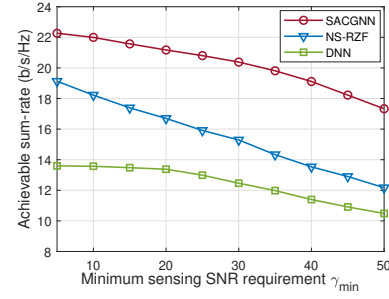

Figure 4: The achievable sum-rate versus the minimum sensing SNR requirement γ_{min}, where $P_{max} = 30$ dBm.

APs and $N_R = 2$ receive APs to cover a 200×200 m^2 area. Considering a three-dimensional (3D) [x, y, z] coordinate, the transmit APs are located at $[0, 100, 20]$ and $[200, 100, 20]$. The receive APs are placed at $[100, 0, 20]$ and $[100, 200, 20]$. Each AP is equipped with a UPA of $M_v = 4$ vertical antennas and $M_h = 16$ horizontal antennas. We consider there are $K = 4$ users. The users and the target are assumed to be randomly distributed in the area, with x and y coordinates ranging from 0 to 200, and the z coordinate ranging from 0 to 35. We adopt the normalized system parameters [2]. In particular, the variance of communication channel gain $\zeta_{i,k}^2$ is set to 0.5, $i = 1, \ldots N_T, k = 1, \ldots, K$. The variance of communication noise σ_k^2 at the k-th user, $k = 1, \ldots, K$, and the variance of sensing noise ξ_j^2 at the j-th receive AP, $j = 1, \ldots, N_R$, are set to 1. The variance of the sensing channel gain $\chi_{i,j}^2$ is set to 0.1, $i = 1, \ldots, N_T, j = 1, \ldots, N_R$. We set the number of hidden layers L to two. The learning rate is set to 10^{-4}. We generate $10,000$ data samples, where $8,000$ of them are used for offline training and $2,000$ are used for testing. We compare the performance of our proposed SACGNN framework with two other baseline methods: (i) null-space projection based beamformer for sensing and regularized zero-forcing beamformer for communication (NS-RZF) proposed in [2]; (ii) a DNN-based scheme which consists of three 3D convolutional neural network layers followed by a fully-connected layer.

In Fig. 4, we show the achievable sum-rate under different sensing SNR requirements. The maximum transmit power P_{max} is set to 30 dBm. We can observe that as the sensing requirement becomes more stringent, the achievable sum-rate decreases since more power is allocated to the sensing beamformer. The results demonstrate that the proposed SACGNN framework outperforms the NS-RZF scheme proposed in [2]. The NS-RZF scheme designs the sensing beamformer to be in the null space of the communication channels and manages inter-user interference through regularized zero-forcing. This scheme may not perform well since the communication channels of users and the target are usually correlated with each other, leading to reduced null space of the communication channels. Moreover, the proposed SACGNN framework also outperforms the DNN-based baseline approach. This is due to the fact that the beamforming design needs to simultaneously address both communication and sensing features, leading to additional complexity in the design process. Typical 3D convolutional neural networks may not be sufficient to capture the intricate interplay between communication and sensing information. The
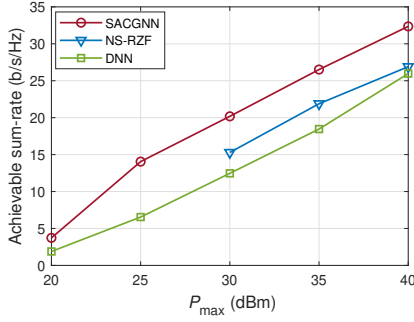
Figure 5: The achievable sum-rate versus the maximum transmit power $P_{\max}$, where $\gamma_{\min} = 30$.

(a) A 3D view of the beam pattern.   (b) The azimuth cut of the beam pattern.

Figure 6: Beam patterns captured at a transmit AP located at $[0, 100, 20]$.

results in Fig. 4 demonstrate the performance gain of the proposed SACGNN framework.

In Fig. 5, we show the achievable sum-rate versus the maximum transmit power, under a fixed sensing SNR requirement, i.e., $\gamma_{\min} = 30$. Similarly, we observe that the proposed approach consistently provides a higher sum-rate compared to the other two baseline methods. In particular, in the low transmit power regime ($P_{\max} = 20, 25$ dBm), the NS-RZF scheme fails to satisfy the sensing SNR requirement. This is because the sensing and communication channels exhibit high correlations, resulting in a limited null space dimension of the communication channels and consequently a marginal sensing beamforming gain.

Finally, the beam patterns generated by one of the transmit APs using our proposed SACGNN framework are shown in Fig. 6, where a 3D view and the azimuth cut are presented. The results are generated based on the following settings. The four users are located at $[40, 40, 30]$, $[140, 40, 20]$, $[140, 140, 20]$, and $[40, 140, 30]$, respectively. The target is located at $[115, 115, 25]$. The view is captured from a transmit AP located at $[0, 100, 20]$. It can be observed from Fig. 6 that the generated beamformer can effectively manage the side lobes and interference, even when the target to be detected is close to a communication user.

5 CONCLUSION

In this paper, we investigated the cooperative ISAC beamforming design in cell-free MIMO systems. We proposed a SACGNN framework which models the cell-free MIMO system for cooperative ISAC as a heterogeneous graph and uses a transformer-based scheme for heterogeneous message passing. We conducted simulations for performance evaluation and included two baseline methods for comparison. Simulation results showed that the proposed SACGNN framework outperforms a conventional null-space projection based scheme and a DNN-based baseline scheme. Moreover, results demonstrated that the proposed SACGNN framework can generate narrow beamformers toward users and the target, where the side lobes can be effectively managed.
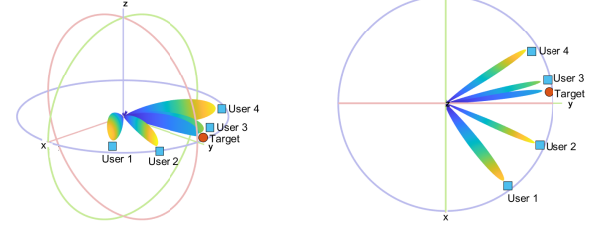
REFERENCES

[1] Zinat Behdad, Özlem Tugfe Demir, Ki Won Sung, Emil Björnson, and Cicek Cavdar. 2024. Multi-static target detection and power allocation for integrated sensing and communication in cell-free massive MIMO. *IEEE Trans. Wireless Commun.* 23, 9 (Sept. 2024), 11580–11596.

[2] Umut Demirhan and Ahmed Alkhateeb. 2023. Cell-free joint sensing and communication MIMO: A max-min fair beamforming approach. In *Proc. of Asilomar Conf. on Signals, Systems, and Computers*. Pacific Grove, CA.

[3] Changfeng Ding, Cheng Zeng, Chuanwen Chang, Jun-Bo Wang, and Min Lin. 2022. Joint precoding for MIMO radar and URLLC in ISAC systems. In *Proc. of ACM MobiCom Workshop on Integrated Sensing and Communications Systems (ISACom)*. Sydney, Australia.

[4] Nuria González-Prelcic, Musa Furkan Keskin, Ossi Kaltiokallio, Mikko Valkama, Davide Dardari, Xiao Shen, Yuan Shen, Murat Bayraktar, and Henk Wymeersch. to appear, 2024. The integrated sensing and communication revolution for 6G: Vision, techniques, and applications. *Proc. IEEE* (to appear, 2024).

[5] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous graph transformer. In *Proc. of The Web Conference (WWW)*. Taipei, Taiwan.

[6] Yi Huang, Yuan Fang, Xinmin Li, and Jie Xu. 2022. Coordinated power control for network integrated sensing and communication. *IEEE Trans. Veh. Tech.* 71, 12 (Dec. 2022), 13361–13365.

[7] Diederik P. Kingma and Jimmy Lei Ba. 2015. Adam: A method for stochastic optimization. In *Proc. Int. Conf. Learn. Representations (ICLR)*. San Diego, CA.

[8] Fan Liu, Yuanhao Cui, Christos Masouros, Jie Xu, Tony Xiao Han, Yonina C. Eldar, and Stefano Buzzi. 2022. Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. *IEEE J. Sel. Areas Commun.* 40, 6 (Jun. 2022), 1728–1767.

[9] Sifan Liu, Rang Liu, Zhiping Lu, Ming Li, and Qian Liu. 2024. Cooperative cell-free ISAC networks: Joint BS mode selection and beamforming design. In *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*. Dubai, United Arab Emirates.

[10] Yifei Shen, Jun Zhang, S. H. Song, and Khaled B. Letaief. 2023. Graph neural networks for wireless communications: From theory to practice. *IEEE Trans. Wireless Commun.* 22, 5 (May 2023), 3554–3569.

[11] Harsh Tataria, Mansoor Shafi, Andreas F. Molisch, Mischa Dohler, Henrik Sjöland, and Fredrik Tufvesson. 2021. 6G wireless systems: Vision, requirements, challenges, insights, and opportunities. *Proc. IEEE* 109, 7 (Jul. 2021), 1166–1199.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proc. of Conf. on Neural Inf. Process. Syst. (NIPS)*. Long Beach, CA.

[13] Xiao Wang, Deyu Bo, Chuan Shi, Shaohua Fan, Yanfang Ye, and S Yu Philip. 2022. A survey on heterogeneous graph embedding: Methods, techniques, applications and sources. *IEEE Trans. Big Data* 9, 2 (Apr. 2022), 415–436.

[14] Xiaojie Wang, Qi Guo, Zhaolong Ning, Lei Guo, Guoyin Wang, Xinbo Gao, and Yan Zhang. 2024. Integration of sensing, communication, and computing for metaverse: A survey. *ACM Comput. Surv.* 56, 10 (May 2024), 1–38.

[15] Zihuan Wang and Vincent W.S. Wong. 2023. Deep learning for ISAC-enabled end-to-end predictive beamforming in vehicular networks. In *Proc. of IEEE Int. Conf. Commun. (ICC)*. Rome, Italy.

[16] Zihuan Wang, Vincent W.S. Wong, and Robert Schober. to appear, 2024. Integrated sensing and communications for end-to-end predictive beamforming design in vehicle-to-infrastructure networks. *IEEE J. Sel. Topics Signal Process.* (to appear, 2024).

[17] Tongyang Xu, Fan Liu, Christos Masouros, and Izzat Darwazeh. 2022. Proof of concept experiments of joint waveform design for integrated sensing and communications. In *Proc. of ACM MobiCom Workshop on Integrated Sensing and Communications Systems (ISACom)*. Sydney, Australia.