Channel-Agnostic Predictive Beamforming for Crowdsourced Bistatic Satellite ISAC With LLM

William D. Lukito, *Graduate Student Member, IEEE*, Wei Xiang, *Senior Member, IEEE*, Chang Liu, *Member, IEEE*, Phu Lai, Peng Cheng, *Member, IEEE*, Weijie Yuan, *Senior Member, IEEE*, and Guoqiang Mao, *Fellow, IEEE*

***Abstract*—Integrated sensing and communications (ISAC) systems promise dual use of spectrum and hardware for data transmission and environmental awareness. However, extending ISAC to satellite networks is challenged by high path loss, long delays, and the overhead of channel estimation. To address these challenges, we propose a channel-agnostic predictive beamforming framework for satellite ISAC (S-ISAC) within a crowdsourced bistatic architecture. Unlike conventional bistatic architectures that require a dedicated sensing receiver, our design aggregates echoes from multiple ground internet of things (IoT) devices (GIDs) in a crowdsourced manner to improve sensing performance without introducing any additional sensing equipment. We propose a model termed Historical Geometric-based LLM (HG-LLM) as a realization of the channel-agnostic predictive beamforming framework. HG-LLM learns to map historical geometric information (HGI) of the satellite, sensing target, and GIDs directly to future beamforming matrices, eliminating the need for channel state information (CSI). We propose two key modules in HG-LLM, namely, the Histogeometric Encoder, which transforms spatial-temporal data into LLM-compatible embeddings, and the TokenBeamformer, which translates the LLM outputs into optimized beamforming weights. Moreover, the backbone LLM is fine-tuned using low-rank adaptation for efficient adaptation to predictive beamforming tasks. Extensive simulations demonstrate that HG-LLM achieves performance levels comparable to channel-based methods across diverse settings, despite relying solely on HGI without requiring explicit CSI.**

***Index Terms*—Integrated sensing and communications, beamforming, satellite, bistatic sensing, large language models, channel-agnostic.**

Received 15 May 2025; revised 13 October 2025; accepted 29 November 2025. Date of publication 8 December 2025; date of current version 20 February 2026. This work was supported in part by the SmartSat CRC funded by Australian Government's CRC Program and in part by Australian Government through Australian Research Council's Discovery Projects Funding Scheme under Project DP220101634. *(Corresponding author: Wei Xiang.)*

William D. Lukito, Wei Xiang, Chang Liu, and Phu Lai are with the School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia (e-mail: w.lukito@latrobe.edu.au; w.xiang@latrobe.edu.au; c.liu6@latrobe.edu.au; p.lai@latrobe.edu.au).

Peng Cheng is with the School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia, and also with The University of Sydney, Sydney, NSW 2006, Australia (e-mail: p.cheng@latrobe.edu.au).

Weijie Yuan is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China (e-mail: yuanwj@sustech.edu.cn).

Guoqiang Mao is with the Research Laboratory of Smart Driving and Intelligent Transportation Systems, Southeast University, Nanjing, Jiangsu 211102, China (e-mail: g.mao@ieee.org).

Digital Object Identifier 10.1109/JSAC.2025.3641581

I. Introduction

INTEGRATED sensing and communications (ISAC) is emerging as a key technology for next-generation wireless networks, i.e., sixth-generation (6G) systems [1], [2], [3]. By enabling a single device and resource to accommodate both communications and sensing, ISAC enhances spectrum efficiency and reduces hardware costs, making it a promising solution for next-generation networks [2], [4]. While most existing research on ISAC focuses on terrestrial networks [5], [6], [7], [8], [9], [10], [11], [12], their limited coverage constrains its potential for wide-area services. In contrast, satellites, which are expected to play a central role in 6G networks [3], [13], [14], offer seamless global coverage and support diverse applications [15], [16], [17]. Despite this, satellite ISAC (S-ISAC) networks remain underexplored, even though they hold significant potential for wide-area sensing and communications services such as satellite internet of things (IoT), disaster monitoring, and remote environmental sensing. In particular, S-ISAC distinguishes itself from terrestrial ISAC by extending sensing and communications capabilities to scenarios that are inaccessible to ground infrastructure.

As a consequence of their vast coverage, satellite links experience significant path loss and propagation delays due to the large transmission distances [14], [15], [16]. For S-ISAC systems, these challenges become more severe when conventional ISAC architectures are adopted,[1] as they rely on the return of the transmitted signal as an echo for sensing [5], [6], [7]. One potential solution is to adopt bistatic ISAC, where the transmitter and receiver are not located at the same physical location [8], [9], [10], [11]. In this configuration, the transmitter is mounted on the satellite, while the receiver is placed on the ground [18], [19], [20]. This setup allows the sensing process to rely on the echo reaching the ground, rather than requiring a reflected echo to return to the satellite, which helps reduce the double path loss and delay specifically associated with sensing. However, existing bistatic S-ISAC studies typically consider a dedicated sensing receiver [19], [20]. Since relying on a dedicated sensing receiver limits scalability and sensing accuracy, we propose to extend the architecture by leveraging multiple ground IoT devices (GIDs),

[1]Conventional ISAC architectures follow a so-called monostatic configuration, where the transmitter and receiver are integrated into the same device, known as a transceiver [9], [11].

which already exist as communications receivers, to also serve as sensing receivers. In this way, echoes are collectively gathered from distributed GIDs in a crowdsourced manner, leading to improved sensing performance without the need for additional sensing equipment. Furthermore, research on bistatic ISAC is still at an early stage, and many important aspects, particularly those related to S-ISAC systems, remain insufficiently explored [8], [9], [10], [11], [19], [20].

For instance, an essential yet underexplored aspect of achieving an efficient and secure bistatic S-ISAC system is the design of an effective transmit beamforming strategy [7], [21], [22], [23]. Such a strategy is vital for directing beamforming energy toward both designated communications users and sensing targets. However, this challenge has not been sufficiently addressed in the context of bistatic S-ISAC. Moreover, most ISAC network design tasks, including transmit beamforming, rely heavily on the availability of channel state information (CSI) [21], which in satellite communications is prone to significant delay and estimation errors due to the large propagation distance [14], [15], [16]. In response, three main alternatives have been explored in the literature: (i) using statistical CSI [19], [22], [23], (ii) predicting CSI and using the predicted values for network design [24], [25], and (iii) employing predictive design strategies that eliminate the need for explicit CSI estimation [5], [6], [26]. Nevertheless, these approaches still involve a large number of input parameters, which scale with the number of antennas and users, making them impractical for large-antenna systems with many users, such as S-ISAC systems. Drawing inspiration from recent advances in large language models (LLMs) [27] and their powerful feature extraction and generation abilities [28], [29], and noting that autoregressive decoders can synthesize high-quality CSI [30], we reduce dependence on CSI in S-ISAC by training an LLM that takes historical geometric information (HGI) as input and directly outputs future beamforming matrices, avoiding channel prediction and feedback.

While LLMs are primarily known for their success in language processing, they are now emerging as a transformative tool in wireless communications and ISAC networks as well [30], [31]. Their potential has been demonstrated in applications such as network optimization [32], [33], [34], [35], [36], [37], beam prediction [38], traffic flow forecasting [39], and channel prediction/feedback [40], [41]. To the best of our knowledge, existing LLM-based optimization studies in wireless communications have so far relied on text-based prompting [42] or CSI inputs [40], both of which are less practical since wireless network optimization is naturally expressed in numerical form and CSI acquisition incurs significant overhead. Furthermore, current text-based methods primarily rely on closed [33], [34], [36] and open vocabularies [32], [35], further constraining their flexibility and scalability in optimization tasks. Additionally, some works lack a clear framework for effectively integrating LLMs into wireless systems [43]. Motivated by these gaps, we propose an LLM-based design framework for S-ISAC that removes the reliance on text-based inputs by using numerical inputs, similar to traditional optimization techniques, which in our case is HGI.

Motivated by the aforementioned gaps, this work proposes an LLM-based predictive transmit beamforming framework for bistatic S-ISAC systems. Unlike existing channel-based methods [6], [19], [22], [23], [24], [25], [26], the proposed framework leverages HGI of the S-ISAC network as input to generate the transmit beamforming matrix for future time slots. By removing any reliance on statistical, predicted, or instantaneous CSI, our approach avoids channel estimation overhead and reduces input dimensionality by depending solely on HGI. This strategic shift not only streamlines the beamforming process but also enhances scalability for large-scale S-ISAC networks. The main contributions of this paper are summarized as follows:

1) We propose a novel crowdsourced sensing system for bistatic S-ISAC networks. Unlike conventional bistatic S-ISAC topologies that employ a single dedicated sensing receiver alongside separate communications terminals [19], [20], our system utilizes existing communications terminals[2] to collaboratively aggregate sensing information. This distributed strategy improves spatial coverage and overall sensing performance, eliminating the need for dedicated sensing equipment.
2) We propose a channel-agnostic predictive beamforming framework for efficient S-ISAC operations. In contrast to conventional methods that rely on CSI, our approach leverages HGI to design the transmit beamforming matrix for future time slots.
3) To realize the predictive beamforming framework, we propose a model termed Historical Geometric-based LLM (HG-LLM). Unlike conventional LLM-based methods that rely on text-based inputs, the proposed HG-LLM directly processes HGI. This is achieved through two proposed key modules, namely, the Histogeometric Encoder, which transforms raw geometric data into LLM-compatible embeddings, and the TokenBeamformer, which maps the LLM outputs into optimized beamforming matrices.
4) We validate the proposed HG-LLM through extensive simulations across various scenarios, including different communications weights, power budgets, noise levels, and prediction horizons. The results show that HG-LLM achieves performance comparable to channel-based methods while consistently satisfying power constraints, showcasing its robustness and efficiency even without relying on channel information.

The remainder of this paper is organized as follows. Section II introduces the system model for the considered bistatic S-ISAC. In Section III, we propose a historical geometry-based predictive beamforming framework along with its problem formulation. Section IV presents HG-LLM as the realization of the proposed framework. Finally, the conclusion of this paper is presented in Section VI.

Notation: Unless otherwise specified, bold uppercase letters, bold lowercase letters, and normal letters/symbols represent matrices, vectors, and scalars, respectively. Constants c and j denote the speed of light ($c \approx 3 \times 10^8$ meters per second)

[2]We will refer to these as the so-called S-ISAC ground terminals.

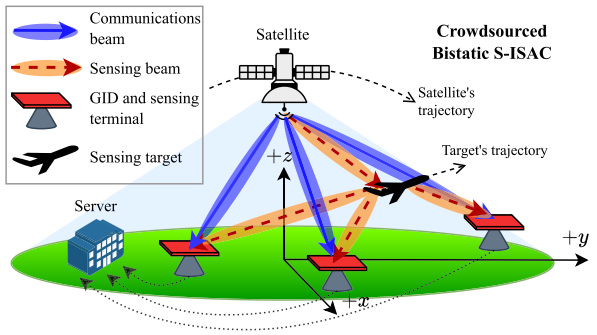


Fig. 1. The considered crowdsourced sensing system operates over a bistatic S-ISAC network, in which the satellite functions as a ISAC transmitter, while the ground devices serve as ISAC receivers. The sensing information collected by all GIDs is aggregated at a centralized server. A detailed illustration of the antenna geometry is provided in Fig. 2.

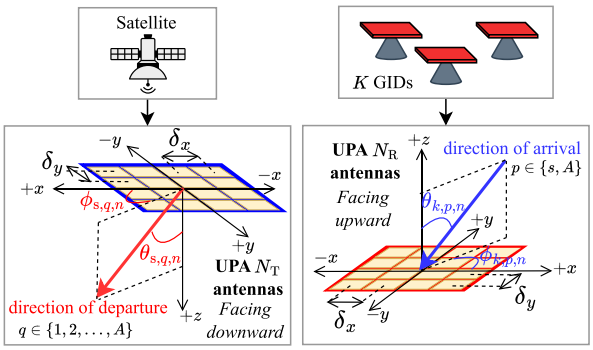


Fig. 2. An illustration of the antenna geometry, where the satellite and GIDs are equipped with UPA antennas aligned with the $x-y$ plane. The satellite antenna faces downward, and the GID antennas face upward.

and the imaginary unit, respectively. Notations $\mathbb{R}^{a\times b}$ and $\mathbb{C}^{a\times b}$ represent real and complex number sets of size a-by-b, respectively. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the circularly symmetric complex Gaussian (CSCG) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathcal{U}(a, b)$ denotes the uniform distribution between a and b. $\mathbf{I}_N$ denotes N-by-N identity matrix. Operators $\mathrm{diag}(\cdot)$ and $\mathrm{vec}(\cdot)$ form a diagonal matrix from a vector and denote vectorization of a matrix, respectively. Superscripts T, *, and H denote transpose, conjugate, and Hermitian matrices/vectors operations, respectively. Operators $\times$ and $\otimes$ indicate scalar multiplication and the Kronecker product, respectively. The Euclidean distance, Frobenius norm, and absolute value are represented by $\|(\cdot)\|$, $\|(\cdot)\|_F^2$, and $|(\cdot)|$, respectively.

II. System Model

We consider a bistatic S-ISAC network [19] within a satellite-based IoT framework [26], [44], [45]. In this setting, the satellite serves as a dual-functioning radar and communications (ISAC) transmitter, while the ground IoT devices (GIDs) act as S-ISAC ground terminals. Both the satellite and the receivers are equipped with uniform planar array (UPA) antennas, comprising $N_{\mathrm{T}} = N_{\mathrm{T}}^x \times N_{\mathrm{T}}^y$ and $N_{\mathrm{R}} = N_{\mathrm{R}}^x \times N_{\mathrm{R}}^y$ elements, respectively.[3] The gain of each transmit and receive antenna element is denoted by G_{T} and G_{R}, respectively. Without loss of generality, we assume the same inter-element spacing for both transmit and receive arrays, with uniform separation along the x- and y-axes given by $\delta_x = 0.5\lambda_{\mathrm{c}}$ and $\delta_y = 0.5\lambda_{\mathrm{c}}$, where $\lambda_{\mathrm{c}} = c/f_{\mathrm{c}}$ is the carrier wavelength corresponding to the carrier frequency f_{c}. The overall system configuration is illustrated in Fig. 1.

A. Temporal & Spatial Models

At the n-th time slot, $n \in \{0, 1, \ldots, N-1\}$, the satellite's position is given by the vector $\mathbf{p}_{\mathrm{S},n} = [x_{\mathrm{S},n}, y_{\mathrm{S},n}, z_{\mathrm{S},n}]$. The satellite is connected to K ground IoT devices (GIDs), each deployed at a fixed location. The position of the k-th GID is denoted by $\mathbf{p}_k = [x_k, y_k, z_k]$, where $k \in \mathcal{K} = \{1, \ldots, K\}$. Additionally, we consider an aerial sensing target, e.g. uncrewed aerial vehicle (UAV), whose trajectory at time slot n is represented by the position vector $\mathbf{p}_{\mathrm{A},n} = [x_{\mathrm{A},n}, y_{\mathrm{A},n}, z_{\mathrm{A},n}]$.

Fig. 2 shows an illustration of the transmit and receive antennas geometry. The transmit antenna forms directions of departure (DoDs) toward the K GIDs and the sensing target. For notational simplicity, we define the unified set $\mathcal{Q} = \mathcal{K} \cup \{\mathrm{A}\}$, where 'A' represents the sensing (aerial) target and $\mathcal{P} = \{\mathrm{S}, \mathrm{A}\}$, where 'S' denotes the satellite. Therefore, the elevation and azimuth DoD angles toward each $q \in \mathcal{Q}$ are denoted by $(\theta_{\mathrm{S},q,n}, \phi_{\mathrm{S},q,n})$. On the other hand, each receive antenna observes two directions of arrival (DoAs): one from the satellite and one from the sensing target. Additionally, the corresponding elevation and azimuth DoA angles at the k-th GID from each $p \in \mathcal{P}$ are denoted by $(\theta_{k,p,n}, \phi_{k,p,n})$. In this work, we assume that both transmit and receive antennas are aligned with x-y plane.[4]

B. Channel Model

To simultaneously support both communications and sensing functionalities, it is necessary to consider the corresponding channels, which are the channel state information (CSI) for communications and the target response matrix (TRM)[5] for sensing.

Remark 1: It is important to note that the channel model described in this section is utilized solely for simulation purposes and offline training. During online inference, the proposed HG-LLM framework operates in a channel-agnostic manner, relying exclusively on HGI to generate the beamforming matrix, without requiring instantaneous or historical CSI and TRM.

1) Communications Model: At the n-th time slot, let the communications channel between the satellite and the k-th GID be represented by the matrix $\mathbf{H}_{k,n} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$. Following the approach in existing studies, the satellite channel is modeled using Rician fading [22], [23], [26], [45]. Thus,

[3]The assumed UPA-equipped GIDs correspond to advanced satellite IoT terminals, where compact antenna arrays are increasingly considered necessary [46].

[4]Since the transmit and receive UPAs are parallel and share a common reference along the z-axis, with the transmit UPA facing downward and the receive UPA facing upward, the DoA is equal to the DoD due to geometric symmetry. Hence, $(\theta_{\mathrm{S},k,n}, \phi_{\mathrm{S},k,n}) = (\theta_{k,\mathrm{S},n}, \phi_{k,\mathrm{S},n})$.

[5]TRM essentially serves as the channel state representation for sensing, analogous to how CSI is utilized for communications.

the channel matrix between the satellite and the k-th can be expanded as follows[6]

$$\mathbf{H}_{k,n} = \mathbf{H}_{k,n}^{\mathrm{LOS}} + \mathbf{H}_{k,n}^{\mathrm{NLOS}}, \tag{1}$$

where $\mathbf{H}_{k,n}^{\mathrm{LOS}}$ and $\mathbf{H}_{k,n}^{\mathrm{NLOS}}$ represent the line-of-sight (LOS) and non-LOS (NLOS) components, respectively. Those can be further expanded as follows [22], [26]

$$\mathbf{H}_{k,n}^{\mathrm{LOS}} = \sqrt{\frac{\kappa_{\mathrm{R}}\varrho_{k,n}}{\kappa_{\mathrm{R}}+1}}\mathbf{b}^{H}\left(\theta_{k,\mathrm{S},n},\phi_{k,\mathrm{S},n}\right)\mathbf{a}\left(\theta_{\mathrm{S},k,n},\phi_{\mathrm{S},k,n}\right), \tag{2}$$

$$\mathbf{H}_{k,n}^{\mathrm{NLOS}} \sim \mathcal{CN}\left(0, \frac{\varrho_{k,n}}{\kappa_{\mathrm{R}}+1}\mathbf{I}_{N_{\mathrm{R}}} \otimes \mathbf{I}_{N_{\mathrm{T}}}\right), \tag{3}$$

where κ_{R} denotes the Rician factor, $\varrho_{k,n} = G_{\mathrm{total}}\left(\frac{c}{4\pi f_{\mathrm{c}} d_{\mathrm{S},k,n}}\right)^2$ represents the large-scale channel gain, $G_{\mathrm{total}} = G_{\mathrm{T}}G_{\mathrm{R}}N_{\mathrm{T}}N_{\mathrm{R}}$ is the combined transmit and receive antenna gain, and $d_{\mathrm{S},k,n} = \|\mathbf{p}_{\mathrm{S},n} - \mathbf{p}_{k,n}\|$ denotes the distance between the satellite and the k-th GID. Additionally, the functions $\mathbf{a}(\cdot)$ and $\mathbf{b}(\cdot)$ represent the transmit and receive array response vectors, respectively. Please refer to Appendix A for its definition.

Remark 2: The large values of $d_{\mathrm{S},k,n}$ in satellite networks significantly reduce $\varrho_{k,n}$, which in turn has a substantial impact on both (2) and (3). Therefore, designing an optimal transmit beamforming matrices is essential to mitigate this degradation and enhance the overall communications performance [5], [6], [12].

2) Bistatic Sensing Model: The TRM at the k-th GID at the n-th time slot can be defined as follows [11], [19][7]

$$\mathbf{G}_{k,n} = \xi_{k,n}\mathbf{b}^{H}\left(\theta_{k,\mathrm{A},n},\phi_{k,\mathrm{A},n}\right)\mathbf{a}\left(\theta_{\mathrm{S},\mathrm{A},n},\phi_{\mathrm{S},\mathrm{A},n}\right), \tag{4}$$

where $\xi_{k,n}$ denotes the reflection coefficient of the target toward the k-th GID at time slot n, and can be expanded as follows

$$|\xi_{k,n}|^2 = G_{\mathrm{total}}\frac{c^2\sigma_{\mathrm{A}}\cos(\beta_{k,n}/2)}{64\pi^3 d_{\mathrm{S},\mathrm{A},n}^2 d_{\mathrm{A},k,n}^2 f_{\mathrm{c}}^2}. \tag{5}$$

Here, $\sigma_{\mathrm{A}}\cos(\beta_{k,n}/2)$ represents the bistatic radar cross section (RCS) of the target, where σ_{A} denotes its corresponding monostatic RCS [19], [49] and $\beta_{k,n}$ denotes the bistatic angle corresponding to the k-th GID.[8] Additionally, $d_{\mathrm{S},\mathrm{A},n}$ and $d_{\mathrm{A},k,n}$ denote the distances from the satellite to the target and from the target to the k-th GID, respectively.

Remark 3: According to (5), the reflection coefficient in a monostatic configuration is generally smaller, as the condition $d_{\mathrm{S},\mathrm{A},n} \triangleq d_{\mathrm{A},k,n}$ leads to greater overall propagation loss compared to the bistatic case where $d_{\mathrm{S},\mathrm{A},n} \gg d_{\mathrm{A},k,n}$. Additionally, despite the reduced path loss in bistatic configurations, the large absolute distances involved still result in significant attenuation, thereby necessitating transmit beamforming to enhance sensing performance.

C. ISAC Signal Model

Inspired by [18] and [19], we allocate $K+1$ data streams by exploiting the available spatial degrees of freedom, with K assigned to communications and one dedicated to sensing. Each stream corresponds to a physical transmit beam, with K beams serving the GIDs and one beam steered for sensing. At the n-th time slot, the symbol vector is defined as $\mathbf{s}_n = [s_{1,n},\ldots,s_{K,n},s_{\mathrm{A},n}]^T \in \mathbb{C}^{(K+1)\times 1}$, where the first K elements correspond to communications with the GIDs, and the last element is dedicated to sensing. The transmitted ISAC signal from the satellite is represented as follows

$$\mathbf{x}_n = \mathbf{F}_n\mathbf{s}_n \in \mathbb{C}^{N_{\mathrm{T}}\times 1}, \tag{6}$$

where $\mathbf{F}_n = [\mathbf{f}_{1,n},\ldots,\mathbf{f}_{K,n},\mathbf{f}_{\mathrm{A},n}] \in \mathbb{C}^{N_{\mathrm{T}}\times(K+1)}$ is the beamforming matrix to be optimized, and $\mathbf{f}_{u,n} \in \mathbb{C}^{N_{\mathrm{T}}\times 1}$ denotes the beamforming vector for each user $u \in \mathcal{U}$. The transmitted signal $\mathbf{x}_n$ propagates through the communications and sensing channels, as defined in (1) and (4), respectively. Thus, at the k-th GID, the received signal $\mathbf{y}_{k,n} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$ is expressed as the superposition of the communications signal, sensing echoes, and additive noise as follows

$$\mathbf{y}_{k,n} = \mathbf{H}_{k,n}\mathbf{x}_n + \mathbf{G}_{k,n}\mathbf{x}_n + \mathbf{n}_{k,n}, \tag{7}$$

where $\mathbf{n}_{k,n} \sim \mathcal{CN}(0,\sigma_k^2\mathbf{I}_{N_{\mathrm{R}}})$ represents the additive CSCG noise for the k-th GID.

To eliminate interference from undesired DoAs, we leverage advanced spatial filtering techniques, e.g., minimum mean-squared error (MMSE) [50]. Let $\mathbf{w}_{k,n} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$ be the receive spatial filter applied to (7), designed according to the row space of $\mathbf{H}_{k,n}$. Since steering vectors at different angles are asymptotically orthogonal [51], the subspaces of $\mathbf{H}_{k,n}$ and $\mathbf{G}_{k,n}$ are nearly orthogonal. Consequently, $\mathbf{w}_{k,n}$ satisfies $\|\mathbf{w}_{k,n}^H\mathbf{G}_{k,n}\| \approx 0$, effectively eliminating sensing interference. Therefore, the communications SINR of the k-th GID can be expressed as follows

$$\gamma_{k,n} = \frac{|\mathbf{w}_{k,n}^H\mathbf{H}_{k,n}\mathbf{f}_{k,n}|^2}{\sum_{\substack{u\in\mathcal{Q}\\u\neq k}}|\mathbf{w}_{k,n}^H\mathbf{H}_{k,n}\mathbf{f}_{u,n}|^2 + \sigma_k^2\|\mathbf{w}_{k,n}\|^2}, \tag{8}$$

where the sum term of denominator iterates for all $\mathcal{Q}$ except k. Consequently, the achievable sum rate at the n-th time slot can be formulated as follows

$$R_{\mathrm{C}}(\mathbf{F}_n) = \sum_{k\in\mathcal{K}}\mathbb{E}_{\mathbf{H}_{k,n}}\{\log_2(1+\gamma_{k,n})\}, \tag{9}$$

where the expectation term $\mathbb{E}_{\mathbf{H}_{k,n}}\{\cdot\}$ is taken over the distribution of NLOS channel realizations.

D. Proposed Crowdsourced Bistatic Sensing

We propose a sensing framework that leverages the estimation results from K distributed GIDs. Accordingly, the overall sensing performance of the system is inherently determined by the individual sensing performance of each GID. To characterize the crowdsourced sensing capability, we aggregate

[6] In practice, the channel between the satellite and each GID is affected by propagation delays and Doppler shifts. However, in this work, we assume that the GIDs are equipped with mechanisms to mitigate these effects [44], [47], [48], ensuring that the channel remains quasi-static within its coherence time.

[7] While [19] considers only a single sensing receiver, our work generalizes this to a multi-receiver setting, where all ISAC receivers also function as sensing receivers.

[8] The bistatic angle corresponding to the k-th GID is defined as the angle subtended at the target between the satellite and the k-th GID [19].

the sensing performance of all GIDs, which can be expressed through a crowdsourced sensing utility function as follows

$$U_{\mathrm{S}}(\mathbf{F}_n) \triangleq f_{\mathrm{S}}\left(U_{\mathrm{B}}\left(\mathbf{y}_{1,n}(\mathbf{F}_n)\right), \ldots, U_{\mathrm{B}}\left(\mathbf{y}_{K,n}(\mathbf{F}_n)\right)\right), \quad (10)$$

where $f_{\mathrm{S}}(\cdot)$ denotes the crowdsourced aggregation function, and $U_{\mathrm{B}}\big(\mathbf{y}_{k,n}(\mathbf{F}_n)\big)$ represents the individual bistatic sensing utility of the k-th GID at time slot n. Following the conventional bistatic S-ISAC approach [18], [19], [20], successive interference cancellation is applied to $\mathbf{y}_{k,n}$ after communications decoding, and the resulting residual is denoted as the residual sensing echo $\mathbf{z}_{k,n}$ at the k-th GID.

Remark 4: The expression in (10) provides a general framework, where the individual bistatic sensing utility can be any sensing utility function defined based on any suitable sensing performance metric, i.e., the sensing rate [12], [26], Cramér-Rao Bound (CRB) [5], [6], estimation error [8], [11], or other metrics depending on the specific application requirements. In addition, the specific form of the crowdsourced aggregation function can be chosen according to the system objective, such as using a sum to maximize overall sensing rate, a minimum operator to ensure fairness, or a weighted sum to prioritize selected devices or regions.

In this work, as a realization of the sensing utility, we adopt the sensing rate to represent the limit of sensing information acquisition [12], [26]. Firstly, the sensing rate at the k-th GID can be obtained by estimating the mutual information (MI) between the residual sensing echo $\mathbf{z}_{k,n}$ and $\mathbf{G}_{k,n}$ given $\mathbf{x}_n$. According to the information theory, MI can be defined as follows

$$I(\mathbf{z}_{k,n}; \mathbf{G}_{k,n}|\mathbf{x}_n) = h(\mathbf{z}_{k,n}|\mathbf{x}_n) - h(\mathbf{z}_{k,n}|\mathbf{G}_{k,n}, \mathbf{x}_n), \quad (11)$$

where $h(\cdot)$ indicates the differential entropy. As derived in Appendix B, given $h(\mathbf{z}_{k,n}|\mathbf{x}_n)$ and $h(\mathbf{z}_{k,n}|\mathbf{G}_{k,n}, \mathbf{x}_n)$, MI can be written as follows

$$I(\mathbf{z}_{k,n}; \mathbf{G}_{k,n}|\mathbf{x}_n) \quad (12)$$
$$= \log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}} + \frac{(\mathbf{x}_n^T \otimes \mathbf{I}_{N_{\mathrm{R}}})\boldsymbol{\Sigma}_{\mathbf{g}_{k,n}}(\mathbf{x}_n^T \otimes \mathbf{I}_{N_{\mathrm{R}}})^H}{\sigma_k^2}\right).$$

Here, $\boldsymbol{\Sigma}_{\mathbf{g}_{k,n}} = \mathbb{E}\left[\mathbf{g}_{k,n}\mathbf{g}_{k,n}^H\right]$ denotes the covariance of the vectorized TRM $\mathbf{g}_{k,n} = \mathrm{vec}(\mathbf{G}_{k,n}) \sim \mathcal{CN}(\boldsymbol{\mu}_{\mathbf{G}_{k,n}}, \boldsymbol{\Sigma}_{\mathbf{G}_{k,n}})$, where $\boldsymbol{\mu}_{\mathbf{G}_{k,n}}$ and $\boldsymbol{\Sigma}_{\mathbf{G}_{k,n}}$ are the mean vector and covariance matrix of $\mathbf{G}_{k,n}$, respectively. Finally, the overall sensing rate can be expressed as follows[9]

$$R_{\mathrm{S}}(\mathbf{F}_n) = \sum_{k\in\mathcal{K}} \mathbb{E}_{\mathbf{G}_{k,n}}\{I(\mathbf{z}_{k,n}; \mathbf{G}_{k,n}|\mathbf{x}_n)\}, \quad (13)$$

where the expectation term $\mathbb{E}_{\mathbf{G}_{k,n}}\{\cdot\}$ corresponds to the ergodic average over $\mathbf{G}_{k,n}$. Accordingly, the sensing rate serves as a realization of the crowdsourced sensing utility function $U_{\mathrm{S}}(\mathbf{F}_n)$, and the summation itself is a realization of the underlying crowdsourced aggregation function $f_{\mathrm{S}}(\cdot)$.

[9]The summation term emphasizes the significance of crowdsourced sensing, as increasing the number of K directly contributes to enhanced performance.

III. Historical Geometric-Based Predictive Beamforming for Bistatic S-ISAC and Problem Formulation

Thus far, both the communications performance in (9) and the sensing performance in (13) are functions of the transmit beamforming matrix $\mathbf{F}_n$. This naturally motivates the design of an optimal beamformer $\hat{\mathbf{F}}_n$ to maximize the overall network utility. However, achieving this objective requires access to the CSI $\mathbf{H}_n \triangleq [\mathbf{H}_{1,n}, \ldots, \mathbf{H}_{K,n}] \in \mathbb{C}^{K\times N_{\mathrm{R}}\times N_{\mathrm{T}}}$ and the TRM $\mathbf{G}_n \triangleq [\mathbf{G}_{1,n}, \ldots, \mathbf{G}_{K,n}] \in \mathbb{C}^{K\times N_{\mathrm{R}}\times N_{\mathrm{T}}}$ [22], [26],[10] whose estimation incurs significant overhead. Recent predictive design approaches [5], [6], [26] attempt to eliminate real-time estimation by using historical CSI and TRM data as input. However, these methods still rely on the prior acquisition of full CSI and TRM datasets, resulting in a large number of required input parameters. In contrast, our proposed framework significantly reduces the input complexity by relying solely on HGI. As shown in (1) and (4), both CSI and TRM are functions of the underlying geometry (through distance-dependent channel gain and DoDs/DoAs). Therefore, it is physically reasonable to design a predictive beamforming framework that eliminates the need for explicit CSI or TRM inputs and instead operates directly on HGI, which represents the trajectories of the satellite, GIDs, and sensing target over time.[11] Predictive strategy mitigates the effect of long propagation delay [26], and in our proposed framework this is realized through predictive beamforming with HGI, which further saves time by eliminating the need for CSI and TRM acquisition.

To model the proposed framework, we define $g_\omega(\cdot)$ as a mapping function parameterized by ω that predicts the transmit beamforming matrix from the HGI inputs. Specifically, the optimal beamforming matrix $\hat{\mathbf{F}}_n$ for time slot n is predicted based on the HGI of the satellite, $\boldsymbol{\Omega}_{n-1}^{\tau} \triangleq \left[\mathbf{p}_{\mathrm{S},n-1}^T, \ldots, \mathbf{p}_{\mathrm{S},n-\tau}^T\right] \in \mathbb{R}^{3\times\tau}$, and the sensing target, $\boldsymbol{\Phi}_{n-1}^{\tau} \triangleq \left[\mathbf{p}_{\mathrm{A},n-1}^T, \ldots, \mathbf{p}_{\mathrm{A},n-\tau}^T\right] \in \mathbb{R}^{3\times\tau}$, together with the current geometric data of the K GIDs, $\mathbf{P}_{\mathrm{G}} \triangleq \left[\mathbf{p}_1^T, \ldots, \mathbf{p}_K^T\right] \in \mathbb{R}^{3\times K}$. Here, τ denotes the number of consecutive historical time slots. Accordingly, the prediction process is expressed as

$$\hat{\mathbf{F}}_n = g_\omega(\boldsymbol{\Omega}_{n-1}^{\tau}, \boldsymbol{\Phi}_{n-1}^{\tau}, \mathbf{P}_{\mathrm{G}}). \quad (14)$$

For notational brevity, the input parameters are grouped into a single matrix $\boldsymbol{\Delta}_{n-1} \triangleq [\boldsymbol{\Omega}_{n-1}^{\tau}, \boldsymbol{\Phi}_{n-1}^{\tau}, \mathbf{P}_{\mathrm{G}}] \in \mathbb{R}^{3\times(2\tau+K)}$, which is then used as prior information in the subsequent problem formulation.[12]

[10]As shown in our system model through (7), optimizing $\mathbf{F}_n$ inherently depends on the availability of $\mathbf{H}_n$ and $\mathbf{G}_n$. Consequently, conventional methods require their explicit estimation.

[11]In this work, we assume that the GIDs remain stationary due to their fixed deployment on the ground. As a result, only the satellite and the sensing target are considered to have trajectories. However, this assumption can be extended to accommodate moving GIDs by appropriately adjusting the input layer of the proposed framework. Please refer to Section IV-A.1.

[12]Since the optimization is performed on the satellite, its historical position is inherently known. The locations of the GIDs are fixed and are already known to the satellite. Therefore, the only uncertainty lies in the historical positions of the sensing targets. For simplicity, we assume this information is acquired onboard the satellite during an initial monostatic detection phase [5], [6], [26].

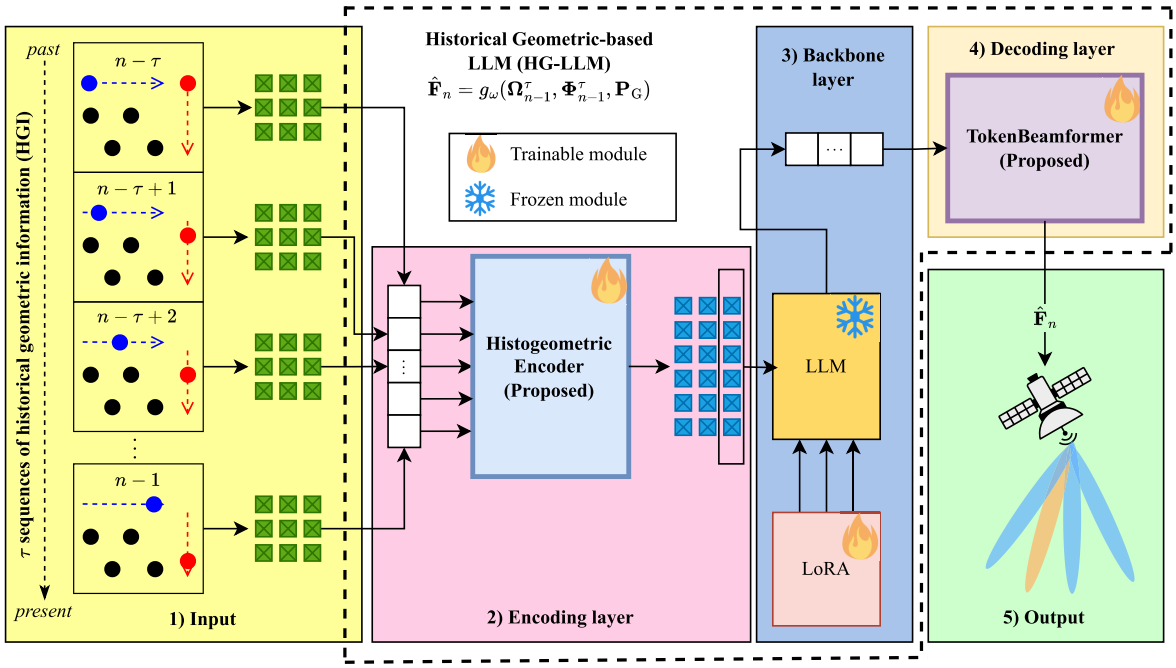


Fig. 3. Proposed HG-LLM predictive beamforming framework, which leverages a length-τ sequence of HGI (the coordinates of the satellite, target, and GIDs) to generate the future optimized beamforming matrix. The detailed architectures of the Histogeometric Encoder and TokenBeamformer are illustrated in Figs. 4 and 5, respectively.

Remark 5: The output of (14) can be extended to predict multiple optimal beamforming matrices for future time slots, represented as $[\hat{\mathbf{F}}_n, \ldots, \hat{\mathbf{F}}_{n+\nu-1}]$, where ν denotes the number of future time slots. However, increasing the prediction horizon may introduce drift over time, necessitating a robust optimization procedure to maintain performance [26]. Furthermore, the formulation in (14) is inherently flexible, allowing adaptation to generate other decision variables, i.e., transmit waveform, power allocation, or other decision variables.

Remark 6: By considering a τ-length historical input, the proposed HGI-based predictive beamforming framework requires only $3(2\tau + K)$ real-valued parameters. In contrast, channel-based schemes require $4\tau K N_\mathrm{T} N_\mathrm{R}$ parameters to represent both the real and imaginary components of CSI and TRM histories. This leads to a substantial reduction in input dimensionality, which becomes particularly advantageous for large antenna arrays. Please refer to Section V-F for a detailed comparison of input dimensionality.

The problem is *formulated* as follows, the main objective of this work is to design a transmit beamformer $\mathbf{F}_n$ for ISAC over a bistatic S-ISAC network by maximizing the network utility function given the HGI. In general, for the n-th time slot, the problem can be formulated as follows

$$\max_{\mathbf{F}_n} \mathbb{E}_{\mathbf{H}_n, \mathbf{G}_n | \mathbf{\Delta}_{n-1}} \Big\{ U\left(U_\mathrm{C}(\mathbf{F}_n, \mathbf{H}_n), U_\mathrm{S}(\mathbf{F}_n, \mathbf{G}_n)\right) \Big\}, \quad (15)$$

$$\text{s.t.} \quad \|\mathbf{F}_n\|_F^2 \le P_\mathrm{max}, \quad (15a)$$

where $U(\cdot)$ and $U_\mathrm{C}(\cdot)$ are the network and communications utilities, respectively. The expectation term $\mathbb{E}_{\mathbf{H}_n, \mathbf{G}_n | \mathbf{\Delta}_{n-1}}\{\cdot\}$ corresponds to the ergodic average over $\mathbf{H}_n$ and $\mathbf{G}_n$, given the HGI $\mathbf{\Delta}_{n-1}$. As for the constraint, P_max denotes the maximum transmit power budget.

Adopted from [26], we define the network utility function as a weighted sum of the normalized communications and sensing rates, expressed as follows

$$U(\mathbf{F}_n) \triangleq \frac{\rho_\mathrm{C} R_\mathrm{C}(\mathbf{F}_n)}{\mu_{\mathrm{C},n}} + \frac{(1-\rho_\mathrm{C}) R_\mathrm{S}(\mathbf{F}_n)}{\mu_{\mathrm{S},n}}, \quad (16)$$

where ρ_C denotes the communications priority, interpreted as the trade-off coefficient between the communications and sensing objectives. Additionally, $\mu_{\mathrm{C},n}$ and $\mu_{\mathrm{S},n}$ represent the upper bounds for communications and sensing rates, respectively, as derived in Appendix C. Unless otherwise specified, we set $\rho_\mathrm{C} = 0.5$ as a default value to indicate equal importance to both objectives.

IV. HG-LLM: Historical Geometric-Based LLM

To this end, it is obvious that obtaining a closed-form solution for (14) by solving the optimization problem in (15) is highly challenging, particularly given the limited availability of prior information. To address this, we propose HG-LLM that leverages the inherent generalization and generative capabilities of LLMs as a solution to accommodate the parameterization of function $g_\omega(\cdot)$. Although LLMs are computationally intensive, their deployment on satellites is feasible through methods such as model compression, knowledge distillation, quantization, and modular architectures, which reduce resource requirements while maintaining performance [52]. Moreover, since LLM inference has already been demonstrated on constrained terrestrial edge devices [39], [53], it is reasonable to expect that satellites can also support such tasks. Generally, the proposed LLM-driven predictive transmit beamforming framework is illustrated in Fig. 3.

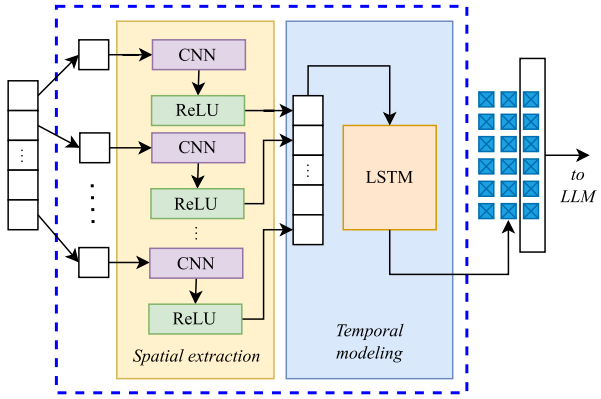


Fig. 4. Architecture of the proposed Histogeometric Encoder that bridges raw geometric inputs to LLM processing without the need for text-based prompts.

A. Proposed Model

Since LLMs are primarily designed for language tasks [27], applying them directly to domain-specific problems such as predictive beamforming is not straightforward [43]. Additionally, fine-tuning an entire LLM is prohibitively resource-intensive due to its massive parameter space. To overcome these challenges, our framework introduces trainable modules: the Histogeometric Encoder and the TokenBeamformer. To adapt the LLM to our S-ISAC domain-specific task, we adopt low-rank adaptation (LoRA) [54], which introduces trainable low-rank matrices into the projection layers. This enables efficient fine-tuning with minimal parameter updates, while leaving the original model weights unchanged. In this design, the Histogeometric Encoder transforms raw geometric inputs into LLM-compatible embeddings, while the TokenBeamformer maps the LLM's output into the final beamforming matrix. This structured approach enables domain-specific adaptation without full-scale LLM retraining, ensuring both efficiency and scalability. The proposed framework is structured as follows:

1) Input: The input to the framework consists of $\mathbf{\Omega}_{n-1}^{\tau}$, $\mathbf{\Phi}_{n-1}^{\tau}$, and $\mathbf{P}_{\mathrm{G}}$. These inputs represent the HGI that is used to predict the optimal beamforming matrices for future time slots. Let us recall that $\mathbf{\Delta}_{n-1}$, as defined in the problem formulation, serves as the input to the framework.

2) Encoding Layer: To enable LLM-driven predictive beamforming, we propose the Histogeometric Encoder, which bridges the gap between raw geometric inputs and LLM processing. Fig. 4 illustrates the architecture of the proposed encoder. Unlike conventional LLM-based methods that require text-based prompts, the Histogeometric Encoder directly maps HGI into structured embeddings, enabling the LLM to reason over spatial-temporal data without prompt conditioning. In addition, in contrast to prior works such as [40], our proposed Histogeometric Encoder directly processes HGI rather than historical CSI, thereby removing the dependence on CSI.

To effectively capture the spatial features from the HGI $\mathbf{\Delta}_{n-1}$, we first employ multiple parallel convolutional neural networks (CNNs), where at each time slot the concatenated 3D positions of the satellite, the sensing target, and all K GIDs are processed as a structured feature map. This parallel processing enables the framework to learn higher-level spatial representations for each time step, capturing relative distances and angular relationships efficiently. The t-th CNN operation can be expressed as follows

$$\mathbf{\Xi}_t = \sigma\left\{\mathrm{CNN}_t\left(\left[\mathbf{p}_{\mathrm{S},n-t}^T, \mathbf{p}_{\mathrm{A},n-t}^T, \mathbf{P}_{\mathrm{G}}^T\right]\right)\right\} \in \mathbb{R}^{(K+2)\times C}, \tag{17}$$

where C denotes the number of CNN's output channels, $t \in \{1,\ldots,\tau\}$ denotes the historical time step index, and $\sigma(\cdot)$ denote the ReLU activation function. Subsequently, the outputs of the parallel CNNs are concatenated as follows

$$\mathbf{\Xi} = [\mathbf{\Xi}_1, \ldots, \mathbf{\Xi}_\tau] \in \mathbb{R}^{\tau\times(K+2)\times C}. \tag{18}$$

After concatenation, the per-slot features are passed to a long short-term memory (LSTM) module, which fuses the τ-step history into a compact temporal state to model temporal dependencies, as follows

$$\mathbf{\Pi} = \mathrm{LSTM}(\mathbf{\Xi}) \in \mathbb{R}^{(2\tau+K)\times d}. \tag{19}$$

Here, the final hidden state of the LSTM serves as the latent representation, denoted as $\boldsymbol{\pi}_n = \mathbf{\Pi}[-1] \in \mathbb{R}^d$, where d represents the hidden dimension size of the LSTM. This temporal modeling stabilizes the input state delivered to the LLM by smoothing variability across the τ-step history and encoding trajectory trend, which improves LLM processing efficiency.

3) Backbone Layer: In this work, we aim to demonstrate the potential of LLMs for optimization in predictive beamforming. To adapt an LLM to the S-ISAC domain, fine-tuning is required, but direct fine-tuning of large models is computationally expensive and memory-intensive. To overcome this limitation, we adopt LoRA [54], which introduces trainable low-rank matrices into the projection layers and enables efficient parameter updates while preserving the pre-trained backbone. Unlike conventional LLM-based methods that rely on text prompts, our framework maps structured latent embeddings from the Histogeometric Encoder directly into the LLM's latent space, allowing predictive reasoning over spatial-temporal data without prompt-based conditioning.

Firstly, the projected latent representation from the encoder can be expressed as follows

$$\boldsymbol{\zeta}_n = \mathbf{P}_p \boldsymbol{\pi}_n \in \mathbb{R}^{\eta_t}, \tag{20}$$

where $\mathbf{P}_p \in \mathbb{R}^{\eta_t \times d}$ is a learned projection matrix, with η_t representing the model's hidden size of LLM. Then, the LLM processes the projected embedding to generate a contextualized latent representation that carries spatial and temporal reasoning as follows

$$\boldsymbol{\xi}_n = \mathrm{LLM}\left(\boldsymbol{\zeta}_n; \{\mathbf{M}'\}\right) \in \mathbb{R}^{\eta_t}, \tag{21}$$

where $\{\mathbf{M}'\}$ denotes the LoRA-adapted internal projection matrices. For each adapted LLM projection $\mathbf{M} \in \mathbb{R}^{\eta_t \times \eta_t}$, we apply LoRA update as follows [54]

$$\mathbf{M}' = \mathbf{M} + \mathbf{B}_r \mathbf{A}_r, \tag{22}$$

where $\mathbf{A}_r \in \mathbb{R}^{r\times\eta_t}$ and $\mathbf{B}_r \in \mathbb{R}^{\eta_t \times r}$ are the trainable low-rank adapter matrices, r denotes the adapter rank, and the base weight $\mathbf{M}$ remains frozen. We apply this update to the linear projections in the self-attention sublayers and set $r = 4$ by

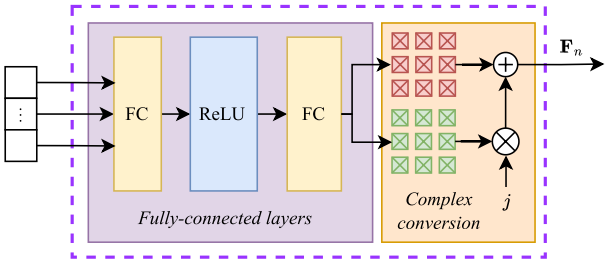


Fig. 5. Architecture of the proposed TokenBeamformer module that translates LLM outputs to the transmit beamforming matrix.

default. Since the pipeline bypasses the tokenizer and vocabulary, therefore no text-based inputs are used. The resulting representations are then processed to generate the optimized beamforming matrix, completing the predictive design loop efficiently.

Remark 7: The backbone of our framework can in principle utilize any suitable LLM or sequence model. Inspired by prior work [37], [40] and considering resource constraints and the relative simplicity of the predictive beamforming task, we deliberately select GPT-2[13] [55]. This lightweight model demonstrates that even a compact LLM can effectively realize the proposed HGI-based channel-agnostic predictive beamforming framework while maintaining computational efficiency.

4) Decoding Layer: To translate the latent representations generated by the LLM into actionable beamforming matrices, we propose the TokenBeamformer module. Fig. 5 illustrates the general architecture of the proposed module. This module is specifically designed to interpret the LLM's output and translate it into the structured format required for predictive beamforming. It consists of fully connected (FC) layers that map the latent space embeddings into real and imaginary components, resulting in a complex-valued beamforming matrix optimized for ISAC tasks. The output of LLM, $\boldsymbol{\xi}_n$ is fed into the TokenBeamformer to generate the beamforming matrix as follows

$$\boldsymbol{\Gamma}_n = \boldsymbol{\Psi}_2 \sigma(\boldsymbol{\Psi}_1 \boldsymbol{\xi}_n + \boldsymbol{\beta}_1) + \boldsymbol{\beta}_2 \in \mathbb{R}^{2N_{\mathrm{T}}(K+1)}, \tag{23}$$

where $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ are learned weight matrices, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are biases. Then, the vector $\boldsymbol{\Gamma}_n$ in (23) is reshaped and split into $\{\boldsymbol{\Gamma}_n^{\mathrm{Re}}, \boldsymbol{\Gamma}_n^{\mathrm{Im}}\} = \mathrm{reshape}(\boldsymbol{\Gamma}_n)$, where $\boldsymbol{\Gamma}_n^{\mathrm{Re}}, \boldsymbol{\Gamma}_n^{\mathrm{Im}} \in \mathbb{R}^{N_{\mathrm{T}} \times (K+1)}$. Therefore, the final beamforming matrix is constructed as follows

$$\mathbf{F}_n = \boldsymbol{\Gamma}_n^{\mathrm{Re}} + j\boldsymbol{\Gamma}_n^{\mathrm{Im}} \in \mathbb{C}^{N_{\mathrm{T}} \times (K+1)}. \tag{24}$$

Moreover, to guarantee compliance with the transmit power budget $P_{\max}$, a penalty term is incorporated during training, as explained in the next subsection. This design effectively bridges the gap between LLM-based reasoning and practical signal processing, enabling end-to-end generation of beamforming solutions directly from spatial-temporal information.

Remark 8: The decoding layer in our framework is designed with a modular structure, which means it can be replaced or extended with alternative modules to support different output representations. This flexibility allows the proposed framework to adapt to a variety of problem settings, rather than being restricted to a single task. In this paper, we demonstrate this capability in the context of predictive beamforming using the TokenBeamformer module.

5) Output: The output of the TokenBeamformer, a complex-valued beamforming matrix, is sent directly to the physical beamformer for real-time adjustment according to (6). This process optimally configures the satellite's transmission for both communications and sensing objectives, completing the predictive design pipeline.

B. Offline Training and Online Inference

To overcome the intractability of finding a closed-form solution to problem (15), we employ a data-driven method to asymptotically approximate the statistical expectation in the objective function. This requires performing offline training on the proposed framework. We begin by reformulating the constrained problem into an unconstrained form, followed by defining the offline training loss function. This process is essential for training the proposed encoder and decoder, as well as guiding the LoRA adaptation to achieve the desired optimization objectives. The main steps are outlined as follows:

1) Constraint Handling: To address the constraint, we employ the penalty method, which transforms the constrained problem into an equivalent unconstrained formulation by adding a penalty term to the objective [56]. This enables our LLM-based model and its LoRA adapters to jointly optimize performance and constraint adherence using standard gradient-based training. Thus, problem (15) can be reformulated as follows

$$\begin{aligned}\max_{\mathbf{F}_n} \; & \mathbb{E}_{\mathbf{H}_n, \mathbf{G}_n | \boldsymbol{\Delta}_{n-1}} \left\{ U\left(U_{\mathrm{C}}(\mathbf{F}_n, \mathbf{H}_n), U_{\mathrm{S}}(\mathbf{F}_n, \mathbf{G}_n)\right)\right\} \\ & - \psi \left[\max\left(0, \|\mathbf{F}_n\|_F^2 - P_{\max}\right)\right]^2, \end{aligned} \tag{25}$$

where ψ denotes the penalty constant applied when the power constrained is violated.[14] Even though the problem is now unconstrained, obtaining a closed-form solution remains intractable.

2) Data-Driven Approximation: To overcome the intractability of finding a closed-form solution to (25), we employ a data-driven approach to asymptotically approximate the statistical expectation in the objective function. Leveraging the powerful generative capabilities of LLM, our framework learns to approximate the solution to unconstrained problem effectively. Let $\mathbb{E}\{f(\mathbf{F}_n)\}$ denote transformed objective function of (25). The approximation can be written as follows

$$\begin{aligned}\mathbb{E}\{f(\mathbf{F}_n)\} &\approx \frac{1}{N_{\mathrm{s}}} \sum_{i=1}^{N_{\mathrm{s}}} f\left(\mathbf{F}_n^{(i)}\right) \\ &= \frac{1}{N_{\mathrm{s}}} \sum_{i=1}^{N_{\mathrm{s}}} f\left(g_\omega\left(\boldsymbol{\Omega}_{n-1}^{\tau(i)}, \boldsymbol{\Phi}_{n-1}^{\tau(i)}, \mathbf{P}_{\mathrm{G}}^{(i)}\right)\right), \end{aligned} \tag{26}$$

[13] As a demonstration, we use DistilGPT-2, a distilled variant of GPT-2 with 6 layers and approximately 82M parameters.

[14] Determining the appropriate value of ψ requires fine-tuning, and it can be either fixed or adaptive [56]. Empirically, larger constant yielded the same constraint satisfaction, affecting only the loss scale [5]. In this work, we choose a fixed value of $\psi = 1$ for simplicity of implementation.

TABLE I
DEFAULT SIMULATION PARAMETERS

Parameters	Default Values
Carrier frequency	$f_{\mathrm{c}} = 11$ GHz
Number of GIDs	$K = 3$ GIDs
Number of historical time slots	$\tau = 5$ time slots
Time slot duration	$\Delta_n = 1$ μs
Number of transmit antennas	$N_{\mathrm{T}}^x = N_{\mathrm{T}}^y = 4$ antennas
Number of receive antennas	$N_{\mathrm{R}}^x = N_{\mathrm{R}}^y = 4$ antennas
Single transmit antenna gain	$G_{\mathrm{T}} = 7$ dBi
Single receive antenna gain	$G_{\mathrm{R}} = 7$ dBi
Power budget	$P_{\max} = 10$ W
Noise power	$\sigma_k^2 = 10^{-9}$ W
Antenna separation	$\delta_x = \delta_y = 0.5\lambda_{\mathrm{c}}$
Radar cross section	$\sigma_{\mathrm{A}} = 10$ m^2

where N_{s} represents the number of training samples [5], [6], [26]. Here, superscript (i) denotes the data sample index, where $i \in \{1, \ldots, N_{\mathrm{s}}\}$. According to the universal approximation theorem, this approximation converges to the true expectation as the number of training samples becomes sufficiently large [57].

3) Loss Function: To convert the expected value approximation into a training loss function, we reformulate the original objective into its negative form as follows

$$\mathcal{L}(\omega) = -\frac{1}{N_{\mathrm{s}}} \sum_{i=1}^{N_{\mathrm{s}}} f\left(g_\omega\left(\mathbf{\Omega}_{n-1}^{\tau(i)}, \mathbf{\Phi}_{n-1}^{\tau(i)}, \mathbf{P}_{\mathrm{G}}^{(i)}\right)\right), \tag{27}$$

where the negative sign is introduced to convert the original maximization objective into its equivalent minimization form, aligning it with standard deep learning optimization techniques. The Histogeometric Encoder, TokenBeamformer, and the fine-tuning of the LLM via LoRA are jointly optimized using this loss function during the offline training process. It is worth noting that S-ISAC propagation characteristics, such as large attenuation, are inherently accounted for in this process since the loss function is defined based on (25), which directly depends on the underlying signal propagation models in Section II.

4) Online Inference: The final solution is obtained by optimizing ω to minimize the loss function through offline training. Once HG-LLM is well-trained, signifying that the loss function in (27) is minimized, the optimized beamforming matrix can be directly obtained by inferring the HGI into the model as follows

$$\hat{\mathbf{F}}_n = g_{\hat{\omega}}(\mathbf{\Omega}_{n-1}^{\tau}, \mathbf{\Phi}_{n-1}^{\tau}, \mathbf{P}_{\mathrm{G}}) \tag{28}$$

where $\hat{\omega} = \arg\min_\omega \mathcal{L}(\omega)$ is the well-trained HG-LLM's parameters, without additional offline training required.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present several simulations to evaluate the performance of the proposed framework. Unless stated otherwise, the default parameters used for the simulations are listed in Table I. The geometric descriptions for each device in the system model are outlined as follows:

- *Satellite*: For each sample realization, the satellite's initial position is uniformly distributed as $x_{\mathrm{S},n-\tau} \sim \mathcal{U}(-1, 1)$ km and $y_{\mathrm{S},n-\tau} \sim \mathcal{U}(0, 1)$ km, while its altitude remains fixed[15] at $z_{\mathrm{S},n} = 500$ km, $\forall n$. The satellite follows a linear trajectory along the x-axis at a constant speed of $v_{\mathrm{S}} = 7.8$ km/s, maintaining its initial y and z coordinates. The trajectory can be modeled as follows
$$x_{\mathrm{S},n+1} = x_{\mathrm{S},n} + \Delta_n v_{\mathrm{S}}. \tag{29}$$
- *Sensing target*: The initial position of the sensing target is randomized for each sample realization, modeled as $x_{\mathrm{A},n-\tau} \sim \mathcal{U}(-1, 1)$ km and $y_{\mathrm{A},n-\tau} \sim \mathcal{U}(0, 1)$ km, with a fixed altitude of $z_{\mathrm{A},n} = 500$ m, $\forall n$. By adopting a linear trajectory [26], we set the sensing target moves along the negative y-axis at a constant speed of $v_{\mathrm{A}} = 10$ m/s. The trajectory can be formulated as follows
$$x_{\mathrm{A},n+1} = x_{\mathrm{A},n} + \Delta_n v_{\mathrm{A}}. \tag{30}$$
- *GIDs*: During each data sample realization, the GIDs are uniformly distributed across the coverage area with coordinates sampled as $x_k \sim \mathcal{U}(-1, 1)$ km, $y_k \sim \mathcal{U}(0, 1)$ km, and fixed at ground level with $z_k = 0$, $\forall k \in \mathcal{K}$.

For offline training,[16] we utilize a dataset of 2000 samples, each containing 10 time slots, with a training-validation split factor of 0.5. The optimization of ω is performed using the AdamW optimizer [59] with an initial learning rate of 5×10^{-4}, adaptively adjusted through the ReduceLROnPlateau method with a reduction factor of 0.5 and a patience value of 10. Both offline training and online inference are executed on an AMD EPYC 7313 CPU and an NVIDIA GeForce RTX 3090 GPU.

To benchmark the performance of the proposed HG-LLM framework, we consider several comparison baselines that can be adapted to our problem setting. However, due to the lack of existing schemes that leverage HGI for predictive beamforming, we adopt several historical channel-based techniques as comparison benchmarks. Importantly, LLM-based optimization methods in the literature predominantly rely on text-based prompting (i.e. closed [33], [34], [36] and open vocabularies [32], [35]), which are not aligned with our approach that utilizes numerical geometric data instead of text-based prompts. To address this, we introduce CSI-HG-LLM as a variant of HG-LLM, specifically adapted to process historical CSI instead of HGI, bypassing the need for prompts while still leveraging LLM-based optimization. The details of these comparison schemes are outlined as follows:

- CSI-HG-LLM: A variant of the proposed HG-LLM that replaces HGI with historical CSI. The Histogeometric Encoder is substituted with a Historical CSI Encoder, while the rest of the architecture remains unchanged.[17]

[15]Since the satellite's movement is relatively small over the considered number of time slots, and the duration of each time slot is short, the effect of Earth's curvature can be neglected.

[16]All data generation and model training are implemented with PyTorch [58]. The generated datasets strictly follow the system model in Section II, ensuring compliance with physical propagation laws.

[17]This method considers an ideal scenario with complete access to all historical channel information, representing full channel state awareness for performance evaluation.

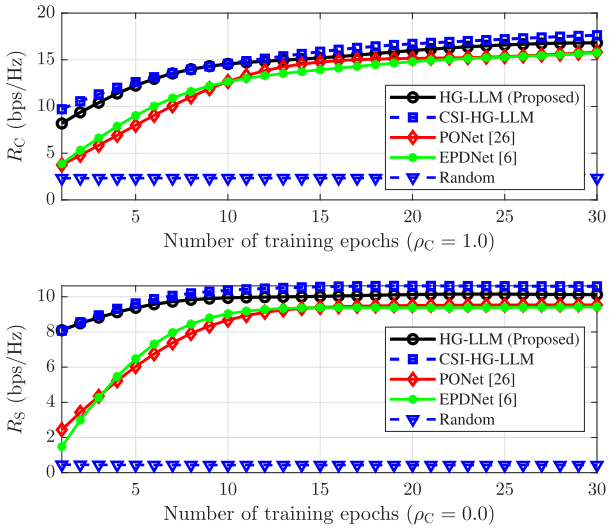


Fig. 6. Testing performance across various comparison schemes, evaluated in terms of communications rates for communications-focused mode ($\rho_C = 1$) and sensing rates for sensing-focused mode ($\rho_C = 0$).

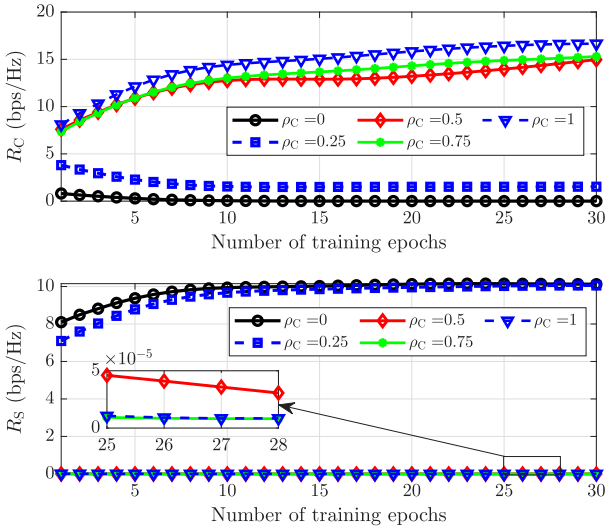


Fig. 7. Testing performance of the proposed HG-LLM, evaluated in terms of communications and sensing rates across different communications weight settings.

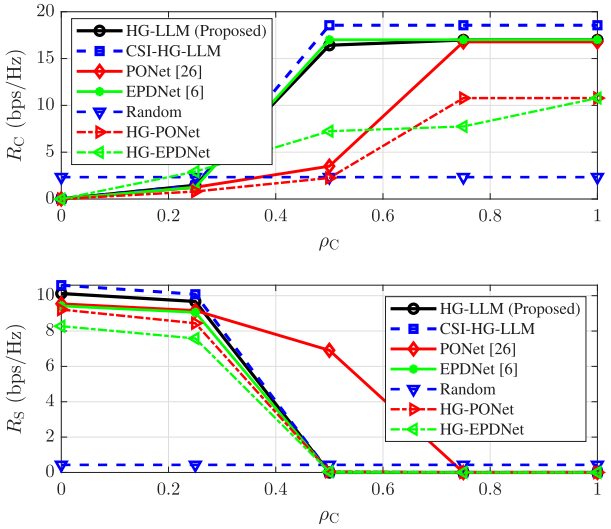


Fig. 8. Communications and sensing rates across different communications weight settings for various comparison schemes.

- PONet [26]: This method leverages historical CSI for predictive beamforming using a combination of CNN and Transformer encoder. Spatial features are extracted via individual CNNs, concatenated, and fed into the Transformer encoder. The final generation layer is modified to produce the beamforming matrix.
- EPDNet [6]: This scheme leverages historical CSI for predictive beamforming. It utilizes parallel CNNs to capture spatial features across multiple time slots, followed by an LSTM to extract temporal dependencies. The combined features are then decoded to generate the beamforming matrix.
- Random: This baseline generates random beamforming matrices and normalizes them to satisfy the power budget, ensuring the power output is always maintained at $P_{\max}$.

A. Testing Performance

To begin with, we evaluate the testing performance in terms of communications and sensing rates under their respective extreme conditions: communications-focused ($\rho_C = 1$) and sensing-focused ($\rho_C = 0$), as illustrated in Fig. 6. Notably, our proposed HG-LLM achieves performance comparable to other channel-based methods, despite relying solely on HGI, which inherently contains less channel-specific detail. This competitive performance is enabled by leveraging the powerful generative capabilities of the LLM, allowing it to infer effective beamforming solutions from geometric data alone. Furthermore, it is evident that CSI-HG-LLM, which directly feeds historical CSI into the HG-LLM architecture, consistently outperforms all other methods across both communications and sensing rates. This is expected, as access to full channel information provides an upper performance bound, demonstrating the advantage of CSI availability in achieving optimal beamforming. Nevertheless, the fact that HG-LLM, using only geometric data, performs closely to other channel-based methods highlights the effectiveness of the proposed framework.

To evaluate the testing performance of HG-LLM across various communications weight settings, we show Fig. 7. The communications and sensing rates show consistent convergence as the number of epochs increases, demonstrating the stability of the proposed framework in the S-ISAC system. However, when the communications weight ρ_C reaches 0.5 or higher, the sensing performance notably decreases. While this drop may appear minor at first glance, a closer look reveals a gradual decline, indicating that the model prioritizes communications as ρ_C increases. The sharp change between $\rho_C = 0.25$ and $\rho_C = 0.5$ suggests a critical trade-off point, emphasizing the need for careful adjustment to maintain a balance. A more detailed trade-off analysis will be presented in the following discussion.

B. Communications and Sensing Trade-off

Fig. 8 illustrates the communications rate and sensing rate across varying communications weight settings. We train a

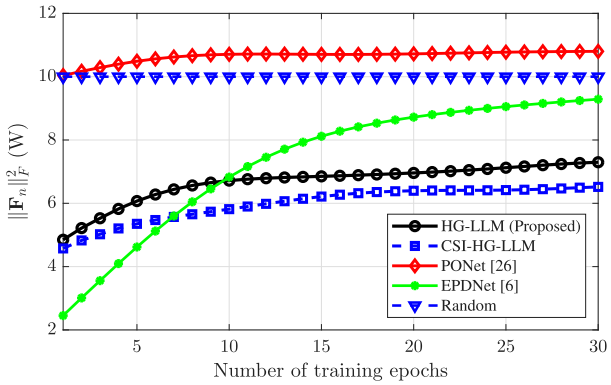


Fig. 9. Beamforming transmission power during the testing phase, evaluated over different number of training epochs and across various comparison schemes.

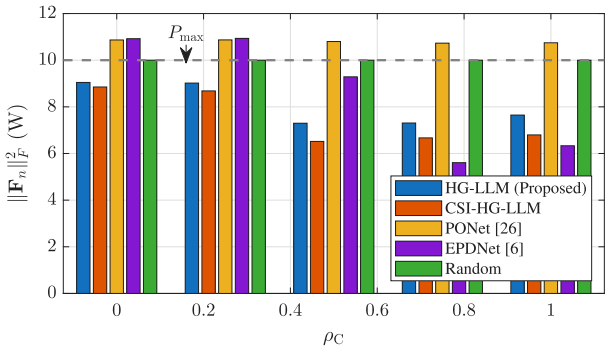


Fig. 10. Actual beamforming transmission power across various comparison schemes under different communications weights. The dashed line represents the power budget of $P_{\max} = 10$ W.

separate model at each ρ_C and plot the resulting trade-off curve. The results demonstrate that the proposed HG-LLM achieves comparable performance with other channel-based methods, despite relying solely on HGI. Furthermore, we include HGI-adapted versions of PONet [26] and EPDNet [6], namely HG-PONet and HG-EPDNet, to assess performance under HGI. As expected, these baselines underperform HG-LLM, since HGI carries less information than historical CSI and conventional models cannot fully exploit HGI. This highlights the benefit of HG-LLM's generative modeling within our framework.

For all methods, as ρ_C increases, the communications rate improves consistently, while the sensing rate correspondingly diminishes, reflecting the inherent trade-off in S-ISAC systems. Interestingly, the balance point between communications and sensing does not occur exactly at $\rho_C = 0.5$, but instead around the range of 0.25 to 0.5. This shift arises because the upper bound for the sensing rate is looser than that of the communications rate, causing the optimization to favor communications even when ρ_C is relatively modest. Ideally, an equal trade-off would be expected at $\rho_C = 0.5$, but the nature of the upper bounds skews this balance point. This highlights the need for proper adjustment of ρ_C to achieve a more equitable distribution of resources between communications and sensing tasks in our S-ISAC systems.

C. Power Constraint Satisfaction and Noise Impact

Since addressing power constraints is crucial, it is important to evaluate the actual power consumption of the beamforming matrix. Fig. 9 presents the beamforming transmission power during the testing phase across different training epochs. The results demonstrate that both the proposed HG-LLM and its variant, CSI-HG-LLM, consistently satisfy the power constraint ($P_{\max} = 10$ W) throughout the testing process. Notably, if the number of training epochs are greater than 11, they also exhibit greater power efficiency compared to EPDNet. This occurs because the TokenBeamformer combines a bounded nonlinearity in (23) with the power penalty in (25), which encourages solutions that meet the objective while staying below the budget. EPDNet, which uses an unbounded linear projection without an explicit power-aware mapping, also satisfies the constraint but tends to operate closer to $P_{\max}$. In contrast, PONet consistently exceeds the power budget, which is expected since it was originally designed with normalization instead of a penalty method for constraint handling. When the normalization layer is removed, PONet shows power budget violations. Although keeping the layer would satisfy the constraint by construction [26], but it would not be a fair comparison. By contrast, the proposed HG-LLM enforces the power limit through the penalty term and does not rely on normalization, which makes it more versatile across scenarios. This highlights the effectiveness of our proposed HG-LLM in addressing power constraints, regardless of whether it relies on HGI or historical CSI.

To further analyze the robustness of the beamforming matrix's actual power consumption across different settings, we present Fig. 10, which illustrates the beamforming power under various communications weight scenarios. It is evident that HG-LLM, both in its historical geometric-based and historical channel-based variants, consistently satisfies the power budget. This is enabled by the generalization capabilities of the LLM, allowing it to adapt across different scenarios effectively. Moreover, the observed consistency stems from the constraint-aware TokenBeamformer, which maintains the realized power within the budget across different weight settings. In contrast, PONet persistently violates the power budget, while EPDNet occasionally exceeds it. These observations highlight the robustness of our proposed framework in adhering to power constraints under any communications weight setting, distinguishing it from other baseline methods.

Fig. 11 illustrates the impact of noise power on communications rates in communications-focused mode and sensing rates in sensing-focused mode across various power budgets. At low noise levels (-110 dBm to -80 dBm), all power budgets achieve near-optimal rates, indicating that noise is not yet a limiting factor. As noise power increases beyond -80 dBm, higher power budgets sustain performance longer, while lower budgets experience rapid degradation. For example, the 50 dBm budget maintains stable rates until -60 dBm, whereas the 10 dBm budget drops off significantly earlier. Beyond -60 dBm, all power settings converge to near-zero rates, signifying that noise dominance overwhelms any power advantage. These results confirm that while higher

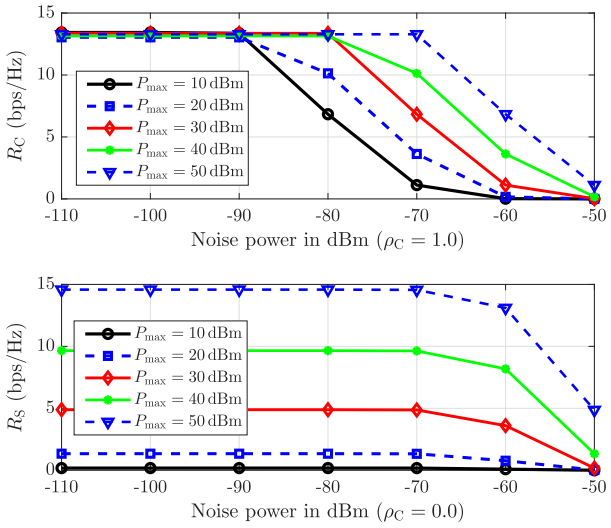


Fig. 11. Effect of noise power on communications rates in communications-focused mode and sensing rates in sensing-focused mode, evaluated across different maximum available power budgets.

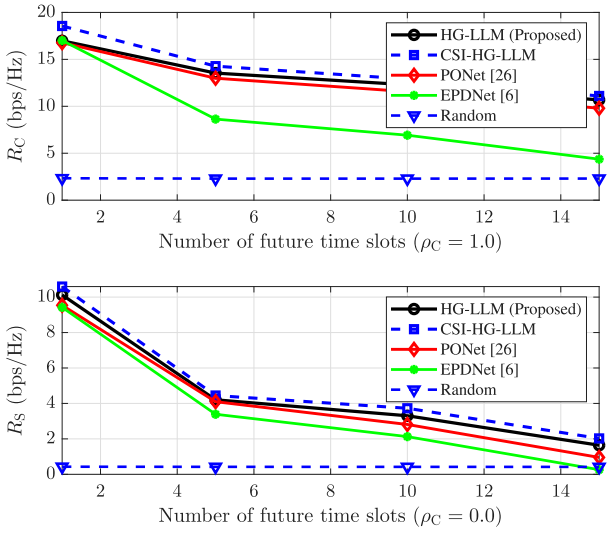


Fig. 12. Communications rates in communications-focused mode and sensing rates in sensing-focused mode versus the number of future time slots, evaluated across various comparison schemes.

power budgets can extend operational range under moderate noise, there is a critical threshold where noise suppression is essential. Notably, HG-LLM maintains robust performance across varying noise levels, demonstrating effective power management and resilience, even with HGI, outperforming other baselines that struggle with power constraints.

D. Prediction Horizon

Fig. 12 shows communications rates under communications-focused mode and sensing rates under sensing-focused mode as the prediction horizon increases. In both cases, all methods exhibit performance degradation over longer horizons, reflecting the prediction drift described in Remark 5. Notably, HG-LLM closely tracks its channel-based counterpart, CSI-HG-LLM, and consistently outperforms PONet at each future time slot, while significantly surpassing EPDNet and the Random baseline. Furthermore, HG-LLM degrades more slowly

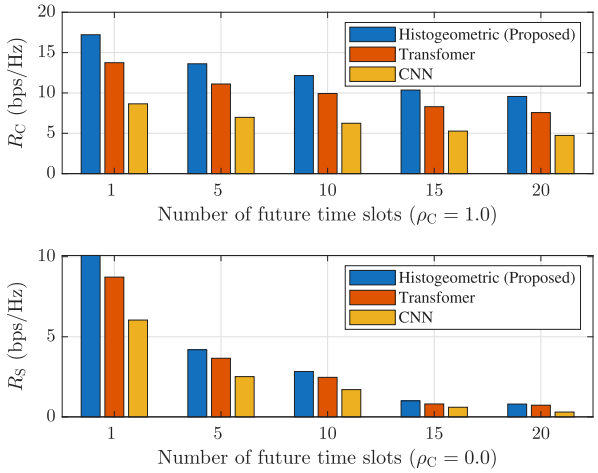


Fig. 13. Communications rates in communications-focused mode and sensing rates in sensing-focused mode versus the number of future time slots, evaluated across various encoder types.

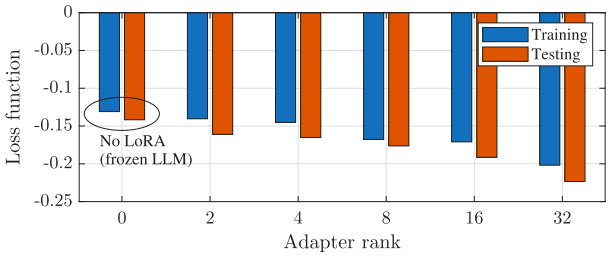
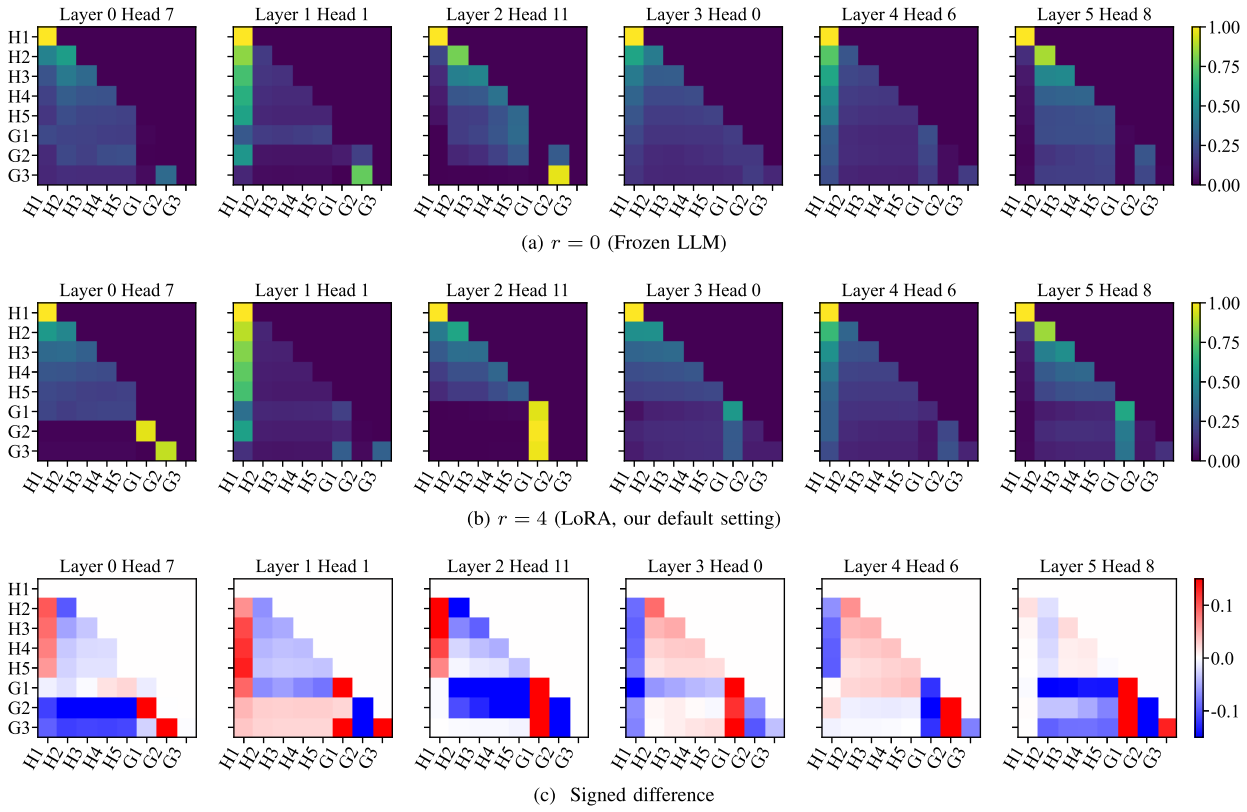


Fig. 14. Effect of LoRA rank on HG-LLM performance during training and testing in the default simulation. Rank 0 disables LoRA and keeps the LLM frozen.

with horizon because the Histogeometric Encoder couples per-step spatial feature extraction via parallel CNNs with temporal modeling implemented with an LSTM. The resulting denoised temporal latent stabilizes the predictor's input and limits error growth over long horizons. This demonstrates HG-LLM's robust predictive capability and its ability to approximate full-CSI performance even as prediction uncertainty grows. To prevent excessive performance decay over very long horizons, it is advisable to periodically refresh the historical input by reacquiring the true geometric positions after a predefined number of predicted steps. This approach helps reset the encoder's context and mitigates drift.

E. Encoder, LoRA, and Backbone Effects

Fig. 13 shows communications and sensing rates for HG-LLM under three encoders in communications- and sensing-focused modes. We compare the proposed Histogeometric encoder, a Transformer encoder, and a CNN encoder. The Histogeometric encoder gives the best results because it combines spatial structure with temporal history, as described in Section IV-A.2. The Transformer variant achieves lower performance because it lacks an explicit spatial inductive bias for the antenna layout. The CNN variant performs worst because its encoder does not model temporal dependence and treats each time step in isolation. Despite this ranking, the decay with longer prediction horizons is similar across encoders, since the



(a) $r = 0$ (Frozen LLM)

(b) $r = 4$ (LoRA, our default setting)

(c) Signed difference

Fig. 15. Layer-wise self-attention maps for the token sequence (H1 − H5, G1 − G3). For clarity, H denotes history tokens and G denotes geometry tokens. Rows show $r = 0$ (top), $r = 4$ (middle), and the signed difference (bottom), while columns show Transformer layers 0 to 5.

shared LLM backbone still models sequence dynamics and governs the horizon behavior.

Regarding fine-tuning of the backbone LLM, we evaluate the effect of LoRA rank on training and testing performance. Fig. 14 shows HG-LLM performance as we vary the LoRA rank r from 0 to 32. Rank 0 disables LoRA and keeps the LLM frozen. Higher ranks reduce training and test losses, suggesting higher task utility for the crowdsourced bistatic S-ISAC system. The gains taper beyond small ranks, with diminishing returns at larger r. Without LoRA, performance is notably worse across metrics. The pretrained backbone was not trained for our scenario, so a frozen model cannot adapt. LoRA adds small adapters that let the backbone adjust to our data. Even low ranks capture the main shift, so performance improves quickly. Balancing accuracy and parameter count, we adopt $r = 4$ as the default. It achieves losses comparable to larger ranks while training fewer parameters.

To further evaluate the impact of LoRA on LLM fine-tuning, we plot layer-wise self-attention in Fig. 15 for the token sequence. We show the attention map for $r = 0$ (frozen LLM), $r = 4$ (our default setting), and the signed difference. For each layer, we display only the head with the largest change to avoid clutter, since averaging across heads and the short as well as causal sequence can hide shifts. The maps reveal small but consistent increases in attention to geometry tokens in mid and late layers after fine-tuning, while early layers remain similar. Attention to recent history strengthens, while long-range context remains actively used. These patterns match our design: the Histogeometric encoder forms history and geometry tokens, the LLM refines them across depth, and the policy head outputs beam and power.

TABLE II

TRAINING AND TESTING NETWORK UTILITIES AT THE FINAL EPOCH ACROSS LLM BACKBONES

Backbones	**Number of parameters**	**Training** $U(\cdot)$	**Testing** $U(\cdot)$
DistilBERT	66×10^6	0.0922	0.099
DistilGPT2 (Default)	82×10^6	0.1461	0.1609
BERT [60]	110×10^6	0.1511	0.1621
GPT-2 [55]	124×10^6	0.1513	0.1633
Llama 3.2-1B [61]	1.23×10^9	0.1617	0.1636

Additionally, since our proposed HG-LLM is a backbone-agnostic framework, we can swap the LLM without other changes. We replace the backbone with DistilBERT (66M), BERT (110M) [60], standard GPT-2 (124M), and Llama-3.2-1B (1.23B) [61], in addition to the default distilled GPT-2 (82M). Table II shows that, in our HGI-only, short-history setting, final training and testing network utilities plateau at DistilGPT-2. Under the same settings and budget, larger backbones change the results only marginally. These findings support using DistilGPT-2 as the default, since it matches larger models with far fewer parameters and lower computation.

F. Complexity and Efficiency

Table III compares the computational complexity of each method in terms of floating point operations (FLOPs), peak

TABLE III

FLOPS, MEMORY USAGE, AND AVERAGE INFERENCE RUNTIME COMPARISON FOR VARIOUS METHODS UNDER $N_{\mathrm{T}} = N_{\mathrm{R}} = 256$

Methods	**FLOPs** ($\times 10^6$)	**Memory** (MB)	**Inference runtime** (ms)
HG-LLM (Proposed)	164.99	295.88	5.22
CSI-HG-LLM	5533.67	1286.56	5.85
PONet [26]	201.56	945.28	4.11
EPDNet [6]	202.06	946.20	4.12

TABLE IV

COMPARISON OF INPUT DIMENSIONALITY WITH $\tau = 5$ ACROSS DIFFERENT NUMBERS OF GIDS

K	**HG-LLM** $3(2\tau + K)$ (Proposed)	**CSI-based** $4\tau K N_{\mathrm{T}} N_{\mathrm{R}}$		
		N_{T}=16 N_{R}=16	N_{T}=64 N_{R}=64	N_{T}=256 N_{R}=256
2	42	1.0×10^4	1.6×10^5	2.6×10^6
4	54	2.0×10^4	3.3×10^5	5.2×10^6
6	66	3.1×10^4	4.9×10^5	7.9×10^6
8	78	4.1×10^4	6.6×10^5	1.0×10^7
10	90	5.1×10^4	8.2×10^5	1.3×10^7

memory usage, and inference runtime (average over 50 runs). Under large arrays, such as $N_{\mathrm{T}} = N_{\mathrm{R}} = 256$, our proposed HG-LLM attains the lowest FLOPs among all methods. This is because channel-based baselines scale with the channel dimension, which grows with $N_{\mathrm{T}}N_{\mathrm{R}}$, while our proposed HG-LLM does not. As noted in Remark 6, the HGI input size $3(2\tau + K)$ grows only with τ and K, whereas the historical CSI input $4\tau K N_{\mathrm{T}} N_{\mathrm{R}}$ grows with both number of antennas and GIDs. Additionally, Table IV confirms that the HGI input remains small with modest growth in τ or K, while the historical CSI input reaches up to 10^7 under the default simulation settings. Moreover, HG-LLM also uses the least memory while maintaining comparable inference runtime. This matters on satellite platforms, where on-board memory is a primary constraint [14], [15], [16]. For this reason we report memory and FLOPs as the main indicators of deployability. The resulting scalability makes HG-LLM well suited to large-array S-ISAC deployments, where CSI-based methods become computationally prohibitive.

VI. CONCLUSION

This paper investigated a crowdsourced bistatic S-ISAC system and proposed a channel-agnostic predictive beamforming framework that relied solely on HGI, eliminating the need for CSI. The framework was designed to maximize a network utility function under power constraints. To realize this framework, we proposed HG-LLM, an LLM-based model that effectively maps HGI to future beamforming matrices. To enable seamless adaptation to LLM processing, we proposed two novel modules: the Histogeometric Encoder, which transformed geometric trajectories into LLM-compatible embeddings without text-based prompts, and the TokenBeamformer, which converted LLM outputs into optimized beamforming matrices. Additionally, we incorporated LoRA for efficient fine-tuning of the LLM. Extensive simulations demonstrated that HG-LLM achieved performance levels comparable to channel-based methods across diverse scenarios, validating its robustness in both communications and sensing tasks while operating without explicit CSI. Finally, future work includes extending the framework to other S-ISAC tasks such as transceiver design to realize a complete S-ISAC pipeline, exploring different LLM architectures to assess their suitability for S-ISAC tasks, and investigating real-world implementation in satellite networks to demonstrate practical feasibility.

APPENDIX A
TRANSMIT AND RECEIVE ARRAY RESPONSE VECTORS

For clarity, all time-slot and user-specific subscripts are omitted in the following expressions. Let $(\theta^{\mathrm{T}}, \phi^{\mathrm{T}})$ and $(\theta^{\mathrm{R}}, \phi^{\mathrm{R}})$ denote the elevation and azimuth DoD and DoA, respectively. The corresponding transmit and receive array response vectors are expressed as

$$\mathbf{a}(\theta^{\mathrm{T}}, \phi^{\mathrm{T}}) = \mathbf{a}^x(\theta^{\mathrm{T}}, \phi^{\mathrm{T}}) \otimes \mathbf{a}^y(\theta^{\mathrm{T}}, \phi^{\mathrm{T}}) \in \mathbb{C}^{1\times N_{\mathrm{T}}}, \quad (31)$$

$$\mathbf{b}(\theta^{\mathrm{R}}, \phi^{\mathrm{R}}) = \mathbf{b}^x(\theta^{\mathrm{R}}, \phi^{\mathrm{R}}) \otimes \mathbf{b}^y(\theta^{\mathrm{R}}, \phi^{\mathrm{R}}) \in \mathbb{C}^{1\times N_{\mathrm{R}}}. \quad (32)$$

Each subarray response along the x- and y-axes, denoted by $\mathbf{v}^{\varsigma}(\theta, \phi)$, is given by

$$\mathbf{v}^{\varsigma}(\theta, \phi) = \frac{1}{\sqrt{N^{\varsigma}}} \left[1, e^{-j\vartheta^{\varsigma}(\theta,\phi)}, \ldots, e^{-j(N^{\varsigma}-1)\vartheta^{\varsigma}(\theta,\phi)}\right], \quad (33)$$

where $\varsigma \in \{x, y\}$ and $\mathbf{v}^{\varsigma}(\theta, \phi) \in \{\mathbf{a}^{\varsigma}(\theta^{\mathrm{T}}, \phi^{\mathrm{T}}), \mathbf{b}^{\varsigma}(\theta^{\mathrm{R}}, \phi^{\mathrm{R}})\}$ denotes either the transmit or receive subarray response, depending on the context. The phase shifts are defined as $\vartheta^x(\theta, \phi) = \pi \sin(\theta)\cos(\phi)$ and $\vartheta^y(\theta, \phi) = \pi \sin(\theta)\sin(\phi)$. As being used in (1) and (4), the above definitions apply to both communications and sensing channels.

APPENDIX B
MI (11) DERIVATION

The sensing rate $R_{\mathrm{S}}(\mathbf{F}_n)$ can be computed by firstly obtain the MI $I(\mathbf{z}_{k,n}; \mathbf{G}_{k,n}|\mathbf{x}_n)$ as described in (11). Firstly, by leveraging matrix multiplication identity [12], $\mathbf{z}_{k,n}$ given $\mathbf{x}_n$ can be written as $\mathbf{z}_{k,n}|\mathbf{x}_n = (\mathbf{x}_n^T \otimes \mathbf{I}_{N_{\mathrm{R}}})\mathbf{g}_{k,n} + \mathbf{n}_{k,n}$, where $\mathbf{g}_{k,n} = \mathrm{vec}(\mathbf{G}_{k,n})$. Consequently, the covariance matrix can be written as $\mathbf{\Sigma}_{\mathbf{z}_{k,n}|\mathbf{x}_n} = (\mathbf{x}_n^T \otimes \mathbf{I}_{N_{\mathrm{R}}})\mathbf{\Sigma}_{\mathbf{g}_{k,n}}(\mathbf{x}_n^T \otimes \mathbf{I}_{N_{\mathrm{R}}})^H + \sigma_k^2 \mathbf{I}_{N_{\mathrm{R}}}$. Therefore, the differential entropy can be expressed as follows

$$h(\mathbf{z}_{k,n}|\mathbf{x}_n) = \log_2 \det\left(\pi e \mathbf{\Sigma}_{\mathbf{z}_{k,n}|\mathbf{x}_n}\right). \quad (34)$$

Secondly, the covariance matrix of $\mathbf{z}_{k,n}$ given $\mathbf{G}_{k,n}$ and $\mathbf{x}_n$ can be written as $\mathbf{\Sigma}_{\mathbf{z}_{k,n}|\mathbf{G}_{k,n},\mathbf{x}_n} = \sigma_k^2 \mathbf{I}_{N_{\mathrm{R}}}$ Accordingly, the differential entropy can be defined as follows

$$h(\mathbf{z}_{k,n}|\mathbf{G}_{k,n}, \mathbf{x}_n) = \log_2 \det(\pi e \sigma_k^2 \mathbf{I}_{N_{\mathrm{R}}}). \quad (35)$$

In the end, we can substitute (34) and (35) into (11) as follows

$$I(\mathbf{z}_{k,n}; \mathbf{G}_{k,n}|\mathbf{x}_n) = \log_2 \det\left(\mathbf{\Sigma}_{\mathbf{z}_{k,n}|\mathbf{x}_n}\right) - \log_2 \det(\sigma_k^2 \mathbf{I}_{N_{\mathrm{R}}}), \quad (36)$$

where the term πe is eliminated as a result of logarithmic properties applied during simplification. Additionally, applying another logarithmic property allows (36) to be rewritten as (12).

APPENDIX C
UPPER BOUND ANALYSIS OF COMMUNICATIONS AND SENSING RATES FOR (16)

Intuitively, the maximum communications rate is achieved under an ideal assumption in which no interference is present. Under the ideal assumption, we can also assume that the receive beamforming $\mathbf{w}_{k,n} \propto \mathbf{H}_{k,n}\mathbf{f}_{k,n}$. Therefore, the upper bound SINR can be simplified as follows

$$\gamma_{k,n}^{\max} = \frac{\|\mathbf{H}_{k,n}\mathbf{f}_{k,n}\|^2}{\sigma_k^2}. \tag{37}$$

To simplify the analysis, we assume the power budget is uniformly allocated, $\|\mathbf{f}_{k,n}\|^2 \leq \frac{P_{\max}}{K+1}$. Then, let $\Sigma_{\mathbf{H}_{k,n}}$ denote the largest singular value of $\mathbf{H}_{k,n}$. Therefore, the maximum communications rate can be expressed as follows

$$\mu_{\mathrm{C},n} = \sum_{k\in\mathcal{K}} \mathbb{E}_{\mathbf{H}_{k,n}} \left\{ \log_2 \left(1 + \frac{\Sigma_{\mathbf{H}_{k,n}} \frac{P_{\max}}{K+1}}{\sigma_k^2} \right) \right\}. \tag{38}$$

For the maximum sensing rate, we begin by finding the eigenvalues of $\mathbf{\Lambda}_{k,n} = \frac{1}{\sigma_k^2}\mathbf{G}_{k,n}\mathbf{F}_n\mathbf{F}_n^H\mathbf{G}_{k,n}^H$. Therefore, the MI can be re-written as follows

$$I(\mathbf{z}_{k,n}; \mathbf{G}_{k,n}|\mathbf{x}_n) \leq N_{\mathrm{R}} \log_2 \left(1 + \frac{1}{N_{\mathrm{R}}} \mathrm{Tr}(\mathbf{\Lambda}_{k,n}) \right), \tag{39}$$

where $\{\lambda_1^{(k,n)}, \ldots, \lambda_{N_{\mathrm{R}}}^{(k,n)}\}$ denote the eigenvalues of $\mathbf{\Lambda}_{k,n}$. Then, $\mathrm{Tr}(\mathbf{\Lambda}_{k,n})$ can be further expanded as follows

$$\mathrm{Tr}(\mathbf{\Lambda}_{k,n}) \leq \frac{\lambda_{\max}^{(k)}(\mathbf{G}_{k,n}^H\mathbf{G}_{k,n})\|\mathbf{F}_n\|_F^2}{\sigma_k^2}, \tag{40}$$

where $\lambda_{\max}^{(k,n)} = \max\{\lambda_1^{(k,n)}, \ldots, \lambda_{N_{\mathrm{R}}}^{(k,n)}\}$. By assuming $\|\mathbf{F}_n\|_F^2 \leq P_{\max}$, the maximum sensing rate can be expressed as follows

$$\mu_{\mathrm{S},n} = \sum_{k\in\mathcal{K}} N_{\mathrm{R}} \log_2 \left(1 + \frac{\lambda_{\max}^{(k,n)} P_{\max}}{N_{\mathrm{R}}\sigma_k^2} \right) \tag{41}$$

where $\lambda_{\max}^{(k,n)}$ denote the maximum eigenvalue of $\mathbf{G}_{k,n}^H\mathbf{G}_{k,n}$. This completes the derivation.

REFERENCES

[1] A. Liu et al., "A survey on fundamental limits of integrated sensing and communication," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 994–1034, 2nd Quart., 2022.

[2] 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.

[3] 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.

[4] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.

[5] 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.

[6] X. Zhang, W. Yuan, C. Liu, J. Wu, and D. W. K. Ng, "Predictive beamforming for vehicles with complex behaviors in ISAC systems: A deep learning approach," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 5, pp. 828–841, Jul. 2024.

[7] M. Huang, F. Gong, G. Li, N. Zhang, and Q.-V. Pham, "Secure precoding for satellite NOMA-aided integrated sensing and communication," *IEEE Internet Things J.*, vol. 11, no. 18, pp. 29533–29545, Sep. 2024.

[8] S. Naoumi, A. Bazzi, R. Bomfin, and M. Chafii, "Complex neural network based joint AoA and AoD estimation for bistatic ISAC," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 5, pp. 842–856, Jul. 2024.

[9] N. K. Nataraja, S. Sharma, K. Ali, F. Bai, R. Wang, and A. F. Molisch, "Integrated sensing and communication (ISAC) for vehicles: Bistatic radar with 5G-NR signals," *IEEE Trans. Veh. Technol.*, vol. 74, no. 4, pp. 1–16, Apr. 2025.

[10] N. Zhao, Q. Chang, X. Shen, Y. Wang, and Y. Shen, "Joint target localization and data detection in bistatic ISAC networks," *IEEE Trans. Commun.*, vol. 73, no. 5, pp. 1–16, May 2025.

[11] C. Luo, A. Tang, F. Gao, J. Liu, and X. Wang, "Channel modeling framework for both communications and bistatic sensing under 3GPP standard," *IEEE J. Sel. Areas Sensors*, vol. 1, pp. 166–176, 2024.

[12] Q. Qi, X. Chen, C. Zhong, C. Yuen, and Z. Zhang, "Deep learning-based design of uplink integrated sensing and communication," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10639–10652, Sep. 2024.

[13] X. Lin, S. Cioni, G. Charbit, N. Chuberre, S. Hellsten, and J.-F. Boutillon, "On the path to 6G: Embracing the next wave of low Earth orbit satellite access," *IEEE Commun. Mag.*, vol. 59, no. 12, pp. 36–42, Dec. 2021.

[14] M. Vaezi et al., "Cellular, wide-area, and non-terrestrial IoT: A survey on 5G advances and the road toward 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1117–1174, 2nd Quart., 2022.

[15] M. Centenaro, C. E. Costa, F. Granelli, C. Sacchi, and L. Vangelista, "A survey on technologies, standards and open challenges in satellite IoT," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1693–1720, 3rd Quart., 2021.

[16] J. A. Fraire, O. Iova, and F. Valois, "Space-terrestrial integrated Internet of Things: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 60, no. 12, pp. 64–70, Dec. 2022.

[17] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 2, pp. 244–251, Mar. 2021.

[18] J. Park, J. Seong, Y. Mao, W. Shin, and B. Ottersten, "A bistatic ISAC framework for LEO satellite systems: A rate-splitting approach," *IEEE Trans. Aerosp. Electron. Syst.*, early access, Aug. 28, 2025, doi: 10.1109/TAES.2025.3603500.

[19] J. Park, J. Seong, J. Ryu, Y. Mao, and W. Shin, "RSMA-based bistatic ISAC framework for LEO satellite systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Denver, CO, USA, Jun. 2024, pp. 1840–1845.

[20] L. Yin and B. Clerckx, "Rate-splitting multiple access for dual-functional radar-communication satellite systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Austin, TX, USA, Apr. 2022, pp. 1–6.

[21] B. S. Mysore, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Precoding for satellite communications: Why, how and what next?," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2453–2457, Aug. 2021.

[22] L. You et al., "Integrated communications and localization for massive MIMO LEO satellite systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11061–11075, Sep. 2024.

[23] L. You et al., "Beam squint-aware integrated sensing and communications for hybrid massive MIMO LEO satellite systems," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 2994–3009, Oct. 2022.

[24] M. Ying, X. Chen, Q. Qi, and W. Gerstacker, "Deep learning-based joint channel prediction and multibeam precoding for LEO satellite Internet of Things," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13946–13960, Oct. 2024.

[25] Y. Zhang, A. Liu, P. Li, and S. Jiang, "Deep learning (DL)-based channel prediction and hybrid beamforming for LEO satellite massive MIMO system," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23705–23715, Dec. 2022.

[26] W. D. Lukito, W. Xiang, C. Liu, P. Lai, P. Cheng, and G. Mao, "Learning to design transceiver for integrated sensing and communications: A satellite communications perspective," *IEEE Trans. Wireless Commun.*, vol. 24, no. 9, pp. 7409–7423, Sep. 2025.

[27] S. Minaee et al., "Large language models: A survey," 2024, *arXiv:2402.06196*.

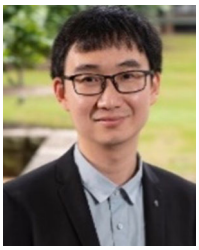
[28] S. Han, J. Yoon, S. O. Arik, and T. Pfister, "Large language models can automatically engineer features for few-shot tabular learning," 2024, *arXiv:2404.09491*.

[29] V. Balek, L. Sýkora, V. Sklenák, and T. Kliegr, "LLM-based feature generation from text for interpretable machine learning," 2024, *arXiv:2409.07132*.

[30] F. Jiang et al., "Large language model enhanced multi-agent systems for 6G communications," *IEEE Wireless Commun.*, vol. 31, no. 6, pp. 48–55, Dec. 2024.
[31] L. Cheng, H. Zhang, B. Di, D. Niyato, and L. Song, "Large language models empower multimodal integrated sensing and communication," *IEEE Commun. Mag.*, vol. 63, no. 5, pp. 1–8, May 2025.
[32] H. Li, M. Xiao, K. Wang, D. I. Kim, and M. Debbah, "Large language model based multi-objective optimization for integrated sensing and communications in UAV networks," *IEEE Wireless Commun. Lett.*, vol. 14, no. 4, pp. 979–983, Apr. 2025.
[33] H. Quan et al., "Large language model agents for radio map generation and wireless network planning," *IEEE Netw. Lett.*, vol. 7, no. 3, pp. 1–5, Sep. 2025.
[34] R. Zhang et al., "Generative AI agents with large language model for satellite networks via a mixture of experts transmission," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 12, pp. 1–16, Dec. 2024.
[35] M. Xu et al., "When large language model agents meet 6G networks: Perception, grounding, and alignment," *IEEE Wireless Commun.*, vol. 31, no. 6, pp. 63–71, Dec. 2024.
[36] K. Qiu, S. Bakirtzis, I. Wassell, H. Song, J. Zhang, and K. Wang, "Large language model-based wireless network design," *IEEE Wireless Commun. Lett.*, vol. 13, no. 12, pp. 3340–3344, Dec. 2024.
[37] Z. Xu, T. Zheng, and L. Dai, "LLM-empowered near-field communications for low-altitude economy," *IEEE Trans. Commun.*, vol. 73, no. 11, pp. 11186–11196, Nov. 2025.
[38] Y. Sheng, K. Huang, L. Liang, P. Liu, S. Jin, and G. Ye Li, "Beam prediction based on large language models," *IEEE Wireless Commun. Lett.*, vol. 14, no. 5, pp. 1–5, May 2025.
[39] Y. Rong, Y. Mao, H. Cui, X. He, and M. Chen, "Edge computing enabled large-scale traffic flow prediction with GPT in intelligent autonomous transport system for 6G network," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 10, pp. 1–18, Oct. 2025.
[40] B. Liu, X. Liu, S. Gao, X. Cheng, and L. Yang, "LLM4CP: Adapting large language models for channel prediction," *J. Commun. Inf. Netw.*, vol. 9, no. 2, pp. 113–125, Jun. 2024.
[41] Y. Cui, J. Guo, C.-K. Wen, S. Jin, and E. Tong, "Exploring the potential of large language models for massive MIMO CSI feedback," 2025, *arXiv:2501.10630*.
[42] H. Zhou et al., "Large language models for wireless networks: An overview from the prompt engineering perspective," *IEEE Wireless Commun.*, vol. 32, no. 4, pp. 98–106, Aug. 2025.
[43] Q. Liu, J. Mu, D. Chen, R. Zhang, Y. Liu, and T. Hong, "LLM enhanced reconfigurable intelligent surface for energy-efficient and reliable 6G IoV," *IEEE Trans. Veh. Technol.*, vol. 74, no. 2, pp. 1830–1838, Feb. 2025.
[44] X. Ding, Y. Ren, X. Xie, Y. Zou, M. Jia, and G. Zhang, "Improving user capacity of satellite Internet of Things via joint user grouping and multi-beam processing," *IEEE Trans. Commun.*, vol. 72, no. 7, pp. 3957–3969, Jul. 2024.
[45] W. D. Lukito et al., "Integrated STAR-RIS and UAV for satellite IoT communications: An energy-efficient approach," *IEEE Internet Things J.*, vol. 12, no. 9, pp. 11356–11371, May 2025.
[46] Z. Xiao et al., "Antenna array enabled space/air/ground communications and networking for 6G," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 2773–2804, Oct. 2022.
[47] Y. Zuo, M. Yue, M. Zhang, S. Li, S. Ni, and X. Yuan, "OFDM-based massive connectivity for LEO satellite Internet of Things," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8244–8258, Nov. 2023.
[48] K. Ying et al., "Quasi-synchronous random access for massive MIMO-based LEO satellite constellations," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1702–1722, Jun. 2023.
[49] R. E. Kell, "On the derivation of bistatic RCS from monostatic measurements," *Proc. IEEE*, vol. 53, no. 8, pp. 983–988, Aug. 1965.
[50] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 271–286, Jan. 2006.
[51] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[52] S. Javaid, R. A. Khalil, N. Saeed, B. He, and M.-S. Alouini, "Leveraging large language models for integrated satellite-aerial-terrestrial networks: Recent advances and future directions," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 399–432, 2025.
[53] X. Zhang et al., "Beyond the cloud: Edge inference for generative large language models in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 24, no. 1, pp. 643–658, Jan. 2025.
[54] E. J. Hu et al., "LoRA: Low-rank adaptation of large language models," 2021, *arXiv:2106.09685*.
[55] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI*, 2019. [Online]. Available: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
[56] A. E. Smith and D. W. Coit, "Constraint-handling technqiues—Penalty functions," in *Handbook of Evolutionary Computation*. Bristol, U.K.: Oxford Univ. Press, 1997.
[57] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[58] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. NeurIPS*, vol. 32, Vancouver, BC, Canada, Dec. 2019, pp. 8026–8037.
[59] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, New Orleans, LA, USA, May 2017, pp. 1–11.
[60] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL*, vol. 1, Jun. 2019, pp. 4171–4186.
[61] A. Grattafiori et al., "The Llama 3 herd of models," 2024, *arXiv:2407.21783*.

William D. Lukito (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from Bandung Institute of Technology (ITB), Bandung, Indonesia, in 2021 and 2022, respectively. He is currently pursuing the Ph.D. degree with La Trobe University, Australia, supported by Australian SmartSat CRC Scholarship. His research interests include wireless communications, the Internet of Things, artificial intelligence, embedded systems, and digital signal processing.

Wei Xiang (Senior Member, IEEE) founded two world-first research centres, namely the Cisco-La Trobe Centre for AI and IoT in 2020 and Australian Centre for AI and Medical Innovation (ACAMI) in 2024. Previously, he was the Foundation Chair and the Head of the Discipline of IoT Engineering, James Cook University, Cairns, Australia. For the past six consecutive years (2020–2025), has consistently ranked among Stanford University World's Top 2% Scientists for both his single-year and career-long impacts. Due to his instrumental leadership in establishing Australia's first accredited Internet of Things Engineering degree program, he was inducted into Pearcy Foundation's Hall of Fame in October 2018. He is currently a La Trobe Distinguished Professor and the Cisco Research Chair of AI and IoT with La Trobe University He is also a TEDx Speaker and an elected fellow of the IET in U.K. and Engineers Australia. He received the TNQ Innovation Award in 2016, Pearcey Entrepreneurship Award in 2017, and Engineers Australia Cairns Engineer of the Year in 2017. He was a co-recipient of four Best Paper Awards at WiSATS'2019, WCSP'2015, IEEE WCNC'2011, and ICWMC'2009. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow (2010–2011) by Queensland Government of Australia, an Endeavour Research Fellow (2012–2013) by the Commonwealth Government of Australia, a Smart Futures Fellow (2012–2015) by Queensland Government of Australia, and a JSPS Invitational Fellow jointly by Australian Academy of Science and Japanese Society for Promotion of Science (2014–2015). He was the Vice Chair of the IEEE Northern Australia Section from 2016 to 2020. He is currently an Associate Editor of IEEE COMMUNICATIONS SURVEYS & TUTORIALS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE INTERNET OF THINGS JOURNAL, IEEE ACCESS, and Nature journal of *Scientific Reports*. He has published over 450 peer-reviewed articles, including three books and nearly 300 journal articles. His research interests include the Internet of Things, wireless communications, machine learning for IoT data analytics, and computer vision. He has severed in a large number of international conferences in the capacity of the general co-chair, the TPC co-chair, and the symposium chair.



Chang Liu (Member, IEEE) received the Ph.D. degree from Dalian University of Technology, China, in 2017. In 2015, he was a Joint Ph.D. Scholar with the University of Tennessee, Knoxville, TN, USA. From 2017 to 2023, he held research appointments at several leading international institutions, including a Post-Doctoral Research Fellow with the University of Electronic Science and Technology of China and The University of New South Wales, Sydney, Australia, an Alexander von Humboldt Research Fellowship at the Friedrich Alexander University Erlangen–Nuremberg, Germany, and a Research Assistant Professor with The Hong Kong Polytechnic University. He is currently a Lecturer (Assistant Professor in U.S. system) with the Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia. To date, he has published more than 70 papers in leading international journals and conferences. He was awarded the prestigious Alexander von Humboldt Research Fellowship and has been consistently recognized among the World's Top 2% Scientists by Stanford University since 2022. His papers have received multiple distinctions, including IEEE Best Paper Awards, ESI Hot Papers, and ESI & SciVal Highly Cited Papers. Several of his works have been featured in IEEE Best Readings, IEEE Selected Ideas in Communications, and IEEE Popular Articles lists. He currently serves as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS. He is also the EDAS Chair of IEEE PIMRC 2026, the Publication Chair of IEEE ISWCS 2026, a Founding Member of the IEEE ComSoc Special Interest Group on Orthogonal Time Frequency Space (OTFS), and the Session Chair of IEEE ICC 2017. His research interests span machine learning in communication and sensing systems, with a focus on the Internet of Things (IoT), Vehicular-to-Everything (V2X) networks, and low-altitude wireless networks (LAWN), supported by enabling technologies including integrated sensing and communication (ISAC), intelligent reflecting surface (IRS), OTFS, and other emerging next-generation wireless techniques.

Phu Lai received the Ph.D. degree in information and communications technology from Swinburne University of Technology, Australia, in 2022. He is currently a Post-Doctoral Research Fellow with Cisco-La Trobe Centre for AI and IoT and Australian Centre for AI in Medical Innovation (ACAMI), where he has been working on a number of research and industry-sponsored projects in retails, wireless communications, bioinformatics, and smart cities. His research interests include applied AI, physical AI, optimization, and cloud computing with applications across different domains.

Peng Cheng (Member, IEEE) received the B.S. and M.S. degrees (Hons.) in communication and information systems from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2006 and 2009, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2013. From 2014 to 2017, he was a Post-Doctoral Research Scientist with CSIRO, Sydney, Australia. From 2017 to 2020, he was an ARC DECRA Fellow/Lecturer with The University of Sydney, Australia. He is currently an Associate Professor with the Department of Computer Science and Information Technology, La Trobe University, Australia, and The University of Sydney. He has published over 140 peer-reviewed research papers in leading international journals and conferences. His current research interests include wireless AI, machine learning, the IoT, millimeter-wave communications, and compressive sensing theory.

Weijie Yuan's (Senior Member, IEEE) research interests include integrated sensing and communications (ISAC), orthogonal time frequency space (OTFS), and the low-altitude wireless networks (LAWN). He currently serves as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, *IEEE Communications Magazine*, *IEEE Communications Standards Magazine*, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, IEEE COMMUNICATIONS LETTERS, IEEE OPEN JOURNAL OF COMMUNICATIONS SOCIETY, *EURASIP Journal of Advances in Signal Processing*, and *npj Wireless Technology*. He has led four special issues in IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and *China Communications*. He was a Guest Editor of IEEE INTERNET OF THINGS JOURNAL and IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY. He also serving/served as the General Co-Chair for ISWCS 2026, the Symposium Co-Chair for IEEE/CIC ICCC 2026, the Special Session Chair for IEEE ISAC 2026, and the Track Co-Chair for IEEE ICC 2025 and IEEE VTC 2025-Spring. He served as an Organizer/Chair for several workshops and special sessions in flagship IEEE and ACM conferences, including IEEE ICC, IEEE VTC, IEEE GlobeCom, IEEE/CIC ICCC, IEEE SPAWC, IEEE WCNC, IEEE ICASSP, and ACM MobiCom. He is the Chair of the IEEE Aerospace and Electronic System Technical Working Group on LAWN and the ComSoc Special Interest Group (SIG) on LAWN. He was a recipient of the Best Editor of IEEE CommL, the Best Paper Award from IEEE ICC 2023, IEEE/CIC ICCC 2023, and IEEE GlobeCom 2024; and the 2025 IEEE Communications Society & Information Theory Society Joint Paper Award.

Guoqiang Mao (Fellow, IEEE) is currently a Chair Professor and the Director of the Center for Smart Driving and Intelligent Transportation Systems, Southeast University. Between 2014 and 2019, he was a Leading Professor and the Founding Director of the Research Institute of Smart Transportation and the Vice Director of the ISN State Key Laboratory, Xidian University. Before that, he was with the University of Technology Sydney and The University of Sydney. He has published 300 papers in international conferences and journals that have been cited more than 15 000 times. His H-index is 57 and is in the list of Top 2% most-cited Scientists Worldwide by Stanford University in 2022, 2023, and 2024 both by Single Year and by Career Impact. His research interests include intelligent transport systems, the Internet of Things, wireless localization techniques, mobile communication systems, and applied graph theory and its applications in telecommunications. He is a fellow of AAIA and IET. He has been serving as the Vice Director for the Smart Transportation Information Engineering Society, Chinese Institute of Electronics, since 2022, and was the Co-Chair for IEEE ITS Technical Committee on Communication Networks (2014–2017). He has been an Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS since 2018, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2014–2019), and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2010–2020); and received "Top Editor" award for outstanding contributions to IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2011, 2014, and 2015. He has served as the chair, the co-chair, and a TPC member in a number of international conferences.