IEEE ComSoc IEEE Open Journal of the Communications Society

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version 11 January, 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.011100

Attention-based Blind Adaptive Receive Beamforming for Interference Limited NGSO Satellite Systems

Almoatssimbillah Saifaldawla *(Graduate Student Member, IEEE)*, **Flor Ortiz** *(Member, IEEE)*, **Eva Lagunas** *(Senior Member, IEEE)*, and **Symeon Chatzinotas** *(Fellow, IEEE)*

[1]SnT–Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, 1855 Luxembourg City, Luxembourg

CORRESPONDING AUTHOR: Almoatssimbillah Saifaldawla (e-mail: moatssim.saifaldawla@uni.lu).

This research was funded by the Luxembourg National Research Fund (FNR) under the project SmartSpace (C21/IS/16193290).

ABSTRACT

This paper presents an interference mitigation framework that can be applied on the user side for Non-Geostationary Satellite Orbit (NGSO) systems that share adjacent, overlapped frequencies to prevent unintentional co-frequency interference (CFI) scenarios. We introduce a novel Attention-based Beamformer (AttBF) model and explore its blind adaptive beamforming capabilities at the user terminals (UTs) side for spatial NGSO-to-NGSO downlink interference nulling, utilizing estimation-free data (e.g., received time-domain signals, frequency-domain representations, and sample covariance matrices (SCMs)) as direct inputs. We present a comprehensive performance evaluation of the proposed AttBF model against traditional deep learning (DL) models across various interference scenarios, encompassing both low spatial correlation (at UT's side-lobe) and high spatial correlation (at UT's main-lobe). To facilitate this research and future investigations into the interference management of NGSO systems, we implement innovative and extensive realistic satellite orbiting simulation and data generation methodologies, introducing new open datasets for the community. The results demonstrate that the proposed AttBF-based beamformer, particularly when employing SCMs input, achieves superior performance in mitigating interference compared to time- and frequency-domain inputs. Our findings highlight the enhanced nulling capabilities of the AttBF-based approach compared to DL-based models, such as convolutional neural networks (CNNs), and traditional methods, including zero forcing beamformer (ZFBF) and sample matrix inversion (SMI), underscoring the potential of advanced DL techniques for improving the reliability and efficiency of NGSO systems.

INDEX TERMS NGSO interference mitigation, Adaptive beamforming (ABF), Transformer-encoder.

I. Introduction

NON-GEOSTATIONARY satellite orbits (NGSOs), particularly the rapidly growing low-Earth orbits (LEOs) mega-constellations, stand as the cornerstone of the global communication ecosystem led by Non-Terrestrial Networks (NTN), their importance constantly amplified by the ever-increasing demand for higher data throughput and the imperative for efficient spectrum utilization [1]. Given the congestion in the lower S and C bands, NGSO systems are increasingly leveraging higher frequency bands, including Ku, K, and Ka, to meet growing bandwidth demands [2]. These systems must manage co-existence strategies for sharing the limited spectrum [3]. Despite proactive efforts by the International Telecommunication Union (ITU) to manage and regulate radio emissions worldwide, the potential for unintentional co-frequency interference (CFI) remains a significant concern among NGSO operators [4]. Thus, advanced interference avoidance and mitigation techniques have become key points in the future NGSO ecosystem.

A. Background

For many years, satellite user terminals (UTs) have relied on dish-type antennas that pointed to a fixed geostationary satellite in the sky [5]. With the advent of mega-LEO constellations, planar antenna arrays are gaining popularity due to their ability to electronically - without mechanical steering parts - steer the receiver main beam towards the moving satellite line of sight (LoS) [6]. This capability, known as beamforming, is a signal processing technique that leverages antenna arrays to spatially direct transmission or reception patterns [7]. By adjusting the complex weighting coefficients of each array element, BF enables precise beam control. Adaptive beamforming techniques (ABF) are those that allow for dynamically optimizing and tuning the antenna weights in real-time. In particular, ABF-based techniques are helpful in interference scenarios, as they can enhance the quality of the received signals, not only by focusing the main beam toward the desired direction of arrivals (DOAs), but they can also place low antenna gain values at unwanted DOAs [8].

Conventional blind ABF methods, such as the Capon beamformer, also known as the Minimum Variance Distortionless Response (MVDR) [9], exploit the inherent statistical properties of the desired signals or the knowledge of the covariance matrix of the received signal. The covariance matrix is typically estimated using a limited number of snapshots "samples" of the received signal, and the observed matrix is called the sample covariance matrix (SCM); moreover, the traditional MVDR solution is replaced by a sample matrix inversion (SMI) algorithm. However, the SMI method requires a large number of signal snapshots to correctly estimate the original covariance matrix and null interference locations, which is not desirable in fast-moving NGSO systems due to limited snapshots per coherent NGSO channel duration. Furthermore, the iterative nature and computational complexity in calculating the array weights (i.e, due to matrix inversion operations) are their main drawbacks [10]. Statistical ABF techniques, such as zero-forcing beamforming (ZFBF), provide optimal CFI interference nulling performance under perfect and full channel state information (CSI) knowledge, i.e., accurately knowing desired and interference DOAs and transmission channel characteristics. However, they usually rely on channel estimation for both desired and interference channels, which is typically difficult to achieve, especially in LEO systems, due to many factors such as: 1) fast time-variation of the LEO channels and the associated Doppler shifts [11], 2) high feedback overhead [12], 3) high computational complexity [13], and 4) imperfect synchronization & unknown interference sources [14]. This motivates the need for blind (estimation-free) beamforming solutions.

In response to these challenges, deep learning (DL)-based beamforming techniques have emerged as a potential solution to the drawbacks of traditional beamforming techniques [10]. DL-based beamforming typically replaces or augments classical weight calculation with a neural network that directly outputs beamforming weights [15]–[17]. Despite the significant time required to train a DL model in a satellite communications (SatCom) scenario, its fast temporal response (inference time) can be exploited for adaptive beamforming. After training, the DL model can perform complex operations in real-time, representing a promising solution for adaptive beamforming, where weights must be repeatedly calculated and conventional solutions are time-consuming.

B. Relevant Work

Achieving optimal beamforming in NGSOs presents more complex challenges compared to terrestrial network systems (TNs), including the need to adapt to highly dynamic environments characterized by moving satellites and UTs, the complex task of managing interference from various sources, and the continuous pursuit of enhanced performance under different operational constraints [18]. The satellite channel itself introduces impairments such as significant path losses, propagation delays, and Doppler shifts, further complicating the implementation of effective beamforming strategies [19]. Some state-of-the-art conventional ABF methods have been applied in the SatCom domain. In [20], and to mitigate the CFI of NGSOs, the authors propose a beamforming and power control-aided interference mitigation (BFPC-IM) scheme for uplink and downlink scenarios. In [21], the authors propose an iterative CFI null algorithm for NGSO systems that addresses an NP-hard unit modulus least squares (UMLS) problem. These state-of-the-art approaches, among others, have an iterative manner and usually assume prior knowledge of the desired steering angle and/or the spatial locations of the interferer, which is typically not the case in CFI scenarios, where no information about the interference locations is available.

Considering that interferers' DOAs are unknown but the received signal carries an impact caused by interferers, DL-based beamforming approaches have been proposed for estimating the array weight vector to steer the antenna array towards the desired signal while nulling interference sources in TNs [22] and signal processing applications [23]. In [22], the authors introduced a deep neural network beamforming (NNBF) model for the TNs uplink sum-rate maximization problem (SRM). Here, the perfect and full CSI was used as input to the NNBF, but in the coexisting CFI case, only the channel component linked to the desired transmitter can be estimated. Furthermore, such a model might be sensitive to channel estimation errors. In [23], the SCM is used as input to a convolutional neural network beamformer (CNNBF). However, the approach in [23] assumes knowledge of the desired beamformer weights. Both [22] and [23] are based on a supervised learning approach, which requires input-output pairs by definition, and obtaining such labelled data can be challenging due to the unpredictable environments of NGSO systems, making the ground truth difficult to determine.

The Attention mechanism, introduced by [24], has been crucial in the development of Transformer-based large language model applications (LLM), which are now the backbone of many AI applications. Methods like convolution, dropout, and batch normalization have been essential for feature extraction and learning in DL models. The Attention mechanism provides a significant degree of interpretability. In particular, the model assigns attention weights to the features of the input that shall gather more focus, offering insights into the model's decision-making process. Transformer-based models have been introduced in the SatCom domain for various tasks, such as resource management [25], interference detection [14], and precoding [26]. In this paper, we propose a novel self-supervised Attention-based beamformer (AttBF) model, which can be deployed on planar antenna arrays at the UT side. It does not rely on any CSI information, and it only uses the SCM of the received signal as an input. The results obtained are encouraging and highlight the potential of advanced GenAI-based techniques in improving conventional beamforming designs in NGSO systems.

C. Contribution

The key contributions of this paper in the SatCom field, particularly in the area of NGSOs coexisting interference nulling via receive beamforming, can be stated as follows,

- We design and develop a novel AttBF-based adaptive beamformer for NGSOs-to-NGSO interference mitigation. We investigate the AttBF model capabilities to point towards the desired satellite while nulling interference locations without the need for any CSI. In particular, our proposed model works with the received signal's SCM, which is calculated within a coherence duration of an NGSO satellite pass.
- The proposed AttBF is a self-trained model, meaning that the learning is done based on Signal-to-Interference-plus-Noise Ratio (SINR) feedback measured at the UT output. Once it is trained, it is deployed in the customer UT, such deployment of the already-trained model ensures fast inference capable of working with short coherence duration and with UTs' power-limited devices.
- Due to the unavailability of real-world data, we implement extensive simulation and data generation methodologies to emulate a realistic NGSO scenario. We test the proposed blind beamformer method, considering interferences from different NGSO satellite orbits affecting the NGSO satellite user. Moreover, all datasets produced in this work are made publicly available at [27] to ensure research reproducibility.
- We present an evaluation of the proposed AttBF model against other beamforming methods for different interference scenarios (e.g., low spatial correlation in side-lobe scenarios and high spatial correlation main-lobe scenarios). In particular, we compare our proposed method with DL-based models [22], [23] and traditional non-learning methods such as ZFBF and SMI.
- The findings reveal that the proposed AttBF-based adaptive beamformer exhibits high performance in mitigating spatial interference instantaneously at the primary and side lobes of the UT, demonstrating superior null capabilities compared to other DL-based models and traditional methods.

D. Notation and Organization

In this study, we adopt the following notation. We indicate matrices in bold uppercase letters $\mathbf{X}$ and column vectors in bold lowercase letters $\mathbf{x}$. $\mathbf{0}$ is a zero column vector and I_M is the identity matrix $M \times M$. We use I when the size of the matrix can be understood by the context. We used $\|\cdot\|$ to indicate the Frobenius norm of the matrix. We use the superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ to indicate the transpose, the Hermitian transpose, and the inverse of the matrix, respectively. $\mathbb{R}^M$ indicates the M-dimensional real space and $\mathbb{C}^M$ indicates the M-dimensional complex space. $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote real and imaginary parts, respectively. $\mathbb{E}_y(\mathbf{x})$ indicates the expected value of $\mathbf{x}$ computed with respect to the probability distribution of y. ∇f denotes the gradient of f. The complex Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{CN}(\mu, \sigma^2)$ and an analogous notation follows for the multivariate complex Gaussian distribution. The time computational complexity is expressed in Big-O notation, denoted as $\mathcal{O}(\cdot)$.

The remainder of this paper is organized as follows. Section II. outlines the proposed system model. Section III. describes the proposed AttBF-based model with SCMs input and its data preprocessing. Section IV. discusses the simulation setup for NGSO-to-NGSO interference scenarios and data generation, analyzes the results obtained from training the models as well as the interference mitigation performance, and provides a comprehensive comparison between the AttBF-based model and traditional approaches. The paper concludes in Section V, which provides a summary, remarks, and suggested future improvements.

II. System Model

This section outlines the interference scenario between NGSO systems, along with the system model and link budget assumptions. It highlights the channel model, the signal model, and the concept of adaptive receive beamforming.

A. Interference Scenario

In Fig. 1 we illustrate the interference scenario, where an NGSO satellite system (green satellite icon) communicating with its UT may eventually and unintentionally be interfered with by other NGSO systems (red satellite icons), more specifically any other NGSO satellites that uses the same frequency bands and falls within the UT field of view (spherical cap) defined by a minimum elevation ($\theta_{\min}$) with respect to the local horizon of the UT, as shown in Fig. 1.

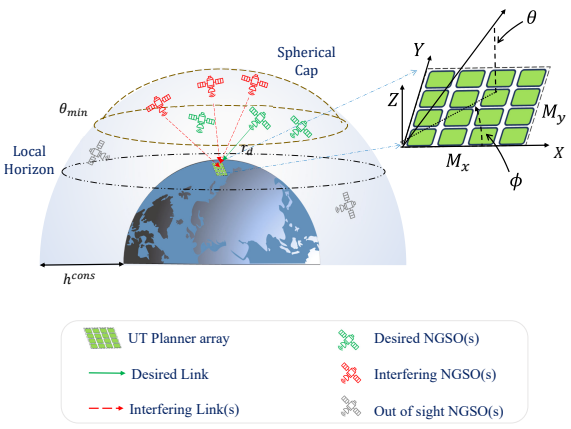

FIGURE 1. Considered scenario with a desired NGSO system (green) and an interfering NGSO system (red).

Our scenario does not account for out-of-sight satellites (gray satellite icons) or those with different frequency bands within the spherical cap because they do not cause any significant CFI. Considering this scenario, we investigate a dynamic interference mitigation solution based on received beamforming that can be deployed on the UT side of the victim NGSO system.

B. Channel Model

In this paper, we assume that the UT is equipped with a uniform planar array (UPA), as shown in Fig. 1. We consider a scenario in which the beam center of the desired satellite transmitters and interference satellite transmitters continuously track the UT location, resulting in worst-case interference scenarios at the UT. This is typically the case for many NGSO systems, where the satellite spot beams are Earth-fixed, meaning that they do not move during the satellite pass. We are only interested in the receiving beamforming at the UT side, and we focus on the downlink from satellites to the UT. By adopting the UPA, we aim to perform beamforming and achieve a high SINR.

Consider the UPA in the UT lying in the xy-plane with $M_x \times M_y$ antenna elements on the x and y axes, respectively, and the total number of antenna elements (antenna size) $M = M_x M_y$. Therefore, the UT array response vector can be written as [28],

$$\mathbf{v}(\phi,\theta) = \frac{1}{\sqrt{M}}\Big[1,\cdots,e^{j\frac{2\pi}{\lambda}d(m_x \sin\phi\cos\theta + m_y \sin\theta)},\cdots, e^{j\frac{2\pi}{\lambda}d((M_x-1)\sin\phi\cos\theta+(M_y-1)\sin\theta)}\Big]^T \in \mathbb{C}^M, \tag{1}$$

where d is the inter-element spacing, $0 \le m_x < M_x$ and $0 \le m_y < M_y$ are the antenna indices in the xy plane, ϕ and θ are the azimuth and elevation angles, respectively, representing the DoAs of the desired satellite signal, and $\lambda = c/f_c$ is the carrier wavelength associated with the frequency carrier f_c. Note that the UT array response vector can be modeled as a plane wave; thus, for notational brevity, we define the pair angles of direction of arrival (DoA) as $\varphi \triangleq (\phi,\theta)$, i.e., the DOAs for the desired satellite (DS) are $\varphi_d \triangleq (\phi_d,\theta_d)$.

We are modeling a scenario where a DS transmits a signal ($s_d \in \mathbb{C}$) with $\mathbb{E}\{s_d s_d^H\} = 1$ to its UT, eventually, $\mathcal{K}(\mathcal{K} < M)$ NGSO satellites interfere with the desired downlink ($\text{IS}_k, \quad k \in 1,\cdots,\mathcal{K}$) and unintentionally transmit an interference signal ($s_{i,k} \in \mathbb{C}$) towards it. The number of main beam-pointing and nulling points should not exceed the number of antenna elements that satisfy the condition $(\mathcal{K}+1 < M)$. A Rician fading channel can be assumed for NGSO communications [29]; however, we assume that the UT is placed in an obstacle-free elevated space, as recommended by many operators (e.g., STARLINK [6]), the LoS component becomes dominant, thus the non-LoS (NLoS) component can be ignored. Doppler shift may introduce rapidly changing phase rotations in the received samples [30], thus, for the sake of simplicity, we assume that the DS Doppler shift is compensated for at the UT using system information (SI) [31]. We define the statistical time-varying channel state information (sCSI) model between the DS satellite and the UT at time instant t and frequency $f_{c,d}$ as follows,

$$\mathbf{h}_d = \chi_d \cdot \mathbf{v}_d(\varphi_d), \tag{2}$$

where χ_m is the desired channel gain, given by,

$$\chi_d = \sqrt{\frac{P_d^{sat} G_d^{ut}(\varphi_d)}{L_d}}, \tag{3}$$

where P_d^{sat} is the DS Equivalent Isotropic Radiated Power (EIRP), and $G_d^{ut}(\varphi_d)$ is the UT receives gain in the DS direction. L_d is the free-space path loss given by,

$$L_d[\text{dB}] = 20\log_{10} r_d + 20\log_{10} f_c + 20\log_{10}\frac{4\pi}{c}, \tag{4}$$

where r_d is the slant range between UT and DS and it can be expressed in terms of elevation angle as,

$$r_d = \Big[(R_E + h^{\text{DScons}})^2 - (R_E + h^{\text{ut}})^2 \cdot \cos^2(\theta_d)\Big]^{\frac{1}{2}} - (R_E + h^{\text{ut}}) \cdot \sin(\theta_d), \tag{5}$$

where R_E denotes the radius of the Earth ($R_E = 6371$km), h^{ut} is the altitude of the UT and h^{DScons} is the altitude of the desired constellation with respect to the Earth's surface, as shown in Fig. 1, the expression $(R_E + h^{\text{DScons}})$ is also known as the semi-major axis SM_d. The effective channel vector for the k^{th} interfering satellite IS_k to the UT receiver can be derived as follows,

$$\mathbf{h}_{i,k} = \chi_{i,k} \cdot \mathbf{v}_k(\varphi_k), \tag{6}$$

and,

$$\chi_{i,k} = \sqrt{\frac{P_k^{sat} G_k^{ut}(\varphi_k)}{L_k}}, \tag{7}$$

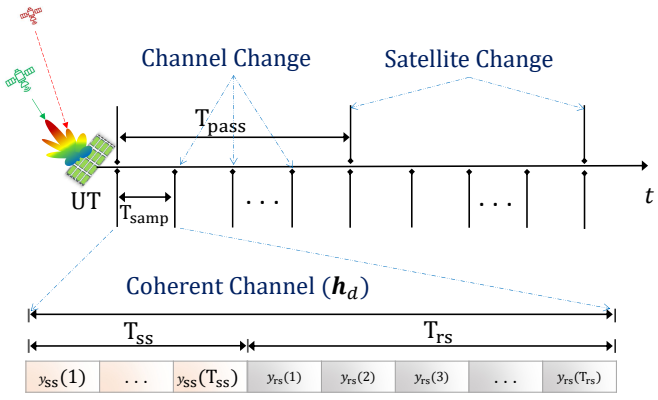

FIGURE 2. Time frame of the simulation scenarios

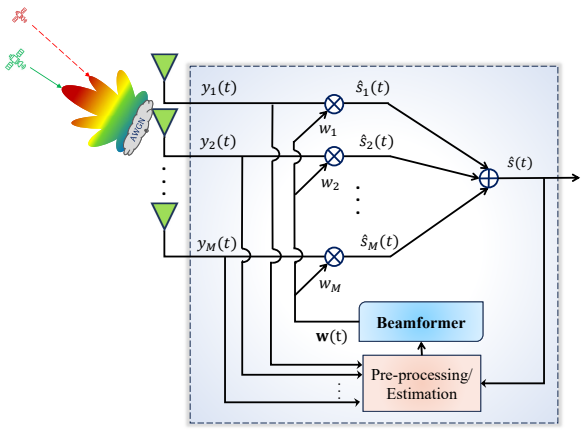

FIGURE 3. Typical structure of ABF

where P_k^{sat} is the IS transmit EIRP, and $G_k^{ut}(\varphi_k)$ is the UT receipt gain in k^{th} satellite IS direction. Similarly to (4) and (5), we can calculate the FSPL from the k^{th} interfering satellite, using the slant range between the UT and IS_k denoted as $r_{i,k}$. The UT array receiving gain $G^{ut}(\varphi)$, at any DOAs φ can be obtained as follows [32],

$$G^{ut}(\varphi) = \eta^{ut}\Xi^{ut}(\varphi). \tag{8}$$

where η^{ut}, Ξ^{ut} are the UPA radiation efficiency and the directivity towards the φ direction. Directivity is the ratio of the radiation intensity in a given direction to the radiation intensity averaged in all directions, and can easily be calculated from the UPA array factor [33]. Intuitively speaking, the directivity should be as low as possible at interferences' DOA, leading to a gain value closer to zero (Ξ^{ut} at null locations $approx 0$). Note that the channel model adopted in (2) and (6) is applicable over a specific coherence time T_{samp}, as shown in Fig. 2, where the relative positions of the satellites and UT do not change significantly, and the physical channel parameters can be considered time invariant. In the next section, we explain the signal model for the entire pass of the desired satellite.

C. Signal Model

Let us assume the DS pass of duration T_{pass}, as shown in Fig. 2, which may encounter interference from the IS satellite. The satellite pass is divided into shorter time periods of duration T_{samp}, which corresponds to the channel coherence time. We assume that the UT array captures a total of T_{samp} snapshots within this coherence time. However, the UT must update the receive beamformer with $T_{\text{ss}} << T_{\text{samp}}$ so that the resulting beamforming weights can be applied for the remaining coherence time $T_{\text{rs}} = T_{\text{samp}} - T_{\text{ss}}$.

The minimum value of T_{ss} that provides enough statistical information about the received signal is said to be at least M independent snapshots [34]. However, in many research articles, it is recommended to use at least double that, $2M$ snapshots, to achieve better beamforming results [35].

Let us now focus on a specific snapshot at a time instance t. In the presence of a single desired satellite and $\mathcal{K}$ interfering satellites, the received signal at the UT planar antenna array at a time instance t can be expressed as,

$$\mathbf{y}(t) = \mathbf{h}_d s_d(t) + \sum_{k=1}^{\mathcal{K}} \mathbf{h}_{i,k} s_{i,k}(t) + \mathbf{z}(t), \tag{9}$$

where $\mathbf{z}(t)$ is the additive white Gaussian noise (AWGN) at the UT receiver $\mathbf{z}(t) \in \mathbb{C}^M$ is i.i.d. $\mathcal{CN}(0, \sigma_z^2)$, with $\sigma_z^2 = \kappa T^{ut} BW_d$ where κ, T^{ut}, and BW_d are the Boltzmann constant, equivalent noise temperature of the UT, and DS signal bandwidth, respectively. Finally, $s_d(t)$ and $s_{i,k}(t)$ denote the independent information symbols transmitted from DS and ISs, respectively, whose average power is assumed to be normalized to one.

For the convenience of mathematical modeling, we will make use of the following matrix notation,

$$\mathbf{H}_i \triangleq [\mathbf{h}_{i,1}\ \mathbf{h}_{i,2}\ \ldots\ \mathbf{h}_{i,\mathcal{K}}] \in \mathbb{C}^{M\times\mathcal{K}}, \tag{10}$$

$$\mathbf{H} \triangleq [\mathbf{h}_d\ \mathbf{H}_i] \in \mathbb{C}^{M\times(\mathcal{K}+1)}, \tag{11}$$

$$\mathbf{S}_i(t) \triangleq [s_{i,1}(t)\ s_{i,2}(t)\ \ldots\ s_{i,\mathcal{K}}(t)] \in \mathbb{C}^{\mathcal{K}\times 1}. \tag{12}$$

D. Traditional Receive Beamforming Techniques

At the front-end of the UT receiver, a beamformer *weight* vector $\mathbf{w} \in \mathbb{C}^{M\times 1}$ is applied to steer the UT main beam towards φ_d direction and eliminating as much interference as possible from φ_k, $k = 1, \ldots, \mathcal{K}$ directions. The estimated signal after applying $\mathbf{w}$ is denoted as $\hat{s} \in \mathbb{C}$ and is given as,

$$\begin{aligned}\hat{s}(t) &= \mathbf{w}^H \mathbf{y}(t) \\ &= \mathbf{w}^H \mathbf{h}_d s_d(t) + \mathbf{w}^H \mathbf{H}_i \mathbf{S}_i(t) + \mathbf{w}^H \mathbf{z}(t).\end{aligned} \tag{13}$$

Based on the above, the SINR after the receive beamforming procedure is expressed as,

$$\gamma_{\text{out}} = \frac{\mathbf{w}^H \mathbf{R}_d \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{i+z} \mathbf{w}} = \frac{\mathbf{w}^H (\mathbf{h}_d \mathbf{h}_d^H) \mathbf{w}}{\mathbf{w}^H (\mathbf{H}_i \mathbf{H}_i^H + \sigma_z^2 \mathbf{I}_M) \mathbf{w}} = \frac{|\mathbf{w}^H \mathbf{h}_d|^2}{\sum_{k=1}^{\mathcal{K}} |\mathbf{w}^H \mathbf{h}_{i,k}|^2 + \sigma_z^2 \mathbf{w}^H \mathbf{w}}, \tag{14}$$

where $\mathbf{R}_d = \mathbf{h}_d \mathbf{h}_d^H \in \mathbb{C}^{M \times M}$ and $\mathbf{R}_{i+z} = \mathbf{H}_i \mathbf{H}_i^H + \sigma_z^2 \mathbf{I}_M \in \mathbb{C}^{M \times M}$ are the covariance matrix of the desired signal and the interference-plus-noise covariance matrix, respectively. Note that we dropped the time index "t" here and in the next parts for simplicity.

The problem of maximizing (14) without knowledge of the interference DOA is mathematically equivalent to the MVDR beamforming problem [36]. In particular, the MVDR beamformer minimizes UT array input power while maintaining unity gain in DS direction,

$$\mathbf{w}_{\text{MVDR}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{v}_d = 1, \tag{15}$$

where $\mathbf{v}_d$ has been used to denote $\mathbf{v}_d(\varphi_d)$ in a shorter way. The optimization problem in (15) leads to the following weights for the beamformer,

$$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{R}^{-1} \mathbf{v}_d}{\mathbf{v}_d^H \mathbf{R}^{-1} \mathbf{v}_d}. \tag{16}$$

Clearly, the covariance matrix $\mathbf{R} = \mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right]$ serves as a fundamental tool in beamforming design, providing a representation of the statistical relationships between the signals received at different elements of an antenna array. It contains information on the power of the signals received at each antenna element and the correlation between the signals received at different pairs of elements. This correlation structure is indicative of the spatial characteristics of the incoming waveforms, including the DOAs of desired and interfering signals. By analyzing the covariance matrix, beamforming algorithms, such as MVDR, can learn insights about the spatial distribution of signal sources and noise, which is essential for designing effective beamforming weights that enhance the reception of the desired signal while suppressing interference. Traditionally, $\mathbf{R}$ is not available and must be estimated from the receive signal by averaging the outer product of the receive signal vector over a finite number of time samples (L), often referred to as *snapshots* [8]. The covariance matrix can be estimated using the UPA snapshots by simply calculating the SCM matrix denoted as $\hat{\mathbf{R}}$, which is also called the autocorrelation matrix, and it is given as follows,

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{y}[l] \mathbf{y}^H[l] = \frac{1}{L} \mathbf{Y} \mathbf{Y}^H, \tag{17}$$

where $\mathbf{Y} \triangleq [\mathbf{y}[1]; \mathbf{y}[2]; \cdots; \mathbf{y}[L]] \in \mathbb{C}^{M \times L}$. When the SCM is used in the MVDR expression, the resulting beamformer is called the SMI beamformer, which solves the following minimum power distortionless response (MPDR) problem [9],

$$\mathbf{w}_{\text{SMI}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{v}_d = 1, \tag{18}$$

where $\mathbf{w}_{\text{SMI}}$ is the *SMI*-based minimum variance beamformer given by,

$$\mathbf{w}_{\text{SMI}} = \frac{\hat{\mathbf{R}}^{-1} \mathbf{v}_d}{\mathbf{v}_d^H \hat{\mathbf{R}}^{-1} \mathbf{v}_d}. \tag{19}$$

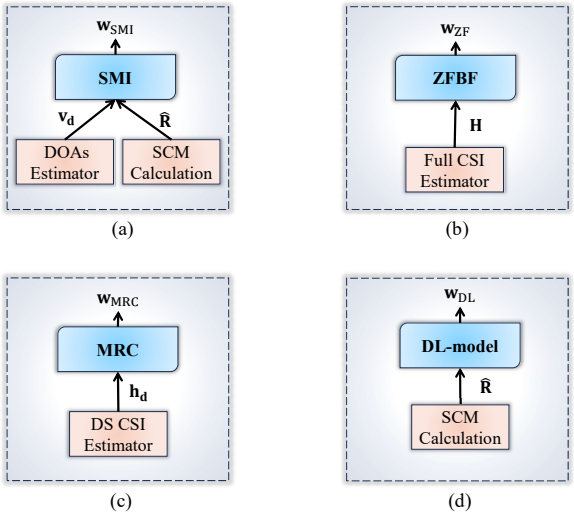

FIGURE 4. Receive Beamforming Techniques: (a) sample matrix inversion (SMI), (b) Zero forcing beamforming (ZFBF), (c) Maximum ratio combining (MRC), and (d) The intended DL-based beamformer

Remark 1. *As L increases, the SCM $\hat{\mathbf{R}}$ converges to the theoretical covariance matrix $\mathbf{R} = \mathbf{R}_d + \mathbf{R}_{i+z}$ and the corresponding* SINR *approximates the optimal value as $L \to \infty$. However, as the size of the available snapshots L decreases, the gap between $\hat{\mathbf{R}}$ and $\mathbf{R}$ increases, dramatically affecting the performance of SMI.*

In Fig. 4, we illustrate the traditional benchmark methods used in this paper, including the SMI beamformer Fig. 4(a), alongside the intended DL-based beamformer, as shown in this figure. In addition to the adaptive SMI beamformer $\mathbf{w}_{\text{SMI}}$, we will compare our DL-approach results with two famous statistical CSI-based beamforming benchmarks that are detailed below.

1) Zero forcing beamformer (ZFBF): Illustrated in Fig. 4(b), ZFBF represents the optimal/near-optimal scenario where UT can place a null towards DOAs of the interfering satellites φ_i. Clearly, ZFBF requires perfect full CSI estimation that includes both DS and ISs DOAs. In other words, the ZFBF uses the pseudoinverse of $\mathbf{H}$ as,

$$\mathbf{w}_{\text{ZF}} = \frac{\mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \boldsymbol{\beta}}{\|\mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \boldsymbol{\beta}\|}. \tag{20}$$

where $\boldsymbol{\beta} = [1; 0; \cdots; 0]^T \in \mathbb{N}^{\mathcal{K}+1}$.

2) Maximum ratio combining (MRC) beamformer: Illustrated in Fig. 4(c), MRC represents the worst-case nulling scenario, where the UT has only access to its own channel estimate corresponding to the desired DOA φ_d and thus focuses the UT main beam towards the DS satellite, without generating nulls towards DOAs of ISs satellites. Assuming perfect knowledge of $\mathbf{h}_d$, MRC is simply derived as,

$$\mathbf{w}_{\text{MRC}} = \frac{\mathbf{h}_d}{\|\mathbf{h}_d\|}. \tag{21}$$

MRC and ZFBF are sensitive to imperfect channel estimates (imCSI). Since we plan to evaluate the impact of the imCSI, we provide the model of CSI estimation errors herein.

The estimation error of the estimated DS channel $\hat{\mathbf{h}}_d$ is calculated as $\mathbf{e}_{d,\text{mrc}} = \hat{\mathbf{h}}_d - \mathbf{h}_d$ for MRC. When considering both the DS and the ISs channels, we model the error as $\mathbf{e}_{d,\text{zf}} = \hat{\mathbf{H}} - \mathbf{H}$, which will be used for ZFBF. Each error can be modeled as a complex Gaussian random variable $\mathbf{e}_d \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2 \mathbf{I}_M)$. Note that perfect CSI (pCSI) is achieved for $\mathbf{e}_d = \mathbf{0}$. All beamformers $\mathbf{w}$ in this study are already or will be normalized, including SMI in (19), so that consistent and controlled beamforming weights are achieved independently of channel gains. In practice, UT might be able to estimate the CSI of the DS link, i.e., $\mathbf{h}_d$, using pilot signals, or at least have prior knowledge of the DS DOAs (to generate the steering vector $\mathbf{v}_d$), but cannot access any information about the ISs links. Estimation of DOA is usually performed using the MUSIC algorithm, where, in this case, the number of signal sources (desired and interference sources $\mathcal{K}+1$) is required. Furthermore, estimation stages and matrix inverse operations (e.g., SMI, ZFBF) add processing time and complexity to the UT design, making the temporal response of fast-moving NGSO satellites a critical concern.

In this paper, we introduce a DL-based model that relies only on the available snapshots, specifically taking the SCM as input to instantly generate good beamformer weights, as illustrated in Fig. 4(d). The DL-based model does not depend on any CSI or is sensitive to its estimation errors, and provides acceptable performance in low and highly spatially correlated scenarios (e.g., in-line interference events).

III. Proposed Attention-Based Beamforming

In this paper, we introduce a novel Attention-based beamformer DL-model, denoted as AttBF. The Attention mechanism, introduced by [24], is the main component behind our proposed AttBF model. This mechanism has been crucial in the development of Transformer-based LLM applications. In the following subsections, we provide a detailed explanation of the design of the proposed AttBF model and the processing carried out in its individual steps, as illustrated in Fig. 5. We also explain the training procedure, then outline other DL-based beamforming methods in the literature that will be used as benchmarks and their computational time complexity.

A. Data Pre-processing

Our proposed AttBF model does not use the raw time-domain signal snapshot matrix $\mathbf{Y}(t)$ directly, so before feeding the proposed model with data, a few preprocessing steps are performed.

1) Sample Covariance Calculation

First, during the coherent duration of a satellite pass, the initial snapshots captured by UT during $L = T_{\text{ss}} << T_{\text{samp}}$ as in Fig. 2, and calculate the SCM, as in (17), more particularly,

$$\hat{\mathbf{R}} = \frac{1}{T_{ss}} \sum_{l=1}^{T_{ss}} \mathbf{y_{ss}}[l] \mathbf{y_{ss}}^H[l] = \frac{1}{T_{ss}} \mathbf{Y_{ss}} \mathbf{Y_{ss}}^H, \tag{22}$$

where $\mathbf{Y}_{ss} \triangleq [\mathbf{y}[1]; \mathbf{y}[2]; \cdots; \mathbf{y}[T_{ss}]] \in \mathbb{C}^{M \times T_{ss}}$. The AttBF model requires the SCMs' data to work in an adaptive way. SCMs represent spatial relationships and correlations within the received signal snapshots.

2) Data Tokenization

We define a tokenization operator,

$$\text{Tok}(\cdot) : x \mapsto [\text{Re}(x); \text{Im}(x)]^T, \tag{23}$$

we simply transform the complex-valued $\hat{\mathbf{R}}$ matrix to a real-valued matrix of tokens $\mathbf{\Gamma}_R \in \mathbb{R}^{M \times 2M}$ as

$$\begin{aligned} \mathbf{\Gamma}_R &= \text{Tok}\left(\hat{\mathbf{R}}\right) \\ &= \begin{bmatrix} \hat{r}_{1,1} & \hat{r}_{1,2} & \cdots & \hat{r}_{1,2M} \\ \hat{r}_{2,1} & \hat{r}_{2,2} & \cdots & \hat{r}_{2,2M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{M,1} & \hat{r}_{M,2} & \cdots & \hat{r}_{M,2M} \end{bmatrix}, \end{aligned} \tag{24}$$

where $\hat{r}_{i,j} \in \mathbb{R}$, each row in $\mathbf{\Gamma}_R$ represents a single input token (in this case an embedded vector $\in \mathbb{R}^{1 \times 2M}$), and M represents the total tokens. Each token corresponds to a row in the matrix, which we represent as a real-valued vector, meaning that there are always $2M$ features per token corresponding to the real and imaginary components of SCM. Unlike LLMs, where tokens represent linguistic units, here, tokens represent spatial relationships encoded in the SCM. This preserves the structural information of the array data while converting it into a form compatible with transformer-based architectures. Effectively, this tokenization step bridges signal processing representations with deep sequence models, enabling the network to learn embeddings that adaptively capture interference patterns for robust beamforming.

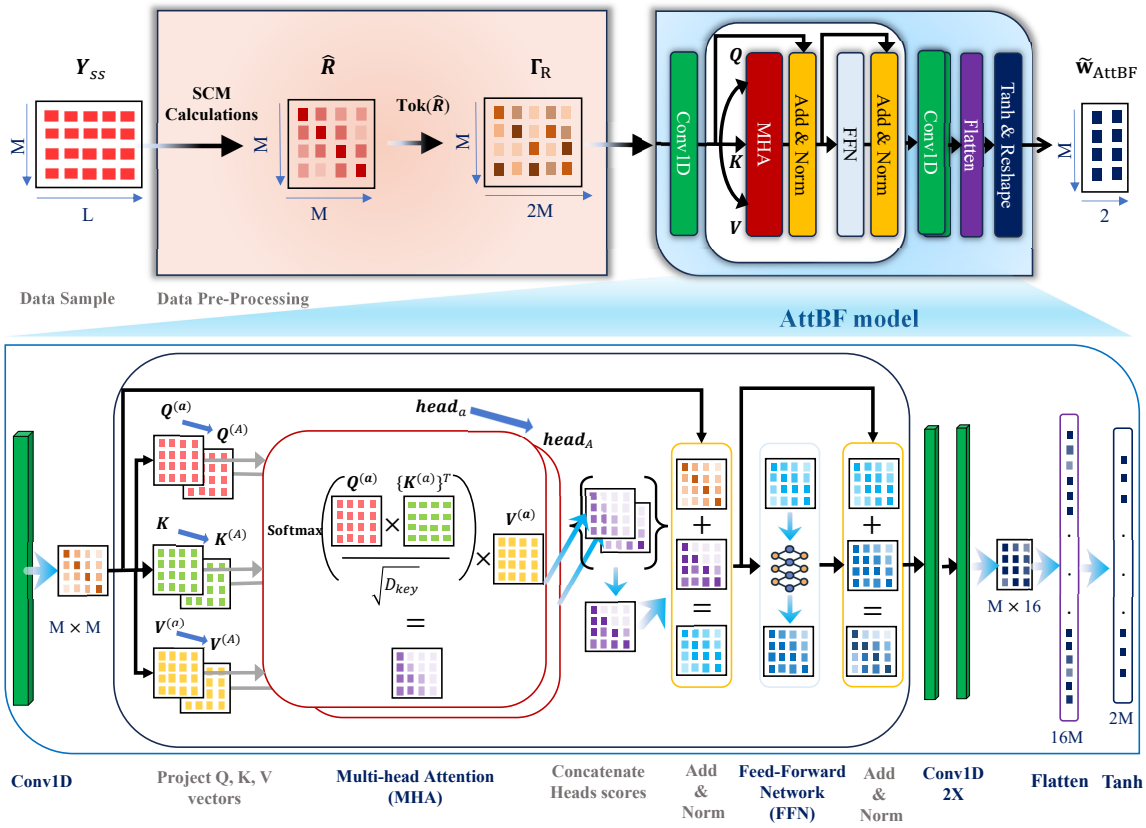

FIGURE 5. **Illustration of the proposed AttBF beamformer.**

B. AttBF Model Architecture

The architecture of the AttBF model is a sequence of DL-based layers and matrix operations that process real-valued SCM $\mathbf{\Gamma}_R$ through successive transformations, as illustrated in Fig. 5, to predict beamforming weights $\mathbf{w}_{\text{AttBF}} \in \mathbb{C}^M$.The Attention mechanism, forming the core Transformer computation, enables the AttBF model to exploit the spatial relationships within the SCM tokens, which are crucial for effective beamforming. Instead of processing everything equally, the Attention mechanism works more dynamically; it efficiently weighs each token's contribution in improving the model's beamforming performance by assigning weights to different parts of the input, and it then self-adjusts its learnable parameters to focus only on the token's parts that are most valuable to the model. We refer to the output of a DL layer as $\mathcal{X}_{\text{Layer Name}}$ and its learnable parameters as $\mathbf{\Theta}_{\text{Layer Name}}$.

1) Feature Extraction Layer

A single 1-dimensional (1D) convolution layer (Conv1D) processes the AttBF model input using M filters, a single kernel, and Gaussian Error Linear Unit (GELU) activation. This layer acts as a dense layer applied independently to each of the $2M$ features per token along the first spatial dimension, reducing it back to M features per token. This can aid in feature extraction (FE) across the real and imaginary components of input data for each antenna element, thereby increasing the computational efficiency of the MHA block.

2) Multi-Head Attention Block

The next part of the AttBF model is the Multi-Head Attention (MHA) block, which consists of $a \in \{1, 2, \cdots, A\}$ Attention heads with key dimensions corresponding to the number of features extracted in each token ($i.e., D_{key} = M$). First, MHA performs a linear projection on the FE output to produce for each head three matrices as:

- Query (Q) matrix: Represents what an antenna element is "looking for" or a query about its correlation with other elements.
- Key (K) matrix: Represents the information an antenna element "holds" or what it has to offer to a query.
- Value (V) matrix: Contains the actual content or features that should be propagated forward if the corresponding key matches a query.

These matrices are formed as $Q^{(a)} = \mathcal{X}_{\text{FE}}\mathbf{\Theta}_Q^{(a)}, K^{(a)} = \mathcal{X}_{\text{FE}}\mathbf{\Theta}_K^{(a)}$, and $V^{(a)} = \mathcal{X}_{\text{FE}}\mathbf{\Theta}_V^{(a)}$, where these projections

$$\text{head}_a = \text{Attention}(Q^{(a)}, K^{(a)}, V^{(a)}) = \text{Dropout}\left(\text{softmax}\left(\frac{Q^{(a)}{K^{(a)}}^\top}{\sqrt{D_{key}}}\right)\right)V^{(a)}. \tag{25}$$

are able to learn parameter matrices $\boldsymbol{\Theta}_Q^{(a)}, \boldsymbol{\Theta}_K^{(a)}, \boldsymbol{\Theta}_V^{(a)} \in \mathbb{R}^{D_{key}\times D_{key}}$. Each head performs a scaled dot-product Attention with Dropout as in (25) to compute its Attention Score.

In (25), for each token in the input, its Query matrix Q is multiplied by the Key matrix K of all other tokens (including itself). This produces a matrix of attention scores. A high score between a query from token i and a key from token j indicates a strong relationship. This is the mechanism by which the model "decides" which token features are most relevant to each other for the purpose of producing beamforming weights. For instance, it might learn that the correlations between certain token features are critical for identifying the desired signal or a strong interferer. The attention scores are then passed through a softmax function to normalize them into a probability distribution (where all weights sum to 1). These normalized weights are then multiplied by the Value matrix V. This step effectively creates a weighted sum of all Value matrices. Tokens with a high attention score (i.e., those deemed most important by the model) contribute more to the output of that attention head. The model thus "focuses" its attention on the most critical parts of the input SCM. For instance, if the SCM contains a strong interferer, the attention mechanism will learn to assign high weights to the correlations associated with that interferer, allowing the model to generate weights that nullify it.

The "multi-head" part of MHA means this entire process is performed multiple times in parallel, with each head using its own independent set of Q, K, and V projection matrices. Each attention head can learn to focus on different aspects of the tokenized SCM. For instance, one head might learn to focus on the desired signal's spatial signature, while another might learn to identify and focus on the spatial signature of a strong interferer. The outputs of all the individual heads are then concatenated and passed through a final linear projection layer as,

$$\begin{aligned}\mathcal{X}_{\text{MHA}} &= \text{MHA}(\mathcal{X}_{\text{FE}}) \\ &= \text{Concat}(\text{head}_1, \cdots, \text{head}_A)\boldsymbol{\Theta}_{\text{MHA}},\end{aligned} \tag{26}$$

where $\boldsymbol{\Theta}_{\text{MHA}}$ are the output projection matrices in the MHA operation, and are also a learnable parameter that enables dimensional alignment and feature fusion between Attention heads.

The MHA block is then followed by a residual connection, and a normalization layer (LayerNorm) to stabilize training and improve generalization by normalizing activations across the M feature dimension,

$$\mathcal{X}_{\text{LN1}} = \text{LayerNorm}\left(\mathcal{X}_{\text{FE}} + \mathcal{X}_{\text{MHA}}\right). \tag{27}$$

3) Feed-forward network (FFN)

The FFN consists of a two-layer transformation with linear and Rectified Linear Unit (ReLU) activation, further processing the features learned by the MHA block, allowing for non-linear transformations of these features as,

$$\mathcal{X}_{\text{FFN}} = (\text{ReLU}(\mathcal{X}_{\text{LN1}}\boldsymbol{\Theta}_{\text{F1}}))\boldsymbol{\Theta}_{\text{F2}}, \tag{28}$$

Afterwards a second residual connection and LayerNorm process layer follow to help stabilize the learning process,

$$\mathcal{X}_{\text{LN2}} = \text{LayerNorm}\left(\mathcal{X}_{\text{LN1}} + \mathcal{X}_{\text{FFN}}\right). \tag{29}$$

4) Beamforming Weights Refining

In the rest of the AttBF architecture, we apply some basic DL-based layers to refine and extract the features that define the beamforming weights. First, two successive Conv1D layers are applied to the output of the second LayerNorm, each with a single kernel, they progressively reduce the number of features and further refine the feature maps along their channel dimension, reducing the number of features from M to $\frac{M}{2}$ and then to 16. Here, the ReLU activation is chosen to introduce non-linearity to the AttBF model,

$$\mathcal{X}_{\text{CN}} = (\text{ReLU}(\mathcal{X}_{\text{LN2}} * \boldsymbol{\Theta}_{\text{C1}}))\boldsymbol{\Theta}_{\text{C2}}. \tag{30}$$

The output tensor is flattened (FL) before a final Dense layer; this flattened vector contains the processed features.

$$\mathcal{X}_{\text{FL}} = \text{vec}(\mathcal{X}_{\text{CN}}), \tag{31}$$

Finally, a Dense layer with $2M$ output units maps the learned features to the size of the complex beamforming weights. A hyperbolic tangent function (Tanh) activation function constrains the output values to the range $[-1, 1]$. The output is then reshaped to $(M \times 2)$, where the last dimension represents the real and imaginary components of the complex weight for each of the M antenna elements,

$$\mathcal{X}_{\text{out}} = \text{Tanh}(\mathcal{X}_{\text{FL}})\boldsymbol{\Theta}_{\text{out}}, \tag{32}$$

$$\tilde{\mathbf{w}}_{\text{AttBF}}^{(n)} = \text{Reshape}(\mathcal{X}_{\text{out}}). \in \mathbb{R}^{M\times 2} \tag{33}$$

The complex-valued beamforming weights $\mathbf{w}_{\text{AttBF}} \in \mathbb{C}^M$ can be constructed from the AttBF output matrix $\tilde{\mathbf{w}}_{\text{AttBF}}$ as,

$$\mathbf{w}_{\text{AttBF}}^{(n)} = \frac{\tilde{\mathbf{w}}_{\text{AttBF}}^{(n)}(:,1) + j\cdot\tilde{\mathbf{w}}_{\text{AttBF}}^{(n)}(:,2)}{\|\tilde{\mathbf{w}}_{\text{AttBF}}^{(n)}(:,1) + j\cdot\tilde{\mathbf{w}}_{\text{AttBF}}^{(n)}(:,2)\|}. \tag{34}$$

This final step preserves the phase and relative amplitude information needed for the beamformer weights.

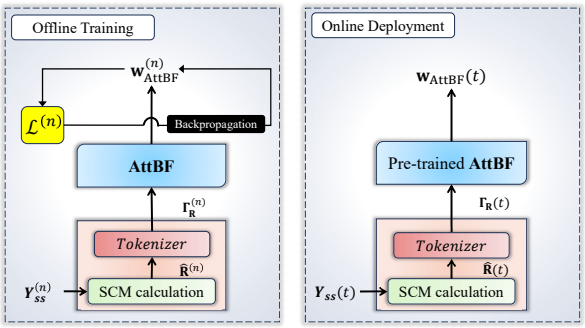

FIGURE 6. AttBF offline training, online deployment objectives.

C. AttBF Model Objective

The basic working concepts of the proposed AttBF design are illustrated in Fig. 6. The proposed learning procedure offers an offline, self-supervised training approach, or reward-like learning, as it learns without requiring explicit, human-labeled data. After training, the model can be used on new, unseen data to perform blind inference (Blind beamforming) without needing any prior knowledge or additional information about the environment. For the sake of notation specific to the proposed model, we add the index n as $\{\cdot\}^{(n)}$ to refer to the processed sample in the training datasets $n = 1, \ldots, \mathcal{N}_{Tr}$ or testing $n = 1, \ldots, \mathcal{N}_{Ts}$.

During the training phase, the model takes signal snapshots matrix $\mathbf{Y}_{ss}^{(n)}$, calculates the SCM $\hat{\mathbf{R}}_{ss}^{(n)}$, and transfers them into a DL-friendly form (i.e, tokens), then learns a beamformer design $\mathbf{w}_{\text{AttBF}}$ that is then fed back into the UT planar array to collect feedback from the environment in the form of a loss function $\mathcal{L}$, this loss function is customized to maximize SINR as,

$$\mathcal{L}_{\mathcal{B}} = -\frac{1}{\mathcal{B}} \sum_{n}^{\mathcal{B}} \gamma_{\text{out}}^{(n)}, \tag{35}$$

where $\mathcal{B}$ is the batch size. AttBF model modifies its learnable parameters according to the feedback of the loss function $\mathcal{L}_{\mathcal{B}}$ through the backpropagation process (Backprop).

Algorithm 1 summarizes the training process for the AttBF. First, the model parameters ($\mathbf{\Theta}_{\text{model}}$) are set to random values at the start. The model inputs a sample $\mathbf{Y}_{ss}^{(n)}$ from the processed mini batch of the training dataset $\mathcal{D}_{\text{Tr}}$, outputs the beamforming weights $\mathbf{w}_{\text{AttBF}}^{(n)}$ for that mini batch, and update the model parameter using the loss function feedback, the loss here is the negative SINR observation, and in this algorithm it is calculated per mini batch using $\mathbf{R}_{d}^{(n)}$ and $\mathbf{R}_{i+z}^{(n)}$ for each training sample. The gradients Gradients (∇_{Θ}) of the loss are calculated per mini batch; they are basically derivatives with respect to the model weights. An optimizer Ω uses these Gradients to update the model's parameters. This is the "learning" step where the model adjusts itself to produce a better beamformer in the next mini-batch. This loop continues until all mini-batches in an epoch are processed, and then it repeats for the next iteration ($epech$). Throughout training, the model keeps track of the best-performing set of parameters ($\mathbf{\Theta}_{\text{model}}^{(\text{best})}$), which is the final output of the training process after maximum iterations $E \rightarrow epoch$. The loss-per-epoch tensor $\mathbf{L}_{\text{epochs}}$ is used later on in the results section to compare the learning process performance with other DL-based benchmarks.

Algorithm 1: The Proposed Model Training Phase.

Input: Training dataset $\mathcal{D}_{\text{Tr}}$, Model f_{Θ}, Optimizer Ω
Output: Optimized model parameters $\mathbf{\Theta}_{\text{best}}$

Initialize: Model parameters $\mathbf{\Theta}_{\text{model}}$, B mini batches $\mathcal{B} \in \mathcal{D}_{\text{Tr}}$
for $epoch = 1$ **to** E **do**
  $\mathcal{L}_{\text{total}} \leftarrow 0$
  for $n = 1 : |\mathcal{B}|$ **do**
    $\mathbf{Y}_{ss}^{(n)}, \mathbf{R}_{d}^{(n)}, \mathbf{R}_{i+z}^{(n)} \leftarrow \mathcal{D}_{\text{Tr}}\{\mathcal{B}\}$ // Inputs
    $\mathbf{\Gamma}_{R}^{(n)} \leftarrow \mathbf{Y}_{ss}^{(n)}$ // Pre-process (23)
    $\mathbf{w}_{\text{AttBF}}^{(n)} \leftarrow f_{\Theta}(\mathbf{\Gamma}_{R}^{(n)})$ // Model outputs
    $\gamma_{\text{out}}^{(n)} \leftarrow \gamma_{\text{out}}(\mathbf{w}_{\text{AttBF}}^{(n)}; \mathbf{R}_{d}^{(n)}; \mathbf{R}_{i+z}^{(n)})$ // Calculate SINR (14)
    $\gamma_{\mathcal{B}} \leftarrow \gamma_{\mathcal{B}} + \gamma_{\text{out}}^{(n)}$ // $\mathcal{B}$ SINR
  $\mathcal{L}_{\mathcal{B}} \leftarrow -\text{mean}(\gamma_{\mathcal{B}})$// $\mathcal{B}$ mean SINR (35)
  $\nabla_{\Theta} \leftarrow \text{Gradient}(\mathcal{L}_{\mathcal{B}}, \mathbf{\Theta}_{\text{model}})$ // Backprop
  $\mathbf{\Theta}_{\text{model}} \leftarrow \Omega(\Theta_{\text{model}}, \nabla_{\Theta})$ // Optimizer
  $\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{total}} + \mathcal{L}_{\mathcal{B}}$ // epoch loss
  $\mathbf{\Theta}_{\text{model}} = \text{Append}(\Theta_{\text{model}})$
  $\mathbf{L}_{\text{epochs}} = \text{Append}(\mathcal{L}_{\text{total}}/B)$
Return $\mathbf{L}_{epochs}$, $\mathbf{\Theta}_{model}^{(best)} \leftarrow \text{Best}(\mathbf{\Theta}_{model})$ // Return loss/epoch values and recall best parameters

Once the AttBF model is trained, it is deployed online, using the model's learned parameters, to provide instant predictions of the beamforming weights $\mathbf{w}_{\text{AttBF}}$ for any input sample (i.e., any $\mathbf{Y}_{ss}$), that focus the UT main beam towards the DS and null ISs directions.

D. Benchmark Models

The main objective of this paper is to introduce a DL-based beamforming model that does not require any CSI or DOAs estimations. The AttBF model can learn to map the spatial correlation matrix $\hat{\mathbf{R}}$ into an appropriate beamforming weight vector $\mathbf{w}_{\text{AttBF}}$ to maximize the received SINR.

Although our AttBF proposal works with the SCM $\hat{\mathbf{R}}$ as input, the same model can be adapted to work directly with the time and frequency domain representations of the received signal snapshots $\mathbf{Y}_{ss}$. We will consider these as benchmark approaches to strengthen the argument that the SCM provides more reliable performance in this context. To differentiate between the proposed model, which will also be referred to as $\text{AttBF}(\text{R})$ from hereafter, and the other models, we will denote the type of input data that will be used for each model. More particularly,

- AttBF(Y) model: Consider time-domain snapshots $\mathbf{Y}_{ss}$ as input after data normalization and tokenization preprocess (i.e., $\mathbf{\Gamma}_Y \in \mathbb{R}^{L\times 2M} \leftarrow \mathbf{Y}_{ss}$).
- AttBF(FFT) model: Consider the fast Fourier transform (FFT) representation of time-domain snapshots $\mathbf{Y}_{ss}$ as input after data tokenization preprocess (i.e., $\mathbf{\Gamma}_{FFT} \in \mathbb{R}^{L\times 2M} \leftarrow \text{FFT}(\mathbf{Y}_{ss}) \leftarrow \mathbf{Y}_{ss}$).

We also compare our proposed model with other traditional DL-based beamformers available in the literature,

- NNBF(H) model: It assumes the NNBF model introduced [22], which relies on full CSI input data (i.e., inputs $\mathbf{H} \in \mathbb{C}^{M\times(\mathcal{K}+1)}$), thus, it is expected to be sensitive to CSI estimation errors.
- CNNBF(R) model: Inspired by the CNNBF model in [23], it uses the SCM inputs (i.e., the real-valued matrix $[\Re(\hat{\mathbf{R}}); \Im(\hat{\mathbf{R}})] \in \mathbb{C}^{M\times 2M}$). This is particularly useful for comparing the traditional CNN-based models with the proposed model using the same input data.

Note that the same loss function and training approach as in Algorithm 1 are used for all benchmark models for a fair comparison. Table 1 shows the difference between all traditional methods and train-based models benchmarks in terms of input type.

TABLE 1. Comparison between our proposed model and other benchmarks

Model/Method	Input	CSI-Free?
Proposed	SCMs ($\hat{\mathbf{R}} \in \mathbb{C}^{M\times 2M}$) (22)	Yes
AttBF(Y)	Time domain samples ($\mathbf{Y_{ss}} \in \mathbb{C}^{L\times 2M}$)	Yes
AttBF(FFT)	Frequency domain samples (FFT($\mathbf{Y_{ss}}$) $\in \mathbb{C}^{L\times 2M}$)	Yes
NNBF(H) [22]	Full CSI matrix ($\mathbf{H} \in \mathbb{C}^{M\times(\mathcal{K}+1)}$) (11)	No
CNNBF(R) [23]	SCMs ($\hat{\mathbf{R}} \in \mathbb{C}^{M\times 2M}$) (22)	Yes
ZFBF	Full CSI matrix ($\mathbf{H} \in \mathbb{C}^{M\times(\mathcal{K}+1)}$) (11)	No
MRC	Desired satellite CSI vector ($\mathbf{h}_d \in \mathbb{C}^M$) (2)	No
SMI	SCMs ($\hat{\mathbf{R}} \in \mathbb{C}^{M\times 2M}$) (22) and Desired satellite response vector ($\mathbf{v}_d \in \mathbb{C}^M$)	Yes

E. AttBF Computational Time Complexity :

The time complexity of an off-line training DL model can be observed from the number of floating point operations (FLOPs) by each layer in inference [37], the complexity per layer for Conv1D: $\mathcal{O}(N_{seq} \cdot D_{hid}^2)$, Attention: $\mathcal{O}(D_{key} \cdot N_{seq}^2)$, and Dense: $\mathcal{O}(N_{in} \cdot N_{out})$, where N_{seq} is the size of the input sequence, D_{hid} is the hidden dimension, D_{key} is the attention key size [24], and N_{in}, N_{out} are the input and output neurons. Assuming the training data have been preprocessed, we can derive the per data sample time complexity for AttBF(Y) and AttBF(FFT) is $\mathcal{O}(L^2)$, and is equal to $\mathcal{O}(M^2)$ for other models for constant $D_{key} = D_{hid} = M$. However, only considering the number of FLOPs cannot be sufficient to draw a fair comparison between DL-based models and traditional methods, so we propose both end-to-end inference time complexity and the pure inference time in the following subsection as a performance indicator.

TABLE 2. Comparison of the End-to-end inference/latency time Complexity

Model/Method	Input
Proposed	$\tau^{\text{cal}}_{\text{SCM}} + \tau^{\text{Tok}}_{\text{SCM}} + \mathcal{O}(M^2)$
AttBF(Y)	$\tau^{\text{Tok}}_{\text{Y}} + \mathcal{O}(L^2)$
AttBF(FFT)	$\tau^{\text{FFT}}_{\text{Y}} + \tau^{\text{Tok}}_{\text{FFT(Y)}} + \mathcal{O}(L^2)$
NNBF(H) [22]	$\tau^{\text{est}}_{\text{fullCSI}} + \tau^{\text{Re/Im}}_{\text{fullCSI}} + \mathcal{O}(M^2)$
CNNBF(R) [23]	$\tau^{\text{cal}}_{\text{SCM}} + \tau^{\text{Re/Im}}_{\text{SCM}} + \mathcal{O}(M^2)$
ZFBF	$\tau^{\text{est}}_{\text{fullCSI}} + \mathcal{O}(M(\mathcal{K}+1)^2 + (\mathcal{K}+1)^3)$
MRC	$\tau^{\text{est}}_{\text{desCSI}} + \mathcal{O}(M)$
SMI	$\tau^{\text{est}}_{\text{desDOA}} + \tau^{\text{cal}}_{\text{SCM}} + \mathcal{O}(M^2L + M^3)$

In Table 2, we show the overall time complexity required by a model/method to perform an inference on a testing data sample. For the DL-based models, the Big-O notation written in the table refer to the model per data sample time complexity, in addition we notated the pre-precess steps time complexity as "τ", for example, our proposed AttBF(R) model includes two data pre-processing steps, which are the SCM calculation time denoted as $\tau^{\text{cal}}_{\text{SCM}} = \mathcal{O}(M^2L)$ as in (17), and then its tokenization -a constant-time operation to separate the real/imaginary parts of SCM- step detonated as $\tau^{\text{Tok}}_{\text{SCM}} = \mathcal{O}(M^2)$ as in (23), finally, the proposed model per data sample inference complexity would be same as the training one, hence, it is equal to $\mathcal{O}(M^2)$.

The rest of the DL-based model pre-precess time complexity notations are as follows: 1) The snapshots matrix $\mathbf{Y_{ss}}$ tokenization $\tau^{\text{tok}}_{\text{Y}} = \mathcal{O}(L^2)$, 2) The standard FFT operation on $\mathbf{Y_{ss}}$ takes $\tau^{\text{FFT}}_{\text{Y}} = \mathcal{O}(ML\log L)$ and its tokenization $\tau^{\text{tok}}_{\text{FFT(Y)}} = \tau^{\text{tok}}_{\text{Y}}$, 3) The full CSI channel $\mathbf{H}$ estimation $\tau^{\text{est}}_{\text{fullCSI}}$ depends on the estimation algorithm used, for example a pilot based method will not be possible for the unknown $\mathcal{K}$ interference channels, overall $\tau^{\text{est}}_{\text{fullCSI}}$ would be significant without mentioning the additional complexity of knowing the number of $\mathcal{K}$ interference sources. Separating Real/Imaginary parts of full CSI channel $\mathbf{H}$ will equal to $\tau^{\text{Re/Im}}_{\text{fullCSI}} = \mathcal{O}(M(\mathcal{K}+1)$, 4) Separating Real/Imaginary parts of SMC is $\tau^{\text{Re/Im}}_{\text{fullCSI}} = \tau^{\text{Tok}}_{\text{SCM}}$.

Similarly, for the traditional methods, the Big-O notation listed in the table refers to the computational time complexity of the benchmark beamforming method per data sample as in (19), (20), and (21). In addition, we noted the time complexity of the pre-calculating/estimation steps as τ. The ZFBF method requires full CSI estimation $\tau^{\text{est}}_{\text{fullCSI}}$, while MRC only requires the desired CSI channel estimation, but also depends on the estimation algorithm (e.g., for least-square (LS) estimation with L pilot signals it will equal to $\mathcal{O}(ML)$). The SMI method involves an SMC calculation $\tau^{\text{cal}}_{\text{SCM}}$ and desired satellite DOA estimation $\tau^{\text{est}}_{\text{desDOA}}$ (e.g., using MUSIC algorithm) for calculating $\mathbf{v}_d$. Note that UT arrays shall be simple for NGSO applications (e.g., Direct to device (D2D)), reducing the overall complexity.

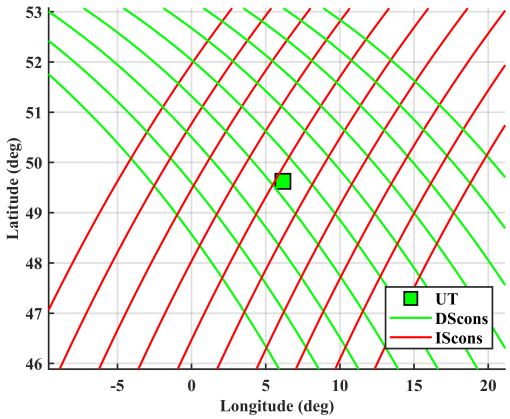

FIGURE 7. **Proposed NGSO systems constellations**

IV. Simulation Results

In this section, we present the simulation setup, including details on data generation, the model training phase, and validation and comparison of our proposed model with the traditional techniques and different DL-based models.

A. Data and Simulation

We run a satellite simulation scenario with one desired NGSO satellite and multiple interfering NGSO satellites using MATLAB satellite toolbox [38]. We adopt a realistic approach to representing satellite motion by assigning each satellite to a defined plane orbit through the six Kepler orbit elements: the semi-major axis, eccentricity, inclination, right ascension of the ascending node (RAAN), argument of perigee, and true anomaly. MATLAB built-in functions provide estimates of latitude, longitude, and altitude over time according to the satellite trajectory. Using this setup, we consider the physical environment (distances, angles, satellites' orbit propagation, etc.) and the operational conditions (frequency, bandwidth). With the correct adjustments of these simulation elements, we can alter SINR values. We can summarize the paper simulation and the datasets generation phase in a few steps,

1) Using MATLAB built-in functions within the Satellite Communication Toolbox, we simulate two co-existing NGSOs constellations and a fixed UT location, as shown in Fig. 7, using parameters in Table 3 and Table 4. We then initiate two simulations $\mathcal{T}_{\text{Tr}}$ and $\mathcal{T}_{\text{Ts}}$ for training and testing datasets generation. Each simulation starts with one DS satellite from a fixed satellite plane in the DScons and one IS satellite ($\mathcal{K}$ = 1) that can be randomly selected from the IScons.
2) The DS satellite pass duration, where the DS is visible to the UT, is defined by the UT minimum elevation angle $\theta_{\text{min}} = 15°$, and in our simulation it equals $T_{\text{pass}} = 300$ seconds. We consider $T_{\text{samp}} = 1$ $seconds, which leads$ to 300 coherent channels (samples/pass). The UT assumes knowledge of DOAs of the desired signal and thus apply an initial beamformer $\mathbf{w}_{\text{in}} = \mathbf{w}_{\text{MRC}}$ [1].
3) During each coherence time T_{samp}, we calculate the link budget for both DS and IS satellite links as in equations (1)-(8), using link budget parameters in Table 5, we also save the first $L = T_{ss}$ snapshots (9) and save the $\mathbf{Y}_{ss}$, γ_{out}, and $\mathbf{H}$, matrices in $\mathcal{D}_{\text{Tr}}$ tensor for training dataset generating phase, or $\mathcal{D}_{\text{Ts}}$ tensor for testing dataset generating phase. For the sake of evaluation, we apply a minimum interference-to-noise ratio (INR) threshold of -12 [dB], to only consider interference where the quality of service (QoS) of the desired link is affected [14].
4) The simulation is precisely configured to stop when reaching $\mathcal{N}_{\text{Tr}} = 9000$ training data samples (30 desired satellite passes) and $\mathcal{N}_{\text{Ts}} = 900$ testing data samples (3 desired satellite passes).
5) The lowest number of snapshots possible is what will make the SCM ($\hat{\mathbf{R}}$) a full rank, which requires at least $L = M$ independent data samples (snapshots) [34], in many research articles it is recommended to use at least $L = 2M$ snapshots when estimating the covariance matrix for better beamforming results [35]. We repeated the above data generation steps and generated three sets of data for when $L = M, 2M,$ and $4M$.

The full datasets in this article are available with more details on the SmartSpace project web page [27]. That said, it will strengthen the proposed AttBF-based model, and possibly other AI-based solutions, if real-world datasets are made available by satellite operators. Authors are committed to disseminate the importance of this issue with relevant industrial partners.

TABLE 3. **Constellations parameters**

Parameters	DScons	IScons
Constellation Altitude (h^{cons})	550 [Km]	950 [Km]
Eccentricity	0.01	0.04
Inclination	55°	120°
RAAN	135°	35°
Argument of Perigee	20°	35°
True anomaly	82°	75°
Δh^{cons}	±50 [Km]	±50 [Km]
ΔRAAN	±12°	±12°

Fig. 8 shows the SINR measured on the UT side for a DS pass T_{pass} assuming that a fixed $\mathbf{w}_{\text{MRC}}$ is applied at the UT planar array for the entire pass of the satellite (non-adaptive beamforming). We can observe several SINR drops due to interference, particularly in the first half of the satellite pass.

[1] In LoS scenarios, and when pCSI exist this yield to $\mathbf{w}_{\text{MRC}} = \frac{\mathbf{h}_d}{\|\mathbf{h}_d\|} \approx \mathbf{v}_d = \mathbf{w}_{\text{in}}$ and $\gamma_{\text{out}}(\mathbf{w}_{\text{MRC}}) \approx \gamma_{\text{in}}$ where γ_{in} is the measured SINR using the initial beamformer $\mathbf{w}_{\text{in}}$.

TABLE 4. UT Parameters

Parameter	Value	Parameter	Value
Latitude	49.6257°	# Elements (M)	$8 \times 8 = 64$
Longitude	6.1598°	Noise Bandwidth	50 MHz
Altitude (h^{ut})	100 [m]	Temperature (T^{ut})	94°K
Min elevation (θ_{min})	15°	Efficiency (η^{ut})	0.99

TABLE 5. Satellite Link Budget Parameters

Parameters	DS Satellite	IS Satellite
Signal bandwidth (BW)	50 MHz	55MHz
Operational frequency (f_c)	11.75 GHz	11.75 GHz
Satellite EIRP (P^{sat})	34 dBW	45 dBW

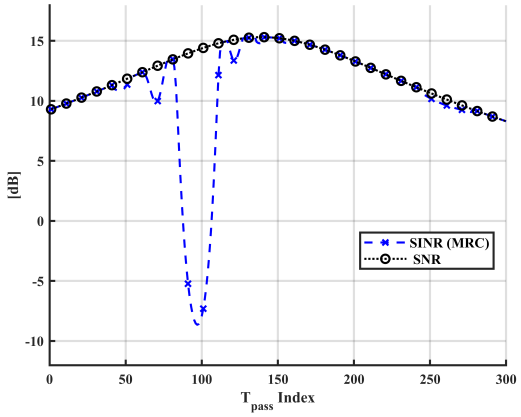

FIGURE 8. A representation link status from the $L = 2M$ test data of a desired satellite pass with nearby interference satellites as received by UT during T_{Pass}

B. Training Phase Results

The TensorFlow and Keras frameworks are used to build all DL models in Python. All models are compiled with $\mathcal{L}$ in (35) as the primary loss function, and Nesterov implemented Adam as the optimizer Ω. The model parameters were updated using the $\mathcal{N}_{\text{Tr}}$ training samples in the training dataset $\mathcal{D}_{\text{Tr}}$. The results were validated using the N_{Ts} testing samples in the testing dataset $\mathcal{D}_{\text{Ts}}$. All three AttBF models have $A = 2$ Attention heads, $D_{key} = D_{hid} = 64$, and Dropout layers (with rate 0.2) are applied to avoid training overfitting. The experiments in the following parts were carried out using an 8-core `AMD Ryzen 7 7700X CPU` unit with 12 GB RAM. During the training phase, as in Algorithm 1, we assume $E = 200$ epochs for all models, a training batch size $\mathcal{B} = 300$ includes all data samples per DS pass T_{pass}, leading to $B = 30$ training batches.

Fig. 9 plots the epochs versus the average SINR per epoch, that is, the negative loss/epochs ($-\mathbf{L}_{\text{epochs}}$) for each model on the $L = 2M$ datasets, demonstrating how beamforming performance quality improves (or stabilizes) as training

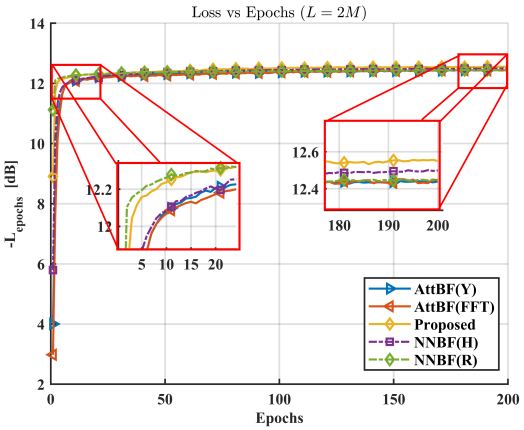

FIGURE 9. Loss vs Training Epoch ($L = 2M$ test data)

progresses. Our AttBF(R) model, labeled as *Proposed*, achieves high convergence stability in early epochs and maintains a lead value throughout the rest of the epochs, indicating faster convergence and superior performance.

TABLE 6. Comparison of models training (# $L = M \mid 2M \mid 4M$)

Model	# Parameters [$\times 10^3$]	Train Time [min]
AttBF(Y)	**184** | 315 | 577	**1.64** | 3.69 | 6.44
AttBF(FFT)	**184** | 315 | 577	1.81 | 3.42 | 6.95
Proposed	**184** | **184** | **184**	3.26 | 3.24 | 3.25
NNBF(H) [22]	203 | 203 | 203	1.75 | **1.77** | **1.74**
CNNBF(R) [23]	511 | 511 | 511	1.77 | **1.73** | **1.75**

In Table 6, we compare the beamforming DL-based beamforming models in terms of their input, trainable parameters in their architectures, and training time efficiency.

- **Number of parameters**: The total trainable parameters (in thousands [$\times 10^3$]) is an indicator of the size of the model. The spatial sharing of weights in Conv1D and Attention drastically reduces the number of parameters compared to Dense layers used in CNNBF(R). Furthermore, NNBF(H) input size is changing with the increment of $\mathcal{K}$, which may require a different architecture for each size of $\mathcal{K}$. With an increasing number of snapshots L, the proposed AttBF(R) model always has a fixed input token sequence M, compared to changing L tokens in AttBF(Y) and AttBF(FFT) models, making it the smallest model in terms of trainable parameters.
- **Train Time**: Our AttBF-based models fall behind conventional CNN models in terms of training time. However, the input size and the type of trainable layers used in each model may suggest that more time may be required for training. Moreover, overhead from operations like softmax, LayerNorm, or Dropout in all AttBF-based models may result in slower training[2],

[2]Backprop is typically implemented using matrix operations, which influence computational and memory requirements for each model.

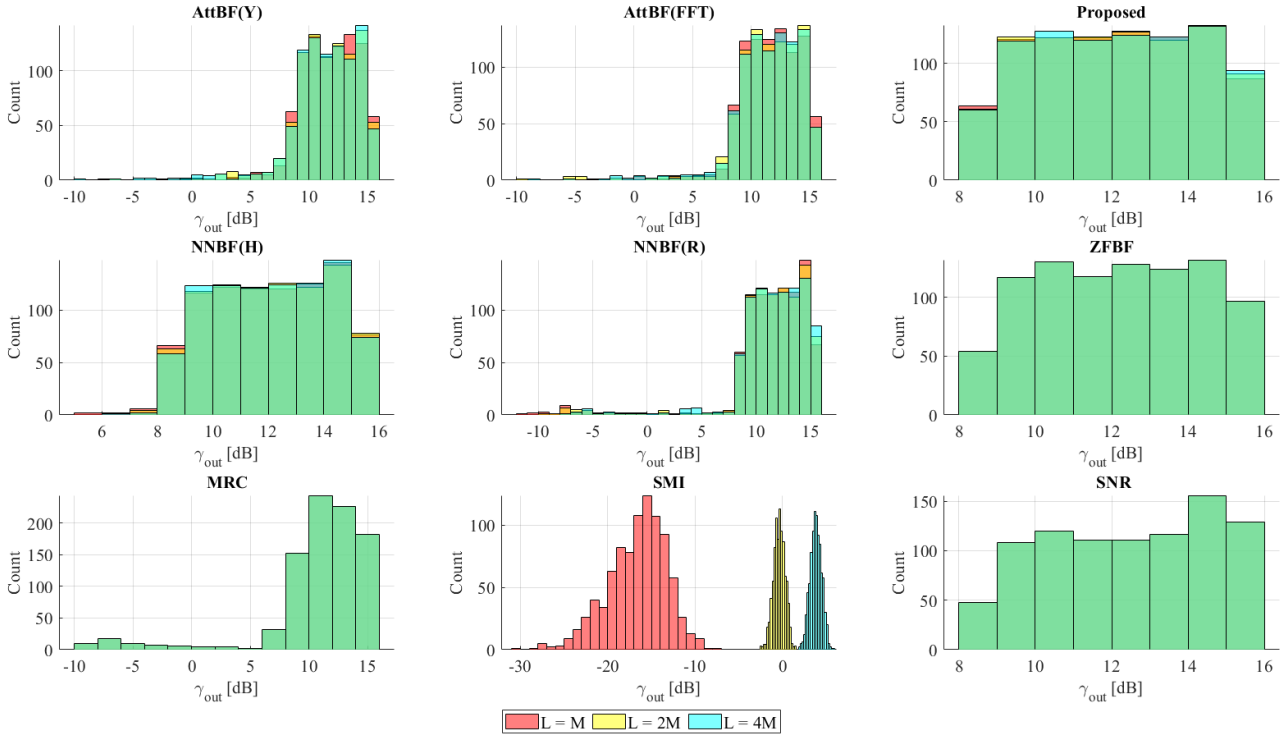

FIGURE 10. **SINR distribution per beamforming method (when $L = M \mid 2M \mid 4M$)**

compared to CNNBF(R) and NNBF(H) models. Despite this, the training time is not very relevant for offline training as it usually depends on the model architecture, training environment, and hardware (e.g., CPU, GPU, accelerators) [39].

C. Beamforming Performance Results

During the testing phase (inference), a test batch size $\mathcal{B} = 1$ includes all data samples per DS T_{pass} pass, leading to $B = 900$ testing batches. In Fig. 10, we show histograms of the output SINR distribution per beamforming method (DL-based models and traditional methods) for the 900 instances in the testing dataset $\mathcal{D}_{\text{Ts}}$, the signal-to-noise (SNR) results are not realizable in real-time and serve as an upper bound for reference only (assuming no interference is present all the time). In each sub-figure, the X-axis (γ_{out} [dB]) represents the output SINR in decibels, a higher range of values, compared to the SNR plot, indicates a stronger desired satellite signal and better null performance in case of interference occurrence, as it means that the desired signal is relatively larger to the combined power of interference and noise. The Y-Axis (Count) shows the number of times a particular SINR value occurred across the entire test dataset. The height of the bars in each histogram indicates the frequency of the corresponding SINR range. Finally, the histograms are stacked and color-coded (i.e., "red" ($L = M$), "yellow" ($L = 2M$), and "cyan" $L = 4M$); other colors within the bars show the contribution of each condition stacked on top of each other. Here, the "Proposed" method's histogram, it can be seen that the majority of the instances are stacked on top of each other (the light green portion), indicating that this method performed similarly well under all L amounts, same as ZFBF, compared to other model/methods which have some wider and some negative γ_{out} values. In Table 7 we summarize the key insights of Fig. 10 results, along with the pure inference time:

- **Pure Inference Time :** Inference time is particularly crucial for fast adaptive beamforming in high-mobility NGSOs, where real-time processing is essential. The pure interface time here is observed without considering the pre-processing duration per model (i.e., for the proposed AttBF(R) model, we assume SCM is calculated and tokenized), and only observes the per-sample model processing time and the calculation time for the traditional methods (e.g., for the ZFBF method we assume full CSI is already estimated). The proposed AttBF(R) model offers the fastest inference time "average on all test data" among learning-based methods ($\sim 4.5\ \mu s$), CNNBF(R) has similar time performance ($\sim 4.6\ \mu s$). Other DL-based models have slightly slower inference ($\sim 6-7\ \mu s$). Regarding classical methods, MRC offer the fastest inference ($\sim 1.2\ \mu s$) but without interference nulling ability, ZFBF comes slightly after ($\sim 2.5\ \mu s$). Still, the SMI method is the slowest among all metrics ($\sim\ 133\ \mu s$), making it not preferable for real-time deployments.

TABLE 7. Comparison of models/methods beamforming performance (# $L = M \mid 2M \mid 4M$)

Model/Method	Pure Inference Time [μs]	γ_{out} [dB]		
		Min	Max	Mean
AttBF(Y)	6.21 | 6.95 | 7.38	-7.04 | -6.46 | -9.23	15.26 | 15.25 | 15.25	11.69 | 11.60 | 11.44
AttBF(FFT)	6.74 | 6.99 | 7.73	-0.75 | -9.49 | -8.14	15.25 | 15.24 | 15.26	11.68 | 11.55 | 11.52
Proposed	**4.36** | 4.70 | **4.57**	**8.00** | **8.03** | **8.06**	**15.29** | **15.30** | **15.30**	**12.09** | **12.11** | **12.13**
NNBF(H) [22]	6.54 | 6.53 | 6.56	5.07 | 7.25 | 6.72	**15.29** | 15.28 | 15.29	12.05 | 12.08 | 12.07
CNNBF(R) [23]	4.67 | **4.50** | 4.71	-11.48 | -9.01 | -7.27	15.27 | 15.28 | 15.28	11.40 | 11.47 | 11.62
ZFBF (optimalBF)	2.14 | 2.80 | 2.77	**8.10** | **8.10** | **8.10**	**15.31** | **15.31** | **15.31**	**12.18** | **12.18** | **12.18**
MRC	**1.31** | **1.16** | **1.08**	-8.65 | -8.65 | -8.65	15.31 | 15.31 | 15.31	10.80 | 10.80 | 10.80
SMI	133.02 | 132.65 | 134.33	-30.65 | -2.53 | 1.88	-9.64 | 1.53 | 5.96	-16.40 | -0.37 | 3.83
SNR	- | - | -	8.30 | 8.30 | 8.30	15.31 | 15.31 | 15.31	12.42 | 12.42 | 12.42

- **Min/Max SINR Stability :** The Min/Max values show the minimum and maximum γ_{out} in the results distribution achieved from each test dataset. The proposed model shows stable values for minima ($sim8$ [dB]) and maximum ($\sim$ 15.30 [dB]), similar to the optimal beamformer ZFBF method. NNBF(H) model shows competitive results. In contrast, other DL models and traditional methods fall behind, especially on minimum SINR, indicating limitations on interference nulling performance in high INR scenarios (i.e., In-Line scenarios).
- **Mean SINR Performance :** The Mean value calculated from the distribution of the results indicates the average γ_{out} achieved from each test dataset. The proposed model demonstrates the highest mean SINR "γ_{out}" in the test data, which is almost aligned with the optimal ZFBF results. NNBF(H) with full CSI inputs (both desired and interference channels) has slightly lower average γ_{out} values compared to the proposed model. All other DL-based models show moderate average γ_{out} values due to wider distribution spread, implying less stable beamforming performance. The classic MRC method falls short under interference conditions; it only indicates the initial average γ_{out}, which is improved by all other methods except SMI, with worse performance due to an insufficient number of snapshots.
- **Impact of Snapshots:** As explained in the data generation steps Section IV.A, we have three main datasets, each with a different snapshot amount ($L = M, 2M, 4M$). The amount of snapshots has no noticeable effects on inference time, except for the AttBF(Y) and AttBF(FFT) models, where it increases with more snapshots, probably because of the model architecture changes. The proposed model and the CNNBF(R) model show some performance improvement in Min/Max/Mean results with more snapshots. The SMI shows noticeable improvements with more snapshots (see Remark. 1).

1) Robustness to high Correlation

It is well-known that the presence of high spatial correlation between desired and interference signals (similar DOAs) makes it harder to separate them, even with accurate channel information. The *mutual coherence* measures how "similar" or "aligned" the normalized steering vectors of two signals are. For a desired steering vector $\mathbf{v}_d$ and $\mathcal{K}$ interfering vectors, mutual coherence measures the highest correlation as [40],

$$\rho_{\max} = \max_{1 \le k \le \mathcal{K}} \left| \mathbf{v}_d^H \mathbf{v}_{i,k} \right|. \tag{36}$$

When $\rho_{\max} = 0$, it means they are perfectly orthogonal (and thus completely uncorrelated). At the same time, a small value $\rho_{\max} \to 0$ shows that they are nearly orthogonal, meaning that they can generally distinguish between the corresponding signals. When $\rho_{\max} \to 1$, the channel matrix $\mathbf{H}$ becomes *poorly conditioned*, posing challenges to methods that involve matrix inversion. A good beamformer should be able to null interfering DOAs while reserving power toward desired DOAs under correlation conditions up to $\rho_{\max} \approx 0.9$ [40].

In Fig. 11 and Fig .12, we present a comparative evaluation of beamforming methods, we show a single DS satellite pass duration T_{pass} from the testing data set $\mathcal{D}_{\text{Ts}}$, containing only 300 testing samples, as shown in the left-upper part of this two figures, this highlights all methods SINR performance per data sample (except AttBF(Y), AttBF(FFT), as proposed AttBF(R) model shows better results than these Attention-based models already, and also to save space). Here, perfect CSI is assumed for the benchmark models/methods. In each figure, we zoomed in on a specific part of that pass, representing an interfering case, and indicated it with a red box. For each zoomed-in part, we plot six sub-figures representing each model/method beam patterns visualization as seen from the UT side for a specific data sample n (assuming that method/model was applied to the UT at that particular sample), in these beam patterns, x-axis: azimuth angle ϕ°, y-axis: elevation θ°, and z-axis: directivity Ξ^{ut}, the DS satellite is highlighted with a black circle ($\circ$), while the IS satellite is indicated by black crosses ($\times$). The observations of the two cases are as follows,

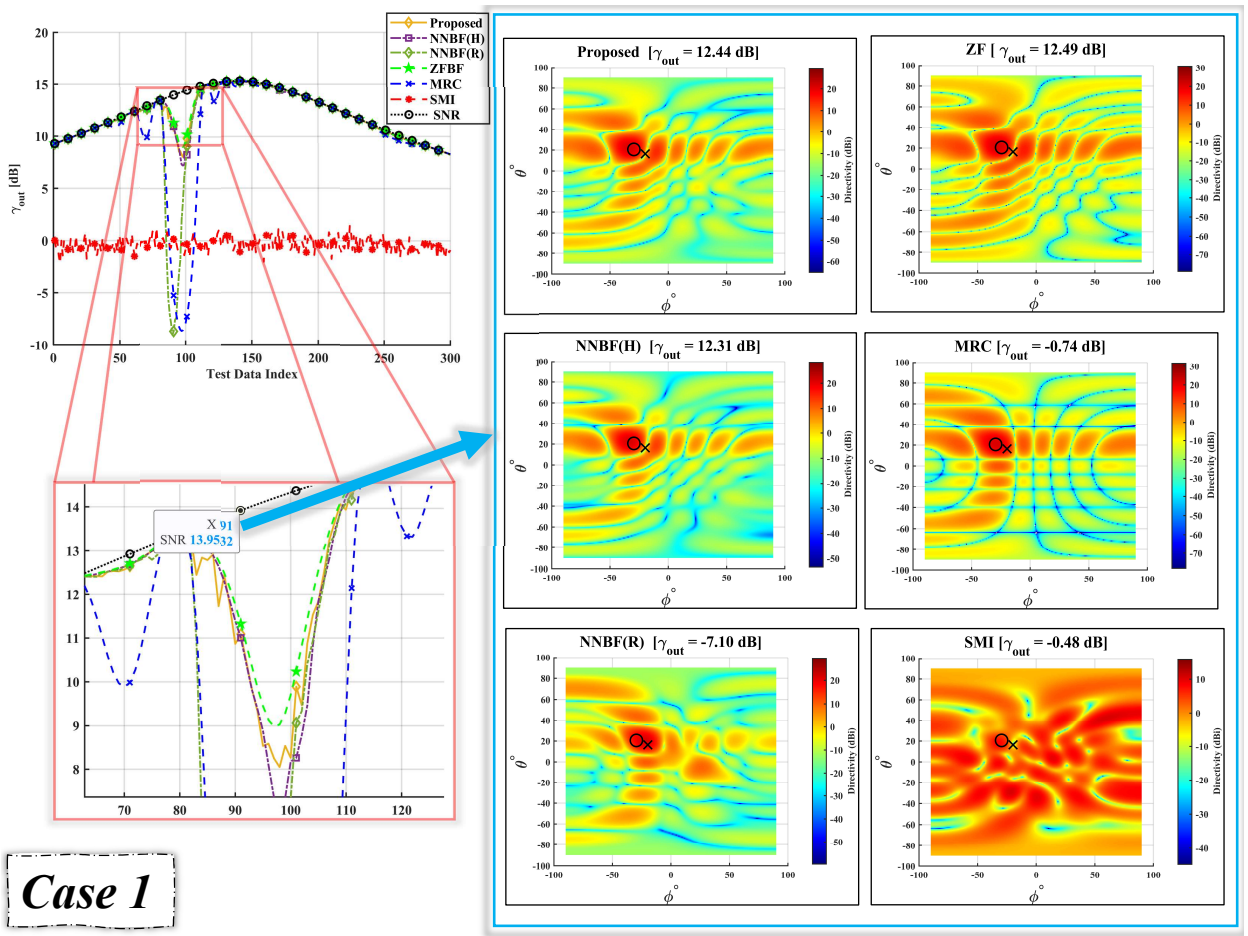

FIGURE 11. **Case1: a main-lobe correlation ($\rho_{\max} = 0.67$)**

- **Case 1 - High Mutual Coherence between desired and interference**
  Looking at the UT array beam pattern obtained from MRC, we can see that the interferer satellite power is captured in the *main-lobe*, thus suffering a severe interference impact (i.e., extremely low SINR level). The DOAs of the DS satellite are $\varphi_d = \left[-27.92°; 19.97°\right]$ and the DOAs of the IS satellite are $\varphi_i = \left[-20.94°; 15.53°\right]$, leading to very high mutual coherence of $\rho_{\max} = 0.67$. Here, the proposed AttBF(R) model creates a well-formed beam pattern and gives the highest γ_{out} compared to other DL-based methods despite the high spatial correlation. NNBF(H) with full CSI input follows behind, and CNNBF(R) seem to have weak nulling capabilities for high $\rho_{\max}$ values. ZFBF has the optimal interference nulling with pCSI, and SMI has the worst performance among all methods, as expected, because with low snapshots, it only reduces power "null" in the DOAs of the IS satellite direction without focusing the power towards the DOAs of the DS satellite.

- **Case 2 - Low Mutual Coherence between desired and interference**
  Noting that both satellites are moving, and we can see with the MRC beamformer, the interferer satellite power is now moved to the *side-lobe*, thus the system suffered a lower interference compared to Case 1. In this instant, the DOAs of the DS satellite are $\varphi_d = \left[-9.56°; 12.38°\right]$ and the DOAs of the IS satellite are $\varphi_i = \left[-31.17°; 7.78°\right]$, which leads to a low mutual coherence of $\rho_{\max} = 0.19$. Here, the proposed AttBF(R) model still provides a well-formed beam pattern, focusing the power toward the DOAs of the DS satellite and reducing the power "null" in the DOAs of the IS satellite direction, which achieves the highest γ_{out} compared to other DL-based methods. NNBF(H) and CNNBF(R) models show competitive SINR and nulling performance with the proposed model in side-lobe instances. ZFBF always has the optimal interference nulling with pCSI, and SMI continues to give the worst performance among all methods.

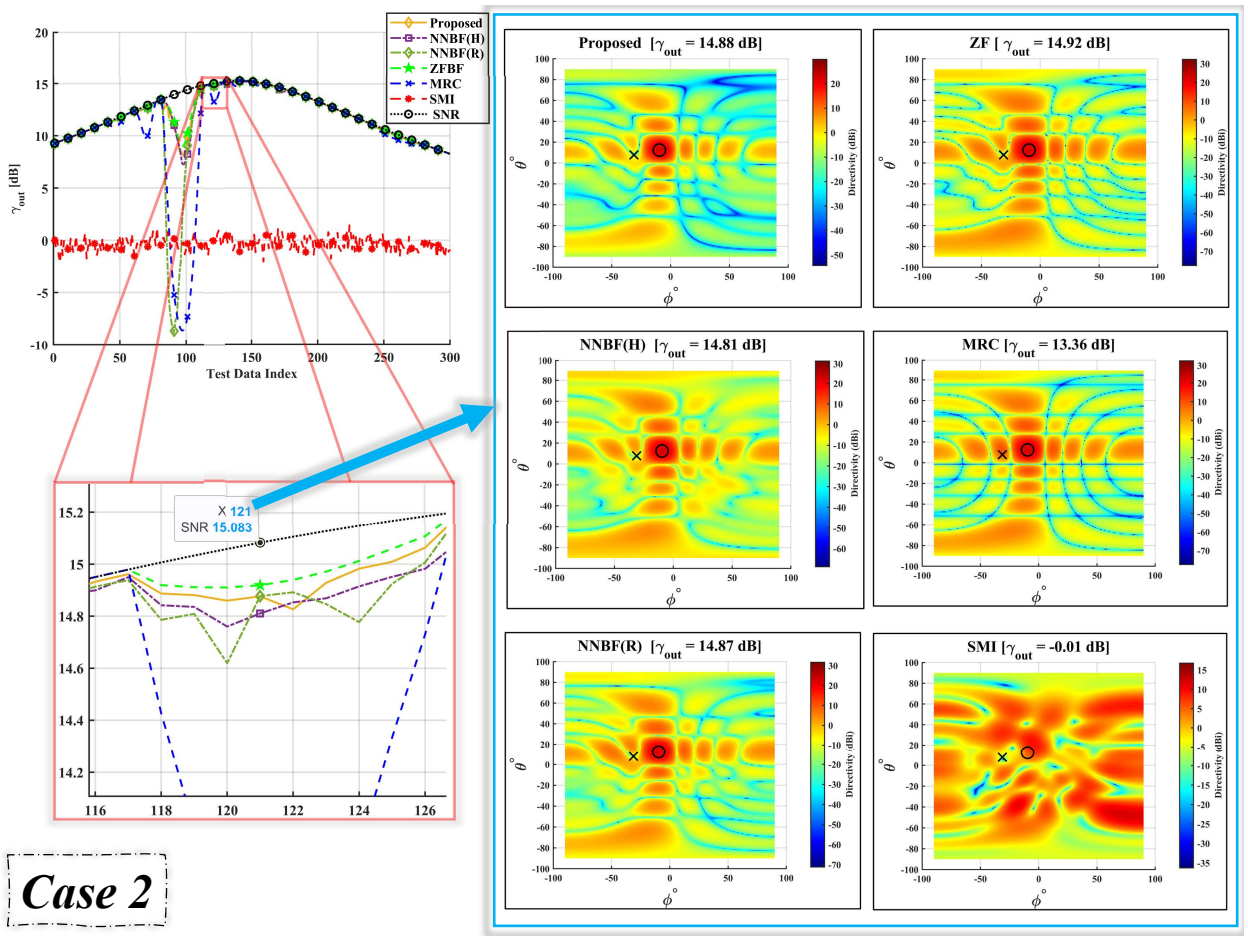

FIGURE 12. **Case2 : a side-lobe correlation ($\rho_{\max} = 0.19$).**

2) Robustness to imCSI

To analyze our model in contrast to CSI-dependent methods, we chosen an interference case in the satellite pass (i.e., a main-lobe instant ($\rho_{\max} = 0.78$)), then show how the γ_{out} value varies with increasing variance of imCSI estimation error (σ_e^2), observed from 500 MontyCarlo realizations, as in Fig. 13. The proposed model AttBF(R) is clearly the most robust and effective, maintaining a high SINR without the need for CSI estimation quality. CNNBF(R) and SMI offer very low but consistent performance, due to their input of the SCM and their lack of reliance on any CSI information. NNBF(H) performs well at small estimate error, but its performance degrades with higher error levels. ZFBF and MRC are not reliable under imperfect practical CSI conditions, as γ_{out} performance gradually degrades with increasing error.

Avoiding CSI estimation and its errors, especially when unknown spatial interference is present, is desirable in NGSO systems, where fast mobility and Doppler effects make CSI acquisition costly and/or inaccurate. This makesDL modes, such as our proposed AttBF(R) model, promising for co-existing NGSO system users.

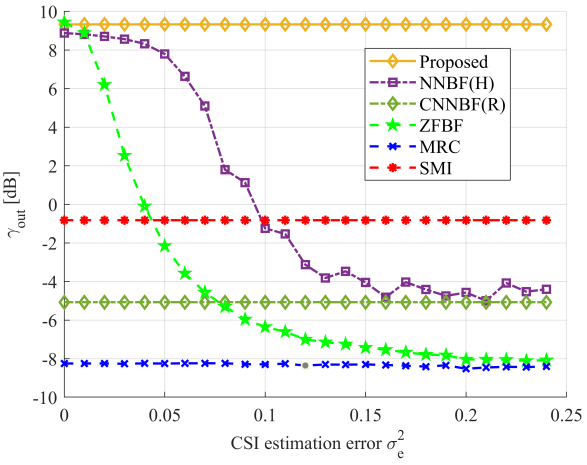

FIGURE 13. **SINR vs CSI estimation error ($\rho_{\max} = 0.78$)**

In conclusion, Table 8 provides a comprehensive summary and the expected trade-off for each beamforming approach discussed in this article.

TABLE 8. Summary Comparison of Beamforming Methods for NGSOs Interference Mitigation

Method	Advantages	Trade-offs
Proposed	• No CSI knowledge required, • High SINR maximization performance, • Fast inference ($\sim 4.5\mu s$).	• Needs SCM $\hat{\mathbf{R}}$ calculation.
$\mathrm{AttBF(Y)}$	• No CSI knowledge required, • Generate a beamformer from time-domin signals inputs.	• Moderate SINR maximization performance.
$\mathrm{AttBF(FFT)}$	• No CSI knowledge required, • Generate a beamformer from frequency-domain inputs.	• Moderate SINR maximization performance. • Requires FFT transformation.
$\mathrm{NNBF(H)}$ [22]	• High SINR performance with pCSI.	• Requires full CSI knowledge, • Performance degrades with imCSI errors.
$\mathrm{CNNBF(R)}$ [23]	• No CSI estimation required, • Fast inference ($\sim 4.5\ \mu s$).	• Moderate SINR maximization performance, • Largest model in terms of trainable parameters, • Needs SCM calculation.
ZFBF	• High SINR maximization performance with pCSI.	• Requires full CSI knowledge, • Performance degrades with imCSI estimation error, • Complex (contain matrix inversion operations).
MRC	• Lowst latency ($\sim 1.1\ \mu s$ inference), • Low complexity.	• Requires desired channel CSI knowledge, • Performance degrades with imCSI errors, • Not suitable for interference nulling tasks.
SMI	• No CSI knowledge required.	• Sensitive to the low accuracy of the estimated $\hat{\mathbf{R}}$, • Requires lots of snapshots ($L \leftarrow T_{\text{samp}}$) to perform optimally, • Highest latency ($\sim 133\ \mu s$), • Needs SCM $\hat{\mathbf{R}}$ calculation.

V. Conclusion and Future Work

In this work, we have presented the model $\mathrm{AttBF(R)}$, a self-supervised DL beamforming method that learns a non-linear mapping from the SCM to the complex beamforming weights. The goal is for this mapping to implicitly learn to identify the spatial characteristics of the desired signal and the interfering signals from the SCMs. By learning to produce appropriate complex weights, the beamformer can steer the main lobe towards the desired user and create nulls in the direction of the interference sources, thus mitigating CFI interference in the NGSOs co-existing environment with latency. The proposed $\mathrm{AttBF(R)}$ model appears to be a promising beamforming technique that balances fast interference null performance, CSI-free requirements, and DL models' computational efficiency, achieving the highest SINR among other DL-based techniques, while traditional methods such as SMI that use the same input as the $\mathrm{AttBF(R)}$ model had the worst performance among all methods due to their sensitivity to the SCM estimation accuracy. Future works may investigate UT mobility scenarios, complex channel model (with NLOS, Doppler effect, etc.), various interfering sources (terrestrial, jamming CFI signals), and real hardware scenarios, which are critical topics for integrating SatComs with the 6G communication systems. In this context,there is a need for collaborations with satellite operators to provide real measurements and generalized datasets, which currently limits the real-world validations for our method (and similar AI-based techniques). It's important to making such data publicly available in future work to accelerate research in this domain.

REFERENCES

[1] E. Lagunas, S. Chatzinotas, and B. Ottersten, "Low-Earth orbit satellite constellations for global communication network connectivity," *Nature Reviews Electrical Engineering*, vol. 1, no. 10, pp. 656–665, 2024.

[2] O. Kodheli, E. Lagunas, N. Maturo, S. K. Sharma, B. Shankar, J. F. M. Montoya, J. C. M. Duncan, D. Spano, S. Chatzinotas, S. Kisseleff, J. Querol, L. Lei, T. X. Vu, and G. Goussetis, "Satellite Communications in the New Space Era: A Survey and Future Challenges," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 70–109, 2021.

[3] Y. He, Y. Li, and H. Yin, "Co-frequency interference analysis and avoidance between NGSO constellations: Challenges, techniques, and trends," *China Communications*, vol. 20, no. 7, pp. 1–14, 2023.

[4] C. Braun, A. M. Voicu, L. Simić, and P. Mähönen, "Should We Worry About Interference in Emerging Dense NGSO Satellite Constellations?" in *2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2019, pp. 1–10.

[5] A. Saifaldawla, F. G. Ortiz, E. Lagunas, and S. Chatzinotas, "Convolutional Autoencoders for Non-Geostationary Satellite Interference Detection," in *IEEE International Conference on Communications (IEEE ICC)*. IEEE, 2024.

[6] Starlink, "Setup guide," https://api.starlink.com/public-files/installation_guide_standard_kit.pdf, (Accessed on 12/05/2025).

[7] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[8] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex Optimization-Based Beamforming," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, 2010.

[9] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2002.

[10] H. A. Kassir, Z. D. Zaharis, P. I. Lazaridis, N. V. Kantartzis, T. V. Yioultsis, and T. D. Xenos, "A Review of the State of the Art and Future Challenges of Deep Learning-Based Beamforming," *IEEE Access*, vol. 10, pp. 80 869–80 882, 2022.

[11] Y. Omid, M. Aristodemou, S. Lambotharan, M. Derakhshani, and L. Hanzo, "Reinforcement learning-based downlink transmit precoding for mitigating the impact of delayed CSI in satellite systems," *arXiv preprint arXiv:2410.21489*, 2024.

[12] M. S. Rahman, E. Onggosanusi, H. Si, and J. Cho, "CSI feedback based on space-frequency compression," in *2020 IEEE 17th Annual*

Consumer Communications & Networking Conference (CCNC), 2020, pp. 1–6.

[13] S. Lin, J. An, L. Gan, M. Debbah, and C. Yuen, "Stacked Intelligent Metasurface Enabled LEO Satellite Communications Relying on Statistical CSI," *IEEE Wireless Communications Letters*, vol. 13, no. 5, pp. 1295–1299, 2024.

[14] A. Saifaldawla, F. Ortiz, E. Lagunas, A. B. M. Adam, and S. Chatzinotas, "GenAI-Based Models for NGSO Satellites Interference Detection," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 904–924, 2024.

[15] H. Yang, K.-Y. Lam, J. Nie, J. Zhao, S. Garg, L. Xiao, Z. Xiong, and M. Guizani, "3D Beamforming Based on Deep Learning for Secure Communication in 5G and Beyond Wireless Networks," in *2021 IEEE Globecom Workshops (GC Wkshps)*, 2021, pp. 1–6.

[16] G. Konstantopoulos and Y. Louët, "Deep Learning Aided Beamforming for Downlink Non-Orthogonal Multiple Access Systems," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4337–4353, 2024.

[17] J. Kim, H. Lee, S.-E. Hong, and S.-H. Park, "Deep Learning Methods for Universal MISO Beamforming," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1894–1898, 2020.

[18] M. Fakharzadeh, S. H. Jamali, P. Mousavi, and S. Safavi-Naeini, "Fast Beamforming for Mobile Satellite Receiver Phased Arrays: Theory and Experiment," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 6, pp. 1645–1654, 2009.

[19] K.-X. Li, L. You, J. Wang, X. Gao, C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Downlink Transmit Design for Massive MIMO LEO Satellite Communications," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1014–1028, 2022.

[20] X. Xie, X. Ding, and G. Zhang, "Interference Mitigation via Beamforming for Spectrum-Sharing LEO Satellite Communication Systems," *IEEE Systems Journal*, vol. 17, no. 4, pp. 5822–5830, 2023.

[21] Z. Chen, L. Kuang, B. Liu, and Z. Ni, "Fast and Efficient Phase-Only Beam Nulling for NGSO Satellite Communication Systems," *IEEE Wireless Communications Letters*, vol. 13, no. 12, pp. 3261–3265, 2024.

[22] C. Vahapoglu, T. J. O'Shea, T. Roy, and S. Ulukus, "Deep learning based uplink multi-user simo beamforming design," in *2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 2024, pp. 329–333.

[23] P. Ramezanpour and M.-R. Mosavi, "Two-Stage Beamforming for Rejecting Interferences Using Deep Neural Networks," *IEEE Systems Journal*, vol. 15, no. 3, pp. 4439–4447, 2021.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is All You Need," *Advances in neural information processing systems*, vol. 30, 2017.

[25] R. Giuliano, E. Innocenti, F. Mazzenga, A. Vizzarri, L. Di Nunzio, P. B. Divakarachari, and I. Habib, "Transformer Neural Network for Throughput Improvement in Non-Terrestrial Networks," in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, 2023, pp. 1–6.

[26] S. Zhang, S. Zhang, W. Yuan, and T. Q. S. Quek, "Transformer-Empowered Predictive Beamforming for Rate-Splitting Multiple Access in Non-Terrestrial Networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 12, pp. 19 776–19 788, 2024.

[27] E. Lagunas, "FNR SmartSpace Project: Leveraging AI to Empower the Next Generation of Satellite Communications," (Accessed on 12/05/2025). [Online]. Available: https://fnr-smartspace-project.uni.lu/datasets/

[28] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[29] L. You, K.-X. Li, J. Wang, X. Gao, X.-G. Xia, and B. Ottersten, "Massive MIMO Transmission for LEO Satellite Communications," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1851–1865, 2020.

[30] W. Zhang, "Two-ray channel models with doppler effects for LEO satellite communications," *ITU Journal on Future and Evolving Technologies*, vol. 5, no. 2, pp. 243–259, 2024.

[31] 3GPP, "NR; Radio Resource Control (RRC); Protocol Specification," 3GPP, Tech. Rep., September 2020.

[32] W. L. Stutzman and G. A. Thiele, *Antenna theory and design*. John Wiley & Sons, 2012.

[33] J. Andrés Vásquez-Peralvo, J. Querol, F. Ortíz, J. Luis González Rios, E. Lagunas, L. Manuel Garcés-Socorrás, J. Carlos Merlano Duncan, M. O. K. Mendonça, and S. Chatzinotas, "Multibeam Beamforming for Direct Radiating Arrays in Satellite Communications Using Genetic Algorithm," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 2343–2357, 2024.

[34] H. Cox, R. Zeskind, and M. Owen, "Robust Adaptive Beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[35] D. Brooker, K. L. Gemba, and L. T. Fialkowski, "Overcoming snapshot-deficient measurements with knowledge-aided approaches," *JASA Express Letters*, vol. 2, no. 5, 2022.

[36] L. C. Godara, *Smart Antennas*. CRC Press, 2004.

[37] P. Zhang, L. Pan, T. Laohapensaeng, and M. Chongcheawchamnan, "Hybrid Beamforming Based on an Unsupervised Deep Learning Network for Downlink Channels With Imperfect CSI," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1543–1547, 2022.

[38] M. Inc., "Interference from satellite constellation on communications link - MATLAB & Simulink - MathWorks Benelux," (Accessed on 12/05/2025). [Online]. Available: https://nl.mathworks.com/help/satcom/ug/interference-from-satellite-constellation-on-comms-link.html

[39] F. Ortiz, V. M. Baeza, L. M. Garces-Socarras, J. A. Vásquez-Peralvo, J. L. Gonzalez, G. Fontanesi, E. Lagunas, J. Querol, and S. Chatzinotas, "Onboard Processing in Satellite Communications Using AI Accelerators," *Aerospace*, vol. 10, no. 2, p. 101, Jan. 2023.

[40] C. Lu, H. Li, and Z. Lin, "Optimized projections for compressed sensing via direct mutual coherence minimization," *Signal Processing*, vol. 151, pp. 45–55, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168418301464

Almoatssimbillah Saifaldawla (Graduate Student Member, IEEE) received his B.Sc. (Honors) in Electronics Engineering Technology from the University of Gezira (UofG), Sudan, in 2018, and his M.Eng. in Communication and Information Engineering from Chongqing University of Posts and Telecommunications (CQUPT), China, in 2022. He is currently a Ph.D. candidate in Computer Science and Computer Engineering with the SIGCOM research group at the Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg. His research focuses on wireless communication, resource management, and the applications of machine learning in satellite communication.

Flor Ortiz (Member, IEEE) received her B.S. degree in telecommunications engineering and her M.S. degree in electrical engineering-telecommunications from the Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico, in 2015 and 2016, respectively. Flor obtained her Ph.D. in Telecommunication Engineering (September 2021) at the Universidad Politecnica de Madrid (UPM), Madrid, Spain. During her Ph.D., she performed a research period at the University of Bologna, Bologna, Italy. She started a close collaboration between the UPM and the University of Bologna, opening a new research line for both groups on applying Machine Learning for radio resource management. She joined as a Research Associate at the Interdisciplinary Centre for Security, Reliability, and Trust (SnT) at the University of Luxembourg. Her research interests are focused on implementing cutting-edge Machine Learning techniques including Continual Learning and and Neuromorphic Computing for operations in Satellite Communications systems.

Eva Lagunas (Senior Member, IEEE) received the MSc and PhD degrees in telecommunications engineering from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 2010 and 2014, respectively. She was Research Assistant within the Department of Signal Theory and Communications, UPC, from 2009 to 2013. In 2009 she was a guest research assistant within the Department of Information Engineering, University of Pisa, Italy. From November 2011 to May 2012 she held a visiting research appointment at the Center for Advanced Communications (CAC), Villanova University, PA, USA. Between 2022-2023, she spent 6 months at Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), Spain. In 2014, she joined the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, where she holds an Assistant Professor position. Her research interests include radio resource management and general wireless networks optimization.

Symeon Chatzinotas (Fellow, IEEE) is currently Full Professor / Chief Scientist I and Head of the research group SIGCOM in the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. In parallel, he is an Adjunct Professor in the Department of Electronic Systems, Norwegian University of Science and Technology, an Eminent Scholar of the Kyung Hee University, Korea and a Collaborating Scholar of the Institute of Informatics & Telecommunications, National Center for Scientific Research "Demokritos". In the past, he has been a Visiting Professor at EPFL, Switzerland and University of Parma, Italy and contributed in numerous R&D projects for the Institute of Telematics and Informatics, Center of Research and Technology Hellas and Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey. He has received the M.Eng. in Telecommunications from Aristotle University of Thessaloniki, Greece and the M.Sc. and Ph.D. in Electronic Engineering from University of Surrey, UK in 2003, 2006 and 2009 respectively. He has authored more than 800 technical papers in refereed international journals, conferences and scientific books and has received numerous awards and recognitions, including the IEEE Fellowship and an IEEE Distinguished Contributions Award. He has served in the editorial board of npj Wireless Technology, IEEE Transactions on Communications, IEEE Open Journal of Vehicular Technology and the International Journal of Satellite Communications and Networking.