Transformer Based Active Sensing for Generalizable Two-Sided Beam Alignment

Zhongze Zhang*,†, Wei Yu*, Jingge Zhu† and Jamie Evans†

*Department of Electrical and Computer Engineering, University of Toronto, Canada

†Department of Electrical and Electronic Engineering, University of Melbourne, Australia

ufo.zhang@mail.utoronto.ca, weiyu@ece.utoronto.ca, {jingge.zhu, jse}@unimelb.edu.au

***Abstract*—Efficient two-sided beam alignment is critical for maximizing wireless communication performance in mmWave systems. Recently, a ping-pong pilot-based active sensing method has been proposed to enable iterative refinement of beam alignment by alternating pilot transmissions between the transmitter (Tx) and receiver (Rx), allowing both ends to update their beam directions based on received signal measurements. This existing state-of-the-art approach relies on using machine learning algorithm based on long short-term memory (LSTM) network to process the sequential updates, but their ability to generalize across diverse channel conditions is still limited. In this work, we propose a transformer-based active sensing framework that leverages self-attention to model complex spatial-temporal relationships and to adapt to varying channel conditions. The proposed method efficiently processes dynamically growing pilot measurements and focuses on the most informative input to enhance generalization without retraining. Experimental results demonstrate that this new approach outperforms the LSTM baseline across a mixture of wireless environments.**

I. Introduction

Millimeter-wave (mmWave) communication systems rely on highly directional beamforming to overcome severe propagation loss. Efficient beam alignment between the transmitter (Tx) and receiver (Rx) is essential for maximizing link quality, especially in dynamic wireless environments where channel characteristics vary significantly due to mobility and blockage. Traditional methods such as exhaustive beam search and hierarchical codebook training incur high overhead and limited adaptability. To address these limitations, active sensing techniques have emerged as a promising solution. In particular, the recent work [1] proposes the ping-pong pilot protocol to enable two-sided beam refinement by alternating pilot transmissions between the Tx and Rx. This sequential feedback-driven scheme allows each side to iteratively update its beamforming strategy based on its own local pilot observations, without requiring channel state information or explicit feedback exchange and has shown to have excellent performance.

To model the sequential nature of the ping-pong protocol, prior work has applied recurrent neural networks (RNNs), particularly the long short-term memory (LSTM) based architectures, to learn beamforming policies from pilot feedback [1]–[4]. However, RNNs face limitations that hinder their ability to generalize across diverse wireless environments. Most notably, they compress the entire observation history into a single hidden state, which limits the model's capacity to selectively preserve useful past information—especially when the environment exhibits sparse and non-stationary structure. Moreover, the recurrence mechanism enforces a strict sequential processing structure and complicates the learning of long-range dependencies, making it harder to reuse information from earlier rounds when it becomes relevant later. These factors may restrict the generalizability of RNN-based sensing policies when evaluated across different channel models, signal-to-noise ratio (SNR) regimes, or propagation sparsities.

Recently, transformers have rapidly become foundational tools across machine learning, demonstrating impressive capabilities in large language models (LLMs). We recognize a compelling analogy between active sensing and interaction with LLM, where LLMs operate by generating responses conditioned on a sequence of prior inputs, dynamically adapting to context as interactions evolve. This iterative, feedback-driven process closely mirrors the class of active sensing problem in wireless communication, where each received pilot signal informs the next beamforming or sensing decision. Just as an LLM tailors its output to a sequence of user queries, an active sensing system tailors its sensing strategy to the sequence of received signals. Furthermore, LLMs have shown remarkdable generalizability for a wide range of tasks.

Motivated by the recent advances in LLMs, we propose a transformer-based active sensing architecture for two-sided beam alignment in mmWave MIMO systems. By framing beam alignment as a sequence-to-decision problem, our approach uses masked self-attention to recursively process the full pilot-beamformer history and infer the next beamforming action at each round. The transformer's ability to attend to all previous rounds—rather than compressing them—enables the model to extract invariant structures from the pilot observations, even as the number of paths, SNR, and angular characteristics vary across environments. We hypothesize that this inductive bias—assigning adaptive importance to past measurements—is key to improving generalization. Recent works have explored the use of transformers in wireless communication, including symbol detection via in-context learning [5]–[7] and the development of foundation models for physical-layer tasks [8]–[10], showing that transformer architectures are well-suited to capturing complex spatiotemporal dependencies and adapting across heterogeneous environments.

The proposed architecture introduces a pair of causal transformer decoders—one at the Tx and one at the Rx—each

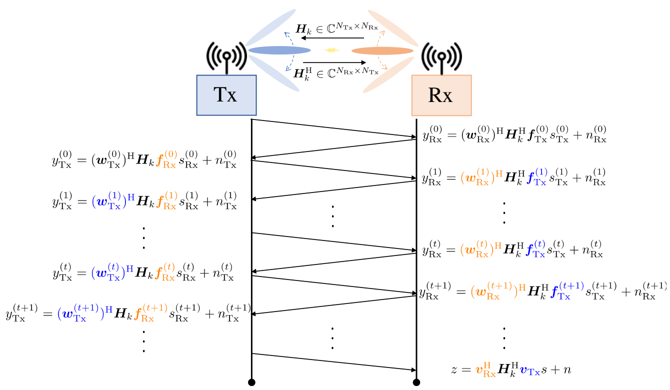

Fig. 1: Beam alignment using ping-pong pilot protocol. The sensing beamformers designed at Tx and Rx are highlighted as blue and orange, respectively. The initial sensing beamformers are fixed, hence not colored.

of which processes a growing sequence of locally observed pilot responses and beamformer vectors. At every pilot round, the model constructs position-indexed input tokens from prior rounds, embeds them into a latent space, and applies masked self-attention to extract context-aware representations of the history. The final attended history is then decoded into a pair of beamformers: one used to transmit the current pilot and one used to receive in the subsequent round. After several rounds of pilot exchange, each side processes the full observation history to synthesize a data-phase beamformer that maximizes end-to-end gain.

To evaluate generalization, we consider a mixture of channel environments, each characterized by a different number of propagation paths and operating SNR. The transformer decoders are trained end-to-end over this mixture to maximize the expected beamforming gain, without access to environment labels. Transformer-based architecture offers two key advantages over recurrent models: i) self-attention allows the model to revisit and prioritize informative pilot-beamformer interactions throughout the sequence, unlike RNNs which compress all past information into a fixed-size hidden state; and ii) the architecture accommodates variable-length input by design, enabling seamless adaptation as the number of pilot rounds increases. Numerical results demonstrate that the proposed method outperforms LSTM-based baselines, particularly in few-pilot regimes under diverse channel conditions.

II. System Model and Problem Formulation

A. System Model

We consider a narrowband mmWave MIMO system consisting of a Tx and a Rx, each equipped with a uniform linear array (ULA) of N_{Tx} and N_{Rx} antennas, respectively. Both transceivers operate under a single radio frequency (RF) chain hybrid beamforming architecture. We use $\boldsymbol{v}_{\mathrm{Tx}} \in \mathbb{C}^{N_{\mathrm{Tx}}}$ and $\boldsymbol{v}_{\mathrm{Rx}} \in \mathbb{C}^{N_{\mathrm{Rx}}}$ to denote the beamforming vectors at the Tx and the Rx, respectively. To ensure a strong communication link, the beamforming pair $\{\boldsymbol{v}_{\mathrm{Tx}}, \boldsymbol{v}_{\mathrm{Rx}}\}$ should be jointly designed based on the channel state information (CSI) to maximize the achievable rate. This joint optimization task is known as the two-sided beam alignment problem.

Here, we denote $\boldsymbol{H} \in \mathbb{C}^{N_{\mathrm{Tx}} \times N_{\mathrm{Rx}}}$ as the uplink channel matrix from the Rx to the Tx, and $\boldsymbol{H}^{\mathrm{H}}$ as the downlink channel, where we assume that the system follows a time-division duplex (TDD) protocol with channel reciprocity. Further, a block-fading channel model is adopted in which the channel coefficients are assumed to remain constant over multiple time frames during a coherence block, but change independently from block to block. We model the mmWave channel as a sparse multipath channel

$$\boldsymbol{H} = \sum_{i=1}^{L_p} \beta_i \boldsymbol{u}_{\mathrm{Tx}}(\theta_i) \boldsymbol{u}_{\mathrm{Rx}}^{\mathrm{H}}(\phi_i), \tag{1}$$

where L_p denotes the number of propagation paths, β_i denotes the complex channel gain, θ_i and ϕ_i denotes the angle-of-departure (AoD) and the angle-of-arrival (AoA) for the i-th path, and $\boldsymbol{u}_{\mathrm{Tx}}(\cdot)$, $\boldsymbol{u}_{\mathrm{Rx}}(\cdot)$ are the transmit and receive steering vectors.

We assume the channel model is drawn from a mixture of dataset. To promote generalization capability across deployment environments, we generate a finite mixture of K fixed environment models, denoted by $\mathcal{E}_1, \cdots, \mathcal{E}_K$. In each environment $\mathcal{E}_k$, the channel is generated from some sparse multipath channel model with a distinct number of propagation paths at some operating SNR. Parameters such as angular distribution and antenna configurations are fixed across environments. The environment model $\mathcal{E}_k, k \in \{1, 2, \cdots, K\}$ is sampled uniformly at random. A channel realization $\boldsymbol{H}$ is then generated accordingly. The alignment strategy must operate without the knowledge of the environment model.

During data transmission, the received signal at the Rx is modeled as:

$$z = \boldsymbol{v}_{\mathrm{Rx}}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{v}_{\mathrm{Tx}} s + n, \tag{2}$$

where $s \in \mathbb{C}$ is the transmit data symbol with $\mathbb{E}[|s|^2] = P$, and $n \sim \mathcal{CN}(0, \sigma^2)$ is the complex Gaussian noise. The beam alignment objective is to maximize the squared beamforming gain $|\boldsymbol{v}_{\mathrm{Rx}}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{v}_{\mathrm{Tx}}|^2$.

If $\boldsymbol{H}$ is known, the optimal transmit and receive beamforming vectors should align with the left and right singular vectors associated with the largest singular value of $\boldsymbol{H}$. However, in practice, $\boldsymbol{H}$ is unknown and must be estimated through pilot training. Conventional pilot training strategies typically use randomly selected beamformers from a codebook to probe the channel, but the performance is restricted by the quality of the codebook [11]. To overcome these limitations, authors of [1] propose a novel ping-pong pilot transmission protocol that operates without feedback and demonstrates strong beam alignment performance with reduced pilot training overhead. We adopt this protocol in this paper.

B. Ping-Pong Pilot Protocol

The ping-pong pilot protocol enables adaptive beam refinement by sending pilots back and forth between the Tx and Rx. Each side iteratively accumulates measurements to update its beampatterns to improve alignment. Here, we remark that the beamforming vectors during the pilot training phase are known as sensing vectors to distinguish them from the beamforming vectors $\{\boldsymbol{v}_{\mathrm{Tx}}, \boldsymbol{v}_{\mathrm{Rx}}\}$ in the data transmission phase.

In each round, the Tx transmits a pilot to the Rx. The Rx, upon receiving the pilot, transmits a return pilot back to the Tx, as shown in Fig. 1. Specifically, in the t-th round, the Tx transmits a known pilot symbol $s_{\mathrm{Tx}}^{(t)}$ (under a power constraint $\mathbb{E}[|s_{\mathrm{Tx}}^{(t)}|^2] \leq P_1$) using a transmit sensing beamforming vector $\boldsymbol{f}_{\mathrm{Tx}}^{(t)} \in \mathbb{C}^{N_{\mathrm{Tx}}}$. The Rx receives the pilot with a receive sensing beamforming vector $\boldsymbol{w}_{\mathrm{Rx}}^{(t)} \in \mathbb{C}^{N_{\mathrm{Rx}}}$. The received pilot at the Rx is given by

$$y_{\mathrm{Rx}}^{(t)} = (\boldsymbol{w}_{\mathrm{Rx}}^{(t)})^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{f}_{\mathrm{Tx}}^{(t)} s_{\mathrm{Tx}}^{(t)} + n_{\mathrm{Rx}}^{(t)}, \ t = 0, \cdots, T-1, \quad (3)$$

where $n_{\mathrm{Rx}}^{(t)} \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive Gaussian noise. Upon receiving the pilot, the Rx transmits a return pilot using its transmit sensing beamforming vector $\boldsymbol{f}_{\mathrm{Rx}}^{(t)}$, received at the Tx with receive sensing beamforming vector $\boldsymbol{w}_{\mathrm{Tx}}^{(t)}$. The received return pilot at the Tx is as follows

$$y_{\mathrm{Tx}}^{(t)} = (\boldsymbol{w}_{\mathrm{Tx}}^{(t)})^{\mathrm{H}} \boldsymbol{H} \boldsymbol{f}_{\mathrm{Rx}}^{(t)} s_{\mathrm{Rx}}^{(t)} + n_{\mathrm{Tx}}^{(t)}, \ t = 0, \cdots, T-1, \quad (4)$$

where $s_{\mathrm{Rx}}^{(t)}$ denotes the return pilot symbol under a power constraint $\mathbb{E}[|s_{\mathrm{Rx}}^{(t)}|^2] \leq P_2$, and $n_{\mathrm{Tx}}^{(t)} \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive Gaussian noise.

As t increases, the Tx and the Rx log their own received pilots and sensing vectors, forming local observation histories:

$$O_{\mathrm{Tx}}^{(t)} = \left\{ \left(y_{\mathrm{Tx}}^{(0)}, \boldsymbol{w}_{\mathrm{Tx}}^{(0)}, \boldsymbol{f}_{\mathrm{Tx}}^{(0)}\right), \cdots, \left(y_{\mathrm{Tx}}^{(t)}, \boldsymbol{w}_{\mathrm{Tx}}^{(t)}, \boldsymbol{f}_{\mathrm{Tx}}^{(t)}\right) \right\}, \quad (5a)$$

$$O_{\mathrm{Rx}}^{(t)} = \left\{ \left(y_{\mathrm{Rx}}^{(0)}, \boldsymbol{w}_{\mathrm{Rx}}^{(0)}, \boldsymbol{f}_{\mathrm{Rx}}^{(0)}\right), \cdots, \left(y_{\mathrm{Rx}}^{(t)}, \boldsymbol{w}_{\mathrm{Rx}}^{(t)}, \boldsymbol{f}_{\mathrm{Rx}}^{(t)}\right) \right\}. \quad (5b)$$

After T rounds of pilot transmission, a total of $2T$ pilot symbols are transmitted between the Tx and the Rx. Each of the Tx and Rx acquires sufficient information from their local histories, $O_{\mathrm{Tx}}^{(T-1)}$ and $O_{\mathrm{Rx}}^{(T-1)}$, to design their beamforming vectors for the data transmission phase.

C. Problem Formulation

The task of two-sided beam alignment is formulated as a sequential decision-making problem, where the Tx and the Rx adaptively design their transmit and receive sensing beamformers over multiple rounds of pilot exchange, as more measurements become available.

At the t-th transmission round, the Tx transmits a pilot to the Rx. The Rx, based on the complete local observation history up to round t, selects a transmit sensing beamforming vector to be used in the same round as well as a receive sensing beamforming vector for use in the subsequent round. These decisions are expressed as follows.

$$\boldsymbol{f}_{\mathrm{Rx}}^{(t)} = \mathcal{G}_{\mathrm{Rx}}^{(t)}\left(O_{\mathrm{Rx}}^{(t)}\right), \quad \boldsymbol{w}_{\mathrm{Rx}}^{(t+1)} = \tilde{\mathcal{G}}_{\mathrm{Rx}}^{(t)}\left(O_{\mathrm{Rx}}^{(t)}\right), \quad (6)$$

where $\mathcal{G}_{\mathrm{Rx}}^{(t)}(\cdot)$ and $\tilde{\mathcal{G}}_{\mathrm{Rx}}^{(t)}(\cdot)$ are functions that map the Rx observation histories up to round t to transmit and receive sensing beamforming vectors, respectively.

Likewise, the Tx designs a transmit sensing beamforming vector as well as a receive sensing beamforming vector for use in the round $t+1$, based on the local observation histories up to round t, as follows

$$\boldsymbol{f}_{\mathrm{Tx}}^{(t+1)} = \mathcal{G}_{\mathrm{Tx}}^{(t)}\left(O_{\mathrm{Tx}}^{(t)}\right), \quad \boldsymbol{w}_{\mathrm{Tx}}^{(t+1)} = \tilde{\mathcal{G}}_{\mathrm{Tx}}^{(t)}\left(O_{\mathrm{Tx}}^{(t)}\right), \quad (7)$$

where $\mathcal{G}_{\mathrm{Tx}}^{(t)}(\cdot)$ and $\tilde{\mathcal{G}}_{\mathrm{Tx}}^{(t)}(\cdot)$ are mapping from the Tx observation histories up to round t to transmit and receive sensing beamforming vectors respectively.

After T pilot rounds, the Tx and the Rx generate their final beamforming vector for data transmission

$$\boldsymbol{v}_{\mathrm{Tx}} = \mathcal{F}_{\mathrm{Tx}}\left(O_{\mathrm{Tx}}^{(T-1)}\right), \quad \boldsymbol{v}_{\mathrm{Rx}} = \mathcal{F}_{\mathrm{Rx}}\left(O_{\mathrm{Rx}}^{(T-1)}\right), \quad (8)$$

where $\mathcal{F}_{\mathrm{Tx}}(\cdot)$ and $\mathcal{F}_{\mathrm{Rx}}(\cdot)$ denote the mapping from full observation histories to the final data beamforming vectors.

The beamforming alignment task is to jointly design the sensing beamformer mapping in (6), (7) and the final beamforming mapping in (8), such that the average beamforming gain is maximized over the mixture of environments:

$$\underset{\mathcal{S}}{\text{maximize}} \quad \mathbb{E}\ |\boldsymbol{v}_{\mathrm{Rx}}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{v}_{\mathrm{Tx}}|^2 \quad (9a)$$

$$\text{subject to} \quad (6),\ (7),\ (8), \quad (9b)$$

where the optimization variables are a set of functions

$$\mathcal{S} : \{\mathcal{F}_{\mathrm{Tx}}(\cdot), \mathcal{F}_{\mathrm{Rx}}(\cdot), \{\mathcal{G}_{\mathrm{Rx}}^{(t)}(\cdot)\}_{t=0}^{T-1}, \{\mathcal{G}_{\mathrm{Tx}}^{(t)}(\cdot), \tilde{\mathcal{G}}_{\mathrm{Tx}}^{(t)}(\cdot), \tilde{\mathcal{G}}_{\mathrm{Rx}}^{(t)}(\cdot)\}_{t=0}^{T-2}\}. \quad (10)$$

III. Proposed Transformer-based Solution

We adopt a data-driven approach to beam alignment by learning a mapping from previously observed pilot responses and beamformer actions to the next beamforming decision. This formulation aligns naturally with the transformer decoder architecture, which is well-suited for autoregressive sequence modeling. Unlike recurrent models such as LSTMs, which compress the entire sequence into a single hidden state, transformers use self-attention to maintain direct access to the full input history. This design is better suited for generalization, particularly in variable environments, because it allows the model to focus on informative parts of the observation history, even if they appear early in the sequence. In our application, this means that the model can dynamically attend to pilot-beamformer interactions that reveal stable channel features—such as dominant paths or angular structure—even as the SNR or path count changes.

A. Overview of the Transformer Model

At the core of the transformer architecture is the query-key-value attention mechanism. The transformer decoder operates by embedding each token into a fixed-dimensional vector space and adding a positional encoding to preserve sequence order. Each embedded token is then projected into query, key, and value vectors. The attention mechanism computes a similarity between the current token's query and all previous keys to determine which values to attend to. Conceptually, the values contain knowledge from prior rounds, the keys determine which pieces of that knowledge are relevant, and the query defines what the model is currently looking for. As

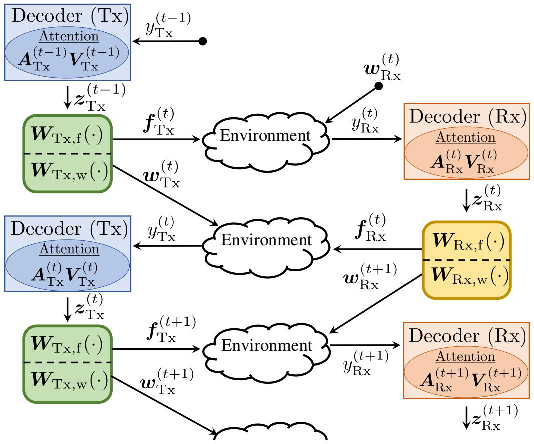

Fig. 2: Neural network architecture.

the sequence grows over time and passes through successive transformer layers, the query, key and value representations evolve from shallow encodings of direct measurements into more abstract and aggregated forms that capture structural channel information. This mechanism enables context-aware decision making without compressing the entire history.

We use a causal transformer decoder, rather than an encoder architecture, because beam alignment is inherently sequential and autoregressive: the received pilot at round t depends on the beamformer chosen at round t, and that beamformer must be selected based on observations from rounds $< t$. A transformer encoder, by contrast, assumes access to the full input sequence and allows bidirectional attention across tokens. The causal mask in the decoder ensures that each output depends only on past and current tokens. As the number of pilot rounds increases, the architecture processes longer sequences while retaining fixed model parameters, enabling the model to adapt naturally to different sensing horizons.

B. Input Representation and Tokenization

We construct observation tokens from observation histories (historical pilot measurements and beamforming vectors) to serve as input to the transformer decoder. Recall that $O_{\mathrm{Tx}}^{(t)}$ denotes the observation histories at the Tx up to round t. We group the information available at the Tx at the i-th round to a tuple

$$T_{\mathrm{Tx}}^{(i)} = (y_{\mathrm{Tx}}^{(i)}, \boldsymbol{w}_{\mathrm{Tx}}^{(i)}, \boldsymbol{f}_{\mathrm{Tx}}^{(i)}), \ \ i = 0, \cdots, t. \tag{11}$$

Subsequently, the tuple is embedded into a real-valued vector space using a learned linear projection. Specifically, let $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ denote the real and imaginary components of a complex value. The embedded observation at round i is

$$\boldsymbol{e}_{\mathrm{Tx}}^{(i)} = \boldsymbol{W}_{\mathrm{Tx,E}}\left(\mathcal{R}(T_{\mathrm{Tx}}^{(i)}), \mathcal{I}(T_{\mathrm{Tx}}^{(i)})\right), \tag{12}$$

where $\boldsymbol{W}_{\mathrm{Tx,E}}(\cdot) : \mathbb{R}^{2+4N_{\mathrm{Tx}}} \to \mathbb{R}^d$ is a learned linear map, and d is the model's hidden dimension. Similarly, $\boldsymbol{W}_{\mathrm{Rx,E}}(\cdot) : \mathbb{R}^{2+4N_{\mathrm{Rx}}} \to \mathbb{R}^d$ is a learned linear map with the same hidden dimension at the Rx transformer decoder.

Finally, as transformers are inherently permutation-invariant, we incorporate sinusoidal positional encodings $\boldsymbol{p}^{(i)} \in \mathbb{R}^d$ into embedded observation to preserve the order of observations across rounds. An observation token is given by

$$\tilde{\boldsymbol{e}}_{\mathrm{Tx}}^{(i)} = \boldsymbol{e}_{\mathrm{Tx}}^{(i)} + \boldsymbol{p}^{(i)}. \tag{13}$$

C. Causal Attention-based Sequence Processing

At the round t, the input to the transformer decoder at the Tx is a sequence of observation tokens

$$\tilde{\boldsymbol{E}}_{\mathrm{Tx}}^{(t)} = [\tilde{\boldsymbol{e}}_{\mathrm{Tx}}^{(0)}, \tilde{\boldsymbol{e}}_{\mathrm{Tx}}^{(1)}, \cdots, \tilde{\boldsymbol{e}}_{\mathrm{Tx}}^{(t)}] \in \mathbb{R}^{(t+1)\times d}. \tag{14}$$

Within the transformer decoder, the self-attention mechanism operates on $\tilde{\boldsymbol{E}}_{\mathrm{Tx}}^{(t)}$ as

$$\boldsymbol{Q}_{\mathrm{Tx}}^{(t)} = \boldsymbol{W}_{\mathrm{Tx,Q}}(\tilde{\boldsymbol{E}}_{\mathrm{Tx}}^{(t)}), \tag{15a}$$
$$\boldsymbol{K}_{\mathrm{Tx}}^{(t)} = \boldsymbol{W}_{\mathrm{Tx,K}}(\tilde{\boldsymbol{E}}_{\mathrm{Tx}}^{(t)}), \tag{15b}$$
$$\boldsymbol{V}_{\mathrm{Tx}}^{(t)} = \boldsymbol{W}_{\mathrm{Tx,V}}(\tilde{\boldsymbol{E}}_{\mathrm{Tx}}^{(t)}), \tag{15c}$$

where $\boldsymbol{W}_{\mathrm{Tx,Q}}, \boldsymbol{W}_{\mathrm{Tx,K}}, \boldsymbol{W}_{\mathrm{Tx,V}} \in \mathbb{R}^{d\times d}$ are learned linear maps. The attention matrix $\boldsymbol{A}_{\mathrm{Tx}}^{(t)}$ is computed:

$$\boldsymbol{A}_{\mathrm{Tx}}^{(t)} = \mathrm{softmax}\left(\frac{\boldsymbol{Q}_{\mathrm{Tx}}^{(t)}(\boldsymbol{K}_{\mathrm{Tx}}^{(t)})^\top}{\sqrt{d}} + \boldsymbol{M}^{(t)}\right), \tag{16}$$

where $\boldsymbol{M}^{(t)} \in \mathbb{R}^{(t+1)\times(t+1)}$ is the causal mask to ensure $[\boldsymbol{A}_{\mathrm{Tx}}^{(t)}]_{ij} = 0$ for $j > i$. We use causal mask to ensure that the token at position i only attends to current and past tokens, preserving the temporal structure of the beam alignment task. This is achieved by constructing the mask as a lower triangular matrix with $-\infty$ above the main diagonal, effectively masking out the upper triangle of the attention matrix.

The attended history matrix at the Tx is as follows:

$$\boldsymbol{Z}_{\mathrm{Tx}}^{(t)} = \boldsymbol{A}_{\mathrm{Tx}}^{(t)}\boldsymbol{V}_{\mathrm{Tx}}^{(t)} \in \mathbb{R}^{(t+1)\times d}. \tag{17}$$

Each row $\boldsymbol{z}_{\mathrm{Tx}}^{(i)} \in \mathbb{R}^d$ represents the attended history vector up to round i. We can interpret $\boldsymbol{z}_{\mathrm{Tx}}^{(i)}$ as an evolving representation of the belief about the channel. The attention matrix $\boldsymbol{A}_{\mathrm{Tx}}^{(t)}$ plays a central role in updating this belief by computing weighted averages over previous observations to filter the more informative observations. The final row $\boldsymbol{z}_{\mathrm{Tx}}^{(t)}$ contains the latest attended history and is used to generate the transmit sensing beamformer and receive sensing beamformer in the round $t+1$:

$$\boldsymbol{f}_{\mathrm{Tx}}^{(t+1)} = \boldsymbol{W}_{\mathrm{Tx,f}}(\boldsymbol{z}_{\mathrm{Tx}}^{(t)}) \in \mathbb{C}^{N_{\mathrm{Tx}}}, \tag{18a}$$
$$\boldsymbol{w}_{\mathrm{Tx}}^{(t+1)} = \boldsymbol{W}_{\mathrm{Tx,w}}(\boldsymbol{z}_{\mathrm{Tx}}^{(t)}) \in \mathbb{C}^{N_{\mathrm{Tx}}}, \tag{18b}$$

where $\boldsymbol{W}_{\mathrm{Tx,f}}$ and $\boldsymbol{W}_{\mathrm{Tx,w}}$ are deep neural networks (DNNs) that perform such a mapping $\mathbb{R}^d \to \mathbb{C}^{N_{\mathrm{Tx}}}$. The transformer decoder at the Rx follows the same methodology from (11)-(18) to design its own sensing beamformer pairs $\{\boldsymbol{f}_{\mathrm{Rx}}^{(t+1)}, \boldsymbol{w}_{\mathrm{Rx}}^{(t+1)}\}$ from its attended history $\boldsymbol{z}_{\mathrm{Rx}}^{(t)}$ up to round t.

Here, we note that the model parameter $\boldsymbol{W}_{\mathrm{Tx},\Delta}$, $\Delta \in \{\mathrm{E, Q, K, V, f, w}\}$ are all independent of t. This design choice allows a single transformer decoder to be applied recurrently, even as the input sequence grows with time. While the input

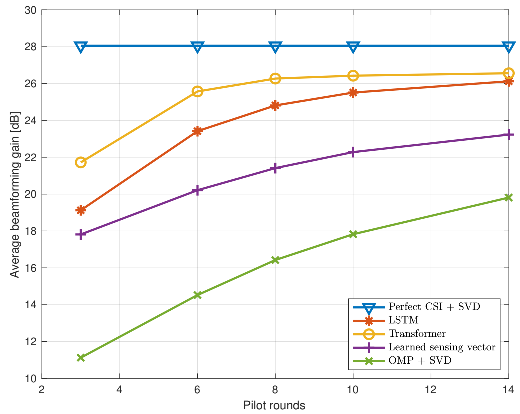

Fig. 3: Beamforming gain vs. pilot rounds, $K = 25$, $N_{\mathrm{Tx}} = 64$, $N_{\mathrm{Rx}} = 32$.

length increases linearly with t, the dimensionality per token stays fixed at d which does not change the model architecture. This is analogous to how LLMs handle user inputs of varying lengths without modifying the underlying architecture.

After T rounds of pilot transmission, the final beamformer for data transmission at the Tx and Rx are produced based on the complete attended histories $\boldsymbol{z}_{\mathrm{Tx}}^{(T-1)}$ and $\boldsymbol{z}_{\mathrm{Rx}}^{(T-1)}$

$$\boldsymbol{v}_{\mathrm{Tx}} = \ell_{\mathrm{Tx}}(\boldsymbol{z}_{\mathrm{Tx}}^{(T-1)}) \in \mathbb{C}^{N_{\mathrm{Tx}}}, \tag{19a}$$

$$\boldsymbol{v}_{\mathrm{Rx}} = \ell_{\mathrm{Rx}}(\boldsymbol{z}_{\mathrm{Rx}}^{(T-1)}) \in \mathbb{C}^{N_{\mathrm{Rx}}}, \tag{19b}$$

where $\ell_{\mathrm{Tx}}(\cdot)$ and $\ell_{\mathrm{Rx}}(\cdot)$ are DNNs. The entire model is trained end-to-end to maximize the final beamforming gain after T rounds of ping-pong pilot exchange over a mixture of dataset with K environments of different number of probation paths and operating SNR.

IV. Numerical Results

In this section, we evaluate the performance of the proposed algorithm for the two-sided beam alignment problem. We consider a system with $N_{\mathrm{Tx}} = 64$ antennas at the Tx and $N_{\mathrm{Rx}} = 32$ antennas at the Rx. The number of paths between the Tx and Rx is set to $\{3, 5, 8, 10, 15\}$. The operating SNR (in dB) ranges in $\{-10, -5, 0, 5, 10\}$. For each channel realization. the AoAs/AoDs are uniformly distributed in $[-60^\circ, 60^\circ]$, and the complex fading coefficients are randomly drawn from the distribution $\mathcal{CN}(0, 1)$.

Compressive sensing with random vector [12]: This method adopts the compressive sensing method in which the orthogonal matching pursuit (OMP) algorithm is used to estimate the channel $\boldsymbol{H}$. The sensing vectors are randomly generated. Given the estimated channel, the beamformer for data transmission are given by SVD method.

DNN with learned RIS configurations: The sequence of sensing beamformers is non-adaptive and is learned from the channel statistics in the training data. Two deep neural network of dimensions $[200, 200, 200, 2N_{\mathrm{Tx}}]$ and $[200, 200, 200, 2N_{\mathrm{Rx}}]$ map received pilots over T time frames to Tx beamformer and Rx beamformer respectively.

When evaluating in a single-environment setting ($K = 1$), the transformer performs similarly to the LSTM, suggesting limited advantage in homogeneous conditions. To evaluate generalization, we consider a multi-environment setting ($K = 25$) where each channel realization is drawn from a mixture of different path counts and SNR values. As shown in Fig. 3, the proposed transformer model outperforms LSTM and non-adaptive baselines across all pilot lengths. The performance gap between the transformer and LSTM persists even as the number of pilot rounds increases, suggesting that the transformer's inductive bias and ability to attend over full observation sequences provide better generalization across heterogeneous environments.

V. Conclusion

This paper presents a transformer-based architecture for two-sided active beam alignment in mmWave MIMO systems. Building on a sequential pilot exchange protocol, we formulate the beam alignment problem as a sequential decision process and propose a pair of causal transformer decoders—one at the Tx and one at the Rx—to generate adaptive sensing strategies. By processing sequences of past pilot observations and beamformers using masked self-attention, the proposed architecture learns to generalize across a mixture of propagation environments with varying SNR and path sparsity.

References

[1] T. Jiang, F. Sohrabi, and W. Yu, "Active sensing for two-sided beam alignment and reflection design using ping-pong pilots," *IEEE J. Sel. Areas Inf. Theory*, vol. 4, pp. 24–39, May 2023.

[2] T. Jiang and W. Yu, "Active sensing for reciprocal MIMO channels," *IEEE Trans. Signal Process.*, vol. 72, pp. 2905–2920, Jun. 2024.

[3] Z. Zhang and W. Yu, "Learning beamforming codebooks for active sensing with reconfigurable intelligent surface," *IEEE Trans. Wireless Commun.*, pp. 1–1, Apr. 2025.

[4] Z. Zhang, T. Jiang, and W. Yu, "Localization with reconfigurable intelligent surface: An active sensing approach," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7698–7711, Jul. 2024.

[5] V. T. Kunde, V. Rajagopalan, C. S. K. Valmeekam, K. Narayanan, S. Shakkottai, D. Kalathil, and J.-F. Chamberland, "Transformers are provably optimal in-context estimators for wireless communications," 2025. [Online]. Available: https://arxiv.org/abs/2311.00226

[6] M. Zecchin, K. Yu, and O. Simeone, "In-context learning for MIMO equalization using transformer-based sequence models," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC-W)*, Jun. 2024, pp. 1573–1578.

[7] H. Zhang, A. B. Sediq, A. Afana, and M. Erol-Kantarci, "Large language models in wireless application design: In-context learning-enhanced automatic network intrusion detection," 2024. [Online]. Available: https://arxiv.org/abs/2405.11002

[8] S. Alikhani, G. Charan, and A. Alkhateeb, "Large wireless model (LWM): A foundation model for wireless channels," 2024. [Online]. Available: https://arxiv.org/abs/2411.08872

[9] A. Aboulfotouh, A. Eshaghbeigi, and H. Abou-Zeid, "Building 6G radio foundation models with transformer architectures," 2024. [Online]. Available: https://arxiv.org/abs/2411.09996

[10] T. Yang *et al.*, "WirelessGPT: A generative pre-trained multi-task learning framework for wireless communication," 2025. [Online]. Available: https://arxiv.org/abs/2502.06877

[11] F. Sohrabi, T. Jiang, W. Cui, and W. Yu, "Active sensing for communications by learning," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1780–1794, Jun. 2022.

[12] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, Jul. 2014.