

Joint Hybrid Delay-Phase Precoding Under True-Time Delay Constraints in Wideband THz Massive MIMO Systems

Dang Qua Nguyen and Taejoon Kim

Department of Electrical Engineering and Computer Science

University of Kansas, Lawrence, KS 66045 USA

Email: {quand, taejoonkim}@ku.edu

Abstract—The beam squint effect that arises in the wideband Terahertz (THz) massive multiple-input multiple-output (MIMO) communication produces a serious array gain loss. True-time delay (TTD)-based hybrid precoding has been considered to compensate for the beam squint effect. By fixing the phase shifter (PS) precoder, a common strategy has been designing TTD precoder under the assumption of unbounded time delay values. In this paper, we present a new approach to the problem of beam squint compensation, based on the joint optimization of the TTD and PS precoders under per TTD device time delay constraints. We first derive a lower bound of the achievable rate. Then, we show that in the large system limit, the ideal analog precoder that completely compensates for the beam squint is equivalent to the one that maximizes the achievable rate lower bound. Unlike the prior approaches, our approach does not require the unbounded time delay assumption; the range of time delay values that a TTD can produce is strictly limited in our approach. Instead of focusing on the design of TTD values only, we jointly optimize both the TTD and PS values to effectively cope with the practical time delay constraints. Taking the advantage of the proposed joint TTD and PS precoder optimization approach, we quantify the minimum number of TTDs required to produce a predefined array gain performance. The simulation results illustrate the substantially improved performance with the array gain performance guarantee of the proposed joint optimization method.

Index Terms—Wideband THz massive multiple-input multiple-output (MIMO), beam squint effect, hybrid precoding, phase shifter (PS), true-time delay (TTD), and joint TTD and PS precoding.

I. INTRODUCTION

The fifth generation (5G)-&-beyond wireless systems are expected to support many new high data rate applications such as augmented reality/virtual reality (AR/VR), eHealth, and holographic telepresence [2]. To facilitate the proposed usecases, communications in the terahertz (THz) band (0.1–10 THz) have recently attracted significant interests from both academia and industry because of an abundance of wideband spectrum resources [3]–[5]; compare to the current millimeter-wave (mmWave) communications in the 5G specifications [6] that utilize a few gigahertz (GHz) bandwidth, the THz communications [3]–[5] utilize tens of GHz bandwidth. Intriguingly,

the data rates on the orders of 10 to 100 Gbps can be achieved using the currently available digital modulation techniques in THz frequencies [5], [7]. However, the envisioned THz communications face numerous challenges due to the high path losses, large power consumption, and inter-symbol interference. To deal with these challenges, the combination of hybrid massive multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) technologies has been popularly discussed [8]–[11] recently.

It is well known that as the number of antennas grows, the MIMO system can be greatly simplified in terms of beamforming and precoding [12]. In view of this, strong theoretical analyses have justified the use of a very large number of antennas at the base station [12]–[14]. This has raised a significant interest in massive MIMO systems at sub-6GHz [12]–[14] and mmWave frequencies [15]–[17]. The underlying assumption behind the works in [12]–[17] was, however, narrowband. Unlike the narrowband systems, wideband high-frequency massive MIMO OFDM systems may suffer from the substantial array gain loss across different OFDM subcarriers as the number of antennas grows due to the spatial-frequency wideband effect [18]–[21], which is also known as the beam squint effect. In particular, beam squint refers to a phenomenon in which the deviation occurs in the spatial direction of each OFDM subcarrier when the wideband OFDM is used in a very large antenna array system. The implication of beam squint is that it causes a serious achievable rate degradation, which potentially demotivates the use of OFDM in the wideband THz massive MIMO systems. Therefore, an efficient beam squint compensation is of paramount importance for the realization of wideband THz massive MIMO communications.

A. Related Works

The beam squint was previously addressed in wideband mmWave massive MIMO [18], [22] by designing the beamforming weights to produce adaptive-beamwidth beams in order to cover the squinted angles. While they show efficacy, these techniques cannot be directly extended to the THz bands due to the extremely narrow pencil beam requirement imposed by one or two orders of magnitude higher carrier frequencies. In the radar community, the beam squint effect of phased array antennas has been independently studied (e.g., see [23]–[25], and references therein). A common method to relieve

Parts of this work was presented in the IEEE International Conference on Communications (ICC) 2022 [1]. This work was supported in part by the National Science Foundation (NSF) under Grant CNS1955561, CNS2225577, CNS2212565, and the Office of Naval Research (ONR) under Grant N00014-21-1-2472.

the beam squint in radar is to employ the true-time delay (TTD) lines instead of using phase shifters (PSs) for analog beamforming [23]–[25]. In contrast with the PS-based analog beamforming that produces frequency-independent phase rotations, the TTD-based analog beamforming generates frequency-dependent phase rotations that can be used for correcting the squinted beams in the spatial domain. However, this method has limitation to be directly used in the massive MIMO systems because it requires a large number of TTD lines. Particularly, each transmit antenna needs to be fed by a dedicated TTD, leading to a high hardware cost and huge power consumption¹

Recently, there has been considerable interest in adopting TTD lines to THz hybrid massive MIMO OFDM systems to cope with beam squint [8]–[10]. Compared with the conventional TTD-based beamforming architectures in radar [23]–[25], these systems [8]–[10] employ a substantially lower number of TTD lines to maintain their power consumption as low as possible; one can think of these methods as combining a small number of TTD lines with a layer of PSs to form an analog precoder so that the capability of generating frequency-dependent phase rotations is maintained while consuming less power than the conventional TTD architecture in radar systems. Nevertheless, these approaches may still suffer from a large amount of power consumption when a large number of TTD lines is required to combat with the extremely severe beam squint present in the wideband THz massive MIMO OFDM systems. Especially, the question of how to quantify the number of TTDs per RF chain to produce a desired beam squint compensation capability remains challenging.

More recently, the TTD-based hybrid precoding architecture has been exploited to solve fast beam training [28], beam tracking [29], and user localization [30] problems. In contrary with the beam squint compensation that focuses the beam directions at every OFDM subcarrier to the same physical direction, these works [28]–[30] take advantage of the beam squint effect by spreading the beams across different OFDM subcarriers simultaneously. By doing so, information such as users' physical locations in different directions can be tracked simultaneously [30].

Most of these prior works have focused on the design of the TTD precoder while fixing the PS precoder, in which the analog precoding design is simplified because it decouples the TTD and PS precoders; the number of design variables is substantially reduced since the number of deployed TTDs is much less than the number of PSs. Furthermore, it was assumed that the TTD values increase linearly with the number of antennas without bounds. In practice, it is difficult to implement these approaches because the unbounded TTD assumption cannot be realizable. The range of time delay values that a TTD can produce is strictly limited (e.g., ≤ 508 ps [26]). Hence, to cope with the practical constraints of TTD, it is much desirable to jointly optimize both the TTD and PS precoders.

¹While the power consumption of a TTD device depends on a specific process technology (e.g., BiCMOS [26], and CMOS [27]), a typical TTD in THz consumes 100 mW [8]. It is worth noting that the power consumption of a typical PS in THz is 20 mW [8], which is much lower than that of a TTD.

B. Overview of Contributions

Motivated by the preceding discussion, we present a joint TTD and PS optimization methodology to mitigate the beam squint effect. The practical TTD constraint is forced where the time delay values are restricted in a given interval. The major contributions of this paper are summarized as follows.

- We approach the general TTD-based hybrid precoder design in the wideband THz massive MIMO OFDM system from an achievable-rate-maximization point of view to present the impact of compensating the beam squint on the achievable rate performance. Treating the product of TTD and PS precoders as a composite analog precoder, we first derive a lower bound of the achievable rate and show that in the large system limit the ideal analog precoder that completely compensates for the beam squint is equivalent to the one that maximizes the achievable rate lower bound.
- Provided the ideal analog precoder identified, we formulate the joint TTD and PS precoder optimization problem based on minimizing the distance between the ideal analog precoder and the product of TTD and PS precoders under the TTD constraints. In contrast with the prior approaches [8]–[10] that only optimize the TTD precoder and assume the unbounded time delay values, we jointly optimize the TTD and PS precoders and the range of time delay values that a TTD can produce in our approach is strictly limited. Although the formulated problem is non-convex and difficult to solve directly, we show that by transforming the problem into the phase domain, the original problem is converted to an equivalent convex problem, which allows us to find a closed form of the global optimal solution. On the basis of the identified optimal solution, our analysis reveals the number of transmit antennas and the amount of time delay required for the best beam squint compensation of our method.
- Leveraging the advantages of the closed-form expressions of our proposed joint optimization approach, a mixed-integer optimization problem is formulated to quantify the minimum number of TTDs required to achieve a predefined array gain performance. Although the formulated mixed-integer problem is intractable, we show that by applying a second-order approximation, the original problem can be relaxed to a tractable form, which enables us to find an optimal solution. Our analysis reveals that when the number of OFDM subcarriers grows, the number of TTDs is linearly increasing with respect to the system bandwidth in order to guarantee a required array gain performance at every OFDM subcarrier.
- We carry out extensive simulations to evaluate the performance of the proposed joint PS and TTD precoding. The simulation results verify that with our optimal design, the beam squint is compensated effectively. Through the simulations, our joint optimization approach outperforms the prior TTD-based precoding approaches that only optimize TTD precoder in terms of array gain and achievable rate. The simulations verify the substantial array gain performance improvement with the minimum number of TTDs informed by our optimization approaches.

Synopsis: The remainder of the paper is organized as follows. Section II presents the channel models of the wideband THz massive MIMO OFDM system and analyzes the array gain loss caused by the beam squint. Section III describes the relationship between beam squint compensation by the ideal analog precoder and the achievable rate lower bound. Then, Section IV derives the closed-form solution of the optimal TTD and PS precoders under the practical TTD constraints and quantifies the minimum number of TTDs that ensures a predefined array gain performance. Section V provides simulation results to corroborate the developed analysis. Finally, the conclusion of this work is drawn in Section VI.

Notation: A bold lower case letter $\mathbf{x}$ is a column vector and a bold upper case letter $\mathbf{X}$ is a matrix. $\mathbf{X}^T$, $\mathbf{X}^H$, $\|\mathbf{X}\|_F$, $\text{tr}(\mathbf{X})$, $\det(\mathbf{X})$, $\mathbf{X}(i, j)$, $\|\mathbf{x}\|_2$, $|x|$, and $\otimes$ are, respectively, the transpose, conjugate transpose, Frobenius norm, trace, determinant, i th row and j th column entry of $\mathbf{X}$, 2-norm of $\mathbf{x}$, modulus of $x \in \mathbb{C}$, and Kronecker product. $\text{blk}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ is an $nN \times N$ block diagonal matrix such that its main-diagonal blocks contain $\mathbf{x}_i \in \mathbb{C}^{n \times 1}$, for $i = 1, 2, \dots, N$ and all off-diagonal blocks are zero. $\mathbf{0}_n$, $\mathbf{1}_n$, and $\mathbf{I}_n$ denote, respectively, the $n \times 1$ all-zero vector, $n \times 1$ all-one vector, and $n \times n$ identity matrix. Given $\mathbf{x} \in \mathbb{R}^{n \times 1}$, $e^{j\mathbf{x}}$ denotes the column vector $[e^{jx_1} \ e^{jx_2} \ \dots \ e^{jx_n}]^T \in \mathbb{C}^{n \times 1}$ obtained by applying $e^{j\cdot}$ element-wise.

II. CHANNEL MODEL AND BEAM SQUINT EFFECT

In this section, we present the channel model of the wideband THz massive MIMO OFDM systems. Then, the array gain loss caused by the beam squint is analyzed.

A. Channel Model

We consider the downlink of a wideband THz hybrid massive MIMO OFDM system where the transmitter is equipped with an N_t -element transmit antenna array with element spacing d . The transmit antenna array is fed by N_{RF} radio frequency (RF) chains to simultaneously transmit N_s data streams to an N_r -antenna receiver. It is assumed that N_t, N_r, N_{RF} , and N_s satisfy $N_s = N_{RF} = N_r \ll N_t$. Herein, we let f_c , B , and K be, respectively, the central (carrier) frequency, bandwidth of the OFDM system, and the number of OFDM subcarriers (an odd number). Then, the k th subcarrier frequency is given by

$$f_k = f_c + \frac{B}{K} \left(k - 1 - \frac{K-1}{2} \right), \text{ for } k = 1, 2, \dots, K. \quad (1)$$

The frequency domain MIMO OFDM channel at the k th subcarrier $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ is

$$\mathbf{H}_k = \sqrt{\frac{N_r N_t}{L}} \sum_{l=1}^L \alpha_l e^{-j2\pi\tau_l f_k} \mathbf{u}_{k,l} \mathbf{v}_{k,l}^H, \quad (2)$$

where L denotes the number of channel (spatial) paths, and $\alpha_l \in \mathbb{C}$ and $\tau_l \in \mathbb{R}$ represent the gain and delay of the l th path of the channel, respectively. For ease of exposition, we assume $L = N_{RF}$, which is equivalent to setting $\alpha_l = 0$, $\forall l > N_{RF}$ in (2). In the remainder of this paper, we use the subscript l for denoting the index of both channel path and RF chain unless specified otherwise. The vectors $\mathbf{v}_{k,l} \in \mathbb{C}^{N_t \times 1}$

and $\mathbf{u}_{k,l} \in \mathbb{C}^{N_r \times 1}$ in (2) are the normalized transmit and receive array response vectors of the k th subcarrier on the l th path, respectively, where the $\mathbf{v}_{k,l}$ is a function of angles of departure (AoDs), i.e., the transmit azimuth $\Psi_l \in [\Psi_{\min}, \Psi_{\max}]$ and elevation $\bar{\Psi}_l \in [\bar{\Psi}_{\min}, \bar{\Psi}_{\max}]$ angles, for $l = 1, 2, \dots, L$. The exactly same definition applies to the normalized receive array response vector $\mathbf{u}_{k,l}$ in (2).

In what follows, we will limit our discussion to the transmit antenna array, keeping in mind that the same applies to the receive array. The antenna geometry of the transmit array is described assuming far-field spatial angles. When the transmit antenna array is uniform rectangular array (URA) located on the yz -plane with N_1 and N_2 elements on the y and z axis, respectively, such that $N_t = N_1 \times N_2$, the transmit array response vector $\mathbf{v}_{k,l}$ in (2) can be modeled by

$$\mathbf{v}_{k,l} = \mathbf{v}_{k,l}^y \otimes \mathbf{v}_{k,l}^z, \quad (3)$$

where the n_1 th entry of $\mathbf{v}_{k,l}^y \in \mathbb{C}^{N_1 \times 1}$ is $\mathbf{v}_{k,l}^y(n_1, 1) = \frac{1}{\sqrt{N_1}} e^{-j\pi \frac{2df_k}{c} (n_1-1) \sin(\Psi_l) \sin(\bar{\Psi}_l)}$, for $n_1 = 1, 2, \dots, N_1$, the n_2 th entry of $\mathbf{v}_{k,l}^z \in \mathbb{C}^{N_2 \times 1}$ is $\mathbf{v}_{k,l}^z(n_2, 1) = \frac{1}{\sqrt{N_2}} e^{-j\pi \frac{2df_k}{c} (n_2-1) \cos(\bar{\Psi}_l)}$, for $n_2 = 1, 2, \dots, N_2$, and $c = 3 \times 10^8$ m/s denotes the speed of light. In particular, $\mathbf{v}_{k,l}^z = \frac{1}{\sqrt{N_2}} [1 \ \dots \ e^{j\pi \frac{2df_k}{c} (N_2-1) \cos(\bar{\Psi}_l)}]^H$ and $\mathbf{v}_{k,l}^y = \frac{1}{\sqrt{N_1}} [1 \ \dots \ e^{j\pi \frac{2df_k}{c} (N_1-1) \sin(\Psi_l) \sin(\bar{\Psi}_l)}]^H$. We define $\psi_{k,l} = \frac{2df_k}{c} \sin(\Psi_l) \sin(\bar{\Psi}_l)$ and $\bar{\psi}_{k,l} = \frac{2df_k}{c} \cos(\bar{\Psi}_l)$ as the spatial directions of the k th subcarrier at the transmitter. Assuming the half wavelength antenna spacing, i.e., $d = \frac{c}{2f_c}$, the spatial directions at the central frequency are simplified to $\psi_l = \sin(\Psi_l) \sin(\bar{\Psi}_l)$ and $\bar{\psi}_l = \cos(\bar{\Psi}_l)$. Hence, setting $\zeta_k = \frac{f_k}{f_c}$ leads to $\psi_{k,l} = \zeta_k \psi_l$, $\bar{\psi}_{k,l} = \zeta_k \bar{\psi}_l$, and

$$\zeta_k = 1 + \frac{B}{f_c} \left(\frac{k-1 - \frac{K-1}{2}}{K} \right), \quad (4)$$

for $k = 1, 2, \dots, K$ and $l = 1, 2, \dots, L$, where (4) follows from the definition of f_k in (1). As a result, the $\mathbf{v}_{k,l}^y$ and $\mathbf{v}_{k,l}^z$ in (3) can be succinctly expressed as $\mathbf{v}_{k,l}^y = \frac{1}{\sqrt{N_1}} [1 \ \dots \ e^{j\pi(N_1-1)\psi_{k,l}}]^H$ and $\mathbf{v}_{k,l}^z = \frac{1}{\sqrt{N_2}} [1 \ \dots \ e^{j\pi(N_2-1)\bar{\psi}_{k,l}}]^H$, respectively.

Note that in the remainder of this paper, we only consider the azimuth angle and set the elevation angle to $\frac{\pi}{2}$ rad, such that we are focusing on receivers that are directly in front of the transmit array. This simplification sets the uniform linear array (ULA) at the transmitter with $N_1 = N_t$, $N_2 = 1$, and $\psi_l = \sin(\Psi_l)$, $\forall l$, which sees the array response vector $\mathbf{v}_{k,l}$ in (3) to be $\mathbf{v}_{k,l} = \frac{1}{\sqrt{N_t}} [1 \ \dots \ e^{j\pi(N_t-1)\psi_{k,l}}]^H \in \mathbb{C}^{N_t \times 1}$. Although this simplification is applied for the ease of exposition, it is straightforward to carry the elevation angle through the following developments presented in the later part of the paper.

B. Beam Squint Effect

As aforementioned in Section I, when the wideband OFDM is employed in a massive MIMO system, a substantial array

gain loss at each subcarrier could occur due to beam squint. To quantify it, we focus on the l th path of the massive MIMO OFDM channel in (2). Denoting the frequency-independent beamforming vector matched to the transmit array response vector with the AoD Ψ_l as $\mathbf{f}^{(l)} = \mathbf{v}_{c,l}$ (i.e., array response vector at the central frequency f_c on the l th path), the array gain at the k th subcarrier of the l th path is given by $g(\mathbf{f}^{(l)}, \psi_{k,l}) = |\mathbf{v}_{k,l}^H \mathbf{f}^{(l)}|$, i.e.,

$$g(\mathbf{f}^{(l)}, \psi_{k,l}) = \frac{1}{N_t} \left| \sum_{n=0}^{N_t-1} e^{jn\pi(\psi_{k,l} - \psi_l)} \right| = \left| \frac{\sin(N_t \Delta_{k,l})}{N_t \sin(\Delta_{k,l})} \right|, \quad (5)$$

where $\Delta_{k,l} = \frac{\pi}{2}(\psi_{k,l} - \psi_l)$. It is not difficult to observe that at the central frequency, the array gain is $g(\mathbf{f}^{(l)}, \psi_{k,l}) = 1$ because $\lim_{x \rightarrow 0} \frac{\sin(N_t x)}{\sin(x)} = N_t$. However, when $f_k \neq f_c$, $\Delta_{k,l}$ deviates from 0; the amount of deviation increases as f_k approaches to f_1 or f_K . As a result, all subcarriers except for the central subcarrier suffer from the array gain loss. The implication in the spatial domain is that the beams at non-central subcarriers may completely split from the one generated at the central subcarrier.

The following proposition quantifies the asymptotic array gain loss as $N_t \rightarrow \infty$.

Proposition 1. Suppose that $\psi_{k,l}$ is the spatial direction at the k th subcarrier ($f_k \neq f_c$) of the l th path. Then, the array gain in (5) converges to 0 as N_t tends to infinity, i.e., $g(\mathbf{f}^{(l)}, \psi_{k,l}) \stackrel{\cdot}{=} 0$, where $\stackrel{\cdot}{=}$ denotes the equality when $N_t \rightarrow \infty$.

Proof. The array gain in (5) can be rewritten as $g(\mathbf{f}^{(l)}, \psi_{k,l}) = \frac{1}{N_t} \left| \frac{\sin(N_t \Delta_{k,l})}{\pi \Delta_{k,l}} \frac{\pi \Delta_{k,l}}{\sin(\Delta_{k,l})} \right|$, where $\Delta_{k,l} \neq 0$ because $f_k \neq f_c$. The proposition follows from the definition of Dirac delta function $\frac{\sin(N_t \Delta_{k,l})}{\pi \Delta_{k,l}} \stackrel{\cdot}{=} \delta(\Delta_{k,l})$ [31], completing the proof. $\square$

Fig. 1a illustrates the convergence trend of Proposition 1. The array gain patterns are calculated for $f_c = 3$ THz, $K = 129$, $\psi_l = 0.8$, and $B = 30$ GHz. As N_t tends to be large, the maximum array gain is only achieved at the central frequency, i.e., 65th subcarrier in Fig. 1a, while other subcarriers suffer from substantial array gain losses. This is quite opposite to the traditional narrowband massive MIMO system in which the array gain grows as $N_t \rightarrow \infty$.

The array gain loss is also numerically understood when the bandwidth B increases. Fig. 1b illustrates those patterns when B grows while using the same parameters as in Fig. 1a except for that $N_t = 256$. As the bandwidth B grows, the maximum array gain is obtained only at the central frequency while other subcarriers experience a large amount of array gain losses. This contrasts with the information-theoretic insight that the system achievable rate grows linearly with the bandwidth while keeping the signal-to-noise-ratio (SNR) fixed.

Wideband THz massive MIMO communication research is in its early stages. In order to truly unleash the potential of THz communications, a hybrid precoding architecture and methods that can effectively compensate for the beam squint effect under practical constraints is of paramount importance.

III. PRELIMINARIES AND ACHIEVABLE RATE LOWER BOUND

We consider a TTD-based hybrid precoding architecture [9], [32], where each RF chain drives M TTDs and each TTD is connected to N PSs as shown in Fig. 1c. The N_t -element ULA is divided into M subarrays with $N = \frac{N_t}{M}$ antennas per subarray. The signal at the k th subcarrier passing through a TTD is delayed by t ($0 \leq t \leq t_{\max}$) in the time domain, which corresponds to the $-2\pi f_k t$ frequency-dependent phase rotation in the frequency domain. The $t_{\max}$ is the maximum time delay value that a TTD device can produce. The k th subcarrier signal at the receiver $\mathbf{y}_k \in \mathbb{C}^{N_r \times 1}$ is then given by [1]

$$\mathbf{y}_k = \sqrt{\rho} \mathbf{H}_k \mathbf{F}_1 \mathbf{F}_{2,k}(\{\mathbf{t}_l\}_{l=1}^{N_{RF}}) \mathbf{W}_k \mathbf{s}_k + \mathbf{n}_k, \quad (6)$$

where $\rho, \mathbf{s}_k \in \mathbb{C}^{N_s \times 1}$, $\mathbf{W}_k \in \mathbb{C}^{N_{RF} \times N_s}$, and $\mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$ are, respectively, the average transmit power, transmit data stream, baseband digital precoder, and normal Gaussian noise vector with each entry being independent and identically distributed (i.i.d.) according to zero mean and variance 1. The $\mathbf{F}_1$ in (6) is the PS precoding matrix, which is the concatenation of N_{RF} PS submatrices $\mathbf{X}_l \in \mathbb{C}^{N_t \times M}$, $l = 1, 2, \dots, N_{RF}$, specifically,

$$\mathbf{F}_1 = \frac{1}{\sqrt{N_t}} [\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_{N_{RF}}] \in \mathbb{C}^{N_t \times M N_{RF}}, \quad (7)$$

where $\mathbf{X}_l = \text{blk}(e^{j\pi \mathbf{x}_1^{(l)}}, e^{j\pi \mathbf{x}_2^{(l)}}, \dots, e^{j\pi \mathbf{x}_M^{(l)}}) \in \mathbb{C}^{N_t \times M}$, $\mathbf{x}_m^{(l)} = [x_{1,m}^{(l)} x_{2,m}^{(l)} \dots x_{N,m}^{(l)}]^T \in \mathbb{R}^{N \times 1}$ is a PS vector, and $x_{n,m}^{(l)}$ is the value of the n th PS that is connected to the m th TTD on the l th RF chain, for $n = 1, 2, \dots, N$, $l = 1, 2, \dots, N_{RF}$, and $m = 1, 2, \dots, M$. The $\mathbf{F}_{2,k}(\{\mathbf{t}_l\}_{l=1}^{N_{RF}}) \in \mathbb{C}^{M N_{RF} \times N_{RF}}$ in (6) is the TTD precoding matrix, which is defined as

$$\mathbf{F}_{2,k}(\{\mathbf{t}_l\}_{l=1}^{N_{RF}}) = \text{blk}(e^{-j2\pi f_k \mathbf{t}_1}, e^{-j2\pi f_k \mathbf{t}_2}, \dots, e^{-j2\pi f_k \mathbf{t}_{N_{RF}}}), \quad (8)$$

where $\mathbf{t}_l = [t_1^{(l)} t_2^{(l)} \dots t_M^{(l)}]^T \in \mathbb{R}^{M \times 1}$ is the l th time delay vector and $t_m^{(l)}$ is the time delay value of the m th TTD on the l th RF chain, for $m = 1, 2, \dots, M$ and $l = 1, 2, \dots, N_{RF}$. For the ease of exposition, in what follows, we omit the time delay vectors $\{\mathbf{t}_l\}_{l=1}^{N_{RF}}$ in the TTD precoder notation $\mathbf{F}_{2,k}(\{\mathbf{t}_l\}_{l=1}^{N_{RF}})$, $\forall k$. We note that the analog precoder corresponds to the product $\mathbf{F}_1 \mathbf{F}_{2,k} \in \mathbb{C}^{N_t \times N_{RF}}$ with the constant modulus constraint

$$|\mathbf{F}_1 \mathbf{F}_{2,k}(i, j)| = \frac{1}{\sqrt{N_t}} \text{ and } \mathbf{F}_1 \mathbf{F}_{2,k} \in \mathcal{F}_{N_t, N_{RF}}, \quad (9)$$

where $\mathcal{F}_{N_t, N_{RF}}$ denotes the set of all matrices $\mathbf{X} \in \mathbb{C}^{N_t \times N_{RF}}$ such that $|\mathbf{X}(i, j)| = \frac{1}{\sqrt{N_t}}$, $\forall i, j$. The data stream $\mathbf{s}_k$ in (6) satisfies $\mathbb{E}[\mathbf{s}_k \mathbf{s}_k^H] = \frac{1}{N_s} \mathbf{I}_{N_s}$. Then, the precoders are normalized such that $\|\mathbf{F}_1 \mathbf{F}_{2,k} \mathbf{W}_k\|_F^2 = N_s$, leading to

$$\mathbb{E}[\|\mathbf{F}_1 \mathbf{F}_{2,k} \mathbf{W}_k \mathbf{s}_k\|_2^2] = 1, \quad k = 1, 2, \dots, K. \quad (10)$$

A. Preliminaries

We first describe the sign invariance property of the array gain in (5) and then identify the ideal analog precoder that completely compensates the beam squint. The sign invariance property and the ideal analog precoder established in this subsection will then be used in Section IV for jointly optimizing TTD and PS precoders.

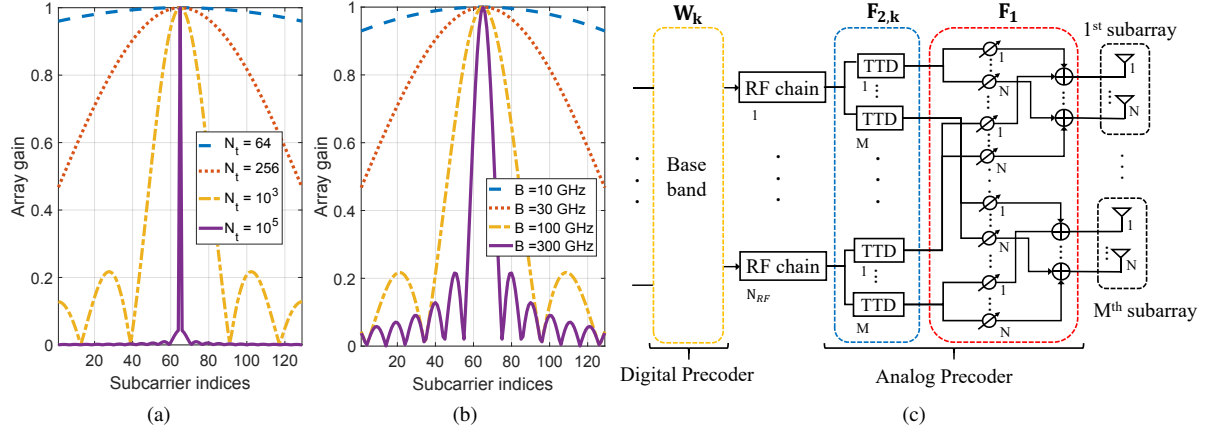


Fig. 1. (a) Array gain vs. subcarrier indices for different numbers of transmit antennas (N_t). (b) Array gain vs. subcarrier indices for different bandwidths (B). (c) TTD-based hybrid precoding architecture

1) *Sign Invariance of Array Gain*: We define the combination of PS precoder $\mathbf{F}_1$ and TTD precoder $\mathbf{F}_{2,k}$ as

$$\mathbf{F}_k = \mathbf{F}_1 \mathbf{F}_{2,k}, \text{ for } k = 1, 2, \dots, K, \quad (11)$$

where the l th column of $\mathbf{F}_k$ is $\mathbf{f}_k^{(l)} = \frac{1}{\sqrt{N_t}} \mathbf{X}_l e^{-j2\pi f_k t_l}$, for $l = 1, 2, \dots, N_{RF}$. The array gain associated with $\mathbf{f}_k^{(l)}$ is then given, based on (5), by

$$g(\mathbf{f}_k^{(l)}, \psi_{k,l}) = \frac{1}{N_t} \left| \sum_{m=1}^M \sum_{n=1}^N e^{j\pi \zeta_k \gamma_{n,m}^{(l)}} e^{j\pi x_{n,m}^{(l)}} e^{-j\pi \zeta_k \vartheta_m^{(l)}} \right|, \quad (12)$$

where $\gamma_{n,m}^{(l)} = ((m-1)N + n - 1)\psi_l$ and $\vartheta_m^{(l)} = 2f_c t_m^{(l)} \in [0, \vartheta_{\max}]$ with $\vartheta_{\max} = 2f_c t_{\max}^{(l)}$. Note that in (12), $e^{j\pi \zeta_k \gamma_{n,m}^{(l)}}$ is frequency-dependent while $e^{j\pi x_{n,m}^{(l)}}$ is frequency-independent, and $e^{-j\pi \zeta_k \vartheta_m^{(l)}}$ depends on both time and frequency. The ability of the beam squint compensation by the m th TTD on the l th RF chain is restricted because the phase rotation $-\pi \zeta_k \vartheta_m^{(l)}$ is only within the interval $[-\pi \zeta_k \vartheta_{\max}, 0]$.

Remark 1. (Sign Invariance Property) The $g(\mathbf{f}_k^{(l)}, \psi_{k,l})$ in (12) is invariant to the multiplication of negative signs to $\gamma_{n,m}^{(l)}$, $x_{n,m}^{(l)}$, and $\vartheta_m^{(l)}$. To be specific, given $\psi_l \geq 0$ (i.e., $\gamma_{n,m}^{(l)} \geq 0$, $\forall m, n$), we denote $\{x_{n,m}^{(l)*}\}$ and $\{\vartheta_m^{(l)*}\}$ as the optimal values of $\{x_{n,m}^{(l)}\}$ and $\{\vartheta_m^{(l)}\}$, respectively, that maximize $g(\mathbf{f}_k^{(l)}, \psi_{k,l})$ in (12). Then, it is not difficult to observe that $\{-x_{n,m}^{(l)*}\}$ and $\{\vartheta_{\max} - \vartheta_m^{(l)*}\}$ also maximize $g(\mathbf{f}_k^{(l)}, -\psi_{k,l})$. Hence, without loss of generality, in what follows, we assume that

$$\psi_l \geq 0, \text{ for } l = 1, 2, \dots, N_{RF}. \quad (13)$$

The sign invariance property will be found to be useful when deriving the solution to our optimization problem for joint time delay and phase shift precoding in Section IV.

2) *Ideal Analog Precoder*: We assume that an ideal analog precoder is the one that can produce arbitrary frequency-dependent phase rotation values to completely mitigate the beam squint effect, which is hereafter denoted as $\tilde{\mathbf{F}}_k^* \in \mathcal{F}_{N_t \times N_{RF}}$, for $k = 1, 2, \dots, K$. The purpose of invoking

the ideal analog precoder is to provide a reference design for the proposed joint TTD and PS precoding method in Section IV. Denoting $\tilde{\mathbf{f}}_k^{(l)}$ as the l th column of $\tilde{\mathbf{F}}_k^*$, the array gain obtained by $\tilde{\mathbf{f}}_k^{(l)}$ is $g(\tilde{\mathbf{f}}_k^{(l)}, \psi_{k,l}) = |\mathbf{v}_{k,l}^H \tilde{\mathbf{f}}_k^{(l)}| = 1$, for $l = 1, 2, \dots, N_{RF}$, resulting in $\tilde{\mathbf{F}}_k^* = [\mathbf{v}_{k,1} \ \mathbf{v}_{k,2} \ \dots \ \mathbf{v}_{k,N_{RF}}]$. Thus, the $((m-1)N+n)$ th row and l th column entry of $\tilde{\mathbf{F}}_k^*$ is

$$\tilde{\mathbf{F}}_k^*((m-1)N+n, l) = \frac{1}{\sqrt{N_t}} e^{-j\pi \zeta_k \gamma_{n,m}^{(l)}}, \forall k, l, m, n. \quad (14)$$

The ideal analog precoder $\tilde{\mathbf{F}}_k^*$ in (14) is achievable with $\mathbf{F}_1 \mathbf{F}_{2,k}$ when each transmit antenna is equipped with a dedicated TTD, i.e., $M = N_t$ and $N = 1$. In this case, the l th column of $\mathbf{F}_1 \mathbf{F}_{2,k}$ is matched exactly to the array response vector of the k th subcarrier on the l th path, $l = 1, 2, \dots, N_{RF}$ and $k = 1, 2, \dots, K$. For instance, the PS and TTD values are designed as $x_{1,m}^{(l)} = 0$ and $t_m^{(l)} = \frac{(m-1)\psi_l}{2f_c}$, respectively, $m = 1, 2, \dots, N_t$ and $l = 1, 2, \dots, N_{RF}$. However, this design is not practically motivated because it requires N_t^2 TTDs, which is very large. Moreover, installing N_t^2 TTDs results in a high hardware complexity and huge power consumption. In Section IV, we address the latter issue by proposing a method to minimize the power consumption of the analog precoder given an array gain performance requirement.

B. Achievable Rate Lower Bound

The impact of beam squint compensation by the ideal analog precoder $\tilde{\mathbf{F}}_k^*$ in (14) on the system achievable rate is of interest. To this end, the performance of hybrid precoding is presented from an achievable-rate-maximization point of view.

Given the analog precoders $\{\mathbf{F}_k\}_{k=1}^K$ in (11), the achievable rate (averaged over the subcarriers) of the channel in (6) is

$$R = \frac{1}{K} \sum_{k=1}^K \log_2 \det \left(\mathbf{I}_{N_s} + \frac{\rho}{N_s} \mathbf{H}_k \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \mathbf{H}_k^H \right). \quad (15)$$

Directly quantifying the impact of the ideal analog precoders $\{\tilde{\mathbf{F}}_k^*\}_{k=1}^K$ in (14) on the achievable rate in (15) is difficult. To be tractable, we derive a lower bound of (15). Defining the singular

value decomposition (SVD) of $\mathbf{H}_k$ as $\mathbf{H}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^H$, where $\mathbf{U}_k \in \mathbb{C}^{N_r \times N_s}$ satisfying $\mathbf{U}_k^H \mathbf{U}_k = \mathbf{I}_{N_s}$, $\mathbf{V}_k \in \mathbb{C}^{N_t \times N_s}$ satisfying $\mathbf{V}_k^H \mathbf{V}_k = \mathbf{I}_{N_s}$, and $\Sigma_k \in \mathbb{R}^{N_s \times N_s}$ is a diagonal matrix of singular values arranged in descending order, and denoting $\Lambda_i(\mathbf{X})$ as the i th eigenvalue of a square matrix $\mathbf{X} \in \mathbb{C}^{n \times n}$, for $i = 1, 2, \dots, n$, the achievable rate R in (15) can be lower bounded by

$$R = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^{N_s} \log_2 \left(1 + \Lambda_i \left(\frac{\rho}{N_s} \mathbf{H}_k \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \mathbf{H}_k^H \right) \right), \quad (16a)$$

$$\geq \frac{1}{K} \sum_{k=1}^K \log_2 \left(1 + \sum_{i=1}^{N_s} \Lambda_i \left(\frac{\rho}{N_s} \mathbf{H}_k \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \mathbf{H}_k^H \right) \right), \quad (16b)$$

$$= \frac{1}{K} \sum_{k=1}^K \log_2 \left(1 + \frac{\rho}{N_s} \sum_{i=1}^{N_s} \Lambda_i \left(\Sigma_k^2 \mathbf{V}_k^H \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \mathbf{V}_k \right) \right), \quad (16c)$$

$$\geq \frac{1}{K} \sum_{k=1}^K \log_2 \left(1 + \rho \left(\det \left(\Sigma_k^2 \mathbf{V}_k^H \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \mathbf{V}_k \right) \right)^{\frac{1}{N_s}} \right), \quad (16d)$$

where (16a) follows from the facts that $\log_2 \det(\mathbf{I}_n + \mathbf{X}) = \sum_{i=1}^n \log_2(\Lambda_i(\mathbf{I}_n + \mathbf{X}))$ and $\Lambda_i(\mathbf{I}_n + \mathbf{X}) = 1 + \Lambda_i(\mathbf{X})$ for $\mathbf{X} \in \mathbb{C}^{n \times n}$, (16b) follows from the fact that $\frac{\rho}{N_s} \mathbf{H}_k \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \mathbf{H}_k^H \succeq 0$ is a positive semi-definite matrix, (16c) is due to the substitution of the SVD for $\mathbf{H}_k$, and (16d) is due to the harmonic-mean-geometric-mean inequality $\sum_{i=1}^n x_i \geq n \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}}$, for non-negative $x_i \in \mathbb{R}$, $\forall i$.

The coupling between $\mathbf{V}_k$ and $\mathbf{F}_k$ in (16d) still restrains any further analysis. It was shown [15], [33] that the ideal analog precoder $\tilde{\mathbf{F}}_k^*$ in (14) converges to an orthogonal basis of $\mathbf{H}_k^H \mathbf{H}_k$ as N_t grows to infinity, i.e., $\tilde{\mathbf{F}}_k^* \tilde{\mathbf{F}}_k^{*H} \doteq \mathbf{I}_{N_{RF}}$, where $\doteq$ denotes the equality when $N_t \rightarrow \infty$. Hence, there exists a right rotation matrix $\mathbf{P}_k$ such that

$$\mathbf{V}_k \doteq \tilde{\mathbf{F}}_k^* \mathbf{P}_k, \quad (17)$$

where $\mathbf{P}_k \in \mathbb{C}^{N_{RF} \times N_{RF}}$ is a unitary matrix, $k = 1, 2, \dots, K$. Substituting (17) into (16d) gives the asymptotic lower bound of R in (15) as

$$R \geq \sum_{k=1}^K \log_2 \left(1 + \rho \left(\det \left(\Sigma_k^2 \mathbf{P}_k^H \tilde{\mathbf{F}}_k^{*H} \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \tilde{\mathbf{F}}_k^* \mathbf{P}_k \right) \right)^{\frac{1}{N_s}} \right), \quad (18)$$

where $\geq$ denotes the greater than or equal inequality when $N_t \rightarrow \infty$.

The following proposition shows the relationship between the ideal analog precoders $\{\tilde{\mathbf{F}}_k^*\}_{k=1}^K$ in (14) and the asymptotic lower bound of R in (18).

Proposition 2. Given Σ_k , $\mathbf{P}_k$, $\tilde{\mathbf{F}}_k^*$, and $\mathbf{W}_k$ in (18), for $k = 1, 2, \dots, K$, the following holds

$$\{\tilde{\mathbf{F}}_k^*\}_{k=1}^K = \underset{\{\mathbf{F}_k\}_{k=1}^K}{\operatorname{argmax}} \lim_{N_t \rightarrow \infty} \sum_{k=1}^K \log_2 \left(1 + \rho \left(\Xi_k \right)^{\frac{1}{N_s}} \right), \quad (19a)$$

$$\text{subject to } \mathbf{F}_k \in \mathcal{F}_{N_t, N_{RF}}, \forall k, \quad (19b)$$

$$\text{where } \Xi_k = \det \left(\Sigma_k^2 \mathbf{P}_k^H \tilde{\mathbf{F}}_k^{*H} \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \tilde{\mathbf{F}}_k^* \mathbf{P}_k \right).$$

Proof. See Appendix A. $\square$

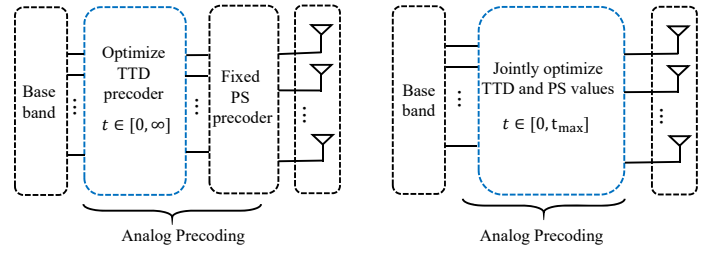


Fig. 2. Comparison between beam squint compensation methods in the prior works [8]–[10] (left) and the proposed method in this work (right).

Proposition 2 reveals that the ideal analog precoder $\tilde{\mathbf{F}}_k^*$ in (14) that completely compensates for the beam squint effect also maximizes the achievable rate lower bound in (18). Therefore, we attempt in Section IV to design TTD precoder $\mathbf{F}_{2,k}$ and PS precoder $\mathbf{F}_1$ in order to best approximate the ideal analog precoder $\tilde{\mathbf{F}}_k^*$, $\forall k$, under the per-TTD time delay constraint.

IV. JOINT DELAY AND PHASE PRECODING UNDER TTD CONSTRAINTS

We propose a novel method to design TTD and PS precoders, which is illustrated in the Fig. 2. In contrary with the prior works [8]–[10] that optimize TTD values while fixing the PS values, we jointly optimize both TTD and PS values. Instead of assuming the ideal TTD that produces any time delay value t (i.e., $0 \leq t \leq \infty$), a finite interval of time delay value (i.e., $0 \leq t \leq t_{\max}$) is taken into account in our approach.

A. Problem Formulation

Ideally speaking, we wish to find $\{t_l\}_{l=1}^{N_{RF}}$ and $\mathbf{F}_1$ satisfying $\mathbf{F}_1 \mathbf{F}_{2,k} = \tilde{\mathbf{F}}_k^*$, $\forall k$. However, given fixed l , m , and n , solving K -coupled matrix equations is an ill-posed problem, because PSs only generate frequency-independent phase values. To overcome this, we approach to formulate a problem that optimizes $\mathbf{F}_1$ and $\{t_l\}_{l=1}^{N_{RF}}$ by minimizing the difference between $\mathbf{F}_1 \mathbf{F}_{2,k}$ and $\tilde{\mathbf{F}}_k^*$, $\forall k$:

$$\min_{\mathbf{F}_1, \{\mathbf{F}_{2,k}\}_{k=1}^K} \frac{1}{K} \sum_{k=1}^K \left\| \tilde{\mathbf{F}}_k^* - \mathbf{F}_1 \mathbf{F}_{2,k} \right\|_F^2, \quad (20a)$$

$$\text{subject to } 0 \leq t_m^{(l)} \leq t_{\max}, \forall l, m, \quad (20b)$$

$$|\mathbf{F}_1(i, j)| \in \left\{ \frac{1}{\sqrt{N_t}}, 0 \right\}, \forall i, j, \quad (20c)$$

$$|\mathbf{F}_{2,k}(p, q)| \in \{0, 1\}, \forall p, q, \quad (20d)$$

$$\mathbf{F}_1 \mathbf{F}_{2,k} \in \mathcal{F}_{N_t, N_{RF}}, \forall k, \quad (20e)$$

where the constraint (20b) indicates the restricted range of the time delay values per-TTD device, the constraint in (20c) is due to the definition of $\mathbf{F}_1$ in (7), the constraint in (20d) is due to the definition of $\mathbf{F}_{2,k}$ in (8), and the constraint in (20e) describes the constant modulus property of the analog precoder in (9). The constraints in (20c)–(20e) in conjunction with the coupling between $\mathbf{F}_1$ and $\mathbf{F}_{2,k}$ in (20a) make the problem difficult to solve. Besides, (20) can be viewed as a matrix factorization problem with non-convex constraints, which has been studied in the context of hybrid analog-digital precoding [15], [17], [34]–[36]. A common approach was applying block coordinate

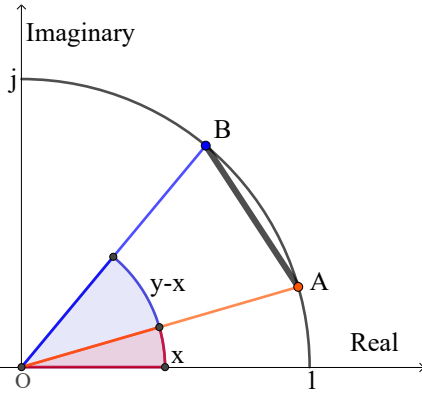


Fig. 3. Conceptually illustration showing the relationship between the distance of two points on the unit modulus circle (the segment AB) and their relative phase difference (the arc corresponding to $y - x$)

descent (BCD) and relaxing the constraints to deal with the non-convexity [15], [17], [34]–[36]. Unlike the prior approaches, we show that the original non-convex problem in (20) can be readily converted into an equivalent convex problem.

Based on the structure of $\mathbf{F}_1$ in (7) and the block diagonal structure of $\{\mathbf{F}_{2,k}\}_{k=1}^K$ in (8), the objective function in (20a) can be rewritten as

$$\frac{1}{K} \frac{1}{N_t} \sum_{k=1}^K \sum_{l=1}^{N_{RF}} \sum_{m=1}^M \sum_{n=1}^N \left| e^{-j\pi\zeta_k\gamma_{n,m}^{(l)}} - e^{j\pi x_{n,m}^{(l)}} e^{-j\pi\zeta_k\vartheta_m^{(l)}} \right|^2. \quad (21)$$

However, it is still difficult to deal with the objective function in (21) due to the unit modulus constraint. The following lemma shows that the optimization on the unit circle is equivalent to the optimization on the corresponding phase domain.

Lemma 1. For $x \in \mathbb{R}$ and $y \in \mathbb{R}$, the following equality holds

$$\underset{y: \text{mod}(y, \pi) \neq x}{\text{argmin}} |e^{jx} - e^{jy}| = \underset{y: \text{mod}(y, \pi) \neq x}{\text{argmin}} |x - y|, \quad (22)$$

where $\text{mod}(y, \pi)$ is y modulo π .

Proof. See Appendix B. $\square$

Fig. 3 graphically visualizes the equivalence in Lemma 1. We let points A and B represent e^{jx} and e^{jy} in (22), respectively. It is not difficult to observe from Fig. 3 that minimizing the length of the segment AB with respect to B is equivalent to minimizing $|y - x|$ with respect to y , in which the latter is a convex optimization problem.

In what follows, Lemma 1 is exploited to convert the non-convex problem in (20) into a convex one. Incorporating Lemma 1 into (21) converts the problem in (20) to the following equivalent problem:

$$\min_{\{x_{n,m}^{(l)}\}, \{\vartheta_m^{(l)}\}} \frac{1}{K} \sum_{k=1}^K \sum_{l=1}^{N_{RF}} \sum_{m=1}^M \sum_{n=1}^N \left| x_{n,m}^{(l)} - \zeta_k \vartheta_m^{(l)} + \zeta_k \gamma_{n,m}^{(l)} \right|^2, \quad (23a)$$

$$\text{subject to } 0 \leq \vartheta_m^{(l)} \leq \vartheta_{\max}, \forall l, m. \quad (23b)$$

Next, we turn (23) to a composite matrix optimization problem. The PS and TTD variables in (23) can be collected into a matrix $\mathbf{A}_l = [\mathbf{a}_1^{(l)} \dots \mathbf{a}_M^{(l)}] \in \mathbb{R}^{(N+1) \times M}$, where $\mathbf{a}_m^{(l)} = [(\mathbf{x}_m^{(l)})^T \ \vartheta_m^{(l)}]^T \in \mathbb{R}^{N+1}$. Containing the analog counterpart in

$\tilde{\mathbf{F}}_k^*$ in a matrix $\mathbf{B}_k^{(l)} \in \mathbb{R}^{N \times M}$, where the n th row and m th column entry of $\mathbf{B}_k^{(l)}$ is $B_k^{(l)}(n, m) = -\zeta_k \gamma_{n,m}^{(l)}$, $\forall k, l, n, m$, the problem (23) becomes

$$\min_{\{\mathbf{A}_l\}_{l=1}^{N_{RF}}} \frac{1}{K} \sum_{k=1}^K \sum_{l=1}^{N_{RF}} \left\| \mathbf{C}_k \mathbf{A}_l - \mathbf{B}_k^{(l)} \right\|_F^2, \quad (24a)$$

$$\text{subject to } \mathbf{0}_M^T \leq \mathbf{e}_{N+1}^T \mathbf{A}_l \leq \vartheta_{\max} \mathbf{1}_M^T, \forall l, \quad (24b)$$

where $\mathbf{C}_k = [\mathbf{I}_N \ -\zeta_k \mathbf{1}_N] \in \mathbb{R}^{N \times (N+1)}$ and $\mathbf{e}_{N+1} \in \mathbb{R}^{N+1}$ is the $(N+1)$ th column of the identity matrix $\mathbf{I}_{N+1}$. The vector inequalities in (24b) is the entry-wise inequalities.

By introducing $\mathbf{C} = \frac{1}{K} \sum_{k=1}^K \mathbf{C}_k^T \mathbf{C}_k \in \mathbb{R}^{(N+1) \times (N+1)}$, $\mathbf{D}_l = \frac{1}{K} \sum_{k=1}^K \mathbf{C}_k^T \mathbf{B}_k^{(l)} \in \mathbb{R}^{(N+1) \times M}$, and $c_B^{(l)} = \frac{1}{K} \sum_{k=1}^K \left\| \mathbf{B}_k^{(l)} \right\|_F^2$, the objective function in (24a) can be rewritten as $\frac{1}{K} \sum_{k=1}^K \sum_{l=1}^{N_{RF}} \left\| \mathbf{C}_k \mathbf{A}_l - \mathbf{B}_k^{(l)} \right\|_F^2 = \sum_{l=1}^{N_{RF}} \sum_{m=1}^M \left((\mathbf{a}_m^{(l)})^T \mathbf{C} \mathbf{a}_m^{(l)} - 2(\mathbf{d}_m^{(l)})^T \mathbf{a}_m^{(l)} \right) + \sum_{l=1}^{N_{RF}} c_B^{(l)}$, where the $\mathbf{d}_m^{(l)}$ is the m th column of $\mathbf{D}_l$. Hence, the problem (24) is equivalently

$$\min_{\mathbf{a}_m^{(l)}} (\mathbf{a}_m^{(l)})^T \mathbf{C} \mathbf{a}_m^{(l)} - 2(\mathbf{d}_m^{(l)})^T \mathbf{a}_m^{(l)}, \quad (25a)$$

$$\text{subject to } 0 \leq \mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} \leq \vartheta_{\max}, \forall l, m. \quad (25b)$$

The problem (25) can be viewed as a decomposition of (24) into MN_{RF} independent problems. This decomposition reveals an alignment with the TTD and PS precoding architecture in Fig. 1c, where each RF feeds M TTDs and each TTD feeds N PSs. Thus, the problem in (25) is equivalent to optimizing the TTD and PS values of each RF chain branch independently.

B. Optimal Closed-form Solution

In this subsection, we find the optimal closed-form solution of the convex problem (25). We start by deriving the expression of $\mathbf{C}$ in (25a). To this end, we first define the constant $\Gamma = N + \eta$, where $\eta = \frac{NB^2(K^2-1)}{f_c^2 12K^2}$. For $\mathbf{C}_k$ in (24), we have $\mathbf{C}_k^T \mathbf{C}_k = \begin{bmatrix} \mathbf{I}_N & -\zeta_k \mathbf{1}_N \\ -\zeta_k \mathbf{1}_N^T & N\zeta_k^2 \end{bmatrix}$, where ζ_k is defined in (4). After some algebraic manipulations, it is readily verified that

$$\sum_{k=1}^K \zeta_k = K, \quad (26a)$$

$$\sum_{k=1}^K \zeta_k^2 = K \left(1 + \frac{\eta}{N} \right). \quad (26b)$$

Then, $\mathbf{C} = \frac{1}{K} \sum_{k=1}^K \mathbf{C}_k^T \mathbf{C}_k$ is simplified, based on (26), to $\mathbf{C} = \begin{bmatrix} \mathbf{I}_N & -\mathbf{1}_N \\ -\mathbf{1}_N^T & \Gamma \end{bmatrix}$. The inverse of $\mathbf{C}$ is given by

$$\mathbf{C}^{-1} = \begin{bmatrix} (\mathbf{I}_N + \frac{1}{\Gamma} \mathbf{1}_N \mathbf{1}_N^T) & \frac{1}{\Gamma} \mathbf{1}_N \\ \frac{1}{\Gamma} \mathbf{1}_N^T & \frac{1}{\Gamma} \end{bmatrix}, \quad (27)$$

which allows us to obtain

$$\mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{e}_{N+1} = \frac{1}{\Gamma}. \quad (28)$$

Defining $\mathbf{b}_{k,m}^{(l)} = [-\zeta_k \gamma_{1,m}^{(l)} \dots -\zeta_k \gamma_{N,m}^{(l)}]^T$ as the m th column of $\mathbf{B}_k^{(l)}$ in (24), it is straightforward that

$$\mathbf{d}_m^{(l)} = \begin{bmatrix} \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{k,m}^{(l)} \\ \frac{1}{K} \sum_{k=1}^K -\zeta_k \mathbf{1}_N^T \mathbf{b}_{k,m}^{(l)} \end{bmatrix} \in \mathbb{R}^{(N+1) \times 1}. \quad (29)$$

From (27) and (29), it is readily verified that

$$\mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{d}_m^{(l)} = \frac{1}{\eta} \frac{1}{K} \sum_{k=1}^K \left((1 - \zeta_k) \sum_{n=1}^N \mathbf{b}_{k,m}^{(l)}(n, 1) \right), \quad (30a)$$

$$= \frac{(2m-1)N-1}{2} \psi_l, \quad (30b)$$

where the $\mathbf{b}_{k,m}^{(l)}(n, 1)$ in (30a) is the n th entry of $\mathbf{b}_{k,m}^{(l)}$, $n = 1, 2, \dots, N$ and (30b) follows from the facts that $\mathbf{b}_{k,m}^{(l)}(n, 1) = -\zeta_k \gamma_{n,m}^{(l)}$ and $\gamma_{n,m}^{(l)} = ((m-1)N + n - 1) \psi_l$.

Based on the results in (28) and (30b), the optimal closed-form solution of the problem in (25) is summarized below.

Theorem 1. The optimal solution $\mathbf{a}_m^{(l)*} = [(\mathbf{x}_m^{(l)*})^T, \vartheta_m^{(l)*}]^T$ to (25) is given by

$$x_{n,m}^{(l)*} = \begin{cases} \frac{N-2n+1}{2} \psi_l, & \text{if } 0 \leq \psi_l \leq \frac{4f_c t_{\max}}{(2m-1)N-1}, \\ \vartheta_{\max} - \gamma_{n,m}^{(l)}, & \text{otherwise,} \end{cases} \quad (31a)$$

$$\vartheta_m^{(l)*} = 2f_c t_m^{(l)*}, \quad (31b)$$

for $l = 1, 2, \dots, N_{RF}$, $n = 1, 2, \dots, N$, $m = 1, 2, \dots, M$, and $\vartheta_m^{(l)*} = 2f_c t_m^{(l)*}$, where the $t_m^{(l)*}$ is

$$t_m^{(l)*} = \begin{cases} \frac{(2m-1)N-1}{4f_c} \psi_l, & \text{if } 0 \leq \psi_l \leq \frac{4f_c t_{\max}}{(2m-1)N-1}, \\ t_{\max}, & \text{otherwise,} \end{cases} \quad (32a)$$

$$(32b)$$

for $l = 1, 2, \dots, N_{RF}$ and $m = 1, 2, \dots, M$.

Proof. See Appendix C. $\square$

Remark 2. The solutions in (32) and (31) differentiate them from prior approaches to solving related problems. For instance, the PS values in [32] were not optimized and given by $x_{n,m}^{(l)} = -(n-1)\psi_l$, $\forall n$. The time delay value of the m th TTD was $t_m^{(l)} = m \frac{N\psi_l}{2f_c}$ in [32]; as m increases, the $t_m^{(l)}$ could be larger than $t_{\max}$, in which case such $t_m^{(l)}$ needs to be floored to the $t_{\max}$, resulting in performance deterioration. As will be discussed in Section V, however, when all TTD values are smaller than $t_{\max}$, the approaches in [9], [32] achieve the same array gain performance as the proposed approach, meaning that the designs in [9], [32] is a special case of Theorem 1.

In the following, the benefits of the proposed joint TTD and PS precoder optimization method are discussed. The optimal condition in (31a) and (32a) can be rewritten as

$$\frac{(2m-1)N-1}{4f_c} \psi_l \leq t_{\max}. \quad (33)$$

When (33) is satisfied, the beam squint is compensated effectively. However, the condition (33) may be violated, for example, when either $t_{\max}$ becomes small or N becomes large (i.e., N_t tends to be large while fixing M , i.e., $N = \frac{N_t}{M}$). Based on the condition in (33), we obtain selection criteria (rule of

thumb) on the required number of transmit antennas (N_t) and the value of maximum time delay ($t_{\max}$) for the best beam squint compensation as follows.

1) *N_t Selection Criterion:* Given the number of TTDs per RF chain M and the $t_{\max}$ values determined by the employed TTD devices, choose N_t such that

$$N_t \leq \frac{M}{2m-1} + \frac{4M}{(2m-1)} \frac{1}{\psi_l} f_c t_{\max}, \forall l, m, \quad (34)$$

where (34) is a result of substituting $N = \frac{N_t}{M}$ into (33). The inequality in (34) can be rewritten as

$$N_t \leq \min_{l,m} \left(\frac{M}{2m-1} + \frac{4M}{(2m-1)} \frac{1}{\psi_l} f_c t_{\max} \right), \quad (35a)$$

$$= \frac{M}{2M-1} + \frac{4M}{(2M-1)} \frac{1}{\max_l \psi_l} f_c t_{\max}. \quad (35b)$$

The $\min_{l,m}$ in (35a) is taken so that (34) holds for all MN_{RF} TTDs. The (35b) follows from substituting $m = M$ and ψ_l by its maximum value $\max_l \psi_l$ into (35a).

2) *$t_{\max}$ Selection Criterion:* Equivalently, given the number of TTDs per RF chain M and the number of transmit antennas N_t , the TTD value should be chosen to satisfy

$$t_{\max} \geq \psi_l \frac{(2m-1)N_t - M}{4M} \frac{1}{f_c}, \forall l, m. \quad (36)$$

The inequality in (36) can be rewritten as

$$t_{\max} \geq \max_{l,m} \left(\psi_l \frac{(2m-1)N_t - M}{4M} \frac{1}{f_c} \right), \quad (37a)$$

$$= (\max_l \psi_l) \frac{(2M-1)N_t - M}{4M} \frac{1}{f_c}. \quad (37a)$$

The system parameter selection criteria in (35b) and (37a) give an idea of the values of N_t and $t_{\max}$ for the best beam squint compensation given a fixed number of TTDs per RF chain M . In practice, it is critical to install an appropriate number of TTDs per RF chain M to reduce the power consumption of the analog precoder because there are total $N_{RF}M$ TTDs and $N_{RF}N_t$ PSs in the analog precoder as shown in Fig. 1c.

C. The Number of TTDs per RF Chain

In this subsection, we further exploit the closed-form solution of the TTD and PS precoders in Theorem 1 to characterize the minimum number of TTDs per RF chain M given a predefined array gain performance while assuming that the constraint in (37a) is always satisfied and the number of transmit antennas N_t is fixed. To this end, we formulate a problem that minimizes M to guarantee a given array gain performance:

$$\min_M M, \quad (38a)$$

$$\text{subject to } N_t = MN, \quad g(\mathbf{f}_k^{(l)}, \psi_{k,l}) \geq g_0, \forall k, l, \quad (38b)$$

where the second constraint in (38b) describes the array gain guarantee with the threshold $0 < g_0 < 1$. We note that the total power consumption of the analog precoder can be modeled as $P_{total} = N_{RF}M P_{TTD} + N_{RF}N_t P_{PS}$, where P_{TTD} and P_{PS} are the power consumption of a TTD and a PS, respectively. Hence, given the fixed values of N_t , P_{TTD} ,

and P_{PS} , the objective of minimizing the power consumption (i.e., $\min_M P_{total}$) is equivalent to the objective in (38).

In solving (38), we first compute the array gain $g(\mathbf{f}_k^{(l)}, \psi_{k,l})$ defined in (12) with the optimal PS values in (31a) and the optimal TTD values in (32a), i.e., $g(\mathbf{f}_k^{(l)}, \psi_{k,l}) = \frac{1}{N_t} \left| \sum_{m=1}^M \sum_{n=1}^N e^{j\pi\zeta_k\gamma_{n,m}^{(l)}} e^{j\pi\frac{N-2n+1}{2}\psi_l} e^{-j\pi\zeta_k\frac{(2m-1)N-1}{2}\psi_l} \right|$, yielding

$$g(\mathbf{f}_k^{(l)}, \psi_{k,l}) = \left| \frac{\sin(N\Delta_{k,l})}{N\sin(\Delta_{k,l})} \right|. \quad (39)$$

Incorporating (39) into the second constraint in (38b) and substituting $\frac{N_t}{M}$ for N , the problem is converted to:

$$\min_M M, \quad (40a)$$

$$\text{subject to} \quad \left| \frac{\sin(\frac{N_t}{M}\Delta_{k,l})}{\frac{N_t}{M}\sin(\Delta_{k,l})} \right| \geq g_0, \forall k, l. \quad (40b)$$

The problem in (40) can be viewed as finding the minimum element in the feasible set $\mathcal{M} = \left(\bigcap_{k=1, l=1}^{K, L} \mathcal{M}_{k,l} \right) \cap \mathcal{N}_t$, where $\mathcal{M}_{k,l} = \left\{ M \in \mathbb{R}: \left| \frac{\sin(\frac{N_t}{M}\Delta_{k,l})}{\frac{N_t}{M}\sin(\Delta_{k,l})} \right| \geq g_0 \right\}$, for $k = 1, 2, \dots, K$ and $l = 1, 2, \dots, L$, and $\mathcal{N}_t = \{m \in \mathbb{N}: m|N_t\}$, where $m|N_t$ means that m is a divisor of N_t . Thus, solving the problem (40) using the greedy search method requires the construction of $KL + 1$ sets $\{\mathcal{M}_{k,l}\}$ and $\mathcal{N}_t$, which is demanding when the number of OFDM subcarriers is large in the wideband THz massive MIMO systems.

To cope with these difficulties, we propose to approximate the constraint in $\mathcal{M}_{k,l}$, i.e., the constraint in (40b) as

$$|q(\Delta_{k,l})| \geq g_0, \forall k, l, \quad (41)$$

where $q(x) = 1 + \frac{1}{6} \left(1 - \frac{N_t^2}{M^2} \right) x^2$ and it is obtained by the Taylor expansion $\frac{\sin(\frac{N_t}{M}x)}{\frac{N_t}{M}\sin(x)} = q(x) + \mathcal{O}(x^4)$ derived at $x = 0$. Incorporating (41) into (40) and noting that $|x| \geq x$, the problem in (40) is relaxed to

$$\min_M M, \quad (42a)$$

$$\text{subject to} \quad M \in \mathcal{N}_t, \quad q(\Delta_{k,l}) \geq g_0, \forall k, l. \quad (42b)$$

About the optimal solution of (42), denoted as M^* , we have the following theorem.

Theorem 2. Given the fixed N_t , B , f_c , and K , the M^* of (42) is

$$M^* = \left\lceil \sqrt{\frac{N_t^2}{1 + \Omega(g_0, B)}} \right\rceil_{N_t}, \quad (43)$$

where $\Omega(g_0, B) = \frac{6(1-g_0)}{\left(\frac{\pi}{2} \frac{B}{f_c} \frac{K-1}{2K}\right)^2 \max_l \psi_l^2}$ and $\lceil x \rceil_{N_t}$ denotes the smallest integer greater than or equal to x that is a divisor of N_t .

Proof. See Appendix D. $\square$

Remark 3. To understand the relationship between the M^* in (43) and the system parameters, we first relax its integer constraint, which leads to $M^* = \sqrt{\frac{N_t^2}{1 + \Omega(g_0, B)}}$. In the regime

of a large number of OFDM subcarriers ($K \gg 1$), we have $\frac{K-1}{2K} \approx \frac{1}{2}$, leading to $M^* \approx \sqrt{\frac{N_t^2 \left(\frac{\pi}{4} \frac{B}{f_c}\right)^2 \max_l \psi_l^2}{6(1-g_0)}} = \left(\frac{\pi N_t}{4f_c} \sqrt{\frac{\max_l \psi_l^2}{6(1-g_0)}}\right) B$. Given a fixed N_t , it reveals that the minimum required number of TTDs that ensures a predefined array gain performance grows linearly with the bandwidth B .

V. SIMULATION RESULTS

In this section, we demonstrate the benefits of the proposed joint TTD and PS precoding by comparing it with the (i) state-of-the-art approach proposed in [8] and (ii) ideal analog precoding $\{\tilde{\mathbf{F}}_k^*\}_{k=1}^K$ in (14) in terms of array gain and achievable rate performance. Throughout the simulation, the system parameters are set as follows unless otherwise stated: the central carrier frequency $f_c = 300$ GHz, bandwidth $B = 30$ GHz, the number of OFDM subcarriers $K = 129$, number of transmit antennas $N_t = 256$, number of receive antennas $N_r = 4$, number of RF chain $N_{RF} = 4$, number of data stream $N_s = 4$, and SNR $\rho = 3$ dB. The number of TTDs per RF chain $M = 16$, the number of PSs per TTD $N = 16$, and the maximum time delay $t_{\max} = 340$ ps. The AoDs and AoAs are uniformly drawn in $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

A. Beam Squint Compensation

To demonstrate the array gain performance, we measure the empirical cumulative distribution function (CDF) of the array gain $G_\psi(x)$ at the central spatial direction ψ , where $G_\psi(x) = \frac{1}{K} \sum_{k=1}^K \mathbb{1}_g(k, x)$, x represents the array gain $0 \leq x \leq 1$ in (5), and the indicator function $\mathbb{1}_g(k, x)$ takes values

$$\mathbb{1}_g(k, x) = \begin{cases} 1, & \text{if } g(\mathbf{f}_k^{(l)}, \psi_k) \leq x, \\ 0, & \text{otherwise,} \end{cases}$$

where $g(\mathbf{f}_k^{(l)}, \psi_k)$ is computed as in (5) with the spatial direction $\psi = 0.8$.

Fig. 4 displays the empirical CDF curves of the array gain when the number of transmit antennas N_t takes the values from $\{128, 256, 512\}$. When $N_t = 128$ the proposed approach and benchmark [8] have the same array gain performance because the designed TTD values in both approaches are smaller than the preset $t_{\max} = 340$ ps. When $N_t = 256$ (Fig. 4b) and $N_t = 512$ (Fig. 4c), the proposed approach and benchmark [8] suffer from the array gain loss because the degree of beam squint increases as N_t grows (i.e., Proposition 1). Nevertheless, the proposed approach provides an enhanced beam squint compensation capability compared to the benchmark. This is due to the fact that as N_t grows some of the time delay values of the benchmark [8] become larger than $t_{\max} = 340$ ps in which case the time delay value needs to be floored to the $t_{\max}$. It can be observed in Fig. 4b that around 80% of OFDM subcarriers achieve the array gain ≥ 0.9 for the proposed approach. For the benchmark [8], on the other hand, there is no OFDM subcarriers that reach 0.9 array gain performance.

Fig. 5 shows the empirical CDF curves of the array gain when the maximum time delay value $t_{\max}$ takes values in $\{320, 340, 400\}$ ps while fixing $N_t = 256$. Seen from Fig. 5, the proposed approach shows the consistent array gain in all

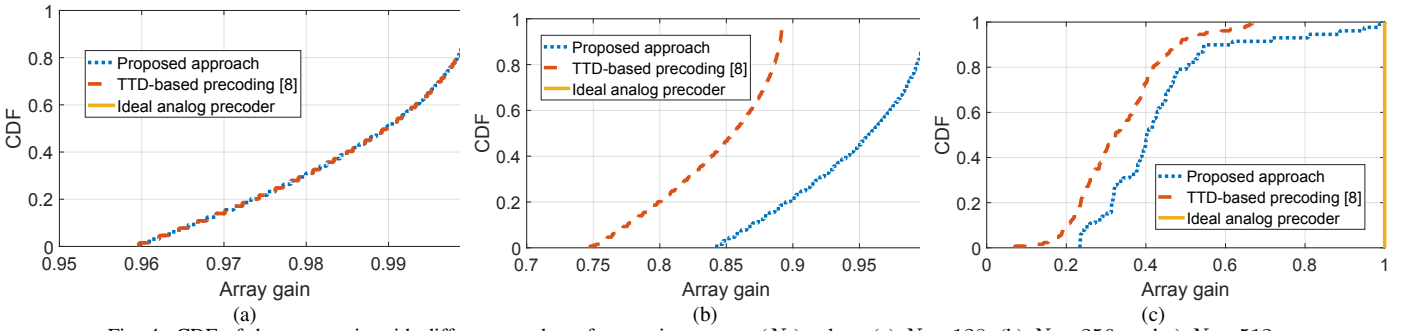


Fig. 4. CDF of the array gain with different number of transmit antennas (N_t) values: (a) $N_t = 128$, (b) $N_t = 256$, and (c) $N_t = 512$.

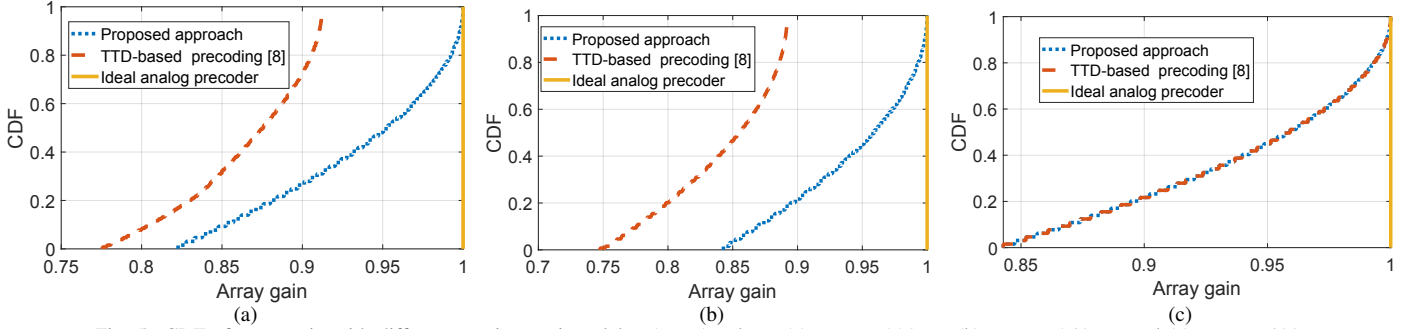


Fig. 5. CDF of array gain with different maximum time delay ($t_{\max}$) values: (a) $t_{\max} = 320$ ps, (b) $t_{\max} = 340$ ps, and (c) $t_{\max} = 400$ ps.

cases while the curve of the benchmark [8] converges to the proposed approach when $t_{\max} = 400$ ps as shown in Fig. 5c. When $t_{\max} = 320$ ps (Fig. 5a) and $t_{\max} = 340$ ps (Fig. 5b), the proposed approach outperforms the benchmark. It is noteworthy to point out that there is no OFDM subcarriers that reach array gain ≥ 0.93 of the benchmark when $t_{\max} \leq 340$ ps.

Overall, the trends in Figs. 4-5 reaffirm Remark 2 indicating that the benchmark is a special cases of the proposed joint optimization. Figs. 4-5 also verify the efficacy of the proposed approach when the N_t and $t_{\max}$ are chosen to meet the criterion in (35b) and (37a) such that $N_t \leq 256$ and $t_{\max} \geq 320$ ps.

B. Achievable Rate Performance

In this subsection, we demonstrate the empirical CDF of the achievable rate per subcarrier $\tilde{R}(x)$, which is defined as $\tilde{R}(x) = \frac{1}{K} \sum_{k=1}^K \mathbb{1}_r(k, x)$ for $x \geq 0$, where the indicator function $\mathbb{1}_r(k, x)$ is given by

$$\mathbb{1}_r(k, x) = \begin{cases} 1, & \text{if } R_k \leq x, \\ 0, & \text{otherwise,} \end{cases}$$

where the R_k is defined by $R_k = \log_2 \det(\mathbf{I}_{N_s} + \frac{\rho}{N_s} \mathbf{H}_k \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \mathbf{H}_k^H)$. The digital precoders $\{\mathbf{W}_k\}_{k=1}^K$ are designed by choosing N_s dominant eigenvectors of $\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k$, where $\mathbf{H}_k = \mathbf{H}_k \mathbf{F}_k$, $\forall k$. Herein, we adopt the same channel model as in [8], where the path delay is uniformly distributed from 0 to 20 ms and the channel path gain is i.i.d following the complex Gaussian distribution with zero mean and variance 1.

Fig. 6 presents the empirical CDF curves of the achievable rate when N_t takes values in $\{128, 256, 512\}$. When $N_t = 128$ in Fig. 6a, the proposed method and the benchmark [8] show the same achievable rate performance because, as discussed in Remark 2, when N_t is relatively small all TTD values are smaller than $t_{\max}$ and the proposed method and the benchmark

become equivalent. As seen in Fig. 6b ($N_t = 256$) and Fig. 6c ($N_t = 512$), on the other hand, the proposed approach shows an improved achievable rate performance compared to the benchmark, which is aligned with the array gain performance trends in Figs. 4b-4c, respectively. The curve of the proposed approach is closer to the ideal achievable rate curve. For instance in Fig. 6b, around 40% of OFDM subcarrier of the proposed approach has the achievable rate ≥ 17 (bits/s/Hz) while the majority of the OFDM subcarriers of the benchmark have the achievable rate < 17 (bits/s/Hz).

Fig. 7 demonstrates the empirical CDF curves of the achievable rate when $t_{\max}$ increases from $t_{\max} = 320$ ps to $t_{\max} = 400$ ps. It can be observed from Fig. 7 that the proposed approach outperforms the benchmark as $t_{\max}$ increases. In Figs. 7a-7b, the proposed method always presents a better achievable rate performance than the benchmark. These align with array gain performance curves in Figs. 5a-5b, respectively.

Figs. 6-7 verify the benefits of the joint TTD and PS optimization of the proposed approach on the achievable rate performance. Similar to Figs. 4-5, Figs. 6-7 demonstrate the efficacy of the proposed joint optimization when N_t and $t_{\max}$ are chosen to meet the criterion in (35b) and (37a) such that $N_t \leq 256$ and $t_{\max} \geq 320$ ps.

C. Array Gain Performance Guarantee of $\tilde{M}^*$ in (43)

In this subsection, we demonstrate the array gain performance guarantee provided by the minimum required number of TTDs per RF chain $\tilde{M}^*$ in (43) of Theorem 2. Similar to Section V-A, we evaluate the empirical CDF of the array gain $G_\psi(x)$ at the spatial direction $\psi = 0.8$ except for that we assume $N_t = 720$ and $t_{\max} = 1000$ ps in this simulation. The array gain threshold is set to $g_0 = 0.9$ in (43).

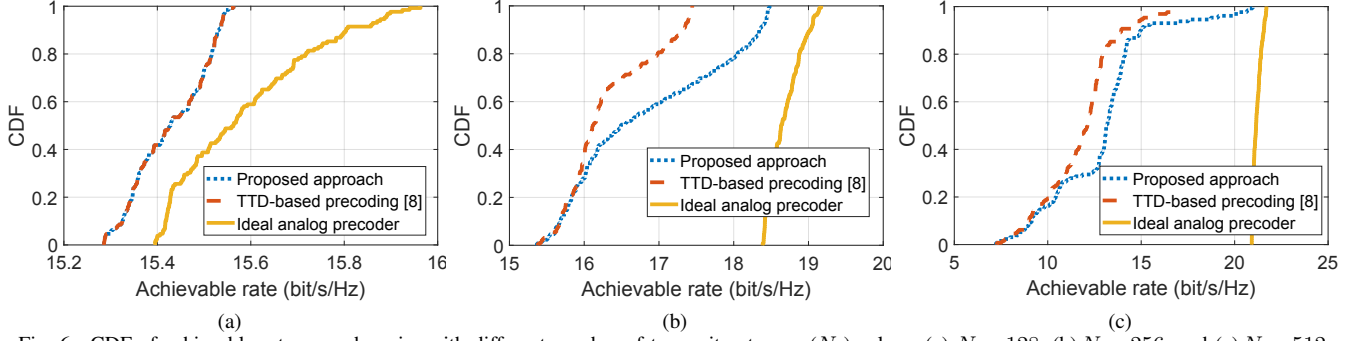


Fig. 6. CDF of achievable rate per subcarrier with different number of transmit antennas (N_t) values: (a) $N_t = 128$, (b) $N_t = 256$, and (c) $N_t = 512$.

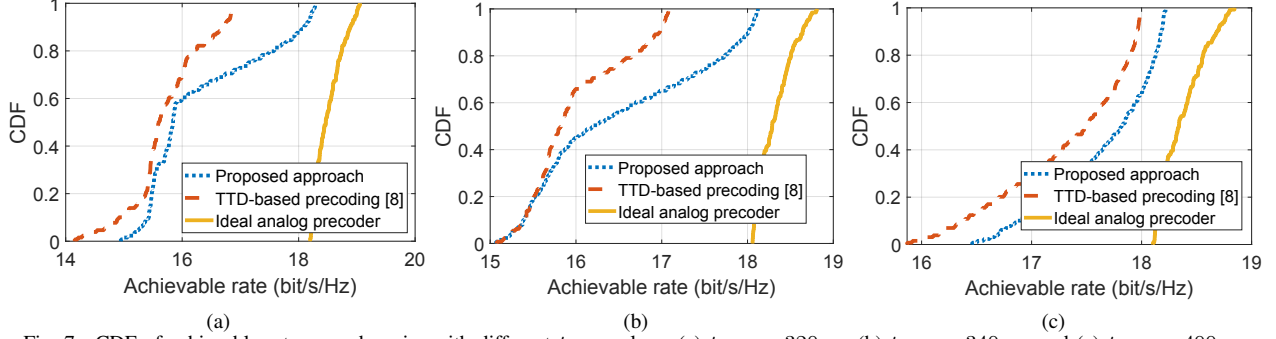


Fig. 7. CDF of achievable rate per subcarrier with different $t_{\max}$ values: (a) $t_{\max} = 320$ ps, (b) $t_{\max} = 340$ ps, and (c) $t_{\max} = 400$ ps.

Incorporating g_0 and the system parameters into (43) yields $\widehat{M}^* = \left\lceil \sqrt{\frac{720^2}{0.9 \cdot 30}} \right\rceil = 60$.

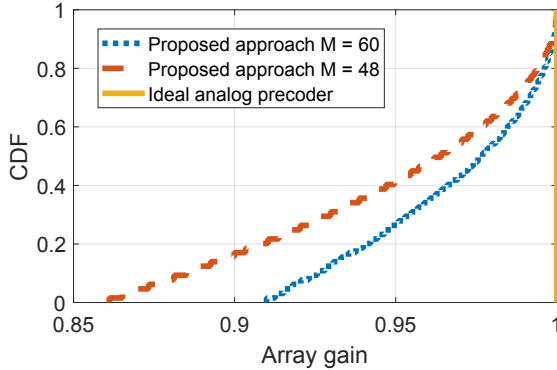


Fig. 8. CDF of array gain of the proposed approach with different number of TTDs per RF chain (M) values.

Fig. 8 illustrates the CDF of the array gain when $M = \widehat{M}^* = 60$ and $M = 48$, where $M = 48$ is chosen to be the largest divisor of N_t that is smaller than $\widehat{M}^* = 60$. We note that the condition in (37a) is satisfied for both values of M when $t_{\max} = 1000$ ps. As shown in Fig. 8, for the optimized $\widehat{M}^* = 60$, every OFDM subcarrier satisfies the array gain ≥ 0.9 . However, when $M = 48$, there are 18% of the OFDM subcarriers that have the array gain < 0.9 . Hence, the curves in Fig. 8 verified that the $\widehat{M}^* = 60$ is the minimum number of TTDs per RF chain to provide the minimum array gain $g_0 = 0.9$.

VI. CONCLUSION

In this paper, we presented a new approach to the problem of compensating the beam squint effect arising in wideband

THz hybrid massive MIMO systems. We showed that the ideal analog precoder that completely compensates for the beam squint is the one that maximizes the achievable rate lower bound. A novel TTD-based hybrid precoding approach was proposed by jointly optimizing the TTD and PS precoders under the per-TTD time delay constraints. The joint optimization problem was formulated in the context of minimizing the distance between the ideal analog precoder and the product of the PS and TTD precoders. By transforming the original problem into the phase domain, the original problem was converted to an equivalent convex problem, which allowed us to find a closed form of the global optimal solution. On the basis of the closed-form expression of our solution, we presented the selection criteria for the required number of transmit antennas and the value of maximum time delay. Taking advantage of the proposed joint TTD and PS precoder optimization approach, we quantified the minimum number of TTDs required to provide an array gain performance guarantee, while minimizing the analog precoder power consumption. Through simulations, we affirmed the superiority of our joint TTD and PS optimization with the array gain performance guarantee.

APPENDIX A PROOF OF PROPOSITION 2

We note that showing Proposition 2 is equivalent to prove

$$\widehat{\mathbf{F}}_k^* = \underset{\mathbf{F}_k}{\operatorname{argmax}} \lim_{N_t \rightarrow \infty} \log_2(\Xi_k), \quad (44a)$$

$$\text{subject to } \mathbf{F}_k \in \mathcal{F}_{N_t, N_{RF}}, \quad (44b)$$

for $k = 1, \dots, K$. The objective in (44a) is decomposed into

$$\log_2 \det \left(\Sigma_k^2 \mathbf{P}_k^H \tilde{\mathbf{F}}_k^H \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \tilde{\mathbf{F}}_k \mathbf{P}_k \right) \\ = \log_2 \det \left(\Sigma_k^2 \mathbf{P}_k^H \mathbf{W}_k \mathbf{W}_k^H \mathbf{P}_k \right) + \log_2 \det \left(\mathbf{F}_k^H \tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k^H \mathbf{F}_k \right). \quad (45)$$

Applying the Hadamard's inequality to $\mathbf{F}_k^H \tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k^H \mathbf{F}_k$ yields

$$\det \left(\mathbf{F}_k^H \tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k^H \mathbf{F}_k \right) \leq \prod_{l=1}^{N_{RF}} \left(\mathbf{f}_k^{(l)} \right)^H \tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k^H \mathbf{f}_k^{(l)} = \prod_{l=1}^{N_{RF}} \left\| \tilde{\mathbf{F}}_k^H \mathbf{f}_k^{(l)} \right\|_2^2. \quad (46)$$

The $\mathbf{f}_k^{(l)}$ in (46) is the l th column of $\mathbf{F}_k$ and $\left\| \mathbf{f}_k^{(l)} \right\|_2^2 = 1$ because of the constant modulus property in (44b). Thus, the following holds

$$\left\| \tilde{\mathbf{F}}_k^H \mathbf{f}_k^{(l)} \right\|_2^2 \leq \Lambda_{\max}(\tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k) \leq 1, \quad (47)$$

where $\Lambda_{\max}(\tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k)$ is the maximum eigenvalue of $\tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k$ and the last equality is due to the fact that $\tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k \preceq \mathbf{I}_{N_{RF}}$. Incorporating (47) into (46) gives $\det \left(\mathbf{F}_k^H \tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k^H \mathbf{F}_k \right) \leq 1$. Therefore, applying (45)

leads to $\log_2 \det \left(\Sigma_k^2 \mathbf{P}_k^H \tilde{\mathbf{F}}_k^H \mathbf{F}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{F}_k^H \tilde{\mathbf{F}}_k \mathbf{P}_k \right) \leq \log_2 \det \left(\Sigma_k^2 \mathbf{P}_k^H \mathbf{W}_k \mathbf{W}_k^H \mathbf{P}_k \right)$, where the equality holds if and only if $\mathbf{F}_k = \tilde{\mathbf{F}}_k^*$. This completes the proof.

APPENDIX B PROOF OF LEMMA 1

Without loss of generality, we assume $0 < |x - y| < \pi$. Then, $\operatorname{argmin}_{0 < |x - y| < \pi} |e^{jx} - e^{jy}| = \operatorname{argmin}_{0 < |x - y| < \pi} \left| \sin \left(\frac{x - y}{2} \right) \right|$
 $= \operatorname{argmin}_{0 < |x - y| < \pi} \sin \left(\left| \frac{x - y}{2} \right| \right) = \operatorname{argmin}_{0 < |x - y| < \pi} |x - y|$,
 where the last step follows from the fact that $\sin(z)$ is an increasing function of $0 < z < \frac{\pi}{2}$. This completes the proof.

APPENDIX C PROOF OF THEOREM 1

To prove Theorem 1, we start by formulating the Lagrangian of (25), which is given by

$$\mathcal{L}(\mathbf{a}_m^{(l)}, \lambda_1, \lambda_2) = \mathbf{a}_m^{(l)T} \mathbf{C} \mathbf{a}_m^{(l)} - 2 \mathbf{d}_m^{(l)T} \mathbf{a}_m^{(l)} + \\ \lambda_1 (\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} - \vartheta_{\max}) + \lambda_2 (-\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)}), \quad (48)$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the Lagrangian multipliers. After incorporating the first order necessary condition for $\mathbf{a}_m^{(l)}$ in (48), the Karush–Kuhn–Tucker (KKT) conditions of (25) are given by:

$$\begin{cases} 2 \mathbf{C} \mathbf{a}_m^{(l)} - 2 \mathbf{d}_m^{(l)} + \lambda_1 \mathbf{e}_{N+1} - \lambda_2 \mathbf{e}_{N+1} & = 0, \\ \lambda_1 (\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} - \vartheta_{\max}) & = 0, \\ \lambda_2 (-\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)}) & = 0, \\ \lambda_1 \geq 0, \lambda_2 \geq 0. \end{cases} \quad (49)$$

It is easy to observe that when $\lambda_1 > 0$ and $\lambda_2 > 0$, there is no solution for (49) because solving (49) leads to $\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} = \vartheta_{\max} > 0$ and $-\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} = 0$, which is a contradiction. When $\lambda_1 = \lambda_2 = 0$, (49) yields $\mathbf{a}_m^{(l)} = \mathbf{C}^{-1} \mathbf{d}_m^{(l)}$ for $0 \leq \mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{d}_m^{(l)} \leq \vartheta_{\max}$ because $0 \leq \mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} \leq \vartheta_{\max}$. When

$\lambda_1 > 0$ and $\lambda_2 = 0$, (49) gives $\mathbf{a}_m^{(l)} = \mathbf{C}^{-1}(\mathbf{d}_m^{(l)} - \frac{1}{2} \lambda_1 \mathbf{e}_{N+1})$ for $\mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{d}_m^{(l)} > \vartheta_{\max}$ because $\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} = \vartheta_{\max}$ and $\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} = \mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{d}_m^{(l)} - \frac{1}{2} \lambda_1 \frac{1}{\eta}$, where the last equality follows from (28). When $\lambda_1 = 0$ and $\lambda_2 > 0$, we obtain $\mathbf{a}_m^{(l)} = \mathbf{C}^{-1}(\mathbf{d}_m^{(l)} + \frac{1}{2} \lambda_2 \mathbf{e}_{N+1})$ for $\mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{d}_m^{(l)} < 0$ because $\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} = 0$ and $\mathbf{e}_{N+1}^T \mathbf{a}_m^{(l)} = \mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{d}_m^{(l)} + \frac{1}{2} \lambda_2 \frac{1}{\eta}$, where the last equality follows from (28). However, by (30b) we have $\mathbf{e}_{N+1}^T \mathbf{C}^{-1} \mathbf{d}_m^{(l)} = \frac{(2m-1)N-1}{2} \psi_l \geq 0$, where the last inequality is due to the sign invariance property in (13). Therefore, the last case leads to a contradiction. In summary, solving (49) gives

$$\mathbf{a}_m^{(l)*} = \begin{cases} \mathbf{C}^{-1} \mathbf{d}_m^{(l)}, & \text{if } 0 \leq \frac{(2m-1)N-1}{2} \psi_l \leq \vartheta_{\max}, \\ \mathbf{C}^{-1} \left(\mathbf{d}_m^{(l)} - \frac{1}{2} \lambda_1 \mathbf{e}_{N+1} \right), & \text{otherwise,} \end{cases} \quad (50a)$$

where $\lambda_1 = 2\eta \left(\frac{(2m-1)N-1}{2} \psi_l - \vartheta_{\max} \right)$.

To further simplify the closed-form solution in (50), we deduce $\mathbf{a}_m^{(l)*}$ in (50a) by simplifying $\mathbf{C}^{-1} \mathbf{d}_m^{(l)}$. Combining (27) and (29) gives

$$\mathbf{C}^{-1} \mathbf{d}_m^{(l)} = \begin{bmatrix} \frac{1}{K} \sum_{k=1}^K \mathbf{b}_{k,m}^{(l)} + \frac{1}{\eta} \frac{1}{K} \sum_{k=1}^K (1 - \zeta_k) \mathbf{1}_N \mathbf{1}_N^T \mathbf{b}_{k,m}^{(l)} \\ \frac{1}{\eta} \frac{1}{K} \sum_{k=1}^K (1 - \zeta_k) \mathbf{1}_N^T \mathbf{b}_{k,m}^{(l)} \end{bmatrix}. \quad (51)$$

Thus, the n th entry of $\tilde{\mathbf{d}}_m^{(l)} \triangleq \mathbf{C}^{-1} \mathbf{d}_m^{(l)}$ in (51), where $1 \leq n \leq N$, is given by

$$\tilde{\mathbf{d}}_m^{(l)}(n, 1) = \frac{1}{K} \gamma_{n,m}^{(l)} \sum_{k=1}^K -\zeta_k + \frac{1}{\eta K} \sum_{n=1}^N \gamma_{n,m}^{(l)} \sum_{k=1}^K (\zeta_k^2 - \zeta_k), \quad (52a)$$

$$= -\gamma_{n,m}^{(l)} + \frac{1}{N} \sum_{n=1}^N \gamma_{n,m}^{(l)}, \quad (52b)$$

$$= \frac{N - 2n + 1}{2} \psi_l, \quad (52c)$$

where (52a) follows from the definition of $\mathbf{b}_{k,m}^{(l)} = [-\zeta_k \gamma_{1,m}^{(l)} \dots - \zeta_k \gamma_{N,m}^{(l)}]^T$, (52b) follows from (26), and (52c) is due to the fact that $\gamma_{n,m}^{(l)} = ((m-1)N + n - 1) \psi_l$. The $(N+1)$ th entry of $\tilde{\mathbf{d}}_m^{(l)}$ in (51) is

$$\tilde{\mathbf{d}}_m^{(l)}(N+1, 1) = \frac{1}{N} \sum_{n=1}^N \gamma_{n,m}^{(l)} = \frac{(2m-1)N-1}{2} \psi_l. \quad (53)$$

Therefore, substituting (52c) and (53) into (50a) leads to (31a) and (32a), respectively. Next, (50b) can be rewritten as

$$\mathbf{C}^{-1} \left(\mathbf{d}_m^{(l)} - \frac{1}{2} \lambda_1 \mathbf{e}_{N+1} \right) = \tilde{\mathbf{d}}_m^{(l)} - \left(\frac{(2m-1)N-1}{2} \psi_l - \vartheta_{\max} \right) \mathbf{1}_{N+1}. \quad (54)$$

Substituting (52c) and (53) into (54) leads to (31b) and (32b), respectively, which completes the proof.

APPENDIX D PROOF OF THEOREM 2

The second constraint in (42b) can be rewritten as

$$g_0 \leq 1 + \frac{1}{6} \left(1 - \frac{N_t^2}{M^2} \right) \max_{k,l} \left(\frac{\pi}{2} \frac{B}{f_c} \left(\frac{k-1 - \frac{K-1}{2}}{K} \right) \psi_l \right)^2, \quad (55a)$$

$$= 1 + \frac{1}{6} \left(1 - \frac{N_t^2}{M^2} \right) \left(\frac{\pi}{2} \frac{B}{f_c} \frac{K-1}{2K} \right)^2 \max_l \psi_l^2, \quad (55b)$$

where (55a) follows from the substitution of $\Delta_{k,l} = \frac{\pi}{2}(\zeta_k - 1)\psi_l$ with ζ_k defined in (4), and (55b) is due to the substitutions $k = K$ and $\psi_l^2 = \max_l \psi_l^2$. Then, the inequality in (55b) can be rewritten as $M \geq \sqrt{\frac{N_t^2}{1+\Omega(g_0, B)}}$, where $\Omega(g_0, B) = \frac{6(1-g_0)}{(\frac{\pi}{2} \frac{B}{f_c} \frac{K-1}{2K})^2 \max_l \psi_l^2}$, leading to $M^* = \left\lceil \sqrt{\frac{N_t^2}{1+\Omega(g_0, B)}} \right\rceil_{N_t}$ because $M \in \mathcal{N}_t$. This completes the proof.

REFERENCES

- [1] D. Q. Nguyen and T. Kim, "Joint delay and phase precoding under true-time delay constraints for THz massive MIMO," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 3496–3501.
- [2] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.
- [3] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Communication*, vol. 12, pp. 16–32, 2014.
- [4] I. F. Akyildiz, C. Han, Z. Hu, S. Nie, and J. M. Jornet, "Terahertz band communication: An old problem revisited and research directions for the next decade," *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 4250–4285, 2022.
- [5] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.
- [6] 3rd Generation Partnership Project (3GPP), *3GPP TR 21.917 - Release 17 Description*, 09 2022, version 1.0.0. [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3937>
- [7] S. Koenig, D. Lopez Diaz, J. Antes, F. Boes, R. Henneberger, A. Leuther, A. Tessmann, R. Schmogrow, D. Hillerkuss, R. Palmer, T. Zwick, C. Koos, W. Freude, O. Ambacher, J. Leuthold, and I. Kallfass, "Wireless sub-THz communication system with high data rate," *Nature Photonics*, vol. 7, pp. 977–981, 10 2013.
- [8] L. Dai, J. Tan, Z. Chen, and H. V. Poor, "Delay-phase precoding for wideband THz massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7271–7286, 2022.
- [9] F. Gao, B. Wang, C. Xing, J. An, and G. Y. Li, "Wideband beamforming for hybrid massive MIMO terahertz communications," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1725–1740, 2021.
- [10] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, "Channel estimation and hybrid combining for wideband terahertz massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1604–1620, 2021.
- [11] L. Yan, C. Han, and J. Yuan, "Energy-efficient dynamic-subarray with fixed true-time-delay design for Terahertz wideband hybrid beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 10, pp. 2840–2854, 2022.
- [12] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [13] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [14] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.
- [15] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.
- [16] 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.
- [17] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, 2016.
- [18] M. Cai, K. Gao, D. Nie, B. Hochwald, J. N. Laneman, H. Huang, and K. Liu, "Effect of wideband beam squint on codebook design in phased-array wireless systems," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.
- [19] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3393–3406, 2018.
- [20] B. Wang, F. Gao, S. Jin, H. Lin, G. Y. Li, S. Sun, and T. S. Rappaport, "Spatial-wideband effect in massive MIMO with application in mmwave systems," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 134–141, 2018.
- [21] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmwave massive MIMO-OFDM systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 5893–5908, 2019.
- [22] X. Liu and D. Qiao, "Space-time block coding-based beamforming for beam squint compensation," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 241–244, 2019.
- [23] R. Mailloux, *Phased Array Antenna Handbook, Third Edition*, 2017.
- [24] M. Longbrake, "True time-delay beamsteering for radar," in *2012 IEEE National Aerospace and Electronics Conference (NAECON)*, 2012, pp. 246–249.
- [25] R. Rotman, M. Tur, and L. Yaron, "True time delay in phased arrays," *Proceedings of the IEEE*, vol. 104, no. 3, pp. 504–518, 2016.
- [26] M.-K. Cho, I. Song, and J. D. Cressler, "A true time delay-based sige bi-directional T/R chipset for large-scale wideband timed array antennas," in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2018, pp. 272–275.
- [27] F. Hu and K. Mouthaan, "A 1–20 ghz 400 ps true-time delay with small delay error in 0.13 μm cmos for broadband phased array antennas," in *2015 IEEE MTT-S International Microwave Symposium*, 2015, pp. 1–3.
- [28] M. Cui, L. Dai, Z. Wang, S. Zhou, and N. Ge, "Near-field rainbow: Wideband beam training for XL-MIMO," 2022. [Online]. Available: <https://arxiv.org/abs/2205.03543>
- [29] J. Tan and L. Dai, "Wideband beam tracking in THz massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1693–1710, 2021.
- [30] H. Luo and F. Gao, "Beam squint assisted user localization in near-field communications systems," 2022. [Online]. Available: <https://arxiv.org/abs/2205.11392>
- [31] R. Bracewell, *The Fourier Transform and its Applications*, 3rd ed. McGraw-Hill, New York, 2000.
- [32] J. Tan and L. Dai, "Delay-phase precoding for THz massive MIMO with beam split," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.
- [33] W. Shen, L. Dai, B. Shim, Z. Wang, and R. W. Heath, "Channel feedback based on AoD-adaptive subspace codebook in fdd massive MIMO systems," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5235–5248, 2018.
- [34] W. Zhang, T. Kim, D. J. Love, and E. Perrins, "Leveraging the restricted isometry property: Improved low-rank subspace decomposition for hybrid millimeter-wave systems," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5814–5827, 2018.
- [35] X. Yu, J. Zhang, and K. B. Letaief, "Alternating minimization for hybrid precoding in multiuser OFDM mmWave systems," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 281–285.
- [36] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1432–1443, 2017.