Explore and Learn: Optimized Two-Stage Search for Millimeter-Wave Beam Alignment

Min Li, Chunshan Liu, Stephen V. Hanly, Iain B. Collings, and Philip Whiting

Abstract

Swift and accurate alignment of transmitter (Tx) and receiver (Rx) beams is one of the fundamental design challenges to enable reliable outdoor millimeter-wave communications. In this paper, we propose a new Optimized Two-Stage Search (OTSS) algorithm for Tx-Rx beam alignment via spatial scanning. In contrast to one-shot exhaustive search, OTSS judiciously divides the training energy budget into two stages. In the first stage, OTSS explores and trains all candidate beam pairs and then discards a set of less favorable pairs learned from the received energy profile. In the second stage, OTSS takes an extra measurement for each of the survived pairs and combines with the previous measurement to determine the best one. For OTSS, we derive an upper bound on its misalignment probability, under a single-path channel model with training codebooks of ideal beam pattern. We also characterize the decay rate function of the upper bound with respect to the training budget and further derive the optimal design parameters of OTSS that maximize the decay rate. OTSS is proved to asymptotically outperform state-of-the-art baselines, and its advantage is also numerically confirmed for limited training budget and with ideal or practically synthesized beams.

Index Terms

Beam alignment, beam training, exhaustive search, hierarchical search, optimized two-stage search, large deviations techniques, millimeter-wave communications.

This work was submitted in part to the 53rd IEEE International Conference on Communications, ICC’19, Shanghai, China, May 20-24, 2019 [1]. The authors are with the School of Engineering, Macquarie University, Sydney, NSW 2109, Australia (e-mail: {min.li, chunshan.liu, stephen.hanly, iain.collings, philip.whiting}@mq.edu.au).

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.
I. INTRODUCTION

Millimeter-wave (mmWave) communications has been recognized as one of the important technologies in the evolving 5G New Radio (NR) [2]–[5]. Owing to the abundant spectrum at mmWave bands (30-300 GHz), mmWave technology has great potential for enabling a variety of data-hungry mobile applications, such as video streaming and vehicle-to-vehicle communications [6]. However, the unfavorable characteristics of mmWave bands, manifesting in severe path loss, sparse scattering and sensitivity to blockage, have posed great challenge in realizing reliable mmWave communications in practice [7].

To combat the significant path loss, directional transmission via beamforming is necessary in particular for outdoor long-range mmWave communications. Establishment of such transmission, however, requires swift and accurate alignment of transmitter (Tx) and receiver (Rx) beams, which is non-trivial to accomplish. In this paper, we focus on this fundamental beam-alignment problem and advance existing studies by developing a new beam-alignment strategy.

Beam training via spatial scanning is a common approach for beam alignment in mmWave communications and it has drawn considerable attention from both academia and industry [7]–[20]. This approach involves a search through pre-defined beam codebooks that cover the scanning space to determine the best beam that aligns with the dominant path for communication. Depending on the application scenarios of mmWave communications, spatial scanning can be performed at one side of the communication link, e.g., at Tx or Rx, to find the best transmit/receive beam for data communication, or be performed at both sides of the communication link to find the best transmit and receive beam pair. In what follows, we discuss spatial scanning schemes by assuming that the search is two-sided. However, the algorithms and analysis will also apply to the one-sided scenario as a special case. Exhaustive search and hierarchical search are seen as two classic strategies of spatial scanning and they distinguish in both codebook construction and search mechanism [7].

In exhaustive search, training codebook is formed by narrow beams with large beamforming gain. Tx and Rx sequentially train each of the beam-pairs in the codebook and find the best one that maximizes a given performance metric (such as combined beamforming gain). On the other hand, in hierarchical search, multi-level codebooks are formed and arranged in a hierarchical manner by using fewer wider beams in the lower level while more narrower beams in the higher
level to cover the same scanning space. In the spirit of bisection search, Tx and Rx first train wide beam-pairs in a lower-level codebook and survive the best one, and then iteratively refine the search using the next-level codebook within the beam subspace associated with the survived beam-pair.

Compared with exhaustive search, hierarchical search reduces the search space and examines fewer beams. However, it may suffer misalignment error propagation, originating from wide beams with small beamforming gain in an early stage. In general, the relative performance between hierarchical search and exhaustive search depends on the pre-beamforming signal-to-noise ratio and the training resource budget (such as training signal power and duration, or effectively training energy). In particular, under the single-path channel model, we have characterized the asymptotic misalignment probability of both hierarchical and exhaustive search and shown that the latter asymptotically outperforms the former, when the training budget goes large.

In addition, the impact of practical beam codebook design on the performance of hierarchical search has also been studied and different beam synthesis techniques have been developed in [11], [13]. Variations of exhaustive search and hierarchical search have also been studied in the presence of favorable beam-pointing side information [17] and in the context of multi-user scenario [21], respectively.

In this paper, we propose a new search algorithm for beam alignment. The proposed algorithm uses the same training codebook as that of exhaustive search. However, in contrast to the classical exhaustive search which is performed in one stage, the proposed algorithm takes place in two stages. In the first stage, the algorithm explores and trains all candidate beam pairs with a fraction of the total training energy budget. Learning from their received energy profile, the algorithm identifies a set of less favorable beam pairs (that are unlikely to align with the dominant path) and eliminates them from further consideration. In the second stage, an extra measurement is taken and coherently combined with the previous measurement for each of the survived beam pairs. Among them, the algorithm recommends the one with the largest combined received energy as the final decision. This algorithm is termed as Optimized Two-Stage Search (OTSS).

Under a line-of-sight single-path channel model and with ideal beam patterns, we derive an upper bound on the misalignment probability of OTSS as a function of key system parameters and verify its tightness numerically. Using large deviations techniques, we also characterize the
decay rate function of the upper bound with respect to the total training energy budget, and further derive the optimal number of less favorable beam pairs eliminated in the first stage ($K^*$) and the optimal fraction of training budget allocated to the first stage ($\alpha^*$) that maximize the decay rate. Both $K^*$ and $\alpha^*$ depend only on the number of candidate beam pairs in the model. This set of analysis not only provides important guidance on the asymptotically optimal choice of the key design parameters of OTSS, but also allows us to conclude that OTSS asymptotically outperforms state-of-the-art baselines (including the classic hierarchical search and exhaustive search), when the training energy budget goes large. The performance advantage of OTSS is also verified numerically when the training budget is finite and when practically synthesized beams are adopted.

**Notation:** Boldface uppercase and lowercase letters denote matrices and vectors, respectively, e.g., $A$ is a matrix and $a$ is a vector. $\|a\|_2$ denotes the $l_2$ norm of vector $a$. Notation $(\cdot)^T$ denotes the matrix transpose, while $(\cdot)^\dagger$ denotes the conjugate transpose. For a pair of integers $(z_1, z_2)$ where $z_1 \leq z_2$, $[z_1:z_2]$ is used to denote the discrete interval $\{z_1, z_1 + 1, \cdots, z_2\}$. Finally, $\mathcal{CN}(0, \sigma^2)$ denotes a complex Gaussian distribution with zero mean and variance $\sigma^2$.

## II. Beam Alignment Problem and Preliminaries

We consider a point-to-point mmWave beam alignment problem, in which a Tx and a Rx wish to align their transmit/reception beams along the dominant path in a mmWave channel. We assume reliable feedback links (via, say, low frequencies) are available for the coordination of the beam search and that the Tx and the Rx are synchronized. We refer the readers to the literature for mmWave synchronization techniques [22], [23].

In particular, beam training via spatial scanning is adopted as in [9], [12], see Fig. 1 for an illustration. Specifically, let $\Psi$ and $\Phi$ be the entire Angle of Departure (AoD) and Angle of Arrival (AoA) scanning interval, respectively. Assume that Tx and Rx is equipped with $N_T$ and $N_R$ antennas, respectively. Let $C_T = \{w_{lT} \in \mathbb{C}^{N_T \times 1}, l_T \in [1 : L_T]\}$ be a set of $L_T$ unit-norm beams at Tx that jointly cover the entire AoD and $C_R = \{f_{lR} \in \mathbb{C}^{N_R \times 1}, l_R \in [1 : L_R]\}$ be a set of $L_R$ unit-norm beams at Rx that jointly cover the entire AoA. The Tx-Rx beam codebook $C$ is therefore given by the cartesian product of $C_T$ and $C_R$ as $C = \{(w, f) : w \in C_T, f \in C_R\}$, with $N \triangleq L_T L_R$ beam pairs in total. For ease of exposition, we simply use $(w_{lT}, f_{lR})$ to denote the $l$th beam pair in $C$, where $l \in [1 : N]$. 
Consider a frequency-flat and block-fading channel model, where the channel remains unchanged during the beam-alignment process. Let $H \in \mathbb{C}^{N_R \times N_T}$ be an arbitrary realization of the mmWave channel between Tx and Rx. The goal of beam alignment is to determine the best beam pair $(w_{l_{\text{opt}}}, f_{l_{\text{opt}}}) \in \mathcal{C}$ that maximizes the effective channel gain after beamforming, i.e.,

$$l_{\text{opt}} = \arg \max_{l \in [1:N]} |f_l^H H w_l|^2.$$  \hspace{1cm} (1)

However, since neither of Tx and Rx has knowledge of $H$, it is necessary to carry out proper beam training and take channel output measurements in order to make such a decision.

Towards this goal, a simple strategy is to train each $(w_l, f_l) \in \mathcal{C}$ and perform exhaustive search. In particular, let $s \in \mathbb{C}$ be the training symbol sent and received on beamforming direction $(w_l, f_l)$, where $|s|^2 = E$ with $E$ being the training energy. The received signal at Rx can be represented as:

$$y_l = f_l^H H w_l s + f_l^H z_l$$
$$= h_l s + z_l, \quad l \in [1:N],$$  \hspace{1cm} (2)

where $h_l \triangleq f_l^H H w_l$ denotes the effective channel after Tx-Rx beamforming, $z_l \in \mathbb{C}^{N_R \times 1}$ is the noise vector (before Rx beamforming) with i.i.d. components $\sim \mathcal{CN}(0, \sigma^2)$ and thus $z_l \sim \mathcal{CN}(0, \sigma^2)$, given that $\|f_l\|_2^2 = 1$.

The received signal is further match-filtered with the training symbol, and the beam pair that leads to the strongest match-filtering output (which essentially captures the received energy at
Rx ([2]) is selected as the best one:

\[
\hat{l}_{ES} = \arg \max_{l \in [1:N]} |y_l s^l|.
\]  

(3)

Consider that the total training energy budget is given as \( E_{tot} \). Since all beam pairs look equally competitive without any prior knowledge, exhaustive search approach would naturally allocate equal training energy to train each of them, i.e., \( E = E_{tot}/N \). If the measurements in (2) were noiseless, a correct decision from (3) would be guaranteed. However, only noisy measurements are obtained in practice, and this renders exhaustive search vulnerable to random noise and results in misalignment events.

Using the same training codebook of exhaustive search, we propose a new beam search algorithm in the next section. We shall further analyze the performance of the proposed algorithm in Section IV and prove that it can outperform exhaustive search under the same training energy budget.

### III. Optimized Two-Stage Search Algorithm

Due to the sparsity of mmWave channels and in particular in the important line-of-sight (LOS) scenario, it is likely that only a small subset of beamforming directions would generate strong received energy at Rx, and the rest mainly capture random noise in the beam training. Therefore, if one could learn about the potentially “good” or “bad” beamforming directions in the process of beam training, the training energy can be spent more effectively by allocating the remaining energy to those directions which are likely the “good” ones in order to better determine the best one.

With this intuition, we now propose a new Optimized Two-Stage Search (OTSS) algorithm that takes place in two stages: in the first stage, a fraction of the total training energy is spent by scanning all candidate beam pairs to learn the potentially “good” ones; at the second stage, the remaining energy is used to find the best one among the “good” ones. Specifically, the total training energy budget is split into two fractions \((\alpha E_{tot}, (1 - \alpha) E_{tot})\), where \( \alpha \in (0, 1] \) is to be optimized later. In the first stage, all candidate beam pairs in \( C \) are explored each with equal training energy

\[
E^{(1)} = \alpha E_{tot}/N,
\]  

(4)
where the superscript “(.)” indexes the stage. Similar to (2), the received signal at Rx can be represented as:

$$y_l^{(1)} = h_l s^{(1)} + z_l^{(1)},$$  \hspace{1cm} (5)$$

and the measured match-filtered output is given by

$$\hat{T}_l^{(1)} = |y_l^{(1)} (s^{(1)})^\dagger|, \ l \in [1 : N],$$  \hspace{1cm} (6)$$

where $s^{(1)} \in \mathbb{C}$ is a training symbol such that $|s^{(1)}|^2 = E^{(1)}$, and the effective noise $z_l^{(1)} \sim \mathcal{CN}(0, \sigma^2)$ as in (2).

Through this exploration, the algorithm then ranks all beam pairs according to their energy statistics $\{\hat{T}_1^{(1)}, \cdots, \hat{T}_N^{(1)}\}$ in an ascending order, identifies the $K$ worst beam pairs that have the smallest energy statistics as the “bad” directions and eliminates them from further consideration. Here, $K \in [1 : N - 1]$ is another key algorithm parameter to be optimized. Without loss of generality, let $\mathcal{B}_G^{(1)}$ be the set of indices of $(N - K)$ survived beam pairs by the end of this stage.

In the second stage, the algorithm evenly splits the remaining training energy $(1 - \alpha)E_{\text{tot}}$ among beam pairs in $\mathcal{B}_G^{(1)}$ as

$$E^{(2)} = (1 - \alpha)E_{\text{tot}}/(N - K),$$  \hspace{1cm} (7)$$

and takes an extra measurement for each pair as

$$y_l^{(2)} = h_l s^{(2)} + z_l^{(2)}, \ l \in \mathcal{B}_G^{(1)},$$  \hspace{1cm} (8)$$

where $s^{(2)} \in \mathbb{C}$ is a training symbol such that $|s^{(2)}|^2 = E^{(2)}$ and noise $z_l^{(2)} \sim \mathcal{CN}(0, \sigma^2)$. By coherently combining the new measurement $y_l^{(2)}$ and its previous measurement $y_l^{(1)}$, the algorithm constructs a set of combined match-filtered outputs:

$$\hat{T}_l^{(2)} = |y_l^{(1)} (s^{(1)})^\dagger + y_l^{(2)} (s^{(2)})^\dagger|, \ l \in \mathcal{B}_G^{(1)}.$$  \hspace{1cm} (9)$$

Finally, the beam pair with the maximum combined output is recommended as the decision, i.e.,

$$\hat{l}_{\text{OTSS}} = \arg \max_{l \in \mathcal{B}_G^{(1)}} \hat{T}_l^{(2)}.$$  \hspace{1cm} (10)$$
TABLE I
OTSS ALGORITHM FOR BEAM ALIGNMENT

| Input: C, beam codebook with N candidate beam pairs; |
| E_{tot}, total energy budget; α, budget fraction for Stage 1; |
| K, the number of beam pairs discarded in Stage 1. |

1) Stage 1:
1.1) For \( l \in [1 : N] \): train \((w_l, f_l)\) to generate \( \hat{y}_l^{(1)} \) sample and compute energy \( \tilde{T}_l^{(1)} \) as in (6).
1.2) Rank all beam pairs based on their energy statistics \( \{\tilde{T}_l^{(1)}\} \), discard the \( K \) worst beam pairs and form survival set \( B_{G}^{(1)} \).

2) Stage 2:
2.1) For \( l \in B_{G}^{(1)} \): retrain \((w_l, f_l)\) to take extra sample \( y_l^{(2)} \) as in (8), and coherently combine \( (y_l^{(2)}, y_l^{(1)}) \) to generate combined statistic \( \tilde{T}_l^{(2)} \) as in (9).

Output: \( l_{OTSS}^* = \arg \max_{l \in B_{G}^{(1)}} \tilde{T}_l^{(2)}, \) as in (10).

The OTSS algorithm described is summarized in Table I. It is clear that given \( N \) candidate beam pairs and a total training energy budget \( E_{tot} \), both parameters \( α \) and \( K \) in OTSS should be properly chosen in order to achieve its best performance. In the next section, we will establish design guideline on this by developing fundamental performance limits of OTSS.

IV. PERFORMANCE ANALYSIS UNDER SINGLE-PATH CHANNEL MODEL AND WITH IDEAL BEAM CODEBOOK

Similar to [9], [12], for tractability, we focus the analysis on a rank-one channel model that captures well the dominant path in a LOS environment. More general models will be numerically investigated in Section V and it will be shown that the insights generated from the analysis here continue to apply therein.

Specifically, assume both Tx and Rx adopt a uniform linear array. The rank-one channel matrix \( H \) is then represented as

\[
H = \gamma u(\phi)v^\dagger(\psi),
\]

(11)

where \(|\gamma|^2\) is the path gain, while \( u(\phi) \in \mathbb{C}^{N_R \times 1} \) and \( v(\psi) \in \mathbb{C}^{N_T \times 1} \) are the steering vectors corresponding to AoA \( \phi \) and AoD \( \psi \) that are defined as

\[
u(\phi) = [1, e^{j2\pi \frac{\phi}{\lambda} \sin(\phi)}, \ldots, e^{j2\pi \frac{\phi}{\lambda} (N_R-1) \sin(\phi)}]^T,
\]

(12)
\[ \mathbf{v}(\psi) = [1, e^{j2\pi \frac{d}{\lambda} \sin(\psi)}, \ldots, e^{j2\pi \frac{d}{\lambda}(N_T-1) \sin(\psi)}]^T, \]  

(13)

respectively, with \( \lambda \) being the wavelength and \( d \) being the antenna spacing. Under this model, effective channel \( h_l \) that accounts for Tx-Rx beams \((w_l, f_l)\) as in (2) is specialized to

\[ h_l = \gamma f_l^\dagger \mathbf{u}(\phi) \mathbf{v}^\dagger(\psi) w_l, \]

(14)

and the corresponding channel gain is thus given by

\[ g_l \triangleq |h_l|^2 = |\gamma f_l^\dagger \mathbf{u}(\phi) \mathbf{v}^\dagger(\psi) w_l|^2 \]

(15)

\[ = |\gamma|^2 F_l(\phi) W_l(\psi), \]

(16)

where we have defined \( W_l(\psi) \triangleq |\mathbf{v}^\dagger(\psi) w_l|^2 \) as the Tx beamforming gain at AoD \( \psi \) and \( F_l(\phi) \triangleq |f_l^\dagger \mathbf{u}(\phi)|^2 \) as the Rx beamforming gain at AoA \( \phi \).

In addition, as established in [12], [13], a desirable beam for beam training purpose should have uniform gain in its intended coverage interval and zero leakage outside the interval. With this ideal beam assumption and supposing all Tx (Rx) beams have equal-size non-overlapping coverage intervals that span the AoD range \( \Psi \) (resp. AoA \( \Phi \)), the Tx (Rx) beamforming gain at \( \psi \in \Psi \) (resp. \( \phi \in \Phi \)) is then quantified by:

\[ W_l(\psi) = \begin{cases} W_T \triangleq \frac{4\pi}{|\Omega_T|/L_T}, & \text{if } \psi \in \Psi_{w_l} \\ 0, & \text{otherwise} \end{cases} \]

(17)

and

\[ F_l(\phi) = \begin{cases} F_R \triangleq \frac{4\pi}{|\Omega_R|/L_R}, & \text{if } \phi \in \Phi_{f_l} \\ 0, & \text{otherwise} \end{cases} \]

(18)

where \( \Psi_{w_l} \) and \( \Phi_{f_l} \) denote the coverage interval of beam \( w_l \) and \( f_l \), while \( \Omega_T \) and \( \Omega_R \) are the solid angles spanned by the entire AoD range \( \Psi \) and AoA range \( \Phi \) [24], respectively. For instance, when \( \Psi = [0, 2\pi] \) and \( L_T = 16 \) beams, we have that \( |\Omega_T| = 4\pi \) and each beam attains constant gain \( W_T = 16 \) within its coverage interval.

With these ideal beam codebooks and under the single-path model with arbitrary AoA \( \phi \) and AoD \( \psi \) given, \( g_l \) of (16) is evaluated to

\[ g_l = \begin{cases} |\gamma|^2 F_R W_T, & \text{if } \psi \in \Psi_{w_l} \text{ and } \phi \in \Phi_{f_l} \\ 0, & \text{otherwise} \end{cases} \]

(19)
Therefore, a perfect alignment through (1) simply chooses the unique beam pair with index \( l_{\text{opt}} \) that leads to non-zero gain.

For the OTSS algorithm proposed, a misalignment event occurs if \( \hat{l}_{\text{OTSS}} \neq l_{\text{opt}} \), and the probability of misalignment is thus defined as \( p_{\text{miss}} = \Pr\{\hat{l}_{\text{OTSS}} \neq l_{\text{opt}}\} \). Without loss of optimality and for notational convenience, \( l_{\text{opt}} = 1 \) is assumed. To further facilitate the analysis, we introduce the following normalized statistics that relate to \( \{\tilde{T}_l^{(1)}\} \) of (6) and \( \{\tilde{T}_l^{(2)}\} \) of (9) as

\[
T_l^{(1)} = \frac{(\tilde{T}_l^{(1)})^2}{\sigma^2 E^{(1)}} , \quad \forall l \in [1 : N],
\]

\[
T_l^{(2)} = \frac{(\tilde{T}_l^{(2)})^2}{\sigma^2 (E^{(1)} + E^{(2)})} , \quad \forall l \in B_G^{(1)},
\]

and define \( T_{(K)}^{(1)} \) as the \( K \)th order statistic (i.e., the \( K \)th smallest value) of \( \{T_2^{(1)}, \cdots, T_N^{(1)}\} \), where recall that \( \sigma^2 \) is the noise variance. By the law of total probability, \( p_{\text{miss}} \) can then be expanded as

\[
p_{\text{miss}} = \Pr\{\hat{l}_{\text{OTSS}} \neq 1\}
= \Pr\{1 \notin B_G^{(1)}\} + \Pr\{\hat{l}_{\text{OTSS}} \neq 1 \text{ and } 1 \in B_G^{(1)}\}
= \Pr\{T_1^{(1)} < T_{(K)}^{(1)}\} + \Pr\{T_1^{(2)} < \max_{l \in B_G^{(1)} \setminus \{1\}} T_l^{(2)}, T_1^{(1)} \geq T_{(K)}^{(1)}\}
= p_{\text{miss}}^{(1)} + p_{\text{miss}}^{(2)},
\]

where \( p_{\text{miss}}^{(1)} \) captures misalignment events that the first beam pair is eliminated in the first stage of OTSS, while \( p_{\text{miss}}^{(2)} \) captures misalignment events that the first beam pair is not chosen at the end of the second stage, though it survives in the first stage.

In what follows, we proceed to study properties of relevant statistics \( \{T_l^{(1)} , l \in [1 : N]\} \) and \( \{T_l^{(2)} , l \in B_G^{(1)}\} \) and develop bounds on \( p_{\text{miss}} \), based on which optimized \( \alpha \) and \( K \) are further derived.

A. Bounds on the Probability of Misalignment

Let \( \chi_k^2(\lambda) \) denote a noncentral chi-squared distribution with degrees of freedom (DoFs) \( k \) and noncentrality parameter \( \lambda \). In the special case with \( \lambda = 0 \), \( \chi_k^2(0) \) becomes a central chi-squared
distribution with DoFs $k$. 

**Lemma 1**: For the OTSS proposed, under the single-path model and the ideal beam codebooks as defined in (17) and (18), we have

$$
\begin{align*}
T_1^{(1)} &\sim \chi^2_2(\lambda^{(1)}_1) \\
T_l^{(1)} &\sim \chi^2_2(0), \quad l \in [2 : N], \\
\end{align*}
$$

(25)

where $\lambda^{(1)}_1 = \frac{2|\gamma|^2 F_R W_T E^{(1)}}{\sigma^2}$, and all $T_l^{(1)}$’s are independent.

**Proof**: Following (19), the received signal $y_l^{(1)}$ as in (5) is specialized to

$$
\begin{align*}
y_l^{(1)} = \begin{cases} 
\gamma \sqrt{F_R W_T} s^{(1)} + z^{(1)}_l, & l = 1 \\
z^{(1)}_l, & l \in [2 : N],
\end{cases}
\end{align*}
$$

(26)

where each $z_l \sim \mathcal{CN}(0, \sigma^2)$. It is immediate to show that each $T_l^{(1)}$ follows the distribution as given by using its definition $T_l^{(1)} = \frac{|y_l^{(1)} s^{(1)}|^2}{2E^{(1)}}$ and $|s^{(1)}|^2 = E^{(1)}$. In addition, $T_l^{(1)}$’s are mutually independent, since each of the received signals is measured at different time under different beam pairs.

Let $T_{(k)}^{(1)}$ be the $k$th order statistic (i.e., the $k$th smallest value) of $\{T_2^{(1)}, \ldots, T_N^{(1)}\}$. Using Lemma 1, it is standard to establish the result as stated in the following corollary [25, p.5].

**Corollary 1**: The probability density function of $T_{(k)}^{(1)}$ is

$$
f_{T_{(k)}^{(1)}}(x) = \frac{(N - 1)!}{2(k - 1)!(N - 1 - k)!} \times \left(1 - \exp(-x/2)\right)^{(k-1)} \exp \left(-\frac{N - k - 1}{2}x\right),
$$

(27)

where $\lambda^{(1)}_1 = \frac{2|\gamma|^2 F_R W_T E^{(1)}}{\sigma^2}$.

**Proof**: See Appendix A for detailed proof.

**Proposition 1**: For the OTSS proposed, misalignment probability $p_{\text{miss}}^{(1)} = \Pr\{T_1^{(1)} < T_{(k)}^{(1)}\}$ at Stage 1 is quantified by

$$
p_{\text{miss}}^{(1)} = \frac{(N - 1)!}{(K - 1)!(N - 1 - K)!} \times \\
\sum_{n=0}^{K-1} \frac{(-1)^n}{(N - K + n)(N - K + n + 1)} \exp \left(-\frac{\lambda^{(1)}_1(N - K + n)}{2(N - K + n + 1)}\right)
$$

(28)

where $\lambda^{(1)}_1 = \frac{2|\gamma|^2 F_R W_T E^{(1)}}{\sigma^2}$.

**Proof**: See Appendix A for detailed proof.
Remark 1: With the expression derived, it is trivial to see that $p_{\text{miss}}^{(1)}$ vanishes, when $\lambda_1^{(1)} \to \infty$ (e.g., $E_{\text{tot}} \to \infty$) on one extreme. On the other extreme with $\lambda_1^{(1)} \to 0$ (e.g., $E_{\text{tot}} \to 0$), we have that $p_{\text{miss}}^{(1)}$ approaches to

$$
\frac{(N-1)!}{(K-1)!(N-K-1)!} \sum_{n=0}^{K-1} (-1)^n \frac{(K-1)}{(N-K+n)(N-K+n+1)} \tag{29}
$$

$$
= \frac{(N-1)!}{(K-1)!(N-K-1)!} \sum_{n=0}^{K-1} (-1)^n \frac{(K-1)}{\Gamma(N-K+2+n)} \tag{30}
$$

$$
= \frac{(N-1)!}{(K-1)!(N-K-1)!} B(K+1, N-K) \tag{31}
$$

$$
= \frac{(N-1)!}{(K-1)!(N-K-1)!} \frac{K!(N-K-1)!}{N!} = \frac{K}{N}, \tag{32}
$$

where (31) follows from [26 Equation 0.160.2] with $B(x, y)$ and $\Gamma(z)$ being Beta and Gamma function, respectively. This result coincides with the general intuition that if no measurements were taken and $K$ out of $N$ beam pairs were randomly eliminated, the probability that the optimal beam pair is eliminated is simply $K/N$. □

As for Stage 2, deriving an exact characterization of $p_{\text{miss}}^{(2)}$ is more challenging, since energy statistic $T_l^{(2)}$ for each of the survived beam pairs is coupled with its previous statistic at Stage 1 through coherent energy combining (9).

In particular, for each survived $l \in B_G^{(1)} \setminus \{1\}$, its $T_l^{(1)}$ must be one of $\{T_{(K+1)}(1), T_{(K+2)}(1), \ldots, T_{(N-1)}(1)\}$, recalling $T_{(k)}$ is the $k$th order statistic of $\{T_2^{(1)}, \ldots, T_{N-1}^{(1)}\}$. Also noting that the extra measurement under each beam-pair $l \in B_G^{(1)} \setminus \{1\}$ at Stage 2 is simply noise, we thus define a set of auxiliary variables as

$$
T_j^{(2)} = \left| \sqrt{T_{(K+1)}^{(1)}} \frac{\sigma^2}{2} E^{(1)} + z_j(s^{(2)}) \right|^2, \tag{33}
$$

where $z_j \sim \mathcal{CN}(0, \sigma^2)$ for $j \in [1 : N - K - 1]$. It is clear that $\{T_j^{(2)}, j \in [1 : N - K - 1]\}$ and $\{T_l^{(2)}, l \in B_G^{(1)} \setminus \{1\}\}$ are statistically equivalent.

We now propose an upper bound on $p_{\text{miss}}^{(2)}$ as

$$
p_{\text{miss}}^{(2)} = \Pr\{T_1^{(2)} < \max_{l \in B_G^{(1)} \setminus \{1\}} T_l^{(2)}, T_1^{(1)} \geq T_{(K)}^{(1)}\}
$$
\[ \Pr\{T_1^{(2)} < \max_{i \in B_G \setminus \{1\}} T_i^{(2)}\} \leq \Pr\{T_1^{(2)} < \max_{j \in [1:N-K-1]} T_j'^{(2)}\} \]

\[ = \Pr\{T_1^{(2)} < \max_{j \in [1:N-K-1]} T_j'^{(2)}\} \leq \sum_{j=1}^{N-K-1} \Pr\{T_1^{(2)} < T_j'^{(2)}\} \]

\[ = \bar{p}^{(2)}_{\text{miss}}, \]  

where (36) follows from the union bound argument.

**Lemma 2:** Variable \( T_1^{(2)} \sim \chi^2_{2} (\lambda_1^{(2)}) \) with noncentrality parameter

\[ \lambda_1^{(2)} = \frac{2|\gamma|^2 F_R W_T (E^{(1)} + E^{(2)})}{\sigma^2}, \]  

while conditioned \( T_{(K+j)}^{(1)} = x \), we have that

\[ \frac{E^{(1)} + E^{(2)}}{E^{(2)}} T_j'^{(2)} \sim \chi^2_{2} \left( \frac{E^{(1)}}{E^{(2)}} x \right), j \in [1 : N - K - 1]. \]  

**Proof:** The proof mainly follows by the construction of each variable and the definition of noncentral chi-squared distribution, see Appendix B.

With this lemma, we are ready to compute \( \bar{p}^{(2)}_{\text{miss}}. \)

**Proposition 2:** For the OTSS proposed, misalignment probability \( p^{(2)}_{\text{miss}} \) at Stage 2 is upper bounded by \( \bar{p}^{(2)}_{\text{miss}} \), which can be computed as

\[ \bar{p}^{(2)}_{\text{miss}} = \sum_{j=1}^{N-K-1} \int_0^\infty F_{(2,2)} \left( \frac{E^{(2)}}{E^{(1)} + E^{(2)}} \right) \left( \lambda_1^{(2)} \left( \frac{E^{(1)}}{E^{(2)}} x \right) \right) f_{T_{(K+j)}^{(1)}} (x) dx, \]  

where \( f_{T_{(K+j)}^{(1)}} (x) \) is given by (27) and \( F_{(n_1, n_2)} (z | \eta_1, \eta_2) \) is the cumulative density function (CDF) of a doubly noncentral F-distribution \( F(n_1, n_2, \eta_1, \eta_2) \) with DoFs \( (n_1, n_2) \) and noncentrality parameters \((\eta_1, \eta_2)\).

**Proof:** It suffices to show that each \( \Pr\{T_1^{(2)} < T_j'^{(2)}\} \) in \( \bar{p}^{(2)}_{\text{miss}} \) can be computed as

\[ \Pr\{T_1^{(2)} < T_j'^{(2)}\} = \mathbb{E}_{T_{(K+j)}^{(1)}} \left[ \Pr\{T_1^{(2)} < T_j'^{(2)} | T_{(K+j)}^{(1)}\} \right] \]

\[ = \mathbb{E}_{T_{(K+j)}^{(1)}} \left[ \Pr \left\{ \frac{T_1^{(2)}}{E^{(1)} + E^{(2)}} < \frac{E^{(2)}}{E^{(1)} + E^{(2)}} \left| T_{(K+j)}^{(1)} \right. \right\} \right] \]
\[ = \mathbb{E}_{F_{(K+j)}} \left[ F_{(2,2)} \left( \frac{E^{(2)}}{E^{(1)} + E^{(2)}} \left| \frac{E^{(1)}}{E^{(2)}} T_{(K+j)} \right|, \lambda^{(2)}_1 \right) \right], \] 

where (41) follows from Lemma 2 and the definition of doubly noncentral \( F \)-distribution.

We therefore establish an upper bound on \( p_{\text{miss}} \) of OTSS (denoted by \( \bar{p}_{\text{miss}} \)) as

\[ \bar{p}_{\text{miss}} = p_{\text{miss}}^{(1)} + p_{\text{miss}}^{(2)}, \]  

where \( p_{\text{miss}}^{(1)} \) is quantified by (28) in Proposition 1 and \( p_{\text{miss}}^{(2)} \) is quantified by (40) in Proposition 2.

Given \( N \) candidate beam pairs and the training energy budget \( E_{\text{tot}} \), we can thus optimize OTSS parameters \( (K, \alpha) \) such that \( \bar{p}_{\text{miss}} \) established is minimized, i.e.,

\[ \min_{K, \alpha} \bar{p}_{\text{miss}} \]  

subject to \( K \in [1 : N - 1] \), \( \alpha \in (0, 1] \).

One naive strategy of finding the optimal solution \( (\bar{K}^*, \bar{\alpha}^*) \) is to carry out a two-dimensional search over the feasible region, e.g., over all possible \( K \)'s and discretized points in \( (0, 1] \).

However, this can be very computationally demanding especially when \( N \) is large. In addition, the optimal solution is generally coupled with \( N, E_{\text{tot}} \) and other system parameters (such as the effective channel gain), which is not a desirable feature from a practical design point of view.

Given these considerations, we study the asymptotic behavior of the upper bound, with the aim to establish a simpler yet useful guideline on the choice of these two parameters of OTSS.

B. Asymptotic Performance Analysis and Further Insights

We focus on understanding how the upper bound decays as the training energy budget goes large.

Proposition 3: Misalignment probability \( p_{\text{miss}}^{(1)} \) satisfies a large deviation principle (LDP) with rate function

\[ - \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log p_{\text{miss}}^{(1)} = \frac{\xi^{(1)}_1}{2 \left( 1 + \frac{1}{N-K} \right)}, \]  

while \( \bar{p}_{\text{miss}}^{(2)} \) satisfies a LDP with rate function

\[ - \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \bar{p}_{\text{miss}}^{(2)} = \frac{\xi^{(2)}_1}{4}, \]
where \( \xi^{(1)}_1 = \frac{2|\gamma|^2 F_RW_T \alpha}{\sigma^2 N} \) and \( \xi^{(2)}_1 = \frac{2|\gamma|^2 F_RW_T}{\sigma^2 \left( \frac{\alpha}{N} + \frac{1-\alpha}{N-K} \right)} \). Therefore, the decay rate of \( \bar{p}_{miss} \) is given by

\[
- \lim_{E_{tot} \to \infty} \frac{1}{E_{tot}} \log \bar{p}_{miss} = \min \left\{ \frac{\xi^{(1)}_1}{2 \left( 1 + \frac{1}{N-K} \right)}, \frac{\xi^{(2)}_1}{4} \right\} \triangleq I_{\bar{p}_{miss}}. \tag{47}
\]

**Proof:** See the proof in Appendix C. \[\square\]

With this result, we further derive the optimal \((K^*, \alpha^*)\) that maximize the decay rate \(I_{\bar{p}_{miss}}\).

**Proposition 4:** For the OTSS proposed, the parameters \((K^*, \alpha^*)\) that maximize rate function \(I_{\bar{p}_{miss}}\) are given by

\[
K^* = N - \text{round}(\sqrt{N}), \tag{48}
\]

\[
\alpha^* = \frac{N(N - K^* + 1)}{K^*(N - K^* + 1) + 2(N - K^*)^2}, \tag{49}
\]

respectively, and the corresponding rate function \(I^*_{\bar{p}_{miss}}\) is

\[
I^*_{\bar{p}_{miss}} = \frac{|\gamma|^2 F_RW_T}{2\sigma^2 \left( N - K^*(N - K^* - 1) \right)}, \tag{50}
\]

**Proof:** See Appendix D. \[\square\]

**Remark 2:** Proposition 4 provides a neat guideline on the choice of OTSS parameters, since \((K^*, \alpha^*)\) established here are only function of \(N\) and do not depend on the training budget and other system parameters. We shall show shortly by numerical results that OTSS with \((K^*, \alpha^*)\) performs close to OTSS with \((\bar{K}^*, \bar{\alpha}^*)\) optimized under each budget \(E_{tot} \) given.

**Remark 3:** Proposition 4 also allows us to conclude that the proposed OTSS outperforms the state-of-the-art beam search baselines under the single-path model considered.

Specifically, in [12], we have proved that exhaustive search asymptotically outperforms hierarchical search\(^1\) in the sense that its misalignment probability has a larger decay rate (with respect to the training energy budget) quantified by

\[
I_{ES} = \frac{1}{N} \frac{2|\gamma|^2 F_RW_T}{4\sigma^2} = \frac{|\gamma|^2 F_RW_T}{2\sigma^2 N}. \tag{51}
\]

\(^1\)Here, hierarchical search includes both the conventional case with equal training energy allocated among beam pairs examined across different stages and the optimized case with training energy allocated in a manner that equalizes the misalignment probability at different stages, see Remark 1 of [12].
Comparing $I_{ES}$ with $I_{\overline{p}_{\text{miss}}}$ of (50), it is immediate to see that $I_{\overline{p}_{\text{miss}}}$ is larger. Therefore, OTSS asymptotically outperforms exhaustive search and thus also hierarchical search.

C. Numerical Evaluation and Comparison

We now provide a numerical example to validate the insights generated from the analysis above.

Specifically, OTSS and exhaustive search both employ the same single-level codebook, with $L_T = 16$ ideal Tx beams covering AoD range $[0, 2\pi]$ and $L_R = 8$ ideal Rx beams covering AoA range $[0, 2\pi]$. Therefore, beamforming gain $W_T = 16$ and $F_R = 8$. The total number of candidate beam pairs is $N = 128$.

Hierarchical search instead employs 4-level Tx-beam codebooks, in which $2^{k_T}$ Tx beams are used at level $k_T \in [1 : 4]$ to cover $[0, 2\pi]$ (see, e.g., [7], [12] for details on hierarchical codebook arrangement) and the last level codebook is the same as that of OTSS for fair comparison. Similarly, 3-level Rx-beam codebooks are used with $2^{k_R}$ Rx beams at level $k_R \in [1 : 3]$ and the last level codebook the same as that of OTSS. Hierarchical search completes in four stages, wherein two Tx-beams and two Rx-beams are scanned at each of the first three stages, while two Tx-beams are scanned at the last stage, with the Rx-beam fixed at the best one determined from the previous stage. Therefore, Tx beamforming gain $W_T^{(k)} = 2^k$ at stage $k \in [1 : 4]$, while Rx beamforming gain $F_R^{(k)} = 2^k$ at stage $k \in [1 : 3]$ and is fixed at 8 for the last stage.

Consider arbitrary fixed single-path channel with path gain $|\gamma|^2$ and vary the total training energy budget $E_{\text{tot}}$ so that $\overline{E} \triangleq \frac{|\gamma|^2 E_{\text{tot}}}{\sigma^2 N} \in [-12 : -7]$ in dB unit. Note that this $\overline{E}$, proportional to $E_{\text{tot}}$, reflects the signal-to-noise ratio at Rx without beamforming.

Fig. 2 compares the performance of different beam alignment algorithms in misalignment probability. In particular, when producing the curve “OTSS: upper bound $p_{\text{miss}}(\bar{K}^*, \bar{\alpha}^*)”$, for each energy budget, we search in the feasible region of $K$ and $\alpha$ and find the pair $(\bar{K}^*, \bar{\alpha}^*)$ that minimizes the upper bound. This associated $(\bar{K}^*, \bar{\alpha}^*)$ are then used to evaluate the misalignment probability of OTSS for the given budget (denoted by $p_{\text{miss}}(\bar{K}^*, \bar{\alpha}^*)$). The tightness of the upper bound is clearly evident in particular when the energy budget is large. In addition, the misalignment probability of OTSS under asymptotically-optimal $(K^*, \alpha^*)$ is also simulated (denoted by $p_{\text{miss}}(K^*, \alpha^*)$). It can be seen that $p_{\text{miss}}(K^*, \alpha^*)$ performs close to $p_{\text{miss}}(\bar{K}^*, \bar{\alpha}^*)$ with only marginal loss. This hence confirms the utility of the asymptotic analysis carried out.
Furthermore, OTSS is shown to significantly outperform all baselines considered when they use the same finite amount of training energy budget. This observation, along with the asymptotic trend we have established, confirm the superiority of OTSS.

Between the baselines, exhaustive search outperforms hierarchical search with equal energy allocation among beam pairs examined and also asymptotically dominates hierarchical search with optimized energy allocation across different stages, as confirmed when $\bar{E}$ increases to $-7$ dB in Fig. 2.

V. FURTHER NUMERICAL STUDIES UNDER PRACTICAL CONSIDERATIONS

In this section, we further evaluate the performance of OTSS under more general channel model and with practically synthesized beams.

Following [12], we evaluate the performance of OTSS using a Rician channel model. The dominant component of the channel is associated with AoD $\psi$ and AoA $\phi$, both uniformly distributed in $[0, 2\pi]$ and the Rician $\mathcal{K}$-factor (i.e., the ratio of the energy of the dominant path to the sum of the energy of the scattering components) is set to 13.2 dB [27]. This corresponds to a line-of-sight scenario in mmWave communications.
As for beam codebooks, consider $L_T = 16$ Tx beams and $L_R = 8$ Rx beams as in the study of Section IV-C. These beams are now practically synthesized beams that might have attenuated gain in its passband and some leakage through its transition band and side lobe. In particular, we use the state-of-the-art flat-beam design technique \cite{28} to synthesize the desired Tx and Rx beams, assuming that Tx and Rx is equipped with $N_T = 64$ and $N_R = 32$ antennas, respectively, and the antenna spacing is half of the wave length. Fig. 3 illustrates the Rx-beam patterns obtained for the simulation.

We now evaluate the performance of OTSS under the channel model and practical beam codebooks described. We also compare its performance with that of exhaustive search. Hierarchical search is not evaluated here, which was previously shown to perform worse than exhaustive search in \cite{12}. Fig. 4 plots the misalignment probability of different schemes. It can be seen that the relative performance trend observed under rank-one channel and ideal beams, as presented in Fig. 2, still remains in this more practical setup: OTSS with $(K^*, \alpha^*)$ outperforms exhaustive search, while OTSS with $(\bar{K}^*, \bar{\alpha}^*)$ leads to additional improvement in the regime of small to moderate training energy budget.

Fig. 5 plots the cumulative density function (CDF) of achievable spectrum efficiency that is calculated according to the post-beamforming effective channel gain (see (16)) using the final
Tx-Rx beam pair determined under different schemes when $\bar{E} = -10$ dB. Consistent with the misalignment probability comparison, OTSS also significantly outperforms exhaustive search in the achievable spectrum efficiency. In particular, OTSS with $(K^*, \alpha^*)$ achieves 10-percentile efficiency of 4.7 bps/Hz, while exhaustive search only achieves 0.06 bps/Hz.

It is finally remarked that all beam-alignment schemes considered here suffer performance loss, as compared to the case with perfect beam pattern in Section IV-C. The main source of the performance degradation is due to the overlapped transition bands of two adjacent beams (see Fig. 3). When either AoA or AoD realization falls into such an overlapped interval, it becomes extremely difficult to make a correct decision based on noisy training measurements of two comparably strong beams. In addition, the side lobes of these imperfect beams introduce additional randomness into beam training measurements and affect the final performance. Given any finite number of Tx or Rx antennas, it is impossible to synthesize perfect beams as desired and hence these negative impacts are inevitable. Some potential remedies might include beam training with dynamic beam codebooks, e.g., intelligently shifting beams or combining adjacent beams as the search progresses, to ensure AoA or AoD realizations are always covered by the main lobe of some beam. However, such designs are beyond the scope of the current work.
Fig. 5. Achievable spectrum efficiency of different schemes under under LOS Rician channel model and with practical imperfect beam codebooks.

VI. CONCLUSIONS

In this paper, we have proposed a new Optimized Two-Stage Search (OTSS) algorithm for mmWave beam alignment. In contrast to exhaustive search where the training energy budget is equally allocated among all candidate beam pairs, the proposed OTSS first explores all beam pairs with a fraction of the budget, identifies a subset as potentially good beam directions and then spends the remaining budget on these directions to determine the best one. Fundamental analysis has been developed to establish the optimal choice for the design parameters of OTSS under a single-path channel model with ideal beam codebooks. It has been further proved that OTSS outperforms state-of-the-art baselines (including hierarchical search and exhaustive search) when the training energy budget goes large. The performance advantage of OTSS has also been confirmed via numerical studies in the regime of limited energy budget and under a more general channel model and practical beam codebooks. As for future work, it is of great interest to establish a new analytical model that captures the essence of imperfect beam patterns and to develop new alignment algorithms with improved robustness to more practical channel models and imperfect beam scenarios.
APPENDIX A

PROOF OF PROPOSITION

Misalignment probability $p_{\text{miss}}^{(1)}$ can be represented as

$$p_{\text{miss}}^{(1)} = \Pr\{T_{1}^{(1)} < T_{(K)}^{(1)}\}$$  \hspace{1cm} (52)

$$= \int_{0}^{\infty} \left[ 1 - Q_{1}\left(\sqrt{\lambda_{1}^{(1)}}, \sqrt{x}\right) \right] f_{T_{(K)}^{(1)}}(x) \, dx$$  \hspace{1cm} (53)

$$= \frac{(N - 1)!}{2(K - 1)!(N - 1 - K)!} \times \int_{0}^{\infty} \left[ 1 - \exp\left(-\frac{x}{2}\right) \right]^{K-1} \exp\left(-\frac{N - K}{2}x\right) \left[ 1 - Q_{1}\left(\sqrt{\lambda_{1}^{(1)}}, \sqrt{x}\right) \right] \, dx$$  \hspace{1cm} (54)

$$= \frac{(N - 1)!}{2(K - 1)!(N - 1 - K)!} \times \left[ \int_{0}^{\infty} \left[ 1 - \exp\left(-\frac{x}{2}\right) \right]^{K-1} \exp\left(-\frac{N - K}{2}x\right) \, dx \right]_{(a)}$$

$$- \int_{0}^{\infty} \left[ 1 - \exp\left(-\frac{x}{2}\right) \right]^{K-1} \exp\left(-\frac{N - K}{2}x\right) Q_{1}\left(\sqrt{\lambda_{1}^{(1)}}, \sqrt{x}\right) \, dx \right]_{(b)}$$  \hspace{1cm} (55)

where (53) follows from Lemma 1 and the cumulative density function of a noncentral chi-squared distribution (as a function of the Marcum Q-function) and (54) follows from Corollary 1.

Now note that

$$(a) = \int_{0}^{\infty} \left[ 1 - \exp\left(-\frac{x}{2}\right) \right]^{K-1} \exp\left(-\frac{N - K}{2}x\right) \, dx$$

$$= \sum_{n=0}^{K-1} (-1)^n \binom{K-1}{n} \int_{0}^{\infty} \exp\left(-\frac{N - K + n}{2}x\right) \, dx$$  \hspace{1cm} (56)

$$= \sum_{n=0}^{K-1} (-1)^n \binom{K-1}{n} \frac{2}{N - K + n}$$  \hspace{1cm} (57)

while

$$(b) = \int_{0}^{\infty} \left[ 1 - \exp\left(-\frac{x}{2}\right) \right]^{K-1} \exp\left(-\frac{N - K}{2}x\right) Q_{1}\left(\sqrt{\lambda_{1}^{(1)}}, \sqrt{x}\right) \, dx$$

$$= \sum_{n=0}^{K-1} (-1)^n \binom{K-1}{n} \int_{0}^{\infty} \exp\left(-\frac{N - K + n}{2}x\right) Q_{1}\left(\sqrt{\lambda_{1}^{(1)}}, \sqrt{x}\right) \, dx$$
\[
= \sum_{n=0}^{K-1} (-1)^n \binom{K-1}{n} \left[ \frac{2}{N-K+n} - \frac{2}{(N-K+n)(N-K+n+1)} \exp \left( -\frac{\lambda_1^{(1)}(N-K+n)}{2(N-K+n+1)} \right) \right],
\]

where (58) uses the fact [29, Equation (16)] that

\[
\int_0^\infty \exp(-px)Q_m(a,b\sqrt{x})dx = \frac{1}{p} - \frac{1}{p} \left( \frac{b^2}{b^2 + 2p} \right)^m \exp \left( -\frac{a^2p}{b^2 + 2p} \right),
\]

where \(Q_m(a,b\sqrt{x})\) is the generalized Marcum Q-function. Plugging (a) and (b) above into (55) thus establishes a characterization of \(p_{\text{miss}}^{(1)}\) as given by (28).

APPENDIX B

PROOF OF LEMMA 2

Under the single-path channel model and perfect beam codebooks considered, \(T_1^{(2)}\) as in (21) is specialized to

\[
T_1^{(2)} = \left| \gamma \sqrt{F_R}W_T \left( E^{(1)} + E^{(2)} \right) + z_1^{(1)} (s^{(1)})^\dagger + z_1^{(2)} (s^{(2)})^\dagger \right|^2
\]

\[
= \left| \gamma \sqrt{2F_R}W_T \left( E^{(1)} + E^{(2)} \right) \sigma^2 + \sqrt{\frac{\sigma^2}{E^{(1)} + E^{(2)}}} \right| \left( z_1^{(1)} (s^{(1)})^\dagger + z_1^{(2)} (s^{(2)})^\dagger \right)^2,
\]

where \(z \sim \mathcal{CN}(0,2)\), since \(z_1^{(1)}\) and \(z_1^{(2)}\) are independent \(\sim \mathcal{CN}(0,\sigma^2)\), \(|s^{(1)}|^2 = E^{(1)}\) and \(|s^{(2)}|^2 = E^{(2)}\). It is immediate to conclude that \(T_1^{(2)}\) follows a noncentral chi-squared distribution with 2 DoFs and noncentrality parameter \(\lambda_1^{(2)}\) as given in (38).

As for \(T_j^{(2)}\) in (33), conditioned on \(T_{(K+j)}^{(1)} = x\),

\[
\frac{E^{(1)} + E^{(1)}}{E^{(2)}} T_j^{(2)} = \left| \sqrt{x \frac{\sigma^2}{E^{(1)} + E^{(2)}}} \left( z_j (s^{(2)})^\dagger \right)^2 \right|
\]

\[
= \left| \sqrt{\frac{E^{(1)}}{E^{(2)}x}} + \sqrt{\frac{\sigma^2}{E^{(2)}}} \right| \left( z_j (s^{(2)})^\dagger \right)^2,
\]

where \(z' \sim \mathcal{CN}(0,2)\), since \(z_j \sim \mathcal{CN}(0,\sigma^2)\) and \(|s^{(2)}|^2 = E^{(2)}\). It is immediate to conclude that \(\frac{E^{(1)} + E^{(1)}}{E^{(2)}} T_j^{(2)}\) follows a noncentral chi-squared distribution with 2 DoFs and noncentrality
parameter $\frac{E^{(1)}}{E^{(2)}} x$, given that $T^{(1)}_{(K+j)} = x$.

**APPENDIX C**

**PROOF OF PROPOSITION 3**

*(Part 1)*: We first consider $p^{(1)}_{\text{miss}} = \Pr\{T^{(1)}_1 < T^{(1)}_{(K)}\}$, where recall that $T^{(1)}_1 \sim \chi^2_2(\lambda^{(1)}_1)$ and $T^{(1)}_{(K)}$ is the $K$th order statistic of $(N - 1)$ i.i.d. variables $\sim \chi^2_2(0)$ as shown by Lemma 1.

Define $\bar{T}^{(1)}_1 \triangleq T^{(1)}_1 E_{\text{tot}}$ and $\bar{T}^{(1)}_{(K)} \triangleq T^{(1)}_{(K)} E_{\text{tot}}$. We thus have $p^{(1)}_{\text{miss}} = \Pr\{T^{(1)}_1 < T^{(1)}_{(K)}\} = \Pr\{\bar{T}^{(1)}_1 < \bar{T}^{(1)}_{(K)}\}$. Since $\bar{T}^{(1)}_1$ and $\bar{T}^{(1)}_{(K)}$ are independent, one can first derive the rate function for each and then combine to characterize the rate function of $p^{(1)}_{\text{miss}}$.

Specifically, following the Gartner-Ellis theorem [30, Chapter 2], we first check that the normalized logarithmic Moment Generating Function (MGF) of $\bar{T}^{(1)}_1$ exists as an extended real number as

$$\Lambda(t_1) = \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \mathbb{E}[e^{E_{\text{tot}} t_1 \bar{T}^{(1)}_1}]$$

(62)

$$= \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \mathbb{E}[e^{T^{(1)}_1 t_1}]$$

(63)

$$= \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \left(\exp\left(\frac{\lambda^{(1)}_1}{1 - 2t_1}\right)\right) = \frac{\xi^{(1)}_1 t_1}{1 - 2t_1}, \quad \text{if } t_1 < \frac{1}{2}$$

(64)

$$+ \infty, \quad \text{otherwise},$$

where (64) follows from $T^{(1)}_1 \sim \chi^2_2(\lambda^{(1)}_1)$, and we have introduced $\xi^{(1)}_1 \triangleq \frac{\lambda^{(1)}_1}{E_{\text{tot}}} = \frac{2|\gamma|^2 F_{\beta W T}}{\sigma^4 N}$.

Further, the origin belongs to the interior of $D_\Lambda = \{t_1 : \Lambda(t_1) < \infty\}$. Therefore, $\bar{T}^{(1)}_1$ satisfies the Gartner-Ellis conditions [30, Assumption 2.3.2]. By the Gartner-Ellis theorem [30, Theorem 2.3.6], the rate function of $T^{(1)}_1$ can thus be calculated as

$$I_{T^{(1)}_1}(u) = \sup_{t_1 \in \mathbb{R}} \left\{u t_1 - \Lambda(t_1)\right\}$$

(65)

$$= \sup_{t_1 < \frac{1}{2}} \left\{u t_1 - \frac{\xi^{(1)}_1 t_1}{1 - 2t_1}\right\}$$

(66)

$$= \frac{(\sqrt{u} - \sqrt{\xi^{(1)}_1})^2}{2}, \quad u \geq 0.$$  

(67)
Similarly, for $\bar{T}^{(1)}_{(K)}$, its normalized logarithmic MGF exists as an extended real number as

$$
\Lambda(t_2) = \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log E[e^{E_{\text{tot}}t_2\bar{T}^{(1)}_{(K)}}] = \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log E[e^{t_2T^{(1)}_{(K)}}] = \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log E[e^{\sum_{j=1}^{K}Z_jt_2}] = \begin{cases} 
lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \left( \prod_{j=1}^{K} \frac{N-j}{\frac{N}{2} - t_2} \right) = 0, & \text{if } t_2 < \frac{N-K}{2} \\
+\infty, & \text{otherwise},
\end{cases}
$$

(68)

(69)

(70)

(71)

where (70) uses the fact that $T^{(1)}_{(K)}$ (the $K$th order statistic of $(N-1)$ i.i.d. variables $\sim \chi^2_2(0)$) is statistically equivalent to $\sum_{j=1}^{K} Z_j$, where $Z_j$ is distributed as an exponential distribution with rate parameter $\frac{N-j}{2}$ (i.e., $Z_j \sim \text{Exp}(\frac{N-j}{2})$) [25, Chapter 1], while (71) follows from the MGF of an exponential distributed variable. Further, the origin belongs to the interior of $D_{\Lambda} = \{ t_2 : \Lambda(t_2) < \infty \}$. Therefore, $\bar{T}^{(1)}_{(K)}$ satisfies the Gartner-Ellis conditions [30, Assumption 2.3.2].

Invoking the Gartner-Ellis theorem, the rate function of $\bar{T}^{(1)}_{(K)}$ can thus be calculated as

$$
I_{\bar{T}^{(1)}_{(K)}}(v) = \sup_{t_2 \in \mathbb{R}} \left\{ vt_2 - \Lambda(t_2) \right\} = \sup_{t_2 < \frac{N-K}{2}} \{ vt_2 \} = \frac{N-K}{2} v, \quad v \geq 0.
$$

(72)

(73)

(74)

With rate functions $I_{\bar{T}^{(1)}_{(K)}}(u)$ and $I_{\bar{T}^{(1)}_{(K)}}(v)$, we thus have that

$$
- \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log p^{(1)}_{\text{miss}} = - \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \Pr \{ \bar{T}^{(1)}_{(K)} < \bar{T}^{(1)}_{(K)} \} = \inf_{0 \leq u \leq v} \left\{ I_{\bar{T}^{(1)}_{(K)}}(u) + I_{\bar{T}^{(1)}_{(K)}}(v) \right\} = \inf_{0 \leq u \leq v} \left\{ \left( \sqrt{u} - \sqrt{\frac{e^{(1)}_{\text{miss}}}{2}} \right)^2 + \frac{N-K}{2} v \right\}.
$$

(75)

(76)

(77)
Using the Karush-Kuhn-Tucker conditions \cite{3} for the minimization problem in (77), it can be shown that the infimum is attained at when 
\[ u^* = v^* = \xi_1(1) \]

Therefore, the decay rate function of \( p_{\text{miss}}^{(1)} \) is given by

\[ \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log p_{\text{miss}}^{(1)} = \frac{\left(\sqrt{u^*} - \sqrt{\xi_1(1)}\right)^2}{2} + \frac{N - K}{2} v^* \]

\[ = \frac{\xi_1(1)}{2(1 + \frac{1}{N-K})} \] \hspace{1cm} (78)

**Part 2:** Now consider \( \bar{p}_{\text{miss}}^{(2)} = \sum_{j=1}^{N-K-1} \Pr\{T_1^{(2)} < T_j^{(2)}\} \), where recall that \( T_1^{(2)} \sim \chi^2_2(\lambda_1^{(2)}) \) and a statistical property of \( T_j^{(2)} \) was established in Lemma \[ \square \]

Define \( T_1^{(2)} \triangleq \frac{T_1^{(2)}}{E_{\text{tot}}} \) and \( T_j^{(2)} \triangleq \frac{T_j^{(2)}}{E_{\text{tot}}} \). We thus have \( \bar{p}_{\text{miss}}^{(2)} = \sum_{j=1}^{N-K-1} \Pr\{T_1^{(2)} < T_j^{(2)}\} = \sum_{j=1}^{N-K-1} \Pr\{T_1^{(2)} < T_j^{(2)}\} \). Since \( \bar{T}_1^{(2)} \) and \( \bar{T}_j^{(2)} \) are independent, one can first derive the rate function for each and then combine to characterize the rate function of \( \bar{p}_{\text{miss}}^{(2)} \).

Specifically, in similar lines of proof for \( \bar{T}_1^{(1)} \), it is standard to show that the rate function of \( \bar{T}_1^{(2)} \) is

\[ I_{\bar{T}_1^{(2)}}(u) = \frac{\left(\sqrt{u} - \sqrt{\xi_1^{(2)}}\right)^2}{2}, \quad u \geq 0 \] \hspace{1cm} (79)

where \( \xi_1^{(2)} = \frac{2|\gamma|^2 F_{R} W_p \left( \frac{\alpha}{N} + \frac{1-\alpha}{N-K} \right)}{\sigma^2} \).

For \( \bar{T}_j^{(2)} \), we first evaluate its logarithmic MGF as

\[ \Lambda(t_2) \]

\[ = \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log E\left[ e^{t_2 T_j^{(2)}} \right] \] \hspace{1cm} (80)

\[ = \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log E\left[ e^{T_j^{(2)} t_2} \right] \] \hspace{1cm} (81)

\[ = \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \mathbb{E}_{T_1^{(1)}(K+j)} \left[ \mathbb{E}\left[ e^{t_2^' \frac{e^{(1)}_{T_1^{(1)}(K+j)}} E_{(2)} T_j^{(2)} } | T_1^{(1)}(K+j) = x \right] \right] \] \hspace{1cm} (82)

\[ \overset{(C.1)}{=} \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \mathbb{E}_{T_1^{(1)}(K+j)}^{x} \left[ \frac{e^{t_2^' \frac{e^{(1)}_{T_1^{(1)}(K+j)}} E_{(2)} } }{1 - 2t_2^'} \right] \] \hspace{1cm} (83)

\[ \overset{(C.1)}{=} \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \frac{1}{1 - 2t_2^'} \mathbb{E}_{T_1^{(1)}(K+j)}^{x} \left[ e^{t_2^' \frac{e^{(1)}_{T_1^{(1)}(K+j)} }{1 - 2t_2^' } x} \right] \] \hspace{1cm} (84)
\[(C.1),(C.2) \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \left[ \log \frac{1}{1 - 2t_2'} \prod_{l=1}^{K+j} \frac{(N - l)/2}{E^{(1)} t_2'} - \frac{E^{(1)} t_2'}{E^{(2)}(1 - 2t_2')} \right] = 0, \quad (85)\]

where we have

- auxiliary variable \( t_2' \triangleq t_2 \frac{E^{(2)}}{E^{(1)} + E^{(2)}}; \)
- condition (C.1) reads as
  \[
  t_2' < \frac{1}{2} \Rightarrow t_2 < \frac{1}{2} \left( 1 + \frac{E^{(1)}}{E^{(2)}} \right); \quad (87)
  \]
- condition (C.2) reads as
  \[
  \frac{E^{(1)} t_2'}{E^{(2)}(1 - 2t_2')} < \frac{N - (K + j)}{2} \Rightarrow t_2 < \frac{1}{2} \left( 1 + \frac{E^{(1)}}{E^{(2)}} \right) \frac{N - (K + j)}{E^{(1)} E^{(2)} + N - (K + j)}, \quad (89)
  \]
  which is stronger than (C.1);
- Equation (83) follows from fact (39) of Lemma 2 and the MGF of a noncentral chi-squared distributed variable;
- Equation (85) uses the fact that \( T_{(K+j)}^{(1)} \) (the \((K+j)\)th order statistic of \((N-1)\) i.i.d. variables \( \sim \chi^2_2(0) \)) is statistically equivalent to \( \sum_{l=1}^{(K+j)} Z_l \), where \( Z_l \sim \text{Exp} \left( \frac{N-1}{2} \right) \) [25, Chapter 1], and then follows from the MGF of an exponential distributed variable.

Therefore, the logarithmic MGF of \( \bar{T}_j^{(2)} \) exists as an extended real number as

\[
\Lambda(t_2) = \begin{cases} 
0, & \text{if } t_2 < \frac{1}{2} \left( 1 + \frac{E^{(1)}}{E^{(2)}} \right) \frac{N - (K + j)}{E^{(1)} E^{(2)} + N - (K + j)} \triangleq \beta_j \\
+\infty, & \text{otherwise.}
\end{cases} \quad (90)
\]

In addition, it is clear that \( 0 \in D_{\Lambda} = \{ t_2 : \Lambda(t_2) < \infty \} \). Therefore, \( \bar{T}_j^{(2)} \) satisfies the Gartner-Ellis conditions.

By the Gartner-Ellis theorem, the rate function of \( \bar{T}_j^{(2)} \) can thus be calculated as

\[
I_{\bar{T}_j^{(2)}}(v) = \sup_{t_2 \in \mathbb{R}} \{ v t_2 - \Lambda(t_2) \} \quad (91)
\]

\[
= \sup_{t_2 < \beta_j} \{ v t_2 \} \quad (92)
\]
\[
= \beta_j v, \quad v \geq 0.
\]

With rate functions \( I_{T_1^{(2)}}(u) \) and \( I_{T_j^{(2)}}(v) \), we thus have

\[
- \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \Pr\{\bar{T}_1^{(2)} < \bar{T}_j^{(2)}\} = \inf_{0 \leq u \leq v} \{I_{T_1^{(2)}}(u) + I_{T_j^{(2)}}(v)\} = \inf_{0 \leq u \leq v} \left\{ \frac{\xi_1^{(2)}}{2} \right\}^2 + \frac{N - K}{2} v
\]

Using the Karush-Kuhn-Tucker conditions \([31]\) for the minimization problem in (96), it can be shown that the infimum is attained at when \( u^* = v^* = \frac{\xi_1^{(2)}}{(1 + 2\beta_j)^2} \). Therefore, the decay rate function of \( \Pr\{\bar{T}_1^{(2)} < \bar{T}_j^{(2)}\} \) is given by

\[
- \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \Pr\{\bar{T}_1^{(2)} < \bar{T}_j^{(2)}\} = \frac{\xi_1^{(2)} \beta_j}{1 + 2\beta_j}.
\]

Consequently, the decay rate function of \( \bar{p}_{\text{miss}}^{(2)} = \sum_{j=1}^{N-K-1} \Pr\{\bar{T}_1^{(2)} < \bar{T}_j^{(2)}\} \) is given by

\[
- \lim_{E_{\text{tot}} \to \infty} \frac{1}{E_{\text{tot}}} \log \bar{p}_{\text{miss}}^{(2)} = \min_{j \in [1:N-K-1]} \frac{\xi_1^{(2)} \beta_j}{1 + 2\beta_j} = \min_{j \in [1:N-K-1]} \frac{\xi_1^{(2)}}{\beta_j} + 2 = \frac{\xi_1^{(2)}}{\beta(N-K-1)} + 2 = \frac{\xi_1^{(2)}}{4}
\]

where (100) follows by the fact that

\[
\frac{1}{\beta_j} = \frac{2}{1 + \frac{E^{(1)}}{E^{(2)}}} + \frac{N - (K + j)}{N - (K + j)},
\]

and

\[
\frac{E^{(1)}}{E^{(2)}} = \frac{2}{1 + \frac{\alpha N - K}{N - (K + j)}} \left( 1 + \frac{\alpha N - K}{N - (K + j)} \right),
\]
which attains its maximum value 2 when \( j = N - K - 1 \).

Finally, taking the minimum of rate functions of \( \bar{p}_{\text{miss}}^{(1)} \) and \( \bar{p}_{\text{miss}}^{(2)} \) established gives a characterization of the rate function of \( \bar{p}_{\text{miss}} \). This completes the proof of Proposition 3.

**APPENDIX D**

**PROOF OF PROPOSITION 4**

Consider the following optimization problem (P.1):

\[
\max_{\alpha,K} I_{\bar{p}_{\text{miss}}} (\alpha, K) \triangleq \min \left\{ \frac{\xi_1^{(1)}}{2 \left( 1 + \frac{1}{N-K} \right)}, \frac{\xi_1^{(2)}}{4} \right\} \tag{103}
\]

subject to \( \alpha \in (0, 1], \ K \in [1 : N - 1] \),

where \( \xi_1^{(1)} = \frac{2|\gamma|^2 F_R W_T \alpha}{\sigma^2 N} \) and \( \xi_1^{(2)} = \frac{2|\gamma|^2 F_R W_T (\frac{\alpha}{N} + \frac{1-\alpha}{N-K})}{\sigma^2} \).

Note that for any given \( N \) and some feasible \( K \), the first term in \( I_{\bar{p}_{\text{miss}}} (\alpha, K) \) monotonically increases as \( \alpha \) increases, while the second term monotonically decreases as \( \alpha \) increases. Therefore, for any given \( K \), \( I_{\bar{p}_{\text{miss}}} (\alpha, K) \) attains its maximum when the two terms equal, i.e.,

\[
\frac{\xi_1^{(1)}}{2 \left( 1 + \frac{1}{N-K} \right)} = \frac{\xi_1^{(2)}}{4}, \tag{105}
\]

which implies that the optimal \( \alpha^* \) under a given \( K \) (i.e., \( \alpha^*(K) \)) is

\[
\alpha^*(K) = \frac{N(N-K+1)}{2(N-K)^2 + K(N-K+1)}. \tag{106}
\]

As a result, solving (P.1) boils down to finding the optimal \( K^* \) for the following maximization problem (P.2):

\[
\max_{K} I_{\bar{p}_{\text{miss}}} (\alpha^*(K), K) = \frac{2|\gamma|^2 F_R W_T}{\sigma^2 N} \frac{\alpha^*(K)}{2 \left( 1 + \frac{1}{N-K} \right)} \tag{107}
\]

\[
\propto \frac{\alpha^*(K)}{\left( 1 + \frac{1}{N-K} \right)} \tag{108}
\]

subject to \( K \in [1 : N - 1] \).

Specifically, with \( \alpha^*(K) \) of (106), the objective (108) can be rewritten as

\[
\frac{N(N-K)}{(N-K)^2 + (N-K)(N-1) + N} \tag{110}
\]
\[
N \leq \frac{N}{N - K} + \frac{N}{N - K} + (N - 1), \quad (111)
\]
\[
\frac{N}{2\sqrt{N} + (N - 1)}, \quad (112)
\]
where the equality above attains when \(N - K = \sqrt{N}\) in general. Given that \(K\) is restricted to an integer in \((P.2)\), then the optimal \(K^*\) is
\[
K^* = N - \text{round}(\sqrt{N}), \quad (113)
\]
and the optimal \(\alpha^*(K^*)\) is evaluated through \((106)\) at \(K^*\). The corresponding optimal rate function can be represented as
\[
I_{\text{phase}}(\alpha^*(K^*), K^*) = \frac{|\gamma|^2 F_{RW} T}{\sigma^2 N} \frac{\alpha^*(K^*)}{2(1 + \frac{1}{N - K^*})} \quad (114)
\]
\[
= \frac{|\gamma|^2 F_{RW} T}{2\sigma^2 \left( N - \frac{K^*(N - K^* - 1)}{2(N - K^*)} \right)}. \quad (115)
\]

REFERENCES

[1] M. Li, C. Liu, S. V. Hanly, I. B. Collings, and P. Whiting, “Beam alignment with two-stage search for millimeter-wave communications,” submitted to IEEE Int. Conf. on Comm. (2019).
[2] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” IEEE Commun. Mag., vol. 49, no. 6, 2011.
[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5G be?” IEEE J. Sel. Areas Comms., vol. 32, no. 6, pp. 1065–1082, 2014.
[4] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, I. Chih-Lin et al., “Millimeter wave communications for future mobile networks,” IEEE J. Sel. Areas Commun., vol. 35, no. 9, pp. 1909–1935, 2017.
[5] J. Lee et al., “Spectrum for 5G: Global status, challenges, and enabling technologies,” IEEE Commun. Mag., vol. 56, no. 3, pp. 12–18, 2018.
[6] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, “Millimeter-wave vehicular communication to support massive automotive sensing,” IEEE Commun. Mag., vol. 54, no. 12, pp. 160–167, 2016.
[7] C. Liu, M. Li, S. V. Hanly, P. Whiting, and I. B. Collings, “Millimeter wave small cells: Base station discovery, beam alignment and system design challenges,” IEEE Wireless Commun. Mag., vol. 25, no. 4, pp. 40–46, August 2018.
[8] J. Wang et al., “Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems,” IEEE J. Sel. Areas Commun., vol. 27, no. 8, pp. 1390–1399, 2009.
[9] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter wave beamforming for wireless backhaul and access in small cell networks,” IEEE Trans. Commun., vol. 61, no. 10, pp. 4391–4403, 2013.
[10] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE J. Sel. Sig. Processing, vol. 8, no. 5, pp. 831–846, 2014.
[11] Z. Xiao, T. He, P. Xia, and X. G. Xia, “Hierarchical codebook design for beamforming training in millimeter-wave communication,” IEEE Trans. Wireless Commun., vol. 15, no. 5, pp. 3380–3392, 2016.

[12] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, “Millimeter wave beam alignment: Large deviations analysis and design insights,” IEEE J. Sel. Areas Commun., vol. 35, no. 7, pp. 1619–1631, July 2017.

[13] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, “Codebook design for beam alignment in millimeter wave communication systems,” IEEE Transactions on Communications, vol. 65, no. 11, pp. 4980–4995, 2017.

[14] S. Haghighatshoar and G. Caire, “The beam alignment problem in mmwave wireless networks,” in Proc. 50th Asilomar Conf. on Signals, Syst. and Comput., 2016, pp. 741–745.

[15] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, “Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation,” IEEE Trans. Signal Process., vol. 65, no. 3, pp. 601–616, 2017.

[16] A. Alkhateeb, Y.-H. Nam, M. S. Rahman, J. Zhang, and R. W. Heath, “Initial beam association in millimeter wave cellular systems: Analysis and design insights,” IEEE Trans. Wireless Commun., vol. 16, no. 5, pp. 2807–2821, 2017.

[17] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, “Inverse multipath fingerprinting for millimeter wave V2I beam alignment,” IEEE Trans. Veh. Technol., vol. 67, no. 5, pp. 4042–4058, 2018.

[18] 3GPP-R1-167672, “Synchronization in NR considering beam sweeping,” 2016.

[19] 3GPP-R1-167543, “Beam management considerations for above 6 GHz NR,” 2016.

[20] V. Raghavan, L. Jianghong, R. N. Challa, L. Qijia, J. Li, A. Y. Gorokhov, and A. Touboul, “Steady-state beam scanning and codebook generation,” 2018, uS Patent App. 15/684,861.

[21] C. Lin, G. Y. Li, and L. Wang, “Subarray-based coordinated beamforming training for mmWave and sub-THz communications,” IEEE J. Sel. Areas Commun., vol. 35, no. 9, pp. 2115–2126, 2017.

[22] C. Liu, M. Li, I. B. Collings, S. V. Hanly, and P. Whiting, “Design and analysis of transmit beamforming for millimeter wave base station discovery,” IEEE Trans. Wireless Commun., vol. 16, no. 2, pp. 797–811, 2017.

[23] C. Barati Nt., S. Hosseini, S. Rangan, P. Liu, T. Korakis, S. Panwar, and T. Rappaport, “Directional cell discovery in millimeter wave cellular networks,” IEEE Trans. Wireless Comms., vol. 14, no. 12, pp. 6664–6678, 2015.

[24] A. Constantine Balanies, Antenna theory analysis and design. John Wiley & Sons Inc., New York, 2005.

[25] N. Balakrishnan and C. R. Rao, Handbook of statistics. v.16: Order statistics: theory and methods. North Holland, Amsterdam (Netherlands), 1998.

[26] I. S. Gradshteyn and I. M. Ryzhik, Table of integrals, series, and products. Academic press, 2014.

[27] Z. Muhi-Eldeen, L. P. Ivrissimtzis, and M. Al-Nuaimi, “Modelling and measurements of millimetre wavelength propagation in urban environments,” IET microwaves, antennas and propagation, vol. 4, no. 9, pp. 1300–1309, 2010.

[28] W. Fan, C. Zhang, and Y. Huang, “Flat beam design for massive mimo systems via riemannian optimization,” IEEE Wireless Commun. Lett., 2018, available via Early Access.

[29] G. Cui, L. Kong, X. Yang, and D. Ran, “Two useful integrals involving generalised marcum q-function,” Electronics letters, vol. 48, no. 16, pp. 1017–1018, 2012.

[30] A. Dembo and O. Zeitouni, Large deviations techniques and applications. Berlin, Germany: Springer, 2009.

[31] S. Boyd and L. Vandenberghe, Convex optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.