Beam alignment via Phase Matching

Chi-Shiang Gau
Department of Electrical and Computer Engineering
University of California, San Diego
cgau@ucsd.edu

Tara Javidi
Department of Electrical and Computer Engineering
University of California, San Diego
tjavidi@ucsd.edu

2024 IEEE Wireless Communications and Networking Conference (WCNC) | 979-8-3503-0358-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/WCNC57260.2024.10570615

***Abstract*—Millimeter-wave (mmWave) with phase array systems are utilized to increase the data rate in wireless communication systems. With the larger phase array, narrower beams can be generated to compensate for the path loss and boost the effective SNR. In order to perform beamforming, however, one must find the Angle of Arrival (AoA). In this paper, we consider the problem of initial beam alignment and CSI acquisition for (sub-)mmWave communication over a single-path channel with a single RF-chain. A three stage adaptive alignment algorithm based on the posterior probability, called Adaptive Phase Matching, is proposed. The algorithm maps the candidate angles of arrivals onto a simple constellation on the complex plane adaptively. The proposed scheme generalizes HiePM to both account for practically feasible antenna patterns as well as utilize the phase information. Furthermore, an upper bound of the expected search time for any given alignment resolution and error probability is derived.**

***Index Terms*—millimeter wave communication, adaptive search, posterior probability, random mapping, feedback channel**

I. Introduction

In order to achieve higher data rates for communication systems, millimeter-wave (mmWave) systems with large frequency bandwidth are deployed. It is well known though that mmWave systems have a large path loss nature. Massive antenna array systems then are, in fact, needed to perform highly directional communications by generating narrow beams to compensate for the large path loss. However, before performing the beamforming technique, the transmitter and receiver must find the AoA. Traditional non-adaptive channel/AoA estimation algorithms such as MUSIC [1] and ESPRIT [2] have high resolution performance. The beam alignment problem can also be solved with a random hashing code-book [3]–[5]. In particualr, in [3], the random hashing beams are combined with voting algorithms, while [4] solve the problem with compressive sensing techniques. The random hashing code book can also be combined with machine learning techniques to find the AoA [5]. However, these algorithms require sufficiently large raw SNR, making them practically unsuitable for mmWave communication.

Drawing on the connection to the problem of search with measurement-dependent noise [6], adaptive beam alignment algorithms, such as SortPM [7] and HiePM [8], select the next beamforming vector in a dynamic fashion according to the posterior probability. These adaptive strategies are shown to outperform the open-loop strategies particularly in low SNR regimes. In particular, [8] provided a lower bound for this performance gain over open-loop strategies. This work has since be generalized by follow up work have relaxed practically relevant assumptions on the model mismatch, transmitter/receive structure, etc. In [10], for instance, error control sounding strategies are proposed. Other adaptive algorithms are also proposed to maximize the data rate instead of finding the precise AoA [11].

Prior adaptive strategies, however, all have relied on an ideal rectangular beams or the approximate rectangular beams [12] in order to distinguish the candidate angles with different magnitude response. In reality, the received beamformed signals are complex-valued vectors. In short, for a given beam, the collection of the responses for different input angles can be viewed as a constellation on the complex plane. We are motivated by the use of phase information in many of the open loop strategies, including antenna selection. In this paper, we generalize SortPM [8] and investigate the potential benefit of using both magnitude and phase information to distinguish the candidate angles by designing some simple constellations. Specifically, for each observation, the agent designs a received beamforming vector that maps the candidate angles onto a constellation on the complex plane. An adaptive mapping algorithm, called Phase Matching, is proposed based on posterior matching [9] and an upper bound for the expected searching time is derived with the Extrinsic Jensen Shannon (EJS) divergence [14] [15].

In this paper, our goal is to generalize the proposed adaptive beamforming and alignment algorithms and to investigate the potential benefit of using the extra phase information. Therefore, we simply follow the setting of [7] and [8] where the channel coefficient is assumed to be known. We note, however, that prior works [10], [13] have proposed methods to allow the use of adaptive schemes despite of unknown channel gains. Thus, we believe utilizing similar techniques to allow for uncertainty in the estimation of channel gains is an important yet straight-forward extension best left for future (and more detailed) work.

Notations: We use boldface letters to represent vectors. $P(x)$ denotes the probability mass function for random variable X and $E[X]$ indicates the expectation of X. $D(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$ denotes the Kullback-Leibler (KL) divergence between two distribution P and Q. We use $\boldsymbol{\pi}^{\downarrow}(t)$ to indicate the sorted version of the vector $\boldsymbol{\pi}(t)$ in descending order elements and $\pi_i^{\downarrow}(t)$ represent the $i-th$ largest element

of $\boldsymbol{\pi}(t)$. We denote the capacity of a Gaussian channel with Q input symbols (a Q points constellation $\mathcal{X}_Q$ on the complex plane) and noise with standard deviation σ as $C(\mathcal{X}_Q, \sigma)$.

II. Problem formulation

Consider a receiver with an antenna array with M antenna elements and a single RF chain. Given a beamforming vector, the received signal can be written as the weighted sum of the signal and noise samples on each antenna element. The structure of the receiver is shown in Fig. 1. Our goal is to design the dynamic selection of the beamforming vector $\boldsymbol{w}$ based on all the past observations to find the true AoA with a minimum number of observations.

A. System Model

The normalized single-path channel model[1] with M antenna elements at the receiver can be written as:

$$\boldsymbol{h} = \boldsymbol{a}(\phi) \tag{1}$$

where $\boldsymbol{a}(\phi)$ denotes the $M \times 1$ steering vector with impinge angle ϕ for a half wavelength spacing antenna array:

$$\boldsymbol{a}(\phi) := [1, e^{-j\pi \sin\phi}, e^{-j2\pi \sin\phi}, ..., e^{-j(M-1)\pi \sin\phi}]^T \tag{2}$$

Here, the angle-of-arrival ϕ is assumed to be fixed during the initial beam alignment process. Assuming $\boldsymbol{w}(t)$ is the $M \times 1$ receiver beamforming vector, the received signal at time t can be written as:

$$y(t) = \boldsymbol{w}^H(t)\boldsymbol{h} + \boldsymbol{w}^H(t)\boldsymbol{n}(t) \tag{3}$$
$$= \boldsymbol{w}^H(t)\boldsymbol{a}(\phi) + \boldsymbol{w}^H(t)\boldsymbol{n}(t) \tag{4}$$
$$= \boldsymbol{w}^H(t)\boldsymbol{a}(\phi) + v(t) \tag{5}$$

where $\boldsymbol{n}(t)$ is the $M \times 1$ AWGN vector with known covariance matrix $\sigma^2 I$ and $v(t) \sim CN(0, ||\boldsymbol{w}(t)||^2\sigma^2)$ is the noise term after beamforming.

In order to calculate the posterior probability, we quantize the angular space $[-\pi/2, \pi/2]$ into M sections denoted as $[\theta_1, \theta_2, ..., \theta_M]$. Also, we assume that the true AoA ϕ belongs to one of the quantized angles:

$$\phi \in \Omega := [\theta_1, \theta_2, ..., \theta_M] \tag{6}$$

Note that the resolution for the quantized angles is defined as $\delta := \frac{1}{M}$. Moreover, in order to get rid of the sin function in (2), here we applied a non-linear quantization for the angle space. Specifically, we let $[\theta_1, \theta_2, ..., \theta_M] = \arcsin(2[0, 1, ..., \frac{M}{2}, -\frac{M}{2}+1, -\frac{M}{2}+2, ...-2, -1]/M)$, then the steering vector for θ_i can be written as:

$$\boldsymbol{a}(\theta_i) = [1, e^{-j2\pi(i-1)/M}, ..., e^{-j2\pi(i-1)(M-1)/M}]^T \tag{7}$$

for $i = 1, 2, ..., M$.

With the above setting, we can stack all the steering vectors together and define the steering matrix $\mathbf{A}(\boldsymbol{\theta})$ as:

$$\mathbf{A}(\boldsymbol{\theta}) := [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), ...\mathbf{a}(\theta_M)] = [DFT]_{M\times M} \tag{8}$$

With this non-uniform quantization, it is shown that the steering matrix for all candidate angles is the $M \times M$ DFT matrix.

[1]We rely on an (implicit) assumption of a fixed complex channel gain.

B. Adaptive Alignment with Fixed Constellations

In this paper, we proposed a three stage adaptive searching algorithm to find the AoA by mapping the candidate angles of arrivals to target constellation in complex plane parameterized by the number of points on the unit circle, K. More precisely, given the design parameter $K \ll M$, we introduce three simple constellations that will be used to design the beamforming vector $\boldsymbol{w}$ for our proposed three stages of operation.

1) Constellation 1:

$$\mathcal{X}_K^{\text{initial}} := [x_1, x_2, x_3, ..., x_K] \tag{9}$$
$$= [1, e^{j\frac{2\pi}{K}}, e^{j\frac{2*2\pi}{K}}, ..., e^{j\frac{(K-1)*2\pi}{K}}] \tag{10}$$

2) Constellation 2:

$$\mathcal{X}_{K+1}^{\text{zoom-in}} := [x_1, x_2, x_3, ..., x_K, x_0] \tag{11}$$
$$= [1, e^{j\frac{2\pi}{K}}, e^{j\frac{2*2\pi}{K}}, ..., e^{j\frac{(K-1)*2\pi}{K}}, 0] \tag{12}$$

3) Constellation 3:

$$\mathcal{X}_2^{\text{confirm}} := [x_1, x_0] = [1, 0] \tag{13}$$

Note that the subscript of $\mathcal{X}$ indicates the distinct points in the constellation with their number K as a design parameter.

Our proposed algorithm at any given time selects a mapping $\gamma : \Omega \to \mathcal{X}$ that maps the M candidate angles onto the constellation points. More precisely, the algorithm selects a constellation $\mathcal{X} \in [\mathcal{X}_K^{\text{initial}}, \mathcal{X}_{K+1}^{\text{zoom-in}}, \mathcal{X}_2^{\text{confirm}}]$ and then maps the M candidate angles onto the constellation points $\mathcal{X}$.

Next, we show that any mapping can be realized by designing the received beamforming vector $\boldsymbol{w}$. Let $d_i(t) := \gamma_t(\theta_i) \in \mathcal{X}$ denote the point that θ_i will be mapped onto at time t. This means the beamforming vector $\boldsymbol{w}(t)$ satisfies the following linear equation:

$$\boldsymbol{w}^H(t)\mathbf{a}(\theta_i) = d_i(t). \tag{14}$$

By combining all M linear equations for all candidate angles, $[\theta_1, \theta_2, ..., \theta_M]$, we can rewrite (14):

$$\boldsymbol{d}_\gamma(t) := [d_1(t), d_2(t), ..., d_M(t)] = \boldsymbol{w}_\gamma^H(t)\boldsymbol{A}(\boldsymbol{\theta}). \tag{15}$$

In other words,

$$\boldsymbol{w}_\gamma(t) = \frac{1}{M}[DFT]\boldsymbol{d}_\gamma(t)^H. \tag{16}$$

Fig. 3 and Fig. 4 illustrate two instances of different mappings where in each figure samples of the (noisy) outputs for all $M = 1024$ candidate angle-of-arrivals are shown. Both mappings map the candidate AoAs to the constellation $\mathcal{X}_{8+1}^{\text{zoom-in}}$. In Fig. 3 (on the left) the beamforming is given as

$$\boldsymbol{w}_{\text{case1}}^H(t)\mathbf{a}(\theta_i) = \begin{cases} x_{mod(\frac{i}{8})+1} & \text{for } i = 1, ..., 64 \\ x_0 & \text{for } i = 65, ..., 1024 \end{cases} \tag{17}$$

while in Fig. 4 the same mapping is applied to a random permutation of candidate AoAs; i.e. $\mathbf{d}(t)$ used for each figure is a column permutation of the other.

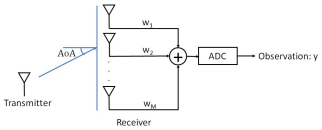

Fig. 1. M antenna elements with 1 RF chain

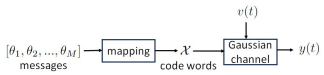

Fig. 2. Gaussian channel

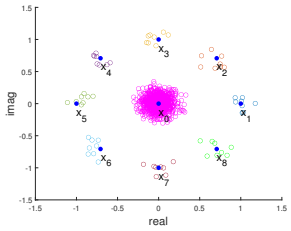

Fig. 3. Received signals (complex) associated with $M = 1024$ candidate AoAs when $\boldsymbol{w}_{\text{case1}}$ is selected. The blue dots show the target constellation $\mathcal{X}_{8+1}^{\text{zoom in}}$.

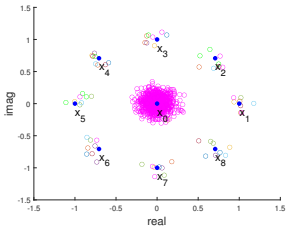

Fig. 4. Received signals (complex) associated with $M = 1024$ candidate AoAs where the beamforming vector is a permutation of $\boldsymbol{w}_{\text{case1}}$.

C. Gaussian Channel

Since we restrict the received beamforming vector $\boldsymbol{w}$ to be the vectors with the form of (16), the observed signal model (5), can be viewed as the output of a complex Gaussian channel with discrete input alphabet $\mathcal{X} \in [\mathcal{X}_K^{\text{initial}}, \mathcal{X}_{K+1}^{\text{zoom-in}}, \mathcal{X}_2^{\text{confirm}}]$ [16] and additive Gaussian noise. Accordingly, one can view our (candidate angles), $[\theta_1, \theta_2, ..., \theta_M]$ as M message with M message mapped into the (coded) input symbol $\boldsymbol{w}^H(t)\boldsymbol{a}(\theta_i) \in \mathcal{X}$ as shown in Fig. 2. Note that since we define three different constellations, $[\mathcal{X}_K^{\text{initial}}, \mathcal{X}_{K+1}^{\text{zoom-in}}, \mathcal{X}_2^{\text{confirm}}]$, here we have three different Gaussian channels which are adaptively selected.

This means that our design of adaptive alignment strategies is equivalent to feedback coding over an additive Gaussian channel with the caveat that the noise variance of our channel depends on our mapping (and more precisely on how many distinct candidate AoAs are mapped onto the unit circle).

D. Posterior probability

Now we define the posterior probability for the candidate angles: $boldsymbol\pi(t) := [\pi_1(t), \pi_2(t), ..., \pi_M(t)]$

where $\pi_i(t)$ is the posterior probability for θ_i: $\pi_i(t) := p(\theta_i | y(1), y(2), ..., y(t))$

The initial prior probability are assumed to be uniformly distributed over all M candidate angles:

$$\pi_i(0) = 1/M, \ \ i = 1, 2, ..., M \tag{18}$$

The updated posterior probability can be calculated with the Bayes rule:

$$\pi_i(t+1) = \frac{\pi_i(t) f(y(t+1)|\theta_i, \boldsymbol{w}(t+1))}{\sum_j \pi_j(t) f(y(t+1)|\theta_j, \boldsymbol{w}(t+1))} \tag{19}$$

where

$$f(y(t+1)|\theta_i, \boldsymbol{w}(t+1)) \sim CN(y(t+1); \boldsymbol{w}^H(t+1)\boldsymbol{a}(\theta_i), \sigma^2) \tag{20}$$

is the likelihood function.

III. Adaptive Phase Matching

In this paper, we propose a three stage adaptive searching algorithm to find the AoA:

1) Initial search: searching with $\mathcal{X}_K^{\text{initial}}$,
2) Zoom in search: searching with $\mathcal{X}_{K+1}^{\text{zoom-in}}$, and
3) Confirmation: searching with $\mathcal{X}_2^{\text{confirm}}$.

In the first two stages, we apply Sorted Posterior Matching (sortPM) [9], [14], [7] to the constellation $\mathcal{X} \in [\mathcal{X}_K^{\text{initial}}, \mathcal{X}_{K+1}^{\text{zoom-in}}]$ for that stage. The main difference between these stages is that in the second stage the noise variance is controlled via a design parameter $\alpha \in [0, 1]$ by ensuring that only αM of candidate AoAs are mapped onto unit circle while the rest on mapped to 0. In the last stage, we simply separate the candidate angle with the largest probability from other candidate angles by mapping them onto x_1 and x_0 in $\mathcal{X}_2^{\text{confirm}}$, respectively. Note that while we have presented the three stages sequentially, the algorithm uses the posterior probability to select the appropriate stage. The decision on how to navigate between stages is determined by the shape of the posterior which is compared to two thresholds: β_0 and β_1.

Next we first describe sortPM algorithm applied to an additive Gaussian channel with discrete complex input alphabet $\mathcal{X}$. We then proceed and detail the choice of thresholds and the adaptive scheme navigating between our stages.

A. Posterior Matching and Sorted Posterior Matching

Posterior Matching (PM) [9] [14] is a random mapping scheme that maps $[\theta_1, \theta_2, ..., \theta_M]$ onto the Q points on the constellation $\mathcal{X}_Q$ according to the posterior probability $\boldsymbol{\pi}(t)$.

Assume we have a Gaussian channel with constellation $\mathcal{X}_Q$ and observation noise with standard deviation σ. Let $C(\mathcal{X}_Q, \sigma)$ represents the capacity of the Gaussian channel with capacity achieving distribution $[\pi_1^*, \pi_2^*, ..., \pi_Q^*]$. Then, for a given posterior probability $\pi_i(t)$ of the $i-th$ candidate angle, θ_i, the Posterior Matching scheme will map θ_i onto the points within a set $S_i :=$

$$\left\{ x_{z_i} \in \mathcal{X}_Q : \sum_{i'=1}^{i-1} \pi_{i'} < \sum_{x' \le x_{z_i}} \pi_{x'}^* \ and \sum_{x' < x_{z_i}} \pi_{x'}^* \le \sum_{i'=1}^{i} \pi_{i'} \right\} \tag{21}$$

and the angle θ_i will be mapped to $x_{z_i} \in S_i$ with probability $\Lambda_{\theta_i, x_{z_i}} :=$

$$\frac{\min\{\sum_{i'=1}^{i} \pi_{i'}, \sum_{x' \le x_{z_i}} \pi_{x'}^*\} - \max\{\sum_{i'=1}^{i-1} \pi_{i'}, \sum_{x' < x_{z_i}} \pi_{x'}^*\}}{\pi_i} \tag{22}$$

Recall that the behavior of mapping the candidate angle θ_i onto the constellation point x_{z_i} at time t is done by assigning $d_i(t) = x_{z_i}$, as shown in (14). Since now we assign x_{z_i} to $d_i(t)$ randomly with probability $\Lambda_{\theta_i, x_{z_i}}$. This means that the vector $\boldsymbol{d}(t)$ in (15) will be assigned with a specific vector $\boldsymbol{d}_\gamma(t) := [x_{z_1}, x_{z_2}, ..., x_{z_M}]$ with probability

$\lambda_\gamma := \Lambda_{\theta_1,x_{z_1}} \times \Lambda_{\theta_2,x_{z_2}} \times ... \times \Lambda_{\theta_M,x_{z_M}}$. Also, the corresponding received beamforming vector $\boldsymbol{w}_\gamma(t)$ is obtained from (15):

$$\boldsymbol{w}_\gamma(t) = \frac{1}{M}[DFT]\boldsymbol{d}_\gamma(t)^H \tag{23}$$

Finally, we select $\boldsymbol{w}(t)$ to be $\boldsymbol{w}_\gamma(t)$ with probability λ_γ. We propose to utilize sortPM [7] [15]: a variation of Posterior Matching where PM mapping is applied to the sorted posterior $\boldsymbol{\pi}^\downarrow(t)$. It is shown in [9] and [7] that when $M \gg K$ this algorithm can achieve the capacity $C(\mathcal{X}_Q, \sigma)$ of the channel. Note that the calculation for sortPM requires: a sorting operation with complexity $O(M)$, finding the mapping regime for M angles, which can be calculated with matrix operation with complexity $O(M \times Q)$, and a random mapping operation with complexity $O(Q)$ for M angles. Therefore, the overall complexity will be $O(M \times Q)$.

B. Thresholds and Capacity for the stages

In the first stage, constellation $\mathcal{X}_K^{\text{initial}}$ is a symmetric pattern with K points around the unit circle. Therefore, the capacity achieving distribution for these K points is $[\pi_1^*, \pi_2^*, ..., \pi_K^*] = [\frac{1}{K}, \frac{1}{K}, ..., \frac{1}{K}]$. Since $d_i(t) \in \mathcal{X}_K^{\text{initial}}$ for all i in this stage, from (15), (16) and (5), we have $v(t) \sim CN(0, ||\boldsymbol{w}(t)||^2\sigma^2) = CN(0, \sigma^2)$. Then, with sortPM, we achieve $C(\mathcal{X}_K^{\text{initial}}, \sigma)$.

In the second stage, we map at most αM angles with high probability onto the points on the unit circle, $[x_1, x_2, ..., x_K]$, and map other candidate angles with low probability onto x_0. The capacity achieving distribution of constellation $\mathcal{X}_{K+1}^{\text{zoom-in}}$ can be written as $[\pi_1^*, \pi_2^*, ..., \pi_K^*, \pi_0^*] = [\frac{P_{outside}(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)}{K}, \frac{P_{outside}(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)}{K}, ..., \frac{P_{outside}(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)}{K}, 1 - P_{outside}(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)]$, where $P_{outside}(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma) := \sum_{i=1}^K \pi_i^*$. Since the power of the noise term, $v(t)$ is at most $\alpha^2\sigma^2$, in this stage, we can achieve $C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)$. Fig. 5 and Fig. 6 show the noisy observation samples for constellation $\mathcal{X}_{8+1}^{\text{zoom-in}}$ with two different values of α. Here we can see that a smaller α ensures a smaller noise variance (visually seen in the deviation from the target constellation point shown in blue) but also reduces the number of distinguishable AoAs.

To chose the threshold β_0, we note that under sortPM, (21) implies that when $\sum_{i=1}^{\alpha M} \boldsymbol{\pi}_i^\downarrow(t) \geq P_{outside}(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)$, there will be at most αM candidate angles mapped onto the unit circle. Therefore, we pick $\beta_0 := P_{outside}(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)$ to be the threshold for entering the zoom-in stage.

The algorithm enters the confirmation stage, when the candidate AoA with the largest (posterior) probability, $\pi_1^\downarrow$, becomes large become larger than threshold $\beta_1 \geq \frac{1}{2}$ [14]. Note that this stage allows for a sequential binary hypothesis test with error exponent E:

$$E := D\Big(P(Y|X = x_1)||P(Y|X = x_0)\Big) \tag{24}$$

ensuring, in effect, a very high reliability for our alignment.

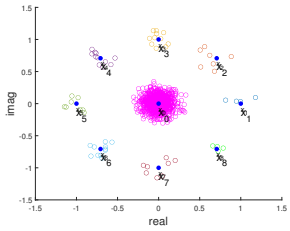

Fig. 5. Received signals (complex) associated with M = 1024 candidate AoAs with $\mathcal{X}_{8+1}^{\text{zoom-in}}$ and $\alpha = 1/16$.

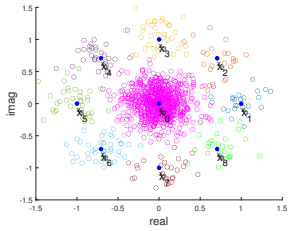

Fig. 6. Received signals (complex) associated with M = 1024 candidate AoAs with $\mathcal{X}_{8+1}^{\text{zoom-in}}$ and $\alpha = 1/4$.

C. Main Result

In our proposed three stage adaptive searching algorithm, the expected searching time for any resolution, $\delta > 0$, and any error probability, $0 < \epsilon < 1$, is upper bounded by:

$$E[\tau_{\epsilon,\delta}] \leq \frac{\log\frac{1}{\delta}}{C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)} + \frac{\log\frac{1}{\epsilon}}{E} + o(\log\frac{1}{\delta\epsilon}) \tag{25}$$

where $C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)$ is the capacity of the constellation $\mathcal{X}_{K+1}^{\text{zoom-in}}$ with noise power $\alpha^2\sigma^2$ and $\alpha \gg \frac{K}{M}$ and E is reliability of hypotheses testing as given by the KL divergence (24). Note that this result is a generalization of [15] by utilizing the phase information. If $K = 1$, (only the second and third stages) we will have the upper bound for the binary searching case as shown in [15].

IV. Analysis

Before we start the analysis, we first introduce the average log-likelihood function and Extrinsic Shannon divergence. We use the average log-likelihood to measure the uncertainty and the reduction of the uncertainty is the Extrinsic Shannon divergence (EJS) [14].

A. Average Log-likelihood

For a given posterior probability at time t, the average log-likelihood $U(t)$ is defined as:

$$U(t) \equiv U(\boldsymbol{\pi}(t)) := \sum_{i=1}^M \pi_i(t)\log\frac{\pi_i(t)}{1 - \pi_i(t)} \tag{26}$$

Property 1: $U(t)$ is a submartingale: $E[U(t+1)|\boldsymbol{\pi}(t)] = U(t) + EJS(\boldsymbol{\pi}(t), \gamma)$, where $EJS(\boldsymbol{\pi}(t), \gamma)$ is defined as:

$$EJS(\boldsymbol{\pi}(t), \gamma) = \sum_{i=1}^M \pi_i(t) D\left(P_{Y|X_i}||\sum_{j\neq i}\frac{\pi_j(t)}{1 - \pi_i(t)}P_{Y|X_j}\right) \tag{27}$$

Here, γ represents the mapping for θ_i onto X_i, $i = 1, 2, ..., M$.
Property 2: if $\pi_i(t) < 1 - \epsilon$, then $U(t) < \log\frac{1-\epsilon}{\epsilon}$

B. Good event and Bad event

In the second stage of the adaptive searching algorithm, we apply the sortPM scheme "trying" to achieve the capacity $C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)$. However, we can only achieve it when we map no more than αM angles on the unit circle. That is

$\sum_{i=1}^{\alpha M} \boldsymbol{\pi}_i^{\downarrow}(t) \geq \beta_0$. Otherwise, we can only achieve the capacity of the first stage $C(\mathcal{X}_K^{\text{initial}}, \sigma)$. Therefore, we can separate these two cases as:

- The "Bad" event: when $\sum_{i=1}^{\alpha M} \boldsymbol{\pi}_i^{\downarrow}(t) < \beta_0$, we do not achieve $C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)$ but achieve $C(\mathcal{X}_K^{\text{initial}}, \sigma)$.
- The "Good" event: when $\sum_{i=1}^{\alpha M} \boldsymbol{\pi}_i^{\downarrow}(t) \geq \beta_0$ we achieve $C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)$.

C. Proof of the Main Result

We denote E_t to be the "bad" event $\sum_{i=1}^{\alpha M} \boldsymbol{\pi}_i^{\downarrow}(t) < \beta_0$, the union of the "bad" event $F_n = \cup_{t=n}^{\infty} E_t$ and the complement of this union which is the "good" event ${F_n}^C = (\cup_{t=n}^{\infty} E_t)^C$. By using the total probability theorem, the union bound and the EJS searching theorem [14] [15], we have:

$$E[\tau_{\epsilon,\delta}] = \int_{\Omega} \tau_{\epsilon,\delta} \; d\mathbb{P} \leq \sum_{t=n}^{\infty} \int_{E_t} E[\tau_{\epsilon,\delta} | \boldsymbol{\pi}(t)] \; d\mathbb{P} + \int_{{F_n}^C} \tau_{\epsilon,\delta} \; d\mathbb{P} \tag{28}$$

$$\leq \sum_{t=n}^{\infty} \mathbb{P}(E_t) \left(t + \frac{\log(1/\delta)}{C(\mathcal{X}_K^{\text{initial}}, \sigma)} + \frac{\log(1/\epsilon)}{E} + f_{C(\mathcal{X}_K^{\text{initial}}, \sigma), E(\epsilon,\delta)} \right) + n + \frac{\log(1/\delta)}{C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)} + \frac{\log(1/\epsilon)}{E} + f_{C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma), E(\epsilon,\delta)} \tag{29}$$

where $f_{C,E(\epsilon,\delta)} := \frac{\log\log\frac{1}{\delta\epsilon}}{C} + \frac{1}{E} + \frac{96}{CE}(C_2)^2$
with $C_2 := \max_{y \in \mathcal{Y}} \frac{\max_{x \in \mathcal{X}} P(Y=y|X=x)}{\min_{x \in \mathcal{X}} P(Y=y|X=x)}$.

Now, if the probability of the "bad" event $P(E_t) \leq k_0 e^{-tE_0}$ and by letting $n = \lceil \log\log\frac{1}{\delta\epsilon} \rceil$ then we have:

$$E[\tau_{\epsilon,\delta}] \leq \frac{\log\frac{1}{\delta}}{C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma)} + \frac{\log\frac{1}{\epsilon}}{E} + g_{R,E}(\epsilon,\delta) \tag{30}$$

where

$$g_{R,E}(\epsilon,\delta) := \frac{k_0 e^{-E_0}}{(1-e^{-E_0})(\log\frac{1}{\delta\epsilon})^{E_0}} \times \left(\lceil \log\log\frac{1}{\delta\epsilon} \rceil + \frac{\log\frac{1}{\delta}}{C(\mathcal{X}_K^{\text{initial}}, \sigma)} + \frac{\log\frac{1}{\epsilon}}{E} + f_{C(\mathcal{X}_K^{\text{initial}}, \sigma), E(\epsilon,\delta)} \right) + \frac{k_0 e^{-2E_0}}{(1-e^{-E_0})^2 (\log\frac{1}{\delta\epsilon})^{2E_0}} + \lceil \log\log\frac{1}{\delta\epsilon} \rceil + f_{C(\mathcal{X}_{K+1}^{\text{zoom-in}}, \alpha\sigma), E(\epsilon,\delta)} \tag{31}$$

D. Probability of the "Bad" event

Now, we want to prove that the probability of the "bad" event $P(E_t) \leq k_0 e^{-tE_0}$. Since, we are comparing the probability of the first αM angles with β_0, we can define a grouping version average log-likelihood and use property 2 to bound this "bad" event.

$$P(E_t) = P\left(\sum_{i=1}^{\alpha M} \pi_i^{\downarrow} < \beta_0 \right) \leq P\left(U_\alpha(t) < \log\frac{1-\beta_0}{\beta_0} \right) \tag{32}$$

Where the grouping version of the average log-likelihood is defined as: (each group αM angles, total $G \equiv \frac{1}{\alpha}$ groups)

$$U_\alpha(t) \equiv U_\alpha(\boldsymbol{\pi}(t)) := \sum_{g=1}^{1/\alpha} \pi_g^\alpha \log\frac{\pi_g^\alpha}{1-\pi_g^\alpha} \tag{33}$$

where $\pi_g^\alpha := \sum_{i \in group(g)} \pi_i(t)$ and $group(g) := [\alpha M(g-1)+1, \alpha M(g-1)+2, ..., \alpha M(g)]$

Similar to property 1, the expected drift of the grouping average log-likelihood is the grouping version EJS:

$$E\left[U_\alpha(t+1)|\pi_g(t)\right] - U_\alpha(t) = \sum_g \pi_g^\alpha(t) D\left(\frac{\sum_{i \in g} \pi_i(t) P_{Y|X_i}}{\pi_g^\alpha(t)} || \frac{\sum_{j \notin g} \pi_j(t) P_{Y|X_j}}{1-\pi_g^\alpha(t)} \right) \tag{34}$$

We can see that the group EJS is the weighted sum of the KL-divergence of each group against the mixture of other groups.

$$Group\ EJS = \sum_g \pi_g^\alpha(t) D\left(P_{mix\ of\ Group\ g} || P_{mix\ of\ non-Group\ g} \right) \tag{35}$$

Now, we will derive the probability of "bad" event. We first derive the bound for the non-grouping version.

$$|U(t+1) - U(t)| = \left| \sum_i \pi_i(t+1) \log\frac{\pi_i(t+1)}{1-\pi_i(t+1)} - \sum_i \pi_i(t) \log\frac{\pi_i(t)}{1-\pi_i(t)} \right| \tag{36}$$

$$\leq \sum_i \pi_i(t+1) \left| \log\frac{\pi_i(t+1)}{1-\pi_i(t+1)} - \log\frac{\pi_i(t)}{1-\pi_i(t)} \right| + \sum_i |\pi_i(t+1) - \pi_i(t)| \left| \log\frac{\pi_i(t)}{1-\pi_i(t)} \right| \tag{37}$$

$$\leq \sum_i \pi_i(t+1) \log C_2 + \sum_i \pi_i(t)[1-\pi_i(t)] C_2 \left| \log\frac{\pi_i(t)}{1-\pi_i(t)} \right| \tag{38}$$

$$\leq \log C_2 + M C_2 \equiv B \tag{39}$$

Next, we use this to bound of the grouping average log-likelihood: $|U_\alpha(t+1) - U_\alpha(t)| \leq |U(t+1) - U(t)| \leq B$

From (34), the expected drift of $U_\alpha(t)$ is the group EJS, which is lower bounded by:

$$Group\ EJS \geq \frac{1}{G} D(P_{mix\ of\ Group1} || P_{mix\ of\ non-Group1}) \tag{40}$$

Since we are using posterior matching, which is a random coding scheme. When calculating the lower bound of group EJS, we must take the average over different mapping γ.

$$Group\ EJS\ with\ Posterior\ Matching \geq \sum_\gamma \lambda_\gamma \frac{1}{G} D(P_{mix\ of\ Group1} || P_{mix\ of\ non-Group1}) \tag{41}$$

$$\geq \frac{K-1}{K} \frac{1}{G} D(\frac{1}{\alpha M} P_{x_a} + \frac{\alpha M - 1}{\alpha M} P_{x_b} || P_{x_b}) \equiv K_\alpha \tag{42}$$

where x_a and x_b are the two closet point of the constellation. Now, with this lower bound, we have a submartingale:

$$E[U_\alpha(t+1)|\boldsymbol{\pi}(t)] - U_\alpha(t) \geq K_\alpha \tag{43}$$

Note that the sorting order for $(t+1)$ and t might be different. However, since $E[U_\alpha^{(\text{sort order for t+1})}(t+1)] \geq E[U_\alpha^{(\text{sort order for t})}(t+1)]$, (43) still hold if the sorting order

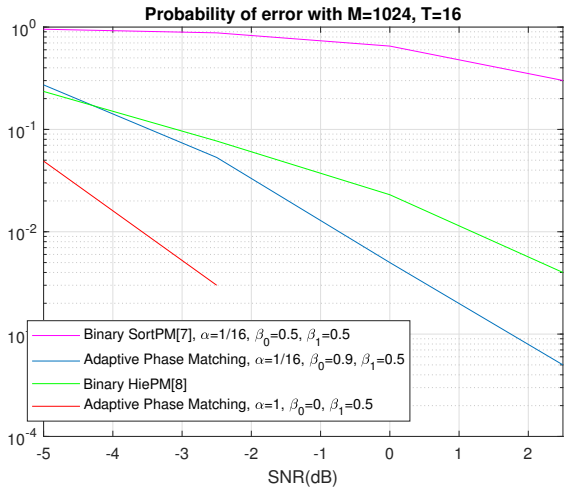

Fig. 7. probability of error for sortPM with and without phase information with $M = 1024, \alpha = 1/16, T = 16$

is different. Next, we define a new bounded submartingale: $V(t) := U_\alpha(t) - tK_\alpha,\ |V(t+1) - V(t)| \leq B + K_\alpha$

Finally, we apply Azuma's inequality to the new submartingale $V(t)$ to bound the probability of the "bad" event:

$$P(E_t) = P\left(\sum_{i=1}^{\alpha M} \pi_i^{\downarrow} < \beta_0\right) \leq P\left(U_\alpha(t) < \log\frac{1-\beta_0}{\beta_0}\right) \tag{44}$$

$$\leq \exp\left(\frac{-[tK_\alpha + U_\alpha(0) - \log\frac{\beta_0}{1-\beta_0}]^2}{2t(B+K_\alpha)^2}\right) \tag{45}$$

V. Simulation Results

In our simulations, the angular space for alignment is quantized into $M = 1024$ candidate AoAs, hence an alignment resolution of 0.18 degrees. Fig. 7 shows the probability of misalignment after a small number $T = 16$ of observations when the design parameters $K = 8$, $\beta_1 = 1/2$ are chosen: Adaptive Phase Matching algorithm is shown to outperforms sortPM without the phase information [7] [15] for all values of α when the raw SNR is above -4dB. In the low SNR regime, on the other hand, the binary HiePM performs better than our proposed three stage algorithm when α is selected to be too small or when the algorithm enters the zoom-in stage late (large β_0). We note that Adaptive Phase Matching outperforms HiePM when $\alpha = 1$ and $\beta_0 = 0$, even thought our theoretical guarantees do not hold.

VI. Conclusion

In this paper, a three stage adaptive searching algorithm is proposed. The algorithm utilizes three target constellation and three stages to provide fast yet highly reliable high-resolution alignment. We analyze the algorithm's superior performance theoretically and numerically. The constellations shown in this paper are chosen to mostly demonstrates the value of phase information in adaptive beam alignment strategies by generalizing the known algorithms in the literature. Note that if there exist some imperfections of the hardware, such as RF impairments, then this will introduce some bias (shifting and rotation) to the constellation points. Although the constellation points might not be perfectly placed on a unit circle, one can still apply the essential ideal of this paper: to separate all the response for all angles according to the posterior probability. The problem of constellation design and optimization remains an important area of future work.

References

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," in IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276-280, March 1986

[2] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 7, pp. 984-995, July 1989

[3] Omid Abari, Haitham Hassanieh, Michael Rodriguez, and Dina Katabi. 2016. Millimeter Wave Communications: From Point-to-Point Links to Agile Network Connections. In Proceedings of the 15th ACM Workshop on Hot Topics in Networks (HotNets '16). Association for Computing Machinery, New York, NY, USA, 169–175.

[4] X. Song, S. Haghighatshoar and G. Caire, "Efficient Beam Alignment for Millimeter Wave Single-Carrier Systems With Hybrid MIMO Transceivers," in IEEE Transactions on Wireless Communications, vol. 18, no. 3, pp. 1518-1533, March 2019.

[5] Han Yan, Benjamin W. Domae, and Danijela Cabric. 2020. MmRAPID: Machine Learning assisted Noncoherent Compressive Millimeter-Wave Beam Alignment. In Proceedings of the 4th ACM Workshop on Millimeter-Wave Networks and Sensing Systems (mmNets'20). Association for Computing Machinery, New York, NY, USA, Article 3, 1–6.

[6] Y. Kaspi, O. Shayevitz, and T. Javidi, "Searching with measurement dependent noise," 2014 IEEE Information Theory Workshop, ITW 2014, pp. 267–271, 2014.

[7] Sung En Chiu and Tara Javidi, "Sequential measurement-dependent noisy search," in 2016 IEEE Information Theory Workshop (ITW), 2016.

[8] Sung-En Chiu, Nancy Ronquillo, and Tara Javidi, "Active Learning and CSI Acquisition for mmWave Initial Alignment," IEEE Journal on Selected Areas in Communications, vol. 37, no. 11, pp. 1–1, 2019.

[9] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," Information Theory, IEEE Transactions on, vol. 57, no. 3, pp. 1186–1222, March 2011.

[10] V. Suresh and D. J. Love, "Single-Bit Millimeter Wave Beam Alignment Using Error Control Sounding Strategies," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 5, pp. 1032-1045, Sept. 2019,

[11] S. Noh, M. D. Zoltowski and D. J. Love, "Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment," in IEEE Transactions on Wireless Communications, vol. 16, no. 9, pp. 5689-5701, Sept. 2017.

[12] A. Alkhateeb, O. El Ayach, G. Leus and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," in IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 831-846, Oct. 2014.

[13] N. Ronquillo, S. -E. Chiu and T. Javidi, "Sequential Learning of CSI for MmWave Initial Alignment," 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2019, pp. 1278-1283.

[14] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic Jensen-Shannon divergence: Applications to variable-length coding," IEEE Transactions on Information Theory, vol. 61, no. 4, pp. 2148–2164, 2015.

[15] S. -E. Chiu and T. Javidi, "Low Complexity Sequential Search With Size-Dependent Measurement Noise," in IEEE Transactions on Information Theory, vol. 67, no. 9, pp. 5731-5748, Sept. 2021

[16] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.