mmRAPID: Machine Learning assisted Noncoherent Compressive Millimeter-Wave Beam Alignment

Han Yan
UCLA
yhaddint@ucla.edu

Benjamin W. Domae
UCLA
bdomae@ucla.edu

Danijela Cabric
UCLA
danijela@ee.ucla.edu

ABSTRACT

Millimeter-wave communication has the potential to deliver orders of magnitude increases in mobile data rates. A key design challenge is to enable rapid beam alignment with phased arrays. Traditional millimeter-wave systems require a high beam alignment overhead, typically an exhaustive beam sweep, to find the beam direction with the highest beamforming gain. Compressive sensing is a promising framework to accelerate beam alignment. However, model mismatch from practical array hardware impairments poses a challenge to its implementation. In this work, we introduce a neural network assisted compressive beam alignment method that uses noncoherent received signal strength measured by a small number of pseudorandom sounding beams to infer the optimal beam steering direction. We experimentally showcase our proposed approach with a 60GHz 36-element phased array in a suburban line-of-sight environment. The results show that our approach achieves post alignment beamforming gain within 1dB margin compared to an exhaustive search with 90.2% overhead reduction. Compared to purely model-based noncoherent compressive beam alignment, our method has 75% overhead reduction.

CCS CONCEPTS

• **Hardware** → **Beamforming**; • **Networks** → *Link-layer protocols*; *Physical links.*

KEYWORDS

beam training, beam alignment, neural networks, compressive sensing, phased array, noncoherent measurement, IEEE 802.11ad/ay

ACM Reference Format:
Han Yan, Benjamin W. Domae, and Danijela Cabric. 2020. mmRAPID: Machine Learning assisted Noncoherent Compressive Millimeter-Wave Beam Alignment. In *4th ACM Workshop on Millimeter-Wave Networks and Sensing Systems (mmNets'20), September 25, 2020, London, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3412060.3418432

1 INTRODUCTION

Millimeter-wave (mmW) communication is a promising technology for future wireless networks, including 5G New Radio and 60 GHz Wi-Fi. Due to abundant spectrum, mmW networks are expected to support ultra-fast data rates. As shown in both theory and prototypes, mmW systems require beamforming (BF) with large antenna arrays and narrow beams at both the transmitter (Tx) and receiver (Rx) to combat severe propagation loss. Before data communication, directional beams must probe the channel to select a beam pair with adequate BF gain. This procedure is referred as beam alignment[1]. Existing mmW systems use analog phased arrays with beam sweeping, an exhaustive search approach, for beam alignment. However, this method introduces high communication overhead. Further, the required number of channel measurements linearly scales with number of antenna elements, which is expected to increase with the evolution of mmW networks. In this work, we present mmRAPID, mmW Random Antenna weight vector based Path Identification without Dictionary. mmRAPID is a novel beam alignment method based on compressive sensing (CS) theory that reduces the number of channel probings to logarithmically scale with antenna array size. We propose a machine learning approach to address a non-trivial CS dictionary mismatch issue due to array hardware impairments. Our implementation and experiments using a 60 GHz testbed demonstrated near perfect beam alignment with 90.2% overhead reduction as compared to exhaustive beam sweeps. To the authors' best knowledge, this is the first work to experimentally demonstrate machine learning based beam alignment using a 60 GHz phased array testbed and real measurement data.

The rest of the paper is organized as follows. Section 2 surveys mmW fast beam alignment designs and proofs-of-concept. We present the problem statement and the motivation for using machine learning in Section 3; the proposed design in Section 4; and the implementation details with our 60 GHz testbed in Section 5. The experimental results are presented in Section 6. Finally, Section 7 concludes the paper.

Scalars, vectors, and matrices are denoted by non-bold, bold lower-case, and bold upper-case letters, respectively. The (i, j)-th element of $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. Similarly, the i-th element of a set $\mathcal{A}$ is denoted by $[\mathcal{A}]_i$. Transpose and Hermitian transpose are denoted by $(.)^{\mathrm{T}}$ and $(.)^{\mathrm{H}}$ respectively. Inner product between $\mathbf{a}$ and $\mathbf{b}$ is denoted as $\langle \mathbf{a}, \mathbf{b} \rangle$. $|\mathbf{a}|$ returns vector with magnitude of each element of $\mathbf{a}$.

2 RELATED WORKS

Beam alignment for mmW is an active research area. While some approaches focus on hardware innovations, e.g., fully-digital array and frequency domain simultaneous beam sweep facilitated by leaky wave antenna [9] or true-time-delay array [5], others rely on signal processing.

mmNets'20, September 25, 2020, London, United Kingdom

ACM ISBN 978-1-4503-8097-3/20/09.
https://doi.org/10.1145/3412060.3418432

[1] It is also referred as beam training, path identification, and path discovery.

Model-based signal processing algorithms for beam alignment mainly rely on the sparsity of mmW channels and a knowledge of the phased array response. State of the art approaches from this class of algorithms, namely hierarchical beam alignment and compressive sensing based beam alignment, have overheads that logarithmically scale with array size. The former uses sounding beams that adapt with previous measurement, bisecting the beam width to reduce the search space [7]. The latter is based on either CS, i.e., with coherent complex sample measurements, or compressive sensing phase retrieval (CPR), i.e., with noncoherent received signal strength (RSS) measurements [12, 15, 21]. Alternatively, UbiG [10] is a model-based algorithm that requires only a constant sounding overhead regardless of array size. UbiG relies on perfect knowledge of all beam patterns to solve for the angle of arrival (AoA)s with a non-convex optimization problem. For all of these methods, the model mismatch due to channel, antenna array and radio hardware impairments introduces non-trivial challenges. Additionally, adaptive codebooks required for hierarchical beam alignment and the genetic optimization proposed for UbiG could add significant complexity or computational overhead to the beam management integrated circuits.

Data-driven signal processing can, from extensive training data, learn to infer the best beam using various low-overhead, in-band measurements and/or out-of-band information. In-band measurements include channel impulse response estimated by omni- directionally received pilots [1] and a proportion of exhaustive beam search results [6, 19]. Out-of-band information includes the terminal's location [13, 19]. To date, these works either use a statistic channel model [6] or ray tracing simulations [13, 19] to generate data.

With the increased availability of mmW testbeds, there are many proofs-of-concept. Work in [3] reports a chip-level demonstration of CS based beam alignment with a channel emulator. [15, 23] showcase fast alignment by solving CPR problems, while [2, 11] design and demonstrate fast beam alignment using multi-lobe sounding beams and combinatorics inspired algorithms. Work in [8] reports experimental work that effectively reduces overhead when more than one spatial stream is used in a hybrid array. Finally, some prototypes also rely on the side information, e.g., sub-6 GHz [14, 18] and visible light [10] measurements, for mmW beam alignment.

This work improves on the existing literature by proposing a novel beam alignment algorithm that maintains low sounding overhead even with model mismatch. Unlike prior algorithms, mmRAPID avoids performance degradation with realistic hardware. To the authors' best knowledge, although traditional compressive beam alignment algorithms have been tested experimentally, this work is also the first to test a machine learning based beam alignment algorithm with experimental data collected from mmW radios.

3 NONCOHERENT COMPRESSIVE BEAM ALIGNMENT

In this section, we start with the mathematical model and problem of noncoherent compressive beam alignment. As a reference, we also describe the state-of-the-art model-based solutions and their limitations.

3.1 System model and problem statement

We consider mmW communication between an access point (AP) Tx and a mobile station (MS) Rx. The AP and MS are each equipped with an analog linear array with N_{T} and N_{R} elements. The channel follows an L-path geometric model $\mathbf{H} = \sum_{l=1}^{L} g_l \mathbf{a}_{\mathrm{R}}(\phi_l)\mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_l)$, where $\mathbf{a}_{\mathrm{T}}(\theta) \in \mathbb{C}^{N_{\mathrm{T}}}$, $\mathbf{a}_{\mathrm{R}}(\phi) \in \mathbb{C}^{N_{\mathrm{R}}}$, and $g_l \in \mathbb{C}$ are the array responses in AP and MS and gain of the l-th path, respectively. Array responses are defined by their n-th element, i.e., $[\mathbf{a}_{\mathrm{R}}(\phi)]_n = \exp(j2\pi(n-1)d/\lambda \sin(\phi))$ and $[\mathbf{a}_{\mathrm{T}}(\theta)]_n = \exp(j2\pi(n-1)d/\lambda \sin(\theta))$, where d, λ, ϕ and θ are the element spacing, carrier wavelength, AoA and angle of departure, respectively. We focus on the MS Rx side by assuming the the AP Tx antenna weight vector (AWV) $\mathbf{v} = \mathbf{a}_{\mathrm{T}}(\theta_1)/\sqrt{N_{\mathrm{R}}}$ is pre-designed. Thus, the channel model in the rest of the paper is

$$\mathbf{h} = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\mathrm{R}}(\phi_l) \approx \alpha_1 \mathbf{a}_{\mathrm{R}}(\phi_1) \triangleq \alpha \mathbf{a}_{\mathrm{R}}(\phi^{\star}), \tag{1}$$

where $\alpha_l = \alpha_l \mathbf{a}_{\mathrm{T}}^{\mathrm{H}}(\theta_l)\mathbf{v}$ is the post-Tx-beam channel gain. The approximation in (1) is from the selection of the Tx beam, which results in $|\alpha_1| \gg |\alpha_l|, l > 1$. We define $\phi^{\star}$ as the true AoA of the channel. When the Rx uses AWV $\mathbf{w}$, the received symbol is

$$y = \mathbf{w}^{\mathrm{H}}\mathbf{h}s + n, \tag{2}$$

where s is the Tx symbol and n is the post-combining thermal noise which is modeled as additive white Gaussian noise with variance σ_{n}^2. Without loss of generality, we let $s = 1$ and define signal-to-noise ratio (SNR) as $\mathrm{SNR} = |\alpha|^2/\sigma_{\mathrm{n}}^2$.

We consider a codebook based communication protocol which consists of two phases: a beam alignment phase and a data communication phase. The channel is unknown to the AP Rx but can be assumed invariant between these two phases. During beam alignment, the MS Rx uses a sounding codebook, $\mathcal{W}_{\mathrm{S}}$ (with $|\mathcal{W}_{\mathrm{S}}| = M$ codewords), to probe the channel. The associated measurements are processed to select the best beam from a fixed directional codebook $\mathcal{W}_{\mathrm{D}}$ (with $|\mathcal{W}_{\mathrm{D}}| = K$ codewords), which is then used in the data communication. Each codeword of directional codebook is a steering vector, i.e., $[\mathcal{W}_{\mathrm{D}}]_k = \mathbf{a}_{\mathrm{R}}(\theta_k)/\sqrt{N_{\mathrm{R}}}$, and these directions $\{\theta_k\}_{k=1}^{K}$ cover an angular region of interest.

Three additional assumptions are relevant to our implementation. Firstly, the sounding codebook $\mathcal{W}_{\mathrm{S}}$ is loaded into hardware in advance and each codeword is applied in a sequential manner. Adaption that uses on-the-fly measurements to change either the codebook or the codeword selection order, e.g., a hierarchy search, is not desired. In fact, we focus on pseudo-random sounding codebooks $\mathcal{W}_{\mathrm{S}}$, a well adopted design from compressive sensing literature when the Rx does not have prior knowledge of the channel [12, 15, 16, 21]. Specifically, the magnitude of each AWV in $\mathcal{W}_{\mathrm{S}}$ is $1/\sqrt{N_{\mathrm{R}}}$ and the phase is randomly picked from the set $\{0, \pi/2, \pi, 3\pi/2\}$. These are referred to as pseudorandom noise (PN) AWV or beams in the remainder of the paper. Secondly, the received symbols in (2) are not directly observable. Instead, each channel measurement is noncoherently taken from the preamble, a sequence of pilot symbols, in the form of RSS. Lastly, the phased array is non-ideal and has realistic hardware impairments. An optimistic assumption is to model these impairments as gain and phase offset, i.e., an unknown multiplicative error $\mathbf{e} \in \mathbb{C}^{N_{\mathrm{R}}}$ independent

of the codewords [23]. With the above assumptions, the M channel probings give RSS $\mathbf{p} = [p_1, \cdots, p_M]^{\mathrm{T}}$ where the m-th probing is

$$p_m = |\tilde{\mathbf{w}}_m^{\mathrm{H}} \mathbf{h}| + n_m. \tag{3}$$

In the equation, $\tilde{\mathbf{w}}_m = \mathrm{diag}(\mathbf{e})[\mathcal{W}_{\mathrm{S}}]_m$ is the receiver combiner with hardware impairment, and n_m is the error in RSS measurement. To that end, the compressive noncoherent beam alignment problem is:

Problem: Design a signal processing algorithm that uses the noncoherent measurements $\mathbf{p}$ from (3) to infer the best directional beam for data communication, i.e., $\hat{\mathbf{w}} \in \mathcal{W}_{\mathrm{D}}$.

The performance metrics are the required number of measurements M and the post-alignment BF gain, i.e., normalized gain in data communication phase $G = |\mathbf{h}^{\mathrm{H}} \mathrm{diag}(\mathbf{e}) \hat{\mathbf{w}}|^2 / \|\mathbf{h}\|^2$. Note that an exhaustive search uses the same codebook for both alignment and communication, i.e., $\mathcal{W}_{\mathrm{S}} = \mathcal{W}_{\mathrm{D}}$. It is straightforward to find the optimal codeword $\mathbf{w}^\star = \max_{\mathbf{w} \in \mathcal{W}_{\mathrm{D}}} G$ with overhead cost $M = K$. The goal is to reduce M while introducing marginal impact to post-alignment BF gain as compared to an exhaustive search[2], e.g., <2dB loss. we further defined overhead reduction ratio as $(K - M)/K$.

3.2 Model based solution and its limitation

The beam alignment with (3) can be formulated as a CPR problem when the error $\mathbf{e}$ is assumed to be negligible, i.e.,

$$\mathbf{p} = |\mathbf{W}^{\mathrm{H}} \mathbf{h}| + \mathbf{n} = |\mathbf{W}^{\mathrm{H}} \mathbf{A}_{\mathrm{R}} \mathbf{g}| + \mathbf{n} \triangleq |\mathbf{\Psi} \mathbf{g}| + \mathbf{n}. \tag{4}$$

In the above equation, $[\mathbf{A}_{\mathrm{R}}]_k = \mathbf{a}_{\mathrm{R}}(\theta_k)$, θ_k are the steering directions in the discrete Fourier transform (DFT) codebook (the AoA hypothesis), and $\mathbf{g} \in \mathbb{C}^K$ is a sparse vector with all-zero elements except the $k^\star$-th being α (i.e. associated with true AoA $k^\star = \{k | \theta_k = \phi^\star\}$). The error vector is $\mathbf{n} = [n_1, \cdots, n_M]^{\mathrm{T}}$. The sensing matrix $\mathbf{W}$ is from error-free sounding AWVs and defined as $[\mathbf{W}]_m \triangleq [\mathcal{W}_{\mathrm{S}}]_m$. The solution to a general CPR model is guaranteed with an adequate number of measurements M, which linearly scales with the sparsity level, i.e., non-zero elements in $\mathbf{g}$, and logarithmically scales with K [17]. Solutions to a general CPR can use convex optimizations [15, 23] or approximate massage passing [17]. Solving CPR in this work, or finding the sparsity level 1 vector $\mathbf{g}$ from $\mathbf{p}$, directly leads to a solution of beam alignment since the non-zero element, say $\hat{k}$-th, can simply be used to select the best beam from the DFT codebook $\hat{\mathbf{w}} = [\mathcal{W}_{\mathrm{D}}]_{\hat{k}}$. In this special case, the heuristic approach of received signal strength matching pursuit (RSS-MP) [16] also applies, where $\hat{k} = \arg\max_k \langle \mathbf{p}, |[\mathbf{\Psi}]_k| \rangle / \|[\mathbf{\Psi}]_k\|$.

The key concern in existing CPR solutions is the required knowledge of dictionary $\mathbf{\Psi}$ (aka magnitude and phase response of AWVs and beam patterns), in (4). With hardware impairments, the sensing matrix $\mathbf{W}$ in (4) is composed of distorted sounding AWVs $[\mathbf{W}]_m = \mathrm{diag}(\mathbf{e})\mathbf{w}_m$. This can be problematic in a practical radio for three reasons. Firstly, in the production of radio hardware, the error $\mathbf{e}$ are due to the combined effects of a systematic offset among all devices and a device dependent random offset [20]. To date, it is generally cost-effective to only calibrate and compensate the common offset. Over-the-air calibration of device-dependent offsets is prohibitively expensive and time consuming. Secondly, the mainlobe of DFT pencil beam is not sensitive to array offset [20]. Although sidelobes are more vulnerable to distortion, they are not directly used in mmW systems during beam sweeping or beam steering. Thus, leaving a device-dependent array offset can be acceptable. Lastly, the beam patterns of PN AWVs are sensitive to array offset, as will be shown in Section 6. Thus, CPR based algorithms are likely to experience model mismatch and degraded performance.

[2]With increased codebook size K, beam steering with AWV $\mathbf{w}^\star$ is asymptotically the same as the steering towards ground truth AoA $\phi_\star$. Thus, we do not directly compare with the latter in this work.

4 MMRAPID DESIGN

To address the model mismatch in solving CPR, we propose a data driven approach for beam alignment.

4.1 Main idea of mmRAPID

The key insight of our approach is that, although analytically solving the noncoherent beam alignment problem using model (4) is difficult, its solution can be easily found by an exhaustive beam sweep. Therefore, we can resort to a data driven approach to learn how to solve the CPR problem (4) with unknown deterministic offset. The proposed system contains two stages, each covering a much longer time scale than the beam alignment or communication phases.

We refer the first stage as the *learning stage*, where the radio uses a concatenated codebook $\mathcal{W} = \mathcal{W}_{\mathrm{D}} \cup \mathcal{W}_{\mathrm{S}}$ for multiple beam alignment phases. Specifically, the sounding results from exhaustive search $\mathcal{W}_{\mathrm{D}}$ provide the solution to the beam alignment problem; the so called *labels* in machine learning terminology. The sounding results from $\mathcal{W}_{\mathrm{S}}$ are treated as the *features*, whose statistical relationship with the *labels* can be extracted by machine learning tools, e.g., neural network (NN) or support vector machine. Admittedly, the beam alignment overhead in this stage is $K + M$, even higher than the overhead K from an exhaustive search. The beam alignment *features* and *labels* must be collected in various environments to reliably generalize their relationship. In a practical system, this would arise from randomness in physical position and orientation of the MS, e.g., a phone held by a human with different posture in different places. In fact, the learning stage can be completely ambient and does not require dedicated interaction from the user [1].

We refer the second stage as the *operation stage*, where MS only uses codebook $\mathcal{W}_{\mathrm{S}}$ for beam alignment, compressing the overhead. The algorithm then only uses the *feature* to predict the *label*, i.e., the best beam $\mathbf{w}^\star$ in data communication phase.

4.2 Neural network design

In this work, we designed a dense NN to predict the optimal DFT beam for a given unknown channel $\mathbf{h}$, i.e., *label*, based on PN beam RSS measurements, i.e., *feature*. The network used 3 fully connected (FC) layers, each using rectified linear unit (ReLU) activation functions. For all tested values of M, we used the same network architecture, with 64, 128, and K units in the first through third FC layers respectively. This relatively shallow NN architecture was designed through testing with extra simulation data. We experimented with deeper networks, but found that only 3 layers were required to achieve maximum performance and thus capture

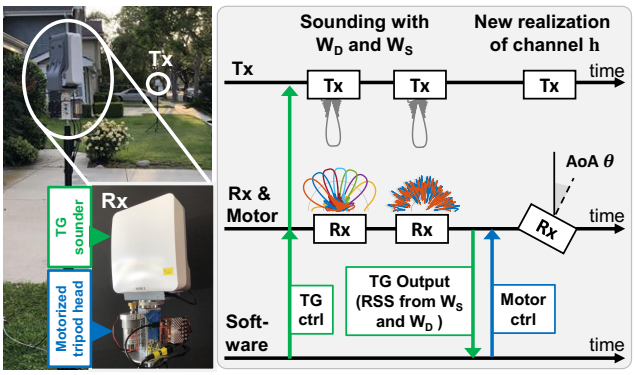


Figure 1: Overview of the testbed, experiment environment, and data capture procedure.

the nonlinear statistics. Additionally, the shallow NN design reduces the computational complexity of the algorithm and the beam management integrated circuits.

Input RSS data was expressed in linear scale and normalized by the maximum value. These feature transformations limit the data to the range $[0, 1]$, prevent activation function saturation, and improve the learning performance. Batch normalization layers were also used just before the ReLU activations in first and second FC layers as feature regularization to improve training efficiency.

Our design used a sparse categorical cross entropy loss function to produce our classification results over the K possible DFT beam physical angle labels. For training, we used the RMSprop optimizer. The network architecture was implemented and trained in Keras/Tensorflow. The total number of trainable parameters in this network depends on the input feature dimension (M) and the label dimension (K): $64M + 129K + 8768$.

5 IMPLEMENTATION IN MMW TESTBED

This section starts with a description of the testbed, followed by the NN based beam alignment implementation.

Our testbed is the Facebook Terragraph (TG) channel sounder, a pair of TG nodes customized for measurements of 60 GHz channels [4]. Each TG node has a 36 by 8 planar phased array. When using pencil beams, the narrowest one-sided 3 dB beamwidth is 1.4° in the azimuth plane. Readers are referred to [4] for more details of the testbed.

The testbed has an application programmable interface (API) that allows a host computer to customize AWVs when transmitting or receiving IEEE802.11ad packets. The API also provides measurements from the received preamble, e.g., received signal strength indicator, short training field (STF) specified SNR, and long-training field specific channel impulse response estimation. Note that although multiple automatic gain control (AGC) amplifiers are involved in preamble measurements, a look-up table is used to make the impact of AGC transparent, which indicates the fidelity of model (3). The array offset was calibrated and compensated using a golden design instead of device-by-device calibration.

5.1 Data Capture Automation

As mentioned in Section 4, it is desired for the NN to learn from beam alignment data collected in various of locations and physical orientations of radios. To automate this procedure, we used a programmable motor on the receiver side. We designed a 3D printed bracket to attach the receiver to a motorized turntable kit controlled by a motor controller, as shown in Figure 1. We mechanically rotated the receiver between −45° and 45° from the transmitter boresight. The procedure achieved pseudo random realizations of AoA of propagation channel $\mathbf{h}$ using pre-programmed motor positions. Figure 1 demonstrates how the receiver collected different physical AoAs using the automated turntable. Note that the motor was not precisely controlled, nor did it provide the true AoA, unlike a turntable required for chamber calibration of the array. The motor's only purpose was to emulate random physical positions and hold the posture of the MS as described in Section 4.1. No such motor is required when generalizing this approach to actual scenarios.

6 EXPERIMENT AND RESULTS

This section describes the experiment details, followed by experimental results and comparison with state-of-the-art.

6.1 Experiment details

We conducted the experiment in a line-of-sight (LoS), suburban outdoor environment, shown in Figure 1a. The radios were mounted on tripods separated by approximately 14 m (91 dB pathloss). The azimuth AoA of the channel is randomly changed by the motor before each capture. Tx directional beam was pointed towards the Rx at all time.

During each collection period, we collected 2,000 points. Each point consisted of 100 RSS measurements using the sounding codebook $\mathcal{W}$, i.e., $K = 64$ DFT beams between −45° and 45° such that adjacent pencil beams overlap by a half of one-sided 3 dB beamwidth and $M_0 = 36$ PN beams. Each data point spends 9 s, including 7 s when the Rx was static[3] and 2 s for the motor movement to create a new LoS propagation direction. Although we collected data for $M_0 = 36$ PN beams, only the first M beams were used for training and testing with compressive beam alignment algorithms. A total of 3 collection periods from three different days were included in this paper's results, each with a different SNR[4]. We changed the SNR by modifying the transmit effective isotropic radiated power (EIRP) which leads to median PN beam SNRs of 10, 10, and 12 dB. After data collection, *labels* with insufficient training data (at least 20 points per SNR) were eliminated, leaving $K = 51$ remaining labels, whose associated AoA are between −26.4° to 43.6°. Of the data associated with the $K = 51$ DFT beam labels, a total of 3,060 data points were used for training and 1898 points were used for evaluation. All data and code is available at [22]. During evaluation, the DFT sounding results were used as the ground truth to measure beam prediction performance from compressive PN probing.

[3]The latency of testbed API is not optimized. Hence, our goal is to achieve alignment with a compressed number of probings M instead of high speed.
[4]Such SNR is measured by STF which has consistent definition with the SNR for (2). Although SNR in (3) cannot be directly measured, it should be larger than the STF-SNR since RSS measurements average a sequence of samples.

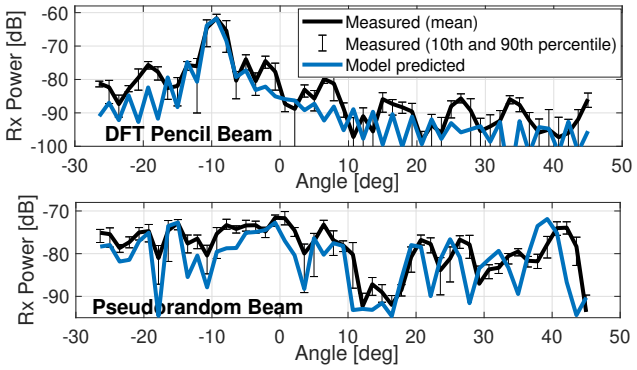


Figure 2: Measured and model predicted beam pattern of a DFT beam $\mathbf{w} \in \mathcal{W}_{\mathbf{D}}$ and a PN beam $\mathbf{w} \in \mathcal{W}_{\mathbf{S}}$.

For fair comparison, the training data was also available to the solution that analytically solves CPR. Note that when using $[\boldsymbol{\Psi}]_k$ in (4) as collected labels of training data, the system can estimate $|\tilde{\mathbf{w}}_m^{\mathrm{H}} \mathbf{a}_{\mathrm{R}}(\theta_k)|$, i.e., the magnitude of dictionary $\boldsymbol{\Psi}_k$. Although such estimates cannot be directly used in CPR as phase information of $\boldsymbol{\Psi}$ are missing, they help the RSS-MP algorithm [16]. Hence, we refer to the vanilla RSS-MP as one that uses only the model predicted dictionary, and dictionary refined RSS-MP as the one enhanced by training data. Also, our best efforts in applying the convex optimization based CPR solution led to unsatisfactory performance, likely due to the imperfect knowledge of the dictionary $\boldsymbol{\Psi}$ when using the PN sounding beam, similar to the finding from [23].

6.2 Experiment results

The captured data allowed us to coarsely evaluate the beam pattern, $|\tilde{\mathbf{w}}^{\mathrm{H}} \mathbf{a}_{\mathrm{R}}(\theta)|^2$ or the magnitude response of AWV at angle θ [5], of the testbed. A comparison between the measured pattern and the model predicted pattern, $|\mathbf{w}^{\mathrm{H}} \mathbf{a}_{\mathrm{R}}(\theta)|^2$, is presented in Figure 2, showing an example DFT pencil beam and an example PN beam[6]. The results verify the arguments in Section 3.2. Although the hardware impairment causes little distortion in the mainlobe of DFT beam, DFT sidelobes and PN beam are susceptible to larger distortion.

Using the NN described in Section 4.2 with the experimental data, we achieved good accuracy for with a compressed numbers of measurements M. Figure 3 (a) shows the test accuracy[7] of the $K = 51$ DFT beams used for $4 \leq M \leq 20$ and different training set sizes. With $M \geq 6$, the test accuracy saturates around 89% for the full training set. Performance does improve with more training data, but NN is already effective with little training data.

The post-alignment BF gain loss (as compared to a full exhaustive search with $K = 51$ measurements) is presented in Figure 3 (b) with a comparison of three algorithms using PN sounding results to predict the best DFT beam for data communication. Vanilla RSS-MP suffered from the mismatched dictionary information and thus had the poor performance. Even with $M = 20$ channel probings, vanilla RSS-MP had more than 2 dB BF gain loss in the 90th percentile. With dictionary estimation in RSS-MP, reasonably good alignment is observed with $M \geq 10$. However, precise alignment (less than 2 dB BF gain in 90 percent of test cases) cannot be achieved until $M = 20$ measurements. The proposed approach provided further savings, requiring only $M = 5$ measurements (90.2% overhead saving) for comparable post-alignment gain.

In Figure 3 (c), we compare the required number of measurements[8] as a function of array size for the RSS-MP and proposed algorithms using experimental and simulation data. The simulation used the same PN sounding AWV realization as in the experiment, but did not include array hardware impairment. The SNR in (3), i.e., ratio between RSS measurements and error variance, is set as 20 dB. We found the following three observations from the results. Firstly, the required overhead of compressive beam alignment scaled logarithmically with array size, an appealing property for future mmW systems. Secondly, the proposed method effectively learned how to solve CPR and provided more accurate beam alignment than the heuristic RSS-MP, even in the simulations without model mismatch. Lastly, the NN had no performance loss in the experimental implementation because the data driven approach is immune to model mismatch due to array imperfection.

7 CONCLUSION AND FUTURE WORKS

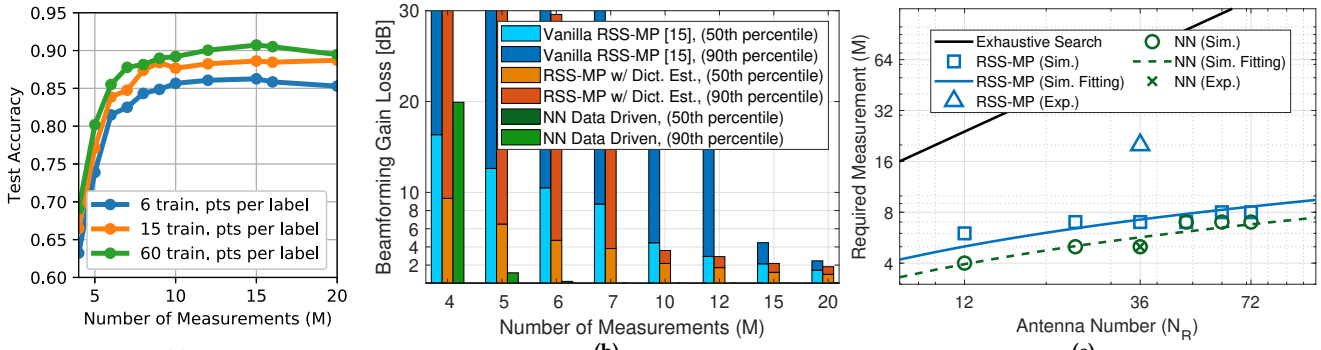
In this work, we presented mmRAPID, a compressive beam alignment scheme that utilizes machine learning to address implementation challenges due to hardware impairments. The results demonstrate that compressive beam alignment can significantly reduce the required number of channel probings. Our implementation on a 60 GHz testbed demonstrates an order of magnitude overhead savings with marginal post-alignment beamforming gain loss, as compared to exhaustive beam sweeps. In the experiment, mmRAPID also outperforms purely model-based compressive methods.

There are still open questions in this area. The approach and results have yet to be generalized to more sophisticated mmW channels, e.g., non-line-of-sight, and the real-time computational cost the algorithm has yet to be thoroughly compared to existing methods. Further, the use of compressed channel probing to predict multiple steering directions in multipath environments have not yet to been explored. Finally, a comparative study of different sounding codebooks, e.g., multi-lobe beams [2, 11], under array impairments and joint design of the codebook and beam alignment algorithm with other machine learning tools are of interest.

8 ACKNOWLEDGMENTS

This work was supported in part by NSF under grant 1718742. This work was also supported in part by the ComSenTer and CONIX Research Centers, two of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. The 60 GHz Terragraph channel sounders were gifts from Telecom Infra Project (TIP).

[5] We experimentally evaluated the coarse beam pattern by categorizing the collected data by AoAs with peak RSS, i.e. the estimated AoA. The RSS measurements in the ith data category then approximate $|\alpha \tilde{\mathbf{w}}^{\mathrm{H}} \mathbf{a}_{\mathrm{R}}(\theta_k)|^2$ and thus the beam pattern at θ_k, assuming the complex path gain α had constant magnitude throughout the experiment.
[6] In the plot, we estimated the magnitude of the complex gain $|\alpha|$ to scale the beam pattern for comparison.
[7] The fraction of DFT beam predictions consistent with the optimal beams from beam sweeps. Incorrect classification may still offer some BF gain.
[8] The minimum M with <2dB BF gain loss compared to a beam sweep in 90% or more of the predictions



(a) (b) (c)

Figure 3: (a) Test accuracy as a function of the number of PN beam measurements M for 3 training set sizes. (b) The post-alignment BF gain loss as function of M. (c) The required M as function of receiver array size N_R.

REFERENCES

[1] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic. 2018. Deep Learning Coordinated Beamforming for Highly-Mobile Millimeter Wave Systems. *IEEE Access* 6 (2018), 37328–37348.

[2] I. Aykin, B. Akgun, and M. Krunz. 2019. Multi-beam Transmissions for Blockage Resilience and Reliability in Millimeter-Wave Systems. *IEEE Journal on Selected Areas in Communications* 37, 12 (2019), 2772–2785.

[3] M. Bajor, T. Haque, G. Han, C. Zhang, J. Wright, and P. R. Kinget. 2019. A Flexible Phased-Array Architecture for Reception and Rapid Direction-of-Arrival Finding Utilizing Pseudo-Random Antenna Weight Modulation and Compressive Sampling. *IEEE Journal of Solid-State Circuits* 54, 5 (2019), 1315–1328.

[4] Lorenzo Bertizzolo, Michele Polese, Leonardo Bonati, Abhimanyu Gosain, Michele Zorzi, and Tommaso Melodia. 2019. MmBAC: Location-Aided MmWave Backhaul Management for UAV-Based Aerial Cells. In *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems* (Los Cabos, Mexico) *(mmNets'19)*. Association for Computing Machinery, New York, NY, USA, 7–12. https://doi.org/10.1145/3349624.3356763

[5] Veljko Boljanovic, Han Yan, Chung-Ching Lin, Soumen Mohapatra, Deukhyoun Heo, Subhanshu Gupta, and Danijela Cabric. 2020. True-Time-Delay Arrays for Fast Beam Training in Wideband Millimeter-Wave Systems. arXiv:2007.08713 [eess.SP]

[6] D. Burghal, N. A. Abbasi, and A. F. Molisch. 2019. A Machine Learning Solution for Beam Tracking in mmWave Systems. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. 173–177.

[7] D. De Donno, J. Palacios, and J. Widmer. 2017. Millimeter-Wave Beam Training Acceleration Through Low-Complexity Hybrid Transceivers. *IEEE Transactions on Wireless Communications* 16, 6 (2017), 3646–3660.

[8] Y. Ghasempour, M. K. Haider, C. Cordeiro, and E. W. Knightly. 2019. Multi-User Multi-Stream mmWave WLANs With Efficient Path Discovery and Beam Steering. *IEEE Journal on Selected Areas in Communications* 37, 12 (2019), 2744–2758.

[9] Yasaman Ghasempour, Rabi Shrestha, Aaron Charous, Edward Knightly, and Daniel M. Mittleman. 2020. Single-shot link discovery for terahertz wireless networks. *Nature Communications* 11, 1 (2020), 1–6.

[10] Muhammad Kumail Haider, Yasaman Ghasempour, Dimitrios Koutsonikolas, and Edward W. Knightly. 2018. LiSteer: MmWave Beam Acquisition and Steering by Tracking Indicator LEDs on Wireless APs. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking* (New Delhi, India) *(MobiCom '18)*. Association for Computing Machinery, New York, NY, USA, 273–288. https://doi.org/10.1145/3241539.3241542

[11] Haitham Hassanieh, Omid Abari, Michael Rodriguez, Mohammed Abdelghany, Dina Katabi, and Piotr Indyk. 2018. Fast Millimeter Wave Beam Alignment. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication* (Budapest, Hungary) *(SIGCOMM '18)*. Association for Computing Machinery, New York, NY, USA, 432–445. https://doi.org/10.1145/3230543.3230581

[12] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed. 2016. An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems. *IEEE Journal of Selected Topics in Signal Processing* 10, 3 (2016), 436–453.

[13] Y. Heng and J. G. Andrews. 2019. Machine Learning-Assisted Beam Alignment for mmWave Systems. In *2019 IEEE Global Communications Conference (GLOBECOM)*. 1–6.

[14] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer. 2015. Steering with eyes closed: Mm-Wave beam steering without in-band measurement. In *2015 IEEE Conference on Computer Communications (INFOCOM)*. 2416–2424.

[15] M. Rasekh and U. Madhow. 2018. Noncoherent compressive channel estimation for mm-wave massive MIMO. *2018 52nd Asilomar Conference on Signals, Systems, and Computers* (Oct 2018).

[16] Maryam Eslami Rasekh, Zhinus Marzi, Yanzi Zhu, Upamanyu Madhow, and Haitao Zheng. 2017. Noncoherent MmWave Path Tracking. In *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications* (Sonoma, CA, USA) *(HotMobile '17)*. Association for Computing Machinery, New York, NY, USA, 13–18. https://doi.org/10.1145/3032970.3032974

[17] P. Schniter and S. Rangan. 2015. Compressive Phase Retrieval via Generalized Approximate Message Passing. *IEEE Transactions on Signal Processing* 63, 4 (2015), 1043–1055.

[18] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. 2017. WiFi-Assisted 60 GHz Wireless Networks. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking* (Snowbird, Utah, USA) *(MobiCom '17)*. Association for Computing Machinery, New York, NY, USA, 28–41. https://doi.org/10.1145/3117811.3117817

[19] V. Va, T. Shimizu, G. Bansal, and R. W. Heath. 2019. Online Learning for Position-Aided Millimeter Wave Beam Training. *IEEE Access* 7 (2019), 30507–30526.

[20] Y. Wang, R. Wu, J. Pang, D. You, A. A. Fadila, R. Saengchan, X. Fu, D. Matsumoto, T. Nakamura, R. Kubozoe, M. Kawabuchi, B. Liu, H. Zhang, J. Qiu, H. Liu, N. Oshima, K. Motoi, S. Hori, K. Kunihiro, T. Kaneko, A. Shirane, and K. Okada. 2020. A 39-GHz 64-Element Phased-Array Transceiver With Built-In Phase and Amplitude Calibrations for Large-Array 5G NR in 65-nm CMOS. *IEEE Journal of Solid-State Circuits* 55, 5 (2020), 1249–1269.

[21] H. Yan and D. Cabric. 2019. Compressive Initial Access and Beamforming Training for Millimeter-Wave Cellular Systems. *IEEE Journal of Selected Topics in Signal Processing* 13, 5 (2019), 1151–1166.

[22] H. Yan and B. Domae. 2020. mmRAPID Simulation Scripts, Algorithm Code, and Experimental Data. https://github.com/bdomae/mmRAPID

[23] Yi Zhang, Kartik Patel, Sanjay Shakkottai, and Robert W. Heath Jr. 2019. Side-Information-Aided Noncoherent Beam Alignment Design for Millimeter Wave Systems. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing* (Catania, Italy) *(Mobihoc '19)*. Association for Computing Machinery, New York, NY, USA, 341–350. https://doi.org/10.1145/3323679.3326532