2ACE: Spectral Profile-driven Multi-resolutional Compressive Sensing for mmWave Channel Estimation

Yiwen Song[†], Changhan Ge[‡], Lili Qiu[‡,§], Yin Zhang

[†] Carnegie Mellon University
Pittsburg, PA, USA

[‡] The University of Texas at Austin
Austin, TX, USA

[§] Microsoft Research Asia
Shanghai, P.R.China

Abstract

Channel estimation is critical to millimeter-wave capability. Unlike sub-6 GHz WiFi, commercial-off-the-shelf 60 GHz WiFi devices adopt a single RF-chain and can only report the combined received signal strength (RSS) instead of the antenna-wise channel state information (CSI). Therefore, recovering the CSI using a limited number of RSS measurements is important but faces the following challenges: (i) solving a non-convex objective is hard and computationally heavy, (ii) the estimation error is high with insufficient RSS measurements, and (iii) channel fluctuates dynamically. To jointly tackle them, we propose 2ACE, an Accelerated and Accurate Channel Estimation approach using spectral profile-driven multiresolutional compressive sensing. Our thorough experiments show that 2ACE yields 2-8 dB reduction in CSI estimation error, 1-5 dB improvement in beamforming performance, and $5° − 10°$ reduction in angle-of-departure estimation error over the existing schemes.

CCS Concepts

• **Theory of computation → Numeric approximation algorithms**; • **Networks → Network protocol design**.

Keywords

Channel Estimation, Wireless Sensing, Millimeter-wave

1 Introduction

With 10 Gbps-level capability brought by its ultra-wide bandwidth, millimeter-wave (mmWave) technologies have ignited a revolution in wireless communication. In its innovative applications, such as environmental sensing [10, 38, 42], health monitoring [16], and imaging [25, 51, 59], channel estimation always plays an important role. While various tools can fetch channel state information (CSI) from sub-6 GHz WiFi network interface controllers (NICs) [17, 54], most of the commercial-off-the-shelf (COTS) 60 GHz WiFi [1,

2] NICs only report the received signal strength (RSS), *i.e.*, the amplitude of the combined signal at the RF chain [35]. Only a few NICs can provide the channel impulse response (CIR) of the strongest path [4], which is still insufficient as beamforming and sensing need CIR of all paths to achieve the best performance. In general, achieving fast and accurate channel estimation for COTS mmWave WiFi remains an open problem.

mmWave communication adopts large phased arrays at both transmitter (TX) and receiver (RX) sides to achieve MIMO diversity-multiplexing gain and enhance beamforming. The signals are transmitted and received with beamweights, called precoders at TX side or combiners at RX side. Hence, the antenna-wise CSI is a matrix of size $N_t × N_r$, where N_t and N_r are the numbers of TX and RX antennas, respectively. Each entry of the matrix represents the complex channel response between a pair of TX and RX antennas. The channel estimation problem, in this case, can be seen as recovering the CSI matrix, including both the magnitude and phase of each element, based on a set of known beamweights and their corresponding RSS measurements. It is essentially a well-known phase retrieval (PR) problem in compressive sensing (CS) that has profoundly impacted many practical applications in other areas, such as X-ray crystallography [18], diffraction [7] or astronomical imaging [11], optics [50], and microscopy [32].

Challenges: Despite recent progress in CS-based phase retrieval, mmWave CSI estimation is still facing a set of challenges.

- Current CS algorithms require a large number of RSS measurements, which increases quadratically with the number of antennas. For example, PhaseLift [9] takes at least $4N_tN_r$ probes to converge. That is $≥ 4096$ probes for $32 × 32$ CSI matrices given most 60 GHz WiFi devices use 32 antennas (*e.g.*, TP-Link AD7200). Such probing takes $≥ 70.4$ ms [52], which is $≥ 70\%$ of the default beacon interval [1]). It is impractical since it leaves no time for data transmission. Moreover, the time-varying channel may change before measurements are complete.

- Existing CS methods suffer from sharp phase transitions under insufficient measurements, *i.e.*, the error could be arbitrarily high before enough measurements have been made. Since it is hard to ensure enough measurements, it is necessary to gracefully adapt the quality of the recovered CSI to the number of measurements.

- We need to enforce the CSI matrix structure. At a high level, we aim to find a solution that minimizes the fitting error *w.r.t.* the measurement while satisfying the known structure in real-world data. While the alternating direction method of multiplier (ADMM) is promising for solving such constrained optimization problems, it requires us to develop new techniques to achieve fast convergence towards a good solution for non-convex objectives.

- Only accurate channel estimation is useful in practice. We seek an algorithm to provide a confidence indicator quantifying the

accuracy of the solution. We accept the answer only when the indicator shows the quality is good enough. Otherwise, we wait for more measurements. Such an indicator is useful when the CSI fluctuates dynamically. We should automate the change detection and adapt the probing to maintain high estimation accuracy.

Methods: We propose *2ACE*, Accelerated and Accurate Channel Estimation with the following powerful techniques.

First, we propose a novel CS algorithm to exploit the spectral profile of a CSI matrix, *i.e.*, a distribution of the singular values in the matrix, for its reconstruction. Compared with the typical low-rank minimization using the nuclear norm, our approach (i) is more flexible to support matrices of different ranks, and (ii) provides more information to overcome insufficient measurements (*e.g.*, how the top singular values decay rather than the number of non-zeros).

Second, we solve non-convex optimization for CSI estimation and develop an efficient solution by going beyond the ADMM: (i) speeding up the convergence using spectral initialization, (ii) increasing the convexity by adapting the weight of the quadratic penalty in the augmented Lagrangian, and (iii) escaping local minimum by yielding multiple solution candidates in parallel with lifting the problem into a higher dimension. In addition, we set spectral profiles according to the measurement budget.

Third, we develop a simple yet effective consistency check to determine if the inference generalizes beyond the observed measurements. It is crucial to quantify errors, especially in the non-convex optimization regime, due to either insufficient RSS measurements or violation of the theoretical condition. Moreover, it allows us to speed up the estimation by detecting and purging stale information.

Fourth, we design a novel multi-resolutional estimation method. We adapt the resolution based on the measurement budget by grouping multiple physical antennas into one virtual antenna and redefining the CSI accordingly.

We evaluate our approach using the CSI matrices from simulation and testbed experiments. Our results show that 2ACE achieves $\leq$ -6 dB error using $1.5N_tN_r$ measurements. Our multi-resolution CSI estimation automatically adapts the granularity based on the measurement budget and achieves good accuracy over a wide range of measurements. Moreover, we show that beamforming using the recovered CSI by 2ACE yields a 1-5 dB improvement over existing algorithms. We further show that the improved CSI accuracy reduces the sensing error – 2ACE reduces the error of the profile generated by the MUSIC algorithm [41] by 10 dB and angle-of-departure estimation error by 5-10 degrees compared with existing schemes. Our code is publicly available at [43]. While this paper focuses on CSI estimation, 2ACE is potentially applicable to matrices in other domains (*e.g.*, traffic or delay matrices) since their spectral profiles may exhibit similar concentrations of variance.

2 Related work

Sparse channel estimation: There are several methods that exploit the sparsity of mmWave channel [39, 40]. They focus on estimating the angle-of-arrival (AoA) and angle-of-departure (AoD) and steering the beam accordingly. [3, 31] estimate the AoA/AoD by solving the optimization problem based on sparse channel assumption. [37] develops a Bayesian learning algorithm, and [22, 29] introduces algorithms based on sparse channel formulation to estimate the channel AoA/AoD profile. SANBA [56] extends PhaseLift [9]

by incorporating possible dominant path AoA as side information. X-Array [52] measures the AoD by applying a matched filter on sector-level-sweeping (SLS) probes [1]. MUST [45] uses WiFi to aid AoA estimation and facilitate beamforming. However, beamforming towards the AoA/AoD of a single path is *not* optimal, multi-armed beams yield better performance as shown in the existing work [23]. In comparison, 2ACE can estimate higher-rank CSI matrices. This capability brings more benefit in rich multi-path scenarios.

Beam sweeping variants: Beam-sweeping samples a few angles to find AoA/AoD [12, 13, 21, 33, 55]. Agile-Link [19] adopts multi-armed beams to sample multiple spatial directions simultaneously to decrease measurement overhead to $O(L \log N_t)$. UbiG [46] proposes an asymptotic beam alignment algorithm using $4K$ probes to determine the sub-optimal beam. BeamSpy [47] learns the channel impulse response by full-beam scanning, and then prunes the beam search space to recover from blockage, while [51] proposes a beam management scheme to maintain stable connection in high-mobility V2X scenarios. Again, beamforming based on 2ACE-recovered high-resolution CSI outperforms approaches mentioned above.

OFDM based estimation: Existing works [6, 26, 36] focusing on mmWave OFDM channel estimation assume that multi-RF chains and complex received signals are available, hence no phase retrieval is needed. However, the commodity 60 GHz WiFi devices have only a single RF chain, reports only the RSS, and does not support OFDM due to high complexities under large bandwidth (*e.g.*, [57]).

Other work: M-Cube [57] ports commodity 60 GHz WiFi card to high-end software-defined radio (SDR) to measure the CSI. Its estimation granularity is limited by analog phase shifters and NIC sampling rate. E-Mi [53] fetches CSI by turning on antennas one by one and requires the RX to obtain the phase information, which is infeasible on COTS 60 GHz WiFi devices. Deep learning has also been used for CSI estimation [14, 20, 24, 44, 58]. But their robustness remains to be seen due to their environmental dependency. In comparison, 2ACE is based on a solid theoretical foundation of compressive sensing and is more lightweight.

3 Background

3.1 Problem formulation

For each measurement, the transmitter uses precoder **f** and the receiver uses combiner **w**. The received signal y can be derived as:

$$b = \mathbf{w}^\top \mathbf{H} \mathbf{f} \gamma + \sigma \qquad (1)$$

where $\mathbf{f} \in \mathbb{C}^{N_t \times 1}$, $\mathbf{w} \in \mathbb{C}^{N_r \times 1}$, γ is the transmitted symbol, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the CSI matrix, and σ is the normalized noise at the RX side, and $^\top$ denotes the transpose. (1) can be further simplified as

$$b = \mathbf{a}^\top \mathbf{x} + \sigma. \qquad (2)$$

where $\mathbf{a} \in \mathbb{C}^{N_r N_t \times 1}$ is the Kronecker product of **w** and **f**, and **x** denote the vectorized CSI matrix **H**. At this point, the CSI estimation can be formulated as recovering **x** based on the beamforming matrix **a** and RSS measurements $|b|$. Take multiple measurements with different weights, and let $\mathbf{b} = [|b_1|, |b_2|, \cdots]^\top$ and corresponding $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots]^\top$. Hereinafter, we denote $n = N_r N_t$. Instead of finding an **x** that strictly satisfies $\forall \mathbf{a}, |b| = |\mathbf{a}^\top \mathbf{x} + \sigma|$, we optimize

$$\min_{\mathbf{x}} \quad \||\mathbf{A}\mathbf{x}| - \mathbf{b}\|_2^2 \qquad (3)$$

The objective function is non-convex *w.r.t.* **x**. Directly solving (3) (*e.g.*, via gradient descent) often get stuck in sub-optimal solutions.

3.2 Existing work on phase retrieval

Recent progress [9] shows that the PR problem in (3) can be turned into a convex optimization problem under certain conditions as:

$$\min_{\mathbf{x}} \quad \||\mathbf{Ax}|^{\circ 2} - \mathbf{b}^{\circ 2}\|_2 \tag{4}$$

where $^{\circ 2}$ computes the per-element square of a matrix. Denote the subscript $_{[i,j]}$ as the element at the i-th row and j-th column of a matrix, and Let $\mathbf{V}_i = \mathbf{a}_i \mathbf{a}_i^\top$, i.e., $\mathbf{V}_i = \mathbf{A}_{[i,\cdot]}^\top \mathbf{A}_{[i,\cdot]}$, and $\mathbf{W} = \mathbf{xx}^\top$. Then, problem (4) is equivalent to

$$\min_{\mathbf{W}} \quad \sum_{i=1}^{m} \left|\left|\text{Tr}(\mathbf{V}_i^\top \mathbf{W})\right| - \mathbf{b}_{[i]}^2\right|$$

$$\text{subject to} \quad \text{rank}(\mathbf{W}) = 1$$

$$\mathbf{W} \succeq 0 \tag{5}$$

where Tr calculates the trace of a matrix.

The compressive sensing theory suggests that when $\mathbf{A}$ contains a sufficient number of rows and satisfies certain technical conditions (e.g., Restricted Isometry Property (RIP) property [8]), one can omit the rank constraint in (5). Then, (5) can be solved using semidefinite programming (SDP) [9]. This result is called over-parameterization. Specifically, $\mathbf{W} \succeq 0$ and $\text{rank}(\mathbf{W}) = 1$ is equivalent to the existence of column vector $\mathbf{x}$ s.t. $\mathbf{W} = \mathbf{xx}^\top$, while $\mathbf{W} \succeq 0$ alone is equivalent to the existence of a matrix $\mathbf{X}$, s.t. $\mathbf{W} = \mathbf{XX}^\top$. Note that we no longer require $\mathbf{X}$ be a column vector. Therefore, the compressive sensing theory states that under proper conditions, we can solve the original phase retrieval problem in which $\mathbf{x}$ is a column vector by over-parameterizing $\mathbf{x}$ into a matrix $\mathbf{X}$. Specifically, one can show that (5) is equivalent to the following convex objective:

$$\min_{\mathbf{X}} \quad \||\mathbf{AX}| - \mathbf{b}||^2 \tag{6}$$

in which $|\cdot| = \sqrt{\cdot \odot \bar{\cdot}}$, where $\bar{\cdot}$ is the conjugate of a complex matrix, and $\odot$ is element-wise multiplication.

Although solving a non-convex objective with a convex optimization is a major breakthrough, it is challenging to apply such theoretical results in practice for the following reasons:

- SDP-based algorithms suffers from a sharp phase transition under insufficient measurements, i.e., the solution can be arbitrarily bad.
- Since $\mathbf{X}$ and $\mathbf{W}$ in (6) has n unknowns and $m = O(n)$ constraints (empirically $m \approx 3n$), SDP is computationally heavy for large n.
- $\mathbf{A}$ may or may not satisfy the RIP, which can happen when there are constraints on the kind of measurements we conduct (e.g., RSS is too low for some probes). Then, we no longer guarantee that the solution $\mathbf{X}$ is rank-1 or globally optimal. Hence, we have no way of knowing how good the recovered $\mathbf{X}$ is.
- CSI changes fast. We should automatically adapt the probing to maintain estimation accuracy in mobile cases.

4 2ACE algorithm

We first formulate the spectral profile-based CSI estimation problem in Sec. 4.1. We then develop an ADMM-based fast and accurate solution in Sec. 4.2. In Sec. 4.3, we propose several novel enhancements: (i) dynamically updating the weight of the quadratic penalty term in the Augmented Lagrangian, (ii) spectral initialization, and (iii) parallel refinement. In Sec. 4.4, we further develop a confidence indicator to support mobile scenarios. In Sec. 4.5, we design a multi-resolution algorithm to automatically adapt the inference granularity according to the measurement budget.

4.1 Spectral profiles based CSI estimation

Designing an effective regularization term is crucial to compressive sensing. In this section, we utilize a unique property of CSI matrices and design a regularization term, called the **spectral profile**. It refers to the approximation error using the best rank-K approximation w.r.t. the squared Frobenius norm, i.e., how much channel energy is captured by the first K paths. Let $P = \{(r_k, f_k)\}$ denote a spectral profile, where the top r_k singular values account for at least f_k of variance. Based on profile P, we formulate the CSI estimation problem as follows:

$$\min_{\mathbf{X}} \quad \||\mathbf{AX}| - \mathbf{b}\|_2^2 + I(\mathbf{X}, P) \tag{7}$$

The indicator function $I(\mathbf{X}, P) = 0$ if CSI matrix $\mathbf{H}$ induced by X satisfies the spectral profile P. Otherwise $I(\mathbf{X}, P) = \infty$.

4.2 ADMM-based phase retrieval

We use ADMM to solve the constrained optimization problem (7). ADMM is an iterative method that optimizes one variable at a time during each iteration while fixing the remaining variables [5]. To apply ADMM, we first reformulate (7) as:

$$\min_{\mathbf{X}} \quad \frac{1}{2}\||\mathbf{Y}| - \mathbf{b}\|_2^2 + I(\mathbf{Z}, P)$$

$$\text{subject to} \quad \mathbf{AX} = \mathbf{Y} \quad \text{and} \quad \mathbf{X} = \mathbf{Z} \tag{8}$$

We empirically find that omitting the regularization term (e.g., $\frac{v}{2}\|\mathbf{X}\|_2^2$) works best. Then we have the Augmented Lagrangian:

$$L(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{M}, \mathbf{N}, \mu) = \frac{1}{2}\||\mathbf{Y}| - \mathbf{b}\|_2^2 + I(\mathbf{Z}, P) + <\mathbf{M}, \mathbf{AX} - \mathbf{Y}>$$

$$+ <\mathbf{N}, \mathbf{X} - \mathbf{Z}> + \frac{\mu}{2}\|\mathbf{AX} - \mathbf{Y}\|_2^2 + \frac{\mu}{2}\|\mathbf{X} - \mathbf{Z}\|_2^2 \tag{9}$$

where $< \cdot, \cdot >$ denotes the inner product of two vectors. Alternating minimization of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ and subsequent update of $\mathbf{M}$ and $\mathbf{N}$ via gradient descent from step t to $t + 1$ yields the following:

$$\mathbf{X}^{(t+1)} = \arg\min_{\mathbf{X}} L(\mathbf{X}, \mathbf{Y}^{(t)}, \mathbf{Z}^{(t)}, \mathbf{M}^{(t)}, \mathbf{N}^{(t)}, \mu) \tag{10a}$$

$$\mathbf{Y}^{(t+1)} = \arg\min_{\mathbf{Y}} L(\mathbf{X}^{(t+1)}, \mathbf{Y}, \mathbf{M}^{(t)}, \mu) \tag{10b}$$

$$\mathbf{Z}^{(t+1)} = \arg\min_{\mathbf{Z}} L(\mathbf{X}^{(t+1)}, \mathbf{Z}, \mathbf{N}^{(t)}, \mu) \tag{10c}$$

$$\mathbf{M}^{(t+1)} = \mathbf{M}^{(t)} + \mu \cdot (\mathbf{AX}^{(t+1)} - \mathbf{Y}^{(t+1)}) \tag{10d}$$

$$\mathbf{N}^{(t+1)} = \mathbf{N}^{(t)} + \mu \cdot (\mathbf{X}^{(t+1)} - \mathbf{Z}^{(t+1)}) \tag{10e}$$

We then show how to solve (10a) – (10c), respectively.

4.2.1 Find $\mathbf{X}$ that $\min L(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{M}, \mathbf{N}, \mu)$: Simplify (10a) to:

$$\min_{\mathbf{X}} \quad \|\mathbf{AX} - \mathbf{Y} + \frac{\mathbf{M}}{\mu}\|_2^2 + \|\mathbf{X} - \mathbf{Z} + \frac{\mathbf{N}}{\mu}\|_2^2. \tag{11}$$

By setting its derivative to 0, we get the closed form solution:

$$\mathbf{X} = (\mathbf{A}^\top \mathbf{A} + \mathbf{I})^{-1} \left(\mathbf{A}^\top (\mathbf{Y} - \frac{\mathbf{M}}{\mu}) + (\mathbf{Z} - \frac{\mathbf{N}}{\mu}) \right). \tag{12}$$

4.2.2 Find $\mathbf{Y}$ that $\min L(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{M}, \mathbf{N}, \mu)$: Simplify (10b) to:

$$\min_{\mathbf{Y}} \quad \frac{1}{2}\||\mathbf{Y}| - \mathbf{b}\|_2^2 + \frac{\mu}{2}\|\mathbf{Y} - \mathbf{C}\|_2^2 \tag{13}$$

where $\mathbf{C} = \mathbf{AX} + \mathbf{M}/\mu$. Let $\mathbf{Y}_{[i]}$ denote the i-th row of $\mathbf{Y}$, we just need to solve the equivalent problem:

$$\min_{\mathbf{Y}_{[i]}} \quad \frac{1}{2}\left||\mathbf{Y}_{[i]}| - \mathbf{b}_{[i]}\right|^2 + \frac{\mu}{2}\left|\mathbf{Y}_{[i]} - \mathbf{C}_{[i]}\right|^2 \tag{14}$$

Algorithm 1 Rescaling SVD.

1: $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{E})$;
2: **for** $k = 1, 2, \ldots$ **do**
3:    $v = \sum_i s_i^2$;
4:    $v_0 = \sum_{i=1}^{r_k} s_i^2$;
5:    **if** $(v_0 < f_k \times v)$ **then**
6:       // scale singular values so first r_k values capture exactly fraction f_k of
7:       // total variance
8:       $s = \sqrt{(v/v_0 - 1)/(1/f_k - 1)}$
9:       **for** $i = r_k + 1, r_k + 2, \ldots$ **do**
10:          $s_i = s_i \times s$;
11:          $\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$
12:       **end for**
13:    **end if**
14: **end for**

Let $\mathbf{y}_{[i]} = \left| \mathbf{Y}_{[i]} \right|$. For any given $\mathbf{y}_{[i]}$ the objective is minimized when $\mathbf{Y}_{[i]} = \mathbf{y}_{[i]} \frac{\mathbf{C}_{[i]}}{|\mathbf{C}_{[i]}|}$. Let $\mathbf{c}_{[i]} = \left| \mathbf{C}_{[i]} \right|$. Then, (14) becomes:

$$\min_{\mathbf{y}_{[i]}} \quad \frac{1}{2} |\mathbf{y}_{[i]} - \mathbf{b}_{[i]}|^2 + \frac{\mu}{2} |\mathbf{y}_{[i]} - \mathbf{c}_{[i]}|^2 \tag{15}$$

The optimal $\mathbf{y}_{[i]}$ is simply $\mathbf{y}_{[i]} = (\mathbf{b}_{[i]} + \mu\mathbf{c}_{[i]})/(1+\mu)$. Therefore, we first compute $|\mathbf{Y}| = (\mathbf{b} + \mu \cdot |\mathbf{C}|)/(1 + \mu)$, and then scale each row $\mathbf{C}_{[i]}$ to have norm $|\mathbf{Y}_i|$.

4.2.3 Find $\mathbf{Z}$ *that* $\min L(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{M}, \mathbf{N}, \mu)$: Let $\mathbf{E} = \mathbf{X} + \mathbf{N}/\mu$. (10c) is equivalent to

$$\min_{\mathbf{Z}} \quad I(\mathbf{Z}, P) + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{E}\|_2^2 \tag{16}$$

$I(\mathbf{Z}, P)$ is non-differentiable, so instead of solving (16) directly, we enforce the spectral profile $P = (r_k, f_k), r_1 \le r_2 \le .. \le r_k$ by a rescaling algorithm. We first reshape $\mathbf{X}$ to $N_t \times N_r$, and find its singular values $\{s_i\}$ sorted in a decreasing order. In Alg. 1, we rescale s to enforce the spectral profile, where SVD referers to singular value decomposition and the outer loop ensures that the top-r_k singular values account for f_k of energy. Line 5-13 re-scales all entries after the top-r_k elements for profile enforcement.

4.2.4 Convergence Criteria: Following the ADMM practice [5], we use the primal/dual residuals for the convergence test, which represent absolute error and relative change, respectively:

$$r_{\text{prim}} = \sqrt{\|\mathbf{A}\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)}\|_2^2 + \|\mathbf{X}^{(t+1)} - \mathbf{Z}^{(t+1)}\|_2^2} \tag{17a}$$

$$r_{\text{dual}} = \mu\sqrt{\|\mathbf{A}^\top(\mathbf{Y}^{(t+1)} - \mathbf{Y}^t)\|_2^2 + \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^t\|_2^2} \tag{17b}$$

The thresholds are related to the differences between pairs of matrices, which depend on their matrix sizes and magnitudes:

$$\epsilon_{\text{prim}} = \epsilon_{\text{abs}}\sqrt{(m + n) \times c} \tag{18a}$$

$$+\epsilon_{\text{rel}}\sqrt{\max(\|\mathbf{A}\mathbf{X}^{(t+1)}\|_2^2, \|\mathbf{Y}^{(t+1)}\|_2^2) + \max(\|\mathbf{X}^{(t+1)}\|_2^2, \|\mathbf{Z}^{(t+1)}\|_2^2)}$$

$$\epsilon_{\text{dual}} = \epsilon_{\text{abs}}\sqrt{2nc} + \epsilon_{\text{rel}}\sqrt{\|\mathbf{A}^\top\mathbf{Y}^{(t+1)}\|_2^2 + \|\mathbf{Z}^{(t+1)}\|_2^2} \tag{18b}$$

where c is the number of column of $\mathbf{X}$, and $\epsilon_{\text{abs}}, \epsilon_{\text{rel}}$ are two hyper-parameters. We stop if $(r_{\text{prim}} \le \epsilon_{\text{prim}}) \wedge (r_{\text{dual}} \le \epsilon_{\text{dual}})$.

4.3 Enhancements

We provide several enhancements to speed up convergence and improve inference accuracy: (i) dynamic update of μ to increase convexity, (ii) spectral initialization of $\mathbf{X}^{(0)}$, (iii) parallel refinement, and (iv) selecting and supporting appropriate spectral profiles.

Algorithm 2 Adaptation of μ

1: **if** $r_{\text{comb}}^{(t+1)} > 0.8 r_{\text{comb}}^{(t)}$ **then**
2:    $\mu^{(t+1)} = 1.03\mu^{(t)}$
3: **else**
4:    $\mu^{(t+1)} = \mu^{(t)}$
5: **end if**

4.3.1 Dynamic Update of μ: The choice of μ significantly impacts the convergence of ADMM. If the optimization is making sufficient progress, we keep the current μ. But if it becomes stagnant, we increase the weight of the penalty term in the augmented Lagrangian to increase the overall convexity. To determine the optimization progress, we leverage the combined residual r_{comb}, which decreases by the ADMM procedure [15]. Evidence shows that r_{comb} also decreases in a non-convex problem [34]. r_{comb} is as follows:

$$r_{\text{comb}}^{(t+1)} = \mu r_{\text{prim}}^2 + \mu(\|\mathbf{Y}^{(t+1)} - \mathbf{Y}^{(t)}\|_2^2 + \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_2^2) \tag{19}$$

where the first term is the primal residual, and the second term relates to the dual residual in (17b) except that it does not multiply by $A^\top$. In Alg. 2, we initialize $\mu = 0.001$ and increase μ dynamically.

4.3.2 Parallel Refinement: In the original problem (8), $\mathbf{X}$ is a $n \times 1$ vector. To further improve the solution, we lift $\mathbf{X}$ to a higher dimension as an $n \times c$ matrix. The benefit of parallel solving multiple candidates for $\mathbf{X}$ is not only getting a better solution, but also decreases computation time by using efficient matrix operations instead of solving (8) for c times. The rationale is that the convex solutions are better connected in high dimension space where one can more easily escape from an inferior local optimal.

Parallel formulation: We formulate the following problem:

$$\min_{\mathbf{X}} \quad \||\mathbf{A}\mathbf{X}| - \mathbf{B}\|_2^2 + I(\mathbf{X}, P) \tag{20}$$

where $\mathbf{B}$ is formed by replicating $\mathbf{b}$ by c times. To solve (20), we simply adapt the solution of $\mathbf{X}$ from (12) and $\mathbf{Z}$ from Alg. 1, and solve $\mathbf{Y}$ element-wise as dipicted in (13) to (15).

4.3.3 Spectral Initialization of $\mathbf{X}$: We need to find a good initial solution $\mathbf{X}^{(0)}$ for $\mathbf{X}$. One approach is spectral initialization, obtained by computing the leading c eigenvectors [30], to support parallel refinement. Recall (5), we can get an estimate of $\mathbf{W} = \mathbf{X}\mathbf{X}^\top$ using the following simple weighted average:

$$\mathbf{W}_{\text{est}} = \frac{\sum_i \text{sim}(\mathbf{W}, \mathbf{V}_i) \cdot \mathbf{V}_i}{\sum_i \text{sim}(\mathbf{W}, \mathbf{V}_i)} = \frac{\sum_i (\mathbf{b}_{[i]}^2/\|\mathbf{a}_i\|_2^2) \cdot \mathbf{V}_i}{\sum_i (\mathbf{b}_{[i]}^2/\|\mathbf{a}_i\|_2^2)} = \frac{\sum_i (\mathbf{a}_i^\top\mathbf{a}_i) \cdot (\mathbf{b}_{[i]}^2/\|\mathbf{a}_i\|_2^2)}{\sum_i (\mathbf{b}_{[i]}^2/\|\mathbf{a}_i\|_2^2)}$$

Intuitively, we project $\mathbf{W}$ to the space defined by $\mathbf{V}_i$ and let the cosine similarity between $\mathbf{W}$ and $\mathbf{V}_i$ define the weight in each projection, which is computed as:

$$\text{sim}(\mathbf{W}, \mathbf{V}_i) = \frac{<\mathbf{W}, \mathbf{V}_i>}{\sqrt{<\mathbf{W}, \mathbf{W}> \cdot <\mathbf{V}_i, \mathbf{V}_i>}} \tag{21}$$

Note that $\mathbf{W}_{\text{est}}$ can be computed by first computing a scaled matrix $\mathcal{A}$ with $\mathcal{A}_{[i]} = \mathbf{A}_{[i]}(\mathbf{b}_{[i]}/\|\mathbf{a}_i\|_2)$. Then we have $\mathbf{W}_{\text{est}} = \mathcal{A}^\top\mathcal{A}/\sqrt{\sum_i (\mathbf{b}_{[i]}^2/\|\mathbf{a}_i\|_2^2)}$. Once we have $\mathbf{W}_{\text{est}}$, we then obtain $\mathbf{X}^{(0)}$ efficiently via SVD as the best rank-c approximation to $\mathbf{W}_{\text{est}}$. Then, we use Alg. 3 to solve (20).

(a) Synthetic traces.   (b) Ray-tracing traces.   (c) Testbed traces [35]
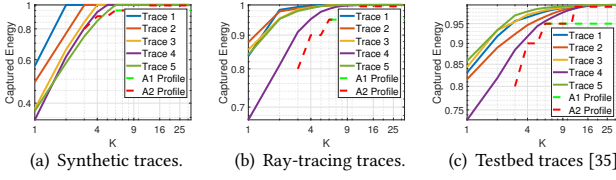
Figure 1: Spectral profile of different 32×32 traces

Algorithm 3 Algorithm for solving (20)

1: $t = 0$
2: $\mu = 0.001$
3: initialize $\mathbf{X}^{(0)}$ via spectral initialization // Sec. 4.3.3
4: initialize: $\mathbf{M}^{(0)} = 0, \mathbf{N}^{(0)} = 0$
5: $\mathbf{Y}^{(0)} = \arg\min_{\mathbf{Y}}(\mathbf{X}^{(0)}, \mathbf{M}^{(0)}, \mu)$;
6: $\mathbf{Z}^{(0)} = \arg\min_{\mathbf{Z}}(\mathbf{X}^{(0)}, \mathbf{N}^{(0)}, \mu)$;
7: **while** $t < $ maxiter **do**
8:     $\mathbf{X}^{(t+1)} = \arg\min_{\mathbf{Z}} L(\mathbf{Y}^{(t)}, \mathbf{Z}^{(t)}, \mathbf{M}^{(t)}, \mathbf{N}^{(t)}, \mu)$ // Sec. 4.2.1
9:     $\mathbf{Y}_{t+1} = \arg\min_{\mathbf{Y}} L(\mathbf{X}^{(t+1)}, \mathbf{M}^{(t)}, \mu)$ // Sec. 4.2.2
10:    $\mathbf{Z}_{t+1} = \arg\min_{\mathbf{Z}} L(\mathbf{X}^{(t+1)}, \mathbf{N}^{(t)}, \mu)$ // Sec. 4.2.3
11:    $\mathbf{M}^{(t+1)} = \mathbf{M}^{(t)} + \mu(\mathbf{A}\mathbf{X}^{(t+1)} - \mathbf{Y}^{(t+1)})$
12:    $\mathbf{N}^{(t+1)} = \mathbf{N}^{(t)} + \mu(\mathbf{X}^{(t+1)} - \mathbf{Z}^{(t+1)})$
13:    **if** convergence is reached // Sec. 4.2.4 **then**
14:       break
15:    **else**
16:       $\mu = \text{Adjust}(\mu)$ // Sec. 4.3.1
17:       $t = t + 1$
18:    **end if**
19: **end while**

4.3.4 Choice of Spectral Profiles :

We consider the following three profiles of selection:

- A0: {} We do not exploit any low-rank property.
- A1: $r_1 = \lceil \sqrt{\min(N_r, N_t)} \rceil$ singular values account for $f_1 = 0.95$ of variance.
- A2: Let $s = \min(N_r, N_t)$. r_i singular values account for f_i of variance, where $r_1 = \lceil 0.5\sqrt{s} \rceil$ for $f_1 = 0.75$, $r_2 = \lceil 0.7\sqrt{s} \rceil$ for $f_2 = 0.85$, $r_3 = \lceil \sqrt{s} \rceil$ for $f_3 = 0.95$, and $r_4 = \lceil 2\sqrt{s} \rceil$ for $f_4 = 0.995$.

We use CSI matrices from simulation, ray tracing, and testbed (inferred by ACO [35]). While imperfect, ACO estimations are not biased since they do not make any rank assumption about the CSI. Fig. 1 shows that the ranks of the CSI matrices are not low as assumed by sparse channel estimation. It also shows that all traces fit the A1/A2 profiles, indicating that we can leverage these spectral profiles for inference by adding a constraint that the estimated CSI should have the top-K singular values accounting for at least these amounts of variances.

Meanwhile, as shown in Alg. 4, when measurements are sufficient, we solve the unknowns without using spectral profiles (A0). Otherwise, we estimate the first singular vector since it contributes the most (A1). Otherwise, we use the A2 profile.

4.4 Confidence indicator

We provide a confidence indicator for our algorithm, where we partition rows of the sensing matrix $\mathbf{A}$ and measurements $\mathbf{b}$ into a training set $(\mathbf{A}_1, \mathbf{b}_1)$ and a testing set $(\mathbf{A}_2, \mathbf{b}_2)$. We first use the training set to estimate the channel $\mathbf{x}$, and then use the testing set to assess how well the inferred amplitude $|\mathbf{A}_2 x|$ matches $\mathbf{b}_2$. We set training-testing split ratio as 19:1. The partition can be

---

Algorithm 4 2ACE Algorithm to incorporate dynamic profile

1: **if** $m >= 3n$ **then**
2:    // no need to use spectral profile w/ enough constraints
3:    $P = \{\}$
4: **else if** $m < n$ **then**
5:    // focus on estimating 1st singular vector w/ too few constraints
6:    $P = \{(r_1, 0.95)\}$
7: **else**
8:    $P = \{(r_1, f_1), (r_2, f_2), (r_3, f_3), (r_4, f_4)\}$
9: **end if**

---

either random, or based on time at which the measurement is conducted. In this way, we can check whether an old $\mathbf{x}$ matches the new measurements; if not, then we purge the oldest measurements and collect more. The confidence indicator benefits both our and traditional methods by determining whether the solution is good, which is especially important when the analytical conditions (*e.g.*, RIP) do not hold.

To sum up, we perform the multi-stage phase retrieval:

1. Partition $(\mathbf{A}, \mathbf{b})$ into $(\mathbf{A}_1, \mathbf{b}_1)$, $(\mathbf{A}_2, \mathbf{b}_2)$.
2. Use $(\mathbf{A}_1, \mathbf{b}_1)$ to solve (8) using Alg. 3.
3. Orthogonalize and normalize columns of $\mathbf{X}$.
4. Parallel refinement (20): Use $\mathbf{X}$ obtained in Step 2 as the initial $\mathbf{X}$, where we still use $(\mathbf{A}_1, \mathbf{b}_1)$. After convergence, we choose the column in $\mathbf{X}$ that minimizes $\|\|\mathbf{A}\mathbf{x}\| - \mathbf{b}\|_2$, denoted by $\mathbf{x}_1$. We call this step parallel refinement since we explore multiple orthogonal initial vectors $\mathbf{X}_{[j]}$ in parallel.
5. We compute our confidence indicator Q as follow:

$$Q = 1 - \frac{\|\|\mathbf{A}_2\mathbf{x}_1\| - \mathbf{b}_2\|_2}{\|\mathbf{b}_2\|_2} \tag{22}$$

   If Q is good enough, we take it as the final answer. Otherwise, we go to the refinement in the next step.

6. Final refinement. Use $\mathbf{x}_1$ as the initial solution, and solve (8) again but now with all measurements. If the new model is similar to the model trained using $\mathbf{x}_1$ after adding new measurements, we take it since it is likely to be accurate. Otherwise, if the measurements are below $2n$, the error is likely due to insufficient measurements and we wait for more data to train using all data together. If the model does not work well for $\mathbf{x}_2$ even with $> 3n$ measurements, the error is likely due to the CSI change, so we purge the measurements from the earliest probing interval and wait for more data.

4.5 Multi-resolution channel estimation

No matter how good the compressive sensing algorithms are, the estimation accuracy can be arbitrarily low if the measurements is insufficient *w.r.t.* the number of unknowns. In such circumstances, we group multiple physical antennas into one *virtual antenna*, and estimate the channel of these virtual antennas. Then, the resolution of CSI matrix can be manipulated as $N_{vt} \times N_{vr}$ based on the measurement budget, where N_{vt} and N_{vr} are the numbers of *virtual* TX and RX antennas, respectively. We achieve it by setting an equal weight, *i.e.*, the same phase shift, on the physical antennas grouped into the same virtual antenna. We feed the sensing matrix $\mathbf{A}$ of size $m \times N_{vt}N_{vr}$ to the inference algorithm and recover a CSI matrix, where each entry represents the "averaged" channel of physical antennas. As the measurement budget increases, we increase N_{vt} and N_{vr} and recover a more fine-grained CSI.

Algorithm 5 Antenna Grouping Algorithm

1: $\mathcal{V} \leftarrow \{1, 2, \cdots, N\}$
2: **while** $\mathcal{V} \neq \emptyset$ **do**
3:     select R_{ref} from $\mathcal{V}$ arbitrarily
4:     **for** $R_i \in \mathcal{V}$ **do**
5:         Calculate $p_{i,\text{ref}}$ between R_{ref} and a_i
6:     **end for**
7:     put $V_k - 1$ antennas with the least $p_{i,\text{ref}}$ with a_{ref} as one group $\mathcal{V}^k$
8:     $\mathcal{V} \leftarrow \mathcal{V} \setminus \mathcal{V}^k$
9: **end while**

Challenges: The CSI of each virtual antenna is the "averaged" CSI of the physical antennas. To make the averaged channel better approximate the individual physical channel, we should group the physical antennas with similar channels into one virtual antenna. We consider two major factors that construct the channels of physical antennas: (i) phase offset yielded by the AoA ϕ, which equals to $\frac{2\pi d}{\lambda} \cos \phi$ for a ULA, and (ii) each antenna has a unique hardware phase offset (*e.g.*, due to a different length of transmission line [57]).

One way to resolve these phase offsets is compensation. If the total positioning and hardware phase offset is O, we can cancel it by multiplying $\exp(-1jO)$ when generating beamweights. However, while the total offset O can be arbitrary, the compensation provided by 2-bit analog phase shifters on commodity 60 GHz WiFi devices can only be 0, $\frac{\pi}{2}$, π, or $\frac{3\pi}{2}$. Thus, there is a residual offset after compensation, which hurts the performance. Therefore, we aim to group physical antennas into virtual antennas such that the residual phase offsets within a group are minimized. Below, we formulate the antenna grouping problem and describe a greedy heuristic.

Antenna grouping Problem: Given a set of antennas, $\mathcal{V} = \{1, 2, \cdots, N\}$, we equally partition them into K non-overlapping groups $\mathcal{V}^k$ with V_k antennas per group ($V_k = \frac{|\mathcal{V}|}{K} \in \mathbb{N}$) such that their total residue phase offsets inside groups is minimized. Namely, we seek a partition $\mathcal{V}^{1,2,..k}$ to reach $\min_{\mathcal{V}^{1,2,..k}} \sum_{k=1}^{K} \sum_{i,j \in \mathcal{V}^k, i \neq j} p_{i,j}$, where $p_{i,j}$ denotes the minimum phase offsets between two antennas i and j after the compensation.

Antenna grouping algorithm: We develop a greedy antenna grouping algorithm. We first probe a few angles to get a coarse AoA. Then we calculate the sum of the positioning and hardware phase offset and feed it to Alg. 5 to derive the grouping. For a given set of N antennas V_a, we select a reference antenna R_{ref} from V_a, and calculate the phase shift $p_{i,\text{ref}}$ between R_{ref} and the other antennas in V_a. Then we select $V_k - 1$ antennas with the least phase shift from R_{ref}, and add them along with R_{ref} to the group. We remove the antennas in the group from the entire antenna set V_a, and find the next group as before until no antennas are left.

Multi-resolution algorithm: To sum up, we first determine the resolution N_{vt} and N_{vr} based on the measurement budget, since our algorithm can recover the CSI of size $N_{\text{vt}} \times N_{\text{vr}}$ using $\sim 1.5N_{\text{vt}}N_{\text{vr}}$ measurements. Then we apply Alg. 5 to group antennas. Next, we probe random beam weights for virtual antennas. We feed the measured RSS along with the corresponding beamweights to derive the virtual CSI using the Alg. 3. The virtual CSI can be used for beamforming and sensing.

5 Evaluation

5.1 Evaluation methodology

We evaluate using three complementary methods: (i) we use multipath simulation to evaluate the impact of different numbers of
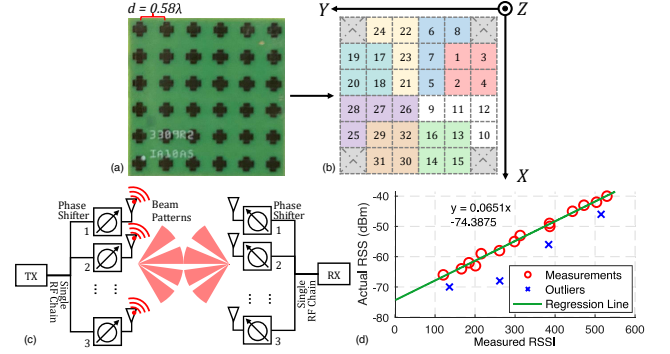
Figure 2: Testbed. (a) UPA, (b) antenna index mapping, (c) hardware schema, (d) RSSI mapping.

multipath systematically; (ii) we use Wireless Insite 3D ray tracing to capture realistic but static environments; and (iii) we use testbed experiments to consider the impact of hardware features, channel fluctuation, and noise. Our algorithms are named *2ACE-Multi* (with multi-resolution) and *2ACE-No-Multi* (without multi-resolution).

Synthetic multipath simulation: We use a uniform linear array (ULA) with N_t Tx antennas and N_r Rx antennas and d antenna spacing. Given L paths in total and the AoA ϕ_l^{A}, AoD ϕ_l^{D}, and complex gain h_l of each paths, the overall channel between n_t-th Tx and n_r-th Rx antenna is calculated as [38]:

$$\mathbf{H}_{[n_t, n_r]} = \sum_{l=1}^{L} h_l \cdot e^{\frac{2\pi j d}{\lambda} ((n_t-1) \cos \phi_l^{\text{D}} + (n_r-1) \cos \phi_l^{\text{A}})} \quad (23)$$

In our evaluation, we set the path number to 6. We set the gain of the first path $|h_1| = 1$ and $0.1 \leq |h_l| \leq 0.2$ for the rest paths. ϕ_l^{A} and ϕ_l^{D} are both randomly generated between $\pi/6$ ands $5\pi/6$.

Wireless Insite: Wireless Insite (WI) is a commercial 3D ray-tracer widely used by the research community [27, 51]. We use a lidar scanner to reconstruct 3D environments in an office and then feed them to WI to generate the CSI.

Testbed: We use two laptops equipped with QCA6320 chipset-based 60 GHz WiFi card. Both TX and RX use a 6×6 uniform planer array (UPA) of 120° field-of-view (FoV), as shown in Fig.2(a,b). The antenna element spacing is 0.58λ [57]. Each antenna has a 1-bit switch (on or off), and a 2-bit phase shifter. Every 4 antennas colored the same share a 1-bit group switch [38]. All antennas share a single RF chain, as shown in Fig.2(c). The central carrier frequency is 60.48 GHz. We follow [38, 57] to calibrate the hardware phase offset on each antenna. The NIC reports the received signal strength indicator (RSSI) of the combined signal across all antennas at the RF chain. We measure a series of per-beam RSSI from NIC and record the corresponding RSS value by triggering the Linux command `iw dev $interface scan`. We then map RSSI to RSS by fitting the measured data via linear regression as shown in Fig.2(d) and find $RSS = 0.0651RSSI - 74.3875$.

Baseline: We compare 2ACE against PhaseLift [9], PLOMP/PL-GAMP/PerfectPhase [56], and sweeping. PLOMP and PLGAMP determine the complex channel from the dominant AoA and AoD. They first use PhaseLift for phase retrieval and use OMP [48] or EMBGAMP [49] to estimate sparse solutions. Although they converge faster than PhaseLift, their accuracy is restricted by its over-simplified sparse channel modeling. Moreover, PLOMP and
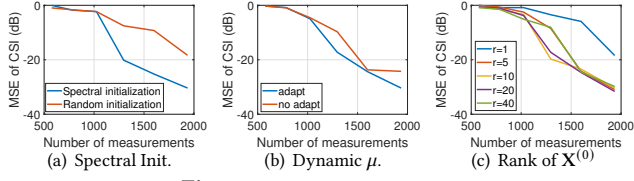
Figure 3: Micro-benchmarks.

PLGAMP only consider the azimuth directions, while PhaseLift and 2ACE support elevation tilting. *PerfectPhase* assumes a perfect phase retrieval and only solves the second step, but is *not* optimal due to the sparse channel representation. It requiresground truth phase information and can only be evaluated in simulation. We also add 2ACE using nuclear norm as a baseline, which promotes low-rank solutions instead of spectral profile for regularization, marked as *Nuclear*. The nuclear norm is a convex relaxation of non-zero eigenvalues (*i.e.*, rank), and is used as a regularization term in the classic compressive sensing low-rank formulation. We also include SLS, which performs TX beam training using 64 probes and fixes the RX to a quasi-omnidirectional beam. There are two enhanced versions (i) *beam sweeping (φ only)* enables both TX and RX beam training on azimuth plane and (2) *beam sweeping (φ & θ)* performs both azimuth and elevation sweeping. *beam sweeping (φ & θ)* is evaluated only in testbed since ULA used in simulation does not support elevation sweeping.

Sensing Matrix: PhaseLift and 2ACE use random beams, which is the general setting for compressive sensing. PLOMP, PLGAMP, and PerfectPhase only work with directional beamweights, which divide the azimuth plane uniformly.

Beamforming: We perform beamforming using the recovered CSI. Following the hardware constraints of existing commodity mmWave cards, in all cases we calculate the TX and RX codebooks as the 2-bit quantization of the first column in $\mathbf{U}$ and $\mathbf{V}$, respectively, where $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ is the SVD of CSI matrix $\mathbf{H}$.

Metrics: We quantify the CSI estimation error in simulation and Wireless Insite using mean square error (MSE), which is computed as $\mathbb{E}[\|\mathbf{x}_{est} - \mathbf{x}_{gt}\|_2^2 / \|\mathbf{x}_{gt}\|_2^2]$. For testbed experiments, due to the lack of the ground-truth CSI, we quantify the estimation error of RSS using random test beams. Moreover, in all cases, we also evaluate the beamforming performance. We run each algorithm 5 times for each configuration and report their average.

Default settings: Unless otherwise specified, we set $N_t = N_r = 16$ and signal-to-noise-ratio (SNR) as 40 dB, which is common for 60 GHz WiFi devices. We further vary the number of antennas and traces to understand their impacts. We use 0 elevations AoA to benefit PLOMP and PLGAMP since they cannot be easily generalized to support non-zero elevation.

5.2 Evaluation results

We first provide micro benchmarks. Then, we compare different algorithms in terms of CSI estimation error, beamforming performance, and sensing. We also evaluate 2ACE in mobile scenarios.

5.2.1 Micro-benchmarks We perform micro-benchmark using Wireless Insite traces to demonstrate each of our optimization techniques helps improve the performance. As shown in Fig.3(a), the spectral initialization outperforms the random initialization. For example, with 1296 measurements, spectral initialization achieves a -20 dB CSI estimation error, whereas random initialization achieves a -8
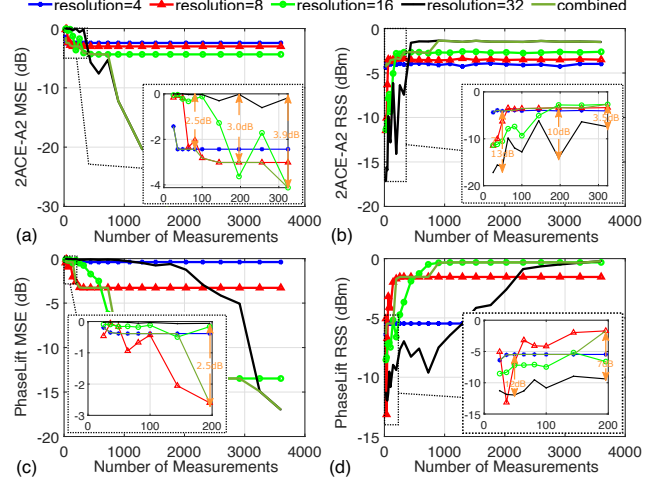

Figure 4: Multi-resolution: 2ACE (a) CSI error and (b) beamforming RSS and PhaseLift (c) CSI error and (d) beamforming RSS.

dB error. Then we compare the performance of μ adaptation in Fig.3 (b). The adapted version achieves lower error and converges faster. Fig.3 (c) shows the performance of parallel refinement. r is the number of initialization. $r = 1$ uses a single initialization and $r > 1$ uses parallel refinement. As we can see, $r > 1$ significantly outperforms $r = 1$, which demonstrates the benefit of parallel refinement.

5.2.2 Multi-resolution Algorithm: We show the benefit of multiresolution using the 32×32 Wireless Insite traces. Fig.4(a) shows the channel estimation error. With a smaller measurement budget, lower resolution yields a lower error, and 2ACE-Multi automatically selects the resolution based on the measurement budget. The dark curve shows that the multiresolution algorithm provides the lowest error in almost all cases. It improves over the 32×32 resolution scheme by 2.5dB, 3.0dB, and 3.9dB when it uses a resolution of 4×4, 8×8, and 16×16, respectively. Fig.4(b) shows that the improved CSI estimation accuracy gives better beamforming performance. 2ACE-Multi improves the RSS of 2ACE-No-Multi by 13dB, 10dB, and 3.5dB when using 4×4, 8×8, and 16×16 resolutions, respectively.

Besides, multi-resolution is general and can be applied to other algorithms. Fig.4 (c) and (d) show it helps PhaseLift reduce MSE by 2.5-12 dB and improve beamforming SNR by 7-12 dB.

5.2.3 CSI Estimation We first compare the performance in multipath simulation. In Fig.5, 2ACE-Multi yields <-10 dB error with $< 2N_t N_r$ measurements. It outperforms PhaseLift, PLOMP, and PLGAMP by 10dB using 500 measurements. Since PLOMP, PLGAMP, and PerfectPhase focus on recovering the dominant path, their channel estimation error remains high even with a large number of measurements. ACO provides a low-resolution estimation of CSI and thus does not provide high channel matrix estimation accuracy. PhaseLift and Nuclear-norm-based have higher errors even with $4N_t N_r$ measurements. 2ACE-Multi outperforms the nuclear norm since it can support matrices that are not strictly low rank. With a small number of measurements (*e.g.*, within 784 probes), 2ACE-Multi yields a 2-3dB improvement over 2ACE-No-Multi. For 32×32 case, 2ACE-Multi consistently performs the best and provides 10dB improvement over all baseline schemes with 2000 measurements.
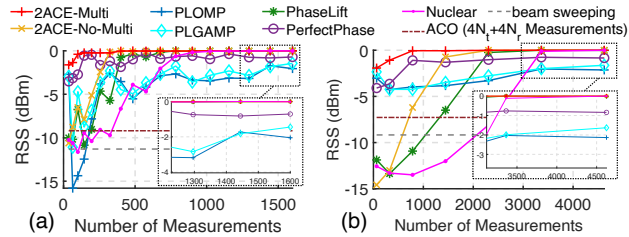
Next, we evaluate using the Wireless Insite traces. Fig.6 shows 2ACE-Multi outperforms all baselines by at least 15 dB with $2N_t N_r$
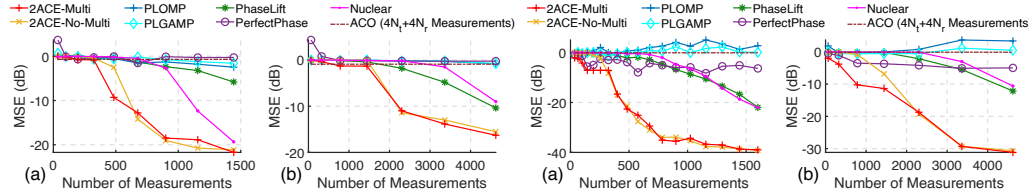
Figure 5: MSE of CSI estimation w/ multipath simulation: (a) 16×16; (b) 32×32.

Figure 6: MSE of CSI estimation w/ Wireless Insite 3D ray-tracing: (a) 16×16; (b) 32×32.

Figure 7: RSS error in testbed.
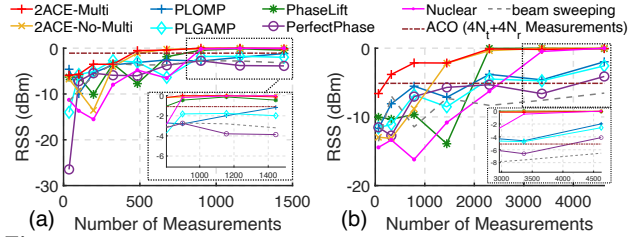
Figure 8: Beamforming based on recovered CSI from multipath simulation: (a) 16×16; (b) 32×32.


Figure 9: Beamforming based on recovered CSI from Wireless Insite 3D ray-tracing: (a) 16×16; (b) 32×32.

Figure 10: Testbed Experiments. (a) Experiment environment. (b-c) corresponds to location (b)-(c) in (a)

CSI matrices. 2ACE-Multi also outperforms 2ACE-No-Multi by 5-10 dB using fewer than 1000 measurements since it dynamically adjusts the estimation resolution based on the number of measurements.

Finally, we evaluate CSI estimation algorithms on our testbed. We collect training probes in a NLoS scenario for 16×16 CSI matrix recovery and quantify the estimation error by comparing the RSS in the testing set. As shown in Fig.7, 2ACE outperforms the others after convergence. 2ACE-Multi provides more than 1dB gain over the other algorithms at around 200 measurements and outperforms 2ACE-No-Multi by 0.7dB with 64 measurements. OMP has convergence problems under a large number of measurements.

5.2.4 Beamforming: We compare the beamforming RSS based on the recovered CSI. Using multipath simulation, Fig.8 shows that 2ACE-Multi quickly converges, achieving 1-5 dB beamforming gain over the baseline algorithms for 16×16 traces across all cases. The baselines take longer to converge and offer lower RSS than 2ACE. PhaseLift takes twice as many measurements as 2ACE. The benefit of 2ACE-Multi is even larger in 32×32 setups, providing 3-5 dB gain in most cases. This is because (i) our spectral profile-based estimation is more effective for a larger matrix, and (ii) the CSI estimation accuracy matters more for beamforming in a larger antenna array, which has more fine-grained beams. Existing compressive sensing works use nuclear norm as a regularization term to enforce low rank. We observe that 2ACE outperforms the nuclear-norm-based method by up to 8 dB because the matrices are not strictly low rank.

Then we evaluate using Wireless Insite traces. Fig.9 shows 2ACE-Multi achieves the best beamforming performance, out-performing PLOMP and PLGAMP by >2 dB. PhaseLift requires 4 times more
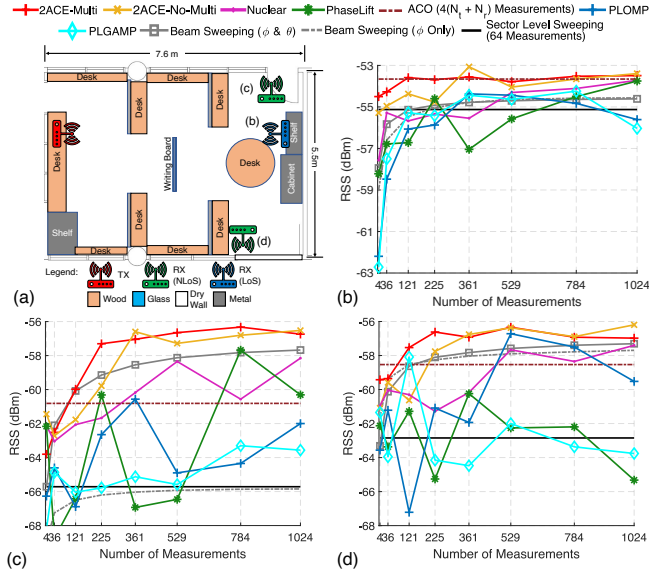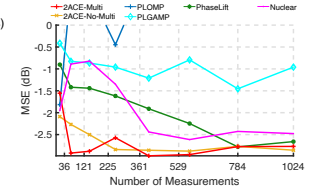
measurements than 2ACE-Multi. Nuclear-norm-based is worse than 2ACE since the channels are not strictly low rank. The performance benefit of our scheme increases to 2-3 dB in 32x32 traces since the spectral profile is more accurate and important for larger matrices.
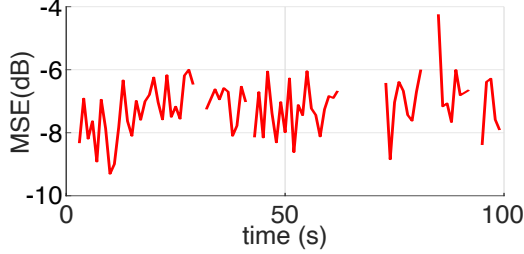
Finally, we evaluate 2ACE on the testbed. We use 16 antennas at both TX and RX. Fig.10 compares the beamforming performance with 3 different receiver locations, among which (d) is in LoS while (b) and (d) are in NLoS. 2ACE-Multi converges the fastest among all algorithms within 361 probes. SLS is always bad due to the missing of Rx-side beam training. Beam sweepings performs well at LoS but underperforms in NLoS scenarios. The codebook design of ACO limits its performance in low SNR scenarios since several bad probing is significantly detrimental to it. As a result, we see that ACO shows benefits in LoS cases but may significantly underperform in NLoS cases. 2ACE-Nuclear, PLOMP, and PLGAMP have serious issues dealing with NLoS scenarios, either due to low-rank assumptions or modeling without elevation angle. PhaseLift tends to converge to the optimal but needs many more measurements than 2ACE. Testbed evaluation shows that 2ACE can handle both LoS and NLoS scenarios and yields an average of $1 \sim 2$ dB beamforming gain compared with baselines.

5.2.5 Sensing: We use the MUSIC algorithm [41] to quantify AoD estimation accuracy. In the simulation, we show the MUSIC profiles and AoD estimation errors of the estimated CSI. We take the three highest peaks in the MUSIC profile as the AoD values and compare the outputs of different algorithms using optimal-matching [28]. On
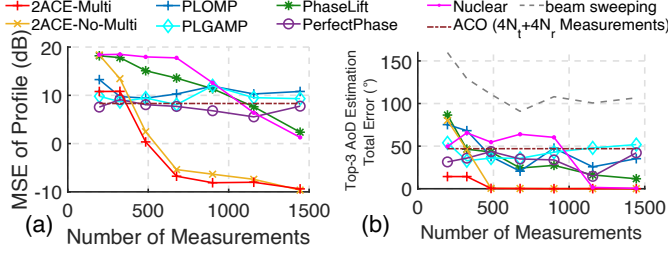
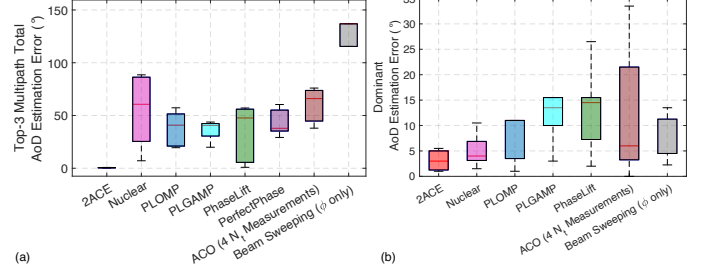**Figure 11: (a) MSE of MUSIC profile (b) Top-3 AoD Total Error in Simulation**

Figure 12: (a) Top-3 Multipath AoD Error in Simulation. (b) Dominant AoD Error on Testbed

Figure 13: Probe estimation error in mobile scenarios, when confidence bound is set at 0.75.

# Measurements	16	64	144	256
2ACE	0.117	0.179	0.52	1.282
2ACE incremental	**0.0065**	**0.0082**	**0.0091**	**0.017**
PLGAMP	1.328	5.677	7.666	12.197
PLOMP	1.009	5.325	7.372	11.752
PhaseLift	60.177	85.898	88.994	101.155

Table 1: Running Time (s) for estimating a 12×12 matrix.

the testbed, we only show the AoD estimation error of the strongest path in LoS scenarios since the missing of oracle CSI.

Using Wireless Insite traces, Fig. 11 (a) shows that 2ACE yields accurate MUSIC profiles with MSE below -10 dB after 512 probes, and outperforms all baselines by at least 10 dB. Fig. 11 (b) shows the AoD estimation error for the first three dominant paths: 2ACE outperforms the baseline algorithms by 5-50 degrees. Fig. 12 (a) shows that 2ACE outperforms all other algorithms with 400 measurements by around 50 degrees.

In the testbed, we place the RX at different locations to create different AoDs. All estimated CSI matrices are recovered with 361 measurements, where most algorithms converge eventually. As shown in Fig.12(b), 2ACE accurately estimates the LoS AoD and outperforms the other methods by 7 degrees on average. Meanwhile, the AoD estimation error using ACO-retrieved CSI is large in a complex indoor environment, which is consistent with the previous observation in X-array (Fig.11 in [52]). This also suggests the relative robustness of 2ACE in complicated environments.

5.2.6 Mobile scenarios: Our confidence indicator in Sec. 4.4 adapts 2ACE well in mobile scenarios. For evaluation, we send 200 probes to measure RSS every communication interval, where 128 probes are used for CSI estimation, among which 2ACE uses 95% of the input frames for training and 5% for deriving the confidence. The rest 72 probes are used for evaluating the system performance. A user holds the RX in hand and randomly moves in a room. During the movement, when the confidence is smaller than 0.75, 2ACE purges the first segment of the input traces and waits for new measurements until the confidence is high enough. Fig.13 shows the RSS error on the testing probes over time. When the user moves fast (*e.g.*, $T = 1 - 2$, $T = 65 - 73$ and $T = 82 - 84$), 2ACE-Multi finds the confidence value too low, so it does not yield the estimates but waits for more measurements. The RSS error of the output estimate is below -4dB all time, and below -6dB for most of the time, which shows the benefit of the confidence estimate.

5.2.7 Running Time: We show the advantage of 2ACE in running time on an Intel Core i7-8700. Using the highest resolution, to recover a 12×12 CSI matrix, 2ACE are 10×-10000× faster than PLGAMP, PLOMP, and PhaseLift. The fastest algorithm is an incremental version of 2ACE, which uses the result from the previous round for initialization. It completes within 17 ms with 256 measurements.

6 Conclusion

In this paper, we develop 2ACE, a novel compressive sensing algorithm that explores the spectral profile of CSI matrices for inference. We propose a series of optimization techniques, including spectral initialization, adapting the convexity of the objective, and parallel refinement. We show they are effective in speeding up the convergence. We further introduce confidence to determine the estimation quality for time-varying channels. Moreover, we develop multi-resolution estimation to automatically adapt the estimation granularity based on the measurement budget. With extensive evaluation, we demonstrate the improvement in the CSI estimation helps enhance beamforming RSS by 1-5 dB and AoD estimation accuracy by at $5°$ - $7°$ for real-world applications. We hope our work could inspire researchers to develop theories for spectral profile-based compressive sensing. More generally, it is interesting to re-visit compressive sensing for real-world matrices that may not be strictly low-rank.

References

[1] 2016. IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Controland Physical Layer Specifications. *IEEE Std 802.11-2016* (2016).

[2] 2021. IEEE Standard for Information Technology—Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks—Specific Requirements Part 11: Wireless LAN Medium Access Control and Physical Layer Specifications Amendment 2: Enhanced Throughput for Operation in License-exempt Bands above 45 GHz. *IEEE Std 802.11ay-2021* (2021).

[3] Ahmed Alkhateeb, Omar El Ayach, Geert Leus, and Robert W. Heath. 2014. Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems. *IEEE J. Sel. Top. Signal. Process.* 8, 5 (2014), 831–846.

[4] Alejandro Blanco, Pablo Jiménez Mateo, Francesco Gringoli, and Joerg Widmer. 2022. Augmenting MmWave Localization Accuracy through Sub-6 GHz on off-the-Shelf Devices. In *Proceedings of the MobiSys '22*. ACM, 477–490.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning* 3, 1 (2011), 1–122.

[6] A. Brighente, M. Cerutti, M. Nicoli, S. Tomasin, and U. Spagnolini. 2020. Estimation of Wideband Dynamic mmWave and THz Channels for 5G Systems and Beyond. *IEEE J. Sel. Areas Commun.* 38, 9 (2020), 2026–2040.

[7] Oliver Bunk, Ana Diaz, Franz Pfeiffer, Christian David, Bernd Schmitt, Dillip K. Satapathy, and J. Friso van der Veen. 2007. Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels. *Acta Crystallographica Section A* 63, 4 (Jul 2007), 306–314.

[8] E.J. Candes and T. Tao. 2005. Decoding by linear programming. *IEEE Trans. Inf. Theory* 51, 12 (2005), 4203–4215.

[9] E. J. Candès, T. Strohmer, and V. Voroninski. 2013. PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. *Communications on Pure and Applied Mathematics* 66, 8 (2013), 1241–1274.

[10] B. Chen, H. Li, Z. Li, X. Chen, C. Xu, and W. Xu. 2020. ThermoWave: A New Paradigm of Wireless Passive Temperature Monitoring via MmWave Sensing. In *Proceedings of the MobiCom '20*. ACM, Article 27, 14 pages.

[11] J. C. Dainty and J. R. Fienup. 1987. Phase retrieval and image reconstruction for astronomy. *Image Recovery: Theory and Application* (1987).

[12] K. P. Dasala, J. M. Jornet, and E. W. Knightly. 2020. SIMBA: Single RF Chain Multi-User Beamforming in 60 GHz WLANs. In *IEEE INFOCOM 2020*. 1499–1508.

[13] K. P. Dasala, J. M. Jornet, and E. W. Knightly. 2021. Uplink Multi-User Beamforming on Single RF Chain mmWave WLANs. In *IEEE INFOCOM 2021*. 1–10.

[14] P. Dong, H. Zhang, G. Ye Li, I. S. Gaspar, and N. NaderiAlizadeh. 2019. Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems. *IEEE J. Sel. Top. Signal. Process.* 13, 5 (2019), 989–1000.

[15] Tom Goldstein, Brendan O'Donoghue, Simon Setzer, and Richard Baraniuk. 2014. Fast Alternating Direction Optimization Methods. *SIAM Journal on Imaging Sciences* 7, 3 (2014), 1588–1623.

[16] U. Ha, S. Assana, and F. Adib. 2020. Contactless Seismocardiography via Deep Learning Radars. In *Proceedings of the MobiCom '20*. ACM, Article 62, 14 pages.

[17] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool Release: Gathering 802.11n Traces with Channel State Information. *ACM SIGCOMM CCR* 41, 1 (Jan. 2011), 53. https://dhalperi.github.io/linux-80211n-csitool/

[18] Robert W. Harrison. 1993. Phase problem in crystallography. *Journal of the Optical Society of America A* 10, 5 (May 1993), 1046–1055.

[19] Haitham Hassanieh, Omid Abari, Michael Rodriguez, Mohammed Abdelghany, Dina Katabi, and Piotr Indyk. 2018. Fast Millimeter Wave Beam Alignment. In *Proceedings of the 2018 ACM SIGCOMM 2018 Conference*. ACM, 432–445.

[20] Hengtao He, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. 2018. Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems. *IEEE Wireless Commun. Lett* 7, 5 (2018), 852–855.

[21] Chen Hu, Xiaodong Wang, Linglong Dai, and Junjie Ma. 2020. Partially Coherent Compressive Phase Retrieval for Millimeter-Wave Massive MIMO Channel Estimation. *IEEE Trans. Signal Process.* 68 (2020), 1673–1687.

[22] Chongwen Huang, Lei Liu, Chau Yuen, and Sumei Sun. 2019. Iterative Channel Estimation Using LSE and Sparse Message Passing for MmWave MIMO Systems. *IEEE Trans. Signal Process* 67, 1 (2019), 245–259.

[23] Ish Kumar Jain, Raghav Subbaraman, and Dinesh Bharadia. 2021. Two Beams Are Better than One: Towards Reliable and High Throughput MmWave Links. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. ACM, 488–502.

[24] Nilesh Kumar Jha and Vincent K. N. Lau. 2021. Online Downlink Multi-User Channel Estimation for mmWave Systems Using Bayesian Neural Network. *IEEE J. Sel. Areas Commun.* 39, 8 (2021), 2374–2387.

[25] Chengkun Jiang, Junchen Guo, Yuan He, Meng Jin, Shuai Li, and Yunhao Liu. 2020. MmVib: Micrometer-Level Vibration Measurement with Mmwave Radar. In *Proceedings of the MobiCom '20*. ACM, Article 45, 13 pages.

[26] Ting Jiang, Maozhong Song, Xuejian Zhang, and Xu Liu. 2021. Channel Estimation for Millimeter Wave Massive MIMO Systems Using Separable Compressive Sensing. *IEEE Access* 9 (2021), 49738–49749.

[27] Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. 2015. WiDeo: Fine-grained Device-free Motion Tracing using RF Backscatter. In *Proceedings of NSDI '15*. USENIX Association, 189–204.

[28] Harold W Kuhn. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics (NRL)* 2, 1-2 (1955), 83–97.

[29] Xingjian Li, Jun Fang, Hongbin Li, and Pu Wang. 2018. Millimeter Wave Channel Estimation via Exploiting Joint Sparse and Low-Rank Structures. *IEEE Trans. Wirel. Commun.* 17, 2 (2018), 1123–1133.

[30] W. Luo, W. Alghamdi, and Y. M. Lu. 2019. Optimal spectral initialization for signal recovery with applications to phase retrieval. *IEEE Trans. Signal Process.* 67, 9 (2019), 2347–2356.

[31] Zhinus Marzi, Dinesh Ramasamy, and Upamanyu Madhow. 2016. Compressive Channel Estimation and Tracking for Large Arrays in mm-Wave Picocells. *IEEE J. Sel. Top. Signal. Process.* 10, 3 (2016), 514–527.

[32] Jianwei Miao, Tetsuya Ishikawa, Qun Shen, and Thomas Earnest. 2008. Extending X-Ray Crystallography to Allow the Imaging of Noncrystalline Materials, Cells, and Single Protein Complexes. *Annu. Rev. Phys. Chem.* 59, 1 (2008), 387–410.

[33] Nicolo Michelusi and Muddassar Hussain. 2018. Optimal Beam-Sweeping and Communication in Mobile Millimeter-Wave Networks. In *2018 IEEE International Conference on Communications (ICC)*. 1–6.

[34] Wenqing Ouyang, Yue Peng, Yuxin Yao, Juyong Zhang, and Bailin Deng. 2020. Anderson Acceleration for Nonconvex ADMM Based on Douglas-Rachford Splitting. *Computer Graphics Forum* 39, 5 (2020), 221–239.

[35] Joan Palacios, Daniel Steinmetzer, Adrian Loch, Matthias Hollick, and Joerg Widmer. 2018. Adaptive Codebook Optimization for Beam Training on Off-the-Shelf 802.11ad Devices. In *Proceedings of MobiCom '18*. ACM, 241–255.

[36] Ranjitha Prasad, Chandra R. Murthy, and Bhaskar D. Rao. 2015. Joint Channel Estimation and Data Detection in MIMO-OFDM Systems: A Sparse Bayesian Learning Approach. *IEEE Trans. Signal Process* 63, 20 (2015), 5369–5382.

[37] Dheeraj Prasanna and Chandra R. Murthy. 2021. mmWave Channel Estimation via Compressive Covariance Estimation: Role of Sparsity and Intra-Vector Correlation. *IEEE Trans. Signal Process* 69 (2021), 2356–2370.

[38] Kun Qian, Xinyu Zhang, Zheng Yang, and Yunhao Liu. 2020. AoD-based localization with Cots millimeter-wave devices. *Sci. Sin. Inf* 51, 1 (2020), 122–.

[39] S. Rangan, T. S. Rappaport, and E. Erkip. 2014. Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges. *Proc. IEEE* 102, 3 (2014), 366–385.

[40] Mathew K. Samimi, George R. MacCartney, Shu Sun, and Theodore S. Rappaport. 2016. 28 GHz Millimeter-Wave Ultrawideband Small-Scale Fading Models in Wireless Channels. In *2016 IEEE 83rd Vehicular Technology Conference*. IEEE.

[41] R. Schmidt. 1986. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation* 34, 3 (1986), 276–280.

[42] E. Soltanaghaei, A. Prabhakara, A. Balanuta, M. Anderson, J. M. Rabaey, S. Kumar, and A. Rowe. 2021. Millimetro: MmWave Retro-Reflective Tags for Accurate, Long Range Localization. In *Proceedings of the MobiCom '21*. ACM, 69–82.

[43] Yiwen Song, Changhan Ge, Lili Qiu, and Yin Zhang. 2023. 2ACE Code Release. https://github.com/gavinsyw/2ACE-mmWave-Channel-Estimation

[44] Suraj Srivastava, Ch Suraj Kumar Patro, Aditya K. Jagannatham, and Lajos Hanzo. 2021. Sparse, Group-Sparse, and Online Bayesian Learning Aided Channel Estimation for Doubly-Selective mmWave Hybrid MIMO OFDM Systems. *IEEE Trans. Commun.* 69, 9 (2021), 5843–5858.

[45] S. Sur, I. Pefkianakis, X. Zhang, and K. Kim. 2017. WiFi-Assisted 60 GHz Wireless Networks. In *Proceedings of the MobiCom '17*. ACM, 28–41.

[46] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. 2018. Towards Scalable and Ubiquitous Millimeter-Wave Wireless Networks. In *Proceedings of the MobiCom '18*. Association for Computing Machinery, 257–271.

[47] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra. 2016. BeamSpy: Enabling Robust 60 GHz Links Under Blockage. In *NSDI 16*. USENIX Association, 193–206.

[48] Joel A. Tropp and Anna C. Gilbert. 2007. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory* 53, 12 (2007), 4655–4666.

[49] Jeremy Vila and Philip Schniter. 2011. Expectation-maximization Bernoulli-Gaussian approximate message passing. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers*. IEEE, 799–803.

[50] Adriaan Walther. 1963. The Question of Phase Retrieval in Optics. *Optica Acta: International Journal of Optics* 10, 1 (1963), 41–49.

[51] Song Wang, Jingqi Huang, and Xinyu Zhang. 2020. Demystifying Millimeter-Wave V2X: Towards Robust and Efficient Directional Connectivity under High Mobility. In *Proceedings of the MobiCom '20*. ACM, Article 51, 14 pages.

[52] Song Wang, Jingqi Huang, Xinyu Zhang, Hyoil Kim, and Sujit Dey. 2020. X-Array: Approximating Omnidirectional Millimeter-Wave Coverage Using an Array of Phased Arrays. In *Proceedings of the MobiCom '20*. ACM, Article 5, 14 pages.

[53] Teng Wei, Anfu Zhou, and Xinyu Zhang. 2017. Facilitating Robust 60 GHz Network Deployment By Sensing Ambient Reflectors. In *14th USENIX Symposium on Networked Systems Design and Implementation*. USENIX Association, 213–226.

[54] Yaxiong Xie, Zhenjiang Li, and Mo Li. 2015. Precise Power Delay Profiling with Commodity WiFi. In *Proceedings of the MobiCom '15*. ACM, 53–64.

[55] Xing Zhang, Lin Zhong, and Ashutosh Sabharwal. 2018. Directional Training for FDD Massive MIMO. *IEEE Trans. Wirel. Commun.* 17, 8 (2018), 5183–5197.

[56] 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 MobiCom '19*. ACM, 341–350.

[57] Renjie Zhao, Timothy Woodford, Teng Wei, Kun Qian, and Xinyu Zhang. 2020. M-Cube: A Millimeter-Wave Massive MIMO Software Radio. In *Proceedings of the MobiCom '20*. ACM, Article 15, 14 pages.

[58] Xuanyu Zheng and Vincent K. N. Lau. 2021. Online Deep Neural Networks for MmWave Massive MIMO Channel Estimation With Arbitrary Array Geometry. *IEEE Trans. Signal Process.* 69 (2021), 2010–2025.

[59] Yanzi Zhu, Yibo Zhu, Ben Y. Zhao, and Haitao Zheng. 2015. Reusing 60GHz Radios for Mobile Radar Imaging. In *Proceedings of the MobiCom '15*. ACM, 103–116.