Atomic-Norm Beamspace Parameter Estimation for Single-BS MIMO–OFDM Positioning

Alireza Pourafzal, *Member, IEEE*, Ruifu Li, Musa Furkan Keskin, *Member, IEEE*, Yu Ge, *Member, IEEE*, Danijela Cabric, *Fellow, IEEE*, and Henk Wymeersch, *Fellow, IEEE*

***Abstract*—Millimeter-wave single-base-station positioning with analog beamforming requires estimating multipath delays and departure/arrival angles from beam–frequency measurements, i.e., compressed observations of the underlying wideband MIMO–OFDM channel. In beamformed settings, the analog precoding/combining operator alters the standard Fourier/Vandermonde structure exploited by classical atomic line-spectral methods, making direct gridless recovery nontrivial. We address this problem by developing a beamspace specialization of coordinate-descent atomic soft-thresholding (CD-AST) that operates directly in the measurement domain through a pushed-forward atomic set induced by the beamforming operator. This yields two estimators: BEAST, a fully coupled 5-D off-grid estimator for joint delay–angle recovery, and EAST, a reduced-complexity delay-first variant that combines one-shot subspace-based delay estimation with per-slice 4-D atomic refinement. Simulations for 400-MHz mmWave MIMO–OFDM with practical beam training show that BEAST achieves delay and angular estimation accuracy close to the corresponding Cramér–Rao bounds and avoids the delay-resolution floor exhibited by subspace and matched-filter baselines, while EAST achieves a favorable accuracy–complexity trade-off by replacing repeated delay searches with one-shot subspace estimation. At 10 dBm total transmit power, BEAST attains centimeter-level positioning accuracy close to the position error bound, while EAST achieves comparable centimeter-level accuracy with reduced runtime.**

***Index Terms*—Millimeter-wave (mmWave), MIMO–OFDM, single-base-station positioning, beamspace channel estimation, gridless super-resolution, atomic norm minimization.**

I. Introduction

Millimeter-wave (mmWave) bands offer unprecedented bandwidth and support the integration of large-scale antenna arrays, making them a promising candidate for high-capacity wireless communication [1], sensing [2] and positioning [3] in emerging 5G New Radio (NR) and future 6G systems. The potential of mmWave to deliver ultra-high data rates and fine spatial resolution has fueled significant interest in physical-layer techniques for joint communication and localization [4]–[6]. Through wide bandwidths and large-aperture arrays, which bring high resolution in delay and angle, respectively, centimeter-level positioning becomes possible at mmWave [7]. However, realizing these benefits in practical multiple-input multiple-output (MIMO)–orthogonal frequency-division multiplexing (OFDM) transceivers remains challenging because (i) analog beamforming exposes only a low-rank projection of the full array response, (ii) joint angle–delay estimation must be carried out from high-dimensional beam-frequency measurements, quickly rendering standard high-resolution estimators computationally prohibitive, and (iii) hardware and pilot overhead constraints limit the number of RF chains, beams and pilot symbols used for channel estimation [8].

Existing channel parameter estimation methods span several distinct categories, each with important limitations for wideband mmWave OFDM scenarios. *Correlation-based estimators* [9] are computationally efficient but fundamentally limited by the ambiguity function, whose main-lobe width is determined by the signal bandwidth and array aperture. In grid-based implementations, they additionally suffer from basis mismatch when the true path parameters fall off the sampled grid. *Maximum likelihood algorithms* such as space-alternating generalized expectation-maximization (SAGE) [10] and RiMAX [11] can be statistically efficient, but lead to prohibitive computational complexity for high-dimensional MIMO–OFDM settings and often require good initialization to avoid convergence to suboptimal local maxima [12]. *Subspace techniques*, including estimation of signal parameters via rotational invariance techniques (ESPRIT) [13] and tensor factorization methods [14], can achieve high resolution for angle and delay estimation under favorable conditions. However, they require accurate model order selection and often rely on structural assumptions (e.g., shift-invariance or Vandermonde-type array responses induced by specific precoder/combiner designs). Their statistical efficiency is attained only when the signal subspace is correctly identified and noise is appropriately whitened [15], and their computational complexity typically scales cubically (or worse) with the effective array or beam–frequency dimensions. Moreover, depending on the adopted factorization technique, an additional pairing stage between angle and delay parameters may be required [16]. Finally, *sparse recovery methods* such as orthogonal matching pursuit (OMP) [17] and sparse Bayesian learning (SBL) [18] relax the model order requirement and offer improved robustness at low signal-to-noise ratio (SNR). Grid-based variants, however, rely on discretization of the

A. Pourafzal and R. Li contributed equally to this work. A. Pourafzal, M. F. Keskin, Y. Ge, and H. Wymeersch are with the Department of Electrical Engineering, Chalmers University of Technology, Sweden (emails: {alireza.pourafzal; furkan; yuge; henkw}@chalmers.se). R. Li and D. Cabric are with the Electrical and Computer Engineering Department, UCLA, USA (emails: doanr37@ucla.edu; danijela@ee.ucla.edu).

This work was supported in part by the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101207620 (MIDAS-6G), and by the Swedish Research Council (VR) through the project 6G-PERCEF under Grant 2024-04390, and by the Wallenberg Foundation and WASP Postdoctoral Scholarship Program. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

continuous parameter space, which degrades estimation accuracy and leads to basis mismatch when the true parameters fall off-grid [19]. To overcome these limitations, gridless sparse approaches based on atomic norm minimization (ANM) have been proposed for mmWave channel estimation [20]–[26], enabling super-resolution recovery directly in the continuous domain without basis mismatch [25], [27]. Most ANM-based channel estimators for mmWave MIMO(–OFDM) rely on semidefinite programming (SDP) representations with Toeplitz, block-Toeplitz, or multi-level Toeplitz constraints, which become computationally heavy in large beam–frequency dimensions and typically require iterative solvers such as alternating direction method of multipliers (ADMM) for tractability. This is explicit, for example, in beamformed full-dimensional MIMO ANM formulations, where the SDP involves high-dimensional positive semidefinite constraints and ADMM is introduced mainly to reduce the cost compared to generic solvers [20], [25], [26], [28]. A more tractable alternative to SDP-based ANM is coordinate-descent atomic-norm soft-thresholding (CD-AST), which replaces semidefinite programming with coordinate-descent updates and yields low per-iteration complexity for Fourier-type atoms [29]. However, the available CD-AST formulation applies to element-space line-spectral estimation with standard Fourier atoms.

In beamformed mmWave MIMO-OFDM channel estimation, the analog precoding/combining and partial sampling operator pushes physical delay-angle atoms into structured beam-frequency atoms in the measurement domain. Therefore, the oracle correlation used by CD-AST is no longer evaluated against standard Fourier atoms, and the fast Fourier transform (FFT)-based maximization available in the original formulation [29] does not apply directly. A naive implementation would then require dense search over the joint delay-angle space, whose cost becomes unfavorable in high dimensions. In addition, the shrinkage and projection steps must be expressed directly for the sensed atoms under the beamforming operator, while preserving the separable structure across subcarriers and beam dimensions for scalability.

In this work, we develop a low-complexity, gridless super-resolution framework for five-dimensional beamspace channel estimation and single-base station (BS) positioning in mmWave MIMO-OFDM. Our key idea is to reformulate atomic soft-thresholding directly in the beam–frequency measurement domain by pushing forward physical path atoms through the beamforming operator and evaluating them only on the observed beam–subcarrier indices. This yields a beamspace atomic formulation. We then apply the core CD-AST primitives, namely a dual-certificate oracle for continuous atom selection and a closed-form complex shrinkage update, without discretization or semidefinite programming. To make the oracle tractable, we exploit the separable structure induced by OFDM and analog beamforming, enabling structured correlation maximization by a decoupled coarse scan followed by local continuous refinement. Building on this formulation, we propose two estimators that offer complementary accuracy–complexity trade-offs. First, we develop BEAST (beamspace coordinate-descent atomic soft-thresholding), a fully gridless five-dimensional beamspace algorithm that operates directly on the complete beam–frequency tensor and jointly estimates delay and angles. Second, we develop EAST (ESPRIT-aided CD-AST), a delay-first decoupling approach motivated by the fine delay resolution in wideband systems, which estimates the dominant delay subspace once via an ESPRIT-type procedure on the mode-3 unfolding, removes the delay dimension by projection, and performs per-slice four-dimensional beamspace CD-AST refinement over the remaining angular parameters. The main contributions are summarized as follows:

- **Beamspace specialization of coordinate-descent atomic soft-thresholding.** We extend CD-AST from line-spectral denoising to single-snapshot beamformed mmWave MIMO-OFDM channel estimation by introducing a pushed-forward beam–frequency atomic set induced by analog precoding and combining, and by deriving the corresponding dual-certificate stopping test and coefficient-update rule directly in the measurement domain.
- **BEAST: a fully gridless 5-D beamspace estimator.** Using the above primitives, we develop a practical beamspace implementation with a structured correlation maximization, enabling joint recovery of delay and angular parameters from the beam–frequency tensor. BEAST attains near-Cramér-Rao bound (CRB) accuracy down to low SNR with tractable computational complexity.
- **EAST: an OFDM-separable delay-first variant.** We further propose a reduced-complexity estimator that exploits the fine-delay resolution of wideband OFDM by estimating the delay support first using subspace methods, projecting onto the delay subspace, and applying per-slice gridless angular refinement. This yields a principled accuracy–complexity trade-off relative to full 5-D recovery. EAST reduces complexity by replacing repeated delay searches with one-shot subspace estimation, at the expense of increased variance under delay-subspace mismatch and error propagation into the angular refinement stage.

Notation. Scalars are x, vectors are $\mathbf{x}$, and multi-dimensional arrays (matrices/tensors) are $\mathbf{X}$; sets and spaces are $\mathcal{X}$. $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$, and $(\cdot)^{\dagger}$ denote transpose, Hermitian transpose, and pseudoinverse. $\langle \cdot, \cdot \rangle$ denotes an inner product, $\otimes$ the Kronecker product, and $\circ$ the outer product. $\| \cdot \|_2$ and $\| \cdot \|_F$ denote the Euclidean and Frobenius norms. $(\cdot)_+$ denotes the thresholding operator such that $\forall x \in \mathbb{R}, (x)_+ = \max(x, 0)$.

II. System Model and Preliminaries

A. Channel Geometry and Array Model

Consider a single BS and a single user equipment (UE), each equipped with a uniform rectangular array (URA) with half-wavelength inter-element spacing. The BS URA has size N_{BS} and the UE URA has size N_{UE} ($N_{\mathrm{BS}} \gg N_{\mathrm{UE}}$). The array reference points are located at $\mathbf{p}_{\mathrm{BS}} \in \mathbb{R}^3$ and $\mathbf{p}_{\mathrm{UE}} \triangleq \mathbf{p}_{0,0} \in \mathbb{R}^3$, respectively. The global coordinate system is aligned with

the BS array so that the BS rotation is $\mathbf{R}_{\mathrm{BS}} = \mathbf{I}_3$, while the UE array has a rotation matrix $\mathbf{R}_{\mathrm{UE}} \in \mathbb{R}^{3\times 3}$ that maps the UE local frame to the global frame. Throughout the paper, $\mathbf{p}_{\mathrm{BS}}$ and $\mathbf{R}_{\mathrm{BS}}$ are known, while the UE position $\mathbf{p}_{0,0}$, UE orientation $\mathbf{R}_{\mathrm{UE}}$, and the common BS–UE clock bias β are unknown deterministic parameters. The propagation environment is described by L planar reflective surfaces $\{\mathcal{S}_\ell\}_{\ell=1}^{L}$, each modeling one cluster (e.g., a building facade, a ground-like reflector). Each surface $\mathcal{S}_\ell$ is characterized by a centroid, a unit normal vector, and a finite rectangular support in 3D space. We denote by $\Gamma_\ell \in \mathbb{C}$ the (effective) Fresnel reflection coefficient of surface $\mathcal{S}_\ell$, with $\Gamma_0 = 1$ for the line-of-sight (LoS) path. Each propagation path is indexed by (ℓ, r), where $\ell = 0$ corresponds to the LoS cluster and $\ell \in \{1, \dots, L\}$ corresponds to a surface. For the LoS cluster, we set $R_0 = 0$. Within cluster $\ell \geq 1$, the index $r = 0$ denotes the specular reflection point on $\mathcal{S}_\ell$ and $r \in \{1, \dots, R_\ell\}$ labels diffuse subpaths on the same surface. Let $\mathbf{p}_{\ell,r} \in \mathbb{R}^3$ denote the anchor point of path (ℓ, r). For $\ell \geq 1$ and $r = 0$, $\mathbf{p}_{\ell,0}$ is obtained as the single-bounce specular intersection (i.e., obtained via image theory for a planar reflector) on $\mathcal{S}_\ell$.[1] For $r \geq 1$, the points $\{\mathbf{p}_{\ell,r}\}_{r=1}^{R_\ell} \subset \mathcal{S}_\ell$ are drawn from a 2D Gaussian density on the surface centered at the specular point and truncated to the rectangular support. This realizes a lobe of diffuse scattering around the specular point and can be viewed as a simple Gaussian approximation of the surface joint angle–delay power spectrum as in [30]. For any path (ℓ, r), we define the unit direction vectors

$$\mathbf{u}_{\mathrm{BS},\ell,r} = \frac{\mathbf{p}_{\ell,r} - \mathbf{p}_{\mathrm{BS}}}{\left\|\mathbf{p}_{\ell,r} - \mathbf{p}_{\mathrm{BS}}\right\|}, \qquad \mathbf{u}^{(\mathrm{glob})}_{\mathrm{UE},\ell,r} = -\frac{\mathbf{p}_{\ell,r} - \mathbf{p}_{0,0}}{\left\|\mathbf{p}_{\ell,r} - \mathbf{p}_{0,0}\right\|}.$$

Angles at the BS are expressed in the global frame as

$$\theta^{\mathrm{az}}_{\ell,r} = \arctan 2\big(\mathbf{u}^{y}_{\mathrm{BS},\ell,r}, \mathbf{u}^{x}_{\mathrm{BS},\ell,r}\big), \quad \theta^{\mathrm{el}}_{\ell,r} = \arcsin\big(\mathbf{u}^{z}_{\mathrm{BS},\ell,r}\big).$$

Angles at the UE are subject to rotation into the UE local frame. The global UE direction is mapped as $\mathbf{u}_{\mathrm{UE},\ell,r} = \mathbf{R}^{-1}_{\mathrm{UE}}\mathbf{u}^{(\mathrm{glob})}_{\mathrm{UE},\ell,r}$, and the UE azimuth and elevation are then obtained as $\phi^{\mathrm{az}}_{\ell,r} = \arctan 2\big(\mathbf{u}^{y}_{\mathrm{UE},\ell,r}, \mathbf{u}^{x}_{\mathrm{UE},\ell,r}\big)$, and $\phi^{\mathrm{el}}_{\ell,r} = \arcsin\big(\mathbf{u}^{z}_{\mathrm{UE},\ell,r}\big)$. The geometric delay of path (ℓ, r) is

$$\tau^{\mathrm{geo}}_{\ell,r} = \frac{\left\|\mathbf{p}_{\ell,r} - \mathbf{p}_{\mathrm{BS}}\right\| + \left\|\mathbf{p}_{\ell,r} - \mathbf{p}_{0,0}\right\|}{c},$$

where c is the speed of light. A common BS–UE clock bias β is present and assumed constant over the considered training burst, such that the total delay is $\tau_{\ell,r} = \beta + \tau^{\mathrm{geo}}_{\ell,r}$.

B. Wideband Channel Assembly

We consider a frequency-selective MIMO channel observed over a bandwidth B around carrier frequency f_{c}. The continuous-frequency BS–UE channel matrix is modeled as a sum of LoS and all non-line-of-sight (NLoS) (specular and diffuse) paths [31][2]

$$\mathbf{H}(f) = \sum_{\ell=0}^{L}\sum_{r=0}^{R_\ell} \alpha_{\ell,r}\, \mathbf{a}_{\mathrm{UE}}\big(\boldsymbol{\phi}_{\ell,r}\big)\, \mathbf{a}^{\mathsf{T}}_{\mathrm{BS}}\big(\boldsymbol{\theta}_{\ell,r}\big)\, \mathrm{e}^{-j2\pi f\, \tau_{\ell,r}}, \quad (1)$$

where the angle vectors are $\boldsymbol{\phi}_{\ell,r} = [\phi^{\mathrm{az}}_{\ell,r}, \phi^{\mathrm{el}}_{\ell,r}]^{\mathsf{T}}$ and $\boldsymbol{\theta}_{\ell,r} = [\theta^{\mathrm{az}}_{\ell,r}, \theta^{\mathrm{el}}_{\ell,r}]^{\mathsf{T}}$, and the URA response admits a separable Kronecker structure. For a generic azimuth–elevation pair $(\varphi^{\mathrm{az}}, \varphi^{\mathrm{el}})$, define the spatial frequencies $\omega_x = \pi \sin(\varphi^{\mathrm{el}})$, and $\omega_y = \pi \cos(\varphi^{\mathrm{el}}) \sin(\varphi^{\mathrm{az}})$, and the steering vector $\mathbf{a}_{N_x\times N_y}(\varphi^{\mathrm{az}}, \varphi^{\mathrm{el}}) = \mathbf{a}_{N_x}(\omega_x) \otimes \mathbf{a}_{N_y}(\omega_y)$. Accordingly, $\mathbf{a}_{\mathrm{UE}}(\boldsymbol{\phi}_{\ell,r}) = \mathbf{a}_{N_1\times N_2}(\phi^{\mathrm{az}}_{\ell,r}, \phi^{\mathrm{el}}_{\ell,r})$, and $\mathbf{a}_{\mathrm{BS}}(\boldsymbol{\theta}_{\ell,r}) = \mathbf{a}_{N_3\times N_4}(\theta^{\mathrm{az}}_{\ell,r}, \theta^{\mathrm{el}}_{\ell,r})$, where $[\mathbf{a}_N(\omega)]_n = \exp(j\omega n)$.

For the LoS and specular components, we adopt a single-bounce Friis-type gain with reflection coefficient Γ_ℓ. Let

$$d_{\ell,0} = \left\|\mathbf{p}_{\ell,0} - \mathbf{p}_{\mathrm{BS}}\right\| + \left\|\mathbf{p}_{\ell,0} - \mathbf{p}_{0,0}\right\|$$

denote the total path length of the specular component. The complex path gain is written as [32] [33]

$$\alpha_{\ell,0} = \frac{\lambda\, \Gamma_\ell}{4\pi d_{\ell,0}}\, \mathrm{e}^{-j\phi_{\ell,0}}, \quad (2)$$

where λ is the carrier wavelength and $\phi_{\ell,0}$ collects reflection-induced phase terms. For the LoS path we set $\Gamma_0 = 1$. The corresponding specular power of surface $\mathcal{S}_\ell$ is $P_{\mathrm{spec},\ell} = |\alpha_{\ell,0}|^2$. For each surface $\mathcal{S}_\ell$, diffuse scattering is modeled by R_ℓ subpaths generated from the Gaussian surface sampling described above. For $r \geq 1$, the delay and angles of path (ℓ, r) are obtained from the geometry as in Section II-A, while the complex gains are modeled as [33]

$$\alpha_{\ell,r} = \frac{\lambda}{(4\pi)^{3/2}} \frac{\sqrt{(1-|\Gamma_\ell|^2)\, \eta_\ell\, \sigma_{\ell,r}/R_\ell}\, \mathrm{e}^{-j\phi_{\ell,r}}}{\left\|\mathbf{p}_{\ell,r} - \mathbf{p}_{\mathrm{BS}}\right\| \left\|\mathbf{p}_{\ell,r} - \mathbf{p}_{0,0}\right\|}, \quad (3)$$

where $\eta_\ell \in [0, 1]$ controls the fraction of non-specular energy converted into diffuse scattering on $\mathcal{S}_\ell$, $\sigma_{\ell,r}$ is an effective radar cross-section of the r-th diffuse point (with prescribed mean and variance), and $\phi_{\ell,r}$ is an independent random phase, uniform over $[0, 2\pi)$. For $r \geq 1$, the diffuse amplitudes in (3) are scaled by $\sqrt{(1-|\Gamma_\ell|^2)\eta_\ell}$, while the specular component scales with $|\Gamma_\ell|$ in (2). As a result, the average powers of the specular, diffuse, and absorbed components of surface $\mathcal{S}_\ell$ scale as $P_{\mathrm{spec},\ell} \propto |\Gamma_\ell|^2$, $P_{\mathrm{diff},\ell} \propto (1-|\Gamma_\ell|^2)\, \eta_\ell$, and $P_{\mathrm{abs},\ell} \propto (1-|\Gamma_\ell|^2)\,(1-\eta_\ell)$, so that $\big(|\Gamma_\ell|^2,\ (1-|\Gamma_\ell|^2)\eta_\ell,\ (1-|\Gamma_\ell|^2)(1-\eta_\ell)\big)$ can be interpreted as the specular, diffuse, and absorbed power fractions of surface $\mathcal{S}_\ell$, in line with surface-based model [30].

C. Signal Model

We consider a single training burst during which the BS transmits pilot symbols over K OFDM subcarriers with spacing Δf around the carrier frequency f_c. The frequency of the kth subcarrier is

$$f_k \triangleq f_c + \Big(k - \frac{K-1}{2}\Big)\Delta f, \qquad k = 0, \dots, K-1.$$

Analog beamforming is implemented via codebook-based beam switching at both the BS and the UE. The BS employs a transmit codebook $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_S] \in \mathbb{C}^{N_{\mathrm{BS}}\times S}$, while the UE employs a receive codebook $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_M] \in \mathbb{C}^{N_{\mathrm{UE}}\times M}$, with unit-norm beams $\|\mathbf{f}_s\|_2 = \|\mathbf{w}_m\|_2 = 1$. During the burst, the BS sequentially fixes a transmit beam $\mathbf{f}_s$ and the UE sweeps through all receive beams $\{\mathbf{w}_m\}_{m=1}^{M}$. This procedure is repeated for all S transmit beams, resulting in MS beam-pair measurements per subcarrier [34].

[1] The single-bounce assumption is not fundamental to the proposed channel estimation framework, which applies to generic multipath collections as long as the resulting components admit distinct angle–delay parameters.

[2] The model assumes frequency-flat array responses over the bandwidth of interest and neglects Doppler effects during the training burst.

With pilot symbol $\bar{p}_{m,s,k}$ transmitted on subcarrier k using beam pair $(\mathbf{f}_s, \mathbf{w}_m)$, the received signal at the UE is

$$\bar{y}_{m,s,k} = \mathbf{w}_m^{\mathsf{H}}\, \mathbf{H}(f_k)\, \mathbf{f}_s\, \bar{p}_{m,s,k} + \bar{n}_{m,s,k}, \tag{4}$$

where $|\bar{p}_{m,s,k}|^2 = P_t$, denoting the transmit power per subcarrier per OFDM symbol, and $\bar{n}_{m,s,k} \sim \mathcal{CN}(0, \sigma_n^2)$ denotes additive noise with variance $\sigma_n^2 = N_0 \Delta f$, corresponding to thermal noise with two-sided power spectral density N_0. All pilot observations over M receive beams, S transmit beams, and K subcarriers are stacked into a third-order tensor:

$$\bar{\mathbf{Y}} \in \mathbb{C}^{M \times S \times K}, \qquad [\bar{\mathbf{Y}}]_{m,s,k} = \bar{y}_{m,s,k}. \tag{5}$$

D. Downlink Positioning

Let $\{\widehat{\boldsymbol{\rho}}_\ell\}_{\ell=0}^{\widehat{L}-1} = \mathsf{est}(\bar{\mathbf{Y}})$ denote the channel-parameter estimator, where $\widehat{L}$ is the estimated number of paths, $\widehat{\boldsymbol{\rho}}_\ell = \big[\widehat{\tau}_\ell, \widehat{\boldsymbol{\theta}}_\ell, \widehat{\boldsymbol{\phi}}_\ell, \widehat{\alpha}_\ell\big]^{\mathsf{T}}$ collects delay, angle of departure (AoD), angle of arrival (AoA), and complex gain of path ℓ. We identify the LoS path as the one with the smallest delay, $\widehat{\ell}_0 \in \arg\min_\ell\ \widehat{\tau}_\ell$, and set $(\widehat{\tau}_0, \widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\phi}}_0) \triangleq (\widehat{\tau}_{\widehat{\ell}_0}, \widehat{\boldsymbol{\theta}}_{\widehat{\ell}_0}, \widehat{\boldsymbol{\phi}}_{\widehat{\ell}_0})$, so that ideally $\widehat{\tau}_0 \approx \tau_{0,0}$. The NLoS candidate is selected as the strongest non-LoS component, $\widehat{\ell}_1 \in \arg\max_{\ell \neq \widehat{\ell}_0} |\widehat{\alpha}_\ell|$, with $(\widehat{\tau}_1, \widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\phi}}_1) \triangleq (\widehat{\tau}_{\widehat{\ell}_1}, \widehat{\boldsymbol{\theta}}_{\widehat{\ell}_1}, \widehat{\boldsymbol{\phi}}_{\widehat{\ell}_1})$. From the estimated azimuth–elevation pairs, define the unit direction mapping $\mathbf{u}(\varphi^{\mathrm{az}}, \varphi^{\mathrm{el}}) \triangleq \big[\cos(\varphi^{\mathrm{el}})\cos(\varphi^{\mathrm{az}}) \quad \cos(\varphi^{\mathrm{el}})\sin(\varphi^{\mathrm{az}}) \quad \sin(\varphi^{\mathrm{el}})\big]^{\mathsf{T}}$. Then $\hat{\mathbf{u}}_{\mathrm{BS},i} = \mathbf{u}(\widehat{\theta}_i^{\mathrm{az}}, \widehat{\theta}_i^{\mathrm{el}})$, and $\hat{\mathbf{u}}_{\mathrm{UE},i}^{(\mathrm{loc})} = \mathbf{u}(\widehat{\phi}_i^{\mathrm{az}}, \widehat{\phi}_i^{\mathrm{el}})$, for $i \in \{0,1\}$. We then define the principal angle separations $\Delta\widehat{\theta}_{1,0} = \arccos\big(\hat{\mathbf{u}}_{\mathrm{BS},1}^{\mathsf{T}} \hat{\mathbf{u}}_{\mathrm{BS},0}\big), \Delta\widehat{\phi}_{1,0} = \arccos\big((\hat{\mathbf{u}}_{\mathrm{UE},1}^{(\mathrm{loc})})^{\mathsf{T}} \hat{\mathbf{u}}_{\mathrm{UE},0}^{(\mathrm{loc})}\big)$. Note that $\Delta\widehat{\phi}_{1,0}$ is invariant to the unknown UE orientation because both AoAs are measured in the same local frame.

With a common clock bias β, $\tau_\ell = \beta + \tau_\ell^{\mathrm{geo}}$ and thus

$$\Delta\widehat{\tau}_{1,0} \triangleq \widehat{\tau}_1 - \widehat{\tau}_0 \approx \tau_1^{\mathrm{geo}} - \tau_0^{\mathrm{geo}}, \tag{6}$$

so that the bias β cancels out. The LoS and NLoS rays define a plane Π (the span of $\hat{\mathbf{u}}_{\mathrm{BS},0}$ and $\hat{\mathbf{u}}_{\mathrm{BS},1}$); the BS–scatter–UE geometry projected onto Π forms a triangle with side lengths equal to the LoS range $d_0 = \|\mathbf{p}_{\mathrm{UE}} - \mathbf{p}_{\mathrm{BS}}\|$ and the NLoS path length $d_1 = d_{\mathrm{BS},1} + d_{\mathrm{UE},1}$. Using $c\,\Delta\widehat{\tau}_{1,0} = d_1 - d_0$ and the law of sines in Π (see also [35], [36]), the LoS range is

$$\widehat{d}_0 = \frac{c\,\Delta\widehat{\tau}_{1,0}\,\sin\big(\Delta\widehat{\theta}_{1,0} + \Delta\widehat{\phi}_{1,0}\big)}{\sin(\Delta\widehat{\theta}_{1,0}) + \sin(\Delta\widehat{\phi}_{1,0}) - \sin\big(\Delta\widehat{\theta}_{1,0} + \Delta\widehat{\phi}_{1,0}\big)}. \tag{7}$$

Given $\widehat{d}_0$, the UE position follows along the LoS BS ray:

$$\widehat{\mathbf{p}}_{\mathrm{UE}} = \mathbf{p}_{\mathrm{BS}} + \widehat{d}_0\, \hat{\mathbf{u}}_{\mathrm{BS},0}. \tag{8}$$

E. Problem Statement

While (7)–(8) assume ideal multipath knowledge, the delay–angle parameters must be estimated from noisy beamformed mmWave MIMO–OFDM measurements. We therefore seek an accurate and computationally efficient estimator operating directly on the beamspace observation tensor in (5).

III. A Primer on Coordinate-Descent Atomic Soft-Thresholding

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{C}$ with a complex inner-product $\langle \cdot, \cdot \rangle$. [3] Let $\mathcal{A} = \{\mathbf{a}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Omega\} \subset \mathcal{H}$ be a continuous atomic set with parameter domain Ω, and define the atomic norm as the gauge of the convex hull $\mathrm{conv}(\mathcal{A})$ [37], [38]

$$\|\mathbf{x}\|_{\mathcal{A}} \triangleq \inf\Big\{ \sum\nolimits_\ell |\alpha_\ell| :\ \mathbf{x} = \sum\nolimits_\ell \alpha_\ell \mathbf{a}(\boldsymbol{\theta}_\ell),\ \mathbf{a}(\boldsymbol{\theta}_\ell) \in \mathcal{A} \Big\}.$$

Consider the atomic-norm denoising problem

$$\min\nolimits_{\mathbf{x} \in \mathcal{H}}\ \|\mathbf{x}\|_{\mathcal{A}} + \frac{\zeta}{2}\|\tilde{\mathbf{y}} - \mathbf{x}\|^2, \qquad \zeta > 0, \tag{9}$$

for a given $\tilde{\mathbf{y}} \in \mathcal{H}$, where ζ controls the sparsity–fit trade-off and implicitly the effective model order. Suppose $\mathbf{x}^\star$ is an optimal solution to (9) and define the residual $\mathbf{r}^\star \triangleq \tilde{\mathbf{y}} - \mathbf{x}^\star$. Since (9) is convex and unconstrained, its first-order optimality is equivalent to the subgradient inclusion [39, Sec. 23]

$$\mathbf{0} \in \partial\|\mathbf{x}^\star\|_{\mathcal{A}} + \zeta(\mathbf{x}^\star - \tilde{\mathbf{y}}) \iff \zeta\,\mathbf{r}^\star \in \partial\|\mathbf{x}^\star\|_{\mathcal{A}}. \tag{10}$$

This yields the dual-feasibility condition

$$\|\mathbf{r}^\star\|_{\mathcal{A}}^* \le \zeta^{-1}, \qquad \|\mathbf{r}\|_{\mathcal{A}}^* \triangleq \sup\nolimits_{\mathbf{a} \in \mathcal{A}} |\langle \mathbf{a}, \mathbf{r} \rangle|, \tag{11}$$

where $\|\cdot\|_{\mathcal{A}}^*$ denotes the dual atomic norm. (11) provides a global stopping test: if the maximum residual-atom correlation falls below ζ^{-1}, no further atom can decrease the objective.

CD-AST [29] solves (9) by maintaining an explicit finite atomic expansion $\mathbf{x} = \sum_{\ell=0}^{L-1} \alpha_\ell \mathbf{a}_\ell$, with $\mathbf{a}_\ell \in \mathcal{A}$, which leads to the atomic regression form over a finite decomposition,

$$\min_{L, \{\alpha_\ell\}, \{\mathbf{a}_\ell \in \mathcal{A}\}}\ \sum\nolimits_{\ell=0}^{L-1} |\alpha_\ell| + \frac{\zeta}{2}\Big\|\tilde{\mathbf{y}} - \sum\nolimits_{\ell=0}^{L-1} \alpha_\ell \mathbf{a}_\ell\Big\|^2. \tag{12}$$

It then applies a block-coordinate descent procedure over (i) coefficients $\{\alpha_\ell\}$ and (ii) continuous atom parameters $\{\boldsymbol{\theta}_\ell\}$ through $\mathbf{a}_\ell = \mathbf{a}(\boldsymbol{\theta}_\ell)$ [29]. The algorithmic primitives are:

1) Dual-certificate Oracle: CD-AST repeatedly evaluates[4]

$$\mathbf{a}^\star \in \arg\max\nolimits_{\mathbf{a} \in \mathcal{A}}\ \big|\langle \mathbf{a}, \mathbf{r} \rangle\big|, \tag{13}$$

identifying the atom most correlated with the current residual.

2) Single-coefficient Update (Complex Shrinkage): Fix all atoms and all coefficients except α_ℓ, and define the residual that excludes the ℓth component $\mathbf{r}_{-\ell} \triangleq \tilde{\mathbf{y}} - \sum_{j \neq \ell} \alpha_j \mathbf{a}_j$. Updating α_ℓ then reduces to the 1-D convex problem

$$\min\nolimits_{\alpha_\ell \in \mathbb{C}}\ |\alpha_\ell| + \frac{\zeta}{2}\|\mathbf{r}_{-\ell} - \alpha_\ell \mathbf{a}_\ell\|^2. \tag{14}$$

Up to constants independent of α_ℓ, (14) is equivalent to

$$\min_{\alpha_\ell \in \mathbb{C}}\ |\alpha_\ell| + \frac{\zeta\|\mathbf{a}_\ell\|^2}{2}\left(|\alpha_\ell|^2 - \frac{2\Re\langle \alpha_\ell \mathbf{a}_\ell, \mathbf{r}_{-\ell} \rangle}{\|\mathbf{a}_\ell\|^2}\right),$$

whose minimizer is complex soft-thresholding [29]

$$\alpha_\ell = \frac{\mathsf{S}_{\zeta^{-1}}(\langle \mathbf{a}_\ell, \mathbf{r}_{-\ell} \rangle)}{\|\mathbf{a}_\ell\|^2}, \qquad \mathsf{S}_\lambda(z) \triangleq \left(1 - \frac{\lambda}{|z|}\right)_+ z. \tag{15}$$

3) Continuous atom refinement: When $\mathcal{A}$ is indexed by a continuous parameter vector $\boldsymbol{\theta}$, each active atom is refined by locally updating $\boldsymbol{\theta}_\ell$ to maximize $|\langle \mathbf{a}(\boldsymbol{\theta}_\ell), \mathbf{r}_{-\ell} \rangle|$. [5]

IV. BEAST: Beamspace CD-AST Algorithm

The CD-AST primitives in Section III are formulated over a finite-dimensional Hilbert space and therefore apply equally to vector, matrix, or tensor-valued observations once

[3] In our finite-dimensional setting, the Hilbert-space structure provides inner product and adjoint identities required for the dual-correlation arguments.

[4] In practice, the maximization in (13) is performed numerically over a continuous parameter space and is therefore implemented approximately (e.g., via coarse initialization followed by local refinement), yielding an ε-accurate maximizer. This numerical approximation does not affect the underlying optimality conditions or the definition of the dual certificate.

[5] In practice, this is done by applying gradient or Newton-type updates to a smooth local surrogate of the correlation objective (e.g., its squared magnitude), which shares the same local maximizers; see [29].

the corresponding inner product is specified. In the present beamformed setting, however, the relevant atomic set is not the original element-space Fourier atomic set, but its image under the linear measurement operator. Thus, extending CD-AST to beamspace requires: *(i)* defining the pushed-forward atoms induced by the analog precoder/combiner, *(ii)* deriving the corresponding measurement-domain dual-certificate oracle and shrinkage update, and *(iii)* implementing this oracle tractably. This formulation does not require $\mathbf{F}$ or $\mathbf{W}$ to have DFT, orthogonal, or shift-invariant structure, although their choice affects the conditioning and the correlation landscape of the resulting beamspace atoms.

A. Beamspace Specialization of CD-AST

1) Beamspace Atom Selection and Shrinkage: Let the physical (element-space) channel tensor be $\mathbf{X} \in \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{BS}} \times K}$ with entries $[\mathbf{X}]_{u,b,k} = [\mathbf{H}(f_k)]_{u,b}$. Consider the sensed atomic-norm regression

$$\min_{\mathbf{X} \in \mathcal{H}_{el}} \ \|\mathbf{X}\|_{\mathcal{A}} + \frac{\zeta}{2} \|\mathbf{Y} - \mathsf{F}(\mathbf{X})\|_{\mathcal{Y}}^2, \tag{16}$$

where $\mathsf{F} : \mathcal{H}_{el} \to \mathcal{H}_{bs}$ is linear in the present finite-dimensional setting (and hence admits an adjoint; see, e.g., [40, Sec. 3.9-4]). In beamformed MIMO–OFDM, we set $\mathcal{H}_{el} = \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{BS}} \times K}$ and $\mathcal{H}_{bs} = \mathbb{C}^{M \times S \times K}$, and define

$$[\mathsf{F}(\mathbf{X})]_{m,s,k} \triangleq \mathbf{w}_m^{\mathrm{H}} \, [\mathbf{X}]_{:,:,k} \, \mathbf{f}_s, \tag{17}$$

so that $\mathbf{Y} = \mathsf{F}(\mathbf{X}) + \mathbf{N}$ with $[\mathbf{Y}]_{m,s,k} = \bar{y}_{m,s,k}/\bar{p}_{m,s,k}$. We equip $\mathcal{Y}$ with the Frobenius inner product

$$\langle \mathbf{U}, \mathbf{V} \rangle \triangleq \sum\nolimits_{m,s,k} [\mathbf{U}]_{m,s,k}^* [\mathbf{V}]_{m,s,k}.$$

a) Pushed-forward Atom: For a single path with parameters $\boldsymbol{\varrho} = (\phi, \boldsymbol{\theta}, \tau)$, define the physical rank-one atom

$$\mathbf{A}(\boldsymbol{\varrho}) \triangleq \mathbf{a}_{\mathrm{UE}}(\phi) \circ \mathbf{a}_{\mathrm{BS}}(\boldsymbol{\theta}) \circ \mathbf{a}_{\mathrm{SC}}(\tau) \in \mathcal{H}, \tag{18}$$

where $\mathbf{a}_{\mathrm{SC}}(\tau) = \mathbf{a}_K(-2\pi \Delta f \, \tau)$. The corresponding pushed-forward atom in the measurement (beamspace) domain is

$$\mathbf{B}(\boldsymbol{\varrho}) \triangleq \mathsf{F}(\mathbf{A}(\boldsymbol{\varrho})) \in \mathcal{Y}, \tag{19}$$

with $[\mathbf{B}(\phi, \boldsymbol{\theta}, \tau)]_{m,s,k} = [\mathbf{a}_{\mathrm{SC}}(\tau)]_k \left(\mathbf{w}_m^{\mathrm{H}} \mathbf{a}_{\mathrm{UE}}(\phi)\right) \left(\mathbf{a}_{\mathrm{BS}}^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{f}_s\right)$. Defining beamspace steering vectors $\mathbf{g}_{\mathrm{UE}}(\phi) \triangleq \mathbf{W}^{\mathrm{H}} \mathbf{a}_{\mathrm{UE}}(\phi) \in \mathbb{C}^M$, and $\mathbf{g}_{\mathrm{BS}}(\boldsymbol{\theta}) \triangleq \mathbf{F}^{\mathrm{H}} \mathbf{a}_{\mathrm{BS}}^*(\boldsymbol{\theta}) \in \mathbb{C}^S$, (19) can be written as $\mathbf{B}(\phi, \boldsymbol{\theta}, \tau) = \mathbf{g}_{\mathrm{UE}}(\phi) \circ \mathbf{g}_{\mathrm{BS}}^*(\boldsymbol{\theta}) \circ \mathbf{a}_{\mathrm{SC}}(\tau)$.

b) Atom-selection Oracle and Stopping Test: Define the measurement residual $\mathbf{R} \in \mathcal{Y}$, as

$$\mathbf{R} \triangleq \mathbf{Y} - \sum\nolimits_{\ell=0}^{L-1} \alpha_\ell \mathbf{B}(\boldsymbol{\varrho}_\ell). \tag{20}$$

At any iteration with residual $\mathbf{R}$, the dual-certificate oracle for (16) can be evaluated directly in measurement space as

$$(\hat{\phi}, \hat{\boldsymbol{\theta}}, \hat{\tau}) \in \arg\max_{\phi, \boldsymbol{\theta}, \tau} \ |\langle \mathbf{B}(\phi, \boldsymbol{\theta}, \tau), \mathbf{R} \rangle| \,, \tag{21}$$

which is equivalent to an adjoint-based oracle by the defining identity $\langle \mathsf{F}(\mathbf{a}), \mathbf{R} \rangle_{\mathcal{Y}} = \langle \mathbf{a}, \mathsf{F}^*(\mathbf{R}) \rangle_{\mathcal{H}}$ [40, Sec. 3.9-4]. Moreover, the global stopping test becomes

$$\sup_{\phi, \theta, \tau} \left| \langle \mathbf{B}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau), \mathbf{R} \rangle \right| \leq \zeta^{-1}. \tag{22}$$

Using (19), the correlation factorizes as

$$\langle \mathbf{B}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau), \mathbf{R} \rangle = \sum_{k=0}^{K-1} [\mathbf{a}_{\mathrm{SC}}(\tau)]_k^* \, \mathbf{g}_{\mathrm{UE}}^{\mathrm{H}}(\phi) \, \mathbf{R}_k \, \mathbf{g}_{\mathrm{BS}}(\boldsymbol{\theta}), \tag{23}$$

where $\mathbf{R}_k \in \mathbb{C}^{M \times S}$ is the kth frontal slice of $\mathbf{R}$ (i.e., $[\mathbf{R}_k]_{m,s} = r_{m,s,k}$).

Algorithm 1 BEAST : Beamspace CD-AST

Input: $\mathbf{Y}$, ζ, tolerance ε, max iterations T
Initialize residual $\mathbf{R} \leftarrow \mathbf{Y}$, tuple set $\mathcal{S} \leftarrow \emptyset$
for $t = 1, \ldots, T$ **do**
  $\boldsymbol{\varrho}^\star \leftarrow \mathsf{Oracle}(\mathbf{R})$ (Alg. 2)
  if $|\langle \mathbf{B}(\boldsymbol{\varrho}^\star), \mathbf{R} \rangle| \leq \zeta^{-1}$ **then**
    break ▷ terminate outer loop and return current $\mathcal{S}$
  end if
  $\alpha \leftarrow \frac{\mathsf{S}_{\zeta^{-1}}(\langle \mathbf{B}(\boldsymbol{\varrho}^\star), \mathbf{R} \rangle)}{\|\mathbf{B}(\boldsymbol{\varrho}^\star)\|^2}$
  $\mathbf{R} \leftarrow \mathbf{R} - \alpha \mathbf{B}(\boldsymbol{\varrho}^\star)$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{(\alpha, \boldsymbol{\varrho}^\star)\}$
  for all $(\alpha_\ell, \boldsymbol{\varrho}_\ell) \in \mathcal{S}$ **do**
    $\mathbf{R}_{-\ell} \leftarrow \mathbf{R} + \alpha_\ell \mathbf{B}(\boldsymbol{\varrho}_\ell)$
    $\boldsymbol{\varrho}_\ell \leftarrow \mathsf{Oracle}(\mathbf{R}_{-\ell})$
    $\mathbf{R} \leftarrow \mathbf{R}_{-\ell} - \alpha_\ell \mathbf{B}(\boldsymbol{\varrho}_\ell)$
  end for
end for
Output: $\mathcal{S}$

c) Sensed Shrinkage: Fix an active parameter $\boldsymbol{\varrho}_\ell$ with atom $\mathbf{B}_\ell = \mathbf{B}(\boldsymbol{\varrho}_\ell)$ and define $\mathbf{R}_{-\ell} \triangleq \mathbf{Y} - \sum_{j \neq \ell} \alpha_j \mathbf{B}(\boldsymbol{\varrho}_j)$. Updating α_ℓ solves

$$\min_{\alpha \in \mathbb{C}} \ |\alpha| + \frac{\zeta}{2} \|\mathbf{R}_{-\ell} - \alpha \mathbf{B}_\ell\|_{\mathcal{Y}}^2, \tag{24}$$

with unique minimizer given by complex soft-thresholding

$$\alpha_\ell = \frac{\mathsf{S}_{\zeta^{-1}}(\langle \mathbf{B}_\ell, \mathbf{R}_{-\ell} \rangle_{\mathcal{Y}})}{\|\mathbf{B}_\ell\|_{\mathcal{Y}}^2}. \tag{25}$$

From the outer-product of $\mathbf{B}(\phi, \theta, \tau)$, its norm factorizes as

$$\begin{aligned} \|\mathbf{B}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)\|_{\mathcal{Y}}^2 &= \|\mathbf{g}_{\mathrm{UE}}(\phi)\|_2^2 \, \|\mathbf{g}_{\mathrm{BS}}(\boldsymbol{\theta})\|_2^2 \, \|\mathbf{a}_{\mathrm{SC}}(\tau)\|_2^2 \\ &= K \, \|\mathbf{W}^{\mathrm{H}} \mathbf{a}_{\mathrm{UE}}(\phi)\|_2^2 \, \|\mathbf{F}^{\mathrm{T}} \mathbf{a}_{\mathrm{BS}}(\boldsymbol{\theta})\|_2^2. \end{aligned} \tag{26}$$

2) Beamspace Regression Interpretation: Substituting the finite expansion $\mathbf{X} = \sum_{\ell=0}^{L-1} \alpha_\ell \mathbf{A}(\boldsymbol{\varrho}_\ell)$ into (16) and $\mathbf{B}(\boldsymbol{\varrho}) = \mathsf{F}(\mathbf{A}(\boldsymbol{\varrho}))$ yields the equivalent beamspace regression

$$\min_{L, \{\alpha_\ell\}, \{\boldsymbol{\varrho}_\ell\}} \ \sum\nolimits_{\ell=0}^{L-1} |\alpha_\ell| + \frac{\zeta}{2} \Big\| \mathbf{Y} - \sum\nolimits_{\ell=0}^{L-1} \alpha_\ell \mathbf{B}(\boldsymbol{\varrho}_\ell) \Big\|_{\mathcal{Y}}^2. \tag{27}$$

Hence, CD-AST operates on the pushed-forward atomic set $\mathcal{B} = \{\mathbf{B}(\boldsymbol{\varrho}) : \boldsymbol{\varrho} \in \Omega\} \subset \mathcal{Y}$, while the physical estimate is recovered as $\widehat{\mathbf{X}} = \sum_\ell \hat{\alpha}_\ell \mathbf{A}(\hat{\boldsymbol{\varrho}}_\ell)$. With residual $\mathbf{R}$ in (20), the stopping rule (22) is the dual-feasibility test $\|\mathbf{R}\|_{\mathcal{B}}^* \leq \zeta^{-1}$. The remaining challenge is then the evaluation of the oracle in (21), requiring continuous optimization over the path parameters $\boldsymbol{\varrho}$. Alg. 1 summarizes the procedure.

B. Oracle Realization

The repeated solution of (21) entails continuous optimization over the 5-D delay–angle parameter space. To reduce the computational burden, BEAST realizes this formulation by a tractable numerical realization of the beamspace oracle, exploiting the separable correlation structure in (23).

1) Structured Decoupled Initialization: We first isolate the delay dimension by correlating the residual with the subcarrier steering vector. For a candidate delay τ, define the delay-contracted residual $\mathbf{Z}_{\mathbf{R}}(\tau) \in \mathbb{C}^{M \times S}$ defined as

$$[\mathbf{Z}_{\mathbf{R}}(\tau)]_{m,s} \triangleq \sum\nolimits_k [\mathbf{a}_{\mathrm{SC}}(\tau)]_k^* \, r_{m,s,k}. \tag{28}$$

The corresponding non-coherently combined delay score is

$$g_{\mathbf{R}}(\tau) \triangleq \|\mathbf{Z}_{\mathbf{R}}(\tau)\|_F^2 = \sum\nolimits_{m,s} \left|[\mathbf{Z}_{\mathbf{R}}(\tau)]_{m,s}\right|^2. \tag{29}$$

This score can be evaluated efficiently by computing, for each beam pair (m, s), the contraction over the subcarrier index via an oversampled FFT, and then summing the resulting squared magnitudes over beam pairs. Its maximizer provides the initial delay estimate $\tau^{(0)}$, where superscript (0) denotes initialization. With $\tau^{(0)}$ fixed, we form the contracted residual matrix $\mathbf{Z}_{\mathbf{R}}(\tau^{(0)})$, after which the remaining dependence is purely angular. The UE-side angle is initialized by maximizing

$$h_{\mathbf{R}}(\boldsymbol{\phi}; \tau^{(0)}) = \sum\nolimits_{s} \left|\mathbf{g}_{\mathrm{UE}}^{\mathsf{H}}(\boldsymbol{\phi})\, [\mathbf{Z}_{\mathbf{R}}(\tau^{(0)})]_{:,s}\right|^2, \tag{30}$$

whose maximizer yields $\boldsymbol{\phi}^{(0)}$. Finally, with $\tau^{(0)}$ and $\boldsymbol{\phi}^{(0)}$ fixed, the BS-side angle is initialized from

$$f_{\mathbf{R}}(\boldsymbol{\theta}; \boldsymbol{\phi}^{(0)}, \tau^{(0)}) = \left|\mathbf{g}_{\mathrm{UE}}^{\mathsf{H}}(\boldsymbol{\phi}^{(0)})\, \mathbf{Z}_{\mathbf{R}}(\tau^{(0)})\, \mathbf{g}_{\mathrm{BS}}(\boldsymbol{\theta})\right|^2, \tag{31}$$

and the corresponding maximizer yields $\boldsymbol{\theta}^{(0)}$.

2) Local Continuous Refinement: The sequential scan above provides a coarse initialization $(\tau^{(0)}, \boldsymbol{\phi}^{(0)}, \boldsymbol{\theta}^{(0)})$ for the continuous oracle in (21). Starting from this point, we refine the selected path parameters by cyclic Newton updates applied to the smooth surrogate

$$\mathsf{C}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau) \triangleq \left|\langle \mathbf{B}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau), \mathbf{R}\rangle\right|^2, \tag{32}$$

which has the same local maximizers as $|\langle \mathbf{B}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau), \mathbf{R}\rangle|$.

At iteration t, refinement proceeds by cyclic block-coordinate Newton updates over $\boldsymbol{\mu} \in \{\boldsymbol{\phi}, \boldsymbol{\theta}, \tau\}$, with the remaining blocks held fixed:

$$\boldsymbol{\mu}^{(t+1)} = \boldsymbol{\mu}^{(t)} - \left[\nabla_{\boldsymbol{\mu}}^2 \mathsf{C}\right]^{-1} \nabla_{\boldsymbol{\mu}} \mathsf{C}\Big|_{(\boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)}, \tau^{(t)})}. \tag{33}$$

For the angular blocks $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, this gives 2×2 Newton updates; for the scalar delay block τ, it reduces to the standard scalar Newton step. The required gradients and Hessians follow from differentiating (23) with respect to the corresponding steering vectors. Since the pushed-forward atoms alter the coherence, curvature, and norm profile of the correlation surface, we do not rely on the global convergence guarantees of the identity-observation CD-AST analysis in [29]. The Newton steps are used here as a local numerical refinement of the oracle candidate, and each candidate update is accepted only if it decreases the beamspace regression objective in (27); otherwise the previous tuple is retained. Alg. 2 summarizes the resulting oracle realization.

C. Implementation Details

1) Numerical Stabilization: To improve numerical conditioning, we normalize $\mathbf{Y}$ to unit Frobenius norm and rescale ζ accordingly, since individual entries of the beamspace observation tensor can be very small in magnitude. We also use a relative stopping tolerance proportional to the current objective value, discard tuples with $|\alpha| < \varepsilon$, and reject refinement updates that increase the beamspace regression objective in (27).

2) Choice of ζ: As in CD-AST, ζ is chosen from the dual threshold, i.e., $\sup_{\boldsymbol{\varrho}} \left|\langle \mathbf{B}(\boldsymbol{\varrho}), \mathbf{N}\rangle\right| \le \zeta^{-1}$. For Gaussian noise, the correlation $\langle \mathbf{B}(\boldsymbol{\varrho}), \mathbf{N}\rangle$ has variance proportional to $\sigma\|\mathbf{B}(\boldsymbol{\varrho})\|_{\mathcal{Y}}^2$, as $\|\mathbf{B}(\boldsymbol{\varrho})\|_{\mathcal{Y}}^2$ expanded in (26). Under unit-norm beams and the steering-vector normalization adopted here, we use the conservative bound $\|\mathbf{B}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)\|_{\mathcal{Y}}^2 \le KMSN_{\mathrm{UE}}N_{\mathrm{BS}}$. Hence, by analogy with the element-space scaling in CD-AST,

$$\zeta \approx \frac{c_\zeta}{\sqrt{\sigma_n^2 / P_t KMSN_{\mathrm{UE}}N_{\mathrm{BS}}}},$$

where c_ζ is a numerical constant.

Algorithm 2 Oracle Realization

Input: $\mathbf{R}$, Q_τ, $\{\boldsymbol{\phi}_i\}_{i=1}^{G_\phi}$, $\{\boldsymbol{\theta}_j\}_{j=1}^{G_\theta}$, I
Output: $\boldsymbol{\varrho} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)$
for τ_q on the $Q_\tau K$ delay grid **do**
  $g_{\mathbf{R}}(\tau_q) \leftarrow \sum_{m,s} \left|\sum_k [\mathbf{a}_{\mathrm{SC}}(\tau_q)]_k^*\, r_{m,s,k}\right|^2$
end for
$\tau^{(0)} \leftarrow \arg\max_{\tau_q} g_{\mathbf{R}}(\tau_q)$
$[\mathbf{Z}_{\mathbf{R}}(\tau^{(0)})]_{m,s} \leftarrow \sum_k [\mathbf{a}_{\mathrm{SC}}(\tau^{(0)})]_k^*\, r_{m,s,k}$ ▷ $\mathbf{Z}_{\mathbf{R}} \in \mathbb{C}^{M \times S}$
for $i = 1, \ldots, G_\phi$ **do**
  $h_{\mathbf{R}}(\boldsymbol{\phi}_i; \tau^{(0)}) \leftarrow \sum_s \left|\mathbf{g}_{\mathrm{UE}}^{\mathsf{H}}(\boldsymbol{\phi}_i)\, [\mathbf{Z}_{\mathbf{R}}(\tau^{(0)})]_{:,s}\right|^2$
end for
$\boldsymbol{\phi}^{(0)} \leftarrow \arg\max_{\boldsymbol{\phi}_i} h_{\mathbf{R}}(\boldsymbol{\phi}_i; \tau^{(0)})$
for $j = 1, \ldots, G_\theta$ **do**
  $f_{\mathbf{R}}(\boldsymbol{\theta}_j; \boldsymbol{\phi}^{(0)}, \tau^{(0)}) \leftarrow \left|\mathbf{g}_{\mathrm{UE}}^{\mathsf{H}}(\boldsymbol{\phi}^{(0)})\, \mathbf{Z}_{\mathbf{R}}(\tau^{(0)})\, \mathbf{g}_{\mathrm{BS}}(\boldsymbol{\theta}_j)\right|^2$
end for
$\boldsymbol{\theta}^{(0)} \leftarrow \arg\max_{\boldsymbol{\theta}_j} f_{\mathbf{R}}(\boldsymbol{\theta}_j; \boldsymbol{\phi}^{(0)}, \tau^{(0)})$
$(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau) \leftarrow (\boldsymbol{\phi}^{(0)}, \boldsymbol{\theta}^{(0)}, \tau^{(0)})$
for $t = 1, \ldots, I$ **do**
  for $\boldsymbol{\mu} \in \{\tau, \boldsymbol{\phi}, \boldsymbol{\theta}\}$ **do**
    $\mathbf{H}_{\boldsymbol{\mu}} \leftarrow \nabla_{\boldsymbol{\mu}}^2 \mathsf{C}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)\big|_{(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)}$
    $\mathbf{g}_{\boldsymbol{\mu}} \leftarrow \nabla_{\boldsymbol{\mu}} \mathsf{C}(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)\big|_{(\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)}$
    $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} - \mathbf{H}_{\boldsymbol{\mu}}^{-1} \mathbf{g}_{\boldsymbol{\mu}}$
  end for
end for
$\boldsymbol{\varrho} \leftarrow (\boldsymbol{\phi}, \boldsymbol{\theta}, \tau)$

3) Computational Remark: Let Q_τ denote the delay oversampling factor (FFT grid size $Q_\tau K$), and let G_ϕ and G_θ denote the number of grid points used in the coarse AoA and AoD scans. One oracle evaluation (coarse initialization + local refinement) has the following cost. *(i) Delay scan:* Evaluating $g_{\mathbf{R}}(\tau)$ for all τ on an oversampled grid can be implemented by computing the $Q_\tau K$-point FFT across the subcarrier dimension for each beam pair (m, s). Hence,

$$T_\tau = \mathcal{O}\big(MS\,(Q_\tau K)\log(Q_\tau K)\big).$$

(ii) Angle scan: After fixing $\hat{\tau}^{(0)}$, we form the contracted residual $\mathbf{Z} \in \mathbb{C}^{M \times S}$ with entries $z_{m,s} \triangleq \sum_k [\mathbf{a}_{\mathrm{SC}}(\hat{\tau}^{(0)})]_k^*\, r_{m,s,k}$. This contraction costs $\mathcal{O}(KMS)$ once and is reused in both angle scans. Each evaluation of the AoA score $h_{\mathbf{R}}(\hat{\tau}^{(0)}, \boldsymbol{\phi})$ reduces to a matrix–vector product $\mathbf{g}_{\mathrm{UE}}^{\mathrm{H}}(\boldsymbol{\phi})\mathbf{Z} \in \mathbb{C}^{1 \times S}$, costing $\mathcal{O}(MS)$. Thus, $T_\phi = \mathcal{O}\big(KMS + G_\phi MS\big)$. After selecting $\hat{\boldsymbol{\phi}}^{(0)}$, we precompute $\mathbf{v}^{\mathsf{H}} = \mathbf{g}_{\mathrm{UE}}^{\mathsf{H}}(\hat{\boldsymbol{\phi}}^{(0)})\mathbf{Z}$ at cost $\mathcal{O}(MS)$, so each AoD evaluation reduces to $\mathbf{v}^{\mathsf{H}}\mathbf{g}_{\mathrm{BS}}(\boldsymbol{\theta})$ with cost $\mathcal{O}(S)$. Hence, $T_\theta = \mathcal{O}(MS + G_\theta S) = \mathcal{O}(G_\theta S)$.

(iii) Local refinement: Let I denote the number of cyclic block-coordinate Newton sweeps. In each sweep, the delay block τ and the angular blocks $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are up-

dated once. For fixed $(\boldsymbol{\phi}, \boldsymbol{\theta})$, forming the scalar sequence $\mathbf{g}_{\mathrm{UE}}^{\mathsf{H}}(\boldsymbol{\phi})[\mathbf{R}]_{:,:,k}\mathbf{g}_{\mathrm{BS}}(\boldsymbol{\theta})\ \forall k$, costs $\mathcal{O}(KMS)$, after which the gradient and Hessian with respect to τ cost only $\mathcal{O}(K)$. Similarly, for fixed $(\boldsymbol{\theta}, \tau)$ or $(\boldsymbol{\phi}, \tau)$, forming the contracted vectors required for the $\boldsymbol{\phi}$- and $\boldsymbol{\theta}$-updates also costs $\mathcal{O}(KMS)$, while the subsequent 2×2 Hessian updates contribute only lower-order terms. Hence, one full Newton sweep costs $\mathcal{O}(KMS)$ up to constant factors, and $T_{\mathrm{Newt}} = \mathcal{O}(IKMS)$.

Combining all operations, one oracle call scales as

$$T_{\mathrm{oracle}} = T_\tau + T_\phi + T_\theta + T_{\mathrm{Newt}}, \tag{34}$$

where the dominant term in typical beam–frequency regimes is the oversampled delay scan $MS\,(Q_\tau K)\log(Q_\tau K)$.

Note that the overall runtime of beamspace CD-AST scales approximately linearly with the number of active tuples L, since each outer iteration performs one oracle call and $\mathcal{O}(L)$ shrinkage/refinement updates.

V. EAST : Delay-First Decoupling with Per-slice 4D CD-AST Refinement

Although BEAST makes the 5D beamspace oracle tractable, it still requires repeated joint delay–angle searches. In wideband mmWave OFDM systems, a batch delay-first decoupling strategy is natural for two reasons: typically $K \gg M, S$, so repeated delay searches dominate the complexity, and the large occupied bandwidth $K\Delta f$ provides fine delay resolution, so dominant paths are often well separated in delay. This motivates estimating the delay support once, removing the delay dimension, and then performing only per-slice angular refinement. EAST combines FLEX-style delay estimation with narrowband beamspace CD-AST for the resulting slices.

A. Delay-first Decoupling (A Background on FLEX [9])

Define the mode-3 unfolding of $\mathbf{Y} \in \mathbb{C}^{M\times S\times K}$ as $\mathbf{Y}_{(3)} \in \mathbb{C}^{K\times(MS)}$, obtained by stacking the MS beam-pair snapshots as columns. Under an L-path specular model, $\mathbf{Y}_{(3)}$ admits the multiple-measurement-vector (MMV) representation [9]

$$\mathbf{Y}_{(3)} = \sum\nolimits_{\ell=0}^{L-1} \alpha_\ell \mathbf{a}_{\mathrm{SC}}(\tau_\ell)\mathbf{v}_\ell^{\mathsf{T}} + \mathbf{N}_{(3)}, \tag{35}$$

where $\mathbf{v}_\ell \in \mathbb{C}^{MS}$ collects the snapshot-dependent angular response of path ℓ, and $\mathbf{N}_{(3)}$ is the unfolded noise matrix.

FLEX [9] estimates the delay support $\{\widehat{\tau}_\ell\}_{\ell=0}^{\widehat{L}-1}$ by applying the delay contraction $\mathbf{Z}_{\mathbf{Y}}(\tau) \in \mathbb{C}^{M\times S}$ to the observation tensor $\mathbf{Y}$, i.e., $[\mathbf{Z}_{\mathbf{Y}}(\tau)]_{m,s} \triangleq \sum_k [\mathbf{a}_{\mathrm{SC}}(\tau)]_k^*\, y_{m,s,k}$, and evaluating the associated delay score $g_{\mathbf{Y}}(\tau) \triangleq \|\mathbf{Z}_{\mathbf{Y}}(\tau)\|_F^2$. The dominant peaks of $g_{\mathbf{Y}}(\tau)$ provide the estimated delay support $\{\widehat{\tau}_\ell\}_{\ell=0}^{\widehat{L}-1}$.

Further, $\mathbf{Y}_{\mathrm{sub}} \in \mathbb{C}^{M\times S\times \widehat{L}}$ is formed slice-wise as

$$[\mathbf{Y}_{\mathrm{sub}}]_{:,:,\ell} \triangleq \mathbf{Z}_{\mathbf{Y}}(\widehat{\tau}_\ell), \quad \ell = 0, \ldots, \widehat{L}-1. \tag{36}$$

When the delay support is accurately recovered, $\mathbf{Y}_{\mathrm{sub}}^{:,:,\ell}$ is dominated by the ℓ-th delay component (up to leakage due to imperfect delay separation), so the corresponding angular parameters can be estimated independently per slice.

B. EAST: ESPRIT-Based Delay Estimation and 4D BEAST

1) Delay Estimation and Decoupling: Building on the MMV representation in (35), EAST estimates the delay support $\{\widehat{\tau}_\ell\}_{\ell=0}^{\widehat{L}-1}$ via an ESPRIT-type[6] shift-invariance method applied to $\mathbf{Y}_{(3)}$.[7] The resulting delay estimates are then used to form the delay-separated tensor $\mathbf{Y}_{\mathrm{sub}}$ as in (36), yielding $\widehat{L}$ spatial slices. For each ℓ, conditioned on $\tau = \widehat{\tau}_\ell$, the slice $[\mathbf{Y}_{\mathrm{sub}}]_{:,:,\ell} \in \mathbb{C}^{M\times S}$ is modeled as a narrowband beamspace channel and processed using a 4D specialization of beamspace CD-AST. After the one-shot delay estimation stage, no further delay search is performed during the angular refinement.

2) Per-slice 4D CD-AST Formulation: For each slice ℓ, define the narrowband atomic set

$$\mathcal{B}_{\mathrm{NB}} = \Big\{\mathbf{B}_{\mathrm{NB}}(\boldsymbol{\phi}, \boldsymbol{\theta}) \triangleq \mathbf{g}_{\mathrm{UE}}(\boldsymbol{\phi})\mathbf{g}_{\mathrm{BS}}^{\mathsf{H}}(\boldsymbol{\theta}) : (\boldsymbol{\phi}, \boldsymbol{\theta}) \in \Omega_\phi \times \Omega_\theta\Big\}.$$

The corresponding per-slice regression problem is

$$\min_{\{\alpha_r, \boldsymbol{\phi}_r, \boldsymbol{\theta}_r\}} \sum_r |\alpha_r| + \frac{\zeta}{2}\left\|[\mathbf{Y}_{\mathrm{sub}}]_{:,:,\ell} - \sum_r \alpha_r \mathbf{g}_{\mathrm{UE}}(\boldsymbol{\phi}_r)\mathbf{g}_{\mathrm{BS}}^{\mathsf{H}}(\boldsymbol{\theta}_r)\right\|_F^2.$$

The oracle and shrinkage steps are inherited from (21)–(25), restricted to the angular variables $(\boldsymbol{\phi}, \boldsymbol{\theta})$. Thus, each delay-separated slice is refined only in the 4D angular domain. EAST does not require an explicit delay–angle pairing step, since each slice is indexed by a specific estimated delay $\widehat{\tau}_\ell$, and angular components extracted from that slice inherit this delay. The complete procedure is summarized in Alg. 3.

3) Computational Complexity of EAST: We summarize the dominant arithmetic costs of EAST by stage (See Alg. 3).

Stage I: Delay-support estimation: We compute the rank-$\widehat{L}$ signal subspace of $\mathbf{Y}_{(3)} \in \mathbb{C}^{K\times(MS)}$ via QR-singular value decomposition (SVD) [15], [42]. For a $K \times (MS)$ matrix with $K \geq 2MS$, the QR-SVD implementation in [42] has arithmetic cost $12K(MS)^2 + 48(MS)^3$ flops. Hence, the dominant scaling is $T_{\mathrm{SVD}} = \mathcal{O}\big(K(MS)^2\big)$. The ESPRIT least-squares operates on $(K-1)\times\widehat{L}$ matrices and costs $\mathcal{O}(K\widehat{L}^2)$, then an eigen-decomposition of a $\widehat{L}\times\widehat{L}$ matrix costing $\mathcal{O}(\widehat{L}^3)$.

Stage II: Delay decoupling: With $\widehat{L}$ delay contractions at cost $\mathcal{O}(KMS)$ each, the total cost is $T_{\mathrm{dec}} = \mathcal{O}(KMS\widehat{L})$.

Stage III: Per-slice 4D CD-AST: Each slice $[\mathbf{Y}_{\mathrm{sub}}]_{:,:,\ell}$ is processed using 4D beamspace CD-AST restricted to $(\boldsymbol{\phi}, \boldsymbol{\theta})$. Let G_ϕ and G_θ denote the angular grid sizes for coarse initialization, and let I_{NB} denote the number of Newton iterations per slice. For one 4D oracle call, the coarse scans cost $\mathcal{O}(G_\phi MS + G_\theta S)$, and local refinement costs $\mathcal{O}(I_{\mathrm{NB}}MS)$. If each slice extracts at most $R_{\max}$ components, the Stage III cost is $T_{\mathrm{4D}} = \mathcal{O}\big(\widehat{L}R_{\max}(G_\phi MS + G_\theta S + I_{\mathrm{NB}}MS)\big)$.

Overall complexity: Collecting terms, the overall complexity of EAST scales as $T_{\mathrm{EAST}} = T_{\mathrm{SVD}} + T_{\mathrm{dec}} + T_{\mathrm{4D}}$. In the typical regime $MS \ll K$ and moderate $\widehat{L}, R_{\max}$, the cost of EAST is dominated by the subspace extraction,

$$T_{\mathrm{EAST}} = \mathcal{O}\big(K(MS)^2\big), \tag{37}$$

[6]Unlike BEAST, which uses FFT-based delay peak detection within a successive pipeline, EAST estimates the delay support once in a batch manner and reuses it across all delay-separated slices. We therefore adopt a super-resolution subspace estimator, since FFT-based batch delay estimation remains resolution-limited without successive component removal.

[7]If the number of paths is unknown, $\widehat{L}$ can be selected via a standard model-order criterion (e.g., information-theoretic tests [41]) or by retaining the dominant delay components whose energy exceeds a fixed fraction of the strongest peak (e.g., relative to the LoS component). In single-BS positioning scenarios, at least two delays are retained to enable geometric reconstruction.

Algorithm 3 EAST: Delay-first decoupling + 4D BEAST

Input: $\mathbf{Y} \in \mathbb{C}^{M\times S\times K}$, $\widehat{L}$, ζ, 4D-oracle settings
Output: $\mathcal{S} = \{(\widehat{\alpha}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}, \widehat{\tau})\}$
Stage I (ESPRIT-based delay estimation)
Form $\mathbf{Y}_{(3)} \in \mathbb{C}^{K\times(MS)}$
Compute $\mathbf{U}_t \in \mathbb{C}^{K\times\widehat{L}}$ (rank-$\widehat{L}$ left signal subspace)
$\mathbf{J}_1 \leftarrow [\mathbf{I}_{K-1}\ \mathbf{0}]$, $\mathbf{J}_2 \leftarrow [\mathbf{0}\ \mathbf{I}_{K-1}]$
$\boldsymbol{\Phi} \leftarrow (\mathbf{J}_2\mathbf{U}_t)^{\dagger}(\mathbf{J}_1\mathbf{U}_t)$
for $\ell = 0$ to $\widehat{L}-1$ **do**
  $\widehat{\omega}_{\tau,\ell} \leftarrow \angle\lambda_\ell(\boldsymbol{\Phi})$
  $\widehat{\tau}_\ell \leftarrow \widehat{\omega}_{\tau,\ell}/(-2\pi\Delta f)$
end for
Stage II (delay decoupling)
for $\ell = 0$ to $\widehat{L}-1$ **do**
  $[\mathbf{Y}_{\mathrm{sub}}]_{:,:,\ell} \leftarrow \mathbf{Z}_{\mathbf{Y}}(\widehat{\tau}_\ell) \in \mathbb{C}^{M\times S}$
end for
Stage III (per-slice 4D CD-AST)
$\mathcal{S} \leftarrow \emptyset$
for $\ell = 0$ to $\widehat{L}-1$ **do**
  $\mathbf{Y}_\ell \leftarrow [\mathbf{Y}_{\mathrm{sub}}]_{:,:,\ell} \in \mathbb{C}^{M\times S}$
  $\mathcal{S}_\ell \leftarrow \mathsf{BEAST}(\mathbf{Y}_\ell; \mathbf{B}_{\mathrm{NB}})$ ▷ 4D angular specialization
  for all $(\widehat{\alpha}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}) \in \mathcal{S}_\ell$ **do**
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{(\widehat{\alpha}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}, \widehat{\tau}_\ell)\}$
  end for
end for
return $\mathcal{S}$

while the delay-decoupling and per-slice refinements contribute lower-order terms. In contrast, BEAST incurs L_{out} outer iterations (oracle calls), each with an oversampled delay scan of cost $\mathcal{O}\big(MS(Q_\tau K)\log(Q_\tau K)\big)$. Comparing the dominant arithmetic terms $L_{\mathrm{out}}MS(Q_\tau K)\log(Q_\tau K)$ and $K(MS)^2$, EAST is computationally advantageous when $L_{\mathrm{out}}Q_\tau\log(Q_\tau K) \gtrsim MS$, that is, when delay searches in BEAST exceed the subspace-extraction cost in EAST.

VI. Simulation Setup and Evaluations

We evaluate the proposed channel estimation and localization framework in two complementary scenarios. A *stress-test* setting isolates individual effects by varying one parameter at a time under a controlled two-path geometry. A *realistic* setting considers clustered reflections with diffuse subpaths to assess robustness under model mismatch.

A. System and Measurement Parameters

1) OFDM and Noise Model: We adopt a 3GPP Frequency Range-2 channel bandwidth of $B_{\mathrm{ch}} = 400$ MHz with subcarrier spacing $\Delta f = 240$ kHz [43, Table 4.2]. Out of this channel, only $K = 1400$ contiguous subcarriers are assumed occupied/processed (guard bands ignored), yielding an effective occupied bandwidth $B_{\mathrm{occ}} = K\Delta f = 336$ MHz. The total transmit power over the occupied bandwidth is denoted by $P_{\mathrm{tx}} = KP_t$. Thermal noise is modeled as spatially and spectrally white with two-sided power spectral density $N_0 = -174$ dBm/Hz. The per-subcarrier noise variance is

$$\sigma_n^2 = 10^{-3}\,10^{N_0/10}\,\Delta f, \tag{38}$$

where N_0 is in dBm/Hz and 10^{-3} converts mW to W.

2) Arrays and Beam Training Dimensions: The BS and UE employ URAs with 24×12 and 12×12 elements, respectively. Beam training uses separable codebooks [34] with $S = 16\times 8$ BS beams and $M = 10\times 10$ UE beams. The resulting measurement tensor follows (4).

B. Beamforming Design Variants

We assess sensitivity to the combining structure by considering two UE combiner codebooks with BS precoder fixed.

1) Structured DFT Beam Training: Let $\mathbf{T}_1 \in \mathbb{C}^{N_{\mathrm{h}}^{\mathrm{BS}}\times M_{\mathrm{h}}^{\mathrm{BS}}}$, $\mathbf{T}_2 \in \mathbb{C}^{N_{\mathrm{v}}^{\mathrm{BS}}\times M_{\mathrm{v}}^{\mathrm{BS}}}$, $\mathbf{T}_3 \in \mathbb{C}^{N_{\mathrm{h}}^{\mathrm{UE}}\times M_{\mathrm{h}}^{\mathrm{UE}}}$, and $\mathbf{T}_4 \in \mathbb{C}^{N_{\mathrm{v}}^{\mathrm{UE}}\times M_{\mathrm{v}}^{\mathrm{UE}}}$ denote separable steering codebooks, where $\mathbf{T}_i$ are truncated DFT matrices and the M-dimensions specify the numbers of sampled beam directions along the horizontal and vertical axes. The codebooks are $\mathbf{F} = \mathbf{T}_2^* \otimes \mathbf{T}_1^*$ and $\mathbf{W} = \mathbf{T}_4 \otimes \mathbf{T}_3$.

2) Dense Gaussian Combiner: For comparison, we replace $\mathbf{W}$ by a fixed dense random matrix $\mathbf{W}_{\mathrm{G}} \in \mathbb{C}^{N_{\mathrm{UE}}\times M}$ with i.i.d. entries $\mathcal{CN}(0,1)$ (drawn once and kept fixed across trials), whose columns are normalized to unit ℓ_2 norm.

C. Algorithms, Inputs, and Metrics

Unless otherwise stated, BEAST uses $Q_\tau = 4$, $G_\phi = 48\times 48$, $G_\theta = 96\times 48$, $I = 50$ Newton steps with tolerance 10^{-5}, $T = 100$, $\varepsilon = 10^{-8}$, $\epsilon_{\mathrm{rel}} = 10^{-4}$, and $c_\zeta = 68.55$ in the dual-threshold scaling of Sec. IV-C2. For EAST, $\widehat{L} = 2$ delay components are retained, and the 4D per-slice CD-AST stage uses the same angular-grid and refinement settings.

We compare BEAST and EAST against representative FFT-based, subspace, tensor, and multidimensional decomposition baselines: FLEX [9], which performs FFT-based coarse delay search followed by ESPRIT-type refinement and sequential pairing; multidimensional ESPRIT (MDE) [16]; SVD-based matched filtering (SMF) [44]; and R-dimensional beamspace ESPRIT (RDB) [45]. Among these methods, BEAST, EAST, MDE, RDB, and SMF are gridless, while FLEX relies on an FFT delay grid. BEAST and FLEX estimate the model order intrinsically, whereas EAST, MDE, RDB, and SMF use the prescribed order $\widehat{L} = 2$. All methods except MDE return paired delay–angle tuples without external pairing.

We report Monte Carlo root mean squared error (RMSE) for angles, delays, position, and clock bias. Along with empirical performance, we report the root CRB (RCRB) with respect to the per-path parameter vector

$$\boldsymbol{\theta}_{\ell,r} = \big[\theta_{\ell,r}^{\mathrm{az}}, \theta_{\ell,r}^{\mathrm{el}}, \phi_{\ell,r}^{\mathrm{az}}, \phi_{\ell,r}^{\mathrm{el}}, \tau_{\ell,r}, \Re\{\alpha_{\ell,r}\}, \Im\{\alpha_{\ell,r}\}\big]^{\mathsf{T}}$$

under the measurement model (4), and the corresponding position error bound (PEB). The CRB is obtained as the inverse FIM [46], [47]. The PEB is computed by transforming the channel-parameter CRB through the Jacobian of the mapping from channel parameters to $(\mathbf{p}_{\mathrm{UE}}, \beta)$ [48, App. A]. In the realistic scenario (Sec. VI-E), the bound is computed using LoS and seven strongest NLoS components, since including all diffuse paths leads to an ill-conditioned FIM.

D. Stress-Test Suite (Controlled Two-Path Geometry)

To isolate estimator behavior under controlled conditions, we adopt a deterministic two-path geometric channel consisting of a LoS component and a single-bounce specular reflection, without diffuse multipath. This setting enables direct attribution of performance differences to algorithmic properties (e.g., SNR threshold, resolution, and accuracy). Unless stated otherwise, the BS and UE positions are fixed at $\mathbf{p}_{\mathrm{BS}} = [0, 0, 6]^{\mathsf{T}}$ m and $\mathbf{p}_{\mathrm{UE}} = [38, -5, 1.5]^{\mathsf{T}}$ m, with UE orientation given by yaw–pitch–roll $(30^\circ, -10^\circ, 5^\circ)$ (ZYX convention) and clock bias $\beta = 15$ ns.

1) SNR Sweep: The specular interaction point is fixed at $\mathbf{p}_0 = [20.75, 4.54, 10.80]^{\mathsf{T}}$ m with reflection coefficient $\Gamma = 1$. The effective SNR is varied by scaling the noiseless measurement tensor by $\sqrt{P_{\mathrm{tx}}}$ while keeping σ_n^2 fixed, and is

$$\mathrm{SNR} = \frac{P_t}{\sigma_n^2} = \frac{P_{\mathrm{tx}}}{K\sigma_n^2}. \tag{39}$$

Each point averages $N_{\mathrm{MC}} = 50$ independent noise realizations with random LoS phase. Fig. 1 shows RMSE versus SNR for LoS AoD, LoS range, and UE position. All estimators exhibit a low-SNR threshold region. BEAST exhibits a relatively higher activation threshold, governed by the dual-threshold parameter ζ. A smaller threshold can activate paths at lower SNR, but increases the risk of false-path detections. FLEX shows a lower threshold in range estimation due to the coherent gain of the FFT-based delay stage, but exhibits a high-SNR error floor caused by leakage and grid quantization.[8] Beyond the threshold region, BEAST approaches the RCRB in both angle and range without an observable high-SNR floor, consistent with gridless refinement. EAST, SMF, RDB, and MDE improve with SNR but retain a gap to the RCRB due to suboptimal weighting in their least-squares/subspace stages [15], which are not statistically efficient under the beam–frequency nonlinear structure. Position RMSE follows the same trend as the range RMSE, confirming that positioning in this geometry is range-dominated.

2) Path Separability Sweep: We fix $P_{\mathrm{tx}} = 30$ dBm ($\mathrm{SNR} = 29.75$ dB) and vary the excess path length Δd between LoS and the specular path by displacing the interaction point orthogonally to the LoS midpoint.[9] Fig. 2 reports RMSE versus Δd for LoS range, specular range, and position. BEAST closely follows the RCRB even in the near-coincident regime (small Δd), consistent with gridless delay refinement and super-resolution beyond FFT-bin granularity. EAST, MDE, SMF, and RDB exhibit similar trends but with a moderate performance gap, remaining clustered together across Δd. This behavior is attributable to the common delay-estimation stage based on least-squares ESPRIT, whose suboptimal weighting limits their accuracy. FLEX degrades when Δd approaches the effective FFT delay resolution. In this regime, the main lobes and sidelobes of the two delay components overlap, leading to constructive and destructive superposition in adjacent bins, which manifests as the observed oscillatory (zigzag) RMSE behavior. The position RMSE follows the same trend, confirming that positioning performance is driven primarily by delay resolvability.

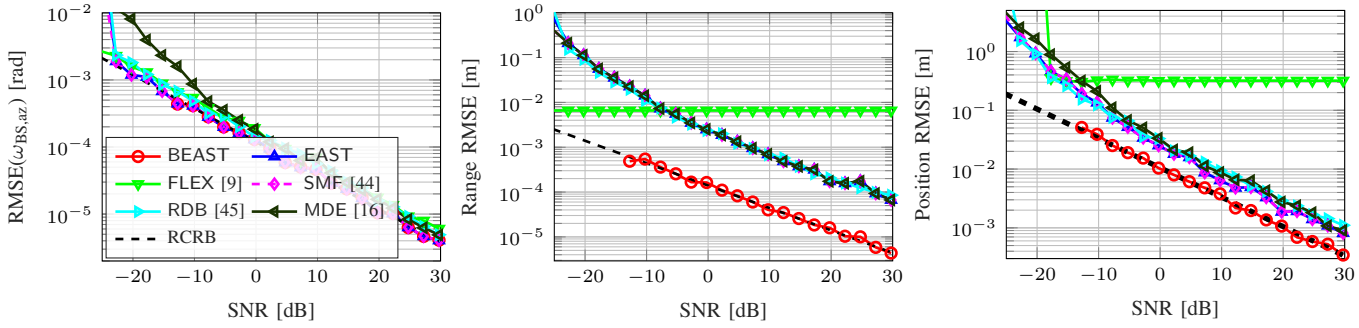
3) Bandwidth Sweep: We sweep $K \in \{128, 256, 512, 1024, 2048, 4096\}$ at fixed $P_{\mathrm{tx}} = 30$ dBm. Since the total noise power over the occupied bandwidth scales as $N_0 K \Delta f$, the effective SNR (cf. (39)) decreases with increasing K. Fig. 3 shows position RMSE versus K. BEAST remains close to the PEB across the sweep, whereas EAST and the other low-complexity baselines (SMF/MDE/RDB) improve monotonically with K and approach the bound at large bandwidth. This confirms that their dominant performance gap at small K is driven by limited range (delay) resolution rather than angular estimation. FLEX also benefits from increasing K, but an offset remains due to residual delay-grid quantization and FFT leakage effects.

4) Robustness to Model Mismatch: To assess robustness to violations of the sparse specular model, we replace the single deterministic reflection by a cluster of $N_{\mathrm{diff}} = 50$ diffuse scatterers distributed over a radius-1 m disk centered at $\mathbf{p}_0$, while preserving total reflected power. The specular magnitude $|\Gamma|$ is swept from 1 (pure specular) to 0 (pure diffuse). As $|\Gamma|$ decreases, the channel departs from the rank-L specular model underlying BEAST/EAST. Fig. 4 shows the resulting position RMSE versus $|\Gamma|$. In this regime, BEAST becomes bias-limited: diffuse energy is approximated by a small number of rank-one atoms, leading to systematic parameter offsets that do not vanish with SNR. Subspace methods (SMF/MDE), relying on second-order structure rather than explicit sparsity, exhibit comparatively smoother behavior.

5) Accuracy–Runtime Trade-off: To characterize the empirical accuracy–complexity envelope, we randomly draw system parameters $K \in \{128, 256, 512, 1024\}$, $M_{\mathrm{UE},h} \in \{6, \ldots, 12\}$, $M_{\mathrm{UE},v} \in \{6, \ldots, 12\}$, $M_{\mathrm{BS},h} \in \{8, \ldots, 16\}$, $M_{\mathrm{BS},v} \in \{6, \ldots, 12\}$, and $P_{\mathrm{tx,dBm}} \in \{10, \ldots, 30\}$, updating the DFT precoder/combiner accordingly while keeping geometry fixed. In this experiment, a snapshot is treated as a gross outlier if the position error e_p is non-finite, non-positive, or exceeds $e_{\max} = 10$ m. These outliers are excluded only from the log-domain scatter and ellipse fit. The corresponding outlier rates are 34.0%, 32.2%, 50.0%, 30.0%, 0.0%, and 0.0% for BEAST, EAST, FLEX, SMF, RDB, and MDE, respectively. Runtime t and position error e_p are then analyzed in the log-domain with $\mathbf{z} = [\log_{10} t,\ \log_{10} e_p]^{\mathsf{T}}$. The empirical dispersion is summarized by the k-sigma ellipse $(\mathbf{z} - \boldsymbol{\mu}_e)^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}_e) = k_e^2$, where $k_e = 1.5$, and $(\boldsymbol{\mu}_e, \boldsymbol{\Sigma})$ denote the sample mean and covariance of the non-missed runs. Fig. 5 reports the joint empirical distribution of runtime per snapshot and position error on log–log axes, together with k-sigma ellipses fitted in the log domain. The plot reveals a clear accuracy–latency frontier. FLEX forms a distinct cluster at very low runtime but high error, corresponding to a fast

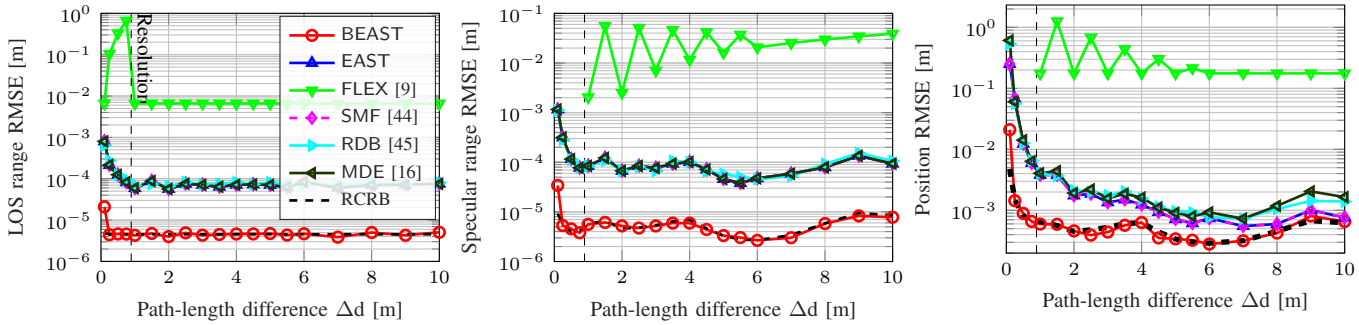
[8] While this floor can be reduced by increasing the FFT grid density or by adding a post-FFT refinement stage, such modifications increase computational complexity and depart from the original low-complexity formulation, and are therefore not considered here.

[9] For reference, in the SNR sweep scenario Sec. VI-D1, $\Delta d \approx 6.7$ m.



(a) RMSE of AoD (b) RMSE of range (c) RMSE of Position

Fig. 1: Stress test (SNR sweep) for the two-path point-cloud channel (LoS + one fixed specular). RMSE versus SNR for (a) BS azimuth AoD, (b) range, and (c) UE position. The dashed curve shows the root CRB (RCRB).



(a) RMSE of LoS Range (b) RMSE of Specular Range (c) RMSE of Position

Fig. 2: Stress test (separation sweep) at fixed transmit power $P_{\text{tx}} = 30$ dBm. RMSE versus excess path-length difference Δd (LoS–specular delay separation) for (a) LoS range, (b) specular range, and (c) UE position.

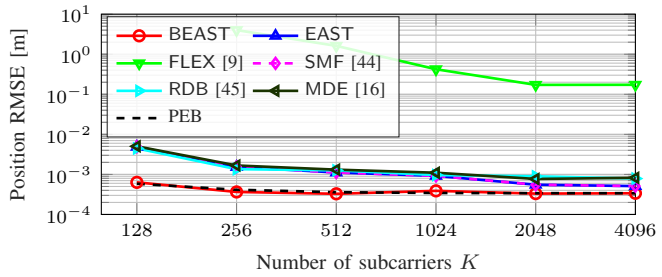


Fig. 3: Stress test (bandwidth sweep) at fixed transmit power $P_{\text{tx}} = 30$ dBm: position RMSE versus number of subcarriers K (bandwidth $B = K\Delta f$) for the two-path channel (LoS + one fixed specular). The dashed curve shows the position error bound (PEB).

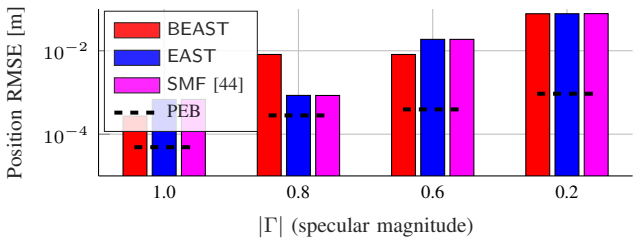


Fig. 4: Stress test (specular-to-diffuse transition): position RMSE versus specular reflection magnitude $|\Gamma|$ for the two-path point-cloud setup where reflected power is gradually transferred from a single deterministic specular component to a diffuse scatterer cloud around the same interaction region. The dashed line indicates the PEB.

but less accurate operating point (bottom-left). Among the non-outlier runs, BEAST reaches the lowest-error regime. Its approximately circular ellipse indicates weak correlation between runtime and accuracy: in well-conditioned snapshots it converges quickly, whereas in ambiguous cases additional iterations do not necessarily yield commensurate accuracy gains. EAST shifts the BEAST cloud leftward, reducing runtime while remaining in the low-error regime, at the cost of a systematic upward shift in error. In contrast, SMF, RDB, and MDE exhibit tilted ellipses, indicating a stronger runtime–accuracy coupling: runs with longer runtime tend to achieve lower position error, suggesting that additional computation is associated with more effective parameter refinement in these methods. Overall, conditioned on non-outlier runs, BEAST provides the lowest position errors and EAST offers a favorable accuracy–runtime compromise, while FLEX achieves the lowest runtime but with higher errors and outlier rate.

6) Stress Test with Gaussian UE Combining: We repeat the SNR sweep while replacing the structured DFT UE combiner by the fixed dense Gaussian design of Sec. VI-B. Only estimators that do not rely on UE shift-invariance (BEAST, EAST, SMF) are considered. Fig. 6 shows that all three estimators remain stable under both combining strategies. Switching from a separable DFT combiner to a dense Gaus-

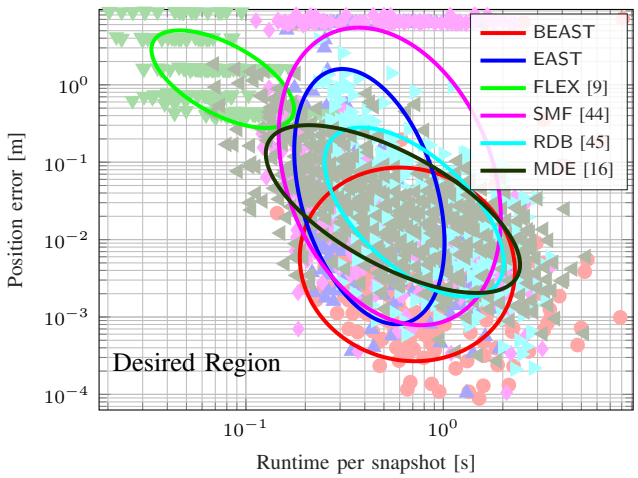


Fig. 5: Runtime-Accuracy trade-off. Each marker is one non-outlier snapshot for a given algorithm, plotted as runtime per snapshot versus position error (both axes in log scale). Samples are excluded from the visualization as gross outliers if t or e_p is non-finite, non-positive, or if $e_p > 10$ m. Solid contours are k_e=1.5 Gaussian ellipses fitted in the $(\log_{10} t, \log_{10} e_p)$ domain and mapped back to linear scale.

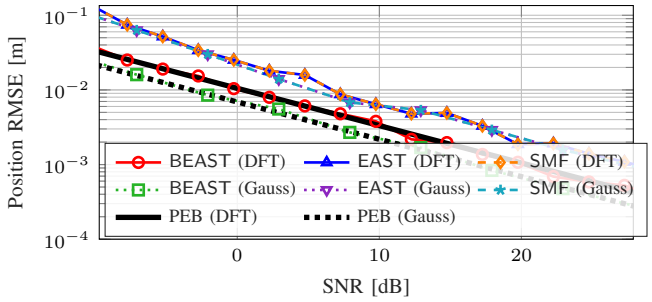


Fig. 6: Impact of UE combiner design on positioning performance in the stress test (LoS + one specular path) under the standard DFT combiner against a fixed dense Gaussian combiner for BEAST, EAST, and the subspace baseline (SMF).

sian combiner reduces the PEB, indicating that, for fixed measurement dimensions, the DFT beam-training operator is more information-lossy in this setup, whereas the dense Gaussian combiner yields a better-conditioned sensing matrix and hence larger Fisher information. BEAST translates this improvement into a corresponding accuracy gain, while EAST and SMF exhibit more limited gains, indicating that their decoupled/subspace processing is less effective in leveraging the improved conditioning of the Gaussian measurement operator. Table I consolidates the qualitative trends across stress dimensions for direct comparison of estimators under controlled perturbations.

E. Realistic Scenario with Clustered Reflections

We next consider a clustered geometric channel consisting of one LoS path and two reflection clusters (wall and ground), each modeled as the superposition of one dominant specular component and multiple diffuse sub-paths (see Fig. 7).

For the wall cluster, we generate $R_{\text{wall}} = 20$ diffuse subpaths in addition to a single specular reflection with magnitude $|\Gamma_{\text{wall}}| = 0.90$. A fraction $\eta_{\text{wall}} = 0.70$ of the reflected

TABLE I: Stress-test relative trends observed in this section.

Test Ref.	BEAST	EAST	FLEX [9]	SMF [44]	MDE [16]	RDB [45]
SNR Sweep	++	+	-	+	$\tilde{0}$	$\tilde{0}$
Bandwidth Sweep	++	+	-	+	$\tilde{0}$	$\tilde{0}$
Runtime	$\tilde{0}$	+	++	$\tilde{0}$	+	+
Specular-to-diffuse	-	$\tilde{0}$	n/a	$\tilde{0}$	n/a	n/a
Gaussian Combiner	++	$\tilde{0}$	n/a	$\tilde{0}$	n/a	n/a

Scores: ++ best, + strong, $\tilde{0}$ competitive, − weak (within each test). n/a: not evaluated or not applicable under that combiner design.

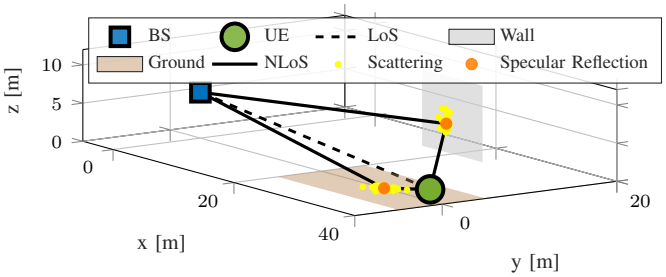
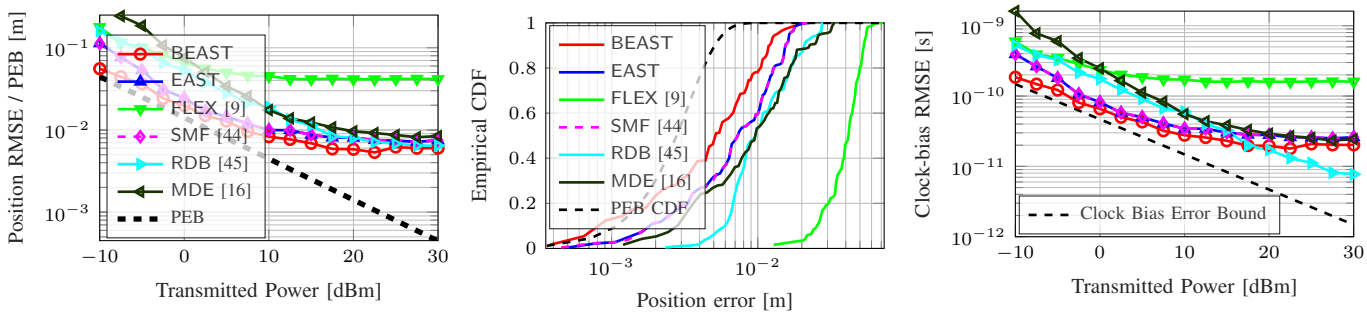


Fig. 7: One realistic channel realization with a LoS path (dashed) and clustered NLoS paths (solid) from wall and ground landmarks, each with one specular point (orange) and multiple diffuse paths (yellow).

power is redistributed over the diffuse components. The diffuse interaction points are drawn in local wall coordinates according to a zero-mean Gaussian distribution with standard deviations $\boldsymbol{\sigma}_{\text{wall}} = [0.75, 1]$ m. The corresponding radar cross section (RCS) are modeled as i.i.d. random variables with mean 0.5 and standard deviation 0.1. For the ground cluster, we generate $R_{\text{ground}} = 35$ diffuse subpaths plus one specular component with magnitude $|\Gamma_{\text{ground}}| = 0.10$, and diffuse fraction $\eta_{\text{ground}} = 0.60$. The diffuse points follow a Gaussian spread with $\boldsymbol{\sigma}_{\text{ground}} = [0.5, 1.2]$ m in local surface coordinates, and RCS values are drawn i.i.d. with mean 0.3 and standard deviation 0.15. In both clusters, total reflected power is preserved by allocating $(1-\eta)$ of the cluster power to the dominant specular component and distributing η proportionally across the diffuse subpaths (see Section II-B).

We sweep the transmitted power by scaling P_{tx} and report position RMSE, clock-bias RMSE, and the position-error CDF. Fig. 8 shows that all estimators improve with increasing power and saturate at high transmission powers, leaving a persistent gap to the PEB and the timing bound. The relative ranking is consistent with the specular stress tests, indicating that the dominant algorithmic mechanisms carry over to the clustered environment. At high powers (SNRs), the common error floor indicates a transition from noise-limited to model-limited behavior. The diffuse intra-cluster spread introduces structured components that are not captured by the sparse specular model underlying the estimators. As a result, constructive and destructive interference within each cluster induces bias in the estimated path parameters, preventing convergence to the PEB even as thermal noise vanishes.

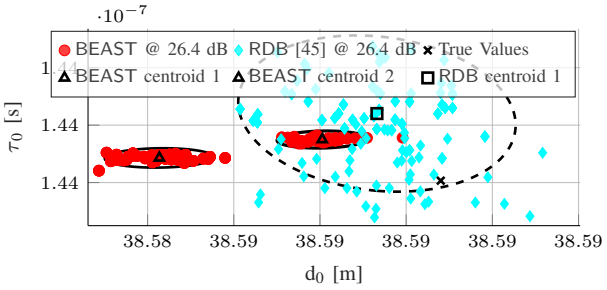
For clock-bias estimation, RDB exhibits a lower high-SNR floor than BEAST and the remaining methods. To understand this behavior, we examine the Monte Carlo scatter of the



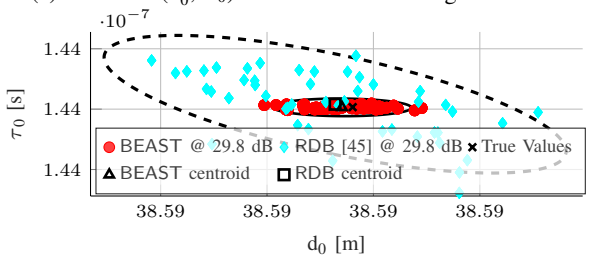
(a) RMSE of position (b) CDF of position at $P_{\text{tx}} = 10$ dBm (c) RMSE of Clock bias estimation

Fig. 8: Realistic scenario: transmitted-power sweep in a two-landmark environment with LoS, wall, and ground reflections, each with one specular component and a diffuse cluster. (a) Position RMSE. (b) Position-error CDF at $P_{\text{tx}} = 10$ dBm. (c) Clock-bias RMSE.

estimated absolute range and delay $(\hat{d}_0, \hat{\tau}_0)$. We form $\mathbf{x} = [\hat{d}_0,\ c\hat{\tau}_0] \in \mathbb{R}^2$ and fit a Gaussian mixture model (GMM) with $K \in \{1, 2\}$ components, selecting K via Bayesian information criterion. Fig. 9 (a) reveals a bimodal behavior for BEAST in the realistic diffuse case: the estimates concentrate into two nearby but distinct clusters, yielding a non-negligible bias when a single solution is selected per run or when aggregating across trials. This is consistent with super-resolution ambiguity in clustered propagation, where closely spaced subpaths compete and the optimizer converges to different plausible explanations depending on the diffuse realization. For comparison, Fig. 9 (b) shows the matched specular case, in which this bimodal behavior is absent. This contrast suggests high-SNR floor of BEAST in clock-bias estimation is driven by mismatch-induced multi-modality rather than by thermal noise alone. Tightening solver hyperparameters reduces outliers but does not remove this mismatch-induced bias, whose magnitude depends on intra-cluster separation.



(a) Absolute (d_0, τ_0) with GMM clustering @ 26.4 dB



(b) GMM clustering of (d_0, τ_0) @ 29.8 dB (stress SNR sweep)

Fig. 9: Absolute $(\hat{d}_0, \hat{\tau}_0)$ scatter with GMM clustering at high SNR for (a) the realistic clustered model with diffuse subpaths and (b) the matched specular stress-test case.

VII. Conclusion and Future Work

We investigated single-BS wideband mmWave positioning from beamspace MIMO–OFDM measurements under unknown UE orientation and clock bias. The focus was on the channel-estimation front-end, since delay and angular errors directly propagate to both position and bias estimates. We proposed two gridless estimators operating directly on the beam–frequency measurement model. BEAST formulates channel estimation as a 5D continuous atomic regression over delay and departure/arrival angles, enabling off-grid super-resolution without discretization. EAST reduces complexity via delay-first decoupling: after subspace-based delay estimation and projection, per-slice 4D atomic refinement is performed only over angular parameters. Across controlled stress tests and a clustered propagation scenario, BEAST closely tracks the CRB in the noise-limited regime after reliable path activation and eliminates grid-induced ranging floors. EAST achieves a favorable accuracy–complexity trade-off with performance close to BEAST under a wide range of conditions. In clustered environments at high SNR, all methods exhibit a model-limited error floor caused by intra-cluster ambiguity, highlighting the role of estimator bias beyond the noise-limited regime.

Future work includes incorporating hybrid specular–diffuse channel models to reduce the high-SNR mismatch floor observed under clustered propagation, and extending the framework to joint delay–angle–Doppler estimation for high-mobility scenarios.

References

[1] T. S. Rappaport *et al.*, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, 2015.

[2] L. Pucci, E. Paolini, and A. Giorgetti, "System-level analysis of joint sensing and communication based on 5G new radio," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2043–2055, 2022.

[3] L. Italiano *et al.*, "A tutorial on 5G positioning," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 3, pp. 1488–1535, 2025.

[4] G. Kwon *et al.*, "Joint communication and localization in millimeter wave networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1439–1454, 2021.

[5] Z. Gong *et al.*, "Simultaneous localization and communications with massive MIMO-OTFS," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 12, pp. 3908–3924, 2023.

[6] G. Kwon *et al.*, "Integrated localization and communication for efficient millimeter wave networks," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 12, pp. 3925–3941, 2023.

[7] Z. Gong *et al.*, "High-accuracy positioning services for high-speed vehicles in wideband mmwave communications," *IEEE Transactions on Signal Processing*, vol. 71, pp. 3867–3882, 2023.

[8] H. Huang, P. Wu, and M. Xia, "Bayesian inference-aided channel estimation for hybrid full-dimensional MIMO systems," *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 1047–1051, 2022.

[9] A. Pourafzal *et al.*, "FLEX: Low-complexity 5D beamspace channel estimation for mmWave MIMO-OFDM," in *33rd European Signal Processing Conference (EUSIPCO)*. Palermo, Italy: IEEE, 2025, pp. 1–5.

[10] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on signal processing*, vol. 42, no. 10, pp. 2664–2677, 2002.

[11] R. Thomä, M. Landmann, and A. Richter, "RIMAX—a maximum likelihood framework for parameter estimation in multidimensional channel sounding," in *Proceedings of the International Symposium on Antennas and Propagation (ISAP'04)*. Citeseer, 2004, pp. 53–56.

[12] C. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of statistics*, pp. 95–103, 1983.

[13] F. Jiang *et al.*, "Beamspace multidimensional ESPRIT approaches for simultaneous localization and communications," *arXiv preprint arXiv:2111.07450*, 2021.

[14] H. Chen *et al.*, "Tensor decompositions in wireless communications and MIMO radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 438–453, 2021.

[15] A. Pourafzal *et al.*, "Low complexity subspace approach for unbiased frequency estimation of a complex single-tone," *Digital Signal Processing*, vol. 145, p. 104304, 2024.

[16] F. Jiang *et al.*, "High-dimensional channel estimation for simultaneous localization and communications," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2021, pp. 1–6.

[17] N.-S. Duong, Q.-T. Nguyen, and T.-M. Dinh-Thi, "OMP-based channel estimation with dynamic grid for mmwave MIMO positioning systems," *IEEE Communications Letters*, vol. 27, no. 10, pp. 2623–2627, 2023.

[18] A. Mishra *et al.*, "Sparse bayesian learning-based channel estimation in millimeter wave hybrid MIMO systems," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2017, pp. 1–5.

[19] M. Ke *et al.*, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Transactions on Signal Processing*, vol. 68, pp. 764–779, 2020.

[20] Y. Tsai, L. Zheng, and X. Wang, "Millimeter-wave beamformed full-dimensional MIMO channel estimation based on atomic norm minimization," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6150–6163, 2018.

[21] H. Chu, L. Zheng, and X. Wang, "Semi-blind millimeter-wave channel estimation using atomic norm minimization," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2535–2538, 2018.

[22] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmwave MIMO systems via atomic norm minimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5786–5797, 2021.

[23] J. Li, M. F. Da Costa, and U. Mitra, "Joint localization and orientation estimation in millimeter-wave MIMO OFDM systems via atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4252–4264, 2022.

[24] Y. Chu *et al.*, "Channel estimation for RIS-aided MIMO systems: A partially decoupled atomic norm minimization approach," *IEEE Transactions on Wireless Communications*, vol. 23, no. 11, pp. 16 048–16 061, 2024.

[25] L. Zhu *et al.*, "A novel gridless uplink/downlink channel estimation method for millimeter wave MIMO-OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 24, no. 5, pp. 3780–3793, 2025.

[26] A. Callejas-Ramos *et al.*, "Atomic norm for parametric estimation of sparse channels," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2025.

[27] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[28] Y. Chu *et al.*, "Downlink-uplink collaborative channel estimation for TDD massive MIMO communications," *IEEE Transactions on Signal Processing*, vol. 73, pp. 3614–3628, 2025.

[29] R. Li and D. Cabric, "A coordinate descent approach to atomic norm denoising," *IEEE Transactions on Signal Processing*, vol. 72, pp. 5077–5090, 2024.

[30] F. Wen *et al.*, "5G positioning and mapping with diffuse multipath," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1164–1174, 2020.

[31] O. El Ayach *et al.*, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[32] V. Degli-Esposti *et al.*, "Measurement and modelling of scattering from buildings," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 1, pp. 143–153, 2007.

[33] J. Ebrahimizadeh *et al.*, "Rcs-based 3-d millimeter-wave channel modeling using quasi-deterministic ray tracing," *IEEE Transactions on Antennas and Propagation*, vol. 72, no. 4, pp. 3596–3606, 2024.

[34] Y. Ge *et al.*, "Experimental validation of single BS 5G mmWave positioning and mapping for intelligent transport," *IEEE Transactions on Vehicular Technology*, 2024.

[35] Y. Chen *et al.*, "Learning to localize with attention: From sparse mmWave channel estimates from a single BS to high accuracy 3D location," *arXiv preprint arXiv:2307.00167*, 2023.

[36] L. Italiano *et al.*, "Holotrace: a location privacy preservation solution for mmwave mimo-ofdm systems," *arXiv preprint arXiv:2509.23444*, 2025.

[37] F. F. Bonsall, "A general atomic decomposition theorem and banach's closed range theorem," *The Quarterly Journal of Mathematics*, vol. 42, no. 1, pp. 9–14, 1991.

[38] V. Chandrasekaran *et al.*, "The convex geometry of linear inverse problems," *Foundations of Computational mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[39] R. T. Rockafellar, "Convex analysis:(pms-28)," 2015.

[40] E. Kreyszig, *Introductory functional analysis with applications*. John Wiley & Sons, 1991.

[41] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[42] B. De Moor, "The singular value decomposition and long and short spaces of noisy matrices," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2826–2838, 1993.

[43] 3GPP, "TS 38.211: NR; physical channels and modulation," 3rd Generation Partnership Project (3GPP), Tech. Rep., release 16.

[44] J. Hong *et al.*, "A tensor-based millimeter wave wideband massive mimo channel estimation technique using uniform planar arrays," *IEEE Wireless Communications Letters*, vol. 13, no. 5, pp. 1458–1462, 2024.

[45] F. Wen *et al.*, "Tensor decomposition based beamspace ESPRIT for millimeter wave MIMO channel estimation," in *2018 IEEE Global Communications Conference (GLOBECOM)*. Abu Dhabi, UAE: IEEE, 2018, pp. 1–7.

[46] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2002.

[47] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5066–5081, 2016.

[48] M. A. Nazari *et al.*, "mmWave 6D Radio Localization With a Snapshot Observation From a Single BS," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 8914–8928, 2023.