Near-field Spatial-domain Channel Extrapolation for XL-MIMO Systems

Jiayi Lu, Jiayi Zhang, Senior Member, IEEE, Hao Lei, Huahua Xiao, Bo Ai, Fellow, IEEE, and Derrick Wing Kwan Ng, Fellow, IEEE

Abstract

Extremely large-scale multiple-input multiple-output (XL-MIMO) systems are pivotal to nextgeneration wireless communications, where dynamic RF chain architectures offer enhanced performance. However, efficient precoding in such systems requires accurate channel state information (CSI) obtained with low complexity. To address this challenge, spatial-domain channel extrapolation has attracted growing interest. Existing methods often overlook near-field spherical wavefronts or rely heavily on sparsity priors, leading to performance degradation. In this paper, we propose an adaptive near-field channel extrapolation framework for multi-subcarrier XL-MIMO systems, leveraging a strategically selected subset of antennas. Subsequently, we develop both on-grid and off-grid algorithms, where the latter refines the former's estimates for improved accuracy. To further reduce complexity, a cross-validation (CV)-based scheme is introduced. Additionally, we analytically formulate the mutual coherence of the sensing matrix and propose a coherence-minimizing-based random pattern to ensure robust extrapolation. Numerical results validate that the proposed algorithms significantly outperform existing methods in both extrapolation accuracy and achievable rate, while maintaining low computational complexity. In particular, our proposed CV ratio offers a flexible trade-off between accuracy and efficiency, and the corresponding off-grid algorithm achieves high accuracy with complexity comparable to conventional on-grid methods.

- J. Lu, H. Lei, and B. Ai are with the State Key Laboratory of Advanced Rail Autonomous Operation, and also with the School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044, P. R. China. (e-mail: jiayilu@bjtu.edu.cn).
- J. Zhang is with the State Key Laboratory of Advanced Rail Autonomous Operation, Nanjing Rongcai Transportation Technology Research Institute Co., Ltd., Nanjing 210000, China, and also with the School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044, P. R. China. (e-mail: jiayizhang@bjtu.edu.cn).
- H. Xiao is with ZTE Corporation and the State Key Laboratory of Mobile Network and Mobile Multimedia Technology. (e-mail: xiao.huahua@zte.com.cn).
- D. W. K. Ng is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: w.k.ng@unsw.edu.au).

Index Terms

Extremely large-scale MIMO, channel extrapolation, near-field communications, mmWave communications.

I. Introduction

Massive multiple-input multiple-output (mMIMO) technology has been widely recognized as one of the core technologies for the fifth-generation (5G) mobile communication systems [1], [2], delivering substantial beamforming and spatial multiplexing gains to achieve higher transmission rates. As a natural evolution of mMIMO, extremely large-scale MIMO (XL-MIMO) is envisioned to satisfy the stringent demands for low latency, high data rate, and enhanced spectral efficiency in emerging sixth-generation (6G) wireless networks [3]. Moreover, to integrate hundreds or even thousands of antennas within a compact XL-MIMO array, the millimeter-wave (mmWave) and terahertz (THz) frequency bands are expected to be important operational scenarios for XL-MIMO systems [3], [4]. Furthermore, XL-MIMO distinguishes itself through its excellent capability of achieving pencil-like energy-focused beamforming, effectively mitigating the severe path loss inherent to high-frequency operations [5].

Notably, XL-MIMO not only scales up the number of antennas but also introduces a paradigm shift in electromagnetic characteristics [6], requiring spherical wave modeling to accurately capture near-field effects within the corresponding Rayleigh distance [7], [8]. Additionally, near-field phenomena enhance the degrees of freedom (DoF) in XL-MIMO channels [3], rendering conventional hybrid architectures with a fixed number of radio frequency (RF) chains inadequate for fully leveraging the available channel DoF. To overcome this challenge, a dynamic RF chain architecture was proposed in [9], where multiple RF chains are dynamically configured through a selection network, achieving near-optimal capacity without incurring additional power consumption. Moreover, to unleash the full performance advantages of dynamic XL-MIMO systems, accurate channel state information (CSI) acquisition with low computational complexity is crucial [10], [11]. To tackle this challenge, spatial-domain channel extrapolation has gained prominence as a viable approach. Specifically, by exploiting the shared propagation environment and the deterministic arrangement of antennas, this technique enables the reconstruction of the complete array's channel leveraging measurements only obtained from a strategically selected subset of antennas.

A. Literature Review

Recently, several studies have explored spatial-domain channel extrapolation through artificial intelligence (AI) and compressed sensing (CS) techniques. In the AI-based category, the authors in [12] demonstrated that a channel-to-channel mapping function exists when the mapping from candidate user locations to the channels measured by the first antenna subset is bijective. Building upon this insight, deep neural networks (DNNs) were utilized to learn this intricate mapping function given sufficient observations [12], [13]. Besides, in [14], an antenna-domain extrapolation network was proposed, leveraging a constrained degradation algorithm to generate a differentiable approximation for discrete antenna selection. In addition, considering both temporal- and spatial-domain channel extrapolation, the study in [15] proposed a novel spatial-domain extrapolation method leveraging self-attention and transfer learning. This method utilizes position embedding to preserve the spatial relationships among antennas and employs temporal-domain channel extrapolation data for model pre-training, thereby capturing essential correlations among antenna CSIs. The model is subsequently fine-tuned adopting spatial-domain extrapolation data, with the initial CSI values estimated via linear interpolation.

The second category encompasses CS-based methods. For instance, in [16], a trivariate Newtonized orthogonal matching pursuit (NOMP)-based multi-domain channel extrapolation scheme was proposed to reconstruct channels across various domains, including time, frequency, polarization, spatial, and user dimensions. Furthermore, an enhanced near-field NOMP estimator was developed in [17], extending the methodology proposed in [16] to reconstruct both user equipment (UE)-to-reconfigurable intelligent surface (RIS) and RIS-to-base station (BS) near-field channels. Moreover, the authors in [17] also investigated different RIS pattern configurations in terms of radiation profiles and parameter estimation accuracy, demonstrating the superiority of a random configuration strategy. In addition, sparse Bayesian learning (SBL) has been recognized as another powerful approach for channel extrapolation. In particular, a tensor-based SBL framework was proposed in [18] to effectively recover the physical channel parameters and reconstruct the cascaded RIS-assisted MIMO-orthogonal frequency-division multiplexing (OFDM) channels.

B. Main Contributions

Despite the promising advances discussed above, several critical limitations remain unaddressed in existing works: Firstly, most studies on spatial-domain channel extrapolation rely heavily on learning-based architectures. However, the limited interpretability of these neural networks often hinders a clear understanding of why the selected antenna elements yield optimal extrapolation results and how these elements are spatially related. Moreover, although extensive research has been conducted on far-field spatial-domain channel extrapolation, spanning a wide range of methods [14]-[16], [18], [19], as well as hardware prototyping and experimental evaluations [20], these approaches are based on the planar wave propagation assumption. Indeed, this assumption becomes inaccurate in near-field scenarios, resulting in noticeable performance degradation. This limitation is particularly critical for CS-based methods, where the energy of a single near-field path component spreads over multiple angles [21], [22], thereby violating the angular-domain sparsity assumption that is fundamental to far-field models. Last but not least, current CS-based methods, e.g., [8], [16], and [23], developed heavily rely on prior knowledge of the number of propagation paths, which is often challenging to determine in advance. As highlighted in [24] and [25], employing a fixed number of iterations is generally a suboptimal strategy for achieving both high accuracy and processing efficiency.

TABLE I: Comparison Of Relevant Literature with this Paper.

References	Near-field	Multi-user	Off-grid	Independent of the	Cross-validation	Deep learning
				number of paths	-based	
[8]	✓	Х	Х	Х	×	Х
[14]	X	✓	-	-	×	✓
[15]	✓	Х	-	-	×	✓
[16]	Х	✓	✓	Х	×	Х
[17]	✓	Х	✓	Х	×	Х
[21]	✓	✓	✓	Х	×	Х
Proposed P-ASOMP	✓	✓	Х	✓	×	Х
Proposed P-ASIGW	✓	✓	✓	✓	×	Х
Proposed CV-P-ASOMP	✓	✓	Х	✓	✓	Х
Proposed CV-P-ASIGW	✓	✓	✓	✓	✓	Х

Motivated by the aforementioned limitations of existing studies, this paper investigates near-field channel extrapolation in a multi-user, multi-subcarrier mmWave XL-MIMO system without relying on prior knowledge of the number of propagation paths. To address the challenges associated with high-dimensional near-field channel reconstruction, we propose a CS-based adaptive channel extrapolation framework. Within this framework, both on-grid and off-grid channel extrapolation algorithms are developed, where the off-grid scheme iteratively refines the coarse on-grid estimates to improve accuracy. Furthermore, a coherence-minimizing antenna selection

strategy is proposed to ensure a well-conditioned sensing matrix and enhance extrapolation robustness. The comparisons of relevant literature with this paper are summarized in TABLE I. Our main contributions are outlined as follows.

- A compressed sensing-based adaptive near-field spatial-domain channel extrapolation framework is proposed for mmWave XL-MIMO systems, operating effectively without relying on prior knowledge of the exact path component configuration. Subsequently, in multi-subcarrier scenarios, we formulate the channel extrapolation problem as a sparse signal recovery task under the multiple measurement vector (MMV) scheme. More importantly, we employ a polar-domain dictionary to capture the correlations among antennas, enabling accurate channel reconstruction by jointly estimating both angle and distance information.
- In the proposed channel extrapolation framework, the sensing matrix is fundamentally determined by the selected antenna elements. Therefore, the antenna selection strategy plays a crucial role in ensuring effective extrapolation performance. To address this, we introduce a novel coherence-minimizing antenna selection strategy, which aims to maintain stable and robust extrapolation performance. Additionally, we evaluate four distinct antenna selection setups, including both deterministic and random antenna positions, by analyzing their radiation profiles and channel extrapolation accuracy. Numerical results are presented to validate the high angular and distance resolution achieved by the proposed strategy.
- Within the proposed adaptive framework, we introduce an on-grid polar-domain adaptive simultaneous orthogonal matching pursuit (P-ASOMP) algorithm for channel extrapolation. To further enhance extrapolation accuracy, we propose an off-grid polar-domain adaptive simultaneous iterative gridless weighted (P-ASIGW) algorithm that refines near-field channel parameters. To reduce the computational complexity of these adaptive extrapolation algorithms, a cross-validation (CV)-based scheme is developed within the MMV framework. Utilizing this scheme, two enhanced algorithms, CV-P-ASOMP and CV-P-ASIGW, are proposed. By reducing the matrix dimension during the correlation process and incorporating the strong Wolfe conditions, the off-grid CV-P-ASIGW method achieves significantly superior performance, while maintaining comparable computational complexity to existing on-grid methods.

Notation: Lower-case and upper-case boldface letters a and A denote a vector and a matrix, respectively. The superscripts $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^\dagger$, and $\|\cdot\|_F$ denote the transpose, conjugate-transpose,

pseudo-inverse, and Frobenius norm, respectively. \odot represents the Hadamard product. The $M \times N$ real-valued matrix and the $M \times N$ complex-valued matrix are denoted by $\mathbb{R}^{M \times N}$ and $\mathbb{C}^{M \times N}$, respectively. $[\mathbf{A}]_{i,j}$ corresponds to the (i,j)-th element of \mathbf{A} . A unit matrix of size N is represented by \mathbf{I}_N . We denote the circularly symmetric complex-valued Gaussian distribution by $\mathcal{CN}(\mu, \sigma^2)$, where μ and σ^2 denote the mean and the variance. The integer ceiling is denoted as $\lceil \cdot \rceil$, and the integer floor is denoted as $\lceil \cdot \rceil$. The number of combinations for selecting K elements from a set of N elements is denoted as \mathbb{C}_N^K . $\mathbb{E}\{x\}$ is the expectation of a random variable x.

II. SYSTEM AND CHANNEL MODEL

In this paper, we consider an uplink time division duplex (TDD) scenario for an mmWave dynamic XL-MIMO system utilizing OFDM to serve U single-antenna users with M subcarriers¹. For the uplink, we assume that the BS receives mutually orthogonal pilot sequences from all U users. This ensures that the channel extrapolation process can be performed independently for each user [21]. Without loss of generality, we focus our analysis on an arbitrary user.

As illustrated in Fig. 1, the BS employs a uniform linear array (ULA) with a dynamic architecture², consisting of N antennas and $K \leq N$ RF chains. Moreover, as the aperture of the XL-MIMO array increases, the near-field region expands, and the amplitude and phase variations across the array become non-negligible, as shown in Fig. 2. Consequently, the conventional far-field planar wave model becomes inadequate and the near-field spherical wave model should be employed instead. Specifically, the near-field multi-path propagation is modeled as comprising L components, including one line-of-sight (LoS) path and L-1 non-line-of-sight (NLoS) paths [26]. Under the spherical wave assumption, the near-field channel vector $\mathbf{h}_m \in \mathbb{C}^{N \times 1}$ can be denoted as [27]

$$\mathbf{h}_{m} = \sum_{l=1}^{L} g_{l,m} \mathbf{c}(\vartheta_{l}, r_{l}) \odot \mathbf{b}_{m}(\vartheta_{l}, r_{l})$$

$$= \sum_{l=1}^{L} g_{l,m} \mathbf{a}_{m}(\theta_{l}, r_{l}),$$
(1)

¹It is worth mentioning that the multicarrier OFDM framework adopted in this study is readily extendable to orthogonal frequency-division multiple access (OFDMA)-based systems. Specifically, for subcarriers assigned to a single user, the signal model and assumptions made in this paper remain valid.

²Moreover, it is worth highlighting that the proposed approach is not restricted to the dynamic RF chain configuration described here; rather, it can be readily extended to more general hybrid analog–digital architectures by employing standard on/off switching circuits.

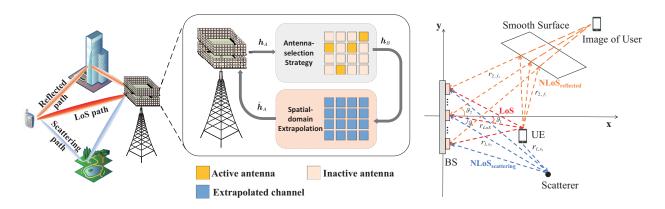


Fig. 1: Illustration of the spatial-domain channel extrapolation in XL- Fig. 2: Near-field XL-MIMO channel MIMO systems.

where $\theta_l = \sin \theta_l$. In (1), the gain correction vector $\mathbf{c}(\theta_l, r_l) \in \mathbb{C}^{N \times 1}$ can be written as $\mathbf{c}(\theta_l, r_l) = 0$ $\left[\frac{r_l}{r_l^{(1)}}, \dots, \frac{r_l}{r_l^{(N)}}\right]^{\mathrm{T}}$, where $r_l^{(n)}$ denotes the actual distance from the scatterer to the *n*-th receiving antenna at the BS in the l-th path. As shown in Fig. 2, for the LoS path, the gain can be expressed as $g_{1,m}=\frac{\lambda_m e^{-jk_mr_1}}{4\pi r_1}$, where $k_m=\frac{2\pi f_m}{c}$ is the wavenumber, λ_m is the wavelength at frequency f_m , and $r_1 = r_{\rm LoS}$ denotes the distance of the LoS path. Besides, the frequency of the m-th subcarrier, given a center frequency f_c and system bandwidth B, is calculated as $f_m = f_c + \frac{B}{M}(m-1-\frac{M-1}{2}), m \in \{1,2,\ldots,M\}.$ Moreover, the NLoS paths are composed of a reflected path and L-2 scattering paths [27]. The reflected path is typically caused by ground or building surface reflections, as illustrated in Fig. 2. The gain of the reflected path is given by $g_{l,m}=\frac{\gamma_f\lambda_m e^{-jk_m(r_{l,f_1}+r_{l,f_2})}}{4\pi(r_{l,f_1}+r_{l,f_2})}$ for l=2, where $|\gamma_f|\in(0,1)$ defines the gain degradation due to reflection; r_{l,f_1} and r_{l,f_2} represent the distance from the user to the smooth surface and the distance from the smooth surface to the reference antenna, respectively. The gains of scattering paths are given by $g_{l,m}=\frac{\gamma_s\lambda_m^2e^{-jk_m(r_{l,s_1}+r_{l,s_2})}}{(4\pi)^2r_{l,s_1}r_{l,s_2}}$ for $l=3,\ldots,L$, where $|\gamma_s|\in(0,1)$ characterizes the reflectivity coefficient of the near-field scatterers; r_{l,s_1} and r_{l,s_2} denote the distance from the user to the scatterer and the distance from the scatterer to the reference antenna, respectively. Therefore, the multipath component gains in the channel model generally satisfy $|g_{\text{LoS}}| > |g_{\text{reflected}}| > |g_{\text{scattering}}|$. The near-field array steering vector $\mathbf{b}_m(\vartheta_l, r_l) \in \mathbb{C}^{N \times 1}$ can be represented by [21]

$$\mathbf{b}_{m}(\vartheta_{l}, r_{l}) = [e^{-jk_{m}(r_{l}^{(1)} - r_{l})}, \cdots, e^{-jk_{m}(r_{l}^{(N)} - r_{l})}]^{\mathrm{T}}.$$
(2)

In particular, for the reflected path, the distance is calculated by $r_l = r_{l,f_1} + r_{l,f_2}$, whereas for scattering paths, it is determined by $r_l = r_{l,s_2}$. By applying the second-order Taylor series

expansion, the distance to the *n*-th antenna, $r_l^{(n)}$, is approximated by $r_l^{(n)} \approx r_l - \delta_n d \sin \vartheta_l + \frac{\delta_n^2 d^2 (1-\sin^2 \vartheta_l)}{2r_l}$, $\delta_n = \frac{2n-N-1}{2}$, $n \in \{1,\ldots,N\}$. Here, ϑ_l represents the physical angle of the user's image relative to the reference antenna in the reflected path, and the physical angle of the last scatterer relative to the reference antenna in the scattering path, as illustrated in Fig. 2.

Owing to the limited transmission paths in mmWave, the inherent sparsity in the polar-domain can be potentially exploited to achieve near-field channel extrapolation with a reduced number of measurements [28], [29]. In near-field scenarios, the channel can be transformed into the polar-domain through a polar-domain dictionary \mathbf{W} . As discussed in [8], $\mathbf{W} \in \mathbb{C}^{N \times N_a N_d}$ is composed of N_d submatrices, where $\mathbf{W} = \begin{bmatrix} \mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_{N_d} \end{bmatrix}$. Each submatrix $\mathbf{W}_{n_d} \in \mathbb{C}^{N \times N_a}$, $n_d \in \{1, 2, \cdots, N_d\}$, contains N_a array steering vectors at the carrier frequency f_c , and can be represented as $\mathbf{W}_{n_d} = [\mathbf{b}(\overline{\vartheta}_1, \overline{r}_{1,n_d}, f_c), \cdots, \mathbf{b}(\overline{\vartheta}_{N_a}, \overline{r}_{N_a,n_d}, f_c)]$. Among them, the on-grid angles are uniformly sampled such that $\sin \overline{\vartheta}_{n_a} = \frac{2(n_a-1)-N_a}{N_a}$, $n_a \in \{1, 2, \cdots, N_a\}$, while on-grid distances are non-uniformly sampled as $\overline{r}_{n_a,n_d} = \frac{D^2 \cos^2 \overline{\vartheta}_{n_a}}{2\beta^2 \lambda_c n_d}$, where β is a predefined parameter³. Then, (1) can be equivalently restated as

$$\mathbf{h}_m = \mathbf{W}\mathbf{h}_p(m),\tag{3}$$

where $\mathbf{h}_p(m) \in \mathbb{C}^{N_a N_d \times 1}$ represents the polar-domain channel at the m-th subcarrier.

Considering the uplink transmission with a single pilot $x_m \in \mathbb{C}$, we denote the received signal at the m-th subcarrier as $\mathbf{r}_m \in \mathbb{C}^{N \times 1}$ at the BS, which can be denoted by

$$\mathbf{r}_m = \mathbf{h}_m x_m + \tilde{\mathbf{n}}_m. \tag{4}$$

In (4), the complex Gaussian noise $\tilde{\mathbf{n}}_m \in \mathbb{C}^{N \times 1}$ follows the distribution $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$, where σ^2 denotes the noise power at each received antenna. To further introduce the near-field spatial-domain channel extrapolation⁴, we define the partially received signal $\mathbf{y}_m \in \mathbb{C}^{K \times 1}$ on the m-th subcarrier as

$$\mathbf{y}_m = \mathbf{A}\mathbf{r}_m = \mathbf{A}(\mathbf{h}_m x_m + \tilde{\mathbf{n}}_m). \tag{5}$$

The binary measurement matrix $\mathbf{A} \in \mathbb{R}^{K \times N}$ is determined by the pattern of antenna selection $\mathbf{\Xi} \in \mathbb{R}^{K \times 1}$, which can be formulated as

³The predefined parameter governs the correlation between near-field steering vectors, which determines the reconstruction accuracy.

⁴To focus on the spatial-domain channel extrapolation task, the effect of phase shifters is deliberately neglected during the CSI acquisition phase.

$$[\mathbf{A}]_{k,n} = \begin{cases} 1, & n \in \Xi(k); \\ 0, & \text{else.} \end{cases}$$
 (6)

For simplicity and without loss of generality, we assume that the pilot symbol $x_m = 1$ and then the post-processed baseband signal can be rewritten as

$$\mathbf{y}_m = \mathbf{A}\mathbf{h}_m + \mathbf{n}_m,\tag{7}$$

where $\mathbf{n}_m = \mathbf{A}\tilde{\mathbf{n}}_m \in \mathbb{C}^{K \times 1}$. Based on (3), we can re-express (7) as

$$\mathbf{y}_{m} = \mathbf{AWh}_{p}(m) + \mathbf{n}_{m}$$

$$= \mathbf{\Psi}\mathbf{h}_{p}(m) + \mathbf{n}_{m},$$
(8)

where the sensing matrix $\Psi \in \mathbb{C}^{K \times N_a N_d}$ is defined by $\Psi = AW$. Moreover, after transmitting over all M subcarriers, the received signal can be expressed as

$$Y = AH + N = \Psi H_p + N, \tag{9}$$

where
$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M] \in \mathbb{C}^{K \times M}$$
, $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_M] \in \mathbb{C}^{N \times M}$, $\mathbf{H}_p = [\mathbf{h}_p(1), \mathbf{h}_p(2), \dots, \mathbf{h}_p(M)] \in \mathbb{C}^{N_a N_d \times M}$, and $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_M] \in \mathbb{C}^{K \times M}$.

III. PROPOSED FRAMEWORK FOR NEAR-FIELD SPATIAL-DOMAIN CHANNEL

EXTRAPOLATION

In Section III-A, we formulate an optimization problem employing a multi-subcarrier joint scheme. Then, in Section III-B, we propose an on-grid adaptive spatial-domain channel extrapolation algorithm, termed P-ASOMP, within the MMV framework. To further enhance accuracy, we introduce an off-grid adaptive channel extrapolation algorithm, named P-ASIGW. Additionally, to mitigate the high computational complexity, we propose a CV-based scheme within the MMV framework in Section III-C, enabling low-complexity near-field channel extrapolation.

A. Problem Formulation

Let $A \in \mathbb{R}^{N \times 1}$ denote the complete set of all antennas and $\mathcal{B} \in \mathbb{R}^{K \times 1}$ represent a subset of A. Accordingly, we define $\mathbf{H}^A \in \mathbb{C}^{N \times M}$ (equivalent to $\mathbf{H} \in \mathbb{C}^{N \times M}$ in (9)) as the comprehensive channel matrix encompassing all antennas, and $\mathbf{H}^{\mathcal{B}} \in \mathbb{C}^{K \times M}$ (equivalent to $\mathbf{Y} \in \mathbb{C}^{K \times M}$ in (9)) as the partially received signal from the channel corresponding to the subset \mathcal{B} . For a given static communication environment, the bijectivity of the position-to-channel mapping function has been demonstrated in [12]. Consequently, a channel-to-channel mapping is established [14]. This correspondence makes it feasible to extrapolate the full CSI from the partially received signal

 $\mathbf{H}^B \in \mathbb{C}^{K \times M}$. As illustrated in Fig. 1, the extrapolated channel is defined by $\hat{\mathbf{H}}^A \in \mathbb{C}^{N \times M}$. Given that the extrapolation problem inherently involves reconstructing a high-dimensional signal from limited observations, we formulate it within the CS framework. By leveraging CS theory, the high-dimensional matrix recovery problem is formulated as an l_0 -norm minimization to identify the sparsest solution. However, due to its nondeterministic polynomial-time (NP)-hard nature [30], a common alternative is to employ l_1 -norm minimization, as detailed in [31]. Under the MMV framework, we further reformulate the sparse signal recovery problem with a row-wise $l_{2,1}$ norm minimization as follows [24]:

$$\hat{\mathbf{H}}_{p} = \arg \min \|\mathbf{H}_{p}\|_{2,1}, \text{ s.t. } \|\mathbf{H}^{\mathcal{B}} - \mathbf{\Psi} \mathbf{H}_{p}\|_{F} \le \epsilon, \tag{10}$$

where ϵ is explicitly determined by the noise level⁵ and $\hat{\mathbf{H}}_p \in \mathbb{C}^{N_a N_d \times M}$ is the estimate of $\mathbf{H}_p \in \mathbb{C}^{N_a N_d \times M}$. Importantly, unlike conventional hybrid architectures, the sensing matrix $\boldsymbol{\Psi}$ in the spatial-domain channel extrapolation problem is determined by the antenna selection pattern rather than the analog combining network. Then, the extrapolated near-field channel can be denoted as

$$\hat{\mathbf{H}}^{\mathcal{A}} = \mathbf{W}\hat{\mathbf{H}}_{p}.\tag{11}$$

As shown in (11), the complete near-field channel is reconstructed exploiting the polar-domain dictionary W.

B. Proposed Polar-domain Adaptive Channel Extrapolation Algorithms

Accurately determining the exact number of multipath components in practical scenarios is inherently challenging and often incurs additional overhead. This uncertainty directly affects the appropriate choice of iteration count in CS-based algorithms. Specifically, adopting insufficient iterations may degrade reconstruction accuracy, while excessive iterations can lead to noise-fitting and increase computational complexity.

To address this issue, we investigate the coherence properties of spatial-domain channels by leveraging a polar-domain dictionary, which effectively captures the inherent angular- and distance-dependent structures of near-field propagation. Motivated by this insight, we propose an on-grid P-ASOMP algorithm with an adaptive halting condition to improve the extrapolation efficiency and robustness, as detailed in **Algorithm 1**.

⁵To ensure accurate recovery, an upper bound on the residual power has been verified in [25], [32]. In this paper, we continue to adopt this limit as the halting condition.

Algorithm 1 Proposed P-ASOMP Algorithm

Require: Partially received signal $H^{\mathcal{B}}$, sensing matrix Ψ , the threshold ϵ .

Ensure: The complete channel $\hat{\mathbf{H}}^{\mathcal{A}}$, number of detected paths \hat{L} .

- 1: Initialize: $\mathbf{R} = \mathbf{H}^{\mathcal{B}}$, $\Gamma = \{\phi\}$, l = 0.
- 2: repeat
- 3: l = l + 1;
- 4: Correlation: $\mathbf{E} = \mathbf{\Psi}^{\mathrm{H}} \mathbf{R} = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_M].$
- 5: Detect new support: $p^* = \arg\max_p \sum_{m=1}^M \|\mathbf{e}_m\|^2$.
- 6: Update support set: $\Gamma_l = \Gamma_{l-1} \cup p^*$.
- 7: Orthogonal projection: $\hat{\mathbf{H}}_p = \mathbf{\Psi}^{\dagger}(:, \Gamma_l)\mathbf{H}^{\mathcal{B}}$.
- 8: Update residual matrix: $\mathbf{R}_l = \mathbf{H}^{\mathcal{B}} \mathbf{\Psi}(:, \mathbf{\Gamma}_l) \hat{\mathbf{H}}_p$.
- 9: **until** $\|\mathbf{R}_l\|_{\mathrm{F}} \leq \epsilon$
- 10: $\hat{\mathbf{H}}^{\mathcal{A}} = \mathbf{W}_{:,\mathbf{\Gamma}_l} \hat{\mathbf{H}}_p, \hat{L} = l.$

We first initialize the residual matrix, the support set, and the number of paths in step 1. Subsequently, the adaptive near-field channel extrapolation process is carried out in steps 2-9. The halting condition is defined by the upper bound on the residual power, given as $\epsilon = \sigma \sqrt{KM}$. Then, in step 4, the projection coefficients of each column of the sensing matrix onto the residual matrix are computed⁶, effectively capturing the spatial correlation through the polar-domain dictionary designed at the center frequency f_c under limited measurements. In step 5, the physical location of the l-th component is determined by the index $p^* = \arg\max_p \sum_{m=1}^M \|\mathbf{e}_m\|^2$. We further add the index p^* to the support set Γ . Following this, the polar-domain channel is estimated using the least squares (LS)⁷ algorithm in step 7. Subsequently, in step 8, the residual matrix is updated by removing the estimated path components. Finally, the complete near-field channel is reconstructed leveraging the polar-domain dictionary in step 10.

⁶Since the center frequency f_c substantially exceeds the system bandwidth, all subcarrier frequencies f_m are approximated as equal to the center frequency, i.e., $f_m \approx f_c$. Consequently, during the detection process, it is assumed that all subcarriers share an identical support set. However, as the bandwidth increases, this approximation becomes invalid, and the beam-split effect becomes pronounced [8]. To address this issue, a frequency-dependent polar-domain dictionary can be designed to mitigate the resulting mismatch, as demonstrated in [24].

⁷In order to ensure that the residuals are orthogonal to the chosen atomic space, LS is widely adopted in existing work [5], [21].

Since the proposed P-ASOMP algorithm operates on a sampled polar-domain (i.e. on-grid angles and distances), its channel extrapolation accuracy is limited. To improve accuracy and better approximate the actual continuous distribution of angles and distances (i.e. off-grid angles and distances), we refine these parameters based on the outcome of the P-ASOMP algorithm. Inspired by the P-SIGW algorithm proposed in [21], we further introduce an off-grid P-ASIGW algorithm, which leverages the number of detected paths \hat{L} from the P-ASOMP algorithm, as detailed in **Algorithm 2**.

Algorithm 2 Proposed P-ASIGW Algorithm

Require: Partially received signal $\mathbf{H}^{\mathcal{B}}$, measurement matrix \mathbf{A} , the minimum distance ρ_{\min} , and number of iterations N_{iter} .

Ensure: The complete channel $\hat{\mathbf{H}}^{\mathcal{A}}$.

- 1: Obtain the initial estimates of the angles $(\hat{\boldsymbol{\theta}}^0)^{\mathrm{T}} = [\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_{\hat{L}}^0]$, the distances $(\hat{\mathbf{r}}^0)^{\mathrm{T}} = [\hat{r}_1^0, \hat{r}_2^0, \dots, \hat{r}_{\hat{L}}^0]$, and the number of detected paths \hat{L} by **Algorithm 1**.
- 2: **for** $n \in \{1, 2, ..., N_{\text{iter}}\}$ **do**
- 3: Choose the strong Wolfe backtracking line search step length l_1 .
- 4: Update the angles by (15).
- 5: Choose the strong Wolfe backtracking line search step length l_2 .
- 6: Update the distances by (16).
- 7: Update the path gains by (13).
- 8: end for

9:
$$\hat{\mathbf{H}}^{\mathcal{A}} = \left[\mathbf{b} \left(\hat{\theta}_{1}^{n}, \hat{r}_{1}^{n} \right), \mathbf{b} \left(\hat{\theta}_{2}^{n}, \hat{r}_{2}^{n} \right), \cdots, \mathbf{b} \left(\hat{\theta}_{\hat{L}}^{n}, \hat{r}_{\hat{L}}^{n} \right) \right] \hat{\mathbf{G}}^{n}.$$

The proposed P-ASIGW algorithm introduces significant innovations in both the initialization stage and refinement stage compared to the P-SIGW algorithm. During the initialization stage, we determine the initial value of the angles $(\hat{\boldsymbol{\theta}}^0)^{\mathrm{T}} = [\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_{\hat{L}}^0] \in \mathbb{C}^{1 \times \hat{L}}$, the distances $(\hat{\mathbf{r}}^0)^{\mathrm{T}} = [\hat{r}_1^0, \hat{r}_2^0, \dots, \hat{r}_{\hat{L}}^0] \in \mathbb{C}^{1 \times \hat{L}}$, and the number of detected paths \hat{L} by executing **Algorithm 1**. Owing to the adaptive halting condition, the reliance on the actual number of path components is eliminated, which enhances usability and avoids excessive iterations. To maximize the likelihood $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}, \hat{\mathbf{G}}) = \arg \max_{\boldsymbol{\theta}, \mathbf{r}, \mathbf{G}} p\left(\mathbf{H}^{\mathcal{B}} \mid \boldsymbol{\theta}, \mathbf{r}, \mathbf{G}\right)$, we alternatively optimize the angles $\hat{\boldsymbol{\theta}}^{\mathrm{T}} \in \mathbb{C}^{1 \times \hat{L}}$, distances $\hat{\mathbf{r}}^{\mathrm{T}} \in \mathbb{C}^{1 \times \hat{L}}$, and path gains $\hat{\mathbf{G}} \in \mathbb{C}^{N_a N_d \times M}$, which is formulated as

$$\min_{\hat{\mathbf{G}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}} \|\mathbf{H}^{\mathcal{B}} - \tilde{\boldsymbol{\Psi}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \hat{\mathbf{G}} \|_F^2.$$
 (12)

Since the optimization problem in (12) is non-convex [21], we utilize the alternating minimization method to address it. Then, the optimal solution for $\hat{\mathbf{G}}^{\text{opt}} \in \mathbb{C}^{N_a N_d \times M}$ can be given by

$$\hat{\mathbf{G}}^{\text{opt}} = \tilde{\mathbf{\Psi}}^{\dagger}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})\mathbf{H}^{\mathcal{B}}.$$
 (13)

Substitute (13) into (12), the maximum-likelihood problem is reformulated as

$$\min_{\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}} \|\mathbf{H}^{\mathcal{B}} - \tilde{\mathbf{\Psi}}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}})\tilde{\mathbf{\Psi}}^{\dagger}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}})\mathbf{H}^{\mathcal{B}}\|_{F}^{2}$$

$$\iff \min_{\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}} \operatorname{Tr} \left\{ \mathbf{H}^{\mathcal{B}^{H}} \left(\mathbf{I} - \mathbf{P}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}) \right)^{H} \left(\mathbf{I} - \mathbf{P}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}) \right) \mathbf{H}^{\mathcal{B}} \right\}$$

$$\iff \min_{\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}} \mathcal{L}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}) = -\operatorname{Tr} \left\{ \mathbf{H}^{\mathcal{B}^{H}} \mathbf{P}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}) \mathbf{H}^{\mathcal{B}} \right\},$$
(14)

where $\mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = \tilde{\boldsymbol{\Psi}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \tilde{\boldsymbol{\Psi}}^{\dagger}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \in \mathbb{C}^{K \times K}$ and (a) is satisfied by $\mathbf{P}^{\mathrm{H}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$. Then, an iterative gradient descent method⁸ is utilized to optimize $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{r}}$ with respect to the new objective function $\mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$. In the n-th iteration of the alternating optimization process, the angles are updated as

$$\hat{\boldsymbol{\theta}}^{n} = \hat{\boldsymbol{\theta}}^{n-1} - l_1 \nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}^{n-1})|_{\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{n-1}}, \tag{15}$$

where l_1 denotes the step length for the angle search. Since the distance-sampling method in the polar-domain dictionary is non-uniform for r but uniform for $\frac{1}{r}$, the distances are updated with respect to $\frac{1}{\hat{r}} = \left[\frac{1}{\hat{r}_1}, \frac{1}{\hat{r}_2}, \dots, \frac{1}{\hat{r}_{\hat{L}}}\right]$ to indirectly update $\hat{\mathbf{r}}$, ensuring stable performance. In the n-th iteration, the distances are updated as

$$\frac{1}{\hat{\mathbf{r}}^n} = \frac{1}{\hat{\mathbf{r}}^{n-1}} - l_2 \nabla_{\frac{1}{\hat{\mathbf{r}}}} \mathcal{L}(\hat{\boldsymbol{\theta}}^n, \hat{\mathbf{r}})|_{\hat{\mathbf{r}} = \hat{\mathbf{r}}^{n-1}}, \tag{16}$$

where l_2 denotes the step length for the distance search. To ensure the non-increasing property, the P-SIGW algorithm updates the step lengths l_1 and l_2 leveraging Armijo backtracking line search [21]. Furthermore, to accelerate convergence, strong Wolfe conditions⁹ are applied in this paper. Taking the angle search as an example, these conditions include two main components: the sufficient decrease condition, which aligns with the Armijo condition in (17), and the curvature condition, as outlined in (18):

$$\mathcal{L}(\hat{\boldsymbol{\theta}}^n) \le \mathcal{L}(\hat{\boldsymbol{\theta}}^{n-1}) + c_1 l_1 \nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}^{n-1})^{\mathrm{T}} d^n, \tag{17}$$

⁸The details of gradient $\nabla \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$ can be found in Appendix B of [21].

⁹When the array size is not extremely large, the strong Wolfe condition leads to faster convergence. However, as the array size increases substantially, its stricter curvature requirement may compromise stability; thus, adopting a relaxed Armijo rule is more effective.

$$|\nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}^n)^{\mathrm{T}} d^n| \le c_2 |\nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}^{n-1})^{\mathrm{T}} d^n|, \tag{18}$$

where $d^n = -\nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}^{n-1})$, $c_1 \in (0, \frac{1}{2})$, and $c_2 \in (c_1, 1)$. Based on (17) and (18), the step lengths for the angle search are determined in step 3. As in step 4, the angles are updated adopting (15). Similarly, the step lengths for the distance search are selected in step 5, and the distances are updated using (16) in step 6. Then, the path gains are updated according to (13).

Finally, after n iterations, the updated near-field path components are exploited to construct the matrix $\tilde{\mathbf{W}}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}}) = \left[\mathbf{b}\left(\hat{\theta}_1^n,\hat{r}_1^n\right),\mathbf{b}\left(\hat{\theta}_2^n,\hat{r}_2^n\right),\cdots,\mathbf{b}\left(\hat{\theta}_{\hat{L}}^n,\hat{r}_{\hat{L}}^n\right)\right]$, and then the complete near-field channel can be reconstructed as $\hat{\mathbf{H}}^{\mathcal{A}} = \tilde{\mathbf{W}}(\hat{\boldsymbol{\theta}},\hat{\mathbf{r}})\hat{\mathbf{G}}^n$ in step 9.

C. Proposed Cross-validation based Polar-domain Adaptive Channel Extrapolation Algorithms

In XL-MIMO systems, the large number of antennas and the near-field effect jointly impose high computational complexity. Although parameter decoupling has been proven effective for complexity reduction in related work, e.g., [33], its application to channel extrapolation remains constricted due to the challenge of estimating the array covariance matrix from limited observations, thereby necessitating novel algorithmic solutions.

Notably, the correlation step in the OMP framework represents a major source of computational complexity, restricting the practicality of adaptive extrapolation algorithms such as those introduced in Section III-B. To mitigate this limitation, we propose a novel CV-based scheme, inspired by the common support set assumption in the MMV framework. Specifically, the goal of this scheme is to accurately identify the common support set by adopting only a minimal subset of training data. The detailed procedure of the resulting on-grid algorithm, termed CV-P-ASOMP, is presented in **Algorithm 3**.

Exploiting the common support set assumption across subcarriers in the MMV framework, our method strategically partitions them into two distinct groups: (1) a training set for support set identification; (2) a validation set for evaluating the termination criterion. In step 1, the CV parameters are determined by the CV ratio α , yielding $T = \lceil \alpha M \rceil$ training subcarriers and V = M - T validation subcarriers. To ensure coverage across the entire bandwidth, the training set is selected via uniform sampling, i.e., $\mathcal{I}_{\text{train}} = [1, 1 + \Delta, \dots, 1 + (T - 1)\Delta]$, where the sampling interval is set as $\Delta = \left\lfloor \frac{M}{T} \right\rfloor$. The validation set is then defined as $\mathcal{I}_{\text{val}} = \mathcal{I}_{\text{all}} \setminus \mathcal{I}_{\text{train}}$, with $\mathcal{I}_{\text{all}} = [1, \dots, M]$. During initialization, we divide the residual matrix into two parts: the training residual matrix $\mathbf{R}_{\text{T}} = \mathbf{H}_{:,\mathcal{I}_{\text{train}}}^{\mathcal{B}} \in \mathbb{C}^{K \times V}$ and the validating residual matrix $\mathbf{R}_{\text{V}} = \mathbf{H}_{:,\mathcal{I}_{\text{val}}}^{\mathcal{B}} \in \mathbb{C}^{K \times V}$.

Next, the adaptive simultaneous orthogonal matching pursuit is executed in steps 2-10, with the same stopping criterion as in **Algorithm 1**. Notably, the correlation operation is computed only exploiting the training residual matrix $\mathbf{R}_T = \mathbf{H}_{:,\mathcal{I}_{\text{train}}}^{\mathcal{B}} \in \mathbb{C}^{K \times T}$, significantly reducing computational complexity. The support set for the l-th component is identified in step 6 and concatenated in step 7. Following the OMP framework, the polar-domain gains are calculated through the LS method [21]. Subsequently, the residual matrices are updated by subtracting the estimated path components in step 9. Finally, in Step 11, the complete near-field channel is extrapolated by leveraging the polar-domain dictionary, which effectively captures the underlying spatial correlation structure.

Algorithm 3 Proposed CV-P-ASOMP Algorithm

Require: Partially received pilot $\mathbf{H}^{\mathcal{B}}$, sensing matrix Ψ , the threshold ϵ , the ratio of cross-validation α .

Ensure: The complete channel $\hat{\mathbf{H}}^{\mathcal{A}}$, number of detected paths \hat{L} .

- 1: Calculate the cross-validation parameters: $T = \lceil \alpha M \rceil$, $\Delta = \lfloor \frac{M}{T} \rfloor$, $\mathcal{I}_{\text{train}} = [1, 1 + \Delta, \dots, 1 + (T-1)\Delta]$, $\mathcal{I}_{\text{val}} = \mathcal{I}_{\text{all}} \setminus \mathcal{I}_{\text{train}}$.
- 2: Initialize: $\mathbf{R}_{\mathrm{T}} = \mathbf{H}_{:,\mathcal{I}_{\mathrm{train}}}^{\mathcal{B}}$, $\mathbf{R}_{\mathrm{V}} = \mathbf{H}_{:,\mathcal{I}_{\mathrm{val}}}^{\mathcal{B}}$, $\Gamma = \{\phi\}$, l = 0.
- 3: repeat
- 4: l = l + 1;
- 5: Correlation: $\mathbf{E} = \mathbf{\Psi}^{\mathrm{H}} \mathbf{R}_{\mathrm{T}} = \left[\mathbf{e}_{1}, \mathbf{e}_{2}, \cdots, \mathbf{e}_{T} \right]$.
- 6: Detect new support: $p^* = \arg\max_p \sum_{v=1}^V \|\mathbf{e}_v\|^2$.
- 7: Update support set: $\Gamma_l = \Gamma_{l-1} \cup p^*$,
- 8: Orthogonal projection: $\hat{\mathbf{H}}_l = \mathbf{\Psi}^\dagger(:, \mathbf{\Gamma}_l) \mathbf{H}^\mathcal{B}_{:,\mathcal{I}_{\text{train}}}, \ \tilde{\mathbf{H}}_l = \mathbf{\Psi}^\dagger(:, \mathbf{\Gamma}_l) \mathbf{H}^\mathcal{B}_{:,\mathcal{I}_{\text{val}}}, \ \hat{\mathbf{H}}_p = \mathbf{\Psi}^\dagger(:, \mathbf{\Gamma}_l) \mathbf{H}^\mathcal{B}.$
- 9: Update residual matrices: $\mathbf{R}_{\mathrm{T}} = \mathbf{H}^{\mathcal{B}}_{:,\mathcal{I}_{\mathrm{train}}} \mathbf{\Psi}(:,\mathbf{\Gamma}_{l})\hat{\mathbf{H}}_{l},\ \mathbf{R}_{\mathrm{V}} = \mathbf{H}^{\mathcal{B}}_{:,\mathcal{I}_{\mathrm{val}}} \mathbf{\Psi}(:,\mathbf{\Gamma}_{l})\tilde{\mathbf{H}}_{l}.$
- 10: **until** $\|\mathbf{R}_{\mathrm{V}}\|_{\mathrm{F}} \leq \epsilon$
- 11: $\hat{\mathbf{H}}^{\mathcal{A}} = \mathbf{W}_{:,\mathbf{\Gamma}_l} \hat{\mathbf{H}}_p, \hat{L} = l.$

To address the error introduced by on-grid sample points, we propose the CV-P-ASIGW algorithm. Similar to **Algorithm 2**, the step lengths are chosen based on the strong Wolfe conditions; however, the key difference is that the initialization information derived from **Algorithm 3**. The off-grid CV-P-ASIGW algorithm is detailed in **Algorithm 4**.

The main difference between the CV-P-ASIGW and P-ASIGW algorithms lies in the initialization strategy. Specifically, in the CV-P-ASIGW algorithm, the initial estimates of the angles $(\hat{\boldsymbol{\theta}}^0)^{\mathrm{T}} = [\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_{\hat{L}}^0] \in \mathbb{C}^{1 \times \hat{L}}$, the distances $(\hat{\mathbf{r}}^0)^{\mathrm{T}} = [\hat{r}_1^0, \hat{r}_2^0, \dots, \hat{r}_{\hat{L}}^0] \in \mathbb{C}^{1 \times \hat{L}}$, and

Algorithm 4 Proposed CV-P-ASIGW Algorithm

Require: Partially received signal $\mathbf{H}^{\mathcal{B}}$, measurement matrix \mathbf{A} , the minimum distance ρ_{\min} , and number of iterations N_{iter} .

Ensure: The complete channel $\hat{\mathbf{H}}^{\mathcal{A}}$.

- 1: Obtain the initial estimates of the angles $(\hat{\boldsymbol{\theta}}^0)^{\mathrm{T}} = [\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_{\hat{L}}^0]$, the distances $(\hat{\mathbf{r}}^0)^{\mathrm{T}} = [\hat{r}_1^0, \hat{r}_2^0, \dots, \hat{r}_{\hat{L}}^0]$, and the number of detected paths \hat{L} by **Algorithm 3**.
- 2: for $n \in \{1, 2, \dots, N_{\text{iter}}\}$ do
- 3: Choose the strong Wolfe backtracking line search step length l_1 .
- 4: Update the angles by (15).
- 5: Choose the strong Wolfe backtracking line search step length l_2 .
- 6: Update the distances by (16).
- 7: Update the path gains by (13).
- 8: end for

9:
$$\hat{\mathbf{H}}^{\mathcal{A}} = \left[\mathbf{b} \left(\hat{\theta}_{1}^{n}, \hat{r}_{1}^{n} \right), \mathbf{b} \left(\hat{\theta}_{2}^{n}, \hat{r}_{2}^{n} \right), \cdots, \mathbf{b} \left(\hat{\theta}_{\hat{L}}^{n}, \hat{r}_{\hat{L}}^{n} \right) \right] \hat{\mathbf{G}}^{n}.$$

the estimated number of paths \hat{L} are all obtained from **Algorithm 3** in Step 1. Benefiting from the tailored adaptive iteration and line search mechanisms, the proposed approach enables efficient reconstruction of the entire near-field channel with reduced computational complexity, while ensuring satisfactory performance.

IV. DESIGN OF PATTERNS FOR ANTENNA SELECTION

During the channel extrapolation phase, a subset of antennas is strategically selected to receive the signals. However, the primary challenge under this framework lies in determining the optimal antenna selection strategy from the entire array [34], [35]. In this section, we address this problem and aim to design a better pattern for effective extrapolation.

For convenience, we define the compression rate as $\eta = \frac{N}{K}$, where K is the number of selected antenna elements. The partially received signal $\mathbf{H}^{\mathcal{B}} \in \mathbb{C}^{K \times M}$ is constructed from the set of indices corresponding to the ULA elements covered by the pattern, as follows:

$$\Xi = \left\{ \xi_0, \xi_1, \dots, \xi_{\frac{N}{\eta} - 1} \right\},\tag{19}$$

where $0 \le \xi_n \le N-1$ holds for $n \in \{0, 1, \dots, \frac{N}{\eta}-1\}$.

A. Deterministic-Position Setups

Deterministic-position setups for spatial-domain channel extrapolation can be interpreted as intuitive sampling strategies across the entire array. Specifically, we consider two types of patterns, namely, the dense uniform pattern and the sparse comb pattern.

1) Dense uniform pattern [16]: In this pattern setup, a contiguous block of $\frac{N}{\eta}$ adjacent elements in the ULA is selected. Without loss of generality, we assume that the starting index of the dense uniform pattern is i, where $i \in \{0, \dots, \eta - 1\}$. Thus, the dense uniform pattern includes the following elements:

$$\Xi_{\mathrm{du},i} = \left\{ \frac{Ni}{\eta}, \frac{Ni}{\eta} + 1, \dots, \frac{N(i+1)}{\eta} - 1 \right\}. \tag{20}$$

2) Sparse comb pattern [16]: This pattern follows a comb-like structure, in which one antenna element is uniformly selected from every η elements. Let i denote the starting index of the sparse comb pattern. Then, the selected antenna elements can be represented as

$$\Xi_{\rm sc} = \{i, i + \eta, \dots, i + N - \eta\}.$$
 (21)

B. Random-Position Setups

In addition to deterministic-position setups with regularly spaced measurement positions, random-position setups can also be employed for antenna selection, analogous to the concept of random matrices in compressed sensing theory [28].

Here, we consider two random antenna selection setups: (i) the sparse random pattern; (ii) the coherence-minimizing-based random pattern. The former is commonly adopted in existing works [16], [17], while the latter is specifically introduced to minimize the mutual coherence, thereby improving extrapolation performance.

- 1) Sparse random pattern [17]: This pattern randomly selects $\frac{N}{\eta}$ elements from the array. Due to its stochastic nature, this selection pattern cannot be represented explicitly in a fixed format. Let $\Xi_{\rm sr}$ denote the sparse random pattern with a size of K, where the elements are selected from $\gamma = [1, \cdots, N]$.
- 2) Coherence-minimizing-based random pattern: For compressed sensing-based methods, the design of the sensing matrix is critical, as it must satisfy the restricted isometry property (RIP) [36], which is a fundamental condition that underpins the success of the technique. Typically, the sensing matrix is designed based on the measurement matrix and the sparse basis matrix. However, in the spatial-domain channel extrapolation scheme considered in (10), since the measurement matrix is constructed exclusively from only a subset of the antenna positions, the DoFs

in the measurement matrix design are limited [28]. Consequently, conventional measurement matrices cannot be directly adopted. Additionally, verifying the RIP is particularly challenging for two main reasons. First, the RIP requires a bounded condition number for all submatrices, where the total number of submatrices grows combinatorially as C_N^K . Second, computing the spectral norm of these matrices is generally non-trivial [36]. As an alternative, mutual coherence serves as a simple yet effective measure to evaluate the suitability of the sensing matrix [28], [30]. It has been demonstrated that lower coherence is associated with improved reconstruction performance [30]. Specifically, the mutual coherence of the sensing matrix is defined as

$$\mu(\boldsymbol{\Psi}) \triangleq \max_{q < p} |\boldsymbol{\Psi}_{:,q}^{H} \boldsymbol{\Psi}_{:,p}| = \max_{q < p} |(\mathbf{A} \mathbf{w}_{q})^{H} (\mathbf{A} \mathbf{w}_{p})|$$

$$= \max_{q < p} \left| \mathbf{w}_{q}^{H} \left(\sum_{k=1}^{K} \mathbf{a}_{k} \mathbf{a}_{k}^{H} \right) \mathbf{w}_{p} \right|$$

$$= \max_{q < p} \left| \sum_{k=1}^{K} \langle \mathbf{w}_{q}, \mathbf{a}_{k} \rangle \langle \mathbf{a}_{k}, \mathbf{w}_{p} \rangle \right| \leq K \cdot C_{max}^{2},$$
(22)

where $p = N_a(n_d - 1) + n_a$, $q = N_a(n_d' - 1) + n_a'$, $\mathbf{w}_q = \mathbf{W}_{:,q}$, $\mathbf{a}_k = \mathbf{A}_{k:}^{\mathrm{H}}$ and $C_{max} = \max_{k,j} |\langle \mathbf{a}_k, \mathbf{w}_j \rangle|$. In (22), the third equality is due to $\mathbf{A}^{\mathrm{H}}\mathbf{A} = \sum_{k=1}^K \mathbf{a}_k \mathbf{a}_k^{\mathrm{H}}$. Moreover, the result in (22) resembles the concept of incoherence [37], which ensures sufficient diversity and randomness in the measurements and avoids concentrating the observations on partial bases. Specifically, incoherence requires that the rows of the measurement matrix should not be able to sparsely represent the columns of the sparse basis matrix, and vice versa [37]. It should be noted that the mutual coherence is jointly determined by the composition of the polar-domain dictionary and the selected antenna indices. The former is influenced by the parameter β , while the latter can be optimized through the antenna selection strategy. With β fixed, the key challenge in designing \mathbf{A} with minimal μ is to determine the optimal selection of $K = \frac{N}{\eta}$ elements for measurement. Accordingly, by taking the polar-domain dictionary as the sparse basis matrix, we further formulate the optimization problem as in (23). To address this problem, we propose a coherence-minimizing-based random pattern generation procedure, as detailed in **Algorithm 5**.

Initially, the set of available indices is defined as $\gamma = [1, \dots, N]$. The goal is to identify a sub-optimal random pattern of length $K = \frac{N}{\eta}$, selected from a total of \mathcal{C}_N^K possible patterns. To ensure computational feasibility, we first generate R random patterns, where R is a positive integer satisfying $R \ll \mathcal{C}_N^K$. Subsequently, in steps 3-7, the coherence of the sensing matrix corresponding to each random pattern is computed. Finally, the random pattern $\Xi_{\rm cmr}$ with the

$$\min_{n(\mathbf{\Xi}_{cmr})=K} \mu(\mathbf{\Psi}) = \min_{n(\mathbf{\Xi}_{cmr})=K} \max_{q < p} |\mathbf{\Psi}_{:,q}^{H} \mathbf{\Psi}_{:,p}| = \max_{q < p} |(\mathbf{A} \mathbf{w}_{q})^{H} (\mathbf{A} \mathbf{w}_{p})|$$

$$= \min_{n(\mathbf{\Xi}_{cmr})=K} \max_{q < p} \left| \sum_{n \in \mathbf{\Xi}} e^{jk_{c} \left(-nd \sin \theta_{n'_{a}} + n^{2}d^{2} \frac{\left(1 - \sin^{2} \theta_{n'_{a}} \right)}{2r_{n'_{d}}} \right)} e^{-jk_{c} \left(-nd \sin \theta_{n_{a}} + n^{2}d^{2} \frac{\left(1 - \sin^{2} \theta_{n_{a}} \right)}{2r_{n_{d}}} \right)} \right|$$

$$= \min_{n(\mathbf{\Xi}_{cmr})=K} \max_{q < p} \left| \sum_{n \in \mathbf{\Xi}_{cmr}} e^{-jk_{c} \left(nd \left(\sin \theta_{n'_{a}} - \sin \theta_{n_{a}} \right) + n^{2}d^{2} \left(\frac{1 - \sin^{2} \theta_{n'_{a}}}{2r_{n'_{d}}} - \frac{1 - \sin^{2} \theta_{n_{a}}}{2r_{n_{d}}} \right)} \right) \right|.$$
(23)

Algorithm 5 Coherence-minimizing-based Random Pattern Generation

Require: $\gamma = [1, \cdots, N], \mathbf{W}$.

Ensure: Ξ_{cmr} .

1: Randomly generate R patterns with length K from γ .

2: **for** $r \in \{1, 2, \cdots, R\}$ **do**

3: **for** k = 1 to K **do**

4: Set $\mathbf{A}_r(k, \mathbf{\Xi}_r(k)) = 1$.

5: end for

6: Compute $\Psi_r = \mathbf{A}_r \mathbf{W}$.

7: Calculate $\mu(\Psi_r)$ according to (22).

8: end for

9: Select Ξ_{cmr} according to (23).

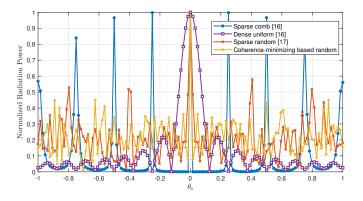
lowest coherence is selected from the R candidate patterns. This approach effectively balances computational efficiency with the goal of minimizing coherence for improved extrapolation performance.

C. Radiation Profile

To determine the most suitable pattern for channel extrapolation, we evaluate the performance of various pattern types by leveraging the radiation profile introduced in [16]. Specifically, the radiation power at any θ_s and r_s is calculated as

$$P(\theta_{\rm s}, r_{\rm s}) = \frac{|\dot{\mathbf{a}}_c^{\rm H}(\theta_{\rm s}, r_{\rm s})\dot{\mathbf{a}}_c(0, R_{\rm min})|}{\|\dot{\mathbf{a}}_c(\theta_{\rm s}, r_{\rm s})\|},\tag{24}$$

where we set $\theta = 0$, $r = R_{\min}$ and design down-sampling steering vector $\grave{\mathbf{a}}_c(\theta_{\rm s}, r_{\rm s})$ based on frequency f_c . In Fig. 3 and Fig. 4, we compare radiation profiles of four sampling patterns: the



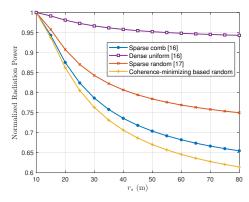


Fig. 3: Radiation profiles of various pattern types with respect to θ_s . Fig. 4: Radiation profiles of various pattern types with respect to r_s .

dense uniform pattern $\Xi_{\rm du}$, sparse comb pattern $\Xi_{\rm sc}$, sparse random pattern $\Xi_{\rm sr}$, and coherence-minimizing-based random pattern $\Xi_{\rm cmr}$, with a compression rate of $\eta=8$. We first examine the angular-domain radiation profile, assuming perfect knowledge of the distance. As shown in Fig. 3, over the angular range $\theta_{\rm s}\in[-1,1]$, a narrower main lobe centered at $\theta_{\rm s}=0$ indicates superior angular resolution. Ideally, side lobes should be significantly lower than the main lobe to effectively mitigate false alarms in the presence of nearby paths [16]. Subsequently, Fig. 4 illustrates the distance-domain radiation profiles under the assumption that the angular direction of each sampling point is known a priori. Over the distance range $r_{\rm s}\in[10,80]$, all patterns exhibit a distinct peak at the reference location with significantly reduced responses elsewhere. Importantly, a more rapid decay from the peak reflects enhanced distance resolution. This characteristic ensures that sampling points at other distances have minimal interference with the desired distance estimate.

- 1) Dense uniform pattern: It is evident that the dense uniform pattern exhibits a significantly wider main lobe compared to the other patterns, which is mainly determined by its aperture [16]. Moreover, this pattern achieves exceptionally low side lobe levels, rendering it highly effective at distinguishing the correct path in multipath environments. However, among the four patterns, it demonstrates the worst distance resolution. Nevertheless, as the number of active antennas increases, its distance resolution performance improves accordingly.
- 2) Sparse comb pattern: This pattern features a significantly narrower main lobe, precisely centered at $\theta_s = 0$. However, the compressed estimation range results in the emergence of η

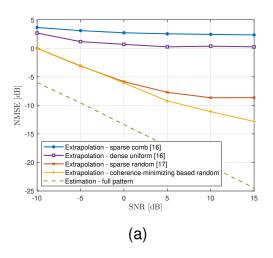
side lobes [16]. While the majority of side lobes remain suppressed, the η side lobes generate significant interference, complicating accurate identification of the desired propagation path. Moreover, unlike the far-field scenario, these side lobes exhibit varying levels of influence in the near-field, a difference attributed to the distinct modeling approaches for far-field and near-field channels [21]. Additionally, the sparse comb pattern demonstrates better distance resolution performance compared to the dense uniform pattern.

- 3) Sparse random pattern: We can clearly observe that this pattern exhibits a relatively narrow main lobe in the correct propagation direction, which is desirable for effective channel extrapolation. However, the maximum heights of its side lobes are higher than those of the dense uniform pattern. Additionally, due to its inherently random selection strategy, its distance resolution performance is unstable and generally worse than that of the sparse comb pattern.
- 4) Coherence-minimizing-based random pattern: By integrating the coherence-minimizing procedure in random pattern generation, this approach achieves a main lobe nearly as narrow as that of the sparse comb pattern, with its peak precisely at the target position $\theta_s = 0$. Furthermore, as illustrated in Fig. 4, it exhibits the best distance resolution performance among the four considered patterns, where the normalized radiation power decays more rapidly than that of the other patterns and demonstrates more stable behavior compared to the sparse random pattern. From Fig. 3 and Fig. 4, it is evident that the coherence-minimizing-based random pattern performs well in both angular and distance resolution, establishing it as a highly effective choice for channel extrapolation.

V. SIMULATION RESULTS

In this section, we evaluate the performance of our proposed near-field spatial-domain channel extrapolation algorithms through simulations under the CS framework. The simulation parameters are $f_c=28$ GHz, B=28 MHz, M=128, N=128, U=3. Therefore, the carrier wavelength is $\lambda_c=\frac{c}{f_c}=10.71$ mm, the antenna spacing is $d=\frac{\lambda_c}{2}=5.36$ mm, and the BS array aperture is $D=(N-1)d\approx Nd=\frac{N\lambda_c}{2}=0.69$ m. Then, it can be calculated that the Rayleigh distance is $\frac{2D^2}{\lambda_c}=88.91$ m [7].

The distances between the BS and near-field user or scatterers are randomly selected from



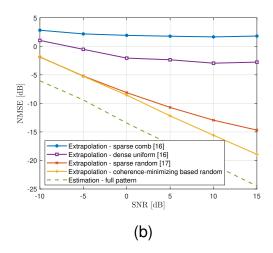


Fig. 5: NMSE performance against the SNR for different antenna selection pattern designs, with (a) $\eta = 8$, (b) $\eta = 4$.

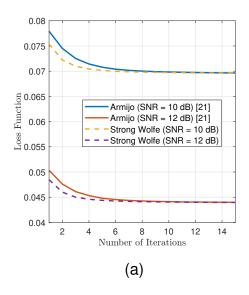
 $\mathcal{U}(10,80)^{10}$ m. In the design of the polar-domain dictionary, a parameter $\beta=1.5$ is employed to ensure adequate incoherence among near-field detection vectors. Throughout all simulation scenarios, the number of candidate patterns is set to R=10. In this paper, the signal-to-noise ratio (SNR) is given by $SNR=\frac{p_t}{\sigma^2}=\frac{1}{\sigma^2}$.

A. Pattern Designs Comparison

In Section IV-C, we evaluate the performance of different pattern types adopting radiation profiles. Next, we evaluate four antenna selection patterns, dense uniform, sparse comb, sparse random, and coherence-minimizing-based random patterns, in the channel extrapolation problem to assess their effectiveness. For a fair comparison, the channel extrapolation corresponding to all patterns is implemented using **Algorithm 1**.

Fig. 5 depicts the NMSE performance for both channel estimation and extrapolation under two settings (a) $\eta = 8$, (b) $\eta = 4$. These numerical results coincide well with the analytical results in Section IV-C. As expected, the random-position setups significantly outperform the deterministic-position setups. Among the deterministic setups, the dense uniform pattern exhibits

¹⁰In this paper, the Rayleigh distance is adopted as the boundary between the far-field and near-field regions. Accordingly, users located within the Rayleigh distance are treated as near-field users, and the study focuses on channel extrapolation by accounting for near-field characteristics. It is worth noting that other definitions for the near-field and far-field boundary also exist, as discussed in [38]. Moreover, hybrid-field scenarios, as a more general and practical case, will be addressed in future work.



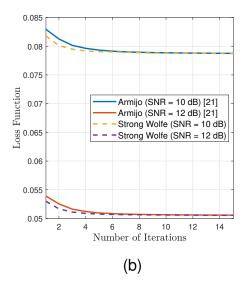


Fig. 6: Convergence analysis of objective function using different line search methods, with (a) $\eta = 8$, (b) $\eta = 4$.

better extrapolation performance than the sparse comb pattern, primarily due to its relatively lower side lobe levels. Although the sparse comb pattern exhibits an extremely narrow main lobe, it still suffers from η strong side lobes, which severely degrade extrapolation performance. Without the coherence-minimizing procedure, the sparse random pattern exhibits instability and generally performs worse than the coherence-minimizing-based random pattern, a result attributed to the inherent randomness of this strategy. Specifically, at an SNR of 15 dB, the performance gap between the sparse random pattern and the coherence-minimizing-based random pattern is 4.17 dB for $\eta=8$ and 4.25 dB for $\eta=4$. It is observed that the coherence-minimizing-based random pattern achieves the best extrapolation accuracy among all sparse patterns, although it remains inferior to channel estimation exploiting the full antenna array. This performance gain stems from its coherence-minimizing strategy, which generates a diverse set of candidate patterns and selects the one with the lowest mutual coherence, thereby enhancing angular and distance resolution and ensuring stable performance.

1) Convergence: The convergence behavior of the proposed P-ASIGW algorithm with different line search methods has been compared in Fig. 6 when (a) $\eta = 8$, (b) $\eta = 4$. The objective function $\|\mathbf{H}^{\mathcal{B}} - \tilde{\Psi}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})\hat{\mathbf{G}}\|_F^2$ exhibits a monotonic decrease over iterations. This indicates that the steps for updating $\hat{\boldsymbol{\theta}}$, $\hat{\mathbf{r}}$, and $\hat{\mathbf{G}}$ are monotonically non-increasing, ensuring that the alternating optimization procedure will converge, as supported by [21]. Note that the line search method based on the

strong Wolfe conditions achieves faster convergence than the Armijo condition-based method, as it incorporates an additional curvature condition, which facilitates the determination of an appropriate step size. Therefore, the strong Wolfe conditions-based method requires only 5 iterations on average, and the Armijo condition-based method necessitates 10 iterations on average to achieve comparable results.

2) Complexity: For the proposed P-ASOMP as stated in Algorithm 1, the computational complexity is dominated by operations of the SOMP procedure, i.e., steps 2-9 in Algorithm 1. Then we calculate the complexity of each step. Since $\Psi \in \mathbb{C}^{K \times N_a N_d}$, $\mathbf{R} \in \mathbb{C}^{K \times M}$, and $\mathbf{H}^{\mathcal{B}} \in \mathbb{C}^{K \times M}$, the computational complexities from steps 3-8 are $\mathcal{O}(1)$, $\mathcal{O}(KN_aN_dM)$, $\mathcal{O}(N_aN_dM)$, $\mathcal{O}(1)$, $\mathcal{O}(\hat{L}^2KM + \hat{L}^3)$, and $\mathcal{O}(\hat{L}KM)$, respectively. Generally, the number of paths \hat{L} is relatively small, thus the computation complexity from steps 3-8 is determined by $\mathcal{O}(KN_aN_dM)$ + $\mathcal{O}(\hat{L}^2KM)$. After \hat{L} iterations, the overall computational complexity is $\mathcal{O}(\hat{L}KN_aN_dM) + \mathcal{O}(\hat{L}^3KM)$. For the P-ASIGW algorithm in Algorithm 2, the total complexity includes the same initialization cost as P-ASOMP and a refinement stage with dominant operations in updating $\hat{\theta}$, $\hat{\mathbf{r}}$, $\hat{\mathbf{G}}$. As shown in (13), the complexity of the update $\hat{\mathbf{G}}$ is $\mathcal{O}(\hat{L}^2KM + \hat{L}^3)$. Furthermore, the complexities of (15) and (16) are determined to be $\mathcal{O}(K^2M + KM^2 + \hat{L}^2KM)$, benefiting from the small value of \hat{L} . After N_{iter} iterations, the complexity in the refinement stage becomes $\mathcal{O}((K^2M+KM^2+KM^2+KM^2))$ $\hat{L}^2KM)N_{\text{iter}}$). Consequently, the overall computational complexity of the proposed P-ASIGW method is $\mathcal{O}(\hat{L}KN_aN_dM) + \mathcal{O}(\hat{L}^3KM) + \mathcal{O}((K^2M + KM^2 + \hat{L}^2KM)N_{\text{iter}})$. Similarly, the proposed CV-P-ASOMP is stated in Algorithm 3 exhibits the same structure as P-ASOMP, but restricts correlation operations to training subcarriers ($T \ll M$), reducing the complexity to $\mathcal{O}(\hat{L}KN_aN_dT+\hat{L}^3KM)$. Finally, the proposed CV-P-ASIGW in **Algorithm 4** combines the initialization cost from CV-P-ASOMP and a refinement process similar to P-ASIGW, resulting in an overall complexity of $\mathcal{O}(\hat{L}KN_aN_dT) + \mathcal{O}(\hat{L}^3KM) + \mathcal{O}((K^2M + KM^2 + \hat{L}^2KM)N_{\text{iter}})$.

B. Evaluation of Channel Extrapolation Algorithms

Having verified the effectiveness of coherence-minimizing-based random pattern design, we further compare our proposed algorithms with state-of-the-art OMP-based methods, such as P-OMP [8], P-SOMP [21], and P-SIGW [21], exploiting this pattern design. The normalized mean square error (NMSE) of the extrapolated channel is calculated by $NMSE = \mathbb{E}(\frac{||\mathbf{H}^{\mathcal{A}} - \hat{\mathbf{H}}^{\mathcal{A}}||_F^2}{||\mathbf{H}^{\mathcal{A}}||_F^2})$.

Moreover, we test the system achievable rate using the extrapolated channel matrix. Let the extrapolated uplink channel matrix on the m-th subcarrier be denoted by $\hat{\mathbf{H}}_m = \left[\hat{\mathbf{h}}_{m1}, \hat{\mathbf{h}}_{m2}, \dots, \hat{\mathbf{h}}_{mU}\right] \in$

 $\mathbb{C}^{N\times U}$, where $\hat{\mathbf{h}}_{mu}\in\mathbb{C}^{N\times 1}$ denotes the extrapolated channel vector between the u-th UE and the BS. The actual channel of the x-th user on the m-th subcarrier is represented by $\mathbf{h}_{mx}\in\mathbb{C}^{N\times 1}$, and the transmitted uplink signal from UE u is $s_{mu}\sim\mathcal{CN}(0,p_t)$, where $p_t=\mathbb{E}[|s_{mu}|^2]$. Specifically, the zero-forcing (ZF) combining vector for the u-th UE is then given by $\mathbf{v}_{mu}=\left[\hat{\mathbf{H}}_m\left(\hat{\mathbf{H}}_m^H\hat{\mathbf{H}}_m\right)^{-1}\right](:,u)$. Based on this, the ergodic achievable uplink rate $R_u\in\mathbb{C}$ for the u-th UE can be formulated as (25), where the expectation is with respect to the channel estimates [39], and $\tilde{\mathbf{h}}_{mu}\in\mathbb{C}^{N\times 1}$ denotes the channel extrapolation error, i.e., $\mathbf{h}_{mu}=\hat{\mathbf{h}}_{mu}+\tilde{\mathbf{h}}_{mu}$.

$$R_{u} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E} \left\{ \log_{2} \left(1 + \frac{p_{mu} |\mathbf{v}_{mu}^{H} \hat{\mathbf{h}}_{mu}|^{2}}{p_{mu} |\mathbf{v}_{mu}^{H} \hat{\mathbf{h}}_{mu}|^{2} + \sum_{x=1, x \neq u}^{U} p_{mx} |\mathbf{v}_{mu}^{H} \mathbf{h}_{mx}|^{2} + \sigma^{2} ||\mathbf{v}_{mu}||^{2}} \right) \right\}.$$
(25)

Fig. 7 and Fig. 8 present the channel extrapolation accuracy and processing efficiency against the ratio of cross-validation α at the SNR of 0 dB, for $\eta = 8$ and $\eta = 4$, respectively. The channel extrapolation accuracy, which quantifies the gap between P-ASOMP and CV-P-ASOMP, is calculated as $(1 - \frac{|\text{NMSE(P-ASOMP)} - \text{NMSE(CV-P-ASOMP)}|}{\text{NMSE(P-ASOMP)}}) \times 100\%$. The processing efficiency is computed as $\frac{t_{\min}}{t_{\alpha}} \times 100\%$, where t_{\min} represents the minimal running time across different CV ratios, and t_{α} denotes the running time for the current CV ratio α . As illustrated in both Fig. 7 and Fig. 8, $\alpha = 0.3$ serves as a critical point, achieving relatively high processing efficiency while maintaining an extrapolation accuracy exceeding 90%. It is evident that larger values of α tend to reduce processing efficiency due to the increased computational cost associated with highdimensional matrix operations. Conversely, excessively small values of α result in insufficient training samples, which may degrade the accuracy of channel extrapolation. Moreover, inadequate training data may cause an overestimation of the number of paths, thereby increasing the number of iterations and further exacerbating computational complexity, as shown in Fig. 8. Hence, to strike an effective balance between computational efficiency and extrapolation accuracy, the cross-validation ratio α is optimally chosen within the range [0.3, 0.5]. In the sequel, we adopt $\alpha = 0.3$ to ensure higher processing efficiency.

The average uplink achievable rates of various on-grid methods are plotted in Fig. 9, while those of off-grid methods are shown in Fig. 10, with U=3 and $\eta=4$, respectively. The first observation is that the proposed P-ASOMP and CV-P-ASOMP algorithms outperform

¹¹To mitigate inter-user interference, ZF combining is employed at the receiver in this work.

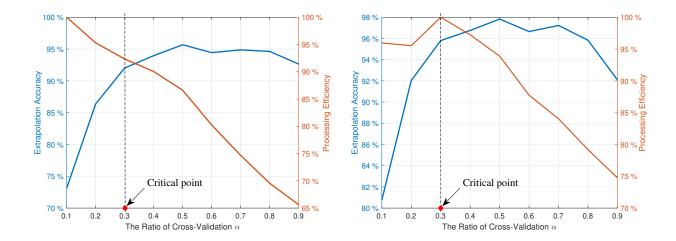


Fig. 7: Comparison of extrapolation accuracy and processing efficiency against the CV ratio, with $\eta = 8$. cessing efficiency against the CV ratio, with $\eta = 4$.

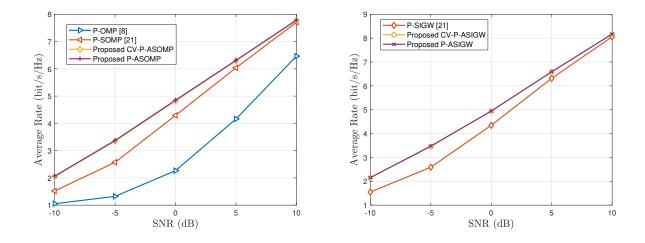


Fig. 9: The average rates of various on-grid methods Fig. 10: The average rates of various off-grid methods against the SNR, with $\eta = 4$, U = 3.

existing on-grid methods in terms of achievable rates across all SNR regimes, demonstrating that the adaptive channel extrapolation scheme can effectively exploit the inherent channel sparsity. Moreover, the proposed P-ASOMP and CV-P-ASOMP algorithms exhibit comparable extrapolation performance to the P-SOMP [21] algorithm under high SNR conditions. This is because the iteration number in the P-SOMP [21] method is specifically set for high SNR scenarios. Based on Fig. 9, it can be observed that the on-grid P-ASOMP and CV-P-ASOMP algorithms exhibit nearly identical performance in terms of average rates. Similarly, as shown

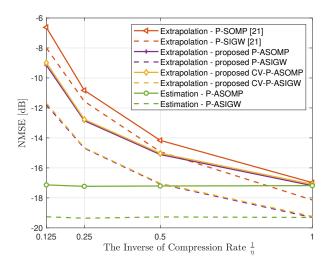
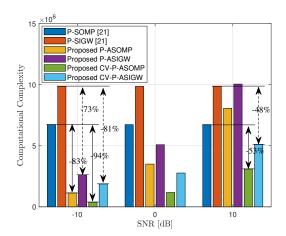


Fig. 11: NMSE values of different channel extrapolation algorithms across various compression rates η .

in Fig. 10, the off-grid P-ASIGW and CV-P-ASIGW algorithms also demonstrate almost the same performance. This result highlights the effectiveness of the proposed CV-based scheme in significantly reducing computational complexity while preserving high extrapolation accuracy.

In Fig. 11, the NMSE performance of the on-grid and off-grid methods is plotted against the compression ratio η at a SNR of 5 dB. As illustrated in Fig. 6, the line search method based on the strong Wolfe conditions exhibits faster convergence compared to its counterpart relying on the Armijo condition [21]. Given that the P-SIGW [21] algorithm employs an Armijo condition-based line search, it requires a higher number of iterations for convergence. In contrast, the proposed P-ASIGW and CV-P-ASIGW algorithms incorporate a strong Wolfe conditionsbased line search strategy, which significantly reduces the required number of iterations for convergence. In this paper, the maximum iteration count is set to 10 for the P-SIGW [21] algorithm and reduced to 5 for the proposed P-ASIGW and CV-P-ASIGW algorithms. By refining angle and distance parameters with significantly higher resolution, the off-grid algorithms achieve substantially improved NMSE performance compared to the on-grid algorithms. Specifically, when the inverse of the compression rate is set to 0.5, the performance gap between P-SOMP [21] and P-SIGW [21] is approximately 0.76 dB. In contrast, the performance gaps between P-ASOMP and P-ASIGW, as well as between CV-P-ASOMP and CV-P-ASIGW, both reach 2.01 dB. Furthermore, it is observed that the proposed adaptive off-grid algorithms, namely P-ASIGW and CV-P-ASIGW, clearly outperform the fixed-iteration off-grid algorithm P-SIGW [21]. This



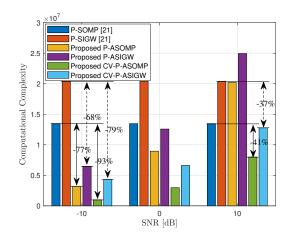


Fig. 12: Computational complexity of various channel extrapolation algorithms, with $\eta = 8$.

Fig. 13: Computational complexity of various channel extrapolation algorithms, with $\eta = 4$.

improvement is primarily attributed to the adaptive mechanism, which enables more accurate initial value estimation, thereby enhancing the overall reconstruction accuracy.

Fig. 12 and Fig. 13 illustrate the computational complexity of various channel extrapolation algorithms, with $\eta = 8$ for Fig. 12, and $\eta = 4$ for Fig. 13. Since the optimal number of iterations varies across the SNR range, the computational complexity of the proposed P-ASOMP, P-ASIGW, CV-P-ASOMP, and CV-P-ASIGW algorithms increases with rising SNR. In low SNR scenarios, the proposed on-grid and off-grid algorithms exhibit superior performance by employing a well-designed stopping criterion, which minimizes the need for extensive iterations. As shown in Fig. 12, at an SNR of -10 dB, the P-ASOMP and CV-P-ASOMP algorithms reduce computational complexity by approximately 83% and 94% compared to the P-SOMP [21] method, respectively. For the off-grid methods, the P-ASIGW and CV-P-ASIGW algorithms achieve reductions of around 73% and 81% compared to the P-SIGW [21] method, respectively. As depicted in Fig. 11 and Fig. 12, the proposed adaptive scheme effectively captures the sparsity condition, while the CV-based adaptive scheme achieves satisfactory extrapolation performance with low computational complexity. At a SNR of 10 dB, the computational complexity of the proposed P-ASOMP and P-ASIGW algorithms increases significantly due to the rise in iteration numbers. However, the CV-based scheme effectively alleviates this issue, achieving comparable extrapolation performance while reducing computational complexity by approximately 50%. Additionally, computational complexity increases as the number of active elements grows, which

is attributed to higher-dimensional matrix computations. In this case, the on-grid and off-grid methods based on the CV scheme still demonstrate significant advantages. As shown in Fig. 13, even in the 10 dB scenarios with a high number of iterations, the CV-P-ASOMP and CV-P-ASIGW algorithms reduce computational complexity by about 41% and 37% compared to the P-SOMP [21] and P-SIGW [21] methods, respectively. Notably, by selecting an appropriate CV ratio, the CV-P-ASOMP algorithm achieves performance comparable to the P-ASOMP algorithm while significantly improving processing efficiency. This is attributed to the shared support sets across all subcarriers, enabling the CV-based algorithms to reduce high-dimensional matrix computations in the correlation step by leveraging shared information in the polar-domain.

VI. CONCLUSION

In this paper, we formulated the channel extrapolation problem as a sparse signal recovery task and proposed a compressed sensing-based spatial-domain channel extrapolation framework for mmWave dynamic XL-MIMO communication systems. Then, we developed both on-grid algorithms, namely P-ASOMP and CV-P-ASOMP, and off-grid algorithms, namely P-ASIGW and CV-P-ASIGW, for effective near-field channel extrapolation with low computational complexity. Under this framework, the measurement matrix was constructed based on a strategically selected subset of antenna elements. By leveraging a pre-designed polar-domain dictionary to accurately characterize the inherent spatial correlation of near-field channels, we observed that the mutual coherence of the sensing matrix is primarily determined by the antenna selection pattern. Since lower mutual coherence directly implies improved reconstruction performance, we introduced a coherence-minimizing-based random pattern design to enhance extrapolation accuracy. Building on this pattern design, we evaluated the proposed on-grid and off-grid algorithms. In particular, the off-grid P-ASIGW and CV-P-ASIGW methods incorporate a strong Wolfe conditions-based line search to accelerate convergence. Notably, the proposed CV-based scheme can maintain satisfactory channel extrapolation accuracy under the MMV framework, while significantly reducing computational complexity compared to existing methods.

REFERENCES

[1] Z. Wang, J. Zhang, E. Björnson, D. Niyato, and B. Ai, "Optimal bilinear equalizer for cell-free massive MIMO systems over correlated Rician channels," *IEEE Trans. Signal Process.*, to appear, 2025.

- [2] Y. Chen, X. Guo, G. Zhou, S. Jin, D. W. K. Ng, and Z. Wang, "Unified far-field and near-field in holographic MIMO: A wavenumber-domain perspective," *IEEE Commun. Mag.*, vol. 63, pp. 30–36, Jan. 2025.
- [3] Z. Wang, J. Zhang, H. Du, D. Niyato, S. Cui, B. Ai, M. Debbah, K. B. Letaief, and H. V. Poor, "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Commun. Surveys Tuts.*, vol. 26, pp. 1560–1605, 3rd Quart. 2024.
- [4] L. U. Khan, W. Saad, D. Niyato, Z. Han, and C. S. Hong, "Digital-twin-enabled 6G: Vision, architectural trends, and future directions," *IEEE Commun. Mag.*, vol. 60, pp. 74–80, Jan. 2022.
- [5] J. Lu, J. Zhang, H. Lei, H. Xiao, Z. Lu, and B. Ai, "Hybrid-field beam-split pattern detection-based channel estimation for THz XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 73, pp. 13944–13949, Sep. 2024.
- [6] Z. Wang, J. Zhang, H. Du, W. E. I. Sha, B. Ai, D. Niyato, and M. Debbah, "Extremely large-scale MIMO: Fundamentals, challenges, solutions, and future directions," *IEEE Wireless Commun.*, vol. 31, pp. 117–124, Jun. 2024.
- [7] K. T. Selvan and R. Janaswamy, "Fraunhofer and Fresnel Distances: Unified derivation for aperture antennas," *IEEE Antennas Propag. Mag.*, vol. 59, pp. 12–15, Aug. 2017.
- [8] M. Cui and L. Dai, "Near-field wideband channel estimation for extremely large-scale MIMO," Sci. China Inf. Sci., vol. 66, p. 172303, Jun. 2023.
- [9] J. Tan and L. Dai, "THz precoding for 6G: Challenges, solutions, and opportunities," *IEEE Wireless Commun.*, vol. 30, pp. 132–138, May 2023.
- [10] Z. Liu, J. Zhang, Z. Liu, H. Du, Z. Wang, D. Niyato, M. Guizani, and B. Ai, "Cell-free XL-MIMO meets multi-agent reinforcement learning: Architectures, challenges, and future directions," *IEEE Wireless Commun.*, vol. 31, pp. 155–162, Aug. 2024.
- [11] J. Lu, J. Zhang, H. Lei, H. Xiao, Z. Lu, and B. Ai, "Subarray-aware beam-split pattern detection for near-field XL-MIMO channel estimation," *IEEE Trans. Veh. Technol.*, vol. 74, pp. 8333–8338, May 2025.
- [12] M. Alrabeiah and A. Alkhateeb, "Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, pp. 1465–1470, Nov. 2019.
- [13] Z. Liu, H. Du, J. Lin, Z. Gao, L. Huang, S. Hosseinalipour, and D. Niyato, "DNN partitioning, task offloading, and resource allocation in dynamic vehicular networks: A Lyapunov-Guided diffusion-based reinforcement learning approach," *IEEE Trans. Mobile Comput.*, vol. 24, pp. 1945–1962, Mar. 2025.
- [14] B. Lin, F. Gao, S. Zhang, T. Zhou, and A. Alkhateeb, "Deep learning-based antenna selection and CSI extrapolation in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, pp. 7669–7681, Nov. 2021.
- [15] Z. Zhang, J. Zhang, Y. Zhang, L. Yu, and G. Liu, "AI-based time-, frequency-, and space-domain channel extrapolation for 6G: Opportunities and challenges," *IEEE Veh. Technol. Mag.*, vol. 18, pp. 29–39, Jan. 2023.
- [16] Y. Han, S. Jin, X. Li, C.-K. Wen, and T. Q. S. Quek, "Multi-domain channel extrapolation for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 69, pp. 8534–8550, Dec. 2021.
- [17] J. Tian, Y. Han, S. Jin, X. Li, J. Zhang, and M. Matthaiou, "Near-field channel reconstruction in sensing RIS-assisted wireless communication systems," *IEEE Trans. Wireless Commun.*, vol. 23, pp. 12223–12238, Sep. 2024.
- [18] X. Xu, S. Zhang, F. Gao, and J. Wang, "Sparse bayesian learning based channel extrapolation for RIS assisted MIMO-OFDM," *IEEE Trans. Commun.*, vol. 70, pp. 5498–5513, Aug. 2022.
- [19] B. Zhou, X. Yang, S. Ma, F. Gao, and G. Yang, "Low-overhead channel estimation via 3D extrapolation for TDD mmwave massive MIMO systems under high-mobility scenarios," *IEEE Trans. Wireless Commun.*, vol. 24, pp. 2797–2813, Apr. 2025.
- [20] H. Li, M. Wang, R. Han, N. Wang, H. Wu, Y. Gu, W. Yuan, and F. Gao, "Deep learning-based channel extrapolation for

- 5G advanced massive MIMO: Hardware prototype and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 24, pp. 1756–1771, Jan. 2025.
- [21] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?," *IEEE Trans. Commun.*, vol. 70, pp. 2663–2677, Apr. 2022.
- [22] S. Yue, S. Zeng, L. Liu, Y. C. Eldar, and B. Di, "Hybrid near-far field channel estimation for holographic MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 23, pp. 15798–15813, Nov. 2024.
- [23] J. Xu, W. Xu, D. W. K. Ng, and A. L. Swindlehurst, "Secure communication for spatially sparse millimeter-wave massive MIMO channels via hybrid precoding," *IEEE Trans. Commun.*, vol. 68, pp. 887–901, Feb. 2020.
- [24] J. Lu, J. Zhang, H. Lei, H. Xiao, and B. Ai, "Adaptive hybrid-field channel estimation in THz XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 74, pp. 11557–11562, Jul. 2025.
- [25] C. Tan, D. Cai, F. Fang, J. Shan, Y. Xu, Z. Ding, and P. Fan, "Hierarchical sparse estimation of non-stationary channel for uplink massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, pp. 6640–6645, Dec. 2023.
- [26] S. Liu, X. Yu, Z. Gao, J. Xu, D. W. K. Ng, and S. Cui, "Sensing-enhanced channel estimation for near-field XL-MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 43, pp. 628–643, Mar. 2025.
- [27] Z. Lu, Y. Han, S. Jin, M. Matthaiou, and T. Q. S. Quek, "Near-field channel reconstruction and user localization for ELAA systems," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, pp. 1–6, Oct. 2022.
- [28] Z. Xiao, S. Cao, L. Zhu, Y. Liu, B. Ning, X.-G. Xia, and R. Zhang, "Channel estimation for movable antenna communication systems: A framework based on compressed sensing," *IEEE Trans. Wireless Commun.*, vol. 23, pp. 11814–11830, Sep. 2024.
- [29] H. Lei, J. Zhang, Z. Wang, H. Xiao, B. Ai, and E. Björnson, "Near-field user localization and channel estimation for XL-MIMO systems: Fundamentals, recent advances, and outlooks," *IEEE Wireless Commun.*, vol. 32, pp. 190–198, Aug. 2025.
- [30] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," Cambridge, MA, USA: Birkhauser, 2013.
- [31] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," C. R. Math., vol. 346, pp. 589–592, May. 2008.
- [32] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmonic Anal.*, vol. 31, pp. 59–73, Jul. 2011.
- [33] C. Huang, J. Xu, W. Xu, X. You, C. Yuen, and Y. Chen, "Low-complexity channel estimation for extremely large-scale MIMO in near field," *IEEE Wireless Commun. Lett.*, vol. 13, pp. 671–675, Mar. 2024.
- [34] T. Gong, S. Zhang, M. Xu, Z. Li, and O. A. Dobre, "Lightweight TCN-based spatial-temporal channel extrapolation for RIS-aided communication," *IEEE Trans. Veh. Technol.*, vol. 73, pp. 12219–12224, Aug. 2024.
- [35] M. Dai, T. Gong, and S. Zhang, "Graph convolutional network-based channel extrapolation for hybrid RIS-aided communication," *IEEE Wireless Commun. Lett.*, vol. 13, pp. 3558–3562, Oct. 2024.
- [36] R. G. Baraniuk, M. A. Davenport, R. A. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, pp. 253–263, Jan. 2008.
- [37] R. G. Baraniuk, "Compressive sensing [Lecture Notes]," IEEE Signal Process. Mag., vol. 24, pp. 118–121, Jul. 2007.
- [38] S. Zeng, B. Di, H. Zhang, Z. Han, and H. Vincent Poor, "Revisiting near-far field boundary in dual-polarized XL-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 24, pp. 7269–7285, Sep. 2025.
- [39] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," Found. Trends Signal Process., vol. 11, no. 3-4, pp. 154–655, 2017.