Compressed Sensing for Sparse Channel Estimation in Next Generation Radios

Addressing Hardware Impairments

Hamed Masoumi

Compressed Sensing for Sparse Channel Estimation in Next Generation Radios

Addressing Hardware Impairments

Compressed Sensing for Sparse Channel Estimation in Next Generation Radios

Addressing Hardware Impairments

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. H. Bijl,
chair of the Board for Doctorates
to be defended publicly on
Monday 30, March 2026 at 10:00

by

Hamed MASOUMI

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Dr. ing. R. Van de Plas,	Delft University of Technology, *promotor*
Dr. N.J. Myers,	Delft University of Technology, *copromotor*
Independent members:	
Prof. dr. ir. A.J. van der Veen,	Delft University of Technology
Prof. dr. ing. N. Llombart Juan,	Delft University of Technology
Prof. dr. H. Wymeersch,	Chalmers University of Technology, Sweden
Prof. dr. U. Spagnolini,	Politecnico di Milano, Italy
Prof. dr. ir. B. de Schutter,	Delft University of Technology, reserve member

Keywords: Compressed sensing, mm-Wave, THz, beamforming, channel estimation, phased array, sector sweep, belief propagation, sparse recovery, hardware impairment, oscillators, phase noise, I/Q imbalance, model mismatch

Printed by: Koninklijke Rijnja

Cover by: Gemini and ChatGPT

ISBN 978-94-6518-286-5

An electronic copy of this dissertation is available at
https://repository.tudelft.nl/.

Contents

Summary

Millimeter wave (mmWave) bands, currently used in 5G, offer significant spectrum to enable Gbps data rates for emerging data-hungry applications. At mmWave and higher bands, however, high scattering and atmospheric attenuation result in low received power. To address this issue, base stations employ large antenna arrays and periodically configure these arrays by learning the propagation environment, called the wireless channel. The use of such large arrays, however, makes the learning process, i.e., wireless channel estimation, extremely challenging. This is because, on the one hand, the training overhead of classical channel estimation methods increases significantly with the array dimensions. On the other hand, compressed sensing (CS)-based methods for fast channel estimation suffer from a poor signal-to-noise ratio (SNR) in the received measurements. Moreover, radio frequency (RF) impairments, which are much more severe in mmWave systems than in low-frequency systems, further complicate channel estimation. For example, low-resolution RF phase shifters, phase noise at the oscillator, and in-phase/quadrature (IQ) mismatch at the down-conversion are among the practically significant impairments.

The first part of this dissertation tackles the poor SNR issue with standard CS-based channel estimation methods using low-resolution phased arrays. To this end, we propose our novel in-sector CS solution that is fully compliant with the IEEE802.11ad/ay compliant devices. The key idea is to partition the angle domain channel into multiple non-overlapping sectors, and then estimate the channel within the sector associated with the highest received power. A big challenge in this approach lies in designing a low-resolution beam codebook to acquire channel measurements in these non-overlapping sectors. To address this, we propose a framework to design a beam codebook that is compatible with low-resolution phased arrays whose resolution can be as low as 1 bit. Since these beams focus energy within specific sectors, they achieve higher SNR within sectors compared to standard wide beams. To acquire CS measurements within the best sector, a convolutional CS-based approach is developed, and the corresponding CS matrix is optimized. With our optimized design, the aliasing artifacts in CS are pushed outside the sector of interest. The proposed codebook results in a higher received SNR than the state-of-the-art sector sweep codebooks, and our optimized CS matrix achieves a better in-sector channel reconstruction than comparable benchmarks.

The second part of this dissertation addresses the problem of channel estimation under phase noise, which arises due to jitters and instabilities at the oscillators. Phase noise perturbs the phase of the acquired channel measurements, causing CS algorithms that ignore such phase errors to fail due to a model mismatch. In practice, the phase errors are similar within a batch of measurements acquired in a short burst, and the errors vary significantly across different batches, resulting in partially coherent measurements. To estimate the channel under such partially coherent phase noise, we develop a graphical

inference method, called message passing-based inference, that exploits the sparse structure of the channel and the phase noise statistics for channel estimation. The nodes in our graph model the unknown wireless channel to be estimated. To deal with phase noise, we absorb the phase error in each batch into the channel, which results in a collection of vectors that are all correlated, i.e., they share the same support and magnitude. Our method incorporates the shared magnitude and the Wiener phase noise statistics in addition to the shared support and sparsity of the channel to infer the channel. The proposed method provides a better channel estimate compared to state-of-the-art methods robust to phase noise.

The last part of this dissertation solves the problem of channel estimation when CS measurements are distorted by in-phase and quadrature-phase (IQ) imbalance at the oscillator. We investigate how the mismatch due to IQ imbalance impacts the channel estimated with a standard CS algorithm. We show that the angle domain channel recovered with convolutional CS has aliased components located at angles that are symmetric about the boresight. Next, we introduce the *augmented CS* model for measurements under IQ imbalance to develop our channel estimation method robust to IQ imbalance. In our *augmented CS* formulation, we construct an auxiliary vector composed of the IQ imbalance parameter and the angle domain channel. The auxiliary vector, whose dimension is twice the dimension of the channel, is referred to as *augmented beamspace*. We show that this vector exhibits a special sparse structure referred to as the *paired support* structure. Then, we develop a custom greedy algorithm called *paired-support orthogonal matching pursuit* (PSOMP) that exploits the *paired support* structure of the augmented beamspace vector for sparse recovery. Our solution is shown to result in a lower channel estimation error than the baseline methods.

1 Introduction

1.1. A Brief Background on Channel Estimation for Millimeter-wave Systems

Millimeter wave (mmWave) radio frequency (RF) systems will be an integral part of the sixth generation (6G) wireless networks [1]. These systems can achieve Gbps of data rates, as they exploit the wide bandwidths available at these frequency bands (30-300 GHz) [1]. To achieve sufficient signal strength at the receiver, in these systems, typically, an array of antennas with a large number of elements is used to focus the transmitter's energy towards the receiver, a procedure known as beamforming [2]. Fortunately, the short wavelengths of the signals at these frequencies allow the incorporation of many antenna elements into a small physical space [3]. An important

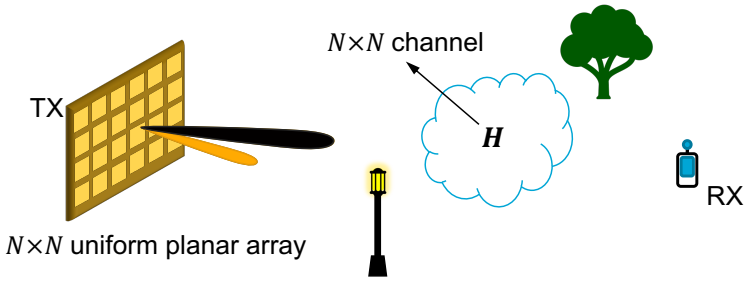


Figure 1.1: A wireless system with an $N \times N$ uniform planar array at the TX, and a single antenna receiver. The $N \times N$ matrix $\mathbf{H}$ with complex-valued entries models the propagation environment, and it is called the channel.

approach to perform beamforming is by learning the propagation environment. In wireless systems, the propagation environment is commonly called a *channel*. Due to the use of a large number of antenna elements, compared to typical low-frequency RF systems, mmWave channels have a higher dimension. An example of a mmWave system with an $N \times N$ uniform planar array (UPA) at the transmitter (TX) and a single antenna receiver (RX) is shown in Fig. 1.1. The $N \times N$ channel matrix $\mathbf{H}$ in Fig. 1.1 models the propagation

environment. The channel dimension N^2, equivalently the number of antenna elements, can be on the order of hundreds to hundreds of thousands of elements [4]. This high channel dimension leads to new challenges in channel estimation in mmWave RF systems.

Channel estimation techniques for mmWave systems differ notably from traditional methods developed for lower frequency systems. This distinction arises because of high scattering at these frequencies, resulting in channels that are sparse in a suitable basis [3]. Specifically, for a typical mmWave system with uniform linear/rectangular arrays, the discrete Fourier transform (DFT) dictionary is a reasonable choice as a sparsifying basis [5]. That means that the 2D-DFT of the channel matrix **H** in Fig. 1.1 contains a few nonzero entries and most of its entries are zero. The DFT of the channel **H** is commonly called the *angle-domain channel* or *beamspace* channel [5]. In Fig. 1.2, an example of a channel obtained from the NYU simulator [6] and its beamspace is shown. It is observed that the beamspace channel is sparse, containing two nonzero entries with significant magnitude.

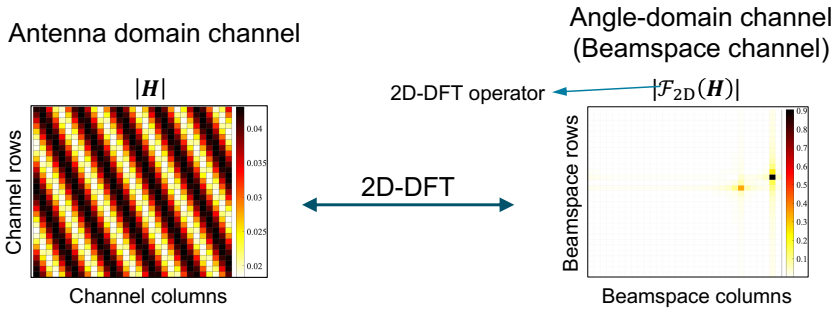


Figure 1.2: An example of a wireless channel for the wireless system in Fig. 1.1 with 32 × 32 uniform planar array at the TX generated from the NYU simulator [6]. The beamspace channel corresponding to **H** is sparse.

The sparse nature of the mmWave channels can be exploited for channel estimation. Methods based on compressive sensing (CS) [7] are popular approaches that leverage the sparse structure for channel estimation using fewer channel measurements than the array dimension. This reduces the training overhead for channel estimation in mmWave systems.

We briefly explain how CS-based mmWave channel estimation is performed. To estimate the channel, the transmitter excites the environment with M beams in sequence. For each of these M beams called training beams, the receiver acquires a channel measurement. During beam training, it is assumed that the channel remains the same. Let the $N^2 \times 1$ vector **x** denote the vector version of the unknown beamspace channel, i.e., **x** is obtained by stacking the columns of the beamspace channel (e.g., the matrix in Fig. 1.2). Since the beamspace **x** is sparse, most of its entries are zero. The set of indices of **x** with nonzero values is called the *support* of **x**. We use **y** to denote the $M \times 1$ vector of CS measurements and **v** to denote the additive Gaussian noise with

independent and identically distributed (IID) entries to write

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{1.1}$$

where $\mathbf{A}$ is a known $M \times N^2$ matrix referred to as the *CS matrix*. This matrix is a function of the 2D-DFT matrix and the beams applied at the UPA to acquire M measurement.

The standard problem of mmWave channel estimation is to estimate $\mathbf{x}$ from M measurements given by $\mathbf{y}$ in (1.1). Since the number of measurements M is smaller than the dimension of the unknown beamspace $\mathbf{x}$, the problem of estimating $\mathbf{x}$ from (1.1) is ill-posed. CS algorithms such as orthogonal matching pursuit (OMP) [8] and approximate message passing (AMP) [9], however, can estimate the unknown beamspace $\mathbf{x}$ by exploiting its sparse structure.

1.2. Hardware impairments and constraints in mmWave systems: Challenges and prior work

We now discuss common hardware impairments in the RF chain of mmWave systems. Specifically, we discuss challenges due to low-resolution phase shifters, phase noise, in-phase/quadrature (IQ) mismatch, carrier frequency offset, power amplifier nonlinearity, and mutual coupling between antenna array elements.

The use of high-resolution phase shifters in mmWave systems is not practically desirable due to their high energy consumption and hardware complexity [10, 11]. For instance, a 1-bit phase shifter may consume around 10 mW [12], whereas a 4-bit phase shifter may need 45 mW for applications in 60 GHz of carrier frequency [13]. Although low-resolution phase shifters are desirable due to their reduced power consumption, their use limits the degrees of freedom in beam design. The reduced degrees of freedom translates to constraints on $\mathbf{A}$ in (1.1). In Chapter 2, we develop a framework for designing directional beams, equivalently $\mathbf{A}$ in (1.1), using low-resolution phase shifters that are suitable for our proposed in-sector sparse channel estimation.

Phase noise is an important RF impairment caused by noise in active and lossy elements of local oscillators [14] leading to jitter in the oscillation period of the generated sinusoidal wave [15]. Phase noise causes random perturbations in the phase of the channel measurements [16]. This effect is more pronounced in systems operating at mmWave and higher carrier frequencies than in systems operating at lower carrier frequencies [17, 18]. Therefore, phase noise in these systems significantly alters the received signal model compared to the standard CS model, which only accounts for additive noise in the linear measurements. We address channel estimation in the presence of phase noise in Chapter 3, where we develop an algorithm that estimates the sparse channel exploiting the statistics of phase noise.

In-phase/quadrature-phase (IQ) demodulator is a key component in the RF front-end of a receiver, responsible for converting RF signals centered at the carrier frequency to complex baseband signals [15]. Ideal IQ demodulators provide two orthogonal channels for the real and imaginary parts of the baseband signal. Mismatches between the I and Q branches of a low-cost RF receiver, however, break down this orthogonality, leading to IQ imbalance. The mismatch is due to non-idealities at the analog components, such as the

local oscillators and the low-pass filters in the IQ branches [19]. This imbalance introduces a signal-dependent perturbation to the received signal [19] that is more prominent at higher carrier frequencies, such as mmWave and THz, than at lower frequencies [20, 21]. Under such impairments, CS-based methods that do not account for IQ imbalance may fail due to the mismatch between the measurement model assumed in the algorithm and the actual received signal model. In Chapter 4, we study how exactly IQ imbalance affects the estimated sparse channel and develop a greedy algorithm for sparse channel recovery that exploits the structure of the CS measurement model under IQ mismatch.

To increase the energy efficiency of mmWave systems, power amplifiers often operate in the nonlinear region [22, 23]. This causes nonlinear distortion in the RF signals transmitted by the antenna elements, which can alter their illumination pattern [24]. Digital predistortion (DPD) is a common method to compensate for the nonlinearity of power amplifiers in phased arrays [22, 25, 26]. In this method, a probe is used to measure the transmitted RF signal, which is then fed back to the baseband to design a DPD that compensates for the nonlinearity of the power amplifiers [25]. In this dissertation, we assume that the power amplifiers driving the antenna elements are properly isolated and a DPD based on methods developed in [22, 25, 26] is used to compensate for their nonlinear effect.

Mutual coupling between the elements of an antenna array is an important hardware impairment that distorts the illumination pattern of the array. Its impact is much severe when the array elements are tightly spaced, e.g., for inter-element spacing of less than half a wavelength [27]. This can occur in scenarios with large antenna arrays constrained to a limited form factor [27], in which the spacing between antenna elements can be much smaller than half a wavelength, leading to significant coupling between the antenna elements. Designing CS matrices and sparse channel estimation algorithms that are robust against mutual coupling, however, is extremely important as a next step, considering the increasing interest in the use of antenna arrays with tightly spaced elements [28, 29]. Throughout the dissertation, we used half a wavelength spacing between the antenna elements, which is commonly considered [30, 31], resulting in small mutual coupling. Therefore, in this dissertation, we have excluded the impact of the mutual coupling in our analysis.

CFO is an important hardware impairment that occurs due to a difference in the carrier frequencies generated by the local oscillators at the transmitter and receiver [32]. CFO causes unknown phase distortions in the measurements that increase linearly with time. In mmWave systems, CFO can be severe as they operate at a high carrier frequency. The phase distortion due to CFO in the measurements can cause sparse channel estimation methods agnostic to this error to fail [33]. Prior work [33–35] developed sparse recovery methods for joint sparse channel and CFO estimation. The tensor-based joint sparse channel and CFO estimation from [33] can be combined with our in-sector sparse channel estimation from chapter 2 for joint CFO and in-sector channel estimation. Also, methods from chapters 3 and 4 of this dissertation can be combined with the trajectory design methods in swift-link from [35] for beamforming robust against CFO. Studying these combinations is beyond the scope of this thesis.

In the dissertation, we explore three important hardware impairments: low-resolution phase shifters, phase noise, and in-phase/quadrature (IQ) mismatch. In the following

subsections, we discuss the challenges posed by each impairment and discuss related prior work.

1.2.1. Low-resolution phase shifters at TX and low SNR at RX

Finding the optimal beam at the mmWave radios through classical channel estimation or beam scanning is challenging due to the use of large arrays and hardware constraints in these systems. For instance, mmWave radios usually employ phased arrays equipped with a single RF chain to achieve a cost-effective and power-efficient architecture [36, 37]. With low-resolution phase shifters at the phased array, also known as low-resolution phased arrays, the acquired measurements are projections of the channel on antenna weight matrices whose entries are constrained to a small alphabet.

An alternative to beam scanning is hierarchical beam search, wherein beams of decreasing beamwidths are sequentially applied. At the end of the search process, a narrow beam that results in the highest received signal-to-noise ratio (SNR) is used for communication [37]. To implement hierarchical search, the design objective is to construct hardware-compatible beams that focus the transmitter's energy on a section of the channel in the angle domain. The sections are referred to as *sectors*, where each sector represents a group of directions in the angle domain. The process of identifying the best sector is called sector-level sweep (SLS). An example of the SLS procedure is shown in Fig. 1.3. Prior work has designed a collection of beams (also known as *codebook*) for SLS with high-resolution phased arrays [37–39] and low-resolution phased arrays [40, 41]. The techniques in [37–41] construct contiguous sectors, i.e., sectors that cover a contiguous band of directions. In Chapter 2 of the dissertation, we explain that realizing such contiguous sectors with extremely low-resolution phased arrays is challenging. For example, the constructed antenna weights in [40] result in significant power leakage outside the sectors with one-bit phased arrays. To address the leakage issue, in Chapter 2, we propose new comb-like sectors that can be realized even with one-bit phased arrays.

In IEEE 802.11ad/ay [42] devices, the transmitter discerns the *sector of interest* by finding the sector that results in the highest received power. Then, the beam refinement protocol (BRP) can be used to obtain channel measurements and subsequently perform beamforming within the sector of interest. Unfortunately, exhaustive beam scanning within this sector, as shown in Fig. 1.4, can still result in a huge overhead in mmWave or THz radios with massive antenna arrays. To address this problem, prior work has exploited the sparse or low-rank characteristics of the angle domain channel [43, 44] for sub-Nyquist channel estimation. Most of these techniques [45–48], however, employ quasi-omnidirectional beams to acquire spatial channel measurements. The use of such wide beams results in a poor SNR in the channel measurements, causing these methods to fail in practice [41]. Therefore, a collection of beams that focus energy only within the sector of interest must be developed for "in-sector" channel estimation using BRP. We further explore this in Chapter 2.

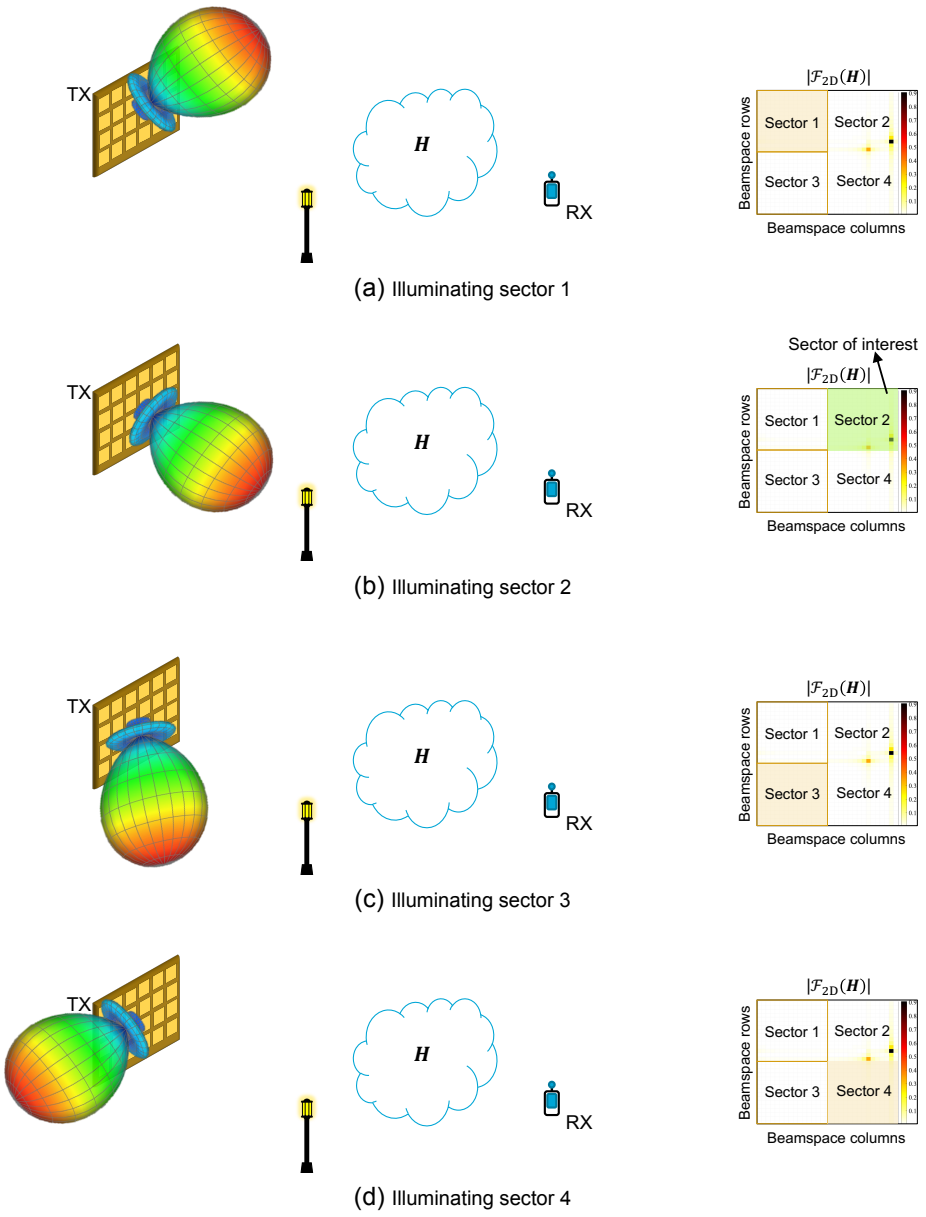


Figure 1.3: An example of SLS procedure to find the sector with the highest received power (also known as the *sector of interest*). In this example, a total of 4 sectors are considered and sector 2 results in the highest SNR.

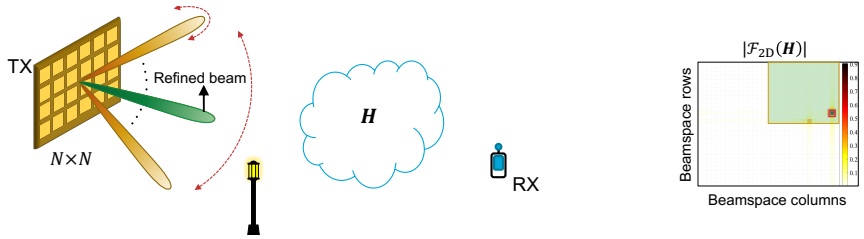


Figure 1.4: An example of BRP within the sector of interest, i.e., sector 2, based on exhaustive beam scanning. For large phased arrays, beam scanning within a sector results in a high training overhead.

Related works on in-sector channel estimation

We now discuss prior work on sparsity-aware channel estimation within a sector [41, 49–52]. In [41], randomly selected DFT columns were used to modulate a spread sequence and generate different beams to obtain channel measurements in different sectors. In [49], a contiguous band of directions is illuminated by sub-arrays. Then, the beam at each sub-array is phase modulated to construct an ensemble of in-sector beams. In [50], the beams were sampled from a large set of random codes based on their capability to focus energy on a sector of interest. We will show in chapter 2 that the techniques in [41, 49] and [50] result in poorer in-sector channel estimates than our method. In [51], linear convolution at the output of a fully digital array is performed, and the steady-state output of the digital array is decimated to reduce the number of measurements and achieve parallel processing. Our method in Chapter 2, unlike [51], performs circular convolution that preserves the channel dimension and obtains compressive measurements of the channel without needing to discard measurements. In [53]-[54], an adaptive beamforming algorithm based on sequential beam search is proposed to reduce the beam scanning time while combating the low SNR issue in initial access. The methods in [53]-[54] only support wireless channels with a single dominant path, while our method exploits sparsity and can support channels with multiple paths.

1.2.2. MmWave channel estimation under phase noise

The spatial channel measurements in mmWave or THz systems are acquired sequentially in time, as shown in Fig. 1.5. In the frame structure shown in Fig. 1.5, TNR stands for training, which denotes the time slot used to acquire a measurement within a batch. The vector of measurements acquired within batch b is denoted by $\mathbf{y}_b$. In practice, the phase of these measurements is perturbed due to phase noise [16]. Phase noise occurs due to random fluctuations in the frequency and phase of an oscillator [15]. Let $\phi_{b,m}$ denote the phase error corresponding to the m^{th} measurement acquired in batch b. As a result, under phase errors, the measurement model in (1.1) is rewritten as

$$\mathbf{y}_b = \text{diag}\left[e^{\mathrm{j}\phi_{b,1}}, e^{\mathrm{j}\phi_{b,2}}, \cdots, e^{\mathrm{j}\phi_{b,M}}\right]\mathbf{A}_b\mathbf{x} + \mathbf{v}_b, \tag{1.2}$$

where $\mathbf{A}_b$ denotes the $M \times N^2$ CS matrix for batch b, the vector of additive white Gaussian noise in batch b is denoted by $\mathbf{v}_b$, and $\text{diag}\left[e^{j\phi_{b,1}}, e^{j\phi_{b,2}}, \cdots, e^{j\phi_{b,M}}\right]$ is a diagonal matrix with $\left[e^{j\phi_{b,1}}, e^{j\phi_{b,2}}, \cdots, e^{j\phi_{b,M}}\right]$ on its main diagonal. As the phase errors are non-zero, the measurement model in (1.2) deviates from the standard CS measurements in (1.1).

In practice, these phase errors are not totally independent, but rather follow a certain process called the Wiener process [55]. In this process, phase noise $\phi_{b,m}$, given the previous phase error (i.e., $\phi_{b,m-1}$ or $\phi_{b-1,M}$ if $m = M$), follows a Gaussian distribution with mean $\phi_{b,m-1}$ or $\phi_{b-1,M}$, and variance that depends on the time stamp of the measurements.

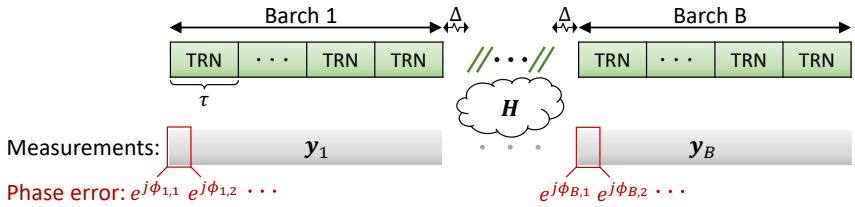


Figure 1.5: The frame structure used to acquire channel measurements. There are B batches in this frame structure with an inter-batch duration of Δ. In practice, measurements within each batch are perturbed by Wiener phase noise that is determined by the time stamp of the measurements. In the IEEE 802.11ad/ay standard τ = 128ns [56].

Phase noise, which is more prominent at higher frequencies [17, 57], significantly alters the received signal model compared to the standard CS model (1.1), which only accounts for additive noise in the linear measurements. Phase noise of an oscillator operating at a frequency of f_c is usually given by the value of its power spectrum at an offset f_m from f_c, denoted by $\mathscr{L}(f_m)$ in dBc/Hz [55]. A large value of $\mathscr{L}(f_m)$ means high phase noise and, therefore, results in high variance in the Wiener phase noise process.

We explain how the model mismatch due to phase noise between (1.1) and (1.2) impacts the standard sparse recovery with the AMP [9]. To this end, in Fig. 1.6, we plot the normalized mean squared error (NMSE) of the estimated channel with phase noise $\mathscr{L}(f_m)$. The AMP algorithm from [9] estimates the channel by ignoring phase errors and assuming that the CS measurements come from the model in (1.5). As observed, the AMP algorithm from [9] that is agnostic to phase errors fails when the variance of phase errors is large. Therefore, the challenge here is to develop a CS algorithm that can recover the channel from measurements that are perturbed by phase noise in (1.2). We note that classical phase noise tracking approaches [58, 59] do not exploit channel sparsity, making them unsuitable for mmWave or THz channel estimation.

Related work on mmWave channel estimation under phase noise

Prior work in [60–62] addresses the phase noise issue by considering only the magnitude of the acquired measurements for sparse channel estimation. Recently, it was shown in [56] that the sparse self-calibration technique from [16] can be used to jointly estimate the phase errors and the sparse channel. While these methods are robust to phase noise [16, 60–62], they overlook the temporal correlation of phase errors, which is common

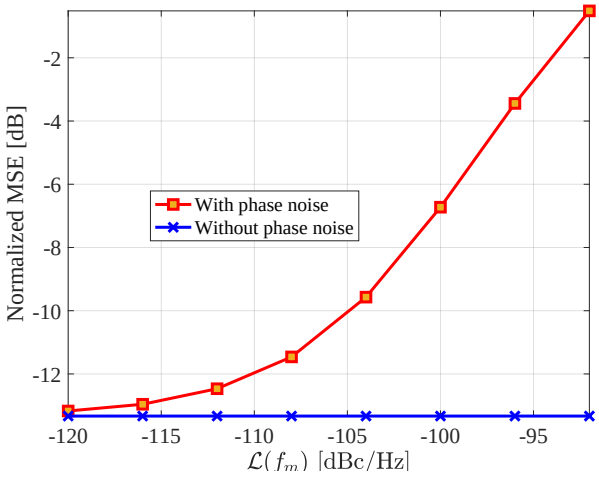


Figure 1.6: Standard sparse channel estimation using [9] fails when the measurements are perturbed by phase noise at f_c = 60 GHz and $\mathscr{L}(f_m)$ dBc/Hz are at f_m = 1 MHz. Here, M = 64 measurements of a 16 × 16 channel were acquired in each of the B = 4 batches.

in practice [55]. One of the early works that exploit the temporal correlation in CS measurements uses an extended Kalman filter to track the phase variations [63]. The method developed in [63], however, only estimates a one-sparse approximation of the channel for single-stream beamforming. A mmWave channel, however, in general, can have multiple non-zero components that can be exploited to achieve a high beamforming gain. This motivates the need to develop techniques that estimate all the sparse entries in the channel.

Recent work in [56, 57, 64, 65] has exploited the partially coherent phase error structure in the CS measurements for channel estimation. Under this structure, the phase errors within short batches of CS measurements are nearly the same while they change considerably across different batches. The sparse Bayesian learning-based method in [57] considers phase errors, but it does not leverage the Wiener structure in the errors within the recovery algorithm. In [64], a two-stage partially coherent compressive phase retrieval (PC-CPR) was proposed to reconstruct the sparse channel from partially coherent CS measurements. The algorithm is initialized using the CPR method from [66], and the phase offsets and the sparse channel are iteratively estimated in a refinement stage. In [65], expectation-maximization generalized approximate message passing (EM-GAMP) [67] is used to first obtain a coarse estimation of the sparse channel, which is subsequently used to estimate the phase offsets. In [56], phase error-aware matching pursuit followed by alternating optimization was developed to recover the sparse channel and the phase errors. The methods in [56, 57, 64] assume that the channel is exactly sparse, which is unrealistic. Further, [56, 64, 65] assume that the number of non-zero entries in the sparse channel is known, which is not practical. To the best of our knowledge, none of the existing methods exploit Wiener phase error statistics in partially

coherent sparse recovery using measurements acquired over multiple batches.

1.2.3. MmWave channel estimation under IQ imbalance

In practice, the radio frequency channel measurements are acquired after down-conversion to the baseband at the receiver. A low-cost RF receiver suffers from a mismatch between the in-phase and the quadrature-phase branches. The mismatch is due to non-idealities at the analog components, such as the local oscillators and the low-pass filters in the IQ branches [19], and is commonly known as IQ imbalance. Typically, IQ imbalance is dominated by imperfections at the oscillator rather than the filter design, as local oscillator design at RF is more challenging [68]. The imbalance distorts the received signal by introducing a signal-dependent perturbation, which is commonly quantified by a metric called image rejection ratio (IRR) [19].

Here, we explain how the CS measurements in (1.1) change under IQ imbalance. For the single antenna RX in Fig. 1.1, IQ imbalance is modeled using a scalar complex number denoted by ξ. The IQ imbalance metric IRR associated with ξ is given by $\text{IRR} = |\xi|^2/|1-\xi^*|^2$. Under IQ imbalance, the CS measurement model in (1.1) is rewritten as

$$\mathbf{y} = \xi\mathbf{A}\mathbf{x} + (1-\xi^c)\mathbf{A}^c\mathbf{x}^c + \xi\mathbf{v} + (1-\xi^c)\mathbf{v}^c. \tag{1.3}$$

where $(\cdot)^c$ returns the conjugate of a complex scalar, vector, or matrix. As the IQ imbalance parameter $\xi \neq 1$, the measurement model in (1.3) deviates from the standard CS measurements in (1.1).

In Fig. 1.7, we plot the NMSE of the estimated channel with IQ imbalance metric IRR. Here, the OMP algorithm from [8] is used for sparse channel estimation. As observed, standard CS algorithms that are agnostic to phase errors fail due to a mismatch in the received measurement model. At high carrier frequencies, such as mmWave and THz, the mismatch in the IQ branches is more severe than at lower frequencies [20, 21]. Note that the IRR is small for a more severe IQ imbalance. For example, [69, 70] reported IRRs between 16 dB and 38 dB at a carrier frequency of 60 GHz. The failure of CS-based channel estimation under IQ imbalance requires rethinking the sparse recovery problem under model mismatch.

Related work on mmWave channel estimation under IQ imbalance

Prior work in [17, 71–75] developed algorithms for single-input single-output (SISO) channel estimation under IQ imbalance. The algorithms in [17, 73–75] solved for the time-domain channel and those in [71, 72, 74] performed frequency-domain channel estimation. The works in [71, 72] investigated channel estimation under IQ imbalance in orthogonal frequency division multiplexing (OFDM) and those in [17, 73–75] studied channel estimation under IQ imbalance for single-carrier transmission schemes. IQ imbalance distorts the channel estimates at mirrored subcarriers, i.e., subcarriers that are equally distant from the carrier frequency, in OFDM systems [71, 72]. In [71], this property is used to develop a least mean squares method for channel estimation in the frequency domain, and [72] developed maximum likelihood and least squares estimators to estimate the channel and the IQ imbalance parameter. In [73], first, a least-squares estimate of an effective channel vector comprising the IQ imbalance parameters is

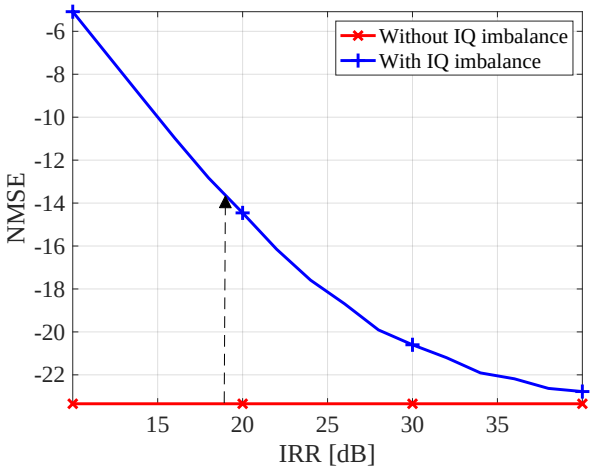


Figure 1.7: NMSE with $10\log_{10}(\text{IRR})$ where $\text{IRR} = |\xi|^2/|1-\xi^*|^2$ represents the level of receiver IQ imbalance. A large IRR indicates a small IQ imbalance. Here, we use $N = 256$, $M = 40$, and random phase shifts at the TX.

obtained. Next, a second least-squares optimization is employed to estimate the IQ imbalance parameter and the channel using alternating minimization. In [75], the received measurements are expressed in terms of the real and imaginary parts of the time-domain channel, and an iterative method based on expectation maximization [76] is developed to estimate the channel. For a terahertz system, [17] estimates an effective channel similar to [73] and then uses a gradient descent method to estimate the channel and the IQ imbalance parameters. The methods in [17, 71–74] do not exploit the sparse structure of channels along any dimension. Although [75] leverages time-domain sparsity of the channel, it does not exploit the spatial sparsity as it considers a SISO system.

We now discuss the prior work in [77–80] on spatial channel estimation under IQ imbalance. These works assume multiple RF chains, each with a different IQ imbalance parameter at the receiver. In [77], first, a phase-shifted version of the IQ imbalance parameters and the magnitude of the channel coefficients are estimated by iteratively solving a maximum likelihood problem. Then, a different set of measurements is used to estimate the phase of the channel coefficients and the IQ imbalance parameters. In [78], each channel measurement is added to its conjugated version to remove the effect of IQ imbalance, and then a linear minimum mean squared error estimate of the channel is obtained from the new measurements. The estimated channel is then used to estimate the IQ imbalance parameters by solving a least-squares problem. As noted in [78], this approach doubles the noise variance, leading to an increased error in the estimated channel. In [79], the product of the channel and the IQ imbalance parameters was estimated. Although the methods in [77–79] solve for the spatial channel, they do not exploit sparsity of the channel in the angle-domain. The work in [80] estimates the IQ imbalance parameter and the channel from the measurements by formulating the problem as a joint calibration and sparse recovery task. The measurement model in

[80, equation (2)], however, ignores the IQ imbalance term associated with the complex conjugate of the channel. In chapter 4, we show that the conjugate term results in significant aliasing artifacts in the channel estimate, which cannot be ignored. To the best of our knowledge, none of the existing works characterize the impact of the IQ imbalance on sparse angle-domain channel estimation.

1.3. Outline and contributions of the dissertation

1.3.1. Chapter 2: In-sector Compressive Beam Acquisition for MmWave and THz Radios

This chapter addresses the issue of low SNR with standard CS-based methods, such as [45–48] that use quasi-omnidirectional illumination pattern for channel training. In this chapter, we adopt a two-stage approach for beamforming. In the first stage, the compressive measurements of the channel within the best sector are used to estimate the channel within this sector. Next, the estimated in-sector channel is used to perform conjugate beamforming. We propose a new approach for in-sector channel estimation by solving two main challenges. The first challenge is to design beams that focus the transmitted energy on distinct non-overlapping sectors within the beamspace. We will see that this problem is extremely challenging with low-resolution phased arrays. The second challenge is to optimize the compressed sensing (CS) matrices to achieve low CS aliasing artifacts in the in-sector channel estimate. We solve both these problems by constructing new in-sector beams, such that the corresponding antenna weight matrices (AWMs) can be applied in low-resolution phased arrays. These AWMs, referred to as base AWMs, each focus power on a distinct sector. In our method, the channel measurements for CS within the sector of interest are then obtained during BRP by circularly shifting the base AWM.

1.3.2. Chapter 3: Message passing-based sparse spatial channel estimation robust to partially coherent phase noise

This chapter addresses the mmWave channel estimation under phase noise from CS measurements in (1.2). In this chapter, we develop a message passing-based sparse recovery algorithm that exploits Wiener statistics of the phase errors in the channel measurements. In message passing-based methods, the unknown variables to be solved, i.e., the sparse channel entries or the phase errors, are represented as nodes in a factor graph. Information about these variables, such as sparsity or measurement-based likelihoods, is incorporated as factors. By iteratively exchanging messages (beliefs) between the factors and the variables, the marginal distributions of the variables are computed, which are subsequently used to estimate the unknowns. We factorize the posterior joint probability density function (PDF) of the phase offsets and the sparse angle domain channel. Our factorization incorporates the temporal correlation in the phase-perturbed batches of CS measurements. Finally, we approximate the messages in our factor graph to develop a tractable channel estimation algorithm.

1.3.3. Chapter 4: Compressed sensing-based sparse spatial channel estimation under IQ imbalance

This chapter addresses the mmWave channel estimation under IQ mismatch from CS measurements in (1.3). In this chapter, we discuss how angle-domain channel reconstruction with standard CS is extremely sensitive to IQ imbalance. Specifically, we observe that sparse beamspace recovery with the standard orthogonal matching pursuit (OMP) [8] fails when IRR is below 30 dB in our setup. To solve this challenge, we introduce an *augmented CS* (AugCS) formulation for sparse recovery under IQ imbalance. In our formulation, we construct an auxiliary vector composed of the IQ imbalance parameter and the angle-domain channel. The auxiliary vector, whose dimension is twice the dimension of the channel, is referred to as *augmented beamspace*. We develop paired-support OMP (PSOMP), a greedy sparse recovery algorithm, that exploits the special sparse structure of *augmented beamspace*. Finally, we formulate a least-squares problem to estimate the beamspace and the IQ imbalance parameter from the recovered augmented beamspace using PSOMP.

Notation: We use a, $\mathbf{a}$, and $\mathbf{A}$ to denote a scalar, vector, and a matrix. The transpose, conjugate, and conjugate-transpose operators are indicated by $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{c}}$, and $(\cdot)^{*}$. , $a[i]$ is the i^{th} entry of $\mathbf{a}$, and $\mathbf{a}_i$ or $\mathbf{a}[i]$ is the i^{th} column of $\mathbf{A}$. We denote the $(i,j)^{\text{th}}$ entry of $\mathbf{A}$ by A_{ij} or $A(i,j)$. The ℓ_2 norm of $\mathbf{a}$ is denoted by $\|\mathbf{a}\|_2$. We use $\|\mathbf{A}\|_{\mathrm{F}}$ to denote the Frobenius norm of $\mathbf{A}$. We use $|\cdot|$ and $\angle\cdot$ to define the element-wise magnitude and phase of a vector or matrix. For an integer N, we define the set $[N] = \{0,\dots,N-1\}$. Also, $\lfloor a \rfloor$ is the floor of a, and $\langle i \rangle_N$ is the modulo-N remainder of i. The flipped version of a vector $\mathbf{a}$ is denoted by $\check{\mathbf{a}}$ where $\check{a}[i] = a[\langle -i \rangle_N]$. We define the vector version of $\mathbf{A}$ as $\mathrm{vec}(\mathbf{A}) = \left[\mathbf{a}_0^{\mathsf{T}},\cdots,\mathbf{a}_{N-1}^{\mathsf{T}}\right]^{\mathsf{T}}$. We define $\mathrm{Diag}(\mathbf{a})$ to be a diagonal matrix with $\mathbf{a}$ on its diagonal. We use $\circledast$, $\otimes$, and $\odot$ to denote the circular convolution, the Kronecker product, and the Hadamard product. The inner product of $\mathbf{A}$ and $\mathbf{B}$ is $\langle \mathbf{A},\mathbf{B} \rangle = \sum_{i,j} A_{ij} B_{ij}^{\mathsf{c}}$. We use $\mathfrak{R}(\kappa)$ and $\mathfrak{I}(\kappa)$ to denote the real and imaginary parts of the complex scalar κ. We use x for a random variable and x for its realization, with $p(x)$ as the PDF of x evaluated at x. The complex Gaussian PDF $\mathscr{CN}(x;a,b)$ has mean a and variance b. Also, $\mathscr{CN}(\mu,\sigma^2)$ denotes a complex Gaussian distribution with mean μ and variance σ^2. Lastly, $\mathbb{R}_{+}^{N\times 1}$ is the set of $N \times 1$ vectors with non-negative real entries, and $\mathsf{j} = \sqrt{-1}$.

2

In-sector Compressive Beam Acquisition for MmWave and THz Radios

Beam acquisition is key in enabling millimeter wave and terahertz radios to achieve their capacity. Due to the use of large antenna arrays in these systems, the common exhaustive beam scanning results in a substantial training overhead. Prior work has addressed this issue by developing compressive sensing (CS)-based methods which exploit channel sparsity for faster beam acquisition. Unfortunately, most CS techniques employ wide beams and suffer from a low signal-to-noise ratio (SNR) in the channel measurements. To solve this challenge, we develop an IEEE 802.11ad/ay compatible technique that takes an in-sector approach for CS. In our method, the angle domain channel is partitioned into several sectors, and the channel within the best sector is estimated and then used for beamforming. The essence of our framework lies in the construction of a low-resolution beam codebook to identify the best sector and in the design of a CS matrix optimized for in-sector channel estimation. Our beam codebook illuminates distinct non-overlapping sectors and can be realized with low-resolution phased arrays. We show that the proposed codebook results in a higher received SNR than the state-of-the-art sector sweep codebooks. Furthermore, our optimized CS matrix achieves a better in-sector channel reconstruction and a higher achievable rate than comparable benchmarks.

This chapter is based on my work in [81]: H. Masoumi, M. Verhaegen and N. J. Myers, "In-sector compressive beam acquisition for mmWave and THz radios," *IEEE Transactions on Communications*., vol. 73, no. 4, pp. 2752–2768, Apr. 2025.

2.1. Channel and System Model

In this section, we consider a narrowband point-to-point wireless system to explain the transmit beam acquisition problem. In section 2.4, we extend our sparse in-sector channel estimation-aided beam acquisition technique to a wideband scenario using the IEEE 802.11ad frame structure.

2.1.1. Channel Model

We consider an $N \times N$ half-wavelength spaced uniform planar array at the transmitter (TX) and a single antenna receiver (RX) as shown in Fig. 2.1(a). We assume isotropic antenna elements at the TX and the RX, for simplicity. In this chapter, we consider the transmit beam acquisition problem where only the TX needs to align its beam towards the best direction [38, 45]. This problem differs from the beam alignment problem in which both the TX and RX need to align their beams to create a directional link between them [49, 82]. We would like to mention, however, that our solution can also be applied to the beam alignment problem by using the designed codebooks at both the TX and the RX. The application of our codebooks is feasible provided that their antenna array dimensions are a power of two. Such array dimensions are commonly used in practical implementations [83–86]. Extending our solution to other array dimensions is beyond the scope of this chapter.

The narrowband multiple-input single-output channel between the TX and RX is modeled as an $N \times N$ matrix $\mathbf{H}$, the matrix representation of the $N^2 \times 1$ vectorized channel. The channel comprises L propagation rays, where the ℓ^{th} ray has a complex gain of β_ℓ, an azimuth angle-of-departure $\theta_{a,\ell}$ and an elevation angle-of-departure $\theta_{e,\ell}$. By defining the beamspace angles as $\omega_{a,\ell} = \pi \sin\theta_{e,\ell} \sin\theta_{a,\ell}$, $\omega_{e,\ell} = \pi \sin\theta_{e,\ell} \cos\theta_{a,\ell}$ [87] and the $N \times 1$ Vandermonde vector $\mathbf{a}_N(\omega)$ as

$$\mathbf{a}_N(\omega) = \left[1, e^{j\omega}, e^{j2\omega}, \cdots, e^{j(N-1)\omega}\right]^{\mathsf{T}}, \tag{2.1}$$

the baseband channel matrix $\mathbf{H}$ is given by

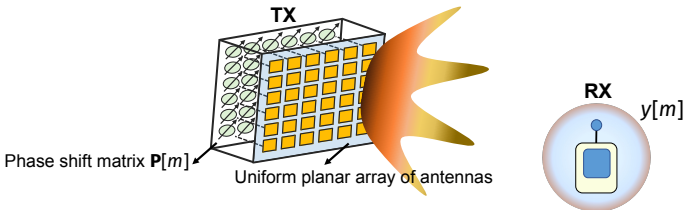
$$\mathbf{H} = \sum_{\ell=1}^{L} \beta_\ell \mathbf{a}_N\left(\omega_{e,\ell}\right) \mathbf{a}_N^{\mathsf{T}}\left(\omega_{a,\ell}\right). \tag{2.2}$$

The channel dimension N^2 can be in the order of hundreds to thousands in typical mmWave or THz access points.

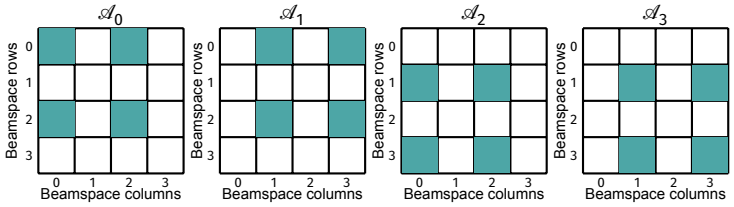
The channel $\mathbf{H}$ is approximately sparse in the angle domain due to high scattering at mmWave and THz wavelengths [1]. To exploit this property during channel reconstruction, $\mathbf{H}$ is expressed in the beamspace (angle domain) where it has a sparse representation. Since we assume a uniform planar array at the TX, the 2D-DFT dictionary is used to represent $\mathbf{H}$ in the beamspace. We use $\mathbf{U}_N$ to denote the standard $N \times N$ unitary DFT matrix and $\mathbf{X}$ to denote the beamspace representation of $\mathbf{H}$. Then, $\mathbf{X}$ and $\mathbf{H}$ are related as

$$\mathbf{H} = \mathbf{U}_N \mathbf{X} \mathbf{U}_N. \tag{2.3}$$

In our analysis, we assume that $\mathbf{X}$ is exactly sparse, i.e., the beamspace angles are exactly



(a) A comb-like beam pattern.



(b) Example of 4 comb-like sectors in the 2D beamspace.

Figure 2.1: An mmWave system with a uniform planar array at the TX employing comb-like beams designed in this work. Each comb-like beam illuminates a beamspace sector shown in (b). In this chapter, the sector that results in the highest received power is identified, and then the channel within that sector is estimated for beamforming.

aligned with any of the N^2 directional 2D-DFT beam directions. In our simulations, we use channels obtained from the NYU channel simulator [6] where $\mathbf{X}$ is only approximately sparse.

2.1.2. System model

The TX uses a phased array wherein the antennas are connected to a single RF chain through a q-bit phase shifter. The set of the possible weights at each antenna is $\mathbb{Q}_q = \left\{e^{\mathrm{j}2\pi\ell/2^q}/N : \ell \in [2^q]\right\}$. As an example, $\mathbb{Q}_1 = \{1/N, -1/N\}$ can be realized with a 1-bit phased array. The TX can apply AWMs to its phased array, which are constrained to be within $\mathbb{Q}_q^{N\times N}$, i.e., the set of all $N \times N$ matrices with entries in $\mathbb{Q}_q$. In this chapter, the AWMs are designed such that their corresponding beams illuminate only a certain set of indices of the beamspace $\mathbf{X}$. A *sector* is defined as a group of indices of the beamspace that are illuminated by a beam. We use S to denote the number of sectors used to partition the $N \times N$ beamspace grid into disjoint subsets $\mathscr{A}_s$, $\forall s \in [S]$. Hence, $\mathscr{A}_s$ denotes the set of all beamspace indices illuminated in sector s. The union of $\{\mathscr{A}_s\}_{s=0}^{S-1}$, i.e., $[N] \times [N]$, spans the entire beamspace. An example of the sectors $\{\mathscr{A}_s\}_{s=0}^{S-1}$ with our comb-like construction proposed in Sec. 2.2 is shown in Fig. 2.1(b) for $S = 4$.

We now explain how our in-sector CS-based beam acquisition method can be integrated into *SLS* and *BRP* of the beam training procedure in IEEE 802.11ad/ay [88] standards. We assume a block fading channel and that the channel remains constant

throughout the SLS and BRP procedures. This assumption is reasonable in practice for a typical 802.11ad/ay application. For instance, the coherence time is about 1 ms [89] in an indoor setup, while the SLS phase takes about 15.76 μs per sector sweep frame [90] and a BRP packet is of up to 46.96 μs [91, Section 20.12.3].

Sector level sweep

During SLS, the TX uses sector sweep frames and illuminates the channel with sectored beam patterns. The standard, however, does not prescribe specific beam patterns to conduct SLS. These beams may be the common contiguous patterns discussed in [39–41] or the comb-like patterns proposed in this work as shown in Fig. 2.1(b). We define $\{\mathbf{P}_s\}_{s=0}^{S-1}$ as the base AWMs applied at the TX for SLS. The base AWM $\mathbf{P}_s$ illuminates sector s. We use y_s^{sls} to denote the measurement acquired when the TX applies $\mathbf{P}_s$ and $v_s \sim \mathcal{CN}(0, \sigma^2)$ as the noise in this measurement, i.e.,

$$y_s^{\text{sls}} = \langle \mathbf{H}, \mathbf{P}_s \rangle + v_s. \tag{2.4}$$

In SLS, the RX calculates the received power associated with (2.4) for each sector. Then, it determines the sector of interest as the one with the highest received power using

$$s_{\text{opt}} = \underset{s \in [S]}{\text{argmax}} \left| y_s^{\text{sls}} \right|^2. \tag{2.5}$$

An illustration of the SLS procedure is shown in Fig. 2.2.

For sector s, we define the set of beamspace indices $\mathcal{A}_s$ as the locations in the beamspace $\mathbf{X}$ that are illuminated by the base AWM $\mathbf{P}_s$. To see how $\mathcal{A}_s$ is exactly determined from $\mathbf{P}_s$, observe that the inner product of two matrices is the same as the inner product of their inverse 2D-DFTs, i.e., $\langle \mathbf{H}, \mathbf{P}_s \rangle = \left\langle \mathbf{X}, \mathbf{U}_N^* \mathbf{P}_s \mathbf{U}_N^* \right\rangle$. When the TX applies $\mathbf{P}_s$ to its array, it generates a beam that illuminates the beamspace at the indices where $\mathbf{U}_N^* \mathbf{P}_s \mathbf{U}_N^*$ is non-zero. Therefore, $\mathcal{A}_s = \left\{(k, \ell) : |(\mathbf{U}_N^* \mathbf{P}_s \mathbf{U}_N^*)_{k,\ell} \neq 0|\right\}$. For the identified sector s_{opt}, we define the corresponding base AWM as $\mathbf{P}_o = \mathbf{P}_{s_{\text{opt}}}$ for ease of notation. The corresponding set $\mathcal{A}_o = \mathcal{A}_{s_{\text{opt}}}$. Our objective in codebook design for SLS is to design $\{\mathbf{P}_s\}_{s=0}^{S-1}$ such that each of them illuminates a distinct set of indices in the beamspace, i.e., $\{\mathcal{A}_s\}_{s=0}^{S-1}$ are disjoint and their union is the entire beamspace. The main challenge in the design of such AWMs is due to the constraint that $\mathbf{P}_s \in \mathbb{Q}_q^{N \times N} \forall s$. In Section 2.2, we present our codebook and show that it outperforms those constructed in prior work [39–41].

SLS uses only the received power to identify the best sector s_{opt} and it does not require estimating the channel $\mathbf{H}$. The antenna weights obtained at the end of the SLS phase, however, usually do not provide sufficient beamforming gain as $S \ll N^2$. To further improve the beamforming gain, a narrow beam can be formed after estimating the spatial channel within the identified sector. The next subsection explains the main challenges to be solved in estimating the in-sector channel.

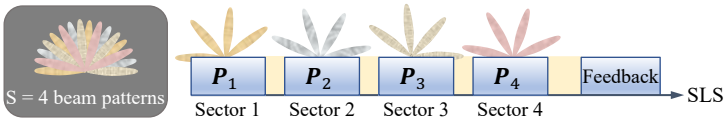


Figure 2.2: Example of SLS to determine the sector of interest using $S = 4$ disjoint comb-like sectors designed with our approach. The RX finds the sector of interest s_{opt} as the one that results in the highest received power and feedbacks this information to the TX. The TX can then estimate the channel within the sector using BRP.

Beam refinement protocol

During BRP, the TX applies a collection of AWMs that illuminate directions within the sector of interest. Under the assumption that the S sectors uniformly partition the beamspace, the in-sector channel is an N^2/S dimension vector. A simple approach to estimate this vector is to apply directional beams that exhaustively scan all directions within the sector. Such a scan, however, results in a substantial training overhead, which is in the order of N^2/S. To avoid this high training overhead, we develop an in-sector CS-based method that obtains compressive measurements of the channel within the sector of interest.

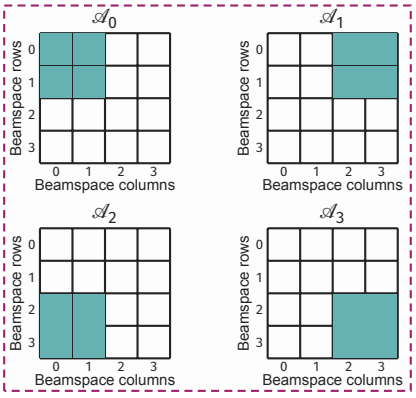
We use M to denote the number of CS measurements within the sector s_{opt}. These measurements are obtained using an ensemble of beams $\{\mathbf{P}_{\text{o}}[m]\}_{m=0}^{M-1}$ that only illuminate the sector of interest, i.e., the matrices $\{\mathbf{U}_N^{\star}\mathbf{P}_{\text{o}}[m]\mathbf{U}_N^{\star}\}_{m=0}^{M-1}$ are non-zero only at the indices in $\mathscr{A}_{\text{o}}$. For a measurement noise $v[m] \sim \mathscr{CN}(0, \sigma^2)$, the m^{th} in-sector measurement $y[m]$ of the channel is

$$y[m] = \langle \mathbf{H}, \mathbf{P}_{\text{o}}[m] \rangle + v[m]. \tag{2.6}$$

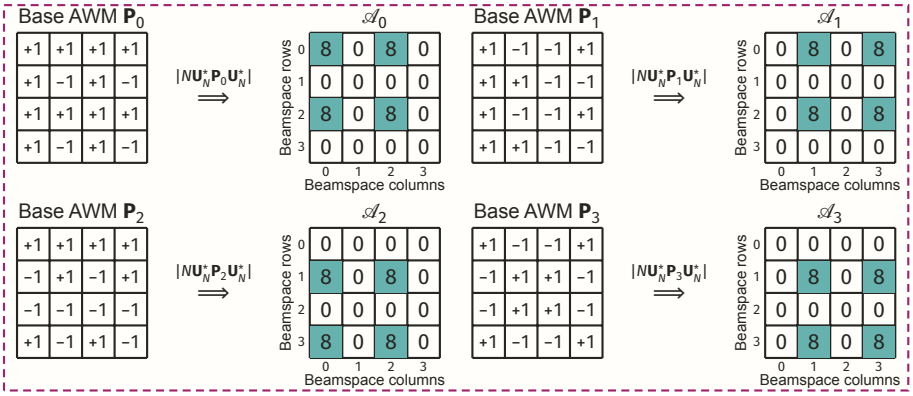
Our goal for in-sector CS is to construct $\{\mathbf{P}_{\text{o}}[m]\}_{m=0}^{M-1}$ such that: i) $\mathbf{P}_{\text{o}}[m] \in \mathbb{Q}_q^{N\times N}$ illuminates only the directions within $\mathscr{A}_{\text{o}}$ for each m and ii) the collection $\{\mathbf{P}_{\text{o}}[m]\}_{m=0}^{M-1}$ results in a low channel estimation error within the sector of interest. In-sector CS is promising over CS techniques [45–48] that employ wide beams, because the SNR of the in-sector measurements is higher than those acquired with a wide beam by about $10\log_{10}S$ dB, which is the sector gain.

2.2. Proposed sectors and base AWMs for SLS

SLS is usually performed using beam patterns that illuminate disjoint and complementary sectors within the beamspace. This is because such illumination patterns concentrate the transmitter's power only over a fraction of the beamspace, enabling a higher received SNR for initial access. In the SLS literature [39, 40] and [92], it is common to consider contiguous sectors such as the ones shown in Fig. 2.3(a). The AWMs that achieve such contiguous patterns, however, cannot be realized with extremely low-resolution phased arrays such as one-bit phased arrays. To see this, let us consider the problem of constructing AWMs that result in illumination patterns shown in Fig. 2.3(a) using a 4×4 antenna array with one-bit phase shifters. Here, the AWMs need to be chosen from $\{\pm 1\}^{4\times 4}$ due to the one-bit constraint. As the matrices in $\{\pm 1\}^{4\times 4}$ are purely real,



(a) $S = 4$ contiguous beam patterns.



(b) $S = 4$ comb-like beams realized with 1-bit phase shifters using our method.

Figure 2.3: Contiguous and comb-like beam patterns to illuminate $S = 4$ disjoint sections of the 2D beamspace with a 4×4 array at the TX. With 1-bit phase shifters, the AWMs corresponding to the beam patterns in (a) cannot be realized, while those corresponding to the patterns in (b) can be achieved using our construction in Sec. 2.2.1.

their inverse 2D-DFTs are mirror symmetric [93], i.e., a pattern that illuminates (k, ℓ) beamspace index must also illuminate the $(\langle -k \rangle_N, \langle -\ell \rangle_N)$ index. We observe that the contiguous patterns in Fig. 2.3(a) do not exhibit a mirror symmetric pattern. As a result, the sectors shown in Fig. 2.3(a) cannot be realized in one-bit phased arrays. To address this challenge, in this chapter, we design comb-like illumination patterns such as the ones shown in Fig. 2.3(b). Such comb-like patterns exhibit mirror symmetry, and the corresponding AWMs can be realized with phased arrays whose resolution can be as low as one bit.

2.2.1. Proposed construction for AWMs in SLS

Now, we explain the structure of the proposed comb-like patterns and discuss our method to design the base AWMs $\{\mathbf{P}_s\}_{s=0}^{S-1}$ with entries in $\mathbb{Q}_q^{N\times N}$ to achieve such patterns.

We drop the sector index s to explain the first step in our technique to construct AWMs for SLS. To design comb-like structures in the beamspace, our method constructs an antenna domain building block $\mathbf{C}$ with N_e rows and N_a columns. We assume that both N_e and N_a divide N, and the product N_eN_a is a power of 2. The building block $\mathbf{C}$ is chosen from the 2D-DFT codebook such that it illuminates just one direction in the $N_e \times N_a$ beamspace. An example for $\mathbf{C}$ is shown in Fig. 2.4(a). With our design, the inverse 2D-DFT of $\mathbf{C}$, i.e., $\mathbf{U}_{N_e}^*\mathbf{C}\mathbf{U}_{N_a}^*$, has just one non-zero entry and $N_eN_a - 1$ zeros. Next, the matrix $\mathbf{C}$ is upsampled by a factor of N/N_e along the row dimension and by a factor of N/N_a along the column dimension. We define these upsampling factors as

$$\rho_e = N/N_e \text{ and} \tag{2.7}$$

$$\rho_a = N/N_a, \tag{2.8}$$

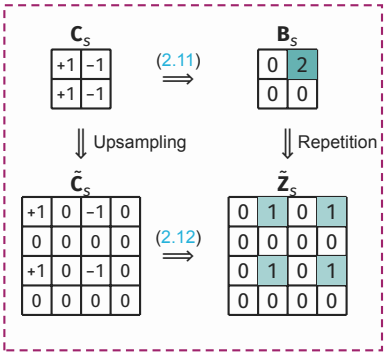
and the upsampled version of $\mathbf{C}$ is defined as $\tilde{\mathbf{C}} \in \mathbb{C}^{N\times N}$. As upsampling a signal results in replication and scaling in the Fourier domain, the beamspace representation of $\tilde{\mathbf{C}}$ is a scaled repetition of the inverse Fourier transform of the building block [94]. This repetitive pattern results in a comb-like structure in the beamspace as shown in Fig. 2.4.

Now, we summarize the final step in base AWM design to construct a phased array compatible matrix which achieves a comb-like beam pattern. Although the upsampled building block $\tilde{\mathbf{C}}$ results in a comb-like pattern, it cannot be realized with a phased array. This is because $\tilde{\mathbf{C}} \notin \mathbb{Q}_q^{N\times N}$ as it has several zeros. To construct a matrix in $\mathbb{Q}_q^{N\times N}$ that achieves a comb-like pattern, we first make use of the property that circularly shifting a matrix does not change the magnitude of its 2D-DFT. As a result, any 2D-circular shift of $\tilde{\mathbf{C}}$ results in the same illumination pattern as $\tilde{\mathbf{C}}$. We also observe that the locations of the zeros are complementary across all the 2D-circular shifts of $\tilde{\mathbf{C}}$. Our method takes a weighted combination of the $\rho_e\rho_a$ 2D-circular shifts of $\tilde{\mathbf{C}}$ to construct a base AWM in $\mathbb{Q}_q^{N\times N}$. The weights are optimized such that the illumination pattern associated with this weighted combination is almost flat within the designed sector. We will show in Sec. 2.3.3 that such flat profiles achieve a low reconstruction error using our sparse recovery guarantees from [95]. Our design procedure is illustrated in Fig. 2.4 for a particular sector.

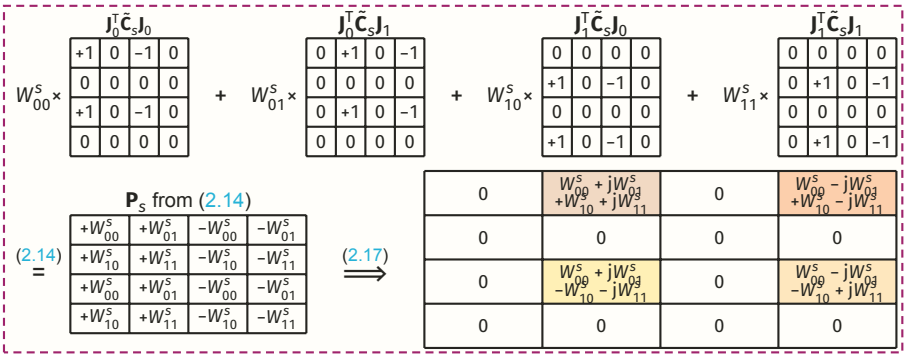
We now explain the mathematical details of our procedure to construct $S = N_aN_e$ sectors. As the entries of the building block are drawn from an $N_e \times N_a$ 2D-DFT codebook, realizing the designed AWMs in a phased array requires $q = \log_2(\max(\{N_a, N_e\}))$-bit phase shifters. We index the N_aN_e sectors in our design using the 2D-index pair (k_e, k_a), where $k_e \in [N_e]$ and $k_a \in [N_a]$. The sector index s is expressed as $s = N_ak_e + k_a$. The set of beamspace indices associated with this sector is

$$\mathscr{A}_s = \{(p, q) : p = nN_e + k_e, q = mN_a + k_a, n = [\rho_e], m = [\rho_a]\}, \tag{2.9}$$

which can be interpreted as a comb-like pattern anchored at (k_e, k_a). A detailed description of our method to construct a base AWM $\mathbf{P}_s$ that illuminates $\mathscr{A}_s$ is given below.



(a) Toy example for (2.11) and (2.12).



(b) Example for constructing the AWM $\mathbf{P}_s$ using (2.14) and its spectral mask (2.17).

Figure 2.4: An example of our proposed framework to construct AWMs that focus energy within sectors with comb-like patterns. Here, $N = 4$, the number of sectors are $S = 4$, $N_e = 2$, and $N_a = 2$. Therefore, $\rho_e = 2$ and $\rho_a = 2$.

- Construct the building block $\mathbf{C}_s \in \mathbb{Q}_q^{N_e \times N_a}$ as an outer product of the k_e^{th} column of $\mathbf{U}_{N_e}$ and the k_a^{th} column of $\mathbf{U}_{N_a}$, i.e.,

$$\mathbf{C}_s = \mathbf{U}_{N_e}(:, k_e)\mathbf{U}_{N_a}^T(:, k_a). \tag{2.10}$$

  Similar to [45], we define $\mathbf{C}_{s,\text{FC}}$ as the flipped and conjugated version of $\mathbf{C}_s$. The angle domain matrix associated with the building block is then

$$\mathbf{B}_s = \mathbf{U}_{N_e}^* \mathbf{C}_{s,\text{FC}} \mathbf{U}_{N_a}^* \tag{2.11}$$

  It can be shown that $B_s(k_e, k_a) = 1$ and $B_s(i, j) = 0\ \forall (i, j) \neq (k_e, k_a)$.

- Upsample $\mathbf{C}_s$ by a factor of ρ_e along the columns and ρ_a along the rows. The

upsampled matrix is obtained by first inserting $\rho_e - 1$ zero-valued row vectors between successive rows of $\mathbf{C}_s$, and then inserting $\rho_a - 1$ zero-valued column vectors between successive columns of the resultant matrix. The upsampled result is an $N \times N$ matrix $\tilde{\mathbf{C}}_s$, whose beamspace representation is defined as

$$\tilde{\mathbf{Z}}_s = \mathbf{U}_N^* \tilde{\mathbf{C}}_{s,\mathrm{FC}} \mathbf{U}_N^* \tag{2.12}$$

$$= \frac{1}{\sqrt{\rho_e \rho_a}} \left[\overbrace{\begin{matrix} \mathbf{B}_s & \cdots & \mathbf{B}_s \\ & & \\ \mathbf{B}_s & \cdots & \mathbf{B}_s \end{matrix}}^{\rho_a \text{ times}} \right] \Big\} \rho_e \text{ times.} \tag{2.13}$$

Here, (2.12) follows from the upsampling property of the 2D-DFT [96]. Observe that $\tilde{\mathbf{Z}}_s$ exhibits a comb-like structure as it contains repetitions of $\mathbf{B}_s$, that has a single non-zero entry. The pattern in $\tilde{\mathbf{Z}}_s$, however, cannot be realized in a phased array as its antenna domain representation $\tilde{\mathbf{C}}_s \notin \mathbb{Q}_q^{N \times N}$.

- Express $\mathbf{P}_s$ as a weighted sum of $\rho_e \rho_a$ distinct 2D-circular shifts of $\tilde{\mathbf{C}}_s$. Let $\mathbf{J} \in \mathbb{R}^{N \times N}$ denote a circulant delay matrix with the first row of $(0, 1, 0, ..., 0)$. The subsequent rows of $\mathbf{J}$ are generated by circularly shifting the previous row by 1 unit. The d circulant delay matrix is then $\mathbf{J}_d = \mathbf{J} \cdot \mathbf{J} \cdots \mathbf{J}$. Then, $\mathbf{P}_s$ in our construction is

$$\mathbf{P}_s = \sum_{\ell=0}^{\rho_e - 1} \sum_{m=0}^{\rho_a - 1} W_{\ell m}^s \mathbf{J}_\ell^T \tilde{\mathbf{C}}_s \mathbf{J}_m. \tag{2.14}$$

We constraint the entries of the $\rho_e \times \rho_a$ weight matrix $\mathbf{W}^s$ in (2.14) to $\mathbb{Q}_q$ so that $\mathbf{P}_s \in \mathbb{Q}_q^{N \times N}$.

A possible choice for $\mathbf{W}^s$ is to set its entries to the elements in $\mathbb{Q}_q$ at random. Although a random choice still illuminates the sector of interest, it does not necessarily lead to a *uniform* illumination pattern across the directions within the sector. Specifically, the entries of $\mathbf{Z}_s$ within the set $\mathscr{A}_s$ will not necessarily have the same magnitude as we can also notice from the example in Fig. 2.5(b). In this chapter, the weights in $\mathbf{W}^s$ are optimized such that the illumination, i.e., the transmitted power, is almost the same across all the directions within each sector. We will show in Section 2.3.3 that having a flat illumination pattern within a sector results in a tight bound on the reconstruction error of the in-sector channel via CS.

We explain how the weights in $\mathbf{W}^s$ are optimized to achieve an almost flat illumination pattern within the sector s. We define an $N \times N$ matrix

$$\mathbf{T}^s = \sum_{\ell=0}^{\rho_e - 1} \sum_{m=0}^{\rho_a - 1} W_{\ell m}^s \mathbf{a}_N(2\pi\ell/N) \mathbf{a}_N^\dagger(2\pi m/N), \tag{2.15}$$

where $\mathbf{a}_N(\omega)$ is the $N \times 1$ Vandermonde vector defined in (2.1). The beamspace representation of the base AWM $\mathbf{P}_s$ in (2.14), referred to as the spectral mask $\mathbf{Z}_s$, is given

by

$$\mathbf{Z}_s = N\mathbf{U}_N^*\mathbf{P}_{s,\mathrm{FC}}\mathbf{U}_N^* \tag{2.16}$$

$$= N[\mathbf{T}^s]^{\mathrm{c}} \odot \tilde{\mathbf{Z}}_s. \tag{2.17}$$

We arrive at (2.17) by first applying flipping and conjugation operations to both sides of (2.14). Then, we use the property that circularly shifting a matrix is equivalent to modulation in the Fourier domain. Next, we observe from (2.13) that $\tilde{\mathbf{Z}}_s$ has equal magnitude entries within the sector s by our construction. For the base AWM $\mathbf{P}_s$ to achieve a uniform illumination over $\mathscr{A}_s$, we notice from (2.17) that the entries of $\mathbf{T}^s$ must have the same magnitude over $\mathscr{A}_s$. We define $\mathbf{T}^s_{\mathscr{A}_s}$ as the $\rho_{\mathrm{e}} \times \rho_{\mathrm{a}}$ submatrix of $\mathbf{T}^s$, comprising entries from $\mathbf{T}^s$ at the indices in $\mathscr{A}_s$. Further, we define an $N \times N$ diagonal matrix $\mathbf{D}_N(\omega) = \mathrm{Diag}(\mathbf{a}_N(\omega))$ that contains the $N \times 1$ Vandermonde vector (2.1) on its diagonal. Then, we can rewrite (2.15) as

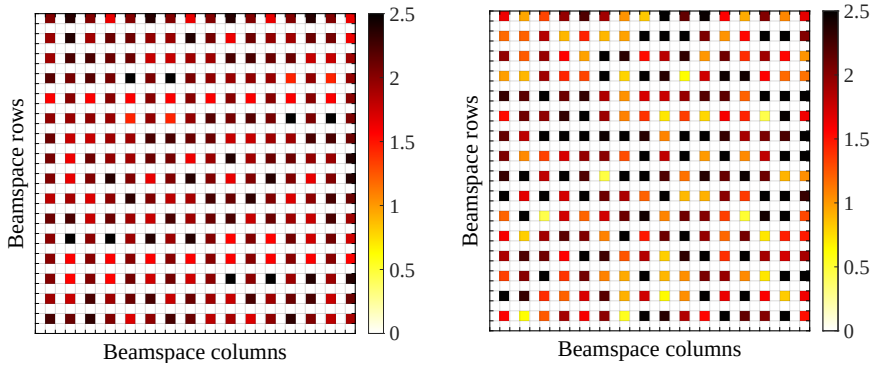
$$\mathbf{T}^s_{\mathscr{A}_s} = \mathbf{U}^*_{\rho_{\mathrm{e}}}\mathbf{D}_{\rho_{\mathrm{e}}}(2\pi k_{\mathrm{e}}/N)\mathbf{W}^s\mathbf{D}_{\rho_{\mathrm{a}}}(2\pi k_{\mathrm{a}}/N)\mathbf{U}^*_{\rho_{\mathrm{a}}}. \tag{2.18}$$

The weight matrix $\mathbf{W}^s \in \mathbb{Q}_q^{\rho_{\mathrm{e}} \times \rho_{\mathrm{a}}}$ should be optimized such that $\mathbf{T}^s_{\mathscr{A}_s}$, i.e., the magnitude of the inverse DFT of $\mathbf{D}_{\rho_{\mathrm{e}}}(2\pi k_{\mathrm{e}}/N)\mathbf{W}^s\mathbf{D}_{\rho_{\mathrm{a}}}(2\pi k_{\mathrm{a}}/N)$ is flat. We can thus write the optimization problem for designing $\mathbf{W}^s$ as

$$\mathscr{O}: \underset{\mathbf{W}^s \in \mathbb{Q}_q^{\rho_{\mathrm{e}} \times \rho_{\mathrm{a}}}, |\mathbf{V}|=\mathbf{1}}{\text{minimize}} \quad \left\| \mathbf{T}^s_{\mathscr{A}_s} - \mathbf{V}/\sqrt{\rho_{\mathrm{e}}\rho_{\mathrm{a}}} \right\|_{\mathrm{F}}. \tag{2.19}$$

The above optimization problem makes sure that the magnitude of entries of $\mathbf{T}^s_{\mathscr{A}_s}$ are as close as possible to $1/\sqrt{\rho_{\mathrm{e}}\rho_{\mathrm{a}}}$ while $\mathbf{W}^s \in \mathbb{Q}_q^{\rho_{\mathrm{e}} \times \rho_{\mathrm{a}}}$. To solve (2.19), we use an alternating minimization approach similar to the PeCAN algorithm proposed in [97]. In the example shown in Fig. 2.5, we observe that the magnitude profile of the spectral mask $|\mathbf{Z}_s|$ with the optimized weights has a more uniform illumination within the sector than the one that uses random weights.

In Fig. 2.6, we compare the received signal strength after SLS using the proposed comb-like sector codebook with those in [39–41]. The AWMs in [39–41] have contiguous beam patterns, and to meet the low-resolution constraint of the phase shifters, we have quantized the corresponding AWMs when necessary. Due to the comb-like structure of our sectors, the indices in $\mathscr{A}_s$ are different with our design and the benchmarks. The received SNR after SLS is proportional to the received power in the best estimated sector. We observe from Fig. 2.6 that our proposed AWMs achieve a higher received power after SLS than the other methods. This is because our comb-like sectors are able to concentrate the transmitter's energy perfectly within the sectors without any leakage, unlike the other methods that suffer due to coarse phase quantization. To measure the power leakage with a codebook, we first define PL_s as the power leakage with AWM $\mathbf{P}_s$ outside $\mathscr{A}_s$. The leakage $\mathrm{PL}_s = 1 - \left(\|\mathbf{Z}_s\{\mathscr{A}_s\}\|_{\mathrm{F}}^2 / \|\mathbf{Z}_s\|_{\mathrm{F}}^2\right)$, where $\mathbf{Z}_s\{\mathscr{A}_s\}$ is a submatrix of the spectral mask $\mathbf{Z}_s$ at the indices in $\mathscr{A}_s$. The power leakage associated with a codebook is then $\mathrm{PL} = (1/S)\sum_{s=1}^{S} \mathrm{PL}_s$. For the codebook in [40], Fig. 2.7 shows the power leakage PL for different resolutions of a 32×32 phased array. Although the set $\mathscr{A}_s$ in our design



(a) Illumination pattern $|\mathbf{Z}_s|$ with the proposed construction and the optimized weights for $\mathbf{W}^s$.

(b) Illumination pattern $|\mathbf{Z}_s|$ with the proposed construction and random weights for $\mathbf{W}^s$.

Figure 2.5: This figure illustrates the spectral masks of our base AWMs designed for SLS. In this example, we use $N = 32$, $N_e = 2$, $N_a = 2$, $k_e = 0$, $k_a = 1$, and $q = 2$ bits. We observe from Fig. (a) and Fig. (b) that using the optimized $\mathbf{W}^s$ in our comb-like construction results in a more uniform illumination pattern than random weights. Here, $\max_{(i,j)\in\mathscr{A}_s} |Z_{s,ij}| / \min_{(i,j)\in\mathscr{A}_s} |Z_{s,ij}|$ is about 1.78 for the pattern in Fig. (a) and is 8.67 for the pattern in Fig. (b).

differs from that in [40], both sets contain the same number of indices to ensure a fair comparison in Fig. 2.7. We observe from Fig. 2.7 that the construction in [40] results in considerable power leakage outside the sectors while our method achieves PL = 0 by construction. It is important to highlight that our design requires a resolution of at least $\log_2(\max\{N_a, N_e\})$ bits at the phase shifters, where N_a and N_e are such that the number of sectors $S = N_a N_e$. After the sector of interest is found using SLS, the TX performs in-sector CS-based channel estimation within this sector for beam acquisition.

2.3. In-sector compressed sensing for channel estimation using the BRP

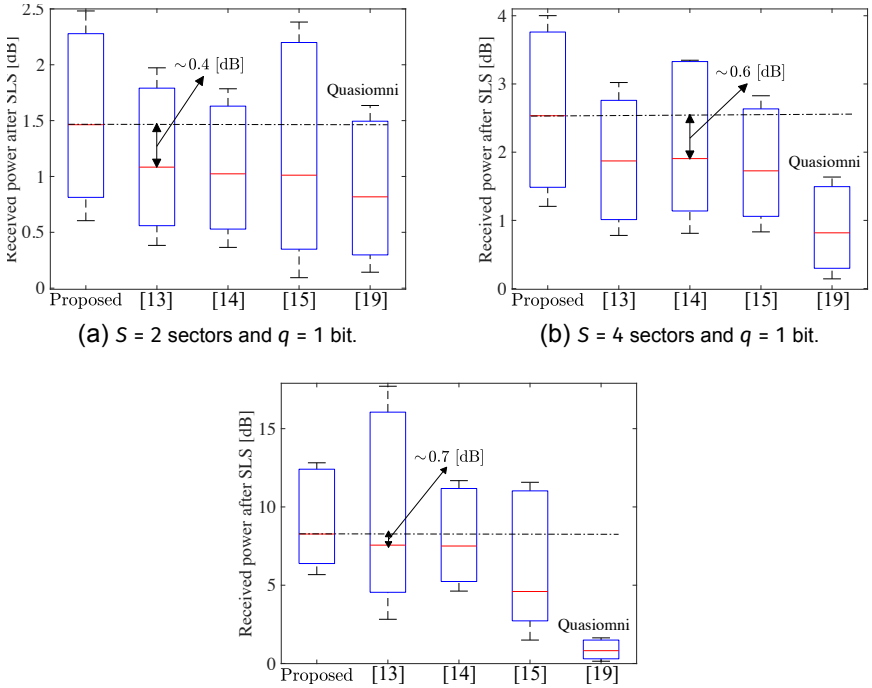
To achieve the full beamforming gain of $10\log_{10}N^2$, the TX needs to further refine the AWM within the sector s_{opt}. This beam refinement can be performed after estimating the N^2/S dimension in-sector channel. In this section, we explain how two-dimensional (2D) CCS (2D-CCS) can be used to acquire channel measurements within the sector s_{opt}. Then, we discuss how to construct an ensemble of AWMs $\{\mathbf{P}_o[m]\}_{m=0}^{M-1}$ that all focus on the sector s_{opt}. Finally, we provide details on the subset of AWMs optimized to reduce aliasing artifacts in CS within the sector.

2.3.1. Preliminaries on 2D-CCS

At the end of SLS, the TX can apply $\mathbf{P}_o$ to illuminate the beamspace indices in $\mathscr{A}_o$, i.e.,

$$\mathscr{A}_o = \left\{(p,q) : (\mathbf{U}_N^* \mathbf{P}_o \mathbf{U}_N^*)_{pq} \neq 0\right\}. \tag{2.20}$$

The use of $\mathbf{P}_o$ at the TX, however, results in a single spatial channel measurement, which is insufficient to estimate the N^2/S entries of the beamspace in $\mathscr{A}_o$. To estimate these entries, multiple AWMs that all focus on the same sector must be designed and applied at the TX. One way to construct such AWMs is to circularly shift $\mathbf{P}_o$ and obtain possibly distinct AWMs. This approach works for in-sector CS because circularly shifting a matrix



(a) $S = 2$ sectors and $q = 1$ bit.

(b) $S = 4$ sectors and $q = 1$ bit.

(c) $S = 16$ sectors and $q = 2$ bits.

Figure 2.6: Our designed base AWMs achieve a higher median received power after SLS than the ones in [39–41] and the quasi-omnidirectional beams in [45]. The benchmarks perform poor for low-resolution phase shifters because their AWMs cannot focus the transmitter's power well within the contiguous sectors. Our AWMs focus power exactly within the comb-like sectors by construction. These plots are obtained for a 32×32 array at the TX. The received power is computed for 100 channel realizations from the NYU channel simulator at 60 GHz and $\sigma^2 = 0$. For $\sigma^2 \neq 0$, the performance of our proposed comb-like illumination pattern relative to other methods will remain the same.

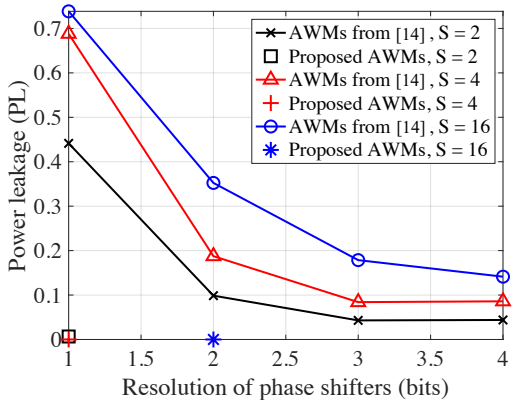


Figure 2.7: For a 32 × 32 phased array, the plot shows that the codebook in [40] results in power leakage outside the defined sectors while our design results in zero leakage. Our codebook can be constructed when the resolution of the phase shifters is at least $\log_2(\max\{N_a, N_e\})$, where the number of sectors can be factorized as $S = N_a N_e$.

does not change the magnitude of its 2D-DFT [93]. The resulting method in which the TX applies circular shifts of an AWM for the RX to acquire channel measurements is called 2D-CCS [45]. Hence, with 2D-CCS, the M in-sector channel measurements $\{y_o[m]\}_{m=0}^{M-1}$ are obtained by applying M distinct circular shifts of $\mathbf{P}_o$ at the TX during BRP.

Now, we discuss mathematical preliminaries on 2D-CCS [98]. With 2D-CCS, the TX can possibly apply N^2 different 2D-circular shifts of $\mathbf{P}_o$ for in-sector channel estimation. As applying all possible circular shifts results in a substantial measurement overhead, the TX only applies $M < N^2$ of these circular shifts to obtain channel measurements. These measurements are used together with a sparse prior for channel estimation. As the AWMs in 2D-CCS are all circular shifts of the base AWM $\mathbf{P}_o$, the channel measurements can be interpreted as a subsampled circular convolution of the channel and $\mathbf{P}_o$ [45]. We use $\Omega = \{(r[m], c[m])\}_{m=0}^{M-1}$ to denote the set of M circular shifts of $\mathbf{P}_o$ applied at the TX to acquire the in-sector channel measurements. Specifically, $\mathbf{P}_o[m]$ is constructed by circularly shifting $\mathbf{P}_o$ by $r[m] \in [N]$ units along its rows, and then circularly shifting the result by $c[m] \in [N]$ units along its columns. We use $\mathbf{H} \circledast \mathbf{P}_o$ to denote the 2D-circular convolution [96] of $\mathbf{H}$ and $\mathbf{P}_o$. We define $\mathscr{P}_\Omega(\mathbf{A})$ as the subsampling operation that returns a vector of size $|\Omega| \times 1$ containing the entries of $\mathbf{A}$ at the indices in Ω. When the TX applies circulant shifts of $\mathbf{P}_o$ according to Ω, the vector of M in-sector channel measurements acquired by the RX is

$$\mathbf{y} = \mathscr{P}_\Omega(\mathbf{H} \circledast \mathbf{P}_o) + \mathbf{v}. \tag{2.21}$$

We observe that the channel measurements in 2D-CCS are determined by the base AWM $\mathbf{P}_o$ and the set of circular shifts Ω.

We now describe the in-sector sensing structure in (2.21) by expressing $\mathbf{H}$ and $\mathbf{P}_o$ in the 2D-DFT domain. Similar to (2.16), the spectral mask $\mathbf{Z}_o$ corresponding to $\mathbf{P}_o$ is defined as

$$\mathbf{Z}_o = N \mathbf{U}_N^* \mathbf{P}_{o,FC} \mathbf{U}_N^*. \tag{2.22}$$

From (2.20) and the properties of the 2D-DFT, it can be shown that $Z_o(k,\ell) \neq 0$ if $(k,\ell) \in \mathscr{A}_o$ and $Z_o(k,\ell) = 0$ if $(k,\ell) \notin \mathscr{A}_o$. We define the masked beamspace $\mathbf{R}_o$ as the entry-wise product of the spectral mask $\mathbf{Z}_o$ and the beamspace $\mathbf{X}$, i.e.,

$$\mathbf{R}_o = \mathbf{X} \odot \mathbf{Z}_o. \tag{2.23}$$

As $\mathbf{Z}_o$ contains only $|\mathscr{A}_o|$ number of non-zeros, the masked beamspace $\mathbf{R}_o$ in (2.23) has utmost $|\mathscr{A}_o|$ non-zeros. Under the assumption that $|\mathscr{A}_o| \sim O(N^2)$, $\mathbf{R}_o$ exhibits a sparse structure due to the sparsity in $\mathbf{X}$. Now, we use the property that the 2D-DFT of the 2D circular convolution of two matrices is equal to the scaled entry-wise product of their 2D-DFTs [96] to rewrite (2.21) as

$$\mathbf{y} = \mathscr{P}_\Omega(\mathbf{U}_N \mathbf{R}_o \mathbf{U}_N) + \mathbf{v}. \tag{2.24}$$

We observe from (2.24) that the channel measurements in 2D-CCS are simply a subsampled 2D-DFT of a sparse masked beamspace. Here, the mask is the discrete beam pattern corresponding to the base AWM $\mathbf{P}_o$, and the subsampling is performed at the indices in Ω.

Now, we explain how in-sector channel estimation is performed with the 2D-CCS-based measurements. Any algorithm that uses the measurements $\mathbf{y} = \mathscr{P}_\Omega(\mathbf{U}_N(\mathbf{X} \odot \mathbf{Z}_o)\mathbf{U}_N)$ in (2.21) can only estimate the entries of $\mathbf{X}$ at the indices in $\mathscr{A}_o$, since the spectral mask $\mathbf{Z}_o$ blanks out entries of $\mathbf{X}$ outside $\mathscr{A}_o$. In this chapter, we use the OMP, a greedy CS algorithm, to obtain the beamspace estimate $\hat{\mathbf{X}}_o$ from the M in-sector measurements. The channel estimate within the sector of interest is then $\hat{\mathbf{H}}_o = \mathbf{U}_N \hat{\mathbf{X}}_o \mathbf{U}_N$.

The success of sparse recovery from partial 2D-DFT measurements in (2.24) depends on the choice of the subsampling indices in Ω. Prior work has demonstrated that subsampling indices chosen at random allow sparse recovery with a high probability [99]. It is also known that random subsampling often results in aliasing artifacts that are almost uniformly spread over the support of the reconstruction, i.e., $[N] \times [N]$ [100]. Many CS algorithms, like the OMP, iteratively cancel the aliasing artifacts to reconstruct the sparse signal. In our in-sector CS problem, a uniformly spread aliasing profile is undesirable because the sparse masked beamspace signal in (2.24) is non-zero only at the indices in $\mathscr{A}_o$. To minimize the reconstruction error, the subsampling set Ω should be constructed such that aliasing artifacts over $\mathscr{A}_o$ are kept small, while the artifacts outside $\mathscr{A}_o$ are not critical since the masked beamspace is known to be zero in those regions. We construct such sets in Sec. 2.3.2.

2.3.2. Proposed circular shifts for in-sector CS

We discuss our procedure to optimize the set of the circular shifts in Ω to reduce the aliasing artifacts within the sector s_{opt}.

To aid our optimization, we write down the partial 2D-DFT-based measurement model, i.e.,

$$\mathbf{y} = \mathscr{P}_\Omega(\mathbf{U}_N(\mathbf{Z}_o \odot \mathbf{X})\mathbf{U}_N) + \mathbf{v}, \tag{2.25}$$

using (2.24) and (2.23). This measurement model is equivalent to a linear measurement model of the $N^2 \times 1$ sparse vector $\mathbf{x} = \text{vec}(\mathbf{X})$. To represent the linear model explicitly,

we replace the sampling operator $\mathscr{P}_\Omega(\cdot)$ with an $M \times N^2$ subsampling matrix $\mathbf{S}$ which is equal to 1 at locations in $\{(m, Nc[m] + r[m])\}_{m=0}^{M-1}$ and zero at the remaining locations. Thus, $\mathbf{S}$ consists of exactly $|\Omega| = M$ ones representing the circular shift indices within Ω. We use $\mathbf{z}_o = \mathrm{vec}(\mathbf{Z}_o)$ to denote the $N^2 \times 1$ vector form of the spectral mask $\mathbf{Z}_o$. The CS matrix that compresses $\mathbf{x}$ to $\mathbf{y}$ can be determined from (2.25) as

$$\mathbf{A}_o = \mathbf{S}(\mathbf{U}_N \otimes \mathbf{U}_N)\,\mathrm{Diag}(\mathbf{z}_o). \tag{2.26}$$

The linear measurement model in (2.24) can be rewritten as

$$\mathbf{y} = \mathbf{A}_o\mathbf{x} + \mathbf{v}. \tag{2.27}$$

Note that the set of the circular shifts $\Omega = \{(r[m], c[m])\}_{m=0}^{M-1}$ determines the subsampling matrix $\mathbf{S}$ and therefore the CS matrix $\mathbf{A}_o$.

The performance of greedy CS algorithms for sparse recovery usually depends on the mutual coherence of the CS matrix [101]. The mutual coherence is simply the maximum of the normalized inner product between the columns of the CS matrix. It was shown in [101] that minimizing the mutual coherence of the CS matrix results in a tight upper bound for the MSE in the sparse estimate. To this end, we optimize the set of the circulant shifts Ω to minimize the mutual coherence of the CS matrix in our problem.

The CS matrix in our problem has a special structure due to our comb-like construction for the sectors. We define $\mathscr{L}_o = \{qN + p : (p, q) \in \mathscr{A}_o\}$ as the set of the 1D indices where the comb-like spectral mask $\mathbf{Z}_o$ is known to be non-zero. Specifically, for the sector of interest s_{opt}, the vectorized spectral mask $\mathbf{z}_o$ is non-zero at the locations in $\mathscr{L}_o$ and zero at the other locations $[N^2]\backslash\mathscr{L}_o$. Therefore, we can equivalently write the measurements (2.27) as

$$\mathbf{y} = \mathbf{A}_{\mathscr{L}_o}\mathbf{x}_{\mathscr{L}_o} + \mathbf{v}. \tag{2.28}$$

In (2.28), $\mathbf{A}_{\mathscr{L}_o}$ is the $M \times \rho_e\rho_a$ submatrix of $\mathbf{A}_o$ obtained by retaining the columns with indices in $\mathscr{L}_o$, and $\mathbf{x}_{\mathscr{L}_o}$ is a subvector of $\mathbf{x}$ with the indices in $\mathscr{L}_o$. So, the in-sector CS problem is to estimate this $\rho_e\rho_a = N^2/S$ dimension sparse vector $\mathbf{x}_{\mathscr{L}_o}$ from the M measurements. Also, sparse recovery in our in-sector CS problem (2.28) depends on the coherence of $\mathbf{A}_{\mathscr{L}_o}$ rather than $\mathbf{A}_o$.

Now, we use the notion of the point spread function (PSF) in partial 2D-DFT CS problem [100] to aid our optimization of the sampling set Ω. We define the $N \times N$ binary matrix $\mathbf{N}_\Omega$ that is 1 at the indices $\Omega = \{(r[m], c[m])\}_{m=0}^{M-1}$ and is zero at other indices. Therefore, $\mathbf{N}_\Omega$ has exactly M entries with the value of 1. The corresponding PSF is [45, 100]

$$\mathrm{PSF}_o = \frac{N}{M}\mathbf{U}_N^*\mathbf{N}_\Omega\mathbf{U}_N^*. \tag{2.29}$$

In a standard partial 2D-DFT CS problem, the coherence of the CS matrix is simply the maximum sidelobe level of the PSF. In our in-sector CS problem, however, only the sidelobes within the sector of interest matter. With our comb-like sectors constructed according to (2.9), it can be shown that the coherence of the effective CS matrix $\mathbf{A}_{\mathscr{L}_o}$ is

$$\mu_o = \max_{\{(i,j)\in\mathscr{T}\}} \left|\mathrm{PSF}_o(i,j)\right|, \tag{2.30}$$

where $\mathscr{T} = \{(i,j)\colon i = nN_e, j = mN_a, (n,m) \in [\rho_e] \times [\rho_a] \setminus (0,0)\}$ for any comb-like sector with the set of beamspace indices defined in (2.9). The maximum sidelobe level of the PSF at the locations in $\mathscr{T}$ determines the mutual coherence of the CS matrix in our in-sector CS problem.

Now, we formulate the coherence minimization problem as

$$\mathscr{P}\colon \begin{cases} \min\limits_{\mathbf{N}_\Omega \in \{0,1\}^{N\times N}} & \max\limits_{(i,j)\in\mathscr{T}} \left|\mathrm{PSF}_o(i,j)\right| \\ \text{s.t.} & \sum\limits_{(i,j)\in[N]\times[N]} N_\Omega(i,j) = M. \end{cases} \tag{2.31}$$

The problem $\mathscr{P}$ is non-convex and hard to solve. In Lemma 1, however, we present an optimal solution $\mathbf{N}_\Omega^{\mathrm{Nyq}}$ that achieves $\mu_o = 0$ when $M = \rho_e\rho_a$. This case corresponds to the Nyquist sampling criterion where the number of measurements M is equal to the sector dimension N^2/S. For the sub-Nyquist regime where $M < \rho_e\rho_a$, our randomized subsampling technique just selects M elements at random from the Nyquist sampling set.

Lemma 1. *For $M = \rho_e\rho_a$, the matrix $\mathbf{N}_\Omega^{\mathrm{Nyq}}$ corresponding to $\Omega = [\rho_e] \times [\rho_a]$, with ρ_e and ρ_a defined in (2.7) and (2.8), is an optimal solution of $\mathscr{P}$ since it achieves $\mu_o = 0$ in (2.30).*

Proof. See Section 2.7.1. □

We now discuss the Nyquist sampling criteria for the N^2 dimension channel estimation problem and the N^2/S dimension in-sector channel estimation problem. In the former case, the Nyquist sampling criterion is $M = N^2$ and the optimal solution $\mathbf{N}_\Omega$ for $\mathscr{P}$ is an $N \times N$ matrix of all ones corresponding to $\Omega = [N] \times [N]$. The PSF corresponding to $\Omega = [N] \times [N]$ can be shown to be zero at all entries except $(0,0)$. For the in-sector CCS problem with our comb-like sectors, the ideal PSF need not be zero at all these locations to achieve $\mu_o = 0$. Instead, it only needs to be zero at the locations defined by $\mathscr{T}$ in (2.30). In Lemma 1, we propose a construction for Ω that achieves $\mu_o = 0$ for $M = N^2/S$. This set simply comprises the 2D-coordinates of any $\rho_e \times \rho_a$ contiguous block on the $[N] \times [N]$ grid.

We use $\mathbf{N}_\Omega^{\mathrm{Nyq}}$ to denote the optimal solution of $\mathscr{P}$ when $M = N^2/S$. In the sub-Nyquist regime, i.e., $M < \rho_e\rho_a$, we propose to select M entries from the Nyquist set of circular shifts in Lemma 1, i.e., $[\rho_e] \times [\rho_a]$, uniformly at random without replacement. We refer to this method as the proposed circular shifts (PCS) for $M < \rho_e\rho_a$, i.e., the subsampling set Ω in PCS is a subset of the optimal set in Lemma 1 for $M = \rho_e\rho_a$. Hence, PCS converges to the optimal solution proposed in Lemma 1 as $M \to \rho_e\rho_a$. We would like to mention that PCS is different from the fully random circular shifts (RCS) used in [45]. With RCS, the M circular shifts are chosen at random from $[N] \times [N]$, while PCS selects the M shifts at random from $[\rho_e] \times [\rho_a]$ as shown in Fig. 2.8.

To illustrate the effectiveness of our PCS in the sub-Nyquist regime $M < \rho_e\rho_a$ over RCS, we examine the mutual coherence of the CS matrices. In Fig. 2.9(a) and 2.9(b), we compare the entry-wise magnitude of the PSF (2.29) obtained from PCS and RCS in the Nyquist regime $M = N^2/S$. We observe that with PCS, the PSF is 0 at the indices of interest, i.e., $\mathscr{T}$, while RCS which selects M indices at random from $[N] \times [N]$ results in a

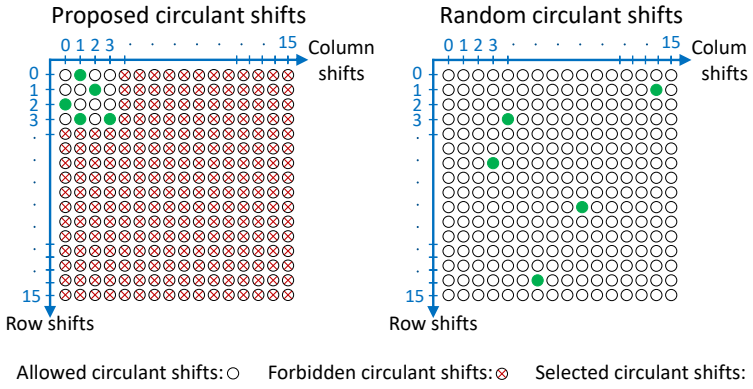


Figure 2.8: One realization of the circular shifts chosen in our PCS scheme and the RCS scheme are shown for $N = 16$, $N_e = 4$, $N_a = 4$, $S = N_e N_a$, $\rho_e = N/N_e$, $\rho_a = N/N_a$, and $M = 5$. While the circular shifts in PCS are chosen at random from $\{0,1,2,3\} \times \{0,1,2,3\}$, the shifts in RCS are chosen at random from $[16] \times [16]$.

higher coherence μ_o. Next, we plot the cumulative distribution function of μ_o for the two sampling schemes in Fig. 2.9(c). We notice from the cumulative distribution function that PCS achieves a smaller coherence for the effective CS matrix $\mathbf{A}_{\mathscr{L}_o}$ than RCS. Finally, we show an instance of PCS and RCS in Fig. 2.8 to distinguish the difference between the two sampling schemes.

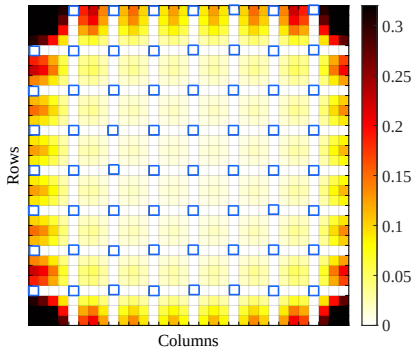
2.3.3. Guarantees on in-sector CS with the OMP

In this section, we discuss our guarantees for OMP-based sparse channel estimation within the sector of interest $\mathscr{A}_o$. Our guarantees are an extension of the coherence-based guarantees in [101], and they help us study the impact of variations in the illumination pattern within the sector. Here, we only provide the main results, and we refer the interested reader to our work in [95] for more details.

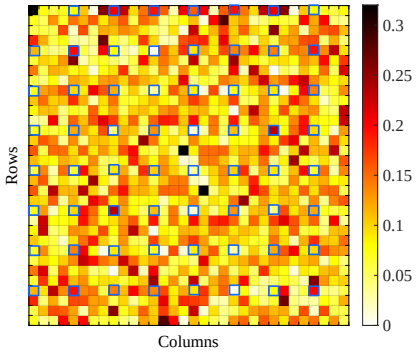
Prior work on coherence-based guarantees for the OMP makes a strong assumption that the CS matrix has equal column norms. In our problem, the norms of the columns within the CS matrix in (2.26) are controlled by $\mathbf{z}_o$. To express our guarantees as a function of the variations in $\mathbf{z}_o$, we define $\mathbf{z}_{\mathscr{L}_o}$ as a $\rho_e\rho_a \times 1$ subvector of the vectorized spectral mask $\mathbf{z}_o$ obtained by retaining indices in $\mathscr{L}_o$. We use d_i to denote the ℓ_2-norm of the i^{th} column of $\mathbf{A}_{\mathscr{L}_o}$. Using the definition of $\mathbf{A}_o$ in (2.26), we can write that

$$d_i = \frac{\sqrt{M}}{N} |z_{\mathscr{L}_o}(i)|, \quad \forall i \in [\rho_e\rho_a], \tag{2.32}$$

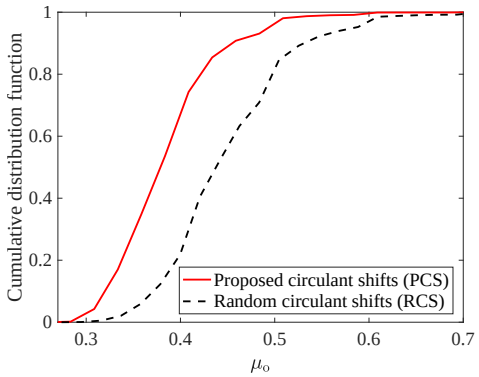
where $z_{\mathscr{L}_o}(i)$ denotes the i^{th} entry of $\mathbf{z}_{\mathscr{L}_o}$. The maximum and the minimum of the column norms in $\mathbf{A}_{\mathscr{L}_o}$ are defined as $d_{\max} = \max_{j \in [\rho_e\rho_a]} d_j$ and $d_{\min} = \min_{j \in [\rho_e\rho_a]} d_j$. As the ℓ_2-norm of $\mathbf{z}_{\mathscr{L}_o}$ is equal to the Frobenius norm of the AWM $\mathbf{P}_o$, i.e., 1, we have $\sum_{i=1}^{\rho_e\rho_a} d_i^2 = M/N^2$.



(a) |PSF| for the proposed circulant shifts (PCS) with $\mu_o = 0$ from (2.30).



(b) |PSF| for the random circulant shifts (RCS) with $\mu_o \approx 0.23$ from (2.30).



(c) Cumulative distribution function of the coherence of the effective CS matrix $\mathbf{A}_{\mathscr{L}_o}$ for PCS and RCS.

Figure 2.9: Proposed subsampling and fully random subsampling for $N = 32$, $N_e = 4$, $N_a = 4$, $S = N_e N_a$, $\rho_e = N/N_e$, $\rho_a = N/N_a$. In (a) and (b), we assume $M = \rho_e \rho_a$ and the entries indicated by □ have indices in $\mathscr{T}$. The PSF for PCS achieves $\mu_o = 0$, which corresponds to zero aliasing artifacts under Nyquist sampling. For the same number of measurements, however, the PSF with RCS results in a positive μ_o. We observe from (c) that the coherence with PCS is lower than that with RCS in the sub-Nyquist regime for $M = 20$.

Due to this energy constraint,

$$0 < d_{\min} \leq d_{\max} < \sqrt{M}/N. \tag{2.33}$$

The entries of the spectral mask in $\mathbf{z}_{\mathscr{L}_o}$ are proportional to the norm of the columns of the CS matrix.

Now, we provide performance guarantees to recover the sparse in-sector beamspace

$\mathbf{x}_{\mathscr{L}_o}$ from the noisy CS measurements in (2.27) using the OMP. The guarantees provided in Theorem 1 set an upper bound on the reconstruction error and a lower bound on the probability of successful support recovery depending on the strength of the weakest component of the sparse signal, the noise power, and the sparsity level.

Theorem 1. *Let* $x_{\mathscr{L}_o,\min} = \min_{j\in\Pi} |x_{\mathscr{L}_o,j}|$, *where* Π *denotes the support of* $\mathbf{x}_{\mathscr{L}_o}$ *with cardinality* $|\Pi| = K$. *If*

$$d_{\min} x_{\mathscr{L}_o,\min} - (2K-1)\mu_o d_{\max} x_{\mathscr{L}_o,\min} \geq 2\gamma, \tag{2.34}$$

for $\gamma > 0$, *with probability exceeding*

$$\Pr\{E\} \geq \left(1 - \sqrt{\frac{2}{\pi}}\cdot\sqrt{\frac{\sigma}{\gamma}}\exp\left(-\frac{\gamma^2}{2\sigma^2}\right)\right)^{2N^2/S}. \tag{2.35}$$

the OMP algorithm successfully recovers the support of $\mathbf{x}_{\mathscr{L}_o}$ *and the MSE of the estimate* $\hat{\mathbf{x}}_{\mathscr{L}_o}$ *is upper bounded as*

$$\|\hat{\mathbf{x}}_{\mathscr{L}_o} - \mathbf{x}_{\mathscr{L}_o}\|_2^2 \leq \left(\frac{d_{\max}}{d_{\min}}\right)^2 \frac{K\gamma^2}{(d_{\min} - (K-1)\mu_o d_{\max})^2}. \tag{2.36}$$

Proof. See Section 2.7.2. □

An important insight from Theorem 1 is that the OMP can identify weak coefficients in the beamspace when $d_{\max}/d_{\min}$ of the CS matrix is small. This observation follows by rewriting (2.34) as

$$x_{\mathscr{L}_o,\min} \geq \frac{2\gamma}{d_{\min} - (2K-1)\mu_o d_{\max}} \tag{2.37}$$

$$= \frac{2\gamma/d_{\min}}{1 - \mu_o(2K-1)(d_{\max}/d_{\min})}. \tag{2.38}$$

The upper bound on the MSE in (2.36) monotonically increases with $d_{\max}/d_{\min}$. In summary, our guarantees on support recovery and the MSE become tight for a small $d_{\max}/d_{\min}$. We notice from (2.33) that the smallest possible $d_{\max}/d_{\min}$ is 1. This is achieved when $\min_{i\in[\rho_e\rho_a]} |z_{\mathscr{L}_o}(i)| = \sqrt{S}/N$, equivalently when $\mathbf{P}_o$ uniformly illuminates the sector of interest. These guarantees motivate our optimization objective in (2.19), which aims to generate AWMs that result in an almost uniform illumination within the sector. Further, it can be concluded from our guarantees that our optimized illumination pattern in Fig. 2.5(a) should achieve better in-sector channel reconstruction than the pattern in Fig. 2.5(b).

2.4. Numerical Results

In this section, we first describe how the proposed beamforming method is extended to a wideband system using the IEEE 802.11ad frame structure. Then, we discuss the

performance of our proposed method in terms of the normalized MSE in the in-sector channel estimate and the achievable rate.

2.4.1. Extension to a wideband system

We denote the L–tap wideband channel as $\{\mathbf{H}[\ell]\}_{\ell=0}^{L-1}$ where $\mathbf{H}[\ell] \in \mathbb{C}^{N\times N}$. The TX first identifies the best sector by applying S different AWMs $\{\mathbf{P}_s\}_{s=0}^{S-1}$ in SLS. The best sector is the one for which the received power, i.e., $\sum_{\ell=0}^{L-1} |\langle \mathbf{H}[\ell], \mathbf{P}_s\rangle + v_\ell|^2$ is the highest and the corresponding base AWM is $\mathbf{P}_o$. Then, the TX acquires CS measurements within this sector by applying different AWMs $\{\mathbf{P}_o[m]\}_{m=0}^{M-1}$ at the beginning of each training (TRN) subfield in a BRP packet. As shown in Fig. 2.10, to obtain the m^{th} measurement, at the beginning of the m^{th} TRN subfield, TX sets the antenna weights at the uniform planar array to $\mathbf{P}_o[m]$. With our design, $\mathbf{P}_o[m]$ is a 2D-circular shift of $\mathbf{P}_o$.

In IEEE 802.11ad, a Golay sequence Ga_{128} of length 128 is used as a guard interval at the beginning of the TRN subfield [102]. This interval provides enough time to change the antenna weights at the TX and also avoids interference across successive subfields. Then, the TX transmits two Golay complementary pairs $\{-\text{Gb}_{128}, \text{Ga}_{128}\}$ and $\{\text{Gb}_{128}, \text{Ga}_{128}\}$ for each AWM, i.e., $\mathbf{P}_o[m]$, applied at the array. The complementary property of Golay sequences can be used to obtain measurements of the equivalent wideband channel seen through the beamformer $\mathbf{P}_o[m]$ [103]. This channel is $\{\langle \mathbf{H}[\ell], \mathbf{P}_o[m]\rangle\}_{\ell=0}^{L-1}$, a vector of length L for each $m \in [M]$. We define $\mathbf{Y}_{\text{blk}} \in \mathbb{C}^{M\times L}$ to denote the set of these ML spatial channel projections within the sector of interest and the measurement noise as $\mathbf{V}_{\text{blk}} \in \mathbb{C}^{M\times L}$. Hence, the $(m,\ell)^{\text{th}}$ entry of $\mathbf{Y}_{\text{blk}}$ is

$$Y_{\text{blk}}(m,\ell) = \langle \mathbf{H}[\ell], \mathbf{P}_o[m]\rangle + V_{\text{blk}}(m,\ell). \tag{2.39}$$

Due to the spreading gain of $N_{\text{seq}} = 256$ with the Golay sequence correlators, the entries of $\mathbf{V}_{\text{blk}}$ are independently and independently distributed as $\mathcal{CN}(0, \sigma^2/N_{\text{seq}})$.

In this chapter, only the DC subcarrier, i.e., the sum of the L taps of the wideband channel $\mathbf{H}_{\text{sum}} = \sum_{\ell=0}^{L-1} \mathbf{H}[\ell]$ is estimated within the sector of interest. Although our method can be applied to estimate all the taps, we only estimate the sum of the channel taps as it requires less computational complexity than wideband channel estimation. This approach is reasonable because the phased array can only apply a frequency-flat beamformer, which may be designed based on a single subcarrier. The in-sector CS measurements are thus $y[m] = \sum_{\ell=0}^{L-1} Y_{\text{blk}}(m,\ell),\ \forall m \in [M]$. Now, similar to the narrowband case, in-sector CS can be performed using $\{y[m]\}_{m=0}^{M-1}$ to obtain an estimate of the DC subcarrier $\hat{\mathbf{H}}_{\text{sum}}^{o}$ within the sector of interest. This estimate can then be used to design the conjugate beamformer at the transmitter to maximize the received power.

2.4.2. System parameters

We consider a 32 × 32 phased array in Fig. 2.1(a) at the TX and we assume 1-bit phase shifters unless otherwise stated. The TX-RX distance is 60 m. The heights of the TX and the RX are 3 m and 1.5 m. The system operates at a carrier frequency of 60 GHz with a bandwidth of 100 MHz that corresponds to a symbol duration of 10 ns. We use SNR_{omni} to denote the SNR without spreading gain and without any beamforming gain. This is

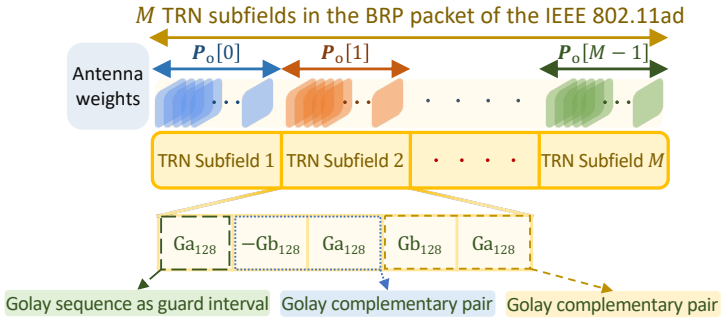


Figure 2.10: BRP packet in IEEE 802.11ad may comprise upto 64 TRN subfields containing Golay complementary pairs [91, Table 20-13, Sec. 20.9.2.2.6], [102]. The in-sector CS measurements can be acquired within the sector of interest s_{opt}, when the TX sequentially applies $\{\mathbf{P}_{\text{o}}[m]\}_{m=0}^{M-1}$ at the beginning of each TRN subfield.

obtained when the TX applies a quasi-omnidirectional beamformer generated by a perfect binary array [45]. We use mmWave channels obtained from the NYU channel simulator [6] for a line-of-sight scenario in an urban micro environment. For 100 independent channel realizations, the omnidirectional root-mean-square delay spread was less than 25 ns in more than 90% of the channel realizations. We model the wideband channel using $L = 10$ taps corresponding to a duration of 100 ns, which is much larger than 25 ns. The simulation results in this section are averaged over the 100 channel realizations.

2.4.3. Performance of the Proposed In-Sector CS-Based Beamforming

We discuss metrics used to evaluate the performance of the proposed method against two other benchmarks. We also assess the performance gain arising from each of our three contributions: (i) Construction of non-overlapping comb-like sectors in SLS, (ii) Optimization of the weight matrix $\mathbf{W}^s$ to achieve an almost uniform illumination pattern, and (iii) Optimization of the circular shifts in PCS for in-sector CS.

To construct the base AWMs in our approach for SLS, we initialize the weight matrix within the iterative optimization algorithm in (2.19) to different matrices that exhibit good autocorrelation properties. This is done because the algorithm (2.19) can converge to different local minima depending on the initialization [104]. To this end, we use perfect binary arrays [105], the quantized DFT matrix, the outer product of two Frank sequences [106], the outer product of two Zadoff-Chu sequences [107], and the outer product of two Golomb sequences [108] for the initialization. Then, the initialization that results in the most uniform magnitude profile within our comb-like construction is used to construct a base AWM for the sector.

For CS-based channel estimation within the sector of interest, we used the OMP algorithm [8] to obtain an estimate of $\mathbf{H}_{\text{sum}}$. Assuming that the $(N^2/S) \times 1$ vector $\mathbf{x}_{\mathscr{L}_{\text{o}}}$ in the in-sector CS measurement model (2.28) is K-sparse with $K \ll \mathcal{O}(N^2/S)$, the computational complexity of the OMP algorithm is dominated by the matrix-vector mul-

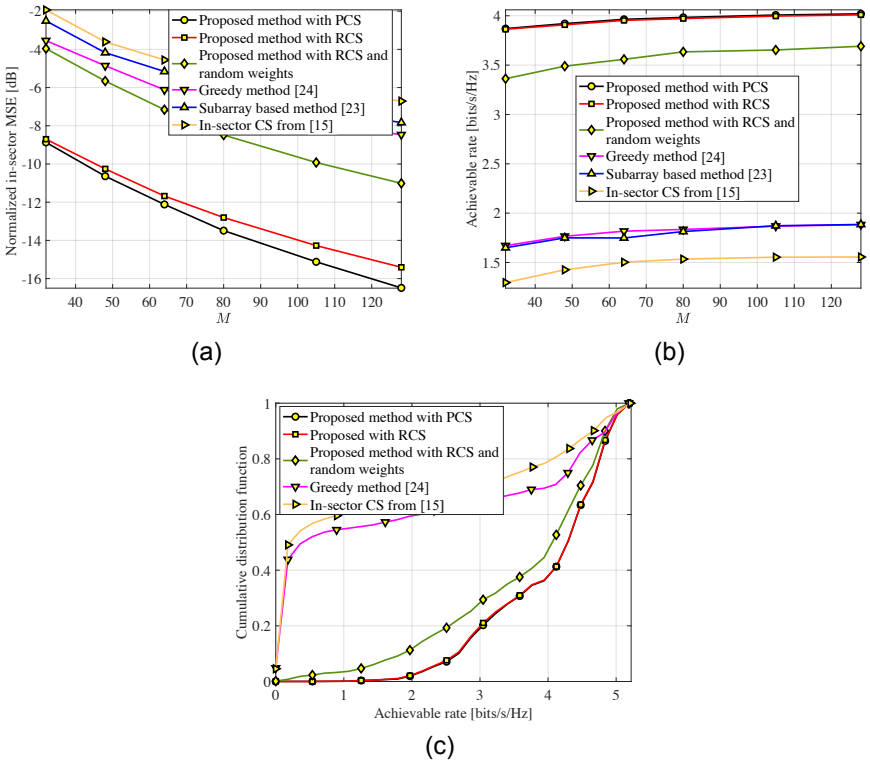


Figure 2.11: The plots show the normalized in-sector MSE, achievable rate, and the cumulative distribution function of the achievable rate using the proposed in-sector CCS-based method as well as the benchmarks. We assume $N_e = 2$, $N_e = 2$ which results $S = 4$ sectors. We use $\mathrm{SNR}_{\mathrm{omni}} = -10$ dB. In (c), we use $M = 80$. It can be noticed that the proposed in-sector CCS-based method outperforms the benchmarks when the weight matrix $\mathbf{W}^s$ in (2.14) is optimized.

tiplication $\mathbf{A}^*_{\mathscr{L}_o}\mathbf{y}$ [109]. The partial 2D-DFT structure of the CS matrix $\mathbf{A}_{\mathscr{L}_o}$ in our case, allows fast implementation of $\mathbf{A}^*_{\mathscr{L}_o}\mathbf{y}$ using the fast Fourier transform. It was discussed in [45] that the complexity of each OMP iteration with such a partial 2D-DFT matrix is $\mathcal{O}\left((N^2/S)\log(N/\sqrt{S})\right)$. For a K-sparse signal, OMP requires at most K iterations to converge [101]. Algorithms based on convex optimization can also be used for sparse recovery. One such algorithm, based on the subgradient method, minimizes an ℓ_1-regularized least squares objective [101]. To achieve ϵ-error in the objective, this method requires $\mathcal{O}(1/\epsilon^2)$ iterations [110]. In our simulations, we observed that the OMP required about 30 iterations, whereas the subgradient method needed several thousand iterations to achieve comparable performance.

We study the performance of 2D-CCS-based with both RCS and the optimized PCS.

We note that in RCS, it is possible to have the same measurement equation corresponding to different $\mathbf{P}_o[m]$s. We remove such measurements to avoid rank deficiency in OMP sparse recovery. Let $\mathbf{H}^o_{\text{sum}}$ denote the spatially filtered version of $\mathbf{H}_{\text{sum}}$ within the sector of interest. The matrix $\mathbf{H}^o_{\text{sum}}$ is obtained by applying a binary mask, that is one at the indices in $\mathscr{A}_o$, over the beamspace representation of $\mathbf{H}_{\text{sum}}$. We then define the normalized in-sector MSE $\mathbb{E}[\|\mathbf{H}^o_{\text{sum}} - \hat{\mathbf{H}}^o_{\text{sum}}\|_F^2]/\mathbb{E}[\|\mathbf{H}^o_{\text{sum}}\|_F^2]$ as a measure of the reconstruction error in $\hat{\mathbf{H}}^o_{\text{sum}}$. In addition, we use SNR_{omni} to denote the signal-to-noise ratio when a quasi-omnidirectional beam is used at the TX.

After the in-sector channel is estimated, the TX applies a beamformer $\mathbf{F} \in \mathbb{Q}_q^{N\times N}$ such that $|\langle \mathbf{F}, \hat{\mathbf{H}}^o_{\text{sum}}\rangle|$ is maximized. Let $\text{phase}(\kappa)$ denote the phase of the complex scalar κ. When there is no constraint on the resolution of the phase shifters, i.e., $q = \infty$, from the dual norm inequality [111], $F^{\text{opt}}_{ij} = \exp\big(\text{jphase}([\hat{H}^o_{\text{sum}}]_{ij})\big)/N$ is the maximizer of $|\langle \mathbf{F}^{\text{opt}}, \hat{\mathbf{H}}^o_{\text{sum}}\rangle|$. Under q-bit phase quantization, the TX applies $\mathbf{F} = \mathscr{Q}_q(\mathbf{F}^{\text{opt}})$ as the transmit beamformer. The effective wideband single-input single-output channel is then $\{\langle \mathbf{H}[\ell], \mathbf{F}\rangle\}_{\ell=0}^{L-1}$. Finally, we use the *waterfilling* algorithm to obtain the achievable rate associated with the equivalent channel[87].

For the benchmarks, we use the methods from [41, 49] and [50] that use contiguous sectors similar to Fig. 2.3(a). These sectors are defined by $\mathscr{B}_s = \{(p,q) : k_e\rho_e \le p < (k_e+1)\rho_e,\ k_a\rho_a \le q < (k_a+1)\rho_a\}$, each representing a $\rho_e \times \rho_a$ block in the beamspace. The M in-sector CS measurements in [41] are acquired using AWMs that are obtained by modulating a Zadoff-Chu-based sequence with M columns of the 2D-DFT dictionary chosen at random. To construct an AWM that illuminates a contiguous sector, [49] partitions the array into multiple subarrays. Then, it stacks the phase-modulated version of the subarray response vectors and takes the outer product of such vectors to construct the AWM [49]. To construct M different AWMs, the angles used for phase modulation are selected at random such that the resultant AWM focuses on the given sector. For the benchmark in [49], we partition the array into 4 subarrays, a parameter tuned to optimize the method in [49] for our channel dataset. Implementing the AWMs in [41] requires $\log_2 N$ resolution phase shifters and the design in [49] needs infinite-resolution phase shifters, which may not be available in practice. For a fair comparison, we quantize the phase of these codewords to q-bits. For the greedy method in [50], we generate one million AWMs from $\mathbb{Q}_q^{N\times N}$ at random. Then, for each sector in $\{\mathscr{B}_s\}_{s=0}^{S-1}$, we select the top AWM that radiates the largest energy within the sector to be used during the SLS. After determining the sector of interest in SLS, the top M AWMs that radiate the largest energy within the sector of interest are selected to obtain in-sector CS measurements with [50].

In Fig. 2.11(a), we compare the normalized in-sector MSE of the estimated channel for the proposed in-sector CCS-based method and the benchmarks. We observe a large gap between the proposed method with PCS or RCS and the benchmarks [41, 49] and [50] for a wide range of CS measurements. This is because our AWMs result in a higher received power within the sector of interest, boosting the received SNR and therefore resulting in a low normalized in-sector MSE. The high received power with our method is due to the concentration of the transmitter's energy only within the sector, while the AWMs in [41, 49] and [50] result in leakage outside the sector. Also, we observe that our method with PCS results in a lower normalized in-sector MSE than with RCS within

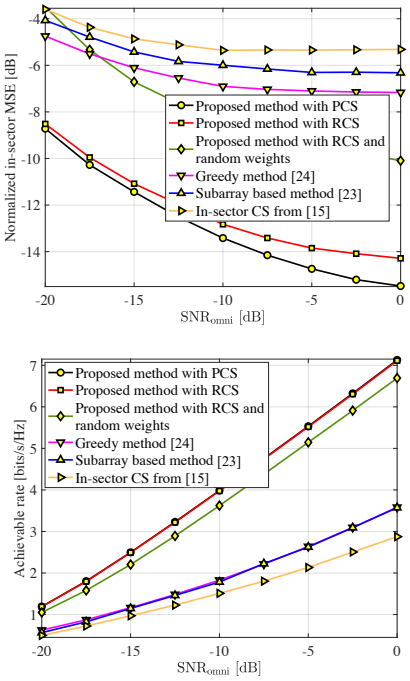


Figure 2.12: Here, we consider $N_e = 2$, $N_e = 2$ that correspond to $S = 4$ sectors, and use $M = 80$ in-sector CS measurements. For in-sector CS with our optimized comb-like sectors, we observe that PCS achieves a lower normalized MSE than RCS. This is because the designed set of circular shifts results in lower aliasing artifacts than random shifts. Furthermore, our method outperforms the benchmarks in [41, 49, 50] in terms of the normalized MSE and the rate.

the sector of interest. This is because PCS leads to lower aliasing artifacts within the sector of interest than the RCS as we observed in Fig. 2.9. We can also observe a similar performance difference between our proposed method using PCS or RCS and the benchmarks in the achievable rate from Fig. 2.11(b) for different CS measurements.

From the cumulative distribution function plot in Fig. 2.11(c), we observe that with a probability exceeding 0.98, our proposed method with the optimized weight matrix $\mathbf{W}^S$ results in an achievable rate of at least about 2 bits/s/Hz. The use of random weights for $\mathbf{W}^S$ results in a lower rate, which motivates the need to use almost uniform comb-like beams designed in this work. For the other in-sector CS benchmarks based on [41, 49] and [50], we observe that the achievable rate is lower than that with our approach. This is because these methods radiate a significant amount of the transmitter's energy outside their intended sectors.

We observe from Fig. 2.11(a) that the proposed method with RCS and a random $\mathbf{W}^S$ results in a higher normalized in-sector MSE than the case with the optimized $\mathbf{W}^S$. This is because a random $\mathbf{W}^S$ is very likely to result in a non-uniform illumination pattern within the comb-like structure as shown in Fig. 2.5(b). The column norms of such a CS matrix

are substantially different with such an illumination, which is expected to result in poor reconstruction according to our analysis in Section 2.3.3.

In Fig. 2.12, we compare the proposed in-sector CS method against the benchmarks for different values of $\mathrm{SNR}_{\mathrm{omni}}$. We observe that our method outperforms the benchmarks over a wide range of $\mathrm{SNR}_{\mathrm{omni}}$. Furthermore, it is evident that at high $\mathrm{SNR}_{\mathrm{omni}}$, PCS outperforms RCS due to a significantly lower in-sector normalized MSE in the channel estimate. While at low SNR levels, additive white Gaussian noise remains the dominant factor, at high SNR levels, the aliasing artifacts arising from subsampling in CS dominate. As the aliasing artifacts with PCS are substantially lower than RCS due to the optimized circular shifts, the disparity in performance between PCS and RCS is discernible at high SNR.

2.5. Addressing implementation challenges with our in-sector beam acquisition technique

A key assumption in this work is that the received SNR within the best sector is sufficient for packet detection. This is reasonable due to the sector gain and the use of low modulation and coding schemes (MCS) such as MCS level 0 during initial access. The RX reports the best identified sector through a sector-sweep feedback packet, which also employs a low modulation and coding scheme [42, 88]. We also assume that perfect timing and carrier synchronization are maintained throughout the process of acquiring in-sector channel measurements within a BRP packet. This synchronization can be established using the *short training field* (STF) at the beginning of a BRP packet in the IEEE 802.11ad standard. In this chapter, we ignore the phase jitter and timing drift within a BRP packet. Extending our work to account for such jitters and drifts is beyond the scope of this chapter.

We discuss how our technique explained for a single user can be extended to a multi-user setting. As we assume a single RF chain in this chapter, the direct application of our sector sweep method to a multi-user setting results in a substantial overhead. This is because the BRP procedure must be performed for different sectors, considering that the best sector varies across the users. This procedure does not scale well with the number of users and can be time-consuming, given that the maximum duration of a BRP packet is 46.96 μs [91, Sec. 20.11.3]. To address this problem, a hybrid beamforming architecture may be used at the TX. In this case, the IEEE 802.11ay standard can be adopted such that different base AWMs can be simultaneously applied across multiple RF chains [42].

Finally, we briefly discuss how our framework can be extended for two types of hybrid beamforming architectures, i.e., (i) fully connected architecture where each RF chain is connected to all the antennas and (ii) subarray-based architecture where each RF chain is connected to a distinct subset of antennas. With fully connected architectures, each RF chain can be used to illuminate a certain sector and the RF chains collectively can illuminate the entire beamspace. This can be achieved by transmitting orthogonal time-domain codes from different RF chains. For the subarray-based architecture, our compressive in-sector channel estimation framework can be applied to each subarray independently. When non-stationarity effects can be ignored, the associated beamspace

submatrices of the subarrays share a common support. In such case, sparse recovery algorithms based on dynamic compressed sensing [112] can be used to exploit the common support structure across beamspace submatrices.

2.6. Conclusions

In this chapter, we proposed a new comb structured codebook for sector level sweep with phased arrays. We also developed a CCS-based method for channel estimation or beam acquisition within the sector of interest. Our approach requires a lower training overhead than the classical exhaustive beam scanning methods and integrates well into the beam refinement protocol of the IEEE 802.11ad/ay standards. Furthermore, it overcomes the poor received SNR issue with CS techniques employing wide beams, by concentrating the transmitter's energy in the sector of interest. We also optimized the CS matrix by designing the circular shifts in CCS to reduce the in-sector channel reconstruction error. Finally, we provided conditions that guarantee the successful recovery of the support of beamspace within the sector of interest. Our results indicate that the reconstruction error in the estimated in-sector channel is low when the antenna weights applied at the transmitter illuminate the sector almost uniformly.

2.7. Appendix

2.7.1. Proof of Lemma 1

Let $\mathbf{1}_{\rho_e \times \rho_a}$ denote a $\rho_e \times \rho_a$ all-ones matrix. We use $\Gamma(\omega_1, \omega_2)$ to denote the 2D discrete-space Fourier transform of $\mathbf{1}_{\rho_e \times \rho_a}$ where $\{\omega_1,\ \omega_2\} \in [0, 2\pi]$ [93] are continuous. The inverse 2D-DFT of $\mathbf{1}_{\rho_e \times \rho_a}$ is $\mathbf{U}^*_{\rho_e} \mathbf{1}_{\rho_e \times \rho_a} \mathbf{U}^*_{\rho_a} = \sqrt{\rho_e \rho_a}\mathbf{E}_{\rho_e \times \rho_a}$ where the $\rho_e \times \rho_a$ matrix $\mathbf{E}_{\rho_e \times \rho_a}$ is zero at all entries except $E_{\rho_e \times \rho_a}(0,0) = 1$. Hence, we have [93, Sec. 5.5]

$$\frac{1}{\sqrt{\rho_e \rho_a}} \Gamma^*\left(\frac{2\pi i}{\rho_e}, \frac{2\pi j}{\rho_a}\right) = \sqrt{\rho_e \rho_a} E_{\rho_e \times \rho_a}(i,j), \forall (i,j) \in [\rho_e] \times [\rho_a]. \tag{2.40}$$

Since $E_{\rho_e \times \rho_a}(i,j) = 0 \ \ \forall (i,j) \in [\rho_e] \times [\rho_a] \backslash (0,0)$, from (2.40), we have

$$\Gamma\left(\frac{2\pi i}{\rho_e}, \frac{2\pi j}{\rho_a}\right) = 0, \quad \forall (i,j) \in [\rho_e] \times [\rho_a] \backslash (0,0). \tag{2.41}$$

Next, we observe that $\mathbf{N}_\Omega^{\text{Nyq}}$ defined in Lemma 1 is an extended version of $\mathbf{1}_{\rho_e \times \rho_a}$ obtained by appending $\rho_e(N_e - 1)$ zeros at the end of each row and then appending $\rho_a(N_a - 1)$ zeros at the end of each column. Now, we use the fact that the 2D discrete-space Fourier transform $\Gamma(\omega_1, \omega_2)$ of the zero-padded signal remains the same, however, its 2D-DFT will be a densely sampled version of $\Gamma(\omega_1, \omega_2)$ [93, Ch. 5]. Hence, the PSF (2.29) of $\mathbf{N}_\Omega^{\text{Nyq}}$ which is a zero-padded version of $\mathbf{1}_{\rho_e \times \rho_a}$ can be written as

$$\text{PSF}(i,j) = \frac{1}{M} \Gamma^*\left(\frac{2\pi i}{N}, \frac{2\pi j}{N}\right), \ \ (i,j) \in [N] \times [N]. \tag{2.42}$$

Finally, from $N = \rho_e N_e$, $N = \rho_a N_a$ and (2.41), we observe that $\mathrm{PSF}(i,j) = 0$ at the desired locations defined by $\mathscr{T}$ in (2.30). Hence, for $\mathbf{N}_{\Omega}^{\mathrm{Nyq}}$, the coherence defined in (2.30) is equal to zero. Therefore, the sampling set Ω given in the lemma is optimal which concludes the proof.

2.7.2. Proof of Theorem 1

To find a guarantee for successful support recovery, we analyze the support detection step in the OMP algorithm [8]. An extended version of this proof is available in our conference paper [95]. We, however, include an outline of this proof here for completeness.

Let $(\mathbf{a}_{\mathscr{L}_o})_j$ denote the j^{th} column of the effective CS matrix $\mathbf{A}_{\mathscr{L}_o}$ in the in-sector CS measurement model (2.28). The noise term $(\mathbf{a}_{\mathscr{L}_o})_j^* \mathbf{v}/d_j$ in the support detection step of the OMP algorithm can lead to incorrect support detection. Therefore, to control support misdetection under noise, we first examine the event

$$E = \left\{ \max_{j \in [\rho_e \rho_a]} \frac{|(\mathbf{a}_{\mathscr{L}_o})_j^* \mathbf{v}|}{d_j} < \gamma \right\}, \tag{2.43}$$

where $\gamma := \sigma\sqrt{2(1+\alpha)\log(N^2/S)}$ and $\alpha > 0$ [101]. In Lemma 2, we provide an upper bound for the probability of occurrence of event E.

Lemma 2. *The probability that event E in (2.43) occurs is lower bounded by*

$$\Pr\{E\} \geq \left(1 - \sqrt{\frac{2}{\pi}} . \sqrt{\frac{\sigma}{\gamma}} \exp\left(-\frac{\gamma^2}{2\sigma^2}\right)\right)^{2N^2/S}. \tag{2.44}$$

Proof. We split $(\mathbf{a}_{\mathscr{L}_o})_j^* \mathbf{v}$ into its real part $\left(\left(\mathbf{a}_{\mathscr{L}_o}\right)_j^* \mathbf{v}\right)$ and its imaginary part $\left(\left(\mathbf{a}_{\mathscr{L}_o}\right)_j^* \mathbf{v}\right)$. Since the entries of $\mathbf{v}$ are independently distributed as $\mathscr{CN}(0, \sigma^2)$, the random variables $\left\{\left(\left(\mathbf{a}_{\mathscr{L}_o}\right)_j^* \mathbf{v}\right)/d_j\right\}$ and $\left\{\left(\left(\mathbf{a}_{\mathscr{L}_o}\right)_j^* \mathbf{v}\right)/d_j\right\} \forall j \in [N^2/S]$ are jointly Gaussian and each distributed as $\mathcal{N}(0, \sigma^2/2)$. For any $\kappa \in \mathbb{C}$, we notice that $|\kappa| < \gamma$ whenever $\Re(\kappa) < \gamma/\sqrt{2}$ and $\Im(\kappa) < \gamma/\sqrt{2}$. Using this observation in (2.43), we can write

$$\begin{aligned}
\Pr\{E\} \geq \Pr\Bigg\{ & \max_{j \in [N^2/S]} \frac{\left|\Re\left(\left(\mathbf{a}_{\mathscr{L}_o}\right)_j^* \mathbf{v}\right)\right|}{d_j} < \frac{\gamma}{\sqrt{2}} \\
& \max_{j \in [N^2/S]} \frac{\left|\Im\left(\left(\mathbf{a}_{\mathscr{L}_o}\right)_j^* \mathbf{v}\right)\right|}{d_j} < \frac{\gamma}{\sqrt{2}} \Bigg\} \\
\overset{(a)}{\geq} & \prod_{j \in [N^2/S]} \left[\Pr\left\{ \frac{\left|\Re\left(\left(\mathbf{a}_{\mathscr{L}_o}\right)_j^* \mathbf{v}\right)\right|}{d_j} < \frac{\gamma}{\sqrt{2}} \right\} \right.
\end{aligned}$$

$$\Pr\left\{\frac{\left|\left((\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{v}\right)\right|}{d_j} < \frac{\gamma}{\sqrt{2}}\right\}\Bigg].$$

In (a), we use Šidák's lemma [113, Theorem 1]. Now, replacing the probability values in the right-hand of the above equation with their upper limits similar to [101], we obtain (2.44). □

Our proof follows a similar structure as the one in [101, Lemma 3]. We begin with the first iteration of the OMP algorithm and analyze the support recovery step. Then, by induction, we show that when the event E occurs and $\mathbf{x}_{\mathscr{L}_o}$ satisfies (2.34), the OMP correctly recovers the support of $\mathbf{x}_{\mathscr{L}_o}$ after K iterations where K is the sparsity level of $\mathbf{x}_{\mathscr{L}_o}$.

Now, considering the first iteration of the OMP algorithm [8] and assuming that the event E occurs, we find conditions for which the index of the selected column, i.e., the one that maximizes $\left|(\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{y}\right|/d_j$, belongs to the support Π of $\mathbf{x}_{\mathscr{L}_o}$. The correct support recovery condition in the first iteration of the OMP algorithm [8] is

$$\max_{j\in\Pi}\frac{\left|(\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{y}\right|}{d_j} > \max_{j\notin\Pi}\frac{\left|(\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{y}\right|}{d_j}. \tag{2.45}$$

For the right-hand side of (2.45), under the event E, we have

$$\begin{aligned}\max_{j\notin\Pi}\frac{|(\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{y}|}{d_j} &= \max_{j\notin\Pi}\frac{|(\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{v}+\sum_{i\in\Pi}x_{\mathscr{L}_o,i}(\mathbf{a}_{\mathscr{L}_o})_j^*(\mathbf{a}_{\mathscr{L}_o})_i|}{d_j}\\ &\overset{(b)}{\le}\max_{j\notin\Pi}\frac{|(\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{v}|}{d_j}+\max_{j\notin\Pi}\frac{\sum_{i\in\Pi}|x_{\mathscr{L}_o,i}(\mathbf{a}_{\mathscr{L}_o})_j^*(\mathbf{a}_{\mathscr{L}_o})_i|}{d_j}\\ &\overset{(c)}{<}\gamma+K\mu d_{\max}x_{\mathscr{L}_o,\max}.\end{aligned} \tag{2.46}$$

We use the triangular inequality in (b). In (c), we use the assumption that the event E occurs and the fact that

$$\frac{|\bar{x}_{\mathscr{L}_s,i}(\mathbf{a}_{\mathscr{L}_o})_j^*(\mathbf{a}_{\mathscr{L}_o})_i|}{d_j}=|x_{\mathscr{L}_o,i}|\frac{|(\mathbf{a}_{\mathscr{L}_o})_j^*(\mathbf{a}_{\mathscr{L}_o})_i|d_i}{d_jd_i}\overset{(d)}{\le}d_{\max}x_{\mathscr{L}_o,\max}\mu.$$

In (d), we use $\mu_o=\max_{j\neq i}\frac{|(\mathbf{a}_{\mathscr{L}_o})_j^*(\mathbf{a}_{\mathscr{L}_o})_i|}{d_jd_i}$ which is an alternative definition for the coherence of the effective CS matrix $\mathbf{A}_{\mathscr{L}_o}$ given in (2.30). Under the event E, for the left-hand side of (2.45), we have

$$\max_{j\in\Pi}\frac{|(\mathbf{a}_{\mathscr{L}_o})_j^*\mathbf{y}|}{d_j}=$$

$$\begin{aligned}
&\max_{j\in\Pi} \frac{|d_j^2 x_{\mathscr{L}_o,j} + (\mathbf{a}_{\mathscr{L}_o})_j^* \mathbf{v} + \sum_{i\in\Pi\setminus\{j\}} x_{\mathscr{L}_o,i} (\mathbf{a}_{\mathscr{L}_o})_j^* (\mathbf{a}_{\mathscr{L}_o})_i|}{d_j} \\
&\geq d_{\min} x_{\mathscr{L}_o,\max} \\
&\quad - \max_{j\in\Pi} \frac{\left|(\mathbf{a}_{\mathscr{L}_o})_j^* \mathbf{v} + \sum_{i\in\Pi\setminus\{j\}} x_{\mathscr{L}_o,i} (\mathbf{a}_{\mathscr{L}_o})_j^* (\mathbf{a}_{\mathscr{L}_o})_i\right|}{d_j} \\
&> d_{\min} x_{\mathscr{L}_o,\max} - \gamma - (K-1)\mu_o d_{\max} x_{\mathscr{L}_o,\max}. \qquad (2.47)
\end{aligned}$$

Note that in the first inequality, we use $\max_{j\in\Pi} |d_j x_{\mathscr{L}_o,j}| \geq d_{\min} x_{\mathscr{L}_o,\max}$. Hence, from (2.46) and (2.47) we can write

$$\begin{aligned}
\max_{j\in\Pi} \frac{|(\mathbf{a}_{\mathscr{L}_o})_j^* \mathbf{y}|}{d_j} &> d_{\min} x_{\mathscr{L}_o,\max} - (2K-1)\mu_o d_{\max} x_{\mathscr{L}_o,\max} \\
&\quad - 2\gamma + \max_{j\notin\Pi} \frac{|(\mathbf{a}_{\mathscr{L}_o})_j^* \mathbf{y}|}{d_j}. \qquad (2.48)
\end{aligned}$$

From (2.48), we observe that under the event E, when

$$d_{\min} x_{\mathscr{L}_o,\max} - (2K-1)\mu_o d_{\max} x_{\mathscr{L}_o,\max} \geq 2\gamma, \qquad (2.49)$$

equation (2.45) holds and therefore the selected entry in the first iteration of the OMP algorithm [8] will belong to the support of $\mathbf{x}_{\mathscr{L}_o}$. We note that (2.34) implies (2.49). Next, by using an induction-based technique as in [101, Theorem. 4], we can show that under (2.34) and the event E, the OMP algorithm will successfully recover the entire support Π of $\mathbf{x}_{\mathscr{L}_o}$ after K iterations. Finally, by following our approach in Theorem 2 of [95], we can derive the upper-bound (2.36) for the MSE in $\hat{\mathbf{x}}_{\mathscr{L}_o}$.

3

Message passing-based sparse channel estimation under partially coherent Wiener phase errors

Compressive sensing (CS) is key to reduce the overhead in estimating sparse high dimensional channels at millimeter wave or terahertz frequencies. The channel measurements in CS are usually perturbed by random phase errors, commonly modeled as a Wiener process, at the oscillators. CS algorithms that ignore such phase errors fail to accurately estimate the channel. In practice, the phase errors are similar within a batch of measurements acquired in a short burst and the errors vary significantly across different batches, resulting in partially coherent measurements. We develop a message passing-based channel estimation algorithm that exploits the sparse structure of the channel together with the Wiener statistics of the phase errors. To this end, we absorb the phase errors into the sparse channel, and introduce three hidden variables to model its support, magnitude, and phase. We derive the message flows between these variables while incorporating Wiener phase noise statistics. Finally, we use alternating optimization to decouple the sparse channel and the phase errors from the vector estimated with our message-passing technique. Using simulations, we show that the proposed algorithm achieves better channel reconstruction than comparable benchmarks.

This chapter is based on my work in [114]: H. Masoumi and N. J. Myers, "Message passing-based sparse channel estimation under partially coherent Wiener phase errors," *IEEE Transactions on Wireless Communications*, September 2025.

3.1. Channel and system model

In this section, we consider a point-to-point narrowband multiple-input single-output (MISO) system and discuss the partially coherent measurement model used for channel estimation. Extending our work to a multi-user scenario is beyond the scope of the dissertation, and we leave it for future work.

3.1.1. Channel model

We consider an $N \times N$ half-wavelength spaced uniform planar array (UPA) at the transmitter (TX) and a single antenna receiver (RX), as shown in Fig. 3.1(a), which is also considered in Chapter 3. The N^2 dimensional narrowband MISO channel between the TX and RX is modeled as an $N \times N$ matrix $\mathbf{H}$. Let L denote the propagation rays in the environment with the ℓ^{th} ray having a complex gain of ζ_ℓ, an azimuth angle-of-departure (AoD) $\vartheta_{a,\ell}$ and an elevation AoD $\vartheta_{e,\ell}$. By defining the beamspace angles as $\omega_{a,\ell} = \pi \sin \vartheta_{e,\ell} \sin \vartheta_{a,\ell}$, $\omega_{e,\ell} = \pi \sin \vartheta_{e,\ell} \cos \vartheta_{a,\ell}$ [87] and the $N \times 1$ Vandermonde vector $\mathbf{a}_N(\omega)$ as

$$\mathbf{a}_N(\omega) = \left[1, e^{\mathrm{j}\omega}, e^{\mathrm{j}2\omega}, \cdots, e^{\mathrm{j}(N-1)\omega}\right]^{\mathsf{T}}, \tag{3.1}$$

the baseband channel matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \sum_{\ell=1}^{L} \zeta_\ell \mathbf{a}_N\left(\omega_{e,\ell}\right) \mathbf{a}_N^{\mathsf{T}}\left(\omega_{a,\ell}\right). \tag{3.2}$$

For typical mmWave or THz access points, the channel dimension N^2 can be in the order of hundreds to thousands.

High scattering at mmWave and THz carrier frequencies results in an approximately sparse channel in the angle domain [1]. By transforming $\mathbf{H}$ into the beamspace (angle domain), we can exploit this sparse structure in channel estimation. Since we assume a UPA at the TX, the 2D-Discrete Fourier Transform (DFT) of $\mathbf{H}$ is used for the beamspace representation. Let $\mathbf{U}_N$ denote the standard $N \times N$ unitary DFT matrix and $\mathbf{X}$ denote the beamspace representation of $\mathbf{H}$. Then, $\mathbf{X}$ and $\mathbf{H}$ are related as

$$\mathbf{H} = \mathbf{U}_N \mathbf{X} \mathbf{U}_N, \tag{3.3}$$

where $\mathbf{X}$ is approximately sparse. The channel $\mathbf{H}$ is unknown and the goal of this chapter is to estimate $\mathbf{H}$ from its phase perturbed measurements acquired at the RX. To acquire one scalar-valued channel measurement $y[m]$, the TX applies a beam training matrix to its array as shown in Fig. 3.1(a).

3.1.2. Frame structure for acquiring measurements

We use Fig. 3.1(b) to explain the frame structure for acquiring measurements to estimate the channel. Our frame structure contains B training batches, and is similar to the structure in IEEE 802.11ad/ay [42, 88]. Each batch consists of a synchronization field and a training field to obtain channel measurements. We define Δ_{sf} as the duration of the synchronization field and Δ_{bf} as the time needed to acquire one channel measurement

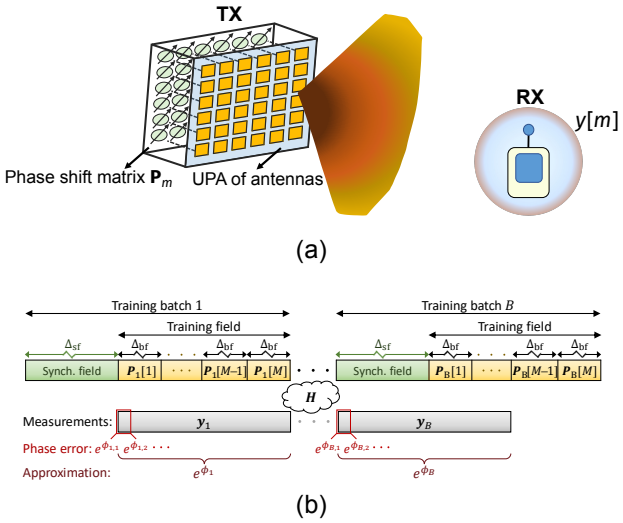


Figure 3.1: A wireless system with a UPA at the TX in (a) and the frame structure for acquiring partially coherent channel measurements in (b). The phase error within each batch is assumed to be constant.

within a batch. In the IEEE 802.11ad standard, Δ_{sf} is several μs while Δ_{bf} is in the order of tens of ns [91, Chapter 20]. We use M to denote the number of distinct beamformers applied in each training batch, equivalently the RX acquires M channel measurements per batch.

We assume a block fading channel that remains constant over B training batches, which is reasonable when the channel coherence time is sufficiently long. For example, in a 60 GHz indoor environment, the coherence time is approximately 1 ms [89], significantly exceeding the training batch duration of about 44 μs with the IEEE 802.11ad/ay standards [56]. We use $\mathbf{P}_b[m] \in \mathbb{C}^{N\times N}$ to denote the beamformer applied at the TX to obtain the m^{th} spatial channel measurement in batch b. This measurement is acquired at the RX and is denoted by $y_b[m]$. The measurements at the RX are perturbed by random phase noise, which is commonly modeled as a Wiener process [58]. Such a process is described using $\mathscr{L}(f_{\text{m}})$, the power spectral density of the phase noise process at an offset f_{m} from the carrier frequency f_{c} [55], i.e., $\mathscr{L}(f_{\text{m}}) = 10\log_{10}\left(f_{\text{c}}^2 c/\left(\pi^2 f_{\text{c}}^4 c^2 + f_{\text{m}}^2\right)\right)$ dBc/Hz. The phase noise parameter c is an oscillator-dependent constant. We assume that this parameter does not change with time [55] as also considered in [56, 57, 63]. An oscillator with a small $\mathscr{L}(f_{\text{m}})$ results in phase errors with low variance.

For a Wiener process, the difference in phase errors is Gaussian with a variance that is directly proportional to the time interval between the sampling instants [55]. As a result, the phase error variations within a batch are smaller when compared to the variation across different batches. To develop our algorithm, we ignore phase error variations within a batch and model only the variations across different batches, while our simulations employ realistic phase errors that also vary within a batch. Under this

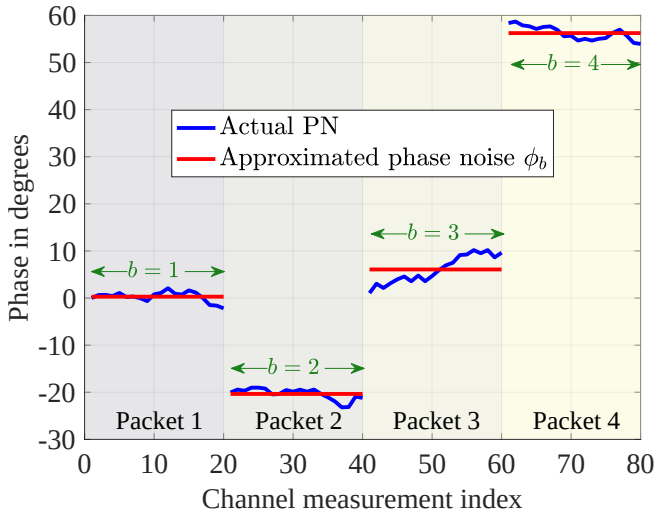


Figure 3.2: A realization of the phase noise process and its approximation for f_c = 60 GHz, B = 4 batches, M = 20 measurements per batch, and phase noise power spectral density $\mathscr{L}(f_m)$ = −101.7 dBc/Hz at f_m = 1 MHz [115]. Here, Δ_{sf} = 44 μs and Δ_{bf} = 128 ns [56].

assumption, we use ϕ_b to model the phase error in all the measurements acquired in batch b. We define $\sigma_{pn}^2 = 4\pi^2 f_c^2 c\Delta$ as the variance of the phase error differences, where Δ is the time duration of a batch. For a Wiener process, $\phi_b|\phi_{b-1} \sim \mathcal{N}\left(\phi_{b-1}, \sigma_{pn}^2\right)$. As a reference, we assume that the initial phase $\phi_1 = 0$. In Fig. 3.2, we show the actual phase noise and the ϕ_bs for a particular realization. The m^{th} channel measurement acquired in batch b is

$$y_b[m] = e^{j\phi_b} \langle \mathbf{H}, \mathbf{P}_b[m] \rangle + w_b[m], \qquad m \in [M], \tag{3.4}$$

where $w_b[m] \sim \mathcal{CN}(0, \sigma^2)$ is additive Gaussian noise. In this chapter, the MB phase-perturbed measurements in (3.4), acquired over B batches, are used to estimate the channel $\mathbf{H}$.

Now, we discuss the partially coherent CS formulation [64] of the spatial measurements in (3.4), to exploit sparsity of the channel in the beamspace. We use $\mathbf{A}_b \in \mathbb{C}^{M \times N^2}$ to denote the CS matrix associated with the M spatial measurements in (3.4) acquired in batch b. The m^{th} row of $\mathbf{A}_b$ is $\text{vec}\left(\mathbf{P}_b^*[m]\right)^{\mathsf{T}} (\mathbf{U}_N \otimes \mathbf{U}_N)$ [45], which depends on the beam training matrix $\mathbf{P}_b[m]$ applied at the TX. We define $\mathbf{y}_b$ as the vector of M measurements $\{y_b[m]\}_{m=1}^M$ acquired in the b^{th} batch, and $\mathbf{w}_b$ as the additive measurement noise vector containing $\{w_b[m]\}_{m=1}^M$. The vector version of the sparse beamspace $\mathbf{X}$ is denoted by the N^2-dimension vector $\mathbf{x}$, i.e., $\mathbf{x} = \text{vec}(\mathbf{X})$. Using $\langle \mathbf{H}, \mathbf{P}_b[m] \rangle = \mathbf{A}_b(m, :)\mathbf{x}$ [45] in (3.4) yields

$$\mathbf{y}_b = e^{j\phi_b} \mathbf{A}_b \mathbf{x} + \mathbf{w}_b. \tag{3.5}$$

The phase errors $\{\phi_b\}_{b=2}^B$ and the sparse vector $\mathbf{x}$ are unknown and have to be estimated

using the acquired measurements $\{\mathbf{y}_b\}_{b=1}^B$ and the known CS matrices $\{\mathbf{A}_b\}_{b=1}^B$. It is also known that $\phi_b|\phi_{b-1} \sim \mathcal{N}(\phi_{b-1}, \sigma_{\text{pn}}^2)$.

3.2. Overview of our message passing-based partially coherent sparse recovery algorithm

As estimating ϕ_bs and $\mathbf{x}$ from the non-linear model in (3.5) is challenging, we first absorb the phase errors into the sparse vector to create B auxiliary vectors, each with the same dimension as $\mathbf{x}$. We then reformulate (3.5) as linear measurements of these auxiliary vectors. We show that these vectors are correlated, i.e., they share the same support and differ only by phase, which follows Wiener statistics. Our message-passing method leverages this correlation, sparsity in the beamspace, and the measurements in (3.5) to estimate $\mathbf{x}$.

3.2.1. Sparse prior and correlated structure of auxiliary vectors

A common sparsity-promoting prior is the Bernoulli-Gaussian (BG) PDF [112, 116], which was used to model sparse channels in [117]. Similar to [117], we model $\mathbf{x}$ as a realization of a random vector x that follows a BG prior. Specifically, x has independent and identically distributed (IID) entries which are zero with probability $1-\lambda$, i.e., $\Pr\{x[n] = 0\} = 1-\lambda$ $\forall n \in [N^2]$, and follow a Gaussian distribution otherwise. We consider a zero-mean complex Gaussian distribution with the same variance μ to write the BG prior as

$$p(x[n]) = (1-\lambda)\delta(x[n]) + \lambda\mathscr{CN}(x[n]; 0, \mu), \tag{3.6}$$

where $\delta(\cdot)$ is the Dirac delta function.

In our method, we absorb the phase error terms $\{e^{\mathrm{j}\phi_b}\}_{b=1}^B$ into the sparse vector $\mathbf{x}$ to construct the auxiliary vectors

$$\mathbf{z}_b = e^{\mathrm{j}\phi_b}\mathbf{x}, \quad \forall b \in [B]. \tag{3.7}$$

As $\mathbf{z}_b$ is just a scalar multiple of $\mathbf{x}$, $\mathbf{z}_b$ is sparse when $\mathbf{x}$ is sparse. Let $\mathcal{S}_{\mathbf{x}}$ denote the support of $\mathbf{x}$, i.e., $\mathcal{S}_{\mathbf{x}} = \{n : x[n] \neq 0\}$, and $\mathcal{S}_{\mathbf{z}_b}$ denote the support of $\mathbf{z}_b$. From (3.7), we observe that $\mathbf{z}_b$ has the same support as $\mathbf{x}$ for each b, i.e., $\mathcal{S}_{\mathbf{z}_1} = \mathcal{S}_{\mathbf{z}_2} = \cdots = \mathcal{S}_{\mathbf{z}_B} = \mathcal{S}_{\mathbf{x}}$. In addition, the entry-wise magnitudes of $\mathbf{x}$ and $\mathbf{z}_b$ are equal, i.e., $|\mathbf{z}_1| = |\mathbf{z}_2| = \cdots = |\mathbf{z}_B| = |\mathbf{x}|$. Furthermore, $\mathbf{z}_b$ and $\mathbf{z}_{b-1}$ differ only by a phase error which follows the Wiener statistics. Using (3.7), the measurements in (3.5) can be rewritten as

$$\mathbf{y}_b = \mathbf{A}_b\mathbf{z}_b + \mathbf{w}_b, \quad \forall b \in [B]. \tag{3.8}$$

Our goal is to develop an algorithm to first estimate $\{\mathbf{z}_b\}_{b=1}^B$ from (3.8) while exploiting the correlation within this collection. Then, we use set of equations in (3.7) to recover $\mathbf{x}$ from $\{\hat{\mathbf{z}}_b\}_{b=1}^B$.

We discuss two methods to reconstruct the auxiliary vectors $\{\mathbf{z}_b\}_{b=1}^B$ from (3.8). The first

method exploits sparsity and the common support across the auxiliary vectors, ignoring the shared magnitude and the Wiener statistics due to phase noise. We refer to this method as *Proposed SuppOnly*, which is solved using the MMV-AMP algorithm [118]. For more details on our this approach, we refer the reader to our conference paper [119]. The emphasis of this chapter is on our second method, which not only exploits sparsity and the shared support in $\{\mathbf{z}_b\}_{b=1}^{B}$ but also leverages the common magnitude structure and the Wiener phase noise statistics in $\{\mathbf{z}_b\}_{b=1}^{B}$.

In our message passing-based method, we model $\mathbf{z}_b$ as a realization of a random vector z_b. The support, magnitude, and the phase of $\mathsf{z}_b[n]$ are modeled using random variables $\mathsf{s}[n] \in \{0,1\}$, $\mathsf{r}[n] \in \mathbb{R}_+$, and $\Theta_b[n] \in \mathbb{R}$ to express $\mathsf{z}_b[n]$ as

$$\mathsf{z}_b[n] = \mathsf{s}[n]\mathsf{r}[n]e^{\mathrm{j}\Theta_b[n]}. \tag{3.9}$$

Equivalently, (3.9) can be expressed as

$$p(z_b[n]|s[n],r[n],\theta_b[n]) = \begin{cases} \delta(z_b[n] - r[n]e^{\mathrm{j}\theta_b[n]}), & \text{if } s[n]=1 \\ \delta(z_b[n]), & \text{if } s[n]=0, \end{cases} \tag{3.10}$$

where $\theta_b[n]$ is a realization of $\Theta_b[n]$. From (3.10), we observe that $z_b[n] = 0$ when $s[n] = 0$, thereby aiding sparse priors. As $\mathsf{z}_b[n]$ has the same support as $\mathsf{x}_b[n]$, it follows from the BG prior on $\mathsf{x}_b[n]$ in (3.6) that [112, 116]

$$p(s[n]) = \lambda^{s[n]}(1-\lambda)^{1-s[n]}, \quad s[n] \in \{0,1\}. \tag{3.11}$$

To determine the prior on $\mathsf{r}[n]$, we first observe that $|\mathsf{z}_b[n]| = |\mathsf{x}[n]|$ from (3.7). Next, it follows from [120, Appendix 5.B] that the magnitude of a random variable distributed as $\mathcal{CN}(x[n];0,\mu)$ has a Rayleigh distribution with parameter $\sqrt{\mu/2}$. Therefore, the PDF of $\mathsf{r}[n]$ is

$$p(r[n]) = \text{Rayleigh}\left(r[n];\sqrt{\mu/2}\right). \tag{3.12}$$

To find the prior on $\Theta_b[n] = \angle\mathsf{z}_b[n]$, we observe from (3.7) that $\angle\mathsf{z}_b[n] = \angle\mathsf{x}[n] + \phi_b$, where ϕ_b models the realization ϕ_b. Since the non-zero entries of x are IID according to $\mathcal{CN}(x[n];0,\mu)$, their phases $\{\angle\mathsf{x}[n]\}_{n=1}^{N^2}$ have IID uniform distribution $\mathcal{U}([-\pi,\pi))$. As we assume $\phi_1 = 0$, we have $\angle\mathsf{z}_1 = \angle\mathsf{x}$ from (3.7). Further, $\{\angle\mathsf{z}_b\}_{b=1}^{B}$ can be modeled as a new Wiener process with $\angle\mathsf{z}_1 \sim \mathcal{U}\left([-\pi,\pi)^{N^2}\right)$, i.e.,

$$\angle\mathsf{z}_b = \angle\mathsf{z}_{b-1} + \mathbf{1}_{N^2}\Phi_b, \tag{3.13}$$

where $\mathbf{1}_{N^2}$ is a vector of ones and $\Phi_b \sim \mathcal{N}\left(0,\sigma_{\text{pn}}^2\right)$. We observe from (3.7) that the same phase error is introduced in all entries of $\mathbf{z}_b$. Therefore, the phase noise innovation $\mathcal{N}\left(0,\sigma_{\text{pn}}^2\right)$ is same for all the elements of z_b in (3.13).

Incorporating a common phase noise innovation across all the elements of z_b is challenging as it leads to numerous loops in the corresponding factor graph. To aid a tractable message passing method, our model assumes that the phase error innovations

within z_b are independent, i.e.,

$$\Theta_b[n] = \angle \mathsf{z}_b[n] = \angle \mathsf{z}_{b-1}[n] + \Phi_b[n], \tag{3.14}$$

where $\Phi_b[n]$ are IID as $\mathcal{N}\left(0, \sigma_{\text{pn}}^2\right)$. Therefore, (3.14) is a relaxation of the true model in (3.13). Under this approximation,

$$p\left(\theta_b[n] \middle| \theta_{b-1}[n]\right) = \mathcal{N}\left(\theta_b[n]; \theta_{b-1}[n], \sigma_{\text{pn}}^2\right), \tag{3.15}$$

where $\theta_1[n] \sim \mathcal{U}([-\pi, \pi))$. Note that our message passing-based method to estimate the auxiliary vectors does not exploit the property that the true $\theta_b[n]$ is invariant with n. Our alternating optimization technique in Sec. 3.3, however, leverages this property to recover $\mathbf{x}$ from the estimates $\{\hat{\mathbf{z}}_b\}_{b=1}^B$.

As the supports of $\{\mathsf{z}_b\}_{b=1}^B$ are modeled by a single random vector s and their magnitudes are modeled by a single random vector r, our method inherently incorporates the shared support and magnitude structure. The Wiener phase noise statistics is incorporated using (3.13), with the limitation that the phase errors within each batch are assumed to be IID.

3.2.2. Factor graph and message scheduling

We use the posterior joint PDF of $\{\mathsf{z}_b\}_{b=1}^B$, s, r and $\{\Theta_b\}_{b=1}^B$ to develop our algorithm for estimating $\{\mathbf{z}_b\}_{b=1}^B$. By using the Bayes' rule and the dependencies between $\{\mathsf{z}_b\}_{b=1}^B$, s, r and $\{\Theta_b\}_{b=1}^B$, we factorize the joint PDF as

$$\begin{aligned} &p\left(\{\mathbf{z}_b\}_{b=1}^B, \mathbf{s}, \mathbf{r}, \{\boldsymbol{\theta}_b\}_{b=1}^B \middle| \{\mathbf{y}_b\}_{b=1}^B\right) \\ &\propto \prod_{b=1}^{B} \left(\prod_{m=1}^{M} p\left(y_b[m] \middle| \mathbf{z}_b\right) \prod_{n=1}^{N} p\left(z_b[n] \middle| s[n], r[n], \theta_b[n]\right) \right. \\ &\left. p\left(\theta_b[n] \middle| \theta_{b-1}[n]\right) \right) \prod_{n=1}^{N} p\left(s[n]\right) \prod_{n=1}^{N} p\left(r[n]\right), \end{aligned} \tag{3.16}$$

where $\propto$ denotes equality up to a constant scale factor, and $p\left(\theta_1[n] \middle| \theta_0[n]\right) = p\left(\theta_1[n]\right)$. See Table 3.1 for the factors in (3.16).

The dependencies among different variables in (3.16) is represented by a *factor graph* shown in Fig. 3.3. The factor graph contains two types of nodes: *factor nodes* shown as rectangles and *variable nodes* shown as circles. Each factor node represents a factor in (3.16) and is connected to the variable nodes corresponding to its arguments. The explicit form of all the factors in our graph are listed in Table 3.1. Our factor graph, shown in Fig. 3.3 comprises B planes, where each plane corresponds to a measurement batch in (3.8). The connections between different planes indicate the correlations across $\{\mathbf{z}_b\}_{b=1}^B$, which follow from (3.9) and (3.14). For example, the n^{th} factor nodes $\{f_b[n]\}_{b=1}^B$ in all planes are connected through $r[n]$ and $s[n]$ in Fig. 3.3(b). This is because $\{\mathsf{z}_b[n]\}_{b=1}^B$

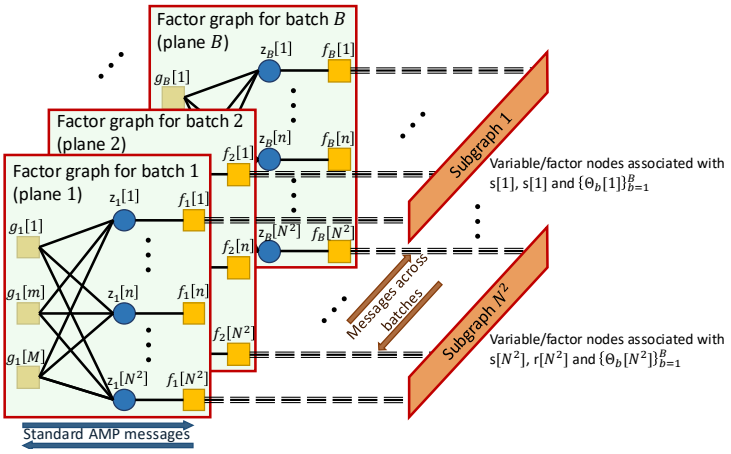
Table 3.1: Factors of the joint PDF in (3.16).

Factor definition	Explicit form
$g_b[m] = p\left(y_b[m]\middle\vert \mathbf{z}_b\right)$	$\mathscr{CN}\left(y_b[m]; \mathbf{a}_b^{\mathsf{T}}[m]\mathbf{z}_b, \sigma^2\right)$
$f_b[n] = p\left(z_b[n]\middle\vert s[n], r[n], \theta_b[n]\right)$	$\delta\left(z_b[n] - s[n]r[n]e^{j\theta_b[n]}\right)$
$q_b[n] = p\left(\theta_b[n]\middle\vert \theta_{b-1}[n]\right)$	$\mathcal{N}\left(\theta_b[n]; \theta_{b-1}[n], \sigma_{\text{pn}}^2\right),\ \theta_1[n] \sim \mathcal{U}([-\pi, \pi))$
$h[n] = p(s[n])$	$\lambda^{s[n]}(1-\lambda)^{1-s[n]},\ s[n] \in \{0, 1\}$
$d[n] = p(r[n])$	$\text{Rayleigh}\left(r[n]; \sqrt{\mu/2}\right) = 2\frac{r[n]}{\mu}e^{-r^2[n]/\mu}$

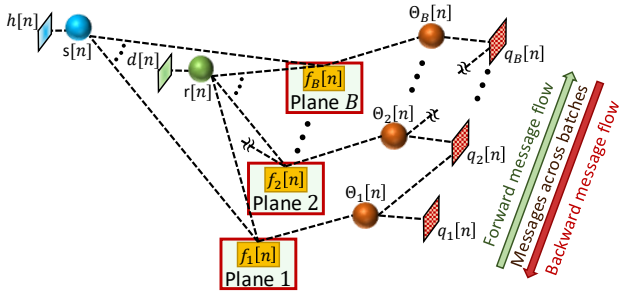
have the same magnitude, i.e., $r[n]$, and are all jointly non-zero when $s[n] = 1$. The evolution of the phase noise process across different batches is modeled by inter-plane connections through variable nodes $\{\Theta_b[n]\}_{b=1}^{B}$ using the factors $\{q_b[n]\}_{b=1}^{B}$. The $q_b[n]$s, shown in Fig. 3.3(b), incorporate Wiener phase noise statistics from (3.15).

Our method aims to find the minimum mean squared error (MMSE) estimate of $\{\mathbf{z}_b\}_{b=1}^{B}$ given the measurements $\{\mathbf{y}_b\}_{b=1}^{B}$, using the factor graph in Fig. 3.3. To this end, the marginal posteriors $p\left(z_b[n]\middle\vert \{\mathbf{y}_b\}_{b=1}^{B}\right)$ are first estimated using the sum-product algorithm [121]. The means of these posteriors then yield the MMSE estimates. In each iteration of the sum-product algorithm, nodes in the factor graph exchange messages, also known as beliefs, which represent probability densities. There are two types of message flows: those sent from variables to neighboring factors, and those sent from factors to neighboring variables. A belief exchanged from a factor to a variable represents the probability distribution of the variable, as perceived from the factor's perspective. In contrast, the belief sent from a variable to a factor represents the variable's perception of its own probability distribution, informed by all the factor nodes it is connected to—except for the one it is messaging. In the sum-product algorithm [121], a message from a variable to a factor is obtained by multiplying all incoming messages from the other neighboring factors. A message from a factor to a variable is calculated by summing (or integrating) the product of all incoming messages over all variables, except for the one to which the message is sent.

Our algorithm propagates messages on a plane-by-plane basis to exploit the Wiener structure in the phase errors. In the first plane ($b = 1$) of Fig. 3.3, the linear regression problem in (3.8) is solved with a sparse prior on z_b. This "intra-plane" step is solved using the standard AMP [9], which approximates the messages from a factor node $g_b[m]$ to a variable node $\mathsf{z}_b[n]$ as Gaussian PDF. This approximation is possible because the factor node $g_b[m]$ depends on $z_b[n]$ through the linear mix $\mathbf{a}_b^{\mathsf{T}}[m]\mathbf{z}_b$ where the elements of $\mathbf{a}_b^{\mathsf{T}}[m]$ have approximately similar norm [9]. After the intra-plane step converges, the marginal posterior of $\mathsf{z}_1[n]$ is estimated for each n. This posterior provides side-information about the support and the magnitude of $\mathsf{z}_2[n]$, a variable to be solved in the next plane ($b = 2$). The Wiener phase noise statistic, combined with the posterior, provides side-information on the phase of $\mathsf{z}_2[n]$. The "out-of-plane" step in our algorithm derives side-information from a plane, while the "inter-plane" step propagates this information to the next plane. In Fig. 3.4, the out-of-plane step shows $f_1[n]$ sending beliefs



(a) Factor graph.



(b) Subgraph n.

Figure 3.3: The factor graph corresponding to the factorization of the posterior joint PDF in (3.16). Figure (b) shows the details of subgraph n in the factor graph. In this chapter, the MMSE estimate of $\{\mathbf{z}_b\}_{b=1}^B$ given the measurements $\{\mathbf{y}_b\}_{b=1}^B$ is computed, which is subsequently used to estimate $\mathbf{x}$. Our factor graph incorporates sparsity, shared support and magnitude structure, and Wiener phase noise statistics. The explicit form of the factors in our graph are listed in Table 3.1.

about the support s[n], magnitude r[n], and phase $\Theta_1[n]$, informed by the message from $z_1[n]$. The support and magnitude messages are directly propagated to the next plane, while the Wiener statistics is used to compute the belief about $\Theta_2[n]$. Finally, the "into-plane" step propagates the support, magnitude, and phase side-information to the next plane. This information is used as a prior to solve for the posterior of z_2 in (3.8) using the AMP. The procedure is continued until an estimate of the marginal posterior PDF for $z_B[n]$ in the last plane, i.e., plane B, is obtained. Because the messages are passed in the direction of the increasing plane index b, we refer to this process as the *forward* message flow. The forward flow can sequentially update the channel estimate with each new measurement batch.

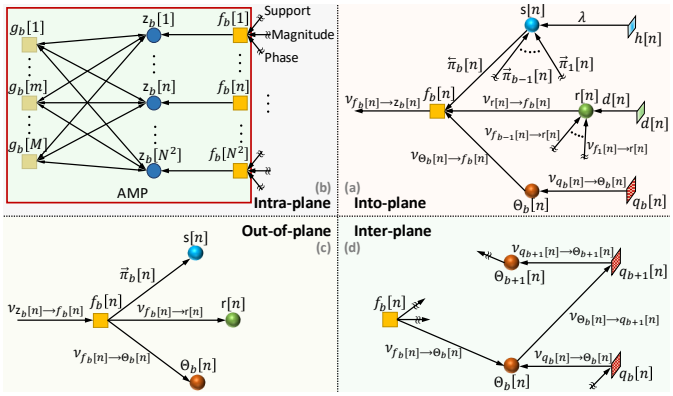


Figure 3.4: Four sections of the factor graph illustrating the four key steps of the serial message schedule. It shows only the messages involved for nodes indexed by $n \in [N^2]$ and the nodes within plane $b \in [B]$.

When the measurements from all the B batches are available for channel estimation, *backward* message flows can also be performed in addition to the forward message flows [112]. The backward flow propagates side-information from plane $b+1$ to plane b, i.e., measurements in the $b+1^{\text{th}}$ batch are used to derive priors needed to solve for $z_b[n]$. After multiple forward-backward message flows, the marginal posterior $p(z_b[n]\big|\{\mathbf{y}_b\}_{b=1}^B)$ is estimated by exploiting measurements from all the batches. The mean of these posteriors gives the auxiliary vector estimates $\{\hat{\mathbf{z}}_b\}_{b=1}^B$. To estimate $\mathbf{x}$ from these vectors, we stack $\{\hat{\mathbf{z}}_b\}_{b=1}^B$ in an $N^2 \times B$ matrix $\hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1, ..., \hat{\mathbf{z}}_B]$. We define $\boldsymbol{\phi} = [\phi_1, ..., \phi_B]^{\mathsf{T}}$, $e^{\mathrm{j}\boldsymbol{\phi}} = \left[e^{\mathrm{j}\phi_1}, ..., e^{\mathrm{j}\phi_B}\right]^{\mathsf{T}}$ and $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_B]$. From (3.7), we observe that $\mathbf{Z}$ is a rank one matrix as $\mathbf{Z} = \mathbf{x}e^{\mathrm{j}\boldsymbol{\phi}^{\mathsf{T}}}$. Hence, we find a rank-one matrix of the form $\mathbf{x}e^{\mathrm{j}\boldsymbol{\phi}^{\mathsf{T}}}$ that is closest to $\hat{\mathbf{Z}}$ in terms of the Frobenius norm distance, i.e.,

$$\{\hat{\mathbf{x}}, \hat{\boldsymbol{\phi}}\} = \underset{\{\mathbf{x},\boldsymbol{\phi}\}}{\operatorname{argmin}} \ \|\hat{\mathbf{Z}} - \mathbf{x}e^{\mathrm{j}\boldsymbol{\phi}^{\mathsf{T}}}\|_{\mathrm{F}}, \tag{3.17}$$

where $\hat{\mathbf{x}}$ is our estimate for the sparse channel $\mathbf{x}$. To solve (3.17), we develop an alternating approach outlined in Algorithm 1.

Algorithm 1: Alternating algorithm to solve (3.17).

Input: $\{\hat{\mathbf{z}}_b\}_{b=1}^B$, Maximum number of iterations $T_{\max}$.
Initialization: $i = 1$, $\hat{\mathbf{x}}^{(1)} = \hat{\mathbf{z}}_1$.
For $i = 1, \cdots, T_{\max}$ **do**:
1. $\hat{\boldsymbol{\phi}}^{(i)} = -\angle\hat{\mathbf{Z}}^*\hat{\mathbf{x}}^{(i)}$. Note: $\angle\mathbf{c}$ returns the angle of $\mathbf{c}$
2. $\hat{\mathbf{x}}^{(i+1)} = \frac{1}{B}\hat{\mathbf{Z}}e^{-\mathrm{j}\hat{\boldsymbol{\phi}}^{(i)}}$.

End For
Output: $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(T_{\max})}$.

3.3. Messages in the proposed algorithm

In this section, we derive the messages exchanged between the nodes within the factor graph in Fig. 3.3. While we primarily focus on forward message flows, the backward flows can be derived through a similar procedure.

We use $v_{i\to j}$ to denote the message sent from node i to node j. Examples of these directional message flows are shown in Fig. 3.4. Note that the message exchanged between the factor $f_b[n]$ and the support variable $\mathrm{s}[n]$ is a Bernoulli probability mass function (PMF), e.g., $v_{f_b[n]\to \mathrm{s}[n]} = (\vec{\pi}_b[n], 1-\vec{\pi}_b[n])$ where $\vec{\pi}_b[n]$ is the belief that $\mathrm{s}[n]$ is 1. For ease of notation, we denote this Bernoulli PMF as $v_{f_b[n]\to \mathrm{s}[n]} = \vec{\pi}_b[n]$. Similarly, $v_{\mathrm{s}[n]\to f_b[n]} = \overleftarrow{\pi}_b[n]$. We use $\lambda = v_{h[n]\to \mathrm{s}[n]}$ to denote the Bernoulli PMF sent from the factor $h[n]$ to the variable $\mathrm{s}[n]$. Finally, as $d[n]$ itself is the belief about the magnitude $\mathrm{r}[n]$, we write $d[n] = v_{d[n]\to \mathrm{r}[n]}$.

Messages exchanged with the AMP subgraph: The forward pass in our algorithm begins with the into-plane step for $b = 1$. As no side-information is available for $b = 1$, $\vec{\pi}_1[n] = h[n]$, $v_{\mathrm{r}[n]\to f_1[n]} = d[n]$, and $v_{\Theta_1[n]\to f_1[n]} = q_1[n]$. These distributions, listed in Table 3.1, are based on the parameters λ and μ, which can either be estimated from a dataset or learned using the expectation maximization algorithm [76]. Since, $h[n]$ is a Bernoulli PMF, $d[n]$ is a Rayleigh PDF and $q_1[n]$ is a uniform PDF over $[-\pi, \pi)$, it follows from (3.9) that into-plane message $v_{f_1[n]\to \mathrm{z}_1[n]}$ is the BG prior

$$v_{f_1[n]\to \mathrm{z}_1[n]} = (1-\lambda)\delta(z_1[n]) + \lambda \mathscr{CN}(z_1[n]; 0, \mu), \tag{3.18}$$

for each n. These into-plane messages provide beliefs on $\{\mathrm{z}_1[n]\}_{n=1}^{N^2}$, which are used in the intra-plane step for $b = 1$.

The intra-plane step in plane b performs standard AMP using the into-plane beliefs and the measurement-based likelihoods, i.e., the $g_b[m]$s. As the CS matrix $\mathbf{A}_b$ is usually dense, the factor graph in plane b has several cycles, which require computing multi-dimensional integrals with the sum-product algorithm. To circumvent this high-complexity integration, AMP approximates $v_{g_b[m]\to \mathrm{z}_b[n]}$ as a Gaussian, by using the central limit theorem and the Taylor series expansion [9]. For the exact messages exchanged between $\{g_b[m]\}_{m=1}^{M}$ and $\mathrm{z}_b[n]$, we refer the reader to [9]. To apply the AMP, the message $v_{f_b[n]\to \mathrm{z}_b[n]}$ is treated as a prior on $\mathrm{z}_b[n]$ for each n. For instance, these messages are the BG prior in (3.18) for $b = 1$. After AMP iterations, the local posterior of $\mathrm{z}_b[n]$ based only on batch b measurements can be obtained. We use $e_b[n]$ to denote the mean and $c_b[n]$ to denote the variance of this posterior. At the end of intra-plane AMP iterations, the messages $\{v_{\mathrm{z}_b[n]\to f_b[n]}\}_{n=1}^{N^2}$ given by

$$v_{\mathrm{z}_b[n]\to f_b[n]} = \mathscr{CN}(z_b[n]; e_b[n], c_b[n]), \tag{3.19}$$

are used to perform the out-of-plane step. Towards the end of this section, we derive the messages $v_{f_b[n]\to \mathrm{z}_b[n]}$ for $b \geq 2$. These messages are deferred to a later point, as they depend on messages involving the phase variables.

Messages with the magnitude variables: In the out-of-plane step, the beliefs on $\mathrm{s}[n]$s, $\mathrm{r}[n]$s, and $\Theta_b[n]$s are updated using the posteriors in (3.19) from plane b. By the sum-product algorithm, $v_{f_b[n]\to \mathrm{r}[n]}$ is derived by first computing the product of $v_{\mathrm{s}[n]\to f_b[n]}$,

the posterior from (3.19) and $v_{\Theta_b[n]\to f_b[n]}$. This product is then integrated over $s[n]$ and $\Theta_b[n]$ to write

$$\begin{aligned} v_{f_b[n]\to r[n]} &= \left(1-\tilde{\pi}_b[n]\right)\mathscr{CN}(0; e_b[n], c_b[n]) \\ &\quad + \tilde{\pi}_b[n]\int_{\theta[n]} v_{\Theta_b[n]\to f_b[n]}\mathscr{CN}\left(r[n]e^{j\theta_b[n]}; e_b[n], c_b[n]\right). \end{aligned} \tag{3.20}$$

The first term in (3.20), corresponding to $s[n] = 0$, is constant with respect to the variable $r[n]$ and this constant makes $v_{f_b[n]\to r[n]}$ an improper distribution which does not integrate to 1 [118]. To address this issue, we use the idea from [118] that considers $s[n] \in \{\epsilon, 1\}$ in the limit $\epsilon \to 0$, and define

$$\Omega\left(\tilde{\pi}_b[n]\right) = \frac{\epsilon^2\tilde{\pi}_b[n]}{\left(1-\tilde{\pi}_b[n]\right) + \epsilon^2\tilde{\pi}_b[n]}. \tag{3.21}$$

For a small ϵ, the message $v_{f_b[n]\to r[n]}$ in (3.20) is modified to

$$\begin{aligned} &\bar{v}_{f_b[n]\to r[n]} = \\ &\left(1-\Omega\left(\tilde{\pi}_b[n]\right)\right)\int_{\theta_b[n]} v_{\Theta_b[n]\to f_b[n]}\mathscr{CN}\left(r[n]e^{j\theta_b[n]}; \frac{1}{\epsilon}e_b[n], \frac{1}{\epsilon^2}c_b[n]\right) \\ &+ \Omega\left(\tilde{\pi}_b[n]\right)\int_{\Theta_b[n]} v_{\Theta_b[n]\to f_b[n]}\mathscr{CN}(r[n]e^{j\theta_b[n]}; e_b[n], c_b[n]), \end{aligned} \tag{3.22}$$

which can be shown to be a proper distribution. To obtain a closed form expression for (3.22), we set $v_{\Theta_b[n]\to f_b[n]}$ to a uniform distribution over $[-\pi, \pi)$. The uniform prior on $\Theta_b[n]$ is exact for $b = 1$, as indicated in Table 3.1. For $b > 1$, however, the prior on $\Theta_b[n]$ may not be uniform over $[-\pi, \pi)$ due to the Wiener phase statistics. While the true prior on $\Theta_b[n]$ is used to compute beliefs on the phase variables, we assume that $\Theta_b[n]$ is uniform over $[-\pi, \pi)$ to compute the belief in (3.22). This assumption, although a limitation for $b > 1$, is made to obtain a closed form expression for (3.22).

The closed form expression for $\bar{v}_{f_b[n]\to r[n]}$, which approximates $v_{f_b[n]\to r[n]}$, is given in (3.48) of Section 3.7.1. This message $\bar{v}_{f_b[n]\to r[n]}$, derived under the assumption that $v_{\Theta_b[n]\to f_b[n]} = 1/2\pi$, includes the modified Bessel function and propagating such a message is computationally expensive [112, 116]. To tackle this challenge, we approximate $\bar{v}_{f_b[n]\to r[n]}$ as a Gaussian PDF, which can be fully characterized by its mean and variance. We define $\varrho_b[n] = |e_b[n]|^2/2c_b[n]$, $t_b[n]$ as the mean of $\bar{v}_{f_b[n]\to r[n]}$ and $v_b[n]$ as its variance. The expressions for $t_b[n]$ and $v_b[n]$, derived in Section 3.7.1, are

$$t_b[n] = \frac{\sqrt{c_b[n]/\pi}}{\epsilon\left(1-\Omega\left(\tilde{\pi}_b[n]\right)\right) + \Omega\left(\tilde{\pi}_b[n]\right)} \times \frac{\exp(\varrho_b[n])}{I_0(\varrho_b[n])}, \tag{3.23}$$

$$v_b[n] = \frac{\epsilon^{-1}\left(1-\Omega\left(\breve{\pi}_b[n]\right)\right)+\Omega\left(\breve{\pi}_b[n]\right)}{\epsilon\left(1-\Omega\left(\breve{\pi}_b[n]\right)\right)+\Omega\left(\breve{\pi}_b[n]\right)} \times \frac{c_b[n]}{2} \times \left(1+2\varrho_b[n]\left(1+\frac{\mathrm{I}_1\left(\varrho_b[n]\right)}{\mathrm{I}_0\left(\varrho_b[n]\right)}\right)\right) - t_b^2[n]. \tag{3.24}$$

In summary, we assumed that $v_{\Theta_b[n]\to f_b[n]} = 1/2\pi$ to approximate the out-of-plane message $v_{f_b[n]\to \mathrm{r}[n]}$ as

$$v_{f_b[n]\to \mathrm{r}[n]} \propto \mathcal{N}(r[n]; t_b[n], v_b[n]). \tag{3.25}$$

We make use of (3.25) to compute the into-plane message $v_{\mathrm{r}[n]\to f_b[n]}$ for $b > 1$. Using the sum-product rule, we can write from the factor graph in Fig. 3.3 that

$$v_{\mathrm{r}[n]\to f_b[n]} = p(r[n]) \prod_{\ell=1}^{b-1} v_{f_b[n]\to \mathrm{r}[n]}. \tag{3.26}$$

Substituting (3.25) and the Rayleigh PDF $d[n] = p(r[n])$ from Table 3.1 in (3.26), we can write

$$v_{\mathrm{r}[n]\to f_b[n]} \propto r[n]\mathcal{N}(r[n]; \eta_b[n], \gamma_b[n]), \quad \forall r[n] \in \mathbb{R}_+, \tag{3.27}$$

where

$$\eta_b[n] = \frac{\sum_{\ell=1}^{b-1} \frac{t_\ell[n]}{v_\ell[n]}}{\frac{2}{\mu} + \sum_{\ell=1}^{b-1} \frac{1}{v_\ell[n]}}, \quad \gamma_b[n] = \frac{1}{\frac{2}{\mu} + \sum_{\ell=1}^{b-1} \frac{1}{v_\ell[n]}}. \tag{3.28}$$

The simplification in (3.27) follows from the fact that the product of $b-1$ Gaussian PDFs is also Gaussian and that the Rayleigh PDF $p(r[n])$ from Table 3.1 can be rewritten as $p(r[n]) = 2\sqrt{\pi/\mu}r[n]\mathcal{N}(r[n]; 0, \mu/2)$ where $r[n] \in \mathbb{R}_+$.

Messages with the support variables: Next, we derive the out-of-plane message $v_{f_b[n]\to \mathrm{s}[n]}$. By the sum-product rule,

$$v_{f_b[n]\to \mathrm{s}[n]} = \int_{z_b[n]} \int_{r[n]} \int_{\theta_b[n]} f_b[n] v_{z_b[n]\to f_b[n]} v_{\mathrm{r}[n]\to f_b[n]} v_{\Theta_b[n]\to f_b[n]}. \tag{3.29}$$

For the Bernoulli PMF $v_{f_b[n]\to \mathrm{s}[n]}$, the belief that $\mathrm{s}[n] = 1$ is

$$\vec{\pi}_b[n] = \frac{[v_{f_b[n]\to \mathrm{s}[n]}]_{\mathrm{s}[n]=1}}{[v_{f_b[n]\to \mathrm{s}[n]}]_{\mathrm{s}[n]=1} + [v_{f_b[n]\to \mathrm{s}[n]}]_{\mathrm{s}[n]=0}}. \tag{3.30}$$

To compute $\vec{\pi}_b[n]$, we derive $v_{f_b[n]\to \mathrm{s}[n]}$ under the assumption that $v_{\Theta_b[n]\to f_b[n]} = 1/2\pi$. The same assumption was made to approximate $\bar{v}_{f_b[n]\to \mathrm{r}[n]}$. To obtain a closed-form expression for $v_{f_b[n]\to \mathrm{s}[n]}$, we use a Rice distributed [122] approximation of $v_{\mathrm{r}[n]\to f_b[n]}$ in

(3.27) with parameters $\rho_b[n]$, $\tau_b[n]$, i.e.,

$$\text{Rice}\left(r[n];\rho_b[n],\tau_b[n]\right) = \frac{r[n]}{\tau_b[n]}\exp\left(-\frac{r^2[n]+\rho_b^2[n]}{2\tau_b[n]}\right)\text{I}_0\left(\frac{\rho_b[n]}{\tau_b[n]}r[n]\right), \tag{3.31}$$

with the same mean and variance as $v_{\text{r}[n]\to f_b[n]}$. The Rice distribution is a reasonable choice as the non-zero entries of $\mathbf{x}$ have a complex Gaussian distribution. The parameters $\rho_b[n]$ and $\tau_b[n]$ can be computed from $v_{\text{r}[n]\to f_b[n]}$ in (3.27), using the RiceFit function in MATLAB [123]. The inputs of this function are the mean and the variance of $v_{\text{r}[n]\to f_b[n]}$, given by (3.61) and (3.62), in Section 3.7.2. Finally, we define $\bar{\tau}_b[n] = (2\tau_b[n] + c_b[n])/2$ and approximate the probability in (3.30) as

$$\vec{\pi}_b[n] \approx \frac{\text{Rice}(|e_b[n]|;\rho_b[n],\bar{\tau}_b[n])}{\frac{2|e_b[n]|}{c_b[n]}\exp\left(-\frac{|e_b[n]|^2}{c_b[n]}\right)+\text{Rice}(|e_b[n]|;\rho_b[n],\bar{\tau}_b[n])}. \tag{3.32}$$

The derivation of (3.32) can be found in Section 3.7.2.

Next, we compute $v_{\text{s}[n]\to f_b[n]}$ for $b > 1$. By the sum-product rule, the probability associated with this Bernoulli PMF is

$$\tilde{\pi}_b[n] = \frac{\lambda\prod_{\ell=1}^{b-1}\vec{\pi}_\ell[n]}{(1-\lambda)\prod_{\ell=1}^{b-1}\left(1-\vec{\pi}_\ell[n]\right)+\lambda\prod_{\ell=1}^{b-1}\vec{\pi}_\ell[n]}, \tag{3.33}$$

where the probability derived in (3.32) is used for $\vec{\pi}_\ell[n]$.

Messages with the phase variables: We now derive $v_{f_b[n]\to\Theta_b[n]}$ and $v_{\Theta_b[n]\to q_{b+1}[n]}$. To obtain a tractable approximation for these messages, we assume that the phase of the intra-plane AMP estimate of $z_b[n]$ is free from errors. Under this assumption, the intra-plane estimate of $z_b[n]$ has a phase equal to that of its mean in (3.19), i.e., $\angle e_b[n]$. Specifically,

$$v_{f_b[n]\to\Theta_b[n]} = \delta\left(\theta_b[n]-\angle e_b[n]\right). \tag{3.34}$$

Due to the Dirac-delta PDF in (3.34), the inter-plane message

$$v_{\Theta_b[n]\to q_{b+1}[n]} = v_{f_b[n]\to\Theta_b[n]}v_{q_b[n]\to\Theta_b[n]},$$

is also a Dirac impulse, i.e., $v_{\Theta_b[n]\to q_{b+1}[n]} \propto \delta(\theta_b[n]-\angle e_b[n])$. The message $v_{q_{b+1}[n]\to\Theta_{b+1}[n]}$ forwarded to the $b+1^{\text{th}}$ plane is then

$$\begin{aligned} v_{q_{b+1}[n]\to\Theta_{b+1}[n]} &= \int_{\theta_b[n]} q_{b+1}[n]v_{\Theta_b[n]\to q_{b+1}[n]} \\ &\propto \mathcal{N}\left(\theta_{b+1}[n];\varphi_{b+1}[n],\sigma_{\text{pn}}^2\right). \end{aligned} \tag{3.35}$$

where $\varphi_{b+1}[n] = \angle e_b[n]$. For $b > 1$, we observe from Fig. 3.4 that the variable $\Theta_b[n]$ connects the factors $q_b[n]$ and $f_b[n]$ in the forward pass. Therefore, $v_{\Theta_b[n]\to f_b[n]} = v_{q_b[n]\to\Theta_b[n]}$, i.e.,

$$v_{\Theta_b[n]\to f_b[n]} \propto \mathcal{N}(\theta_b[n];\varphi_b[n],\sigma_{\text{pn}}^2). \tag{3.36}$$

Into-plane message $v_{f_b[n]\to z_b[n]}$: For the first plane $b = 1$, the message $v_{f_b[n]\to z_b[n]}$ is simply the BG prior in (3.18) which enforces sparsity. The belief on $z_b[n]$ in the subsequent planes ($b \geq 2$) could incorporate more information than a simple BG prior. This additional information is due to the statistics of the support, magnitude, and phase learned from all preceding planes in the forward pass.

From the sum-product rule, the message $v_{f_b[n]\to z_b[n]}$ is computed as

$$v_{f_b[n]\to s[n]} = \int_{r[n]} \int_{s[n]} \int_{\theta_b[n]} f_b[n] v_{\Theta_b[n]\to f_b[n]} v_{s[n]\to f_b[n]} v_{r[n]\to f_b[n]},$$

where the individual terms are listed in (3.36), (3.33) and (3.27). As this integration is challenging, we approximate it as a Bernoulli-Gaussian Mixture (BGM) distribution. Such an assumption aids computationally tractable messages for subsequent planes. An example of the true $v_{f_b[n]\to z_b[n]}$ and our BGM approximation is shown in Fig. 3.5(a). This belief, originating from the preceding plane, comprises non-zero probabilities concentrated on an annular segment as illustrated in Fig. 3.5(a). The radius of this segment is governed by the magnitude prior on $z_2[1]$, while its angular spread is determined by the uncertainty in $z_1[1]$ as well as the variance of the incremental phase error due to the Wiener process.

Our BGM approximation of $v_{f_b[n]\to z_b[n]}$ uniformly spaces the Gaussians on the annular segment shown in Fig. 3.5(a). We assume that the means of the complex Gaussians in our BGM are of the form

$$\beta_{b,k}[n] = \bar{\beta}_b[n] e^{\mathrm{j}\angle\beta_{b,k}[n]}, \tag{3.37}$$

to achieve a non-zero probability density over the annular segment. In (3.37), we use $\bar{\beta}_b[n]$ to denote the magnitude of $\beta_{b,k}[n]$. The magnitude of means, i.e., $\bar{\beta}_b[n]$, is the same for all Gaussian components in our BGM. We set

$$\bar{\beta}_b[n] = \mathrm{mean}\big(v_{r[n]\to f_b[n]}\big), \tag{3.38}$$

and its expression is derived in (3.61). We assume the same variance of $\xi_b[n]$ for all Gaussian components in our BGM. We use $K_b[n]$ to denote the number of complex Gaussians in our BGM approximation and assume that $K_b[n]$ is odd. The weight of the k^{th} GM component is denoted by $\omega_{b,k}[n] \in [0, 1]$, with $\sum_{k=1}^{K_b[n]} \omega_{b,k}[n] = 1$. Our BGM approximation takes the form

$$\begin{aligned} v_{f_b[n]\to z_b[n]} \approx\ & (1 - \bar{\pi}_b[n])\delta(z_b[n]) \\ & + \bar{\pi}_b[n] \sum_{k=1}^{K_b[n]} \omega_{b,k}[n] \mathscr{CN}\big(z_b[n]; \beta_{b,k}[n], \xi_b[n]\big). \end{aligned} \tag{3.39}$$

The parameters in our BGM, i.e., the $\omega_{b,k}[n]$s, $K_b[n]$, and $\xi_b[n]$ are learned from the true $v_{f_b[n]\to z_b[n]}$.

Next, we observe that the Gaussian mixture in (3.39) approximates

$$\left[v_{f_b[n]\to z_b[n]}\right]_{s[n]\neq 0} = \bar{\pi}_b[n] \iint_{r[n]\theta_b[n]} \delta(z_b[n] - r[n]e^{j\theta_b[n]}) v_{r[n]\to f_b[n]} \times v_{\Theta_b[n]\to f_b[n]}. \tag{3.40}$$

From (3.40), (3.27) and (3.36), it can be seen that $\left[v_{f_b[n]\to z_b[n]}\right]_{s[n]\neq 0}$ is exactly zero at $z_b[n] = 0$. To make sure that the mixture in (3.39) is nearly zero at $z_b[n] = 0$, the variance $\xi_b[n]$ is chosen such that the means, i.e., $\beta_{b,k}[n]$s, are at least $2\sqrt{\xi_b[n]/2}$ away from $z_b[n] = 0$. Also, the variance $\xi_b[n]$ should be large enough to encompass the annular region with a limited number of Gaussian mixture components, with this limit imposed for computational tractability. We use

$$\xi_b[n] = 2\min\left\{\left(\frac{\text{mean}\left(v_{r[n]\to f_b[n]}\right)}{2}\right)^2, \text{var}\left(v_{r[n]\to f_b[n]}\right)\right\}, \tag{3.41}$$

where $\text{var}\left(v_{r[n]\to f_b[n]}\right)$, derived in (3.62), is the variance of the true belief $v_{r[n]\to f_b[n]}$. When $\xi_b[n]$ from (3.41) is small, a substantial number of GM components may be needed to encompass the annular region. To avoid the high computational complexity in such a case, we use $\max\{\xi_b[n], \Xi_b[n]\}$ as the variance of the GM components in our approximation. Here, $\Xi_b[n]$ is chosen to make sure that $K_b[n]$ is smaller than 60.

The phase values of $\beta_{b,k}[n]$s are symmetrically placed around $\beta_{b,(K_b[n]+1)/2}[n]$, which is set to the mean of $v_{\Theta_b[n]\to z_b[n]}$, i.e., $\angle\beta_{b,(K_b[n]+1)/2}[n] = \varphi_b[n]$. When $\sigma^2_{\text{pn}} \leq \xi_b[n]$, we use just $K_b[n] = 1$ complex Gaussian component to cover the annular region. When $\sigma^2_{\text{pn}} > \xi_b[n]$, the number of GM components $K_b[n]$ are chosen such that the Euclidean distance between any two neighboring $\beta_{b,k}[n]$s is at least $2\sqrt{\xi_b[n]/2}$ as shown in Fig. 3.5(b). To satisfy this condition, we set

$$K_b[n] = 2\left\lfloor \frac{\pi - \sin^{-1}\left(\frac{\sqrt{\xi_b[n]/2}}{\bar{\beta}_b[n]}\right)}{2\sin^{-1}\left(\frac{\sqrt{\xi_b[n]/2}}{\bar{\beta}_b[n]}\right)} \right\rfloor + 1 \text{ and,} \tag{3.42}$$

$$\angle\beta_{b,k}[n] = \varphi_b[n] + \left(\frac{K_b[n]+1}{2} - k\right) \frac{2\pi - 2\sin^{-1}\left(\frac{\sqrt{\xi_b[n]/2}}{\bar{\beta}_b[n]}\right)}{K_b[n]-1}. \tag{3.43}$$

The derivations of (3.42) and (3.43 are provided in 3.7.3.

Now, we explain how we determine the weights $\omega_{b,k}[n]$s that scale the GM components in (3.39). As seen from Fig. 3.5(a), the probability mass in the annular region tapers off around the central angle, due to Gaussian statistics of the phase error innovations. As a result, we assign the largest weight to the $(K_b[n]+1)/2$-th GM component with

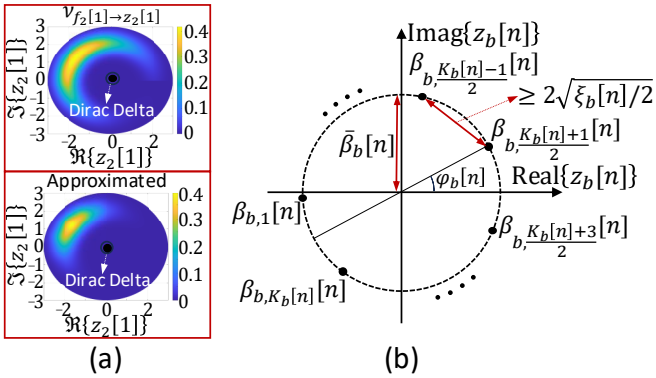


Figure 3.5: In the forward pass, solving for auxiliary vector entries up to plane b provides side information (belief) on the entry in plane b. We approximate this belief, denoted by $v_{f_b[n]\to z_b[n]}$, as a Bernoulli-Gaussian mixture, both of which are shown in (a). We assume that the Gaussian mixture components have the same variance $\xi_b[n]$, and that their means have the same magnitude denoted by the dots in (b).

mean $\bar{\beta}_b[n]e^{j\varphi_b[n]}$ and small weight to the components whose mean is farther away from $\beta_{b,(K_b[n]+1)/2}[n]$. We define $\{\tilde{\omega}_{b,k}[n]\}_{k=1}^{K_b[n]}$ as

$$\tilde{\omega}_{b,k}[n] = \mathcal{N}\big(\beta_{b,k}[n]; \beta_{b,(K_b[n]+1)/2}[n], \sigma^2_{\text{pn}}[n]\big), \tag{3.44}$$

to denote the unnormalized weights assigned to the GM components. The GM weights $\omega_{b,k}[n]$s are computed by normalizing the weights in (3.44) as

$$\omega_{b,k}[n] = \frac{\tilde{\omega}_{b,k}[n]}{\sum_{k=1}^{K_b[n]} \tilde{\omega}_{b,k}[n]}, \tag{3.45}$$

We notice from Fig. 3.5(a) that our approximated BGM belief in (3.39) well approximates the true $v_{f_b[n]\to z_b[n]}$. We observed empirically that our message passing algorithm can successfully reconstruct the channel even with the approximated belief shown in Fig. 3.5(a).

After performing the "Intra-plane" step of plane B in the first forward message flow, the backward message flow can be performed. To this end, the messages in the "Out-of-plane" step of plane B are computed in the same way as in the forward message flow. These messages are then used to compute the messages for the "Inter-plane" and the "Into-plane" step of plane $B-1$. Next, the proposed GM approximation is used to obtain local priors for $\{z_{B-1}[n]\}_{n=1}^{N^2}$ which are used to perform AMP iterations in the "Intra-plane" step. This procedure continues until it performs the "Intra-plane" step in plane 1 to obtain a belief about $\{z_1[n]\}_{n=1}^{N^2}$. In the forward message flow, the messages related to plane b are updated based on the information received from planes 1 up to plane $b-1$ while in the backward message flow the messages related to plane b are updated based on the

Algorithm 2: Channel estimation with one forward pass

Input: Parameters λ and μ of the prior distribution (3.6) of $\mathbf{x}$, phase perturbed CS measurements $\{\hat{\mathbf{y}}_b\}_{b=1}^{B}$ (3.4), CS matrices $\{\mathbf{A}_b\}_{b=1}^{B}$, phase noise variance σ_{pn}^2 and the measurement noise σ^2.

While $b \leq B$ **do**:

If $b = 1$ **then**

1. Estimate $\hat{\mathbf{z}}_1[n]$ with the AMP [9] using prior in (3.18).
2. Determine $\vec{\pi}_1[n]$s assuming $\rho_b[n] = 0, \tau_b[n] = \mu/2, \ \forall n \in [N^2]$ using (3.32) for the first batch.
3. Find $\{t_1[n]\}_{n=1}^{N^2}$ using (3.23) and $\{v_1[n]\}_{n=1}^{N^2}$ using (3.24).

Else

4. Calculate $\{\tilde{\pi}_b[n]\}_{n=1}^{N^2}$ using (3.33).
5. Set $\varphi_b[n] = \angle e_{b-1}[n], \forall n \in [N^2]$.
6. Compute parameters $\{\omega_{b,k}[n]\}_{k=1}^{K_b[n]}$ from (3.45), $\{\beta_{b,k}[n]\}_{k=1}^{K_b[n]}, \forall n \in [N^2]$ from (3.37), (3.38), (3.43), and $\{\xi_b[n]\}_{n=1}^{N^2}$ from (3.41) for the BGM approximation in (3.39).
7. Use learned BGM prior in (3.39) to estimate $\hat{\mathbf{z}}_b[n]$ with AMP.
8. Compute parameters $\{\rho_b[n]\}_{n=1}^{N^2}$ and $\{\tau_b[n]\}_{n=1}^{N^2}$ of the Rice distribution (3.31) fit to the magnitude-related messages in (3.27).
9. Find $\{\vec{\pi}_b[n]\}_{n=1}^{N^2}$ using (3.32).
10. Compute the parameters $\{\eta_b[n]\}_{n=1}^{N^2}$ and $\{\gamma_b[n]\}_{n=1}^{N^2}$ of the magnitude related message (3.27) going into the batch.
11. Compute $\{t_b[n]\}_{n=1}^{N^2}$ using (3.23) and $\{v_b[n]\}_{n=1}^{N^2}$ using (3.24).

End If

12. $b \leftarrow b + 1$.

End While

13. Determine $\hat{\mathbf{x}}$ using Algorithm 1 based on the estimates $\{\hat{\mathbf{z}}_b\}_{b=1}^{B}$ obtained from the above procedure.

Output: $\hat{\mathbf{x}}$.

information received from planes B down to plane $b+1$. Multiple rounds of forward and backward message flows can be used to refine the estimates $\{\hat{z}_b\}_{b=1}^{B}$.

Now, we discuss the computational complexity of our developed message passing-based method outlined in Algorithm 2. As our algorithm performs AMP within each plane, and we limit the number of Gaussian mixture components similar to [67], the computational complexity of calculating parameters for the intra-plane messages, i.e., steps **1** and **7** in Algorithm 2, is dominated by matrix multiplication by $\mathbf{A}_b$ resulting in $\mathcal{O}\left(MN^2\right)$ complexity [67]. Furthermore, the parameters of the into-plane, out-of-plane, and inter-plane messages, i.e., steps **2**–**6** and **8**–**11**, are computed independently for each n and do not depend on the number of measurements, resulting in an overall

computational complexity of $\mathcal{O}\left(N^2\right)$. Therefore, assuming T^{amp} AMP iterations per plane, the overall complexity per plane is dominated by $\mathcal{O}\left(T^{\text{amp}}MN^2\right)$. Finally, for B planes and T^{fb} forward and backward message propagation, the overall computational complexity is $\mathcal{O}\left(T^{\text{fb}}BT^{\text{amp}}MN^2\right)$. The computational complexity of the standard AMP from [9] for sparse recovery using BM measurements is $\mathcal{O}\left(BT^{\text{amp}}MN^2\right)$. We observe that the complexity of our approach is higher than that of the AMP by a factor of T^{fb}. This is due to exploiting the Wiener statistics in the phase noise, which results in better channel reconstruction. The complexity of both methods, however, scales only linearly with the number of measurements M and the array dimension N^2.

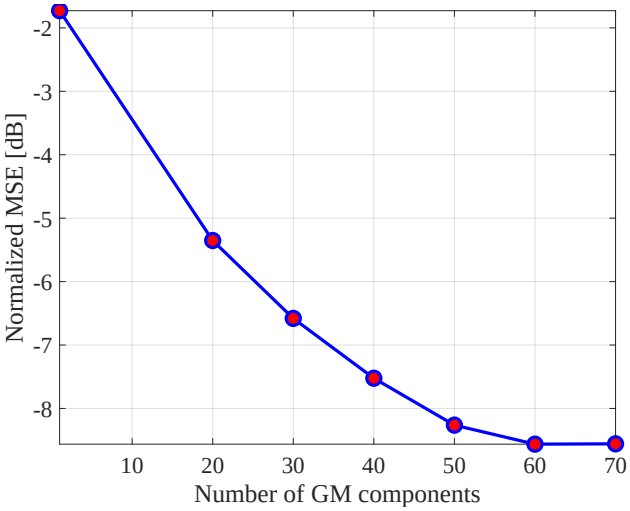


Figure 3.6: NMSE with the number of Gaussian components used to approximate $v_{f_b[n]\to z_b[n]}$ in (3.39). Here, we use $M = 50$, $B = 4$, and SNR = −15 dB. We limit the maximum number of Gaussian components to 60 as the NMSE remains unchanged when more components are used."

Due to the uniform prior $\mathcal{U}([-\pi, \pi])$ on the phase variable in plane 1, i.e., Θ_1, the number of GM components used in the "Into-plane" step of plane 1 is higher than that of other planes. Fewer GM components are needed in the subsequent planes as the prior is concentrated (as shown in Fig. 3.5(a)). In Fig. 3.6, we investigate the impact of the number of GM components used to approximate $v_{f_b[n]\to z_b[n]}$ in (3.39). As can be seen, increasing the number of GM components in the approximation of $v_{f_b[n]\to z_b[n]}$ using (3.39) improves the NMSE of the estimate as it improves the approximation of $v_{f_b[n]\to z_b[n]}$. It is observed, however, that the NMSE remains almost the same when the number of GM components is greater than 60. Therefore, we limit the number of GM components to 60 to avoid high computational complexity.

3.4. Simulation Results

We consider a narrowband system with a 16 × 16 UPA at the TX for f_c = 60 GHz. The TX-RX distance is 15 m. Unless otherwise specified, phase noise at f_m = 1 MHz offset from f_c is $\mathscr{L}(f_m)$ = −101.7 dBc/Hz [115]. The duration to obtain each measurement is Δ_{bf} = 128 ns and the spacing between successive batches is Δ_{sf} = 44 μs [56], which results in $\sigma^2_{pn} \approx 0.12$ [55]. We use B = 4 batches. The beam training matrices $\{\mathbf{P}_b[m]\}_{m=1}^{M}$ $\forall b \in \{1, \cdots, B\}$ applied at the UPA are random circular shifts of a 16 × 16 perfect binary array [45]. For each beam training matrix at the TX, spread sequences of duration 128 ns are transmitted within the training fields in Fig. 3.1(b) [45], resulting in a spreading gain of about 20 dB. We use 100 urban micro line-of-sight channels from the NYU simulator [6]. Therefore, the AoDs can be off-grid resulting in approximately sparse beamspace channels. This channel model is more realistic than the model in [56] which assumes on-grid AoDs and known sparsity. Finally, Wiener phase noise is introduced considering time durations of Δ_{sf} and Δ_{bf}. This induces phase errors in the CS measurements acquired both within and across batches. Although our algorithm models only one phase error per batch (see (3.5)), the proposed method works even when the measurements within a batch are affected by varying phase errors.

The channel dataset is used to calculate the prior parameters λ and μ in (3.6). To compute the sparsity rate λ, we first find the strongest entries in the angle domain channel that account for 95 percent of the energy in the channel. Denoting this number by K, we then calculate λ as the mean of K/N^2 over all channels within the dataset. Next, the sample variance of the K strongest entries from all the channels is used to compute μ. This resulted in $\lambda \approx 0.07$ and $\mu \approx 13.3$ for our dataset. An expectation-maximization-based approach can be used in practice to learn these parameters; however, this extension is beyond the scope of the dissertation.

We use $\hat{\mathbf{H}}$ to denote the channel estimate, which is derived from the angle domain estimate $\hat{\mathbf{X}}$ as $\hat{\mathbf{H}} = \mathbf{U}_N \hat{\mathbf{X}} \mathbf{U}_N$. The normalized mean squared error (NMSE) is defined as $\mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2 / \|\mathbf{H}\|_F^2]$. The signal-to-noise ratio (SNR) is given by SNR = $\sum_{b=1}^{B} \|\mathbf{A}_b \mathbf{x}\|^2 / MB\sigma^2$. To avoid numerical issues at high SNRs, we use only one Gaussian mixture component, i.e., the one with the largest weight, to approximate $\nu_{f_b[n] \to z_b[n]}$. We use T_{max} = 5 in Algorithm 1 and $\epsilon = 10^{-6}$ in (3.21). To further limit the number of GM components, we remove the GM components whose corresponding non-normalized weight is 10^{-3} times smaller than the maximum non-normalized weight $\tilde{\omega}_{b,(K_b[n]+1)/2}[n]$ in (3.44).

We benchmark our method against our *SuppOnly* method [119], which solves (3.5) by exploiting only the common support across $\{z_b\}_{b=1}^{B}$. Unlike [119], the technique proposed in this chapter incorporates Wiener phase noise statistics and the common magnitude structure across $\{z_b\}_{b=1}^{B}$, in addition to leveraging the support structure. We also benchmark against partially coherent sparse recovery based on *SparseLift* [16]. The application of *SparseLift* to solve (3.5) was discussed in [56], wherein the lifted matrix $\boldsymbol{\Gamma} = e^{j\boldsymbol{\phi}} \mathbf{x}^T$ is first estimated. For a fair comparison, we use AMP [9] to solve for $\boldsymbol{\Gamma}$. Then, the singular-value decomposition of the estimate $\hat{\boldsymbol{\Gamma}}$ is used to determine $\hat{\mathbf{x}}$ [56]. We also use the partially coherent matching pursuit (PCMP) algorithm from [56] to benchmark our method. PCMP is a greedy algorithm that iteratively detects the support of $\mathbf{x}$, similar to the matching pursuit algorithm. Then, the phase error and the coefficient corresponding

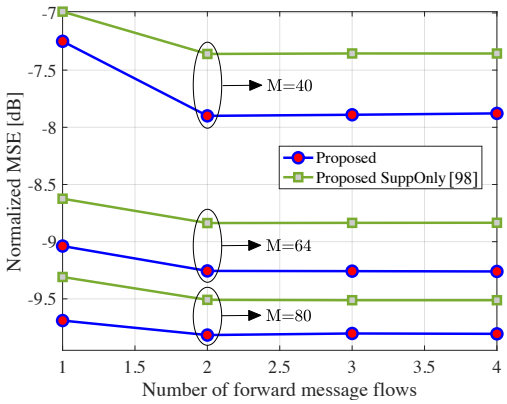


Figure 3.7: NMSE with the number of forward passes for $B = 4$, SNR = −15 dB. NMSE stabilizes beyond the second pass, as measurements from all batches are used starting with the second pass.

to the detected support are estimated using an alternating optimization. PCMP requires knowledge of the sparsity level for $\mathbf{x}$. Therefore, we use PCMP algorithms with two different sparsity levels as benchmarks, denoted by PCMP1 and PCMP2. PCMP1 uses a sparsity level of $\lceil \lambda N^2 \rceil = 19$ and PCMP2 uses a sparsity level of $2\lceil \lambda N^2 \rceil = 38$ with $\lceil \cdot \rceil$ denoting the ceil operator. We also use a baseline, called *IgnoredPN*, that just stacks the CS matrices across all batches and employs standard AMP over the stacked phase perturbed measurements [9]. Finally, we compare our algorithm against ideal case, called *ZeroPN*, which uses phase error-free measurements for sparse recovery with the stacked CS matrices.

From Fig. 3.7, we observe that the NMSE with the proposed method and the method from [119] decrease until two forward message flows. This is because, to estimate z_b in the first message flow, only measurements until the current batch, i.e., $\{\mathbf{y}_\ell\}_{\ell=1}^{b}$, are used. By the end of the second forward message flow, two forward and one backward message flows are completed, ensuring that measurements from all the B batches are used for estimation. Additional message flows beyond this point did not reduce the estimation error.

We notice from Fig. 3.8 that our method results in a lower NMSE than the benchmarks when $\mathscr{L}(f_{\mathrm{m}})$, the phase noise power spectral density, is smaller than −96 dBc/Hz. Note that *SparseLift* does not incorporate Wiener phase noise statistics. At high $\mathscr{L}(f_{\mathrm{m}})$, however, the phase errors within a batch become large enough to perturb the auxiliary estimates in each plane. We observe that our *proposed* method is severely impacted by these errors when compared to the *proposed SuppOnly* method. This is because our *proposed* method propagates not only the beliefs about support, but also beliefs about both magnitude and phase to subsequent planes, which are inaccurate under strong phase noise perturbations, i.e., $\mathscr{L}(f_{\mathrm{m}}) > -96$ dBc/Hz. These inaccuracies arise because our model assumes no phase noise variation within a batch, which is invalid at high $\mathscr{L}(f_{\mathrm{m}})$. These distorted phase and magnitude beliefs are propagated across the

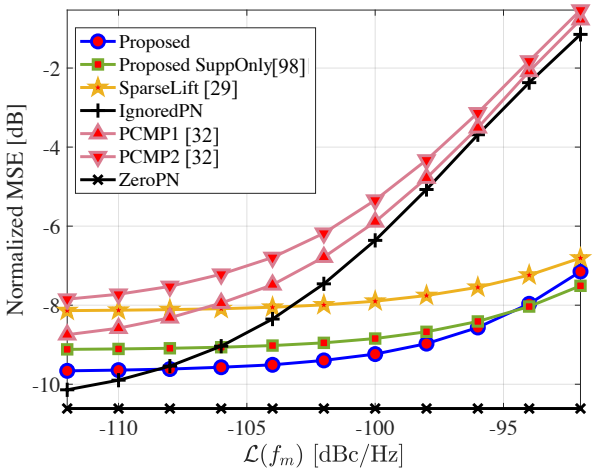


Figure 3.8: NMSE with phase noise power spectral density $\mathscr{L}(f_m)$ for $B = 4$, $M = 64$ and SNR = −15 dB. Our method outperforms SparseLift [16] and SuppOnly [119] methods for $\mathscr{L}(f_m) < -96$ dBc/Hz.

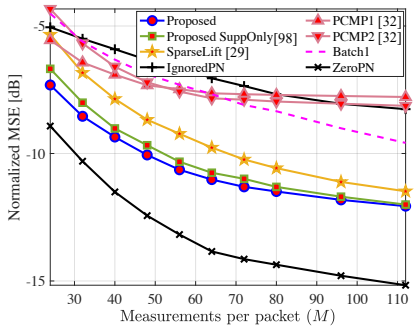
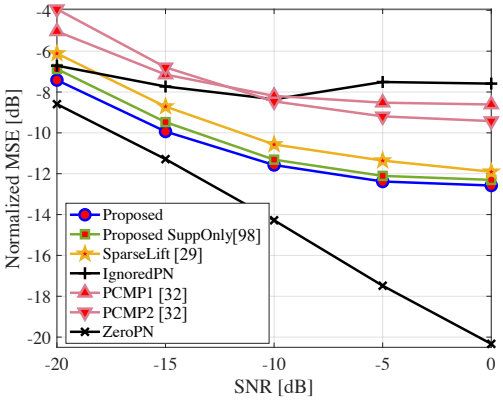


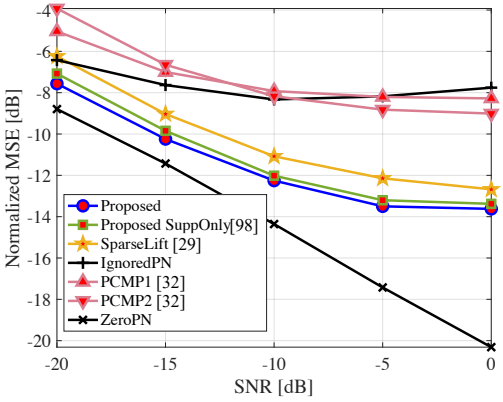
Figure 3.9: NMSE with M for $B = 4$, SNR = −10 dB. The proposed method results in a smaller NMSE than benchmarks with phase-perturbed measurements.

planes, ultimately leading to a poorer channel estimate with the *proposed* method than with the *proposed SuppOnly* method. This means that for low-quality oscillators with $\mathscr{L}(f_m) > -96$ dBc/Hz, our *SuppOnly* method is suitable, while the *proposed* method is suitable when $\mathscr{L}(f_m) < -96$ dBc/Hz. The derived method in Algorithm 2 is still valuable as the oscillators in [115, 124, 125] have $\mathscr{L}(f_m) < -96$ dBc/Hz. Finally, we observe that the *IgnoredPN* method that just stacks all the measurements performs better than our method when $\mathscr{L}(f_m) < -108$ dBc/Hz. At 60 GHz, the oscillator in [115, 124, 125] generates a phase noise between −99.4 dBc/Hz and −104.6 dBc/Hz, for which our method provides the lowest NMSE. Note that the range of $\mathscr{L}(f_m)$ for which our method performs best varies with the choice of Δ_{sf} and Δ_{bf}.

We observe from Fig. 3.9 that the proposed method achieves about 0.6 dB improve-



(a) Actual phase noise (varies within and across batches).



(b) Phase noise consistent with our model in (3.5).

Figure 3.10: NMSE with SNR for $B = 4$, $M = 80$. Our method results in a smaller NMSE than the benchmarks for the phase noise in (a) and (b). In (a), the measurements are corrupted by the actual phase noise, which varies both within and across batches. In (b), consistent with our model in (3.5), the phase noise is constant over each batch, and it varies across batches. Approximating the actual phase noise with a constant phase noise within each batch results in a loss of about 0.9 dB in the NMSE at high SNR.

ment in the NMSE compared to our *SuppOnly* method. This is because our factor graph in Fig. 3.3 exploits additional structure in the auxiliary vectors, i.e., magnitude and phase correlation, while the *SuppOnly* method exploits only the common support structure. The gap between our proposed technique and the *SuppOnly*, however, reduces at a large M as the beliefs enforcing common support structure have a stronger impact than those incorporating magnitude and phase correlation. Finally, we observe that there is almost no improvement in the NMSE with our method over *SuppOnly*, when $M \geq 95$. This is because our assumption that the phase noise is almost the same over a batch

breaks down for a long batch, equivalently a large M. In Fig. 3.9, we also included *Batch1* method wherein standard AMP from [9] is used to estimate the beamspace using only measurements from batch 1. Specifically, the *Batch1* method solves for the sparse $\mathbf{x}$ from $\mathbf{y}_1 = \mathbf{A}_1\mathbf{x} + \mathbf{w}_1$ and ignores all the other measurements. Simply stacking the other measurements (*IgnoredPN*) does not work well compared to the *Batch1* method due to inter-batch phase errors. Our *proposed* method leverages useful information from other batches even under these phase errors, resulting in a lower NMSE than the *Batch1* method.

Fig. 3.10 shows that our technique achieves a lower NMSE than the *SuppOnly*, especially at a low SNR, as it exploits additional structure. As SNR increases, NMSE reduces with the proposed method and the *SuppOnly* method. The mismatch between the batch-based phase noise model considered in the measurement model (3.5) in our algorithm and the realistic model in our simulations results in the NMSE floor in Fig. 3.10(a). The error floor due to phase noise is also evident from the observation that the NMSE with *IgnoredPN* does not improve at high SNR. By comparing Fig. 3.10(b) and Fig. 3.10(a), we observe that approximating the actual phase noise with a constant phase noise within each batch results in a loss of about 0.9 dB in the NMSE at high SNR. It is also seen that the NMSE saturates beyond a certain SNR, even with the same phase noise within each batch. The NMSE saturation is due to the use of the relaxed version of the true phase model in (3.13), given by (3.14), to construct our factor graph. Finally, we observe that PCMP1 and PCMP2 do not perform well compared to the proposed, the *SuppOnly* and the *SparseLift* methods. This is because PCMP is designed for exactly sparse channels with known sparsity levels, while the channels in our simulations are approximately sparse.

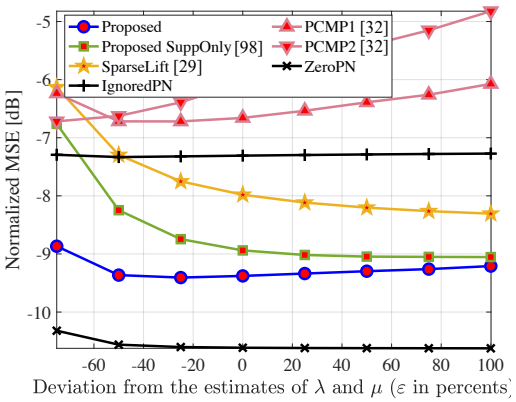


Figure 3.11: NMSE with the deviation from the estimates of the prior parameters λ and μ in percentages. Here, we use $M = 64$ for $B = 4$, SNR $= -15$ dB. Our proposed method is robust to inaccuracies in the prior parameters λ and μ, when compared to the benchmarks.

In Fig. 3.11, we investigate the robustness of our method to inaccuracies in prior parameters estimation. To this end, we add perturbations to the estimated priors, i.e.,

λ and γ, as $(1+\varepsilon)\lambda$ and $(1+\varepsilon)\mu$ with $\varepsilon \in [-1, 1]$. As it is observed from Fig. 3.11, our *proposed* approach is more robust to mismatch in the estimated prior parameters than the benchmarks.

3.5. Potential research directions to address challenges with the proposed method

In this chapter, a total of BM spatial channel measurements are acquired using B training batches with M measurements per batch for channel estimation. The use of a large number of training batches results in a low spectral efficiency due to a high training overhead and a high computational complexity. One way to address this challenge is to use a single batch, i.e., $B = 1$. With only a single batch, however, as seen from Fig. 9, the channel estimation error is high due to the limited number of measurements acquired using a single batch. To reduce the training overhead while achieving a low estimation error, we suggest progressively decreasing the number of acquired measurements with the batch index b. The use of high measurements initially provides a good channel estimate. This estimate may be further refined using fewer phase-mismatched measurements. In our framework, this is equivalent to having CS matrices $\{\mathbf{A}_b\}_{b=1}^{B}$ with a decreasing number of rows as the batch index b increases to B. Optimizing the number of rows across the B batches is beyond the scope of this dissertation. In our work, we assume that the same number of measurements is acquired in any batch, although our method can also be applied to a more general setting with different numbers of measurements per batch.

In this work, we assume that the channel remains the same during the B training batch transmissions. A potential future research direction is to extend the proposed method to the problem of channel tracking [126, 127], where the channel also changes across batches. In this case, the beliefs about the magnitude, support, and phase of the tracked channel from the previous batch can be used as side information to estimate the channel in the current batch. The channel tracking problem, however, is beyond the scope of this dissertation.

3.6. Conclusions

In this chapter, we developed a message-passing-based technique that enables sparse channel estimation under partially coherent Wiener phase noise. Our approach absorbs the phase errors into the sparse channel to define a collection of phase-perturbed auxiliary sparse vectors. The proposed method operates in two stages. The first stage estimates the auxiliary vectors by exploiting their common support and magnitude structure, Wiener phase noise statistics, and sparsity. To this end, we developed a message passing-based approach to leverage these properties at inference. The second stage in our method uses alternating optimization to reconstruct the sparse channel from the auxiliary vector estimates. Numerical results demonstrate that our method achieves a lower normalized mean-squared error than comparable benchmarks.

3.7. Appendix

3.7.1. Derivation of $\bar{v}_{f_b[n]\to \mathrm{r}[n]}$ and its mean and variance

We first provide a closed form approximation for $\bar{v}_{f_b[n]\to \mathrm{r}[n]}$ assuming $v_{\Theta_b[n]\to f_b[n]} = 1/2\pi$, which is exact for $b = 1$. Then, we compute the mean and variance of $\bar{v}_{f_b[n]\to \mathrm{r}[n]}$ using this approximation. To this end, we replace the complex Gaussian PDFs in (3.22) with their explicit form to obtain

$$\begin{aligned}
&\bar{v}_{f_b[n]\to \mathrm{r}[n]} = \\
&\frac{\epsilon^2\left(1-\Omega\left(\tilde{\pi}_b[n]\right)\right)}{\pi c_b[n]}\exp\left(-\frac{|r[n]|^2+\epsilon^{-2}|e_b[n]|^2}{\epsilon^{-2}c_b[n]}\right)\\
&\quad\frac{1}{2\pi}\int_{\theta_b[n]}\exp\left(\Re\left(\frac{2r[n]e_b[n]}{\epsilon^{-1}c_b[n]}e^{-\mathrm{j}\theta_b[n]}\right)\right)\\
&+\frac{\Omega\left(\tilde{\pi}_b[n]\right)}{\pi c_b[n]}\exp\left(-\frac{|r[n]|^2+\epsilon^{-2}|e_b[n]|^2}{\epsilon^{-2}c_b[n]}\right)\\
&\quad\frac{1}{2\pi}\int_{\theta_b[n]}\exp\left(\Re\left(\frac{2r[n]e_b[n]}{c_b[n]}e^{-\mathrm{j}\theta_b[n]}\right)\right).
\end{aligned} \tag{3.46}$$

To compute the above integral, we use the identity

$$\int_{-\pi}^{\pi}\exp\left(\Re\left(\kappa e^{-\mathrm{j}\theta}\right)\right)d\theta = 2\pi \mathrm{I}_0(|\kappa|), \tag{3.47}$$

where $\mathrm{I}_n(\cdot)$ is the modified Bessel function of the first kind of n^{th} order. The identity in (3.47) holds for any constant $\kappa \in \mathbb{C}$ [128], with $\Re(\kappa)$ denoting the real part of κ. This results in

$$\begin{aligned}
&\bar{v}_{f_b[n]\to \mathrm{r}[n]} \approx \\
&\frac{1-\Omega\left(\tilde{\pi}_b[n]\right)}{\pi\epsilon^{-2}c_b[n]}\exp\left(-\frac{r^2[n]+\epsilon^{-2}|e_b[n]|^2}{\epsilon^{-2}c_b[n]}\right)\mathrm{I}_0\left(\frac{2\,|e_b[n]|}{\epsilon^{-1}c_b[n]}r[n]\right)\\
&+\frac{\Omega\left(\tilde{\pi}_b[n]\right)}{\pi c_b[n]}\exp\left(-\frac{r^2[n]+|e_b[n]|^2}{c_b[n]}\right)\mathrm{I}_0\left(\frac{2\,|e_b[n]|}{c_b[n]}r[n]\right).
\end{aligned} \tag{3.48}$$

We use $t_b[n]$ to denote the mean and $v_b[n]$ to denote the variance of $\bar{v}_{f_b[n]\to \mathrm{r}[n]}$ in (3.48), which are given by

$$t_b[n] = \frac{\int_0^\infty r[n]\bar{v}_{f_b[n]\to \mathrm{r}[n]}dr[n]}{\int_0^\infty \bar{v}_{f_b[n]\to \mathrm{r}[n]}dr[n]}, \tag{3.49}$$

$$v_b[n] = \frac{\int_0^\infty r^2[n]\bar{v}_{f_b[n]\to \mathrm{r}[n]}dr[n]}{\int_0^\infty \bar{v}_{f_b[n]\to \mathrm{r}[n]}dr[n]} - t_b^2[n]. \tag{3.50}$$

To derive closed-form expressions for $t_b[n]$ and $v_b[n]$ in (3.49) and (3.50), we define $\varrho_b[n] = |e_b[n]|^2 / 2c_b[n]$ and use [129, Equation 10.43.24] and (3.48) to obtain

$$\int_0^\infty \bar{\nu}_{f_b[n]\to r[n]} dr_b[n] \approx \frac{\epsilon\left(1-\Omega\left(\bar{\pi}_b[n]\right)\right)+\Omega\left(\bar{\pi}_b[n]\right)}{2\sqrt{\pi c_b[n]}} \exp(-\varrho_b[n])\, I_0(\varrho_b[n]). \tag{3.51}$$

We simpify the numerator of $t_b[n]$ in (3.49) using the Rice distribution [122] defined in (3.31) and (3.48) as

$$r_b[n]\bar{\nu}_{f_b[n]\to r[n]} \approx \frac{1-\Omega\left(\bar{\pi}_b[n]\right)}{2\pi}\mathrm{Rice}\left(r[n];\epsilon^{-1}|e_b[n]|, \frac{1}{2}\epsilon^{-2}c_b[n]\right) + \frac{\Omega\left(\bar{\pi}_b[n]\right)}{2\pi}\mathrm{Rice}\left(r[n];|e_b[n]|, \frac{1}{2}c_b[n]\right). \tag{3.52}$$

By using (3.52), the numerator of $t_b[n]$ can be expressed as

$$\int_0^\infty r[n]\bar{\nu}_{f_b[n]\to r[n]} dr[n] \approx \frac{1}{2\pi}. \tag{3.53}$$

By substituting (3.53) and (3.51) in (3.49), we obtain the closed-form expression (3.23) for $t_b[n]$.

Next, we use (3.52) to rewrite the numerator in (3.50) as

$$\int_0^\infty r_b^2[n]\bar{\nu}_{f_b[n]\to r[n]} dr[n] \approx \frac{1-\Omega\left(\bar{\pi}_b[n]\right)}{2\pi}\int_0^\infty r[n]\mathrm{Rice}\left(r[n];\epsilon^{-1}|e_b[n]|, \frac{1}{2}\epsilon^{-2}c_b[n]\right)dr[n] + \frac{\Omega\left(\bar{\pi}_b[n]\right)}{2\pi}\int_0^\infty r[n]\mathrm{Rice}\left(r[n];|e_b[n]|, \frac{1}{2}c_b[n]\right)dr[n]. \tag{3.54}$$

As the integrands in (3.54) are the Rice distribution means,

$$\int_0^\infty r_b^2[n]\bar{\nu}_{f_b[n]\to r[n]} dr[n] \approx \frac{\epsilon^{-1}\left(1-\Omega\left(\bar{\pi}_b[n]\right)\right)+\Omega\left(\bar{\pi}_b[n]\right)}{4}\sqrt{\frac{c_b[n]}{\pi}}\, L_{1/2}\left(-\frac{|e_b[n]|^2}{c_b[n]}\right). \tag{3.55}$$

where $L_{1/2}(\cdot)$ denotes a Laguerre polynomial given by

$$L_{1/2}(\alpha) = \exp\left(\frac{\alpha}{2}\right)\left[(1-\alpha)I_0\left(-\frac{\alpha}{2}\right) - \alpha I_1\left(-\frac{\alpha}{2}\right)\right]. \tag{3.56}$$

By using (3.56), (3.55) and (3.51) in (3.50), we obtain the closed-form expression (3.24) for $v_b[n]$.

3.7.2. Derivation of $\vec{\pi}_b[n]$

We first compute the mean and the variance of $v_{\mathrm{r}[n]\to f_b[n]}$, which are then used to obtain $\rho_b[n]$ and $\tau_b[n]$ of the Rice distribution in (3.31). We define $\varsigma_b[n] = \eta_b[n]/\sqrt{2\gamma_b[n]}$, use $\iota_b[n]$ to denote the mean of $v_{\mathrm{r}[n]\to f_b[n]}$ and $\varkappa_b[n]$ to denote its variance. We have

$$\iota_b[n] = \frac{\int_0^\infty r[n] v_{\mathrm{r}[n]\to f_b[n]}}{\int_0^\infty v_{\mathrm{r}[n]\to f_b[n]}}, \quad \varkappa_b[n] = \frac{\int_0^\infty r^2[n] v_{\mathrm{r}[n]\to f_b[n]}}{\int_0^\infty v_{\mathrm{r}[n]\to f_b[n]}} - \iota_b^2[n]. \tag{3.57}$$

To derive closed-form expressions for (3.57), we denote the error function as $\mathrm{erf}(a) = (2/\sqrt{\pi})\int_0^a \exp(-t^2)\,dt$. Now, by substituting $v_{\mathrm{r}[n]\to f_b[n]}$ with its approximation from (3.27) and applying integral by parts, we obtain

$$\int_0^\infty v_{\mathrm{r}[n]\to f_b[n]} dr[n] \overset{(3.27)}{\approx} \frac{\eta_b[n]}{2}\left[1 + \mathrm{erf}\left(\frac{\eta_b[n]}{\sqrt{2\gamma_b[n]}}\right)\right] + \sqrt{\frac{\gamma_b[n]}{2\pi}}\exp\left(-\frac{\eta_b^2[n]}{2\gamma_b[n]}\right), \tag{3.58}$$

$$\int_0^\infty r[n] v_{\mathrm{r}[n]\to f_b[n]} dr[n] \overset{(3.27)}{\approx} \frac{\eta_b^2[n] + \gamma_b[n]}{2}\left[1 + \mathrm{erf}\left(\frac{\eta_b[n]}{\sqrt{2\gamma_b[n]}}\right)\right] + \eta_b[n]\sqrt{\frac{\gamma_b[n]}{2\pi}}\exp\left(-\frac{\eta_b^2[n]}{2\gamma_b[n]}\right), \tag{3.59}$$

$$\int_0^\infty r^2[n] v_{\mathrm{r}[n]\to f_b[n]} dr[n] \overset{(3.27)}{\approx} \frac{\eta_b[n](\eta_b^2[n] + 3\gamma_b[n])}{2}\left[1 + \mathrm{erf}\left(\frac{\eta_b[n]}{\sqrt{2\gamma_b[n]}}\right)\right] + (7\eta_b^2[n] + 2\gamma_b[n])\sqrt{\frac{\gamma_b[n]}{2\pi}}\exp\left(-\frac{\eta_b^2[n]}{2\gamma_b[n]}\right). \tag{3.60}$$

Substituting (3.58), (3.59), and (3.60) in (3.57), we have

$$\begin{aligned}\iota_b[n] &= \frac{\int_0^\infty r[n] v_{\mathrm{r}[n]\to f_b[n]}}{\int_0^\infty v_{\mathrm{r}[n]\to f_b[n]}} \overset{(3.27)}{\approx} \eta_b[n] \\ &+ \frac{\gamma_b[n]\left(1+\operatorname{erf}\left(\varsigma_b[n]\right)\right)}{\eta_b[n]\left(1+\operatorname{erf}\left(\varsigma_b[n]\right)\right)+\sqrt{\frac{2\gamma_b[n]}{\pi}}\exp\left(-\varsigma_b^2[n]\right)},\end{aligned} \tag{3.61}$$

$$\begin{aligned}\varkappa_b[n] &= \frac{\int_0^\infty r^2[n] v_{\mathrm{r}[n]\to f_b[n]}}{\int_0^\infty v_{\mathrm{r}[n]\to f_b[n]}} - \iota_b^2[n] \overset{(3.27)}{\approx} \eta_b^2[n] + 2\gamma_b[n] - \iota_b^2[n] \\ &+ \gamma_b[n]\eta_b[n]\frac{1+\operatorname{erf}\left(\varsigma_b[n]\right)+\frac{12}{\sqrt{\pi}}\varsigma_b[n]\exp\left(-\varsigma_b^2[n]\right)}{\eta_b[n]\left(1+\operatorname{erf}(\varsigma_b[n])\right)+\sqrt{\frac{2\gamma_b[n]}{\pi}}\exp\left(-\varsigma_b^2[n]\right)},\end{aligned} \tag{3.62}$$

Now, we use the RiceFit function in MATLAB [130] to find $\rho_b[n]$ and $\tau_b[n]$ of the Rice distribution in (3.31). These parameters are tuned such that the Rice distribution's mean is equal to $\iota_b[n]$ (3.61) and its variance is equal to $\varkappa_b[n]$ (3.62).

Next, we derive an approximation of $[v_{f_b[n]\to \mathrm{s}[n]}]_{\mathrm{s}[n]=1}$ assuming $v_{\Theta_b[n]\to f_b[n]} = 1/2\pi$. To this end, we use (3.47) to compute the integral with respect to $\theta_b[n]$ in (3.29) when $s[n] = 1$. This results in

$$\begin{aligned}&[v_{f_b[n]\to \mathrm{s}[n]}]_{\mathrm{s}[n]=1} \overset{(3.29)}{=} \\ &\frac{1}{\pi c_b[n]}\int_{r[n]}\exp\left(-\frac{r^2[n]+|e_b[n]|^2}{c_b[n]}\right)\mathrm{I}_0\left(\frac{2\,|e_b[n]|}{c_b[n]}r[n]\right)v_{\mathrm{r}[n]\to f_b[n]}.\end{aligned} \tag{3.63}$$

By substituting $v_{\mathrm{r}[n]\to f_b[n]}$ in (3.29) with the Rice distribution in (3.31) and using [129, Equation 10.43.28] in computing the integral with respect to $r[n]$, we have

$$\begin{aligned}&[v_{f_b[n]\to \mathrm{s}[n]}]_{\mathrm{s}[n]=1} \\ &\propto \frac{1}{2\pi\,|e_b[n]|}\mathrm{Rice}\left(|e_b[n]|\,;\,\rho_b[n],\frac{2\tau_b[n]+c_b[n]}{2}\right).\end{aligned} \tag{3.64}$$

Similarly, $[v_{f_b[n]\to \mathrm{s}[n]}]_{\mathrm{s}[n]=0}$ is

$$v_{f_b[n]\to \mathrm{s}[n]}\big|_{\mathrm{s}[n]=0} \propto \mathscr{CN}(0; e_b[n], c_b[n])\,. \tag{3.65}$$

By substituting (3.64) and (3.65) in (3.30), we can derive (3.32).

3.7.3. Derivation of $K_b[n]$ in (3.42) and the angles in (3.43)

In this section, we find the number of Gaussian components for the GM model (3.39) such that the ℓ_2-norm distance between the means of any two neighboring components is at least $2\sqrt{\xi_b[n]/2}$ as shown in Fig. 3.12. We place the components with means

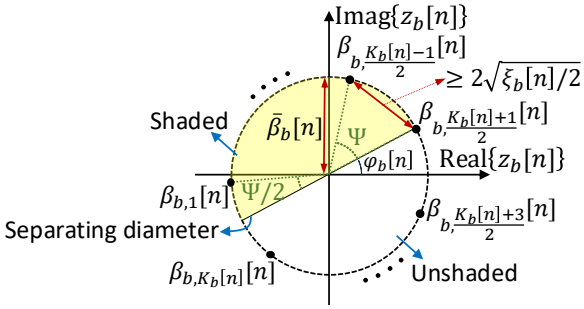


Figure 3.12: We use mixture of $K_b[n]$ complex Gaussian PDFs to approximate $v_{f_b[n]\to z_b[n]}$. We assume that these Gaussian PDFs share a common variance, denoted by $\xi_b[n]$, and that their means have the same magnitude, denoted by $\bar{\beta}_b[n]$.

$\{\beta_{b,k}\}_{b=1}^{(K_b[n]-1)/2}$ in one half of the circle indicated by the shaded region. Also, components with means $\{\beta_{b,k}\}_{b=(K_b[n]-1)/2}^{K_b[n]}$ are in unshaded half of the circle, and the component with mean $\beta_{b,(K_b[n]+1)/2} = \bar{\beta}_b[n]e^{j\varphi_b[n]}$ lies on a diameter of the circle that separates these two half circles. We set the means of the components in the unshaded part to be the mirror of those in the shaded part with respect to the separating diameter. Hence, it is sufficient to identify the means in the shaded part to determine the rest of the means.

Let $\Psi \in (0, \pi)$ denote the angle between the line segments connecting the center of the circle to $\beta_{b,(K_b[n]+1)/2}$ and $\beta_{b,(K_b[n]-1)/2}$ as shown in Fig. 3.12. From geometry, we can infer from Fig. 3.12 that for $\Psi \geq 2\sin^{-1}\left(\sqrt{\xi_b[n]/2}/\bar{\beta}_b[n]\right)$, we have $\|\beta_{b,(K_b[n]-1)/2} - \beta_{b,(K_b[n]+1)/2}\| \geq 2\sqrt{\xi_b[n]/2}$. To meet this condition on Ψ, we can find that the maximum number of components in the shaded half is

$$\frac{K_b[n]-1}{2} = \left\lfloor \frac{\pi - \Psi/2}{\Psi} \right\rfloor. \tag{3.66}$$

Now, (3.42) immediately follows from (3.66).

To find the phase of the means $\beta_{b,k}$s, we set $\angle\beta_{b,1}[n] = \varphi_b[n] + (\pi - \frac{\Psi}{2})$, without loss of generality. Next, we distribute the remaining components evenly by an angular distance of

$$\angle\beta_{b,k}[n] - \angle\beta_{b,k+1}[n] = \frac{2\pi - \Psi}{K_b[n]-1}, \quad k = 1, \cdots, K_b[n]-1, \tag{3.67}$$

which results in (3.43).

4

Compressed sensing-based sparse spatial channel estimation under IQ imbalance

Compressive sensing (CS) enables fast spatial channel estimation in millimeter-wave and terahertz systems by leveraging the sparsity of the channel in the angle-domain. CS measurements, however, are often distorted by in-phase and quadrature-phase (IQ) imbalance at the oscillator, leading to a model mismatch. In this chapter, we study how this mismatch impacts the channel estimated with a standard CS algorithm. Next, we develop an augmented CS model to jointly estimate the sparse channel and the IQ imbalance parameter. The sparse vector in our model comprises the channel as well as the IQ imbalance parameter. We show that this vector exhibits group sparsity, which is exploited using our custom paired-support orthogonal matching pursuit (PSOMP) algorithm. Finally, the estimate is decomposed to determine the channel and the IQ imbalance parameter. We provide support recovery guarantees for our PSOMP algorithm, highlighting the impact of IQ imbalance on channel recovery. Numerical results show that our method achieves better support recovery and lower error in the estimated channel than the baselines.

This chapter is based on my work in [131]: H. Masoumi and N. J. Myers, "Compressed sensing-based sparse spatial channel estimation under IQ imbalance," submitted to the *IEEE Transactions on Signal Processing*, May 2025.

4.1. Channel and system model

In this section, we consider a narrowband multiple-input single-output (MISO) system and discuss how IQ imbalance affects sparse channel estimation with the OMP. To this end, we employ the IQ imbalance model used in [77–80, 132]. Our method can still be extended to the wideband case, but this requires employing a blind deconvolution approach together with our method. This is beyond the scope of the dissertation, and we leave it for future work. Also, throughout this Chapter, we consider IQ imbalance only at the receiver and assume that the transmitter has no IQ imbalance. Our assumption is reasonable in settings where the transmitter is a base station that uses a precise and expensive oscillator than the mobile user. Additionally, methods based on predistortion [133] can also be employed to compensate for the transmitter's IQ imbalance.

4.1.1. Channel model

We consider a uniform linear array (ULA) with N half-wavelength spaced antenna elements at the transmitter (TX) and a single antenna receiver (RX) as shown in Fig. 4.1(a). The N dimensional narrowband MISO channel between the TX and RX is modeled as an $N \times 1$ vector $\mathbf{h}$. Let L denote the propagation rays in the environment with the ℓ^{th} ray having a complex gain of ζ_ℓ and an angle-of-departure (AoD) θ_ℓ. By defining the beamspace angle as $\omega_\ell = \pi \sin \theta_\ell$ [87] and the $N \times 1$ Vandermonde vector $\mathbf{a}_N(\omega)$ as

$$\mathbf{a}_N(\omega) = \left[1, e^{j\omega}, e^{j2\omega}, \cdots, e^{j(N-1)\omega}\right]^{\mathsf{T}}, \tag{4.1}$$

the baseband channel vector $\mathbf{h}$ is given by

$$\mathbf{h} = \sum_{\ell=1}^{L} \zeta_\ell \mathbf{a}_N(\omega_\ell). \tag{4.2}$$

For typical mmWave or THz access points, the number of antenna elements N can be in the order of hundreds to thousands [4]. Although we consider a ULA for simplicity of exposition, our method can be applied to rectangular arrays by incorporating an appropriate array response in (4.2).

At mmWave and THz frequencies, the high scattering results in channels that are approximately sparse in the angle-domain [1]. To exploit this sparsity, $\mathbf{h}$ is transformed into the beamspace (angle-domain). To this end, the Discrete Fourier Transform (DFT) of $\mathbf{h}$ is used for the beamspace representation since we consider a ULA at the TX. Let $\mathbf{U}_N$ denote the $N \times N$ unitary DFT matrix and $\mathbf{x}$ denote the beamspace representation of $\mathbf{h}$. Then, $\mathbf{x}$ and $\mathbf{h}$ are related as

$$\mathbf{h} = \mathbf{U}_N \mathbf{x}. \tag{4.3}$$

We assume that $\mathbf{x}$ is exactly sparse in our analysis, i.e., the beamspace angles are exactly aligned with any of the N directional beams from the DFT codebook. In the simulations, we use both exactly sparse channels and channels that are obtained from the NYU simulator [6], in which case $\mathbf{x}$ is only approximately sparse. The channel $\mathbf{h}$ is unknown and the goal of this chapter is to estimate $\mathbf{h}$ from its measurements acquired under an IQ imbalance at the RX.

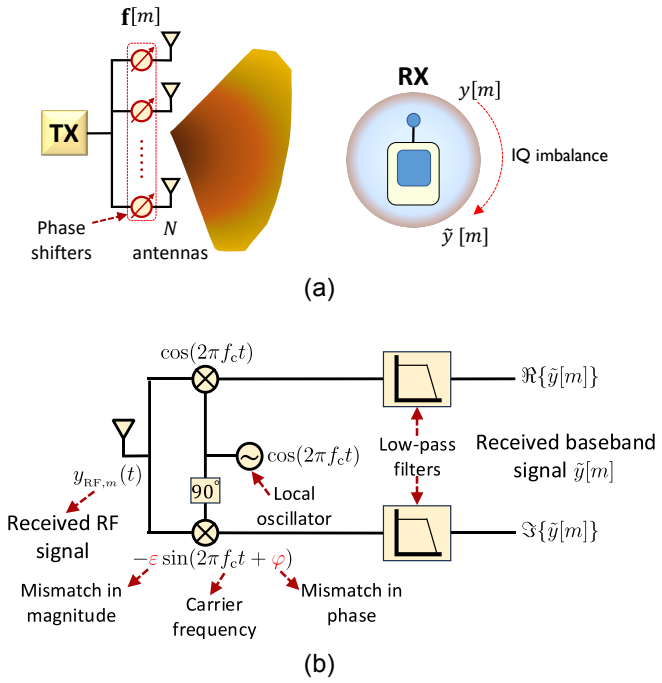


Figure 4.1: (a) A wireless system with a N element ULA at the TX and a single antenna RX. (b) Receiver architecture showing how IQ imbalance is induced in the measurements. Here, ε is the magnitude mismatch and φ is the phase mismatch between the I and the Q branches. In an ideal case (no IQ imbalance), $\varepsilon = 1$ and $\varphi = 0$.

4.1.2. System model and channel measurements

In this section, we first explain the channel measurement model in the absence of IQ imbalance. Next, we use this model to explain how the channel measurements are perturbed under an IQ imbalance at the receiver.

The TX employs a phased array in which the antennas are connected to a single RF chain through phase shifters as shown in Fig. 4.1(a). The set of possible weights at any antenna is $\left\{e^{i\phi}/\sqrt{N} : \phi \in [-\pi, \pi)\right\}$. CS-based channel estimation methods typically apply a collection of random phase shifts to acquire channel measurements. Such a choice results in quasi-omnidirectional beams [46], which are good for sparse beamspace recovery. We use $M \ll N$ to denote the total number of channel measurements. We define an $N \times 1$ vector $\mathbf{f}[m]$ comprising the N antenna weights used at the TX to acquire the m^{th} channel measurement. We use $y[m]$ to denote the m^{th} CS measurement acquired in an IQ imbalance-free RX, when the TX applies $\mathbf{f}[m]$. When there is no IQ imbalance at the RX, the received measurement in the baseband is

$$y[m] = \mathbf{f}^{\mathsf{T}}[m]\mathbf{h} + v[m], \tag{4.4}$$

where $v[m] \sim \mathcal{CN}(0, \sigma^2)$ is additive Gaussian noise.

IQ imbalance is introduced in the received measurements, due to imperfections at the mixers and filters used in downconversion [132]. The mismatch between the in-phase and quadrature branches can be modeled using two parameters in a narrowband system [19, Ch. 4], [132]. We use $\varepsilon \in [0, 1]$ to model the mismatch in magnitude and $\varphi \in [-\pi, \pi)$ to model the mismatch in phase. We use $\tilde{y}[m]$ to denote the m^{th} measurement in the baseband under the receiver IQ imbalance, as shown in Fig. 4.1. To model this imbalance, we define

$$\xi = \frac{1}{2}\left(1 + \varepsilon e^{-\mathrm{j}\varphi}\right). \tag{4.5}$$

We refer to ξ as the IQ imbalance parameter, indicating the overall imbalance due to the magnitude and phase mismatches between the IQ branches. For a receiver without IQ imbalance, $\varepsilon = 1$ and $\varphi = 0$ [132] so that $\xi = 1$. The m^{th} received measurement under an IQ imbalance is given by [132]

$$\tilde{y}[m] = \xi y[m] + (1 - \xi^*) y^*[m], \tag{4.6}$$

where $\tilde{y}[m]$ is the channel measurement obtained when the TX employs $\mathbf{f}[m]$ as the antenna weights.

The RX acquires M CS measurements $\{\tilde{y}[m]\}_{m=0}^{M-1}$ under an unknown IQ imbalance. These measurements are obtained when the TX sequentially applies M different training vectors $\{\mathbf{f}[m]\}_{m=0}^{M-1}$ at its phased array. Let $\mathbf{F}$ denote the $N \times M$ matrix containing M training vectors in its columns, $\tilde{\mathbf{y}}$ denote an $M \times 1$ vector containing the M CS measurements from (4.6), and $\mathbf{v}$ denote an $M \times 1$ vector with independent and identically distributed (IID) noise samples $\{v[m]\}_{m=0}^{M-1}$. Under an IQ imbalance at the RX, the vector of M channel measurements in (4.6) can be expressed as

$$\tilde{\mathbf{y}} = \xi \mathbf{F}^{\mathsf{T}}\mathbf{h} + (1 - \xi^*)\mathbf{F}^*\mathbf{h}^{\mathsf{c}} + \underbrace{\xi \mathbf{v} + (1 - \xi^*)\mathbf{v}^{\mathsf{c}}}_{=\mathbf{w}}. \tag{4.7}$$

The entries of the composite noise term $\{w[m]\}_{m=0}^{M-1}$ are IID distributed with a mean of zero. The real and imaginary parts of $w[m]$, however, are correlated due to the IQ imbalance as shown in Section 4.6.2. In practice, the correlation is small as ξ is close to 1 [134]. Therefore, for simplicity, we neglect this correlation and model $w[m]$ as $w[m] \sim \mathcal{CN}\left(0, \sigma_{\mathrm{w}}^2\right)$ with

$$\sigma_{\mathrm{w}}^2 = \mathbb{E}\left[|w[m]|^2\right] = \left(|\xi|^2 + |1 - \xi^*|^2\right)\sigma^2, \tag{4.8}$$

denoting the variance of $w[m]$, $\forall m \in \{0, \cdots, M-1\}$.

As observed from (4.7), the imperfect CS measurements boil down to the perfect CS measurements from (4.4), i.e., $\mathbf{y} = \mathbf{F}^{\mathsf{T}}\mathbf{h} + \mathbf{v}$, when there is no IQ imbalance, i.e., $\xi = 1$. We refer to $\mathbf{F}^{\mathsf{T}}\mathbf{h}$ as the desired part. Due to the receiver IQ imbalance, the imperfect CS measurements from (4.7) contain two important terms, i.e., $\xi\mathbf{F}^{\mathsf{T}}\mathbf{h}$ and $(1 - \xi^*)\mathbf{F}^*\mathbf{h}^{\mathsf{c}}$. The first term, $\xi\mathbf{F}^{\mathsf{T}}\mathbf{h}$, is the scaled and rotated version of the desired part by the IQ imbalance parameter ξ. The second term, $(1 - \xi^*)\mathbf{F}^*\mathbf{h}^{\mathsf{c}}$, referred to as the mismatch term,

is proportional to the conjugate of the desired part, scaled and rotated by $1 - \xi^*$. A high imbalance between the in-phase and quadrature branches results in a small ξ, leading to a large mismatch. A common measure of IQ imbalance is the IRR [19], which is a function of the IQ imbalance parameter ξ, and defined as

$$\text{IRR} = \frac{|\xi|^2}{|1 - \xi^*|^2}. \tag{4.9}$$

A small IRR means a high imbalance between the in-phase and quadrature branches, while a large IRR corresponds to a small IQ imbalance. Typical values of IRR at a carrier frequency of 60 GHz vary from about 16 dB to 38 dB [69, 70].

To exploit sparsity of the unknown channel $\mathbf{h}$ in the beamspace, we use (4.3) and (4.7) to rewrite the received channel measurements in $\tilde{\mathbf{y}}$ as

$$\tilde{\mathbf{y}} = \xi \mathbf{F}^{\mathsf{T}} \mathbf{U}_N \mathbf{x} + (1 - \xi^*) \mathbf{F}^* \mathbf{U}_N^{\mathsf{c}} \mathbf{x}^{\mathsf{c}} + \mathbf{w}. \tag{4.10}$$

We define

$$\mathbf{A} = \mathbf{F}^{\mathsf{T}} \mathbf{U}_N \tag{4.11}$$

to denote the measurement matrix in CS-based channel estimation. In a typical CS setting, the number of rows in $\mathbf{A}$, i.e., M, is smaller than the number of columns N. Using the definition of the CS matrix in (4.11), we can rewrite (4.10) as

$$\tilde{\mathbf{y}} = \xi \mathbf{A} \mathbf{x} + (1 - \xi^*) \mathbf{A}^{\mathsf{c}} \mathbf{x}^{\mathsf{c}} + \mathbf{w}. \tag{4.12}$$

The goal of this work is to estimate the sparse beamspace $\mathbf{x}$ from its measurements ($\tilde{\mathbf{y}}$ in (4.12)), which are perturbed by an unknown IQ imbalance.

To motivate this research, we study how IQ imbalance impacts sparse beamspace recovery with the standard OMP. The standard OMP assumes $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{x} + \mathbf{w}$ which is different from (4.12), thereby leading to a model mismatch. We use K to denote the sparsity level of $\mathbf{x}$, i.e., the number of its nonzero components. The K indices are chosen uniformly at random, and the corresponding non-zero coefficients are sampled from $\mathcal{CN}(0, 1)$. The normalized mean squared error (NMSE) of the channel estimate $\hat{\mathbf{x}}$, obtained with standard OMP, is given by

$$\text{NMSE} = \mathbb{E}\left[\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|^2}{\|\mathbf{x}\|^2}\right], \tag{4.13}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator. The NMSE defined in (4.13) is calculated by averaging $\|\mathbf{x} - \hat{\mathbf{x}}\|^2 / \|\mathbf{x}\|^2$, equivalently $\|\mathbf{h} - \hat{\mathbf{h}}\|^2 / \|\mathbf{h}\|^2$, over 200 independent realizations of the channel. As seen from Fig. 4.2, the OMP algorithm [8] with CS measurements under IQ imbalance results in poor channel reconstruction for small IRR, i.e., a severe IQ imbalance. This is because for small IRR, the magnitude of $(1 - \xi^*)$ is large according to (4.9), leading to a strong mismatch term $(1 - \xi^*)\mathbf{A}^{\mathsf{c}}\mathbf{x}^{\mathsf{c}}$ in (4.12). In Section 4.2, we analyze how IQ imbalance causes standard sparse recovery algorithms to fail due to model mismatch.

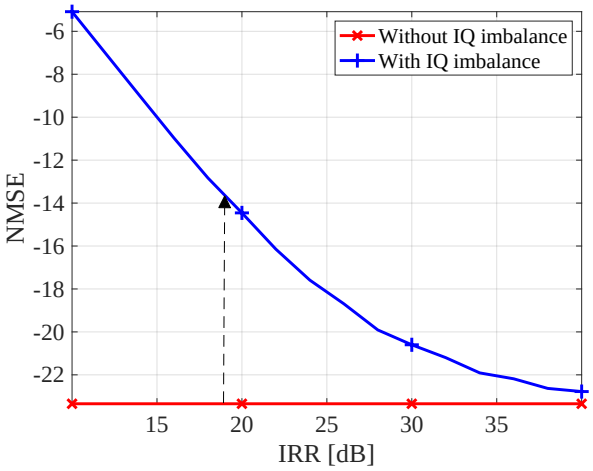


Figure 4.2: NMSE with $10\log_{10}(\text{IRR})$ where $\text{IRR} = |\xi|^2/|1-\xi^*|^2$ from (4.9) represents the level of receiver IQ imbalance. A large IRR indicates a small IQ imbalance. Here, we use $N = 256$, $M = 40$, $\sigma^2 = 1$, and random phase shifts are used at the TX.

4.2. Impact of IQ imbalance on sparse recovery

In this section, we study the impact of the IQ imbalance when convolutional CS measurement matrices [45, 135] are used for sparse recovery. We show that any CS algorithm agnostic to the IQ imbalance results in aliased beamspace components due to model mismatch. Furthermore, we show that these components are located at circular folded indices of the true beamspace locations, i.e., if a beamspace component is located at the n^{th} index, IQ imbalance leads to an aliased component at the $\langle -n \rangle_N^{\text{th}}$ index due to model mismatch.

Measurement model with convolutional CS: In convolutional CS, a measurement is acquired by circularly shifting a fixed beamformer $\mathbf{f}_{\text{b}}$ at the transmitter [136]. We use $c[m] \in [N]$ to denote the circular shift applied over $\mathbf{f}_{\text{b}}$ to acquire the m^{th} channel measurement. As $M \ll N$ in CS, the RX acquires a subsampled circular convolution of the unknown $\mathbf{h}$ and the known $\mathbf{f}_{\text{b}}$.

To model this subsampled circular convolution, we define $\mathbf{S}$ as an $M \times N$ binary matrix whose entries are one only at the coordinates $\{(m, c[m])\}_{m=0}^{M-1}$ [45]. Let $\check{\mathbf{f}}_{\text{b}}$ denote the flipped version of $\mathbf{f}_{\text{b}}$ with its n^{th} entry given by $\check{f}_{\text{b}}[n] = f_{\text{b}}[\langle -n \rangle_N]$ and $\mathbf{g}$ denote the spectral mask defined as $\mathbf{g} = \sqrt{N}\left(\mathbf{U}_N \check{\mathbf{f}}_{\text{b}}\right)^{\text{c}}$. The CS matrix from (4.11) can be rewritten as $\mathbf{A} = \mathbf{S}\mathbf{U}_N \text{diag}(\mathbf{g})$ in convolutional CS [45], where $\text{diag}(\mathbf{g})$ is a diagonal matrix with $\mathbf{g}$ on its main diagonal. Therefore, IQ imbalance free CS measurements from (4.4) in convolutional CS are given by [45]

$$\mathbf{y} = \mathbf{S}\mathbf{U}_N \text{diag}(\mathbf{g})\,\mathbf{x} + \mathbf{v}. \tag{4.14}$$

The CS matrix in (4.14) is $\mathbf{A} = \mathbf{S}\mathbf{U}_N \text{diag}(\mathbf{g})$, which is a subsampled DFT matrix with a

spectral mask $\mathbf{g}$. Using this definition of $\mathbf{A}$ in (4.12), we can express the convolutional CS measurements under an IQ imbalance as

$$\tilde{\mathbf{y}} = \xi \mathbf{S}\mathbf{U}_N \mathrm{diag}(\mathbf{g})\,\mathbf{x} + (1-\xi^*)\mathbf{S}\mathbf{U}_N^c \mathrm{diag}(\mathbf{g}^c)\,\mathbf{x}^c + \mathbf{w}. \tag{4.15}$$

The subsampled DFT structure of $\mathbf{A}$ in convolutional CS allows us to study the impact of IQ imbalance on sparse recovery.

Standard CS techniques consider a linear measurement model, which is different from our measurement model in (4.15). Such algorithms consider $\mathbf{A} = \mathbf{S}\mathbf{U}_N \mathrm{diag}(\mathbf{g})$ as the CS matrix and estimate a sparse $\tilde{\mathbf{x}}$ that satisfies

$$\tilde{\mathbf{y}} = \mathbf{S}\mathbf{U}_N \mathrm{diag}(\mathbf{g})\,\tilde{\mathbf{x}} + \mathbf{w}. \tag{4.16}$$

In Theorem 2, we show that this estimate $\tilde{\mathbf{x}}$ comprises aliased components at mirrored locations, which makes it hard to identify whether the non-zero components in $\tilde{\mathbf{x}}$ are due to aliases or a beamspace component.

Theorem 2. *Let $\check{\mathbf{x}}$ denote the flipped version of the beamspace $\mathbf{x}$, i.e., $\check{x}[n] = x[\langle -n \rangle_N]$, and $\check{\mathbf{g}}$ denote the flipped version of the spectral mask $\mathbf{g}$ with non-zero entries. We define*

$$\bar{\mathbf{x}} = \left[\mathrm{diag}(\mathbf{g})\right]^{-1} \mathrm{diag}\left(\check{\mathbf{g}}^c\right) \check{\mathbf{x}}^c. \tag{4.17}$$

Then, $\tilde{\mathbf{x}}$ which is consistent with the mismatched model in (4.16) is

$$\tilde{\mathbf{x}} = \xi \mathbf{x} + (1-\xi^*)\bar{\mathbf{x}}. \tag{4.18}$$

Proof. See Section 4.6.1. □

As seen in (4.18), IQ imbalance affects sparse recovery with standard CS algorithms in two ways: (i) scaling of the true beamspace $\mathbf{x}$ by the IQ imbalance parameter ξ, and (ii) the introduction of sparse aliased components resulting from $(1-\xi^*)\bar{\mathbf{x}}$. In Remark 1, we discuss the impact of the IQ imbalance on the support of the effective beamspace $\tilde{\mathbf{x}}$.

Remark 1. *Let $\mathcal{S}$ denote the support of a K-sparse beamspace $\mathbf{x}$, i.e., $|\mathcal{S}| = K$, $\bar{\mathcal{S}}$ denote the support of $\bar{\mathbf{x}}$ from (4.17), and $\tilde{\mathcal{S}}$ denote the support of $\tilde{\mathbf{x}}$ from (4.18). The support of the aliased sparse component vector $(1-\xi^*)\bar{\mathbf{x}}$ is*

$$\bar{\mathcal{S}} = \{\langle -n \rangle_N : n \in \mathcal{S}\}, \quad \textit{and therefore} \quad \left|\bar{\mathcal{S}}\right| = K. \tag{4.19}$$

The support of $\tilde{\mathbf{x}}$ is the union of $\mathcal{S}$ and $\bar{\mathcal{S}}$, i.e., $\tilde{\mathcal{S}} = \mathcal{S} \cup \bar{\mathcal{S}}$, with $|\tilde{\mathcal{S}}| \leq 2K$.

Proof. As the entries of $\mathbf{g}$ are non-zero, from (4.17), it is observed that $\bar{\mathbf{x}}$ and $\check{\mathbf{x}}$ have the same support. Since $\check{\mathbf{x}}$ is the flipped version of $\mathbf{x}$, i.e., $\check{x}[n] = x[\langle -n \rangle_N]$, the support of $\check{\mathbf{x}}$ and therefore $\bar{\mathbf{x}}$ is $\bar{\mathcal{S}} = \{\langle -n \rangle_N : n \in \mathcal{S}\}$. For a specific instance, the indices in $\bar{\mathcal{S}}$ are illustrated in Fig. 4.3. We observe from (4.18) that $\tilde{\mathcal{S}} = \mathcal{S} \cup \bar{\mathcal{S}}$. As each of these sets have K indices, $|\tilde{\mathcal{S}}| \leq 2K$. □

Remark 1 indicates that the beamspace reconstructed using the mismatched linear CS model comprises aliased components when $\xi \neq 1$. The aliases appear at the circular

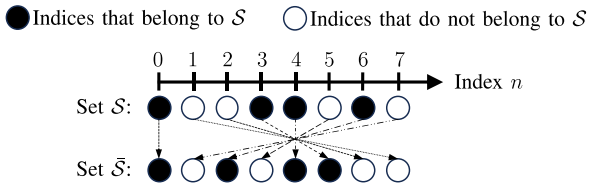


Figure 4.3: An example for the support of $\mathbf{x}$ given by $\mathcal{S} = \{0, 3, 4, 6\}$, and its circular folded version given by $\bar{\mathcal{S}} = \{0, 2, 4, 5\}$ representing the support of $\bar{\mathbf{x}}$. Here, $N = 8$, $K = 4$. In this example, we observe that $|\tilde{\mathcal{S}}| = |\mathcal{S} \cup \bar{\mathcal{S}}| = 6$.

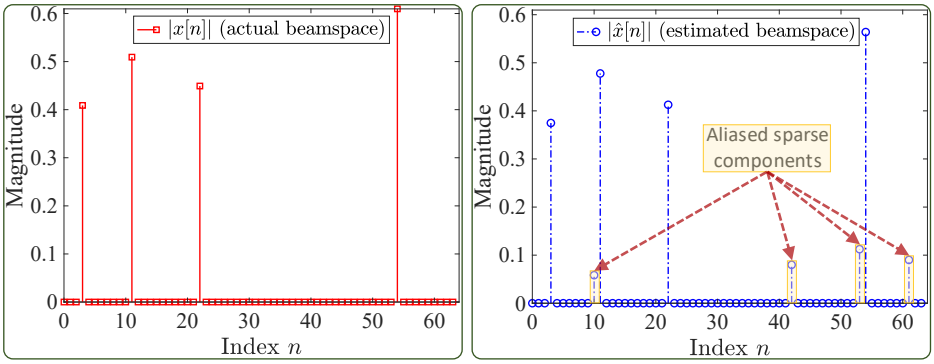


Figure 4.4: The plot shows the beamspace recovered from $M = 32$ CS measurements in (4.15), which are acquired under receiver IQ imbalance [8]. Here, a linear CS model was used for the OMP to study the impact of model mismatch. We use $N = 64$, IRR = 15 dB corresponding to $\epsilon = 0.85$ and $\varphi \approx 18^{\circ}$ from (4.5). Model mismatch due to IQ imbalance leads to aliased sparse components in the reconstructed beamspace.

folded indices corresponding to the original beamspace indices. Fig. 4.4 shows the beamspace reconstructed by applying the standard OMP over a mismatched linear CS model. Here, the support of original beamspace $\mathbf{x}$ is $\mathcal{S} = \{3, 11, 22, 54\}$. The aliased sparse components occur at the corresponding circularly folded indices in $\bar{\mathcal{S}} = \{10, 42, 53, 61\}$. Under IQ imbalance, we observe from Fig. 4.4 that standard OMP recovers a sparse signal supported on $\tilde{\mathcal{S}} = \mathcal{S} \cup \bar{\mathcal{S}}$. Also, the magnitudes of the reconstructed entries at $\mathcal{S}$ are smaller than the corresponding entries in $\mathbf{x}$, due to the scaling ξ in (4.18). Due to the model mismatch problem, it is not possible to determine whether the non-zero entries arise from aliasing or from components of the original beamspace. This non-identifiability issue prevents the application of convolutional CS for joint estimation of the sparse channel and IQ imbalance. Therefore, we consider fully random CS matrices in the subsequent sections.

4.3. Proposed sparse recovery algorithm using CS measurements corrupted by IQ imbalance

In this section, we present an equivalent formulation of the CS measurements in (4.7), called the augmented CS (AugCS) measurement model. To arrive at this formulation, we define an auxiliary sparse vector called the augmented beamspace. This vector is twice the dimension of $\mathbf{x}$, and is a function of the actual beamspace $\mathbf{x}$ and the IQ imbalance parameter ξ. We develop our greedy paired-support OMP (PSOMP) algorithm to estimate the augmented beamspace by exploiting its special structure and provide support recovery guarantees for the PSOMP. The joint estimates of the true beamspace and the IQ imbalance parameter are obtained from the estimated augmented beamspace by solving a least-squares problem.

4.3.1. Augmented CS formulation of CS under IQ imbalance

To obtain the AugCS measurement model, we construct a new $M \times 2N$ CS matrix by concatenating $\mathbf{A}$ and its conjugate $\mathbf{A}^c$. This matrix is denoted by $\mathbf{A}_{\text{aug}}$ and expressed as

$$\mathbf{A}_{\text{aug}} = \left[\mathbf{A},\ \mathbf{A}^c\right]. \tag{4.20}$$

We refer to $\mathbf{A}_{\text{aug}}$ as the augmented CS matrix. Next, we define augmented beamspace as a $2N \times 1$ vector denoted by $\mathbf{z}$ which is given by

$$\mathbf{z} = \left[\xi\mathbf{x}^{\mathsf{T}},\ (1-\xi^*)\mathbf{x}^*\right]^{\mathsf{T}}. \tag{4.21}$$

From (4.21), it is observed that if $\mathbf{x}$ is sparse, its augmented version, $\mathbf{z}$, is also sparse. Specifically, if $\mathbf{x}$ has K nonzero components, then $\mathbf{z}$ contains $2K$ nonzero components.

We now discuss the paired support structure in $\mathbf{z}$. Let $\mathcal{Q}$ denote the support of $\mathbf{z}$, i.e.,

$$\mathcal{Q} = \{i\ :\ \langle i\rangle_N \in \mathcal{S}, i \in [2N]\}. \tag{4.22}$$

We notice from (4.21) or (4.22) that the non-zero entries in $\mathbf{z}$ occur in pairs. Specifically, whenever $z[n]$ is nonzero, $z[n+N]$, $\forall n \in [N]$ is also nonzero. This structure can be equivalently expressed as

$$n \in \mathcal{Q} \iff (n+N) \in \mathcal{Q},\ \ \forall n \in [N], \tag{4.23}$$

which we call the *paired support* structure. The vector $\mathbf{z}$, whose non-zero components occur in groups, is called block-sparse [137] or group-sparse [138].

We use (4.12), (4.20), and (4.21) to obtain the equivalent augmented CS measurement model for (4.7) as

$$\tilde{\mathbf{y}} = \mathbf{A}_{\text{aug}}\mathbf{z} + \mathbf{w}. \tag{4.24}$$

It is observed that (4.24) is in the form of a standard CS measurement model as in (4.14). Hence, the standard OMP algorithm [8] with $\mathbf{A}_{\text{aug}}$ as the CS matrix can be used for the sparse recovery of $\mathbf{z}$ in (4.24). This approach, however, does not exploit the block-sparsity and paired support structure of $\mathbf{z}$. In Section 4.3.3, we explain our proposed

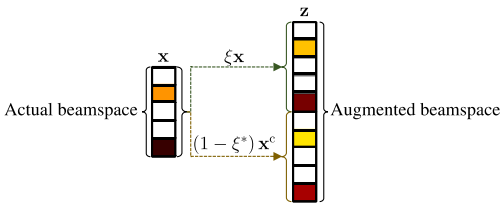


Figure 4.5: An illustration of constructing the augmented beamspace $\mathbf{z}$ from the actual beamspace $\mathbf{x}$. The augmented beamspace is block sparse, and its support has a paired support structure defined in (4.23).

PSOMP algorithm that exploits this structure for the sparse recovery of $\mathbf{z}$.

4.3.2. Proposed PSOMP algorithm

Our PSOMP algorithm exploits the paired support structure in $\mathbf{z}$ to obtain the estimate $\hat{\mathbf{z}}$ from $\tilde{\mathbf{y}}$. Our approach, outlined in Algorithm 3, is an adaptation of the OMP [8]. In the i^{th} iteration of the PSOMP Algorithm 3, we use $\mathcal{Q}^i$ to denote the estimated support, $\hat{\mathbf{z}}^i$ to denote the estimated sparse vector, and $\mathbf{r}^i$ to denote the residue. Also, we use $\hat{\mathbf{z}}^i_{\mathcal{Q}^i}$ to denote a subvector of $\hat{\mathbf{z}}^i$ indexed by $\mathcal{Q}^i$ and $\left[\mathbf{A}_{\text{aug}}\right]_{\mathcal{Q}^i}$ to denote a matrix obtained by retaining columns of $\mathbf{A}$ indexed by $\mathcal{Q}^i$. Unlike the OMP, which selects one column of $\mathbf{A}_{\text{aug}}$ in each iteration, PSOMP selects two columns. In step **2** of Algorithm 3, the second column is selected to exploit the paired support structure in $\mathbf{z}$, i.e., the structure discussed in (4.23). Then, an estimate of the entries of the sparse vector corresponding to the selected columns is obtained in step **3** and the residue is computed in step **4**. These steps are repeated until a stopping criterion is met.

Algorithm 3: Proposed PSOMP algorithm to recover $\mathbf{z}$ from our augmented CS formulation (4.24).

Input: CS measurements $\tilde{\mathbf{y}}$ and CS matrix $\mathbf{A}_{\text{aug}}$ for the augmented CS formulation (4.20) and the noise power σ^2;

Initialize: $i \leftarrow 1$, $\mathcal{Q}^0 \leftarrow \emptyset$, $\hat{\mathbf{z}}^0 \leftarrow \mathbf{0}_{2N\times 1}$ and $\mathbf{r}^0 \leftarrow \tilde{\mathbf{y}}$;

while *stopping criterion is not met* **do**

1. $v \leftarrow \underset{j\in[N]}{\text{argmax}} \left\{ \left| \mathbf{a}_j^* \mathbf{r}^{i-1} \right|, \left| \mathbf{a}_j^{\mathsf{T}} \mathbf{r}^{i-1} \right| \right\}$;
2. $\mathcal{Q}^i \leftarrow \mathcal{Q}^{i-1} \cup \{v,\ v+N\}$;
3. $\hat{\mathbf{z}}^i_{\mathcal{Q}^i} \leftarrow ([\mathbf{A}_{\text{aug}}]^*_{\mathcal{Q}^i}[\mathbf{A}_{\text{aug}}]_{\mathcal{Q}^i})^{-1}[\mathbf{A}_{\text{aug}}]^*_{\mathcal{Q}^i}\tilde{\mathbf{y}}$;
4. $\mathbf{r}^i \leftarrow \tilde{\mathbf{y}} - \mathbf{A}_{\text{aug}}\hat{\mathbf{z}}^i$;
5. $i \leftarrow i+1$;

end

Result: $\hat{\mathbf{z}} = \hat{\mathbf{z}}^i$

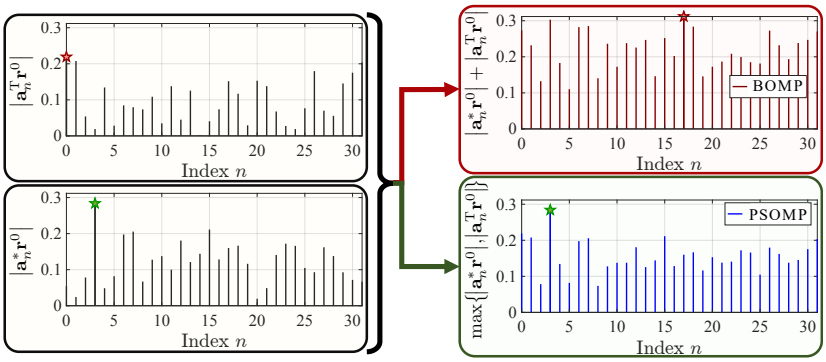


Figure 4.6: An example of support detection with PSOMP (Algorithm 3) and BOMP/GOMP for $K = 1$, $N = 32$, $M = 5$. The support of the original beamspace is $\mathcal{S} = \{3\}$ in $\mathbf{x}$. Our PSOMP algorithm successfully detects this support, while BOMP/GOMP fails.

The proposed PSOMP algorithm also differs from BOMP [137] and Group-OMP (GOMP) [138] algorithms that select a group of columns in each iteration. In particular, the BOMP [137] and GOMP [138] algorithms use

$$v_{\text{BOMP/GOMP}} \leftarrow \underset{j\in[N]}{\operatorname{argmax}} \left\{ \left| \mathbf{a}_j^* \mathbf{r}^{i-1} \right| + \left| \mathbf{a}_j^{\mathsf{T}} \mathbf{r}^{i-1} \right| \right\}, \tag{4.25}$$

as the support detection rule, which is different from the rule we use in step **1** of our PSOMP algorithm. We call $\left|\mathbf{a}_j^* \mathbf{r}^{i-1}\right|$ and $\left|\mathbf{a}_j^{\mathsf{T}} \mathbf{r}^{i-1}\right|$ matching terms as they indicate how well $\mathbf{r}^{i-1}$ is aligned with each column of $\mathbf{A}_{\text{aug}}$. In an ideal setting with zero additive noise, $M = N$, and an orthogonal CS matrix, the matching terms are non-zero only for $j \in \mathcal{S}$, thereby aiding correct support identification.

As CS matrices have fewer rows than columns, their columns are not orthogonal, resulting in non-zero matching terms for $j \notin \mathcal{S}$. We refer to these terms as matching artifacts. With our approach, the maximum of $\left|\mathbf{a}_j^* \mathbf{r}^{i-1}\right|$ and $\left|\mathbf{a}_j^{\mathsf{T}} \mathbf{r}^{i-1}\right|$ is used for support detection, while with BOMP and GOMP, their sum is used. The success of our support detection step is determined by the strength of the matching artifacts. As an example, consider a 1-sparse beamspace $\mathbf{x}$ with support $\mathcal{S} = \{3\}$ for which the matching terms are shown in Fig. 4.6. As we observe, these terms have matching artifacts outside the original support $\mathcal{S} = \{3\}$. We observe from Fig. 4.6 that the matching artifacts outside the original support are low with our method than BOMP/GOMP, thereby resulting in better support recovery. This is because the sum of the matching artifacts can result in a false alarm with BOMP/GOMP.

In Fig. 4.7, we study the empirical probability of exactly recovering the sparse support of the beamspace $\mathbf{x}$ from the CS measurements under IQ imbalance with the proposed PSOMP algorithm (Algorithm 3) and the BOMP algorithm [137]. As observed, the proposed PSOMP requires fewer measurements than the BOMP algorithm to achieve the same probability of exact support recovery [137]. For instance, to exactly recover the

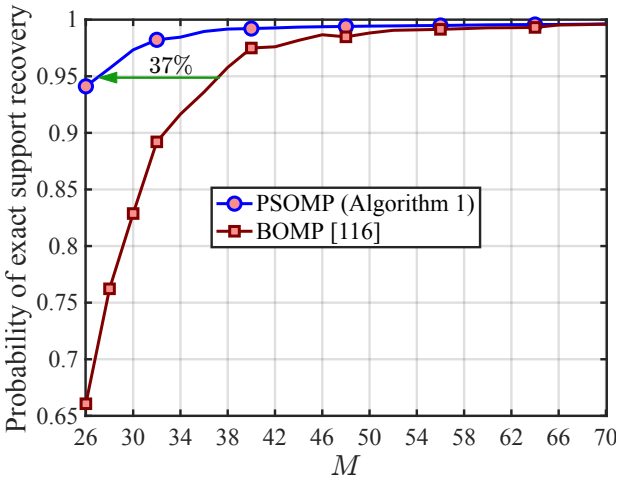


Figure 4.7: Probability of exactly recovering the support of the sparse beamspace with our PSOMP algorithm (Algorithm 3) and with the BOMP/GOMP [137, 138] algorithms. The proposed PSOMP algorithm requires fewer measurements than the BOMP/GOMP to achieve the same probability of exact support recovery. Here, we use $N = 256$, $K = 5$, SNR = 5 dB, and IRR = 20 dB.

support of $\mathbf{x}$ in 95% of cases, BOMP [137] requires approximately 37 measurements, whereas our PSOMP algorithm requires around 27 measurements. This is because the support detection rule in step **1** of our PSOMP algorithm is more robust against matching artifacts than the one in (4.25) used in the BOMP algorithm, as is seen from the example in Fig. 4.6. After obtaining $\hat{\mathbf{z}}$, the augmented beamspace estimate, our goal now is to estimate $\mathbf{x}$ and ξ from $\hat{\mathbf{z}}$.

4.3.3. Least-squares based joint sparse recovery and IQ imbalance parameter estimation

We explain how to obtain the beamspace estimate $\hat{\mathbf{x}}$ and the IQ imbalance estimate $\hat{\xi}$ from $\hat{\mathbf{z}}$. As seen from (4.21), $\hat{\mathbf{z}}$ must be of the form $\left[\xi\mathbf{x}^{\mathsf{T}},\ (1-\xi^*)\mathbf{x}^*\right]^{\mathsf{T}}$, where $\mathbf{x}$ and ξ are unknowns to be estimated. To estimate the unknowns, we minimize the squared error cost

$$\mathcal{O}(\mathbf{x},\xi) = \left\| \hat{\mathbf{z}} - \begin{bmatrix} \xi\mathbf{x} \\ (1-\xi^*)\mathbf{x}^c \end{bmatrix} \right\|^2 \tag{4.26}$$

over $(\mathbf{x},\xi)$ to obtain $(\hat{\mathbf{x}},\hat{\xi})$, i.e.,

$$\{\hat{\mathbf{x}},\hat{\xi}\} = \underset{\{\mathbf{x},\xi\}}{\operatorname{argmin}}\ \mathcal{O}(\mathbf{x},\xi)\,. \tag{4.27}$$

Due to the product of $\mathbf{x}$ and ξ, the objective function $\mathcal{O}(\mathbf{x}, \xi)$ is not convex in $(\mathbf{x}, \xi)$. In Theorem 3, we provide a solution for (4.27) in which $\left(\hat{\mathbf{x}}, \hat{\xi}\right)$ is a stationary point of $\mathcal{O}(\mathbf{x}, \xi)$.

Theorem 3. *Let $\hat{\mathbf{z}}_1$ and $\hat{\mathbf{z}}_2$ denote $N \times 1$ subvectors of $\hat{\mathbf{z}}$ such that $\hat{\mathbf{z}}_1$ contains the first N elements of $\hat{\mathbf{z}}$, and $\hat{\mathbf{z}}_2$ contains the next N elements of $\hat{\mathbf{z}}$, i.e., $\hat{\mathbf{z}} = \left[\hat{\mathbf{z}}_1^{\mathsf{T}}, \hat{\mathbf{z}}_2^{\mathsf{T}}\right]^{\mathsf{T}}$. We define the scalars*

$$\alpha = \|\hat{\mathbf{z}}_1\|^2, \quad \beta = \|\hat{\mathbf{z}}_2\|^2, \quad \gamma = \hat{\mathbf{z}}_1^{\mathsf{T}}\hat{\mathbf{z}}_2. \tag{4.28}$$

A stationary point for $\mathcal{O}(\mathbf{x}, \xi)$ is given by

$$\hat{\xi} = \frac{\alpha - \beta - 2\gamma + \sqrt{(\alpha - \beta)^2 + 4|\gamma|^2}}{2(\alpha - \beta + \gamma^* - \gamma)}, \tag{4.29}$$

$$\hat{\mathbf{x}} = \frac{\hat{\xi}^*\hat{\mathbf{z}}_1 + (1 - \hat{\xi})\hat{\mathbf{z}}_2^{\mathsf{c}}}{\left|\hat{\xi}\right|^2 + \left|1 - \hat{\xi}\right|^2}. \tag{4.30}$$

This stationary point is obtained by solving

$$\left[\nabla_{\mathbf{x}}\mathcal{O}(\mathbf{x}, \xi)\right]_{\left(\hat{\mathbf{x}}, \hat{\xi}\right)} = \mathbf{0}_{N\times 1}, \tag{4.31}$$

$$\left[\frac{\partial}{\partial \xi}\mathcal{O}(\mathbf{x}, \xi)\right]_{\left(\hat{\mathbf{x}}, \hat{\xi}\right)} = 0, \tag{4.32}$$

where $\left[\nabla_{\mathbf{x}}(\cdot)\right]_{\left(\hat{\mathbf{x}}, \hat{\xi}\right)}$ and $\left[\frac{\partial}{\partial \xi}(\cdot)\right]_{\left(\hat{\mathbf{x}}, \hat{\xi}\right)}$ denote the gradient and the partial derivative with respect to $\mathbf{x}$ and ξ evaluated at $\left(\hat{\mathbf{x}}, \hat{\xi}\right)$. Here, $\mathbf{0}_{N\times 1}$ is the $N \times 1$ vector of all zeros.

Proof. See Section 4.6.3. □

Our approach uses (4.28)-(4.30) from Theorem 3 to obtain the estimates $\hat{\mathbf{x}}$ and $\hat{\xi}$ from $\hat{\mathbf{z}}$.

4.3.4. Guarantees with our proposed PSOMP algorithm

In this section, we establish guarantees for support recovery of the augmented beamspace, $\mathbf{z}$, or equivalently, the beamspace $\mathbf{x}$, using our PSOMP algorithm. Our approach is inspired by the guarantees presented in [101] for the standard OMP [8]. We extend the analysis in [101] to our PSOMP algorithm, which also accounts for the IQ imbalance. These guarantees are based on the concept of *mutual coherence* of the CS matrix, helping us study the impact of the IQ imbalance on sparse support recovery.

In this section we assume that $\mathbf{A}$ from (4.11) has unit norm columns, i.e., $\{\|\mathbf{a}_n\| = 1\}_{n=1}^{N}$, for simplicity. As a result, the augmented CS matrix $\mathbf{A}_{\text{aug}}$ from (4.20) also has unit norm columns. Our results, however, can be extended to generic CS matrices with different column norms using our framework in [95]. The mutual coherence of $\mathbf{A}$ is an indication of its quality for sparse recovery. We use μ to denote the mutual coherence of $\mathbf{A}$ defined as

$$\mu = \max_{\{(j,\ell):j\neq\ell, j\in[N], \ell\in[N]\}} |\mathbf{a}_j^*\mathbf{a}_\ell|. \tag{4.33}$$

A CS matrix with a small μ is favorable for sparse recovery [101]. We use μ_{aug} to denote mutual coherence of the augmented CS matrix $\mathbf{A}_{\text{aug}}$. From the definition of the mutual coherence in (4.33) and from (4.20), we obtain μ_{aug} as

$$\mu_{\text{aug}} = \max_{\{(j,\ell,i):j\neq\ell,j\in[N],\ell\in[N],i\in[N]\}} \left\{ |\mathbf{a}_j^* \mathbf{a}_\ell|,\ |\mathbf{a}_j^\mathsf{T} \mathbf{a}_i| \right\}. \tag{4.34}$$

The second term $|\mathbf{a}_j^\mathsf{T} \mathbf{a}_i|$ in (4.34) accounts for the inner product of the columns of $\mathbf{A}$ with the columns of the conjugate of $\mathbf{A}$ in $\mathbf{A}_{\text{aug}}$ from (4.20). From (4.34) and (4.33), it can be shown that $\mu_{\text{aug}} \geq \mu$, where μ is the mutual coherence of $\mathbf{A}$.

In Theorem 4, we provide conditions under which the support of $\mathbf{z}$, i.e. $\mathcal{Q}$, equivalently the support of $\mathbf{x}$, i.e., $\mathcal{S}$, can be successfully detected using PSOMP and BOMP algorithms. The conditions presented in Theorem 4 establish a lower-bound on the magnitude of the weakest nonzero component in $\mathbf{x}$, denoted by $x_{\min}$:

$$x_{\min} = \min_{j\in\mathcal{S}} \left| x[j] \right|. \tag{4.35}$$

Our guarantees quantify how IQ imbalance level IRR, mutual coherence of $\mathbf{A}$ and $\mathbf{A}_{\text{aug}}$, sparsity level K of $\mathbf{x}$, and the additive noise variance σ^2 impact the support recovery of $\mathbf{x}$.

Theorem 4. *Our PSOMP algorithm in Algorithm 3 successfully recovers the support of* $\mathbf{x}$ *with a probability exceeding*

$$\left(1 - \frac{1}{(2N)^{1+\delta}\sqrt{(1+\delta)\log(2N)}}\right)^{4N} \tag{4.36}$$

for some constant $\delta > 0$ *if*

$$x_{\min} - \left(2K\left(1 + \frac{1}{\sqrt{\text{IRR}}}\right) - 1\right)\mu_{\text{aug}} x_{\min} \geq 2\bar{\rho}, \tag{4.37}$$

where $\bar{\rho} = \sigma\sqrt{2\left(1+\frac{1}{\text{IRR}}\right)(1+\delta)\log(2N)}$. *Furthermore, if*

$$x_{\min} - (4K-1)\,\mu_{\text{aug}} x_{\min} \geq 4\bar{\rho}\left(1 + \frac{1}{\sqrt{\text{IRR}}}\right)^{-1}, \tag{4.38}$$

and (4.36) are satisfied, with a probability exceeding (4.36), the BOMP algorithm [137] successfully recovers the support of $\mathbf{x}$.

Proof. See Section 4.6.4. □

The condition in (4.37) provides a lower bound on the weakest non-zero coefficient of $\mathbf{x}$ for successful support identification with PSOMP from CS measurements acquired

under IQ imbalance. By rewriting (4.37), this lower bound is

$$x_{\min} \geq \frac{2\bar{\rho}}{1 - \left(2k\left(1 + \frac{1}{\sqrt{\text{IRR}}}\right) - 1\right)\mu_{\text{aug}}}. \tag{4.39}$$

It follows from (4.39) that a severe IQ imbalance (low IRR) imposes a more stringent condition on $\mathbf{x}$, i.e., requires a higher threshold on $x_{\min}$, for successful support recovery.

Remark 2. *Under zero additive noise, i.e., $\sigma = 0$, Algorithm 3 correctly recovers the support $\mathcal{S}$ of $\mathbf{x}$ if*

$$k \leq \frac{1}{2\left(1 + \frac{1}{\sqrt{\text{IRR}}}\right)}\left(1 + \frac{1}{\mu_{\text{aug}}}\right). \tag{4.40}$$

When $\text{IRR} \to \infty$, (4.40) reduces to the one derived in [101].

Proof. The proof follows from (4.37) by setting $\sigma = 0$. □

Remark 3. *Our PSOMP algorithm enables successful support recovery with a smaller $x_{\min}$ than the one required by BOMP.*

Proof. We rewrite (4.38) as

$$x_{\min} \geq \frac{2\bar{\rho}}{\frac{1}{2}\left(1 + \frac{1}{\sqrt{\text{IRR}}}\right)\left(1 - (4K - 1)\,\mu_{\text{aug}}\right)}. \tag{4.41}$$

Since $0.5\left(1 + 1/\sqrt{\text{IRR}}\right) \leq 1$, the right-hand side of (4.39) is smaller than that of (4.41), which completes the proof. □

Remark 3 shows that our PSOMP achieves better support recovery than the BOMP algorithm [137]; this is also observed from Fig. 4.8. Fig. 4.8 shows the impact of IQ imbalance on the support recovery of $\mathbf{x}$. We observe that support recovery improves with an increasing IRR, which aligns with our findings in Theorem 4. The improvement in the probability of exact support recovery saturates for $\text{IRR} \geq 25$ as the image components in $\mathbf{z}$ from (4.21), i.e., non-zero components in $\mathbf{z}$ due to $(1 - \xi^*)\mathbf{x}^c$, become negligible. Finally, we notice from Fig. 4.6 that support detection with PSOMP is robust to matching artifacts than that of BOMP, which coincides with Remark 3.

4.4. Simulation Results

We consider a narrowband system with a ULA of 256 antenna elements at the TX for $f_c = 60$ GHz. The TX-RX distance is 15 m. The channel estimate $\hat{\mathbf{h}}$ is obtained from the beamspace estimate using $\hat{\mathbf{h}} = \mathbf{U}_N\hat{\mathbf{x}}$. We use the NMSE in (4.13) to evaluate channel estimation performance and the average relative error $\mathbb{E}\left[\left|\left(\xi - \hat{\xi}\right)/\xi\right|\right]$ to quantify the error in the estimate $\hat{\xi}$. We use Golay complementary sequence [102] of length N_{seq} for transmission, resulting in a spreading gain of $10\log N_{\text{seq}}$ [45]. Here we use, $N_{\text{seq}} = 256$.

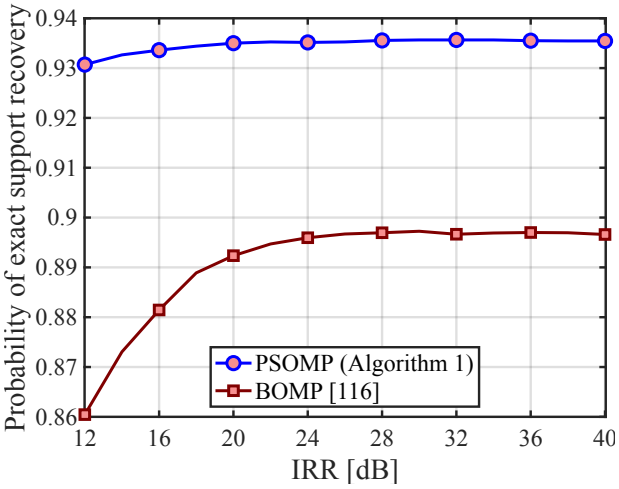


Figure 4.8: Probability of exactly recovering the support of the sparse beamspace with IRR. Our PSOMP algorithm (Algorithm 3) achieves a better support recovery than the BOMP/GOMP algorithms [137, 138]. Here, we use $N = 256$, $K = 5$, and SNR = −5 dB.

The signal-to-noise ratio (SNR) without spreading gain is given by SNR = $\|\mathbf{Ax}\|^2/M\sigma^2$. We provide numerical results for both exactly sparse beamspace channels and those with off-grid AoDs. For evaluation, we use 200 sparse beamspace channels with K nonzero components, following the same methodology described in Fig. 4.2. Also, we use 200 urban micro line-of-sight channels from the NYU simulator [6]. Note that the beamspace channels generated by the NYU simulator can have off-grid AoDs, resulting in approximately sparse beamspace channels.

We benchmark our method against four baselines. We use *StdCS ZeroIQ* for standard OMP applied to an ideal linear measurement model, i.e., $\xi = 1$, using the same CS matrix $\mathbf{A}$. The second baseline *StdCS IgnoredIQ* employs the same sparse recovery method as in *StdCS ZeroIQ*, but the measurements are corrupted by the IQ imbalance. Therefore, *StdCS IgnoredIQ* suffers from model mismatch. Finally, we use *AugCS OMP* and *AugCS BOMP* to denote the last two baselines. These two baselines serve as benchmarks for our proposed PSOMP algorithm in estimating the augmented beamspace $\mathbf{z}$ from the augmented CS measurement model. This means that both baselines, *AugCS OMP* and *AugCS BOMP*, use (4.29) and (4.30) from Theorem 3 to estimate ξ and $\mathbf{x}$ from the estimate $\hat{\mathbf{z}}$. The third baseline uses OMP to recover the augmented beamspace $\mathbf{z}$, while the fourth baseline uses BOMP from [137] to recover $\mathbf{z}$. Lastly, we use *AugCS PSOMP* to indicate our method, which uses Algorithm 3 to recover $\mathbf{z}$, and employs equations (4.29) and (4.30) from Theorem 3 to compute the estimates $\hat{\xi}$ and $\hat{\mathbf{x}}$.

The M antenna weight vectors $\{\mathbf{f}[\mathbf{m}]\}_{m=1}^{M}$ are constructed by setting their entries to IID samples drawn from $\left\{e^{\mathrm{j}\phi}/\sqrt{N} : \phi \in [-\pi, \pi)\right\}$[1]. As a result, the columns of the resulting

[1]Unlike convolutional CS that uses circularly shifted $\mathbf{f}[m]$s, the use of random $\mathbf{f}[m]$s does not lead to the

CS matrix $\mathbf{A}$ from (4.11) can have different ℓ_2 norms. To account for this variation, we scale the terms in the support detection step similar to the approach in [8]. Therefore, we use

$$v \leftarrow \underset{j\in[N]}{\operatorname{argmax}} \left\{ \frac{\left|\mathbf{a}_j^* \mathbf{r}^{i-1}\right|}{\|\mathbf{a}_j\|}, \frac{\left|\mathbf{a}_j^{\mathsf{T}} \mathbf{r}^{i-1}\right|}{\|\mathbf{a}_j\|} \right\},$$

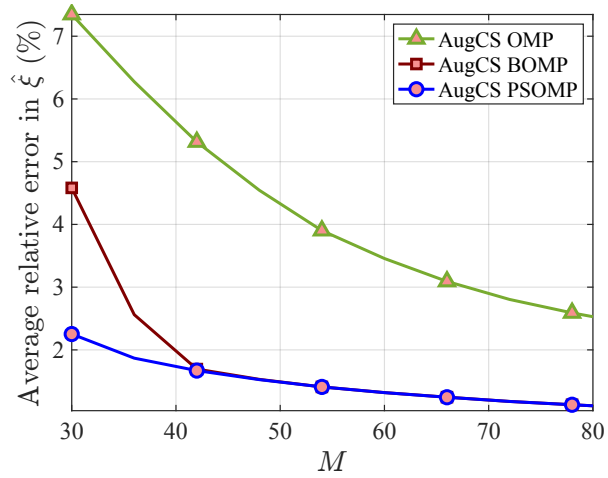
in the support detection step of Algorithm 3. In [95], we showed that normalizing the columns of $\mathbf{A}$ in the support detection of OMP results in better support detection and a low NMSE in the estimate. Using our framework in [95], a similar result can be obtained for the PSOMP algorithm.

For exactly sparse beamspace channels, we observe from Fig. 4.9(b) that employing our augmented CS formulation in (4.24) results in a lower channel estimation NMSE with OMP or PSOMP compared to *StdCS IgnoredIQ*. With the BOMP, the NMSE in $\hat{\mathbf{h}}$ is worse than that of *StdCS IgnoredIQ* for a small number of measurements, but the NMSE with the BOMP approaches that of our proposed PSOMP algorithm when the number of measurements is large. This is because the support detection step of the BOMP algorithm given in (4.25) suffers from higher matching artifacts than the support detection step in the proposed PSOMP algorithm, as discussed in Fig. 4.6. Therefore, support detection with the BOMP is poor when the number of measurements is small. The proposed *AugCS PSOMP* outperforms *AugCS OMP* in the NMSE plot because the proposed PSOMP exploits the paired support structure of $\mathbf{z}$ from (4.23). From Fig. 4.9(a), we observe that the proposed *AugCS PSOMP* method results in a small error in the estimated $\hat{\xi}$.

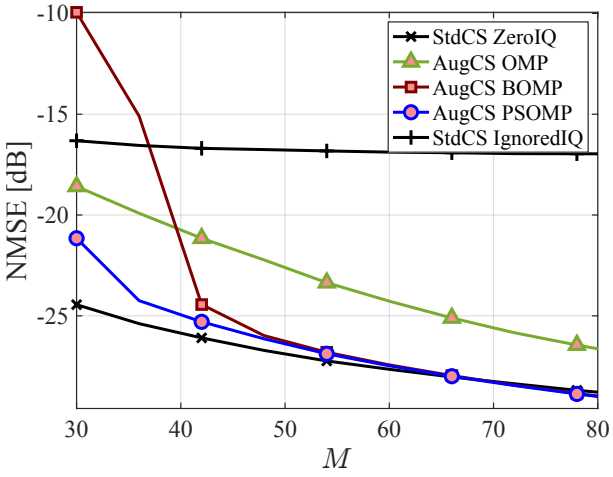
For channels from the NYU simulator [6], we observe from Fig. 4.10(b) that the proposed *AugCS PSOMP* results in the lowest NMSE under IQ imbalance when 4× overcomplete DFT dictionary is used for the beamspace representation. This is because an overcomplete dictionary uses finer grid angles than the standard DFT dictionary to represent beamspace, reducing leakage due to off-grid AoDs. With the standard DFT dictionary, high leakage from off-grid AoDs results in a high NMSE with all methods compared to the case with the 4× overcomplete DFT dictionary. The leakage deteriorates support detection with all the methods. Since PSOMP or BOMP select two elements per iteration, a misdetection in support results in two incorrectly selected elements, leading to a higher NMSE than the OMP. Fig. 4.10(a) shows that the proposed *AugCS PSOMP* method results in a low estimation error in $\hat{\xi}$ as it exploits the structure in support of $\mathbf{z}$ and employs a support detection rule that is more robust against matching artifacts than the one used in the BOMP.

From Fig. 4.11, we observe that NMSE decreases with increasing IRR, i.e., decreasing IQ imbalance, for all methods. NMSE with the proposed *AugCS PSOMP* and the *AugCS BOMP* saturate at values different from the NMSE achieved with *StdCS ZeroIQ*. This is because IQ imbalance at large IRR, i.e., $\xi \sim 1$, is negligible, due to which the paired-support structure within $\mathbf{z}$ in (4.21) is lost. When a 4×overcomplete DFT dictionary is used, our *AugCS PSOMP* results in the lowest NMSE among all methods for IRR < 30 dB. With the standard DFT dictionary, however, BOMP and PSOMP perform worse than

non-identifiability issue.



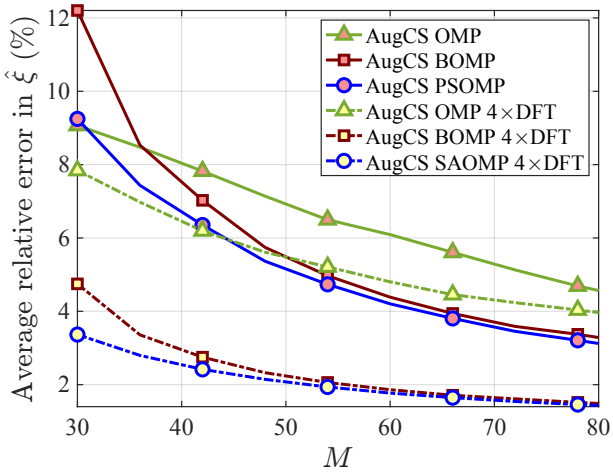
(a) Average relative error in $\hat{\xi}$



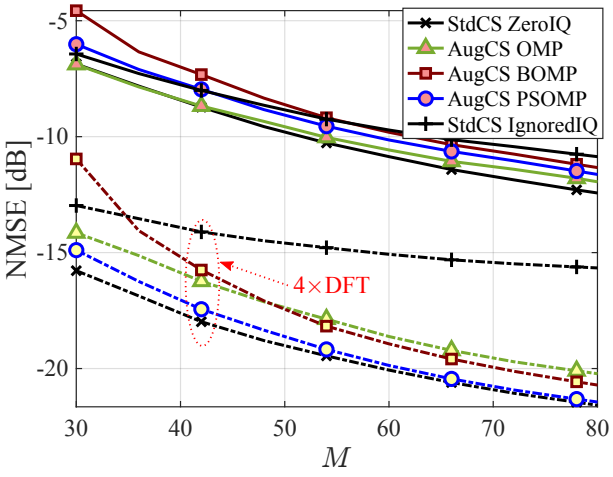
(b) NMSE in the estimate $\hat{\mathbf{h}}$

Figure 4.9: Estimation error in $\hat{\xi}$ and $\hat{\mathbf{h}}$ with number of measurements M for exactly sparse channels with $K = 5$. Sparse recovery with the proposed augmented CS formulation and the proposed PSOMP algorithm (Algorithm 3) achieves a lower estimation error in $\hat{\mathbf{h}}$ and $\hat{\xi}$ than with the OMP and BOMP [137] algorithms. Here, we use $N = 256$, $K = 5$, IRR = 20 dB, and SNR = −5 dB.

OMP in NMSE due to leakage effects from off-grid AoDs. The leakage causes false alarms in support detection due to dense matching terms. Since BOMP and PSOMP select two support elements in each iteration, false alarms in support detection with BOMP and PSOMP are significant than that of OMP, resulting in a worse NMSE with BOMP and PSOMP. Another observation is that with the standard DFT dictionary, the



(a) Average relative error in $\hat{\xi}$



(b) NMSE in the estimate $\hat{\mathbf{h}}$

Figure 4.10: Estimation error in $\hat{\xi}$ and $\hat{\mathbf{h}}$ with number of measurements M for channels from NYU simulator [6] whose beamspace contain off-grid AoDs. Sparse recovery with the proposed AugCS formulation and the proposed PSOMP algorithm (Algorithm 3) under IQ imbalance achieves the lowest estimation error in $\hat{\mathbf{h}}$ when 4× overcomplete DFT dictionary is used for the beamspace representation. Here, we use $N = 256$, IRR = 20 dB, and SNR = −5 dB.

NMSE with *AugCS OMP* is lower than that with *StdCS ZeroIQ* at a high IRR. This is because *AugCS OMP* is based on the augmented CS formulation, which uses a larger CS matrix $\mathbf{A}_{\text{aug}}$ than *StdCS ZeroIQ* that employs just $\mathbf{A}$.

In Fig.4.12, we observe that NMSE decreases with an increasing SNR. The proposed

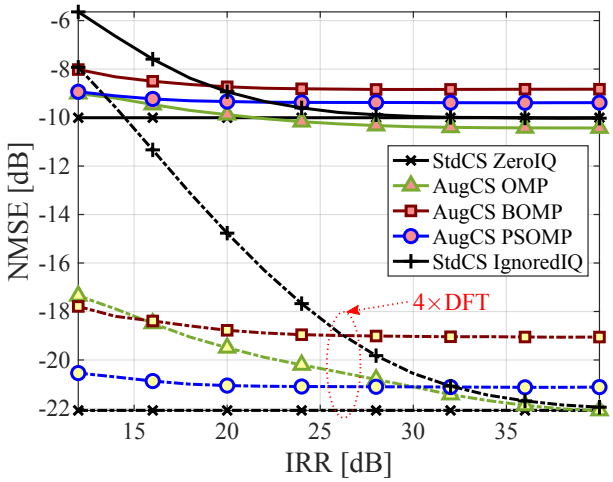


Figure 4.11: NMSE with IRR for channels from NYU simulator [6], M = 50 and SNR = 0 dB. The proposed *AugCS PSOMP* results in the lowest NMSE among different methods under IQ imbalance when the IQ imbalance is high, i.e., a small IRR, and 4× overcomplete DFT dictionary is used.

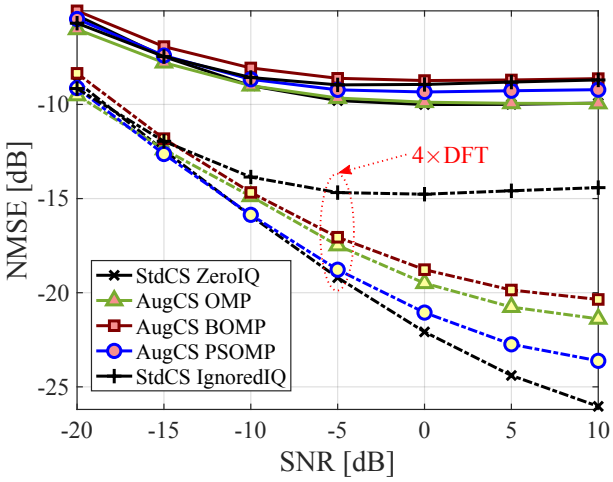


Figure 4.12: NMSE with SNR for channels from NYU simulator [6], M = 50 and IRR = 20 dB. The proposed *AugCS PSOMP* results in the lowest NMSE with 4× overcomplete DFT dictionary among different methods under IQ imbalance.

AugCS PSOMP results in the lowest NMSE among other methods under IQ imbalance with 4× overcomplete DFT dictionary. At low SNRs, the performance is noise-limited, and therefore, the NMSE with all the methods is similar.

4.5. Conclusions

In this chapter, we characterized the impact of receiver IQ imbalance on the recovered beamspace channel with standard sparse recovery algorithms that are agnostic to this mismatch. Our analysis showed that with convolutional CS, IQ imbalance leads to aliased beamspace components at circular folded locations of the true beamspace locations. To jointly recover the IQ imbalance parameter and the sparse beamspace, we introduced the augmented CS formulation that solves for a block-sparse vector exhibiting a paired support structure. We developed the PSOMP algorithm to exploit this structure and estimate the augmented beamspace. The augmented beamspace was then decomposed to obtain an estimate of the IQ imbalance parameter and the beamspace. For our PSOMP algorithm, we also provided support recovery guarantee as a function of the IQ imbalance parameter.

4.6. Appendix

4.6.1. Proof of Theorem (2)

We factor out $\mathbf{S}\mathbf{U}_N\mathrm{diag}(\mathbf{g})$ in (4.15) to arrive at the standard CS measurement model (4.16) and obtain $\tilde{\mathbf{x}}$ as

$$\tilde{\mathbf{x}} = \xi\mathbf{x} + (1-\xi^*)\left[\mathrm{diag}(\mathbf{g})\right]^{-1}\mathbf{U}_N^c\mathbf{U}_N^c\mathrm{diag}(\mathbf{g}^c)\,\mathbf{x}^c. \tag{4.42}$$

We define $\mathbf{r} = \mathbf{U}_N\mathrm{diag}(\mathbf{g})\,\mathbf{x}$ and use $\mathbf{s}$ to denote the inverse DFT of $\mathbf{r}$, i.e., $\mathbf{s} = \mathrm{diag}(\mathbf{g})\,\mathbf{x} = \mathbf{U}_N^c\mathbf{r}$, where we used $\mathbf{U}_N^c\mathbf{U}_N = \mathbf{I}_N$ in $\mathbf{U}_N^c\mathbf{r}$ to simplify the result with $\mathbf{I}_N$ denoting an $N \times N$ identity matrix. From the property of DFT that the inverse DFT of the conjugate of $\mathbf{r}$ is equal to the flipped and conjugated version of its inverse DFT given by $\mathbf{s}$ [94], we have $\check{\mathbf{s}}^c = \mathrm{diag}\left(\check{\mathbf{g}}\right)^c\check{\mathbf{x}}^c = \mathbf{U}_N^c\mathbf{r}^c$ where $\check{s}[n] = s[\langle -n\rangle_N]$ defines the flipped version of $\mathbf{s}$. Therefore, we can express (4.42) as

$$\tilde{\mathbf{x}} = \xi\mathbf{x} + (1-\xi^*)\left[\mathrm{diag}(\mathbf{g})\right]^{-1}\mathrm{diag}\left(\check{\mathbf{g}}\right)^c\check{\mathbf{x}}^c, \tag{4.43}$$

which concludes the proof.

4.6.2. Statistical properties of $w[m]$ from (4.7)

We construct a 2×1 vector with the real and imaginary parts of $w[m]$ denoted by $\tilde{\mathbf{w}}_m$ as

$$\tilde{\mathbf{w}}_m = \begin{bmatrix} w_{\mathrm{r}}[m] \\ w_{\mathrm{i}}[m] \end{bmatrix} = \begin{bmatrix} v_{\mathrm{r}}[m] \\ 2\xi_{\mathrm{i}}v_{\mathrm{r}}[m] + (2\xi_{\mathrm{r}}-1)\,v_{\mathrm{i}}[m] \end{bmatrix}, \tag{4.44}$$

with the subscripts $(\cdot)_{\mathrm{r}}$ and $(\cdot)_{\mathrm{i}}$ denoting the real and imaginary parts of the complex scalars $w[m]$, $v[m]$, and ξ. Since $v[m] \sim \mathcal{CN}\left(0,\sigma^2\right)$, $\tilde{\mathbf{w}}_m$ is normally distributed with mean $\mathbb{E}\left[\tilde{\mathbf{w}}_m\right] = [0,0]^{\mathsf{T}}$ and covariance matrix

$$\mathbb{E}\left[\tilde{\mathbf{w}}_m\tilde{\mathbf{w}}_m^{\mathsf{T}}\right] = \begin{bmatrix} \frac{1}{2}\sigma^2 & \xi_{\mathrm{i}}\sigma^2 \\ \xi_{\mathrm{i}}\sigma^2 & \left(|\xi|^2 + |1-\xi^*|^2 - \frac{1}{2}\right)\sigma^2 \end{bmatrix}. \tag{4.45}$$

For ξ close to one, i.e., $\xi_r \to 1$ and $\xi_i \to 0$, the off-diagonal entries are small, meaning that the correlation between $w_r[m]$ and $w_i[m]$ is small.

4.6.3. Proof of Theorem 3: Stationary points of $\mathcal{O}(\mathbf{x}, \xi)$

We use $\hat{\mathbf{z}} = \left[\hat{\mathbf{z}}_1^T,\ \hat{\mathbf{z}}_2^T\right]^T$ to expand the objective function $\mathcal{O}(\mathbf{x}, \xi)$ as

$$\begin{aligned}\mathcal{O}(\mathbf{x}, \xi) &= \left\|\hat{\mathbf{z}}_1 - \xi\mathbf{x}\right\|^2 + \left\|\hat{\mathbf{z}}_2 - (1-\xi^*)\mathbf{x}^c\right\|^2 \\ &= \left\|\hat{\mathbf{z}}_1\right\|^2 - \xi\hat{\mathbf{z}}_1^*\mathbf{x} - \xi^*\mathbf{x}^*\hat{\mathbf{z}}_1 + |\xi|^2\|\mathbf{x}\|^2 + \left\|\hat{\mathbf{z}}_2\right\|^2 \\ &\quad - (1-\xi^*)\hat{\mathbf{z}}_2^*\mathbf{x}^c - (1-\xi)\mathbf{x}^T\hat{\mathbf{z}}_2 + |1-\xi|^2\|\mathbf{x}\|^2. \end{aligned} \tag{4.46}$$

To find the stationary point $\left(\hat{\mathbf{x}}, \hat{\xi}\right)$ of $\mathcal{O}(\mathbf{x}, \xi)$, we solve the system of equations given by

$$\begin{cases} \nabla_{\mathbf{x}}\mathcal{O}(\mathbf{x}, \xi) &= \mathbf{0}_{N\times 1}, \\ \frac{\partial}{\partial\xi}\mathcal{O}(\mathbf{x}, \xi) &= 0, \end{cases} \tag{4.47}$$

for $\mathbf{x}$ and ξ. The gradient of $\mathcal{O}(\mathbf{x}, \xi)$ from (4.46) with respect to $\mathbf{x}$ is

$$\nabla_{\mathbf{x}}\mathcal{O}(\mathbf{x}, \xi) = -\xi\hat{\mathbf{z}}_1^* + |\xi|^2\mathbf{x}^* - (1-\xi)\hat{\mathbf{z}}_2^T + |1-\xi|^2\mathbf{x}^*. \tag{4.48}$$

The solution of $\nabla_{\mathbf{x}}\mathcal{O}(\mathbf{x}, \xi) = \mathbf{0}_{N\times 1}$ for $\mathbf{x}$ is

$$\nabla_{\mathbf{x}}\mathcal{O}(\mathbf{x}, \xi) = \mathbf{0}_{N\times 1} \Longrightarrow \mathbf{x} = \frac{\xi^*\hat{\mathbf{z}}_1 + (1-\xi)\hat{\mathbf{z}}_2^c}{|\xi|^2 + |1-\xi|^2}. \tag{4.49}$$

Next, we compute the derivative of $\mathcal{O}(\mathbf{x}, \xi)$ with respect to ξ as

$$\frac{\partial}{\partial\xi}\mathcal{O}(\mathbf{x}, \xi) = -\hat{\mathbf{z}}_1^*\mathbf{x} + \xi^*\|\mathbf{x}\|^2 + \mathbf{x}^T\hat{\mathbf{z}}_2 - (1-\xi^*)\|\mathbf{x}\|^2. \tag{4.50}$$

The solution of $\partial\left(\mathcal{O}(\mathbf{x}, \xi)\right)/\partial\xi = 0$ for ξ is

$$\frac{\partial}{\partial\xi}\mathcal{O}(\mathbf{x}, \xi) = 0 \Longrightarrow \xi = 0.5 + \frac{\mathbf{x}^*\left(\hat{\mathbf{z}}_1 - \hat{\mathbf{z}}_2^c\right)}{2\|\mathbf{x}\|^2}. \tag{4.51}$$

Using the definitions of α, β, γ from (4.28) and substituting $\mathbf{x}$ in (4.51) with the expression from (4.49) , we can rewrite (4.51) as

$$\xi = 0.5 + \frac{\left(|\xi|^2 + |1-\xi|^2\right)\left(\xi(\alpha-\gamma^*) + (1-\xi^*)(\gamma-\beta)\right)}{2\left(|\xi|^2\alpha + |1-\xi|^2\beta + \xi(1-\xi)\gamma^* + \xi^*(1-\xi^*)\gamma\right)}. \tag{4.52}$$

After simplifying (4.52), we obtain an expression that is a quadratic function of ξ as

$$(\alpha - \beta + \gamma^* - \gamma)\xi^2 + (2\gamma + \beta - \alpha)\xi - \gamma = 0. \tag{4.53}$$

Let $\{\xi_1, \xi_2\}$ denote the two solutions of (4.53) given by

$$\xi_1 = \frac{\alpha - \beta - 2\gamma + \sqrt{(\alpha-\beta)^2 + 4|\gamma|^2}}{2(\alpha - \beta + \gamma^* - \gamma)}, \tag{4.54}$$

$$\xi_2 = \frac{\alpha - \beta - 2\gamma - \sqrt{(\alpha-\beta)^2 + 4|\gamma|^2}}{2(\alpha - \beta + \gamma^* - \gamma)}. \tag{4.55}$$

To find the solution that is valid, we examine ξ_1 and ξ_2 when the estimate $\hat{\mathbf{z}}$ is exact. To this end, we use $\hat{\mathbf{z}}^{\text{exact}}$ to denote such an estimate, i.e., $\hat{\mathbf{z}}^{\text{exact}} = \mathbf{z} = \left[\xi \mathbf{x}^{\mathsf{T}},\ (1-\xi^*)\mathbf{x}^*\right]^{\mathsf{T}}$ from (4.21). It can be shown that

$$[\xi_1]_{\hat{\mathbf{z}}=\hat{\mathbf{z}}^{\text{exact}}} = \xi, \tag{4.56}$$

$$[\xi_2]_{\hat{\mathbf{z}}=\hat{\mathbf{z}}^{\text{exact}}} = \frac{\xi^* - 1}{2\xi^* - 1}. \tag{4.57}$$

Since ξ_1 reduces to the exact solution for $\hat{\mathbf{z}} = \hat{\mathbf{z}}^{\text{exact}}$, we set $\hat{\xi} = \xi_1$. Finally, we substitute ξ with $\hat{\xi}$ in (4.49) to conclude.

4.6.4. Proof of Theorem 4: Support recovery guarantees for the proposed PSOMP algorithm (Algorithm 3)

The proof extends the approach used in [101] for the standard OMP to our PSOMP algorithm. First, we examine the noise term in the support detection step of PSOMP, i.e., step **1** of Algorithm 3. To this end, we define $\bar{\mathbf{z}}^i = \mathbf{z} - \hat{\mathbf{z}}^i$ to denote the difference between the original sparse vector $\mathbf{z}$ and its estimated version in the i^{th} iteration of our PSOMP algorithm. Hence, in the i^{th} iteration, the residue $\mathbf{r}^i$ is

$$\mathbf{r}^i = \mathbf{A}_{\text{aug}}\bar{\mathbf{z}}^i + \mathbf{w}, \tag{4.58}$$

and the support recovery criterion in step **1** is equivalent to

$$v = \underset{j\in[N]}{\operatorname{argmax}} \left\{ |\mathbf{a}_j^* \mathbf{A}_{\text{aug}} \bar{\mathbf{z}}^{i-1} + \mathbf{a}_j^* \mathbf{w}|, |\mathbf{a}_j^{\mathsf{T}} \mathbf{A}_{\text{aug}} \bar{\mathbf{z}}^{i-1} + \mathbf{a}_j^{\mathsf{T}} \mathbf{w}| \right\}. \tag{4.59}$$

From (4.59), we observe that the noise terms $\mathbf{a}_j^*\mathbf{w}$ and $\mathbf{a}_j^{\mathsf{T}}\mathbf{w}$ can cause incorrect support detection. To control support misdetection under noise, we first examine the event

$$E = \left\{ \max_{j\in[N]} \left\{ |\mathbf{a}_j^*\mathbf{w}|,\ |\mathbf{a}_j^{\mathsf{T}}\mathbf{w}| \right\} < \rho \right\} \tag{4.60}$$

with $\rho = \sigma_{\text{w}}\sqrt{2(1+\delta)\log(2N)}$ and $\delta > 0$ [101]. In Lemma 3, we provide a lower bound on the probability of the event E. A similar lemma was derived in [101] for a real-valued CS problem. Our result, however, applies to a more general CS setting with complex-valued entries.

Lemma 3. *The probability of the event E in (4.60) is lower bounded as*

$$\Pr\{E\} \geq \left(1 - \frac{1}{(2N)^{1+\delta}\sqrt{(1+\delta)\log(2N)}}\right)^{4N}. \tag{4.61}$$

Proof. Real and imaginary parts of $\mathbf{a}_j^*\mathbf{w}$ and $\mathbf{a}_j^\mathsf{T}\mathbf{w}$ are jointly Gaussian with zero mean. The variance of each is

We use $\Re(\mathbf{a}_j^*\mathbf{w})$ and $\Im(\mathbf{a}_j^*\mathbf{w})$ to denote the real and imaginary parts of $\mathbf{a}_j^*\mathbf{w}$. Similarly, $\Re(\mathbf{a}_j^\mathsf{T}\mathbf{w})$ and $\Im(\mathbf{a}_j^\mathsf{T}\mathbf{w})$ denote the real and imaginary parts of $\mathbf{a}_j^\mathsf{T}\mathbf{w}$. Since the entries of $\mathbf{w}$ are independently distributed as $\mathcal{CN}(0, \sigma_w^2)$ for ξ close to 1 , the random variables $\Re(\mathbf{a}_j^*\mathbf{w})$, $\Im(\mathbf{a}_j^*\mathbf{w})$, $\Re(\mathbf{a}_j^\mathsf{T}\mathbf{w})$, and $\Im(\mathbf{a}_j^\mathsf{T}\mathbf{w})$ $\forall j \in [N]$ are jointly Gaussian and each distributed as $\mathcal{N}(0, \sigma_w^2/2)$. For any $\kappa \in \mathbb{C}$, we notice that $|\kappa| < \rho$ whenever $\Re(\kappa) < \rho/\sqrt{2}$ and $\Im(\kappa) < \rho/\sqrt{2}$. Using this observation in (4.60), we can write

$$\begin{aligned}
\Pr\{E\} \geq \Pr\Bigg\{ & \max_{j\in[N]} |\Im(\mathbf{a}_j^*\mathbf{w})| < \frac{\rho}{\sqrt{2}} \cap \max_{j\in[N]} |\Re(\mathbf{a}_j^*\mathbf{w})| < \frac{\rho}{\sqrt{2}} \\
& \cap \max_{j\in[N]} |\Im(\mathbf{a}_j^\mathsf{T}\mathbf{w})| < \frac{\rho}{\sqrt{2}} \cap \max_{j\in[N]} |\Re(\mathbf{a}_j^\mathsf{T}\mathbf{w})| < \frac{\rho}{\sqrt{2}} \Bigg\} \\
\overset{(a)}{\geq} & \prod_{j\in[N]} \Pr\left\{|\Im(\mathbf{a}_j^*\mathbf{w})| < \frac{\rho}{\sqrt{2}}\right\} \Pr\left\{|\Re(\mathbf{a}_j^*\mathbf{w})| < \frac{\rho}{\sqrt{2}}\right\} \\
& \Pr\left\{|\Im(\mathbf{a}_j^\mathsf{T}\mathbf{w})| < \frac{\rho}{\sqrt{2}}\right\} \Pr\left\{|\Re(\mathbf{a}_j^\mathsf{T}\mathbf{w})| < \frac{\rho}{\sqrt{2}}\right\}.
\end{aligned}$$

In (a), we use Šidák's lemma [113, Theorem 1]. By replacing the probabilities in the right-hand side of the above equation with their upper limits as in [101], we obtain (4.61). □

Next, we derive a condition that ensures the successful recovery of the support of $\mathbf{x}$ when the event E in (4.60) occurs. To this end, we begin with the first iteration of Algorithm 3 and analyze step **1**. Then, using induction, we show that when event E occurs and $\mathbf{x}$ satisfies (4.37), the PSOMP algorithm correctly recovers the support of $\mathbf{z}$, and consequently, the support of $\mathbf{x}$ after K iterations.

Considering the first iteration of Algorithm 3 and assuming that event E occurs, we derive conditions under which the index j of the selected column, i.e., the one that maximizes $\left\{\left|\mathbf{a}_j^*\mathbf{r}^0\right|, \left|\mathbf{a}_j^\mathsf{T}\mathbf{r}^0\right|\right\}$, belongs to the support, $\mathcal{S}$, of $\mathbf{x}$. Noting that $\mathbf{r}^0 = \mathbf{y}$, this condition can be equivalently written as

$$\max_{j\in\mathcal{S}} \left\{|\mathbf{a}_j^*\mathbf{y}|, |\mathbf{a}_j^\mathsf{T}\mathbf{y}|\right\} > \max_{j\notin\mathcal{S}} \left\{|\mathbf{a}_j^*\mathbf{y}|, |\mathbf{a}_j^\mathsf{T}\mathbf{y}|\right\}. \tag{4.62}$$

Under event E, an upper bound for the right-hand of (4.62) is

$$\max_{j\notin\mathcal{S}} \left\{|\mathbf{a}_j^*\mathbf{y}|, |\mathbf{a}_j^\mathsf{T}\mathbf{y}|\right\}$$

$$
\begin{aligned}
&= \max_{j\notin\mathcal{S}}\Bigg\{\bigg|\mathbf{a}_j^*\mathbf{w} + \sum_{i\in\mathcal{S}}\Big(\mathbf{a}_j^*\mathbf{a}_i z[i] + \mathbf{a}_j^*\mathbf{a}_i^c z[i+N]\Big)\bigg|, \\
&\qquad \bigg|\mathbf{a}_j^\mathsf{T}\mathbf{w} + \sum_{i\in\mathcal{S}}\Big(\mathbf{a}_j^\mathsf{T}\mathbf{a}_i z[i] + \mathbf{a}_j^\mathsf{T}\mathbf{a}_i^c z[i+N]\Big)\bigg|\Bigg\} \\
&\overset{(a)}{\leq} \max_{j\notin\mathcal{S}}\Bigg\{|\mathbf{a}_j^*\mathbf{w}| + \sum_{i\in\mathcal{S}}\Big(\Big|\mathbf{a}_j^*\mathbf{a}_i z[i]\Big| + \Big|\mathbf{a}_j^*\mathbf{a}_i^c z[i+N]\Big|\Big), \\
&\qquad |\mathbf{a}_j^\mathsf{T}\mathbf{w}| + \sum_{i\in\mathcal{S}}\Big(\Big|\mathbf{a}_j^\mathsf{T}\mathbf{a}_i z[i]\Big| + \Big|\mathbf{a}_j^\mathsf{T}\mathbf{a}_i^c z[i+N]\Big|\Big)\Bigg\} \\
&\overset{(b)}{\leq} \max_{j\notin\mathcal{S}}\Big\{|\mathbf{a}_j^*\mathbf{w}|,\ |\mathbf{a}_j^\mathsf{T}\mathbf{w}|\Big\} \\
&\qquad + \max_{j\notin\mathcal{S}}\Bigg\{\sum_{i\in\mathcal{S}}\Big(\Big|\mathbf{a}_j^*\mathbf{a}_i z[i]\Big| + \Big|\mathbf{a}_j^*\mathbf{a}_i^c z[i+N]\Big|\Big), \\
&\qquad \sum_{i\in\mathcal{S}}\Big(\Big|\mathbf{a}_j^\mathsf{T}\mathbf{a}_i z[i]\Big| + \Big|\mathbf{a}_j^\mathsf{T}\mathbf{a}_i^c z[i+N]\Big|\Big)\Bigg\} \\
&\overset{(c)}{\leq} \rho + k(|\xi| + |1-\xi^*|)\mu_{\text{aug}}x_{\max}, \qquad (4.63)
\end{aligned}
$$

where $x_{\max} = \max_{\{j\in\mathcal{S}\}}\big|x[j]\big|$ denotes the magnitude of the strongest component in $\mathbf{x}$. In (a), we use the triangular inequality. In (b), we use the fact that $\max\{\tau_1+\tau_2, \tau_3+\tau_4\} \leq \max\{\tau_1,\tau_3\} + \max\{\tau_2,\tau_4\}$ for real non-negative τ_1, τ_2, τ_3, and τ_4. In (c), we use the assumption that the event E occurs, and the fact that

$$
\Big|\mathbf{a}_j^\mathsf{T}\mathbf{a}_i z[i]\Big| = |z[i]||\mathbf{a}_j^\mathsf{T}\mathbf{a}_i| \overset{(4.21)}{=} |\xi x[i]||\mathbf{a}_j^\mathsf{T}\mathbf{a}_i| \overset{(4.34)}{\leq} |\xi|\mu_{\text{aug}}x_{\max}.
$$

and similarly $\Big|\mathbf{a}_j^*\mathbf{a}_i z[i]\Big| \leq |\xi|\mu_{\text{aug}}x_{\max}$, $\Big|\mathbf{a}_j^*\mathbf{a}_i^c z[i+N]\Big| \leq |1-\xi^*|\mu_{\text{aug}}x_{\max}$, and $\Big|\mathbf{a}_j^\mathsf{T}\mathbf{a}_i^c z[i+N]\Big| \leq |1-\xi^*|\mu_{\text{aug}}x_{\max}$.

Under the event E, for the left-hand side of (4.62), we have

$$
\begin{aligned}
\max_{j\in\mathcal{S}}\Big\{|\mathbf{a}_j^*\mathbf{y}|, |\mathbf{a}_j^\mathsf{T}\mathbf{y}|\Big\} = \max_{j\in\mathcal{S}}\Bigg\{&\bigg|\mathbf{a}_j^*\mathbf{a}_j z[j] + \mathbf{a}_j^*\mathbf{a}_j^c z[j+N] \\
&+ \mathbf{a}_j^*\mathbf{w} + \sum_{i\in\mathcal{S}\setminus\{j\}}\Big(\mathbf{a}_j^*\mathbf{a}_i z[i] + \mathbf{a}_j^*\mathbf{a}_i^c z[i+N]\Big)\bigg|, \\
&\bigg|\mathbf{a}_j^\mathsf{T}\mathbf{a}_j z[j] + \mathbf{a}_j^\mathsf{T}\mathbf{a}_j^c z[j+N] + \mathbf{a}_j^\mathsf{T}\mathbf{w} \\
&+ \sum_{i\in\mathcal{S}\setminus\{j\}}\Big(\mathbf{a}_j^\mathsf{T}\mathbf{a}_i z[i] + \mathbf{a}_j^\mathsf{T}\mathbf{a}_i^c z[i+N]\Big)\bigg|\Bigg\}
\end{aligned}
$$

$$\overset{(d)}{\geq} \max_{j\in\mathcal{S}} \Bigg\{ |z[j]| - |\mathbf{a}_j^* \mathbf{a}_j^c z[j+N]| - |\mathbf{a}_j^* \mathbf{w}| - \sum_{i\in\mathcal{S}\setminus\{j\}} \Big(|\mathbf{a}_j^* \mathbf{a}_i z[i]| + |\mathbf{a}_j^* \mathbf{a}_i^c z[i+N]| \Big), |z[j+N]| - \left| \mathbf{a}_j^{\mathsf{T}} \mathbf{a}_j z[j] \right| - |\mathbf{a}_j^{\mathsf{T}} \mathbf{w}| - \sum_{i\in\mathcal{S}\setminus\{j\}} \Big(|\mathbf{a}_j^{\mathsf{T}} \mathbf{a}_i z[i]| + |\mathbf{a}_j^{\mathsf{T}} \mathbf{a}_i^c z[i+N]| \Big) \Bigg\}$$

$$\overset{(e)}{>} \max_{j\in\mathcal{S}} \Bigg\{ |z[j]| - |1-\xi^*| \mu_{\text{aug}} x_{\max} - \rho - \sum_{i\in\mathcal{S}\setminus\{j\}} \Big(|\xi| \mu_{\text{aug}} x_{\max} + |1-\xi^*| \mu_{\text{aug}} x_{\max} \Big), |z[j+N]| - |\xi| \mu_{\text{aug}} x_{\max} - \rho - \sum_{i\in\mathcal{S}\setminus\{j\}} \Big(|\xi| \mu_{\text{aug}} x_{\max} + |1-\xi^*| \mu_{\text{aug}} x_{\max} \Big) \Bigg\}$$

$$\overset{(f)}{=} |\xi| x_{\max} - \rho - ((k-1)|\xi| + k|1-\xi^*|) \mu_{\text{aug}} x_{\max}. \tag{4.64}$$

In (d), we use $\|\mathbf{a}_j\| = 1$ and the triangular inequality. In (e), we use identities similar to the ones in (d) from (4.63). In (f), we use $|\xi| > |1-\xi^*|$, as IRR > 0 dB in practice, and therefore $\max_{j\in\mathcal{S}} |z[j]| = |\xi| x_{\max} \geq \max_{j\in\mathcal{S}} |z[j+N]| = |1-\xi^*| x_{\max}$. From (4.63) and (4.64), we can write

$$\max_{j\in\mathcal{S}} \Big\{ |\mathbf{a}_j^* \mathbf{y}|, |\mathbf{a}_j^{\mathsf{T}} \mathbf{y}| \Big\} > |\xi| x_{\max} - 2\rho + \max_{j\notin\mathcal{S}} \Big\{ |\mathbf{a}_j^* \mathbf{y}|, |\mathbf{a}_j^{\mathsf{T}} \mathbf{y}| \Big\} - ((2k-1)|\xi| + 2k|1-\xi^*|) \mu_{\text{aug}} x_{\max}. \tag{4.65}$$

From (4.65), we observe that under the event E and

$$|\xi| x_{\max} - ((2k-1)|\xi| + 2k|1-\xi^*|) \mu_{\text{aug}} x_{\max} \geq 2\rho, \tag{4.66}$$

equation (4.62) holds. Hence, the selected entry in step **1** of the first iteration of Algorithm 3 will belong to $\mathcal{S}$, and the selected pair in step **2** will belong to $\mathcal{Q}$. We define $\bar{\rho} = \sigma\sqrt{2\left(1+\frac{1}{\text{IRR}}\right)(1+\delta)\log(2N)}$ and divide both sides of (4.66) by $|1-\xi^*|\sqrt{\text{IRR}}$ to obtain

$$x_{\max} - \left(2k \left(1 + \frac{1}{\sqrt{\text{IRR}}} \right) - 1 \right) \mu_{\text{aug}} x_{\max} \geq 2\bar{\rho}. \tag{4.67}$$

Since (4.37) implies (4.67) and the residue is orthogonal to the selected columns, using the induction-based technique from [101, Theorem. 4], we observe that under (4.37) and the event E, Algorithm 3 will successfully recover the entire support of $\mathbf{x}$, i.e., $\mathcal{S}$, and therefore $\mathcal{Q}$ after K iterations. We follow a similar procedure to derive the condition in

(4.38) for BOMP.

5

Conclusions and Recommendations

In this dissertation, novel methods for millimeter wave channel estimation were developed that exploit their sparse structure in the angle domain, while being robust to hardware impairments and constraints. The proposed methods addressed the low SNR issue with standard sparse recovery by designing channel acquisition methods that are compatible with low-resolution phased arrays. Also, new sparse channel estimation algorithms were developed that are robust against phase noise and IQ mismatch at the receiver's front-end.

5.1. Summary

In Chapter 2 of the dissertation, we designed a novel comb-structured codebook for sector-level sweep with phased array antennas. We also developed a convolutional CS (CCS) that uses the designed comb-structured beam codebook for channel estimation within the sector of interest. Compared to conventional exhaustive beam scanning approaches, the proposed method significantly reduces training overhead by exploiting the sparse nature of mmWave channels. Importantly, the proposed approach seamlessly aligns with the beam refinement procedures defined in the IEEE 802.11ad/ay standards. A key advantage of our approach lies in addressing the poor received SNR typically observed in CS techniques that rely on wide beams. The poor SNR issue was addressed by making use of our designed comb-structured beams to focus the transmitter's energy within the sector of interest, resulting in an improved SNR in acquiring channel measurements. Furthermore, we optimized the CS matrix by carefully selecting the circular shifts used in CCS to push the aliasing artifacts outside the sector of interest, to ultimately reduce the channel estimation error within the sector of interest. In this Chapter, we also derived theoretical conditions under which the support of the beamspace channel within the sector of interest can be successfully recovered. Our simulation results confirmed that the reconstruction error remains low when the antenna weight vectors at the transmitter are such that they illuminate the sector nearly uniformly.

The CS measurements used for sparse recovery of mmWave channels are perturbed by phase noise, which is severe at higher frequencies. To address this challenge, in Chapter 3 of the dissertation, we developed a message passing-based technique for sparse channel estimation assuming a partially coherent Wiener phase noise process. This assumption is due to the observation that the phase noise variation within batches is negligible, while it changes drastically according to a Wiener process from one packet to another. Our approach absorbs the common phase error for each batch into the sparse channel to define a collection of phase-perturbed auxiliary sparse vectors corresponding to the batches. The acquired channel measurements are then linear projections of the unknown auxiliary vectors. The proposed method consists of two main stages, each carefully designed to exploit the underlying structure of the problem. In the first stage, the auxiliary vectors are estimated by leveraging their shared support and magnitude patterns, as well as the statistical properties of Wiener phase noise and the sparsity of the angle-domain channel. To this end, we developed a specialized message passing-based algorithm that incorporates these structures during inference. In the second stage, we employ an alternating optimization strategy to reconstruct the original sparse angle-domain channel and the phase errors using the auxiliary vector estimates obtained from the first stage. The results showed that our proposed two-stage method achieves a lower normalized mean-squared error than comparable benchmarks.

In Chapter 4 of the dissertation, we addressed another important RF impairment called IQ imbalance. To this end, we first characterized the impact of receiver IQ imbalance on the recovered beamspace channel with standard sparse recovery, agnostic to this mismatch. Through analytical and numerical analysis, we showed that with convolutional CS, IQ imbalance introduces aliased components into the recovered beamspace channel. Specifically, these artifacts appear at circularly folded positions relative to the true beamspace locations, which can significantly degrade channel estimation performance. To address this challenge, we proposed an augmented CS framework capable of jointly estimating both the IQ imbalance parameter and the sparse beamspace channel. This approach reformulates the problem into recovering a block-sparse vector with a paired support structure that reflects the symmetry induced due to the IQ imbalance. Building on this formulation, we developed the PSOMP algorithm, which explicitly leverages the paired support structure to estimate the augmented beamspace vector. Next, we developed a least-squares method that decomposes the recovered augmented beamspace using PSOMP to obtain an estimate of the IQ imbalance parameter and the beamspace channel. Finally, we provided support recovery guarantees for the proposed PSOMP algorithm, indicating that a high IQ mismatch severely impacts the support recovery.

5.2. Recommendations and future directions

This section outlines some of the limitations of this dissertation and highlights areas that require further investigation. It also suggests potential research directions on the topic of channel estimation for millimeter-wave and terahertz radios.

5.2.1. Addressing out-of-sector leakage with extremely low-resolution phased arrays

In Chapter 2, the constructed comb-like sectors require phase shifters whose resolution scales logarithmically with the number of sectors. In extremely large antenna arrays, a high number of sectors may be employed to achieve a high signal power within the sector. In such a case, our comb-like construction requires high-resolution phase shifters, which may not be affordable for a low hardware complexity. Applying our beams to extremely low-resolution arrays results in a new problem of out-of-sector leakage, which must be accounted for in our channel estimation technique.

The out-of-sector leakage problem is best understood when considering contiguous sectors, which are commonly used. In our work in [139], we developed a CS-based channel estimation with ideal sectors in Fig. 5.1(a), assuming phase and magnitude control at the array. Under low-resolution phase arrays, however, as we observed in Chapter 2, these contiguous sectors result in leakage outside the sector (see Fig. 5.1(b)). Therefore, it is not possible to partition the angle-domain channel into multiple non-overlapping sections with these sectors. Now, the question is how do we construct low-resolution codes that minimize the out-of-sector leakage as much as possible. The challenge here is to develop a channel estimation method that is robust against the out-of-sector leakage. Also, for a contiguous sector shown in Fig. 5.1(a), we have developed an in-sector CS algorithm in [139] that pushes the aliasing artifact outside the sector. Now, the question is how our in-sector CS algorithm in [139] can be extended to be aware of aliasing artifacts when solving this problem over multiple non-ideal sectors?

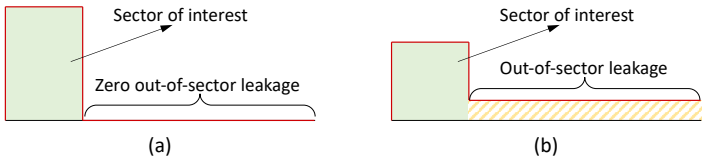


Figure 5.1: A sketch of an ideal contiguous illumination pattern in (a) with no out-of-sector leakage, and a non-ideal contiguous illumination pattern in (b) with out-of-sector leakage.

5.2.2. The case of strong phase noise

In Chapter 3 of the dissertation, it is assumed that the phase error for measurements obtained within a packet is the same, and the phase error changes from one packet to another follows a Wiener process. At high carrier frequencies such as THz, however, phase noise is more severe [17], resulting in a high variance for the Wiener phase noise process. This causes the same phase error assumption within packets to break down. Furthermore, the high variance causes the inter-packet Wiener phase noise structure to break down. It is also observed in Chapter 3 that our proposed method does not perform well when the phase noise is strong. A challenging problem in such a scenario is to develop a new sparse recovery algorithm that is robust against strong phase noise. Denoting the total number of packets B as shown in Fig. 5.2, the number

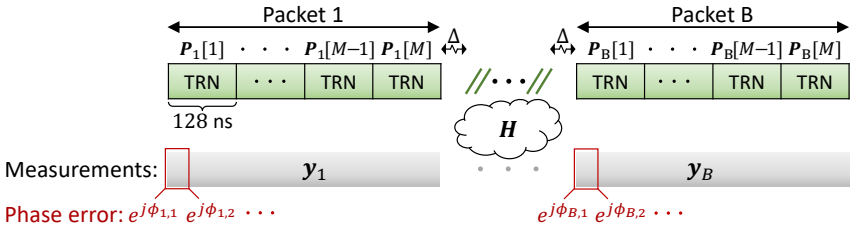


Figure 5.2: The frame structure used to acquire channel measurements with strong phase noise. Measurements within each packet are perturbed by Wiener phase noise, and they change drastically from one packet to another due to strong phase noise.

of measurements acquired in packet b by M, the m^{th} beamformer applied at packet b by $\mathbf{P}_b[m]$, the corresponding phase error by $\phi_{b,m}$, and the vector of M measurements in packet b by $\mathbf{y}_b$, and the $M \times 1$ vector of additive white Gaussian noise in packet b by $\mathbf{w}_b$, we have

$$\mathbf{y}_b = \text{diag}\left[e^{\text{j}\phi_{b,1}}, e^{\text{j}\phi_{b,2}}, \cdots, e^{\text{j}\phi_{b,M}}\right]\mathbf{A}_b\mathbf{x} + \mathbf{w}_b, \tag{5.1}$$

where $\mathbf{A}_b$ denotes the $M \times 1$ CS matrix for packet b and $\mathbf{x}$ denotes the $N^2 \times 1$ beamspace channel.

In Chapter 3, the diagonal matrix in (5.1) was replaced by a scalar, assuming that the phase errors were approximately the same for all measurements acquired in the batch. When phase noise is strong, this assumption no longer holds and requires rethinking the sparse channel estimation under phase noise. A first step to solve the problem in (5.1) could be to exploit the shared support structure as shown in Fig. 5.3. Another interesting problem in this case is to optimize the sensing matrices so that the sparse recovery problem is robust to phase noise.

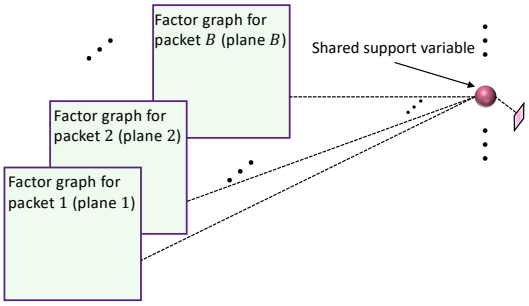


Figure 5.3: A sketch of a factor graph that exploits shared support structure for the case of a severe phase noise.

5.2.3. Sparse channel estimation under IQ imbalance with hybrid analog-digital architecture at the receiver

Our developed method for sparse channel estimation under IQ imbalance in Chapter 4 assumed that the receiver has a single RF chain. The receiver, however, may employ a hybrid analog-digital architecture with multiple RF chains, where each RF chain leads to a different IQ mismatch [80], as shown in Fig. 5.4. In particular, let R denote the number of RF chains at the receiver, $\{\xi_r\}_{r=1}^{R}$ denote the IQ imbalance parameter associated with R RF chains, $\mathbf{x}$ denote the beamspace channel, $\mathbf{A}$ denote the CS matrix and $\mathbf{w}$ denote the additive Gaussian noise. The vector of CS measurements denoted by $\mathbf{y}$ is

$$\mathbf{y} = \boldsymbol{\Xi}\mathbf{A}\mathbf{x} + (1 - \boldsymbol{\Xi}^*)\,\mathbf{A}^c\mathbf{x}^c + \mathbf{w}, \tag{5.2}$$

where $\boldsymbol{\Xi}$ is a block diagonal matrix whose r^{th} block is a diagonal matrix with ξ_r on its main diagonal entries. The CS measurements in (5.2) are different from the CS measurements in Chapter 4. Specifically, unlike Chapter 4, wherein the impact of IQ imbalance was modeled using a scalar, in (5.2) the IQ imbalance modeled by $\boldsymbol{\Xi}$ is a diagonal matrix. Because of this, extending our method in Chapter 4 to the case with multiple RF chains at the receiver is not trivial. Now, research questions similar to those in Chapter 4 can be

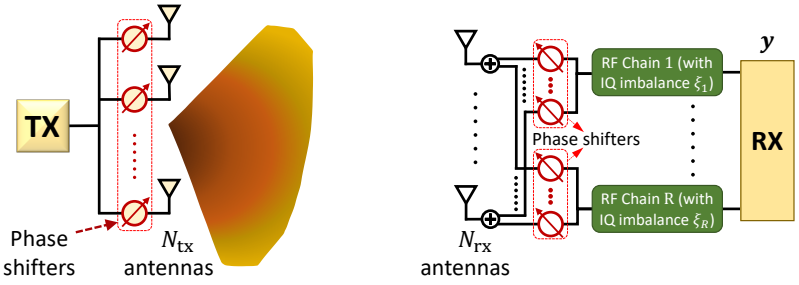


Figure 5.4: A wireless system with an N_{tx} element ULA at the TX and a hybrid analog-digital architecture at RX with N_{rx} antennas and R RF chains. The IQ imbalance parameter associated with r^{th} RF chain is given by ξ_r.

investigated. For instance, how exactly does the collection of IQ imbalance parameters impact the channel reconstruction using standard CS, agnostic to IQ imbalance? Also, extension of our approach to hybrid arrays, i,e. Fig. 5.4, joint IQ imbalance parameter vector and sparse channel estimation, is another interesting direction to be investigated.

Bibliography

[1] H. Sarieddeen, M.-S. Alouini and T. Y. Al-Naffouri. 'An overview of signal processing techniques for terahertz communications'. In: *Proc. IEEE* 109.10 (2021), pp. 1628–1665.

[2] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb and G. C. Trichopoulos. 'Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond'. In: *IEEE Access* 7 (2019), pp. 78729–78757.

[3] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh and A. M. Sayeed. 'An overview of signal processing techniques for millimeter wave MIMO systems'. In: *IEEE J. Sel. Topics Signal Process.* 10.3 (2016), pp. 436–453.

[4] I. F. Akyildiz, C. Han and S. Nie. 'Combating the distance problem in the millimeter wave and terahertz frequency bands'. In: *IEEE Commun. Mag.* 56.6 (2018), pp. 102–108.

[5] J. Brady, N. Behdad and A. M. Sayeed. 'Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements'. In: *IEEE Trans. Antennas Propag.* 61.7 (2013), pp. 3814–3827.

[6] S. Sun, G. R. MacCartney and T. S. Rappaport. 'A novel millimeter-wave channel simulator and applications for 5G wireless communications'. In: *Proc. IEEE Intl. Conf. Commun. (ICC)*. 2017, pp. 1–7.

[7] E. J. Candès and M. B. Wakin. 'An introduction to compressive sampling'. In: *IEEE Signal Process. Mag.* 25.2 (2008), pp. 21–30.

[8] M. Gharavi-Alkhansari and T. S. Huang. 'A fast orthogonal matching pursuit algorithm'. In: *Proc. IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*. Vol. 3. 1998, pp. 1389–1392.

[9] S. Rangan. 'Generalized approximate message passing for estimation with random linear mixing'. In: *Proc. IEEE Int. Symp. Inf. Theory*. IEEE. 2011, pp. 2168–2172.

[10] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb and R. W. Heath. 'Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?' In: *IEEE access* 4 (2016), pp. 247–267.

[11] A. S. Poon and M. Taghivand. 'Supporting and enabling circuits for antenna arrays in wireless communications'. In: *Proceedings of the IEEE* 100.7 (2012), pp. 2207–2218.

[12] X. Gao, L. Dai, Y. Sun, S. Han and I. Chih-Lin. 'Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems'. In: *Proc. IEEE International Conference on Communications (ICC)*. IEEE. 2017, pp. 1–6.

[13] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic and R. W. Heath. 'Channel estimation and hybrid combining for mmWave: Phase shifters or switches?' In: *Proc. Information Theory and Applications Workshop (ITA)*. IEEE. 2015, pp. 90–97.

[14] X. Gao, E. A. Klumperink, P. F. Geraedts and B. Nauta. 'Jitter analysis and a benchmarking figure-of-merit for phase-locked loops'. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 56.2 (2009), pp. 117–121.

[15] G. Hueber and R. B. Staszewski. 'RF Impairment Compensation for Future Radio Systems'. In: *Multi-Mode / Multi-Band RF Transceivers for Wireless Communications: Advanced Techniques, Architectures, and Trends*. 2011, pp. 451–496. DOI: 10.1002/9780470634455.ch15.

[16] S. Ling and T. Strohmer. 'Self-calibration and biconvex compressive sensing'. In: *Inverse Problems* 31.11 (2015), p. 115002.

[17] Z. Sha and Z. Wang. 'Channel estimation and equalization for terahertz receiver with RF impairments'. In: *IEEE J. Sel. Areas Commun.* 39.6 (2021), pp. 1621–1635.

[18] Y. Zou, P. Zetterberg, U. Gustavsson, T. Svensson, A. Zaidi, T. Kadur, W. Rave and G. Fettweis. 'Impact of major RF impairments on mm-wave communications using OFDM waveforms'. In: *Proc. IEEE Globecom Workshops (GC Wkshps)*. IEEE. 2016, pp. 1–7.

[19] B. Razavi and R. Behzad. *RF microelectronics*. Vol. 2. Prentice hall New York, 2012.

[20] D. Parveg, M. Varonen, A. Safaripour, P. Kangaslahti, M. Kantanen, T. Tikka, T. Gaier, K. A. Halonen and A. Hajimiri. 'An mm-wave CMOS I–Q subharmonic resistive mixer for wideband zero-IF receivers'. In: *IEEE Microw. Wireless Compon. Lett.* 30.5 (2020), pp. 520–523.

[21] X. Yang, M. Matthaiou, J. Yang, C.-K. Wen, F. Gao and S. Jin. 'Hardware-constrained millimeter-wave systems for 5G: Challenges, opportunities, and solutions'. In: *IEEE Commun. Mag.* 57.1 (2019), pp. 44–50.

[22] B. Khan, N. Tervo, M. Jokinen, A. Pärssinen and M. Juntti. 'Statistical digital predistortion of 5G millimeter-wave RF beamforming transmitter under random amplitude variations'. In: *IEEE Transactions on Microwave Theory and Techniques* 70.9 (2022), pp. 4284–4296.

[23] N. N. Moghadam, G. Fodor, M. Bengtsson and D. J. Love. 'On the energy efficiency of MIMO hybrid beamforming for millimeter-wave systems with nonlinear power amplifiers'. In: *IEEE Transactions on Wireless Communications* 17.11 (2018), pp. 7208–7221.

[24] N. Tervo, M. E. Leinonen and A. Parssinen. ‘On the Impact of RF impairments in 6G systems when scaling up the frequency and bandwidth’. In: *presentation in Workshop titled “International workshop on Wireless Communications in Terahertz (IWCT)” in 2023 EuCNC & 6G Summit*. 2023.

[25] A. B. Ayed, Y. Cao, P. Mitran and S. Boumaiza. ‘Digital predistortion of millimeter-wave arrays using near-field based transmitter observation receivers’. In: *IEEE Transactions on Microwave Theory and Techniques* 70.7 (2022), pp. 3713–3723.

[26] S. Lee, M. Kim, Y. Sirl, E.-R. Jeong, S. Hong, S. Kim and Y. H. Lee. ‘Digital predistortion for power amplifiers in hybrid MIMO systems with antenna subarrays’. In: *Proc. 2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*. IEEE. 2015, pp. 1–5.

[27] M. Akrout, V. Shyianov, F. Bellili, A. Mezghani and R. W. Heath. ‘Super-wideband massive MIMO’. In: *IEEE journal on selected areas in communications* 41.8 (2023), pp. 2414–2430.

[28] P. Zheng, S. Tarboush, H. Sarieddeen and T. Y. Al-Naffouri. ‘Mutual coupling-aware channel estimation and beamforming for RIS-assisted communications’. In: *IEEE Transactions on Wireless Communications* 25 (2026), pp. 4374–4390.

[29] T. Qiu, R. Li, C. Pan, T. Zhang, D. Xia, C. Wang and H. Ren. ‘Mutual Coupling Aware Channel Estimation for RIS-Aided Multi-User mmWave Systems’. In: *arXiv preprint arXiv:2511.08112* (2025).

[30] D. A. Sehrai, J. Khan, M. Abdullah, M. Asif, M. Alibakhshikenari, B. Virdee, W. A. Shah, S. Khan, M. Ibrar, S. Jan *et al.* ‘Design of high gain base station antenna array for mm-wave cellular communication systems’. In: *Scientific Reports* 13.1 (2023), p. 4907.

[31] R. Ansems, G. Federico, A. B. Smolders and D. Caratelli. ‘Multimode phased antenna array for mm-Wave user terminals with ultrawide-angle scanning capabilities’. In: *IEEE Transactions on Antennas and Propagation* 72.1 (2023), pp. 1021–1026.

[32] R. W. Heath Jr. *Introduction to wireless digital communication: a signal processing perspective*. Prentice Hall, 2017.

[33] N. J. Myers and R. W. Heath. ‘A compressive channel estimation technique robust to synchronization impairments’. In: *Proc. IEEE Intl. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*. 2017, pp. 1–5.

[34] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal and R. W. Heath. ‘Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems’. In: *IEEE Trans. Wireless Commun.* 17.5 (2018), pp. 2946–2960.

[35] N. J. Myers, A. Mezghani and R. W. Heath. ‘Swift-link: A compressive beam alignment algorithm for practical mmWave radios’. In: *IEEE Trans. Signal Process.* 67.4 (2018), pp. 1104–1119.

[36] I. P. Roberts, J. G. Andrews, H. B. Jain and S. Vishwanath. 'Millimeter-Wave Full Duplex Radios: New Challenges and Techniques'. In: *IEEE Wireless Commun.* 28.1 (2021), pp. 36–43. DOI: 10.1109/MWC.001.2000221.

[37] Z. Xiao, T. He, P. Xia and X.-G. Xia. 'Hierarchical Codebook Design for Beamforming Training in Millimeter-Wave Communication'. In: *IEEE Trans. Wireless Commun.* 15.5 (2016), pp. 3380–3392. DOI: 10.1109/TWC.2016.2520930.

[38] S. Noh, M. D. Zoltowski and D. J. Love. 'Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment'. In: *IEEE Trans. Wireless Commun.* 16.9 (2017), pp. 5689–5701. DOI: 10.1109/TWC.2017.2713357.

[39] Z. Xiao, H. Dong, L. Bai, P. Xia and X.-G. Xia. 'Enhanced Channel Estimation and Codebook Design for Millimeter-Wave Communication'. In: *IEEE Trans. Veh. Technol.* 67.10 (2018), pp. 9393–9405. DOI: 10.1109/TVT.2018.2854369.

[40] P. Raviteja, Y. Hong and E. Viterbo. 'Analog Beamforming With Low Resolution Phase Shifters'. In: *IEEE Wireless Commun. Lett.* 6.4 (2017), pp. 502–505. DOI: 10.1109/LWC.2017.2709306.

[41] C.-R. Tsai and A.-Y. Wu. 'Structured random compressed channel sensing for millimeter-wave large-scale antenna systems'. In: *IEEE Trans. Signal Process.* 66.19 (2018), pp. 5096–5110.

[42] Y. Ghasempour, C. R. Da Silva, C. Cordeiro and E. W. Knightly. 'IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi'. In: *IEEE Commun. Mag.* 55.12 (2017), pp. 186–192.

[43] P. A. Eliasi, S. Rangan and T. S. Rappaport. 'Low-Rank Spatial Channel Estimation for Millimeter Wave Cellular Systems'. In: *IEEE Trans. Wireless Commun.* 16.5 (2017), pp. 2748–2759. DOI: 10.1109/TWC.2017.2662687.

[44] A. Alkhateeb, G. Leus and R. W. Heath. 'Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?' In: *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process. (ICASSP)*. 2015, pp. 2909–2913.

[45] N. J. Myers, A. Mezghani and R. W. Heath. 'FALP: Fast beam alignment in mmWave systems with low-resolution phase shifters'. In: *IEEE Trans. Commun.* 67.12 (2019), pp. 8739–8753.

[46] Z. Marzi, D. Ramasamy and U. Madhow. 'Compressive channel estimation and tracking for large arrays in mm-wave picocells'. In: *IEEE J. Sel. Topics Signal Process.* 10.3 (2016), pp. 514–527.

[47] E. Vlachos, G. C. Alexandropoulos and J. Thompson. 'Massive MIMO Channel Estimation for Millimeter Wave Systems via Matrix Completion'. In: *IEEE Signal Process. Lett.* 25.11 (2018), pp. 1675–1679. DOI: 10.1109/LSP.2018.2870533.

[48] W. Ma, C. Qi, Z. Zhang and J. Cheng. 'Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO'. In: *IEEE Trans. Commun.* 68.5 (2020), pp. 2838–2849.

[49] W. Wang and W. Zhang. ‘Jittering Effects Analysis and Beam Training Design for UAV Millimeter Wave Communications’. In: *IEEE Trans. Wireless Commun.* (2021).

[50] A. Ali, N. González-Prelcic and R. W. Heath. ‘Millimeter wave beam-selection using out-of-band spatial information’. In: *IEEE Trans. Wireless Commun.* 17.2 (2017), pp. 1038–1052.

[51] P.-C. Chen and P. Vaidyanathan. ‘Convolutional beamspace for linear arrays’. In: *IEEE Trans. Signal Process.* 68 (2020), pp. 5395–5410.

[52] P.-C. Chen and P. Vaidyanathan. ‘Hybrid Convolutional Beamspace for DOA Estimation of Millimeter Wave Sources’. In: *Proc. Asilomar Conf. Signals Syst. Comput.* 2022, pp. 86–90.

[53] S.-E. Chiu, N. Ronquillo and T. Javidi. ‘Active learning and CSI acquisition for mm-Wave initial alignment’. In: *IEEE J. Sel. Areas Commun.* 37.11 (2019), pp. 2474–2489.

[54] S.-E. Chiu and T. Javidi. ‘Low complexity sequential search with size-dependent measurement noise’. In: *IEEE Trans. Inf. Theory* 67.9 (2021), pp. 5731–5748.

[55] A. Demir, A. Mehrotra and J. Roychowdhury. ‘Phase noise in oscillators: a unifying theory and numerical methods for characterization’. In: *IEEE Trans. on Circuits and Syst. I: Fundamental Theory and Applications* 47.5 (2000), pp. 655–674.

[56] W. Yi, N. J. Myers and G. Joseph. ‘Sparse Millimeter Wave Channel Estimation from Partially Coherent Measurements’. In: *Proc. IEEE Intl. Conf. Commun. (ICC)*. 2024, pp. 1897–1902.

[57] R. Zhang, B. Shim and H. Zhao. ‘Downlink compressive channel estimation with phase noise in massive MIMO systems’. In: *IEEE Trans. Commun.* 68.9 (2020), pp. 5534–5548.

[58] G. Colavolpe, A. Barbieri and G. Caire. ‘Algorithms for iterative decoding in the presence of strong phase noise’. In: *IEEE J. Sel. Areas Commun.* 23.9 (2005), pp. 1748–1757.

[59] S. Shayovitz and D. Raphaeli. ‘Message passing algorithms for phase noise tracking using Tikhonov mixtures’. In: *IEEE Trans. Commun.* 64.1 (2015), pp. 387–401.

[60] M. E. Rasekh and U. Madhow. ‘Noncoherent compressive channel estimation for mm-wave massive MIMO’. In: *Proc. Asilomar Conf. Signals Syst. Comput.* 2018, pp. 889–894.

[61] X. Li, J. Fang, H. Duan, Z. Chen and H. Li. ‘Fast beam alignment for millimeter wave communications: A sparse encoding and phaseless decoding approach’. In: *IEEE Trans. Signal Process.* 67.17 (2019), pp. 4402–4417.

[62] K.-H. Liu, X. Li, H. Zhao and G. Fan. ‘Structured Phase Retrieval-aided Channel Estimation for Millimeter-Wave/Sub-Terahertz MIMO Systems’. In: *Proc. 96th IEEE Veh. Tech. Conf. (VTC-Fall)*. 2022, pp. 1–5.

[63] H. Yan and D. Cabria. ‘Compressive sensing based initial beamforming training for massive MIMO millimeter-wave systems’. In: *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*. 2016, pp. 620–624.

[64] C. Hu, X. Wang, L. Dai and J. Ma. ‘Partially coherent compressive phase retrieval for millimeter-wave massive MIMO channel estimation’. In: *IEEE Trans. Signal Process.* 68 (2020), pp. 1673–1687.

[65] C. Wei, H. Jiang, J. Dang, L. Wu and H. Zhang. ‘Accurate channel estimation for mmWave massive MIMO with partially coherent phase offsets’. In: *IEEE Commun. Lett.* 26.9 (2022), pp. 2170–2174.

[66] L. Zhang, G. Wang, G. B. Giannakis and J. Chen. ‘Compressive phase retrieval via reweighted amplitude flow’. In: *IEEE Trans. Signal Process.* 66.19 (2018), pp. 5029–5040.

[67] J. P. Vila and P. Schniter. ‘Expectation-maximization Gaussian-mixture approximate message passing’. In: *IEEE Trans. Signal Process.* 61.19 (2013), pp. 4658–4672.

[68] Z. Zhu, X. Huang, M. Caron and H. Leung. ‘Blind self-calibration technique for I/Q imbalances and DC-offsets’. In: *IEEE Trans. Circuits Syst. I: Reg. Papers* 61.6 (2013), pp. 1849–1859.

[69] K. Okada, N. Li, K. Matsushita, K. Bunsen, R. Murakami, A. Musa, T. Sato, H. Asada, N. Takayama, S. Ito *et al.* ‘A 60-GHz 16QAM/8PSK/QPSK/BPSK direct-conversion transceiver for IEEE 802.15.3c’. In: *IEEE J. Solid-State Circuits* 46.12 (2011), pp. 2988–3004.

[70] R. Wu, R. Minami, Y. Tsukui, S. Kawai, Y. Seo, S. Sato, K. Kimura, S. Kondo, T. Ueno, N. Fajri *et al.* ‘64-QAM 60-GHz CMOS transceivers for IEEE 802.11ad/ay’. In: *IEEE J. Solid-State Circuits* 52.11 (2017), pp. 2871–2891.

[71] A. Tarighat, R. Bagheri and A. H. Sayed. ‘Compensation schemes and performance analysis of IQ imbalances in OFDM receivers’. In: *IEEE Trans. Signal Process.* 53.8 (2005), pp. 3257–3268.

[72] K.-Y. Sung and C.-c. Chao. ‘Estimation and compensation of I/Q imbalance in OFDM direct-conversion receivers’. In: *IEEE J. Sel. Topics Signal Process.* 3.3 (2009), pp. 438–453.

[73] W. Hou and M. Jiang. ‘Enhanced joint channel and IQ imbalance parameter estimation for mobile communications’. In: *IEEE Commun. Lett.* 17.7 (2013), pp. 1392–1395.

[74] X. Zhang, H. Li, W. Liu and J. Qiao. ‘Iterative IQ imbalance compensation receiver for single carrier transmission’. In: *IEEE Trans. Veh. Techol.* 66.9 (2017), pp. 8238–8248.

[75] X. Cheng, Y. Yang and S. Li. ‘Joint compensation of transmitter and receiver I/Q imbalances for SC-FDE systems’. In: *IEEE Trans. Veh. Techol.* 69.8 (2020), pp. 8483–8498.

[76] A. P. Dempster, N. M. Laird and D. B. Rubin. ‘Maximum likelihood from incomplete data via the EM algorithm’. In: *J. Roy. Statist. Soc.* 39.1 (1977), pp. 1–22.

[77] S. Wang and L. Zhang. ‘Signal processing in massive MIMO with IQ imbalances and low-resolution ADCs’. In: *IEEE Trans. Wireless Commun.* 15.12 (2016), pp. 8298–8312.

[78] N. Kolomvakis, M. Matthaiou and M. Coldrey. ‘IQ imbalance in multiuser systems: Channel estimation and compensation’. In: *IEEE Trans. Commun.* 64.7 (2016), pp. 3039–3051.

[79] Y. Chen, L. You, A.-A. Lu and X. Gao. ‘Widely-linear processing for the uplink of the massive MIMO with IQ imbalance: Channel estimation and data detection’. In: *IEEE Trans. Signal Process.* 69 (2021), pp. 4685–4698.

[80] Y. Teng, L. Jia, A. Liu and V. K. Lau. ‘Joint estimation of channel and I/Q imbalance in massive MIMO: A two-timescale optimization approach’. In: *IEEE Trans. Wireless Commun.* 18.10 (2019), pp. 4723–4737.

[81] H. Masoumi, M. Verhaegen and N. J. Myers. ‘In-sector compressive beam acquisition for mmWave and THz radios’. In: *IEEE Transactions on Communications* 73.4 (2025), pp. 2752–2768.

[82] E. Khordad, I. B. Collings, S. V. Hanly and G. Caire. ‘Compressive sensing-based beam alignment schemes for time-varying millimeter-wave channels’. In: *IEEE Trans. Wireless Commun.* 22.3 (2022), pp. 1604–1617.

[83] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer and M. Hollick. ‘Compressive millimeter-wave sector selection in off-the-shelf IEEE 802.11 ad devices’. In: *Proc. ACM Intl. Conf. on emerging Networking EXperiments and Technologies*. 2017, pp. 414–425.

[84] V. Raghavan, A. Partyka, A. Sampath, S. Subramanian, O. H. Koymen, K. Ravid, J. Cezanne, K. Mukkavilli and J. Li. ‘Millimeter-wave MIMO prototype: Measurements and experimental results’. In: *IEEE Commun. Mag.* 56.1 (2018), pp. 202–209.

[85] A. Alhamed, G. Gültepe and G. M. Rebeiz. ‘64-element 16–52-GHz transmit and receive phased arrays for multiband 5G-NR FR2 operation’. In: *IEEE Trans. Microwave Theory and Techniques* 71.1 (2022), pp. 360–372.

[86] S. Zihir, O. D. Gurbuz, A. Kar-Roy, S. Raman and G. M. Rebeiz. ‘60-GHz 64-and 256-elements wafer-scale phased-array transmitters using full-reticle and subreticle stitching techniques’. In: *IEEE Trans. on Microwave Theory and Techniques* 64.12 (2016), pp. 4701–4719.

[87] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[88] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia and J. C. Widmer. ‘IEEE 802.11 ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi’. In: *IEEE Commun. Mag.* 52.12 (2014), pp. 132–141.

[89] P. F. Smulders. ‘Statistical characterization of 60-GHz indoor radio channels’. In: *IEEE Trans. Antennas Propag.* 57.10 (2009), pp. 2820–2829.

[90] O. Kome, L. Cen, L. Hanqing and Y. Rui. *Further Details on Multi-Stage, Multi-Resolution Beamforming Training in 802.11 ay, doc*. 2016.

[91] 'IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications'. In: *IEEE Std* (2021), pp. 1–4379. DOI: 10.1109/IEEESTD.2021.9363693.

[92] M. Giordani, M. Mezzavilla and M. Zorzi. 'Initial Access in 5G mmWave Cellular Networks'. In: *IEEE Commun. Mag.* 54.11 (2016), pp. 40–47. DOI: 10.1109/MCOM.2016.1600193CM.

[93] A. K. Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[94] D. G. Manolakis and V. K. Ingle. *Applied digital signal processing: theory and practice*. Cambridge university press, 2011.

[95] H. Masoumi, M. Verhaegen and N. J. Myers. 'Analysis of Orthogonal Matching Pursuit for Compressed Sensing in Practical Settings'. In: *Proc. IEEE Statistical Signal Process. Workshop (SSP)*. 2023, pp. 170–174. DOI: 10.1109/SSP53291.2023.10207984.

[96] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[97] P. Stoica, H. He and J. Li. 'On designing sequences with impulse-like periodic correlation'. In: *IEEE Signal Process. Lett.* 16.8 (2009), pp. 703–706.

[98] K. Li, L. Gan and C. Ling. 'Convolutional compressed sensing using deterministic sequences'. In: *IEEE Trans. Signal Process.* 61.3 (2012), pp. 740–752.

[99] E. Candes and J. Romberg. 'Sparsity and incoherence in compressive sampling'. In: *Inverse Problems* 23.3 (2007), p. 969.

[100] M. Lustig, D. Donoho and J. M. Pauly. 'Sparse MRI: The application of compressed sensing for rapid MR imaging'. In: *Magn. Reson. Med.* 58.6 (2007), pp. 1182–1195.

[101] Z. Ben-Haim, Y. C. Eldar and M. Elad. 'Coherence-based performance guarantees for estimating a sparse vector under random noise'. In: *IEEE Trans. Signal Process.* 58.10 (2010), pp. 5030–5043.

[102] M. Golay. 'Complementary series'. In: *IRE Trans. Inf. Theory* 7.2 (1961), pp. 82–87.

[103] K. V. Mishra and Y. C. Eldar. 'Sub-Nyquist channel estimation over IEEE 802.11ad link'. In: *Proc. Intl. Conf. Sampling Theory Appl. (SampTA)*. 2017, pp. 355–359.

[104] L. Zhao, J. Song, P. Babu and D. P. Palomar. 'A unified framework for low autocorrelation sequence design via majorization–minimization'. In: *IEEE Trans. Signal Process.* 65.2 (2016), pp. 438–453.

[105] J. Jedwab, C. Mitchell, F. Piper and P. Wild. 'Perfect binary arrays and difference sets'. In: *Discr. Math.* 125.1-3 (1994), pp. 241–254.

[106] R. Frank, S. Zadoff and R. Heimiller. 'Phase shift pulse codes with good periodic correlation properties (corresp.)' In: *IRE Trans. Inf. Theory* 8.6 (1962), pp. 381–382.

[107] D. Chu. ‘Polyphase codes with good periodic correlation properties’. In: *IEEE Trans. Inf. Theory* 18.4 (1972), pp. 531–532.

[108] N. Zhang and S. W. Golomb. ‘Polyphase sequence with low autocorrelations’. In: *IEEE Trans. Inf. Theory* 39.3 (1993), pp. 1085–1089.

[109] W. Dai and O. Milenkovic. ‘Subspace pursuit for compressive sensing signal reconstruction’. In: *IEEE Trans. Inform. Theory* 55.5 (2009), pp. 2230–2249.

[110] S. Boyd and J. Park. ‘Subgradient methods’. In: *lecture notes of EE364b, Stanford University* (2014).

[111] S. Boyd, S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[112] J. Ziniel and P. Schniter. ‘Dynamic compressive sensing of time-varying signals via approximate message passing’. In: *IEEE Trans. Signal Process.* 61.21 (2013), pp. 5270–5284.

[113] Z. Šidák. ‘Rectangular confidence regions for the means of multivariate normal distributions’. In: *J. Amer. Stat. Assoc.* 62.318 (1967), pp. 626–633.

[114] H. Masoumi and N. J. Myers. ‘Message passing-based sparse channel estimation under partially coherent Wiener phase errors’. In: *IEEE Transactions on Wireless Communications* (2025).

[115] J. Rimmelspacher, R. Weigel, A. Hagelauer and V. Issakov. ‘A Quad-core 60 GHz push-push 45 nm SOI CMOS VCO with -101.7 dBc/Hz phase noise at 1 MHz offset, 19% continuous FTR and -187 dBc/Hz FoMT’. In: *Proc. IEEE 44th European Solid State Circuits Conf. (ESSCIRC)*. 2018, pp. 138–141.

[116] S. Rangan, A. K. Fletcher, V. K. Goyal, E. Byrne and P. Schniter. ‘Hybrid approximate message passing’. In: *IEEE Trans. Signal Process.* 65.17 (2017), pp. 4577–4592.

[117] J. Mo, P. Schniter and R. W. Heath. ‘Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs’. In: *IEEE Trans. Signal Process.* 66.5 (2017), pp. 1141–1154.

[118] J. Ziniel and P. Schniter. ‘Efficient high-dimensional inference in the multiple measurement vector problem’. In: *IEEE Trans. Signal Process.* 61.2 (2012), pp. 340–354.

[119] H. Masoumi and N. J. Myers. ‘Message passing-based sparse spatial channel estimation robust to partially coherent phase noise’. In: *Proc. IEEE Vehicular Technology Conference (VTC)*. 2025.

[120] A. F. Molisch. *Wireless communications*. Vol. 34. John Wiley & Sons, 2012.

[121] F. R. Kschischang, B. J. Frey and H.-A. Loeliger. ‘Factor graphs and the sum-product algorithm’. In: *IEEE Trans. Inf. Theory* 47.2 (2001), pp. 498–519.

[122] S. O. Rice. ‘Mathematical analysis of random noise’. In: *The Bell System Technical Journal* 23.3 (1944), pp. 282–332.

[123] J. C. Lagarias, J. A. Reeds, M. H. Wright and P. E. Wright. 'Convergence properties of the Nelder–Mead simplex method in low dimensions'. In: *SIAM Journal on optimization* 9.1 (1998), pp. 112–147.

[124] J. Rimmelspacher, R. Ciocoveanu, G. Steffan, M. Bassi and V. Issakov. 'Low power low phase noise 60 GHz multichannel transceiver in 28 nm CMOS for radar applications'. In: *Proc. IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. 2020, pp. 19–22.

[125] H. Jia, P. Guan, W. Deng, Z. Wang and B. Chi. 'A low-phase-noise quad-core millimeter-wave fundamental VCO using circular triple-coupled transformer in 65-nm CMOS'. In: *IEEE J. Solid-State Circuits* 58.2 (2022), pp. 371–385.

[126] L. Lian, A. Liu and V. K. N. Lau. 'Exploiting Dynamic Sparsity for Downlink FDD-Massive MIMO Channel Tracking'. In: *IEEE Trans. Signal Process.* 67.8 (2019), pp. 2007–2021.

[127] Y. Wei, M.-M. Zhao, A. Liu and M.-J. Zhao. 'Channel Tracking and Prediction for IRS-Aided Wireless Communications'. In: *IEEE Trans. Wireless Commun.* 22.1 (2023), pp. 563–579.

[128] K. V. Mardia and P. E. Jupp. *Directional statistics*. John Wiley & Sons, 2009.

[129] F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl and M. A. McClain. *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/10.43, Release 1.2.1 of 2024-06-15. URL: https://dlmf.nist.gov/10.43.

[130] G. Ridgway. *Rice/Rician distribution*. MATLAB Central File Exchange, https://www.mathworks.com/matlabcentral/fileexchange/14237-rice-rician-distribution, Retrieved August 4, 2024. URL: https://www.mathworks.com/matlabcentral/fileexchange/14237-rice-rician-distribution.

[131] H. Masoumi and N. J. Myers. 'Compressive sensing-based sparse spatial channel estimation under IQ imbalance'. In: *submitted to the IEEE Transactions on Signal Processing* (2025).

[132] M. Valkama, M. Renfors and V. Koivunen. 'Advanced methods for I/Q imbalance compensation in communication receivers'. In: *IEEE Trans. Signal Process.* 49.10 (2001), pp. 2335–2344.

[133] Z. Zhu, X. Huang and H. Leung. 'Joint I/Q mismatch and distortion compensation in direct conversion transmitters'. In: *IEEE Trans. Wireless Commun.* 12.6 (2013), pp. 2941–2951.

[134] E. M. Al Seragi, Y. L. Rajendra, W. Ahmad, M. Kaynak, P. F. Goldsmith and S. Zeinolabdedinzadeh. 'A 220–320 GHz high image rejection sideband separating receiver for space-borne observatories'. In: *IEEE Trans. THz Sci. Techol.* 14.2 (2024), pp. 162–177.

[135] H. Rauhut. 'Compressive sensing and structured random matrices'. In: *Theoretical foundations and numerical methods for sparse recovery* 9 (2010), pp. 1–92.

[136] N. J. Myers, A. Mezghani and R. W. Heath. ‘Spatial Zadoff-Chu modulation for rapid beam alignment in mmwave phased arrays’. In: *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*. 2018, pp. 1–6.

[137] Y. C. Eldar, P. Kuppinger and H. Bolcskei. ‘Block-sparse signals: Uncertainty relations and efficient recovery’. In: *IEEE Trans. Signal Process.* 58.6 (2010), pp. 3042–3054.

[138] G. Swirszcz, N. Abe and A. C. Lozano. ‘Grouped orthogonal matching pursuit for variable selection and prediction’. In: *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*. 2009, pp. 1150–1158.

[139] H. Masoumi, N. J. Myers, G. Leus, S. Wahls and M. Verhaegen. ‘Structured Sensing Matrix Design for In-sector Compressed mmWave Channel Estimation’. In: *Proc. IEEE Intl. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*. 2022, pp. 1–5. DOI: 10.1109/SPAWC51304.2022.9833949.

[140] H. Masoumi and N. J. Myers. ‘Compressive Spatial Channel Estimation under IQ Imbalance’. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2026.

[141] H. Masoumi and N. J. Myers. ‘Message passing-based sparse channel estimation from partially coherent measurement vectors’. In: *Proc. Joint European Conference on Networks and Communications and 6G Summit (EuCNC/6G Summit)*. 2025.

Acknowledgements

I would like to thank my co-promotor and supervisor, Dr. Nitin Jonathan Myers, for giving me the opportunity and for his support along the way. I really enjoyed our discussions and I appreciate your valuable contributions that improved our work and me as a researcher. I also thank my promoters, Dr. Ing. Raf Van de Plas, Prof. dr. Michel Verhaegen, and Prof. dr. Sander Wahls, for overseeing my PhD trajectory and for their valuable input. Also, I thank the independent members of my doctoral committee for their valuable feedback on this thesis.

I want to thank all my DCSC colleagues for making the workplace a pleasant and welcoming environment. I also thank many of my friends outside the TU Delft for their support, which indirectly helped me.

Finally, I sincerely thank my family for their support with everything in my life.

Hamed Masoumi
Delft, March 2026

List of Publications

9. H. Masoumi and N. J. Myers. ‘Compressive sensing-based sparse spatial channel estimation under IQ imbalance’. major revision with the *IEEE Transactions on Signal Processing* (2025)

8. H. Masoumi and N. J. Myers. ‘Message passing-based sparse channel estimation under partially coherent Wiener phase errors’. In: *IEEE Transactions on Wireless Communications* (2025)

7. H. Masoumi, M. Verhaegen and N. J. Myers. ‘In-sector compressive beam acquisition for mmWave and THz radios’. In: *IEEE Transactions on Communications* 73.4 (2025), pp. 2752–2768

6. H. Masoumi, P. B. Groen, B. F. J. Kamminga, L. J. Hoogland, B. V. Schaik, A. C. V. D. Veen, E. Focante and N. J. Myers. ‘In-sector direction-of-arrival estimation using a switched receive array’. minor revision with the *IEEE Sensors Letters* (2026)

5. H. Masoumi and N. J. Myers. ‘Compressive Spatial Channel Estimation under IQ Imbalance’. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2026

4. H. Masoumi and N. J. Myers. ‘Message passing-based sparse spatial channel estimation robust to partially coherent phase noise’. In: *Proc. IEEE Vehicular Technology Conference (VTC)*. 2025

3. H. Masoumi and N. J. Myers. ‘Message passing-based sparse channel estimation from partially coherent measurement vectors’. In: *Proc. Joint European Conference on Networks and Communications and 6G Summit (EuCNC/6G Summit)*. 2025

2. H. Masoumi, M. Verhaegen and N. J. Myers. ‘Analysis of Orthogonal Matching Pursuit for Compressed Sensing in Practical Settings’. In: *Proc. IEEE Statistical Signal Process. Workshop (SSP)*. 2023, pp. 170–174. DOI: 10.1109/SSP53291.2023.10207984

1. H. Masoumi, N. J. Myers, G. Leus, S. Wahls and M. Verhaegen. ‘Structured Sensing Matrix Design for In-sector Compressed mmWave Channel Estimation’. In: *Proc. IEEE Intl. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*. 2022, pp. 1–5. DOI: 10.1109/SPAWC51304.2022.9833949