Low-resolution compressed sensing and beyond for communications and sensing
Trends and opportunities

Joseph, Geethu; Gandikota, Venkata; Bhandari, Ayush; Choi, Junil; Kim, In-soo; Lee, Gyoseung; Matthaiou, Michail; Murthy, Chandra R.; Ngo, Hien Quoc; More Authors

DOI
10.1016/j.sigpro.2025.110020

Publication date
2025

Document Version
Final published version

Published in
Signal Processing

Citation (APA)
Joseph, G., Gandikota, V., Bhandari, A., Choi, J., Kim, I., Lee, G., Matthaiou, M., Murthy, C. R., Ngo, H. Q., & More Authors (2025). Low-resolution compressed sensing and beyond for communications and sensing: Trends and opportunities. *Signal Processing*, *235*, Article 110020. https://doi.org/10.1016/j.sigpro.2025.110020

Signal Processing 235 (2025) 110020

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Low-resolution compressed sensing and beyond for communications and sensing: Trends and opportunities

Geethu Joseph [a,*], Venkata Gandikota [b], Ayush Bhandari [c], Junil Choi [d], In-soo Kim [e], Gyoseung Lee [d], Michail Matthaiou [f], Chandra R. Murthy [g], Hien Quoc Ngo [f], Pramod K. Varshney [b], Thakshila Wimalajeewa [h], Wei Yi [i], Ye Yuan [i], Guoxin Zhang [i]

[a] Delft University of Technology, Netherlands
[b] Syracuse University, USA
[c] Imperial College London, United Kingdom
[d] Korea Advanced Institute of Science and Technology, Republic of Korea
[e] Qualcomm Technologies, Inc., USA
[f] Queen's University Belfast, United Kingdom
[g] Indian Institute of Science, Bangalore, India
[h] Johns Hopkins University Applied Physics Laboratory, USA
[i] The University of Electronic Science and Technology of China, China

ARTICLE INFO

Keywords:
One-bit compressed sensing
Sparsity
Low-bit quantization
Wireless channel estimation
Massive MIMO
Reconfigurable intelligent surfaces
Task-based quantization
Variational Bayesian
Sparse Bayesian learning
Maximum likelihood
Cramér–Rao bound
Target localization and tracking
Unlimited sensing
Modulo ADC

ABSTRACT

This survey paper examines recent advancements in low-resolution signal processing, emphasizing quantized compressed sensing. Rising costs and power demands of high-sampling-rate data acquisition drive the interest in quantized signal processing, particularly in wireless communication systems and Internet of Things sensor networks, as 6G aims to integrate sensing and communication within cost-effective hardware. Motivated by this urgency, this paper covers novel signal processing algorithms designed to address practical challenges arising from quantization and modulo operations, as well as their impact on system performance. We begin by introducing the framework of one-bit compressed sensing and discuss relevant theories and algorithms. We explore the application of quantized compressed sensing algorithms to sensor networks, radar, cognitive radio, and wireless channel estimation. We highlight how generic methods can be tailored to an application using specific examples from wireless channel estimation. Additionally, we review other low-resolution techniques beyond one-bit compressed sensing along with their applications. We also provide a brief overview of the emerging concept of unlimited sampling. While this paper does not aim to be exhaustive, it selectively highlights results to inspire readers to appreciate the diverse algorithmic tools (convex optimization, greedy methods, and Bayesian approaches) and sampling techniques (task-based quantization and unlimited sampling).

1. Introduction

Quantization of signals is a critical aspect of modern digital signal processing applications, including sensing, communication, and inference. Ideally, measurements should have high resolution, but implementing such analog-to-digital converters (ADCs) in practical systems presents significant challenges. These devices often become bottlenecks due to their power consumption, size, and manufacturing costs, which increase exponentially with the number of bits [1–3]. This complexity has sparked a growing interest in quantized signal processing, particularly for applications such as 6G wireless communication and sensing systems [4,5].

In 6G applications, high sampling rates pose significant challenges. For instance, millimeter-wave (mmWave) multiple-input and multiple-output (MIMO) technology demands large bandwidths, requiring increased sampling rates for ADCs. However, manufacturing high-resolution (e.g., over 8 bits) and fast ADCs is expensive and power-hungry. Similarly, in applications like spectral sensing and cognitive radio networks, which also require high sampling rates, the cumulative costs and power consumption of the high-resolution fast ADCs can be prohibitive and impractical. Furthermore, these applications involve edge intelligence that demands low-power hardware and signal processing capabilities. One immediate solution to these challenges is to employ low-resolution ADCs, which offers robustness, memory efficiency, and

* Corresponding author.
E-mail address: G.Joseph@tudelft.nl (G. Joseph).

https://doi.org/10.1016/j.sigpro.2025.110020
Received 22 November 2024; Received in revised form 23 February 2025; Accepted 26 March 2025
Available online 10 April 2025

simplicity in hardware implementation, particularly in sensor design. However, coarse quantization in ADCs can undermine the performance of traditional signal processing techniques that assume high-resolution quantization and often require customized algorithms to achieve adequate system performance. Motivated by these challenges, this paper studies low-resolution signal processing techniques, particularly the low-rate signal acquisition method of compressed sensing (CS), with an emphasis on their applications in wireless communication and sensing.

CS is a signal acquisition technique that requires fewer samples than the Nyquist rate [6–10]. It exploits the sparsity or compressibility of natural signals, such as images, audio signals, and communication signals, which can be expressed using a few nonzero coefficients in a suitable basis. For example, images are sparse in a Fourier or wavelet basis [6]. Without loss of generality, in the rest of the paper, we assume that the coefficient vector is sparse in the standard canonical basis unless otherwise specified. The sparsity of a coefficient vector is measured using the ℓ_0-norm, which counts the number of nonzero entries,

$$\|\boldsymbol{x}\|_0 = |\{i : \boldsymbol{x}[i] \neq 0\}|. \tag{1}$$

A vector $\boldsymbol{x} \in \mathbb{R}^n$ is said to be sparse if $\|\boldsymbol{x}\|_0 \ll n$ and s-sparse if $\|\boldsymbol{x}\|_0 = s < n$. The compressive sensing measurements $\boldsymbol{y}$ are acquired by projecting the sparse vector onto a set of basis vectors, called a dictionary matrix or measurement matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Here, $\boldsymbol{A}$ has fewer rows than columns $m < n$, leading to compression. Thus, the CS problem is

$$\arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{2}$$

Since the above problem is NP-hard, the sparse signal $\boldsymbol{x}$ is estimated from $\boldsymbol{y}$ using various algorithmic approaches, including convex relaxation, greedy algorithms, and Bayesian methods. The quality of the reconstruction depends on the signal's sparsity (or compressibility), the reconstruction algorithm used, and the properties of the measurement matrix. One of the celebrated results in CS is that randomly generated dictionaries are likely to be incoherent with any fixed sparsity basis, making them ideal for CS in terms of the measurements m required for successful recovery of an s-sparse n-length $\boldsymbol{x}$. For more details on CS theory and algorithms, see [11,12].

CS achieved via random projections at the sampling stage is valuable in several resource-constrained systems. However, in many applications like Internet of Things (IoT), which operate under severe resource limitations, further compression and quantization of compressed measurements may be necessary. Implementing coarse quantization can significantly reduce bandwidth requirements and computational costs at local nodes, making it desirable in highly resource-scarce environments. Low-bit quantization results in information loss, making it challenging for the traditional CS methods to perform effectively. Early works on CS assume that quantization error is negligible, treating it as a noisy CS problem. This approach implicitly assumes high-resolution quantizers, which are often unrealistic or inefficient. Moreover, it yields pessimistic reconstruction error estimates that cannot exceed the noise floor, particularly the quantization error. To address the challenges posed by coarse quantization, the field of quantized CS has advanced significantly, showing that tailored algorithms can substantially improve performance over traditional methods [13] to solve the quantized CS problem,

$$\arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \boldsymbol{y} = Q_B(\boldsymbol{A}\boldsymbol{x}), \tag{3}$$

where Q_B represents B-bit quantization. These efficient schemes have demonstrated that quantized CS-based systems can significantly enhance processing speeds for large volumes of sparse data while simultaneously lowering communication costs and simplifying hardware implementation. Quantized CS has recently found widespread use in a range of applications, including wireless sensor networks [14–17], cognitive radio [18–23], wireless communication [24–27], radar [28–33], image processing [34,35], and medical technologies [36,36,37]. This paper consolidates various results from the literature, focusing on applying low-bit quantized CS in wireless communication and sensing.

This survey is not exhaustive and aims to strike a balance between theoretical rigor and practical insights. The key results and solution strategies presented in this paper are as follows. We start by discussing one-bit compressed sensing (1bCS), the most widely studied form of quantized CS. One-bit quantization is especially attractive for hardware implementations and has demonstrated resilience to nonlinear distortions, as well as dynamic range limitations. Section 2 presents the general framework of 1bCS, detailing the conditions on the measurement matrix required for the successful recovery of sparse vectors. Notably, the results demonstrate that the minimal number of measurements required scales linearly with the sparsity level and logarithmically with the size of the sparse vector. Additionally, this section explores extensions and connections to efficient machine learning frameworks. While Section 2 emphasizes the underlying theoretical principles, Section 3 of the paper shifts to practical applications in wireless communications and sensing. We cover applications across various domains such as IoT, wireless sensor networks, radar, cognitive radio, and wireless channel estimation. Section 3 also delves into the specific application of MIMO channel estimation, showing how quantized CS algorithms can be adapted to enhance performance. For example, quantized CS algorithms used for channel estimation can be integrated with data decoding or can be combined with task-based quantizers that optimize the quantizer based on the estimation task. In Section 4, we examine alternative low-resolution signal processing techniques beyond CS. This section introduces non-sparse techniques in IoT, wireless sensor networks, and radar with an emphasis on localization and tracking algorithms that use low-bit quantization in conjunction with maximum likelihood (ML) approaches. It also covers low-resolution techniques in cognitive radio networks and wireless channel estimation. We also present the unlimited sensing framework (USF), a new paradigm designed to address saturation issues in digital acquisition systems. The paper concludes by highlighting several avenues for future research. The roadmap of the paper is given in Fig. 1.

Notation

Notation is usually introduced when it first appears. Boldface small letters denote vectors, and boldface capital letters denote matrices. The symbols $\boldsymbol{a}[i]$, $\boldsymbol{A}_i$, and $\boldsymbol{A}[i,j]$ represent the ith entry of vector $\boldsymbol{a}$, the ith column of matrix $\boldsymbol{A}$, and the entry on the ith row and jth column of matrix $\boldsymbol{A}$, respectively. The operator Q_B represents B-bit quantization.

We denote the all-one vector of length a as $\mathbf{1}_a$ and the all-zero vector as $\mathbf{0}$. The symbol $\mathbb{1}\{\cdot\}$ represents the indicator function, which takes the value one if the condition in the argument is true and zero otherwise. The sets of $M\times N$ real and complex matrices are represented by $\mathbb{R}^{M\times N}$ and $\mathbb{C}^{M\times N}$, respectively. The symbol $\boldsymbol{I}$ denotes the identity matrix. Also, $\mathbb{Z}$ represents the set of integers.

We use $\|\cdot\|_0$ and $\|\cdot\|$ to denote the ℓ_0 norm and the ℓ_2 norm, respectively. For any signal $\boldsymbol{a} \in \mathbb{R}^n$, let $\text{supp}(\boldsymbol{a}) \triangleq \{i \in \{1,2,\dots,n\} \mid \boldsymbol{a}[i] \neq 0\}$ denote the support of $\boldsymbol{a}$. The symbols $(\cdot)^{-1}$, $(\cdot)^{\mathsf{T}}$, $(\cdot)^*$, $(\cdot)^{\mathsf{H}}$, $\text{tr}(\cdot)$, and $|\cdot|$ are the matrix operations of inverse, transpose, conjugate, conjugate transpose, trace, and determinant respectively. Also, $\otimes$, $\diamond$, and $\odot$ represent the Kronecker, Khatri–Rao, and Hadamard products respectively. The operator $\text{diag}(\cdot)$ returns a (block) diagonal matrix with the argument along the diagonal.

We use $\mathcal{N}(\boldsymbol{a},\boldsymbol{A})$ and $\mathcal{CN}(\boldsymbol{a},\boldsymbol{A})$ to denote the real and complex Gaussian distributions respectively, with mean $\boldsymbol{a}$ and covariance $\boldsymbol{A}$. Also, $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ denote the real and imaginary parts, respectively.

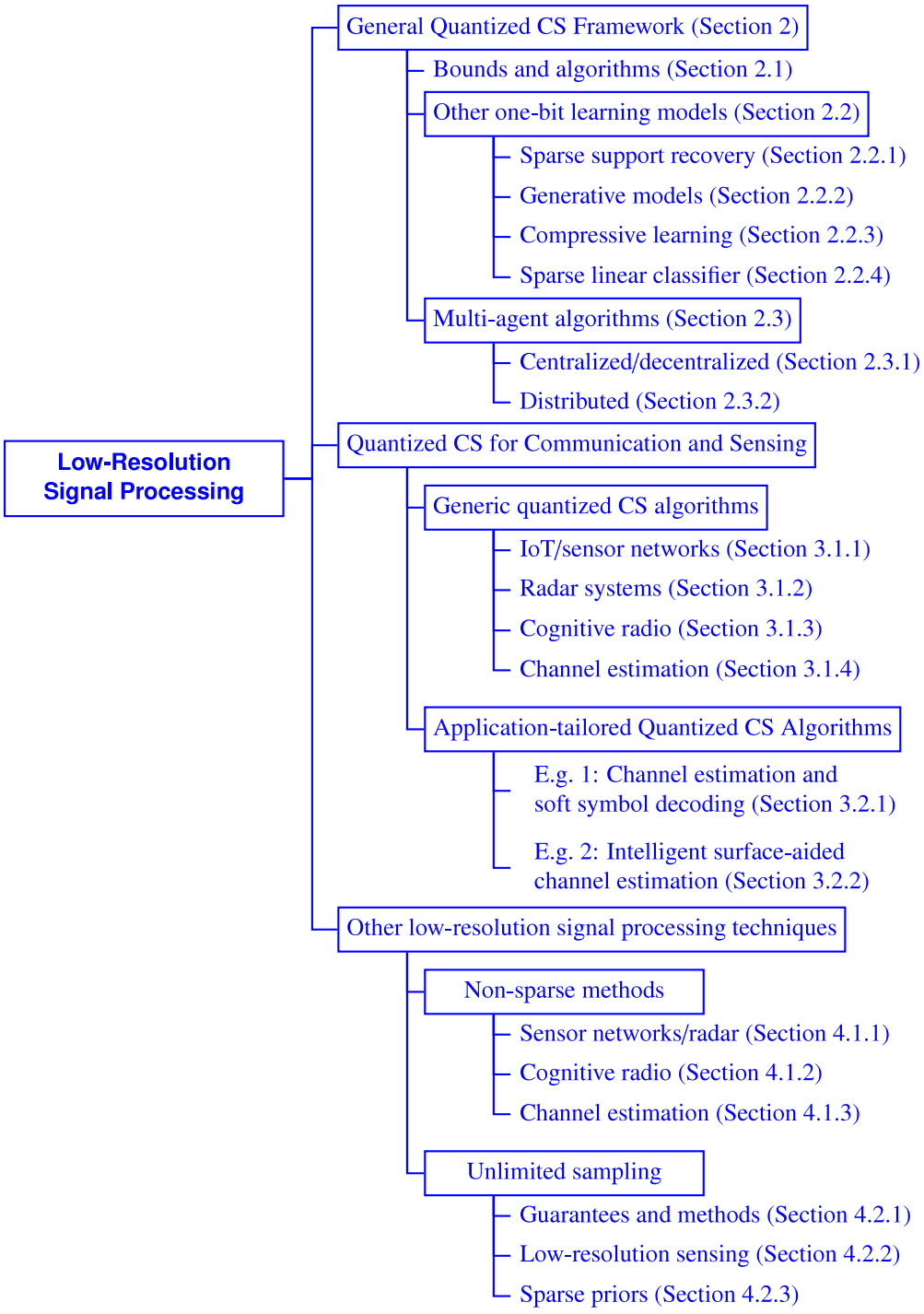


Fig. 1. Roadmap of the paper.

2. Quantized compressed sensing: Theory and general algorithmic approaches

As discussed in Section 1, quantized CS seeks to recover a sparse vector from its quantized linear measurements. While increasing the bit-depth reduces the quantization error, the improvement in reconstruction accuracy saturates beyond 3–5 bits per measurement. The findings in [13] suggest that dithering can enhance low-bit quantization performance and that the optimal bit-depth depends on the task; lower bit-depths are sufficient for classification, while full reconstruction requires higher bit-depths. For scalar quantization, the primary focus of this survey, [38] presents a trade-off between bit-depth and measurement rate. They recommend using fewer bits with more measurements in low-signal-to-noise ratio (SNR) settings and higher bit-depth with fewer measurements in high-SNR conditions. Notably, among low-bit-depth methods, the most extreme form, 1bCS, has gained significant attention due to its superior performance guarantees across a wide range of scenarios. The popularity of 1bCS also arises from its simple and cost-effective quantizer, as well as the reduced bit requirements for storage and transmission. In this section, we present results on the performance of sparse signal reconstruction and parameter estimation using noisy one-bit measurements, highlighting the theoretical foundations.

1bCS acquires 1-bit measurements of the form,

$$\boldsymbol{y} = Q_1(\boldsymbol{A}\boldsymbol{x}) \triangleq \operatorname{sign}(\boldsymbol{A}\boldsymbol{x}) \in \{-1, 1\}^m, \tag{4}$$

where the $\operatorname{sign}(\cdot)$[1] function is applied coordinate-wise to the linear measurements $\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$ for a chosen measurement matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Since the prescribed quantization function is invariant under positive scaling of the signal $\boldsymbol{x}$, the magnitude information of the signal is lost. Hence, without loss of generality, we assume that $\boldsymbol{x}$ lies on the unit sphere, and the recovery algorithms usually focus on obtaining an estimate $\hat{\boldsymbol{x}}$ on the unit sphere.

The main goal in 1bCS is to accurately recover the underlying s-sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ given m ($< n$) well-chosen 1-bit measurements such that the recovered vector $\hat{\boldsymbol{x}}$ satisfies a given error bound ϵ, i.e., $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \leq \epsilon$. To address noise during signal acquisition and post-quantization errors, we adopt a generalized measurement model,

$$\boldsymbol{y} \triangleq \boldsymbol{\eta} \odot \operatorname{sign}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}) \in \{-1, 1\}^m, \tag{5}$$

where $\boldsymbol{w} \in \mathbb{R}^n$ is the additive noise, usually considered to be independent and identically distributed (iid) Gaussian entries with a fixed variance of σ^2, and $\boldsymbol{\eta} \in \{-1, 1\}^m$ is the bit-flip noise with bounded number of negative entries applied coordinate-wise to the quantized output.

2.1. Measurement bounds and algorithms

In recent decades, the impressive recovery guarantees achievable even under extreme quantization have spurred extensive research into 1bCS. These results have led to the development of efficient recovery algorithms that require an optimal number of measurements and remain robust to noise. Some of the key approaches in 1bCS include convex program-based methods [39–41], greedy or iterative methods, including binary iterative hard thresholding (BIHT) [42–45], Bayesian techniques [46–51], generative model-based methods [34,52–56] (see Table 1). For a comprehensive review, we direct readers to an excellent survey on this topic [57]. The following sections present recent advancements and new directions beyond this survey, emphasizing their provable recovery guarantees.

One of the early works established that any ϵ-approximate recovery algorithm for an s-sparse signal $\boldsymbol{x}$ requires at least $\Omega(s\log\frac{n}{s} + \frac{s}{\epsilon} - s^{3/2})$ measurements [42]. This bound is subsequently refined in [72] to $\Omega(s\log\frac{n}{s} + \frac{s}{\epsilon})$. Following this result, research has focused on designing efficient and theoretically robust recovery algorithms that achieve this improved measurement bound. In pursuit of provable recovery guarantees, a class of consistent reconstruction algorithms is developed in the 1bCS literature that finds a unit-norm vector $\hat{\boldsymbol{x}}$ with minimal sparsity that aligns with the measurement outcomes [42]. Specifically, the underlying optimization problem for recovery can be defined as follows:

$$\min \|\hat{\boldsymbol{x}}\|_0 \quad \text{s.t.} \quad \operatorname{sign}(\boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{y} \quad \text{and} \quad \|\hat{\boldsymbol{x}}\| = 1. \tag{6}$$

This work also demonstrates that random Gaussian measurement matrices, where each entry of $\boldsymbol{A}$ is chosen independently from a standard Gaussian distribution with $m = O(\frac{s}{\epsilon}\log\frac{n}{s})$ rows, can achieve an ϵ-approximate solution to (6). Similar results are also known to hold with sub-Gaussian [73] and partial circulant matrices [74].

To ensure robustness against noise, both in the form of random additive noise before quantization and bounded bit-flip errors after quantization, it has been shown that $m = O\left(\frac{s}{\epsilon^2}\log\frac{n}{s}\right)$ Gaussian measurements are sufficient to achieve a uniform embedding of s-sparse unit vectors onto the vertices of the Boolean hypercube $\{-1, 1\}^m$ [42]. The uniform embedding property implies that the pairwise ℓ_2 distances (equivalent to spherical or angular distances) between any two s-sparse unit vectors are approximately preserved when mapped to $\{-1, 1\}^m$. Therefore, the original distances on the unit sphere can be estimated by Hamming distances in the Boolean hypercube, allowing a stable recovery of sparse vectors using a small number of measurements. This approach extends beyond s-sparse vectors; uniform tessellation results have also been generalized to arbitrary subsets of the unit sphere. These generalizations involve using measurement matrices with entries drawn from Gaussian, sub-Gaussian, and heavy-tailed distributions, allowing for flexibility in applications, though they often require affine measurements [75–77].

Table 1
Summary of sparse recovery methods in Section 2.

Approach	Key algorithms
Optimization-based methods	ℓ_1-norm based program [39,58]
	Sparse logistic regression [40]
	Penalized ℓ_1-norm [59]
	Weighted ℓ_1-norm [60]
	Smoothly clipped absolute deviation, minimax concave penalty [61,62]
	Schur-concave functions [63]
	k-support norm [64]
	Pinball loss minimization [65]
	ℓ_2-constrained least squares [41]
	ℓ_1-total variation norm [66]
Greedy or iterative methods	Matched sign pursuit [44]
	Gradient support pursuit [43]
	BIHT[42,67]
	Adaptive outlier pursuit [45]
	Normalized BIHT and variants [67]
Bayesian techniques	Message passing algorithm [68]
	Approximate message passing [46–48]
	Sparse Bayesian learning [50,51]
Generative model-based methods	Rectified linear unit (ReLU) with dithering [69]
	Extension of BIHT [52]
	Non-convex optimization [34,70]
	Diffusion models (score-based) [71]

While the above-mentioned results establish the theoretical guarantees for signal recovery, they do not resolve the algorithmic challenge of provable and efficient recovery. To tackle the intractability of (6) due to the ℓ_0 minimization objective and the unit norm constraint, several works have focused on solving a relaxed version of this problem [39,40,44,58,65,78]. An early approach uses a greedy algorithm called matched sign pursuit, which combines the principle of consistent reconstruction with compressive sampling matching pursuit from the standard CS literature [44]. Using a similar approach, 1bCS algorithm robust to random noise added pre-quantization is also devised [43]. An alternative approach uses convex relaxation that reformulates the problem as a linear program, yielding an efficient ϵ-approximate recovery algorithm that requires $m = O(\frac{s}{\epsilon^5}\log^2\frac{n}{s})$ Gaussian measurements [39],

$$\min \|\hat{\boldsymbol{x}}\|_1 \quad \text{s.t.} \quad \operatorname{sign}(\boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{y} \quad \text{and} \quad \|\boldsymbol{A}\hat{\boldsymbol{x}}\|_1 = m. \tag{7}$$

To address adversarial bit-flip errors occurring after quantization, a variant of (7) has been developed,

$$\min -\boldsymbol{y}^\top\boldsymbol{A}\hat{\boldsymbol{x}} \quad \text{s.t.} \quad \|\hat{\boldsymbol{x}}\|_1 \leq s \quad \text{and} \quad \|\hat{\boldsymbol{x}}\|_2 \leq 1, \tag{8}$$

that ensures reliable recovery using $m = O\left(\frac{s}{\epsilon^6}\log\frac{n}{s}\right)$ Gaussian measurements [40]. This approach seeks to find a signal $\hat{\boldsymbol{x}}$ within the convex hull of all s-sparse unit vectors that maximizes consistency between $\boldsymbol{y}$ and $\operatorname{sign}(\boldsymbol{A}\hat{\boldsymbol{x}})$.

Further, various regularization techniques enhanced the convex programming approach by promoting sparsity and improving model

[1] For any $a \in \mathbb{R}$, $\operatorname{sign}(a) = 1$ if $a \geq 0$, and $\operatorname{sign}(a) = -1$ if $a < 0$.

robustness. Penalized ℓ_1-norm methods [59] leads to an efficient algorithm with a closed-form solution:

$$\min\ -\boldsymbol{y}^\top \boldsymbol{A}\hat{\boldsymbol{x}} + \omega\|\hat{\boldsymbol{x}}\|_1 \quad \text{s.t.} \quad \|\hat{\boldsymbol{x}}\|_2 \le 1, \tag{9}$$

where ω is the weight parameter. The weighted ℓ_1-norm techniques [60] extend this approach to a weighted ℓ_1-norm formulation and derive analytical solutions under specific weights. Non-convex penalties, such as the smoothly clipped absolute deviation and minimax concave penalty, further enforce sparsity [61,62]. One example is [61]

$$\min\ -\boldsymbol{y}^\top \boldsymbol{A}\hat{\boldsymbol{x}} + h_{\text{non-convex}}(\hat{\boldsymbol{x}}) + \frac{\omega}{2}\|\hat{\boldsymbol{x}}\|_2^2 \quad \text{s.t.} \quad \|\hat{\boldsymbol{x}}\|_2 \le 1, \tag{10}$$

where $h_{\text{non-convex}}$ represents the minimax concave penalty, and $\omega \ge 0$ controls the smoothness. Further extensions to other non-convex penalties with analytical solutions are explored in [62]. Another non-convex penalty is normalized ℓ_1 Shannon entropy function, a Schur-concave measure of concentration, which achieves sparse solutions at its minima [63]. This nonconvex problem is reformulated as weighted ℓ_1-norm subproblems, solved iteratively using a generalized fixed-point continuation algorithm. Moreover, other regularization include k-support norm, pinball loss minimization [65], ℓ_2-constrained least squares [41], ℓ_1-total variation norm [66]. The k-support norm-based method demonstrates that the new estimator admits a closed-form solution, eliminating the need for optimization. The paper also establishes its consistency and provides recovery guarantees for both Gaussian and sub-Gaussian random measurements. The alternative approach based on pinball loss minimization is a convex approach, solved using dual-coordinate ascent algorithms. The ℓ_2-constrained least squares approach in [41] recovers solutions with higher sparsity levels and the resulting optimization problem is solved using weighed primal and dual active set algorithm. Recently, an ℓ_1-total variation method for 1bCS estimates signals that are both element and gradient sparse, providing a closed-form solution and an adaptive dither vector quantization scheme [66].

Although these methods achieve efficient recovery with a near-optimal number of measurements, their dependence on the error parameter ϵ is not optimal [39,40]. To address this issue, an alternative approach based on projected gradient descent, specifically designed for 1bCS is introduced [42]. This algorithm, called BIHT, is the 1bCS version of the iterative hard thresholding algorithm [79]. Starting with a random point $\hat{\boldsymbol{x}}^0$ on the unit sphere, every iteration of the BIHT algorithm takes a small step in the negative (sub)gradient direction, followed by projection onto the set of s-sparse vectors through thresholding. This algorithm has shown excellent empirical performance and is detailed in Algorithm 1, where the function Threshold$(\cdot)$ retains the s largest-magnitude entries and sets the remaining entries to zero. Later, adaptive outlier pursuit extends BIHT to handle noisy measurement, and it reduces to BIHT when measurements are noiseless [45]. In this algorithm, every iteration estimates the sparse signal like BIHT but excludes potentially corrupted measurements, and then updates the list of likely corrupted measurements.

Algorithm 1 BIHT Algorithm

Initialize $\hat{\boldsymbol{x}}^0$ to have unit norm
for $t = 1, 2, \dots, L$ **do**
  $\tilde{\boldsymbol{x}}^t \leftarrow \hat{\boldsymbol{x}}^{t-1} + \frac{\sqrt{2\pi}}{2m}\boldsymbol{A}^\top(\boldsymbol{y} - \text{sign}(\boldsymbol{A}\hat{\boldsymbol{x}}^{t-1}))$
  $\hat{\boldsymbol{x}}^t \leftarrow$ Threshold$(\tilde{\boldsymbol{x}}^t)$.
end for
Return $\frac{\hat{\boldsymbol{x}}^L}{\|\hat{\boldsymbol{x}}^L\|}$.

It has been shown that the BIHT algorithm converges within an ϵ radius of the true signal in just one step given $m = O(\frac{s}{\epsilon^2}\log n)$ Gaussian measurements [13,42]. A later study shows that this estimate remains close to the true signal and does not diverge in subsequent iterations [80]. The first formal convergence result for BIHT is based on a modified version, known as the normalized BIHT algorithm [67]. This algorithm, in every iteration, normalizes the projected vector to obtain a unit norm vector. This approach achieves convergence within an ϵ ball of the true signal with $m > \max\{(s\log\frac{n}{s})^{10}, 2^{48}\frac{1}{\epsilon}(s\log\frac{n}{s})^{7/2}\}$ measurements. While its dependence on ϵ is optimal, the constant factors and the dependence on s are quite far from optimal. In a sequence of recent groundbreaking works, convergence and robustness of the normalized BIHT variant is proven using an optimal $O(\frac{s}{\epsilon}\log n)$ Gaussian measurements [81,82]. The convergence proof relies on a finer property of measurement matrices known as the restricted approximate invertibility condition, ensuring that the sign measurements behave similarly to scaled linear measurements. The core of the convergence proof involves demonstrating that Gaussian matrices with an optimal number of rows satisfy this condition with high probability [81].

Another algorithmic approach that offers a promising alternative is Bayesian inference methods, especially under the assumption of Gaussian-distributed measurement matrices and noiseless outputs. In a Bayesian framework, the goal is to maximize the posterior distribution given the observed value $p(\boldsymbol{x}|\boldsymbol{y})$. When the true signal distribution is known, the Bayesian inference is theoretically optimal [68]; however, it often suffers from high computational complexity. To handle the computational difficulty of exact Bayesian inference, generalized approximate message passing algorithms have been developed, building on belief propagation techniques assuming a hierarchical prior structure on the signal $\boldsymbol{x}$ such as the *Gaussian-inverse Gamma* prior that promotes sparsity [46,49]. For signals corrupted by Gaussian noise, variational inference techniques, like variational message passing [48] with a two-layer hierarchical prior, are adopted to encourage the sparsity of the signal. Alternatively, [47] provides a robust approach by modeling quantization noise explicitly and optimizing the posterior distribution of the signal via variational expectation maximization. Another approach uses Bussgang-like decomposition using which the 1bCS problem can be approximated as a standard linear model and use the standard sparse Bayesian learning (SBL) algorithm for sparse vector recovery [51]. Further, [50] uses the SBL framework where the correlation matrix is approximated using the arcsin law. These innovations collectively enable Bayesian methods to effectively address signal recovery challenges, even in cases complicated by quantization and noise. However, the sample complexity of Bayesian algorithms is largely unknown. Table 2 summarizes the known measurement bounds together with the bounds of generative models discussed in Section 2.2.2.

While the above-described techniques have mostly focused on the sparse signal recovery problem, they have also extended the 1bCS framework to address a broader range of learning problems, which we discuss next.

2.2. Connections to broader learning problems

Recent works have expanded the 1bCS framework to accommodate other low-dimensional structures and learning problems. We discuss a new two-stage algorithmic paradigm for signal recovery in Section 2.2.1, explore extensions for recovering non-sparse signals generated by a deep neural network in Section 2.2.2, and examine various extensions of 1bCS for efficient machine learning in Section 2.2.3 and Section 2.2.4.

2.2.1. Support recovery

The two-stage algorithmic paradigm has recently been investigated for the recovery of sparse signals from a small number of one-bit measurements. In these algorithms, the first stage focuses on recovering the approximate support set of the sparse signal, which is limited to a size of at most $O(s)$. In the second stage, the algorithm employs additional $O\left(\frac{s}{\epsilon}\log\frac{n}{s}\right)$ Gaussian measurements to approximately recover the magnitudes of the nonzero entries of the sparse vector, based on

Table 2
A comparison of sample complexities m of 1bCS algorithms to recover an s-sparse vector in $\mathbb{R}^N$ within an error of ϵ (for generative models, r is the radius of latent space and κ is the number of layers).

Approach	Measurement Complexity	Remarks and References
Lower Bounds		
	$\Omega(s\log\frac{n}{s}+\frac{s}{\epsilon}-s^{3/2})$	Early result [42]
	$\Omega(s\log\frac{n}{s}+\frac{s}{\epsilon})$	Improved bound [72]
Upper Bounds		
Optimization	$O\left(\frac{s}{\epsilon^5}\log^2\frac{n}{s}\right)$	ℓ_1-optimization (noiseless) [39]
	$O\left(\frac{s}{\epsilon^4}\log n\right)$	Sparse logistic regression [40,73,74]
	$O\left(\frac{s}{\epsilon^2}\log n\right)$	Penalized ℓ_1-norm [59]
		Nonconvex penalty [61]
	$O\left(\frac{s}{\epsilon^2}\right)$	Nonconvex penalty (strong signals) [61]
BIHT	$O\left(\frac{s}{\epsilon}\log n\right)$	BIHT (noiseless) [42]
		Normalized BIHT [81,82]
Generative models	$O(\frac{s\kappa}{\epsilon^2}\log n\log^2(\frac{1}{\epsilon}))$	ReLU network with dithering [83]
	$O(\frac{s\kappa}{\epsilon^2}\log(r/\epsilon))$	Generic generative model [52,54]
	$O(\frac{s}{\epsilon^2}(r^2+\kappa))$	Optimization-based [34]

the identified support. It is important to note that while the algorithm operates in two stages, the entire measurement matrix is designed in advance and is not adaptively modified during the measurement process. The required number of measurements for these algorithms is largely determined by the first stage, which is discussed next.

For any signal $\boldsymbol{x}\in\mathbb{R}^n$, let $\mathrm{supp}(\boldsymbol{x})\triangleq\{i\in\{1,2,\dots,n\}\mid \boldsymbol{x}[i]\neq 0\}$ denote the support of $\boldsymbol{x}$. The goal of the support recovery problem is to identify $\mathrm{supp}(\boldsymbol{x})$ using a minimal number of one-bit measurements. For this problem, one approach uses techniques derived from the heavy-hitters streaming algorithm to obtain a *sign-sketch* that accurately recovers the support of a fixed signal $\boldsymbol{x}$ using $O(s\log n)$ measurements with high probability [84]. The result also holds in the presence of pre-quantization noise and for compressible signals. However, these techniques necessitate modifications to the measurement matrix for each new signal, which is impractical.

In contrast, a universal measurement matrix for 1bCS is a single set of measurements that can be applied to all sparse signals. Universal support recovery is initially explored using techniques from coding theory and group testing, establishing sufficient conditions for exact support recovery [85]. Later advancements refine these conditions to achieve improved bounds for both exact and approximate support recovery [72,86–88]. These results all build upon the union-free property of the measurement matrices, which plays a key role in support recovery.

Definition 1 (*Robust Union Free Family (RUFF)*)**.** A binary matrix $\boldsymbol{E}\in\{0,1\}^{m\times n}$ is said to be (d,s,α)-robust union free if each column of $\boldsymbol{E}$ has weight exactly equal to d, and for every set of $s+1$ columns of $\boldsymbol{E}$ indexed by $\{i_0,i_1,\dots,i_s\}\subset\{1,2,\dots,n\}$, there exist at least $(1-\alpha)d$ rows indexed by $\mathcal{T}\subset\{1,2,\dots,m\}$ such that $\boldsymbol{E}[i_0,j]=1$, and $\boldsymbol{E}[i_l,j]=0$ for all l and $j\in\mathcal{T}$.

A Robust union free family (RUFF) measurement matrix ensures that each support set of size at most s generates a unique outcome signature. Existence of (d,s,α)-RUFFs with $m=O(\frac{s^2}{\alpha^2}\log n)$ rows, and $d=O(\frac{s}{\alpha}\log n)$ is known using random constructions [89,90].

The use of RUFF properties in measurement matrices, particularly with $\alpha=1$ enables exact recovery of non-negative sparse vectors [72]. Further, they show the resilience of the algorithm to $(\frac{1}{2}-\alpha)d$ adversarial bit-flips when using RUFF with $\alpha<1$. To handle all s-sparse signals, an expander-based measurement matrix construction with $m=O(s^2\log n)$ rows is also introduced in this work. The same recovery algorithm is later shown in [72] to work for all s-sparse signals using RUFF with $\alpha=\frac{1}{2}$, thereby establishing an upper bound of $m=O(s^2\log n)$. Furthermore, the authors also show a lower bound of $m=\Omega(s^2\frac{\log n}{\log s})$ measurements for exact support recovery by establishing that $(d,s-1,1)$ RUFF property is necessary.

Furthermore, to reduce the dependence of $O(s^2)$ for the approximate recovery problem using a two-stage algorithm, several works have investigated the approximate support recovery problem [86–88]. In particular, it is shown that $O((s^{3/2}+\frac{s}{\epsilon})\log\frac{n}{s})$ measurements from a generalization of RUFFs (called list-RUFF) are sufficient if one needs to recover a small superset of the true support of size $O(s)$, and $O(\frac{s}{\epsilon}\log\frac{n}{s})$ measurements are sufficient if the algorithm is allowed to return support with ϵs errors (both false positives and false negatives) [88]. This work also establishes a nearly matching lower bound of $m=\Omega(\frac{s}{\epsilon}(\log\frac{s}{\epsilon})^{-1}\log\frac{n}{\epsilon s})$ for approximate support recovery.

In a non-universal setting, exact support recovery achieves a tight bound of $m=\Theta(s\log n)$ measurements [91] whereas ϵ-approximate support recovery needs $m=\Theta(\frac{s}{\epsilon}\log n)$ measurements [87] for signals with a bounded dynamic range, $\frac{\max_{\boldsymbol{x}[i]\neq 0}|\boldsymbol{x}[i]|}{\min_{\boldsymbol{x}[i]\neq 0}|\boldsymbol{x}[i]|}$. Additionally, when the support of the signal is restricted to certain small groups of indices, only $O(s\log n)$ measurements are necessary [92].

2.2.2. Generative priors

The study of 1bCS has been extended to recovering structured signals that go beyond the standard sparsity assumption. An early approach inspired by the success of deep generative models is to explore the recovery of signals within the range of a Λ-Lipschitz continuous generative model [93]. Formally, let $\boldsymbol{x}$ be a signal in the range of $G:\mathcal{B}_2^s(r)\to\mathbb{R}^n$, where G is a generative model that is Λ-Lipschitz, and $\mathcal{B}_2^s(r)$ denotes an ℓ_2 ball of radius $r>0$ centered at zero in $\mathbb{R}^s$. We note that κ-layer deep neural networks with ReLU, Hyperbolic tangent, or sigmoid activation functions are known to be $\Lambda=n^{\Theta(\kappa)}$-Lipschitz continuous. Recovering such signals can use a gradient descent approach in the latent space that aims to minimize an empirical recovery error function. This approach needs roughly $m=O(\frac{s}{\epsilon^2}\log(\Lambda r))$ random Gaussian measurements for ϵ-approximate signal recovery, and relies on a stronger variant of the restricted eigenvalue condition which is satisfied by random Gaussian matrices. Several follow-up works have provided better algorithmic guarantees and established tight information-theoretic lower bounds on the number of measurements necessary for accurate recovery. See [94] for a survey on related results. Several works built upon this approach achieve results comparable to classical sparsity-based methods while requiring significantly fewer measurements [52–56,70,71,83,95–98].

Generalizing the recovery conditions for sparse signals discussed earlier [42], recent work extends the results to provide an information-theoretic characterization for the recovery of signals with generative priors [34,52]. The authors of [52] establish an upper bound, demonstrating that $O(\frac{s}{\epsilon}\log\frac{\Lambda r}{\epsilon})$ independent Gaussian measurements are sufficient with high probability to distinguish any two signals that are ϵ separated. The lower bounds indicate that at least $O(\frac{s}{\epsilon}+s\log(\Lambda r))$ measurements are necessary for ϵ-approximate recovery of the signal. Furthermore, a modified BIHT algorithm archives excellent empirical performance, where the hard threshold function is replaced by a projection operation onto the range of the generative model. However, the measurement bound in [52] is generic and not directly connected to BIHT version. Concurrently, [34] introduces an optimization-based algorithm combined with the generative model, which achieves a similar measurement complexity.

Provable recovery algorithms with an almost optimal number of measurements are first introduced for ReLU-based generative models without an offset [83]. In particular, the authors design an empirical risk minimization algorithm that recovers bounded target vectors produced by the model from $O(\frac{s\kappa}{\epsilon^2}\log n\log^2(\frac{1}{\epsilon}))$ quantized noisy sub-exponential measurements with a uniform dither. The analysis relies on the piecewise linearity of ReLU networks that allows for a

more efficient tessellation of the range space using affine hyperplanes. The authors also establish an information-theoretic lower bound that matches the achievability up to ϵ^{-1} factor for shallow networks. Subsequent works have extended the provable recovery algorithms to signals generated using Λ-Lipschitz generative models with $O(\frac{s}{\epsilon^2}\log \Lambda r)$ measurements [34,53–56]. Notably, a least absolute shrinkage and selection operator (LASSO)-based approach for non-uniform recovery is first introduced [53], which is later extended to obtain uniform recovery guarantees [56]. This approach circumvents the need for Gaussian measurements by using structured partial Gaussian circulant matrices for recovery. Additionally, a one-shot projection algorithm for non-uniform recovery of generative signals is developed, utilizing an almost optimal number of Gaussian measurements [55]. These results also hold for more general single-index measurement models with minimal assumptions on the quantization function. Another study investigates a diffusion-based generative model, which shows promise but is limited by the large training dataset and high computational cost, leading to latency issues, especially on edge devices [71]. The study also lacks theoretical analyses of reconstruction performance and measurement requirements. Finally, since generative models generally require a large volume of training data, algorithms based on untrained deep neural networks are also available in the 1bCS literature [70].

2.2.3. Compressive learning with one-bit measurements

In many practical applications, such as in IoT, exact reconstruction of signals is not necessary. Instead, the focus is on computing some function of the original signal rather than recovering it. This approach allows for a substantial reduction in implementation complexity by inferring information directly from compressed measurements without the need for reconstruction. This area of research, known as compressive learning, has recently gained significant attention [99,100]. One of the first studies in this direction estimates the ℓ_2 norm of a s-sparse signal $\boldsymbol{x}$ with $r^2 \le \|\boldsymbol{x}\|^2 \le R^2$ using only $O(\frac{R^4}{r^2})$ affine measurements [101]. Random linear projections onto low-dimensional subspaces, which are well-studied in the context of CS and dimensionality reduction, are known to preserve certain properties of high-dimensional datasets, such as pairwise Euclidean distances between data points. Consequently, compressed datasets – often referred to as sketches – retain sufficient information that many compressive learning algorithms utilize to achieve efficient signal classification and clustering [102–105]. Numerous studies have also extended these techniques to develop effective learning algorithms from one-bit quantized sketches.

In one-bit quantized learning, the goal is to infer properties of signals (that are not necessarily sparse) using a few one-bit measurements given by (4). Geometrically, each one-bit measurement depicts the side of the chosen hyperplane in which the data point (signal) lies. This binary embedding preserves pairwise angular distances between data points, making it useful for developing efficient learning algorithms. One technique is to train an algorithm using a few labeled and one-bit quantized data points, enabling efficient classification of newly acquired signals. The algorithm computes a "score function" based on binary measurements during the training phase and classifies new data by maximizing a class probability function. A separate but related line of work provides an upper bound on the misclassification error when the signals follow a Gaussian mixture model [106]. This work designs an accurate correlation estimator using a few one-bit sign measurements that are utilized by the classification algorithm. Fundamental performance limits are investigated, extending analogous works in the compressive sensing setting to the one-bit quantized setting. Further, extending the analogous works of [105] in CS setting to the one-bit quantized setting, techniques are designed to estimate the Chernoff and KL divergence distances between the probability density functions (PDFs) of measurements based on their membership in different classes [107]. It is noted that the misclassification error, bounded using these distance measures, decreases exponentially with an increasing number of measurements.

Another important area of study focuses on clustering, which aims to extract patterns from a set of unlabeled data points. Compressive clustering algorithms seek to provide efficient clustering solutions using compressed datasets without the need for reconstruction [104]. However, the one-bit quantized variant of clustering is less explored. The existing research investigates this problem by estimating data distribution through random sampling of its characteristic function at randomly drawn frequencies [108] and its extension also demonstrate differential privacy [109].

2.2.4. Mixture of sparse linear classifiers

The signal recovery in the 1bCS model is equivalent to *learning* a sparse linear classifier in the active query model. Here, each designed measurement can be considered as querying the classifier for its label, and the goal is to recover the linear classifier using a minimal number of such queries. The mixture of sparse linear classifiers generalizes this problem to simultaneously recover L sparse signals using a minimal number of queries. In such a model, the output of a query (or measurement) is derived stochastically from one of the signals at random.

Let $\mathcal{X} = \{\boldsymbol{x}_1 \ldots, \boldsymbol{x}_L\} \subseteq \mathbb{R}^n$ be the set of sparse signals with $\|\boldsymbol{x}_i\|_0 \le s$, and the outcome with respect to a query (or measurement) vector $\boldsymbol{a} \in \mathbb{R}^n$ is derived as:

$$y_j = \text{sign}(\langle \boldsymbol{a}, \boldsymbol{x}_j \rangle), \qquad \boldsymbol{x}_j \sim_D \mathcal{X}, \tag{11}$$

for some fixed distribution D (usually taken to be uniform) over the set of unknown signals $\mathcal{X}$. Mixtures of simple machine learning models, such as mixtures of distributions [110] and regression [111], have been extensively studied over decades to address heterogeneous data in areas like machine translation [112], health [113], medicine [114], and object recognition [115], etc. Mixtures of classifiers, in particular, are well-suited for modeling categorical data prevalent in these applications. The mixture of linear classifiers has been studied rigorously in the literature, providing algorithms for approximating the subspace spanned by component classifiers and for making predictions based on feature and label inputs [116].

In the active query model, some researchers have looked at the approximate recovery of L signals that are s-sparse within the mixture model [117]. The authors provide a two-stage non-adaptive algorithm that first recovers the support of all the signals using combinatorial matrices related to RUFFs (see Section 2.2.1 for details) and then approximately recovers the non-zero entries of each s-sparse signal using additional Gaussian queries. Note that similar to the two-stage recovery algorithms discussed in Section 2.2.1, the queries are completely non-adaptive, while the recovery algorithm proceeds in two stages. The algorithm uses $m = O(\epsilon^{-1}L^{L+3}s^{L+2}\log^2 n)$ measurements for ϵ-approximate recovery under the assumption that each signal contains one unique identifying feature. Other approaches forego this assumption and provide recovery with $m = O(L^2(Ls)^{\log L}\log^2 n)$ and $O(L^3 s(s)^L \log^2 n)$ measurements, respectively, and are also tolerant to noise in the query outputs [118,119]. One method uses tensor decomposition based on queries [118], while another provides both single-stage and two-stage algorithms [119]. However, these sample complexities remain distant from a basic counting-based lower bound of $\Omega(sL\frac{\log n}{\log(L+1)})$ and require runtime exponential in L and s. Moreover, in a different model, where queries cannot be designed but samples are drawn from a fixed distribution, approximate recovery can be achieved with as few as $O(L^3\log(Ls^{\log L}n))$ samples [120].

In addition to signal reconstruction and learning from a single measurement vector, 1bCS has also been extended to the recovery of multiple sparse vectors and adapted to accommodate various sparsity structures, incorporating both centralized and decentralized algorithms. These advancements are particularly significant for distributed sensor networks (DSNs). The following section explores this topic further.

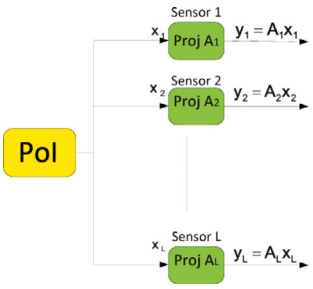


Fig. 2. Acquisition of compressed measurements of observations with temporal sparsity (from [121]), where PoI refers to the phenomenon of interest. These measurements are quantized to address communication constraints.

2.3. Centralized and decentralized multi-agent algorithms

1bCS has emerged as a promising approach for handling high-dimensional sparse signals in DSNs to reduce communication and extend network lifetime in power-constrained sensor networks monitoring various phenomena. In this section, we discuss some algorithms for quantized CS, motivated by resource-constrained DSNs.

One immediate application of quantized CS in DSNs is compressive data acquisition. When merging quantized CS and data acquisition from distributed sensing nodes, one needs to design processing at the compression side as well as at the reconstruction side, under communication constraints. On the compression side, the practical design of compression schemes via random projections depends on how the sparsity is exploited. On the reconstruction side, suitably extended quantized CS techniques for single and multiple measurement vector models to account for communication-related aspects are used. The quantized CS-based data acquisition problem can be formulated as a sparse signal reconstruction problem that may exploit temporal, spatial, and spatiotemporal sparsity. Here, we only briefly discuss the temporal sparsity case, and for other cases, the reader may refer to [121].

2.3.1. Temporal sparsity-aware algorithms

Consider a DSN shown in Fig. 2, where L sensor nodes observe the phenomenon of interest. The time samples collected at the ith node are represented by the vector $\boldsymbol{x}_i \in \mathbb{R}^n$. These time samples are assumed to be sparse in a suitable orthonormal basis. CS is applied to compress temporal sparse data in that only a small number of random projections are obtained via $\boldsymbol{y}_i = \boldsymbol{A}_i \boldsymbol{x}_i$ at the ith node for $i = 1, \dots, L$, where $\boldsymbol{A}_i \in \mathbb{R}^{m \times n}$ with $m < n$ is the measurement matrix used at the ith node. This compression via random projections is carried out independently at each node. The goal is to reconstruct $[\boldsymbol{x}_1, \dots, \boldsymbol{x}_L]$ based on their 1-bit quantized compressed versions $\boldsymbol{z}_i = Q_1(\boldsymbol{y}_i)$ communicated through the network where the reconstruction techniques depend on the specific communication architecture, centralized or decentralized, used to combine $\boldsymbol{z}_i$'s. For the sake of simplicity, only noiseless cases are considered in this section.

Using matrix notation, we consider recovering $\boldsymbol{X} \equiv [\boldsymbol{x}_1, \dots, \boldsymbol{x}_L]$ jointly based on quantized observations $\boldsymbol{Z} = Q_1(\boldsymbol{Y})$ with $\boldsymbol{Y} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_L]$. The multiple measurement vector model can be represented in matrix form with the same projection matrix, such that $\boldsymbol{A}_i = \boldsymbol{A}$ for $i = 1, \dots, L$,

$$\boldsymbol{Z} = Q_1(\boldsymbol{Y}) = Q_1(\boldsymbol{A}\boldsymbol{X}). \tag{12}$$

One may also consider the more general case with different projection matrices at the individual nodes, the observation matrix at the fusion center can then be expressed as

$$\boldsymbol{Z} = Q_1(\boldsymbol{Y}) = Q_1([\boldsymbol{A}_1 \boldsymbol{x}_1 \dots \boldsymbol{A}_L \boldsymbol{x}_L]). \tag{13}$$

When considering the simultaneous sparse approximation framework, $\boldsymbol{X}$ needs to be reconstructed when $\boldsymbol{Y}$ and $\boldsymbol{A}$ (with the same projection matrix) or $\boldsymbol{A}_1, \dots, \boldsymbol{A}_L$ (with different matrices) are given. Joint estimation of $\boldsymbol{X}$ can leverage joint or structured sparsity common in sensor networks [122].

A widely used sparsity model is the joint sparsity model [122]. In this model, the sparse signals observed at multiple nodes, $\boldsymbol{x}_i$'s, have the same but unknown sparsity pattern with respect to the same basis. However, the corresponding amplitudes can be different in general. The joint sparsity model with the same measurement matrix as in (12) is commonly termed as the multiple measurement vector model [123, 124]. While developing algorithms and evaluating performance with the joint sparsity model, several measures have been defined. To represent the number of nonzero elements of $\boldsymbol{X}$, we can use $\|\cdot\|_{\text{row-0}}$ norm where

$$\|\boldsymbol{X}\|_{\text{row-0}} = \{i \in \{1, \dots, n\} : \exists j \text{ s.t.} \boldsymbol{X}[i, j] \neq 0\}. \tag{14}$$

The natural approach to solve for sparse $\boldsymbol{X}$ from $\boldsymbol{Z}$ in (12) is to solve the following optimization problem

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_{\text{row-0}} \quad \text{s.t.} \quad \boldsymbol{Z} = Q_1(\boldsymbol{A}\boldsymbol{X}). \tag{15}$$

Centralized algorithms

This setting assumes that the nodes transmit their compressed measurements to a central fusion center with single-hop communication, and the fusion center solves the problem in (15). One approach to solving the problem is to use the ML decoder developed for sparsity pattern recovery [125]. The minimum number of one-bit compressive measurements that should be obtained per node to perform sparsity pattern recovery with a vanishing probability of error is determined via the ML approach. The results in [125] establish that $mL = \Omega\left(C_s s \log\left(\frac{n}{s}\right)\right)$ measurements are necessary for the noiseless setting, where C_s is a function of the sparsity level. Interestingly, the bound is on mL, implying that only a single measurement ($m = 1$) from some sparse vectors is enough to reliably recover the joint sparsity pattern if L is large enough. The implementation of the ML decoder becomes intractable as the signal dimension and the number of sensors increase. Yet, the performance bounds obtained through the ML algorithm serve as a benchmark for comparing suboptimal, computationally tractable algorithms.

Since the ML decoder is not computationally feasible, tractable algorithms for sparsity pattern recovery in a centralized setting are proposed in the literature [125]. In one algorithm, an optimization problem is formulated that minimizes the ML function and the $l_{1,\infty}$ quasi-norm of a matrix and uses the iterative shrinkage-thresholding algorithm. Specifically, it solves

$$\arg\min_{\boldsymbol{X}} f_{\text{ML}}(\boldsymbol{A}\boldsymbol{X}) + \bar{\lambda}\|\boldsymbol{X}\|_{\text{row-0}}, \tag{16}$$

where f_{ML} is the negative log-likelihood function (without imposing the sparsity constraint), $\bar{\lambda}$ is the penalty parameter, and $\|\cdot\|_{\text{row-0}}$ is defined in (14). Since optimizing $\|\cdot\|_{\text{row-0}}$ is hard, one often solves it by using the mixed norm approach. The relaxed convex problem is

$$\arg\min_{\boldsymbol{X}} f_{\text{ML}}(\boldsymbol{A}\boldsymbol{X}) + \bar{\lambda}\|\boldsymbol{X}\|_{1,\infty}, \tag{17}$$

where $\|\boldsymbol{X}\|_{1,\infty} = \sum_{i=1}^{n} \max_{1 \le j \le L} \boldsymbol{X}[i, j]$ is the $\ell_{1,\infty}$ quasi-norm of a matrix. Here, both $f_{\text{ML}}(\boldsymbol{A}\boldsymbol{X})$ and $\bar{\lambda}\|\boldsymbol{X}\|_{1,\infty}$ are convex functions and (17) can be solved using the standard convex solvers, such as the iterative shrinkage-thresholding algorithms.

The other algorithm extends the BIHT algorithm to the multiple measurement vector model available at the fusion center. Similar to BIHT discussed earlier in Section 2, in each iteration, the gradient of the cost function is evaluated using the previous iterate, and a step proportional to the gradient norm is taken in the negative direction to minimize the cost function. This process continues until a stopping criterion is met, with the final iteration providing the estimated support.

While centralized solutions discussed above are attractive in terms of performance, their power consumption due to direct communication with the fusion center may be high. So, they may not be practical in large or resource-constrained networks and we explore decentralized approaches for sparsity recovery.

Decentralized algorithms

In a decentralized setting, all the nodes in the network send their one-bit quantized measurement vectors to their one-hop neighbors. To define the notion of neighborhood, the decentralized network is modeled as an undirected graph, where the vertices represent the sensor nodes and the communication links correspond to the edges of the graph. Consequently, a node i can send its measurements to another node only if there is an edge connecting the ith node to the other node.

Two decentralized algorithms have been proposed in the literature, each featuring distinct stages that strategically embed the fusion of measurement information and decisions among nodes [125]. These collaborations among the nodes can be structured into two distinct stages, namely the information fusion stage and the index fusion stage. During the information fusion stage, each node estimates the support set using any centralized algorithm, convex relaxation- or BIHT-based methods described earlier. Because each node only has access to measurements from its neighbors, the problem size is smaller than that of fusion center-based recovery. In the index fusion stage, the nodes send their estimates to all other nodes in the network, which then arrive at the final support estimate using simple fusion methods, such as the majority fusion rule, which counts the s-most frequently occurring indices among the estimates. These algorithms assume the prior knowledge of the sparsity level of the signal. If the sparsity levels of signals are not known in advance, they need to be estimated using methods proposed in the literature. A comparison of the performance of these proposed decentralized algorithms with those of centralized ones and their real-valued counterparts shows excellent performance, demonstrating their value in resource-constrained environments.

2.3.2. Distributed inference

DSNs are also employed in various inference problems. In such applications, 1bCS techniques can again be used as a means of data compression where the goal is to solve an inference problem and not signal reconstruction. In inference problems such as detection [126–128], classification [61,91,106], and parameter estimation [129–131], it is sufficient to construct a reliable decision statistic based on compressed data without recovering the original signal. Beyond the standard 1bCS framework, this requires the investigation of different metrics for performance analysis and quantification of the amount of information preserved under compression to obtain a reliable inference decision. In this section, we briefly discuss the compressive detection problem with one-bit measurements. For discussion on classification and parameter estimation problems, see [121].

Event or object detection is an important task performed by DSNs [132]. In order to solve a detection problem efficiently in a multi-sensor setup, one needs to process the signals at the sensing nodes that maximally contribute to the overall decision-making prior to the transmission of its processed information to the fusion center. In order to minimize the amount of information to be transmitted, we may employ quantized CS. When the number of active events is much less than the number of all possible events (rare events), the event detection problem can be formulated as a sparse recovery problem. Consider the scenario shown in Fig. 3, where there are n sources scattered over a region out of which s are active simultaneously. The scenario can arise in various applications such as radar surveillance [133] and cognitive radios [134]. The quantized measurement vector at m active sensors has the following form

$$\boldsymbol{z} = Q_1(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}), \tag{18}$$

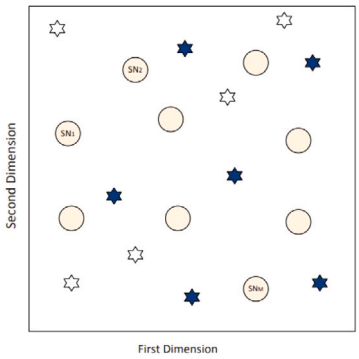


Fig. 3. Sparse event monitoring; [135], dark stars-active sources, void stars-inactive sources, circles-sensor nodes.

where $\boldsymbol{x}$ is a sparse vector with $\boldsymbol{x}[i] \in 0, 1$. The (i, j)th element of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}[i, j] = r_{i,j}^{-\tilde{\alpha}/2} |h_{i,j}|, \tag{19}$$

where $r_{i,j}$ denotes the distance from the ith sensor to the jth source, $\tilde{\alpha}$ is the propagation loss factor and $h_{i,j}$ is the fading coefficient of the channel between the ith sensor and the jth source. In this framework, the sparse event monitoring problem reduces to estimating the sparse vector $\boldsymbol{x}$ from (18) when the elements of $\boldsymbol{A}$ are as given in (19). Here, $\boldsymbol{A}$ is not a user-defined random matrix satisfying restricted isometric properties as the standard CS framework desires. Rather, the randomness arises due to the random locations and fading coefficients.

The problem of distributed detection of sparse stochastic signals can be framed as a binary hypothesis testing problem as

$$\mathcal{H}_0 : \boldsymbol{z} = Q_1(\boldsymbol{w}), \tag{20}$$

$$\mathcal{H}_1 : \boldsymbol{z} = Q_1(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}). \tag{21}$$

The study in [126] has introduced a one-bit detector using the Generalized Likelihood Ratio Test for distributed detection of sparse deterministic signals, which requires full signal reconstruction. Meanwhile, [127] presents a one-bit detector based on the Locally Most Powerful Test for distributed detection of sparse stochastic signals, where sparse signals are modeled using a Bernoulli-Gaussian distribution. This work avoids the need for full signal reconstruction, making it more efficient. The approach in [128] improved this distributed detector by quantizing the likelihood ratios to generate the one-bit data instead of directly quantizing the analog observations.

This concludes our discussion on generalized quantized CS algorithms. So far, we have examined the theory and algorithms for 1bCS, exploring various approaches. Over the last decade, quantized CS has gained traction across various fields, including medical imaging, machine learning, computer science, and statistics. These general theories and algorithms can be further refined and adapted for specific applications, with additional enhancements through the integration of application-specific tasks. The next section focuses on the customized quantized CS approaches for wireless communication and sensing applications.

3. Quantized CS algorithms for wireless communication and sensing

In wireless sensing, quantized CS is commonly used, with 1bCS standing out for its simplicity, cost-effectiveness, low resource requirements, and robustness to certain linear and nonlinear distortions, such as saturation errors. This section explores its diverse applications in both wireless communication and sensing, highlighting key advancements, methodologies, and their impact on system performance and efficiency.

Table 3
Summary of different 1bCS algorithms discussed in Section 2 applied to wireless communication and sensing applications.

Domain	Methods
IoT/Sensor Networks	ℓ_1-normed based [16,136]
	Log-sum penalty function-based [15]
	Blind BIHT [14]
Radar Imaging	Sparse logistic regression [137,138]
	BIHT variants [30,139,140]
	Bayesian method [28]
Radar Direction Finding	Norm optimization [141–143]
	BIHT variants [144,145]
	Bayesian methods [146,147]
Cognitive Radio Networks	Norm minimization [23]
	BIHT variants [18,148,149]
Wireless Channel Estimation	Norm optimization [150,151]
	Approximate message passing [25,152,153]
	SBL [154–156]
Feedback-based Channel Estimation	BIHT variants [26,157–159]
	Bayesian methods [24]

3.1. Generic quantized CS algorithms

Quantized CS is widely used in wireless sensor networks, cognitive radio, radar, and direction of arrival estimation, to address bandwidth constraints, energy efficiency, and reliable sensing under limited resources. Its ability to operate with low-resolution measurements while maintaining useful signal recovery makes it particularly valuable in these applications. Below, we review the relevant literature, summarized in Table 3, that focuses on adapting the general framework of quantized CS algorithms across these domains.

3.1.1. IoT and wireless sensor networks

IoT and wireless sensor networks consist of small, low-power sensors that monitor the environment, periodically collecting data and transmitting it to the fusion center via shared wireless channels. Since sparsity is a common feature of many signals (temperature, humidity, illumination, etc.) monitored by the sensors, CS frameworks are widely used in wireless sensor networks. Given the limited computational and energy resources of sensor nodes, 1bCS is particularly suitable as it enables accurate signal reconstruction while minimizing storage costs and hardware complexity.

One of the earliest works to apply 1bCS for data gathering in sensor networks appears in [14]. This work adapts BIHT to estimate the sparsity level of unknown data by analyzing the variance among solutions of the BIHT algorithm at different sparsity levels and selecting the best one. This modified version, called blind BIHT, improves compression efficiency and reduces communication costs. In [15], 1bCS is utilized for source localization in wireless sensor networks, by introducing a Gaussian entropy (log-sum penalty function)-based method for sparse signal recovery. A memory-efficient 1bCS algorithm using a circulant random bipolar measurement matrix is proposed in [136], to demonstrate that 1bCS can lead to less data traffic in sensor networks. Additionally, [16] examines 1bCS in noisy wireless sensor networks affected by channel-induced bit-flipping errors. To mitigate these errors, an amplitude-aided signal reconstruction scheme is proposed, improving accuracy in low-SNR conditions or when the number of sensors is limited.

Another important aspect of using quantized CS in wireless sensor networks is analyzing how efficiently sparse signals are represented, particularly from an information-theoretic perspective. This involves determining whether quantized CS measurements provide an effective representation and how to optimize the quantizers themselves. Early works in quantized CS, such as [152,160,161], optimize scalar quantizers for CS reconstruction, while quantization-aware decoding algorithms for fixed encoders are proposed in [162]. Although these methods improve performance over quantization-unaware versions, they are suboptimal for minimizing the mean squared error in signal reconstruction. The issues include optimizing only the encoder or decoder, using scalar quantization instead of vector quantization, and minimizing measurement quantization distortion, which does not always minimize reconstruction distortion due to non-linearities. To address this, joint optimization of vector quantization-based encoder–decoder pairs is proposed in [163,164], aiming to minimize the mean squared error in quantized CS. However, these methods are computationally expensive, hindering practical implementation. Furthermore, the early results in [165] and further analysis in [166] explore lossy CS, while additional work on distortion–rate bounds and remote compression appear in [167,168]. Further, [169] analyzes the rate–distortion performance of quantized CS measurements in wireless sensor networks using three reconstruction methods: ℓ_1-norm minimization, A*OMP, and LASSO. This study presents an incremental transmission scheme that refines sensor measurements at the fusion center on demand, reducing energy consumption and achieving rate–distortion performance close to the optimum. Later, [170] studies the rate–distortion performance of various single-sensor quantized CS schemes for compressing sparse signals using noisy quantized measurements. They propose three practical methods: compress-and-estimate, estimate-and-compress, and support-estimation-and-compress. Further, [171] presents practical symbol-by-symbol quantizer-based methods for different compression strategies, assessing the compression limit of quantized CS through an analytical lower bound and numerical approximation. Using high-resolution functional scalar quantization, they show noticeable improvements in operational distortion–rate performance.

3.1.2. Radar systems

In synthetic aperture radar systems, low-bit encoding is used to reduce transmission costs through quantized CS. Early research on one-bit radar imaging with conventional methods, such as matched filtering [172,173], shows that one-bit quantization can lead to ghost targets due to the loss of magnitude information from fixed threshold ADCs. Since targets with strong scattering coefficients are sparsely distributed, 1bCS methods provide an alternative for improved one-bit radar imaging by leveraging the sparse representation of radar echoes without increasing hardware complexity. These methods either compensate for information loss caused by quantization through random and adaptive quantization thresholds [129,174] or leverage additional information during the imaging process [28]. For example, a maximum a posteriori approach in [28] effectively suppresses ghost targets and artifacts. In [30], an enhanced one-bit radar imaging algorithm based on the BIHT framework exploits two-level block sparsity to account for clustering and joint sparsity patterns. However, this method's performance is often degraded by noise in the data sampling and transmission process. To address this, robust 1bCS algorithms based on BIHT are developed to handle sign flip errors effectively [139]. These algorithms introduce an adaptive quantization level, iteratively updated with the imaging result using a relaxed quantization consistency condition to reduce noise and improve reconstruction quality. Additionally, an adversarial sample-based BIHT method [140] uses adversarial samples to train and adapt quantization level parameters, ensuring quantization consistency. However, for large-scale one-bit imaging, these approaches are time-consuming, and so, sparse logistic regression-based one-bit synthetic aperture radar imaging is introduced [137]. This approach combines efficient sparse logistic regression-solving techniques with CS imaging constraints to achieve good results with low run time and excellent convergence. More recently, this framework has been enhanced by utilizing two-level structured sparsity [138] that encourages clustered sparsity patterns and suppresses high-intensity artifacts and clutter caused by sign flips.

Direction estimation is another key problem that leverages quantized CS. The sparse vector recovery-based direction of arrival estimators, such as fixed-point continuation [58,141] and BIHT, perform well even with limited snapshots. BIHT has been extended to complex-valued and multi-snapshot cases, leading to the complex-valued BIHT algorithm [144]. Another work exploits the generalized linear model inference problem and extends the generalized SBL algorithm for multiple snapshots, achieving better estimation accuracy than BIHT by exploiting the joint sparsity of the real and imaginary components [146]. More recently, deep learning-inspired algorithms like the deep unfolded version of fixed point continuation [142] have been proposed for one-bit direction-of-arrival estimation. These approaches are categorized as on-grid methods since they determine the direction by searching within predefined grid points. However, they suffer from errors when the actual direction of arrival falls between grid points (off-grid), necessitating denser grids, which increase computational complexity and reduce stability. To address this, off-grid DOA estimation techniques have been developed. One approach enhances the complex BIHT algorithm with gradient descent for signal estimation and bases-updating and backtracking strategies to enhance off-grid estimation accuracy and convergence [145]. Other approaches include atomic norm minimization using the alternating direction method of multipliers [143], and an off-grid iterative Bayesian algorithm within the block successive upper-bound minimization framework [147].

3.1.3. Cognitive radio networks

Cognitive radio networks consist of wireless users equipped with sensing capabilities to improve the efficiency of spectrum utilization. It enhances spectrum efficiency by allowing secondary users to transmit on unused portions of the spectrum. However, the secondary users must vacate the spectrum when the primary user reoccupies it. The key feature of cognitive radios is spectrum sensing, which enables them to accurately determine the availability of spectrum [175,176]. In many applications, wideband channels must be monitored, requiring high-speed sampling. However, traditional spectrum sensing methods typically rely on high-precision quantization for optimal performance, which leads to significant energy consumption, motivating low-resolution (particularly one-bit) spectrum sensing.

Several studies have demonstrated that quantized CS can reduce spectrum sensing costs [23]. For example, [23] examines the tradeoff between computation cost and compression performance, highlighting the communication cost savings of 1bCS in spectrum sensing for networked systems. It also proposes a block reconstruction algorithm that leverages the block sparsity of the signals. Further, to tackle the high sampling rate challenges in cognitive radio devices, a modulated wideband converter is typically used to sample sparse multiband signals. Using the 1bCS framework, [18] proposes an alternative sub-Nyquist sampling system using comparators for efficient space utilization and low bit-budget, demonstrating its advantages over modulated wideband converter (especially at low input SNR).

The framework has also been extended to distributed collaborative spectrum sensing, where secondary users share their measurements and make a common decision [148,149]. These approaches leverage joint sparsity and spatial diversity through average consensus, guiding local signal reconstruction with the weighted BIHT algorithm [149]. Here, local reconstruction and fusion alternate until reliable spectrum detection is achieved. Additionally, sub-Nyquist sampling at the output of wide receive front-end filters, combined with group testing concepts, has been explored for wideband spectrum sensing [177]. In this context, the combinatorial group testing problem parallels 1bCS, with the unknown sparse spectrum occupancy and measurement matrix approximated as a binary vector and matrix, respectively. Further, in wideband spectrum sensing scenarios where multi-user cooperation is infeasible and prior sparsity information is unavailable, non-cooperative spectrum sensing is needed. In this context, [178] uses a multicoset sampling framework to achieve sub-Nyquist sampling without requiring sparsity knowledge and introduces a subspace-aided 1bCS algorithm for spectrum support estimation without signal reconstruction.

3.1.4. Wireless channel estimation

In wireless communications, massive MIMO is a key enabling technology to meet the increasing demand for data rate and energy efficiency [179,180]. However, the advantages of massive MIMO come at the cost of increased power consumption and hardware complexity due to the large number of radio frequency (RF) chains, high precision ADCs, etc. In particular, the power consumption of ADCs grows exponentially with the number of quantization bits per sample [181–183]. Also, full precision ADCs require correspondingly high rate data processing at the receiver. This has spurred interest in employing low-resolution ADCs in the base station (BS) of a massive MIMO system [5,151,184–186]. In such systems, quantized CS has paved the way for innovative applications that capitalize on signal sparsity, such as interference cancellation, angle-of-arrival estimation, channel estimation, and symbol detection.

An example of a quantized sparse estimation problem in wireless communication is channel estimation, which is arguably the most extensively studied application in the field. Channel estimation with low-resolution ADCs explores the sparse representation of the wireless channel in either the delay domain or angular domain. One algorithmic technique involves approximating the nonlinear quantization effects of one-bit ADCs into a linear form using the Bussgang decomposition, enabling linear minimum mean squared error (MMSE) channel estimation [25,150,151]. Another approach is the generalized approximate message passing algorithm, initially developed in CS to handle quantized measurements [152], which has been adapted for channel estimation [153]. Additionally, some studies have explored the SBL framework for channel estimation [154] and time-varying channel tracking [155,156].

Furthermore, in frequency division duplex massive MIMO systems, channel reciprocity is unavailable due to the uncorrelated uplink and downlink spectral bands. Therefore, the downlink channel state information (CSI) is fed back to the BS through the uplink channel [187]. While initial codebook-based methods reduce feedback overhead, 1bCS approaches have enhanced CSI accuracy by exploiting channel sparsity and user correlation. Notable methods include complex 1-bit Bayesian CS [24] and BIHT-based distributed 1bCS, where received symbols are quantized to one bit per dimension [157]. The method in [157] is enhanced using partial amplitude information with algorithms such as quantized partially joint orthogonal matching pursuit and quantized partially joint iterative hard thresholding [158]. It is further extended to off-grid channel estimation in [159]. To save uplink bandwidth, a 1-bit CS-based CSI feedback method assisted by superimposed coding is proposed. In this method, downlink CSI is superimposed on uplink user data sequences and fed back to the BS. The 1bCS channel estimation algorithms at the BS with superimposed coding include BIHT-based recovery [26] and deep learning enhancements [188,189].

3.2. Application-tailored quantized CS algorithms

While several studies have explored quantized CS for wireless sensing and communication applications, most of them are limited to 1bCS. This is not surprising, as the general theory of quantized CS has often gravitated toward one-bit quantization. A notable exception is wireless channel estimation, where low-resolution CS techniques, beyond one-bit quantization, are also widely explored. Many of these studies do not employ generalized algorithms for quantized CS, as the unique characteristics of each context demand more tailored approaches. In the following, we investigate two distinct flavors of MIMO channel estimation problems with low-resolution (more than one-bit) ADCs. Section 3.2.1 delves into the integrated challenges of channel estimation and data decoding in multi-user MIMO-orthogonal frequency-division multiplexing (OFDM) systems, while Section 3.2.2 focuses on channel estimation for reconfigurable intelligent surfaces (RIS)-aided MIMO systems in narrowband scenarios. Each problem presents its own formulation, motivating the need for specialized algorithms that

address the domain-specific needs of the underlying application. The aim of this section is not to provide a comprehensive survey on quantized CS algorithms in wireless communication but to illustrate how generalized theories and approaches can be enhanced through the use of tailored algorithms, optimizing performance based on the particular requirements of each application.

3.2.1. Example 1: Channel estimation and soft symbol decoding

In this subsection, we present an iterative delay domain *sparse channel estimation* and *soft symbol decoding* algorithm for a massive MIMO-OFDM system with low-resolution ADCs. Three challenges arise in using low-resolution ADCs in multi-user MIMO-OFDM systems: degraded performance of traditional detectors due to non-linearities of coarse quantization [190], the need for long pilot sequences for accurate channel estimation reducing spectral efficiency as pilot signals are also quantized [151,182,191], and the requirement for bit log likelihood ratio (LLR)s for channel decoders, necessitating the computation of posterior beliefs of data symbols (also known as *soft symbols*) based on the quantized observations. We exploit the sparsity of the channels in their time-domain representation to obtain the posterior distributions of the channels. We also present a quantized variational Bayesian (QVB) *soft symbol decoding* procedure that uses the estimated channels to obtain the posterior beliefs of the data symbols. We iterate between channel estimation and soft-symbol detection to further refine the channel and soft-symbol estimates.

Description of the massive MIMO-OFDM system and problem statements

We consider the uplink of a single cell massive MIMO-OFDM system with M antennas at the BS and K single antenna user equipments (UEs), where $M \geq K$. Each UE encodes its information bits and maps them to constellation symbols. The symbols are then mounted onto subcarriers and OFDM modulated using an inverse discrete Fourier transform (DFT). The OFDM modulated data symbols are parallel-to-serial converted, a cyclic prefix is added, RF up-converted to the passband, and transmitted over a frequency-selective wireless channel to the BS. At the BS, the received RF signal is down-converted to baseband, sampled, and quantized using B-bit ADCs to obtain the received complex baseband signal.

Each UE transmits τ_{p} pilot OFDM symbols followed by τ_{d} data OFDM symbols. We assume that the coherence interval of the channel is at least $\tau_{\mathrm{p}}+\tau_{\mathrm{d}}$ OFDM symbols. We denote the number of subcarriers by N_{c}. After some algebra, it can be shown that the received time-domain pilot OFDM symbols $\boldsymbol{Z}^{(\mathrm{p})} \in \mathbb{C}^{\tau_{\mathrm{p}} N_{\mathrm{c}} \times M}$ can be expressed as [192][2]

$$\boldsymbol{Z}^{(\mathrm{p})} = \begin{bmatrix} (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \boldsymbol{X}^{(\mathrm{p})}(1) \left(\boldsymbol{I}_K \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}},n} \right), \\ \vdots \\ (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \boldsymbol{X}^{(\mathrm{p})}(\tau_{\mathrm{p}}) \left(\boldsymbol{I}_K \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}},n} \right) \end{bmatrix} \boldsymbol{H} + \boldsymbol{W}^{(\mathrm{p})} \triangleq \boldsymbol{\Phi}^{(\mathrm{p})} \boldsymbol{H} + \boldsymbol{W}^{(\mathrm{p})}, \tag{22}$$

where $\otimes$ denote the Kronecker product, $\boldsymbol{\Phi}^{(\mathrm{p})} \in \mathbb{C}^{\tau_{\mathrm{p}} N_{\mathrm{c}} \times Kn}$, $\boldsymbol{\Psi}_{N_{\mathrm{c}}}$ is the DFT matrix, $\boldsymbol{\Psi}_{N_{\mathrm{c}},n}$ and the n column truncated DFT matrix. Also, $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_M] \in \mathbb{C}^{Kn \times M}$ is a *row-sparse* channel matrix in the time domain with the channel between each user and each receive antenna represented as an n-tap, s-sparse complex vector, $\boldsymbol{X}^{(\mathrm{p})}(t) = \mathrm{diag}(\boldsymbol{X}_1^{(\mathrm{p})}(t), \ldots, \boldsymbol{X}_K^{(\mathrm{p})}(t)) \in \mathbb{C}^{KN_{\mathrm{c}} \times KN_{\mathrm{c}}}$, with $\boldsymbol{X}_k^{(\mathrm{p})}(t)$ is a diagonal matrix with the pilots loaded on the subcarriers as its entries, and $\boldsymbol{W}^{(\mathrm{p})}$ is the additive white Gaussian noise matrix. The channel matrix $\boldsymbol{H}$ has Kn rows, with each set of n consecutive rows representing the n potential delay taps per user. Only s taps have nonzero coefficients, making $\boldsymbol{H}$ row-sparse, which we exploit in the channel estimation procedure.

[2] The interested reader is referred to [192] for complete details on the material presented in this subsection. Further results on variational Bayesian channel estimation and data detection can be found in [193–199].

The received signal is quantized using a uniform quantizer with intervals $\xi^{(l)} = (-B/2 + l)\Delta$ for $l = 1, \ldots, B-1$, and quantization levels at $(\xi^{(l)} + \xi^{(l+1)})/2$. The dynamic range is set using the expected signal power P_R, with $\xi^{(0)} = -2.5\sqrt{P_R/2}$ and $\xi^{(B)} = 2.5\sqrt{P_R/2}$. The values outside this range are clipped to $\pm(B-1)\Delta/2$. The received pilots are quantized using B-bit ADCs as

$$\boldsymbol{Y}^{(\mathrm{p})} = \mathcal{Q}_B(\boldsymbol{Z}^{(\mathrm{p})}) = \mathcal{Q}_B(\boldsymbol{\Phi}^{(\mathrm{p})} \boldsymbol{H} + \boldsymbol{W}^{(\mathrm{p})}) \in \mathbb{C}^{\tau_{\mathrm{p}} N_{\mathrm{c}} \times M}. \tag{23}$$

Our first goal to estimate $\boldsymbol{H}$ given $\boldsymbol{Y}^{(\mathrm{p})}$ and $\boldsymbol{\Phi}^{(\mathrm{p})}$ in (23). After estimating $\boldsymbol{H}$, our goal is to decode the data symbols in the data transmission phase.

In the data transmission phase, we can vectorize and stack the signal received over the M receive antennas and τ_{d} OFDM data symbols to obtain the frequency-domain received data symbols $\boldsymbol{Z}^{(\mathrm{d})} \in \mathbb{C}^{MN_{\mathrm{c}} \times \tau_{\mathrm{d}}}$ as

$$\boldsymbol{Z}^{(\mathrm{d})} = \begin{bmatrix} (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \boldsymbol{H}_1^{\mathrm{freq}} \\ (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \boldsymbol{H}_2^{\mathrm{freq}} \\ \vdots \quad \ddots \quad \vdots \\ (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \boldsymbol{H}_M^{\mathrm{freq}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^{(\mathrm{d})}(\tau_{\mathrm{p}}+1) & \ldots & \boldsymbol{x}^{(\mathrm{d})}(\tau_{\mathrm{p}}+\tau_{\mathrm{d}}) \end{bmatrix} + \boldsymbol{W}^{(\mathrm{d})} = \bar{\boldsymbol{D}} \boldsymbol{X}^{(\mathrm{d})} + \boldsymbol{W}^{(\mathrm{d})}, \tag{24}$$

where $\bar{\boldsymbol{D}} \in \mathbb{C}^{MN_{\mathrm{c}} \times KN_{\mathrm{c}}}$ is the measurement matrix for data detection, $\boldsymbol{H}_{n_r}^{\mathrm{freq}} \in \mathbb{C}^{KN_{\mathrm{c}} \times KN_{\mathrm{c}}}$ is a diagonal matrix containing the frequency-domain representation of the channel between the K users and the n_rth receive antenna, $\boldsymbol{X}^{(\mathrm{d})} \in \mathbb{C}^{KN_{\mathrm{c}} \times \tau_{\mathrm{d}}}$ is the transmit data matrix, and $\boldsymbol{W}^{(\mathrm{d})}$ is the additive white Gaussian noise matrix during the data phase. Now, we quantize the received signal (24) using the B-bit ADCs to obtain

$$\boldsymbol{Y}^{(\mathrm{d})} = \mathcal{Q}_B(\boldsymbol{Z}^{(\mathrm{d})}) = \mathcal{Q}_B\left(\bar{\boldsymbol{D}} \boldsymbol{X}^{(\mathrm{d})} + \boldsymbol{W}^{(\mathrm{d})}\right). \tag{25}$$

Our goal in this part is to decode the data symbols $\boldsymbol{X}^{(\mathrm{d})}$ given $\boldsymbol{Y}^{(\mathrm{d})}$ and $\bar{\boldsymbol{D}}$. Once we estimate the posterior posterior distribution of $\boldsymbol{X}^{(\mathrm{d})}$, we will perform data-aided channel estimation to refine the channel estimates, as described next.

Recall that the unquantized received pilot signal can be expressed as $\boldsymbol{Z}^{(\mathrm{p})} = \boldsymbol{\Phi}^{(\mathrm{p})} \boldsymbol{H} + \boldsymbol{W}^{(\mathrm{p})}$. Similar to the pilot reception phase, if we consider the decoded data as known virtual pilot symbols, then we can write the received data signal as

$$\boldsymbol{Z}^{(\mathrm{d})}(t) = (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \left\langle \boldsymbol{X}^{(\mathrm{d})}(t) \right\rangle \left(\boldsymbol{I}_K \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}},n} \right) \boldsymbol{H} + \boldsymbol{W}^{(\mathrm{d})}(t), \tag{26}$$

where $t = \{\tau_{\mathrm{p}}+1, \ldots, \tau_{\mathrm{p}}+\tau_{\mathrm{d}}\}$ with $\left\langle \boldsymbol{X}^{(\mathrm{d})}(t) \right\rangle \in \mathbb{C}^{KN_{\mathrm{c}} \times KN_{\mathrm{c}}}$ given by

$$\left\langle \boldsymbol{X}^{(\mathrm{d})}(t) \right\rangle = \mathrm{diag}\left(\left\langle \boldsymbol{X}_1^{(\mathrm{d})}(t) \right\rangle, \ldots, \left\langle \boldsymbol{X}_K^{(\mathrm{d})}(t) \right\rangle \right). \tag{27}$$

Here, $\left\langle \boldsymbol{X}_k^{(\mathrm{d})}(t) \right\rangle = \mathrm{diag}\left(\left\langle \boldsymbol{x}_k^{(\mathrm{d})}(t) \right\rangle \right) \in \mathbb{C}^{N_{\mathrm{c}} \times N_{\mathrm{c}}}$ with $\left\langle \boldsymbol{x}_k^{(\mathrm{d})}(t) \right\rangle$ being the posterior means of the decoded data symbols of the kth user during the tth OFDM symbol. We stack $\boldsymbol{Z}^{(\mathrm{p})}$ and $\boldsymbol{Z}^{(\mathrm{d})}(t)$ to obtain an expression for the unquantized received signal over one coherence interval as

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}^{(\mathrm{p})} \\ \boldsymbol{Z}^{(\mathrm{d})}(\tau_{\mathrm{p}}+1) \\ \vdots \\ \boldsymbol{Z}^{(\mathrm{d})}(\tau_{\mathrm{p}}+\tau_{\mathrm{d}}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}^{(\mathrm{p})} \\ (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \left\langle \boldsymbol{X}^{(\mathrm{d})}(\tau_{\mathrm{p}}+1) \right\rangle \left(\boldsymbol{I}_K \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}},n} \right) \\ \vdots \\ (\mathbf{1}_K^\top \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}}}^{\mathsf{H}}) \left\langle \boldsymbol{X}^{(\mathrm{d})}(\tau_{\mathrm{p}}+\tau_{\mathrm{d}}) \right\rangle \left(\boldsymbol{I}_K \otimes \boldsymbol{\Psi}_{N_{\mathrm{c}},n} \right) \end{bmatrix} \boldsymbol{H} + \boldsymbol{W} \triangleq \boldsymbol{\Phi} \boldsymbol{H} + \boldsymbol{W}, \tag{28}$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{(\tau_{\mathrm{p}}+\tau_{\mathrm{d}})N_{\mathrm{c}} \times Kn}$ is the augmented measurement matrix and $\boldsymbol{W} \in \mathbb{C}^{(\tau_{\mathrm{p}}+\tau_{\mathrm{d}})N_{\mathrm{c}} \times M}$ is the additive white Gaussian noise matrix. The B-bit quantized received signal is given by

$$\boldsymbol{Y} = \mathcal{Q}_B(\boldsymbol{Z}) = \mathcal{Q}_B(\boldsymbol{\Phi} \boldsymbol{H} + \boldsymbol{W}) \in \mathbb{C}^{(\tau_{\mathrm{p}}+\tau_{\mathrm{d}})N_{\mathrm{c}} \times M}. \tag{29}$$

Our goal is to estimate $\boldsymbol{H}$ given $\boldsymbol{Y}$ and $\boldsymbol{\Phi}$. Once we estimate $\boldsymbol{H}$, we use it to obtain $\bar{\boldsymbol{D}}$ as in (25), which in turn is used to refine the posterior beliefs of the D-quadrature amplitude modulated data symbols in the following data decoding iteration.

In the next three subsections, we present a QVB-based solution to the above channel estimation and data detection problem.

Quantized variational Bayesian algorithm

To develop the algorithm, we first derive the channel estimation and soft symbol decoding algorithms, and then combine them into an iterative algorithm for joint channel estimation and soft symbol decoding.

Quantized variational Bayesian channel estimation. Our goal is to infer the posterior distributions of the channels and the LLRs of the data symbols, given the quantized pilot and data observations. We adopt a probabilistic graphical model for this statistical inference, but exact posterior computation is intractable due to high-dimensional integrals over the channels and data symbols. So, we use approximate inference techniques, replacing the exact posterior distribution with a tractable distribution close to the original. An excellent introduction to variational Bayesian inference can be found in [200].

In order to exploit the lag-domain sparsity in the channel, as in [50], we use a two-stage hierarchical prior on $\boldsymbol{H}$, i.e., $\forall\, i$, $\boldsymbol{h}_i \sim \mathcal{CN}(\boldsymbol{h}_i; \boldsymbol{0}, \boldsymbol{P}^{-1})$, where the precision matrix $\boldsymbol{P}$ is diagonal and contains the hyperparameters $\boldsymbol{\alpha} = [\boldsymbol{\alpha}[1], \ldots, \boldsymbol{\alpha}[Kn]]^{\top}$ as its diagonal elements. Further, we impose a Gamma hyperprior on $\boldsymbol{\alpha}$. This results in a Student's t-distributed prior on $\boldsymbol{h}_i$, which is known to promote sparse channel estimates [201]. We express the logarithm of the joint probability distribution of the observations and latent variables as

$$\begin{aligned}&\ln p(\boldsymbol{Y}^{(\mathrm{p})}, \boldsymbol{Z}^{(\mathrm{p})}, \boldsymbol{H}, \boldsymbol{\alpha}; \boldsymbol{\Phi}^{(\mathrm{p})}, \sigma_w^2, a, b)\\ &= \ln p(\boldsymbol{Y}^{(\mathrm{p})} \mid \boldsymbol{Z}^{(\mathrm{p})}) + \ln p(\boldsymbol{Z}^{(\mathrm{p})} \mid \boldsymbol{H}; \boldsymbol{\Phi}^{(\mathrm{p})}, \sigma_w^2) + \ln p(\boldsymbol{H} \mid \boldsymbol{P}) + \ln\, p(\boldsymbol{\alpha}; a, b),\end{aligned} \tag{30}$$

where the prior distributions of $\boldsymbol{H}$ and $\boldsymbol{\alpha}$ are given by

$$\begin{aligned}p(\boldsymbol{H} \mid \boldsymbol{P}) \quad &= \prod_{i=1}^{M} \frac{|\boldsymbol{P}|}{\pi^{Kn}} \exp\left(-\boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{P} \boldsymbol{h}_i\right) \text{ and}\\ p(\boldsymbol{\alpha}; a, b) \quad F &= \prod_{k=1}^{Kn} \frac{b^a}{\Gamma(a)} \boldsymbol{\alpha}[k]^{a-1} \exp\left(-b\boldsymbol{\alpha}[k]\right),\end{aligned} \tag{31}$$

and $\Gamma(\cdot)$ is the Gamma function. We set a and b to small values (say, 10^{-4}) such that the hyperprior $p(\boldsymbol{\alpha}; a, b)$ is non-informative. We approximate the posterior distribution of the latent variables as the factorized distribution:

$$p(\boldsymbol{Z}^{(\mathrm{p})}, \boldsymbol{H}, \boldsymbol{\alpha} \mid \boldsymbol{Y}^{(\mathrm{p})}; \boldsymbol{\Phi}^{(\mathrm{p})}, \sigma_w^2, a, b) \approx q_{\boldsymbol{H}}(\boldsymbol{H}) q_{\boldsymbol{Z}}(\boldsymbol{Z}^{(\mathrm{p})}) q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}), \tag{32}$$

$$= \prod_{i=1}^{M} q_{\boldsymbol{h}_i}(\boldsymbol{h}_i) \prod_{i=1}^{M} q_{\boldsymbol{z}_i}(\boldsymbol{z}_i^{(\mathrm{p})}) \prod_{k=1}^{Kn} q_{\boldsymbol{\alpha}[k]}(\boldsymbol{\alpha}[k]), \tag{33}$$

where we define $\boldsymbol{Z}^{(\mathrm{p})} \triangleq [\boldsymbol{z}_1^{(\mathrm{p})}, \ldots, \boldsymbol{z}_M^{(\mathrm{p})}]$. Next, we express the conditional probability distributions of the observations and latent variables that are needed to compute the posterior distributions under the factorized structure as

$$p(\boldsymbol{Y}^{(\mathrm{p})} \mid \boldsymbol{Z}^{(\mathrm{p})}) \triangleq \prod_{i=1}^{M} \mathbb{I}\left(\boldsymbol{z}_i^{(\mathrm{p})} \in \left(\boldsymbol{z}_i^{(\mathrm{lo})}, \boldsymbol{z}_i^{(\mathrm{up})}\right)\right), \tag{34}$$

$$p(\boldsymbol{Z}^{(\mathrm{p})} \mid \boldsymbol{H}; \boldsymbol{\Phi}^{(\mathrm{p})}, \sigma_w^2) = \prod_{i=1}^{M} \frac{1}{(\pi\sigma_w^2)^{\tau_{\mathrm{p}} N_{\mathrm{c}}}} \exp\left(-\frac{1}{\sigma_w^2} \|\boldsymbol{z}_i^{(\mathrm{p})} - \boldsymbol{\Phi}^{(\mathrm{p})} \boldsymbol{h}_i\|^2\right), \tag{35}$$

where $\mathbb{I}(\cdot)$ is the indicator function, $\boldsymbol{z}_i^{(\mathrm{lo})}$ and $\boldsymbol{z}_i^{(\mathrm{up})}$ are the lower and upper quantization thresholds corresponding to the ith column of $\boldsymbol{Y}^{(\mathrm{p})}$, respectively. The posterior distributions of the latent variables are computed by finding the expectations of the logarithm of the joint distribution (30) with respect to the latent variables, and are provided in closed form in the following.

The posterior distribution $q_{\boldsymbol{H}}(\boldsymbol{H})$ is complex normal with the covariance matrix of each of its columns and mean given by

$$\boldsymbol{\Sigma}_{\boldsymbol{H}} = \left(\frac{1}{\sigma_w^2} \boldsymbol{\Phi}^{(\mathrm{p})\mathsf{H}} \boldsymbol{\Phi}^{(\mathrm{p})} + \langle \boldsymbol{P} \rangle\right)^{-1} \quad \text{and} \quad \langle \boldsymbol{H} \rangle = \frac{1}{\sigma_w^2} \boldsymbol{\Sigma}_{\boldsymbol{H}} \boldsymbol{\Phi}^{(\mathrm{p})\mathsf{H}} \left\langle \boldsymbol{Z}^{(\mathrm{p})} \right\rangle, \tag{36}$$

respectively. Here, $\langle \boldsymbol{P} \rangle = \mathrm{diag}(\langle \boldsymbol{\alpha} \rangle)$, and $\left\langle \boldsymbol{Z}^{(\mathrm{p})} \right\rangle$ and $\langle \boldsymbol{\alpha} \rangle$ are the posterior means of $q_{\boldsymbol{Z}}(\boldsymbol{Z}^{(\mathrm{p})})$ and $q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$, respectively. The posterior distribution $q_{\boldsymbol{Z}}(\boldsymbol{Z}^{(\mathrm{p})})$ is truncated complex normal with mean

$$\left\langle \boldsymbol{Z}^{(\mathrm{p})} \right\rangle = \boldsymbol{\Phi}^{(\mathrm{p})} \langle \boldsymbol{H} \rangle + \frac{\sigma_w}{\sqrt{2}} \frac{f\left(\frac{\boldsymbol{Z}^{(\mathrm{lo})} - \boldsymbol{\Phi}^{(\mathrm{p})} \langle \boldsymbol{H} \rangle}{\sigma_w/\sqrt{2}}\right) - f\left(\frac{\boldsymbol{Z}^{(\mathrm{up})} - \boldsymbol{\Phi}^{(\mathrm{p})} \langle \boldsymbol{H} \rangle}{\sigma_w/\sqrt{2}}\right)}{F\left(\frac{\boldsymbol{Z}^{(\mathrm{up})} - \boldsymbol{\Phi}^{(\mathrm{p})} \langle \boldsymbol{H} \rangle}{\sigma_w/\sqrt{2}}\right) - F\left(\frac{\boldsymbol{Z}_n^{(\mathrm{lo})} - \boldsymbol{\Phi}^{(\mathrm{p})} \langle \boldsymbol{H} \rangle}{\sigma_w/\sqrt{2}}\right)}, \tag{37}$$

where $\boldsymbol{Z}^{(\mathrm{lo})}$ and $\boldsymbol{Z}^{(\mathrm{up})}$ are the lower and upper quantization levels corresponding to the observation $\boldsymbol{Y}^{(\mathrm{p})}$, respectively, and $\langle \boldsymbol{H} \rangle$ is the posterior mean of $q_{\boldsymbol{H}}(\boldsymbol{H})$. Also, $f(\cdot)$ and $F(\cdot)$ are the PDF and cumulative distribution function of a standard normal random variable, respectively, computed element-wise on the real and imaginary parts of the argument. The division operation in (37) is also performed element-wise. Finally, the posterior distribution $q_{\boldsymbol{\alpha}[k]}(\boldsymbol{\alpha}[k])$ follows a Gamma distribution with mean, shape and rate parameters given by

$$\langle \boldsymbol{\alpha}[k] \rangle = \frac{a + M}{b + \sum_{i=1}^{m} \langle |\boldsymbol{H}[k,i]|^2 \rangle}, \quad \widetilde{a}_k = a + M, \quad \text{and} \quad \widetilde{b}_k = b + \sum_{i=1}^{M} \langle |\boldsymbol{H}[k,i]|^2 \rangle. \tag{38}$$

Quantized variational Bayesian soft symbol decoding. Next, we develop a QVB algorithm for soft symbol decoding in MIMO-OFDM systems using the model in (25). We consider the unquantized received data signal as a latent variable and express the logarithm of the joint probability distribution of the observations and the latent variables as

$$\ln\, p\left(\boldsymbol{Y}^{(\mathrm{d})}, \boldsymbol{Z}^{(\mathrm{d})}, \boldsymbol{X}^{(\mathrm{d})} | \bar{\boldsymbol{D}}, \sigma_w^2\right) = \ln p(\boldsymbol{Y}^{(\mathrm{d})} | \boldsymbol{Z}^{(\mathrm{d})}) + \ln p\left(\boldsymbol{Z}^{(\mathrm{d})} | \boldsymbol{X}^{(\mathrm{d})}, \bar{\boldsymbol{D}}, \sigma_w^2\right) + \ln p(\boldsymbol{X}^{(\mathrm{d})}). \tag{39}$$

We factorize the posterior distribution of $\boldsymbol{Z}^{(\mathrm{d})}$ and $\boldsymbol{X}^{(\mathrm{d})}$ as

$$p\left(\boldsymbol{Z}^{(\mathrm{d})}, \boldsymbol{X}^{(\mathrm{d})} \mid \boldsymbol{Y}^{(\mathrm{d})}, \bar{\boldsymbol{D}}, \sigma_w^2\right) \approx q_{\boldsymbol{Z}}\left(\boldsymbol{Z}^{(\mathrm{d})}\right) \prod_{t=\tau_{\mathrm{p}}+1}^{\tau_{\mathrm{p}}+\tau_{\mathrm{d}}} \prod_{k=1}^{KN_{\mathrm{c}}} q_{x_{kt}}\left(x_{kt}^{(\mathrm{d})}\right), \tag{40}$$

where $\boldsymbol{Z}^{(\mathrm{d})} = [\boldsymbol{z}_{\tau_{\mathrm{p}}+1}^{(\mathrm{d})}, \ldots, \boldsymbol{z}_{\tau_{\mathrm{p}}+\tau_{\mathrm{d}}}^{(\mathrm{d})}]$, and $x_{kt}^{(\mathrm{d})}$ is the kth component of $\boldsymbol{x}^{(\mathrm{d})}(t)$. We write the conditional probability distributions in (39) as follows:

$$p(\boldsymbol{Y}^{(\mathrm{d})} \mid \boldsymbol{Z}^{(\mathrm{d})}) = \mathbb{I}\left(\boldsymbol{Z}^{(\mathrm{d})} \in (\boldsymbol{Z}^{(\mathrm{lo})}, \boldsymbol{Z}^{(\mathrm{up})})\right), \tag{41}$$

$$p(\boldsymbol{Z}^{(\mathrm{d})} \mid \boldsymbol{X}^{(\mathrm{d})}; \bar{\boldsymbol{D}}, \sigma_w^2) = \prod_{t=\tau_{\mathrm{p}}+1}^{\tau_{\mathrm{p}}+\tau_{\mathrm{d}}} \frac{1}{(\pi\sigma_w^2)^{MN_{\mathrm{c}}}} \exp\left(-\frac{1}{\sigma_w^2} \left\|\boldsymbol{z}^{(\mathrm{d})}(t) - \bar{\boldsymbol{D}} \boldsymbol{x}^{(\mathrm{d})}(t)\right\|_2^2\right), \tag{42}$$

where $\boldsymbol{Z}^{(\mathrm{lo})}, \boldsymbol{Z}^{(\mathrm{up})}$ are the entry-wise lower and upper quantization intervals of the real and imaginary components of $\boldsymbol{Y}^{(\mathrm{d})}$.

The posteriors, in this case, are given as follows. The posterior $q_{x_{kt}}\left(x_{kt}^{(\mathrm{d})}\right)$ follows a Boltzmann distribution with the probability mass function

$$q_{x_{kt}}\left(x_{kt}^{(\mathrm{d})} = \zeta_i\right) = \frac{\exp\left(f_{kt}(\zeta_i)\right)}{\sum_{\zeta' \in \mathbb{D}} \exp\left(f_{kt}(\zeta')\right)}, \tag{43}$$

for $i = 1, \ldots, D$, where $k \in \{1, \ldots, KN_{\mathrm{c}}\}$, $t \in \{\tau_{\mathrm{p}} + 1, \ldots, \tau_{\mathrm{p}} + \tau_{\mathrm{d}}\}$, $\mathbb{D} = \{\zeta_1, \ldots, \zeta_D\}$ is the signal constellation set of cardinality D, and

$$f_{kt}(\zeta) = -\frac{1}{\sigma_w^2}\left(\|\bar{\boldsymbol{D}}_k\|^2 |\zeta|^2 - 2\mathrm{Re}\left[\bar{\boldsymbol{D}}_k^{\mathsf{H}}\left(\left\langle \boldsymbol{z}^{(\mathrm{d})}(t) \right\rangle - \sum_{\substack{k'=1 \\ k' \neq k}}^{KN_{\mathrm{c}}} \bar{\boldsymbol{D}}_{k'} \left\langle x_{k't}^{(\mathrm{d})} \right\rangle\right)\zeta^*\right]\right) + \ln p(x_{kt}^{(\mathrm{d})} = \zeta), \tag{44}$$

where $\mathrm{Re}(\cdot)$ and $(\cdot)^*$ denote the real part and complex conjugate operators, respectively, $\bar{\boldsymbol{D}}_k$ is the kth column of $\bar{\boldsymbol{D}}$, $\left\langle \boldsymbol{z}^{(\mathrm{d})}(t) \right\rangle$ and $\left\langle x_{k't}^{(\mathrm{d})} \right\rangle$ are the posterior means of $q_{\boldsymbol{z}_t}(\boldsymbol{z}^{(\mathrm{d})}(t))$ and $q_{x_{k't}}(x_{k't}^{(\mathrm{d})})$, respectively. We compute the mean and mean square value of $q_{x_{kt}}(x_{d,kt})$ as

$$\left\langle x_{kt}^{(\mathrm{d})} \right\rangle = \sum_{\zeta \in \mathbb{D}} \zeta\, q_{x_{kt}}(\zeta), \quad \text{and} \quad \left\langle |x_{kt}^{(\mathrm{d})}|^2 \right\rangle = \sum_{\zeta \in \mathbb{D}} |\zeta|^2\, q_{x_{kt}}(\zeta). \tag{45}$$

Also, the posterior distribution $q_{\boldsymbol{Z}}\left(\boldsymbol{Z}^{(\mathrm{d})}\right)$ is truncated complex normal, with mean

$$\left\langle \boldsymbol{Z}^{(\mathrm{d})} \right\rangle = \bar{\boldsymbol{D}} \left\langle \boldsymbol{X}^{(\mathrm{d})} \right\rangle + \frac{\sigma_w}{\sqrt{2}} \frac{f\left(\frac{\boldsymbol{Z}^{(\mathrm{lo})} - \bar{\boldsymbol{D}}\langle \boldsymbol{X}^{(\mathrm{d})}\rangle}{\sigma_w/\sqrt{2}}\right) - f\left(\frac{\boldsymbol{Z}^{(\mathrm{up})} - \bar{\boldsymbol{D}}\langle \boldsymbol{X}^{(\mathrm{d})}\rangle}{\sigma_w/\sqrt{2}}\right)}{F\left(\frac{\boldsymbol{Z}^{(\mathrm{up})} - \bar{\boldsymbol{D}}\langle \boldsymbol{X}^{(\mathrm{d})}\rangle}{\sigma_w/\sqrt{2}}\right) - F\left(\frac{\boldsymbol{Z}^{(\mathrm{lo})} - \bar{\boldsymbol{D}}\langle \boldsymbol{X}^{(\mathrm{d})}\rangle}{\sigma_w/\sqrt{2}}\right)}, \quad (46)$$

where $\boldsymbol{Z}^{(\mathrm{lo})}$ and $\boldsymbol{Z}^{(\mathrm{up})}$ are defined in (41), $\left\langle \boldsymbol{X}^{(\mathrm{d})} \right\rangle$ contains the posterior means of $q_{x_{kt}}(x_{kt}^{(\mathrm{d})})\ \forall k, t$ as its entries.

The QVB algorithm starts by randomly initializing the latent variables. It iteratively computes the posterior distributions of data symbols in (43), which in turn yields the bit LLRs. Next, we describe the data-aided channel estimation procedure.

Iterative quantized variational Bayesian channel estimation and soft symbol decoding. Here, we merge the channel estimation and soft symbol decoding into an iterative algorithm. We utilize the data-aided channel estimation system model to refine the channel estimates in an iterative fashion. We first compute an initial channel estimate using the pilot symbols. Then, we utilize the posterior means of the decoded data symbols to form a new measurement matrix $\boldsymbol{\Phi}$ that is input to the channel estimation block. Following a procedure similar to that described for channel estimation, we obtain the posterior statistics of the latent variables as shown below:

$$\boldsymbol{\Sigma}_{\boldsymbol{H}} = \left(\frac{1}{\sigma_w^2}\boldsymbol{\Phi}^{\mathsf{H}}\boldsymbol{\Phi} + \langle \boldsymbol{P} \rangle\right)^{-1}, \langle \boldsymbol{H} \rangle = \frac{1}{\sigma_w^2}\boldsymbol{\Sigma}_{\boldsymbol{H}}\boldsymbol{\Phi}^{\mathsf{H}}\langle \boldsymbol{Z} \rangle, \quad (47)$$

$$\langle \boldsymbol{\alpha}[k] \rangle = \frac{a+M}{b + \sum_{n=1}^{M}\langle |h_{kn}|^2 \rangle}, \qquad k = 1, \ldots, Kn, \quad (48)$$

$$\langle \boldsymbol{Z} \rangle = \boldsymbol{\Phi}\langle \boldsymbol{H} \rangle + \frac{\sigma_w}{\sqrt{2}} \frac{f\left(\frac{\boldsymbol{Z}^{(\mathrm{lo})} - \boldsymbol{\Phi}\langle \boldsymbol{H}\rangle}{\sigma_w/\sqrt{2}}\right) - f\left(\frac{\boldsymbol{Z}^{(\mathrm{up})} - \boldsymbol{\Phi}\langle \boldsymbol{H}\rangle}{\sigma_w/\sqrt{2}}\right)}{F\left(\frac{\boldsymbol{Z}^{(\mathrm{up})} - \boldsymbol{\Phi}\langle \boldsymbol{H}\rangle}{\sigma_w/\sqrt{2}}\right) - F\left(\frac{\boldsymbol{Z}^{(\mathrm{lo})} - \boldsymbol{\Phi}\langle \boldsymbol{H}\rangle}{\sigma_w/\sqrt{2}}\right)}, \quad (49)$$

where $\boldsymbol{Z}^{(\mathrm{lo})}$ and $\boldsymbol{Z}^{(\mathrm{up})}$ are the lower and upper quantization thresholds corresponding to $\boldsymbol{Y}$. We repeat this process of channel estimation and data decoding for a fixed number of iterations. Finally, we use the posterior distribution of the transmit symbols to obtain the bit LLRs, which are deinterleaved and input to the channel decoder. The pseudocode for the iterative QVB channel estimation and soft symbol decoding explained above can be found in [192].

Numerical evaluation

In this subsection, we show the coded bit error rate (BER) performance of the iterative QVB channel estimation and soft symbol decoding algorithm described above. The data bits are generated iid uniformly distributed from the unit-energy 4-quadrature amplitude modulation constellation distribution. At each UE, the data bits are encoded with an LDPC channel code as per 3GPP 5G New Radio specifications [202]. We use the parity check matrix from LDPC base graph 0 with a lifting size Z_c set to 8 and set index 0, which results in 176 message bits and 544 coded bits per block. The coded bits are interleaved by a random interleaver known to both the UE and the BS, mapped to the constellation, OFDM modulated, and transmitted over frequency-selective wireless channels. We define the SNR as $1/\sigma_w^2$. We set the maximum number of iterations for QVB channel estimation and data detection algorithms to 25, and the total number of outer iterations to four.

The left plot in Fig. 4 shows the coded BER of the QVB algorithm with the ADC resolution set to $\{1, 2, 3\}$ bits. The system bandwidth is set to 2 GHz, so the sampling period T_s is 0.5 ns. We set the cyclic prefix length to the maximum delay spread of $n = 32$ symbols. The number of nonzero taps s is set to 8, with the corresponding delays generated uniformly at random between 0 and $(n-2)T_s$. The channel gains of the nonzero taps are iid complex normal with zero mean and unit variance. The figure shows that the performance dramatically improves as the number of quantization bits is increased. While not shown in the plot, the improvement in further increasing the number of quantization bits is marginal.

The right plot in Fig. 4 compares the coded BER performance of the QVB algorithm with that of the MMSE channel estimator and soft-detector [203]. For the quantized MMSE receiver, we compute the DFT after the quantization and perform the equalization. An advantage of the QVB algorithm is that it can recover the channel with only one pilot OFDM symbol. However, for a fair comparison, we set $\tau_{\mathrm{p}} = 8$ because the conventional OFDM receiver cannot estimate the channel in an underdetermined setting. We see that, at a BER of 10^{-4}, the QVB algorithm (labeled "QVB $\tau_{\mathrm{p}} = 8$, EstCSIR") outperforms the conventional OFDM receiver with *unquantized* observations and channels estimated using $\tau_{\mathrm{p}} = 8$ pilot OFDM symbols by around 13 dB. In fact, it *even outperforms* the conventional OFDM receiver with *unquantized observations and perfect CSIR* by 2.5 dB, underscoring the importance of directly inferring the posterior distributions of the data symbols.

To summarize, this section presented a flexible variational Bayesian algorithm for pilot-based channel estimation and soft symbol decoding, achieving superior performance by leveraging channel sparsity in the lag/delay domain, even with minimal pilot symbols and coarsely quantized samples. We next look at RIS-aided channel estimation with quantized measurements.

3.2.2. Example 2: Reconfigurable intelligent surface-aided channel estimation

5G New Radio serves the demand for high data rate use cases by using the mmWave band in the 30–300 GHz range, typically known as FR2 [204]. However, the severe path loss experienced in the mmWave band significantly limits the coverage compared to the lower bands. To compensate for the severe path loss, RIS is recently proposed in [205], an array of metamaterial-based passive reflecting elements capable of adjusting the amplitude and phase of the impinging signal as intended. The RIS is attractive due to the low power consumption resulting from a large number of passive reflecting elements.

To further lower the power consumption of RIS-aided mmWave massive MIMO systems, some RF components can be replaced with energy-efficient components. For example, the BS can be equipped with a large number of low-resolution ADCs that coarsely quantize the received signal [206]. In case the RIS is equipped with a small number of active sensors capable of receiving the impinging signal unlike the passive reflecting elements, the RF chains of the active sensors can be equipped with low-resolution ADCs as well. The motivation for deploying a small number of active sensors at the RIS is to observe and estimate the UE-RIS and RIS-BS links separately [207,208].

For the RIS to generate a favorable propagation condition by aligning the reflection amplitude and phase shift with the channel, accurate CSI is necessary. The distinct feature of RIS channel estimation is that the channel is composed of the UE-RIS link of size NK and RIS-BS link of size MN, which in turn form a UE-RIS-BS link of size MNK where M, N, and K are the numbers of the BS antennas, RIS elements, and single-antenna users, respectively. Since the size of the UE-RIS-BS link is proportional to M and N, the channel estimation overhead is typically large, which necessitates an efficient yet accurate channel estimator. RIS channel estimation becomes more challenging when RIS-aided mmWave massive MIMO systems are equipped with low-resolution ADCs, which coarsely quantize the received signal.

To accurately estimate the channel of RIS-aided mmWave massive MIMO systems in the presence of low-resolution ADCs, the prior knowledge of the channel sparsity in the mmWave band should be exploited. Furthermore, the cascaded channel structure composed of the UE-RIS and RIS-BS links should also be taken into account, which calls for a quantized CS approach designed explicitly for quantized RIS mmWave channel estimation. Since there are various scenarios on where to deploy low-resolution ADCs, the quantized RIS mmWave channel estimation problem yields many challenging yet exciting topics.

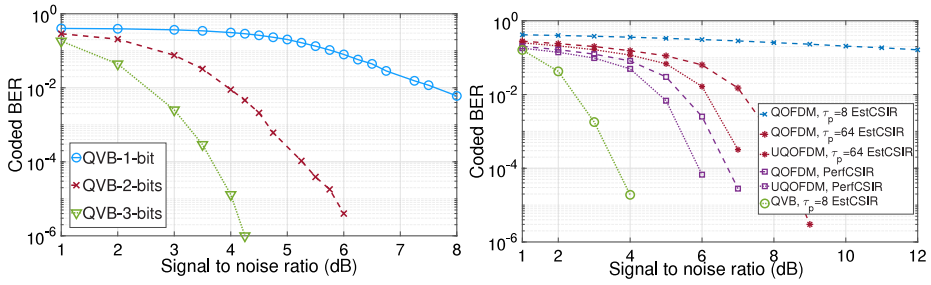


Fig. 4. Coded BER as a function of SNR (dB), with $M = 32$, $K = 8$, $N_c = 256$, $n = 32$, $s = 8$, $\tau_d = 10$. The left plot shows the performance for three values of the number of quantization bits. The right plot shows the performance of conventional OFDM processing (curves labeled "UQOFDM" and "QOFDM") [203], the QVB algorithm where curves labeled "EstCSIR" use the estimated channel while those labeled "PerfCSIR" assume perfect channel estimates.

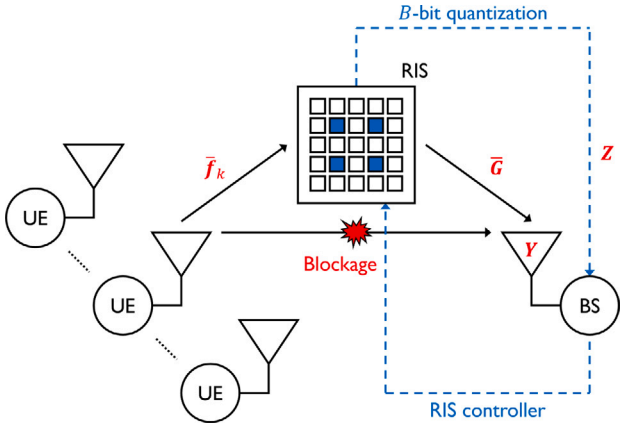


Fig. 5. The uplink of an RIS-aided mmWave massive MIMO system with low-resolution ADC-based active sensors.

This section addresses two related quantized RIS-aided mmWave channel estimation problems: one using fixed low-resolution ADCs and the other utilizing a task-based quantizer. We begin by discussing the quantized RIS mmWave channel estimation problem where the RIS is equipped with a small number of low-resolution ADC-based active sensors [209].

Quantized reconfigurable intelligent surface-aided channel estimation

Consider the uplink of an RIS-aided mmWave massive MIMO system with an M-antenna BS and K single-antenna users as depicted in Fig. 5. The RIS is equipped with N elements, which are partitioned into N_p passive reflecting elements and $N_a = N - N_p \ll N$ active sensors. The passive reflecting elements are dedicated to reflecting the impinging signal as intended to generate a favorable propagation condition [210–212]. Meanwhile, the active sensors are capable of receiving the impinging signal, which is forwarded to the BS for channel estimation. Since deploying high-resolution ADCs at the RIS is not practical due to the high power consumption, the active sensors are implemented by connecting the RIS elements to N_a RF chains with B-bit ADCs that coarsely quantize the received signal [207].

In essence, our goal is to estimate the UE-RIS and RIS-BS links separately based on the received signal at the BS and additional information acquired from the low-resolution ADC-based active sensors at the RIS. Then, the BS can avoid estimating the UE-RIS-BS link of size MNK and only estimate the UE-RIS link of size NK in the subsequent coherence blocks because the RIS-BS link remains constant across multiple coherence blocks in practice [213]. As a result, the channel estimation overhead is reduced by estimating the UE-RIS and RIS-BS links separately in the first coherence block and replacing UE-RIS-BS link estimation with UE-RIS link estimation in the subsequent blocks. Here, we focus on the first coherence block, where all the links are unknown. In addition, we assume that the UE-BS link is blocked.

As a preliminary to formulating the system model, let $\boldsymbol{\Omega} \in \{0,1\}^{N\times T}$ and $\boldsymbol{\Omega}^c = \mathbf{1}_{N\times T} - \boldsymbol{\Omega}$ denote the index matrices of the active sensors and passive reflecting elements that constitute the RIS over the channel estimation phase of length T. For example, the N_a rows of $\boldsymbol{\Omega}$ that correspond to the active sensors are all-one vectors, while the N_p remaining rows that correspond to the passive reflecting elements are all-zero vectors.

Then, the received signal $\boldsymbol{Y} \in \mathbb{C}^{M\times T}$ at the BS over the channel estimation phase of length T is

$$\boldsymbol{Y} = \bar{\boldsymbol{G}}\left(\underbrace{\boldsymbol{\Omega}^c \odot \boldsymbol{V}}_{=\boldsymbol{S}} \odot \bar{\boldsymbol{F}}\boldsymbol{X}\right) + \boldsymbol{W}_B, \tag{50}$$

where $\odot$ is the Hadamard product, $\bar{\boldsymbol{F}} \in \mathbb{C}^{N\times K}$ and $\bar{\boldsymbol{G}} \in \mathbb{C}^{M\times N}$ are UE-RIS and RIS-BS links. Meanwhile, $\boldsymbol{V} \in \mathbb{C}^{N\times T}$ is the passive reflection matrix with the reflection amplitude $|\boldsymbol{V}[i,j]| \le 1$ and phase shift $\angle\boldsymbol{V}[i,j] \in [0.2\pi)$, $\boldsymbol{X} \in \mathbb{C}^{K\times T}$ is the reference signal, and $\boldsymbol{W}_B \in \mathbb{C}^{M\times T}$ is the Gaussian noise with iid $\mathcal{CN}(0,\sigma_B^2)$ elements. Likewise, the quantized received signal $\boldsymbol{Z} \in \mathbb{C}^{N\times T}$ at the active sensors forwarded to the BS is

$$\boldsymbol{Z} = \mathcal{Q}_B(\boldsymbol{U}) = \mathcal{Q}_B(\boldsymbol{\Omega} \odot \bar{\boldsymbol{F}}\boldsymbol{X} + \boldsymbol{W}_R), \tag{51}$$

where $\boldsymbol{U} \in \mathbb{C}^{N\times T}$ is the unquantized received signal at the active sensors, while $\boldsymbol{W}_R \in \mathbb{C}^{N\times T}$ is the Gaussian noise with iid $\mathcal{CN}(0,\sigma_R^2)$ elements. The B-bit quantizer $\mathcal{Q}_B(\cdot)$ is applied to the real and imaginary parts elementwise as

$$\boldsymbol{Z}[i,j] = \mathcal{Q}_B(\boldsymbol{U}[i,j]) \iff \begin{cases} \mathrm{Re}(\boldsymbol{Z}^{(\mathrm{lo})}[i,j]) \le \mathrm{Re}(\boldsymbol{U}[i,j]) < \mathrm{Re}(\boldsymbol{Z}^{(\mathrm{up})}[i,j]) \\ \mathrm{Im}(\boldsymbol{Z}^{(\mathrm{lo})}[i,j]) \le \mathrm{Im}(\boldsymbol{U}[i,j]) < \mathrm{Im}(\boldsymbol{Z}^{(\mathrm{up})}[i,j]), \end{cases} \tag{52}$$

where $\boldsymbol{Z}^{(\mathrm{lo})} \in \mathbb{C}^{N\times T}$ and $\boldsymbol{Z}^{(\mathrm{up})} \in \mathbb{C}^{N\times T}$ are the lower and upper thresholds associated with $\boldsymbol{Z}$. In other words, the real and imaginary parts of $\boldsymbol{Z}$, $\boldsymbol{Z}^{(\mathrm{lo})}$, and $\boldsymbol{Z}^{(\mathrm{up})}$ correspond to one of the 2^B quantization thresholds. Before moving on, note that the N_p rows of $\boldsymbol{Z}$ that correspond to the passive reflecting elements are dummy variables that have no meaningful information due to the N_p all-zero rows of $\boldsymbol{\Omega}$.

To take the channel sparsity in the mmWave band into account, the virtual channel representation is adopted, which transforms the dense channel $\{\bar{\boldsymbol{F}}, \bar{\boldsymbol{G}}\}$ into the sparse channel $\{\boldsymbol{F}, \boldsymbol{G}\}$ by [214]

$$\bar{\boldsymbol{F}} = \boldsymbol{A}_R\boldsymbol{F}, \quad \text{and} \quad \bar{\boldsymbol{G}} = \boldsymbol{A}_B\boldsymbol{G}\boldsymbol{A}_R^H, \tag{53}$$

where $\boldsymbol{A}_B \in \mathbb{C}^{M\times M_g}$ and $\boldsymbol{A}_R \in \mathbb{C}^{N\times N_g}$ are the overcomplete dictionaries that satisfy $M_g \ge M$ and $N_g \ge N$. In practice, the overcomplete dictionaries are selected based on the array geometry of the BS and RIS. Finally, applying the virtual channel representation to (50) and (51) gives

$$\boldsymbol{Y} = \boldsymbol{A}_B\boldsymbol{G}\boldsymbol{A}_R^H(\boldsymbol{S} \odot \boldsymbol{A}_R\boldsymbol{F}\boldsymbol{X}) + \boldsymbol{W}_B, \quad \text{and} \quad \boldsymbol{Z} = \mathcal{Q}_B(\boldsymbol{\Omega} \odot \boldsymbol{A}_R\boldsymbol{F}\boldsymbol{X} + \boldsymbol{W}_R). \tag{54}$$

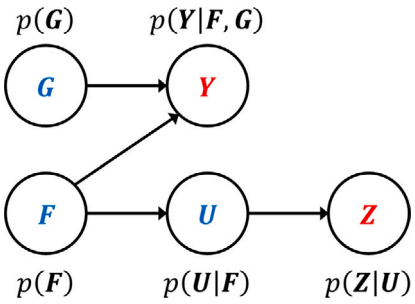


Fig. 6. The Bayesian network of a hierarchical Bayesian model where $\{\boldsymbol{Y}, \boldsymbol{Z}\}$ is the measurement and $\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{U}\}$ is the hidden variable. The arrows represent the conditional dependence between two random variables.

Our goal is to estimate sparse $\{\boldsymbol{F}, \boldsymbol{G}\}$ from $\{\boldsymbol{Y}, \boldsymbol{Z}\}$, which can be interpreted as a combination of low-rank matrix factorization [215] and low-rank matrix completion [216] problems. Here, the high-level idea of the SBL approach is introduced for tackling the problem, which is a Bayesian approach to the quantized CS problem. The interested readers are referred to [209] for a deep dive into SBL-based channel estimation for RIS-aided mmWave massive MIMO systems with low-resolution ADC-based active sensors.

In the SBL approach [201,217–219], the goal is to perform approximate posterior inference on $\{\boldsymbol{F}, \boldsymbol{G}\}$ so that the posterior mean of $\{\boldsymbol{F}, \boldsymbol{G}\}$ can be computed, which is the MMSE estimate. To capture the interaction between $\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{U}, \boldsymbol{Y}, \boldsymbol{Z}\}$, all the variables are treated as random variables that constitute a hierarchical Bayesian model as illustrated in Fig. 6. The conditional distributions of $\{\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{Z}\}$ that define the hierarchical Bayesian model are

$$p(\mathrm{vec}(\boldsymbol{Y})|\boldsymbol{F}, \boldsymbol{G}) = \mathcal{CN}(\mathrm{vec}(\boldsymbol{Y})|\mathrm{vec}(\boldsymbol{A}_{\mathrm{B}}\boldsymbol{G}\boldsymbol{A}_{\mathrm{R}}^{\mathrm{H}}(\boldsymbol{S} \odot \boldsymbol{A}_{\mathrm{R}}\boldsymbol{F}\boldsymbol{X})), \sigma_{\mathrm{B}}^2\boldsymbol{I}), \tag{55}$$

$$p(\mathrm{vec}(\boldsymbol{U})|\boldsymbol{F}) = \mathcal{CN}(\mathrm{vec}(\boldsymbol{U})|\mathrm{vec}(\boldsymbol{\Omega} \odot \boldsymbol{A}_{\mathrm{R}}\boldsymbol{F}\boldsymbol{X}), \sigma_{\mathrm{R}}^2\boldsymbol{I}), \tag{56}$$

$$p(\mathrm{vec}(\boldsymbol{Z})|\boldsymbol{U}) = \mathbb{I}(\mathrm{Re}(\boldsymbol{Z}^{(\mathrm{lo})}) \preceq \mathrm{Re}(\boldsymbol{U}) \prec \mathrm{Re}(\boldsymbol{Z}^{(\mathrm{up})})) \times \mathbb{I}(\mathrm{Im}(\boldsymbol{Z}^{(\mathrm{lo})}) \preceq \mathrm{Im}(\boldsymbol{U}) \prec \mathrm{Im}(\boldsymbol{Z}^{(\mathrm{up})})), \tag{57}$$

which are obtained from the measurement model defined in (52) and (54). Here, $\mathcal{CN}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{C})$ is the PDF of a complex Gaussian random vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{C}$. The prior distributions of $\{\boldsymbol{F}, \boldsymbol{G}\}$ are assumed to be the Student's-t distributions, which are sparsity-promoting prior distributions that make approximate posterior inference tractable [220]. Then, approximate posterior inference is carried out by finding the posterior distributions of $\{\boldsymbol{F}, \boldsymbol{G}\}$ that maximize the negative variational free energy based on the conditional distributions of $\{\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{Z}\}$ defined in (55)–(57) and prior distributions of $\{\boldsymbol{F}, \boldsymbol{G}\}$. After obtaining the posterior distributions, computing the MMSE estimates of $\{\boldsymbol{F}, \boldsymbol{G}\}$ is straightforward.

The SBL approach is *approximate* posterior inference in the sense that (1) the prior distributions of $\{\boldsymbol{F}, \boldsymbol{G}\}$ are assumed to be the Student's-t distributions, (2) posterior distributions of $\{\boldsymbol{F}, \boldsymbol{G}\}$ are assumed to be independent, which is known as the mean-field approximation, and (3) the posterior distributions of $\{\boldsymbol{F}, \boldsymbol{G}\}$ that maximize the negative variational free energy are found in an alternating fashion instead of jointly. These are the key assumptions that make the negative variational free energy maximization problem tractable at the expense of attaining a local maximum of the negative variational free energy.

To demonstrate that the UE-RIS and RIS-BS links can be estimated separately by deploying low-resolution ADC-based active sensors at the RIS, a simulation result is provided in Fig. 7. The simulation setup is such that there are $M = 16$ antennas at the BS, $N = 8 \times 8$ elements at the RIS with N_{a} active sensors quantized by $B = 4$-bit ADCs, and $K = 4$ users in the 28 GHz band, while the length of the channel estimation phase is $T = 400$ symbols. The system parameters are configured as specified in the Dense Urban-eMBB scenario in ITU-R M.2412-0 [221].

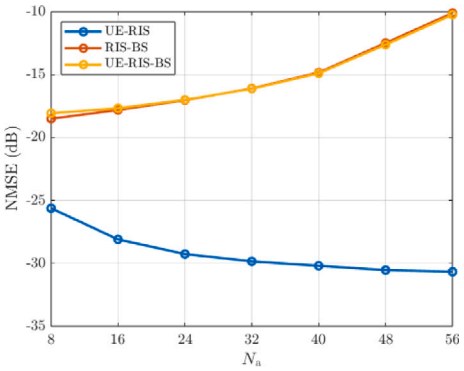


Fig. 7. Normalized mean squared errors (NMSEs) of all the links as a function of the number of active sensors N_{a}.

In Fig. 7, the normalized mean squared error (NMSE) of the UE-RIS link decreases as N_{a} increases, which is not as surprising. In contrast, the NMSE of the RIS-BS link increases, unlike the UE-RIS link. The reason for such a phenomenon is that the RIS-BS link can only be observed through the reflected signal. Therefore, more active sensors mean less passive reflecting elements or equivalently less RIS-BS link measurements. In addition, another interesting point is that the NMSE of the UE-RIS-BS link composed of the UE-RIS and RIS-BS links is lower-bounded by the worst NMSE of the UE-RIS and RIS-BS links. The figure clearly shows that it is possible to obtain sufficiently low NMSE results with not-so-large N_{a}, making the SBL-based channel estimation for RIS-aided mmWave massive MIMO systems highly practical.

Up to this point, we have discussed the quantized channel estimation problem with a fixed quantizer. Next, we delve into task-based quantization, where the quantizer is custom-designed to extract the information relevant to the specific task at hand.

Task-based quantized reconfigurable intelligent surface-aided channel estimation

Task-based quantization [222–226], allows for jointly designing a quantizer with hybrid analog and digital architectures to minimize task recovery errors. The key idea of this approach is to extract lower-dimensional information from the signals in the analog domain based on a specific system task, which can dramatically reduce the overall number of bits required, thereby minimizing memory requirements and power consumption. Now, we discuss how the task-based quantization approach can be applied to channel estimation in RIS-aided mmWave massive MIMO systems, specifically for estimating the cascaded UE-RIS-BS link channel [227]. Note that, although the dimension of the cascaded channel would be large, using the task-based quantization approach, it is feasible to project the received signals onto a lower-dimensional space whose size can be determined by the number of propagation paths. By leveraging the sparse nature of mmWave channels, which have a limited number of propagation paths, this dimensionality can be significantly reduced.

In the case of cascaded channel estimation, deploying active sensors at the RIS is not necessary. Therefore, we consider the system shown in Fig. 5 without active sensors, i.e., $N = N_{\mathrm{p}}$ and $\boldsymbol{S} = \boldsymbol{V}$, while we consider finite-resolution ADCs at the BS in this case. In the considered setup, the total number of time slots T for the channel estimation phase is divided into T_{s} blocks, with each block consisting of $\bar{\tau}$ time slots. Within each block, the RIS configurations remain unchanged, and all UEs transmit repeated orthogonal pilot sequences, e.g., DFT sequences, throughout the T_{s} blocks. Let $\boldsymbol{y}[t, u] = \boldsymbol{y}[(t-1)\bar{\tau} + u]$ be the uplink received signal at the BS in the uth time slot of the tth block. The measurement matrix at the BS, obtained by stacking the received

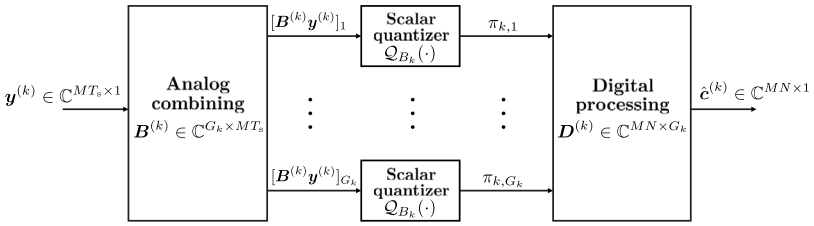


Fig. 8. An illustration of the hardware-limited task-based quantization system to estimate the vectorized cascaded channel $\boldsymbol{c}^{(k)}$.

signals during $\bar{\tau}$ time slots of the tth block, is given by

$$\boldsymbol{Y}[t] = [\boldsymbol{y}[t,1], \ldots, \boldsymbol{y}[t,\bar{\tau}]] = \sum_{k=1}^{K} \bar{\boldsymbol{G}} \mathrm{diag}(\boldsymbol{S}_t) \bar{\boldsymbol{f}}_k \sqrt{P_k} \boldsymbol{x}^{(k)\mathrm{T}} + \boldsymbol{W}_{\mathrm{B}}[t], \tag{58}$$

where $\boldsymbol{x}^{(k)\mathrm{T}} = [x_k[t,1], \ldots, x_k[t,\bar{\tau}]] \in \mathbb{C}^{1\times\bar{\tau}}$ is the pilot sequence from the kth UE satisfying $\|x_k[t,u]\|^2 = 1$, P_k is UE transmit power, and $\boldsymbol{W}_{\mathrm{B}}[t] = [\boldsymbol{n}_{\mathrm{B}}[t,1], \ldots, \boldsymbol{n}_{\mathrm{B}}[t,\bar{\tau}]] \in \mathbb{C}^{M\times\bar{\tau}}$ is the Gaussian noise. Right multiplying $\frac{1}{\bar{\tau}}\boldsymbol{x}^{(k)*}$ by (58) and collecting these signals over T_{s} blocks leads to

$$\boldsymbol{Y}^{(k)} = \frac{1}{\bar{\tau}}[\boldsymbol{Y}[1]\boldsymbol{x}^{(k)*}, \ldots, \boldsymbol{Y}[T_{\mathrm{s}}]\boldsymbol{x}^{(k)*}] = \sqrt{P_k}\bar{\boldsymbol{G}}\mathrm{diag}(\bar{\boldsymbol{f}}_k)\boldsymbol{S} + \boldsymbol{W}^{(k)}, \tag{59}$$

where $\boldsymbol{W}^{(k)} = \frac{1}{\bar{\tau}}[\boldsymbol{W}_{\mathrm{B}}[1]\boldsymbol{x}^{(k)*}, \ldots, \boldsymbol{W}_{\mathrm{B}}[T_{\mathrm{s}}]\boldsymbol{x}^{(k)*}] \in \mathbb{C}^{M\times T_{\mathrm{s}}}$. Using the identity $\mathrm{vec}(\boldsymbol{JKL}) = (\boldsymbol{L}^{\mathrm{T}} \otimes \boldsymbol{J})\mathrm{vec}(\boldsymbol{K})$, we can vectorize $\boldsymbol{Y}^{(k)}$ in (59) as

$$\mathrm{vec}(\boldsymbol{Y}^{(k)}) = \sqrt{P_k}(\boldsymbol{S}^{\mathrm{T}} \otimes \boldsymbol{I})\mathrm{vec}(\bar{\boldsymbol{G}} \diamond \bar{\boldsymbol{f}}_k^{\mathrm{T}}) + \mathrm{vec}(\boldsymbol{W}^{(k)}) = \boldsymbol{S}^{(k)}\boldsymbol{c}^{(k)} + \boldsymbol{n}^{(k)} \triangleq \boldsymbol{y}^{(k)}, \tag{60}$$

where $\diamond$ denotes the Khatri–Rao product, $\boldsymbol{S}^{(k)} = \sqrt{P_k}(\boldsymbol{S}^{\mathrm{T}} \otimes \boldsymbol{I})$ is the training matrix, $\boldsymbol{c}^{(k)} = \mathrm{vec}(\bar{\boldsymbol{G}} \diamond \bar{\boldsymbol{f}}_k^{\mathrm{T}})$ is the vectorized cascaded channel for the kth UE, and $\boldsymbol{n}^{(k)} = \mathrm{vec}(\boldsymbol{W}^{(k)})$.

Here, our goal is to design a channel estimator for $\boldsymbol{c}^{(k)}$ from $\boldsymbol{y}^{(k)}$ using task-based quantization with identical scalar ADCs, referred to as hardware-limited task-based quantization [222]. The considered channel estimator is depicted in Fig. 8, where $\boldsymbol{y}^{(k)}$ is first projected onto a lower-dimensional space $\mathbb{C}^{G_k\times 1}$ satisfying $G_k \le MT_{\mathrm{s}}$ using an analog combining matrix $\boldsymbol{B}^{(k)} \in \mathbb{C}^{G_k\times MT_{\mathrm{s}}}$. Subsequently, the real and imaginary parts of each element in $\boldsymbol{B}^{(k)}\boldsymbol{y}^{(k)}$ are quantized using B_k-bit identical scalar ADCs, denoted as $\mathcal{Q}_{B_k}(\cdot)$, with the support γ_k. Denoting the output of the scalar ADCs as $\boldsymbol{\pi}^{(k)} = [\pi_{k,1}, \ldots, \pi_{k,G_k}]^{\mathrm{T}} \in \mathbb{C}^{G_k\times 1}$, $\boldsymbol{c}^{(k)}$ is reconstructed by a digital processing matrix $\boldsymbol{D}^{(k)} \in \mathbb{C}^{MN\times G_k}$, with its estimate given by $\hat{\boldsymbol{c}}^{(k)} = \boldsymbol{D}^{(k)}\boldsymbol{\pi}^{(k)}$. The objective in this system is to design $\boldsymbol{B}^{(k)}$, $\boldsymbol{D}^{(k)}$, and γ_k to minimize the mean square error distortion between $\boldsymbol{c}^{(k)}$ and its quantized estimate, leading to the following optimization problem:

$$\min_{\boldsymbol{B}^{(k)},\boldsymbol{D}^{(k)},\gamma_k} \mathbb{E}\left[\|\boldsymbol{c}^{(k)} - \hat{\boldsymbol{c}}^{(k)}\|^2\right] \overset{(a)}{=} \mathbb{E}\left[\|\boldsymbol{c}^{(k)} - \tilde{\boldsymbol{c}}^{(k)}\|^2\right] + \min_{\boldsymbol{B}^{(k)},\boldsymbol{D}^{(k)},\gamma_k} \mathbb{E}\left[\|\tilde{\boldsymbol{c}}^{(k)} - \hat{\boldsymbol{c}}^{(k)}\|^2\right], \tag{61}$$

where (a) follows from the orthogonality principle denoting $\tilde{\boldsymbol{c}}^{(k)} = \mathbb{E}[\boldsymbol{c}^{(k)}|\boldsymbol{y}^{(k)}]$ as the MMSE estimate of $\boldsymbol{c}^{(k)}$ from $\boldsymbol{y}^{(k)}$, and $\hat{\boldsymbol{c}}^{(k)}$ is the quantized representation. Since explicitly deriving $\tilde{\boldsymbol{c}}^{(k)}$ is challenging, we approximate $\tilde{\boldsymbol{c}}^{(k)}$ as the linear MMSE estimate, which can be derived in a closed-form expression as in [227]. Based on this approximation, $\boldsymbol{B}^{(k)}$, $\boldsymbol{D}^{(k)}$, and γ_k for the problem in (61) can be derived according to Lemma 1 and Theorem 1 from [222].

The NMSE performance according to the overall number of quantization bits available at the BS B_{T} is depicted in Fig. 9 with $M = 16$, $N = 8\times 8$, $K = 4$, and $T_{\mathrm{s}} = 5$ with $\bar{\tau} = K$ in the 28 GHz band. The digital-only approach, which is task-ignorant, applies the linear MMSE estimator to quantized observations in the digital domain with the fixed ADC resolution $\lfloor 2\frac{B_{\mathrm{T}}}{2MT} \rfloor$. It is observed that the channel estimator using hardware-limited task-based quantization effectively approaches the linear MMSE estimate achievable without quantization

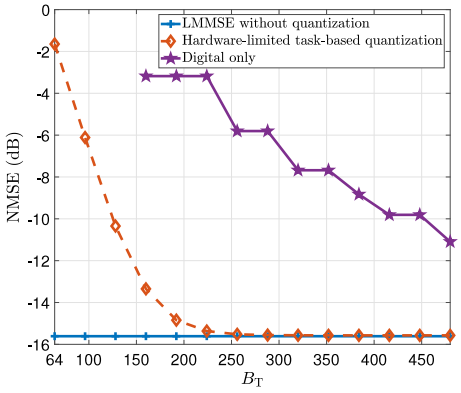


Fig. 9. NMSE comparison of UE-RIS-BS link channels as a function of number of quantization bits B_{T}.

with a small number of bits compared to the digital-only approach, suggesting that incorporating system tasks in the analog domain can significantly enhance the system performance.

4. Other low-resolution signal processing techniques

The previous sections discussed the capabilities and accomplishments of low-resolution CS, addressing scenarios from the extreme case of one-bit resolution to multi-bit quantization. Low-bit quantization is a broad field that extends beyond CS techniques, and not all low-resolution signal processing methods are based on CS principles. So, we turn our attention to two other low-resolution signal processing techniques. First, we explore low-bit quantization algorithms (without necessarily leveraging sparsity) in the context of localization and tracking applications. Next, we examine the emerging paradigm of unlimited sampling, which aims to address the saturation challenges commonly encountered in digital signal acquisition.

4.1. Signal recovery with low-bit quantization

Non-sparse signal recovery with low-bit quantization is a broad and diverse field with applications across a broad range of applications. Several studies have examined this problem from a classical statistical perspective, with the ML-based recovery being a key approach for non-sparse signal recovery from quantized measurements [228–230]. An example area of interest is the recovery of the frequency and phase of temporal and spatial sinusoidal signals using only 1-bit information with fixed quantization thresholds, which is thoroughly explored in [231,232]. Also, the recovery of general signals with high-dimensional parameters based on sign comparison information is examined [233,234]. A more recent study has looked at an efficient signal estimation and threshold design algorithm for recovering signals from 1-bit noisy measurements, handling both time-varying and fixed signals under various noise conditions [235–237]. The deep learning-aided models using techniques like deep unfolding and architectures

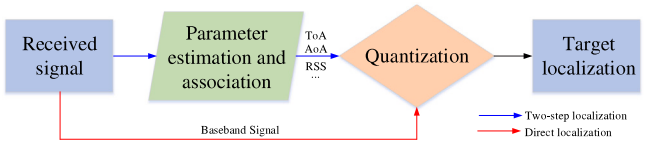


Fig. 10. Target localization architecture using low-bit quantized measurements.

like multi-layer perceptron and long short-term memory are also explored and shown near-optimal performance compared to traditional methods [96,238].

Extensive surveys on low-bit quantization in various areas, including image processing [239–241], audio processing and language models [242–244], model quantization in deep neural networks [245,246], are available in the literature. However, the resulting algorithms differ significantly and do not share obvious commonalities. Given the breadth of this field and the need to adapt techniques to specific structures, we focus on some concrete applications in wireless communication and sensing areas, reviewing some of the state-of-the-art algorithms.

4.1.1. IoT, wireless sensor networks, and radar

Parameter estimation algorithms based on low-resolution data in wireless sensor networks and radar is a well-studied topic. While parameter estimation benefits from more signal samples and sensors [247], they also increase power, storage, and transmission bandwidth requirements [248]. Low-resolution quantization reduces processing costs by using fewer bits per sample, lowering data volume [249–251]. We review target localization and tracking algorithms, though this framework easily extends to parameter estimation and parameter tracking, respectively.

Parameter estimation: Target localization

Target localization determines a target's position using signals from multiple sensors connected to a fusion center via wireless channels, each collecting signal from the target source. Low-bit quantization in target localization is categorized into two-step localization [250–253] and direct localization [254–258]. In two-step localization, sensors first estimate measurement parameters such as time of arrival or received signal strength and transmit these estimates to a fusion center, which then solves geometric equations to determine the target's location. Low-bit quantization reduces bandwidth by compressing these parameters before transmission. In contrast, direct localization bypasses measurement association, estimating the target's location directly from the quantized received signals from the target. By leveraging the constraint that all signals originate from the same target, it offers higher accuracy and robustness under low SNR conditions. The key difference lies in estimator construction: two-step localization uses only measurement parameters, while direct localization processes full signals. Consequently, two-step localization is more computationally efficient at the fusion center. Fig. 10 summarizes low-bit quantized localization methods.

We now present a generalized low-bit localization formulation. Consider joint localization with m sensors that are connected to a common fusion center through wireless communication channels. Each sensor observes n samples of the signal from the target source.[3] We define $\boldsymbol{Z} = \bar{\boldsymbol{Z}} + \boldsymbol{W} \in \mathbb{C}^{m\times n}$ as the observed sample matrix, where $\bar{\boldsymbol{Z}}$ represents noise-free measurements and $\boldsymbol{W}$ is an iid noise matrix with elements in its ith row drawn from $\mathcal{CN}(0,\sigma_i^2)$. The B-bit quantized observed sample matrix $\boldsymbol{Y} \in \mathbb{C}^{m\times n}$ is[4]

$$\mathrm{Re}(\boldsymbol{Y}[i,j]) = Q_B(\mathrm{Re}(\boldsymbol{Z}[i,j])) = \begin{cases} V_1 & \text{if } \xi^{(0)}[i,j] < \mathrm{Re}(\boldsymbol{Z}[i,j]) < \xi^{(1)}[i,j] \\ \vdots & \vdots \\ V_{\bar{L}} & \text{if } \xi^{(\bar{L}-1)}[i,j] \le \mathrm{Re}(\boldsymbol{Z}[i,j]) < \xi^{(\bar{L})}[i,j], \end{cases} \tag{62}$$

where $B = \log_2 \bar{L}$. We can quantize the imaginary part of the signal similarly, replacing $\mathrm{Re}(\cdot)$ with $\mathrm{Im}(\cdot)$. The terms $\xi^{(l)}[i,j]$ for $l = 0,\dots,\bar{L}$ are the quantization thresholds and $\{V_l,\ l = 1,\dots,\bar{L}\}$ denotes the output binary code words with length B.

Assuming the real and imaginary parts of the elements in $\boldsymbol{Y}$ are independent, the probability mass function of the low-bit quantized data $\boldsymbol{Y}[i,j]$ can be expressed as

$$p(\boldsymbol{Y}[i,j]) = p(\mathrm{Re}(\boldsymbol{Y}[i,j]))\, p(\mathrm{Im}(\boldsymbol{Y}[i,j])) \tag{63}$$

$$= \prod_{l=1}^{\bar{L}} p\left(\mathrm{Re}(\boldsymbol{Y}[i,j]) = V_l\right)^{\mathbb{I}(\mathrm{Re}(\boldsymbol{Y}[i,j])=V_l)} p\left(\mathrm{Im}(\boldsymbol{Y}[i,j]) = V_l\right)^{\mathbb{I}(\mathrm{Im}(\boldsymbol{Y}[i,j])=V_l)}, \tag{64}$$

Since $\boldsymbol{Z}$ is typically assumed to follow a Gaussian distribution [254–256], we have

$$p\left(\mathrm{Re}(\boldsymbol{Y}[i,j]) = V_l\right) = F\left(\frac{\sqrt{2}(\xi^{(l-1)}[i,j] - \mathrm{Re}(\bar{\boldsymbol{Z}}[i,j]))}{\sigma_i}\right) - F\left(\frac{\sqrt{2}(\xi^{(l)}[i,j] - \mathrm{Re}(\bar{\boldsymbol{Z}}[i,j]))}{\sigma_i}\right), \tag{65}$$

where the term $F(\cdot)$ is the complementary cumulative distribution function of the standard normal distribution. Similar results can be derived for the probability $p\left(\mathrm{Im}(\boldsymbol{Y}[i,j]) = V_l\right)$ by replacing the term $\mathrm{Re}(\boldsymbol{Y}[i,j])$ with $\mathrm{Im}(\boldsymbol{Y}[i,j])$. Invoking the independence between observed samples, the ML estimate of the unknown target location $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \ln p(\boldsymbol{Y}) = \arg\max_{\boldsymbol{\theta}} \ln \left[\prod_{i=1}^{m}\prod_{j=1}^{n} p(\mathrm{Re}(\boldsymbol{Y}[i,j]))\, p(\mathrm{Im}(\boldsymbol{Y}[i,j]))\right]. \tag{66}$$

The estimator that maps observations to the location differs for two-step or direct localization based on sensor capabilities and observation types, such as time of arrival [251,252] or received signal strength [250,253]. They represent localization as the intersection of localization lines or surfaces in the location state space corresponding to the observation information and the resulting optimization problem is solved via methods, such as semidefinite programming [251], majorization–minimization [257], and exhaustive search [254,255].

The position estimator in (66) depends on quantization thresholds, which significantly impact performance due to information loss. Based on threshold generation, quantizers are classified as uniform (evenly spaced thresholds) or non-uniform (designed based on specific criteria). Non-uniform quantization generally outperforms uniform quantization by minimizing quantization distortion, often modeled as Gaussian noise. A well-known example is the Lloyd-Max quantizer [259], which jointly designs thresholds and outputs. This approach has been explored in two-step localization methods [260]. Another approach optimizes quantization based on performance metrics. For target localization, given the likelihood function in (66), quantization aims to maximize localization accuracy. The Cramér–Rao lower bound (CRLB), a lower bound on estimation accuracy, is closely tied to quantization thresholds [250,254,255]. Thus, an effective strategy is to select

[3] In direct localization, n denotes discrete baseband signal samples, while in two-step localization, it represents measurement parameters (e.g., time of arrival, signal strength) and is typically $n = 1$, requiring less communication bandwidth.

[4] In two-step localization, quantization applies to real-valued scalars. For direct localization, which involves phase information, real and imaginary parts are quantized separately. Thus, the two-step quantizer is a special case of the direct localization quantizer.

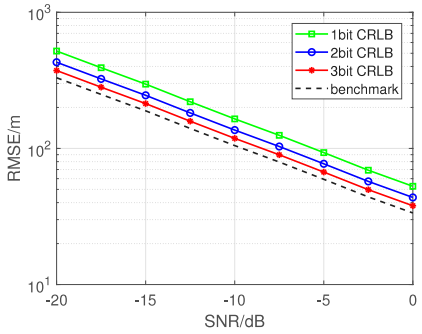


Fig. 11. Comparison of low-bit localization performance under different quantization bit depths, where the unquantized CRLB is used as a benchmark.

thresholds that minimize the CRLB or maximize Fisher information,

$$\hat{\boldsymbol{\xi}} = \arg\min_{\boldsymbol{\xi}} \operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{\theta}}\} \quad \text{s.t.} \quad -\infty < \xi^1[i,j] < \xi^2[i,j] < \cdots < \xi^{\bar{L}-1}[i,j] < +\infty, \forall i, \forall j$$
$$\boldsymbol{\xi} = \left[\xi^{(1)}[1,1], \ldots, \xi^{(1)}[m,n], \ldots \xi^{(\bar{L}-1)}[m,n], \ldots, \xi^{(\bar{L}-1)}[m,n]\right], \tag{67}$$

where $\boldsymbol{C}_{\boldsymbol{\theta}}$ is the CRLB of the localization performance of the system and the estimate $\hat{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ represents the optimized quantization thresholds. Saturated quantization is commonly assumed in the literature [254,255,261], meaning these extreme thresholds are not optimized with $\xi^0[i,j] = -\infty$ and $\xi^{\bar{L}}[i,j] = +\infty$. The problem in (67) is a high-dimensional optimization problem, typically solved using particle swarm optimization [254,255] or heuristic optimization methods [250, 251].

We next show a numerical example to demonstrate the localization performance gap between low-bit and full-precision quantization. We consider a target at $(-1,-0.5)$ km that emits linear frequency-modulated signals, detected by four receivers at $(-1.5,-1.5)$ km, $(-0.75,-1.75)$ km, $(0.75,-1.75)$ km, and $(1.5,-1.5)$ km. The source signal has a 1 MHz bandwidth and a 40 MHz sampling rate. Taking direct localization as an example, Fig. 11 shows a comparison of the CRLB for low-bit target localization algorithms with different quantization bits and full-precision quantization. We see that 3-bit quantization nearly matches full-precision performance, a trend also observed in two-step localization. This highlights the benefits of low-bit quantization in maintaining accuracy while reducing communication overhead. However, a detailed quantitative analysis of performance loss across different quantization levels remains unexplored. The low-bit position estimator in (66) is formulated for the localization problem, but it extends to other settings. For example, incorporating transmission uncertainty in $\boldsymbol{Y}$ enables channel-aware localization for imperfect channels [251,253,254]. Additionally, performance analysis in (67) can go beyond quantization thresholds. By varying quantization bit depths across sensors, (67) transforms into a bit allocation optimization problem [262,263]. Overall, these methods establish a foundational framework for low-bit parameter estimation, adaptable to more complex scenarios.

Parameter monitoring: Target tracking

In multi-sensor data fusion problems in wireless sensor networks, low-bit parameter tracking is a significant area of research. We explore this area, focusing on multi-sensor target tracking scenarios. Mathematically, target tracking can be viewed as a nonlinear Bayesian state filtering problem. Consider n sensors collaboratively tracking one target. At a given discrete time $k = 0, 1, 2, \ldots$, the state of the target $\boldsymbol{x}_k$ and the measurement $\boldsymbol{z}_k$ of each sensor $i = 1, 2, \ldots, n$ is given by the tracking kernel functions [264],

$$\boldsymbol{x}_k = f\left(\boldsymbol{x}_{k-1}, \boldsymbol{n}_k\right) \quad \text{and} \quad \boldsymbol{z}_k^i = h^i\left(\boldsymbol{x}_k, \boldsymbol{w}_k^i\right). \tag{68}$$

Here, the target state $\boldsymbol{x}_k$ is updated as a first-order Markov process with the state transition function $f(\boldsymbol{x}_{k-1}, \boldsymbol{n}_k)$ and process noise $\boldsymbol{n}_k$. The measurement $\boldsymbol{z}_k^i$ at the ith sensor is obtained from the target state $\boldsymbol{x}_k$ through function h^i, contaminated by the measurement noise $\boldsymbol{w}_k^i$.

To estimate $\boldsymbol{x}_k$ over time, target tracking implements a Bayesian estimation framework consisting of predict and update processes given by [265]

$$p\left(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}\right) = \int p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}\right) p\left(\boldsymbol{x}_{k-1}|\boldsymbol{z}_{1:k-1}\right) \mathrm{d}\boldsymbol{x}_{k-1}, \tag{69}$$

$$p\left(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}\right) = \frac{p\left(\boldsymbol{z}_k|\boldsymbol{x}_k\right) p\left(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}\right)}{p\left(\boldsymbol{z}_k|\boldsymbol{z}_{1:k-1}\right)}. \tag{70}$$

The state estimate is characterized by the posterior PDF $p(\boldsymbol{x}_k|\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_k)$. During the update process indicated in (70), the likelihood function $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$ updates the *a prior* PDF given by $p(\boldsymbol{x}_k|\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{k-1})$, yielding the posterior PDF. Assuming the measurements are conditionally independent, the likelihood function becomes the product of local likelihoods

$$p\left(\boldsymbol{z}_k|\boldsymbol{x}_k\right) = \prod_{i=1}^{n} p\left(\boldsymbol{z}_k^i|\boldsymbol{x}_k\right), \tag{71}$$

where the local likelihood $p(\boldsymbol{z}_k^i|\boldsymbol{x}_k)$ corresponds to the PDF of the measurements from the ith sensor indicated in (68). To implement the update process in a sensor network, local sensors send their measurements $\boldsymbol{z}_k^i$ for $i = 1, 2, \ldots, n$ to a fusion node to establish the global likelihood. The primary difference between low-bit tracking and traditional approaches lies in the type of measurement data transmitted to the fusion node and the subsequent filtering and fusion processes. Traditional multi-sensor target tracking assumes analog measurements are transmitted over ideal channels, achieving "what you send is what you get". However, practical limitations, such as limited communication bandwidth and bit errors, necessitate bit quantization in low-bit target tracking.

In low-bit tracking, measurements from each sensor $\boldsymbol{z}_k^i$ are quantized into discrete bit data. Consider a $\log_2 \bar{L}$ bit quantizer, characterized by [261]

$$m_k^i = \begin{cases} 0 & \xi_k^{(0),i} < Q\left(\boldsymbol{z}_k^i\right) < \xi_k^{(1),i} \\ \vdots & \vdots \\ \bar{L}-1 & \xi_k^{(\bar{L}-1),i} < Q\left(\boldsymbol{z}_k^i\right) < \xi_k^{(\bar{L}),i}, \end{cases} \tag{72}$$

where $Q(\boldsymbol{z}_k^i)$ is a quantization function converting the measurement into a scalar, and we define $\boldsymbol{\xi}_k^i = [\xi_k^{(0),i}, \xi_k^{(1),i}, \ldots, \xi_k^{(\bar{L}),i}]^\top$ as the data quantizer. Channel fading and noise may introduce bit errors, resulting in transmitted measurements $\bar{m}_k^i = C(m_k^i)$. These quantized measurements $\bar{m}_k^i$ $(i = 1, 2, \ldots, n)$ are used to reconstruct the quantized measurement-based likelihood functions $p(\bar{m}_k^i|\boldsymbol{x}_k, \boldsymbol{\xi}_k^i)$ to replace $p(\boldsymbol{z}_k^i|\boldsymbol{x}_k)$ for state filtering and data fusion.

Fig. 12 illustrates the low-bit target tracking framework, where the key design components are likelihood function and filter design. The likelihood function is influenced by the statistics of quantized measurements, the quantization process, and CSI, and these nonlinear effects complicate the filter design. Currently, most research focuses on single-time state estimation problems like target detection and localization [266–270]. For tracking, some work has established a likelihood function for one-bit tracking under a simple binary symmetric channel model [271], extending it to generalized asymmetric channels and multi-bit quantization scenarios [272]. For filter design, particle filters are commonly used due to the nonlinear effects of bit quantization. Another approach uses a channel-aware particle filter algorithm [272], offering improved accuracy but increased complexity. Further, some researchers explore fitting the Gaussian error function as a closed-form to accelerate the process [273].

Performance metrics are essential for evaluating and guiding tracking system design [274,275]. The work in [272] provided the posterior CRLB for low-bit tracking, showing that the tracking performance depends on the number of bits and quantizer thresholds. This insight has led to research optimizing quantization bits and thresholds [273,276,

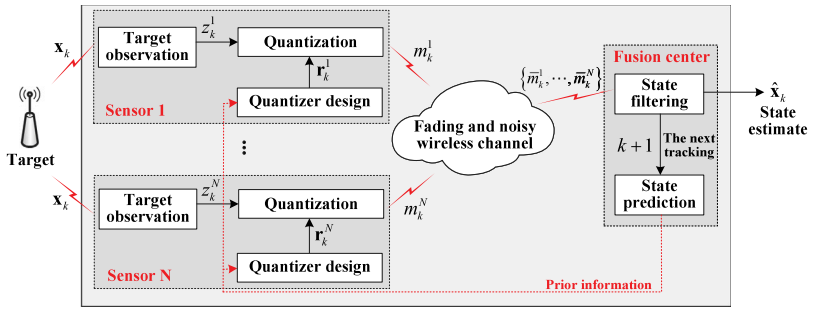
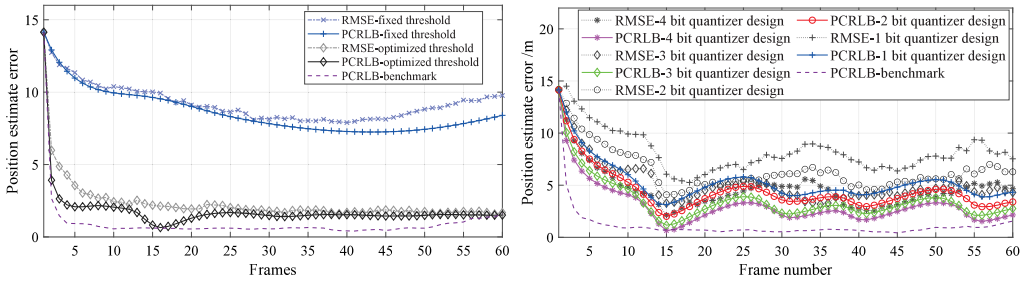


Fig. 12. Network configuration and data processing flow of the low-bit target tracking (From Fig. 1 of [273]).



(a) Comparision of fixed threshold strategy and the dynamic quantizer design (From Fig. 4 of [273]).

(b) Dynamic quantizer design with varying quantization bits (From Fig. 9 of [273]).

Fig. 13. Comparison of position RMSEs and posterior CRLBs.

277]. For example, Fig. 13(a) shows that threshold optimization significantly improves the tracking accuracy for one-bit tracking compared to fixed quantizer thresholds. Fig. 13(b) illustrates that increasing the number of bits allows low-bit tracking to approach ideal non-quantized tracking results.

4.1.2. Cognitive radio networks

Spectrum sensing in cognitive radar networks has been explored through various non-sparse approaches. One such method is an autocorrelation-based wideband sensing technique, where a single-bit quantizer is employed to preserve spectrum occupancy, as shown in [278,279]. Extending this idea, one-bit quantizers are used in [280] to reduce the power consumption of a fast Fourier transform-based wideband sensing approach, where the power spectral density is estimated using DFT. This approach is further extended to multi-antenna cognitive radio receivers and hard-decision cooperative sensing, where local hard decisions are shared with a fusion center [281,282]. The framework is also adapted for uncoordinated environments, supporting both synchronous and asynchronous networks [283,284].

Furthermore, the above methods assume prior knowledge of parameters such as noise power, channel characteristics, or signal properties, which may not always be practical. To address this, one-bit spectrum sensing in the absence of prior information, also known as blind spectrum sensing, has been explored. Designing detectors under this constraint often requires numerical techniques, such as the generalized likelihood ratio test [266]. However, numerical methods increase computational time and complexity, undermining the goal of simple and efficient spectrum sensing [285]. To overcome this, some algorithms rely on the arcsine-based method for reconstructing the covariance matrix, which links the autocorrelation function of an unquantized stationary signal to its quantized counterpart through a nonlinear, invertible arcsine function. This approach led to a closed-form one-bit eigenvalue moment ratio detector proposed in [286]. It is shown that this detector suffers a 3 dB performance loss compared to its infinite-bit counterpart, though the degradation reduces to 2 dB at low SNR. Another study introduced a detector for one-bit observations based on Rao's test, leveraging circularity to enhance performance [287], leading to the second-order eigenvalue moment ratio of the one-bit complex-valued sample covariance matrix. However, these techniques are vulnerable to model inaccuracies and may fail in analytically intractable noise scenarios. More recently, a data-driven approach has been proposed, utilizing Gustafson–Kessel fuzzy c-means clustering for detecting primary user signals in white and correlated noise. This method identifies quantization-invariant features to form decision vectors, which are then clustered to estimate model parameters and classify channel occupancy.

4.1.3. Wireless channel estimation

Wireless channel estimation from low-resolution measurements predates CS and evolves significantly from early pilot-based methods to more advanced techniques such as adaptive quantization and joint estimation. Initially, pilot-based channel estimation focuses on ML-based expectation–maximization approaches, explored using both parametric and non-parametric methods [288,289]. A simpler least squares channel estimation method is also studied in [182,290], providing an alternative to more complex approaches. Additionally, channel estimation using the maximum a posteriori approach is investigated in [291,292]. Other models, such as one-bit channel estimation with an unknown threshold, are considered in [293], highlighting the challenges associated with threshold uncertainty. On the theoretical front, the CRLB for channel estimation in one-bit quantized MIMO systems is examined in [294].

While many of these methods primarily focus on point-to-point communication, it is possible to extend these algorithms to distributed reception or multi-user scenarios. In distributed settings, channel estimation algorithms achieve near-optimal performance with relatively simple receiver operations as the training length increases, even without assuming sparsity [295]. However, these estimators often rely on ML algorithms or high-complexity iterative methods. To overcome these challenges, the ML estimators are extended to near-ML estimators in [296], providing lower complexity while offering theoretical guarantees for performance.

For one-bit massive MIMO systems, obtaining reliable CSI typically requires a long training sequence [182]. To address this issue, a Bayes-optimal joint channel and data estimation scheme is introduced in [297], where the estimated data symbols assist in channel estimation. This scheme shows performance comparable to that with perfect CSI, but the computational complexity of the joint technique remains too high to be practical for commercial systems. To reduce complexity, low-complexity estimators that model quantization noise as independent noise are proposed in [298], significantly lowering the computational burden. Later, more generalized models of quantization noise are incorporated into the Bussgang linear MMSE estimator, which reduces the complexity of ML-based methods but is initially limited to Gaussian-distributed channels in OFDM systems [151]. The generalized Bussgang linear MMSE estimator is further explored in [299], and the equivalence between Bussgang linear MMSE and optimal MMSE channel estimators for noisy and noiseless single input multiple output systems is established in [300]. In [301], these results extend to optimal MMSE estimation for Rayleigh fading MIMO channels with additive white Gaussian noise, with the findings also applicable to multi-user uplink systems.

A key challenge in low-resolution channel estimation is the design of an optimal quantizer. Many of the above studies assume a fixed or typically zero quantization threshold, which, while convenient, may not be optimal for all scenarios [182,297]. To address this, [302] develops an optimal design for quantization thresholds, introducing two schemes: an adaptive quantization scheme that adjusts thresholds dynamically based on the channel characteristics, and a random quantization scheme that generates thresholds using statistical knowledge of the channel. Both of these schemes result in a significant performance improvement compared to fixed quantization, while also reducing training overhead and maintaining high estimation accuracy.

In summary, non-sparse signal recovery with low-bit quantization spans diverse applications in wireless communication and sensing, using approaches from classical statistical methods and ML estimation to deep learning models and optimal quantizer design.

So far, we have focused on quantization error, which arises when real-valued measurements are mapped to a finite set of digital bit strings. In practice, low-resolution ADC also lead to *saturation error* when measurements exceed the quantizer's range, causing clipping. While 1bCS is unaffected by saturation error,[5] all other low-bit quantization methods are vulnerable to it. Scaling down signals to prevent saturation increases quantization noise, degrading signal quality. Although some work addresses CS and low-bit signal recovery under saturation, existing architectures face a trade-off between dynamic range and resolution due to a fixed bit budget [303,304]. Recently, *unlimited sensing* has emerged as a promising alternative, which we discuss next. While one-bit sampling is indifferent to dynamic range, unlimited sampling offers a natural way to generate time-varying thresholds for ADCs to handle dynamic range limitations.

4.2. Unlimited sensing: Pushing the limits of low-resolution acquisition

USF is a radically different approach to digital acquisition, recently introduced in [305–308]. The USF is based on a simple yet powerful observation that,

signals can be recovered from quantization noise.

The starting point of the USF is a novel representation of measurements defined by:

$$y = g - Q_B(g) \qquad \text{(Quantization Noise)}. \tag{73}$$

Note that in conventional literature, digital representation of a discrete-time signal, say $g \in \mathbb{R}$, is obtained by restricting its range; this is done by quantizing it via $Q_B(g)$ (e.g., (51), (23) and (25)). In contrast to this strategy, the USF advocates quantization of y. As we shall see, this approach addresses and overcomes fundamental limitations present in conventional digital sensing methods.

We begin our discussion with some context. Conventional approaches assume that the signal being quantized is ideal or that the ADC functions without imperfections. However, in practice, a *fundamental limitation* of ADCs is their dynamic range [181], denoted by λ, which represents the maximum voltage range the ADC can record or handle. When the input signal exceeds this threshold, the ADC saturates, leading to a *permanent loss of information* as the signal is clipped. This has been a widely reported challenge in literature [309–314]. In such cases, even in the absence of noise, the resulting measurements are distorted, thereby compromising the performance of recovery algorithms that rely on the assumption of ideal quantization.

When designing ADCs, there is an inherent *trade-off* between dynamic range and digital resolution within a given bit-budget[6] [181]. Given the pervasive use of digital acquisition in today's world, this trade-off has significant implications. While it is possible to achieve both high dynamic range (HDR) and high dynamic resolution (HDRes) by designing high-resolution ADCs, this approach is typically undesirable and unsustainable due to the exponential increase in power consumption with the number of bits [181,315].

The HDR-HDRes trade-off in ADCs is frequently put to the test in various scenarios and applications. For example, in the design of communication systems, co-located transmitter and receiver setups in full-duplex systems experience severe self-interference, reaching up to 100 dB [316]. Similarly, balancing receiver saturation and power consumption in the RF chains of massive MIMO systems imposes stringent design constraints. These challenges have driven the development of low-resolution ADC-based solutions [5,317]. However, low-resolution ADCs in massive MIMO systems face significant challenges in signal detection due to quantization noise and the non-linearity introduced by low-resolution quantization, leading to sub-optimal system performance[7].

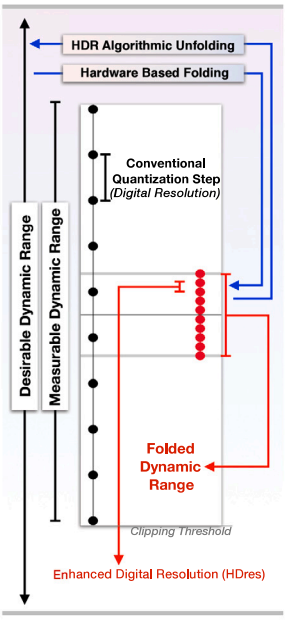
The innovative approach at the core of the USF, depicted in Fig. 14(a), involves transforming HDR signals into a low dynamic range representation by intentionally *folding* them in the analog domain (see Fig. 14(b).[8]), prior to sampling or quantization. This process, similar to winding a thread on a spool or twirling spaghetti on a fork, is accomplished using specialized hardware. Mathematically, this amounts to injecting modulo non-linearity in the hardware. By folding the signals before quantization, the USF effectively prevents clipping and saturation. In scenarios with a fixed bit budget and HDR input, conventional signals may suffer from coarse quantization or even clipping, but the folded samples in the USF maintain HDRes However, these folded signals require a decoding process to be unfolded, introducing a new class of inverse problems. Depending on the folding non-linearity and

[5] In multi-bit quantization, saturation errors occur due to a limited dynamic range. However, one-bit sampling remains unaffected by these limitations since it records only the sign of the signal. Beyond the comparison bit, it does not capture additional information, such as the distance between the signal value and the threshold.

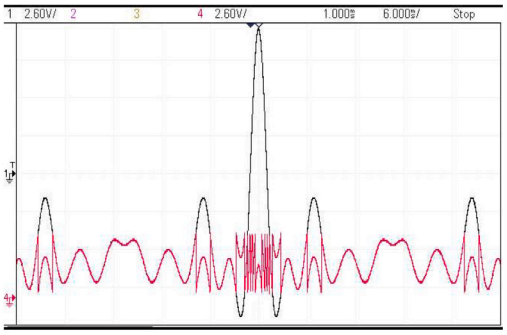
[6] This, in turn, controls the quantization noise floor. The higher the number of bits, the lower the quantization noise.

[7] A common approach to mitigate these issues is to use low-order modulation schemes (e.g., binary phase shift keying) to increase the inter-symbol distance in the constellation diagram, simplifying signal detection. However, when higher symbol rates are required, higher-order modulations must be employed, which complicates signal detection and further degrades system performance.

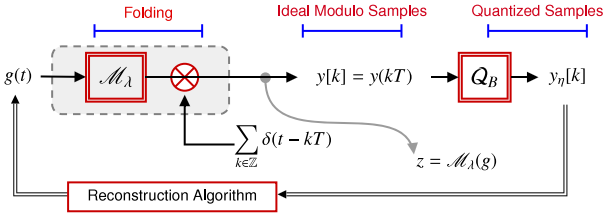
[8] A live YouTube demonstration is available at https://youtu.be/prV40WlzHh4



(a) USF's analog domain folding prior to sampling and quantization preserves HDR features with high digital resolution. Recovery algorithms then decode the encoded HDR signals.



(b) Modulo ADC hardware [305, 306, 308]. An oscilloscope screenshot plotting a continuous-time function (ground truth, black) and its folded version (pink). The input signal with dynamic range 20V ($\approx 10\lambda$) is folded into a 4.025V signal.



(c) Sampling architecture

Fig. 14. USF in a nutshell with an illustration and architecture.

the input signal characteristics (e.g., sparse, smooth, bandlimited, or parametric), specific signal priors are employed to guide the unfolding process in a mathematically principled manner.

Thus by leveraging a co-design of hardware (for folding) and algorithms (for unfolding), the USF is able to recover HDR signals with HDRes. For a given bit budget, the USF's ability to achieve both HDR and HDRes simultaneously presents a promising low-power solution to digital signal representation challenges. Essentially, the folding non-linearity incorporated into the hardware lowers the quantization noise floor, thereby improving the quality of the measurements.

The pipeline underlying the USF is shown in Fig. 14(c). The fundamental principles are similar to the conventional setting of point-wise sampling except for the presence of non-linear modulo mapping in the analog domain. A breakdown of the steps is as follows. We begin with an input signal g to be sampled.

Using innovative folding hardware [308,318], the input function g is folded in the range $[-\lambda, \lambda]$ via centered modulo non-linearity defined by,

$$\mathscr{M}_\lambda : g \mapsto 2\lambda \left(\left[\!\left[\frac{g}{2\lambda} + \frac{1}{2} \right]\!\right] - \frac{1}{2} \right), \tag{74}$$

where we define

$$[\![g]\!] \triangleq g - \lfloor g \rfloor, \quad \text{with } \lfloor g \rfloor = \sup\{\rho \in \mathbb{Z} | \rho \leqslant g\}. \tag{75}$$

This results continuous-time folded signal (see Fig. 14(b)), $z(t) = \mathscr{M}_\lambda(g(t))$. Note that one may equivalently write $z = g - Q_{\lambda'}(g)$, for some λ' related to λ, establishing the fact that modulo representation is essentially the *quantization noise* in conventional digital representation.

The folded function $z(t)$ is sampled using impulse-train, $\oplus_{kT} = \sum_{k\in\mathbb{Z}} \delta(t - kT)$, with sampling rate $T > 0$ yielding uniform samples,

$$y[k] \triangleq z(kT) = \mathscr{M}_\lambda(g(kT)), \quad k \in \mathbb{Z}. \tag{76}$$

as shown in Fig. 14(c). When considering the case of signal quantization with a budget of B bits per sample, each modulo sample $y[k]$ is rounded to the closest element in the set $C_{B,\lambda} = \left\{ \pm \frac{(2n+1)}{2^B} \lambda \,\middle|\, n \in \{0, \dots, 2^{B-1} - 1\} \right\}$. The resulting quantized measurements are defined by,

$$y_\eta[k] \triangleq Q_B(y[k]). \tag{77}$$

4.2.1. Recovery guarantees and methods

Since introducing the USF, various signal classes have been studied. Some notable examples include, (a) bandlimited [305,306,308,319,320] and bandpass [321] signals, (b) sparse signals [322–324], (c) sinusoidal mixtures [325,326], and (d) splines [327].

To maintain backward compatibility with the Shannon–Nyquist sampling theory, we begin our discussion with bandlimited signal classes [328]; the case of square-integrable functions, compactly supported in the Fourier with maximum frequency Ω. We denote such functions by $g \in \mathsf{PW}_\Omega$ (Paley–Wiener Space). For such functions, it is natural to ask: under what sampling density conditions do modulo samples uniquely characterize a bandlimited function? The answer to this question is that any sampling rate faster than critical sampling, i.e., $0 < T < \pi/\Omega$ allows for a one-to-one mapping between a bandlimited function, $g \in \mathsf{PW}_\Omega$ and modulo samples $y[k] = \mathscr{M}_\lambda(g(kT))$. Without loss of generality, normalizing the bandwidth such that $\Omega = \pi$ with the critical sampling rate is $T = 1$ allows us to use oversampled representations. In that case, we denote the sampling rate by $T_{\bar{\epsilon}} = \frac{\pi}{\pi + \bar{\epsilon}}, \bar{\epsilon} > 0$, implying that $0 < T_{\bar{\epsilon}} < 1$. A formal statement of the injectivity conditions is presented in [329], and a subsequent alternative form of proof can be found in [319].

For any $g \in \mathsf{PW}_\Omega$, the *unlimited sampling theorem* proven in [305,306] provided the first sampling criterion for recovery from modulo samples. The formal statement is as follows.

Theorem 2 (*Unlimited Sampling Theorem [305,306]*). *Let $g(t)$ be a continuous-time function with maximum frequency Ω. Then, a sufficient condition for the recovery of $g(t)$ from its modulo samples (up to an additive constant) taken every T seconds apart is $T \leqslant 1/(2\Omega e)$ where e is Euler's number.*

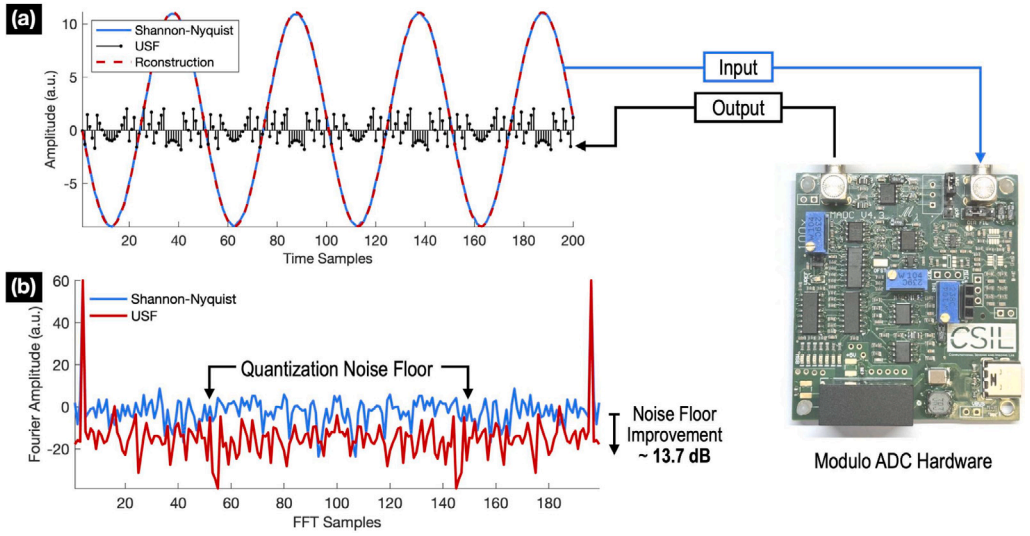


Fig. 15. Hardware experiments demonstrating the practical advantages of USF. (a) An input signal (sine wave) with amplitude swing of 20.1925 V is folded within the range $[-\lambda, \lambda]$, $\lambda \approx 2.01$ using the modulo ADC presented in [308]. HDR reconstruction is then performed using the algorithm (US-Alg) detailed in [306]. (b) Both signals are quantized using the same number of bits. Reconstruction from modulo samples results in an improvement in the quantization noise floor by approximately 13.7 dB, enhancing the digital precision of the system. This improvement translates to better detection and reconstruction accuracy across various applications, including communication systems [330,331], tomography [332,333] and radars [334].

The above theorem establishes a recovery principle similar to the *Shannon–Nyqvist* theorem. Specifically, a constant factor of oversampling is sufficient for recovery, regardless of λ. It is well established that bandlimited signals cannot be recovered if the spectrum is aliased [328]. However, counter-intuitively, even though modulo samples correspond to a non-bandlimited function, i.e., $\mathscr{M}_\lambda(g) \notin \mathsf{PW}_\Omega$, which results in an aliased spectrum, a bounded time-bandwidth product $\Omega T \leq (2e)^{-1}$ ensures the inversion of $\mathscr{M}_\lambda(\cdot)$.

Further, in the case of quantization, the modulo samples $y[k]$ are affected by noise η of amplitude bounded by a some constant $b_0 > 0$. That is,

$$\forall k \in \mathbb{Z}, \quad y_\eta[k] \triangleq Q_B(y[k]) \equiv y[k] + \eta[k], \quad |\eta[k]| \leqslant b_0. \tag{78}$$

In the presence of noise, it is possible that $y_\eta[k] \notin [-\lambda, \lambda]$. However, for b_0 below a certain fixed threshold, the USF recovery method can provably reconstruct the noisy bandlimited samples $\widetilde{\gamma}[k]$ from the corresponding noisy modulo samples $y_\eta[k]$, albeit up to an unknown constant. In this scenario, the noise present in the reconstructed samples matches the noise affecting the modulo samples on an entry-wise basis, or, $\widetilde{\gamma}[k] = \gamma[k] + \eta[k] + 2\rho\lambda$ with $\rho \in \mathbb{Z}$.

Theorem 3 (*Unlimited Sampling Theorem with Quantization*). *Let $g(t) \in PW_\Omega$ and assume that $\beta_g \in 2\lambda\mathbb{Z}$ is known with $\|g\|_\infty \leqslant \beta_g$. For the dynamic range, we work with the normalization $DR = \beta_g/\lambda$. Let the noisy modulo samples be of the form (78) with a noise bound given in terms of the dynamic range as*

$$\|\eta\|_\infty \leqslant \frac{\lambda}{4}(2 \cdot DR)^{-\frac{1}{\nu}}, \qquad \nu \in \mathbb{Z}, \nu > 0. \tag{79}$$

Then a sufficient condition for recovery of $\widetilde{\gamma}[k] = \gamma[k] + \eta[k] + 2\rho\lambda$ with $\rho \in \mathbb{Z}$ is that

$$T \leqslant \frac{1}{2^\nu \Omega e}. \tag{80}$$

The proofs of Theorems 2 and 3 are constructive and led to the reconstruction algorithm, *US-Alg*, introduced in [306]. For further results on quantization and rate–distortion analysis in the USF context, we refer to [320].

Recovery from folded samples can also be performed in the Fourier domain via the *Fourier-Prony* method [308]. Beyond the *time* and *frequency* domain approaches, alternatives such as Wavelet-based reconstruction [335], reconstruction method with sampling slightly above the Nyquist rate [319], optimization-based recovery [336,337], have been considered in the literature.

Furthermore, the hardware experiment illustrated in Fig. 15 clearly demonstrates how USF can simultaneously achieve HDR and HDRes capabilities. The USF strategy results in an improvement of the quantization noise floor ($\approx$ 13.7 dB). This improvement allows for the implementation of low-resolution sensing approaches without compromising HDR signal features and translates to better detection and reconstruction accuracy across various applications. Concrete examples include communication systems [330,331], tomography [332,333] and radars [334].

Given the pivotal role played by digital sensing pipelines, the simple yet powerful USF philosophy of *folding before sampling* can be applied in various ways. Specifically, USF can be combined with different signal models (e.g., sparse, smooth, parametric, and time-invariant) or integrated with various architectures (e.g., compressive, one-bit, time-encoded, and multi-channel sensing). In this context, we briefly discuss two variations that align with the theme of this paper, namely, (i) low-resolution sensing and (ii) sparse priors.

4.2.2. Low-resolution sensing

As illustrated by the hardware example in Fig. 15 and the experiments considered in [330–334], the USF naturally serves as a low-resolution sensing technique; for a fixed bit-budget, it offers clear advantages over traditional samplers. When considering the extreme case of 1-bit quantization, USF can be combined with existing approaches to prevent distortion in such methods. We consider three different architectures for 1-bit sensing.

- *Sigma-Delta of $\Sigma\Delta$ Quantization:* Conventional $\Sigma\Delta$ converters take advantage of the fact that oversampling with fewer bits is more cost-effective to implement in hardware. The $\Sigma\Delta$ scheme operates on the principle of "noise shaping". By oversampling the input signal [338,339], noise arising from the coarse quantization can be pushed to higher parts of the Fourier spectrum, thus encoding the input bandlimited function with minimal distortion. As in the conventional setting, if the input signal's dynamic range significantly exceeds a preset threshold, the $\Sigma\Delta$ quantizer saturates, leading to a breakdown of the pipeline. However, by combining USF with $\Sigma\Delta$, this limitation is overcome because the folded

signal remains within the sensing threshold. For further details, we refer to [340].

- *Time-Encoding or Event-Driven Sampling:* Time encoding [341] provides an alternative to conventional uniform sampling by transforming continuous-time signals into streams of trigger times, which form the basis of Event-Driven Sampling (EDS) models. Since one only records the trigger time stamps, the resulting signal can be considered as a one-bit stream of data. EDS offers significant advantages in reducing power consumption and enhancing time resolution, drawing inspiration from how biological nervous systems encode information. However, if an analog signal surpasses a predefined dynamic range, EDS may not produce the required distribution of trigger times, leading to recovery distortion due to aliasing. This is the equivalent of "saturation" in conventional ADCs. As shown in [342], by injecting modulo non-linearity prior to EDS, this pitfall can be avoided. Hardware validation of USF inspired EDS indeed demonstrates the advantages of this approach.
- *Sign-Based One-Bit:* One of the simplest one-bit architectures records only the sign of the measurement, but this approach can result in significant information loss. In their recent work, *unlimited one-bit sampling* [343], the authors introduce a one-bit architecture where the input signal is first modulo folded and then converted into a one-bit stream. Instead of using a fixed threshold[9] for comparison, the authors employ time-varying sampling thresholds, leading to improved signal estimation.

4.2.3. Sparse priors

Over the past two decades, the concept of sparsity and the application of sparse priors have proven to be highly effective data models. Sparse priors play a vital role in efficiently capturing key features and signatures within a signal. This naturally makes the interaction between USF and sparse signal models very interesting, and it has been explored in various flavors.

- *Compressive Sensing:* CS with modulo measurements is first applied in HDR imaging systems, limited to noiseless modulo folding with two periods [344,345]. Later, this model has been expanded to general noisy modulo folding, introducing approximate message-passing algorithms for Bernoulli-Gaussian distributed sparse signals with known parameters [346,347]. The modulo-CS model is applied to line spectral estimation using a two-stage recovery algorithm combining dynamic programming and orthogonal matching pursuit [348]. Convex relaxation-based methods are also studied, developing a mixed integer linear program for sparse signal recovery from modulo measurements [349]. Here, the authors derive conditions on the minimum number of measurements necessary for the unique recovery of sparse vectors. However, mixed integer programming is computationally intensive to solve and becomes even more complicated with mixed integer quadratic programming in the presence of noise. Therefore, [350] presents a variant of iterative hard thresholding, which has linear complexity and is robust to noise.
- *Sparse Super-Resolution:* Recovering spikes or Dirac-impulses from filtered measurements is a classic problem in engineering [351, 352]. In this context, the interaction of spikes with pulses leads to pulsating behavior, often resulting in signals with HDR features. To address this, one can either allocate bits to capture the full dynamic range or focus on recording measurements with high digital resolution, as the accuracy of spike recovery depends on data quality [351,352]. However, USF can effectively overcome this trade-off. Both time-domain [322] and frequency-domain [323] methods for sparse super-resolution have been explored in this context. A hardware example based on time-of-flight imaging, as detailed in [323], offers a compelling demonstration of the practical benefits of USF. Furthermore, by leveraging the sparsity of input signals, it is possible to design recovery strategies that are independent of any sampling rate criteria [324].
- *Fourier-Domain Sparsity:* Estimating the parameters of a sparse mixture of sinusoids is a well-known problem with numerous applications. This model corresponds to a sparse signal in the Fourier domain. The earliest solution to this problem dates back to Prony's work [353], which laid the foundation for the field of *spectral estimation* [354]. Utilizing USF for the estimation of sinusoidal parameters offers significant benefits, as it can detect both *strong* and *weak* components by enhancing digital resolution while preventing clipping [325]. For instance, in the case of radars, which follow a sinusoidal model, a 10 dB improvement in sensitivity has been demonstrated in real-world experiments [334]. Additionally, when combined with multi-channel measurements, USF enables the implementation of sub-Nyquist sampling schemes [326] without compromising dynamic range.

5. Concluding remarks and future directions

In this paper, we explored low-resolution signal processing theory, algorithms, and applications in wireless communication and sensing. The first part was dedicated to 1bCS, where we presented information-theoretic measurement bounds and various algorithmic tools, along with extensions and connections to related areas such as mixture models and recovery methods. We then discussed customized quantized CS algorithms for wireless communication systems. We first introduced a simple and flexible QVB algorithm for channel estimation and soft symbol decoding. This algorithm achieved accurate estimates using a minimal number of pilot symbols, even from coarsely quantized samples. Additionally, we examined SBL algorithms for channel estimation aided by RIS with fixed low-resolution ADCs as well as cascaded channel estimation utilizing task-based quantizers. Our discussion also covered low-resolution signal processing techniques beyond CS. We analyzed an ML approach for target localization, showing that non-uniform quantization improves performance. Using the CRLB, we found that low-bit quantization reduces communication overhead while preserving localization accuracy. We also explored the new concept of USF, which enabled ADCs to reset instead of saturating, producing modulo samples and facilitating sub-Nyquist sampling without compromising dynamic range. Overall, this paper consolidated various results spread out in the literature and emphasized algorithmic tools and sampling strategies for low-resolution signal processing and applications.

This discussion opens up several exciting research possibilities in low-resolution signal processing. We list a few possible directions here.

1. **Cross-pollination of ideas**: Promising directions include combining algorithmic tools like convex relaxation, QVB, and sparse Bayesian learning with sampling techniques such as task-based, CRLB-optimized, and modulo quantization, as well as processing strategies like centralized, decentralized, and distributed approaches. These could be applied to areas like wireless channel estimation, distributed sensing, target localization, and tracking.
2. **Limitations of traditional CS methods**: Most quantized CS approaches rely on conventional CS frameworks, which limits their ability to handle more complex signal structures beyond sparsity. Furthermore, many algorithms primarily exploit sparsity, while other forms of structured sparsity, such as block, hierarchical, and piecewise sparsity, remain underexplored [355–357]. Developing new algorithms that address these challenges presents a significant opportunity for future research.
3. **Theoretical gaps**: Recent studies on 1bCS utilizing generative priors and deep unfolding are generally agnostic to sparse structures, allowing them to accommodate various forms of structure.

[9] For instance, for the usual sign-based one-bit, the threshold is zero.

However, they lack robust theoretical guarantees. Similarly, empirically effective and computationally simpler algorithms, such as those based on QVB and SBL, also need to establish theoretical guarantees, which remains an open challenge. Moreover, the asymptotic and complexity analyses of these algorithms have not been effectively translated into practical system design guidelines.

4. **Channel-aware algorithms and federated learning**: Research on quantization, driven by bandwidth constraints in systems like IoT and wireless sensor networks, demands the development of algorithms that can account for channel impairments that often degrade performance. The development of channel-aware algorithms and decentralized schemes for distributed sensing networks is a promising avenue. For example, task-based quantization could be integrated with distributed CS in wireless sensor networks. Additionally, further exploration is required to create communication-efficient schemes for federated learning with 1bCS in decentralized sensing networks.
5. **Extension to B-bit quantized measurements and mixed ADCs**: While 1bCS is the most extensively studied problem within the quantized CS framework, most 1bCS algorithms are not easily extendable to higher-bit quantization. Thus, a significant gap exists between 1bCS and B-bit quantized CS. Developing theory and algorithms for a quantization-agnostic setting remains an open challenge. Another important area is CS with a mixed-ADC architecture, combining high- and low-resolution ADCs to improve system performance. However, a key challenge lies in determining the optimal ratio of high- to low-resolution ADCs, considering that low-resolution ADCs perform better in low SNR regions but struggle in high SNR scenarios. This architecture must balance energy efficiency, signal detection performance, and SNR conditions.
6. **Modulo ADC integration**: The introduction of USF and modulo ADCs shows great promise in handling the limited dynamic range of ADCs. However, the integration of modulo ADCs with CS remains underdeveloped, with few algorithms effectively addressing noise handling and limited theoretical progress. The impact of finite-bit quantization in modulo CS also presents a new research direction.

In summary, CS-aided low-resolution signal processing is a rapidly advancing field from theoretical, algorithmic, and practical standpoints, presenting numerous challenges and opening intriguing avenues for future research.

CRediT authorship contribution statement

Geethu Joseph: Writing – review & editing, Writing – original draft, Conceptualization. **Venkata Gandikota:** Writing – original draft. **Ayush Bhandari:** Writing – original draft. **Junil Choi:** Supervision. **Insoo Kim:** Writing – original draft. **Gyoseung Lee:** Writing – original draft. **Michail Matthaiou:** Supervision. **Chandra R. Murthy:** Writing – original draft. **Hien Quoc Ngo:** Supervision. **Pramod K. Varshney:** Writing – original draft. **Thakshila Wimalajeewa:** Writing – review & editing. **Wei Yi:** Supervision. **Ye Yuan:** Supervision. **Guoxin Zhang:** Writing – original draft.

Acknowledgment

The authors gratefully acknowledge the discussions with and assistance from Sai Subramanyam Thoota in developing the contents of Section 3.2.1.

Data availability

No data was used for the research described in the article.

References

[1] B. Le, T.W. Rondeau, J.H. Reed, C.W. Bostian, Analog-to-digital converters, IEEE Signal Process. Mag. 22 (6) (2005) 69–77.
[2] N. Shlezinger, Y.C. Eldar, M.R. Rodrigues, Hardware-limited task-based quantization, IEEE Trans. Signal Process. 67 (20) (2019) 5223–5238.
[3] A. Kipnis, Y.C. Eldar, A.J. Goldsmith, Fundamental distortion limits of analog-to-digital compression, IEEE Trans. Inform. Theory 64 (9) (2018) 6013–6033.
[4] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, K. Niu, Pattern division multiple access - A novel nonorthogonal multiple access for fifth-generation radio networks, IEEE Trans. Veh. Technol. 66 (4) (2016) 3185–3196.
[5] Z. Jiayi, L. Dai, X. Li, Y. Liu, L. Hanzo, On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems, IEEE Commun. Mag. 56 (7) (2018) 205–211.
[6] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306.
[7] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.
[8] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inform. Theory 52 (12) (2006) 5406–5425.
[9] D.L. Donoho, For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution, Comm. Pure Appl. Math. 59 (7) (2006) 907–934.
[10] G. Joseph, P.K. Varshney, Measurement bounds for compressed sensing in sensor networks with missing data, IEEE Trans. Signal Process. 69 (2021) 905–916.
[11] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser, 2013.
[12] M.A. Davenport, M.F. Duarte, Y.C. Eldar, G. Kutyniok, Introduction to Compressed Sensing, Cambridge University Press, 2012.
[13] P.T. Boufounos, L. Jacques, F. Krahmer, R. Saab, Quantization and compressive sensing, in: Compressed Sensing and Its Applications, Birkhäuser, 2015, pp. 193–237.
[14] J. Xiong, Q. Tang, 1-Bit compressive data gathering for wireless sensor networks, J. Sensors 2014 (1) (2014) 177—-183.
[15] Y. Shen, J. Fang, H. Li, One-bit compressive sensing and source localization in wireless sensor networks, in: Proceedings of IEEE China Summit and International Conference on Signal and Information Processing, 2013, pp. 379–383.
[16] C.-H. Chen, J.-Y. Wu, Amplitude-aided 1-bit compressive sensing over noisy wireless sensor networks, IEEE Wirel. Commun. Lett. 4 (5) (2015) 473–476.
[17] H. Zayyani, M. Korki, F. Marvasti, A distributed 1-bit compressed sensing algorithm robust to impulsive noise, IEEE Commun. Lett. 20 (6) (2016) 1132–1135.
[18] N. Fu, L. Yang, J. Zhang, Sub-Nyquist 1 bit sampling system for sparse multiband signals, in: Proceedings of European Signal Processing Conference, 2014, pp. 736–740.
[19] F. Salahdine, N. Kaabouch, H. El Ghazi, One-bit compressive sensing vs. multi-bit compressive sensing for cognitive radio networks, in: Proceedings of IEEE International Conference on Industrial Technology, 2018, pp. 1610–1615.
[20] S.K. Sharma, E. Lagunas, S. Chatzinotas, B. Ottersten, Application of compressive sensing in cognitive radio communications: A survey, IEEE Commun. Surv. & Tutorials 18 (3) (2016) 1838–1860.
[21] F. Salahdine, N. Kaabouch, H. El Ghazi, A survey on compressive sensing techniques for cognitive radio networks, Phys. Commun. 20 (2016) 61–73.
[22] Z. Tian, G.B. Giannakis, Compressed sensing for wideband cognitive radios, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 4, 2007, pp. IV–1357–IV–1360.
[23] D. Lee, T. Sasaki, T. Yamada, K. Akabane, Y. Yamaguchi, K. Uehara, Spectrum sensing for networked system using 1-bit compressed sensing with partial random circulant measurement matrices, in: Proceedings of IEEE Vehicular Technology Conference (Spring), 2012, pp. 1–5.
[24] W. Lu, B. Deng, W. Zhang, J. Wang, L. Zhong, S. Peng, Channel feedback based on complex 1-bit Bayesian compressed sensing in FDD massive MIMO systems, in: Proceedings of International Conference on Artificial Intelligence, Information Processing and Cloud Computing, 2019, pp. 1–6.
[25] J. Mo, P. Schniter, N.G. Prelcic, R.W. Heath, Channel estimation in millimeter wave MIMO systems with one-bit quantization, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2014, pp. 957–961.
[26] C. Qing, Q. Yang, B. Cai, B. Pan, J. Wang, Superimposed coding-based CSI feedback using 1-bit compressed sensing, IEEE Commun. Lett. 24 (1) (2019) 193–197.
[27] Y. Seo, N.Y. Yu, Optimum modulation orders for 1-bit compressively sampled signals in multicarrier transmission, in: Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications, 2019, pp. 1–5.

[28] X. Dong, Y. Zhang, A MAP approach for 1-bit compressive sensing in synthetic aperture radar imaging, IEEE Geosci. Remote. Sens. Lett. 12 (6) (2015) 1237–1241.

[29] M. Demir, E. Erçelebi, One-bit compressive sensing with time-varying thresholds in synthetic aperture radar imaging, IET Radar, Sonar & Navig. 12 (12) (2018) 1517–1526.

[30] X. Wang, G. Li, Y. Liu, M.G. Amin, Enhanced 1-bit radar imaging by exploiting two-level block sparsity, IEEE Trans. Geosci. Remote Sens. 57 (2) (2018) 1131–1141.

[31] S.J. Zahabi, M.M. Naghsh, M. Modarres-Hashemi, J. Li, Compressive pulse-Doppler radar sensing via 1-bit sampling with time-varying threshold, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2017, pp. 3419–3423.

[32] S.J. Zahabi, M.M. Naghsh, M. Modarres-Hashemi, J. Li, One-bit compressive radar sensing in the presence of clutter, IEEE Trans. Aerosp. Electron. Syst. 56 (1) (2019) 167–185.

[33] T. Feuillen, One Bit at a Time: the Use of Quantized Compressive Sensing in Radar Signal Processing (Ph.D. thesis), Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2021.

[34] G. Joseph, S. Kafle, P.K. Varshney, One-bit compressed sensing using generative models, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2020, pp. 3437–3441.

[35] J. Fang, Y. Shen, L. Yang, H. Li, Adaptive one-bit quantization for compressed sensing, Signal Process. 125 (2016) 145–155.

[36] E.G. Allstot, A.Y. Chen, A.M. Dixon, D. Gangopadhyay, H. Mitsuda, D.J. Allstot, Compressed sensing of ECG bio-signals using one-bit measurement matrices, in: Proceedings of IEEE International New Circuits and Systems Conference, 2011, pp. 213–216.

[37] J. Haboba, M. Mangia, R. Rovatti, G. Setti, An architecture for 1-bit localized compressive sensing with applications to EEG, in: Proceedings of IEEE Biomedical Circuits and Systems Conference, 2011, pp. 137–140.

[38] J.N. Laska, R.G. Baraniuk, Regime change: Bit-depth versus measurement-rate in compressive sensing, IEEE Trans. Signal Process. 60 (7) (2012) 3496–3505.

[39] Y. Plan, R. Vershynin, One-bit compressed sensing by linear programming, Comm. Pure Appl. Math. 66 (8) (2013) 1275–1297.

[40] Y. Plan, R. Vershynin, Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach, IEEE Trans. Inform. Theory 59 (1) (2012) 482–494.

[41] Q. Fan, C. Jia, J. Liu, Y. Luo, Robust recovery in 1-bit compressive sensing via ℓ_q-constrained least squares, Signal Process. 179 (2021) 107822.

[42] L. Jacques, J.N. Laska, P.T. Boufounos, R.G. Baraniuk, Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors, IEEE Trans. Inform. Theory 59 (4) (2013) 2082–2102.

[43] S. Bahmani, P.T. Boufounos, B. Raj, Robust 1-bit compressive sensing via gradient support pursuit, 2013, arXiv preprint arXiv:1304.6627.

[44] P.T. Boufounos, Greedy sparse signal reconstruction from sign measurements, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2009, pp. 1305–1309.

[45] M. Yan, Y. Yang, S. Osher, Robust 1-bit compressive sensing using adaptive outlier pursuit, IEEE Trans. Signal Process. 60 (7) (2012) 3868–3875.

[46] Z. Yang, L. Xie, C. Zhang, Variational Bayesian algorithm for quantized compressed sensing, IEEE Trans. Signal Process. 61 (11) (2013) 2815–2824.

[47] F. Li, J. Fang, H. Li, L. Huang, Robust one-bit Bayesian compressed sensing with sign-flip errors, IEEE Signal Process. Lett. 22 (7) (2014) 857–861.

[48] C. Zhou, Z. Zhang, F. Liu, Robust 1-bit compressive sensing via variational Bayesian algorithm, Digit. Signal Process. 50 (2016) 84–92.

[49] O. Musa, G. Hannak, N. Goertz, Generalized approximate message passing for one-bit compressed sensing with AWGN, in: 2016 IEEE Global Conference on Signal and Information Processing, GlobalSIP, 2016, pp. 1428–1432.

[50] Y. Ding, S.-E. Chiu, B.D. Rao, Sparse recovery with quantized multiple measurement vectors, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2017, pp. 845–849.

[51] J. Zhu, L. Han, X. Meng, Z. Xu, Binary sparse Bayesian learning algorithm for one-bit compressed sensing, 2018, arXiv preprint arXiv:1805.03043.

[52] Z. Liu, S. Gomes, A. Tiwari, J. Scarlett, Sample complexity bounds for 1-bit compressive sensing and binary stable embeddings with generative priors, in: Proceedings of International Conference on Machine Learning, 2020, pp. 6216–6225.

[53] Z. Liu, J. Scarlett, The generalized lasso with nonlinear observations and generative priors, in: Advances in Neural Information Processing Systems, 2020, pp. 19125–19136.

[54] Z. Liu, S. Ghosh, J. Scarlett, Robust 1-bit compressive sensing with partial Gaussian circulant matrices and generative priors, in: Proceedings of IEEE Information Theory Workshop, 2021, pp. 1–6.

[55] J. Liu, Z. Liu, Non-iterative recovery from nonlinear observations using generative models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 233–243.

[56] J. Chen, J. Scarlett, M. Ng, Z. Liu, A unified framework for uniform signal recovery in nonlinear generative compressed sensing, in: Advances in Neural Information Processing Systems, 2024, pp. 8224–8252.

[57] Z. Li, W. Xu, X. Zhang, J. Lin, A survey on one-bit compressed sensing: Theory and applications, Front. Comput. Sci. 12 (2018) 217–230.

[58] P.T. Boufounos, R.G. Baraniuk, 1-bit compressive sensing, in: Proceedings of Annual Conference on Information Sciences and Systems, 2008, pp. 16–21.

[59] L. Zhang, J. Yi, R. Jin, Efficient algorithms for robust one-bit compressive sensing, in: Proceedings of International Conference on Machine Learning, 2014, pp. 820–828.

[60] P. Xiao, B. Liao, Robust one-bit compressive sensing with weighted ℓ_1-norm minimization, Signal Process. 164 (2019) 380–385.

[61] R. Zhu, Q. Gu, Towards a lower sample complexity for robust one-bit compressed sensing, in: Proceedings of International Conference on Machine Learning, 2015, pp. 739–747.

[62] X. Huang, M. Yan, Nonconvex penalties with analytical solutions for one-bit compressive sensing, Signal Process. 144 (2018) 341–351.

[63] P. Xiao, B. Liao, J. Li, One-bit compressive sensing via Schur-concave function minimization, IEEE Trans. Signal Process. 67 (16) (2019) 4139–4151.

[64] S. Chen, A. Banerjee, One-bit compressed sensing with the k-support norm, in: Artificial Intelligence and Statistics, 2015, pp. 138–146.

[65] X. Huang, L. Shi, M. Yan, J.A. Suykens, Pinball loss minimization for one-bit compressive sensing: Convex models and algorithms, Neurocomputing 314 (2018) 275–283.

[66] J. Hou, X. Liu, 1-bit compressed sensing via an l1-TV regularization method, IEEE Access 10 (2022) 116473–116484.

[67] M.P. Friedlander, H. Jeong, Y. Plan, Ö. Yılmaz, NBIHT: An efficient algorithm for 1-bit compressed sensing with optimal error decay rate, IEEE Trans. Inform. Theory 68 (2) (2021) 1157–1177.

[68] Y. Xu, Y. Kabashima, L. Zdeborová, Bayesian signal reconstruction for 1-bit compressed sensing, J. Stat. Mech. Theory Exp. 2014 (11) (2014) P11015.

[69] S. Qiu, X. Wei, Z. Yang, Robust one-bit recovery via ReLU generative networks: Near-optimal statistical rate and global landscape analysis, in: Proceedings of International Conference on Machine Learning, 2020, pp. 7857–7866.

[70] S. Kafle, G. Joseph, P.K. Varshney, One-bit compressed sensing using untrained network prior, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2021, pp. 2875–2879.

[71] X. Meng, Y. Kabashima, Quantized compressed sensing with score-based generative models, 2022, arXiv preprint arXiv:2211.13006.

[72] J. Acharya, A. Bhattacharyya, P. Kamath, Improved bounds for universal one-bit compressive sensing, in: Proceedings of IEEE International Symposium on Information Theory, 2017, pp. 2353–2357.

[73] A. Ai, A. Lapanowski, Y. Plan, R. Vershynin, One-bit compressed sensing with non-Gaussian measurements, Linear Algebra Appl. 441 (2014) 222–239.

[74] S. Dirksen, S. Mendelson, Robust one-bit compressed sensing with partial circulant matrices, 2018, arXiv preprint arXiv:1812.06719.

[75] Y. Plan, R. Vershynin, Dimension reduction by random hyperplane tessellations, Discrete Comput. Geom. 51 (2) (2014) 438–461.

[76] S. Oymak, B. Recht, Near-optimal bounds for binary embeddings of arbitrary sets, 2015, arXiv preprint arXiv:1512.04433.

[77] S. Dirksen, S. Mendelson, Non-Gaussian hyperplane tessellations and robust one-bit compressed sensing, J. Eur. Math. Soc. 23 (9) (2021) 2913–2947.

[78] J.N. Laska, Z. Wen, W. Yin, R.G. Baraniuk, Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements, IEEE Trans. Signal Process. 59 (11) (2011) 5289–5301.

[79] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (3) (2009) 265–274.

[80] D. Liu, S. Li, Y. Shen, One-bit compressive sensing with projected subgradient method under sparsity constraints, IEEE Trans. Inform. Theory 65 (10) (2019) 6650–6663.

[81] N. Matsumoto, A. Mazumdar, Binary iterative hard thresholding converges with optimal number of measurements for 1-bit compressed sensing, in: Proceedings of IEEE Annual Symposium on Foundations of Computer Science, 2022, pp. 813–822.

[82] N. Matsumoto, A. Mazumdar, Robust 1-bit compressed sensing with iterative hard thresholding, in: Proceedings of Annual ACM-SIAM Symposium on Discrete Algorithms, 2024, pp. 2941–2979.

[83] S. Qiu, X. Wei, Z. Yang, Robust one-bit recovery via ReLU generative networks: Improved statistical rates and global landscape analysis, 2019, arXiv preprint arXiv:1908.05368.

[84] J. Haupt, R. Baraniuk, Robust support recovery using sparse compressive sensing matrices, in: Proceedings of Annual Conference on Information Sciences and Systems, 2011, pp. 1–6.

[85] S. Gopi, P. Netrapalli, P. Jain, A. Nori, One-bit compressed sensing: Provable support and vector recovery, in: Proceedings of International Conference on Machine Learning, 2013, pp. 154–162.

[86] L. Flodin, V. Gandikota, A. Mazumdar, Superset technique for approximate recovery in one-bit compressed sensing, in: Proceedings of International Conference on Neural Information Processing Systems, 2019, pp. 10387–10396.

[87] A. Mazumdar, S. Pal, Support recovery in universal one-bit compressed sensing, 2021, arXiv preprint arXiv:2107.09091.

[88] N. Matsumoto, A. Mazumdar, S. Pal, Improved support recovery in universal one-bit compressed sensing, IEEE Trans. Inform. Theory (2023).

[89] P. Erdös, P. Frankl, Z. Füredi, Families of finite sets in which no set is covered by the union of two others, J. Combin. Theory Ser. A 33 (2) (1982) 158–166.

[90] R. De Wolf, Efficient data structures from union-free families of sets, Unpubl. Manuscr. (2012).

[91] A. Gupta, R. Nowak, B. Recht, Sample complexity for 1-bit compressed sensing and sparse classification, in: Proceedings of IEEE International Symposium on Information Theory, 2010, pp. 1553–1557.

[92] S. Bulusu, V. Gandikota, P.K. Varshney, 1-bit compressed sensing with local sparsity patterns, in: Proceedings of IEEE International Symposium on Information Theory, 2023, pp. 1172–1177.

[93] A. Bora, A. Jalal, E. Price, A.G. Dimakis, Compressed sensing using generative models, in: Proceedings of International Conference on Machine Learning, 2017, pp. 537–546.

[94] J. Scarlett, R. Heckel, M.R. Rodrigues, P. Hand, Y.C. Eldar, Theoretical perspectives on deep learning methods in inverse problems, IEEE J. Sel. Areas Inf. Theory 3 (3) (2022) 433–453.

[95] X. Wei, Z. Yang, Z. Wang, On the statistical rate of nonlinear recovery in generative models with heavy-tailed data, in: Proceedings of International Conference on Machine Learning, 2019, pp. 6697–6706.

[96] S. Khobahi, N. Naimipour, M. Soltanalian, Y.C. Eldar, Deep signal recovery with one-bit quantization, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 2987–2991.

[97] S. Kafle, T. Wimalajeewa, P.K. Varshney, Noisy one-bit compressed sensing with side-information, IEEE Trans. Signal Process. (2024) 3792–3804.

[98] J. Chen, Z. Liu, Efficient algorithms for non-Gaussian single index models with generative priors, Proc. the AAAI Conf. Artif. Intell. 38 (10) (2024) 11346–11354.

[99] H. Boche, G. Caire, R. Calderbank, M. März, G. Kutyniok, R. Mathar, Compressed Sensing and Its Applications: Second International MATHEON Conference 2015, Birkhäuser, 2018.

[100] V. Schellekens, Extending the Compressive Statistical Learning Framework: Quantization, Privacy and Beyond (Ph.D. thesis), Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2021.

[101] K. Knudson, R. Saab, R. Ward, One-bit compressive sensing with norm estimation, IEEE Trans. Inform. Theory 62 (5) (2016) 2748–2758.

[102] M.A. Davenport, P.T. Boufounos, M.B. Wakin, R.G. Baraniuk, Signal processing with compressive measurements, IEEE J. Sel. Top. Signal Process. 4 (2) (2010) 445–460.

[103] H. Reboredo, F. Renna, R. Calderbank, M.R. Rodrigues, Bounds on the number of measurements for reliable compressive classification, IEEE Trans. Signal Process. 64 (22) (2016) 5778–5793.

[104] N. Keriven, N. Tremblay, Y. Traonmilin, R. Gribonval, Compressive K-means, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2017, pp. 6369–6373.

[105] T. Wimalajeewa, H. Chen, P.K. Varshney, Performance limits of compressive sensing-based signal classification, IEEE Trans. Signal Process. 60 (6) (2012) 2758–2770.

[106] Q. Liu, W. Xu, Y. Dang, X. Wang, Compressive classification based on one-bit measurements, in: Proceedings of IEEE/CIC International Conference on Communications in China, 2019, pp. 101–105.

[107] W. Xu, Q. Liu, Y. Wang, X. Bian, Performance limits of one-bit compressive classification, Signal Process. 178 (2021) 107808.

[108] V. Schellekens, L. Jacques, Quantized compressive k-means, IEEE Signal Process. Lett. 25 (8) (2018) 1211–1215.

[109] V. Schellekens, A. Chatalic, F. Houssiau, Y.-A. De Montjoye, L. Jacques, R. Gribonval, Differentially private compressive k-means, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 7933–7937.

[110] D.A. Reynolds, et al., Gaussian mixture models, Encycl. Biom. 741 (659–663) (2009).

[111] R.D. De Veaux, Mixtures of linear regressions, Comput. Statist. Data Anal. 8 (3) (1989) 227–245.

[112] P. Liang, A. Bouchard-Côté, D. Klein, B. Taskar, An end-to-end discriminative approach to machine translation, in: Proceedings of International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics, 2006, pp. 761–768.

[113] P. Deb, A.M. Holmes, Estimates of use and costs of behavioural health care: A comparison of standard and finite mixture models, Heal. Econ. 9 (6) (2000) 475–489.

[114] E. Blackwell, C.F.M. De Leon, G.E. Miller, Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine, Psychosom. Med. 68 (6) (2006) 870–878.

[115] A. Quattoni, M. Collins, T. Darrell, Conditional random fields for object recognition, in: Advances in Neural Information Processing Systems, 2004, pp. 1097–1104.

[116] Y. Sun, S. Ioannidis, A. Montanari, Learning mixtures of linear classifiers, in: Proceedings of International Conference on Machine Learning, 2014, pp. 721–729.

[117] V. Gandikota, A. Mazumdar, S. Pal, Recovery of sparse linear classifiers from mixture of responses, in: Advances in Neural Information Processing Systems, 2020, pp. 14688–14698.

[118] S. Pal, A. Mazumdar, V. Gandikota, Support recovery of sparse signals from a mixture of linear measurements, in: Advances in Neural Information Processing Systems, 2021, pp. 19082–19094.

[119] N. Polyanskii, On learning sparse vectors from mixture of responses, in: Advances in Neural Information Processing Systems, 2021, pp. 19876–19887.

[120] S. Pal, A. Mazumdar, On learning mixture models with sparse parameters, in: International Conference on Artificial Intelligence and Statistics, 2022, pp. 9182–9213.

[121] T. Wimalajeewa, P.K. Varshney, Application of compressive sensing techniques in distributed sensor networks: A survey, 2017, arXiv preprint arXiv:1709.10401.

[122] D. Baron, M.F. Duarte, M.B. Wakin, S. Sarvotham, R.G. Baraniuk, Distributed compressive sensing, 2009, arXiv preprint arXiv:0901.3403.

[123] J. Chen, X. Huo, Theoretical results on sparse representations of multiple-measurement vectors, IEEE Trans. Signal Process. 54 (12) (2006) 4634–4643.

[124] S.F. Cotter, B.D. Rao, K. Engan, K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, IEEE Trans. Signal Process. 53 (7) (2005) 2477–2488.

[125] S. Kafle, V. Gupta, B. Kailkhura, T. Wimalajeewa, P.K. Varshney, Joint sparsity pattern recovery with 1-b compressive sensing in distributed sensor networks, IEEE Trans. Signal Inf. Process. over Networks 5 (1) (2018) 15–30.

[126] H. Zayyani, F. Haddadi, M. Korki, Double detector for sparse signal detection from one-bit compressed sensing measurements, IEEE Signal Process. Lett. 23 (11) (2016) 1637–1641.

[127] X. Wang, G. Li, P.K. Varshney, Detection of sparse stochastic signals with quantized measurements in sensor networks, IEEE Trans. Signal Process. 67 (8) (2019) 2210–2220.

[128] C. Li, Y. He, X. Wang, G. Li, P.K. Varshney, Distributed detection of sparse stochastic signals via fusion of 1-bit local likelihood ratios, IEEE Signal Process. Lett. 26 (12) (2019) 1738–1742.

[129] C. Gianelli, L. Xu, J. Li, P. Stoica, One-bit compressive sampling with time-varying thresholds for sparse parameter estimation, in: Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop, 2016, pp. 1–5.

[130] C. Gianelli, L. Xu, J. Li, P. Stoica, One-bit compressive sampling with time-varying thresholds for multiple sinusoids, in: Proceedings of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2017, pp. 1–5.

[131] S. Huang, T.D. Tran, 1-bit compressive sensing via approximate message passing with built-in parameter estimation, 2020, arXiv preprint arXiv:2007.07679.

[132] P.K. Varshney, Distributed Detection and Data Fusion, Springer Science & Business Media, 2012.

[133] G. Shi, J. Lin, X. Chen, F. Qi, D. Liu, L. Zhang, UWB echo signal detection with ultra-low rate sampling based on compressed sensing, IEEE Trans. Circuits Syst. II: Express Briefs 55 (4) (2008) 379–383.

[134] S. Hong, Direct spectrum sensing from compressed measurements, in: Proceedings of Military Communications Conference, 2010, pp. 1187–1192.

[135] J. Meng, H. Li, Z. Han, Sparse event detection in wireless sensor networks using compressive sensing, in: Proceedings of Annual Conference on Information Sciences and Systems, 2009, pp. 181–185.

[136] T. Sakdejayont, D. Lee, Y. Peng, Y. Yamashita, H. Morikawa, Evaluation of memory-efficient 1-bit compressed sensing in wireless sensor networks, in: Proceedings of IEEE Region 10 Humanitarian Technology Conference, 2013, pp. 326–329.

[137] S. Ge, D. Feng, S. Song, J. Wang, X. Huang, Sparse logistic regression based one-bit SAR imaging, IEEE Trans. Geosci. Remote Sens. (2023).

[138] S. Ge, N. Jiang, D. Feng, S. Song, J. Wang, J. Zhu, X. Huang, Enhanced one-bit SAR imaging method using two-level structured sparsity to mitigate adverse effects of sign flips, IEEE Trans. Geosci. Remote Sens. (2024).

[139] J. Han, G. Li, X.-P. Zhang, One-bit radar imaging via adaptive binary iterative hard thresholding, IEEE Trans. Comput. Imaging 7 (2021) 1005–1017.

[140] J. Han, G. Li, K. Wang, M. Duan, X.-P. Zhang, 1-bit radar imaging based on adversarial samples, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–13.

[141] X. Huang, P. Xiao, B. Liao, One-bit direction of arrival estimation with an improved fixed-point continuation algorithm, in: Proceedings of International Conference on Wireless Communications and Signal Processing, 2018, pp. 1–4.

[142] P. Xiao, B. Liao, N. Deligiannis, Deepfpc: A deep unfolded network for sparse signal recovery from 1-bit measurements with application to DOA estimation, Signal Process. 176 (2020) 107699.

[143] Z. Wei, W. Wang, F. Dong, Q. Liu, Gridless one-bit direction-of-arrival estimation via atomic norm denoising, IEEE Commun. Lett. 24 (10) (2020) 2177–2181.

[144] C. Stöckle, J. Munir, A. Mezghani, J.A. Nossek, 1-bit direction of arrival estimation based on compressed sensing, in: Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications, 2015, pp. 246–250.

[145] P. Wang, H. Yang, Z. Ye, 1-bit direction of arrival estimation via improved complex-valued binary iterative hard thresholding, Digit. Signal Process. 120 (2022) 103265.

[146] X. Meng, J. Zhu, A generalized sparse Bayesian learning algorithm for 1-bit DOA estimation, IEEE Commun. Lett. 22 (7) (2018) 1414–1417.

[147] L. Feng, L. Huang, Q. Li, Z.-Q. He, M. Chen, An off-grid iterative reweighted approach to one-bit direction of arrival estimation, IEEE Trans. Veh. Technol. 72 (6) (2023) 8134–8139.

[148] S. Yan, Collaborative spectrum sensing for cognitive radio networks by using 1-bit compressed sensing, in: Proceedings of the International Conference on Communication and Information Processing, 2018, pp. 289–293.

[149] S. Yan, M. Liu, J. Si, Distributed collaborative spectrum sensing using 1-bit compressive sensing in cognitive radio networks, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. 103 (1) (2020) 382–388.

[150] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcic, R.W. Heath, Adaptive one-bit compressive sensing with application to low-precision receivers at mmWave, in: Proceedings of IEEE Global Communications Conference, 2015, pp. 1–6.

[151] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A.L. Swindlehurst, L. Liu, Channel estimation and performance analysis of one-bit massive MIMO systems, IEEE Trans. Signal Process. 65 (15) (2017) 4075–4089.

[152] U.S. Kamilov, V.K. Goyal, S. Rangan, Message-passing de-quantization with applications to compressed sensing, IEEE Trans. Signal Process. 60 (12) (2012) 6270–6281.

[153] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, P. Ting, Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs, IEEE Trans. Signal Process. 64 (10) (2015) 2541–2556.

[154] B. Srinivas, P. Priya, D. Sen, S. Chakrabarti, Channel estimation in sub-6 GHz and hybrid millimeter wave MIMO systems with low-resolution ADCs, IEEE Trans. Green Commun. Netw. 7 (2) (2022) 707–723.

[155] J. Ma, S. Zhang, H. Li, F. Gao, Z. Han, Time-varying downlink channel tracking for quantized massive MIMO networks, IEEE Trans. Wirel. Commun. 19 (10) (2020) 6721–6736.

[156] W. Fan, Y. Xia, C. Li, Y. Huang, Bayesian channel tracking and AoA acquisition in millimeter wave MIMO systems with low-resolution ADCs, in: Proceedings of Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 2022, pp. 1355–1360.

[157] Z. Zhou, X. Chen, D. Guo, M.L. Honig, Sparse channel estimation for massive MIMO with 1-bit feedback per dimension, in: Proceedings of IEEE Wireless Communications and Networking Conference, 2017, pp. 1–6.

[158] F. Kulsoom, A. Vizziello, H.N. Chaudhry, P. Savazzi, Joint sparse channel recovery with quantized feedback for multi-user massive MIMO systems, IEEE Access 8 (2020) 11046–11060.

[159] C. Qian, X. Fu, N.D. Sidiropoulos, Amplitude retrieval for channel estimation of MIMO systems with one-bit ADCs, IEEE Signal Process. Lett. 26 (11) (2019) 1698–1702.

[160] J.Z. Sun, V.K. Goyal, Optimal quantization of random measurements in compressed sensing, in: Proceedings of IEEE International Symposium on Information Theory, 2009, pp. 6–10.

[161] A. Shirazinia, S. Chatterjee, M. Skoglund, Analysis-by-synthesis quantization for compressed sensing measurements, IEEE Trans. Signal Process. 61 (22) (2013) 5789–5800.

[162] A. Zymnis, S. Boyd, E. Candes, Compressed sensing with quantized measurements, IEEE Signal Process. Lett. 17 (2) (2009) 149–152.

[163] A. Shirazinia, S. Chatterjee, M. Skoglund, Joint source-channel vector quantization for compressed sensing, IEEE Trans. Signal Process. 62 (14) (2014) 3667–3681.

[164] A. Shirazinia, S. Chatterjee, M. Skoglund, Distributed quantization for measurement of correlated sparse sources over noisy channels, 2014, arXiv preprint arXiv:1404.7640.

[165] V.K. Goyal, A.K. Fletcher, S. Rangan, Compressive sampling and lossy compression, IEEE Signal Process. Mag. 25 (2) (2008) 48–56.

[166] A. Kipnis, G. Reeves, Y.C. Eldar, A.J. Goldsmith, Compressed sensing under optimal quantization, in: Proceedings of IEEE International Symposium on Information Theory, 2017, pp. 2148–2152.

[167] G. Coluccia, A. Roumy, E. Magli, Operational rate-distortion performance of single-source and distributed compressed sensing, IEEE Trans. Commun. 62 (6) (2014) 2022–2033.

[168] C. Weidmann, M. Vetterli, Rate distortion behavior of sparse sources, IEEE Trans. Inform. Theory 58 (8) (2012) 4969–4992.

[169] F. da Rocha Henriques, L. Lovisolo, E.A. Barros da Silva, Rate-distortion performance and incremental transmission scheme of compressive sensed measurements in wireless sensor networks, Sensors 19 (2) (2019) 266.

[170] M. Leinonen, M. Codreanu, M. Juntti, Practical compression methods for quantized compressed sensing, in: Proceedings of IEEE Conference on Computer Communications Workshops, 2019, pp. 756–761.

[171] M. Leinonen, M. Codreanu, M. Juntti, Signal reconstruction performance under quantized noisy compressed sensing, in: Proceedings of Data Compression Conference, 2019, pp. 586–586.

[172] G. Franceschetti, S. Merolla, M. Tesauro, Phase quantized SAR signal processing: Theory and experiments, IEEE Trans. Aerosp. Electron. Syst. 35 (1) (1999) 201–214.

[173] G. Fornaro, V. Pascazio, G. Schirinzi, Synthetic aperture radar interferometry using one bit coded raw and reference signals, IEEE Trans. Geosci. Remote Sens. 35 (5) (1997) 1245–1253.

[174] C. Gianelli, L. Xu, J. Li, P. Stoica, One-bit compressive sampling with time-varying thresholds: Maximum likelihood and the Cramér-Rao bound, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2016, pp. 399–403.

[175] L. Wei, P. Dharmawansa, O. Tirkkonen, Multiple primary user spectrum sensing in the low SNR regime, IEEE Trans. Commun. 61 (5) (2013) 1720–1731.

[176] L. Huang, Y.-H. Xiao, Q. Zhang, Robust spectrum sensing for noncircular signal in multiantenna cognitive receivers, IEEE Trans. Signal Process. 63 (2) (2014) 498–511.

[177] A. Sharma, C.R. Murthy, Group testing-based spectrum hole search for cognitive radios, IEEE Trans. Veh. Technol. 63 (8) (2014) 3794–3805.

[178] J. Yang, Z. Song, H. Zhang, Y. Gao, Compressive spectrum sensing with 1-bit ADCs, 2024, arXiv preprint arXiv:2411.04611.

[179] T. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas, IEEE Trans. Wirel. Commun. 9 (11) (2010) 3590–3600.

[180] F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, F. Tufvesson, Scaling up MIMO: Opportunities and challenges with very large arrays, IEEE Signal Process. Mag. 30 (1) (2013) 40–60.

[181] R.H. Walden, Analog-to-digital converter survey and analysis, IEEE J. Sel. Areas Commun. 17 (4) (1999) 539–550.

[182] C. Risi, D. Persson, E.G. Larsson, Massive MIMO with 1-bit ADC, 2014, arXiv.

[183] B. Murmann, Energy limits in A/D converters, in: Proceedings of IEEE Faible Tension Faible Consommation, 2013, pp. 1–4.

[184] Y. Li, C. Tao, A.L. Swindlehurst, A. Mezghani, L. Liu, Downlink achievable rate analysis in massive MIMO systems with one-bit DACs, IEEE Commun. Lett. 21 (7) (2017) 1669–1672.

[185] A. Mezghani, A.L. Swindlehurst, Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs, IEEE Trans. Signal Process. 66 (11) (2018) 2972–2983.

[186] H. Kim, J. Choi, Channel estimation for spatially/temporally correlated massive MIMO systems with one-bit ADCs, EURASIP J. Wirel. Commun. Netw. 2019 (1) (2019) 267.

[187] R. Zhang, H. Zhao, J. Zhang, Distributed compressed sensing aided sparse channel estimation in FDD massive MIMO system, IEEE Access 6 (2018) 18383–18397.

[188] C. Qing, Q. Ye, B. Cai, W. Liu, J. Wang, Deep learning for 1-bit compressed sensing-based superimposed CSI feedback, PLoS One 17 (3) (2022) e0265109.

[189] C. Qing, Q. Ye, W. Liu, J. Wang, Fusion learning for 1-bit CS-based superimposed CSI feedback with bi-directional channel reciprocity, IEEE Commun. Lett. 26 (4) (2022) 813–817.

[190] J. Liu, Z. Luo, X. Xiong, Low-resolution ADCs for wireless communication: A comprehensive survey, IEEE Access 7 (2019) 91291–91324.

[191] J. Ma, L. Ping, Data-aided channel estimation in large antenna systems, IEEE Trans. Signal Process. 62 (12) (2014) 3111–3124.

[192] S.S. Thoota, C.R. Murthy, Massive MIMO-OFDM systems with low resolution ADCs: Cramér-Rao bound, sparse channel estimation, and soft symbol decoding, IEEE Trans. Signal Process. 70 (2022) 4835–4850.

[193] R. Prasad, C.R. Murthy, B.D. Rao, Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach, IEEE Trans. Signal Process. 63 (20) (2015) 5369–5382.

[194] G. Joseph, C.R. Murthy, A noniterative online Bayesian algorithm for the recovery of temporally correlated sparse vectors, IEEE Trans. Signal Process. 65 (20) (2017) 5510–5525.

[195] K.P. Arunkumar, C.R. Murthy, Iterative sparse channel estimation and data detection for underwater acoustic communications using partial interval demodulation, IEEE Trans. Signal Process. 66 (19) (2018) 5041–5055.

[196] K. Arunkumar, C.R. Murthy, Soft symbol decoding in sweep-spread-carrier underwater acoustic communications: A novel variational Bayesian algorithm and its analysis, IEEE Trans. Signal Process. 68 (2020) 2435–2448.

[197] D. Prasanna, C.R. Murthy, mmWave channel estimation via compressive covariance estimation: Role of sparsity and intra-vector correlation, IEEE Trans. Signal Process. 69 (2021) 2356–2370.

[198] S.S. Thoota, C.R. Murthy, Variational Bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low resolution ADCs, IEEE Trans. Commun. 69 (5) (2021) 3467–3481.

[199] Y. He, G. Joseph, Bayesian algorithms for Kronecker-structured sparse vector recovery with application to IRS-MIMO channel estimation, IEEE Trans. Signal Process. (2024) 1–15.

[200] C.M. Bishop, Pattern Recognition and Machine Learning, Springer New York, 2006.

[201] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.

[202] 3GPP, TS 38.212 V15.7.0 (2019-09); NR; Multiplexing and channel coding, 2019.

[203] N.K. Chavali, A.R. Yalla, A soft-demapper for coded MIMO-OFDM system, in: Proceedings of International Conference on Contemporary Computing and Informatics, 2014, pp. 451–457.

[204] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, G. Chen, Millimeter wave communication: A comprehensive survey, IEEE Commun. Surv. & Tutorials 20 (3) (2018) 1616–1653.

[205] Q. Wu, R. Zhang, Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network, IEEE Commun. Mag. 58 (1) (2020) 106–112.

[206] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, C. Studer, Throughput analysis of massive MIMO uplink with low-resolution ADCs, IEEE Trans. Wirel. Commun. 16 (6) (2017) 4038–4051.

[207] X. Hu, R. Zhang, C. Zhong, Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications, IEEE Trans. Wirel. Commun. 21 (2) (2022) 1132–1142.

[208] Y. Jin, J. Zhang, X. Zhang, H. Xiao, B. Ai, D.W.K. Ng, Channel estimation for semi-passive reconfigurable intelligent surfaces with enhanced deep residual networks, IEEE Trans. Veh. Technol. 70 (10) (2021) 11083–11088.

[209] I.-S. Kim, M. Bennis, J. Oh, J. Chung, J. Choi, Bayesian channel estimation for intelligent reflecting surface-aided mmWave massive MIMO systems with semi-passive elements, IEEE Trans. Wirel. Commun. 22 (12) (2023) 9732–9745.

[210] Q. Wu, R. Zhang, Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming, IEEE Trans. Wirel. Commun. 18 (11) (2019) 5394–5409.

[211] S. Abeywickrama, R. Zhang, Q. Wu, C. Yuen, Intelligent reflecting surface: Practical phase shift model and beamforming optimization, IEEE Trans. Commun. 68 (9) (2020) 5849–5863.

[212] Q. Wu, S. Zhang, B. Zheng, C. You, R. Zhang, Intelligent reflecting surface-aided wireless communications: A tutorial, IEEE Trans. Commun. 69 (5) (2021) 3313–3351.

[213] C. Hu, L. Dai, S. Han, X. Wang, Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications, IEEE Trans. Commun. 69 (11) (2021) 7736–7747.

[214] A. Sayeed, Deconstructing multiantenna fading channels, IEEE Trans. Signal Process. 50 (10) (2002) 2563–2579.

[215] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (8) (2009) 30–37.

[216] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, Found. Comput. Math. 9 (6) (2009) 717–772.

[217] D. Wipf, B. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.

[218] D.G. Tzikas, A.C. Likas, N.P. Galatsanos, The variational approximation for Bayesian inference, IEEE Signal Process. Mag. 25 (6) (2008) 131–146.

[219] M.J. Wainwright, M.I. Jordan, Graphical models, exponential families, and variational inference, Found. Trends® Mach. Learn. 1 (1–2) (2008) 1–305.

[220] P.D. Hoff, A First Course in Bayesian Statistical Methods, Springer, 2009.

[221] Report ITU-R M.2412-0, Guidelines for Evaluation of Radio Interface Technologies for IMT-2020, Tech. Rep., ITU-R, 2017.

[222] N. Shlezinger, Y.C. Eldar, M.R.D. Rodrigues, Hardware-limited task-based quantization, IEEE Trans. Signal Process. 67 (20) (2019).

[223] N. Shlezinger, Y.C. Eldar, M.R.D. Rodrigues, Asymptotic task-based quantization with application to massive MIMO, IEEE Trans. Signal Process. 67 (15) (2019) 3995–4012.

[224] N. Shlezinger, Y.C. Eldar, Task-based quantization with application to MIMO receivers, Commun. Inf. Syst. 20 (2020) 131–162.

[225] S. Salamatian, N. Shlezinger, Y.C. Eldar, M. Médard, Task-based quantization for recovering quadratic functions using principal inertia components, in: Proceedings of IEEE International Symposium on Information Theory, 2019, pp. 390–394.

[226] N.I. Bernardo, J. Zhu, Y.C. Eldar, J. Evans, Design and analysis of hardware-limited non-uniform task-based quantizers, IEEE Trans. Signal Process. 71 (2023) 1551–1562.

[227] G. Lee, I.-s. Kim, Y.C. Eldar, A.L. Swindlehurst, J. Choi, Channel estimation for RIS-aided communication systems: A task-based quantization approach, in: Proceedings of International Symposium on Wireless Communication Systems, 2024, pp. 1–6.

[228] E. Masry, The reconstruction of analog signals from the sign of their noisy samples, IEEE Trans. Inform. Theory 27 (6) (1981) 735–745.

[229] N. Naimipour, M. Soltanalian, Graph clustering using one-bit comparison data, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2018, pp. 1998–2001.

[230] C. Li, R. Zhang, J. Li, P. Stoica, Bayesian information criterion for signed measurements with application to sinusoidal signals, IEEE Signal Process. Lett. 25 (8) (2018) 1251–1255.

[231] A. Host-Madsen, P. Handel, Effects of sampling and quantization on single-tone frequency estimation, IEEE Trans. Signal Process. 48 (3) (2000) 650–662.

[232] O. Bar-Shalom, A.J. Weiss, DOA estimation using one-bit quantized measurements, IEEE Trans. Aerosp. Electron. Syst. 38 (3) (2002) 868–884.

[233] O. Dabeer, E. Masry, Multivariate signal parameter estimation under dependent noise from 1-bit dithered quantized data, IEEE Trans. Inform. Theory 54 (4) (2008) 1637–1654.

[234] O. Dabeer, A. Karnik, Signal parameter estimation using 1-bit dithered quantization, IEEE Trans. Inform. Theory 52 (12) (2006) 5389–5405.

[235] S. Khobahi, M. Soltanalian, Signal recovery from 1-bit quantized noisy samples via adaptive thresholding, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2018, pp. 1757–1761.

[236] J. Chen, C.-L. Wang, M.K. Ng, D. Wang, High dimensional statistical estimation under uniformly dithered one-bit quantization, IEEE Trans. Inform. Theory 69 (8) (2023) 5151–5187.

[237] F. Mayer, C. Vogel, An optimization-based approach to one-bit quantization, in: Proceedings of IEEE International Symposium on Circuits and Systems, 2024, pp. 1–5.

[238] R. Huang, X. Chen, Deep blind one-bit signal recovery, in: Proceedings of IEEE International Workshop on Machine Learning for Signal Processing, 2020, pp. 1–6.

[239] P.C. Cosman, R.M. Gray, M. Vetterli, Vector quantization of image subbands: A survey, IEEE Trans. Image Process. 5 (2) (1996) 202–225.

[240] S.A. Lashari, R. Ibrahim, N. Taujuddin, N. Senan, S. Sari, Thresholding and quantization algorithms for image compression techniques: A review, Asia- Pac. J. Inf. Technol. Multimed. 7 (1) (2018) 83–89.

[241] W. Lin, L. Dong, Adaptive downsampling to improve image compression at low bit rates, IEEE Trans. Image Process. 15 (9) (2006) 2513–2521.

[242] W.C. Chu, Vector quantization of harmonic magnitudes in speech coding applications - A survey and new technique, EURASIP J. Adv. Signal Process. 2004 (2004) 1–13.

[243] O. Derrien, P. Duhamel, M. Charbit, G. Richard, A new quantization optimization algorithm for the MPEG advanced audio coder using a statistical subband model of the quantization noise, IEEE Trans. Audio, Speech, Lang. Process. 14 (4) (2006) 1328–1339.

[244] R. Gong, Y. Ding, Z. Wang, C. Lv, X. Zheng, J. Du, H. Qin, J. Guo, M. Magno, X. Liu, A survey of low-bit large language models: Basics, systems, and algorithms, 2024, arXiv preprint arXiv:2409.16694.

[245] B. Rokh, A. Azarpeyvand, A. Khanteymoori, A comprehensive survey on model quantization for deep neural networks in image classification, ACM Trans. Intell. Syst. Technol. 14 (6) (2023) 1–50.

[246] Y. Guo, A survey on methods and theories of quantized neural networks, 2018, arXiv preprint arXiv:1808.04752.

[247] H. Godrich, A.M. Haimovich, R.S. Blum, Target localization accuracy gain in MIMO radar-based systems, IEEE Trans. Inform. Theory 56 (6) (2010) 2783–2803.

[248] A. Ribeiro, G. Giannakis, Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case, IEEE Trans. Signal Process. 54 (3) (2006) 1131–1143.

[249] X. Yang, R. Niu, E. Masazade, P.K. Varshney, Channel-aware tracking in multi-hop wireless sensor networks with quantized measurements, IEEE J. Sel. Areas Commun. 49 (4) (2013) 2353–2368.

[250] R. Niu, P.K. Varshney, Target location estimation in sensor networks with quantized data, IEEE Trans. Signal Process. 54 (12) (2006) 4519–4528.

[251] Y. Yan, G. Yang, H. Wang, X. Shen, Semidefinite relaxation for source localization with quantized ToA measurements and transmission uncertainty in sensor networks, IEEE Trans. Commun. 69 (2) (2020) 1201–1213.

[252] Z. Wang, Q. He, R.S. Blum, Quantized time delay for target localization in cloud MIMO radar, in: Proceedings of IEEE Radar Conference, 2021, pp. 1–5.

[253] O. Ozdemir, R. Niu, P.K. Varshney, Channel aware target localization with quantized data in wireless sensor networks, IEEE Trans. Signal Process. 57 (3) (2009) 1190–1202.

[254] G. Zhang, W. Yi, M. Matthaiou, P.K. Varshney, Direct target localization with low-bit quantization in wireless sensor networks, IEEE Trans. Signal Process. 72 (2024) 3059–3075.

[255] G. Zhang, W. Yi, P.K. Varshney, L. Kong, Direct target localization with quantized measurements in noncoherent distributed MIMO radar systems, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–18.

[256] L. Ni, D. Zhang, Y. Sun, N. Liu, J. Liang, Q. Wan, Detection and localization of one-bit signal in multiple distributed subarray systems, IEEE Trans. Signal Process. 71 (2023) 2776–2791.

[257] Z. Wang, B. Tan, E.S. Lohan, L. Ni, M. Valkama, Q. Wan, Majorization-minimization-based direct localization using one-bit channel measurements, IEEE Wirel. Commun. Lett. 13 (5) (2024) 1389–1393.

[258] A. Weiss, G.W. Wornell, One-bit direct position determination of narrowband Gaussian signals, in: Proccedings of IEEE Workshop on Statistical Signal Processing, 2021, pp. 466–470.

[259] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inform. Theory 28 (2) (1982) 129–137.

[260] W. Li, M. Li, A. Liu, T.X. Han, Design and optimization of cooperative sensing with limited backhaul capacity, in: Proceedings of IEEE Vehicular Technology Conference, 2023, pp. 1–6.

[261] F. Gao, L. Guo, H. Li, J. Liu, J. Fang, Quantizer design for distributed GLRT detection of weak signal in wireless sensor networks, IEEE Trans. Wirel. Commun. 14 (4) (2014) 2032–2042.

[262] A. Ababneh, Low-complexity bit allocation for RSS target localization, IEEE Sensors J. 19 (17) (2019) 7733–7743.

[263] Q. An, J. Wang, Z. Zhang, Y. Shen, Information-based bit allocation for cooperative visual sensing in vehicular networks, IEEE Trans. Veh. Technol. 72 (2) (2023) 2365–2380.

[264] X.R. Li, V.P. Jilkov, Survey of maneuvering target tracking. Part I. Dynamic models, IEEE Trans. Aerosp. Electron. Syst. 39 (4) (2003) 1333–1364.

[265] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2) (2002) 174–188.

[266] J. Fang, Y. Liu, H. Li, S. Li, One-bit quantizer design for multisensor GLRT fusion, IEEE Signal Process. Lett. 20 (3) (2013) 257–260.

[267] R. Jiang, B. Chen, Fusion of censored decisions in wireless sensor networks, IEEE Trans. Wirel. Commun. 4 (6) (2005) 2668–2673.

[268] R. Niu, B. Chen, P.K. Varshney, Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks, IEEE Trans. Signal Process. 54 (3) (2006) 1018–1027.

[269] B. Chen, L. Tong, P.K. Varshney, Channel-aware distributed detection in wireless sensor networks, IEEE Signal Process. Mag. 23 (4) (2006) 16–26.

[270] B. Liu, B. Chen, Channel-optimized quantizers for decentralized detection in sensor networks, IEEE Trans. Inform. Theory 52 (7) (2006) 3349–3358.

[271] O. Ozdemir, R. Niu, P.K. Varshney, Channel aware particle filtering for tracking in sensor networks, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2006, pp. 290–294.

[272] O. Ozdemir, R. Niu, P.K. Varshney, Tracking in wireless sensor networks using particle filtering: Physical layer considerations, IEEE Trans. Signal Process. 57 (5) (2009) 1987–1999.

[273] Y. Yuan, W. Yi, W. Choi, L. Kong, Dynamic quantizer design for target tracking for wireless sensor network with imperfect channels, IEEE Trans. Wirel. Commun. 22 (3) (2022) 1695–1711.

[274] P. Tichavsky, C.H. Muravchik, A. Nehorai, Posterior Cramér-Rao bounds for discrete-time nonlinear filtering, IEEE Trans. Signal Process. 46 (5) (1998) 1386–1396.

[275] Y. Yuan, W. Yi, P.K. Varshney, Exponential mixture density based approximation to posterior Cramér-Rao lower bound for distributed target tracking, IEEE Trans. Signal Process. 70 (2022) 862–877.

[276] S. Liu, E. Masazade, X. Shen, P.K. Varshney, Adaptive non-myopic quantizer design for target tracking in wireless sensor networks, in: Proceedings of Asilomar Conference on Signals, Systems, and Computers, 2013, pp. 1085–1089.

[277] A. Vempaty, H. He, B. Chen, P.K. Varshney, On quantizer design for distributed Bayesian estimation in sensor networks, IEEE Trans. Signal Process. 62 (20) (2014) 5359–5369.

[278] A. Ali, W. Hamouda, Power-efficient wideband spectrum sensing for cognitive radio systems, IEEE Trans. Veh. Technol. 67 (4) (2017) 3269–3283.

[279] A. Ali, W. Hamouda, A novel one-bit quantization design for correlation-based low-power wideband sensing, in: Proceedings of IEEE International Conference on Communications, 2016, pp. 1–6.

[280] A. Ali, W. Hamouda, Low power wideband sensing for one-bit quantized cognitive radio systems, IEEE Wirel. Commun. Lett. 5 (1) (2015) 16–19.

[281] J.C. Merlano-Duncan, S.K. Sharma, S. Chatzinotas, B. Ottersten, X. Wang, Multi-antenna based one-bit spatio-temporal wideband sensing for cognitive radio networks, in: Proceedings of IEEE International Conference on Communications, 2017, pp. 1–7.

[282] A. Ali, W. Hamouda, Cooperative low-power wideband sensing based on 1-bit quantization, IEEE Commun. Lett. 22 (2) (2017) 368–371.

[283] A. Ali, W. Hamouda, Generalized FFT-based one-bit quantization system for wideband spectrum sensing, IEEE Trans. Commun. 68 (1) (2019) 82–92.

[284] D. Ciuonzo, G. Papa, G. Romano, P.S. Rossi, P. Willett, One-bit decentralized detection with a Rao test for multisensor fusion, IEEE Signal Process. Lett. 20 (9) (2013) 861–864.

[285] T. Yucek, H. Arslan, A survey of spectrum sensing algorithms for cognitive radio applications, IEEE Commun. Surv. & Tutorials 11 (1) (2009) 116–130.

[286] Y. Zhao, X. Ke, B. Zhao, Y. Xiao, L. Huang, One-bit spectrum sensing based on statistical covariances: Eigenvalue moment ratio approach, IEEE Wirel. Commun. Lett. 10 (11) (2021) 2474–2478.

[287] P.-W. Wu, L. Huang, D. Ramírez, Y.-H. Xiao, H.C. So, One-bit spectrum sensing for cognitive radio, IEEE Trans. Signal Process. (2024).

[288] T.M. Lok, V.-W. Wei, Channel estimation with quantized observations, in: Proceedings of IEEE International Symposium on Information Theory, 1998, p. 333.

[289] M.T. Ivrlac, J.A. Nossek, On MIMO channel estimation with single-bit signal-quantization, in: Proceedings of ITG Mmart Antenna Workshop, 2007, pp. 1–7.

[290] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, C. Studer, One-bit massive MIMO: Channel estimation and high-order modulations, in: Proceedings of IEEE International Conference on Communication Workshop, 2015, pp. 1304–1309.

[291] A. Mezghani, F. Antreich, J.A. Nossek, Multiple parameter estimation with quantized channel output, in: Proceedings of International ITG Workshop on Smart Antennas, 2010, pp. 143–150.

[292] C. Studer, G. Durisi, Quantized massive MU-MIMO-OFDM uplink, IEEE Trans. Commun. 64 (6) (2016) 2387–2399.

[293] M.S. Stein, S. Bar, J.A. Nossek, J. Tabrikian, Performance analysis for channel estimation with 1-bit ADC and unknown quantization threshold, IEEE Trans. Signal Process. 66 (10) (2018) 2557–2571.

[294] S. Rao, A. Mezghani, A.L. Swindlehurst, Channel estimation in one-bit massive MIMO systems: Angular versus unstructured models, IEEE J. Sel. Top. Signal Process. 13 (5) (2019) 1017–1031.

[295] J. Choi, D.J. Love, D.R. Brown, Channel estimation techniques for quantized distributed reception in MIMO systems, in: Proceedings of Asilomar Conference on Signals, Systems and Computers, 2014, pp. 1066–1070.

[296] J. Choi, J. Mo, R.W. Heath, Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs, IEEE Trans. Commun. 64 (5) (2016) 2005–2018.

[297] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, P. Ting, Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs, IEEE Trans. Signal Process. 64 (2016) 2541–2556.

[298] C. Mollen, J. Choi, E.G. Larsson, R.W. Heath, Uplink performance of wideband massive MIMO with one-bit ADCs, IEEE Trans. Wirel. Commun. 16 (1) (2016) 87–100.

[299] Q. Wan, J. Fang, H. Duan, Z. Chen, H. Li, Generalized bussgang LMMSE channel estimation for one-bit massive MIMO systems, IEEE Trans. Wirel. Commun. 19 (6) (2020) 4234–4246.

[300] B. Fesl, M. Koller, W. Utschick, On the mean square error optimal estimator in one-bit quantized systems, IEEE Trans. Signal Process. 71 (2023) 1968–1980.

[301] M. Ding, I. Atzeni, A. Tölli, A.L. Swindlehurst, On optimal MMSE channel estimation for one-bit quantized MIMO systems, IEEE Trans. Signal Process. (2025).

[302] F. Wang, J. Fang, H. Li, Z. Chen, S. Li, One-bit quantization design and channel estimation for massive MIMO systems, IEEE Trans. Veh. Technol. 67 (11) (2018) 10921–10934.

[303] J.N. Laska, P.T. Boufounos, M.A. Davenport, R.G. Baraniuk, Democracy in action: Quantization, saturation, and compressive sensing, Appl. Comput. Harmon. Anal. 31 (3) (2011) 429–443.

[304] S. Foucart, T. Needham, Sparse recovery from saturated measurements, Inf. Inference: A J. the IMA 6 (2) (2017) 196–212.

[305] A. Bhandari, F. Krahmer, R. Raskar, On unlimited sampling, in: Proceedings of International Conference on Sampling Theory and Applications, 2017, pp. 31–35.

[306] A. Bhandari, F. Krahmer, R. Raskar, On unlimited sampling and reconstruction, IEEE Trans. Signal Process. 69 (2020) 3827–3839.

[307] A. Bhandari, F. Krahmer, R. Raskar, Methods and apparatus for modulo sampling and recovery, 2020, US10651865B2.

[308] A. Bhandari, F. Krahmer, T. Poskitt, Unlimited sampling from theory to practice: Fourier-Prony recovery and prototype ADC, IEEE Trans. Signal Process. (2021) 1131–1141.

[309] J.V. Vleck, D. Middleton, The spectrum of clipped noise, Proc. IEEE 54 (1) (1966) 2–19.

[310] B.F. Logan, Signals designed for recovery after clipping-I. Localization of infinite products, AT& T Bell Lab. Tech. J. 63 (2) (1984) 261–285.

[311] J. Abel, J. Smith, Restoring a clipped signal, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 1991, pp. 1745–1748.

[312] K.-P. Ho, Maximum-likelihood power estimation for clipped signals, IEEE Trans. Signal Process. 16 (7) (2009) 569–571.

[313] F. Esqueda, S. Bilbao, V. Valimaki, Aliasing reduction in clipped signals, IEEE Trans. Signal Process. 64 (20) (2016) 5255–5267.

[314] A. Nardecchia, V. Motto-Ros, L. Duponchel, Saturated signals in spectroscopic imaging: Why and how should we deal with this regularly observed phenomenon? Anal. Chim. Acta 1157 (338389) (2021) 1–9.

[315] B. Murmann, The race for the extra decibel: A brief review of current ADC performance trajectories, IEEE Solid- State Circuits Mag. 7 (3) (2015) 58–66.

[316] A. Sabharwal, P. Schniter, D. Guo, D.W. Bliss, S. Rangarajan, R. Wichman, In-band full-duplex wireless: Challenges and opportunities, IEEE J. Sel. Areas Commun. 32 (9) (2014) 1637–1652.

[317] J. Zhang, L. Dai, S. Sun, Z. Wang, On the spectral efficiency of massive MIMO systems with low-resolution ADCs, IEEE Commun. Lett. 20 (5) (2016) 842–845.

[318] D. Florescu, F. Krahmer, A. Bhandari, The surprising benefits of hysteresis in unlimited sampling: Theory, algorithms and experiments, IEEE Trans. Signal Process. 70 (2022) 616–630.

[319] E. Romanov, O. Ordentlich, Above the Nyquist rate, modulo folding does not hurt, IEEE Signal Process. Lett. 26 (8) (2019) 1167–1171.

[320] O. Ordentlich, G. Tabak, P.K. Hanumolu, A.C. Singer, G.W. Wornell, A modulo-based architecture for analog-to-digital conversion, IEEE J. Sel. Top. Signal Process. (2018) 825–840.

[321] G. Shtendel, D. Florescu, A. Bhandari, Unlimited sampling of bandpass signals: Computational demodulation via undersampling, IEEE Trans. Signal Process. (2023) 1–12.

[322] A. Bhandari, F. Krahmer, R. Raskar, Unlimited sampling of sparse signals, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2018, pp. 1–5.

[323] A. Bhandari, Back in the US-SR: Unlimited sampling and sparse super-resolution with its hardware validation, IEEE Trans. Signal Process. 29 (2022) 1047–1051.

[324] R. Guo, A. Bhandari, Unlimited sampling of FRI signals independent of sampling rate, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2023, pp. 1–6.

[325] A. Bhandari, F. Krahmer, R. Raskar, Unlimited sampling of sparse sinusoidal mixtures, in: Proceedings of IEEE International Symposium on Information Theory, 2018, pp. 336–340.

[326] Y. Zhu, R. Guo, P. Zhang, A. Bhandari, Frequency estimation via Sub-Nyquist unlimited sampling, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2024, pp. 9636–9640.
[327] A. Bhandari, F. Krahmer, HDR imaging from quantization noise, in: Proceedings of IEEE International Conference on Image Processing, 2020, pp. 101–105.
[328] M. Unser, Sampling - 50 years after Shannon, Proc. IEEE 88 (4) (2000) 569–587.
[329] A. Bhandari, F. Krahmer, On identifiability in unlimited sampling, in: Proceedings of International Conference on Sampling Theory and Applications, 2019, pp. 1–4.
[330] Z. Liu, A. Bhandari, B. Clerckx, λ-MIMO: Massive MIMO via modulo sampling, IEEE Trans. Commun. (2023) 6301–6315.
[331] L.G. Ordonez, P. Ferrand, M. Duarte, M. Guillaud, G. Yang, On full-duplex radios with modulo-ADCs, IEEE Open J. Commun. Soc. 2 (2021) 1279–1297.
[332] M. Beckmann, A. Bhandari, F. Krahmer, The modulo Radon transform: Theory, algorithms, and applications, SIAM J. Imaging Sci. 15 (2) (2022) 455–490.
[333] M. Beckmann, A. Bhandari, M. Iske, Fourier-domain inversion for the modulo Radon transform, IEEE Trans. Comput. Imaging 10 (2024) 653–665.
[334] T. Feuillen, B. Shankar, A. Bhandari, Unlimited sampling radar: Life below the quantization noise, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2023, pp. 1–5.
[335] S. Rudresh, A. Adiga, B.A. Shenoy, C.S. Seelamantula, Wavelet-based reconstruction for unlimited sampling, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2018, pp. 4584–4588.
[336] R. Guo, A. Bhandari, ITER-SIS: Robust unlimited sampling via iterative signal sieving, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2023, pp. 1–5.
[337] E. Azar, S. Mulleti, Y.C. Eldar, Residual recovery algorithm for modulo sampling, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2022, pp. 5722–5726.
[338] H. Inose, Y. Yasuda, A unity bit coding method by negative feedback, Proc. IEEE 51 (11) (1963) 1524–1535.
[339] I. Daubechies, R. DeVore, Approximating a bandlimited function using very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order, Ann. Math. 158 (2) (2003) 679–710.
[340] O. Graf, A. Bhandari, F. Krahmer, One-bit unlimited sampling, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 5102–5106.
[341] A. Lazar, L. Toth, Perfect recovery and sensitivity analysis of time encoded bandlimited signals, IEEE Trans. Circuits Syst. I. Regul. Pap. 51 (10) (2004) 2060–2073.
[342] D. Florescu, A. Bhandari, Time encoding via unlimited sampling: Theory, algorithms and hardware validation, IEEE Trans. Signal Process. (2022) 1–13.
[343] A. Eamaz, K.V. Mishra, F. Yeganegi, M. Soltanalian, UNO: Unlimited sampling meets one-bit quantization, IEEE Trans. Signal Process. 72 (2024) 997–1014.
[344] V. Shah, C. Hegde, Signal reconstruction from modulo observations, in: Proceedings of IEEE Global Conference on Signal and Information Processing, 2019, pp. 1–5.
[345] V. Shah, C. Hegde, Sparse signal recovery from modulo observations, EURASIP J. Adv. Signal Process. 2021 (1) (2021).
[346] O. Musa, P. Jung, N. Goertz, Generalized approximate message passing for unlimited sampling of sparse signals, in: Proceedings of IEEE Global Conference on Signal and Information Processing, 2018, pp. 336–340.
[347] O. Musa, P. Jung, G. Caire, On approximate message passing algorithms for unlimited sampling of sparse signals, in: Proceedings of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2023, pp. 131–135.
[348] Q. Zhang, J. Zhu, F. Qu, D.W. Soh, Line spectral estimation via unlimited sampling, IEEE Trans. Aerosp. Electron. Syst. 60 (5) (2024) 7214–7231.
[349] D. Prasanna, C. Sriram, C.R. Murthy, On the identifiability of sparse vectors from modulo compressed sensing measurements, IEEE Signal Process. Lett. 28 (2020) 131–134.
[350] G. Joseph, Noise-resilient unlimited sampling and recovery of sparse signals, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2025, pp. 1–5.
[351] R.J.P. De Figueiredo, C.-L. Hu, Waveform feature extraction based on Tauberian approximation, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-4 (2) (1982) 105–116.
[352] E.J. Candès, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Comm. Pure Appl. Math. 67 (6) (2013) 906–956.
[353] G. du Prony, Essai experimental et analytique sur les lois de la dilatabilite de fluides elastiques et sur celles da la force expansion de la vapeur de l'alcool, a differentes temperatures, J. de L' Éc. Polytech. 1 (22) (1795) 24–76.
[354] P. Stoica, R.L. Moses, Spectral Analysis of Signals, Pearson Prentice Hall, 2005.
[355] M.F. Duarte, Y.C. Eldar, Structured compressed sensing: From theory to applications, IEEE Trans. Signal Process. 59 (9) (2011) 4053–4085.
[356] K. Li, C.R. Rojas, T. Yang, H. Hjalmarsson, K.H. Johansson, S. Cong, Piecewise sparse signal recovery via piecewise orthogonal matching pursuit, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, pp. 4608–4612.
[357] P. Sprechmann, I. Ramirez, G. Sapiro, Y.C. Eldar, C-HiLasso: A collaborative hierarchical sparse modeling framework, IEEE Trans. Signal Process. 59 (9) (2011) 4183–4198.