Rate-Distortion Performance of Lossy Compressed Sensing of Sparse Sources

Markus Leinonen*, *Student Member, IEEE,* Marian Codreanu*, *Member, IEEE,* Markku Juntti*, *Senior Member, IEEE,* and Gerhard Kramer[†], *Fellow, IEEE*

Abstract—We investigate lossy compressed sensing (CS) of a hidden, or remote, source, where a sensor observes a sparse information source indirectly. The compressed noisy measurements are communicated to the decoder for signal reconstruction with the aim to minimize the mean square error distortion. An analytically tractable lower bound to the remote rate-distortion function (RDF), i.e., the conditional remote RDF, is derived by providing support side information to the encoder and decoder. For this setup, the best encoder separates into an estimation step and a transmission step. A variant of the Blahut-Arimoto algorithm is developed to numerically approximate the remote RDF. Furthermore, a novel entropy coding based quantized CS method is proposed. Numerical results illustrate the main rate-distortion characteristics of the lossy CS, and compare the performance of practical quantized CS methods against the proposed limits.

Index Terms—Remote source coding, conditional rate-distortion theory, side information, Blahut-Arimoto algorithm, wireless sensor networks.

I. INTRODUCTION

The proliferation of wireless sensor networks (WSNs) calls for energy-efficient communications in environmental, industrial, healthcare, military, and many other applications [1], [2]. A typical monitoring task consists of sensing and encoding at a resource-limited sensor, followed by a rate-limited information transmission to the decoder for signal reconstruction. Since wireless access is typically the main contributor to energy consumption [2], source compression [3]–[5] is crucial. Moreover, as most sensor signals are continuous-valued, quantization [6] is inevitable. Rate-distortion (RD) theory [7, Ch. 2] [8, Ch. 10] [9, Ch. 9] provides an information-theoretic tool for performance analysis and benchmarking of practical coding methods. In particular, the RD function (RDF) represents the best achievable compression performance for a given distortion fidelity.

The source signals of many WSN applications are *sparse*, e.g., environmental monitoring [10]–[12], source localization [13], and cognitive radio communications including spectrum sensing and direction of arrival estimation [14]. Sparsity can be utilized by compressed sensing (CS) [15]–[23] that enables accurate reconstruction of sparse (or compressible) signals of length-N from $M \leq N$ noisy linear measurements. The practical necessity of converting real-valued measurements into bits for data transmission/storage launched the framework of quantized CS (QCS) [24]–[27], referred to as *lossy CS* herein. Since the encoder of a CS based sensor accesses the source indirectly, the compression setup gives rise to *remote source coding*, also known as indirect/noisy coding [28] [7, Sect. 3.5, 4.5.4] [29]–[32], first introduced by Dobrushin and Tsybakov in [28]. Other works on remote compression in non-CS setups can be found in, e.g., [33]–[35].

The information-theoretic treatment of lossy CS is incomplete, i.e., *the remote RDF* is not known [24]. Using the replica method, the remote RDF of a lossy CS setup was derived in a general form under a large system limit (i.e., $M, N \to \infty$) in [36]; however, the remote RDF was not analytically solved. Most existing QCS analyses focus on the *operational* RD performance in systems where the encoder is restricted to be a quantizer operating on a single input symbol (either scalar or vector) at a time. Typically, these assume perfect knowledge of the sparse signal support, and study high rate quantization or/and large signal dimensions [37]–[41] [42, Ch. 3]. Scalar quantization (SQ) based QCS setups were studied in, e.g., [25], [43], [44]. Empirical performance of QCS algorithms that optimize either the encoder or decoder can be found in, e.g., [25]–[27], [38], [45], [46]; joint optimization of encoder-decoder pair(s) was studied in [41], [47]–[50]. RD bounds for directly compressing sparse sources were derived in [51]–[53], and the compression of (sparse) Bernoulli generalized Gaussian sources via uniform SQ was addressed in [54].

Deriving a RDF/remote RDF in closed form seems difficult in general. Remote RDFs have been derived only for few well-behaved source/observation distributions; see, e.g., the works in [28], [30], [7, Sect. 4.5.4], [33]–[35] assuming Gaussian distributions, and [55] considering a remote binary source. When a closed-form solution is unattainable, one may use numerical approximation. For example, the Blahut-Arimoto (BA) algorithm [56], [57] has been adapted to remote sensing scenarios under joint compression and classification in [58], and the CEO problem in [59]. Extensions were proposed to approximate the RDF of a continuous-amplitude source in [60], and the capacity of a continuous channel in [61].

Alternative techniques include a mapping approach akin to deterministic annealing in [62], and Lagrange duality based convex optimization in [63].

The encoder and/or decoder can possess various types and degrees of prior knowledge, i.e., *side information* (SI), on the signal of interest, e.g., support, magnitude variations, or occupied frequency bands. Added SI sometimes allows to derive RD lower bounds in closed form. In particular, compression with shared SI at the encoder and decoder follows conditional RD theory introduced by Gray [64], [65]. Compression with correlated, but not necessarily identical SI at the encoder and decoder was studied in, e.g., [66]–[68]. For different non-CS SI-aided compression variants, see, e.g., [7, Sect. 6.1] [69]–[71] [72, Sect. 11.1] [73, Sect. 5.6].

Contributions: The objective of this paper is to address the RD performance of CS used in compressing remote sources. We focus on sparse Gaussian vector sources and the mean square error (MSE) distortion. The closed-form solution to the remote RDF seems difficult to find. An analytically tractable lower bound to the remote RDF, termed *the conditional remote RDF*, is derived by providing support SI to the encoder and decoder. The best encoder of this setup separates into a rate independent minimum MSE estimation step, and a transmission step of the resulting estimate. A modified BA algorithm is developed to numerically approximate the remote RDF. Numerical results illustrate the main RD characteristics of the lossy CS, assess the tightness of the proposed lower bound, and exemplify the compression performance of various practical QCS methods against the proposed curves.

The main contributions of this paper are summarized as follows:

- Information-theoretic formulation of the lossy CS problem;
- Characterization of the remote RDF, defining the compression limit of the lossy CS;
- Derivation of the conditional remote RDF, i.e., an analytically tractable lower bound to the remote RDF;
- Development of a numerical approximation method for the remote RDF;
- Empirical validation of the proposed RD curves and comparisons to several practical QCS methods;
- Development of a novel entropy coding based QCS method which approaches the optimal performance given by the remote RDF.

The proposed lower bound, and the numerically evaluated remote RDF, shed light on general RD characteristics of lossy CS. We remark that the treatment of CS as an integral part of the coding system is general in the sense that no restrictive assumptions are made for the signal parameters, and the decoder is not limited to any standard CS signal reconstruction algorithm. Furthermore, the decoded estimates are not constrained to be sparse.

Related works: To the best of our knowledge, this is the first work that approaches lossy CS comprehensively from the information-theoretic remote source coding perspective. Initial results were published in [74]; herein we extend them by providing 1) more detailed and insightful derivations, 2) a numerical approximation method for the remote RDF, 3)

design of a novel entropy coding based QCS method, and 4) more extensive empirical performance results. As a difference to the QCS analyses in [37]–[39], [41], our coding system is not limited to a quantizer operating on a single input symbol at a time. Coluccia *et al.* [40] derived a distortion-rate lower bound under the support SI at the decoder, but assume 1) high-rate quantization, 2) a large system regime as $N \to \infty$, and 3) noiseless measurements; we consider noisy CS of a finite length vector source encompassing a composite structure, and the derived results hold for any rate. In a parallel work by Kipnis *et al.* [36], the derivation of the remote RDF is different to ours as they 1) assume a large system limit $M, N \to \infty$, 2) consider a Bernoulli-Gauss (scalar) source, and 3) use the replica method. While they provided curves via the BA algorithm, the requisite alphabet discretization was not addressed. Whereas most previous BA approaches, including [36], treat the discretization implicitly via "fine discretization", or via SQ, we propose an optimized vector quantizer (VQ[1]) based discretization method that takes the remote nature of the lossy CS into account via modified distortion measures, and that yields accurate approximations of the remote RDF.

Organization: The paper is organized as follows. Section II presents the system model and formulates the lossy CS problem. A lower bound to the remote RDF is derived in Section III. A numerical approximation method for computing the remote RDF is developed in Section IV. A novel QCS method is developed in Section V. Simulation results are provided in Section VI, and Section VII concludes the paper.

Notations: Italic capital letters denote random variables (X); boldface non-italic capital letters denote random vectors ($\mathbf{X}$); boldface non-italic small letters denote realizations of random vectors ($\mathbf{x}$); boldface italic small and capital letters denote deterministic vectors ($\boldsymbol{a}$) and matrices ($\boldsymbol{A}$), respectively; calligraphy letters denote sets/alphabets ($\mathcal{B}$). A block of m consecutive random vectors $\{\mathbf{X}_1, \ldots, \mathbf{X}_m\}$ is denoted as $\mathbf{X}^m \triangleq \{\mathbf{X}_n\}_{n=1}^m$. The following operators are used: $(\cdot)^{\mathrm{T}}$ denotes the matrix transpose; $\odot$ denotes the Hadamard product, i.e., ith entry of vector $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$ is $C_i = A_i B_i$; $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix; $\mathrm{rank}(\cdot)$ denotes matrix rank; $\max\{\cdot\}$ denotes the maximum element; $\mathrm{supp}(\mathbf{a})$ denotes the support of vector $\mathbf{a}$; $|\mathcal{B}|$ denotes the cardinality of set $\mathcal{B}$; $\|\cdot\|_0$ counts the number of non-zero entries of a vector. The set of symmetric positive semi-definite (definite) $N \times N$-matrices are denoted as $\mathbb{S}_+^N$ ($\mathbb{S}_{++}^N$). A diagonal matrix is denoted as $\mathrm{diag}(a_1, \ldots, a_N)$, an identity matrix as $\boldsymbol{I}$, and a matrix with all entries zeros as $\mathbf{0}$. The ℓ_2-norm is denoted as $\|\cdot\|_2$. The logarithms $\log(\cdot)$ are of base 2.

II. LOSSY CS VIA REMOTE SOURCE CODING

We investigate the RD performance of the model depicted in Fig. 1, where the information source is observed via noisy compressed measurements, encoded with a lossy source code, and communicated to the decoder for signal reconstruction. The transmissions from encoder E to decoder D are assumed to be error-free. The compression task is classified as *remote*

---

[1] The acronym "VQ" will be interchangeably used for "vector quantization" and "vector quantizer".
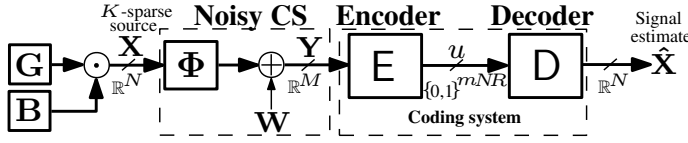
Fig. 1. Lossy CS of a sparse source.

source coding because the encoder accesses the source only through noisy measurements. The source and the CS measurement model are defined next, followed by the formal statement of the problem.

A. Source Model

Let $\{\mathbf{X}_n\}_{n=1}^{\infty}$ be a discrete-time memoryless vector source sequence[2]. Each vector $\mathbf{X}_n = [X_{n,1} \cdots X_{n,N}]^{\mathrm{T}}$ is K-sparse[3], $K \leq N$, i.e., it takes on values in the continuous source alphabet $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_0 = K\}$. The set $\mathcal{X}$ thus consists of the union of $\binom{N}{K}$ subspaces, i.e., the signal model is nonlinear [75], [76]. We further assume that the source sequence is generated from the memoryless sequence of tuples $\{(\mathbf{G}_n, \mathbf{B}_n)\}_{n=1}^{\infty}$ such that $\mathbf{X}_n = \mathbf{G}_n \odot \mathbf{B}_n$, where $\odot$ denotes the Hadamard product; $\mathbf{G}_n$ is a length-N zero mean Gaussian random vector $\mathbf{G}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{G}})$ with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{G}} \in \mathbb{S}_{++}^N$; $\mathbf{B}_n$ is a length-N binary support random vector, independent of $\mathbf{G}_n$, with the discrete alphabet $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_{|\mathcal{B}|}\}$, where $|\mathcal{B}| = \binom{N}{K}$ is the number of all possible *sparsity patterns*. Each $\mathbf{b}_s = [b_{s,1} \cdots b_{s,N}]^{\mathrm{T}} \in \mathcal{B}$ is unique, contains K ones and $N - K$ zeros, and is associated with the *a priori* probability $p(\mathbf{b}_s) \triangleq \Pr(\mathbf{B} = \mathbf{b}_s)$ with $p(\mathbf{b}_s) \in [0, 1]$ and $\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) = 1$.

B. Noisy CS

Let $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ be a fixed and known CS measurement matrix, $K \leq M \leq N$. The sensor (i.e., the encoder) observes $\{\mathbf{X}_n\}_{n=1}^{\infty}$ indirectly [15], [16], [18], [23] as

$$\mathbf{Y}_n = \boldsymbol{\Phi}\mathbf{X}_n + \mathbf{W}_n, \ n = 1, 2, \ldots \quad (1)$$

where $\mathbf{W}_n$, $n = 1, 2, \ldots$ are length-M i.i.d. Gaussian random measurement noise vectors independent of $\{\mathbf{X}_n\}_{n=1}^{\infty}$, and each $\mathbf{Y}_n$ is a length-M measurement random vector that takes values in the measurement vector space $\mathcal{Y}$. We use $\mathbf{W}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{W}})$ with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{W}} \in \mathbb{S}_{++}^M$.

CS signal recovery performance is strongly affected by the structure of $\boldsymbol{\Phi}$. The restricted isometry property (RIP) and the coherence of $\boldsymbol{\Phi}$ establish guarantees on stable and accurate CS signal reconstruction, and they are often used to assess the quality of $\boldsymbol{\Phi}$ [77, Ch. 5,6]. For instance, $\boldsymbol{\Phi}$ with i.i.d. Gaussian entries satisfies the RIP with high probability if $M \geq CK\log(N/K)$, where C is a positive constant [77, Ch. 1]. Nonetheless, we make no restricting assumptions on $\boldsymbol{\Phi}$.

---

[2]Due to the independence over time, the time index n will often be suppressed for brevity whenever not explicitly needed.

[3]With slight abuse of terminology, a K-sparse signal contains *exactly* (instead of *at most*) K non-zero elements.

C. Lossy CS Problem

Let $\mathbf{X}^m \triangleq \{\mathbf{X}_n\}_{n=1}^m$ and $\mathbf{x}^m \triangleq \{\mathbf{x}_n\}_{n=1}^m$ denote the blocks of m consecutive source random vectors and the corresponding realizations, respectively. Let $\mathcal{X}^m$ denote the m-fold Cartesian product of $\mathcal{X}$. Analogous notations are used for the other vectors. Let $\hat{\mathbf{X}}$ be the reproduction random vector at the decoder output, taking values in the reproduction alphabet $\hat{\mathcal{X}}$. Finally, define the average per-letter mean square error (MSE) distortion between vectors $\mathbf{x} = [x_1 \cdots x_N]^{\mathrm{T}} \in \mathcal{X}$ and $\hat{\mathbf{x}} = [\hat{x}_1 \cdots \hat{x}_N]^{\mathrm{T}} \in \hat{\mathcal{X}}$ as

$$d(\mathbf{x}, \hat{\mathbf{x}}) \triangleq N^{-1} \sum_{k=1}^N (x_k - \hat{x}_k)^2 \quad (2)$$

and the average per-letter MSE distortion blocks $\mathbf{x}^m \in \mathcal{X}^m$ and $\hat{\mathbf{x}}^m \in \hat{\mathcal{X}}^m$ as

$$d(\mathbf{x}^m, \hat{\mathbf{x}}^m) \triangleq (mN)^{-1} \sum_{n=1}^m \sum_{k=1}^N (x_{n,k} - \hat{x}_{n,k})^2. \quad (3)$$

The lossy source coding system in Fig. 1 operates as follows [78, Sect. 2.1] [8, Sect. 10.2] [72, Sect. 3.5, 3.6]. The encoder E observes a block of CS measurements $\mathbf{y}^m \in \mathcal{Y}^m$, and compresses it to a message represented by an index $u \in \mathcal{U}$ of rate mNR bits using an encoder mapping

$$g_{\mathsf{E}}^m : \mathcal{Y}^m \to \mathcal{U} \triangleq \{1, \ldots, 2^{mNR}\} \quad (4)$$

where the rate R is defined as the bits/entry of $\mathbf{X}$. The decoder D uses the index to reconstruct an estimate of $\mathbf{x}^m \in \mathcal{X}^m$ via a decoder mapping

$$g_{\mathsf{D}}^m : \mathcal{U} \to \hat{\mathcal{X}}^m. \quad (5)$$

A pair (R, D) for distortion $D \geq 0$ is achievable if there exists a sequence of $(2^{mNR}, m)$-RD codes with mappings g_{E}^m and g_{D}^m such that $\lim_{m \to \infty} \mathbb{E}\big[d(\mathbf{X}^m, g_{\mathsf{D}}^m\{g_{\mathsf{E}}^m(\mathbf{Y}^m)\})\big] \leq D$. Let $\mathcal{R}$ be the closure of the set of achievable (R, D) pairs.

Definition 1. (*Lossy CS source coding problem*) Amongst all E-D pairs of mappings (4) and (5), determine the infimum of (achievable) rates R such that $\mathbf{X}$ can be reproduced with the average distortion satisfying $\mathbb{E}\big[d(\mathbf{X}, \hat{\mathbf{X}})\big] \leq D$, i.e., define [8, Sect. 10.2]

$$R_{\mathbf{X}}^{\mathrm{rem}}(D) = \inf_{(R,D) \in \mathcal{R}} R. \quad (6)$$

We call $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ *the remote RDF* of source $\mathbf{X}$.

We remark that the used MSE distortion implies, in general, non-sparse reconstruction, which might be undesirable in certain applications. Next, we derive the mathematical expression for $R_{\mathbf{X}}^{\mathrm{rem}}(D)$.

D. Remote RDF

The general expression of the remote RDF for a discrete memoryless source with discrete memoryless observations has been derived in [7, Eqs. (3.5.1) – (3.5.5)]. Adapting the result to continuous-valued signals $\mathbf{X}$ and $\mathbf{Y}$, $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ in (6) can be expressed as

$$R_{\mathbf{X}}^{\mathrm{rem}}(D) = \min_{f(\hat{\mathbf{x}}|\mathbf{y}) : \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \leq D} \frac{1}{N} I(\mathbf{Y}; \hat{\mathbf{X}}) \quad (7a)$$

where the optimization is over the conditional probability density function (PDF) $f(\hat{\mathbf{x}}|\mathbf{y})$, commonly referred to as *the*

test channel, and $d(\mathbf{x}, \hat{\mathbf{x}})$ is the distortion in (2). The mutual information between $\mathbf{Y}$ and $\hat{\mathbf{X}}$ is

$$I(\mathbf{Y}; \hat{\mathbf{X}}) = \int_{\mathbf{y}} \int_{\hat{\mathbf{x}}} f(\mathbf{y}) f(\hat{\mathbf{x}}|\mathbf{y}) \log \frac{f(\hat{\mathbf{x}}|\mathbf{y})}{f(\hat{\mathbf{x}})} d\mathbf{y} d\hat{\mathbf{x}} \qquad (7b)$$

and the average MSE distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ is

$$\mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \stackrel{(a)}{=} \int_{\mathbf{x}} \int_{\mathbf{y}} \int_{\hat{\mathbf{x}}} f(\mathbf{x}) f(\mathbf{y}|\mathbf{x}) f(\hat{\mathbf{x}}|\mathbf{y}) d(\mathbf{x}, \hat{\mathbf{x}}) d\mathbf{x} d\mathbf{y} d\hat{\mathbf{x}} \qquad (7c)$$

where (a) follows from $f(\hat{\mathbf{x}}|\mathbf{y}) = f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{x})$ because $\mathbf{X} \to \mathbf{Y} \to \hat{\mathbf{X}}$ forms a Markov chain. Note that the remote sensing mechanism is captured by the conditional PDF $f(\mathbf{y}|\mathbf{x})$, governed by the CS measurements in (1).

Due to the time-varying sparsity of $\{\mathbf{X}_n\}_{n=1}^{\infty}$ through $\{\mathbf{B}_n\}_{n=1}^{\infty}$, the PDFs of $\mathbf{X}$, and consequently, of $\mathbf{Y}$ are mixture distributions, which seems to make the optimization over $f(\hat{\mathbf{x}}|\mathbf{y})$ in (7a) difficult. Hence, we confine our treatment of the lossy CS problem of Definition 1 to the following two approaches. In Section III, we derive an analytically tractable *lower bound* to $R_{\mathbf{X}}^{\mathrm{rem}}(D)$, whereas in Section IV we develop a method to numerically approximate $R_{\mathbf{X}}^{\mathrm{rem}}(D)$. Note that the difficulty resides also in the *direct* compression of $\mathbf{X}$ for which only RD bounds have been derived [51]–[53].

III. RATE-DISTORTION LOWER BOUND FOR LOSSY CS

We derive a lower bound to $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ in (7a) by considering the compression setup of Fig. 2, where the encoder E_{si} and decoder D_{si} possess side information (SI) on sequence $\{\mathbf{B}_n\}_{n=1}^{\infty}$. Having support SI at the decoder is often optimistic in practice, but sometimes the encoder may acquire SI on $\mathbf{B}$ (i.e., an estimate $\hat{\mathbf{B}}$) from the measurements $\mathbf{Y}$ at a moderate cost via a sparse signal reconstruction algorithm (see, e.g., the algorithm listing in [79, Sect. 1.6], and analysis in [80]). Nevertheless, the shared support SI lets us derive an analytically tractable lower bound to $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ which sheds light on the RD behavior of the original setup in Fig. 1, and establishes a benchmark for practical coding methods. The associated RD problem is formulated below.

A. Lossy CS Problem with Support SI

Owing to the support SI, an *informed* lossy source code is defined as follows [64], [66], [69] [70, Sect. 2.3.1]. The encoder E_{si} observes a block of CS measurements $\mathbf{y}^m \in \mathcal{Y}^m$ along with the SI $\mathbf{b}^m \in \mathcal{B}^m$, and compresses it to a message index $u \in \mathcal{U}$ using an encoder mapping

$$g_{\mathsf{E}_{\mathrm{si}}}^m : \mathcal{Y}^m \times \mathcal{B}^m \to \mathcal{U}. \qquad (8)$$

The decoder D_{si} uses the index, and the common SI $\mathbf{b}^m$ to reconstruct an estimate of $\mathbf{x}^m \in \mathcal{X}^m$ via a decoder mapping

$$g_{\mathsf{D}_{\mathrm{si}}}^m : \mathcal{U} \times \mathcal{B}^m \to \hat{\mathcal{X}}^m. \qquad (9)$$

A pair (R, D) for distortion $D \geq 0$ is achievable if there exists a sequence of informed $(2^{mNR}, m)$-rate-distortion codes with mappings $g_{\mathsf{E}_{\mathrm{si}}}^m$ and $g_{\mathsf{D}_{\mathrm{si}}}^m$ such that $\lim_{m \to \infty} \mathbb{E}[d(\mathbf{X}^m, g_{\mathsf{D}_{\mathrm{si}}}^m \{g_{\mathsf{E}_{\mathrm{si}}}^m(\mathbf{Y}^m, \mathbf{B}^m), \mathbf{B}^m\})] \leq D$. Let $\mathcal{R}_{\mathrm{si}}$ be the closure of the set of such achievable (R, D) pairs.
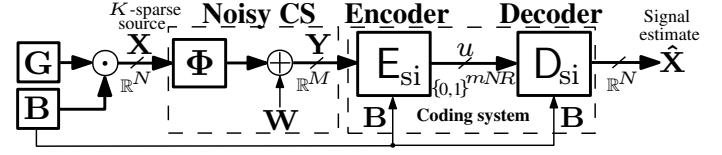

Fig. 2. Lossy CS of a sparse source with support SI.

Definition 2. (*Lossy CS source coding problem with support SI*) Amongst all E_{si}-D_{si} pairs of mappings (8) and (9), determine the infimum of (achievable) rates R such that $\mathbf{X}$ can be reproduced with the average distortion satisfying $\mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \leq D$, i.e., define

$$R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D) = \inf_{(R, D) \in \mathcal{R}_{\mathrm{si}}} R. \qquad (10)$$

We call $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ *the conditional remote RDF* of source $\mathbf{X}$. Clearly, $\mathcal{R} \subseteq \mathcal{R}_{\mathrm{si}}$, and $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ establishes a lower bound to the best possible compression performance of the lossy CS as

$$R_{\mathbf{X}}^{\mathrm{rem}}(D) \geq R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D). \qquad (11)$$

The remainder of the section is devoted to deriving $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$.

B. Conditional Remote RDF

The conditional RDF for a discrete source along with the respective coding theorems is given in [64]. Extending the results to a *remote* compression setup, the conditional remote RDF $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ can be expressed as

$$R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D) = \min_{\{f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{b}_s)\}_{s=1}^{|\mathcal{B}|} : \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] \leq D} \frac{1}{N} I(\mathbf{Y}; \hat{\mathbf{X}}|\mathbf{B}) \quad (12a)$$

where the optimization is over the $|\mathcal{B}|$ different test channels $f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{b}_s)$, $s = 1, \ldots, |\mathcal{B}|$, the conditional mutual information between $\mathbf{Y}$ and $\hat{\mathbf{X}}$ given $\mathbf{B}$ is

$$I(\mathbf{Y}; \hat{\mathbf{X}}|\mathbf{B}) = \sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) I(\mathbf{Y}; \hat{\mathbf{X}}|\mathbf{B} = \mathbf{b}_s) \qquad (12b)$$

and the average MSE distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$ is

$$\mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] = \sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})|\mathbf{B} = \mathbf{b}_s] \qquad (12c)$$

where, as compared to (7c), the expectation is also taken over $\mathbf{B}$.

Observe that (12b) and (12c) decompose with respect to realizations $\mathbf{B} = \mathbf{b}_s$, $s = 1, \ldots, |\mathcal{B}|$. $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ in (12a) can thus be expressed as the weighted sum minimization [64, Theorem 5]

$$R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D) = \min_{\substack{\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) D_s = D \\ D_s \geq 0,\, s=1,\ldots,|\mathcal{B}|}} \sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s) \quad (13)$$

with optimization variables D_s, $s = 1, \ldots, |\mathcal{B}|$, where $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$ is *the conditional marginal remote RDF* of source $\mathbf{X}$ for a fixed realization $\mathbf{B} = \mathbf{b}_s$ and distortion $D_s \geq 0$, i.e., we have

$$R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s) = \min_{f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{b}_s) : \mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})|\mathbf{B}=\mathbf{b}_s] \leq D_s} \frac{1}{N} I(\mathbf{Y}; \hat{\mathbf{X}}|\mathbf{B} = \mathbf{b}_s)$$
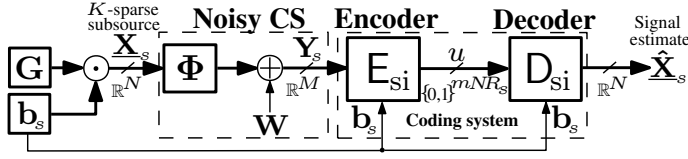$$(14a)$$

Fig. 3. Lossy CS of a sparse *subsource* with support SI.

where the mutual information between $\mathbf{Y}$ and $\hat{\mathbf{X}}$, conditioned on $\mathbf{B} = \mathbf{b}_s$, is

$$I(\mathbf{Y}; \hat{\mathbf{X}}|\mathbf{B} = \mathbf{b}_s) = \int_{\mathbf{y}} \int_{\hat{\mathbf{x}}} f(\mathbf{y}|\mathbf{b}_s) f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{b}_s) \log \frac{f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{b}_s)}{f(\hat{\mathbf{x}}|\mathbf{b}_s)} \mathrm{d}\mathbf{y} \mathrm{d}\hat{\mathbf{x}} \quad (14b)$$

and the average MSE distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}$, conditioned on $\mathbf{B} = \mathbf{b}_s$, is

$$\mathbb{E}\big[d(\mathbf{X}, \hat{\mathbf{X}})|\mathbf{B} = \mathbf{b}_s\big] \overset{(a)}{=} \int_{\mathbf{x}} \int_{\mathbf{y}} \int_{\hat{\mathbf{x}}} f(\mathbf{x}|\mathbf{b}_s) f(\mathbf{y}|\mathbf{x}, \mathbf{b}_s) f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{b}_s) d(\mathbf{x}, \hat{\mathbf{x}}) \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{y} \mathrm{d}\hat{\mathbf{x}} \quad (14c)$$

where in (a) we used $f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{x}, \mathbf{b}_s) = f(\hat{\mathbf{x}}|\mathbf{y}, \mathbf{b}_s)$ because $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow \hat{\mathbf{X}}$ forms a Markov chain when conditioned on $\mathbf{B}$. Owing to the support SI, all PDFs above are equivalent to those in (7), except conditioned on a realization $\mathbf{B} = \mathbf{b}_s$.

Regarding the above formulations, the characterization of $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ in (13) boils down to deriving each $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$, $s = 1, \ldots, |\mathcal{B}|$, in (14a). This is carried out in the next section.

C. Conditional Marginal Remote RDF

Fundamentally, the conditional marginal remote RDF $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$, $s = 1, \ldots, |\mathcal{B}|$, in (14a) determines the minimum (achievable) rate R_s such that $\mathbf{X}$ can be reproduced with the average distortion satisfying $\mathbb{E}\big[d(\mathbf{X}, \hat{\mathbf{X}})|\mathbf{B} = \mathbf{b}_s\big] \leq D_s$ in the setup depicted in Fig. 3, where $\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s) R_s = R$. As preliminaries for deriving $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$, we introduce three definitions.

Definition 3. (*Subsource*) Let $\{\underline{\mathbf{X}}_{s,n}\}_{n=1}^{\infty} = \{\mathbf{G}_n \odot \mathbf{b}_s\}_{n=1}^{\infty}$ be the memoryless sequence of the sth *subsource*, consisting of K-sparse source vectors $\{\mathbf{X}_n\}_{n=1}^{\infty}$ restricted to a fixed realization $\mathbf{B} = \mathbf{b}_s$, $s = 1, \ldots, |\mathcal{B}|$. Each subsource $\underline{\mathbf{X}}_s$ comprises of two parts:

$$\underline{\mathbf{X}}_s : \begin{cases} \mathbf{X}_s \triangleq \mathbf{G}_{\mathrm{supp}(\mathbf{b}_s)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{X}_s}) \\ \mathbf{0}_{N-K} \end{cases} \quad (15)$$

where 1) $\mathbf{X}_s$ is a length-K random vector that extracts the entries of $\underline{\mathbf{X}}_s$ (i.e., the entries of $\mathbf{G}$) restricted to the support of $\mathbf{b}_s$, where $\mathrm{supp}(\mathbf{b}_s) \triangleq \{k \in \{1, \ldots, N\} | b_{s,k} \neq 0\}$ denotes the support of vector $\mathbf{b}_s$, $\mathbf{G}_{\mathrm{supp}(\mathbf{b}_s)}$ extracts the entries G_k from $\mathbf{G}$ for indices $k \in \mathrm{supp}(\mathbf{b}_s)$, and the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}_s} \in \mathbb{S}_{++}^K$ extracts the entries $\Sigma_{\mathbf{G}}(k, k')$ from $\boldsymbol{\Sigma}_{\mathbf{G}}$ for indices $k, k' \in \mathrm{supp}(\mathbf{b}_s)$; 2) $\mathbf{0}_{N-K}$ is the all-zero vector corresponding to the entries of $\underline{\mathbf{X}}_s$ for indices $k \in \mathrm{supp}(\mathbf{b}_s)^{\mathrm{c}}$, where $\mathrm{supp}(\mathbf{b}_s)^{\mathrm{c}} \triangleq \{\{1, \ldots, N\} \setminus \mathrm{supp}(\mathbf{b}_s)\}$ is the complement of $\mathrm{supp}(\mathbf{b}_s)$.

We emphasize that although the subsources $\underline{\mathbf{X}}_s$, $s = 1, \ldots, |\mathcal{B}|$, are virtual, i.e., not actually present in the system, they have an instructive role in our derivations. In light of decomposability, the subsources bear a resemblance with composite sources[4] [7, Sect. 6.1.1] [70].

Definition 4. (*CS measurements of a subsource*) Let $\{\mathbf{Y}_{s,n}\}_{n=1}^{\infty}$ be the memoryless sequence of the CS measurements of form (1) restricted to a fixed realization $\mathbf{B} = \mathbf{b}_s$, i.e., the CS measurements of subsource $\underline{\mathbf{X}}_s$, $s = 1, \ldots, |\mathcal{B}|$, in (15), defined as

$$\mathbf{Y}_s \triangleq \boldsymbol{\Phi}\underline{\mathbf{X}}_s + \mathbf{W} = \boldsymbol{\Phi}_s \mathbf{X}_s + \mathbf{W} \quad (16)$$

where matrix $\boldsymbol{\Phi}_s \in \mathbb{R}^{M \times K}$ extracts the K columns of $\boldsymbol{\Phi}$ with indices $k \in \mathrm{supp}(\mathbf{b}_s)$; consequently, we have $\mathbf{Y}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Y}_s})$ with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}_s} = \boldsymbol{\Phi}_s \boldsymbol{\Sigma}_{\mathbf{X}_s} \boldsymbol{\Phi}_s^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathbf{W}} \in \mathbb{S}_{++}^M$.

Definition 5. (*MMSE estimator of a subsource*) Let $\underline{\mathbf{Z}}_s$ be a length-N random vector representing the minimum mean square error (MMSE) estimator of source $\mathbf{X}$ given $\mathbf{Y}$ for a fixed realization $\mathbf{B} = \mathbf{b}_s$, i.e., the MMSE estimator of subsource $\underline{\mathbf{X}}_s$ in (15) given $\mathbf{Y}_s$ in (16). Each $\underline{\mathbf{Z}}_s$ is given by the conditional expectation as [81, Sect. 8.2]

$$\underline{\mathbf{Z}}_s \triangleq \mathbb{E}[\mathbf{X}|\mathbf{Y}, \mathbf{B} = \mathbf{b}_s], \ s = 1, \ldots, |\mathcal{B}|, \quad (17)$$

which, owing to the sparsity of $\underline{\mathbf{X}}_s$ (cf. (15)), decomposes into two parts:

$$\underline{\mathbf{Z}}_s : \begin{cases} \mathbf{Z}_s \triangleq \mathbb{E}[\mathbf{X}_s|\mathbf{Y}, \mathbf{B} = \mathbf{b}_s] \\ \quad = \boldsymbol{\Sigma}_{\mathbf{X}_s \mathbf{Y}_s} \boldsymbol{\Sigma}_{\mathbf{Y}_s}^{-1} \mathbf{Y}_s = \boldsymbol{F}_s \mathbf{Y}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Z}_s}) \\ \mathbf{0}_{N-K} \end{cases} \quad (18)$$

where 1) $\mathbf{Z}_s$ is the length-K random vector that represents the MMSE estimator of $\mathbf{X}_s$ given $\mathbf{Y}$ and $\mathbf{B} = \mathbf{b}_s$, and 2) $\mathbf{0}_{N-K}$ corresponds to the MMSE estimator of the zero part of $\underline{\mathbf{X}}_s$; for jointly Gaussian random vectors, $\mathbf{Z}_s$ is linear [82, Sect. 10.2], where the cross-covariance matrix is $\boldsymbol{\Sigma}_{\mathbf{X}_s \mathbf{Y}_s} = \boldsymbol{\Sigma}_{\mathbf{X}_s} \boldsymbol{\Phi}_s^{\mathrm{T}} \in \mathbb{R}^{K \times M}$, $\boldsymbol{F}_s \triangleq \boldsymbol{\Sigma}_{\mathbf{X}_s \mathbf{Y}_s} \boldsymbol{\Sigma}_{\mathbf{Y}_s}^{-1} \in \mathbb{R}^{K \times M}$, and $\mathbf{Z}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Z}_s})$ with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}_s} = \boldsymbol{F}_s \boldsymbol{\Sigma}_{\mathbf{X}_s \mathbf{Y}_s}^{\mathrm{T}} \in \mathbb{S}_+^K$.

$R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$ in (14a) can be characterized by a two-stage encoding structure: the encoder first optimally estimates the subsource $\underline{\mathbf{X}}_s$ (see (15)) from measurements $\mathbf{Y}_s$ (see (16)), and then optimally encodes the constructed estimator $\underline{\mathbf{Z}}_s$ in (18). This is elaborated next.

1) MMSE Distortion Separation: Let $\hat{\underline{\mathbf{X}}}_s$ be a length-N random vector representing the reproduction of subsource $\underline{\mathbf{X}}_s$ at the decoder output (see Fig. 3). Using the MMSE orthogonality principle [81, Ch. 8.2.1], the average conditional MSE distortion $\mathbb{E}\big[d(\underline{\mathbf{X}}_s, \hat{\underline{\mathbf{X}}}_s)\big] \triangleq \mathbb{E}\big[d(\mathbf{X}, \hat{\mathbf{X}})|\mathbf{B} = \mathbf{b}_s\big]$ in (14c) separates as

$$\mathbb{E}\big[d(\underline{\mathbf{X}}_s, \hat{\underline{\mathbf{X}}}_s)\big] = D_{\mathbf{Z}|\mathbf{b}_s} + \mathbb{E}\big[d(\underline{\mathbf{Z}}_s, \hat{\underline{\mathbf{X}}}_s)\big] \quad (19)$$

---

[4]The sequence pair $\{(\mathbf{X}_n, \mathbf{B}_n)\}_{n=1}^{\infty}$ forms a jointly stationary and ergodic regenerative composite source with stationary memoryless subsource processes; $\{\mathbf{B}_n\}_{n=1}^{\infty}$ is the hidden switch sequence that controls the output process $\{\mathbf{X}_n\}_{n=1}^{\infty}$ by randomly activating the subsources $\{\underline{\mathbf{X}}_{s,n}\}_{n=1}^{\infty}$ according to probabilities $p(\mathbf{b}_s)$, $s = 1, \ldots, |\mathcal{B}|$ [7, Sect. 6.1.1] [70].

where we denoted the (rate dependent) average MSE distortion between $\underline{\mathbf{Z}}_s$ and $\hat{\underline{\mathbf{X}}}_s$ as $\mathbb{E}\big[d(\underline{\mathbf{Z}}_s, \hat{\underline{\mathbf{X}}}_s)\big] = N^{-1}\mathbb{E}\big[\|\underline{\mathbf{Z}}_s - \hat{\underline{\mathbf{X}}}_s\|_2^2\big]$, and defined the (rate independent) average MMSE estimation error with respect to subsource $\underline{\mathbf{X}}_s$ as [82, Sect. 10.2]

$$D_{\mathbf{Z}|\mathbf{b}_s} = N^{-1}\mathrm{Tr}\big(\boldsymbol{\Sigma}_{\mathbf{X}_s} - \boldsymbol{\Sigma}_{\mathbf{Z}_s}\big). \qquad (20)$$

Separation similar to (19) appears also in, e.g., [28]–[30].

2) Reduced Distortion: By the decomposability of (18), the last term in (19) splits as

$$\mathbb{E}\big[d(\underline{\mathbf{Z}}_s, \hat{\underline{\mathbf{X}}}_s)\big] = \mathbb{E}\big[d(\mathbf{Z}_s, \hat{\mathbf{X}}_s)\big] + \mathbb{E}\big[d(\mathbf{0}_{N-K}, \hat{\underline{\mathbf{X}}}_{\mathrm{supp}(\mathbf{b}_s)^c})\big] \qquad (21)$$

where $\hat{\mathbf{X}}_s$ is the length-K reproduction random vector associated with $\mathbf{X}_s$, and $\hat{\underline{\mathbf{X}}}_{\mathrm{supp}(\mathbf{b}_s)^c}$ is the reproduction random vector associated with the zero part of $\underline{\mathbf{X}}_s$. Since a RDF is a monotonic nonincreasing function of the distortion [7, Sect. 2], it is optimal for $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$ to set $\hat{\underline{\mathbf{X}}}_{\mathrm{supp}(\mathbf{b}_s)^c} = \mathbf{0}_{N-K}$, and thus the distortion in (19) reduces to

$$\mathbb{E}\big[d(\underline{\mathbf{X}}_s, \hat{\underline{\mathbf{X}}}_s)\big] = D_{\mathbf{Z}|\mathbf{b}_s} + \mathbb{E}\big[d(\mathbf{Z}_s, \hat{\mathbf{X}}_s)\big]. \qquad (22)$$

Let $D_s' \geq 0$ be a *reduced* distortion criterion for sth subsource as

$$D_s' \triangleq D_s - D_{\mathbf{Z}|\mathbf{b}_s} \geq 0, \ s = 1, \dots, |\mathcal{B}|, \qquad (23)$$

where $D_s \geq 0$ is the distortion criterion in (14a), and $D_{\mathbf{Z}|\mathbf{b}_s}$ is given in (20). Note that by (22), $\mathbb{E}\big[d(\mathbf{Z}_s, \hat{\mathbf{X}}_s)\big] \leq D_s'$ implies $\mathbb{E}\big[d(\underline{\mathbf{X}}_s, \hat{\underline{\mathbf{X}}}_s)\big] \leq D_s$.

3) Estimate-and-Compress Separation: Let $R_{\mathbf{Z}|\mathbf{b}_s}^{\mathrm{dir}}(D_s')$ denote the *direct* RDF of the MMSE estimator $\mathbf{Z}_s$ defined in (18) for reduced distortion D_s' in (23), i.e., define

$$R_{\mathbf{Z}|\mathbf{b}_s}^{\mathrm{dir}}(D_s') = \min_{f(\hat{\mathbf{z}}_s|\mathbf{z}_s):\mathbb{E}[d(\mathbf{Z}_s,\hat{\mathbf{Z}}_s)] \leq D_s'} \frac{1}{N}I(\mathbf{Z}_s; \hat{\mathbf{Z}}_s) \qquad (24a)$$

where the minimization is over the test channel $f(\hat{\mathbf{z}}_s|\mathbf{z}_s)$, $\hat{\mathbf{Z}}_s$ is a length-K reproduction random vector for $\mathbf{Z}_s$, the average mutual information between $\mathbf{Z}_s$ and $\hat{\mathbf{Z}}_s$ is

$$I(\mathbf{Z}_s; \hat{\mathbf{Z}}_s) = \int_{\mathbf{z}_s}\int_{\hat{\mathbf{z}}_s} f(\mathbf{z}_s)f(\hat{\mathbf{z}}_s|\mathbf{z}_s)\log\frac{f(\hat{\mathbf{z}}_s|\mathbf{z}_s)}{f(\hat{\mathbf{z}}_s)}\mathrm{d}\mathbf{z}_s\mathrm{d}\hat{\mathbf{z}}_s \qquad (24b)$$

and the average MSE distortion between $\mathbf{Z}_s$ and $\hat{\mathbf{Z}}_s$ is

$$\mathbb{E}\big[d(\mathbf{Z}_s, \hat{\mathbf{Z}}_s)\big] = \int_{\mathbf{z}_s}\int_{\hat{\mathbf{z}}_s} f(\mathbf{z}_s)f(\hat{\mathbf{z}}_s|\mathbf{z}_s)d(\mathbf{z}_s, \hat{\mathbf{z}}_s)\mathrm{d}\mathbf{z}_s\mathrm{d}\hat{\mathbf{z}}_s. \qquad (24c)$$

The RDF $R_{\mathbf{Z}|\mathbf{b}_s}^{\mathrm{dir}}(D_s')$ can be derived by decorrelating the Gaussian (effective) source $\mathbf{Z}_s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Z}_s})$ via the Karhunen-Loève transform, and applying reverse water-filling [8, Sect. 10.3.3]. Accordingly, let $\boldsymbol{\Sigma}_{\mathbf{Z}_s} = \boldsymbol{Q}_s\boldsymbol{\Lambda}_s\boldsymbol{Q}_s^{\mathrm{T}}$, where $\boldsymbol{\Lambda}_s \triangleq \mathrm{diag}(\lambda_{s,1}, \dots, \lambda_{s,K})$ contains the eigenvalues $\lambda_{s,1} \geq \dots \geq \lambda_{s,K} \geq 0$ of $\boldsymbol{\Sigma}_{\mathbf{Z}_s} \in \mathbb{S}_+^K$, and the columns of $\boldsymbol{Q}_s \in \mathbb{R}^{K\times K}$ are the corresponding eigenvectors. Consequently, we have

$$R_{\mathbf{Z}|\mathbf{b}_s}^{\mathrm{dir}}(D_s') = \min_{\substack{\sum_{k=1}^K D_{s,k}' = D_s' \\ D_{s,k}' \geq 0,\, k=1,\dots,K}} \frac{1}{N}\sum_{k=1}^K \max\left\{0, \frac{1}{2}\log\frac{\lambda_{s,k}}{D_{s,k}'}\right\} \qquad (25)$$

where $D_{s,k}'$, $k = 1, \dots, K$, are the optimization variables.

The following proposition, which follows from the proofs in [28], [30], gives an expression for the conditional marginal remote RDF $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$.

Proposition 1. *The conditional marginal remote RDF of* $\underline{\mathbf{X}}_s$ *in* (14a) *is given as the (direct) RDF of the MMSE estimator* $\mathbf{Z}_s$ *in* (25)*, i.e., we have*

$$R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s) = R_{\mathbf{Z}|\mathbf{b}_s}^{\mathrm{dir}}(D_s'), \ s = 1, \dots, |\mathcal{B}|, \qquad (26)$$

where $D_s' = D_s - D_{\mathbf{Z}|\mathbf{b}_s} \geq 0$ *is the reduced distortion in* (23)*, and* $D_{\mathbf{Z}|\mathbf{b}_s}$ *is given in* (20)*.*

By Proposition 1, the remote source coding problem of Definition 2 separates into 1) the MMSE estimation of $\underline{\mathbf{X}}_s$ given $\mathbf{Y}_s$, and 2) the derivation of the RDF of the resultant estimator. On this account, the best encoder E_{si} comprises of the MSE-optimal extraction of the subsources $\underline{\mathbf{X}}_s$ from the noisy linear measurements $\mathbf{Y}_s$ in (16), $s = 1, \dots, |\mathcal{B}|$, followed by the optimal coding of the extracted messages. The estimate-and-compress separation is illustrated in Fig. 4.

Remark 1. The optimal conditional PDF $f(\hat{\mathbf{z}}_s|\mathbf{z}_s)$ for (24a), and the optimal variables for (25) are interrelated with Gaussian "forward channels" $\hat{Z}_{s,k}' \triangleq \theta_{s,k}Z_{s,k}' + V_{s,k}$, $k = 1, \dots, K$, in Fig. 4 via parameters $\theta_{s,k} = \frac{\lambda_{s,k} - D_{s,k}'}{\lambda_{s,k}}$, and $\sigma_{V_{s,k}}^2 = \theta_{s,k}D_{s,k}'$, where $Z_{s,k}'$ is the kth element of the decorrelated MMSE estimator $\mathbf{Z}_s' = \boldsymbol{Q}_s^{\mathrm{T}}\mathbf{Z}_s$, and $V_{s,k} \sim \mathcal{N}(0, \sigma_{V_{s,k}}^2)$ is a zero mean Gaussian random variable independent of $\hat{Z}_{s,k}'$ [7, Theorem 4.3.2].

Remark 2. A proof of the optimality of the two-step coding structure is implicitly present in the seminal work by Dobrushin and Tsybakov [28, Sect. 5] for the case with frequency-weighted MSE distortion where the source and observable processes are jointly Gaussian and stationary. Furthermore, they proved such optimality explicitly for the MSE distortion in the case where the observations are noisy versions of the signal (i.e., no dimension reduction) [28, Sect. 7]. Later, Wolf and Ziv [30] addressed a distortion-rate framework, and proved that separation holds for the MSE distortion under more general conditions (i.e., Gaussianity is not needed). Consequently, the decomposition principle of Proposition 1 is also valid for non-Gaussian sources/observations; however, finding analytical expressions for $R_{\mathbf{Z}|\mathbf{b}_s}^{\mathrm{dir}}(D_s')$ and $D_{\mathbf{Z}|\mathbf{b}_s}$ may be difficult. Similar separation results appear in, e.g., [31], [34], [35], [7, Ch. 4.5.4].

Remark 3. $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$ is an upper bound to the conditional marginal remote RDF of a subsource $\underline{\tilde{\mathbf{X}}}_s = \tilde{\mathbf{G}} \odot \mathbf{b}_s$, where $\tilde{\mathbf{G}}$ is a non-Gaussian random vector with covariance matrix $\boldsymbol{\Sigma}_{\tilde{\mathbf{G}}} = \boldsymbol{\Sigma}_{\mathbf{G}}$ [7, p. 130].

D. Characterization of the Conditional Remote RDF

Let $D_{\mathbf{Z}|\mathbf{B}} \geq 0$ denote the total average MMSE estimation error over all subsources $\underline{\mathbf{X}}_s$, $s = 1, \dots, |\mathcal{B}|$, with support SI, i.e., we have

$$\begin{aligned} D_{\mathbf{Z}|\mathbf{B}} &\triangleq \sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)D_{\mathbf{Z}|\mathbf{b}_s} \\ &\overset{(a)}{=} N^{-1}\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)\mathrm{Tr}\big(\boldsymbol{\Sigma}_{\mathbf{X}_s} - \boldsymbol{\Sigma}_{\mathbf{Z}_s}\big) \end{aligned} \qquad (27)$$
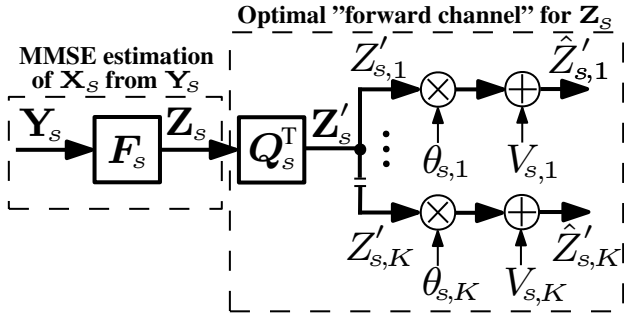
Fig. 4. The optimal compression structure with respect to the conditional marginal remote RDF $R_{\mathbf{X}|\mathbf{b}_s}^{\mathrm{rem}}(D_s)$ in (14a).

where (a) follows from (20). The conditional remote RDF $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ is given by the following theorem.

Theorem 1. *For distortion range $D_{\mathbf{Z}|\mathbf{B}} \leq D < \frac{1}{N}\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)\mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{X}_s})$, $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ is positive, and can be evaluated via the convex minimization problem as*

$$R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D) = \min_{\substack{\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)\sum_{k=1}^{K} D'_{s,k} = D - D_{\mathbf{Z}|\mathbf{B}} \\ D'_{s,k} \geq 0,\, k=1,\dots,K,\, s=1,\dots,|\mathcal{B}|}} \frac{1}{N}\sum_{s=1}^{|\mathcal{B}|}\sum_{k=1}^{K} \cdots$$
$$p(\mathbf{b}_s)\max\left\{0, \frac{1}{2}\log\frac{\lambda_{s,k}}{D'_{s,k}}\right\} \tag{28}$$

where $\mathbf{\Sigma}_{\mathbf{X}_s}$ is the covariance matrix of $\mathbf{X}_s$ in (15); $\lambda_{s,1},\dots,\lambda_{s,K}$ are the eigenvalues of covariance matrix $\mathbf{\Sigma}_{\mathbf{Z}_s}$ of $\mathbf{Z}_s$ in (18); $D'_{s,k}$ are the optimization variables, $k = 1,\dots,K$, $s = 1,\dots,|\mathcal{B}|$. If the distortion values satisfy $D \geq \frac{1}{N}\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)\mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{X}_s})$, then $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ is zero.

Proof. Applying Proposition 1, substituting (25) into (13), and recasting the sum distortion constraint in terms of variables $D'_{s,k}$, $k = 1,\dots,K$, $s = 1,\dots,|\mathcal{B}|$, along with (27) results in (28). A valid distortion requires $D \geq D_{\mathbf{Z}|\mathbf{B}} \geq 0$. In particular, for all $D \geq \frac{1}{N}\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)\mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{X}_s})$, we have $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D) = 0$. If the encoder sends no information (i.e., $R = 0$), then the decoder can set $\hat{\mathbf{X}} = \mathbf{0}_N$, resulting in an admissible distortion because $\mathbb{E}[d(\mathbf{X}, \hat{\mathbf{X}})] = \frac{1}{N}\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2] = \frac{1}{N}\mathbb{E}[\|\mathbf{X}\|_2^2] = \frac{1}{N}\sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)\mathrm{Tr}(\mathbf{\Sigma}_{\mathbf{X}_s}) \leq D$. $\square$

Remark 4. $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ is an upper bound to the conditional remote RDF of a source $\tilde{\mathbf{X}} = \tilde{\mathbf{G}} \odot \mathbf{B}$, where $\tilde{\mathbf{G}}$ is a non-Gaussian random vector with covariance matrix $\mathbf{\Sigma}_{\tilde{\mathbf{G}}} = \mathbf{\Sigma}_{\mathbf{G}}$ [7, p. 130].

By Theorem 1, $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ is determined by a weighted sum of the RDFs of the MMSE estimators $\mathbf{Z}_s$ under the reduced distortion criterion, where the weights – the prior probabilities of the sparsity patterns $p(\mathbf{b}_s)$, $s = 1,\dots,|\mathcal{B}|$ – represent the "appearance frequencies" of such estimators. In particular, (28) involves finding the optimal allocation of the distortion components not only across the $|\mathcal{B}|$ different sparsity patterns, but also across the K entries of each decorrelated random vector $\mathbf{Z}'_s$. This type of weighted minimization is discernibly a consequence of the composite source structure.

$R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ reflects the remote sensing nature of the lossy CS: regardless of the rate, the lowest achievable distortion is ultimately dictated by $D_{\mathbf{Z}|\mathbf{B}}$ – the constant term solely governed by the noisy measurement model in (16). This unavoidable compression performance degradation caused by the indirect observations of the source distinguishes the lossy CS from directly compressing $\mathbf{X}$; RD bounds for compressing sparse sources have been derived in, e.g., [51]–[53]. Note that a constant distortion floor occurs whether or not the support SI is available – only the respective levels for $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ and $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ are different. This is demonstrated by the numerical results in Section VI.

IV. NUMERICAL APPROXIMATION OF THE REMOTE RDF

Since finding an *analytical solution* for the lossy CS problem of Definition 1 seems to be elusive, we develop a method based on the BA algorithm [56], [57] to *numerically approximate* the remote RDF $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ in (7a). As for lossy CS, the standard algorithm must be adapted to handle 1) continuous-valued signals $\mathbf{X}$ and $\mathbf{Y}$, and 2) the remote compression setup. The former is accomplished by a VQ-optimized alphabet discretization method, and the latter by appropriately modifying the distortion measure.

A. Discretization of Signal Alphabets

As the BA algorithm accepts discrete input/output alphabets, the measurement vector space $\mathcal{Y}$ and the reproduction alphabet $\hat{\mathcal{X}}$ are discretized via a VQ. Let $\mathcal{V} \triangleq \{1,\dots,|\mathcal{V}|\}$ be an index set. The $|\mathcal{V}|$-level VQ is determined by 1) the encoder regions $\mathcal{S}_v$, $v \in \mathcal{V}$, which partition the measurement space, i.e., $\mathcal{S}_v \subseteq \mathcal{Y}$, $\mathcal{S}_v \cap \mathcal{S}_{v'} = \emptyset$, for any $v \neq v'$, and $\bigcup_{v=1}^{|\mathcal{V}|} \mathcal{S}_v = \mathcal{Y}$; 2) the reconstruction codebook $\hat{\mathcal{X}}^{\mathrm{q}} \triangleq \{\hat{\boldsymbol{x}}_1,\dots,\hat{\boldsymbol{x}}_{|\mathcal{V}|}\}$ with codevectors $\hat{\boldsymbol{x}}_v \in \mathbb{R}^N$, $v \in \mathcal{V}$. The VQ encoder is a mapping $\mathsf{E}^{\mathrm{vq}} : \mathcal{Y} \to \mathcal{V}$ such that for an input $\mathbf{y} \in \mathcal{S}_v$, it produces an index $\mathsf{E}^{\mathrm{vq}}(\mathbf{y}) = v \in \mathcal{V}$; the VQ decoder performs an inverse mapping $\mathsf{D}^{\mathrm{vq}} : \mathcal{V} \to \hat{\mathcal{X}}^{\mathrm{q}}$ as $\mathsf{D}^{\mathrm{vq}}(v) = \hat{\boldsymbol{x}}_v \in \hat{\mathcal{X}}^{\mathrm{q}}$. The random variable V represents the VQ output.

The next section illuminates the role of the VQ in RD approximation, and the specific optimization of the VQ is deferred until Section IV-C.

B. Modified Blahut-Arimoto Algorithm for Lossy CS

Consider a VQ as described above with $p(v) \triangleq \Pr(V = v) = \int_{\mathbf{y} \in \mathcal{S}_v} f(\mathbf{y})\mathrm{d}\mathbf{y}$, $v \in \mathcal{V}$. Consequently, index $v \in \mathcal{V}$ represents all measurement vectors that belong to VQ region $\mathcal{S}_v$. Similarly, let $\hat{\mathbf{X}}^{\mathrm{q}}$ be a discrete reproduction random vector at the output of decoder D with alphabet $\hat{\mathcal{X}}^{\mathrm{q}} = \{\hat{\boldsymbol{x}}_1,\dots,\hat{\boldsymbol{x}}_{|\mathcal{V}|}\}$ (i.e., the VQ codebook). Replacing $\mathbf{Y}$ with V, and $\hat{\mathbf{X}}$ with $\hat{\mathbf{X}}^{\mathrm{q}}$ in (7a), $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ can be approximated as

$$R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D) = \min_{p(\hat{\boldsymbol{x}}_j|v):\mathbb{E}[d(\mathbf{X},\hat{\mathbf{X}}^{\mathrm{q}})] \leq D} \frac{1}{N}I(V;\hat{\mathbf{X}}^{\mathrm{q}}) \tag{29a}$$

where the optimization is over the conditional probabilities $p(\hat{\boldsymbol{x}}_j|v) \triangleq \Pr(\hat{\mathbf{X}}^{\mathrm{q}} = \hat{\boldsymbol{x}}_j|V = v)$, $v, j \in \mathcal{V}$. The mutual information between V and $\hat{\mathbf{X}}^{\mathrm{q}}$ is

$$I(V;\hat{\mathbf{X}}^{\mathrm{q}}) = \sum_{v=1}^{|\mathcal{V}|}\sum_{j=1}^{|\mathcal{V}|} p(v)p(\hat{\boldsymbol{x}}_j|v)\log\frac{p(\hat{\boldsymbol{x}}_j|v)}{p(\hat{\boldsymbol{x}}_j)} \tag{29b}$$

and the average distortion between $\mathbf{X}$ and $\hat{\mathbf{X}}^{\mathrm{q}}$ is

$$\mathbb{E}\big[d(\mathbf{X}, \hat{\mathbf{X}}^{\mathrm{q}})\big] = \sum_{v=1}^{|\mathcal{V}|} \sum_{j=1}^{|\mathcal{V}|} p(v)p(\hat{\boldsymbol{x}}_j|v)\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v) \quad (29c)$$

where $\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v) \geq 0$ is the *modified* distortion measure, defined as the average per-letter MSE distortion between $\mathbf{X}$ and $\hat{\boldsymbol{x}}_j$ conditioned on $V = v$, i.e.,

$$\begin{aligned}
\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v) &\triangleq \mathbb{E}\big[d(\mathbf{X}, \hat{\boldsymbol{x}}_j)|V = v\big], \ v, j \in \mathcal{V} \\
&= \frac{1}{N} \int_{\mathbf{y}} \mathbb{E}\big[\|\mathbf{X} - \hat{\boldsymbol{x}}_j\|_2^2 | V = v, \mathbf{Y} = \mathbf{y}\big] f(\mathbf{y}|v)\mathrm{d}\mathbf{y} \\
&= \frac{1}{N} \int_{\mathbf{y}} \frac{p(v|\mathbf{y})}{p(v)} \cdots \\
&\qquad \mathbb{E}\big[\|\mathbf{X} - \hat{\boldsymbol{x}}_j\|_2^2 | V = v, \mathbf{Y} = \mathbf{y}\big] f(\mathbf{y})\mathrm{d}\mathbf{y} \\
&\overset{(a)}{=} \frac{1}{N} \frac{1}{p(v)} \int_{\mathbf{y} \in \mathcal{S}_v} \mathbb{E}\big[\|\mathbf{X} - \hat{\boldsymbol{x}}_j\|_2^2 | \mathbf{Y} = \mathbf{y}\big] f(\mathbf{y})\mathrm{d}\mathbf{y}
\end{aligned}$$
$$(29d)$$

where (a) follows from the Markov chain $\mathbf{X} \to \mathbf{Y} \to V$, and from $p(v|\mathbf{y}) = 1$, if $\mathbf{y} \in \mathcal{S}_v$, $v \in \mathcal{V}$, and 0 otherwise. Note that pre-calculated $|\mathcal{V}|^2$ quantities $\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v)$ remain fixed in the algorithm. In the context of *discrete* remote sources, a distortion measure similar to (29d) appears in, e.g., [7, Sect. 3.5] [31], [58].

Consider a Lagrangian for (29a) as

$$\begin{aligned}
&\mathcal{L}\big(\{p(\hat{\boldsymbol{x}}_j|v)\}_{v,j=1}^{|\mathcal{V}|}, \lambda, \{\nu_v\}_{v=1}^{|\mathcal{V}|}\big) = \\
&\frac{1}{N} \sum_{v=1}^{|\mathcal{V}|} \sum_{j=1}^{|\mathcal{V}|} p(v)p(\hat{\boldsymbol{x}}_j|v)\log\frac{p(\hat{\boldsymbol{x}}_j|v)}{p(\hat{\boldsymbol{x}}_j)} + \\
&\lambda \sum_{v=1}^{|\mathcal{V}|} \sum_{j=1}^{|\mathcal{V}|} p(v)p(\hat{\boldsymbol{x}}_j|v)\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v) + \\
&\sum_{v=1}^{|\mathcal{V}|} \nu_v \sum_{j=1}^{|\mathcal{V}|} p(\hat{\boldsymbol{x}}_j|v)
\end{aligned} \quad (30)$$

where $\lambda > 0$ is the Lagrange multiplier associated with the sum distortion constraint, and ν_v, $v \in \mathcal{V}$, are the Lagrange multipliers associated with the valid conditional probability constraints $\sum_{j=1}^{|\mathcal{V}|} p(\hat{\boldsymbol{x}}_j|v) = 1$, $\forall v \in \mathcal{V}$. Following the standard BA procedure, an (R, D) point of $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$ in (29a) is obtained by sequentially updating the conditional probabilities $p(\hat{\boldsymbol{x}}_j|v)$, and the reproduction probabilities $p(\hat{\boldsymbol{x}}_j)$ for a fixed λ at each iteration $t = 1, 2, \ldots$ as [8, Sect. 10.8]

$$p(\hat{\boldsymbol{x}}_j|v)^{t+1} := \frac{p(\hat{\boldsymbol{x}}_j)^t \exp\big[-\lambda\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v)\big]}{\sum_{j'=1}^{|\mathcal{V}|} p(\hat{\boldsymbol{x}}_{j'})^t \exp\big[-\lambda\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_{j'}|v)\big]}, \ v, j \in \mathcal{V}$$
$$(31a)$$

$$p(\hat{\boldsymbol{x}}_j)^{t+1} := \sum_{v=1}^{|\mathcal{V}|} p(\hat{\boldsymbol{x}}_j|v)^{t+1} p(v), \ j \in \mathcal{V}, \quad (31b)$$

until convergence, and evaluating the rate according to (29b), and the distortion according to (29c). Hence, different values for λ sweep the curve for $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$, which approximates the remote RDF $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ in (7a) with an accuracy that increases with the number $|\mathcal{V}|$.

The proposed method is summarized in Algorithm 1. The algorithm can be terminated when the quantities do not significantly change, e.g., when $\sum_{j=1}^{|\mathcal{V}|} \big(p(\hat{\boldsymbol{x}}_j)^t - p(\hat{\boldsymbol{x}}_j)^{t-1}\big)^2 < \epsilon_{\mathrm{ba}}$ for a pre-defined positive constant $\epsilon_{\mathrm{ba}} > 0$. The algorithm inputs $p(v)$, $\hat{\boldsymbol{x}}_v$, and $\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v)$, $v, j \in \mathcal{V}$, are the outcomes of the VQ optimization. The optimization is carried out next.

C. Vector Quantization Optimization

In Algorithm 1, the accuracy of distortion evaluation through (29c) is ultimately limited by the $|\mathcal{V}|^2$ fixed quantities

---

Algorithm 1 Modified Blahut-Arimoto algorithm for approximating the remote RDF $R_{\mathbf{X}}^{\mathrm{rem}}(D)$

Inputs: a) Lagrange multipliers $\lambda > 0$; b) codevectors $\hat{\boldsymbol{x}}_v$, and index probabilities $p(v)$, $v \in \mathcal{V}$, obtained as described in Section IV-C; c) modified distortion measures $\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v)$, $v, j \in \mathcal{V}$, of form (29d).

Initializations: a) Set $t := 1$; b) set $p(\hat{\boldsymbol{x}}_j)^t := 1/|\mathcal{V}|$, $j \in \mathcal{V}$.

for a given λ
repeat
  1) Update the conditional probabilities $p(\hat{\boldsymbol{x}}_j|v)^{t+1}$, $v, j \in \mathcal{V}$, according to (31a).
  2) Update the reproduction probabilities $p(\hat{\boldsymbol{x}}_j)^{t+1}$, $j \in \mathcal{V}$, according to (31b).
  3) Set $t := t + 1$.
until a pre-defined stopping criterion is met.
4) Compute the rate R_λ according to (29b), and the distortion D_λ according to (29c).
end for

Output: $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$ curve determined by the (R_λ, D_λ) pairs.

$\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v)$, $v, j \in \mathcal{V}$, of (29d). Taking this into account, the E^{vq}-D^{vq} pair, that is, the encoder regions $\mathcal{S}_v$, and codevectors $\hat{\boldsymbol{x}}_v$, $v \in \mathcal{V}$, are optimized to minimize the average MSE distortion between the source $\mathbf{X}$ and its $|\mathcal{V}|$-level reproduction $\hat{\mathbf{X}}^{\mathrm{q}}$, i.e., we choose

$$\begin{aligned}
\{\mathcal{S}_v^*, \hat{\boldsymbol{x}}_v^*\}_{v\in\mathcal{V}} &:= \operatorname*{argmin}_{\{\mathcal{S}_v, \hat{\boldsymbol{x}}_v\}_{v\in\mathcal{V}}} \frac{1}{N}\mathbb{E}\big[\|\mathbf{X} - \hat{\mathbf{X}}^{\mathrm{q}}\|_2^2\big] \\
&\overset{(a)}{:=} \operatorname*{argmin}_{\{\mathcal{S}_v, \hat{\boldsymbol{x}}_v\}_{v\in\mathcal{V}}} \frac{1}{N} \sum_{v\in\mathcal{V}} \cdots \\
&\qquad \int_{\mathbf{y}\in\mathcal{S}_v} \mathbb{E}\big[\|\mathbf{X} - \hat{\boldsymbol{x}}_v\|_2^2 | \mathbf{Y} = \mathbf{y}\big] f(\mathbf{y})\mathrm{d}\mathbf{y} \\
&\overset{(b)}{:=} \operatorname*{argmin}_{\{\mathcal{S}_v, \hat{\boldsymbol{x}}_v\}_{v\in\mathcal{V}}} \sum_{v\in\mathcal{V}} p(v)\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_v|v)
\end{aligned}$$
$$(32)$$

where in (a) we used the Markov chain $\mathbf{X} \to \mathbf{Y} \to V$, and $p(v|\mathbf{y}) = 1$, if $\mathbf{y} \in \mathcal{S}_v$, $v \in \mathcal{V}$, and 0 otherwise; (b) follows from (29d).

Remark 5. Besides $\bar{d}(\mathbf{X}, \hat{\boldsymbol{x}}_j|v)$, $v, j \in \mathcal{V}$, our VQ affects the final distortion in (29c) through the conditional probabilities $p(\hat{\boldsymbol{x}}_j|v)$, $v, j \in \mathcal{V}$ – the variables to be optimized in iterative steps (31a) and (31b). In addition, the VQ affects the rate approximation in (29b) through the index probabilities $p(v)$, $v \in \mathcal{V}$. Therefore, we expect to achieve a better approximation of $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ by incorporating the VQ optimization in the iterative loop of Algorithm 1, and, thus, generating a unique VQ for each λ. For example, the mapping approach in [62] adapts the reproduction points within the optimization loop. Nevertheless, the proposed non-adaptive discretization seems to yield decent accuracy for all distortion values D, as demonstrated in Section VI.

Remark 6. Setting the conditional probabilities (31a) as $p(\hat{\boldsymbol{x}}_j|v) = 1$ for $v = j$, and 0 otherwise, results in reproduction probabilities $p(\hat{\boldsymbol{x}}_j) = p(v)$ for $v = j$, and 0 otherwise in (31b). The distortion (29c) for $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$ then becomes equal

to the VQ distortion in (32), and the rate (29b) becomes $R = -\sum_{v=1}^{|\mathcal{V}|} p(v)\log(p(v))$, i.e., the entropy of quantization index V. $R_{\mathbf{X},\text{ba}}^{\text{rem}}(D)$ is thus a "noisy VQ" that randomizes the mapping $V \rightarrow \hat{\mathcal{X}}^{\text{q}}$ via conditional probabilities $p(\hat{\boldsymbol{x}}_j|v)$ acting as a noisy channel between the encoder output and decoder input.

The joint optimization over $\mathcal{S}_v$ and $\hat{\boldsymbol{x}}_v$, $v \in \mathcal{V}$, in (32) seems to be intractable, and so we use a common alternating minimization to derive necessary optimality conditions. Accordingly, the proposed VQ is equivalent to the VQ in [41] designed for noiseless channels. The optimal encoder regions for fixed codevectors satisfy a generalized nearest-neighbor condition

$$\mathcal{S}_v^* = \left\{ \mathbf{y} : \|\mathbf{z} - \hat{\boldsymbol{x}}_v\|_2^2 \leq \|\mathbf{z} - \hat{\boldsymbol{x}}_{v'}\|_2^2, \forall v' \neq v \right\}, \; v \in \mathcal{V}, \tag{33}$$

where $\mathbf{z} \in \mathbb{R}^N$ is the MMSE estimate of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$, defined as [83]–[85]

$$\begin{aligned} \mathbf{z} &\triangleq \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] \\ &= \sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s|\mathbf{y})\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}, \mathbf{B} = \mathbf{b}_s] \\ &= \sum_{s=1}^{|\mathcal{B}|} \frac{p(\mathbf{b}_s)f(\mathbf{y}|\mathbf{b}_s)}{\sum_{s'=1}^{|\mathcal{B}|} p(\mathbf{b}_{s'})f(\mathbf{y}|\mathbf{b}_{s'})} \underline{\mathbf{z}}_s \end{aligned} \tag{34}$$

where the conditional PDF $f(\mathbf{y}|\mathbf{b}_s)$ is Gaussian as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Y}_s})$ (see Definition 4), and $\underline{\mathbf{z}}_s \triangleq \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}, \mathbf{B} = \mathbf{b}_s]$ is the MMSE estimate of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$ and $\mathbf{B} = \mathbf{b}_s$, which, by Definition 5, comprises of vectors $\mathbf{z}_s = \boldsymbol{F}_s \mathbf{y}_s \in \mathbb{R}^K$ and $\mathbf{0}_{N-K}$. Similarly, the optimal codevectors for fixed encoder regions satisfy a generalized centroid condition

$$\hat{\boldsymbol{x}}_v^* = \frac{1}{p(v)} \int_{\mathbf{y} \in \mathcal{S}_v} \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}] f(\mathbf{y}) d\mathbf{y}, \; v \in \mathcal{V}. \tag{35}$$

The VQ can be trained offline in the spirit of the iterative Lloyd algorithm [86]–[88] by successively applying the necessary optimality conditions (33) and (35) for training data sets.

V. "ESTIMATE-AND-COMPRESS" QCS METHOD

Approaching $R_{\mathbf{X}}^{\text{rem}}(D)$ in (7a) requires encoding (large) blocks of vectors. We instead propose a symbol-by-symbol QCS method that follows the optimal "estimate-and-compress" principle underlying $R_{\mathbf{X}}^{\text{rem}}(D)$, and is empirically shown to approach $R_{\mathbf{X}}^{\text{rem}}(D)$ in Section VI.

The proposed method, termed ECVQ-CS, relies on entropy-constrained VQ [89], and minimizes a weighted distortion-rate cost function

$$\begin{aligned} &(1 - \mu^{\text{ec}}) \sum_{i \in \mathcal{I}} p(i) \mathbb{E}\big[\|\mathbf{X} - \boldsymbol{c}_i\|_2^2 | I = i\big] - \\ &\mu^{\text{ec}} \sum_{i \in \mathcal{I}} p(i) \log(p(i)) \end{aligned} \tag{36}$$

where I is the quantization index with index set $\mathcal{I} \triangleq \{1, \ldots, |\mathcal{I}|\}$, $\boldsymbol{c}_i \in \mathbb{R}^N$ are the reconstruction codevectors, and the parameter $\mu^{\text{ec}} \in [0, 1]$ adjusts the distortion-rate trade-off. Using alternating optimization [49], [50], [89], ECVQ-CS can be trained via a three-step iterative algorithm, where 1)

the encoder regions are formed as

$$\begin{aligned} \mathcal{S}_i^* = \Big\{ \mathbf{y} : &(1 - \mu^{\text{ec}})\|\mathbf{z} - \boldsymbol{c}_i\|_2^2 - \mu^{\text{ec}}\log(p(i)) \leq \\ &(1 - \mu^{\text{ec}})\|\mathbf{z} - \boldsymbol{c}_{i'}\|_2^2 - \mu^{\text{ec}}\log(p(i')), \forall i' \neq i \Big\}, \; i \in \mathcal{I}, \end{aligned} \tag{37}$$

where $\mathbf{z} = \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$ is the MMSE estimate of (34); 2) the rate terms $-\log(p(i))$, $i \in \mathcal{I}$, are updated given the new regions; 3) the codevectors are set equivalently to (35). Finally, the index probabilities $p(i) = \int_{\mathbf{y} \in \mathcal{S}_i} f(\mathbf{y}) d\mathbf{y}$, $i \in \mathcal{I}$, are used to generate a binary source codebook with average codeword length close to the index entropy via, e.g., Huffman coding [90]. The name "estimate-and-compress" describes the two main steps: the sensor compresses each measurement realization $\mathbf{y}$ by 1) forming the MMSE estimate $\mathbf{z} = \mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$, and 2) obtaining the optimal encoding index as $i^* = \arg\min_{i \in \mathcal{I}} (1 - \mu^{\text{ec}})\|\mathbf{z} - \boldsymbol{c}_i\|_2^2 - \mu^{\text{ec}}\log(p(i))$.

VI. NUMERICAL RESULTS

A. Simulation Setup

Consider setups with $\boldsymbol{\Sigma}_{\mathbf{G}} = \sigma_G^2 \boldsymbol{I}_N$ with $\sigma_G^2 = 1$, and $\boldsymbol{\Sigma}_{\mathbf{W}} = \sigma_W^2 \boldsymbol{I}_M$ with $\sigma_W^2 = 0.01$. The following curves and QCS methods are evaluated:

1) $R_{\mathbf{X}|\mathbf{B}}^{\text{rem}}(D)$: the conditional remote RDF of Theorem 1.
2) $R_{\mathbf{X},\text{ba}}^{\text{rem}}(D)$: a numerically approximated remote RDF of Algorithm 1.
3) $R_{\mathbf{X}|\mathbf{B}}^{\text{dir}}(D)$: *the conditional direct RDF of* $\mathbf{X}$, corresponding to lossy compression of $\mathbf{X}$ with $\mathbf{B}$ available as SI at the encoder and decoder (see [52, Sect. VII-A]). Clearly, we have $R_{\mathbf{X}|\mathbf{B}}^{\text{dir}}(D) \leq R_{\mathbf{X}|\mathbf{B}}^{\text{rem}}(D)$.
4) $R_{\mathbf{X},\text{ba}}^{\text{dir}}(D)$: a numerically approximated *direct RDF of* $\mathbf{X}$ which represents lossy compression of $\mathbf{X}$ without support SI, and is obtained by applying the discretization of Section IV-C and Algorithm 1 with using $\mathbf{Y} = \mathbf{X}$. Clearly, we have $R_{\mathbf{X}|\mathbf{B}}^{\text{dir}}(D) \leq R_{\mathbf{X},\text{ba}}^{\text{dir}}(D) \leq R_{\mathbf{X},\text{ba}}^{\text{rem}}(D)$.
5) ECVQ-CS: the proposed "estimate-and-compress" method of Section V with $\mu^{\text{ec}} = 0.1/\log(|\mathcal{I}|)$, and Huffman codewords.
6) VQ-CS: the fixed-rate QCS method in [41], where the $|\mathcal{I}|$-level VQ is optimized for noiseless channels.
7) VQ-CE: a baseline fixed-rate QCS method that performs the two stages of ECVQ-CS in the reverse (suboptimal) order: the encoder optimally quantizes $\mathbf{Y}$ in the MSE-sense, unaware of $\mathbf{X}$, and the decoder estimates $\mathbf{X}$ from these quantized measurements. The encoder of VQ-CE is an $|\mathcal{I}|$-level VQ that minimizes the distortion $\sum_{i \in \mathcal{I}} p(i)\mathbb{E}\big[\|\mathbf{Y} - \hat{\boldsymbol{y}}_i\|_2^2 | I = i\big]$ with nearest-neighbor vectors $\hat{\boldsymbol{y}}_i \in \mathbb{R}^M$, and its decoder consists of the MSE-optimal codevectors of form (35). This "compress-and-estimate" approach underlies many early QCS methods, cf. [27], [38].
8) $D_{\mathbf{Z}|\mathbf{B}}$: the average MMSE estimation error in (27) (known $\mathbf{B}$).
9) $D_{\mathbf{Z}}$: numerically evaluated average MMSE estimation error of $\mathbf{X}$ given $\mathbf{Y}$, i.e., $D_{\mathbf{Z}} \triangleq N^{-1}\mathbb{E}\big[\|\mathbf{X} - \mathbf{Z}\|_2^2\big]$; estimator $\mathbf{Z} \triangleq \mathbb{E}[\mathbf{X}|\mathbf{Y}]$ takes values according to (34) (unknown $\mathbf{B}$).

The quantization rate is set as $\log|\mathcal{V}| = 12$ bits for $R_{\mathbf{X},\text{ba}}^{\text{rem}}(D)$, and varied as $\log|\mathcal{I}| = 1, \ldots, 12$ bits for the QCS methods. The measurement matrix $\boldsymbol{\Phi}$ is generated by taking

the first M rows of an $N \times N$ discrete cosine transform matrix, and normalizing the columns as $\| \cdot \|_2 = 1$. The distortion is measured as $10\log_{10}\left(\mathbb{E}[d(\mathbf{X}, \mathbf{X}^{\mathrm{est}})]/N^{-1}\mathbb{E}[\|\mathbf{X}\|_2^2]\right)$ dB, where $\mathbf{X}^{\mathrm{est}}$ is the method-dependent decoded estimate of $\mathbf{X}$, and $\mathbb{E}\left[\|\mathbf{X}\|_2^2\right] = \sum_{s=1}^{|\mathcal{B}|} p(\mathbf{b}_s)\mathrm{Tr}(\boldsymbol{\Sigma}_{\mathbf{X}_s})$. The rate is measured as R bits/entry of $\mathbf{X}$. The convex minimization problems are solved via CVX [91].

Complexity: As the complexities of Algorithm 1 and the QCS methods increase exponentially with the number of VQ levels and $|\mathcal{B}|$, the experiments are confined to moderate signal dimensions and quantization rates; the most complex setup used in the paper involves $N = 20$, $M = 8$, $K = 2$, and $\log|\mathcal{V}| = \log|\mathcal{I}| = 12$ bits. As for any discretized BA algorithm, we remark that the same complexity issue due to a large number of variables remains regardless of the quantization method. In fact, due to the VQ advantages [92], the proposed algorithm enjoys a superior trade-off between the approximation accuracy and the complexity as compared to, e.g., SQ. Complexities of the QCS methods can be decreased via using SQ or low-complexity VQ variants like tree-structured, multi-step, and lattice VQs [93]. An alternative is to reconstruct the signal from quantized measurements via, e.g., standard ℓ_1-minimization or a greedy reconstruction method, albeit with low compression performance [24]. The complexity of ECVQ-CS could be reduced by approximating the MMSE estimates (34) by, e.g., the randomized orthogonal matching pursuit [83]. These considerations, which could allow more realistic setups with N and M being hundreds, are left for future study.

B. Rate-Distortion Behavior of Lossy CS

Consider a setup with $N = 7$, $M = 5$, $K = 1$, and equal support probabilities $p(\mathbf{b}_s) = 1/|\mathcal{B}|$, $\forall s = 1, \ldots, |\mathcal{B}|$. Fig. 5(d) depicts the average distortion versus the average rate for different compression schemes.

Consider first the SI aided lower bounds $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$, $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D)$, and $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ to the remote RDF $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ in (7a). Owing to the direct observations with support SI, $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$ appears as the line (in log scale) with slope $-6\frac{N}{K} = -42$ dB/bit [52], and yields the lowest R for all values of D, as expected. The substantially increased rate for $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D)$ as compared to $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$ is caused by the necessity of conveying the support of $\mathbf{X}$ to the decoder. While $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$ and $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ nearly coincide at high distortion, the curves diverge for moderate to low distortion values. The gradually increasing gap between $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ and $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$ for low values of D is a consequence of the remote sensing. Note that whereas arbitrarily small distortion is achievable at asymptotically high rates for $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$ and $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D)$ (i.e., $\lim_{D \to 0} R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D) = \infty$ and $\lim_{D \to 0} R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D) = \infty$), the lowest achievable distortion for $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ is the MMSE estimation error floor $D_{\mathbf{Z}|\mathbf{B}}$ (i.e., $\lim_{D \to D_{\mathbf{Z}|\mathbf{B}}} R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D) = \infty$).

Focus now on the approximate remote RDF $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$, i.e., the best achievable performance of any QCS method. The gap between $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ and $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$ represents the compression loss induced by the random measurements taken

without knowing the sparse support in a QCS setup [24]. The tightness of the lower bound is heavily influenced by the signal setup parameters, as will be exemplified in the subsequent experiments. Despite the gap, the proposed lower bound $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ captures the main peculiarities of lossy CS: the curve has an almost linear distortion region at low rates, whereas for high rates, the distortion saturates to the MMSE estimation error floor $D_{\mathbf{Z}}$. We remark that due to the sparsity, the slope of $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$ at low rates is steeper than the conventional -6 dB/bit. Note that for small R, the rate is the most limiting factor to achievable distortion, and, thus, $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D)$ nearly coincides with $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$; for higher rates, the impact of noisy CS measurements increases, thereby degrading the performance of $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$. Regarding the approximation accuracy of $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D)$ and $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$, observe that the highest obtained rate is $R \approx 1.3$ bits, so the "over-sampling ratio" of the VQ discretization is at least $|\mathcal{V}|/2^{NR} \approx 7.5$.

As the encoder of VQ-CE is CS-blind, its performance is the worst amongst the QCS methods. The advantages of entropy coding are shown by the proposed ECVQ-CS curves which, for moderate rates, approach the compression limit $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$. As a proof of validity, VQ-CS eventually saturates to $D_{\mathbf{Z}}$, which is expected to also happen for the other QCS methods at sufficiently high rates.

C. Effect of Number of Measurements

For the setup of Section VI-B, Figs. 5(a) – (d) illustrate the influence of different numbers of measurements $M = \{2, 3, 4, 5\}$ on the compression performance. As M increases, i.e., the signal-to-noise ratio increases, the level of $D_{\mathbf{Z}}$ decreases, and the performance of each method without support SI moves closer to the lower bound $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$. The largest gain is achieved when M is increased from $M = 2$ to $M = 3$, whereas the difference between $M = 4$ and $M = 5$ is almost negligible. This matches the CS philosophy: increasing M beyond the value that suffices for accurate CS signal recovery does not bring significant gains. In this respect, provided that M is already at this satisfactory level, it pays off to primarily invest in rate R to meet the given distortion fidelity D. Note that the convergence of the curves to their respective distortion floors is rather similar for all M, and that $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$, $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D)$, $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$, and $D_{\mathbf{Z}|\mathbf{B}}$ remain unaltered.

D. Effect of Support Probabilities

Consider a setup with $N = 20$, $M = 8$, $K = 2$, and unequal support probabilities as $p(\mathbf{b}_s) = \alpha_{\mathrm{pl}}^s / \sum_{s'=1}^{|\mathcal{B}|} \alpha_{\mathrm{pl}}^{s'}$, $s = 1, \ldots, |\mathcal{B}|$, where $0 < \alpha_{\mathrm{pl}} \leq 1$ is a parameter that adjusts the concentration of the probability mass function (PMF) of $\mathbf{B}$, and $1 > p(\mathbf{b}_1) \geq \cdots \geq p(\mathbf{b}_{|\mathcal{B}|}) > 0$. For small values of α_{pl}, the PMF concentrates around a fraction of elements in $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_{|\mathcal{B}|}\}$, and vice versa. $\alpha_{\mathrm{pl}} = 1$ corresponds to the uniform distribution, whereas $\alpha_{\mathrm{pl}} \to 0$ approaches remote compression of only a single 2-sparse vector. The vectors $\mathbf{b}_s$ in alphabet $\mathcal{B}$ are ordered such that the decimal number of a binary string represented by $\mathbf{b}_{s+1}$ is greater than that of $\mathbf{b}_s$, $s = 1, \ldots, |\mathcal{B}| - 1$ ($|\mathcal{B}| = 190$).
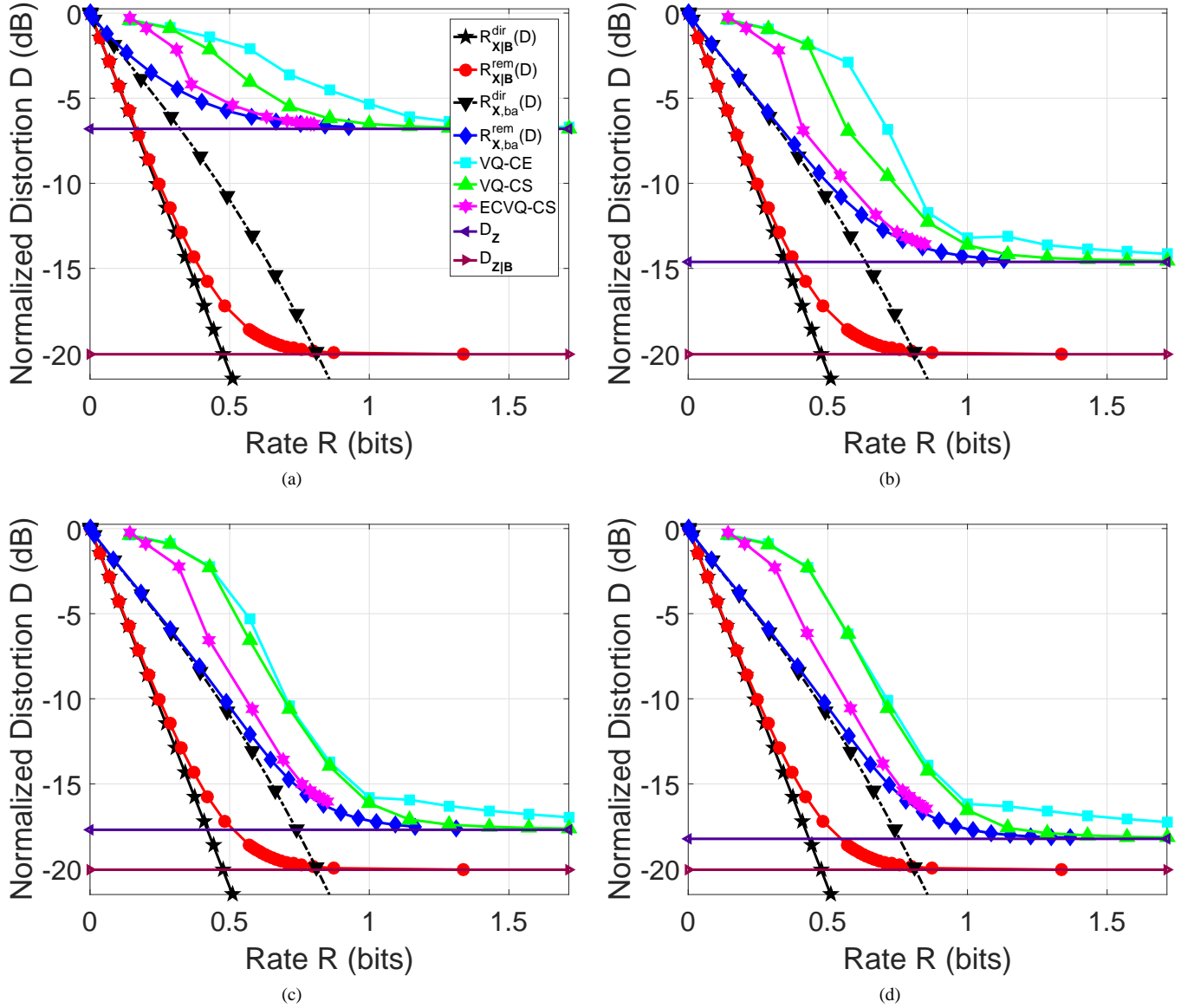
Fig. 5. RD performance of lossy CS schemes with equal support probabilities for $N = 7$, $K = 1$, and the number of measurements (a) $M = 2$, (b) $M = 3$, (c) $M = 4$, and (d) $M = 5$. The colors and markers of the curves in (b), (c), and (d) are equivalent to those in (a).

Fig. 6 shows the average distortion versus the average rate for $\alpha_{\mathrm{pl}} = \{0.98, 0.90, 0.72\}$. Decreasing α_{pl} reduces the uncertainty of the signal support which improves the compression efficiency. This is seen as the increased decay rate of D for the non-SI schemes, the shift of $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$ towards $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$ and $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{dir}}(D)$ towards $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{dir}}(D)$, and the reduction of the gap between $D_{\mathbf{Z}}$ and $D_{\mathbf{Z}|\mathbf{B}}$, which is related to the best possible support recovery for a given setup. This exemplifies that, for a sufficiently concentrated PMF of $\mathbf{B}$, the proposed ECVQ-CS efficiently encodes sparse vectors from noisy CS measurements: its performance approaches the best achievable performance of a support unaware QCS method (i.e., $R_{\mathbf{X}}^{\mathrm{rem}}(D)$). The result illustrates the MSE separation principle (cf. (19)): an efficient QCS method implicitly (successfully) recovers $\mathbf{X}$ from $\mathbf{Y}$, and encodes the resulting estimates optimally.

VII. CONCLUSIONS AND FUTURE WORK

We addressed lossy compression of single-sensor CS from the remote source coding perspective. By giving support side information to the encoder and decoder, we derived the conditional remote RDF to establish a compression lower bound for a finite-rate CS setup. The best such encoder separates into an MMSE estimation step and an optimal transmission step. A modified BA algorithm was proposed to numerically approximate the remote RDF, serving as the best attainable compression performance of any practical QCS method. The main RD characteristics of the lossy CS were demonstrated by comparing the performance of various practical QCS methods – including the devised near-optimal entropy coding based QCS method – with the proposed limits.

Finding a closed-form expression for $R_{\mathbf{X}}^{\mathrm{rem}}(D)$ remains the ultimate goal. Alternatively, one could try to accurately
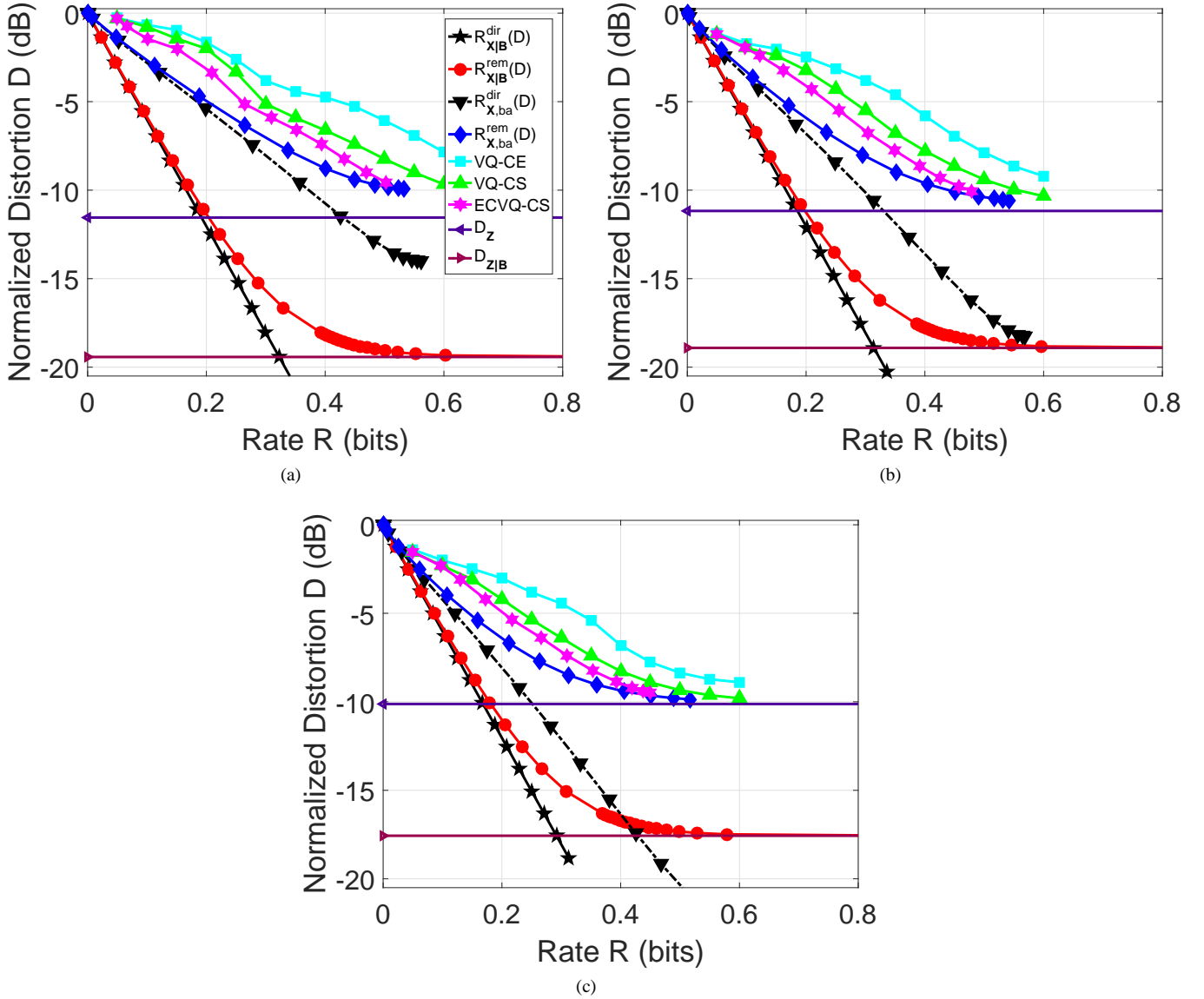
Fig. 6. RD performance of lossy CS schemes for $N = 20$, $M = 8$, $K = 2$, and power law type support probabilities with parameters (a) $\alpha_{\mathrm{pl}} = 0.98$, (b) $\alpha_{\mathrm{pl}} = 0.90$, and (c) $\alpha_{\mathrm{pl}} = 0.72$. The colors and markers of the curves in (b) and (c) are equivalent to those in (a).

approximate $R_{\mathbf{X}}^{\mathrm{rem}}(D)$, or find a substitute for the shared support SI to derive a tighter lower bound as $R_{\mathbf{X}|\mathbf{B}}^{\mathrm{rem}}(D)$. As for the proposed BA algorithm, one could 1) compare it to a method based on the mapping approach in [62], 2) assess the performance gain (i.e., approximation accuracy vs. the computational complexity) of the VQ discretization against regular grid quantization, and 3) modify the algorithm to make it applicable to high-dimensional signal setups. Other open problems include finding theoretical results for 1) the (almost) linear slope of $R_{\mathbf{X},\mathrm{ba}}^{\mathrm{rem}}(D)$ at low rates, and 2) the gap between $D_{\mathbf{Z}}$ and $D_{\mathbf{Z}|\mathbf{B}}$, which is ultimately defined by support recovery performance. It has been shown for direct source compression that if a quantizer is optimized for a memoryless Gaussian source, and the quantizer is then used to compress a non-Gaussian source, the resulting distortion is as bad as if the source were actually Gaussian [94]. If this can be shown

for remote compression, the achieved curves for the QCS methods are upper bounds to sparse sources $\tilde{\mathbf{X}} = \tilde{\mathbf{G}} \odot \mathbf{B}$, where $\tilde{\mathbf{G}}$ is a non-Gaussian random vector with covariance matrix $\Sigma_{\tilde{\mathbf{G}}} = \Sigma_{\mathbf{G}}$.

REFERENCES

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[2] M. Vecchio, R. Giaffreda, and F. Marcelloni, "Adaptive lossless entropy compressors for tiny IoT devices," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 1088–1100, Feb. 2014.

[3] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[4] ——, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec*, vol. 7, pp. 142–163, 1959.

[5] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Mag.*, vol. 21, no. 5, pp. 80–94, Sep. 2004.

[6] R. Gray and D. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[7] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*, ser. Prentice-Hall Series in Information and System Sciences. Prentice Hall, 1971.

[8] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, USA: John Wiley, 2006.

[9] R. Gallager, *Information Theory and Reliable Communication*. New York, USA: John Wiley & Sons, Inc., 1968.

[10] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, "Distributed compressed sensing," Dept. Elect. Comput. Eng., Rice University, Houston, TX, Tech. Rep. ECE-0612, Nov. 2005. [Online]. Available: http://webee.technion.ac.il/people/drorb/pdf/DCS112005.pdf

[11] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3447–3461, Oct. 2012.

[12] M. Leinonen, M. Codreanu, and M. Juntti, "Sequential compressed sensing with progressive signal reconstruction in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1622–1635, Mar. 2015.

[13] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[14] S. K. Sharma, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Application of compressive sensing in cognitive radio communications: A survey," *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 3, pp. 1838–1860, Third quarter 2016.

[15] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[16] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[17] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[18] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[19] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[20] E. Candes, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3, Madrid, Spain, Aug. 22–30 2006, pp. 1433–1452.

[21] M. Duarte, M. Wakin, D. Baron, and R. Baraniuk, "Universal distributed sensing via random projections," in *Proc. IEEE Int. Symp. on Inform. Proc. in Sensor Networks*, New York, NY, USA, 2006, pp. 177–185.

[22] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118 –121, Jul. 2007.

[23] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[24] V. Goyal, A. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 48–56, 2008.

[25] J. Sun and V. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, Jun. 28 – Jul.3 2009, pp. 6–10.

[26] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Processing Lett.*, vol. 17, no. 2, pp. 149–152, 2010.

[27] L. Jacques, D. Hammond, and J. Fadili, "Dequantizing compressed sensing: When oversampling and non-Gaussian constraints combine," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 559–571, 2011.

[28] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inform. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[29] D. Sakrison, "Source encoding in the presence of random disturbance (corresp.)," *IEEE Trans. Inform. Theory*, vol. 14, no. 1, pp. 165–167, Jan. 1968.

[30] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. 16, no. 4, pp. 406–411, Jul. 1970.

[31] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inform. Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.

[32] T. Flynn and R. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. 33, no. 6, pp. 773–787, Nov. 1987.

[33] C. Tian and J. Chen, "Remote vector Gaussian source coding with decoder side information under mutual information and distortion constraints," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4676–4680, Oct. 2009.

[34] I. D. Schizas, G. B. Giannakis, and N. Jindal, "Distortion-rate bounds for distributed estimation using wireless sensor networks," *EURASIP J. Advances Signal Processing*, vol. 1, p. 12, 2008, article ID: 748605.

[35] A. Kipnis, A. J. Goldsmith, T. Weissman, and Y. C. Eldar, "Distortion rate function of sub-Nyquist sampled Gaussian sources," *IEEE Trans. Inform. Theory*, vol. 62, no. 1, pp. 401–429, Jan. 2016.

[36] A. Kipnis, G. Reeves, Y. Eldar, and A. Goldsmith, "Compressed sensing under optimal quantization," in *Proc. IEEE Int. Symp. Inform. Theory*, Aachen, Germany, Jun. 25-30 2017, pp. 2148–2152.

[37] A. Fletcher, S. Rangan, and V. Goyal, "On the rate-distortion performance of compressed sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Honolulu, HI, USA, Apr. 15-20 2007, pp. III–885 – III–888.

[38] W. Dai and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1857–1866, 2011.

[39] A. Shirazinia, S. Chatterjee, and M. Skoglund, "Performance bounds for vector quantized compressive sensing," in *Proc. IEEE Int. Symp. Inform. Theory and its Applications*, Honolulu, HI, USA, Oct. 28-31 2012, pp. 289–293.

[40] G. Coluccia, A. Roumy, and E. Magli, "Operational rate-distortion performance of single-source and distributed compressed sensing," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 2022–2033, Jun. 2014.

[41] A. Shirazinia, S. Chatterjee, and M. Skoglund, "Joint source-channel vector quantization for compressed sensing," *IEEE Trans. Signal Processing*, vol. 62, no. 14, pp. 3667–3681, Jul. 2014.

[42] G. Coluccia, C. Ravazzi, and E. Magli, *Compressed Sensing for Distributed Systems*, 1st ed., ser. SpringerBriefs in Electrical and Computer Engineering. Springer Singapore, 2015.

[43] P. Boufounos, "Universal rate-efficient scalar quantization," *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1861–1872, Mar. 2012.

[44] J. Laska and R. Baraniuk, "Regime change: Bit-depth versus measurement-rate in compressive sensing," *IEEE Trans. Signal Processing*, vol. 60, no. 7, pp. 3496–3505, Jul. 2012.

[45] U. Kamilov, V. Goyal, and S. Rangan, "Message-passing de-quantization with applications to compressed sensing," *IEEE Trans. Signal Processing*, vol. 60, no. 12, pp. 6270–6281, Dec. 2012.

[46] A. Shirazinia, S. Chatterjee, and M. Skoglund, "Analysis-by-synthesis quantization for compressed sensing measurements," *IEEE Trans. Signal Processing*, vol. 61, no. 22, pp. 5789–5800, Nov. 2013.

[47] ——, "Distributed quantization for measurement of correlated sparse sources over noisy channels," Jul. 2015, available at http://arxiv.org/abs/1404.7640v2.

[48] M. Leinonen, M. Codreanu, and M. Juntti, "Channel-robust compressed sensing via vector pre-quantization in wireless sensor networks," in *Proc. IEEE Global Conf. on Signal and Inform. Proc.*, Orlando, Florida, USA, Dec. 14–16 2015, pp. 383–387.

[49] ——, "Distributed variable-rate quantized compressed sensing in wireless sensor networks," in *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.,* invited paper, Edinburgh, UK, Jul. 3–7 2016, pp. 1–5.

[50] ——, "Distributed distortion-rate optimized compressed sensing in wireless sensor networks," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1609–1623, Apr. 2018.

[51] C. Chang, "On the rate distortion function of Bernoulli Gaussian sequences," in *Proc. IEEE Int. Symp. Inform. Theory*, Austin, TX, Jun. 13-18 2010, pp. 66–70.

[52] C. Weidmann and M. Vetterli, "Rate distortion behavior of sparse sources," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 4969–4992, Aug. 2012.

[53] L. Palzer and R. Timo, "A lower bound for the rate-distortion function of spike sources that is asymptotically tight," in *Proc. IEEE Inform. Theory Workshop*, Cambridge, UK, Sep. 11–14 2016, pp. 1–5.

[54] A. Fraysse, B. Pesquet-Popescu, and J.-C. Pesquet, "On the uniform quantization of a class of sparse sources," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3243–3263, Jul. 2009.

[55] A. Kipnis, S. Rini, and A. J. Goldsmith, "The indirect rate-distortion function of a binary i.i.d source," in *Proc. IEEE Inform. Theory Workshop*, Jeju Island, Korea, Oct. 11-15 2015, pp. 352–356.

[56] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[57] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[58] Y. Dong, S. Chang, and L. Carin, "Rate-distortion bound for joint compression and classification with application to multiaspect scattering," *IEEE Sensors J.*, vol. 5, no. 3, pp. 481–492, Jun. 2005.

[59] G. Ku, J. Ren, and J. M. Walsh, "Computing the rate distortion region for the CEO problem with independent sources," *IEEE Trans. Signal Processing*, vol. 63, no. 3, pp. 567–575, Feb. 2015.

[60] W. Finamore and W. Pearlman, "Optimal encoding of discrete-time continuous-amplitude memoryless sources with finite output alphabets," *IEEE Trans. Inform. Theory*, vol. 26, no. 2, pp. 144–155, Mar. 1980.

[61] C.-I. Chang and L. D. Davisson, "On calculating the capacity of an infinite-input finite (infinite)-output channel," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1004–1010, Sep. 1988.

[62] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1939–1952, Nov. 1994.

[63] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *IEEE Trans. Inform. Theory*, vol. 50, no. 2, pp. 245–258, Feb. 2004.

[64] R. Gray, "Conditional rate-distortion theory," Stanford Electron. Lab, Stanford University, Tech. Rep., Oct. 1972.

[65] ——, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 480–489, Jul. 1973.

[66] T. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1629–1638, Jun. 2002.

[67] S. Cheng, V. Stankovic, and Z. Xiong, "Computing the channel capacity and rate-distortion function with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4418–4425, Dec. 2005.

[68] B. Guler, E. MolavianJazi, and A. Yener, "Remote source coding with two-sided information," in *Proc. IEEE Int. Symp. Inform. Theory*, Jun. 2015, pp. 2176–2180.

[69] B. Leiner and R. Gray, "Rate-distortion theory for ergodic sources with side information (corresp.)," *IEEE Trans. Inform. Theory*, vol. 20, no. 5, pp. 672–675, Sep. 1974.

[70] D. Goblirsch, "Quantization systems for hidden markov sources," Ph.D. dissertation, University of Maryland, 1989. [Online]. Available: http://drum.lib.umd.edu/handle/1903/4942

[71] T. Linder, R. Zamir, and K. Zeger, "On source coding with side-information-dependent distortion measures," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2697–2704, Nov. 2000.

[72] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York, USA: Cambridge University Press, 2012.

[73] J. Gibson, *Information Theory and Rate Distortion Theory for Communications and Compression*, 1st ed., ser. Synthesis Lectures on Communications. Morgan & Claypool, 2013.

[74] M. Leinonen, M. Codreanu, M. Juntti, and G. Kramer, "Rate-distortion lower bound for compressed sensing via conditional remote source coding," in *Proc. IEEE Inform. Theory Workshop*, Cambridge, UK, Sep. 11–14 2016, pp. 275–279.

[75] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[76] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge University Press, Nov. 2012, ch. 1, pp. 1–64.

[77] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. Springer New York, 2013.

[78] R. Gray, *Source Coding Theory*, ser. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 1990, vol. 83.

[79] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[80] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 3065–3092, May 2012.

[81] A. Oppenheim and G. Verghese, *Signals, Systems and Inference*. NJ, USA: Prentice Hall, 2015.

[82] R. Gallager, *Stochastic Processes: Theory for Applications*. New York, USA: Cambridge University Press, 2013.

[83] M. Elad and I. Yavneh, "A plurality of sparse representations is better than the sparsest one alone," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4701–4714, Oct. 2009.

[84] M. Protter, I. Yavneh, and M. Elad, "Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Signal Processing*, vol. 58, no. 7, pp. 3471–3484, Jul. 2010.

[85] J. S. Turek, I. Yavneh, and M. Elad, "On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Signal Processing*, vol. 59, no. 8, pp. 3526–3535, Aug. 2011.

[86] T. Fine, "Properties of an optimum digital system and applications," *IEEE Trans. Inform. Theory*, vol. 10, no. 4, pp. 287–296, Oct. 1964.

[87] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[88] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[89] P. Chou, T. Lookabaugh, and R. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 1, pp. 31–42, Jan. 1989.

[90] D. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.

[91] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.22," http://cvxr.com/cvx, Aug. 2012.

[92] T. Lookabaugh and R. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1020–1033, Sep. 1989.

[93] R. Gray, "Vector quantization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 1, no. 2, pp. 4–29, Apr. 1984.

[94] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inform. Theory*, vol. 43, no. 1, pp. 38–47, Jan. 1997.

Markus Leinonen (S'11) received his B.Sc. (EE) and M.Sc. (EE) degrees from University of Oulu, Oulu, Finland in 2010 and 2011, respectively. In 2010, he joined Centre for Wireless Communications (CWC) at University of Oulu, where he is currently working toward Dr.Sc. (Tech.) degree. During 2013 and 2014, he has been a guest researcher at Technical University of Munich, Germany. His research interests include compressed sensing, lossy compression, rate-distortion analysis, and distributed cross-layer optimization in energy-efficient data gathering wireless sensor networks.

Marian Codreanu (S'02-M'07) received the M.Sc. degree from the University Politehnica of Bucharest, Romania, in 1998, and the Ph.D. degree from University of Oulu, Finland, in 2007. From 1998 to 2002 he was a Teaching Assistant at the Telecommunications Department of Politehnica University of Bucharest. In 2002 he joined the Centre for Wireless Communications at University of Oulu where he is currently an Adjunct Professor and holds an Academy Research Fellow position. In 2008 he was a visiting postdoctoral researcher at the University of Maryland, College Park, USA. His research interests include optimization, compressed sensing, information theory, and signal processing for wireless communication systems and networks.

Dr. Codreanu received the best doctoral thesis prize within the area of all technical sciences in Finland in 2007. In 2013 he was nominated Academy Research Fellow by the Academy of Finland. He was a Co-Chair of the Technical Program Committee (TPC) of The First Nordic Workshop on Cross-Layer Optimization in Wireless Networks in 2010 and a Co-Chair of the TPC of the Nordic Workshop on System and Network Optimization for Wireless (SNOW) in 2011, 2013, 2014 and 2016. Currently, he is serving as Vice Chair of the IEEE Finland Communications and Information Theory Joint Societies Chapter.

Markku Juntti (S'93-M'98-SM'04) received his M.Sc. (EE) and Dr.Sc. (EE) degrees from University of Oulu, Oulu, Finland in 1993 and 1997, respectively.

Dr. Juntti was with University of Oulu in 1992–98. In academic year 1994–95, he was a Visiting Scholar at Rice University, Houston, Texas. In 1999–2000, he was a Senior Specialist with Nokia Networks. Dr. Juntti has been a professor of communications engineering since 2000 at University of Oulu, Centre for Wireless Communications (CWC), where he leads the Communications Signal Processing (CSP) Research Group. He also serves as Head of CWC – Radio Technologies (RT) Research Unit. His research interests include signal processing for wireless networks as well as communication and information theory. He is an author or co-author in some 450 papers published in international journals and conference records as well as in books *WCDMA for UMTS*, *Signal Processing Handbook* and *5G Wireless Technologies*. Dr. Juntti is also an Adjunct Professor at Department of Electrical and Computer Engineering, Rice University, Houston, Texas, USA.

Dr. Juntti is an Editor of IEEE Transactions on Communications and was an Associate Editor for IEEE Transactions on Vehicular Technology in 2002–2008. He was Secretary of IEEE Communication Society Finland Chapter in 1996–97 and the Chairman for years 2000–01. He has been Secretary of the Technical Program Committee (TPC) of the 2001 IEEE International Conference on Communications (ICC 2001), and the Co-Chair of the Technical Program Committee of 2004 Nordic Radio Symposium and 2006 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2006), and the General Chair of 2011 IEEE Communication Theory Workshop (CTW 2011). He has served as Co-Chair of the Signal Processing for Communications Symposium of Globecom 2014 and IEEE GlobalSIP 2016 Symposium on Transceivers and Signal Processing for 5G Wireless and mm-Wave Systems.

Gerhard Kramer (S'91-M'94-SM'08-F'10) received the Dr. sc. techn. degree from ETH Zurich in 1998. From 1998 to 2000, he was with Endora Tech AG in Basel, Switzerland, and from 2000 to 2008 he was with the Math Center at Bell Labs in Murray Hill, NJ, USA. He joined the University of Southern California, Los Angeles, CA, USA, as a Professor of Electrical Engineering in 2009. He joined the Technical University of Munich (TUM) in 2010, where he is currently Alexander von Humboldt Professor and Chair of Communications Engineering. His research interests include information theory and communications theory, with applications to wireless, copper, and optical fiber networks. Dr. Kramer served as the 2013 President of the IEEE Information Theory Society.