Distributed Reception in the Presence of Gaussian Interference

by

Christian Chapman

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved August 2019 by the
Graduate Supervisory Committee:

Daniel W. Bliss, Chair
Christ D. Richmond
Oliver Kosut
Cihan Tepedelenlioglu

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

An analysis is presented of a network of distributed receivers encumbered by strong in-band interference. The structure of information present across such receivers and how they might collaborate to recover a signal of interest is studied. Unstructured (random coding) and structured (lattice coding) strategies are studied towards this purpose for a certain adaptable system model. Asymptotic performances of these strategies and algorithms to compute them are developed. A jointly-compressed lattice code with proper configuration performs best of all strategies investigated.

DEDICATION

Dedicated to my father, mother and sister. Your support was indispensable.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

iv

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Figure 1.1: High-level sketch of the scenario. Distributed observers seek to recover information about a signal of interest. Observers each broadcast a limited amount of information to each other via a local array network (LAN). In-band interfering signals are present and obscure the signal of interest from each individual observer.

Consider a situation where multiple receivers have the common goal of recovering a broadcast in the presence of much stronger interfering signals. If the listening nodes are able to form a local array network (LAN) among one another, the nodes can mitigate interference and recover the signal of interest by sharing information and processing the messages they share. This scenario could arise, for instance, when a group of local nodes must aid a neighbor's reception, when a group of cellular nodes have the common goal of receiving a code-domain-multiple-access transmission in a crowded environment, or when an adversary jams a broadcast node meant to serve multiple users. A high level model of this scenario is shown in Figure 1.1. When the amount of information transmission among listeners is limited (as it well may be if they must conserve power or bandwidth), they must take care to only forward novel information to their neighbors. Several approaches to the problem are presented
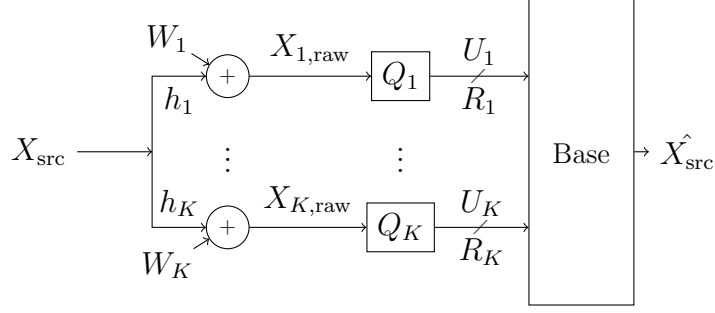
Figure 1.2: A block diagram of the system in question.

culminating in a strategy involving jointly-compressed lattice codes. The strategy is potentially practically realizable.

This scenario is studied in terms of a reduced version of the system where instead of communicating to one another, all listeners instead forward digital messages over a reliable LAN link to a base node (possibly virtual, representing knowledge common to all receivers). Interfering signals are modeled as a correlated Gaussian component added to all observers' discrete-time observations. A detailed description of this model is presented in Section 1.1 along with a rationale for how it may arise in an actual wireless communications setting. A block diagram is shown in Figure 1.2.

The coding strategies presented address either the problem of recovering a signal of interest to low mean-squared-error or the problem of decoding a message modulated into the signal of interest. Though closely related, the two problems are theoretically distinct as the decoding problem imposes message structure on the source's transmission which can be exploited at the observers.

1.1 System Model

This section presents the mathematical model for the problem studied and an example of how the model might arise in the scenario just described. A list of mathematical notation used is given in Table 2.1. A message is modulated and broadcast

as a real source X_{src} with power constraint $\mathbb{E}[\|X_{\mathrm{src}}\|^2] \leq 1$. The source is observed by K receivers as $X_{1,\mathrm{raw}}, \ldots, X_{K,\mathrm{raw}}$, where

$$X_{k,\mathrm{raw}} = h_k X_{\mathrm{src}} + W_k, \ \ k = 1, \ldots, K, \tag{1.1}$$

$\vec{h} \in \mathbb{R}^K$ and noise terms $\vec{W}_{[K]}$ are multivariate Gaussian $\mathcal{N}(0, \Sigma_{\mathrm{noise}})$, noise covariance $\Sigma_{\mathrm{noise}} \in \mathbb{R}^{K \times K}$ positive definite. Vectors formed by n repeated uses of this channel are indicated with superscript. Each of the K receivers forward information to a base node at rate $R_k > 0$ through a side channel.

Definition 1. A source-to-base communication rate R is said to be *achievable* if for any $\varepsilon > 0$ then for large enough n the following objects can be constructed to satisfy a low-error-probability condition:

- Modulation scheme $\phi : [2^{n(R-\varepsilon)}] \to \mathbb{R}^n$ with

$$2^{-n(R-\varepsilon)} \cdot \sum_m \frac{1}{n} \|\phi(m)\|^2 < 1$$

- Relay encoding functions $(\mathrm{enc}_k)_{k \in [K]}$, $\mathrm{enc}_k : \mathbb{R}^n \to [2^{nR_k}]$

- Base decoder $\mathrm{dec} : \prod_{k \in [K]} [2^{nR_k}] \to [2^{n(R-\varepsilon)}]$.

The low-error-probability condition is as follows. Taking:

$$M \sim \mathrm{unif} \, [2^{n(R-\varepsilon)}],$$

$$X_{\mathrm{src}}^n := \phi(M),$$

$$\hat{M} := \mathrm{dec} \left(\mathrm{enc}_1(X_{1,\mathrm{raw}}^n), \ldots, \mathrm{enc}_K(X_{K,\mathrm{raw}}^n) \right),$$

then $\mathbb{P}\left(\hat{M} \neq M \right) < \varepsilon$.

Definition 2. For a source X_{src}^n with each component i.i.d. $\mathcal{N}(0, 1)$, a mean-squared-error distortion D is said to be *achievable* if for any $\varepsilon > 0$ then for large enough n the following objects can be constructed to satisfy a low-error-probability condition:

3

- Relay encoding functions $(\text{enc}_k)_{k \in [K]}$, $\text{enc}_k : \mathbb{R}^n \to [2^{nR_k}]$

- Base decoder $\text{dec} : \prod_{k \in [K]} [2^{nR_k}] \to \mathbb{R}^n$.

The low-error-probability condition is as follows. Taking:

$$\hat{X_{\text{src}}}^n := \text{dec}\left(\text{enc}_1(X_{1,\text{raw}}^n), \ldots, \text{enc}_K(X_{K,\text{raw}}^n)\right),$$

then $\mathbb{P}\left(\frac{1}{n}\|X_{\text{src}}^n - \hat{X_{\text{src}}}^n\|^2 > D\right) < \varepsilon$.

1.1.1 Model Ontology

This section describes how the model above may arise among wireless radios.

LAN

The LAN link between observers is a side channel over which each observer can broadcast and recover fellow observers' messages. It may correspond to some multiple-access frequency band unencumbered by interfering sources. Redistribution among observers of bandwidth or time-slot resource along such a band corresponds to redistribution of the receiver-to-base bitrates $R_1, \ldots, R_K$. The LAN has been modeled such that for each channel use, each receiver forwards a fixed number of bits to the base. Each observer may forward to the base at some predetermined rate, or observers may adapt their rates within LAN resources. Both situations are considered throughout the study.

Jointly-Gaussian Discrete-Time Observations

Take the source of interest to be a single-antenna broadcaster modulating its information into a train of orthogonal-frequency-division-multiplexing (OFDM) symbols as discussed in (Bliss and Govindasamy, 2013, §10.5.3). Make the following assumptions:

- Interferers broadcast white Gaussian noise over subcarrier bands, and this noise is independent of the source's modulation at each subcarrier.

- The channels between broadcasters and observers have approximately flat frequency response within each subcarrier band and many OFDM symbols fit within these channels' coherence times.

- Each broadcaster's signal is observed with roughly equal Doppler shift at all the observers[1]

Now the ensemble of observer responses at a particular subcarrier frequency, averaged over each symbol duration, are approximately represented by a complex version of the given model: the source X_{src} now takes on complex values within its unit power constraint, the real multivariate normal noise distribution in Equation (1.1) is replaced with a circularly-symmetric complex multivariate normal, and channel coefficients $\vec{h}$ are complex.[2] In the sequel the problem of coding for the real or complex component of an individual subcarrier is considered. Adaptations of the main theorems to the complex case can be obtained by decomposing the channel into a real one with the proper symmetries. This is done in Appendix F. The complex case is not treated directly since most lattice theory results and intuitions are in terms of real geometry.

---

[1]The Doppler and slow channel evolution assumptions can possibly relaxed although the problem of designing robust codes is considerably complicated. Avenue for further study of these aspects is outlined in the conclusion, Section 6.1.4.

[2]Vector $\vec{X}_{\Re} + j\vec{X}_{\Im}$ is *circularly-symmetric complex multivariate normal* about 0 with positive-semidefinite covariance $\mathbf{M} \in \mathbb{C}^{K \times K}$ if $(\vec{X}_{\Re}, \vec{X}_{\Im}) \sim \mathcal{N}(0, \begin{bmatrix} \Re\{\mathbf{M}\} & -\Im\{\mathbf{M}\} \\ \Im\{\mathbf{M}\} & \Re\{\mathbf{M}\} \end{bmatrix})$

Channel State

All coding schemes studied presume access to perfect channel state information at all observe nodes. That is, observe nodes have access to the tuple

$$(\vec{h}, \mathbf{\Sigma}_{\text{noise}}),$$

quantities averaged over many OFDM symbols. Initial numerical investigations (presented later) suggest existence of configurations of the investigated schemes which are tolerant to poor channel state estimates. However this has not been deeply studied and presents an avenue for further research discussed in Chapter 6.

One method by which channel state estimates can be made available to observers is as follows, assuming source and observe nodes are synchronized enough that each observer properly captures each transmitted OFDM symbol. Take the first few OFDM symbols as a training sequence of which all observers are informed. Observers repeat quantizations of the observed phase and amplitude at each subcarrier frequency for each training symbol over the LAN. These quantizations can be used to build an estimate of the channel state.

1.2   Background

This section provides context for existing work on the problem.

The problem of finding the best achievable mean-squared-error distortion as given in Definition 2 has been studied as the *quadratic Gaussian many-help-one distributed source coding* problem Tavildar and Wagner (2009). Strategies for the source coding problem (Definition 2) can be adapted to yield an achievable source-to-base rate for the communication problem (Definition 1) by first performing a source coding stage at the encoders and base, then channel coding on the resulting point-to-point channel

from source to base. This adaptation is sub-optimal in general since the source coding stage as described may not take advantage of the source's codebook structure.

Both problems are similar in structure to the CEO problem Berger *et al.* (1996). The CEO problem is of structure similar to the present network: distributed observers provide a base node side information at some rate about a source of interest which the base attempts to estimate. Unfortunately it is unusual for studies on the CEO problem to take noise across observers as correlated. For this reason, most studies on the CEO problem cannot directly be adapted to an effective strategy for the present problems.

1.2.1   Unstructured Coding Strategies

By 'unstructured coding' it is meant that codebooks involved are populated by drawing codewords from a distribution and do not have any designed internal algebraic structure. Much work on unstructured codes for similar networks can be adapted to the present problems. For some covariance structures, the best possible performance of the source coding problem in Definition 2 is such a coding technique Tavildar and Wagner (2009), but in general much better performance is possible. The best possible performance for the case of two observing nodes is bounded in Maddah-Ali and Tse (2010).

The major study on the application unstructured codebooks for the communications problem in Definition 1 for *general channel* is Sanderovich *et al.* (2008). A specialization to the present channel is Chapman *et al.* (2018) and the main results of this study are the topic of Chapter 3. Although the studies treat the communications problem directly, their analytic strategies are still fundamentally built off of (and limited by) a distributed source coding perspective and do not directly exam-

7

ine performance gains that can be gotten by imposing joint-structure on source and observer codebooks. This technicality is examined further in Chapters 2 and 3.

Unstructured coding schemes for the source-coding problem are known, although given results from studies on structured codes like those in Wagner (2011), Nazer and Gastpar (2008), Krithivasan and Pradhan (2011) and Lim *et al.* (2018) it seems likely that strategies imposing more structure on the codebooks can outperform unstructured strategies.

1.2.2 Lattice-Based Strategies

Many recent developments in structured distributed source coding for Gaussian channels involve lattice structures. Lattices have many useful properties which make them attractive structures for building distributed source codes. Throughout this document, terminology and basic lattice theory results are taken from Zamir (2014), and the necessary concepts are introduced in proper in Sections 2.4, 2.5. A *lattice* in the context of this study is a collection $\Lambda \subset \mathbb{R}^n$ of separated vectors closed under integer linear combinations. A lattice Λ can be endowed with a *base region* $B \subset \mathbb{R}^n$, which is any set with zero moment where $(B + \lambda)_{\lambda \in \Lambda}$ is a partition of $\mathbb{R}^n$. Fixing Λ and B, a modulo-Λ operation can be defined as follows, for any $a \in \mathbb{R}^n$:

$$\mathrm{mod}_\Lambda(a) := a + \lambda^*$$

where $\lambda^* \in \Lambda$ is the unique choice so that $a + \lambda^* \in B$. It is straightforward to demonstrate that this operation satisfies the following relation:

$$\mathrm{mod}_\Lambda \left[\mathrm{mod}_\Lambda(a) + \mathrm{mod}_\Lambda(b) \right] = \mathrm{mod}_\Lambda(a + b). \tag{1.2}$$

This property is useful for distributed compression of correlated signals: say one observer receives a and another receives b, and the base seeks to recover $(a + b)$. Say

a and b correlated so that the variance of $(a + b)$ is much smaller than that of either a or b. If Λ and B are designed so that $(a + b)$ probably lies within B (i.e. with high probability $\text{mod}_\Lambda(a + b) = a + b$) then instead of quantizing the entire observation, each observer can instead quantize the *modulo* of its observation, and the decoder can still recover its component of interest through Equation (1.2). In many cases the modulo of the signal will have much smaller dynamic range, making quantization fidelity cheaper. The decoder can recombine the quantized modulo results using the above identity to recover $(a + b)$ with better precision than available had the relays not performed this modulo operation.

The MIMO processing methods presented in Ordentlich and Erez (2017), extended in He and Nazer (2016)[3] are closely related to the present scenario. These papers examine application of a lattice code exploiting properties such as those described above to a slightly different problem: design a lattice code such that the decoder recovers an estimate for all the observers' signals, not just a single component of interest within them. Relaxing this constraint allows for a larger rate region for the many-help-one problem investigated in Chapter 4. Lattice-coding-based strategies perform well for similar problems and in certain settings provide optimal performance or performance strictly better than known unstructured codebook strategies Zamir *et al.* (2002), Nazer *et al.* (2016), (Zamir, 2014, Chapter 12), Cheng *et al.* (2018). A prominent development in lattice-based distributed compression called Integer Forcing Source Coding (IFSC) Ordentlich and Erez (2017) deals with a variation of the present problem where multiple users have a message for a decoder informed by distributed antennas. Lattice coding schemes can be designed for MIMO communication to have

---

[3]Extensions appears in conference proceedings in dates before the initial results since initial results were made available in arχv.org some time before their appearance in a Journal.

universality properties such that a transmitter need not know the channel Campello *et al.* (2018), Ordentlich and Erez (2014).

1.2.3   Second-Order Lattice-Based Strategies

Lattice-modulo-encoded messages such as those described in the previous often carry redundant information. Reference Wagner (2011) provides an upper and lower bound on conditional entropies between two such messages. Reference Yang and Xiong (2011) realized an according compression scheme for such encodings using further lattice processing on them in the context of an insightful 'coset planes' abstraction. It was further noticed in Yang and Xiong (2014) that improvement towards the many-help-one problem is gotten by splitting helper messages into two parts: one part a coarse quantization of the signal, compressed across helpers via Slepian-Wolf joint-compression (these message parts corresponding to the 'high bit planes'), and another a lattice-modulo-encoding representing signal details (corresponding to 'low bit planes'). Chapter 5 extends these ideas to a general quantity of helpers, providing upper and lower bounds for the joint-entropy of such encodings. It also treats a case where a component of the sources has lattice structure.

A joint-compression scheme for lattice encodings called 'Generalized Compute Compress and Forward' was introduced in Cheng *et al.* (2018) towards coding for a multi-user additive white Gaussian noise channel where a decoder seeks to recover all user's messages and is informed by helpers. The scheme exploits the same idea from Yang and Xiong (2014), this time splitting lattice encodings into an arbitrary amount of messages at different coset planes. The scheme in Chapter 5 work follows along the same lines, although for a network with one user and where many interferers without codebook structure are also present.

1.2.4 Other Techniques

Other bodies of work that deal with distributed source coding with this problem structure but are not deeply investigated in this study include:

- Joint-compression schemes for discrete sources, such as Stankovic *et al.* (2006) and Shirani *et al.* (2016). When the relays first quantize their broadcast and then use these codes, their performance is subsumed by the ones presented in Section 2.3.

- Distributed source coding using syndromes Pradhan and Ramchandran (1999), a general approach to practical distributed source coding. Choi *et al.* (2015) and related works develop relay processing strategies to preserve specific constellations patterns (BPSK, etc.) and investigate coded bit error rates after syndrome fusion.

- Distributed lossy arithmetic coding, where distributed sources use arithmetic codes with overlapping intervals. Its theory is developed in detail in (Wang *et al.*, 2017, Chapter 6).

1.3 Overview and Contributions

Chapter 2 covers the necessary technical results used throughout the rest of the study. Chapter 3, based on Chapman *et al.* (2018) provides the best known bounds on the system's achievable rate using unstructured coding schemes. Chapter 4, based on Chapman *et al.* (2019) discusses a lattice coding strategy for the problem. Chapter 5, based on Chapman and Bliss (2019) examines inefficiencies in the strategy from Chapter 4, yielding description of a performance-improving joint-compression stage. Chapter 6 describes viable directions for improvement and future study.

Considered jointly with time-sharing, performance of the strategies introduced by this study are the strongest currently known for their scenarios. It is unknown whether the achievable communications rates for the strategy described in Chapter 5 subsume those given in Chapter 3. In contrast the source coding strategy described in Chapter 5 does subsume the one described in Chapter 3.

Mathematical notation is described in Table 2.1. Preliminaries for each section as well as a local description of variables is given in a table at the beginning of each chapter. Long proofs are delayed to appendices.

TECHNICAL PRELIMINARIES

This chapter introduces some technical concepts and constructions used in the remainder of the document. A table of general mathematical notation is given in 2.1.

$a := b$	Define a to equal b
$[n]$	Integers from 1 to n
$\log(\cdot)$	Logarithm, base-2
$\boldsymbol{A}, \boldsymbol{a}, \vec{a}, \vec{A}$	Matrix, column vector, vector, random vector
$j; \Re, \Im$	Complex unit; real, imaginary component
$\boldsymbol{A}^\dagger, \boldsymbol{a}^\dagger$	Conjugate transpose
$[\boldsymbol{A}]_{S,T}$	Submatrix corresponding to rows S, columns T of $\boldsymbol{A}$
$\vec{Y}_S$	an $|S|$-vector, the sub-vector of $\vec{Y}$ including components with indices in S. If S has order then this vector respects S's order.
$\mathbf{I}_K$	$K \times K$ identity matrix
0_K	$K \times 1$ zero vector
$\operatorname{diag} \vec{a}$	Square diagonal matrix with diagonals $\vec{a}$
$\operatorname{pinv}(\cdot)$	Moore-Penrose pseudoinverse
$\mathcal{N}(0, \boldsymbol{\Sigma})$	Normal distribution with zero mean, covariance $\boldsymbol{\Sigma}$
$X \sim f$	X is a random variable distributed like f
$X^n, f(x^n)$	Vector of n independent trials of a random variable distributed like X, a function interpreted to take inputs such as the variable just discribed

$\mathbb{E}[X]$	Expectation of X
$\text{var}(a)$	Variance (or covariance matrix) of (components of) a (averaged over time index if applicable).
$\text{var}(a\|b)$	Conditional variance (or covariance matrix) of (components of) a given observation b (averaged over time index if applicable).
$\text{cov}(a,b), \text{cov}(a,b\|c)$	Covariance between a and b,, covariance between a and b conditioned on c (averaged over time index if applicable).
$\mathcal{E}(a\|b)$	Linear MMSE estimate of a given observations b
$\mathcal{E}_\perp(a\|b)$	Complement of $\mathcal{E}(a\|b)$, i.e., $\mathcal{E}_\perp(a\|b) := a - \mathcal{E}(a\|b)$. An important property is that $\mathcal{E}(a\|b)$ and $\mathcal{E}_\perp(a\|b)$ are uncorrelated.
$H(X\|Y),\ h(X\|Y),\ I(X;Y)$	Shannon entropy, differential entropy, mutual information
$\text{round}_L(\cdot), \text{mod}_L(\cdot)$	Lattice round, modulo to a lattice L (when it is clear what base region is associated with L).

Table 2.1: Symbols and Notation

2.1 Basic Information Theorems

2.2 Slepian-Wolf Lossless Joint-Compression

The encoder rates needed for lossless distributed compression of discrete sources is given by the Slepian-Wolf theorem (El Gamal and Kim (2011)).

Theorem 1. *Take $\vec{S}^n = (S_1^n, \ldots, S_K^n)$ to be n independent identically distributed trials of a random source $(S_1, \ldots, S_K)$ with distribution $P_{\vec{S}}$ over finite support $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_K$. 'Compression rates' $R_1, \ldots, R_K > 0$ are said to be 'achievable' if any*

$\varepsilon > 0$ *has* n *large enough so that 'encoder' maps* $\mathrm{enc}_k : \mathcal{S}_k^n \to [2^{nR_k}], \ k = 1, \ldots, K$ *and a 'decoder' map* dec *exist with*

$$\mathbb{P}(\mathrm{dec}(\mathrm{enc}_1(S_1^n), \ldots, \mathrm{enc}_K(S_K^n)) \neq \vec{S}^n) \leq \varepsilon.$$

Compression rates $R_1, \ldots, R_K > 0$ *are achievable if and only if:*

$$\sum_{k \in T} R_k > H(S_T | S_{[K] \setminus T}) \ \forall T \subset [K]$$

A common method for demonstrating achievability of the bound is to select each encoder uniformly at random from the collection of possible encoding maps. Provided rates are as large as prescribed, then for long enough blocklength it is unlikely that any source vector in the inverse map of the encodings is typical according to the source's distribution, other than the source vector itself. This strategy is known as 'random binning.' Unfortunately, practically realizable encoding maps are often difficult to realize in practice.

2.3   Berger-Tung Lossy Joint-Compression

The Berger-Tung inner and outer bounds give approximate rate-distortion regions for a closely related problem of lossy distributed source coding where each observer's reception must be recovered at the base to within some distortion criterion. In particular they deal with the following problem (stated for two encoders):

> A source $(X_{1,t}, X_{2,t})$ is i.i.d. in time t. Receiver i $(i = 1, 2)$ observes X_i T times as $(X_{i,t})_{t \in [T]}$ and processes it through an encoder f_i into a message $M_i = f_i((X_{i,t})_{t \in [T]})$ with rate R_i (equivalently $H(M_i) \leq TR_i$). Receiver i forwards M_i to a base. The base processes receiver messages $(M_i)_{i=1,2}$ through decoders $(g_i)_{i=1,2}$ into estimates
>
> $$(g_i((M_j)_{j=1,2}))_i = ((\hat{X}_{i,t})_{t \in [T]})_{i=1,2}.$$

For distortion measures $d_1(x_1, \hat{x}_1), \; d_2(x_2, \hat{x}_2) \geq 0$, what rate-distortion pairs $(R_1, D_1), \; (R_2, D_2)$ are achievable?

For distortion measures $d_1(\cdot), d_2(\cdot)$, then a rate-distortion pair ((R_i, D_i))$_{i=1,2}$ is said to be "achievable" if for any $\varepsilon > 0$ and T large enough then encoders $(f_i)_{i=1,2}$ and decoders $(g_i)_{i=1,2}$ exist so that for $i = 1, 2$, receiver i forming $M_i = f_i((X_{i,t})_{t \in [T]})$ and base forming $(\hat{X}_{i,t})_{t \in [T]} = g_i((M_j)_{j=1,2})$,

- $H(M_i) \leq TR_i$,

- $P\left(\frac{1}{T} \sum_{\ell=1}^{T} d_i(X_{i,\ell}, \hat{X}_{i,\ell}) \geq D_i \right) < \varepsilon$.

Berger and Tung provide an achievable rate region and an outer bound for this problem (El Gamal and Kim (2011)):

Theorem 2. (Berger-Tung inner bound) $(R_i, D_i)_{i=1,2}$ *is achievable if there are random variables* (U_1, U_2, Q) *where:*

- $R_1 > I(X_1; U_1 | U_2, Q)$,

- $R_2 > I(X_2; U_2 | U_1, Q)$,

- $R_1 + R_2 > I(X_1, X_2; U_1, U_2 | Q)$,

- *Some decoder functions* $g_i^*(u_1, u_2, q)$ *have* $d_i(X_i, g_i^*(U_1, U_2, Q)) < D_i$ *for* $i = 1, 2$,

- *The PMF of* (X_1, X_2, U_1, U_2, Q) *can be factored:* $P_{X_1, X_2} \cdot P_Q \cdot P_{U_1 | X_1, Q} \cdot P(U_2 | X_2, Q)$

(It suffices to check U_i *with* $\text{Image}(U_i) \leq \text{Image}(X_i) + 4$*)*

One strategy for achieving the Berger-Tung inner bound is to first form quantizations of the observed variables (quantization possibly piloted by additional artificial

commonly known side-information Q), then jointly compress these quantizations using the Slepian-Wolf theorem described above. The inner bound is deficient in general as described below.

Theorem 3. (Berger-Tung outer bound)*: For $(R_i, D_i)_{i=1,2}$ to be achievable, then there must be some random variables (U_1, U_2) which satisfy:*

- $R_1 > I(X_1, X_2; U_1 | U_2)$,

- $R_2 > I(X_1, X_2; U_2 | U_1)$,

- $R_1 + R_2 > I(X_1, X_2; U_1, U_2)$,

- *Some decoder functions $g_i^*(u_1, u_2)$ have $d_i(X_i, g_i^*(U_1, U_2)) < D_i$ for $i = 1, 2$,*

- *The tuple (X_1, X_2, U_1, U_2) satisfies Markov conditions: $U_1 \leftrightarrow X_1 \leftrightarrow X_2$ and $X_1 \leftrightarrow X_2 \leftrightarrow U_2$.*

2.3.1   Sub-optimality of Berger-Tung Inner and Outer Bounds

The inner and outer Berger-Tung bounds are not tight in general for their source coding problem. The specific reason for this is discussed briefly in (El Gamal and Kim, 2011, Section 12.5) and we expound on it here. Because of the correlation between X_1 and X_2, Receiver 1's observation of X_1 contains some information about the other source's realization, X_2. This means Receiver 1 may have some information about the result of Receiver 2's processing on X_2. In particular Receiver 1 may be able to derive some knowledge of U_2, Receiver 2's message content. It can use this knowledge of U_2's realization to improve its own message content U_1.

The inner bound's PMF factoring restriction is too strong. It enforces that Receiver 1's message be independent of all the other receivers' messages after conditioning on Receiver 1's observation. This disallows Receiver 1 from using its knowledge

of Receiver 2's message realization to improve its own. An example where this helps is given in (El Gamal and Kim, 2011, Section 12.5).

The outer bound's Markov constraint is too lax. It allows for arbitrary dependence between messages after conditioning on their respective receiver's observations, even when such dependence is not possible. For example, if X_1, X_2 are jointly Gaussian then the Markov constraint allows $U_1 = X_1 + N_1 + N_c$, $U_2 = X_2 + N_c$ where N_1, N_c are independent Gaussian distortions. Indeed, one can verify that the Markov conditions $U_2 \to X_2 \to X_1$ and $U_1 \to X_1 \to X_2$ are satisfied and that noise powers can be made for the outer bound's entropy conditions to hold, but clearly this scheme can't be realized outside some trivial edge cases.

The discrepancy between the inner and outer bound is that depending on the channel, Receiver 1 varies in its ability to observe what specific parts of its source are lost or retained during Receiver 2's lossy reduction of X_2 into U_2. (And vice versa for Receiver 2).

There is further loss in Berger-Tung source coding in its application to the problems in Definitions 1, 2 due to the strategy's aspect of recovering all the receiver's observations rather than only the component of interest. This is examined further in Chapter 3.

2.4   Lattices

Definition 3. A *lattice* $L \subset \mathbb{R}^n$ is a countably infinite set of discrete points closed under addition and subtraction characterized by a *lattice basis*, a finite collection $\{\ell_1, \ldots, \ell_k\} \subset \mathbb{R}^n$ of not-linearly-dependent vectors whose integer linear combinations span L. The dimension of L is the size of the basis.

The dimension of lattices described in this study will have the dimension of their underlying space unless explicitly mentioned.
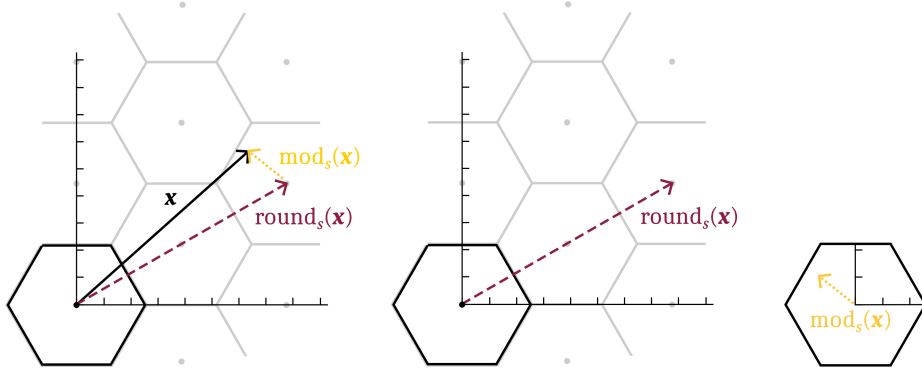
Figure 2.1: A graphical example of lattice modulo and rounding operations on a vector $x \in \mathbb{R}^2$ for a particular lattice (drawn as dots) and hexagonal base region s (outlined in black, translates outlined by solid gray lines). Notice $x = \text{round}_s(x) + \text{mod}_s(x)$, that the image of $\text{round}_s(\cdot)$ is the lattice, and that the image of $\text{mod}_s(\cdot)$ is s.

Definition 4. A region $s \subset \mathbb{R}^n$ is a *base region* for a lattice L if the sets $(l + s)_{l \in L}$ are all disjoint, their union forms $\mathbb{R}^n$, and if s has its moment at the origin.

One can define modulo and rounding operations relative to a base region s for a lattice L:

$$\text{round}_s : \mathbb{R}^n \to L,$$

$$\text{round}_s(x) := l \in L \text{ where } (x - l) \in s,$$

$$\text{mod}_s : \mathbb{R}^n \to s,$$

$$\text{mod}_s(x) := x - \text{round}_s(x).$$

Note that round_s is well defined since s being a *base region* implies there is one and only one lattice point l that satisfies its prescription. These operations have some useful properties. A graphical example of the lattice round and modulo operations for a lattice in $\mathbb{R}^2$ and a hexagonal base region s is shown in Figure 2.1.

Property 1. (mod is the identity within s) $x \in s$ *has* $\text{mod}_s(x) = x$.

Proof: $\text{mod}_s(x) = x - \text{round}_s(x) = x$. $\square$

Property 2. $x \in \mathbb{R}^n, l \in L$ *have* $\text{round}_s(x + l) = \text{round}_s(x) + l$.

Proof: $(x+l) - (\text{round}_s(x)+l) = (x - \text{round}_s(x)) \in s$. So by definition $\text{round}_s(x+l) = \text{round}_s(x) + l$. $\square$

Property 3. $x \in \mathbb{R}^n, l \in L$ *have* $\text{mod}_s(x + l) = \text{mod}_s(x)$.

Proof: $\text{mod}_s(x+l) = (x+l) - \text{round}_s(x+l) = (x+l) - \text{round}_s(x) - l = x - \text{round}_s(x) = \text{mod}_s(x)$. The second equality is by Property 2. $\square$

Property 4. (Lattice modulo is distributive) $x, y \in \mathbb{R}^n$ *have:*

$$\text{mod}_s(\text{mod}_s(x) + y) = \text{mod}_s(x + y).$$

Proof: $\text{mod}_s(\text{mod}_s(x) + y) = \text{mod}_s(x + y - \text{round}_s(x)) = \text{mod}_s(x + y)$. The last equality is by Property 3. $\square$

2.5   Asymptotic Lattice Constructions

Lattices in high dimension can have useful properties for coding problems. These properties are described in terms of a sequence of lattices increasing in dimension. The essential properties exploited in this study are presented in this section using terminology slightly adapted from Zamir (2014). The definitions in this section have many equivalent characterizations, and only the characterizations most immediately applicable to their utility here are presented.

Definition 5. A sequence of lattices $L^{(n)} \in \mathbb{R}^n$ with base regions $B^{(n)}$ are said to be *good for quantization* (synonymous with *Rogers-good*) with mean-squared-error distortion σ^2 if for $X \sim \text{unif } B^{(n)}$ and any $\varepsilon > 0$, then

$$\mathbb{P}(X \notin B^{(n)}(0, \sqrt{n(\sigma^2 + \varepsilon)})) \to_n 0.$$

Definition 6. A sequence of lattices $L^{(n)} \in \mathbb{R}^n$ with base regions $B^{(n)}$ are said to be *good for coding* (synonymous with *Penrose-good*) in noise of power below σ^2 if for any $\varepsilon > 0$ and X uniform in $B(0, \sqrt{n(\sigma^2 - \varepsilon)})$, then

$$\mathbb{P}(X \notin B^{(n)}) \to_n 0$$

Definition 7. For lattices $L_c^{(n)}, L_f^{(n)}$ with $L_c^{(n)} \subset L_f^{(n)}$ with respective base regions $B_c^{(n)}, B_f^{(n)}$ (subscript c standing for 'coarse' and f 'fine'), then if

$$\frac{1}{n} \log(|B_c^{(n)} \cap L_f^{(n)}|) \to_n r$$

then $L_f^{(n)}$ is said to have *nesting rate r* in $B_c^{(n)}$.

The nesting rate characterizes the highest possible entropy-rate for a signal that has support $B_1^{(n)} \cap L_2^{(n)}$. A pictoral example for a two-dimensional lattice is shown in Figure 2.2.

A construction essential to the lattice codes developed in this study is provided in Ordentlich and Erez (2016):

Theorem 4. (Ordentlich and Erez, 2016, Theorem 2) *For any $\sigma_1^2 > \sigma_2^2 > \cdots > \sigma_K^2 > 0$ there exists a sequence of nested lattices:*

$$L_1^{(n)} \subset \cdots \subset L_K^{(n)} \subset \mathbb{R}^n$$

with respective base regions $B_1^{(n)}, \ldots, B_K^{(n)} \subset \mathbb{R}^n$ (formed by their lattice's Voronoi partition) with the following properties:

- *$L_k^{(n)}$ is good for coding in noise of power below σ_k^2.*

- *$L_k^{(n)}$ is good for quantization with mean-squared-error distortion σ_k^2.*

- *For $k < m$ then $L_m^{(n)}$ has nesting rate $\frac{1}{2} \log \frac{\sigma_k^2}{\sigma_m^2}$ in $B_k^{(n)}$.*
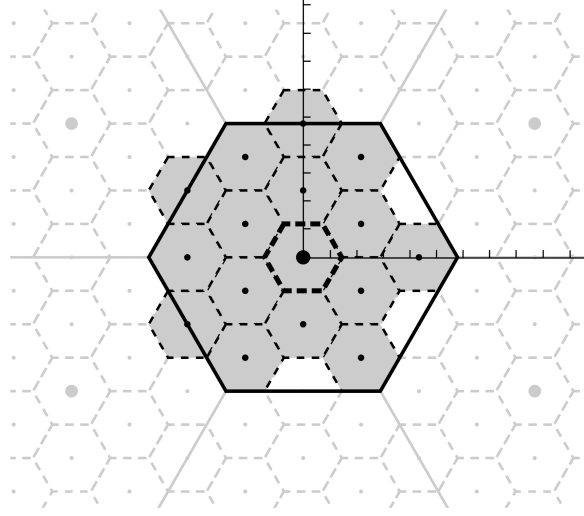
Figure 2.2: An example of nested lattices in $\mathbb{R}^2$. Points from a coarse lattice L_c are drawn as large dots. Points from a fine lattice $L_k \supset L_c$ besides those in L_c are drawn as small dots. A base region B_c for L_c and its transaltes $B_c + L_c$ are outlined by solid lines, and a base region B_k for L_k and its translates $B_k + L_k$ are outlined by dashed lines. The base regions for points in $L_k \cap B_c$ are shaded. Notice that $L_k \cap B_c$ has 2^4 points, so 4 bits are required to describe a general point in $L_k \cap B_c$, or 2 bits per real sample.

Loosely, such a construction is gotten by forming the generator matrix for a random block code over a large finite field, embedding its rows in real space and using the resulting vectors as a lattice basis.

Chapter 3

UNSTRUCTURED SCHEMES

This chapter presents some unstructured coding schemes to bound the problem's achievable rate. These bounds are compared in Figure 3.1.

3.1 Cut-set Upper Bound

The achievable rate is upper bounded by the relaxation of the problem where any collection S of relays can provide unlimited information to the decoder. On the other hand, no more than a total of $\sum_{k \in S} R_k$ bits of information can be provided to the decoder by the relays in $S \subset [K]$, so the achievable rate cannot be greater than

$$R_{CS} := \min_{S \subset [K]} I(X_{\mathrm{src}}; (X_{k,\mathrm{raw}})_{k \in S})) + \sum_{k \in S^C} R_k,$$

This upper bound is an application of the *cut-set bound* described in (El Gamal and Kim, 2011, Chapter 20).

3.2 Plain Quantization

One strategy for each observer would be to encode its reception with an approximately-rate-distortion optimal source coder for its observed source at the prescribed observer rate R_k. This yields an achievable rate Chapman *et al.* (2018):

$$R_{QF} := I(X_{\mathrm{src}}; (X_{k,\mathrm{raw}} + Z_k)_{k \in [K]}), \quad Z_k \sim \mathcal{N}(0, \mathrm{var}\, X_{k,\mathrm{raw}}/(2^{2R_k} - 1)) \qquad (3.1)$$

and an achievable distortion

$$D_{QF} := \mathrm{var}(X_{\mathrm{src}} | (X_{k,\mathrm{raw}} + Z_k)_{k \in [K]}), \quad Z_k \sim \mathcal{N}(0, \mathrm{var}\, X_{k,\mathrm{raw}}/(2^{2R_k} - 1)).$$
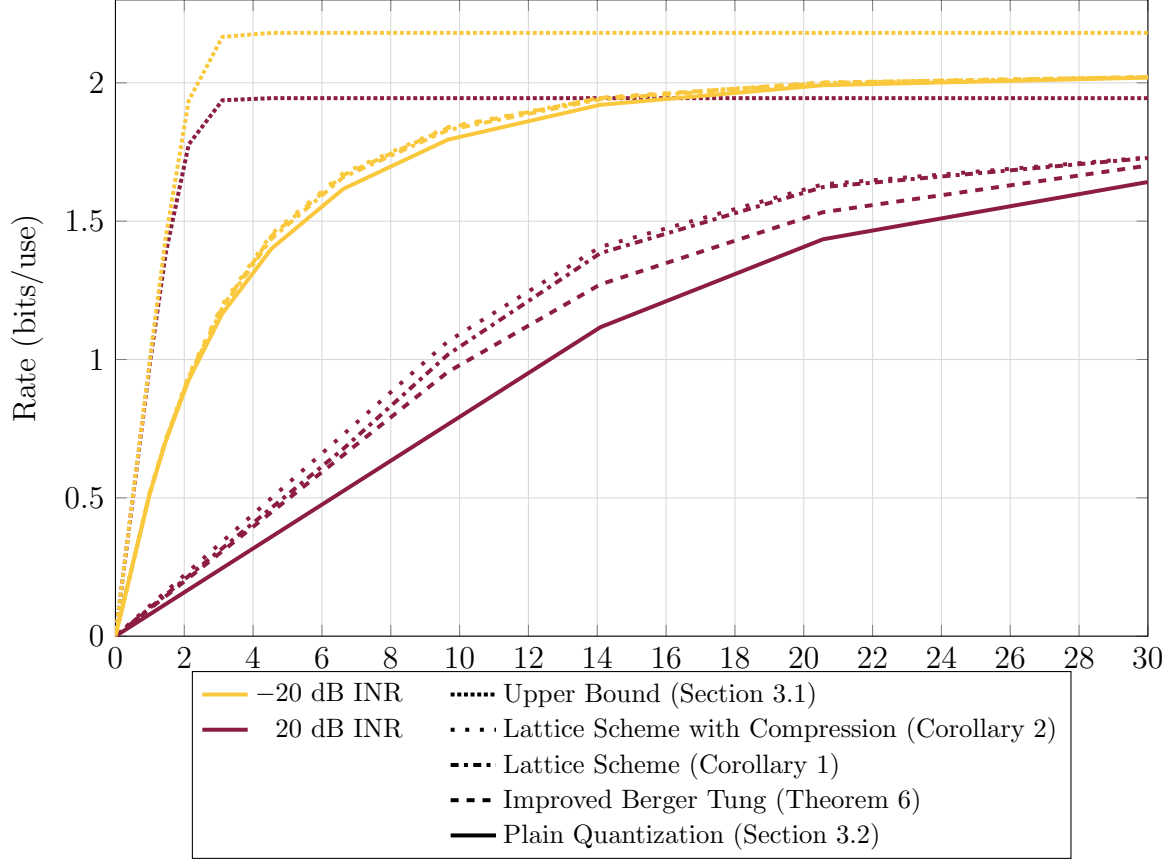
Figure 3.1: Bound performance versus combined relay rate (that is, the total rate available to hub from all the helpers combined). 4 relays with 0 dB average receiver SNR, averaged over 200 channels with gains $\sim \mathcal{CN}(0, I)$. A single-rank interferer present at each receiver with a uniform random phase and ± 20 dB INRs. Single-user quantize-and-forward is most greatly affected by interference (Compare the Broadcast curves from -20 dB to 20 dB INR). Distributed compression offers the most significant benefits over Gaussian compress-and-forward in strong interference and when relay-to-base resources are scarce. Lattice compression schemes (curves for Corollaries 1, 2) are covered in Chapters 4, 5. Computational details are provided in Appendix B.2

Source encoders which operate near the rate-distortion limit for Gaussian sources tend to be very computationally expensive. One strategy of reasonable complexity which comes within half a bit of this limit is to first bin each sample and then perform arithmetic coding on sequences of binned outputs. This is discussed in more detail in Popat (1990).

3.3 Berger-Tung-Based Strategies

The strategy used to achieve the Berger-Tung inner bound discussed in Section 2.3 can be adapted to the present problem.

Theorem 5. (Berger-Tung with Gaussian distortion for distributed receive, (Chapman *et al.*, 2018, Theorem 2)) *Fix non-negative numbers* $\sigma_1^2, \ldots, \sigma_K^2$ *so that the following conditions hold. Defining* $U_k^* \triangleq X_{k,raw} + W_{Q,k}$ *and* $W_{Q,k} \sim \mathcal{N}(0, \sigma_k^2)$ *independent, then all* $S \subset [K]$ *have:*

$$\sum_{k \in S} R_k \geq I((X_{k,raw})_{k \in S}; (U_k^*)_{k \in [K]} | (U_k^*)_{k \in S^C}).$$

Now any source-to-base rate

$$R_{BT} < I(X_{src}; \vec{U}_{[K]}^*)$$

or distortion

$$D_{BT} > \mathrm{var}\left(X_{src} \Big| \vec{U}_{[K]}^*\right) \tag{3.2}$$

is achievable.

Proof for the achieved distortion is provided in Appendix B.1.

3.4 Improvements to the Berger-Tung Strategy

Theorem 3 from Chapman *et al.* (2018) gives an inner bound on the achievable rate for this problem through a realization of a modification the Berger-Tung distributed

25

source coding strategy in El Gamal and Kim (2011) Theorem 12.1, restricted to Gaussian distortions.

In this strategy, each relay produces a quantization of its observation with Gaussian distortion, then the relays jointly compress their quantizations with the Slepian-Wolf theorem so that there is only one source message jointly typical with the ensemble of compressions. The generalization over plain Berger-Tung compression comes from an improvement that binning rates in joint compression only need be chosen to retain information pertaining to the source, not all the quantizations.

Theorem 6. (Improved Berger-Tung with Gaussian distortion for distributed receive) *Fix non-negative numbers $\sigma_1^2, \ldots, \sigma_K^2, \lambda$ so that the following conditions hold. Defining $U_k^* \triangleq X_{k,raw} + W_{Q,k}$, $W_{Q,k} \sim \mathcal{N}(0, \sigma_k^2)$, then $\forall S \subset [K]$:*

$$\sum_{k \in S} R_k \geq I((X_{k,raw})_{k \in S}; (U_k^*)_{k \in [K]} | (U_k^*)_{k \in S^C}) - \lambda.$$

Now any source-to-base rate

$$R_{\bar{BT}} < I(X_{src}; \vec{U}_{[K]}^*) - \lambda \tag{3.3}$$

is achievable.

Proof for the achieved communications rate is provided in Appendix B.1.

A more general version of this bound (in particular, without the restriction to Gaussian distortion) is the main subject of Sanderovich *et al.* (2008). This more general bound does not have closed form, even for the Gaussian case. This is the strongest bound known for the present problem. The extent to which the improvement fully overcomes the inefficiencies in application of the plain Berger-Tung strategy from Theorem 5 is not known. Inefficiencies are both those discussed in Section 2.3.1 and the ones previously mentioned in Section 1.2 due to relays not decoding when X_{src}^n has codebook structure.

26

Chapter 4

STRUCTURED SCHEME

4.1   Introduction

In this chapter a scheme for recovery of a signal by distributed listeners in the presence of Gaussian interference is constructed by exhausting an 'iterative power reduction' property of lattice codes. When the amount of information shared among observers is limited (as it well may be if they must conserve power or bandwidth, see Section 1.1.1), they must take care to only forward novel information to their neighbors. Broadly, the codes described in this chapter provide a method of controlling the specificity of information provided by each observer's message to this end. The strategy is called in this chapter 'Successive Integer-Forcing Many-help-one' (SIFM).

An upper bound for the coding scheme's achieved mean-squared-error distortion for the base's estimate of X_{src} is derived. The strategy exposes a parameter search
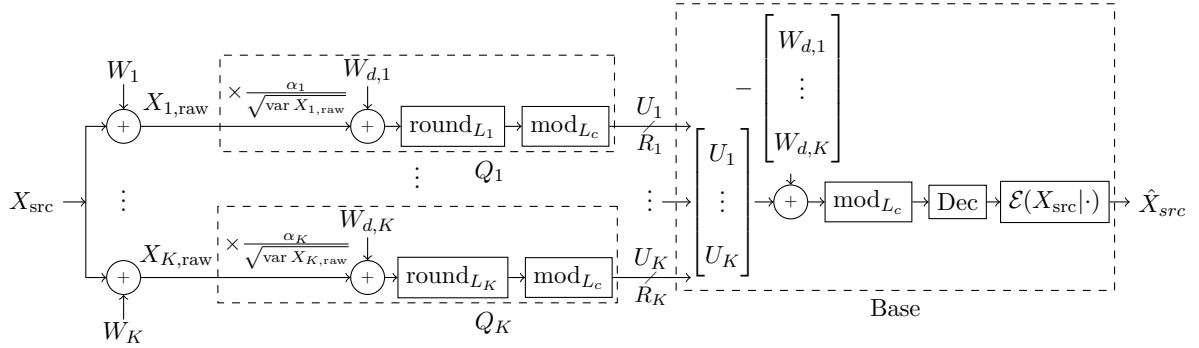
Figure 4.1: Block diagram of structured lattice coding strategy, expanded from the general block diagram in Figure 1.2. See the beginning of Section 4.2 for details.

problem which, when solved, results in a scheme which outperforms others of its kind. Performance of a blocklength-one scheme is simulated and is seen to improve over plain source coding without compression in the presence of many interferers, and experiences less outages over ensembles of channels. The asymptotic version of the strategy presented in this chapter still has a performance gap to random coding bounds from Chapter 3 in some regimes. This gap is partially remedied in Chapter 5 whose topic is joint-compression of the encodings produced in this chapter.

This is a strategy for the *Gaussian many-help-one source coding problem* Tavildar and Wagner (2009), allowing for the case where the 'receiver being helped' cannot provide side information to the decoder. The asymptotic version of the scheme is a direct application of general results and ideas from Nazer *et al.* (2016) to the many-help-one problem, although derivation differs, yielding a different characterization of its achieved rate region in terms of Algorithm 1. There is a lot of existing literature on lattice signal processing for this scenario, and similar techniques have been applied to closely related problems. The procedures which comprise SIFM have also been described notionally in Ordentlich and Erez (2017) and are contained within more involved MIMO communication scenarios in Zamir (2014)[Chapter 12].

SIFM is a strict generalization of a scheme by Krithvasan and Pradhan (KP) in Krithivasan and Pradhan (2007). Although conceptually very similar, the generalizations of SIFM over KP necessitate a near ground-up re-description of the strategy. This parametrization of the KP scheme in terms of SIFM is shown in Appendix C.4. Unfortunately, SIFM replaces KP's formulaic choice of certain parameters with a difficult continuous-domain nonlinear parameter search problem described in Section 4.4.2. This comes with the benefit of SIFM reliably outperforming KP in the regimes tested in Section 4.4.

It was recognized in References Wagner (2011), Yang and Xiong (2011), Yang and Xiong (2014) and Cheng *et al.* (2018) that the lattice messages involved in such schemes are still non-trivially jointly distributed and could be further compressed to yield performance improvements. Most results on this correlation Wagner (2011), Yang and Xiong (2011) focus on the case of two observers, and the compression described in Cheng *et al.* (2018) is inexhaustive. SIFM suffers from the same problem, and the amount of redundancy still present in observer messages has not yet been totally characterized. Analysis of the network scheme in the method presented here enables further study of these encoding redundancies. An initial investigation of such properties is presented in Chapter 5.

4.1.1   Outline

A coding scheme for recovery of a Gaussian source using side information from many receivers in the presence of interference is presented in Section 4.2. This scheme outperforms others of its kind by exhausting a power-reduction property. The asymptotic version of the scheme is presented in Section 4.3, yielding Theorem 7. An algorithm for calculating the asymptotic scheme's performance in Theorem 7 is presented as Algorithm 1.

The coding scheme is described in terms of events of successful decoding (Definition 8) constructed in terms of arbitrary lattices. The definition allows for an upper bound on achieved distortion depending on scheme parameters and the lattices used (Lemma 2.) In limit with blocklength and certain choice of lattices, Lemma 2 yields an asymptotic performance (Corollary 1.) This bound, along with several others is plotted over various regimes in Section 4.4.

4.2 Successive Integer Forcing Many-Help-One Scheme

A block diagram of the scheme is shown in Figure 4.1. Broadly, the strategy operates as follows.

Design: First a blocklength n is chosen and a 'coarse lattice' $L_c \subset \mathbb{R}^n$ is chosen according to some design to be specified. One assumes the LAN is established and allows for reliable communication from receivers to the base node, each k-th receiver at some rate R_k. Given these rates, a 'fine lattice' $L_k \supset L_c$ is chosen for each k-th receiver according to some design to be specified. At each k-th receiver a quantization dither[1] $W_{d,k}$ is selected randomly uniformly over the base region of the receiver's fine lattice, L_k.

One assumes that the covariance between the transmitter's signal and all the receivers' observations is known at all the receivers and stable over n observations. As a function of rates $R_1, \ldots, R_K > 0$ and the channel covariance matrix, some scale parameters $\alpha_1, \ldots, \alpha_K > 0$ are designed.

Operation:

1. Receiver k, labeled Q_k in Figure 4.1, normalizes all its observed sequence of n samples by $\alpha_k / \sqrt{\operatorname{var} X_{k,\mathrm{raw}}}$ so that expected-power-per-sample is 1.

2. Receiver k quantizes the result from the previous step by adding dither $W_{d,k}$, then rounding the result onto a nearby point on its fine lattice L_k.

---

[1]The quantization step involves dither. The dither causes quantization noise to manifest as additive, independent of the input signal, and uniform over the base of L_k. These properties are demonstrated via the *crypto lemma* (Zamir, 2014, Theorem 4.1.1). Use of dither is described in detail in the Appendix proofs and must be included here for a complete description of the scheme's operation but is inessential to an initial broad understanding of the scheme.

3. Receiver k takes the modulo of the rounded result onto to L_c, producing a point in L_c's 'modulo-space.' L_k must be designed with respect to L_c such that the result of this step has entropy-rate less than R_k. Receiver k forwards this result to the base. See Figure 2.2 for an example design of L_c and L_k for blocklength $n = 2$ and receiver rate $R_k = 2$ bits per sample.

4. The base receives all the receiver messages, each message being some point in L_c-modulo-space, and removes the dither by subtracting the chosen dithers from each message, and taking the L_c-modulo of the result.

5. The base recombines them in a way that the result is no longer in modulo-space, but some linear combination of the quantizations.

6. The base recombines this recovered component from all the original modulo-space messages to produce new modulo-space points with the just-recovered component removed. The removal process is illustrated in Figure 4.2.

7. This removal allows for new recombinations to allow recovery of different components. This process is repeated until no more components can be recovered. All repetitions of steps 5-7 are included in 'Dec' block in Figure 4.1.

8. All recovered components are used to estimate the source.

This is described in full precision below. To simplify exposition, this chapter instead works over variables

$$X_k := \frac{1}{\sqrt{\operatorname{var} X_{1,\mathrm{raw}}}} X_{k,\mathrm{raw}}, \ \ k = 1, \ldots, K$$

so that normalization is not necessary.

An example of how Property 4 can be used is shown in Figure 4.2. The efficacy of the lattice strategy presented in this chapter is in exhaustive use of this technique.
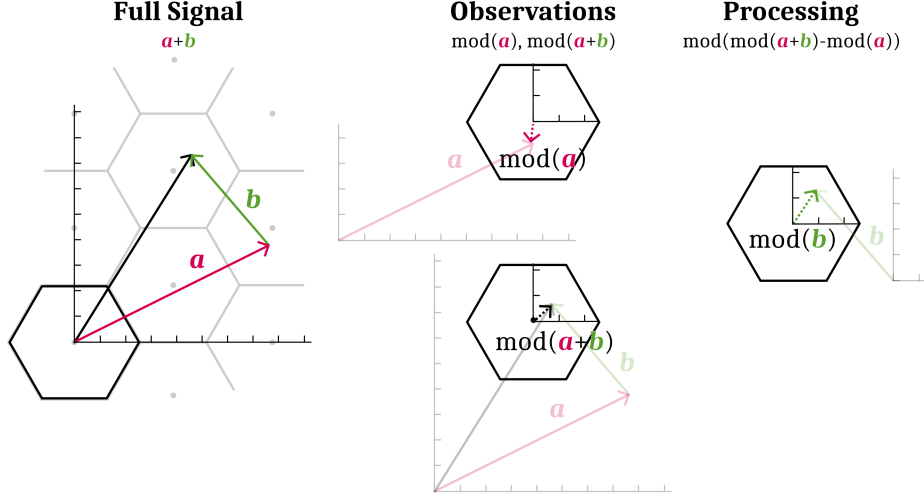
Figure 4.2: A graphical instance of removal of a vector component $\boldsymbol{a}$ from modulo $\text{mod}(\boldsymbol{a} + \boldsymbol{b})$. Drawn in the same context as Figure 2.1. The full signal, $\boldsymbol{a} + \boldsymbol{b}$ sans any modulo operations, is plotted on the left. Say that available for processing are the vectors $\text{mod}(\boldsymbol{a})$ and $\text{mod}(\boldsymbol{a} + \boldsymbol{b})$, illustrated in the middle. A particular processing of $\text{mod}(\boldsymbol{a})$ and $\text{mod}(\boldsymbol{a} + \boldsymbol{b})$ produces $\text{mod}(\boldsymbol{b})$, shown on the right. Non-modulo components are drawn lightly under each modulo.

4.2.1 Lattice Scheme Description

Fix the following parameters:

- Blocklength n

- Receiver scales $\vec{\alpha} = (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K$,

- 'Fine' lattices $L_k \subset \mathbb{R}^n$ each with a base region $B_k \subset \mathbb{R}^n$, $k \in [K]$

- A 'coarse' lattice $L_c \subset \mathbb{R}^n$ with a base region B_c, where $L_k \supset L_c$ and $\frac{1}{n} \log |B_c \cap L_k| \leq R_k$ for each $k \in [K]$.

- Functions $\phi_k : (B_c \cap L_k) \to [2^{nR_k}]$ which enumerate their domain's points

- Dither variables $\vec{W}_d = (W_{d,1}, \ldots, W_{d,K})$, with $W_{d,k} \sim \text{unif } B_k$ independent over index k.

For brevity we neglect to denote application of the enumeration ϕ_k and its inverse when it is clear from context where it should be applied.

Encoders

Each encoder first scales its observation then quantizes the result by rounding with dither onto its fine lattice L_k. This discretizes the source's observation onto a countable collection of points, but it may still be too high-rate to forward to the base directly. The encoder wraps the discretization onto the coarse lattice's base region B_c by applying mod_{B_c}. The domain reduction from all of L_k to only points within $B_c \cap L_k$ reduces the discretization's entropy enough to forward it to the decoder.

Construct the encoder for receiver k as $\text{enc}_k : \mathbb{R}^n \to [2^{nR_k}]$:

$$\text{enc}_k(x_k^n) := \phi_k(\text{mod}_{L_c}(\text{round}_{L_k}(\alpha_k x_k^n + W_{d,k})))$$

Decoders

The decoder produces estimates of particular integer linear combinations $\mathbf{a}_1, \ldots, \mathbf{a}_K$ of the source observations by processing the encodings in stages. In stage k, the decoder recovers combination $\mathbf{a}_k$, and in all future stages the $\mathbf{a}_k$ component is used to aid recovery. This is be described in more detail later. To construct each stage it is necessary to describe the covariance between receiver quantizations. By (Zamir, 2014, Theorem 4.1.1),

$$\text{round}_{L_k}(\alpha_k X_k^n + W_{d,k}) - W_{d,k} = \alpha_k X_k^n - \tilde{W}_{d,k} \tag{4.1}$$

and $\tilde{W}_{d,k} \sim \text{unif } B_k$ is independent of X_k^n. Denote $\tilde{W}_d = (\tilde{W}_{d,1}, \ldots, \tilde{W}_{d,K})$. Also denote the scaled receiver observations as $\overrightarrow{\alpha X} = (\alpha_1 X_1, \ldots, \alpha_K X_K)$, and similarly for

time-expanded $\vec{X}^n$. Then on average over time, the receivers' dithered rounding to their fine lattices effectively adds noise of the following covariance to the observations:

$$\mathbf{\Sigma}_Q := \mathrm{diag}\left(\left(\mathbb{E}\left[\frac{1}{n}\|\tilde{W}_{d,k}\|^2\right]\right)_{k\in[K]}\right).$$

A covariance matrix $\mathbf{C}$ between the sources after dithered rounding to the fine lattices can be written as the time-averaged covariance of vectors in (4.1):

$$\mathbf{C} := \mathbb{E}\left[\frac{1}{n}(\overrightarrow{\alpha X}^n - \tilde{W}_d)^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d)\right].$$

$$= (\mathrm{diag}\,\vec{\alpha})\mathbf{\Sigma}(\mathrm{diag}\,\vec{\alpha})^\dagger + \mathbf{\Sigma}_Q.$$

Now some events over outcomes of $(\overrightarrow{\alpha X}^n - \tilde{W}_d)$ are constructed.

Definition 8. For some $K' \in [K]$, fix an integer matrix $\mathbf{A} \in \mathbb{Z}^{K \times K'}$, call its columns $\mathbf{a}_1, \ldots, \mathbf{a}_{K'} \in \mathbb{Z}^K$, and take $\mathbf{A}_k$ to be the first k columns of $\mathbf{A}$.

In terms of $\mathbf{A}$ define[2] matrices for each $k = 1, \ldots, K'$:

$$\mathbf{S}_k(\mathbf{v}) := \underset{\mathbf{u}\in\mathbb{R}^k}{\arg\min}\,\mathrm{var}\left((\mathbf{v} - \mathbf{A}_k\mathbf{u})^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d)\right)$$

$$\mathbf{R}_k(\mathbf{v}) := [\mathbf{I}_{K \times K} - \mathbf{A}_k\mathbf{S}_k]\mathbf{v}.$$

Also in terms of $\mathbf{A}$ define events $\mathcal{M}(\mathbf{A}_1), \ldots, \mathcal{M}(\mathbf{A}_{K'}) \subset (\mathbb{R}^n)^K$:

$$\mathcal{M}(\mathbf{A}_1) := \{\mathbf{z} \in \mathbb{R}^{K \times n} : \mathbf{a}_1^\dagger\mathbf{z} \in B_c\},$$

$$\mathcal{M}(\mathbf{A}_k) := \mathcal{M}(\mathbf{A}_{k-1}) \cap \left\{\mathbf{z} \in \mathbb{R}^{K \times n} : (\mathbf{R}_{k-1}\mathbf{a}_k)^\dagger\mathbf{z} \in B_c\right\}.$$

The events designate when a particular processing of the encodings successfully produces an estimate of the observations *without modulo*:

Lemma 1. *Fix* $\mathbf{A} \in \mathbb{Z}^{K \times K'}$ *as in Definition 8. Take* $U_k = \mathrm{enc}_k(X_k^n)$ *for* $k \in [K]$. *There is a function* f *where* $f(\vec{U}) = \mathbf{A}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d)$ *whenever* $(\overrightarrow{\alpha X}^n - \tilde{W}_d) \in \mathcal{M}(\mathbf{A})$.

---

[2]The arg min is derived in closed form in Appendix C.3

The functions $f_1, \ldots, f_{K'}$ for each output of f, $f_{[k]}$ denoting the first k outputs, are given as follows, all mod *taken with respect to coarse base region B_c.*

$$f_1\left(\vec{U}\right) := \mathrm{mod}\left(\mathbf{a}_1^\dagger(\vec{U} - \vec{W}_d)\right),$$

$$f_m\left(\vec{U}\right) := \mathrm{mod}\left\{\mathbf{a}_m^\dagger(\vec{U} - \vec{W}_d) - \ldots \right.$$

$$\left. \mathrm{mod}\left([\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger f_{[m-1]}(\vec{U})\right)\right\} + \ldots$$

$$[\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger f_{[m-1]}(\vec{U}).$$

Proof is delayed to Appendix C.1.

This aspect of general reuse of all previously recovered components for the recovery of a new one is the source of benefit of the present scheme over Krithivasan and Pradhan (2007). Comments in Zamir (2014) among other documents describe such a compression strategy.

A decoder can be realized from each event from Definition 8 by using a linear estimator on the output of f gotten from that event in Lemma 1. The likelihood of each event is quite sensitive to channel covariance $\boldsymbol{\Sigma}$ receiver scalings $\vec{\alpha}$ and integer vector $\mathbf{a}_\ell$ and as layers of component recovery are added the events become increasingly unlikely. Thus only a few decoders perform reliably.

Decoder Performance

We now bound the worst-case performance of such decoders.

Definition 9. In the context of Lemma 1, take $\mathbf{e}_{X|\mathbf{A}}$ to be the coefficients of the best linear unbiased estimator for X given $\mathbf{A}^\dagger(\overrightarrow{\alpha X} - \vec{W}_Q)$, $\vec{W}_Q \sim \mathcal{N}(0, \boldsymbol{\Sigma}_Q)$. Then for $\Delta > \mathrm{var}\left(X|\mathbf{A}^\dagger(\overrightarrow{\alpha X} - \vec{W}_Q)\right)$ define a decoder:

$$\mathrm{dec}\left(\vec{U}; \mathbf{A}\right) := \mathbf{e}_{X|\mathbf{A}}^\dagger f(\vec{U}), \tag{4.2}$$

or 0 if the observed average power of (4.2) is greater than Δ.

The mean-squared-error distortion each such decoder achieves can be upper bounded in terms of the probability of the event $\mathcal{M}(\mathbf{A})$:

Lemma 2. *Take a decoder from Definition 9 and define*

$$d^2 := 1 - \mathrm{var}(X | \mathbf{e}^\dagger_{X_{src}|\mathbf{A}}(\overrightarrow{\alpha X} + \vec{W}_Q)).$$

Take $\mathcal{E}$ to be the event where the decoding is zero. Then no worse than the following mean-squared-error is achieved in estimating X^n_{src}:

$$d^2 + (\Delta + 1)^2 \cdot \sqrt{3\mathbb{P}(\mathcal{M}^C)} + \sqrt{3\mathbb{P}(\mathcal{E})}$$

Proof. The goal is to approximate the integral:

$$\int \frac{1}{n} \| \mathrm{dec}(\vec{U}; \mathbf{A}) - X^n_{src} \|^2 d\mathbb{P} = \int_{\mathcal{M}} \frac{1}{n} \| \mathrm{dec}(\vec{U}; \mathbf{A}) - X^n_{src} \|^2 d\mathbb{P} + \dots \qquad (4.3)$$
$$\int_{\mathcal{M}^C} \frac{1}{n} \| \mathrm{dec}(\vec{U}; \mathbf{A}) - X^n_{src} \|^2 d\mathbb{P}.$$

Bound the first summand of (4.3):

$$\int_{\mathcal{M}} \frac{1}{n} \| \mathrm{dec}(\vec{U}; \mathbf{A}) - X^n_{src} \|^2 d\mathbb{P} = \frac{1}{n} \int_{\mathcal{M}} \mathbf{1}_{\mathcal{E}^C} \| E[X^n_{src} | \mathbf{A}(\overrightarrow{\alpha X} - \vec{W}_Q)] - X^n_{src} \|^2 + \mathbf{1}_{\mathcal{E}} \| X^n_{src} \|^2 d\mathbb{P}$$

$$< \frac{1}{n} \int \| E[X^n_{src} | \mathbf{A}(\overrightarrow{\alpha X} - \vec{W}_Q)] - X^n_{src} \|^2 d\mathbb{P} + \frac{1}{n} \int_{\mathcal{E}} \| X^n_{src} \|^2 d\mathbb{P}$$

$$= d^2 + \frac{1}{n} \int_{\mathcal{E}} \| X^n_{src} \|^2 d\mathbb{P}.$$

where the first equality follows by Lemma 1 and choice of decoder. Bound the second summand of (4.3):

$$\int_{\mathcal{M}^C} \frac{1}{n} \| \mathrm{dec}(\vec{U}; \mathbf{A}) - X^n_{src} \|^2 d\mathbb{P} \le \frac{1}{n} \int_{\mathcal{M}^C} \| (1 + \Delta) X^n_{src} \|^2 d\mathbb{P}$$

$$= (1 + \Delta)^2 \frac{1}{n} \int_{\mathcal{M}^C} \| X^n_{src} \|^2 d\mathbb{P}.$$

Applying Hölder's inequality as follows to the two bounds yields the result:

$$\int_S \|X_{\text{src}}^n\| d\mathbb{P} \leq \left[\int \mathbf{1}_S^2 d\mathbb{P}\right]^{1/2} \cdot \left[\int \|X_{\text{src}}\|^4 d\mathbb{P}\right]^{1/2}$$

$$= \sqrt{3\mathbb{P}(S)}. \qquad \square$$

The bound on distortion suffices for the asymptotic analysis in Section 4.3, but is quite coarse in low dimension.

4.3   Asymptotic Scheme

Analysis of the scheme in limit with blocklength over particular choice of $L_c, (L_k)_{k\in[K]}$ yields a nice characterization of its performance. Theorem 7 demonstrates that in limit with blocklength and certain lattice design, there is essentially one decoder which performs at least as well as any others. The subspace of receiver observations it reliably recovers is characterized. First, a matrix definition is needed.

Definition 10. For $\varepsilon > 0$ define covariance matrices $\mathbf{\Sigma}_{Q\infty}, \mathbf{C}_\infty \in \mathbb{R}^{K\times K}$

$$\mathbf{\Sigma}_{Q\infty} := \text{diag}\left(2^{-2R_1+\varepsilon}, \ldots, 2^{-2R_K+\varepsilon}\right). \tag{4.4}$$

$$\mathbf{C}_\infty := (\text{diag}\,\vec{\alpha})^\dagger \mathbf{\Sigma}(\text{diag}\,\vec{\alpha}) + \mathbf{\Sigma}_{Q\infty}.$$

$\mathbf{C}_\infty$ represents the covariance between quantized observations achieved in limit with blocklength when the lattices $(L_k)_{k\in[K]}$ are chosen well.

Theorem 7. *Take $\varepsilon > 0$ small and $\mathbf{C}_\infty$ from (4.4). Define S to be the smallest subspace in $\mathbb{R}^K$ with the property that all integer vectors $\mathbf{v} \in \mathbb{Z}^K$ have either $\min_{\mathbf{s}\in S}(\mathbf{v} - \mathbf{s})^\dagger \mathbf{C}_\infty(\mathbf{v} - \mathbf{s}) \geq 1$ or $\mathbf{v} \in S$. Fix $\mathbf{P}_\infty$ as the projection onto S. Then for large enough blocklength n and certain encoders, some processing f of the encodings has with high probability*

$$f(\vec{U}) = \mathbf{P}_\infty(\overrightarrow{\alpha X}^n + \vec{W}_Q'),$$

where $\vec{W}_Q' = (W_{Q,1}', \ldots W_{Q,K}')$ has independent components and $\mathbb{E}(\frac{1}{n}\|W_{Q,k}'\|^2) < 2^{-2R_k+2\varepsilon}$.

37

Corollary 1. *A decoder provided side information as in Theorem 7 can achieve the following MSE distortion in estimating X_{src}:*

$$d_L^2 := 1 - \vec{\alpha}^\dagger \operatorname{pinv}(\mathbf{P}_\infty \mathbf{C}_\infty \mathbf{P}_\infty) \vec{\alpha}.$$

Proof. Apply a linear estimation for the source on f's output. $\qquad\qquad\square$

Proof is given in Appendix C.2. The theorem is demonstrated by observing that if lattices are chosen well, then events from Definition 8 approach probability zero or one, and that $\mathbf{P}_\infty$'s image coincides with the span of the high-probability-events' vectors. Computation of the projection $\mathbf{P}_\infty$ can be done via repeated reduction of the covariance $\mathbf{C}_\infty^{1/2}$:

Algorithm 1 Compute projection $\mathbf{P}_\infty$ and processing stages $\mathbf{A}$ from $\mathbf{C}_\infty$.

$\quad \mathbf{A} \leftarrow [\,], \; \mathbf{a} \leftarrow \operatorname{SLVC}(\mathbf{C}_\infty^{1/2}), \; \mathbf{R} \leftarrow \mathbf{I}_{K \times K},$

$\quad$**while** $0 < (\mathbf{Ra})^\dagger \mathbf{C}_\infty \mathbf{Ra} < 1$ **do**

$\qquad \mathbf{A} \leftarrow [\mathbf{A}, \mathbf{a}]$

$\qquad \mathbf{R} \leftarrow \mathbf{I}_{K \times K} - \mathbf{A}\operatorname{pinv}(\mathbf{A}^\dagger \mathbf{C}_\infty \mathbf{A})\mathbf{A}^\dagger \mathbf{C}_\infty$

$\qquad \mathbf{a} \leftarrow \operatorname{SLVC}(\mathbf{C}_\infty^{1/2}\mathbf{R})$

$\quad$**end while**

$\quad \mathbf{P}_\infty \leftarrow \mathbf{A}(\mathbf{A}^\dagger \mathbf{A})^{-1}\mathbf{A}^\dagger$

$\quad$**return** $\mathbf{P}_\infty, \; \mathbf{A}$

In Algorithm 1, the subroutine $\operatorname{SLVC}(\mathbf{B})$, 'Shortest Lattice Vector Coordinates' returns the nonzero integer vector $\mathbf{a}$ which minimizes the norm of $\mathbf{Ba}$ while $\mathbf{Ba} \neq 0$. $\operatorname{SLVC}(\cdot)$ can be implemented using a lattice enumeration algorithm like one in Schnorr and Euchner (1994) together with the LLL algorithm to convert a set of spanning lattice vectors into a basis Buchmann and Pohst (1987). Algorithm 1 indeed returns $\mathbf{P}_\infty$ since it is lifted from $\mathbf{P}_\infty$'s construction in the proof of Theorem 7 given in Appendix C.2.

4.3.1 Complexity of Scheme

The complexity of the scheme in operation is identical to that of codes with similar structure. Examples include IFSC and KP as discussed in the introduction. In particular if $g(n)$ grows like the amount of operations used to evaluate round$_L$ (L the involved lattice for which round$_L$ is hardest to compute), then it follows from the scheme description that the time complexity of each encoder is $O(n + g(n))$ and the time complexity of the base's decoder is $O(nK + K^2 \cdot g(n)))$. Practicality of such a scheme is then primarily dependent on the existence of lattices which both satisfy the 'goodness' properties described here and have rounding, modulo operations of tractable complexity. Several propositions for such lattice structures exist, for instance *LDPC lattices* ($O(n)$) da Silva and Silva (2018) and *polar lattices* ($O(n \log n)$) Liu *et al.* (2018).

Configuration of the scheme is also a difficult computational problem. Determination of the optimal configuration is a non-smooth continuous-domain search for the objective given in Corollary 1 over choice of encoder scalings $\vec{\alpha} \in \mathbb{R}^+$. Each objective evaluation involves computing Algorithm 1. Due to this algorithm involving several shortest-lattice-vector-problems in dimension up to K, one implementation of this algorithm has time complexity $O(K2^K)$. This complexity reduces considerably if the strict shortest-lattice-vector problem is relaxed to allow approximate solutions such as those provided by the LLL algorithm. No method for global optimization of the objective is known. Computationally tractable approximations can provide strong performance as seen in Section 4.4.

4.4   Performance

Here the performance of the scheme is compared to existing results. All numerical results deal with complex circularly-symmetric Gaussian channels. Although the SIFM derivations shown are in terms of real channels for easier exposition, they apply just as well to the complex case by considering each complex observation as two appropriately correlated real observations. The explicit extension to the complex case is presented in Appendix F. Each data point was computed on average over 200 randomly generated channel covariances selected with the prescribed statistics. An outer bound representing the performance if the base had access to the receivers' observations in full precision is also shown.

4.4.1   Quantize and Forward

One strategy much simpler than SIFM is for each observer forward its own quantized representation of its observations to the base. A 'saturating uniform quantizer' was simulated, where each real observation was clipped to some interval, then rounded onto a quantization step using the prescribed data rate. This curve is labeled '1D Quantize & Forward.'

Performance when observers used rate-distortion quantizers for Gaussian sources was also computed and is labeled 'HD Quantize & Forward.' Performance for the rate-distortion quantizer strategy when observer bitrates were allowed to vary within a sum-rate was also plotted as 'HD Quantize & Forward (Variable-Bitrate).'

These techniques are strong when the LAN rate is severely limited or the interference is low, as seen in 4.3 and 4.5. The function of joint compression in these regimes is more subtle since there are less superfluous correlated components among receiver observations to eliminate. Quantize & Forward schemes are weak relative to

the others in the presence of many strong interferers, as seen in Figures 4.6, 4.3, 4.4, and 4.5.

4.4.2   Asymptotic Scheme

The distortion SIFM achieves is dependent on choice of observer scales $\vec{\alpha}$ and is non-convex in terms of them due to the discontinuity of Algorithm 1's outputs. For this reason some form of search for good scale parameters is required for each fixed channel covariance and observer rate vector $\vec{r}$. For performance evaluation, the search problem was solved approximately.

Recall that the asymptotic scheme described here includes one by Krithivasan and Pradhan (KP, see Introduction) as a special case. Details of the parametrization of the KP scheme in terms of SIFM are shown in Appendix C.4. It was observed empirically that scaling all receiver's observation by the same constant can yield strong performance. The strongest between uniform and KP scaling was chosen as an initial guess and was improved by taking random steps.

4.4.3   One-Shot Scheme

If the involved lattices are all one-dimensional (i.e. nested intervals) then the likelihood of each event in Definition 8 is straightforward to approximate via Monte-Carlo simulation. This enables a slight modification of Algorithm 1 to be used for identifying a strong decoding strategy for given receiver scalings. Although the upper bound in Lemma 2 applies to the one-shot strategy, it is often too weak in low dimension to be informative of a scheme's true distortion. Distortion was estimated by simulation instead. In plots this scheme is labeled '1D SIFM.'

Unfortunately, the best identified configurations for the one-shot scheme did not consistently provide significant improvement over uniform quantization as the high

dimensional scheme had. This is thought to be because even in extreme observations, the uniform quantizer saturates and still provides a reasonable representation of the source. In contrast the same observation in SIFM creates a lattice wrapping distortion which significantly affects the final decoding result. Reducing the likelihood of such errors in low dimension requires conservative choice of scales which further degrades performance. However, such problems diminish in limit with dimension.

Performance for the asymptotic SIFM strategy when observer bitrates were allowed to vary within a sum-rate was also plotted as 'HD SIFM (Variable-Bitrate).'

4.4.4   Versus Increasing Receiver Rates

Performance for the various schemes considered is shown in Figure 4.3 for 5 receivers, each messaging to the base at rate varying from 2 bits-per-complex-sample up through 16 bits-per-complex-sample, in the presence of 3 interferers each appearing at each observer at an average of 20 dB above an 0 dB signal of interest.

Coincidentally the performance of the KP scheme with equal bitrates was observed in all plots to closely match the performance of the variable-bit-rate Quantize & Forward scheme.

As seen in Figure 4.4 there is still a significant gap at low bitrates between the performance of SIFM and the bound given in Chapman *et al.* (2018). This bound is based on non-structured joint-compression of quantizations (like Berger-Tung source coding El Gamal and Kim (2011) but binned for recovery of a source rather than all the observers' quantizations). The gap in performance could be due to a combination of factors. First, the scheme from Chapman *et al.* (2018) is designed so that receiver messages are independent of one another, while in SIFM some inter-message dependences are still present after lattice processing. This means SIFM messages could
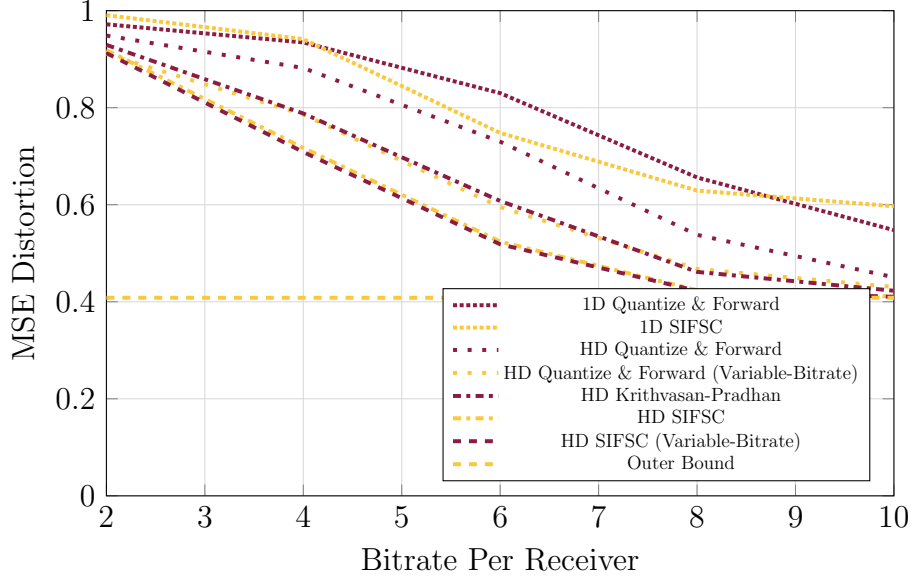
Figure 4.3: MSE performance versus receiver bitrate for five receivers the presence of three 20 dB interferers, all observing a 0 dB source. SIFM consistently significantly outperforms the rest of the schemes, and does not appear to benefit much from redistribution of observer bit-rates.

be jointly compressed to improve performance. Another contributing factor could be poor solution of the search problems mentioned in Section 4.4.2.

4.4.5   Versus Adding Interferers

Performance for the various schemes is is shown in Figure 4.5 for a system of 5 receivers, each messaging to the base at a rate of 6 bits-per-complex-sample, in the presence of 0 through 5 independent interferers each appearing at each observer an average of 20 dB above a 0 dB signal of interest.
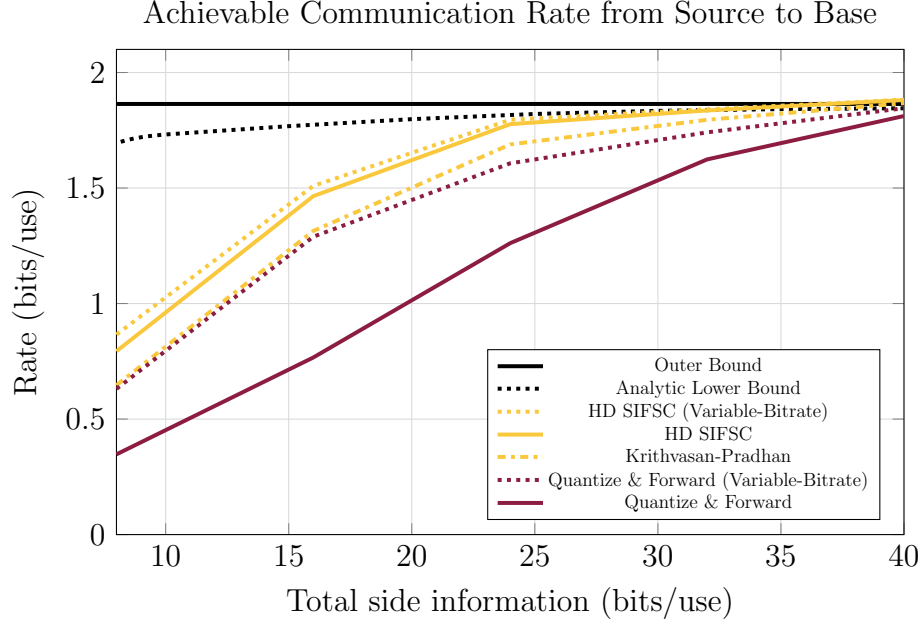
Figure 4.4: Achievable end-to-end communication rate between source and base for various sources, including an achievable rate using Berger-Tung-like lossy distributed source coding. Four receivers each observing a source at 0 dB in the presence of one 20 dB interferer. Notice that at low bitrates there is a large gap between this bound and the achieved SIFM rate. This could either be due to the fact that SIFM messages are not independent of each other and could be further jointly compressed, or because the SIFM messages are sub-optimally configured.

4.4.6   One-Shot Outage Probability

One may be interested in likelihood of being able to recover a signal to above some acceptable noise threshold over an ensemble of channels. Although in low-interference regimes the one-shot SIFM scheme is reliably outperformed by the much simpler one-shot Quantize & Forward scheme (Figure 4.5), one-shot SIFM is much less likely to perform exceptionally poorly when averaged over the current channel model. See Figure 4.6.
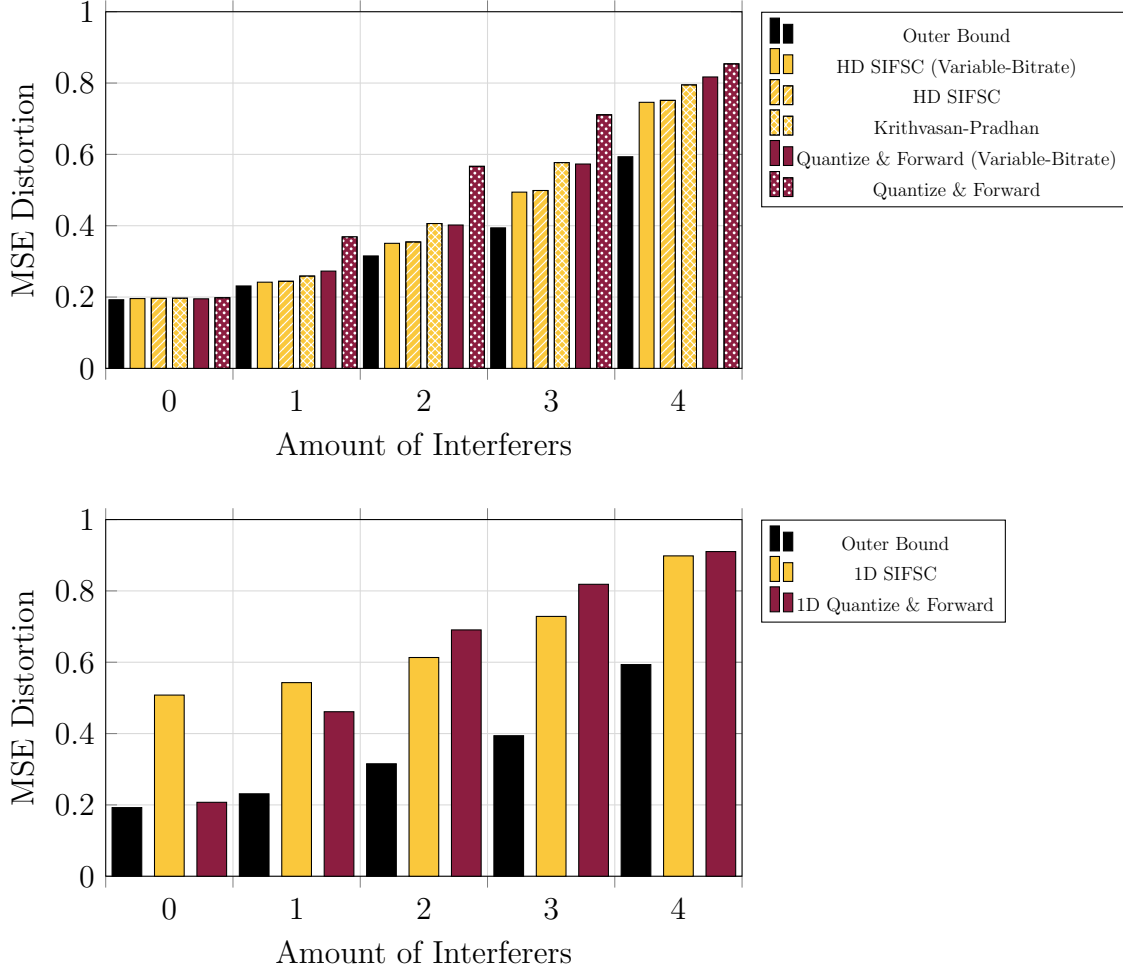
Figure 4.5: MSE performance versus amount of 20 dB interferers, when each of 5 observer encodes at a rate of 6 bits per observation (or 30 bits total shared among observers for Variable-Rate strategies). SIFM consistently outperforms the rest of the strategies and was not observed to improve much by reconfiguring observer bitrates. One-shot SIFM did not consistently outperform much simpler one-shot quantizers.
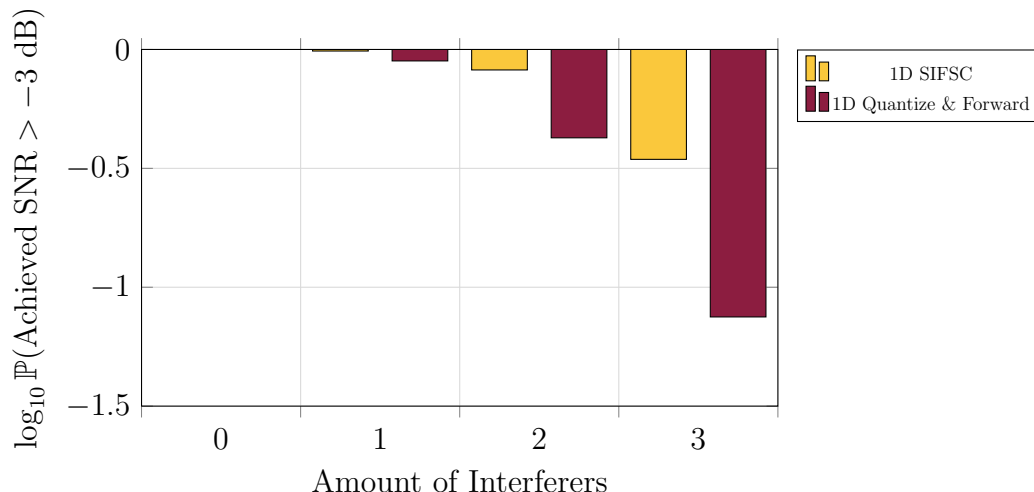
Figure 4.6: Likelihood of scheme providing recovery of X_{src} better than -3 dB SNR versus amount of 20 dB interferers, when each of 5 observer encodes at a rate of 6 bits per observation. For this threshold and ensemble of channels, one-shot SIFM performs roughly as reliably as one-shot Quantize & Forward with one less 20 dB interferer present.

4.5 Conclusion

Successive Integer-Forcing Many-Help-One (SIFM) is a lattice-algebra-based strategy for distributed coding of a Gaussian source in correlated noise. For good choice of parameters, SIFM consistently outperforms many of the strategies it generalizes. Finding good parameters for SIFM is a difficult non-convex search problem but reasonably strong solutions can be found through well-initialized random search.

A one-shot implementation of the scheme usually outperforms plain uncompressed quantization when multiple interferers are present, but often performs worse when there are few. This is probably due to SIFM's heavy dependence on the absence of tail events that are somewhat common in low blocklength versions of the scheme. In spite of this, the one-shot scheme is more typically above low-SNR thresholds in

certain ensembles of channels. It is expected that performance would improve greatly if higher dimensional nested lattices were used.

There is still some gap between the best achievable rate and SIFM as seen in Figure 4.4. This is at least partially due to redundancies in SIFM messages. As noted by Wagner (2011) and Yang and Xiong (2011), some redundancies still exist between SIFM messages. This indicates that further compression of the messages is possible. This is investigated in Chapter 5.

STRUCTURED SCHEME WITH JOINT COMPRESSION

Full Signal **Support** **Observation** **Support**
(pre-conditioning) **(post-conditioning)**
$\beta \cdot \vec{a} + \vec{b}$ $\mathrm{mod}(\beta \cdot \vec{a} + \vec{b})$ $\mathrm{mod}(\vec{a})$ $\mathrm{mod}(\beta \cdot \vec{a} + \vec{b})$
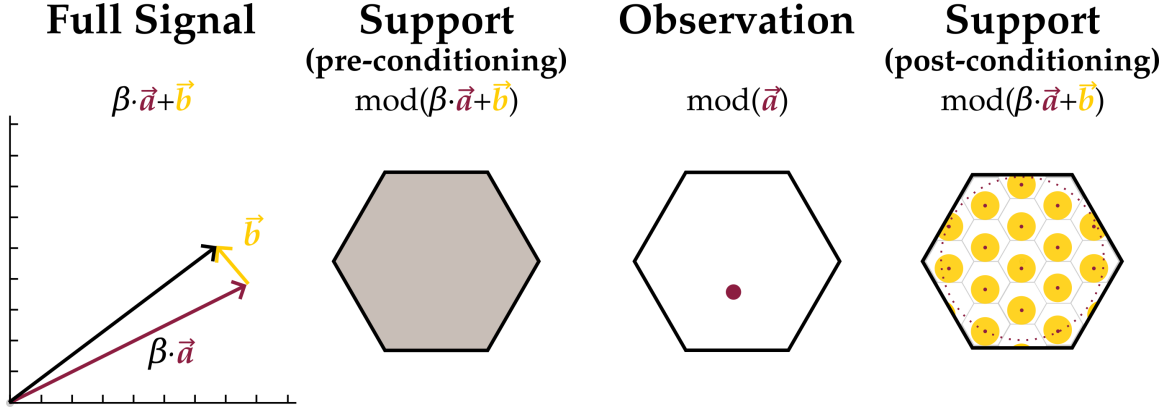
Figure 5.1: Collapse of the support of a random signal's modulo after conditioning on the modulo of a related signal. Modulo is shown to some lattice L with base region B. Consider a signal comprised of two independent random components, $\vec{a}$ and $\vec{b}$, equaling $\beta\vec{a} + \vec{b}$. A possible outcome is drawn on the far left. Unconditioned, the support for $\mathrm{mod}(\beta\vec{a} + \vec{b})$ is the entire base region B, shown fully shaded in gray. Once $\mathrm{mod}(\vec{a})$ is observed, the component $\beta\vec{a}$ is known up to an additive factor in βL. If further the powers of $\vec{a}$ and $\vec{b}$ are bounded above, this leaves feasible points for $\mathrm{mod}(\beta\vec{a} + \vec{b})$ as a subset of those of the unconditioned variable. This subset is shaded yellow on the far right.

As seen in Chapter 4, lattices provide useful structure for distributed coding of correlated sources. The basic lattice encoder construction investigated in Chapter 4 is to first round an observed sequence to a 'fine' lattice with dither, then produce the result's modulo to a 'coarse' lattice as the encoding. Such encodings may be jointly-dependent. In this chapter a class of upper and lower bounds is established on the

conditional entropy-rates of such encodings when sources are correlated and Gaussian and the lattices involved are a from an asymptotically-well-behaved sequence. The upper bounds guarantee existence of a joint–compression stage which can increase encoder efficiency. The bounds exploit the property that the amount of possible values for one encoding collapses when conditioned on other sufficiently informative encodings. The bounds are applied to the scenario of communicating through a many-help-one network in the presence of strong correlated Gaussian interferers, and such a joint–compression stage is seen to compensate for some of the inefficiency in certain simple encoder designs.

5.1 Introduction

Lattice codes are a useful tool for information theoretic analysis of communications networks. Sequences of lattices can be designed to posess certain properties which make them useful for noisy channel coding or source coding in limit with dimension. These properties have been termed 'good for channel coding' and 'good for source coding' Zamir (2014). Sequences posessing both such properties exist, and an arbitrary number of sequences can be nested Ordentlich and Erez (2016). One application of 'good' sequences of nested lattices is in construction of distributed source codes for Gaussian signals. Well designed codes for such a scenario built off of such lattices enables encoders to produce a more efficient representation of their observations than would be possible without joint code design Chapman *et al.* (2019). Such codes can provide optimal or near-optimal solutions to coding problems Erez and Zamir (2004); Ordentlich *et al.* (2013); Ordentlich and Erez (2014). Despite their demonstrated ability to compress signals well in these cases, literature has identified redundancies across lattice encodings in other contexts Wagner (2011); Yang and Xiong (2011, 2014); Cheng *et al.* (2018). In these cases, further compression of en-

codings is possible. This chapter studies the correlation between lattice encodings of a certain design.

A class of upper bounds on the conditional Shannon entropies between lattice encodings of correlated Gaussian sources is produced by exploiting linear relations between lattice encodings and their underlying signals' covariances. The key idea behind the analysis is that when the lattice-modulo of one random signal is conditioned on the lattice-modulo of a related signal, the region of feasible points for the first modulo collapses. A sketch of this support reduction is shown in Figure 5.1. This process is repeated until all information from the conditionals is integrated into the estimate of the support set. The upper bound establishes stronger performance limits for such coding structures since it demonstrates that encoders are able to convey the same encodings at lower messaging rates.

5.1.1   Background

The redundancy of lattice-modulo-encoded messages has been noticed before, usually in the context of the following many-help-one problem: many 'helpers' observe correlated Gaussian signals and forward messages to a decoder which is interested in recovering a linear combination of said signals. Towards this end, Wagner in Wagner (2011) provides an upper and lower bound on conditional entropies such as those here for a case with two lattice encodings. Yang in Yang and Xiong (2011) realized a similar compression scheme for such encodings using further lattice processing on them and presents an insightful 'coset planes' abstraction. It was further noticed by Yang in Yang and Xiong (2014) that improvement towards the many-help-one problem is obtained by splitting helper messages into two parts: one part a coarse quantization of the signal, compressed across helpers via Slepian–Wolf joint–compression (these message parts corresponding to the 'high bit planes'), and another a lattice-modulo-

encoding representing signal details (corresponding to 'low bit planes'). This chapter extends these ideas to a general quantity of helpers, and treats a case where a single component of the observations is known to have lattice structure.

Most recently, a joint–compression scheme for lattice encodings called 'Generalized Compute Compress and Forward' was introduced in Cheng *et al.* (2018), towards coding for a multi-user additive white Gaussian noise channel where a decoder seeks to recover all user's messages and is informed by helpers. The scheme in Cheng *et al.* (2018) makes use of concepts from Yang and Xiong (2014). In the scheme each lattice message is split into a combination of multiple components, each component from a different coset plane. Design of which coset planes are used yields different performance results. Section 5.3 follows along the same lines, although for a network with one user and where many interferers without codebook structure are also present.

Throughout the chapter, terminology and basic lattice theory results are taken from Zamir (2014). The lattice encoders studied are built from an ensemble of nested lattices, all both 'good for quantization' (Rogers-good) and 'good for coding' (Poltyrev-good). Such a construction is provided in Ordentlich and Erez (2016). An algorithm from Chapman *et al.* (2019) is also used which takes as an argument the structure of some lattice modulo encodings and returns linear combinations of the underlying signals recoverable by a certain type of processing on such encodings. This algorithm is listed here as STAGES*$(\cdot)$ and is shown in Appendix D.1.

5.1.2 Outline

The main theorem providing upper bounds on conditional entropies of lattice messages, along with an overview of its proof is stated in Section 5.2. An adaptation of the result to a lower bound is provided in Section 5.2.1. The theorem is slightly strengthened for an application to the problem of communicating over a many-help-

one network in Section 5.3. A numerical analysis of the bounds is given in Section 5.3.2. A conclusion and discussion on the bound's remaining inefficiencies is given in Section 5.4. A table of notation is provided in Table 2.1. A key for the interpretation of significant named variables is given in Table A.1.

5.2 Main Results

The main results are as follows:

Theorem 8. *For covariance $\Sigma \in \mathbb{R}^{K \times K}$, take $\vec{X}^n = (X_1^n, \ldots, X_K^n)$ to be n independent draws from the joint-distribution $\mathcal{N}(0, \Sigma)$. Take rates $r_1, \ldots, r_K > 0$ and any $\varepsilon > 0$. If n is large enough, an ensemble of nested lattices $L_c \subset L_1, \ldots, L_K$ (with base regions $B_c \supset B_1, \ldots, B_K$) from (Ordentlich and Erez, 2016, Theorem 1) can be designed so that the following holds. First fix independent dithers $W_k \sim \text{unif } B_k$. These dithers have $\text{var } W_k = 2^{-2r_k}$. Also fix $Y_k := \text{round}_{B_k}(X_k^n + W_k) - W_k$ and lattice modulo encodings $U_k := \text{mod}_{B_c}(\text{round}_{B_k}(X_k^n + W_k))$.*

Now for any $\boldsymbol{\nu}_0 \in \mathbb{Z}^{K-1}$, number $n_0 \in \mathbb{N}$, basis $\{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_K\} \subset \mathbb{Z}^K$, fix variables:

$$Y_0 := Y_K + \tfrac{1}{n_0}\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]},$$

$$\vec{Y}_c := (Y_0 - Y_K, Y_1, \ldots, Y_{K-1}),$$

$$\delta_0^2 := n_0^2,$$

$$\sigma_k^2 := \text{var}\left(Y_0 \middle| \text{STAGES}^*\left(\text{var}\left(\vec{Y}_c \middle| (\boldsymbol{\nu}_j^\dagger \vec{Y}_c)_{0<j\leq k}\right)\right)^\dagger \vec{Y}_c\right), k \in \{0\} \cup [K],$$

$$\delta_k^2 := \text{var}\left(\boldsymbol{\nu}_k^\dagger \vec{Y}_c \middle| \text{STAGES}^*\left(\text{var}\left(\vec{Y}_c \middle| (\boldsymbol{\nu}_j^\dagger \vec{Y}_c)_{0<j<k}\right)\right)^\dagger \vec{Y}_c\right), k \in [K].$$

Then the conditional entropy-rate is bounded:

$$\frac{1}{n}H\left(\vec{U}_K \middle| \vec{U}_{[K-1]}, \vec{W}\right) \leq \min_{k \in \{0\} \cup [K]} \left[r_K + \frac{1}{2}\log \sigma_k^2 + \sum_{j=0}^{k} \max\{\tfrac{1}{2}\log \delta_j^2, 0\} \right] + K^2 \cdot \varepsilon.$$

Bounds of this form hold simultaneously for any subset and reordering of message indices $1, \ldots, K$.

Proof for Theorem 8 is given in Appendix D.2. The proof is built from (Chapman *et al.*, 2019, Theorem 1), its associated algorithm $\textsc{Stages}^*(\cdot)$ (listed here in Appendix D.1) and two lemmas which provide useful decompositions of the involved quantities.

Lemma 3. *Take variables as in the statement of Theorem 8. Then, the ensemble of lattices described can include an 'auxiliary lattice' $\hat{L}' \subset L_K$ with base region $\hat{B}'$, nesting ratio $\frac{1}{n} \log |\hat{B}' \cap L_K| \to \frac{1}{2} \log \sigma^2 + \varepsilon$ so that*

$$U_K = \mathrm{mod}_{B_c} \left(\mathcal{C} + \frac{1}{n_0} \tilde{Y} + \tilde{Y}_\perp \right),$$

where $\mathcal{C}, \mathcal{D}$ are functions of $(\vec{U}_{[K]}, \vec{W})$, and with high probability

$$\tilde{Y} = -\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]} \in (\mathcal{D} + L_c),$$
$$\tilde{Y}_\perp = \mathcal{E}_\perp \left(Y_0 \middle| \vec{A} \right) \in \hat{B},$$
$$\vec{A} = \textsc{Stages}^* \left(\mathrm{var} \, \vec{Y}_c \right)^\dagger \vec{Y}_c.$$

In addition, $\sigma^2 = \max\{2^{-2r_K}, \mathrm{var} \, \tilde{Y}_\perp\}$.

Lemma 4. *Take variables as in the statement of Theorem 8. Then, the ensemble of lattices described can include 'auxiliary lattices' $\hat{L} \subset L_c, \hat{L}' \subset L_K$ with base regions $\hat{B}, \hat{B}'$, nesting ratios $\frac{1}{n} \log |\hat{B} \cap L_c| \to \frac{1}{2} \log \delta^2 + \varepsilon, \frac{1}{n} \log |\hat{B}' \cap L_K| \to \frac{1}{2} \log \sigma^2 + \varepsilon$ so that, for any linear combination Y of $\vec{Y}_{[K]}$, vector $\vec{\alpha} \in \mathbb{Z}^K$, matrix $\boldsymbol{A} \in \mathbb{R}^{* \times K}$ and $\vec{A} = \boldsymbol{A} \vec{Y}_c$, then*

$$Y = \mathcal{C} + \beta \tilde{Y} + \tilde{Y}_\perp,$$

where $\mathcal{C}, \mathcal{D}$ are functions of $(\vec{A}, \mathrm{mod}_{n_0 B_c}(Y_0), \vec{U}_{[K]}, \vec{W})$, β is some scalar estimation coefficient, and with high probability

$$\tilde{Y} = \mathcal{E}_\perp \left(\vec{\alpha}^\dagger Y_c \middle| \vec{A} \right) \in (\mathcal{D} + L_c) \cap \hat{B},$$

$$\tilde{Y}_\perp = \mathcal{E}_\perp\left(Y\middle|\vec{A}, \tilde{Y}\right) \in \hat{B}'.$$

In addition, $\delta^2 = \operatorname{var}\tilde{Y}$, $\sigma^2 = \max\{2^{-2r_K}, \operatorname{var}\tilde{Y}_\perp\}$.

Proofs for Lemmas 3, 4 are given in Appendix D.2. These lemmas do not strictly require that the sources be multivariate normal. This technical generalization is relevant in the application to the communication strategy in Section 5.3. Broadly, the proof of Theorem 8 goes as follows.

1. Choose some $\boldsymbol{\nu}_0 \in \mathbb{Z}^{K-1}$, $n_0 \in \mathbb{N}$. Apply Lemma 3 to U_K. Call $\tilde{Y}_\perp$ a 'residual.'

2. Choose some $\boldsymbol{\nu} \in \mathbb{Z}^K$. Apply Lemma 4 to the residual to break the residual $\tilde{Y}_\perp$ up into the sum of a lattice part due to $\boldsymbol{\nu}^\dagger \vec{Y}_{[K-1]}$ and a new residual, whatever is left over.

3. Repeat the previous step until the residual vanishes (up to $K-1$ times). Notice that this process has given several different ways of writing U_K; by stopping at any amount of steps, U_K is the modulo sum of several lattice components and a residual.

4. Design the lattice ensemble for the encoders such that the log-volume contributed to the support of U_K by each component can be estimated. The discrete parts will each contribute log-volume $\frac{1}{2}\log\delta^2$ and residuals will contribute log-volume $r_K + \frac{1}{2}\log\sigma^2$.

5. Recognize the entropy of U_K is no greater than the log-volume of its support. Choose the lowest support log-volume estimate of those just found.

Notice that each lemma application involves choice of some integer parameters. Choices which yield the strongest bound are unknown. Possible schemes for these

decisions are the subroutines Alpha0($\cdot$), Alpha($\cdot$), listed in Appendix D.1. As implemented, Alpha0($\cdot$) chooses $n_0 = 1$ and the integer linear combination $\boldsymbol{\nu}_0$ which leaves the least residual. As implemented, Alpha($\cdot$) chooses the integer linear combination $\boldsymbol{\nu}$ for which $\boldsymbol{\nu}^\dagger \vec{Y}_{[K-1]}$ is closest to being recoverable from current knowledge at each lemma application. It produces the combination for which the entropy $\frac{1}{2} \log \delta^2$ of the unknown part of $\boldsymbol{\nu}^\dagger \vec{Y}_{[K-1]}$ is minimized. This may be a suboptimal choice since, while such combinations are close to recoverable, they may not be very pertinent to a description of U_K. Nonetheless, it is still a good enough rule to produce nontrivial entropy bounds, as seen in Section 5.3.2.

5.2.1   Lower Bound

Entropies for the involved variables can be rearranged to adapt the upper bound in Theorem 8 into a lower bound. Compute:

$$
\begin{aligned}
H\left(U_K \middle| \vec{U}_{[K-1]}, \vec{W}\right) &= H\left(U_K \middle| \vec{Y}_{[K-1]}, \vec{W}\right) + I\left(U_K; \vec{Y}_{[K-1]} \middle| \vec{U}_{[K-1]}, \vec{W}\right) \\
&= H\left(U_K \middle| \vec{Y}_{[K-1]}, \vec{W}\right) + H\left(\vec{Y}_{[K-1]} \middle| \vec{U}_{[K-1]}, \vec{W}\right) - \ldots \\
&\quad H\left(\vec{Y}_{[K-1]} \middle| \vec{U}_{[K]}, \vec{W}\right) \\
&= H\left(U_K \middle| \vec{Y}_{[K-1]}, \vec{W}\right) + H\left(\vec{Y}_{[K-1]} \middle| \vec{W}\right) - \ldots \\
&\quad H\left(\vec{U}_{[K-1]} \middle| \vec{W}\right) - H\left(\vec{Y}_{[K-1]} \middle| \vec{U}_{[K]}, \vec{W}\right). \\
&= H\left(Y_s \middle| \vec{Y}_{[K-1]}, \vec{W}\right) + H\left(\vec{Y}_{[K-1]} \middle| \vec{W}\right) - \ldots \\
&\quad H\left(Y_s \middle| U_K, \vec{Y}_{[K-1]}, \vec{W}\right) - \ldots \\
&\quad H\left(\vec{U}_{[K-1]} \middle| \vec{W}\right) - H\left(\vec{Y}_{[K-1]} \middle| \vec{U}_{[K]}, \vec{W}\right) \\
&= H\left(\vec{Y}_{[K]} \middle| \vec{W}\right) - H\left(Y_s \middle| U_K, \vec{Y}_{[K-1]}, \vec{W}\right) - \ldots \qquad (5.1) \\
&\quad H\left(\vec{U}_{[K-1]} \middle| \vec{W}\right) - H\left(\vec{Y}_{[K-1]} \middle| \vec{U}_{[K]}, \vec{W}\right).
\end{aligned}
$$

55

The entropy-rate of the summand in Equation (5.1) is given by Lemma 8. Each subtracted summand requires an upper bound. These are provided by Theorem 8 and Theorem 11. Lemma 8 and Theorem 11 are provided in Appendix E.

5.3 Lattice-Based Strategy for Communication via Decentralized Processing

Consider a scenario where a decoder seeks to decode a message from a single-antenna broadcaster in an additive white Gaussian noise (AWGN). The decoder does not observe a signal directly but instead is provided information by a collection of distributed observers ('helpers') which forward it digital information, each observer-to-decoder link supporting a different communications rate. This network is depicted in Figure 1.1. A block diagram is shown in Figure 1.2. This is the problem of a single-antenna transmitter communicating to a decoder informed out-of-band by a network of helpers in the presence of additive white Gaussian noise and interference.

Note that this problem is different from the problem of distributed source coding of a linear function Tavildar and Wagner (2009); Wagner (2011); Yang and Xiong (2014, 2011); Chapman *et al.* (2019). In contrast to the source coding problem, the signal being preserved by the many-help-one network in the present case has a codebook structure. This structure can be exploited to improve the source-to-decoder communications rate. This problem has been studied Sanderovich *et al.* (2008); Chapman *et al.* (2018), but the best achievable rate is still unknown. In this section, we present a strategy that takes advantage of this codebook structure.

The core of the strategy is to apply a slight modification of Theorem 8 to the network. The transmitter modulates its communications message using a nested lattice codebook such as one in Erez and Zamir (2004). The helpers employ lattice encoders such as those from Theorem 8, and then perform Slepian–Wolf distributed lossless compression (El Gamal and Kim, 2011, Theorem 10.3) on their encodings to further

reduce their rate. Because the codeword appears as a component of all the helper's observations, the bound on the message's joint entropy obtained from Theorem 8 can be strengthened, allowing one to use a more aggressive post-compression stage.

5.3.1 Description of the Communication Scheme

It is well known that a nested lattice codebook with dither achieves Shannon information capacity in a point-to-point AWGN channel with a power-constrained transmitter Erez and Zamir (2004). One interesting aspect of the point-to-point communications scheme described in Erez and Zamir (2004) is that decoding of the noisy signal is done in modulo space. We will see in this section how lattice encodings like those in Theorem 8 can be used to provide such a decoder enough information to recover a communications message.

Without loss of generality, assume that the transmitter is limited to have average transmission power 1. The scheme's codebook is designed from nested lattices $L_{f,\mathrm{msg}} \supset L_{c,\mathrm{msg}}$ with base regions $B_{f,\mathrm{msg}}, B_{c,\mathrm{msg}}$. $L_{f,\mathrm{msg}}$ is chosen to be good for coding and $L_{c,\mathrm{msg}}$ good for quantization. The messaging rate of this codebook is determined by the nesting ratio of $L_{c,\mathrm{msg}}$ in $L_{f,\mathrm{msg}}$:

$$R_{\mathrm{msg}} := \frac{1}{n} \log |L_{f,\mathrm{msg}} \cap B_{c,\mathrm{msg}}|.$$

Lattices can be designed with nesting ratios such that any rate above zero can be formed. Taking a message $M \in L_{f,\mathrm{msg}} \cap B_{c,\mathrm{msg}}$ and choosing a dither $W_{\mathrm{msg}} \sim -B_{c,\mathrm{msg}}$ of which the decoder is informed, then the codeword associated with M is:

$$X^n_{\mathrm{msg}}(M) := \frac{\mathrm{mod}_{L_{c,\mathrm{msg}}} (M + W_{\mathrm{msg}})}{\sqrt{\mathrm{var}\, W_{\mathrm{msg}}}} \in \frac{B_{L_{c,\mathrm{msg}}}}{\sqrt{\mathrm{var}\, W_{\mathrm{msg}}}} \subset \mathbb{R}^n.$$

We now describe observations of such a signal by helpers in the presence of AWGN interferers. For covariance $\mathbf{\Sigma}_{\mathrm{noise}} \in \mathbb{R}^{K \times K}$, take

$$\vec{X}^n_{\mathrm{noise}} = (X^n_{\mathrm{noise},1}, \ldots, X^n_{\mathrm{noise},K}) \in (\mathbb{R}^n)^K$$

to be n independent draws from the joint-distribution $\mathcal{N}(0, \boldsymbol{\Sigma}_{\text{noise}})$. In addition, take a random vector X^n_{msg} as described at the beginning of Section 5.3.1 and a vector $\boldsymbol{c}_{\text{msg}} \in \mathbb{R}^K$ and define $\boldsymbol{\Sigma}_{\text{msg}} := \boldsymbol{c}_{\text{msg}} \boldsymbol{c}^\dagger_{\text{msg}}$. Now, the k-th helper observes the vector:

$$X^n_k = [\boldsymbol{c}_{\text{msg}}]_k X^n_{\text{msg}} + X^n_{\text{noise},k} \in \mathbb{R}^n.$$

Form an observations vector:

$$\vec{X}^n := \boldsymbol{c}_{\text{msg}}(X^n_{\text{msg}}) + \vec{X}^n_{\text{noise}} \in (\mathbb{R}^n)^K,$$

and finally form a cumulative time-averaged covariance matrix as

$$\boldsymbol{\Sigma} := \text{var}\, \vec{X}^n = \boldsymbol{c}_{\text{msg}} \boldsymbol{c}^\dagger_{\text{msg}} + \boldsymbol{\Sigma}_{\text{noise}} \in \mathbb{R}^{K \times K}.$$

If helpers are informed of message dither W_{msg}, then they are informed of the codebook for X_{msg} and its lattice structure. Using lattice encoders such as those described in Theorem 8, this codebook information can be used to strengthen the upper bound on conditional entropies between the messages.

Unfortunately, the lower bound from Theorem 8 cannot be strengthened in the same way. With the added dither term W_{msg}, an argument that a particular inter-action information equals zero (Equations (E.3)-(E.4)) no longer holds. It is unclear whether or not this complication is fundamental to the lower bound argument used. The usual lower bound from Theorem 8 still holds but is looser in this context.

Theorem 9. *In the context of the channel description given in Section 5.3.1, entropy bounds identical to those from Theorem 8 hold for its described observer encodings. The bounds also hold re-defining:*

$$Y_0 := X_{msg},$$

defining the rest of the variables in the theorem as stated. The bounds also hold instead re-defining:

$$Y_c := (Y_0 - Y_K, Y_1, \ldots, Y_{K-1}, X_{src}),$$

vectors $\{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{K+1}\} \subset \mathbb{Z}^{K+1}$ a basis where all vectors but one $\boldsymbol{\nu}_s, s \in [K+1]$ have 0 as their $(K+1)$-th component and $\boldsymbol{\nu}_s = [0, 0, \ldots, 0, 1]^\dagger$, taking

$$\vec{a}_{\mathbb{R}}^{(msg)} \in \text{image STAGES}^* \left(\text{var} \left([\vec{Y}_c]_{[K]} \middle| (\boldsymbol{\nu}_j^\dagger \vec{Y}_c)_{0<j<s} \right) \right),$$

$$\vec{a}_{\mathbb{Z}}^{(msg)} \in \mathbb{Z}^K,$$

$$\lambda^{(msg)} := \text{cov}(X_{msg}^n, (\vec{a}_{\mathbb{R}}^{(msg)} + \vec{a}_{\mathbb{Z}}^{(msg)})^\dagger [\vec{Y}_c]_{[K]}),$$

$$Y_\perp^{(msg)} := \mathcal{E} \left((\vec{a}_{\mathbb{R}}^{(msg)} + \vec{a}_{\mathbb{Z}}^{(msg)})^\dagger [\vec{Y}_c]_{[K]} \middle| X_{msg}^n \right),$$

$$\delta_{(msg)}^2 := \left(\frac{\lambda^{(msg)}}{\gamma_n} - 1 \right)^2 + \text{var} \, Y_\perp^{(msg)},$$

$$\delta_s^2 := \max\{1, \frac{\delta_{(msg)}^2}{2^{-2r_{msg}}} + \varepsilon\},$$

and taking the rest of the variables in the theorem as stated over range $k \in [K+1]$.

A sketch for Theorem 9 is provided in Appendix D.3. The theorem's statement can be broadly understood in terms of the proof of Theorem 8. After a number of steps s in the support analysis for Theorem 8, the codebook component X_{msg}^n can be partially decoded yielding tighter estimation of that component's contribution to the support of U_K. The variables $\lambda^{(msg)}, \vec{a}_{\mathbb{R}}^{(msg)}, \vec{a}_{\mathbb{Z}}^{(msg)}$ are parameters for this partial decoding. Lattice modulo messages such as those described in Theorem 9 can be recombined in a useful way:

Lemma 5. *For $\varepsilon > 0$ and vectors $\mathbf{a}_{\mathbb{Z}} \in \mathbb{Z}^K$, $\mathbf{a}_{\mathbb{R}} \in \text{image STAGES}^*(\boldsymbol{\Sigma}) \subset \mathbb{R}^K$, then lattice modulo encodings $\vec{U}_{[K]}$ from Theorem 9 can be processed into:*

$$U_{proc} := \text{mod}_{L_{c,msg}} \left(\lambda X_{msg} + Y_{noise} \right), \tag{5.2}$$

where $\lambda \in \mathbb{R}$ is some constant:

$$\lambda := \mathrm{cov}\left(X_{msg}^n, (\mathbf{a}_\mathbb{Z} + \mathbf{a}_\mathbb{R})^\dagger \vec{Y}_{[K]}\right)$$

and the noise term Y_{noise} has the following properties:

- $\sigma_{noise}^2 := \mathrm{var}\, Y_{noise} = \mathrm{var}\left((\mathbf{a}_\mathbb{Z} + \mathbf{a}_\mathbb{R})^\dagger \vec{Y}_{[K]} \Big| X_{msg}^n\right),$

- $Y_{noise} \perp (X_{msg}, M, W_{msg}),$

- Y_{noise} *is with high probability in the base cell of any lattice good for coding semi norm-ergodic noise up to power $\sigma_{noise}^2 + \varepsilon$.*

Lemma 5 is demonstrated in Appendix D.4. Notice that Equation (5.2) is precisely the form of signal processed by the communications decoder described in Erez and Zamir (2004). The following result summarizes the performance of this communications strategy.

Corollary 2. *Fix a codebook rate $r_{msg} > 0$. As long as helper-to-decoder messaging rates $R_1, \ldots, R_K > 0$ satisfy all the following criteria:*

$$\forall S \subset [K], \ \sum_{k \in S} R_k > \tilde{H}(S|[K]\backslash S) + \varepsilon, \tag{5.3}$$

each $\tilde{H}(S|[K]\backslash S)$ being any entropy-rate bound obtained from Theorem 9, then the following communications rate from source to decoder is achievable, taking $\mathbf{a}_\mathbb{Z}, \mathbf{a}_\mathbb{R}, \lambda, \sigma_{noise}^2$ from their definitions in Lemma 5:

$$R_{msg} < \min\left\{r_{msg}, \ \sup_{\mathbf{a}_\mathbb{Z}, \ \mathbf{a}_\mathbb{R}} \max_{\gamma^2 \in (0,1]} \frac{1}{2}\log\left[\frac{\gamma^2}{(\lambda - \gamma)^2 + \sigma_{noise}^2}\right]\right\}. \tag{5.4}$$

Proof for Corollary 2 is given in Appendix D.5, and evaluation of the achieved communications rates for certain lattice code designs is shown in Section 5.3.2.

5.3.2   Numerical Results

The achievable rate given in Corollary 2 depends on the design of the lattice encoding scheme at the helpers. Identification of the best such lattice encoders for such a system is closely tied to a receivers' covariance structure Chapman *et al.* (2019). For this reason and for the purpose of evaluating the effect of joint compression stage, we restrict our attention to a particular channel structure and lattice encoder design.

The line-of-sight configuration shown in Figure 5.2 is considered. It yields helper observations with the following covariance structure, in Figure 5.2 labeling interfering sources W_{I1}, W_{I2}, and W_{I3} and helpers from top to bottom:

$$X_{raw,1} = \frac{\sqrt{P_S}}{\|1 + (\frac{2}{3})e^{i\pi \cdot 1/2}\|} X_{msg} + W_1 + \ldots$$
$$\frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 1/2} - e^{i\pi \cdot 2/3})\|} W_{I1} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 1/2} - e^{i\pi \cdot 1})\|} W_{I2} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 1/2} - e^{i\pi \cdot 4/3})\|} W_{I3},$$

$$X_{raw,2} = \frac{\sqrt{P_S}}{\|1 + (\frac{2}{3})e^{i\pi \cdot 5/6}\|} X_{msg} + W_2 + \ldots$$
$$\frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 5/6} - e^{i\pi \cdot 2/3})\|} W_{I1} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 5/6} - e^{i\pi \cdot 1})\|} W_{I2} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 5/6} - e^{i\pi \cdot 4/3})\|} W_{I3},$$

$$X_{raw,3} = \frac{\sqrt{P_S}}{\|1 + (\frac{2}{3})e^{i\pi \cdot 7/6}\|} X_{msg} + W_3 + \ldots$$
$$\frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 7/6} - e^{i\pi \cdot 2/3})\|} W_{I1} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 7/6} - e^{i\pi \cdot 1})\|} W_{I2} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 7/6} - e^{i\pi \cdot 4/3})\|} W_{I3},$$

$$X_{raw,4} = \frac{\sqrt{P_S}}{\|1 + (\frac{2}{3})e^{i\pi \cdot 3/2}\|} X_{msg} + W_4 + \ldots$$
$$\frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 3/2} - e^{i\pi \cdot 2/3})\|} W_{I1} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 3/2} - e^{i\pi \cdot 1})\|} W_{I2} + \frac{\sqrt{P_I}}{\|(\frac{2}{3})(e^{i\pi \cdot 3/2} - e^{i\pi \cdot 4/3})\|} W_{I3},$$

$W_k \sim \mathcal{N}(0,1)$ i.i.d.
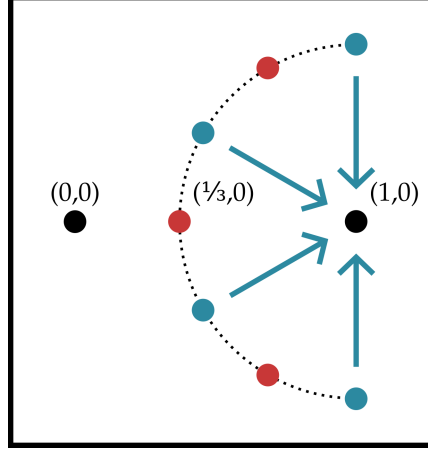
Figure 5.2: The line-of-sight channel considered. A black transmit node at $(0,0)$ seeks to communicate with a black decoder node at $(1,0)$. Three red 'interferer' nodes broadcast an independent Gaussian signal, each interferer has its own signal. The decoder does not observe any signal directly but is forwarded messages from four blue 'helper' nodes which observe signals through a line-of-sight additive-white-Gaussian noise channel. The interferers and helpers are oriented alternatingly and equispaced about a radius-$2/3$ semicircle towards the encoder with center $(1,0)$.

where $P_S, P_I > 0$ are signal, interferer powers, respectively. Choice of this channel is arbitrary but provides an instance where the decoder would not be able to recover the signal of interest if it observed directly without the provided helper messages.

Communications Schemes

First, we describe a class of lattice encoders the four helpers could employ:

- Fix some $c \in (0,3)$. If helper $k \in [4]$ in the channel from Figure 5.2 observes $X_{\mathrm{raw},k}^n$, then it encodes a normalized version of the signal:

$$X_k^n := \frac{c}{\sqrt{\mathrm{var}\, X_{\mathrm{raw},k}^n}} X_{\mathrm{raw},k}^n.$$

- Fix equal lattice encoding rates per helper $r = r_1 = r_2 = r_3 = r_4$, and take lattice encoders as described in Theorem 8. These rates may be distinct from the helper-to-base messaging rates $R_1, \ldots, R_4$ if the encodings are compressed afterwards; after compression $R_k \leq r_k$.

Communications schemes involving lattice encoders of this form are compared in Figure 5.3 over an ensemble of choices for lattice encoder rates r and scales $c \in (0, 3)$. Achieved transmitter-to-decoder communication rate versus sum-rate from helpers to decoder are plotted. The following quantities are plotted:

- *Upper Bound*: An upper bound on the achievable transmitter-to-decoder communications rate, corresponding to helpers which forward with infinite rate. This bound is given by the formula $I(X_{msg}; (X_{\mathrm{raw},k})_{k \in [4]})$.

- *Corollary 2* The achievable communications rate from Corollary 2, where each helper computes the lattice encoding described above, then employs a joint–compression stage to reduce its messaging rate. The sum-helpers-to-decoder rate for this scheme is given by Equation (5.3), taking $S = [4]$. The achieved messaging rate is given by the right-hand-side of Equation (5.4).

- *Uncompressed Lattice*: The achievable communications rate from Corollary 2, with each helper forwarding to the decoder its entire lattice encoding without joint–compression. The sum-helpers-to-decoder rate for this scheme is $4r$ since in this scheme each helper forwards to the base at rate $R_k = r$. The achieved messaging rate is given by the right-hand-side of Equation (5.4).

- *Quantize & Forward*: An achievable communications rate where helper-to-decoder rates $R_k, k \in [4]$ are chosen so that $R_1 + R_2 + R_3 + R_4 = R_{sum}$ and each helper forwards a rate-distortion-optimal quantization of its observation to the

decoder. The decoder processes these quantizations into an estimate of X_{msg} and decodes. This is discussed in more detail in Chapman *et al.* (2018). The sum-helpers-to-decoder rate for this scheme is R_{sum}. The achieved messaging rate is $I(X_{msg}; (X_{raw,k} + Z_k)_{k \in [4]})$, where $Z_k \sim \mathcal{N}(0, \mathrm{var}(X_{raw,k}) \cdot 2^{-2R_k})$.

Performance of these strategies for different broadcaster powers is shown in Figure 5.3. In each subplot the transmitter broadcasts with power such that the average SNR seen across helpers is the given 'transmitter' dB figure. Each interferer broadcasts its own signal with its power the given 'interferer' dB stronger than the transmitter's power. Notice that, although the uncompressed lattice scheme is often outperformed by plain Quantize & Forward for the same helper message rates, adding a properly configured compression stage can more than make up for the sum-rate difference. In certain regimes, even the compressed lattice scheme performs worse or practically the same as Quantize & Forward, indicating the given lattice encoder design is weak; uncompressed lattice encoders can be configured to implement the Quantize & Forward scheme. Notice that none of the strategies produce convex rate regions, indicating that time-sharing can be used to achieve better rates in some regimes.

In all figures shown, the gap between achieved rates from the joint–compression bound given from Theorem 8 and Theorem 9 (the latter being an improvement) were often nonzero but too small to noticeably change the graphs in Figure 5.3. For this reason only, achievable rates for the strategy from Corollary 2 are plotted. The gain from involving codebook knowledge in lattice encoding compression is either insignificant for the tested scenario, or choices in computing the upper bounds are too poor to reveal its performance gains. Sub-optimality of the algorithm implementations here are all summarized and discussed in Section 5.4.
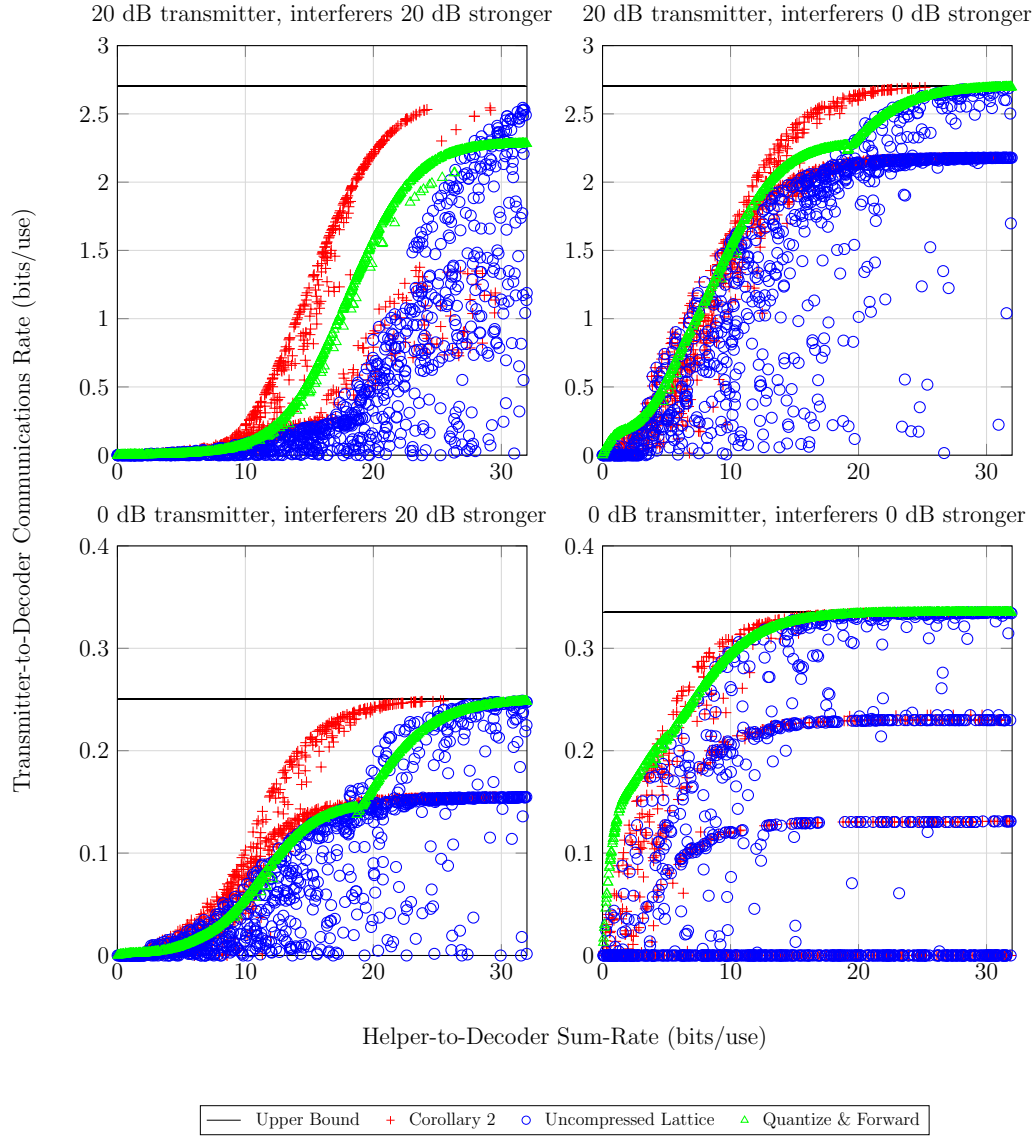
Figure 5.3: Communications rate versus helper-sum-rate for 1000 randomly chosen encoding schemes as described in Section 5.3.2 in the line-of-sight channel from Figure 5.2, Equation 5.3.2.

5.4 Conclusion

A class of upper bounds on the joint entropy of lattice-modulo encodings of correlated Gaussian signals was presented in Theorem 8. Proof of these bounds involves reducing the problem to the entropy of one lattice message, say, U_K conditioned on the rest, $\vec{U}_{[K-1]}$. The upper bound for this reduced case involves an iterative construction where in each step a suitable integer vector is chosen. Choice of vectors in these steps determines the order in which the observed lattice-modulo components are integrated into an estimate of U_K's support. Different choice of vectors at each step yields a different bound, and the strongest sequence of choices is unknown. For numerical results in Section 5.3.2, a certain suboptimal was used although there is no guarantee that this choice is optimal.

The upper bounds were applied to the problem of communicating through a many-help-one network, and these bounds were evaluated for a rendition of the problem using lattice codes of simple structure. The bounds in 8 can be strengthened in this scenario by integrating codebook knowledge. This strengthening is described in Theorem 9.

In spite of the suboptimal lattice encoder designs analyzed, it was seen in Section 5.3.2 that jointly-compressed lattice encoders are able to significantly outperform more basic schemes in the presence of heavy interference, even when the joint compression stage uses the weaker entropy bounds from Theorem 8. In the numerical experiments tried, the strengthening in Theorem 9 was not seen to significantly improve compression. Whether this is typically true or just an artifact of poor design of the joint-compression stage is unknown. In either case, the simpler joint-compression strategy without codebook knowledge was seen to improve performance.

The most immediate forwards steps to the presented results is in characterization of the search problems posed by Theorems 8, 9. Although not discussed, corner-points of joint compression described here are implementable using further lattice processing on the encodings $U_1, \ldots, U_K$ and their dithers $\vec{W}$. Such a process might mimic the compression procedure described in Cheng *et al.* (2018). Tightness arguments from this work may also apply to the present less structured channel.

Finally, according to the transmission method in Cheng *et al.* (2018), the achievable rate in Corollary 2 may be improvable by breaking the transmitter's message M up into a sum of multiple components, each from a finer lattice. Joint–compression for such a transmission could integrate codebook information from each message component separately, allowing for more degrees of freedom in the compression stage's design, possibly improving the achievable rate. This is an extension of the argument in Appendix D.4. These improvements are out of this chapter's scope but provide meaningful paths forward.

Chapter 6

CONCLUSION

The problem of communicating in the presence of correlated Gaussian noise was studied. Analytic techniques involving unstructured coding are presented in Chapter 3. A first-order lattice strategy which subsumes others of its type is presented in Chapter 4. Evaluation of the first-order scheme's asymptotic performance involves a difficult search problem over lattice encoder configurations, but numerical experiments demonstrate that it reliably outperforms related schemes. A one-dimensional simulation of the lattice strategy was simulated and seen to perform more reliably over ensembles of channels than a simpler encoding scheme.

A second-order strategy involving joint compression of messages from Chapter 4 is presented in Chapter 5. The joint-compression stage is developed by describing upper bounds on the joint-entropy of lattice encodings. The upper bound can be strengthened when it is known that the source of interest uses a lattice codebook. The strategy for upper-bounding the joint entropy can be adapted into a lower bound. Evaluation of these joint-entropies for a given lattice encoder configuration introduces another search problem. Rather than attempt to solve the joint-search to evaluate the strategy's performance, certain configuration designs were studied. In spite of

the sub-optimal choices the asymptotic joint-compression scheme could significantly outperform simpler encoding schemes.

6.1 Discussion and Avenues for Further Study

6.1.1 Improvement of Joint-Entropy Bound

Though the present study gives handles for computing performance and entropy figures for the schemes described, the analytic behavior of these figures is poorly understood. The gap between the upper bound in Theorem 8 and the lower bound provided by Section 5.2.1 is unknown. Search space reductions for the problems involved in numerical evaluation of the bounds have not been investigated.

6.1.2 Implementation of Structured Schemes and Sensitivity to Channel State Estimation Error

The essential ingredients to implementation of the structured scheme described in Chapter 4 is a high-dimensional lattice good for coding and quantization with readily computable rounding and modulo operations. Much effort has been devoted to such a construction Liu *et al.* (2018), da Silva and Silva (2018), Conway and Sloane (1982), Ingber *et al.* (2012). Given such a lattice, certain nesting constructions are easy to obtain. For instance, for a a whole number, any lattice L in dimension n is nested within the scaled lattice $(\frac{1}{a})L$ with nesting ratio $\log a$.

Sensitivity of actual implementations to low-dimensional, imperfect lattice design has not been investigated, though simulation of numerical results in Chapter 4 demonstrates such techniques can provide improvement even in one dimension. Numerical experiments in Section 5.3.2 suggest existence of lattice encoder configurations which, though sub-optimal for particular channels, have low outage rate over a wide ensemble of channels.

The description of the conditional entropy upper bounds in Chapter 5 implicitly describes a scheme for joint-compression of such messages. The derived compression and decompression schemes involve only further lattice processing of the messages. The decompression stage would form lattice of judicious recombinations of the described rounding results.

Proofs in D.2 involve somewhat intricate nested lattice and typical set constructions, but many of properties are either never used or are only of theoretical importance. For instance, the property that fine lattices are nested within one another is never used. Roughly speaking, the entire utility of auxillary lattices $\hat{L}_k$ is in upperbounding the amount of points in L_c which occur within a ball mostly contained in $\hat{B}_k$. Estimating the amount of lattice points L_c occuring within a sphere is analytically involved, but counting those in $\hat{B}_k$ is easy given that $\hat{L}_k$ was constructed. For practical purposes, existence of $\hat{L}_k$ is likely unnecessary.

6.1.3   Improvement of Joint Compression Stage

The joint-compression stage for lattice encodings described in Chapter 5 supports some immediate improvements. For the communications problem described in Definition 1, instead of designing the compression stage such that the lattice encodings are first recovered and then processed into a message, one could instead design the compression stage such that only the message M is preserved, recovered directly from the compressions provided by the observers. Similarly for the source-coding problem described in Definition 2, one could design the compression stage such that only the lattice recombination of interest is recovered, rather than the lattice encodings first followed by lattice processing. The rate region of this lossless coding problem is treated by Gelfand and Pinsker in Gel'fand and Pinsker (1979) (result stated in english in Theorem 1, Wolf *et al.* (2017)). However, in this situation the path to im-

70

plementation as described in Section 6.1.2 is less immediate. More granular structure is needed to organize each source sequence.

6.1.4   MIMO Problems and Channels with Doppler

A very strong assumption about observed Doppler was made in Section 1.1.1. It may be possible to relax this assumption by assuming that although Doppler shifts may distinct between a transmitter and each observer, shifts still evolve slowly over a period of many OFDM symbols. In this situation noise may appear correlated over different subcarriers across receivers. Such a situation would require an expansion of the described lattice schemes to handle joint processing of multiple subcarriers. Such extensions would also be relevant to a setting where the base seeks to recover multiple sources, or where receivers observe through multiple antennas.

REFERENCES

Berger, T., Z. Zhang and H. Viswanathan, "The ceo problem [multiterminal source coding]", IEEE Transactions on Information Theory **42**, 3, 887–902 (1996).

Bliss, D. W. and S. Govindasamy, *Adaptive wireless communications: MIMO channels and networks* (Cambridge University Press, 2013).

Buchmann, J. and M. Pohst, "Computing a lattice basis from a system of generating vectors", in "European Conference on Computer Algebra", pp. 54–63 (Springer, 1987).

Campello, A., C. Ling and J.-C. Belfiore, "Universal lattice codes for mimo channels", IEEE Transactions on Information Theory **64**, 12, 7847–7865 (2018).

Chapman, C. and D. W. Bliss, "Upper bound on the joint entropy of correlated sources encoded by good lattices", (in review, submitted to Entropy by MDPI) **N/A**, N/A, N/A (2019).

Chapman, C., M. Kinsinger, A. Agaskar and D. W. Bliss, "Distributed recovery of a gaussian source in interference with multi-stage lattice decoding", (Accepted with minor revisions, to appear in Entropy by MDPI) **N/A**, N/A, N/A (2019).

Chapman, C. D., H. Mittelmann, A. R. Margetts and D. W. Bliss, "A decentralized receiver in gaussian interference", Entropy **20**, 4, 269 (2018).

Cheng, H., X. Yuan and Y. Tan, "Generalized compute-compress-and-forward", IEEE Transactions on Information Theory **65**, 1, 462–481 (2018).

Choi, J., D. J. Love and T. P. Bidigare, "Coded distributed diversity: A novel distributed reception technique for wireless communication systems", IEEE Transactions on Signal Processing **63**, 5, 1310–1321 (2015).

Conway, J. and N. Sloane, "Fast quantizing and decoding and algorithms for lattice quantizers and codes", IEEE Transactions on Information Theory **28**, 2, 227–232 (1982).

Cover, T. M. and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).

da Silva, P. R. B. and D. Silva, "Multilevel ldpc lattices with efficient encoding and decoding and a generalization of construction d'", IEEE Transactions on Information Theory **65**, 5, 3246–3260 (2018).

Diggavi, S. N. and T. M. Cover, "The worst additive noise under a covariance constraint", IEEE Transactions on Information Theory **47**, 7, 3072–3081 (2001).

El Gamal, A. and Y.-H. Kim, *Network information theory* (Cambridge university press, 2011).

Erez, U., S. Litsyn and R. Zamir, "Lattices which are good for (almost) everything", IEEE Transactions on Information Theory **51**, 10, 3401–3416 (2005).

Erez, U. and R. Zamir, "Achieving $\frac{1}{2}\log(1 + \mathrm{SNR})$ on the awgn channel with lattice encoding and decoding", IEEE Transactions on Information Theory **50**, 10, 1 (2004).

Gel'fand, S. I. and M. S. Pinsker, "Coding of sources on the basis of observations with incomplete information", Problems of Information Transmission **15**, 2, 45–57 (1979).

Ghasemmehdi, A. and E. Agrell, "Faster recursions in sphere decoding", IEEE Transactions on Information Theory **57**, 6, 3530–3536 (2011).

Hastad, J., B. Just, J. C. Lagarias and C.-P. Schnorr, "Polynomial time algorithms for finding integer relations among real numbers", SIAM Journal on Computing **18**, 5, 859–881 (1989).

He, W. and B. Nazer, "Integer-forcing source coding: Successive cancellation and source-channel duality", in "2016 IEEE International Symposium on Information Theory (ISIT)", pp. 155–159 (IEEE, 2016).

Ihara, S., "On the capacity of channels with additive non-gaussian noise", Information and Control **37**, 1, 34–39 (1978).

Ingber, A., R. Zamir and M. Feder, "Finite-dimensional infinite constellations", IEEE Transactions on Information Theory **59**, 3, 1630–1656 (2012).

Jones, E., T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python", URL `http://www.scipy.org/`, [Online; accessed ¡today¿] (2001–).

Krithivasan, D. and S. S. Pradhan, "Lattices for distributed source coding: Jointly gaussian sources and reconstruction of a linear function", in "International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes", pp. 178–187 (Springer, 2007).

Krithivasan, D. and S. S. Pradhan, "Distributed source coding using abelian group codes: A new achievable rate-distortion region", IEEE Transactions on Information Theory **57**, 3, 1495–1519 (2011).

Lim, S. H., C. Feng, A. Pastore, B. Nazer and M. Gastpar, "A joint typicality approach to compute–forward", IEEE Transactions on Information Theory **64**, 12, 7657–7685 (2018).

Liu, L., Y. Yan, C. Ling and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices", IEEE Transactions on Communications **67**, 2, 915–928 (2018).

Maddah-Ali, M. A. and D. Tse, "Interference neutralization in distributed lossy source coding", in "Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on", pp. 166–170 (IEEE, 2010).

Nazer, B., V. R. Cadambe, V. Ntranos, and G. Caire, "Expanding the compute-and-forward framework: Unequal powers, signal levels, and multiple linear combinations", IEEE Transactions on Information Theory **62**, 9, 4879–4909 (2016).

Nazer, B. and M. Gastpar, "The case for structured random codes in network capacity theorems", European Transactions on Telecommunications **19**, 4, 455–474 (2008).

Ordentlich, O. and U. Erez, "Precoded integer-forcing universally achieves the mimo capacity to within a constant gap", IEEE Transactions on Information Theory **61**, 1, 323–340 (2014).

Ordentlich, O. and U. Erez, "A simple proof for the existence of "good" pairs of nested lattices", IEEE Transactions on Information Theory **62**, 8, 4439–4453 (2016).

Ordentlich, O. and U. Erez, "Integer-forcing source coding", IEEE Transactions on Information Theory **63**, 2, 1253–1269 (2017).

Ordentlich, O., U. Erez and B. Nazer, "Successive integer-forcing and its sum-rate optimality", in "2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)", pp. 282–292 (IEEE, 2013).

Popat, A. C., *Scalar Quantization with Arithmetic Coding*, Ph.D. thesis, Massachusetts Institute of Technology (1990).

Pradhan, S. S. and K. Ramchandran, "Distributed source coding using syndromes (discus): Design and construction", in "Data Compression Conference, 1999. Proceedings. DCC'99", pp. 158–167 (IEEE, 1999).

Sanderovich, A., S. Shamai, Y. Steinberg and G. Kramer, "Communication via decentralized processing", IEEE Transactions on Information Theory **54**, 7, 3008–3023 (2008).

Schnorr, C.-P. and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems", Mathematical Programming **66**, 1-3, 181–199 (1994).

Shirani, F., M. Heidari and S. S. Pradhan, "Quasi linear codes: Application to point-to-point and multi-terminal source coding", in "2016 IEEE International Symposium on Information Theory (ISIT)", pp. 730–734 (IEEE, 2016).

Stankovic, V., A. D. Liveris, Z. Xiong and C. N. Georghiades, "On code design for the slepian-wolf problem and lossless multiterminal networks", IEEE transactions on Information Theory **52**, 4, 1495–1507 (2006).

Tavildar, P. V., Saurabha and A. B. Wagner, "The gaussian many-help-one distributed source coding problem", IEEE Transactions on Information Theory **56**, 1, 564–581 (2009).

Wagner, A. B., "On distributed compression of linear functions", IEEE Transactions on Information Theory **57**, 1, 79–94 (2011).

Wang, S., Y. Fang and S. Cheng, *Distributed Source Coding: Theory and Practice* (John Wiley & Sons, 2017).

Wolf, A., D. C. González, M. Dörpinghaus, G. Fettweis *et al.*, "On the binary lossless many-help-one problem with independently degraded helpers", arXiv preprint arXiv:1701.06416 (2017).

Yang, Y. and Z. Xiong, "An improved lattice-based scheme for lossy distributed compression of linear functions", in "2011 Information Theory and Applications Workshop", pp. 1–5 (IEEE, 2011).

Yang, Y. and Z. Xiong, "Distributed compression of linear functions: Partial sum-rate tightness and gap to optimal sum-rate", IEEE Transactions on Information Theory **60**, 5, 2835–2855 (2014).

Zamir, R., *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory* (Cambridge University Press, 2014).

Zamir, R. and M. Feder, "On universal quantization by randomized uniform/lattice quantizers", IEEE Transactions on Information Theory **38**, 2, 428–436 (1992).

Zamir, R. and M. Feder, "On lattice quantization noise", IEEE Transactions on Information Theory **42**, 4, 1152–1159 (1996).

Zamir, R., S. Shamai and U. Erez, "Nested linear/lattice codes for structured multiterminal binning", IEEE Transactions on Information Theory **48**, 6, 1250–1276 (2002).

APPENDIX A

KEY OF VARIABLES

K	Number of lattice encodings in current context.
n	Scheme blocklength
W_k	Lattice dither k
U_k	Lattice encoding k
$X_{k,\text{raw}}$	Observation at receiver k, before normalizing.
X_k	Normalized observation at receiver k, i.e. $X_k / \operatorname{var} X_k$
Y_k	Quantization of X_k^n
$\mathbf{\Sigma}$	$K \times K$ time-averaged covariance between observations $X_1^n, \ldots, X_K^n$
$\mathbf{\Sigma}_Q$	$K \times K$ time-averaged covariance between quantizations $Y_1, \ldots, Y_K$
$R_1, \ldots, R_K$	Side information messaging rates for helpers in the Section 5.3 communications scenario
$\alpha_1, \ldots, \alpha_K$	Scales for lattice encoders to use on normalized observations
L_c	Central coarse lattice, good for quantization and coding with scale 1
$r_1, \ldots, r_K$	Nesting ratios for fine lattices $L_1, \ldots, L_K$ in the coarse lattice base region B_c, equivalent to the encoding rates of lattice codes when joint compression is not used
r_{msg}	Nesting ratio for codebook coarse lattice $L_{c,\text{msg}}$ in codebook fine lattice $L_{f,\text{msg}}$ in Section 5.3, equivalent to codebook rate
$\boldsymbol{c}_{\text{msg}}$	Covariance between codeword and quantizations in Section 5.3
$\boldsymbol{\nu}_s$	Integer combination of $\vec{Y}_c$ to analyze in step s of Appendix D.2
δ_s^2	Variance of $\boldsymbol{\nu}_s^\dagger \vec{Y}_c$ after removing prior knowledge in Appendix D.2
σ_s^2	Variance of Y_K uncorrelated with prior knowledge and $\boldsymbol{\nu}_s^\dagger \vec{Y}_c$ in Appendix D.2
β_s	Regression coefficient for $\boldsymbol{\nu}_s^\dagger \vec{Y}_c$ in Y_K after including prior knowledge at step s in Appendix D.2

Table A.1: General Description of Common Variables in Body and Appendices

APPENDIX B

SUPPLEMENTS TO UNSTRUCTURED BOUNDS

B.1 Acheivability Proofs for Unstructured Strategies

Several lemmas are needed to prove bound achievability. The following lemma demonstrates that (non-overflowed) lattice encodings become close in information content to the observations plus Gaussian noise. Lemma 6 demonstrates the joint differential entropies between source estimates from the encodings are close to the equivalent differential entropies in the Gaussian distortion system. Lemma 7 demonstrates the encodings themselves have joint Shannon entropies that are close to mutual informations in the Gaussian distortion system.

Lemma 6. Take helper observations $\vec{X}^n_{[K],\text{raw}}$ and independent dithers $\vec{W}_{Q,[K]}$ where for each k, the dither vector $\vec{W}_{Q,k}$ is distributed along the Voronoi base cell of a good-for-quantization lattice in $\mathbb{R}^n$ with mean-squared-error distortion $\sigma^2_{Q,k}$. For any $A \subset [K]$, $\varepsilon > 0$ and n large enough,

$$\left| \frac{1}{n} h(\vec{X}^n_{A,\text{raw}} - \vec{W}_{Q,A}) - h(\vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A}) \right| < \varepsilon$$

for $\tilde{W}_{Q,k} \sim \mathcal{N}(0, \sigma^2_{Q,k})$, $k = 1, \ldots, K$ independent.

Proof. Follows similarly to Theorem 3 from Zamir and Feder (1996). Notice that for any additive-independent-noise channel $x \to x + w$ then $I(x; x + w) = h(x + w) - h(x + w|x) = h(x + w) - h(w)$ so that $h(x + w) = h(w) + I(x; x + w)$. Then

$$\frac{1}{n} h(\vec{X}^n_{A,\text{raw}} - \vec{W}_{Q,A}) = \frac{1}{n} [h(\vec{W}_{Q,A}) + I(\vec{X}^n_{A,\text{raw}}; \vec{X}^n_{A,\text{raw}} - \vec{W}_{Q,A})]$$

$$= \frac{1}{n} [h(\vec{W}_{Q,A}) - nh(\tilde{W}_{Q,A}) + nh(\tilde{W}_{Q,A}) + \ldots$$

$$I(\vec{X}_{A,\text{raw}}; \vec{X}_{A,\text{raw}} - \vec{W}_{Q,A})]$$

$$= \frac{1}{n} [[\sum_{k \in A} h(\vec{W}_{Q,k}) - nh(\tilde{W}_{Q,k})] + nh(\tilde{W}_{Q,A}) + \ldots \quad \text{(B.1)}$$

$$I(\vec{X}^n_{A,\text{raw}}; \vec{X}^n_{A,\text{raw}} - \vec{W}_{Q,A})]$$

$$\geq \frac{1}{n} [[\sum_{k \in A} h(\vec{W}_{Q,k}) - nh(\tilde{W}_{Q,k})] + nh(\tilde{W}_{Q,A}) + \ldots \quad \text{(B.2)}$$

$$nI(\vec{X}_{A,\text{raw}}; \vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A})]$$

$$= \frac{1}{n} [[\sum_{k \in A} h(\vec{W}_{Q,k}) - nh(\tilde{W}_{Q,k})] + nh(\vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A})]$$

$$= \frac{1}{n} [\sum_{n \in A} D(P_{\vec{W}_{Q,k}} \| P^n_{\tilde{W}_{Q,k}})] + h(\vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A})$$

$$\overset{n \uparrow \infty}{\to} \varepsilon + h(X_{A,\text{raw}} + W_A). \quad \text{(B.3)}$$

where (B.1) follows since the terms are independent, (B.2) follows by the worst additive noise lemma (Diggavi and Cover (2001)), and the substitution and convergence

in (B.3) follow by Section III of Zamir and Feder (1996). Also:

$$\frac{1}{n}h(\vec{X}_{A,\text{raw}} - \vec{W}_{Q,A}) = \frac{1}{n}[h(\vec{W}_{Q,A}) + I(\vec{X}^n_{A,\text{raw}}; \vec{X}^n_{A,\text{raw}} - \vec{W}_{Q,A})]$$

$$\leq \frac{1}{n}h(\vec{W}_{Q,A}) + I(\vec{X}_{A,\text{raw}}; \vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A}) + \dots \quad \text{(B.4)}$$

$$\sum_{k \in A} D(P_{\vec{W}_{Q,k}} \| \prod_{k \in A} P^n_{\tilde{W}_{Q,k}})$$

$$< I(\vec{X}_{A,\text{raw}}; \vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A}) + \sum_{k \in A} \frac{1}{n}h(\vec{W}_{Q,k}) + \varepsilon/|A| \quad \text{(B.5)}$$

$$\leq I(\vec{X}_{A,\text{raw}}; \vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A}) + \sum_{k \in A} h(\tilde{W}_{Q,k}) + \varepsilon \quad \text{(B.6)}$$

$$= I(\vec{X}_{A,\text{raw}}; \vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A}) + h(\tilde{W}_{Q,A}) + \varepsilon$$

$$= h(\vec{X}_{A,\text{raw}} + \tilde{W}_{Q,A}) + \varepsilon$$

where (B.4) follows by Lemma 1 from Ihara (1978) and (B.5),(B.6) follow since the terms in $\vec{W}_{Q,A}$ are independent and by Section III of Zamir and Feder (1996). $\square$

Lemma 7. Take length-n helper observations $\vec{X}^n_{[K],\text{raw}}$, dithered lattice encodings of said observations $\vec{U}_{[K]} = \text{round}_{\mathcal{L}^n_k}(\vec{X}^n_{k,\text{raw}} + \vec{W}_{Q,k})$ (dithers $\vec{W}_{Q,[K]}$ independent and uniform over their fine lattices $\check{\mathcal{L}}^n_k$, each lattice good for quantization to noise of power $\sigma^2_{Q,k}$). Then for any $\varepsilon > 0$, n large enough has for all $S \subseteq [K]$:

$$\left| \frac{1}{n}H(\vec{U}_S|\vec{W}_{Q,S}) - I(\vec{X}_{S,\text{raw}}; \vec{Y}_S) \right| < \varepsilon. \quad \text{(B.7)}$$

In (B.7), $\vec{Y}_S := \vec{X}_{S,\text{raw}} + \tilde{W}_{Q,S}$, with $\vec{X}_{S,\text{raw}}$ distributed the same as helper observations at a single time index, and

$$\tilde{W}_Q = (\tilde{W}_1, \dots, \tilde{W}_K) \sim \mathcal{N}\left(0, \text{diag}\left(\sigma^2_{Q,1}, \dots, \sigma^2_{Q,N}\right)\right).$$

Proof. Say $S = \{s_1, \dots, s_{|S|}\}$. Now

$$\frac{1}{n}H(\vec{U}_S|\vec{W}_{Q,S}) = \sum_{t=1}^{|S|} \frac{1}{n}H(U_{s_t}|\vec{U}_{\{s_1,\dots,s_{t-1}\}}, \vec{W}_{Q,S})$$

$$= \sum_{t=1}^{|S|} \frac{1}{n}H(U_{s_t}|\vec{U}_{\{s_1,\dots,s_{t-1}\}}, \vec{W}_{Q,\{s_1,\dots,s_t\}}).$$

Apply Lemma 8 to each of the summands, identifying $\vec{W}_{Q,k} \leftrightarrow W_{Q,k}$. $\square$

With these lemmas we are prepared to prove achievability of performances in Chapter 3.

Theorem 10. Communications rate R_{QF} from Equation (3.1) and distortion D_{QF} from Equation 3.2 are achievable.

Proof. Achieve D_{QF} by applying linear estimation for X_{src} on decodings from rate-distortion source coders as described in Cover and Thomas (2012). Strategy for achieving communications rate R_{QF} is described in detail below.

If some $k \in [K]$ has $R_k = 0$, the system is equivalent to the case where the kth receiver's observation is not present, so without loss of generality assert that each $R_k > \varepsilon$. Fix some rate $R_{\text{msg}} < R_{QF}$ and a block length $N = n^2 \in \mathbb{N}$. Form a message M uniform on $[2^{nR_{\text{msg}}}]$.

Outline

Form a rate-R_{msg} codebook with length-n codewords. At receiver $k \in [K]$, form a length-n lattice encoder with coarse, fine lattices chosen such that encodings have rate r_n. Combine encodings at base and find the typical codeword. Observe that probability of error is low by recognizing that each n-length encoding from a helper approximately contains the information content of the helper's observation plus Gaussian distortion.

Transmitter setup

Generate a codebook mapping $\phi : [2^{NR_{\text{msg}}}] \to \mathbb{R}^{N \times 1}$ where:

$$\phi(m) = (x_{m,1}, \ldots, x_{m,N}) \in \mathbb{R}^{N \times 1}$$

with each component $x_{m,\ell}$ drawn iid from $\mathcal{N}(0,1)$. Reveal ϕ to the transmitter and the base.

Helper encoder setup

For each helper k, $k \in [K]$, generate a dither vector $\vec{W}_{Q,k} = (\vec{W}_{Q,k,\ell})_{\ell \in [n]} \in (\mathbb{R}^{n \times 1})^n$, where each successive n-segment is uniform in the base region $B_{f,k}$ of a lattice $\mathcal{L}_k^n \subset \mathbb{R}^n$ good for quantization to MSE distortion

$$G^n(\mathcal{L}_k^n) = \frac{\text{var } X_{k,\text{raw}}}{2^{r_k - \varepsilon} - 1}.$$

Reveal $\vec{W}_{Q,k}$ to the base and helper n.

Transmission

To send message $M \in [2^{nR_{\text{msg}}}]$, the transmitter broadcasts

$$X_{\text{src}}^N = (X_{\text{src}}^{(1)}, \ldots, X_{\text{src}}^{(n)}) = \phi(M).$$

Helper encoding and forwarding

Helper k ($k \in [K]$) observes a sequence of length N, $\vec{X}_{k,\text{raw}}^N = (\vec{X}_{k,\text{raw}}^{(\ell)})_{\ell \in [n]} \in (\mathbb{R}^{n \times 1})^n$. Now $X_{k,\text{raw}} - \vec{W}_{Q,k} = (X_{k,\text{raw}}^\ell - \vec{W}_{Q,k}^{(\ell)})_{\ell \in [n]} \in \mathbb{R}^N = (\mathbb{R}^n)^n$ is composed of n consecutive length-n segments. For the ℓ-th length-n segment ($\ell \in [n]$), form a quantization $U_n^\ell \in \mathcal{L}_k^n$ by finding the point in $\mathcal{L}_k^n$ associated with the region $\ell + B_{f,k}$ in which that segment resides. The properties of such U_n are the subject of References Zamir and Feder (1992, 1996). By the extension of Theorem 1 in Appendix A of Reference Zamir and Feder (1992),

$$H(U_k^\ell | \vec{W}_{Q,k}) = \frac{1}{n} I(X_{k,\text{raw}}; X_{k,\text{raw}} - \vec{W}_{Q,k}).$$

Further, by Theorem 3 in Reference Zamir and Feder (1996)

$$\frac{1}{n} I\left(X_{k,\ell,\mathrm{raw}}^n; X_{k,\ell,\mathrm{raw}}^n - \vec{W}_{Q,k}\right) \xrightarrow[n\uparrow\infty]{} I(X_{k,\mathrm{raw}};\ X_{k,\mathrm{raw}} + \tilde{W}_{Q,k}) = r_k - \varepsilon$$

where

$$\tilde{W}_{Q,k} \sim \mathcal{N}(0, G^n(\mathcal{L}_k^n, B_{f,k})),\ \ k \in [K]. \tag{B.8}$$

Thus the n encoded messages produced by helper k, $(U_k^\ell)_{\ell \in [n]}$, are within the LAN rate constraint for large enough blocklength N.

Helper $k \in [K]$ forwards $(U_k^\ell)_{\ell \in [n]}$ to the base.

Decoding

At the base receive all of the lattice quantizations,

$$\vec{U} := ((U_k^\ell)_{\ell \in [n]})_{k \in [K]}.$$

Take $A_\varepsilon^N(X_{\mathrm{src}}, \vec{U} | \vec{W}_{Q,[K]})$ to be the collection of

$$(\boldsymbol{x}_m, \boldsymbol{u}) \in \mathrm{Range}(\phi) \times \left(\prod_{k \in [K]} \mathcal{L}_k^n\right)^n$$

which are jointly ε-weakly-typical with respect to the joint distribution of $(X_{\mathrm{src}}^N, \vec{U})$, conditional on all the dithers $(\vec{W}_{Q,k})_{k \in [K]}$. Weak- and joint-typicality are defined in Cover and Thomas (2012).

At the base, find a message estimate $\hat{m} \in [2^{NR_{\mathrm{msg}}}]$ where $(\phi(\hat{m}), \vec{U}) \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U} | \vec{W}_{Q,[K]})$. Declare error events $\mathcal{E}_0$ if M is not found to be typical with $\vec{U}$, and $\mathcal{E}_1$ if there is some $\hat{m} \neq M$ where $(\phi(\hat{m}), \vec{U}) \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U} | \vec{W}_{Q,[K]})$.

Error analysis

By typicality and the law of large numbers, $P(\mathcal{E}_0) \to 0$ as $n \to \infty$. Also,

$$P(\mathcal{E}_1) \leq \sum_{\hat{m} \neq M} P\left(\{\phi(\hat{m}) : (\phi(\hat{m}), \vec{U}) \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U} | \vec{W}_{Q,[K]})\}\right)$$

$$\leq \sum_{\hat{m} \neq M} 2^{-t \cdot (I(X_{\mathrm{src}}^{(1)}; (\vec{U}_k^{(1)})_{k \in [K]} | \vec{W}_{Q,[K]}^{(1)}) - 3\varepsilon)}$$

$$< 2^{-n \cdot (I(X_{\mathrm{src}}^{(1)}; (\vec{U}_k^1)_{k \in [K]} | \vec{W}_{Q,[K]}^{(1)}) - 3\varepsilon) + NR_{\mathrm{msg}}} \tag{B.9}$$

$$= 2^{-n \cdot (h(X_{\mathrm{src}}^{(1)}) - h(X_{\mathrm{src}}^{(1)} | (\vec{U}_k^1)_{k \in [K]}, \vec{W}_{Q,[K]}^{(1)}) - 3\varepsilon - nR_{\mathrm{msg}})}$$

$$\leq 2^{-n \cdot (h(X_{\mathrm{src}}^{(1)}) - h(X_{\mathrm{src}}^{(1)} | X_{[K],\mathrm{raw}}^{(1)} - \vec{W}_{Q,[K]}^{(1)}) - 3\varepsilon - nR_{\mathrm{msg}})} \tag{B.10}$$

$$\leq 2^{-n \cdot (n \cdot I(X_{\mathrm{src}}; \vec{X}_{[K],\mathrm{raw}} + \tilde{W}_{Q,[K]}) - 4\varepsilon - nR_{\mathrm{msg}})} \tag{B.11}$$

$\tilde{W}_{Q,[K]}$ is defined as in Equation (B.8). Equation (B.10) follows by the data processing inequality. Equation (B.11) follows by combining the entropies into a mutual information, then using the worst additive noise lemma Diggavi and Cover (2001). So if R_{msg} is chosen less than $I(X_{\mathrm{src}}; \vec{X}_{[K],\mathrm{raw}} + \tilde{W}_{Q,[K]}) - 4\varepsilon$ then $P(\mathcal{E}_0 \cup \mathcal{E}_1) \to 0$ when $T \to \infty$. $\square$

Theorem. Communications rate $R_{\tilde{B}T}$ from Equation (3.3) is achievable, and distortion D_{BT} from Equation (3.2) is achievable.

Proof. For the distortion bound, fix variables as in the statement of Theorem 5. Apply Slepian-Wolf joint-compression on the rate-distortion encodings described in Theorem 10, then apply linear estimation on the base's decodings. The achievability strategy for the communications rate is detailed below.

Fix the following variables such that the statement in Theorem 6 is satisfied. $\lambda > 0$. Fix $\sigma_1^2, \ldots, \sigma_K^2 > 0$,

$$\rho_1 := \frac{1}{2}\log(\operatorname{var} X_{1,\mathrm{raw}}/\sigma_1^2), \ldots, R_K := \frac{1}{2}\log(\operatorname{var} X_{K,\mathrm{raw}}/\sigma_K^2),$$

and messaging rate R_{msg}. Take block length $N = n^2 \in \mathbb{N}$.

If some $k \in [K]$ has $R_k = 0$ or $\rho_k = 0$, the system is equivalent to the case where the kth helper is not present, so without loss of generality assert that each $R_n, \rho_n > \varepsilon$.

Form a message M uniform on $[2^{NR_{\mathrm{msg}}}]$.

Outline

Form a rate-R_{msg} codebook with length-n codewords. At receiver $k \in [K]$, form a length-n regular lattice encoder with lattice chosen coarse enough that its encodings have rate ρ_n. Randomly bin the encodings down to rate R_n. At the base find the codeword jointly typical with the binned encodings. Observe that probability of error is low by recognizing that the un-binned n-length encodings from helpers approximately contain the joint-information content of the helper's observations plus Gaussian distortion.

Transmitter setup

Generate a codebook mapping $\phi : [2^{nR_{\mathrm{msg}}}] \to \mathbb{R}^{n \times 1}$ where

$$\phi(m) = (x_{m,1}, \ldots, x_{m,n}) \in \mathbb{R}^{n \times 1}$$

with each component $x_{m,\ell}$ drawn i.i.d. from $\mathcal{N}(0,1)$. Reveal ϕ to the transmitter and the base.

Helper encoder setup

For each helper k, $k \in [K]$, generate a dither vector $\vec{W}_{Q,k} = (\vec{W}_{Q,k}^{(\ell)})_{\ell \in [k]} \in (\mathbb{R}^{n \times 1})^n$, where each successive n-segment is uniform in the base region $B_{f,k}$ of a lattice $\mathcal{L}_k^n \subset \mathbb{R}^n$ good for quantization to MSE distortion σ_k^2. Reveal $\vec{W}_{Q,k}$ to the base and helper n.

For each receiver $k \in [K]$, form a random mapping $\mathsf{Index}_k : \mathcal{L}_k^n \to [2^{nR_k}]$ which takes on iid values at each $u_n \in \mathcal{L}_k^n$:

$$\mathsf{Index}_k(u_k) \sim \mathsf{Uniform}\left([2^{nR_k}]\right)$$

Index_k is the binning scheme used by receiver k. Distribute each map Index_k to helper k and the base.

Transmission

To send a message $M \in [2^{NR_{\mathrm{msg}}}]$, have the transmitter broadcast

$$\boldsymbol{X} = (X_{\mathrm{src}}^{(1)}, \ldots, X_{\mathrm{src}}^{(n)}) = \phi(M).$$

Helper encoding

Helper k ($k \in [K]$) observes a sequence of length N,

$$X_{k,\text{raw}}^N = (X_{k,\text{raw}}^\ell)_{\ell \in [n]} \in (\mathbb{R}^{n \times 1})^n$$

Now $X_{k,\text{raw}}^N - \vec{W}_{Q,k} = (X_{k,\text{raw}}^\ell - \vec{W}_{Q,k}^{(\ell)})_{\ell \in [n]} \in \mathbb{R}^N = (\mathbb{R}^n)^n$ is composed of n consecutive length-n segments. For the ℓ-th length-n segment, $\ell \in [n]$, form a quantization $U_k^\ell \in \mathcal{L}_k^n$ by finding the point in $\mathcal{L}_k^n$ associated with the region $B_{f,k}$ in which that segment resides.

Applying Lemma 7, then for large enough n, any $S \subseteq [K]$ has

$$\frac{1}{n} H(U_S^\ell | U_{S^C}^\ell, (\vec{W}_{Q,k})_{k \in [K]}) \underset{n \uparrow \infty}{\to} I(X_{S,\text{raw}} + \tilde{W}_{Q,S}; X_{S,\text{raw}} | \tilde{W}_{Q,S^C}) < \sum_{k \in S} R_k$$

where the right-hand inequality comes from choice of $(\rho_1, \ldots, \rho_K)$.

Helper joint-compression and forwarding

At receiver $k \in [K]$, form binned encodings $(V_k^\ell)_{\ell \in [n]}$ where for each ℓ,

$$V_k^\ell := \mathsf{Index}_k(U_k^\ell).$$

Note that $|\mathsf{Range}(\mathsf{Index}_k)| \leq 2^{n \cdot R_k}$ so $H(V_k^\ell) \leq n \cdot R_k$ and $H((V_k^\ell)_{\ell \in [n]}) \leq n \cdot R_k$ so that the LAN messaging constraint is satisfied.

Each receiver forwards $(V_k^\ell)_{\ell \in [n]}$ to the base.

Decoding

At the base receive all all the binned helper encodings $\boldsymbol{V} = ((V_k^\ell)_{\ell \in [n]})_{k \in [K]}$.

Take $A_\varepsilon^N(X_{\text{src}}, \vec{U} | \vec{W}_{Q,[K]})$ to be the set of

$$\left(\boldsymbol{x}_m, (\boldsymbol{u}^1, \ldots, \boldsymbol{u}^K)\right) \in \mathsf{Range}(\phi) \times \left(\prod_{k \in [K]} \mathcal{L}_k^n\right)^n$$

which are jointly ε-weakly-typical with respect to the joint distribution of $(\boldsymbol{X}, \vec{U})$, conditional on all the dithers $(\vec{W}_{Q,k})_{k \in [K]}$. Weak- and joint-typicality are defined in Reference Cover and Thomas (2012).

For an ensemble $\boldsymbol{v}$ of bin indices from all the receivers,

$$\boldsymbol{v} := ((v_k^\ell)_{\ell \in [n]})_{k \in [K]} \in \prod_{k \in [K]} [2^{n \cdot R_n}]^n,$$

define:

$$B_{\boldsymbol{v}} \triangleq \left\{ \boldsymbol{u} \in \left(\prod_{k \in [K]} \mathcal{L}_k^n\right)^n \middle| \mathsf{Index}_k(u_k^\ell) = v_k^\ell, k \in [K], \ell \in [n] \right\}.$$

Each $B_{\boldsymbol{v}}$ represents the set of helper lattice quantizations represented by the ensemble of helper bin indices $\boldsymbol{v}$. At the base, find $\hat{m}$ for which there is some $\hat{\boldsymbol{u}} \in B_{\boldsymbol{V}}$ where $(\phi(\hat{m}), \hat{\boldsymbol{u}}) \in A_\varepsilon^N(X_{\text{src}}, \vec{U} | \vec{W}_{Q,[K]})$. Declare $\hat{m}$ as the decoded message.

Error analysis

We have the following error events:

- $\mathcal{E}_{JT}$: $\boldsymbol{X}$ is not typical with anything in $B_{\boldsymbol{V},[K]}$

- $\mathcal{E}_{\hat{m},S}$: For $S \subseteq [K]$, there is some $\hat{m} \neq M$ and $\hat{\boldsymbol{u}} = (\hat{\boldsymbol{u}}_S, \hat{\boldsymbol{u}}_{S^C}) \in B_{\boldsymbol{V}}$ where $(\phi(\hat{m}), \hat{\boldsymbol{u}}) \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U}|\vec{W}_{Q,[K]})$ and $\hat{\boldsymbol{u}}_{S^C} = \vec{U}_{S^C}$, but any $k \in S$ has $\hat{\boldsymbol{u}}_k \neq \vec{U}_k$.[1]

By the law of large numbers and properties of typical sets, as n becomes large, eventually $P(\mathcal{E}_{JT}) < \varepsilon$. Now compute:

$$\sum_{\hat{m} \neq M} P(\mathcal{E}_{\hat{m},S}) \leq 2^{NR_{\mathrm{msg}}} \cdot \sum_{\substack{\tilde{\boldsymbol{u}} \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U}|\vec{W}_{Q,[K]}): \\ \tilde{\boldsymbol{u}}_{S^C} = \vec{U}_{S^C}}} \mathbf{1}_{B_{\boldsymbol{V}}}(\tilde{\boldsymbol{u}}) \cdot P\left((\tilde{X}_{\mathrm{src}}, \tilde{\boldsymbol{u}}) \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U}|\vec{W}_{Q,[K]})\right)$$

where $\tilde{X}_{\mathrm{src}}$ is a random variable independent of and distributed identically as X_{src}. Then with high probability for large enough t,

$$\sum_{\hat{m} \neq M} P(\mathcal{E}_{\hat{m},S}) \leq 2^{NR_{\mathrm{msg}}} \cdot \sum_{\substack{\tilde{\boldsymbol{u}} \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U}|\vec{W}_{Q,[K]}): \\ \tilde{\boldsymbol{u}}_{S^C} = \vec{U}_{S^C}}} \mathbf{1}_{B_{\boldsymbol{V}}}(\tilde{\boldsymbol{u}}) \cdot 2^{-n \cdot (I(X_{\mathrm{src}}; \vec{U}_{[K]}|\vec{W}_{Q,[K]}) - 3\varepsilon)}$$

$$\leq 2^{NR_{\mathrm{msg}}} \cdot \left|\{\tilde{\boldsymbol{u}} \in A_\varepsilon^N(X_{\mathrm{src}}, \vec{U}|\vec{W}_{Q,[K]})|\tilde{\boldsymbol{u}}_{S^C} = \vec{U}_{S^C}\}\right| \cdot \ldots \qquad \text{(B.12)}$$

$$(1 + \varepsilon) \cdot P(\vec{U} \in B_{\boldsymbol{V}}) \cdot 2^{-n \cdot (I(X_{\mathrm{src}}; \vec{U}_{[K]}|\vec{W}_{Q,[K]}) - 3\varepsilon)}$$

$$\leq 2^{NR_{\mathrm{msg}}} \cdot 2^{n \cdot (H(\vec{U}_S|\vec{U}_{S^C}, \vec{W}_{Q,[K]}) + \varepsilon)} \cdot \ldots$$

$$2^{-n \cdot (n \cdot \sum_{k \in S} R_k - \varepsilon)} \cdot 2^{-n \cdot (I(X_{\mathrm{src}}; \vec{U}_{[K]}|\vec{W}_{Q,[K]}) - 3\varepsilon)}$$

where (B.12) follows by construction of $B_{\boldsymbol{V}}$. Taking the log,

$$\log\left(\sum_{\hat{m} \neq M} P(\mathcal{E}_{\hat{m},S})\right) \leq NR_{\mathrm{msg}} + n \cdot (H(\vec{U}_S|\vec{U}_{S^C}, \vec{W}_{Q,[K]}) + \varepsilon) - \ldots$$

$$n \cdot (n \cdot \sum_{k \in S} R_k - \varepsilon) - n \cdot \left(I(X_{\mathrm{src}}; \vec{U}_{[K]}|\vec{W}_{Q,[K]}) - 3\varepsilon\right) + \varepsilon$$

$$\leq n \cdot (H(\vec{U}_S|\vec{U}_{S^C}, \vec{W}_{Q,[K]}) - n \cdot \sum_{k \in S} R_k + nR_{\mathrm{msg}} - \ldots$$

$$I(X_{\mathrm{src}}; \vec{U}_{[K]}|\vec{W}_{Q,[K]}) + 6\varepsilon)$$

$$\leq n^2 \cdot (I(X_{S,\mathrm{raw}}; X_{S,\mathrm{raw}} + \tilde{W}_{Q,S}|X_{S^C,\mathrm{raw}} + \tilde{W}_{Q,S^C}) - \ldots \qquad \text{(B.13)}$$

$$\sum_{k \in S} R_k + R_{\mathrm{msg}} - I(X_{\mathrm{src}}; X_{[K],\mathrm{raw}} + \tilde{W}_{Q,[K]}) + 8\varepsilon/n).$$

Equation (B.13) follows by using Lemma 7 on $H(\vec{U}_S|\vec{U}_{S^C})/n$ and using reasoning identical to (B.9) through (B.11) on $I(X_{\mathrm{src}}; \vec{U}|\vec{W}_{Q,[K]})$. Recall that R_{msg} was chosen

---

[1] $\mathcal{E}_{\hat{m},S}$ denotes the situation where the base identifies the wrong quantizations for all the receivers in S.

so that $R_{\text{msg}} < I(X_{\text{src}}; X_{[K],\text{raw}} + \tilde{W}_{Q,[K]}) - \lambda$. Then (B.13) will approach $-\infty$ as $n \to \infty$ (thereby $\mathbb{P}(\cup_{\hat{m},S}\mathcal{E}_{\hat{m},S})$ approaches 0) if all $S \subseteq [K]$ satisfy:

$$I(X_{S,\text{raw}}; X_{S,\text{raw}} + \tilde{W}_{Q,S} | X_{S^C,\text{raw}} + \tilde{W}_{Q,S^C}) < \lambda + \sum_{k \in S} R_k - 8\varepsilon. \tag{B.14}$$

By assumption (B.14) holds for all S for ε small enough. Since all error events approach 0, then a rate of $R_{\tilde{B}T}$ is achievable. $\qquad\square$

B.2   Numerical Optimization Details

Evaluation of the Gaussian distortion and distributed compression bounds in each environment requires optimization of a quasi-convex objective (Chapman et al. (2018)) over a non-convex domain. To overcome this, an iterative interior point method was used to find the maximum: each constraint $f(x) < 0$ is replaced with a stricter constraint, $f(x) + \beta < 0$ for some $\beta > 0$ and a minimization of the objective is performed under the new constraint. The optimal point is passed as the initial guess for the next iteration, where the objective function is minimized with updated constraints $f(x) + \beta' < 0$ with $0 < \beta' < \beta$. The method is iterated until the constraint is practically equivalent to $f(x) < 0$ or the optimal point converged. Each individual minimization was performed using sequential least-squares programming (SLSQP) through SciPy Jones et al. (01).

APPENDIX C

SUPPLEMENTS FOR STRUCTURED CODE

C.1 Proof of Lemma 1 Construction of Mod-Elementary Decoders

Proof. Assume $\mathcal{M}(\mathbf{A}_j)$ occurs. Then $\mathcal{M}(\mathbf{A}_1) \supset \cdots \supset \mathcal{M}(\mathbf{A}_{j-1})$ have occurred also. Compute:

$$f_1\left(\vec{U}\right) = \mathrm{mod}\left(\mathbf{a}_1^\dagger(\mathrm{mod}(\alpha_k X_k^n) - \tilde{W}_{d,k})_k\right) \tag{C.1}$$

$$= \mathrm{mod}\left(\mathbf{a}_1^\dagger(\mathrm{mod}(\alpha_k X_k^n) - \mathrm{mod}(\tilde{W}_{d,k}))_k\right) \tag{C.2}$$

$$= \mathrm{mod}\left(\mathbf{a}_1^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d)\right) \tag{C.3}$$

$$= \mathbf{a}_1^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d). \tag{C.4}$$

where (C.1) is due to (Zamir, 2014, Theorem 4.1.1) (the crypto lemma), (C.2) is since $W'_{d,k} \in B_c$, (C.3) is by Property 4 and (C.4) since $\mathcal{M}(\mathbf{A}_1)$ occurred.

Now for any $m > 1$ say $f_{[m-1]}$ has:

$$f_{[m-1]}(\vec{U}) = \mathbf{A}_{[m-1]}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_{d,k}).$$

Then compute (similarly to the base case):

$$
\begin{aligned}
f_m\left(\vec{U}\right) = \mathrm{mod}\Big\{&\mathrm{mod}\left(\mathbf{a}_m^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d)\right) - \ldots \\
&\mathrm{mod}\left([\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger \mathbf{A}_{[m-1]}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d)\right)\Big\} + \ldots \\
&[\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger \mathbf{A}_{[m-1]}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d) \\
= \mathrm{mod}\Big\{&\mathbf{a}_m^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d) - \ldots \\
&[\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger \mathbf{A}_{[m-1]}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d)\Big\} + \ldots \\
&[\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger \mathbf{A}_{[m-1]}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d) \\
= \mathrm{mod}\Big\{&[\mathbf{R}_{m-1}\mathbf{a}_m]^\dagger(\overrightarrow{\alpha X}^n + \tilde{W}_d)\Big\} + \ldots \\
&[\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger \mathbf{A}_{[m-1]}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d).
\end{aligned}
\tag{C.5}
$$

Since $\mathcal{M}(\mathbf{A}_m)$ has ocurred, then $[\mathbf{R}_{m-1}\mathbf{a}_m]^\dagger(\overrightarrow{\alpha X}^n + \tilde{W}_d) \in B_c$, hence by Property 1, (C.5) reduces:

$$
\begin{aligned}
f_m\left(\vec{U}\right) &= [\mathbf{R}_{m-1}\mathbf{a}_m]^\dagger(\overrightarrow{\alpha X}^n + \tilde{W}_d) + \ldots \\
&\quad [\mathbf{S}_{m-1}\mathbf{a}_m]^\dagger \mathbf{A}_{[m-1]}^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d) \\
&= \mathbf{A}_m^\dagger(\overrightarrow{\alpha X}^n - \tilde{W}_d).
\end{aligned}
$$

So by induction the described mod-elementary functions behave as desired.   □

C.2 Proof of Theorem 7, Asymptotic Lattice Scheme Performance

Proof. Italicized words here are technical terms taken from the specified source. Using the construction in (Ordentlich and Erez, 2016, Theorem 2), take sequences of lattices $L_c^{(n)}, (L_k^{(n)})_{k \in [K]}$ with base regions $B_c^{(n)}, (B_k^{(n)})_k$ where:

- $L_c^{(n)}$ with $B_c^{(n)}$ is *good for channel coding in the presence of semi norm-ergodic* noise, (Ordentlich and Erez, 2016, Definition 4) scaled to handle *semi norm-ergodic* (Ordentlich and Erez, 2016, Definition 2) noise of power less than 1.

- Each $L_k^{(n)} \subset L_c^{(n)}$ is *good for mean squared error quantization* (Ordentlich and Erez, 2016, Definition 5), with $\lim_n \frac{1}{n} \mathbb{E}[\|u_k^{(n)}\|^2] < 2^{-2r_k + \varepsilon}$, taking each $u_k^{(n)} \sim$ unif $B_k^{(n)}$ independent.

- Eventually in n, $\frac{1}{n} \log(|L_c^{(n)} \cap L_k^{(n)}|) < r_k$.

For each $k \in [K]$ take $Z_k^n := \alpha_k X_k^n - \tilde{W}_{d,k}$, where $(\tilde{W}_{d,k})_{k \in [K]}$ is distributed like $(u_k^{(n)})_{k \in [K]}$ and is independent of $(X_k^n)_{k \in [K]}$. Each event $\mathcal{M}(\mathbf{A}_j)$ has associated with it some linear combination $\mathbf{R}_{j-1} = I$ if $j = 1$ or as in Definition 8 otherwise. Then by (Krithivasan and Pradhan, 2007, Appendix V) the linear combination $(\mathbf{R}_{j-1}\mathbf{a}_j)^\dagger (Z_k^n)_{k \in [K]}$ probably lands in the base of a lattice cell as long as that lattice is good for coding for powers greater than

$$p = \frac{1}{n} \mathbb{E}[\|(\mathbf{R}_{j-1}\mathbf{a}_j)^\dagger (Z_k^n)_{k \in [K]}\|^2] = (\mathbf{R}_{j-1}\mathbf{a}_j)^\dagger \mathbf{C}_\infty (\mathbf{R}_{j-1}\mathbf{a}_j).$$

Since $L_c^{(n)}$ is *good for channel coding in the presence of semi norm-ergodic noise* with power less than 1 Ordentlich and Erez (2016) then (by definition) Z_k^n occurs in $\mathcal{M}(\mathbf{A}_j)$ with arbitrarily high probability eventually in n if $p < 1$, and with arbitrarily low probability if $p > 1$. The case where $p = 1$ never occurs when using $\varepsilon > 0$ that affects $\mathbf{C}_\infty$ so that every $(\mathbf{R}_{j-1}\mathbf{a}_j)^\dagger \mathbf{C}_\infty (\mathbf{R}_{j-1}\mathbf{a}_j)$ is irrational (only countable $\mathbf{R}_{j-1}, \mathbf{a}_j$ are possible, so ε small enough always exist).

We now demonstrate that in this situation, among the high-probability events from Definition 8, there are some whose conditions cannot be strengthened. Any event $\mathcal{M}(\mathbf{A}_j)$ with probability eventually high and associated projection $\mathbf{R}_{j-1}$ either has some vector $\mathbf{a} \in \mathbb{Z}^K$ with $\mathbf{a}$ not in $\mathbf{R}_{j-1}$'s null space and $\frac{1}{n} \mathbb{E}[\|(\mathbf{R}_{j-1}\mathbf{a})^\dagger (Z_k^n)_{k \in [K]}\|^2] < 1$, or no such vector $\mathbf{a}$ exists. If there is such an $\mathbf{a}$ then taking $\mathbf{a}_{j+1} = \mathbf{a}$, the event $\mathcal{M}([\mathbf{A}_j, \ \mathbf{a}_{j+1}])$ will also have eventually high probability. Repeating the argument, such $\mathbf{a}$'s can only be found up to $m < K$ times: by then the matrix $\mathbf{A}_j$ has column basis for all $\mathbb{R}^K$, or all choice of $\mathbf{a} \in \mathbb{Z}^K$ yields $\frac{1}{n} \mathbb{E}[\|(\mathbf{R}_{j-1}\mathbf{a})^\dagger (Z_k^n)_{k \in [K]}\|^2] > 1$.

The result of this maximal strengthening of $\mathcal{M}(\mathbf{A}_j)$'s conditions to $\mathcal{M}(\mathbf{A}_{j+m})$ has associated with it a projection $\mathbf{R}_-$ with the property that any integer vector $\mathbf{a} \in \mathbb{Z}^K$ has either $\mathbf{a}$ in $\mathbf{R}_-$'s null space or $\frac{1}{n} \mathbb{E}[\|(\mathbf{R}_-\mathbf{a})^\dagger (Z_k^n)_{k \in [K]}\|^2] = (\mathbf{R}_-\mathbf{a})^\dagger \mathbf{C}_\infty (\mathbf{R}_-\mathbf{a}) > 1$. Any projections created in such a way from two different events must be equal, since if the first strengthened event's projection's null space had a vector the second's did not, then the second strengthened event could be further strengthened using the vectors in the first event's sequence.

By above, if $\mathbf{R}_-$ has nonzero range, and if a nonzero subspace U does not contain (image $\mathbf{R}_-$) then some nonzero integer vector $\mathbf{a} \in \mathbb{Z}^K$ away from U has $\min_{\mathbf{s} \in U}(\mathbf{a} - \mathbf{s})^\dagger \mathbf{C}_\infty(\mathbf{a} - \mathbf{s}) < 1$. So $S = \text{image}(\mathbf{R}_-)$ is the smallest of all subspaces with the property that any $\mathbf{a} \in \mathbb{Z}^K$ has either $\min_{\mathbf{s} \in S}(\mathbf{a} - \mathbf{s})^\dagger \mathbf{C}_\infty(\mathbf{a} - \mathbf{s}) \geq 1$ or $\mathbf{a} \in S$ otherwise.

Finally apply Lemma 1 and Corollary 2 to any high-probability event associated with projection $\mathbf{R}_-$. Since the event eventually has arbitrarily high probability, choosing $\Delta > 1$ eventually gives expected distortion arbitrarily close to d_L^2. $\qquad\square$

C.3   Closed Form for $\mathbf{S}_k(\mathbf{v})$

For exposition the definition of $\mathbf{S}_k(\mathbf{v})$ is shown semantically, as a minimizer, as opposed to in closed form. Its closed form is readily derived:

$$
\underset{\mathbf{u} \in \mathbb{R}^k}{\arg\min} \ (\mathbf{v} - \mathbf{A}^{(k)}\mathbf{u})^\dagger \mathbf{C}(\mathbf{v} - \mathbf{A}^{(k)}\mathbf{u})
$$

$$
= \underset{\mathbf{u} \in \mathbb{R}^k}{\arg\min} \ -2\left\{\mathbf{v}^\dagger \mathbf{C}\mathbf{A}^{(k)}\mathbf{u}\right\} + \mathbf{u}^\dagger \mathbf{A}^{(k)\dagger}\mathbf{C}\mathbf{A}^{(k)}\mathbf{u}
$$

$$
= \left[\text{pinv}(\mathbf{A}^{(k)\dagger}\mathbf{C}\mathbf{A}^{(k)})\mathbf{A}^{(k)\dagger}\mathbf{C}\right]\mathbf{v}.
$$

C.4   KP Parametrization

Here the parametrization the KP scheme Krithivasan and Pradhan (2007) provides in terms of the asymptotic scheme and Lemma 1 is shown. First fix some $\mathbf{c} \in \mathbb{R}^K$ and an ordered partition $(\Theta_1, \ldots, \Theta_p)$ of $[K]$. With these variable choices, the KP scheme specifies use of $\mathbf{A}$ with columns $\mathbf{a}_m = \mathbf{1}_{\Theta_m}$. A matrix with this structure is denoted with subscript as $\mathbf{A}_{KP}$.

It also specifies receiver scaling coefficients $\vec{\alpha}$ defined in parts over the partition sets. Starting at $m = 1$ and up through $m = p$, then the components $(\alpha_k)_{k \in \Theta_m}$ are specified by a minimization problem dependent on the result of previous steps. Take $\vec{W}_{Q\infty} \sim \mathcal{N}(0, \mathbf{\Sigma}_{Q\infty})$. Also take $\mathbf{c}_{\Theta_k}$ to be the $|\Theta_k|$-vector with coefficients taken from $\mathbf{c}$ at indices in Θ_k. Now starting at $m = 1$ and up through $m = p$ define, if possible:

$$
(\alpha_k)_{\Theta_m} := \sqrt{\frac{1 - \text{var}\left(\mathbf{a}_k^\dagger \vec{W}_{Q\infty}\right)}{\text{var}\left(\mathbf{c}_{\Theta_m}\vec{X} \,\middle|\, \mathbf{A}_{KP,m-1}(\overrightarrow{\alpha X} + \vec{W}_{Q\infty})\right)}}\, \mathbf{c}_{\Theta_m}. \tag{C.6}
$$

If the square-root in (C.6) does not exist (i.e. if the numerator is not positive), then a KP scheme with the partition $(\Theta_1, \ldots, \Theta_p)$ cannot be designed for specified $\mathbf{c}$ and encoder rates $(R_1, \ldots, R_K)$. For any $\mathbf{c}$ and nonzero rates, there is always some partition that works. For example, a singleton partition works. A well-formed KP scheme guarantees achievable average distortion:

$$
\hat{d}_{KP}^2 = \text{var}\left(X_{\text{src}} \,\middle|\, \mathbf{A}_{KP}^\dagger(\overrightarrow{\alpha X} + \vec{W}_{Q\infty})\right).
$$

APPENDIX D

SUPPLEMENTS FOR STRUCTURED CODE JOINT COMPRESSION

D.1 Subroutines

Algorithm 2 Compute recoverable linear combinations $\boldsymbol{A} \in \mathbb{R}^{K \times m}$ from modulos of lattice encodings with covariance $\boldsymbol{\Sigma}_Q \in \mathbb{R}^{K \times K}$.

---

function $\text{STAGES}^*(\boldsymbol{\Sigma})$
    $\boldsymbol{A} \leftarrow [\,], \ \vec{a} \leftarrow \text{SLVC}\left(\boldsymbol{\Sigma}_Q^{1/2}\right), \ \boldsymbol{R} \leftarrow \boldsymbol{I}_K,$
    while $0 < (\boldsymbol{R}\vec{a})^\dagger \boldsymbol{\Sigma}_Q (\boldsymbol{R}\vec{a}) < 1$ do
        $\boldsymbol{A} \leftarrow [\boldsymbol{A}, \vec{a}]$
        $\boldsymbol{R} \leftarrow \boldsymbol{I}_K - \boldsymbol{A}\,\text{pinv}(\boldsymbol{A}^\dagger \boldsymbol{\Sigma}_Q \boldsymbol{A})\boldsymbol{A}^\dagger \boldsymbol{\Sigma}_Q$
        $\vec{a} \leftarrow \text{SLVC}\left(\boldsymbol{\Sigma}_Q^{1/2} \boldsymbol{R}\right)$
    end while
    $\boldsymbol{A} \leftarrow [\boldsymbol{A}, \text{LATTICEKERNEL}\left(\boldsymbol{\Sigma}_Q^{1/2} \boldsymbol{R}, \boldsymbol{A}\right)]$
    return $\boldsymbol{A}$
end function

Here, we provide a list of subroutines involved in a statement of the results:

- $\text{STAGES}^*(\cdot)$ is a slight modification of an algorithm from Chapman et al. (2019), reproduced here in Algorithm 2. The original algorithm characterizes the integral combinations $\boldsymbol{A}^\dagger \vec{Y}$ which are recoverable with high probability from lattice messages $\vec{U}$ and dithers $\vec{W}$, excluding those with zero power. The exclusion is due to the algorithm's use of $\text{SLVC}(\cdot)$ as just defined. Such linear combinations never arose in the context of Chapman et al. (2019), although it provides justification for them being recoverable; in the paper, the algorithm's argument is always full-rank. This is not true in the present context. The version here includes these zero-power subspaces by including a call to $\text{LATTICEKERNEL}(\cdot)$ before returning.

- $\text{SLVC}(\boldsymbol{B})$, 'Shortest Lattice Vector Coordinates' returns the nonzero integer vector $\vec{a}$ which minimizes the norm of $\boldsymbol{B}\vec{a}$ while $\boldsymbol{B}\vec{a} \neq 0$, or the zero vector if no such vector exists. $\text{SLVC}(\cdot)$ can be implemented using a lattice enumeration algorithm like one in Schnorr and Euchner (1994) together with the LLL algorithm to convert a set of spanning lattice vectors into a basis Buchmann and Pohst (1987).

- $\text{LATTICEKERNEL}(\boldsymbol{B}, \boldsymbol{A})$, for $\boldsymbol{B} \in \mathbb{R}^{K \times d}$, $\boldsymbol{A} \in \mathbb{Z}^{d \times a}$ returns the integer matrix $\boldsymbol{A}_\perp \in \mathbb{Z}^{d \times b}$ whose columns span the collection of all $\vec{a} \in \mathbb{Z}^K$ where $\boldsymbol{B}\vec{a} = 0$ while $\boldsymbol{A}^\dagger \vec{a} = 0_a$. In other words, it returns a basis for the integer lattice in $\ker \boldsymbol{B}$ whose components are orthogonal to the lattice $\boldsymbol{A}$. This can be implemented using an algorithm for finding 'simultaneous integer relations' as described in Hastad et al. (1989).

- $\text{ICQM}(\boldsymbol{M}, \vec{v}, c)$ is an "Integer Convex Quadratic Minimizer." It provides a solution for the NP-hard problem: "Minimize $(\vec{x}^\dagger \boldsymbol{M} \vec{x} + 2\vec{v}^\dagger \vec{x} + c)$ over $\vec{x}$ with integer components." Although finding the optimal solution is exponentially

difficult in input size, algorithms are tractable for low dimension. (Ghasemmehdi and Agrell, 2011, Algorithm 5, Figure 2)

- CVARCOMPONENTS$(\mathbf{\Sigma}_Q, \boldsymbol{A})$ returns certain variables $\{\boldsymbol{M}, \vec{v}, c\}$ involved in computing

$$\mathrm{var}\left(Y_K - \boldsymbol{\nu}^\dagger \vec{Y}_{[K-1]} \,\Big|\, \boldsymbol{A}\vec{Y}_{[K-1]}\right)$$

when $\vec{Y} = (Y_1, \ldots, Y_K)$ has covariance $\mathbf{\Sigma}_Q$. Write some matrices in block form:

$$\mathbf{\Sigma}_Q = \begin{bmatrix} \boldsymbol{M}_1 & \vec{v}_1 \\ \vec{v}_1^\dagger & \varsigma_1^2 \end{bmatrix},$$

$$\mathbf{\Sigma}_Q \begin{bmatrix} \boldsymbol{A} \\ 0 \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{A} \\ 0 \end{bmatrix}^\dagger \mathbf{\Sigma}_Q \begin{bmatrix} \boldsymbol{A} \\ 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{A} \\ 0 \end{bmatrix}^\dagger \mathbf{\Sigma}_Q = \begin{bmatrix} \boldsymbol{M}_2 & \vec{v}_2 \\ \vec{v}_2^\dagger & \varsigma_2^2 \end{bmatrix}.$$

Then, taking $\boldsymbol{M} = (\boldsymbol{M}_1 - \boldsymbol{M}_2)$, $\boldsymbol{v} = -(\vec{v}_1 - \vec{v}_2)$, $c = (\varsigma_1^2 - \varsigma_2^2)$, one can check:

$$\mathrm{var}\left(Y_K - \boldsymbol{\nu}^\dagger \vec{Y}_{[K-1]} \,\Big|\, \boldsymbol{A}\vec{Y}_{[K-1]}\right) = \boldsymbol{\nu}^\dagger \boldsymbol{M} \boldsymbol{\nu} + 2\vec{v}^\dagger \boldsymbol{\nu} + c.$$

- CVAR$(\boldsymbol{M}_1 | \boldsymbol{M}_2; \mathbf{\Sigma})$ computes the conditional covariance matrix of $\boldsymbol{M}_1^\dagger \vec{Z}$ conditioned on $\boldsymbol{M}_2^\dagger \vec{Z}$ for $\vec{Z} \sim \mathcal{N}(0, \mathbf{\Sigma})$. This is given by the formula:

$$\mathrm{CVAR}(\boldsymbol{M}_1 | \boldsymbol{M}_2; \mathbf{\Sigma}) := \boldsymbol{M}_1^\dagger \mathbf{\Sigma} \boldsymbol{M}_1 - \boldsymbol{M}_1^\dagger \mathbf{\Sigma} \boldsymbol{M}_2 \,\mathrm{pinv}(\boldsymbol{M}_2^\dagger \mathbf{\Sigma} \boldsymbol{M}_2) \boldsymbol{M}_2^\dagger \mathbf{\Sigma} \boldsymbol{M}_2.$$

- ALPHA0$(\mathbf{\Sigma}_Q, \boldsymbol{A})$ in Algorithm 3 implements a strategy for choosing $\boldsymbol{\nu}_0$ in Theorems 8, 9.

- ALPHA$(\mathbf{\Sigma}, \boldsymbol{A})$ in Algorithm 4 implements a strategy for choosing $\boldsymbol{\nu}_s$ in Theorems 8, 9.

Algorithm 3 Strategy for choosing $\boldsymbol{\nu}_0$ for Theorems 8, 9

---

function ALPHA0$(\mathbf{\Sigma})$ ▷ Find $\boldsymbol{\nu}_0$ which minimizes $\mathrm{var}\left(Y_K - \boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]} \,\Big|\, \boldsymbol{A}\vec{Y}_{[K-1]}\right)$ for
$\mathbf{\Sigma} = \mathrm{var}\, \vec{Y}_{[K-1]}, \boldsymbol{A} = \mathrm{STAGES}^*(\mathbf{\Sigma})$.
    $\boldsymbol{A} \leftarrow \mathrm{STAGES}^*(\mathbf{\Sigma})$.
    $\{\boldsymbol{M}, \vec{v}, c\} \leftarrow \mathrm{CVARCOMPONENTS}(\mathbf{\Sigma}, \boldsymbol{A})$
    $\{\boldsymbol{\nu}, \sigma^2\} \leftarrow \mathrm{ICQM}(\boldsymbol{M}, \vec{v}, c)$
    $n_0 \leftarrow 1$
    return $\{n_0, \boldsymbol{\nu}, \sigma^2\}$
end function

Algorithm 4 Strategy for picking $\boldsymbol{\nu}_s$ for Theorems 8, 9.

---

    function ALPHA($\boldsymbol{\Sigma}, \boldsymbol{A}$)   ▷ Entropy-greedy implementation: choose $\boldsymbol{\nu}$ where the unknown part of $\boldsymbol{\nu}^\dagger \vec{Y}_{[K-1]}$ has the least entropy among any combination with an unknown part. Expects $\boldsymbol{\Sigma} = \text{var}\, \vec{Y}_c$, $\boldsymbol{A} = \text{STAGES}^*\!\left(\text{var}\left(\vec{Y}_c \middle| [\boldsymbol{\nu}_0, \ldots, \boldsymbol{\nu}_{s-1}]^\dagger \vec{Y}_c\right)\right)$.

        if rank $\boldsymbol{A} = K$ **then**

            $\boldsymbol{\nu} \leftarrow 0$

        else

            $\boldsymbol{\Sigma}_{\text{reduced}} \leftarrow \text{CVAR}(\boldsymbol{I}_K | \boldsymbol{A};\ \boldsymbol{\Sigma})$, $\boldsymbol{\nu} \leftarrow \text{SLVC}(\boldsymbol{\Sigma}_{\text{reduced}})$

        end if

        return $\boldsymbol{\nu}$

    end function

D.2   Proof of Lemmas 3, 4, Theorem 8

Proof. (Lemma 3)

Take $\mathcal{D} := \text{mod}_{B_c}(\boldsymbol{\nu}_0^\dagger(\vec{U}_{[K-1]} - \vec{W}_{[K-1]}))$. Then, by modulo's distributive property $\mathcal{D} = \text{mod}_{B_c}(\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]})$ so that $\tilde{Y} = -\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]} \in (\mathcal{D} + L_c)$. Compute:

$$\frac{1}{n_0}\,\text{mod}_{n_0 B_c}\,\tilde{Y} = \frac{1}{n_0}\,\text{mod}_{n_0 B_c}(-\vec{\alpha}_0^\dagger \vec{Y}_{[K-1]}).$$

$$= \text{mod}_{B_c}\left(-\frac{1}{n_0}\vec{\alpha}_0^\dagger \vec{Y}_{[K-1]}\right).$$

Now:

$$U_K = \text{mod}_{B_c}\left(W_k + Y_k + \frac{1}{n_0}\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]} - \frac{1}{n_0}\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]}\right)$$

$$= \text{mod}_{B_c}\left(W_k + Y_0 - \frac{1}{n_0}\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]}\right)$$

$$= \text{mod}_{B_c}\left(W_k + \mathcal{E}\left(Y_0 \middle| \vec{A}\right) + \mathcal{E}_\perp\left(Y_0 \middle| \vec{A}\right) - \frac{1}{n_0}\boldsymbol{\nu}_0^\dagger \vec{Y}_{[K-1]}\right)$$

$$= \text{mod}_{B_c}\left(W_k + \mathcal{E}\left(Y_0 \middle| \vec{A}\right) + \tilde{Y}_\perp + \frac{1}{n_0}\tilde{Y}\right).$$

By (Chapman et al., 2019, Theorem 1), $\vec{A}$ can be recovered by processing $(\vec{U}_{[K-1]}, \vec{W}, \tilde{Y})$, hence $\mathcal{E}\left(Y_0 \middle| \vec{A}\right)$ can also be recovered. Choose $\mathcal{C} := -W_k + \mathcal{E}\left(Y_0 \middle| \vec{A}\right)$ so that the claim holds applying modulo's distributive property. $\qquad\square$

Proof. (Lemma 4)

Take $U_0 = n_0 U_K + \boldsymbol{\nu}_0^\dagger U_{[K-1]}$, $W_0 = n_0 W_K + \boldsymbol{\nu}_0^\dagger W_{[K-1]}$, $\vec{U}_c = (U_0, \vec{U}_{[K]}), \vec{W}_c = (W_0, \vec{W})$. Take $\mathcal{C} := \mathcal{E}\left(\boldsymbol{\nu}^\dagger \vec{Y}_c \middle| \vec{A}\right)$ and $\mathcal{D} := \text{mod}_{B_c}(\boldsymbol{\nu}^\dagger(\vec{U}_c - \vec{W}_c) - \mathcal{E}\left(\boldsymbol{\nu}^\dagger Y_c \middle| \vec{A}\right)) = \text{mod}_{B_c}(\mathcal{E}_\perp\left(\boldsymbol{\nu}^\dagger Y_c \middle| \vec{A}\right))$. Choose $\beta := \frac{\text{cov}\left(Y, \tilde{Y} \middle| \vec{A}\right)}{\delta^2}$.

Include good-for-coding auxillary lattices with the prescribed scales in the lattice ensemble from Theorem 8. With high probability since $\hat{L}$ is good for coding semi norm-ergodic noise of power $\delta^2 + \varepsilon$ Ordentlich and Erez (2016) and applying (Krithivasan and Pradhan, 2007, Appendix V) to $\tilde{Y}$, $\tilde{Y}_\perp$ yields the result. □

Proof. (Theorem 8)

D.2.1 Upper Bound for Singleton S

Take a nested lattice construction from (Erez et al., 2005, Theorem 1), involving the following sets:

- Coarse and fine encoding lattices $L_c, L_1, \ldots, L_K$ (base regions $B_c, B_1, \ldots, B_K$) with each k has $L_c \subset L_k$ designed with nesting ratio $\frac{1}{n} \log |B_c \cap L_k| \to r_k$.

- Discrete part auxiliary lattices $\hat{L}_1, \ldots, \hat{L}_K$ (base regions $\hat{B}_1, \ldots, \hat{B}_K$) with each $\hat{L}_k \subset L_c$ having nesting ratio $\frac{1}{n} \log |B_c \cap \hat{L}_k| \to \frac{1}{2} \log \delta_k^2$.

- Initial residual part auxiliary lattice $\hat{L}'_0$ (base region $\hat{B}'_0$) with $\hat{L}'_0 \subset L_K$, nesting ratio $\frac{1}{n} \log |\hat{B}'_0 \cap L_K| \to \frac{1}{2} \log \sigma_0^2$.

- Residual part auxiliary lattices $\hat{L}'_1, \ldots, \hat{L}'_K$ (base regions $\hat{B}'_1, \ldots, \hat{B}'_K$) with each $\hat{L}'_k \subset L_K$, having nesting ratio $\frac{1}{n} \log |\hat{B}'_k \cap L_K| \to \frac{1}{2} \log \sigma_k^2$.

The specified nesting ratios for the auxiliary lattices, $\sigma_0^2, \sigma_1^2, \ldots, \sigma_K^2, \delta_1^2, \ldots, \delta_K^2$ will be specified later.

Initialization

Apply Lemma 3 to U_K, and label the resulting variables $\tilde{Y}_0 := \tilde{Y}$, $\tilde{Y}_{0\perp} := \tilde{Y}_\perp$, $(\hat{L}'_0, \hat{B}'_0) := (\hat{L}', \hat{B}')$, $\mathcal{D}_0 := n_0 \mathcal{D}$, $\mathcal{C}_0 := \mathcal{C}$, $\sigma_0^2 := \sigma^2$. In addition, define $\delta_0^2 := n_0^2$, $\beta_0 = \frac{1}{n_0}$, $\hat{B}_0 := n_0 B_c$ Now,

$$U_K = \text{mod}_{B_c} \left(\mathcal{C}_0 + \beta_0 \tilde{Y}_0 + \tilde{Y}_{0\perp} \right)$$

so the support of U_K is contained within:

$$\begin{aligned}
\mathcal{S}_0 :=& [\mathcal{C}_0 + \hat{B}'_0 + (L_c/n_0 + \mathcal{D}_0)] \cap (B_c \cap L_K) \\
=& [\mathcal{C}_0 + \hat{B}'_0 + \beta_0[(\mathcal{D}_0 + L_c) \cap \hat{B}_0] \cap (B_c \cap L_K).
\end{aligned}$$

Support Reduction

Iterate over steps $s = 1, \ldots, K$. For step s, condition on any event of the form $\tilde{Y}_{(s-1)} = \ell_s \in (\mathcal{D}_{s-1} + L_c) \cap \hat{B}_{s-1}$, of which there are no more than $2^{n \cdot (\log(\delta_{s-1}^2) + \varepsilon)}$ choices due to the nesting ratio for $\hat{L}_{s-1}$ in L_c. Take $\boldsymbol{A}_s := \text{STAGES}^* \left(\text{var} \left(\vec{Y}_c \Big| [\boldsymbol{\nu}_0, \ldots, \boldsymbol{\nu}_{s-1}]^\dagger \vec{Y}_c \right) \right)$. By (Chapman et al., 2019, Theorem 1), $\vec{A}_s := \boldsymbol{A}_s^\dagger \vec{Y}_c$ is recoverable by processing $(\vec{A}_{s-1}, \text{mod}_{n_0 B_c} Y_0 \vec{U}_{[K]}, \vec{W})$.

Now, apply Lemma 4 to $(Y, \boldsymbol{\nu}, \boldsymbol{A}) = (\tilde{Y}_{(s-1)\perp}, \boldsymbol{\nu}_s, \boldsymbol{A}_s)$, and label the resulting variables $\tilde{Y}_s := \tilde{Y}$, $\tilde{Y}_{s\perp} := \tilde{Y}_\perp$, $(\hat{L}_s, \hat{B}_s) := (\hat{L}, \hat{B})$, $(\hat{L}'_s, \hat{B}'_s) := (\hat{L}', \hat{B}')$, $\mathcal{D}_s := \mathcal{D}$, $\mathcal{C}_s := \mathcal{C}$, $\beta_s := \beta$, $\sigma_s^2 := \sigma^2$, $\delta_s^2 := \delta^2$. Now,

$$U_K = \mathrm{mod}_{B_c}\left(\left[\sum_{t=0}^s \mathcal{C}_t + \beta_t \tilde{Y}_t\right] + \tilde{Y}_{s\perp}\right)$$

so the support of U_K is contained within:

$$\mathcal{S}_s := \left[\left[\sum_{t=0}^s \mathcal{C}_t + \beta_t\left[(\mathcal{D}_t + L_c) \cap \hat{B}_t\right]\right] + \hat{B}'_s\right] \cap (B_c \cap L_K).$$

Count Points in Estimated Supports

By design, there are no more than $\prod_{t=0}^s 2^{n \cdot (\frac{1}{2}\log(\delta_t^2) + \varepsilon)}$ possible choices for $\mathcal{S}_s$. Each $\mathcal{S}_s$ has no more than $|\hat{B}'_s \cap (B_c \cap L_K)| \leq 2^{n \cdot (r_K + \frac{1}{2}\log(\sigma_s^2) + \varepsilon)}$ points. Then,

$$H(U_K|\vec{U}_{[K-1]}, \vec{W}) \leq \min_{s \in \{0\} \cup [K]} n \cdot \left(r_s + \frac{1}{2}\log(\sigma_s^2) + \sum_{t=0}^s \frac{1}{2}\log(\delta_t^2) + K\varepsilon\right).$$

Bound Simultaneity

An argument is given in Section D.2.1 for an upper bound on the singleton case. The argument uses a Zamir-good nested lattice construction with a finite amount of nesting criteria, and conditions on a finite amount of high-probability events. Then, the argument holds for all cases of this form simultaneously by using a Zamir-good nested lattice construction satisfying all of each case's nesting criteria and conditioning on all of each case's high-probability events.

The entropy for the general case $S = \{s_1, \ldots, s_{|S|}\}$, $T = \{t_1, \ldots, t_{|T|}\}$ can be rewritten using the chain rule:

$$H\left(\vec{U}_S \middle| \vec{U}_T, \vec{W}\right) = \sum_{p=1}^{|S|} H\left(\vec{U}_{s_p} \middle| \vec{U}_{\{s_m : m < p\} \cup T}, \vec{W}\right). \quad \square$$

D.3 Sketch of Theorem 9 for Upper Bound on Entropy-Rates of Decentralized Processing Messages

Proof. (Sketch) Proceed identically as in the proof of Theorem 8 in Appendix D.2 up until either Section D.2.1 Initialization if definition for Y_0 was changed, or repetition s where $\boldsymbol{\nu}_s = [0, 0, \ldots, 0, 1]$ in Section D.2.1 Support Reduction if definition for $(\boldsymbol{\nu}_k)_k$ changed. In this portion, perform the following analysis instead. Compute:

$$\begin{aligned}
\mathcal{D}_{(msg)} &:= \mathrm{mod}_{B_{c,\mathrm{msg}}}((\vec{a}_{\mathbb{R}}^{(msg)} + \vec{a}_{\mathbb{Z}}^{(msg)})^\dagger \vec{Y}_{[K-1]} - W_{\mathrm{msg}}) \\
&= \mathrm{mod}_{B_{c,\mathrm{msg}}}(\lambda^{(msg)} X_{\mathrm{msg}}^n + Y_\perp^{(msg)} - W_{\mathrm{msg}})
\end{aligned}$$

$$= \mathrm{mod}_{B_{c,\mathrm{msg}}} \left(\frac{\lambda^{(msg)}}{\gamma_n} \mathrm{mod}_{B_{c,\mathrm{msg}}}(M + W_{\mathrm{msg}}) + Y_{\perp}^{(msg)} - W_{\mathrm{msg}} \right)$$

$$= \mathrm{mod}_{B_{c,\mathrm{msg}}} \left(\left(1 + \frac{\lambda^{(msg)}}{\gamma_n} - 1 \right) \mathrm{mod}_{B_{c,\mathrm{msg}}}(M + W_{\mathrm{msg}}) + Y_{\perp}^{(msg)} - W_{\mathrm{msg}} \right)$$

$$= \mathrm{mod}_{B_{c,\mathrm{msg}}} \left(M + \left(\frac{\lambda^{(msg)}}{\gamma_n} - 1 \right) \mathrm{mod}_{B_{c,\mathrm{msg}}}(M + W_{\mathrm{msg}}) + Y_{\perp}^{(msg)} \right). \quad \text{(D.1)}$$

The additive terms in Equation (D.1) are independent of one another, and the terms besides M have observed power $\delta^2_{(msg)}$. Choose the nesting ratio for $L_{f,\mathrm{msg}}$ in $\hat{B}_s$ as

$$\hat{r}_s := \frac{1}{2} \log \left(\delta_s^2 \right).$$

Then, with high probability since $\hat{L}_s$ is good for coding semi norm-ergodic noise below power δ_s^2 Ordentlich and Erez (2016) and applying (Krithivasan and Pradhan, 2007, Appendix V) to the derivation in Equation (D.1),

$$M \in \mathcal{L}_{(msg)} := (L_{f,\mathrm{msg}} \cap B_{c,\mathrm{msg}}) \cap \mathrm{mod}_{B_{c,\mathrm{msg}}} \left(\mathcal{D}_{(msg)} + \hat{B}_s \right), \quad \text{(D.2)}$$

where $\mathcal{D}_{(msg)}$ is computable by processing $(\vec{U}_{[K-1]}, \vec{W}, (\tilde{Y}_t)_{[s-1]})$ and

$$\frac{1}{n} \log |\mathcal{L}_{(msg)}| \leq \hat{r}_s + \varepsilon.$$

Rearranging Equation (D.2),

$$X_{\mathrm{msg}}^n = \mathrm{mod}_{B_{c,\mathrm{msg}}}(M + W_{\mathrm{msg}}) \in \mathcal{L}_s := \mathrm{mod}_{B_{c,\mathrm{msg}}} \left(\mathcal{L}_{(msg)} + W_{\mathrm{msg}} \right).$$

Now define:

$$\tilde{Y}_s := X_{\mathrm{msg}}^n,$$
$$\tilde{Y}_{s\perp} := \mathcal{E}_{\perp} \left(\tilde{Y}_{(s-1)\perp} \Big| X_{\mathrm{msg}}^n \right),$$
$$\mathcal{C}_s := \mathcal{E} \left(\tilde{Y}_{s-1} \Big| \vec{A}_s \right),$$
$$\beta_s := \frac{\mathrm{cov} \left(\tilde{Y}_{(s-1)\perp}, Y_s \Big| \vec{A}_s \right)}{\delta_s^2},$$
$$\sigma_s^2 := \mathrm{var}(\tilde{Y}_{s\perp}).$$

By construction, $\tilde{Y}_{s\perp}$ is the components in $\tilde{Y}_{(s-1)\perp}$ uncorrelated with X_{msg}^n:

$$\tilde{Y}_{(s-1)\perp} = \beta_s \tilde{Y}_s + \tilde{Y}_{s\perp} + \mathcal{C}_s,$$
$$\tilde{Y}_s \in \mathcal{L}_s.$$

Proceed as in proof of Theorem 8. $\qquad \square$

D.4   Proof of Lemma 5 for Recombination of Decentralized Processing Lattice Modulos

Proof. By (Chapman et al., 2019, Theorem 1), a processing of $\vec{U}_{[K]}$ with high probability outputs

$$\mathbf{a}_{\mathbb{R}}^{\dagger} \vec{Y}_{[K]},$$
$$\mathbf{a}_{\mathbb{R}} \in \text{image STAGES}^{*}(\mathbf{\Sigma}) \subset \mathbb{R}^{K}.$$

One can assume the nested lattices for the message transmitter, $L_{f,\text{msg}} \supset L_{c,\text{msg}}$, are part of the lattice ensemble from Theorem 8, in particular ones finer than the main coarse lattice L_c so that $L_c \subseteq L_{c,\text{msg}}$ and:

$$\frac{1}{n} \log |L_{c,\text{msg}} \cap B_c| \to \hat{r}_{c,\text{msg}} \geq 0.$$

With this structure, then, for any $\mathbf{a}_{\mathbb{Z}} \in \mathbb{Z}^{K}$, the encodings can be processed to produce (using lattice modulo's distributive and subgroup properties)

$$\text{mod}_{L_{c,\text{msg}}} \left(\text{mod}_{L_c} \left(\mathbf{a}_{\mathbb{R}}^{\dagger} \vec{Y}_{[K]} + \mathbf{a}_{\mathbb{Z}}^{\dagger} (\vec{U}_{[K]} - \vec{W}_{[K]}) \right) \right) = \dots$$
$$\text{mod}_{L_{c,\text{msg}}} \left(\text{mod}_{L_c} \left(\mathbf{a}_{\mathbb{R}}^{\dagger} \vec{Y}_{[K]} + \mathbf{a}_{\mathbb{Z}}^{\dagger} \vec{Y}_{[K]} \right) \right) = \dots$$
$$\text{mod}_{L_{c,\text{msg}}} \left(\mathbf{a}_{\mathbb{R}}^{\dagger} \vec{Y}_{[K]} + \mathbf{a}_{\mathbb{Z}}^{\dagger} \vec{Y}_{[K]} \right) = \dots$$
$$\text{mod}_{L_{c,\text{msg}}} \left(\lambda X_{\text{msg}} + Y_{\text{noise}} \right),$$

where, in Equation (5.2), $\lambda \in \mathbb{R}$ and Y_{noise} is the conglomerate of noise terms independent of X_{msg} that are left over.

For channels with additive Gaussian noise, Y_{noise} is a mixture of Gaussians and independent components uniform over good-for-quantization lattice base regions, so Y_{noise} will probably, for long enough blocklength, land inside the base of any coarse enough good-for-coding lattice (Krithivasan and Pradhan, 2007, Appendix V).   □

D.5   Proof of Corollary 2 for Achievability of the Decentralized Processing Rate

Proof. Fix any $r_{\text{msg}}, \mathbf{a}_{\mathbb{Z}}, \mathbf{a}_{\mathbb{R}}, \lambda, \sigma_{\text{noise}}^2$ from their definitions in Lemma 5 and any $\gamma^2 \in (0,1]$. Choose a communications rate R_{msg} satisfying the criterion in the statement. Form an ensemble of lattices such as those described in Theorem 9, with nesting ratio for L_c in L_{msg} as $\frac{1}{2} \log(1/\gamma^2)$ for $\gamma \in (0,1)$ and $L_{\text{msg}} = L_c$ if $\gamma^2 = 1$. This design means $\gamma_n^2 := \text{var mod}_{B_{c,\text{msg}}}(X_{\text{msg}} + W_{\text{msg}}) \to_n \gamma^2$.

Have the transmitter encode its message M into a modulation X_{msg}^n as described at the beginning of Section 5.3.1 using a dither W_{msg} of which all helpers and the decoder are informed. Have each k-th helper, $k = 1, \dots, K$, process its observation vector into a lattice modulo encoding U_k as described in Theorem 9 using a dither W_k of which the decoder is informed.

By Theorem 9, there exists a Slepian–Wolf binning scheme such that each k-th helper can process its message U_k into a compression U_k^* with $\frac{1}{n} H(U_k^*) < R_k$, and

where a decoder can with high probability process the ensemble of compressions $(U_1^*, \ldots, U_K^*)$ along with dither side information $(\vec{W}, W_{\text{msg}})$ into $(U_1, \ldots, U_K)$. Employ this binning scheme at each of the receivers, and have them each forward their compressions U_k^* to the decoder.

Have the decoder decompress $(U_1^*, \ldots, U_K^*)$ into $(\hat{U}_1, \ldots, \hat{U}_K)$. By the previous statement, with high probability, $(\hat{U}_1, \ldots, \hat{U}_K) = \vec{U}$. Use the processing obtained from Lemma 5 on $(\hat{U}_1, \ldots, \hat{U}_K)$, with high probability producing a signal:

$$U_{\text{proc}} := \text{mod}_{L_{c,\text{msg}}} \left(\lambda X_{\text{msg}} + Y_{\text{noise}} \right).$$

Decoding

Decoding proceeds similar to Erez and Zamir (2004). At the decoder, compute:

$$
\begin{aligned}
U'_{\text{proc}} :=& \text{mod}_{L_{c,\text{msg}}} \left(U_{\text{proc}} - W_{\text{msg}} \right) = \ldots \\
& \text{mod}_{L_{c,\text{msg}}} \left(\frac{\lambda}{\gamma_n} \text{mod}_{L_{c,\text{msg}}} \left(M + W_{\text{msg}} \right) + Y_{\text{noise}} - W_{\text{msg}} \right) = \ldots \\
& \text{mod}_{L_{c,\text{msg}}} \left(\left(1 + \frac{\lambda}{\gamma_n} - 1 \right) \text{mod}_{L_{c,\text{msg}}} \left(M + W_{\text{msg}} \right) + Y_{\text{noise}} - W_{\text{msg}} \right) = \ldots \\
& \text{mod}_{L_{c,\text{msg}}} \left(M + \left(\frac{\lambda}{\gamma_n} - 1 \right) \text{mod}_{L_{c,\text{msg}}} \left(M + W_{\text{msg}} \right) + Y_{\text{noise}} \right).
\end{aligned} \tag{D.3}
$$

Recall that the fine codebook lattice $L_{f,\text{msg}}$ has been designed to be good for coding and so that the coarse codebook lattice $L_{c,\text{msg}}$ has a nesting ratio within it as R_{msg}. This means that $L_{f,\text{msg}}$ is good for coding semi norm-ergodic noise with power less than $\gamma_n^2 2^{-2R_{\text{msg}}}$.

Notice $M \perp \text{mod}_{L_{c,\text{msg}}} \left(M + W_{\text{msg}} \right) \perp Y_{\text{noise}}$, where the first independence is by the crypto lemma Zamir (2014). This is to say that additive terms other than M in Equation (D.3) are noise with power

$$
\begin{aligned}
\text{var} & \left\{ \left(\frac{\lambda}{\gamma_n} - 1 \right) \text{mod}_{L_{c,\text{msg}}} \left(M + W_{\text{msg}} \right) + Y_{\text{noise}} \right\} = \ldots \\
& \gamma_n^2 \cdot (1 - \lambda/\gamma_n)^2 + \sigma_{\text{noise}}^2.
\end{aligned} \tag{D.4}
$$

Furthermore, by (Krithivasan and Pradhan, 2007, Appendix V) on the noise, then it is probably in the base region of any lattice good for coding semi norm-ergodic noise with power less than Equation (D.4). Then, $\text{round}_{L_{f,\text{msg}}}(U'_{\text{proc}}) = M$ with high probability if

$$\gamma_n^2 \cdot (1 - \lambda/\gamma_n)^2 + \sigma_{\text{noise}}^2 < \gamma_n^2 2^{-2R_{\text{msg}}},$$

or, rearranging,

$$R_{\text{msg}} < \frac{1}{2} \log \left\{ \frac{\gamma_n^2}{(\lambda - \gamma_n)^2 + \sigma_{\text{noise}}^2} \right\}. \tag{D.5}$$

The limit of the right side of Equation (D.5) equals Equation (5.4). $\qquad \square$

APPENDIX E

LOWER BOUND PROOFS

Lemmas for calculation of entropies of the form $H(\vec{Y}_T|\vec{W})$ are needed to establish the lower bound for Theorem 8. Writing $T = \{t_1, \ldots, t_{|T|}\}$ then using the chain rule

$$H(\vec{Y}_T|\vec{W}) = \sum_{p=1}^{|T|} H(Y_{t_p}|\vec{Y}_{\{t_q : q < a\}p}, \vec{W}).$$

Thus it is enough to treat a case of the form $H(Y_K|\vec{Y}_{[K-1]}, \vec{W})$.

Lemma 8. Take $\vec{Y}, \vec{W}$ as described in the statement of Theorem 8. Then, the Zamir-good lattices in the statement can be designed such that eventually

$$\left| \frac{1}{n} H(Y_K|\vec{Y}_{[K-1]}, \vec{W}) - \left[r_K - \frac{1}{2} \log \frac{|[\mathbf{\Sigma}_Q]_{[K-1],[K-1]}|}{|\mathbf{\Sigma}_Q|} \right] \right| \leq \varepsilon.$$

Proof. The entropy-rate in question will be upper and lower bounded to within ε.
Upper Bound:
 By the data processing inequality,

$$\frac{1}{n} H(Y_K|\vec{Y}_{[K-1]}, \vec{W}) \leq \frac{1}{n} H\left(Y_K \Big| W_K, \mathcal{E}\left(Y_K \Big| \vec{Y}_{[K-1]} \right) \right)$$
$$= \frac{1}{n} H\left(Y_K - \mathcal{E}\left(Y_K \Big| \vec{Y}_{[K-1]} \right) \Big| W_K, \mathcal{E}\left(Y_K \Big| \vec{Y}_{[K-1]} \right) \right).$$

Take

$$\sigma^2 := \mathrm{var}\left(X_K + W_K^* \Big| \vec{X}_{[K-1]} + \vec{W}_{[K-1]}^* \right) + \varepsilon.$$

Note $\sigma^2 \geq \mathrm{var}\left(Y_K \Big| \mathcal{E}\left(Y_K \Big| \vec{Y}_{[K-1]} \right) \right)$ by consideration of the involved variables' covariance matrices.
 Now assume the lattice ensemble from the statement in Theorem 8 has a good lattice $L_{\text{lem}} \subset L_K$ with base region B_{lem} and nesting ratio for L_{lem} in L_K approaching $\frac{1}{2} \log \sigma^2 / 2^{-2r_K}$. Since L_{lem} is good for coding semi norm-ergodic noise of power σ^2 Ordentlich and Erez (2016), applying (Krithivasan and Pradhan, 2007, Appendix V) to $Y_K - \mathcal{E}\left(Y_K \Big| \vec{Y}_{[K-1]} \right)$, then with high probability $Y_K - \mathcal{E}\left(Y_K \Big| \vec{Y}_{[K-1]} \right) \in B_{\text{lem}}$. Then, taking $\boldsymbol{v}^\dagger = \mathrm{cov}(X_K, \vec{X}_{[K-1]} + \vec{W}_{[K-1]})$,

$$\frac{1}{n} H(Y_K|\vec{Y}_{[K-1]}, \vec{W}) \leq \frac{1}{2} \log \frac{\sigma^2}{2^{-2r_K}} + \varepsilon$$
$$= \frac{1}{2} \left[\log \frac{\mathrm{var}\, X_K + W_K^*}{2^{-2r_K}} - \ldots \right.$$
$$\left. \log(1 - \boldsymbol{v}^\dagger \, \mathrm{var}(\vec{X}_{[K-1]} + \vec{W}_{[K-1]})^{-1} \boldsymbol{v}) \right] + \varepsilon$$
$$= \frac{1}{2} \log \frac{\mathrm{var}\, X_K + W_K^*}{2^{-2r_K}} - I(X_K + W_K^*; \vec{X}_{[K-1]} + \vec{W}_{[K-1]}^*) + \varepsilon.$$

101

Lower Bound:

$$
\begin{aligned}
H(Y_K|\vec{Y}_{[K-1]}, \vec{W}) &= H(Y_K|\vec{W}) - I(Y_K; \vec{Y}_{[K-1]}|\vec{W}) \\
&= H(Y_K|\vec{W}) - I(Y_K; \vec{Y}_{[K-1]}) + I(Y_K; \vec{Y}_{[K-1]}; \vec{W}) \\
&= H(Y_K|\vec{W}) - I(Y_K; \vec{Y}_{[K-1]}) \qquad\qquad\qquad\text{(E.1)} \\
&= H(Y_K|\vec{W}) - h(Y_K) - h(\vec{Y}_{[K-1]}) + h(\vec{Y}) \\
&= H(Y_K|W_K) - h(Y_K) - h(\vec{Y}_{[K-1]}) + h(\vec{Y}) \\
&\geq n \cdot \left(\frac{1}{2} \log \frac{\operatorname{var} Y_K}{2^{-2r_K}} - \varepsilon \right) - h(Y_K) - h(\vec{Y}_{[K-1]}) + h(\vec{Y}). \quad\text{(E.2)}
\end{aligned}
$$

where Equation (E.1) is because (noting $W_k = -\operatorname{mod}_{L_k}(Y_k)$)

$$
\begin{aligned}
I(Y_K; \vec{Y}_{[K-1]}; \vec{W}) &= h(Y_K) + h(\vec{Y}_{[K-1]}) + h(\vec{W}) \dots \qquad\qquad\text{(E.3)} \\
&\quad - h(Y_K, \vec{Y}_{[K-1]}) - h(Y_K, \vec{W}) - h(\vec{Y}_{[K-1]}, \vec{W}) \dots \\
&\quad + h(Y_K, \vec{Y}_{[K-1]}, \vec{W}) \\
&= h(Y_K) + h(\vec{Y}_{[K-1]}) + h(\vec{W}) \dots \\
&\quad - h(Y_K, \vec{W}) - h(\vec{Y}_{[K-1]}, \vec{W}) \\
&= h(Y_K) + h(\vec{Y}_{[K-1]}) + h(\vec{W}) \dots \\
&\quad - h(Y_K) - h(W_{[K-1]}) - h(\vec{Y}_{[K-1]}) - h(W_K) \\
&= h(Y_K) + h(\vec{Y}_{[K-1]}) \dots \\
&\quad - h(Y_K) - h(\vec{Y}_{[K-1]}) \\
&= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(E.4)}
\end{aligned}
$$

and Equation (E.2) is because generating a dither W_K and encoding X_K as $(Y_K, \vec{W})$ is a rate-distortion-efficient quantization scheme for that source and rate r_k. By (Zamir and Feder, 1996, Equation (22)) on each independent $W_1, \dots, W_K$ and the vector entropy power inequality, eventually:

$$
\begin{aligned}
\frac{1}{n} H(Y_K|\vec{Y}_{[K-1]}, \vec{W}) &\geq -2\varepsilon + \frac{1}{2} \log \frac{\operatorname{var} X_K + W_K^*}{2^{-2r_K}} - \dots \\
& h(X_K + W_K^*) - h(\vec{X}_{[K-1]} + \vec{W}_{[K-1]}^*) + h(\vec{X}^* + \vec{W}^*)
\end{aligned}
$$

with $\vec{X}, \vec{W}^*$ as in the statement. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Theorem 11. Take $\vec{Y}, \vec{U}, \vec{W}$ as described in the statement of Theorem 8. Then for disjoint $S, T \subset [K-1]$ define

- $\vec{U}' := \left(\vec{U}_S, \left(\operatorname{mod}_{L_c} \frac{Y_t}{\sqrt{\operatorname{var} Y_t + \varepsilon}} \right)_{t \in T}, \frac{Y_K}{\sqrt{\operatorname{var} Y_K + \varepsilon}} \right).$

- $\vec{Y}' := \left(\vec{Y}_S, \left(\frac{Y_t}{\sqrt{\operatorname{var} Y_t + \varepsilon}} \right)_{t \in T}, \frac{Y_K}{\sqrt{\operatorname{var} Y_K + \varepsilon}} \right).$

- $r'_K := r_K + \frac{1}{2} \log \operatorname{var} Y_K$.

then the Zamir-good lattices in the statement can be designed such that

$$\frac{1}{n} H(Y_K | \vec{U}_{S \cup \{K\}}, \vec{Y}_T, \vec{W}) \leq \frac{1}{n} H\left(U'_{|S|+|T|+1} \middle| \vec{U}'_{S \cup T}, \vec{W}\right) + \varepsilon.$$

and fixing variables $(\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_K, n_0)$ as described in Theorem 8 then the upper bound described therein holds on the variables $(\vec{U}', \vec{Y}', r'_K, \vec{W})$.

Proof. Eventually in n, with high probability each $U'_k = Y'_k$ for $k = |S| + 1, \ldots, |S| + |T| + 1$. Also $\frac{1}{n} H(Y_K | \vec{U}_{S \cup \{K\}}, \vec{Y}_T, \vec{W}) \leq \frac{1}{n} H(Y_K | U_K)$ In the nested lattice ensemble from the statement, include lattice good-for-coding within noise power $\operatorname{var} Y_K$ with base region B and $\frac{1}{n} \log |L_K \cap B| \to_n r_K + \frac{1}{2} \log \operatorname{var} Y_K$. Then with high probability $Y_K \in B$. Take $\mathcal{E}$ as the intersection of all these high-probability events. Then:

$$\frac{1}{n} H(Y_K | \vec{U}_{S \cup \{K\}}, \vec{Y}_T, \vec{W}) \leq \frac{1}{n} [H(Y_K | \vec{U}_{S \cup \{K\}}, \vec{Y}_T, \vec{W}, \mathcal{E}) + H(\mathbf{1}_{\mathcal{E}})]$$

$$\leq \frac{1}{n} H(Y_K | \vec{U}_{S \cup \{K\}}, \vec{Y}_T, \vec{W}, \mathcal{E}) + \frac{1}{n}$$

Then the first inequality in the statement holds for large enough n. The rest of the proof follows identically to the provided one for Theorem 8 on variables $(\vec{U}', \vec{Y}', r'_K, \vec{W})$. $\square$

APPENDIX F

APPLICATION OF MAIN RESULTS TO COMPLEX CHANNEL STRUCTURE

The model described in Section 1.1.1 is complex, but the rest of the body treats a real model. The model in Section 1.1.1 is equivalent to the real channel where the k-th observer receives two real signals:

$$X_{k,\mathrm{raw}}^{\Re} := \Re(h_k X_{\mathrm{src}} + W_k)$$
$$X_{k,\mathrm{raw}}^{\Im} := \Im(h_k X_{\mathrm{src}} + W_k)$$

where the vector $(X_{[K],\mathrm{raw}}^{\Re}, X_{[K],\mathrm{raw}}^{\Im})$ has covariance structure:

$$(X_{[K],\mathrm{raw}}^{\Re}, X_{[K],\mathrm{raw}}^{\Im}) \sim \mathcal{N}\left(0, \begin{bmatrix} \Re(\boldsymbol{\Sigma}_{\mathrm{noise}} + \vec{h}\vec{h}^{\dagger}) & -\Im(\boldsymbol{\Sigma}_{\mathrm{noise}} + \vec{h}\vec{h}^{\dagger}) \\ \Im(\boldsymbol{\Sigma}_{\mathrm{noise}} + \vec{h}\vec{h}^{\dagger}) & \Re(\boldsymbol{\Sigma}_{\mathrm{noise}} + \vec{h}\vec{h}^{\dagger}) \end{bmatrix}\right). \tag{F.1}$$

Apply encoding strategies as described in the body of the study to this equivalent real channel as follows. Treat each observer as two virtual ones, one observing $X_{k,\mathrm{raw}}^{\Re}$ and the other $X_{k,\mathrm{raw}}^{\Im}$. Take equal scales $\alpha_k^{\Re} = \alpha_k^{Im}$ in the structured schemes. When observer k is designated observer-to-base rate $R_k > 0$ each of the virtual observers forms a rate-$\frac{1}{2}R_K$ encoding of its observation.

Due to the block-antisymmetric structure of the covariance matrix in (F.1) and evenly split message rates, any linear combination of the form

$$[a_1, \ldots, a_K, \ b_1, \ldots, b_K]$$

of the encodings has equal recovery probability from Lemma 1 as the combinations

$$[b_1, \ldots, b_K, \ a_1, \ldots, a_K],$$
$$[a_1, \ldots, a_K, \ -b_1, \ldots, -b_K].$$

Then the real and imaginary components of the underlying complex signals are recovered with equal probability to identical distortion levels through the described schemes. Two options for treating the complex channel model are:

- Re-derive each result using complex lattices and different source coding formulae when appropriate.

- Apply the reduction to real channels as described above.

The first option can yield some performance gains in low-blocklength implementations since an n-dimensional complex lattice can be designed to have at least as good performance characteristics as those of a real n-dimensional lattice.