Applied and Computational Harmonic Analysis **10**, 203–233 (2001)
doi:10.1006/acha.2000.0340, available online at http://www.idealibrary.com on IDEAL®

Quantized Frame Expansions with Erasures [1]

Vivek K. Goyal and Jelena Kovačević

Mathematics of Communication Research, Bell Labs, Lucent Technologies,
600 Mountain Avenue, Murray Hill, New Jersey 07974
E-mail: v.goyal@ieee.org, jelena@bell-labs.com

and

Jonathan A. Kelner [2]

Harvard University, 38 Leverett Mail Center, Cambridge, Massachusetts 02138
E-mail: kelner@fas.harvard.edu

Communicated by Henrique S. Malvar

Frames have been used to capture significant signal characteristics, provide numerical stability of reconstruction, and enhance resilience to additive noise. This paper places frames in a new setting, where some of the elements are deleted. Since proper subsets of frames are sometimes themselves frames, a quantized frame expansion can be a useful representation even when some transform coefficients are lost in transmission. This yields robustness to losses in packet networks such as the Internet. With a simple model for quantization error, it is shown that a normalized frame minimizes mean-squared error if and only if it is tight. With one coefficient erased, a tight frame is again optimal among normalized frames, both in average and worst-case scenarios. For more erasures, a general analysis indicates some optimal designs. Being left with a tight frame after erasures minimizes distortion, but considering also the transmission rate and possible erasure events complicates optimizations greatly.

1. INTRODUCTION

Signal representations using frames—redundant sets of vectors in contrast to bases—are used for a variety of reasons, including resilience to additive noise [6], resilience to quantization [14], numerical stability of reconstruction [6], and greater freedom to capture significant signal characteristics [1, 2, 28]. The redundancy of a frame can also mitigate the effect of losses in packet-based communication systems [10, 12, 13]. This paper details

[1] Some of the results in this paper were reported at the IEEE Data Compression Conference (Snowbird, UT, March 1999).

[2] J. A. Kelner contributed this work as an intern with Bell Labs, Lucent Technologies.

1063-5203/01 $35.00

this relatively new application of frames and related results pertaining to finite-dimensional frames.

A modern communication network, be it the public Internet or a private network, provides a means to transport packets of data from one device to another. These packets are sequences of information bits of a certain length surrounded by error-control, addressing, and timing information that assure that the packet is delivered without errors to the intended recipient with the identity or sequencing of the packet intact. Packets are either delivered without error or not delivered at all, with failures due primarily to buffer overflows at intermediate nodes in the network. We may abstract the behavior of the network as delivering a packet with some probability of failure and some random, though relatively predictable, delivery time.

To most users, the behavior of a packet network is characterized not by random losses, but by unpredictable transport time. This is due to a protocol, invisible to the user, that retransmits lost packets. The detection of a missing packet and the subsequent retransmission of the packet generally takes much longer than a successful packet transmission, generating the highly variable delay. In many applications, large delay in unacceptable. Thus retransmission of lost packets is not feasible; instead, one must make due with whatever is received.

If a lost packet is independent of all other transmitted data, then the information in the lost packet is indeed unrecoverable at the receiver. On the other hand, dependencies between transmitted packets facilitate complete or partial recovery in the face of losses. Deterministic dependencies lead to conventional channel coding [19], while statistical dependencies give the techniques proposed in [30] and generalized in [11].

The encoding structure considered in this paper combines elements of deterministic and statistical redundancy. Refer to Fig. 1. We denote the $\mathbb{R}^N$-valued information to be communicated by x. The source vector is represented through a frame expansion with frame operator F, yielding $y = Fx \in \mathbb{R}^M$. The scalar quantization of the frame expansion coefficients gives $\hat{y}$ lying in a discrete subset of $\mathbb{R}^M$. We abstract the effect of the network to be the erasure of some components of $\hat{y}$. This implies that the components of $\hat{y}$ are placed in more than one packet, for otherwise all of $\hat{y}$ would be lost at once. If they are placed in M separate packets, then any subset of the components of $\hat{y}$ may be received; otherwise only certain subsets will be possible. The reconstruction process is unspecified for now. Neglecting quantization, the redundancy is completely deterministic; modeling the quantization error as random additive noise or using randomized quantization makes the redundancy statistical.

To contrast this with a classical approach, consider the system shown in Fig. 2. The source vector x is quantized in its original basis representation. Then a block channel code C adds redundant (parity) symbols to aid in reconstruction in the case of transmission

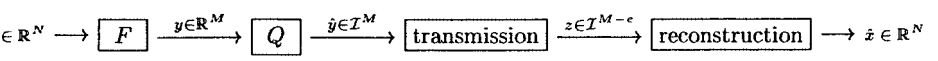

FIG. 1. Abstraction of a communication system using a quantized frame expansion. The signal vector $x \in \mathbb{R}^N$ is expanded with a frame operator F to give the frame coefficient vector $y \in \mathbb{R}^M$. The scalar quantization of y gives $\hat{y}$, which is transmitted over a network that erases some components. A reconstruction $\hat{x} \in \mathbb{R}^N$ is computed from the received vector z.

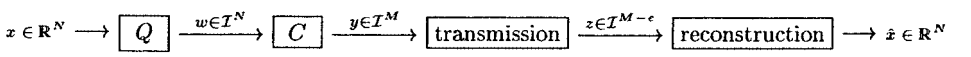

FIG. 2. A communication system using a block channel-coded basis representation, to contrast with Fig. 1. Instead of introducing redundancy with a frame in $\mathbb{R}^N$, a parity symbol in $\mathcal{I}^{M-N}$ is appended to the message symbol w. Up to $M - N$ erasures can be completely corrected, but beyond this number the performance drops sharply, as shown in Fig. 3.

losses. Such a system works well when there are at most $M - N$ lost symbols because the quantized representation w can be recovered. With more erasures, the performance sharply deteriorates because many quantized values are compatible with the received vector. There is no benefit from very few erasures because the parity information is completely redundant. Qualitatively, the classical approach is good when there are approximately $M - N$ erasures, as shown in Fig. 3.

EXAMPLE 1.1. Suppose we have vectors $\{x^{(i)}\}_{i=1}^K \subset \mathbb{R}^N$ to communicate to a receiver using M packets sent over a network. Suppose further that $N = 2$ and each vector is independent and has independent, zero-mean, unit-variance Gaussian components. A typical packet size for a network using Internet Protocol, Version 6 (IPv6) is 576 bytes, since packets of this length must be handled without fragmentation [7]. With 40 bytes allocated to headers, each packet has a payload of 536 bytes. We will send $K = 536$ vectors in $M = 3$ packets.

The classical approach would be to apply an eight-bit quantizer to each component of each vector, yielding two-tuples of eight-bit strings $\{w^{(i)}\}_{i=1}^K$. Reconstruction R (inverse quantization) yields mean-square error (MSE) $E|x_j^{(i)} - R(w_j^{(i)})|^2 \approx 8.8 \cdot 10^{-5}$ for all i and j. To provide robustness against the loss of a packet, three packets are produced

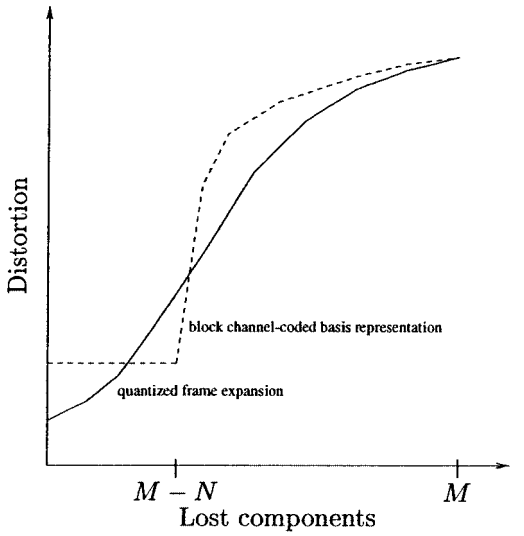

FIG. 3. Qualitative comparison of encoding and transmission using a quantized frame expansion (see Fig. 1) versus a block channel-coded basis representation (see Fig. 2). The distortion of the basis representation jumps sharply with more than $M - N$ lost symbols. This paper considers only zero to $M - N$ erasures.

in the following manner: Packet 1 contains $\{w_1^{(i)}\}_{i=1}^K$, Packet 2 contains $\{w_2^{(i)}\}_{i=1}^K$, and Packet 3 contains $\{w_1^{(i)} \oplus w_2^{(i)}\}_{i=1}^K$, where $\oplus$ denotes bitwise exclusive-or or $\mathbb{Z}_2$ addition. The sequences $\{w_1^{(i)}\}_{i=1}^K$ and $\{w_2^{(i)}\}_{i=1}^K$ can be recovered from any two of the three packets. However, receiving all the packets is no better than receiving two, and receiving only Packet 3 is essentially useless.

A typical use of a quantized frame expansion would first use

$$F = \begin{bmatrix} 1 & 0 \\ -1/2 & \sqrt{3}/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix}$$

to compute $y^{(i)} = Fx^{(i)}$, $i = 1, 2, \ldots, K$. (We will see that this choice of F is optimal in several ways.) Then each component of each $y^{(i)}$ is quantized with an eight-bit quantizer to obtain $\{\hat{y}^{(i)}\}_{i=1}^K$. Three packets are formed with Packet j containing $\{\hat{y}_j^{(i)}\}_{i=1}^K$. The advantages of this representation are the symmetry of the three packets and the fact that each packet contains information that cannot be inferred completely from the other two.

The per component MSE distortions for all of the possible combinations of received packets are given in the following table:

Packets received	$\{1, 2, 3\}$	Any two	$\{1\}$ or $\{2\}$	$\{3\}$	$\emptyset$
Classical	$8.8 \cdot 10^{-5}$	$8.8 \cdot 10^{-5}$	0.5	1	1
Quantized frame	$5.8 \cdot 10^{-5}$	$1.2 \cdot 10^{-4}$	0.5	0.5	1

The quantized frame system is worse in one situation and better in two others; the best choice depends on the likelihoods of these events and/or other design criteria. Note that K is immaterial, so our analysis is based simply on components of $\hat{y}$ being erased.

In this paper we consider only what happens when there are at most $M - N$ lost components. This allows us, under a very mild condition, to show that the received vector z specifies the source vector x to within a bounded set. Then by using a statistical model for the quantization error $\hat{y} - y$, the reconstruction error does not depend on the source vector x, but instead only on properties of the frame operator F. To analyze cases with more than $M - N$ lost components requires a statistical model for the source vector [13].

The structure of the paper is to add one block at a time: First, in Section 2, we consider just expanding with a frame and reconstructing. In this case, most matrices F allow the signal vector x to be reconstructed exactly, so unlike in other sections, the reconstruction error is not an issue. In Section 3 we add a quantization block and analyze the reconstruction error with different noise models and reconstruction methods. In Section 4 we introduce the possibility of lost coefficients, and thus address the full system of Fig. 1.

2. FRAME EXPANSIONS

To begin our study of quantized frame expansions with erasures, let us look at only the first ingredient: the frame expansion, as depicted in Fig. 4. As long as F has full rank, the signal vector x can be recovered exactly from the expansion vector y; the accuracy of the reconstruction $\hat{x}$ is not an issue. Nevertheless, many interesting properties of F may

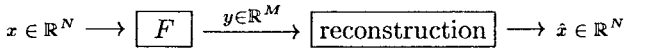

FIG. 4. The frame system. F stands for the frame operator.

be characterized. In this section we introduce notation and terminology and establish basic properties of frames.

2.1. Frame Fundamentals

This introduction to frames is adapted from [6, Chap. 3]. Since our application is the communication of real vectors, we consider only frames in $\mathbb{R}^N$, though we retain notation that makes extensions to $\mathbb{C}^N$ obvious. All of the concepts apply in any Hilbert space, though, e.g., minima and maxima may have to be replaced by infima and suprema.

Let $\Phi = \{\varphi_k\}_{k=1}^M \subset \mathbb{R}^N$. Φ is called a *frame* if there exist $A > 0$ and $B < \infty$ such that

$$A\|x\|^2 \leq \sum_{k=1}^{M} |\langle x, \varphi_k \rangle|^2 \leq B\|x\|^2, \qquad \text{for all } x \in \mathbb{R}^N. \tag{1}$$

A and B are called the *frame bounds*. The lower bound in (1) is equivalent to requiring that Φ spans $\mathbb{R}^N$, so a frame will always have $M \geq N$. Also, notice that one can choose $B = \sum_{k=1}^M |\varphi_k|^2$ whenever $M < \infty$; thus, any finite set of vectors that spans $\mathbb{R}^N$ is a frame. We will refer to $r = M/N$ as the *redundancy* of the frame. We will encode exclusively with *uniform frames*, that is, frames with $\|\varphi_k\| = 1$ for $k = 1, \ldots, M$.

Given a frame $\Phi = \{\varphi_k\}_{k=1}^M$ in $\mathbb{R}^N$, the associated *frame operator* F is the linear operator from $\mathbb{R}^N$ to $\mathbb{R}^M$ defined by

$$(Fx)_k = \langle x, \varphi_k \rangle, \qquad \text{for } k = 1, 2, \ldots, M. \tag{2}$$

Denoting vectors as columns, this operation is a left matrix multiplication where F is an $M \times N$ matrix with the kth row equal to φ_k^*.[3] Using the frame operator F, (1) can be rewritten as

$$AI_N \leq F^*F \leq BI_N, \tag{3}$$

where I_N is the $N \times N$ identity matrix and the matrix inequality means

$$x^*AI_Nx \leq x^*F^*Fx \leq x^*BI_Nx, \qquad \text{for all } x \in \mathbb{R}^N. \tag{4}$$

Considering x as an eigenvector of F^*F in (4) gives the following property:

PROPERTY 2.1. *For any frame, the eigenvalues of F^*F lie in the interval $[A, B]$.*

[3] The superscript * denotes a transpose; as suggested by the notation, the Hermitian transpose should be used in $\mathbb{C}^N$.

Another elementary condition on the eigenvalues is the following:

PROPERTY 2.2. *For any frame, the sum of the eigenvalues of F^*F equals the sum of the lengths of the frame vectors. In particular, for a uniform frame the sum of the eigenvalues equals M.*

Proof. Denote the eigenvalues by $\{\lambda_i\}_{i=1}^N$. Using elementary properties of the trace and the definition of F,

$$\sum_{i=1}^{N} \lambda_i = \operatorname{tr}(F^*F) = \operatorname{tr}(FF^*) = \sum_{i=1}^{M} \varphi_i^* \varphi_i = \sum_{i=1}^{M} \|\varphi_i\|^2. \quad \blacksquare$$

It follows from Property 2.1 that F^*F is invertible (all of its eigenvalues are nonzero), and furthermore

$$B^{-1} I_N \leq (F^*F)^{-1} \leq A^{-1} I_N. \tag{5}$$

The *dual frame* of Φ is a frame defined as $\tilde{\Phi} = \{\tilde{\varphi}_k\}_{k=1}^M$, where

$$\tilde{\varphi}_k = (F^*F)^{-1} \varphi_k, \qquad \text{for } k = 1, 2, \ldots, M. \tag{6}$$

Noting that $\tilde{\varphi}_k^* = \varphi_k^* (F^*F)^{-1}$ and stacking $\tilde{\varphi}_1^*, \tilde{\varphi}_2^*, \ldots, \tilde{\varphi}_M^*$ in a matrix, the frame operator associated with $\tilde{\Phi}$ is

$$\tilde{F} = F(F^*F)^{-1}. \tag{7}$$

Since $\tilde{F}^*\tilde{F} = (F^*F)^{-1}$, (5) shows that B^{-1} and A^{-1} are frame bounds for $\tilde{\Phi}$.

2.2. Tight Frames

A frame Φ is called a *tight frame* if the frame bounds in (1) can be taken to be equal. Tight frames constitute an important class of frames. Since they are self-dual, they have some desirable reconstruction properties that also extend to frames with B/A close to one. In the context of this work, optimality properties of tight frames are established in Subsections 3.2 and 4.2.1.

For a tight frame, (1) reduces to something similar to Parseval's equality: a tight frame operator scales the energy of an input by a constant factor A. Looking at (3), we can say that $F^*F = AI_N$ if and only if Φ is a tight frame. Moreover, Property 2.1 simplifies to the following, where the value of A follows from Property 2.2:

PROPERTY 2.3. *For a tight frame, F^*F has eigenvalue A with multiplicity N. If the frame is also uniform, $A = M/N = r$.*

The fact that F^*F is diagonal means that the columns of F are orthogonal. (Recall that it is the *rows* of F that constitute the frame; these are not orthogonal unless $r = 1$.) Viewing the columns of F as vectors in $\mathbb{R}^M$, it is obvious that F can be extended to an orthogonal basis:

PROPERTY 2.4. *For a tight frame, F is the first N columns of an $M \times M$ matrix with orthogonal columns.*

Reversing our point of view to start from the higher-dimensional space, a tight frame is a projection of an orthogonal basis for $\mathbb{R}^M$ to an N-dimensional subspace. With a uniform tight frame, we have a more specific characterization:

PROPERTY 2.5. *For a uniform tight frame, F is the first N columns of $\sqrt{M/N} \cdot U$ for some orthogonal matrix U.*

Proof. Since $F^*F = (M/N)I_N$, each column of F has Euclidean norm $\sqrt{M/N}$. Using the Gram–Schmidt process, append $M - N$ columns to F such that the columns are orthogonal and have norm $\sqrt{M/N}$; call the resulting matrix V. Now $U = \sqrt{N/M} \cdot V$ is an orthogonal matrix. ■

2.2.1. Classifying Uniform Tight Frames. As they form a useful subset of the set of frames, it is of interest to classify the uniform tight frames that exist for a given N and M. This can be useful in designing a tight frame for a particular application.

Define an equivalence relation for frames by bundling a frame with all frames that can result from rigid rotations or reflections of the entire frame, as well as the negation of some individual vectors (see Fig. 5). Of course, since the equivalence is between sets, the numbering or permutation of elements is irrelevant. Along with preserving tightness, this equivalence relation preserves the geometric arrangement of lines obtained by extension of the frame vectors. For example, suppose we rotate, reflect, and negate the frame Φ to obtain the frame Γ. Then the new frame vectors are

$$\gamma_k = \sigma_k U \varphi_k, \qquad \text{for } k = 1, 2, \ldots, M,$$

where $\sigma_k = \pm 1$ and U is some unitary matrix. The new frame operator G can be written as

$$G = \Sigma F U^*, \qquad \text{where } \Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_M).$$

It is now easy to see that the tightness of Φ implies the tightness of Γ since

$$G^*G = UF^* \underbrace{\Sigma^*\Sigma}_{I_M} FU^* = U \underbrace{F^*F}_{AI_N} U^* = AI_N.$$

Furthermore, the uniformity of Φ implies the uniformity of Γ.

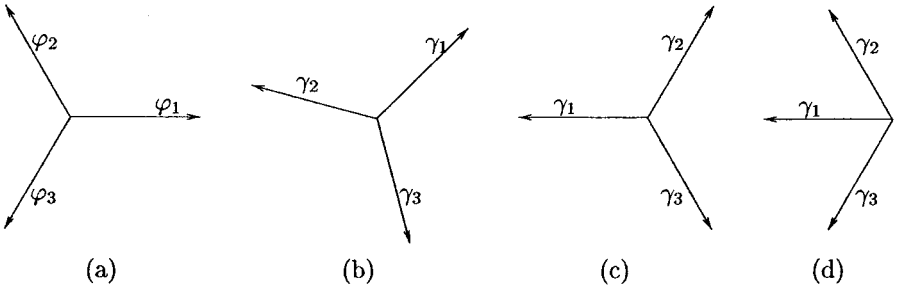

FIG. 5. Examples of the equivalence relation for tight frames. $\Phi = \{\varphi_1, \varphi_2, \varphi_3\}$ is a frame with $M = 3$ elements in $\mathbb{R}^2$. The other frames are in the same equivalence class. (a) Original tight frame. (b) Rotation of the frame by 45 degrees. (c) Reflection of the frame around the vertical axis. (d) Negation of a single vector φ_1.

With this concept of equivalence, tight frames with $M = N + 1$ are essentially unique:

THEOREM 2.6. *There is exactly one equivalence class of uniform tight frames for* $M = N + 1$.

Proof. See Appendix A.1. ■

For example, the tight frame shown in Fig. 5a and used in Example 1.1 describes all possible tight frames with $N = 2$ and $M = 3$.

Unfortunately, when M exceeds $N + 1$, there are uncountably many equivalence classes of the type described above; thus, we cannot systematically obtain all uniform tight frames. However, at least for $N = 2$, the uniform tight frames still have a simple characterization:

THEOREM 2.7. *The following are equivalent:*

(1) $\{\varphi_k = (\cos\alpha_k, \sin\alpha_k)\}_{k=1}^{M}$ *is a uniform tight frame.*
(2) $\sum_{k=1}^{M} z_k = 0$, *where* $z_k = e^{j2\alpha_k}$ *for* $k = 1, 2, \ldots, M$.

Proof. See Appendix A.2. ■

With this characterization, one set of solutions has z_k's equal to the Mth roots of unity. These give examples of the *harmonic frames* that are defined in Subsection 2.2.2. A uniform tight frame that is the union of L smaller uniform tight frames is obtained when not only $\sum_{k=1}^{M} z_k = 0$, but also there is a partition of $\{1, 2, \ldots, M\}$ into $S_1, S_2, \ldots, S_L$ such that $\sum_{k \in S_i} z_k = 0$ for $i = 1, 2, \ldots, L$.

For example, the only solution for $M = 3$ is the "Mercedes-Benz" frame (see Fig. 5). For $M = 4$, the only possibility is that the frame is a union of two orthonormal bases. However, there are infinitely many solutions parameterized by the angle between the two bases. For $M = 5$, the solutions include the harmonic tight frame and frames obtained as the union of an orthonormal basis and a Mercedez-Benz tight frame.

Notice that all frames built as unions of smaller frames must have redundancy at least 2. Our final attempt to classify uniform tight frames relates the possibility of repeated vectors to the redundancy. Repeated vectors are undesirable for encoding because they lead to transform coefficients (inner products) carrying no new information.

THEOREM 2.8. *Let* $\{\varphi_j\}_{j=1}^{M}$ *be a uniform tight frame for* $\mathbb{R}^N$ *or* $\mathbb{C}^N$ *and let* $r = M/N$. *Then at most* $K = \lfloor r \rfloor$ *elements of the frame can be equal. If* $K = M/N$, *these elements are orthogonal to the rest. There can be* N *such* K*-tuples of equal elements and each* K*-tuple is orthogonal to the span of the rest.*

Proof. See Appendix A.3. ■

2.2.2. Construction of Uniform Tight Frames. For the encoding applications developed in this paper, we would like to be able to construct a uniform tight frame for any given N and M. Citing Property 2.5, it is tempting to claim that taking the first N columns of any $M \times M$ orthogonal matrix does the trick. While doing so will give a tight frame, it will generally not give a *uniform* tight frame. For example, the first N columns of I_M are the frame operator associated with a tight frame, but it is clearly not uniform since $M - N$ vectors in the frame are zero. For some orthogonal matrices, taking the first N columns works; however, we have no useful parameterization of these matrices.

One very useful family of frames in $\mathbb{C}^N$—harmonic frames—can be obtained starting with an $M \times M$ discrete Fourier transform (DFT) basis by projecting it onto an

N-dimensional space. This family is given by

$$(\varphi_{k+1})_{i+1} = \frac{1}{\sqrt{N}} W_M^{ki}, \qquad i = 0, \ldots, N-1,\ k = 0, \ldots, M-1, \tag{8}$$

where $W_M = e^{j2\pi/M}$ is the Mth root of unity. Computing an expansion with this frame is like computing a DFT; it can be done with a fast Fourier transform-like algorithm, which can be a great savings over a general matrix multiplication.

Real harmonic tight frames can be defined similarly: if N is even, let

$$\varphi_{k+1} = \sqrt{\frac{2}{N}} \left[\cos\frac{k\pi}{M},\ \cos\frac{3k\pi}{M},\ \ldots,\ \cos\frac{(N-1)k\pi}{M}, \right.$$
$$\left. \sin\frac{k\pi}{M},\ \sin\frac{3k\pi}{M},\ \ldots,\ \cos\frac{(N-1)k\pi}{M} \right]^T \tag{9}$$

for $k = 0, 1, \ldots, M-1$; if N is odd, let

$$\varphi_{k+1} = \sqrt{\frac{2}{N}} \left[\frac{1}{\sqrt{2}},\ \cos\frac{2k\pi}{M},\ \cos\frac{4k\pi}{M},\ \ldots,\ \cos\frac{(N-1)k\pi}{M}, \right.$$
$$\left. \sin\frac{2k\pi}{M},\ \sin\frac{4k\pi}{M},\ \ldots,\ \cos\frac{(N-1)k\pi}{M} \right]^T \tag{10}$$

for $k = 0, 1, \ldots, M-1$. Exhibiting this specific family of tight frames gives the following application of Theorem 2.6:

COROLLARY 2.9. *Any uniform tight frame with $M = N + 1$ is equivalent to a harmonic tight frame.*

Despite the difficulty of computing many distinct uniform tight frames, almost all uniform frames with high redundancy are approximately tight. More precisely, a sequence of random uniform frames with increasing redundancy will asymptotically approach a tight frame:

THEOREM 2.10 (Tightness of Random Frames [14]). *Let $\{\Phi_M\}_{M=N}^{\infty}$ be a sequence of frames in $\mathbb{R}^N$ such that Φ_M is generated by choosing M vectors independently with a uniform distribution on the unit sphere in $\mathbb{R}^N$. Let F_M be the frame operator associated with Φ_M. Then, in the mean squared sense,*

$$\frac{1}{M} F_M{}^* F_M \to \frac{1}{N} I_N \text{ elementwise as } M \to \infty.$$

2.3. Reconstruction from Frame Coefficients

The frame representation y, as applied in this paper, is a tool for communication of the information signal x; thus, the reconstruction or estimation of x from y is of fundamental interest.

One possibility is to use the *Moore–Penrose generalized inverse* or *pseudoinverse* of F [17],

$$F^\dagger = (F^* F)^{-1} F^*, \tag{11}$$

where the frame condition ensures that the inverse in (11) exists. It is easy to check that $F^\dagger F = I_N$, so $F^\dagger(Fx) = x$ for all $x \in \mathbb{R}^N$. The pseudoinverse is not the only linear

reconstruction operator with this property. Suppose we wish to find all $G \in \mathbb{R}^{N\times M}$ such that $G(Fx) = x$ for all $x \in \mathbb{R}^N$. Since Fx lies in a particular N-dimensional subspace, it does not matter what G does to vectors orthogonal to this subspace; this gives many additional degrees of freedom. Denote the singular value decomposition of F by

$$F = \underbrace{U}_{M\times M}\ \Sigma\ \underbrace{V^*}_{N\times N}, \qquad \text{where } \Sigma = \begin{bmatrix} \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_N) \\ 0_{(M-N)\times N} \end{bmatrix},$$

and U and V are unitary. Then the pseudoinverse is given by

$$F^\dagger = V^*\Sigma^\dagger U, \qquad \text{where } \Sigma^\dagger = [\operatorname{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_N^{-1}),\ 0_{N\times(M-N)}].$$

However, with any $W \in \mathbb{R}^{N\times(M-N)}$,

$$G = V^*[\operatorname{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_N^{-1}),\ W]U$$

satisfies $GF = I_N$. There is something special about the pseudoinverse, though: when $y = Fx$ is not known exactly, it eliminates the influence of errors that are orthogonal to the range of F. Specifically, if instead of having access to $y = Fx$ we have $\hat{y} = y + \eta$, then $F^\dagger\hat{y}$ will be equivalent to the reconstruction from the orthogonal projection of $\hat{y}$ onto the range of F.

The pseudoinverse is precisely the transpose of the frame operator associated with the dual frame (7), that is,

$$F^\dagger = \tilde{F}^*. \tag{12}$$

Thus reconstructing with the pseudoinverse is equivalent to using the frame expansion coefficients as weights in a linear combination of dual frame elements:

$$x = F^\dagger Fx = \tilde{F}^*Fx = \sum_{k=1}^{M} \langle x, \varphi_k\rangle\tilde{\varphi}_k. \tag{13}$$

As in the previous matrix-oriented formulation, other reconstruction formulas are possible, but using the dual frame minimizes reconstruction errors; for details the reader is referred to [6, Sect. 3.2].

If M is much larger than N, computing the dual frame may be prohibitively expensive. An iterative procedure for approximating the dual frame while avoiding matrix inversion is given in [6, Sect. 3.2].

3. QUANTIZED FRAME EXPANSIONS

Quantization is to approximate a signal varying continuously in amplitude by one whose amplitude is restricted to a prescribed set of discrete values. Since digital communication systems transmit only discrete values, any digital communication of continuous-valued information includes quantization. Introducing a quantization block gives us the system shown in Fig. 6 and brings us one step closer to the overall system depicted in Fig. 1.

To understand this and the following sections, it suffices to consider *uniform quantization*: rounding to the nearest multiple of a fixed quantization step size Δ. When a vector is

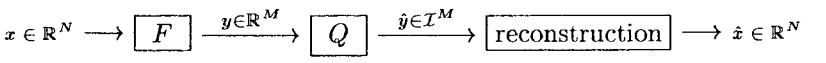

FIG. 6. Schematic representation of a quantized frame expansion with reconstruction.

quantized, the rounding is done separately for each component of the vector. Decreasing Δ makes the quantized representation more accurate, but increases the amount of data that must be transmitted. The choice of Δ is arbitrary; we will consider optimizing performance by varying the frame and the reconstruction method.

Here the details of quantization are unimportant because we make strong assumptions about the quantization error $\hat{y} - y$; this is described in Subsection 3.1. For readers wishing to learn more about quantization, Gray and Neuhoff have written an excellent, comprehensive introduction to quantization with many historical details [15].

3.1. Quantization Models

One can always consider quantization to be the addition of a noise signal $\eta = \hat{y} - y$, but η is not just any noise. "Noise" connotes unpredictability and perhaps a mysterious origin; here the origin is not at all mysterious and the noise signal itself is usually a deterministic function of the input. Nevertheless, modeling quantization noise stochastically leads to tractable analyses and useful results.

The calculations in Subsections 3.2 and 4.2 are based on the following assumptions:

(a) Each noise component η_i has mean zero and variance σ^2.
(b) The noise components are uncorrelated; i.e., η_i and η_j are uncorrelated for $i \neq j$.

These can be expressed as

$$E[\eta_i] = 0 \qquad \text{and} \qquad E[\eta_i \eta_j] = \delta_{ij}\sigma^2 \qquad \text{for all } i, j. \tag{14}$$

In Subsection 3.3, assumption (a) is replaced by the following stronger assumption:

(a′) Each noise component η_i is uniformly distributed on $[-\Delta/2, \Delta/2]$.

Assumption (a) follows from (a′) with $\sigma^2 = \Delta^2/12$.

In typical deterministic quantization, these assumptions do not hold, but they may be approximately true. Consider uniform quantization of a random vector x. If the probability density of x is smooth and Δ is small, then each marginal density can be well-approximated by a piecewise constant function that is constant on every interval $((k - \frac{1}{2})\Delta, (k + \frac{1}{2})\Delta)$, $k \in \mathbb{Z}$. Under this approximate density, each noise component is uniformly distributed as in (a′). Furthermore, if a multidimensional analog of this holds for the density itself—instead of marginals—assumption (b) also holds.

Note that despite the digression to justify the model, in subsequent analyses a stochastic model is used for the noise only—not for the source; all expectations are with respect to η. If the source is random, we must assume η is independent of x.

Assumptions (a′) and (b) are strictly accurate if *uniform subtractive dithered quantization* is used. Denote a uniform scalar quantizer (rounding operation) with step size Δ by $Q_{\text{uniform},\Delta}$. A subtractive dithered version is then defined by

$$Q_{\text{dither},\Delta}(y) = Q_{\text{uniform},\Delta}(y + z) - z,$$

where z is uniformly distributed on $[\Delta/2, \Delta/2)^M$ and is generated independently each time the quantizer is applied. With this quantizer, assumptions (a′) and (b) hold [16, 20]. [4] Since the output of $Q_{\text{dither},\Delta}$ is not discrete, this quantizer can actually be used only if z is pseudorandom and known at the decoder. In this case $Q_{\text{uniform},\Delta}(y+z)$ is transmitted and the subtraction of z is done only at the decoder.

3.2. Linear Reconstruction

In Subsection 2.3, we saw that there are linear operators that recover the original signal x from the frame expansion y. In this section we find the reconstruction error in using such a linear operator to reconstruct from the quantized representation $\hat{y}$, under the quantization noise model (14). We will find that given a frame, the dual frame operator should be used in reconstruction, and that among frames of any given dimension, tight frames are best.

Suppose we wish to approximate $x \in \mathbb{R}^N$ given $\hat{y} = Fx + \eta$, where $F \in \mathbb{R}^{M\times N}$ is a frame operator and $\eta \in \mathbb{R}^M$ satisfies (14). With no further information about η, it makes sense to choose $\hat{x}$ to minimize the residual $\|F\hat{x} - \hat{y}\|_2$. It is well known that the pseudoinverse provides the solution to this problem [17],

$$\hat{x} = F^\dagger \hat{y}, \tag{15}$$

with $F^\dagger$ as in (11). We refer to (15) as a *linear reconstruction* because a linear operator $F^\dagger$ is used. This reconstruction is equivalent to the identity (13), except the exact expansion coefficients are replaced by noisy ones,

$$\hat{x} = F^\dagger \hat{y} = F^\dagger(Fx+\eta) = \tilde{F}^*(Fx+\eta) = \sum_{k=1}^{M} (\langle x, \varphi_k\rangle + \eta_k)\tilde{\varphi}_k, \tag{16}$$

where we used (12).

Let us now calculate the error of a linear reconstruction:

$$x - \hat{x} = \sum_{k=1}^{M} \langle x, \varphi_k\rangle \tilde{\varphi}_k - \sum_{k=1}^{M} (\langle x, \varphi_k\rangle + \eta_k)\tilde{\varphi}_k = -\sum_{k=1}^{M} \eta_k \tilde{\varphi}_k.$$

The expected squared-ℓ_2 error per component (mean-squared error) is

$$\begin{aligned}
\text{MSE} &= \frac{1}{N} E\|x-\hat{x}\|^2 = \frac{1}{N} E\left\| \sum_{k=1}^{M} \eta_k \tilde{\varphi}_k \right\|^2 \\
&= \frac{1}{N} E\left[\sum_{i=1}^{M}\sum_{k=1}^{M} \eta_i \eta_k \tilde{\varphi}_i^* \tilde{\varphi}_k \right] = \frac{1}{N} \sum_{i=1}^{M}\sum_{k=1}^{M} \delta_{ik}\sigma^2 \tilde{\varphi}_i^* \tilde{\varphi}_k \qquad (17) \\
&= \frac{1}{N}\sigma^2 \sum_{k=1}^{M} \|\tilde{\varphi}_k\|^2, \qquad (18)
\end{aligned}$$

where (17) results from evaluating expectations using the model (14) for η. Further simplifications can be made using (7) and basic properties of the trace,

[4] Furthermore, the noise components are mutually independent and independent of y.

$$\text{MSE} = \frac{1}{N}\sigma^2 \sum_{k=1}^{M} \|\tilde{\varphi}_k\|^2 = N^{-1}\sigma^2 \operatorname{tr}(\tilde{F}\tilde{F}^*) = N^{-1}\sigma^2 \operatorname{tr}((F^*F)^{-1})$$
$$= N^{-1}\sigma^2 \operatorname{tr}(V\Lambda^{-1}V^*) = N^{-1}\sigma^2 \operatorname{tr}(\Lambda^{-1}),$$

where $F^*F = V\Lambda V^*$ is the spectral decomposition of F^*F. With the $\{\lambda_i\}_{i=1}^N$ denoting the eigenvalues of F^*F, we have

$$\text{MSE} = \frac{1}{N}\sigma^2 \sum_{i=1}^{N} \frac{1}{\lambda_i}. \tag{19}$$

The characterization of the MSE in terms of the spectrum of F^*F allows us to consider the choice of a frame to minimize the MSE, leading to the following theorem:

THEOREM 3.1. *When encoding with a uniform frame and decoding with linear reconstruction* (16)*, under the noise model* (14)*, the MSE is minimum if and only if the frame is tight.*

Proof. Recall that the sum of the eigenvalues of F^*F is constant and equal to M (Property 2.2). Thus, we are attempting to minimize the MSE given in (19) by the sum $\sum_{i=1}^{M} \lambda_i^{-1}$ subject to the constraint that the sum of the λ_i's is constant. This occurs when all of the eigenvalues are equal, which, in turn, is true if and only if the frame is tight. ■

These computations and a couple of simple consequences are summarized by the following theorem:

THEOREM 3.2. *Consider linear reconstruction* (16) *with noise* η *satisfying* (14) *and define the mean-squared error* (MSE) *by* $N^{-1}E\|x - \hat{x}\|^2$. *For any frame, the* MSE *is given by* (19) *and satisfies*

$$B^{-1}\sigma^2 \le \text{MSE} \le A^{-1}\sigma^2. \tag{20}$$

For a uniform frame,

$$\frac{N\sigma^2}{M} \le \text{MSE} \le A^{-1}\sigma^2. \tag{21}$$

For a uniform tight frame,

$$\text{MSE} = \frac{N}{M}\sigma^2 = r^{-1}\sigma^2. \tag{22}$$

Proof. See Appendix A.4. ■

3.2.1. Gaussian Case. In the beginning of this section, linear reconstruction was justified by the minimization of a residual. This merely intuitive appeal can be replaced by an optimality claim if the signal x and received vector $\hat{y}$ are jointly Gaussian. This occurs if x and η are independent Gaussian vectors.

Denote an estimation function $\hat{x} = R(\hat{y})$. This estimator minimizes the mean-squared error $E\|x - \hat{x}\|^2$ if it is the conditional expectation [22]: $R(\hat{y}) = E[x|\hat{y}]$. If x and $\hat{y}$ are jointly Gaussian, the conditional expectation is a linear function [22]. Specifically, the best estimator is the dual frame reconstruction (16).

As the reader might suspect, the Gaussian case would not be singled out if linear reconstruction were more generally optimal. Since quantization error is almost always

far from Gaussian, x and $\hat{y}$ are usually not jointly Gaussian, and we would be remiss to not mention ways to improve upon the performance of linear reconstruction. The next subsection discusses this.

3.3. Consistent Reconstruction

3.3.1. Deterministic Quantization. A deterministic scalar quantizer is a partitioning of the real numbers into intervals along with a labeling of the intervals. Knowing a quantized value $\hat{w} = Q(w)$ gives a hard constraint on the value of unknown scalar variable w; w must lie in the interval $Q^{-1}(\hat{w})$.[5] The quantized value of the vector of frame expansion coefficients $\hat{y} = Q(y)$ gives M of these constraints, which can be written as

$$\alpha_k \le y_k < \beta_k, \qquad k = 1, 2, \ldots, M.$$

Recalling that $y_k = \langle x, \varphi_k \rangle$, each of the $2M$ inequalities corresponds to an $(N-1)$-dimensional hyperplane that x must lie above or below. Together, these hyperplanes demarcate a convex *consistent set*. An estimate $\hat{x}$ in the consistent set computed from $\hat{y}$ is called a *consistent estimate* [25].

Consistent estimates can be obtained by alternating projections onto convex sets (POCS) [31] or by linear programming. Both of these techniques are described in [14]. An efficient algorithm that works for some frames and gives similar performance is given in [3].

When F is a frame and the quantization intervals are finite (the α_k's and β_k's are finite), the consistent set is bounded. Intuitively, this boundedness means that we know more about the value of x than is revealed by a noise model like (14), since the noise model allows unbounded η. This intuition is supported by various results from [4, 5, 14, 23, 25–27]. With a variety of analysis techniques and considering various families of frames, these papers establish that consistent reconstruction techniques give MSE that can be asymptotically approximated as cr^{-2}, where c is a constant that depends on the source and quantization and the redundancy r is approaching infinity. This is the best possible asymptotic decay of the MSE as a function of the redundancy, so consistency is a sufficient condition for performance within a constant factor of the best possible reconstruction algorithm.

Comparing cr^{-2} to the $O(r^{-1})$ expression (22) suggests that using a consistent reconstruction algorithm can greatly reduce the MSE when the redundancy of the frame is high. This is indeed true, but naturally many caveats have been omitted to permit this simple comparison.

3.3.2. Randomized Quantization. With subtractive dithered quantization, assumption (a′) of Subsection 3.1 holds. As in case of deterministic quantization, there are hard bounds on the quantization error:

$$\hat{y}_k - \frac{\Delta}{2} \le y_k < \hat{y}_k + \frac{\Delta}{2}, \qquad k = 1, 2, \ldots, M.$$

[5] $Q^{-1}(\hat{w})$ is an interval for any *regular* quantizer [15]. This interval will usually have finite extent. If it is infinite at one end, subsequent references to "two hyperplanes" should be taken as a single hyperplane. The interval will be infinite at both ends only in the degenerate case of a zero-bit quantizer, which we implicitly disallow.

Consistent reconstruction for this case is discussed in [23]. A simple recursive algorithm is given that attains optimal $O(r^{-2})$ MSE. This is again within a constant factor of the performance of an optimal algorithm.

3.3.3. Notes. A disappointing aspect of the aforementioned $O(r^{-2})$ MSE results for consistent reconstruction algorithms is that they do not indicate how to compute the constant factors in the MSE; thus, they provide no guidance on how to design the frame. Fortunately, (19) proves to be predictive of the performance of consistent reconstruction, with an additional multiplicative factor. This statement is made clear by Fig. 7.

Three MSE calculations were made for each of 500 random frames with $N = 4$ and $M = 32$: the MSE predicted by (19), the average MSE of linear reconstruction (15) for 1000 random source vectors, and the average MSE of a consistent reconstruction algorithm based on linear programming for the same random source vectors. A quantization step size of $\Delta = 1/10$ is used. The observed MSEs are plotted against the predicted MSE. For linear reconstruction, a line through the origin with unit slope fits the data very well; this reaffirms the MSE expression (19).

There are no theoretical results to indicate any relationship between the MSE (19) and the performance of consistent reconstruction. However, we find that a line through the origin with slope 0.607 provides a good fit to the data. While we have no analytical mechanism for determining the constant factor 0.607, we can infer that (19) provides a reasonable design criterion independent of the reconstruction method. For the remainder of the paper, we return to the noise model (14) and linear reconstruction (15).

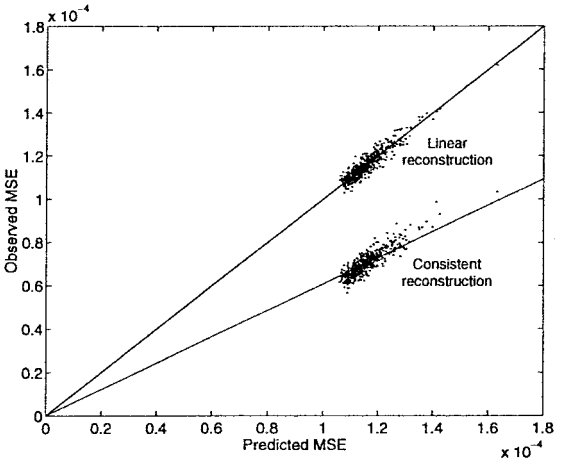

FIG. 7. Results of experiment to assess the predictive value of (19). For each of 500 random frames with $M = 32$ elements, 1000 random unit-norm source vectors in $\mathbb{R}^4$ were expanded with the frame and quantized with $\Delta = 1/10$. The MSE of linear reconstruction (15) and a consistent reconstruction computed with a linear program are compared to the predicted MSE (19). The solid lines are linear fits. The experiment indicates that (19) is a useful design criterion even if consistent reconstruction is used.

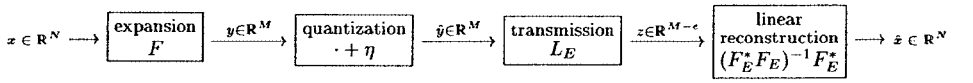

FIG. 8. The full system, as considered in Section 4. A signal expansion is computed with frame operator F. The expansion coefficients are quantized, which is modeled as the addition of a vector η satisfying (14). The deletion of some quantized expansion coefficients in the transmission is represented by operator L_E. An estimate of the original vector is computed using the linear MMSE estimator.

4. INTRODUCING ERASURES

We are now prepared to consider the overall system shown in Fig. 8. As in the previous section, the signal vector x is expanded with a quantized frame expansion to get $\hat{y}$. Now we introduce the transmission to the mix, which we abstract as the erasure of some components of $\hat{y}$. The received vector is denoted z. Quantization is indicated by the addition of the noise term η, which we assume satisfies (14). The signal is estimated using the linear reconstruction method of Subsection 3.2. With linear reconstruction, the MSE can be computed using the noise model without incorporating any information about the source.

Denote the index set of erasures by E; i.e., $\{\hat{y}_k\}_{k\in E}$ are lost. To the decoder it is as if a quantized frame expansion were computed with the frame $\Phi_E = \{\varphi_k\}_{k\notin E}$, assuming Φ_E is a frame. (To emphasize properties that depend only on the number of erasures $e = |E|$, the notation Φ_e is also used.) The effective frame operator is $F_E = L_E F$, where L_E captures the losses; L_E is the $(M - e) \times M$ matrix obtained by deleting the E-numbered rows from an $M \times M$ identity matrix.

For any particular set of erased components E, the results of Subsection 3.2 can be used to compute the MSE of the estimate $\hat{x}$—provided Φ_E is a frame. Thus the first order of business is to study conditions on Φ that maximize the number of erasures that can be withstood before Φ_E fails to span $\mathbb{R}^N$; this follows in Subsection 4.1. In Subsection 4.2 the effect of the erasures on the MSE is studied.

4.1. Effect of Erasures on the Structure of a Frame

To effectively reconstruct after e erasures using the techniques described in Subsection 3.2 or 3.3, it is necessary that Φ_E be a frame. If not, the dual frame linear reconstruction is not well-defined because the inverse in (6) does not exist. Moreover, hard bounds on the quantization error fail to give a bounded consistent set, so a consistent estimate may have very high error. [6]

After the deletion of several of the vectors, is what was originally a frame still a frame? Clearly the deletion of more than $M - N$ vectors leaves too few to span the space. This applies to all frames and gives no insight into the selection of a good frame.

With a bad choice of a frame, the situation can be even worse. For example, for any N and M one can construct a frame such that there is a particular deletion that leaves a set that is no longer a frame; all it takes is for one vector to be orthogonal to all of the rest. Thus we can be left with much more than N elements, but still not have a frame. We obviously

[6] Reconstruction when Φ_E is not a frame is discussed in [13].

want to avoid such situations to have robustness to as many combinations of erasures as possible.

Since uniform tight frames optimize robustness to quantization noise, as shown by Theorem 3.1, we might hope that they also have special properties with respect to erasures. Uniform tight frames do indeed have some special properties, but are not all created equal.

One erasure from a uniform tight frame with $M > N$ cannot destroy the property of being a frame, as shown by the following theorem:

THEOREM 4.1. *Let $\Phi = \{\varphi_k\}_{k=1}^M \subset \mathbb{R}^N$ be a uniform tight frame, with $M > N$. For any k, $\Phi_1 = \Phi_{\{k\}} = \Phi \backslash \varphi_k$ is a frame. Φ_1 has a lower frame bound $A_1 = M/N - 1$ and an upper frame bound $B_1 = M/N$.*

Proof. See Appendix A.5. ■

Extending the proof given in Appendix A.5, it is easy to check that e erasures can decrease the lower frame bound of a uniform frame by at most e. Thus, e erasures will leave a frame when $M/N > e$. This is a far cry from being able to guarantee that $M - N$ erasures leaves a basis for $\mathbb{R}^N$. In fact, even a uniform *tight* frame can fail to remain a frame after $M - N$ erasures. An example is given by

$$F = \begin{pmatrix} 0 & \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{3}} \\ 0 & -\sqrt{\frac{1}{3}} & \sqrt{\frac{2}{3}} \\ 0 & 1 & 0 \\ \sqrt{\frac{5}{6}} & 0 & \sqrt{\frac{1}{6}} \\ -\sqrt{\frac{5}{6}} & 0 & \sqrt{\frac{1}{6}} \end{pmatrix}.$$

One can verify that the rows of F (the frame elements) have unit length and that the columns of F are orthogonal. Deleting the last two rows of F leaves a rank-deficient matrix; thus, $\Phi_{\{4,5\}}$ is not a frame.

Though all uniform tight frames do not have the desirable property that Φ_{M-N} is a frame, there do exist such frames for any N and M. We have not found a useful parameterization of these frames, but we can demonstrate their existence by proving that harmonic frames suffice.

THEOREM 4.2. *Let $\Phi = \{\varphi_k\}_{k=1}^M$ be a complex harmonic frame in $\mathbb{C}^N$ given by* (8) *or a real harmonic frame in $\mathbb{R}^N$ given by* (9) *or* (10). *Then any subset of N or more vectors from Φ forms a frame, i.e., Φ_{M-N} is a frame.*

Proof. See Appendix A.6. ■

4.2. Effect of Erasures on the MSE

For the source-independent reconstruction techniques used in this paper, Φ_E must be a frame. Since, as argued in the previous section, it is possible to find uniform frames Φ such that Φ_E is a frame for any erasures of up to $M - N$ components, we assume such frames Φ for the remainder of the paper. We ultimately want to design frames that give good MSE performance, for which the first step is to compute the effect of erasures on the MSE.

Recall the quantization noise model (14) and linear reconstruction (16). The optimal reconstruction uses the dual $\tilde{\Phi}_E$ of the frame Φ_E, not the dual of the original frame Φ. Denote the MSE with erasure set E by MSE_E. Using the frame operator $F_E = L_E F$ associated with Φ_E, the MSE has been determined in (19),

$$\mathrm{MSE}_E = \frac{\sigma^2}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i(F_E^* F_E)}, \tag{23}$$

where $\{\lambda_i(F_E^* F_E)\}_{i=1}^N$ is the set of eigenvalues of $F_E^* F_E$. A useful equivalent form is

$$\mathrm{MSE}_E = N^{-1}\sigma^2 \operatorname{tr}((F_E^* F_E)^{-1}). \tag{24}$$

Can the MSE be expressed simply in terms of the original frame? Better yet: Are there expressions that depend only on the number of erasures $e = |E|$? Which frames minimize the average MSE_E over different erasure patterns E, and which minimize the worst-case MSE_E?

The best case scenario for MSE_E is clear: Φ_E is a uniform frame with $M - e$ elements so, by Theorem 3.1, the minimum MSE is $(M-e)^{-1}N\sigma^2$, attained if and only if Φ_E is tight. It is certainly possible for Φ_E to be tight, but for any original frame Φ, few erasure patterns will leave a frame Φ_E that is tight:

THEOREM 4.3. *Let Φ be a uniform frame in $\mathbb{R}^N$ with $N > 1$, and let $e \in \mathbb{Z}^+$. It is not possible for every Φ_E with $|E| = e$ to be tight.*

Proof. See Appendix A.7. ■

Since all the Φ_E's are not tight, the average of MSE_E (and, of course, the maximum) is greater than $(M-e)^{-1}N\sigma^2$.

We will address the problem of determining general expressions for MSE_E only for *tight* frames; this appears in Subsection 4.2.2. First, we consider the effect of a single erasure. As shown in Subsection 4.2.1, the average and worst-case MSE_1 are minimized if and only if a tight frame is used. Combined with the zero-erasure optimality of tight frames, this provides further justification for the later focus on tight frames.

4.2.1. MSE with One Erasure. Rather then starting immediately with the general case, we first compute the MSE when there is one erasure from a *tight* frame. The calculations are simple with one erasure, and the MSE has a rather remarkable property that it depends not on the particular tight frame nor on the position of the erasures—just on the size of the frame.

Let Φ be a uniform tight frame with $M > N$. Since the numbering of the frame elements is arbitrary, we can assume that the erased component is $\langle x, \varphi_1 \rangle$. Denote the frame operator associated with $\Phi_{\{1\}} = \Phi \backslash \varphi_1$ by F_1. By Theorem 4.1, we know that $\Phi_{\{1\}}$ is a frame.

The MSE can be determined from the trace of $(F_1^* F_1)^{-1}$, which has a simple form. Since $F^*F = \sum_{k=1}^M \varphi_k \varphi_k^* = (M/N) I_N$ and $F_1^* F_1 = \sum_{k=2}^M \varphi_k \varphi_k^*$,

$$F_1^* F_1 = \frac{M}{N} I_N - \varphi_1 \varphi_1^*. \tag{25}$$

The identity

$$(A - BCD)^{-1} = A^{-1} + A^{-1}B(C^{-1} - DA^{-1}B)^{-1}DA^{-1} \tag{26}$$

yields

$$(F_1^* F_1)^{-1} = \left(\frac{M}{N} I - \varphi_1 \varphi_1^*\right)^{-1} = \frac{N}{M} I + \frac{N^2}{M(M-N)} \varphi_1 \varphi_1^*.$$

The trace is now easy to compute using linearity:

$$\begin{aligned} \operatorname{tr}(F_1^* F_1)^{-1} &= \operatorname{tr}\left(\frac{N}{M} I + \frac{N^2}{M(M-N)} \varphi_1 \varphi_1^*\right) \\ &= \frac{N}{M} \underbrace{\operatorname{tr}(I_N)}_{N} + \frac{N^2}{M(M-N)} \underbrace{\operatorname{tr}(\varphi_1 \varphi_1^*)}_{1} = \frac{N^2}{M}\left(1 + \frac{1}{M-N}\right). \end{aligned}$$

Substituting in (24) and comparing to (22) gives

$$\mathrm{MSE}_1 = \left(1 + \frac{1}{M-N}\right)\frac{N}{M}\sigma^2 = \left(1 + \frac{1}{M-N}\right)\mathrm{MSE}_0. \tag{27}$$

This result is remarkably simple, both in its form and in its independence from φ_1.

Note that deleting one element from a uniform tight frame fails to leave a frame if and only if the original frame is a basis, i.e., the redundancy ratio is one. In this case $M = N$ and (27) breaks down, but this should be expected because the analysis using the dual frame does not apply.

The MSE (27) obtained when the original frame is tight is both average-case and minimax optimal, formalized by the following theorem:

THEOREM 4.4. *Consider encoding with a uniform frame and decoding with linear reconstruction* (16)*, under noise model* (14)*. The MSE averaged over all possible erasures of one frame element,*

$$\overline{\mathrm{MSE}}_1 = \frac{1}{M}\sum_{k=1}^{M} \mathrm{MSE}_{\{k\}},$$

is minimum if and only if the original frame is tight. Also, a tight frame minimizes the maximum distortion caused by one erasure

$$\max_{k=1,2,\ldots,M} \mathrm{MSE}_{\{k\}}.$$

Proof. See Appendix A.8. ■

EXAMPLE 4.1. Consider the uniform frames $\Phi = \{\varphi_k\}_{k=1}^3$ and $\Psi = \{\psi_k\}_{k=1}^3$ in $\mathbb{R}^2$ given by

$$\varphi_1 = \psi_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \varphi_2 = \begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}, \qquad \varphi_3 = \begin{bmatrix} -1/2 \\ -\sqrt{3}/2 \end{bmatrix},$$

$$\psi_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \psi_3 = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}.$$

Φ is the tight frame from Example 1.1 and Ψ is not tight.

Each single erasure from Φ gives a nontight frame with $\lambda(F_1^* F_1) = \{\frac{3}{2}, \frac{1}{2}\}$. The sum of the reciprocals of these eigenvalues is $\frac{8}{3}$.

The erasure of ψ_3 from Ψ leaves a tight frame with $\lambda(F_{\{3\}}^* F_{\{3\}}) = \{1, 1\}$. The sum of reciprocals is 2, which is better than what was obtained for Φ. However, the erasure of ψ_1 or ψ_2 is much worse. One can compute $\lambda(F_{\{1\}}^* F_{\{1\}}) = \lambda(F_{\{2\}}^* F_{\{2\}}) = \{1 + \sqrt{\frac{1}{2}}, 1 - \sqrt{\frac{1}{2}}\}$, so the sum of reciprocals is 4. One of the single-element erasure events is better for Ψ, but Ψ is worse in the average- and worst-case.

4.2.2. MSE with e Erasures. It is not possible to extend Theorem 4.4 to more than one erasure; e.g., all tight frames do not minimize the two-erasure MSE. However, because of the importance of tight frames in the zero- and one-erasure cases, we limit our attention to tight frames for the remainder of the paper. We now turn to the computation of the MSE when a tight frame is subject to an arbitrary number of erasures.

Consider that erasures have occurred at positions in the index set E. Assume that Φ_E is a frame and denote the associated frame operator by F_E. The MSE is now proportional to the trace of $(F_E^* F_E)^{-1}$. The matrix inversion is more difficult to compute because instead of having a rank-1 perturbation to a scaled identity as in (25), we have

$$F_E^* F_E = \frac{M}{N} I_N - \varphi \varphi^*,$$

where φ is an $N \times e$ matrix with columns $\{\varphi_k\}_{k \in E}$.

Using (26) we can write

$$\begin{aligned}(F_E^* F_E)^{-1} &= \frac{N}{M} I_N + \frac{N}{M} I_N \varphi \left(I_e - \varphi^* \frac{N}{M} I_N \varphi \right)^{-1} \varphi^* \frac{N}{M} I_N \\ &= \frac{N}{M} I_N + \frac{N^2}{M^2} \varphi \left(I_e - \frac{N}{M} \varphi^* \varphi \right)^{-1} \varphi^*. \end{aligned} \tag{28}$$

This form is simpler because it involves the inverse of an $e \times e$ matrix instead of an $N \times N$ matrix. (With one erasure—as in Subsection 4.2.1—we are left with only a scalar to invert, so the ensuing analysis is very easy.) Employing a series expansion of the matrix inversion in (28),

$$\left(I_e - \frac{N}{M} \varphi^* \varphi \right)^{-1} = \sum_{k=0}^{\infty} \left(\frac{N}{M} \varphi^* \varphi \right)^k,$$

leads to the calculation

$$\begin{aligned} \operatorname{tr}\left[\varphi \left(I_e - \frac{N}{M} \varphi^* \varphi \right)^{-1} \varphi^* \right] &= \operatorname{tr}\left[\left(I_e - \frac{N}{M} \varphi^* \varphi \right)^{-1} \varphi^* \varphi \right] \\ &= \operatorname{tr}\left[\sum_{k=0}^{\infty} \left(\frac{N}{M} \varphi^* \varphi \right)^k \varphi^* \varphi \right] \\ &= \sum_{k=0}^{\infty} \left(\frac{N}{M} \right)^k \operatorname{tr}\left((\varphi^* \varphi)^{k+1} \right). \end{aligned} \tag{29}$$

Substituting (28) and (29) in (24) yields

$$\begin{aligned}\mathrm{MSE}_E &= \frac{\sigma^2}{N}\left(\frac{N^2}{M} + \frac{N^2}{M^2}\sum_{k=0}^{\infty}\left(\frac{N}{M}\right)^k \operatorname{tr}\left((\varphi^*\varphi)^{k+1}\right)\right)\\ &= \frac{N\sigma^2}{M}\left(1 + \frac{1}{M}\sum_{k=0}^{\infty}\left(\frac{N}{M}\right)^k \operatorname{tr}\left((\varphi^*\varphi)^{k+1}\right)\right)\\ &= \left(1 + \frac{1}{M}\sum_{k=0}^{\infty}\left(\frac{N}{M}\right)^k \operatorname{tr}\left((\varphi^*\varphi)^{k+1}\right)\right)\mathrm{MSE}_0. \end{aligned} \tag{30}$$

Note that with one erasure $\varphi^*\varphi = 1$; the series in (30) is geometric and the MSE simplifies to (27). When there are more erasures, (30) simplifies not to a geometric series, but to a sum of geometric series.

Denote the eigenvalues of $\varphi^*\varphi$ by $\{\mu_i\}_{i=1}^e$. Then $\operatorname{tr}(\varphi^*\varphi)^{k+1} = \sum_{i=1}^e \mu_i^{k+1}$ (see [17, p. 43]). Thus, the series in (30) becomes a sum of geometric series: [7]

$$\begin{aligned}\sum_{k=0}^{\infty}\left(\frac{N}{M}\right)^k \operatorname{tr}\left((\varphi^*\varphi)^{k+1}\right) &= \sum_{k=0}^{\infty}\left(\frac{N}{M}\right)^k \sum_{i=1}^{e}\mu_i^{k+1} = \sum_{i=1}^{e}\mu_i \sum_{k=0}^{\infty}\left(\frac{N}{M}\mu_i\right)^k\\ &= \sum_{i=1}^{e}\frac{\mu_i}{1-(N/M)\mu_i}. \end{aligned} \tag{31}$$

Substituting (31) in (30) gives

$$\mathrm{MSE}_E = \left(1 + \sum_{i=1}^{e}\frac{\mu_i}{M - N\mu_i}\right)\mathrm{MSE}_0. \tag{32}$$

4.3. Frame Designs Issues

We would like to use the results of Subsection 4.2 to deduce optimal frame designs. We have already found that any tight frame is optimal for zero or one erasure (see Theorems 3.1 and 4.4). For more than one erasure, computations are significantly simplified if the original frame is tight, so we have limited attention to this case.

How can we minimize the distortion with e erasures from a tight frame, as given in (32)? The expression (32) is similar to the zero-erasure expression (19) in that the sum can be written as $\sum_{i=1}^e f(\mu_i)$ where $f(\cdot)$ is convex and $\sum_{i=1}^e \mu_i$ is constrained to a constant. [8] In this case, $\sum_{i=1}^e \mu_i = \operatorname{tr}(\varphi^*\varphi) = \operatorname{tr}(\varphi\varphi^*) = e$. The minimum of (32) is obtained when each μ_i is equal to 1—provided this is feasible.

If $e \le N$, it is indeed possible to have $\mu_i = 1$, $i = 1, 2, \ldots, e$. This occurs if and only if the erased vectors are pairwise orthogonal. Then $\varphi^*\varphi = I_e$ and (32) gives

$$\mathrm{MSE}_{e \text{ orthogonal erasures}} = \left(1 + \frac{e}{M-N}\right)\mathrm{MSE}_0.$$

[7] Whenever Φ_E is a frame, the common ratio $|(N/M)\mu_i|$ is less than 1 and the last equality holds; a proof is given in Appendix A.9.

[8] See also the proof of Theorem 4.4 in Appendix A.8 for another analogous minimization.

If $e > N$, it is not possible to have e eigenvalues equal to 1 because there will be at most N nonzero eigenvalues. Denoting the nonzero eigenvalues $\{\mu_i\}_{i=1}^{N}$,

$$\mathrm{MSE}_E = \left(1 + \sum_{i=1}^{N} \frac{\mu_i}{M - N\mu_i}\right) \mathrm{MSE}_0.$$

This MSE is minimized when $\mu_i = e/N$, $i = 1, 2, \ldots, N$, which occurs when the *erased elements* form a tight frame. [9]

When arbitrary erasure events (subsets $E \subset \{1, 2, \ldots, M\}$) are allowed, it is not possible to design a frame that will always achieve the minimum MSE (see Theorem 4.3). However, the packetization may be designed so that all erasure patterns are not possible, in which case optimal performance (minimum MSE distortion given the number of received coefficients) is possible. Specifically, if each packet contains coefficients corresponding to a tight frame, then—since the union of tight frames is a tight frame—any set of received packets gives the minimum MSE possible for that number of received coefficients. This solution requires each packet to carry at least N coefficients.

EXAMPLE 4.2. Suppose we wish to design a uniform frame with $N = 3$ and $M = 7$ for the situation in which $\{\hat{y}_k\}_{k=1}^{3}$ is sent in Packet 1 and $\{\hat{y}_k\}_{k=4}^{7}$ is sent in Packet 2. Because each packet carries at least N components, the design is very easy. One can choose $\{\varphi_k\}_{k=1}^{3}$ to be *any* orthonormal basis and $\{\varphi_k\}_{k=4}^{7}$ to be *any* uniform tight frame. Then whether Packet 1, Packet 2, or both are received, the effective frame is tight and the MSE is as low as possible for the number of received components.

EXAMPLE 4.3. Again with $N = 3$ and $M = 7$, suppose the packetization has $\{\hat{y}_k\}_{k=1}^{3}$, $\{\hat{y}_k\}_{k=4}^{5}$, and $\{\hat{y}_k\}_{k=6}^{7}$ in Packets 1, 2, and 3, respectively. The theory developed here indicates that $\{\varphi_k\}_{k=1}^{3}$ should be an orthonormal basis and that we should have $\varphi_4 \perp \varphi_5$ and $\varphi_6 \perp \varphi_7$. The remaining degrees of freedom affect the performance, but can only be resolved with a numerical optimization. The result of the optimization depends on the weights (relative importance) assigned to each possible combination of received packets.

Theorem 2.10 indicates that asymptotically as $M \to \infty$, optimal performance is possible for any small erasure event, independent of the packetization. Suppose a frame is generated with M unit vectors selected independently according to a uniform distribution, with M large. If after e erasures $M - e$ is large, the remaining frame is approximately tight, so the MSE is nearly minimum.

5. CONCLUDING COMMENTS

This paper has demonstrated a new application for frame expansions: providing robustness to losses in packet-based network communication. Only finite-dimensional frames have been considered. This is consistent with the intended application, where each packet carries a finite amount of data, as outlined in Example 1.1. However, when the

[9] If the original frame and the erased elements are both tight frames, the remaining elements are also a tight frame. This is the optimality condition mentioned at the beginning of Subsection 4.2.

dimension of a data set is large (but still finite), it is a practical necessity to use structured signal expansions.

As noted earlier, real and complex harmonic frames are available for any size and dimension. Expansions with respect to these frames can be computed efficiently with FFT-like algorithms. Just as discrete wavelet transforms (DWTs) asymptotically require less operations than discrete Fourier transforms, oversampled filter banks provide efficient implementations of frame expansions [29]. More importantly, one can expect that these expansions would have advantages over Fourier techniques for many types of practical signals [8]. In particular, wavelet and wavelet-like bases have proven very effective in image compression.

Given the machinery of the DWT and its implementation through iterated filter banks, a simple way to obtain a frame expansion is to remove all the downsampling. However, such undecimated DWTs have large redundancies; an N sample vector is expanded to $N \log L$ samples for a depth-L tree. Design techniques for wavelet frames with lower redundancy, especially redundancy 2, have received recent attention [9, 18, 24]. The applicability of these techniques to multiple description coding is studied in [9]. Numerical optimization of finite-dimensional frames for multiple description coding was considered in [21]. We continue to investigate the analytical design of good frames for small redundancies M/N.

APPENDIX

A.1. Proof of Theorem 2.6

Let F be the operator associated with a uniform tight frame, with $M = N + 1$. We will show that the tightness condition and normalization make F essentially unique, i.e., unique up to the equivalence relation described in the text.

By Property 2.5, F consists of the first N columns of a scaled $M \times M$ orthogonal matrix $\overline{F}$. The normalization of each row of $\overline{F}$ to $\sqrt{M/N}$ implies that

$$\sum_{j=1}^{N+1} \overline{F}_{ij}^2 = \frac{N+1}{N}, \qquad \text{for } i = 1, 2, \ldots, N+1. \tag{A.1}$$

Furthermore, since our tight frame is normalized so that $\|\varphi_k\| = 1$, $k = 1, \ldots, N+1$, we have that

$$\sum_{j=1}^{N} \overline{F}_{ij}^2 = 1, \qquad \text{for } i = 1, 2, \ldots, N+1. \tag{A.2}$$

Subtracting (A.2) from (A.1) gives

$$\overline{F}_{i,N+1}^2 = \frac{1}{N}, \qquad \text{for } i = 1, 2, \ldots, N+1, \tag{A.3}$$

that is, the last column of $\overline{F}$ is $(\pm N^{-1/2}, \pm N^{-1/2}, \ldots, \pm N^{-1/2})$ for some choice of signs. From this it follows that the span of the first N columns is the orthogonal complement of the vector $\sigma = (\pm 1, \pm 1, \ldots, \pm 1)$ for some choice of signs.

Any given choice of signs in σ determines an N-dimensional subspace. Since *orthonormal bases* for a subspace are unitarily equivalent, the possible tight frames corresponding to a single choice of σ are in the same equivalence class. Flipping a sign in σ reflects the subspace, and hence also yields tight frames in the same equivalence class. Thus, we have shown that all uniform tight frames belong to the same equivalence class. The existence of harmonic frames establishes that the class is nonempty.

A.2. Proof of Theorem 2.7

Form the frame operator matrix

$$F = \begin{bmatrix} \cos\alpha_1 & \sin\alpha_1 \\ \cos\alpha_2 & \sin\alpha_2 \\ \vdots & \vdots \\ \cos\alpha_M & \sin\alpha_M \end{bmatrix}. \tag{A.4}$$

For the frame to be tight is to have $F^*F = (M/2)I_2$, which leads to

$$\sum_{k=1}^{M} (\cos\alpha_k)^2 = \frac{M}{2}, \tag{A.5}$$

$$\sum_{k=1}^{M} (\sin\alpha_k)^2 = \frac{M}{2}, \tag{A.6}$$

$$\sum_{k=1}^{M} \cos\alpha_k \sin\alpha_k = 0. \tag{A.7}$$

Subtracting (A.6) from (A.5) gives

$$\sum_{k=1}^{M} \cos 2\alpha_k = 0, \tag{A.8}$$

while multiplying (A.7) by 2 yields

$$\sum_{k=1}^{M} \sin 2\alpha_k = 0. \tag{A.9}$$

Finally, adding (A.8) to j times (A.9) gives

$$\sum_{k=1}^{M} z_k = 0, \qquad \text{where } z_k = e^{j2\alpha_k}. \tag{A.10}$$

A.3. Proof of Theorem 2.8

Suppose the hypothesis is not true, that is, there are $(K+1)$ elements which are equal, $\varphi_1 = \varphi_2 = \cdots = \varphi_{K+1}$. For any frame

$$F^*F = \sum_{k=1}^{M} \varphi_k \varphi_k^*, \tag{A.11}$$

while for a uniform tight frame

$$F^*F = AI_N = (M/N)I_N. \tag{A.12}$$

Thus, using (A.12),

$$\varphi_1^* F^* F \varphi_1 = A\varphi_1^* \varphi_1 = A\|\varphi_1\|^2. \tag{A.13}$$

On the other hand, using (A.11),

$$\varphi_1^* F^* F \varphi_1 = \varphi_1^* \left(\sum_{i=1}^{M} \varphi_i \varphi_i^* \right) \varphi_1 = (K+1)\|\varphi_i\|^4 + \sum_{i=K+2}^{M} |\varphi_1^* \varphi_i|^2 \tag{A.14}$$

$$\geq (K+1)\|\varphi_1\|^4. \tag{A.15}$$

Equating (A.13) and (A.15) implies that

$$A\|\varphi_1\|^2 - (K+1)\|\varphi_1\|^4 \geq 0.$$

Since $A = M/N$ and $\|\varphi_1\| = 1$, this implies

$$\frac{M}{N} \geq \left\lfloor \frac{M}{N} \right\rfloor + 1, \tag{A.16}$$

which is a contradiction.

Assuming now that K elements are equal and $K = M/N$, repeat the above derivation of (A.14) with K replacing $(K+1)$. This leads to

$$\frac{M}{N}\|\varphi_1\|^4 + \sum_{i=K+1}^{M} |\varphi_1^* \varphi_i|^2 = \frac{M}{M}\|\varphi_1\|^2,$$

which in turn yields

$$\sum_{i=K+1}^{M} |\varphi_1^* \varphi_i|^2 = 0.$$

This means that φ_1 (and thus $\varphi_2, \ldots, \varphi_K$) is orthogonal to the span of $\{\varphi_j\}_{j=K+1}^{M}$.

A.4. Proof of Theorem 3.2

The calculations yielding (19) are given in the text. The bound (20) is a consequence of Property 2.1.

For a uniform frame, Theorem 3.1 implies that the MSE cannot be lower than that achieved by a uniform tight frame; this yields the lower bound of (21).

For a uniform tight frame, $A = B = M/N$; thus, (20) simplifies to (22). Alternatively, one can note that by Property 2.3, every $\lambda_i = M/N$, $i = 1, \ldots, N$. Substituting in (19) gives the desired result.

It is shown in [14] that, for any uniform frame,

$$\frac{M\sigma^2}{NB^2} \leq \text{MSE} \leq \frac{M\sigma^2}{NA^2}.$$

But since $A \leq M/N \leq B$ for any uniform frame, this bound is weaker than (20).

A.5. Proof of Theorem 4.1

For a tight frame, we know that

$$\sum_{k=1}^{M} |\langle x, \varphi_k \rangle|^2 = A \qquad \text{for all } x \text{ such that } \|x\| = 1.$$

Suppose φ_i is deleted from the frame, and thus $|\langle x, \varphi_i \rangle|^2$ is subtracted from the sum. Then since $0 \leq |\langle x, \varphi_i \rangle| \leq 1$,

$$A - 1 \leq \sum_{\substack{k=1 \\ k \neq i}}^{M} |\langle x, \varphi_k \rangle|^2 \leq A \qquad \text{for all } x \text{ such that } \|x\| = 1. \tag{A.17}$$

For a uniform tight frame with $M > N$, $A = M/N > 1$. Therefore, (A.17) shows that Φ_1 is a frame with frame bounds $A - 1$ and A.

A.6. Proof of Theorem 4.2

First note that if a finite set of vectors has a subset that is a frame, then the original set is also a frame. Thus it suffices to consider subsets with N vectors; since all of these will be shown to be frames, larger subsets are also frames.

Consider first the complex harmonic frame given by (8). Pick an arbitrary subset $\{k_1, k_2, \ldots, k_N\}$ of $\{1, 2, \ldots, M\}$ and denote the operator associated with this subset by $F_{M,N}$. By inspection of (8),

$$F_{M,N} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & W_M^{k_1 - 1} & \cdots & W_M^{(N-1)(k_1-1)} \\ 1 & W_M^{k_2 - 1} & \cdots & W_M^{(N-1)(k_2-1)} \\ \vdots & \vdots & & \vdots \\ 1 & W_M^{k_N - 1} & \cdots & W_M^{(N-1)(k_N-1)} \end{bmatrix}.$$

We can recognize $F_{M,N}$ as a scaled Vandermonde matrix with determinant (see [17, p. 29])

$$\det F_{M,N} = N^{-N/2} \prod_{\substack{i,j=1 \\ i>j}}^{N} (W_M^{k_i - 1} - W_M^{k_j - 1}).$$

The above determinant is nonzero *if and only if* all the values $W_M^{k_i}$ are distinct. Since these are distinct roots of unity, $\det F_{M,N}$ is nonzero, and the chosen set of N vectors is a frame.

The real case is similar, but messier; we consider here only the case of N even and frames given by (9). Pick any subset $\{k_1, k_2, \ldots, k_N\}$ of $\{0, 1, \ldots, M-1\}$ and define F by choosing $\varphi_{k_1+1}^*, \varphi_{k_2+1}^*, \ldots, \varphi_{k_N+1}^*$ as the rows. We would like to show $\det F \neq 0$, which will be clear after a sequence of elementary operations on F. Rather than writing the entire matrix F, the operations will be demonstrated with the jth row of F. The scale factor $\sqrt{2/N}$ is omitted.

First note that, using Euler's formula, the jth row is

$$\left[\frac{1}{2}(W_{2M}^{k_j} + W_{2M}^{-k_j}),\ \frac{1}{2}(W_{2M}^{3k_j} + W_{2M}^{-3k_j}),\ \ldots,\ \frac{1}{2}(W_{2M}^{(N-1)k_j} + W_{2M}^{-(N-1)k_j}),\right.$$
$$\left.\frac{1}{2i}(W_{2M}^{k_j} - W_{2M}^{-k_j}),\ \frac{1}{2i}(W_{2M}^{3k_j} - W_{2M}^{-3k_j}),\ \ldots,\ \frac{1}{2i}(W_{2M}^{(N-1)k_j} - W_{2M}^{-(N-1)k_j})\right].$$

Multiply the last $N/2$ columns by i; add the last $N/2$ columns to the corresponding first $N/2$ columns; add $-\frac{1}{2}$ times the first $N/2$ columns to the last $N/2$ columns; and multiply the last $N/2$ columns by -2; the resulting jth row is

$$\left[W_{2M}^{k_j},\ W_{2M}^{3k_j},\ \ldots,\ W_{2M}^{(N-1)k_j},\ W_{2M}^{-k_j},\ W_{2M}^{-3k_j},\ldots,\ W_{2M}^{-(N-1)k_j}\right].$$

Now factoring $W_{2M}^{-(N-1)k_j}$ from row j gives

$$\left[W_{2M}^{Nk_j},\ W_{2M}^{(N+2)k_j},\ \ldots,\ W_{2M}^{2(N-1)k_j},\ W_{2M}^{(N-2)k_j},\ W_{2M}^{(N-4)k_j},\ \ldots,\ W_{2M}^{0}\right].$$

After reordering columns—with $(N/2) + \lfloor N/4 \rfloor$ exchanges—we have a Vandermonde matrix. Accounting for the scaling and elementary operations,

$$\begin{aligned}\det F &= (-1)^{((N/2)+\lfloor N/4 \rfloor)}\left(\frac{2}{N}\right)^{N/2} i^{-N/2}(-2)^{-N/2} \\ &\quad\times \prod_{\ell=1}^{N} W_{2M}^{-(N-1)k_\ell} \prod_{\substack{i,j=1 \\ i>j}}^{N} (W_{2M}^{2k_i} - W_{2M}^{2k_j}) \\ &= (-1)^{((N/2)+\lfloor N/4 \rfloor)} N^{-N/2} i^{N/2} W_{2M}^{-(N-1)\sum_\ell k_\ell} \prod_{\substack{i,j=1 \\ i>j}}^{N} (W_{2M}^{2k_i} - W_{2M}^{2k_j}).\end{aligned}$$

This determinant is nonzero, so the selected vectors form a frame.

A.7. Proof of Theorem 4.3

Consider first the case $e = 1$. Denote the frame operator associated with $\Phi\backslash\{\varphi_k\}$ by F_k. By hypothesis, $\Phi\backslash\{\varphi_k\}$ is tight for any k, so

$$F_k^* F_k = N^{-1}(M-1) I_N, \qquad k = 1, 2, \ldots, M.$$

These matrices are also related by $F_k^* F_k = F^* F - \varphi_k \varphi_k^*$. Therefore, picking any pair of indices i and j and subtracting gives

$$0 = F_i^* F_i - F_j^* F_j = \varphi_j \varphi_j^* - \varphi_i \varphi_i^*. \tag{A.18}$$

This implies φ_i and φ_j are collinear, and by extension that all the frame elements are collinear. This contradicts the assumption that Φ is a frame. Therefore it is not possible for every single erasure from a uniform frame to leave a tight frame.

For the case of general e, Eq. (A.18) is obtained by considering any pair of erasure events that differ only in that one includes φ_j in place of φ_i. Thus the same contradiction as before is obtained.

A.8. Proof of Theorem 4.4

Let Φ be a uniform frame. Consider first the optimization of Φ for average-case MSE; the minimax optimization will follow easily.

We need to consider simultaneously each of the M possible erasures of one element of Φ. Define

$$H_i = \sum_{\substack{k=1 \\ k \neq i}}^{M} \varphi_k \varphi_k^* = F^* F - \varphi_i \varphi_i^*.$$

Since $\mathrm{MSE}_{\{i\}} = N^{-1} \sigma^2 \operatorname{tr}(H_i^{-1})$, the average MSE with one erasure is

$$\overline{\mathrm{MSE}}_1 = \frac{1}{M} \sum_{i=1}^{M} \frac{\sigma^2}{N} \operatorname{tr}\big(H_i^{-1}\big). \tag{A.19}$$

Using (26),

$$H_i^{-1} = (F^* F)^{-1} + (F^* F)^{-1} \varphi_i l [1 - \varphi_i^* (F^* F)^{-1} \varphi_i]^{-1} \varphi_i^* (F^* F)^{-1}.$$

Now noting that $[1 - \varphi_i^* (F^* F)^{-1} \varphi_i]$ is a scalar and using the invariance of a trace of a product to the cyclic permutation of factors,

$$\begin{aligned} \operatorname{tr}(H_i^{-1}) &= \operatorname{tr}\big((F^* F)^{-1}\big) + [1 - \varphi_i^* (F^* F)^{-1} \varphi_i]^{-1} \operatorname{tr}\big(\varphi_i^* (F^* F)^{-2} \varphi_i\big) \\ &= \operatorname{tr}\big((F^* F)^{-1}\big) + \frac{\varphi_i^* (F^* F)^{-2} \varphi_i}{1 - \varphi_i^* (F^* F)^{-1} \varphi_i}. \end{aligned}$$

Substituting in (A.19) gives

$$\begin{aligned} \overline{\mathrm{MSE}}_1 &= \frac{\sigma^2}{MN} \sum_{i=1}^{M} \left(\operatorname{tr}\big((F^* F)^{-1}\big) \frac{\varphi_i^* (F^* F)^{-2} \varphi_i}{1 - \varphi_i^* (F^* F)^{-1} \varphi_i} \right) \\ &= \frac{\sigma^2}{N} \operatorname{tr}\big((F^* F)^{-1}\big) + \frac{\sigma^2}{MN} \sum_{i=1}^{M} \frac{\varphi_i^* (F^* F)^{-2} \varphi_i}{1 - \varphi_i^* (F^* F)^{-1} \varphi_i}. \end{aligned} \tag{A.20}$$

We know from Theorem 3.1 that the first term of (A.20) is minimized if and only if Φ is tight; it thus suffices to consider the minimization of the second term.

We can use a technique similar to that used in the proof of Theorem 3.1. Recall that in the earlier proof we had a constraint $\sum_{i=1}^{N} \lambda_i = N$ and we wanted to minimize $\sum_{i=1}^{N} f(\lambda_i)$, where $f(z) = z^{-1}$. Since $f(\cdot)$ is a convex function, the minimum occurs when each term contributes equally, provided this is feasible—which is in fact the case. In the present proof, $v_i = \varphi_i^*(F^*F)^{-1}\varphi_i$ plays the role of λ_i.

First note that the sum of v_i's is constrained:

$$\sum_{i=1}^{M} v_i = \sum_{i=1}^{M} \varphi_i^*(F^*F)^{-1}\varphi_i = \sum_{i=1}^{M} \operatorname{tr}\left(\varphi_i^*(F^*F)^{-1}\varphi_i\right) = \sum_{i=1}^{M} \operatorname{tr}\left((F^*F)^{-1}\varphi_i\varphi_i^*\right)$$

$$= \operatorname{tr}\left(\sum_{i=1}^{M}(F^*F)^{-1}\varphi_i\varphi_i^*\right) = \operatorname{tr}\left((F^*F)^{-1}\underbrace{\sum_{i=1}^{M}\varphi_i\varphi_i^*}_{F^*F}\right) = \operatorname{tr}(I_N) = N.$$

The remainder of the proof relies on the following simple lemma:

LEMMA A.1. *Let M be a square matrix and let w be a compatibly dimensioned unit vector. Then $w^*M^*Mw \geq (w^*Mw)^2$, with equality if and only if w is an eigenvector of M.*

Proof. The matrix $I - ww^*$ is positive semidefinite. Thus $(Mw)^*(I - ww^*)(Mw) \geq 0$. Expanding the left-hand side and rearranging gives $w^*M^*Mw \geq (w^*Mw)^2$. Furthermore, w is an eigenvector of $I - ww^*$ corresponding to eigenvalue 0. The remaining eigenvalues are all 1. Therefore equality holds if and only if Mw is parallel to w, or w is an eigenvector of M. ■

Now we wish to express the summation in (A.20) in terms of the v_i's. Applying Lemma A.1 gives

$$\varphi_i^*(F^*F)^{-2}\varphi_i \geq \left(\varphi_i^*(F^*F)^{-1}\varphi_i\right)^2 = v_i^2.$$

Thus we can bound the critical term of (A.20) as

$$\sum_{i=1}^{M} \frac{\varphi_i^*(F^*F)^{-2}\varphi_i}{1 - \varphi_i^*(F^*F)^{-1}\varphi_i} \geq \sum_{i=1}^{M} \frac{v_i^2}{1 - v_i}. \tag{A.21}$$

Since $z^2/(1-z)$ is a convex function, the right-hand side of (A.21) is minimized when each term contributes equally, provided this is feasible; i.e., $v_i = N/M$, $i = 1, 2, \ldots, M$.

Making each v_i equal N/M is indeed feasible; it is easy to verify that it occurs whenever Φ is a tight frame. At the same time, equality holds in (A.21), so the average MSE is minimized and tightness is *sufficient* for optimally of the frame. To complete the proof, we will show that tightness is *necessary* for optimality. Specifically, $v_i = N/M$, $i = 1, 2, \ldots, M$, and equality in (A.21) together imply that Φ is tight.

Equality in (A.21) implies that φ_i is an eigenvector of $(F^*F)^{-1}$. Denote the corresponding eigenvalue by ν_i. The eigenvalue–eigenvector property gives $\varphi_i^*(F^*F)^{-1}\varphi_i = \nu_i$, which means ν_i and v_i are identical! Since the φ_i's span $\mathbb{R}^N$, all the eigenvalues are obtained in this manner, and since all the eigenvalues are equal, Φ is tight.

The minimax optimality is clear because the average-case MSE is minimized while keeping every term in (A.19) equal. Obviously, the maximum term of (A.19) cannot be smaller than the mean.

A.9. Notes on the Convergence of (31)

In the final equality of (31), we have used the formula for the sum of a convergent geometric series. The geometric series is convergent if and only if $|(N/M)\mu_i| < 1$. We establish here that this inequality holds whenever Φ_E is a frame. In particular, we show that $|(N/M)\mu_i| \leq 1$, with equality if and only if Φ_E fails to be a frame.

Recall that the eigenvalues of $\varphi^*\varphi$ are denoted $\{\mu_i\}_{i=1}^{e}$ and note that the nonzero eigenvalues of $\varphi^*\varphi$ equal the nonzero eigenvalues of $\varphi\varphi^*$. The matrices F^*F and $\varphi\varphi^*$ are closely related; each can be written as a sum of outer products of frame elements, but F^*F contains more terms:

$$F^*F = \sum_{k=1}^{M} \varphi_k \varphi_k^* \qquad \text{and} \qquad \varphi\varphi^* = \sum_{k=1}^{e} \varphi_k \varphi_k^*.$$

If v is the normalized eigenvector of $\varphi\varphi^*$ associated with eigenvalue μ_i, then

$$\begin{aligned} v^*F^*Fv &= v^*\left(\sum_{k=1}^{M} \varphi_k\varphi_k^*\right)v = v^*\varphi\varphi^*v + v^*\left(\sum_{k=e+1}^{M} \varphi_k\varphi_k^*\right)v \\ &= \mu_i + \sum_{k=e+1}^{M} |v^*\varphi_k|^2 \geq \mu_i. \end{aligned} \tag{A.22}$$

Since v^*F^*Fv is bounded from above by the largest eigenvalue of F^*F, we have $\mu_i \leq M/N$.

Equality holds in (A.22) if and only if the eigenvector v is orthogonal to $\{\varphi_k\}_{k=e+1}^{M}$, i.e., all of the vectors remaining in Φ_E. If v is orthogonal to all of the vectors in Φ_E, then Φ_E does not span $\mathbb{R}^N$ and is not a frame. The analysis of Subsection 4.2 is not intended to apply to this case.

ACKNOWLEDGMENT

We thank Peter Casazza and Janet Tremain for suggesting the current version of Theorem 2.8, which allows K to be arbitrary. Also, they independently obtained Theorem 2.7; we have formalized it as per their suggestion. The proof of Theorem 4.4 is in large part due to James Mazo. We thank Ingrid Daubechies and Radu Balan for fruitful discussions. Thanks also to Pier Luigi Dragotti and anonymous reviewers for their comments.

REFERENCES

1. J. J. Benedetto and D. Colella, Wavelet analysis of spectrogram seizure chirps, *in* "Proc. SPIE Wavelet Appl. in Signal and Image Proc. III," Vol. 2569, pp. 512–521, San Diego, CA, July 1995.
2. J. J. Benedetto and G. E. Pfander, Wavelet periodicity detection algorithms, *in* "Proc. SPIE Wavelet Appl. in Signal and Image Proc. VI," Vol. 3459, pp. 48–55, San Diego, CA, July 1998.
3. Z. Cvetković, Source coding with quantized redundant expansions: Accuracy and reconstruction, *in* "Proc. IEEE Data Compression Conf., Snowbird, UT" (J. A. Storer and M. Cohn, Eds.), pp. 344–353, IEEE Comput. Soc., Los Alamitos, CA, 1999.
4. Z. Cvetković and M. Vetterli, Discrete-time wavelet extrema representation: Design and consistent reconstruction, *IEEE Trans. Signal Process.* **43** (1995), 681–693.
5. Z. Cvetković and M. Vetterli, Error-rate characteristics of oversampled analog-to-digital conversion, *IEEE Trans. Inform. Theory* **44** (1998), 1961–1964.
6. I. Daubechies, "Ten Lectures on Wavelets," SIAM, Philadelphia, 1992.

7. S. Deering and R. Hinden, Internet Protocol, version 6 (IPv6) specification, Network Working Group Request for Comments 1883, December 1995, `ftp://ftp.isi.edu/in-notes/rfc1883.txt`.
8. D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, Data compression and harmonic analysis, *IEEE Trans. Inform. Theory* **44** (1998), 2435–2476.
9. P. L. Dragotti, J. Kovačević, and V. K. Goyal, Quantized oversampled filter banks with erasures, *in* "Proc IEEE Data Compression Conf., Snowbird, UT" (J. A. Storer and M. Cohn, Eds.), pp. 173–182, IEEE Comput. Soc., Los Alamitos, CA, 2001.
10. V. K. Goyal, "Beyond Traditional Transform Coding," Ph.D. thesis, University California, Berkeley, 1998, published as Univ. California, Berkeley, Electron. Res. Lab. Memo. No. UCB/ERL M99/2, Jan. 1999.
11. V. K. Goyal and J. Kovačević, Optimal multiple description transform coding of Gaussian vectors, *in* "Proc. IEEE Data Compression Conf., Snowbird, UT" (J. A. Storer and M. Cohn, Eds.), pp. 388–397, IEEE Comput. Soc., Los Alamitos, CA, 1998.
12. V. K. Goyal, J. Kovačević, and M. Vetterli, Multiple description transform coding: Robustness to erasures using tight frame expansions, *in* "Proc IEEE Int. Symp. Inform. Th., Cambridge, MA, August 1998," p. 408.
13. V. K. Goyal, J. Kovačević, and M. Vetterli, Quantized frame expansions as source-channel codes for erasure channels, *in* "Proc. IEEE Data Compression Conf., Snowbird, UT" (J. A. Storer and M. Cohn, Eds.), pp. 326–335, IEEE Comput. Soc., Los Alamitos, CA, 1999.
14. V. K. Goyal, M. Vetterli, and N. T. Thao, Quantized overcomplete expansions in $\mathbb{R}^N$: Analysis, synthesis, and algorithms, *IEEE Trans. Inform. Theory* **44** (1998), 16–31.
15. R. M. Gray and D. L. Neuhoff, Quantization, *IEEE Trans. Inform. Theory* **44** (1998), 2325–2383.
16. R. M. Gray and T. G. Stockham, Jr., Dithered quantizers, *IEEE Trans. Inform. Theory* **39** (1993), 805–812.
17. R. A. Horn and C. R. Johnson, "Matrix Analysis," Cambridge Univ. Press, Cambridge, UK, 1985; reprinted with corrections, 1987.
18. N. Kingsbury, Image processing with complex wavelets, *Philos. Trans. Roy. Soc. London Ser. A.* **357** (1999), 2543–2560.
19. S. Lin and D. J. Costello, "Error Control Coding: Fundamentals and Applications," Prentice Hall, Englewood Cliffs, NJ, 1983.
20. S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, Quantization and dither: A theoretical survey, *J. Audio Engrg. Soc.* **40** (1992), 355–375.
21. S. Mehrotra and P. A. Chou, On optimal frame expansions for multiple description quantization, *in* "Proc. IEEE Int. Symp. Inform. Th., Sorrento, Italy, June 2000," p. 176.
22. A. Papoulis, "Probability, Random Variables, and Stochastic Processes," 3rd ed., McGraw-Hill, New York, 1991.
23. S. Rangan and V. K. Goyal, Recursive consistent estimation with bounded noise, *IEEE Trans. Inform. Theory* **47** (2001), 457–464.
24. I. W. Selesnick and L. Sendur, Iterated oversampled filter banks and wavelet frames, *in* "Proc. SPIE Wavelet Appl. in Signal and Image Proc. VIII," Vol. 4119, San Diego, CA, July–August 2000.
25. N. T. Thao and M. Vetteri, Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimates, *IEEE Trans. Signal Process.* **42** (1994), 519–531.
26. N. T. Thao and M. Vetteri, Reduction of the MSE in R-times oversampled A/D conversion from $O(1/R)$ to $O(1/R^2)$, *IEEE Trans. Signal Process.* **42** (1994), 200–203.
27. N. T. Thao and M. Vetteri, Lower bound on the mean-squared error in oversampled quantization of periodic signals using vector quantization analysis, *IEEE Trans. Inform. Theory* **42** (1996), 469–479.
28. M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Trans. Image Process.* **4** (1995), 1549–1560.
29. M. Vetterli and J. Kovačević, "Wavelets and Subband Coding," Prentice Hall, Englewood Cliffs, NJ, 1995.
30. Y. Wang, M. T. Orchard, and A. R. Reibman, Multiple description image coding for noisy channels by pairing transform coefficients, *in* "Proc. IEEE Workshop on Multimedia Sig. Proc., Princeton, NJ, June 1997," pp. 419–424.
31. D. C. Youla, Mathematical theory of jmage restoration by the method of convex projections, *in* "Image Recovery: Theory and Application" (H. Stark, Ed.), Academic Press, San Diego, 1987.