On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener

G. David Forney, Jr.[1]

Abstract. This paper explains why MMSE estimation arises in lattice-based strategies for approaching the capacity of linear Gaussian channels, and comments on its properties.

1 Introduction

Recently, Erez and Zamir [11, 13, 14, 32] have cracked the long-standing problem of achieving the capacity of additive white Gaussian noise (AWGN) channels using lattice codes and lattice decoding. Their method uses Voronoi codes (nested lattice codes), dither, and an MMSE estimation factor α that had previously been introduced in more complex multiterminal scenarios, such as Costa's "dirty-paper channel" [8]. However, they give no fundamental explanation for why an MMSE estimator, which is seemingly an artifact from the world of analog communications, plays such a key role in the digital communications problem of achieving channel capacity.

The principal purpose of this paper is to provide such an explanation, in the lattice-based context of a mod-Λ AWGN channel model. We discuss various properties of MMSE-based schemes in this application, some of which are unexpected.

MMSE estimators also appear as part of capacity-achieving solutions for more complicated digital communication scenarios involving linear Gaussian channels. While some of the explanation for the role of MMSE estimation in these more complex situations is no doubt information-theoretic (see, *e.g.,* [20]), the observations of this paper may provide some clues as to why MMSE estimation and lattice-type coding also work well in these more general applications.

2 Lattice-based coding for the AWGN channel

Consider the real discrete-time AWGN channel $Y = X + N$, where $\mathsf{E}[X^2] \leq S_x$ and N is independent[2] zero-mean Gaussian noise with variance S_n. The capacity is $C = \frac{1}{2}\log_2(1 + \mathrm{SNR})$ bits per dimension (b/d), where $\mathrm{SNR} = S_x/S_n$. Following Erez and Zamir [11, 13, 14, 32], we will show that lattice-based transmission systems can approach the capacity of this channel.

2.1 Lattices and spheres

Geometrically, an N-dimensional lattice Λ is a regular array of points in $\mathbb{R}^N$. Algebraically, Λ is a discrete subgroup of $\mathbb{R}^N$ which spans $\mathbb{R}^N$. A Voronoi region $\mathcal{R}_V(\Lambda)$ of Λ represents the quotient group $\mathbb{R}^N/\Lambda$ by a set of minimum-energy coset representatives for the cosets of Λ in $\mathbb{R}^N$. For any $\mathbf{x} \in \mathbb{R}^N$, "$\mathbf{x}$ mod Λ" denotes the unique element of $\mathcal{R}_V(\Lambda)$ in the coset $\Lambda + \mathbf{x}$. Geometrically, $\mathbb{R}^N$ is the disjoint union of the translated Voronoi regions $\{\mathcal{R}_V(\Lambda) + \boldsymbol{\lambda}, \boldsymbol{\lambda} \in \Lambda\}$. The volume $V(\Lambda)$ of $\mathcal{R}_V(\Lambda)$ is therefore the volume of $\mathbb{R}^N$ associated with each point of Λ.

[1] MIT, Cambridge, MA 02139 USA. E-mail: `forneyd@attbi.com`.

[2] Note that without the independence of N, the "additive" property is vacuous, since for any real-input, real-output channel we may define $N = Y - X$, and then express Y as $Y = X + N$. We exploit this idea later.

As $N \to \infty$, the Voronoi regions of some N-dimensional lattices can become more or less spherical, in various senses. As $N \to \infty$, an N-sphere (ball) of squared radius $N\rho^2$ has normalized volume (per two dimensions)

$$V_{\otimes}(N\rho^2)^{2/N} \overset{N\to\infty}{\longrightarrow} 2\pi e \rho^2.$$

The average energy per dimension of a uniform probability distribution over such an N-sphere goes to $P_{\otimes}(N\rho^2) = \rho^2$. The probability that an iid Gaussian random N-tuple with zero mean and symbol variance S_n falls outside the N-sphere becomes arbitrarily small for any $S_n < \rho^2$.

It is known that there exist high-dimensional lattices whose Voronoi regions are quasi-spherical in the following second moment sense. The *normalized second moment* of a compact region $\mathcal{R} \subset \mathbb{R}^N$ of volume $V(\mathcal{R})$ is defined as

$$G(\mathcal{R}) = \frac{P(\mathcal{R})}{V(\mathcal{R})^{2/N}},$$

where $P(\mathcal{R})$ is the average energy per dimension of a uniform probability distribution over $\mathcal{R}$. The normalized second moment of $\mathcal{R}$ exceeds that of an N-sphere. The normalized second moment of an N-sphere decreases monotonically with N and approaches $\frac{1}{2\pi e}$ as $N \to \infty$. A result of Poltyrev, presented by Feder and Zamir [30], is that there exist lattices Λ such that $\log 2\pi e G(\Lambda)$ is arbitrarily small, where $G(\Lambda)$ denotes the normalized second moment of $\mathcal{R}_V(\Lambda)$. Such lattices are said to be "good for quantization," or "good for shaping."

Poltyrev [26] has also shown that there exist high-dimensional lattices whose Voronoi regions are quasi-spherical in the sense that the probability that an iid Gaussian noise N-tuple with symbol variance S_n falls outside the Voronoi region $\mathcal{R}_V(\Lambda)$ is arbitrarily small as long as

$$S_n < \frac{V(\Lambda)^{2/N}}{2\pi e}.$$

Such lattices are said to be "good for AWGN channel coding," or "sphere-bound-achieving" [19].

2.2 Mod-lattice transmission and capacity

We now show that the mod-Λ transmission system shown in Figure 1 can approach the channel capacity $C = \frac{1}{2}\log_2(1 + S_x/S_n)$ b/d arbitrarily closely, provided that $G(\Lambda) \approx 1/(2\pi e)$ and $f(\mathbf{Y})$ is a MMSE estimator of $\mathbf{X}$.

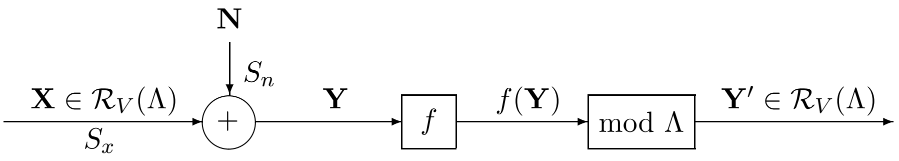

N
S_n
$\mathbf{X} \in \mathcal{R}_V(\Lambda)$
S_x
+
$\mathbf{Y}$
f
$f(\mathbf{Y})$
mod Λ
$\mathbf{Y}' \in \mathcal{R}_V(\Lambda)$

Figure 1. Mod-Λ transmission system over an AWGN channel.

This system is based on an N-dimensional lattice Λ whose Voronoi region $\mathcal{R}_V(\Lambda)$ has volume $V(\Lambda)$, average energy per dimension $P(\Lambda) = S_x$ under a uniform probability distribution over $\mathcal{R}_V(\Lambda)$, and thus normalized second moment $G(\Lambda) = P(\Lambda)/V(\Lambda)^{2/N}$.

The N-dimensional input vector $\mathbf{X}$ is restricted to the Voronoi region $\mathcal{R}_V(\Lambda)$. The output vector $\mathbf{Y}$ is mapped by some function f to another vector $f(\mathbf{Y}) \in \mathbb{R}^N$, which is then mapped modulo Λ to $\mathbf{Y}' = f(\mathbf{Y})$ mod Λ, also in the Voronoi region $\mathcal{R}_V(\Lambda)$.

2.3 MMSE estimation and capacity

Our main result is that capacity can be approached in the system of Figure 1 if and only if the lattice Λ is "good for shaping" and the function $f(\mathbf{Y})$ is an MMSE estimator. (The sufficiency of these conditions was shown in [13].)

As a first step, we derive a lower bound:

Theorem 1 (Mod-Λ channel capacity and MMSE estimation) *The capacity $C(\Lambda, f)$ of the mod-Λ transmission system of Figure 1 is lowerbounded by*

$$C(\Lambda, f) \geq C - \frac{1}{2}\log_2 2\pi e G(\Lambda) - \frac{1}{2}\log_2 \frac{S_{e,f}}{S_e} \quad b/d,$$

where $C = \frac{1}{2}\log_2(1+\mathrm{SNR})$ b/d is the capacity of the underlying AWGN channel, $G(\Lambda)$ is the normalized second moment of $\mathcal{R}_V(\Lambda)$, and $S_{e,f}$ and S_e are the average energies per dimension of $\mathbf{E}_f = f(\mathbf{Y}) - \mathbf{X}$ and of $\mathbf{E} = \hat{\mathbf{X}}(\mathbf{Y}) - \mathbf{X}$, respectively, where $\hat{\mathbf{X}}(\mathbf{Y})$ is the linear MMSE estimator of $\mathbf{X}$ given $\mathbf{Y}$.

The key to the proof of this theorem is the introduction of a dither variable $\mathbf{U}$ that is known to both transmitter and receiver, and whose probability distribution is uniform over the Voronoi region $\mathcal{R}_V(\Lambda)$, as in [12, 13]. Given a data vector $\mathbf{V} \in \mathcal{R}_V(\Lambda)$, the channel input is taken as

$$\mathbf{X} = \mathbf{V} + \mathbf{U} \bmod \Lambda.$$

This makes $\mathbf{X}$ a uniform random variable over $\mathcal{R}_V(\Lambda)$, *statistically independent of* $\mathbf{V}$. This property follows from the following lemma:[3]

Lemma 2 (Crypto lemma) *Let G be a compact abelian group[4] with group operation $+$, and let $Y = X + N$, where X and N are random variables over G and N is independent of X and uniform over G. Then Y is independent of X and uniform over G.*

Proof. Since $y - x$ runs through G as y runs through G and $p_N(n)$ is constant over $n \in G$, the distribution $p_{Y|X}(y|x) = p_N(y - x)$ is constant over $y \in G$ for any $x \in G$. □

One effect of the dither $\mathbf{U}$ is therefore to ensure that the channel input $\mathbf{X} = \mathbf{V} + \mathbf{U} \bmod \Lambda$ is uniform over $\mathcal{R}_V(\Lambda)$ and thus has average energy per dimension $P(\Lambda) = S_x$. A second and more important effect is to make $\mathbf{X}$ and thus also $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ independent of $\mathbf{V}$.

The dither may be subtracted out at the output of the channel, mod Λ, to give

$$\mathbf{Z} = f(\mathbf{Y}) - \mathbf{U} \bmod \Lambda.$$

The end-to-end channel is illustrated in Figure 2.

[3] We call this the crypto lemma because if we take X as plaintext, N as a cryptographic key, and $Y = X + N$ as the encrypted message, then the encrypted message is independent of the plaintext provided that the key is uniform, so no information can be obtained about the plaintext from the encrypted message without the key. On the other hand, given the key, the plaintext may be easily recovered from the encrypted message via $X = Y - N$. This is the principle of the one-time pad, which, as Shannon showed, is essentially the only way to achieve perfect secrecy in a cryptographic system.

[4] The group G is required to be compact so that its Haar (translation-invariant) measure $\mu(G)$ is finite and thus normalizable to a uniform probability distribution over G. However, G need not be abelian.

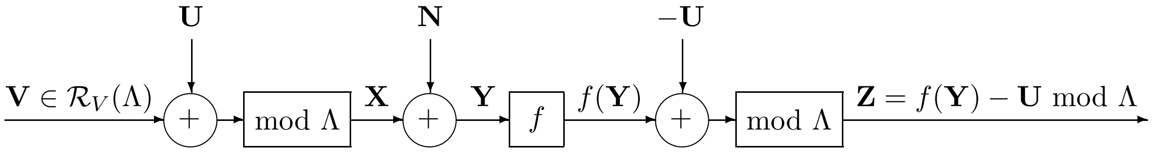


Figure 2. Creation of a mod-Λ channel $\mathbf{Z} = f(\mathbf{Y}) - \mathbf{U}$ mod Λ using dither.

Now let us regard $f(\mathbf{Y})$ as an estimator of $\mathbf{X}$, and define the estimation error as $\mathbf{E}_f = f(\mathbf{Y}) - \mathbf{X}$. Since $\mathbf{Y}$ and $\mathbf{X}$ are independent of $\mathbf{V}$, so is $\mathbf{E}_f$. Then

$$\mathbf{Z} = \mathbf{X} + \mathbf{E}_f - \mathbf{U} = \mathbf{V} + \mathbf{E}_f \text{ mod } \Lambda.$$

In short, we have created a mod-Λ additive noise channel $\mathbf{Z} = \mathbf{V} + \mathbf{E}_f$ mod Λ, where $\mathbf{E}_f$ is independent of $\mathbf{V}$. This equivalent channel is illustrated in Figure 3.

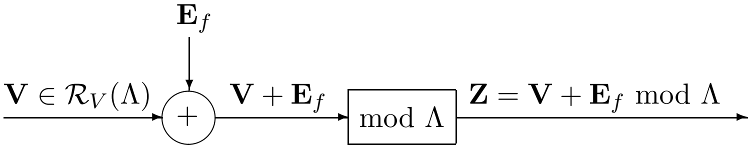

$\mathbf{E}_f$
$\mathbf{V} \in \mathcal{R}_V(\Lambda)$
$+$
$\mathbf{V} + \mathbf{E}_f$
mod Λ
$\mathbf{Z} = \mathbf{V} + \mathbf{E}_f$ mod Λ

Figure 3. Equivalent mod-Λ additive noise channel $\mathbf{Z} = \mathbf{V} + \mathbf{E}_f$ mod Λ.

As is well known, the capacity of an additive-noise channel $\mathbf{Z} = \mathbf{V} + \mathbf{E}_f$ mod Λ is achieved when the input distribution is uniform over $\mathcal{R}_V(\Lambda)$, in which case the output distribution is uniform as well, by the crypto lemma. In bits per dimension, it is equal to

$$C(\Lambda, f) = \frac{1}{N}(h(\mathbf{Z}) - h(\mathbf{Z} \mid \mathbf{V})) = \frac{1}{N}(\log_2 V(\Lambda) - h(\mathbf{E}'_f)) \quad \text{b/d},$$

where $h(\mathbf{Z}) = \log_2 V(\Lambda)$ is the differential entropy of a uniform distribution over a region of volume $V(\Lambda)$, and $h(\mathbf{E}'_f)$ is the differential entropy of the Λ-aliased additive noise $\mathbf{E}'_f = \mathbf{E}_f$ mod Λ. Now since $\mathbf{E}'_f$ is the result of applying the many-to-one mod-Λ map to $\mathbf{E}_f$, we have

$$h(\mathbf{E}'_f) \leq h(\mathbf{E}_f).$$

Moreover, if $\mathbf{E}_f$ has average energy per dimension $S_{e,f}$, then we have

$$h(\mathbf{E}_f) \leq \frac{N}{2} \log_2 2\pi e S_{e,f},$$

the differential entropy of an iid zero-mean Gaussian distribution with the same average energy. Putting this all together, and using $V(\Lambda)^{2/N} = P(\Lambda)/G(\Lambda)$ and $P(\Lambda) = S_x$, we have

$$C(\Lambda, f) \geq \frac{1}{N} \log_2 V(\Lambda) - \frac{1}{2} \log_2 2\pi e S_{e,f} = \frac{1}{2} \log_2 \frac{S_x}{S_{e,f}} - \frac{1}{2} \log_2 2\pi e G(\Lambda) \quad \text{b/d}.$$

The linear MMSE estimator $\hat{\mathbf{X}}(\mathbf{Y})$ of $\mathbf{X}$ is $\hat{\mathbf{X}}(\mathbf{Y}) = \alpha \mathbf{Y}$, where

$$\alpha = \frac{S_x}{S_x + S_n} = \frac{\text{SNR}}{1 + \text{SNR}}.$$

By the orthogonality principle of MMSE estimation theory, the linear MMSE estimation error $\mathbf{E} = \mathbf{X} - \alpha\mathbf{Y} = (1-\alpha)\mathbf{X} - \alpha\mathbf{N}$ is then uncorrelated with $\mathbf{Y}$.[5] The average energy of the estimation error per dimension becomes

$$S_e = (1-\alpha)^2 S_x + \alpha^2 S_n = \frac{S_x S_n}{S_x + S_n} = \alpha S_n.$$

Finally, since $S_x/S_e = S_y/S_n = 1 + \mathrm{SNR}$, we have

$$C(\Lambda, f) \geq \frac{1}{2}\log_2 \frac{S_x}{S_e} + \frac{1}{2}\log_2 \frac{S_e}{S_{e,f}} - \frac{1}{2}\log_2 2\pi e G(\Lambda) = C - \frac{1}{2}\log_2 2\pi e G(\Lambda) - \frac{1}{2}\log_2 \frac{S_{e,f}}{S_e} \quad \text{b/d.}$$

This completes the proof of Theorem 1. □

Remark 1 (dither is unnecessary). Evidently a channel $\mathbf{Z} = \mathbf{V} + \mathbf{u} + \mathbf{E}_f \bmod \Lambda$ with a fixed dither vector $\mathbf{u} \in \mathcal{R}_V(\Lambda)$ has the same capacity $C(\Lambda, f)$. Therefore introducing the random dither variable $\mathbf{U}$ is just a tactic to prove Theorem 1; dither is not actually needed to achieve $C(\Lambda, f)$. However, dither is key to decoupling the Shannon and the Wiener problems.[6] □

Remark 2 (MMSE estimation and bias). Notice that the signal-to-noise ratio of the channel $\mathbf{Z} = \mathbf{V} + \mathbf{E} \bmod \Lambda$ is $S_x/S_e = S_y/S_n = 1 + \mathrm{SNR}$, and moreover this channel has no bias. Thus the MMSE factor α and random dither increase the effective signal-to-noise ratio from SNR to $1 + \mathrm{SNR}$ without introducing bias. This is evidently a different way of approaching capacity than that given in [4], where the apparent $\mathrm{SNR}_{\mathrm{MMSE-DFE}}$ was discounted to $\mathrm{SNR}_{\mathrm{MMSE-DFE,U}} = \mathrm{SNR}_{\mathrm{MMSE-DFE}} - 1$ to account for bias. We will discuss bias further below. □

Remark 3 ("dirty-paper" capacity). This approach easily extends to give a constructive proof of Costa's result [8] that channel interference known to the transmitter does not reduce capacity; see, *e.g.,* [32, 1]. Let the channel model be $\mathbf{Y} = \mathbf{X} + \mathbf{N} + \mathbf{S}$, where $\mathbf{S}$ is an arbitrary interference

[5] These relations are illustrated by the "Pythagorean" right triangle shown in Figure 4(a) below, which follows from interpreting covariances as inner products of vectors in a two-dimensional Hilbert space. Since $\mathsf{E}[XN] = 0$, the two vectors corresponding to X and N are orthogonal. Their squared lengths are given by $\mathsf{E}[X^2] = S_x$ and $\mathsf{E}[N^2] = S_n$. The hypotenuse corresponds to the sum $Y = X + N$, and has squared length $S_y = S_x + S_n$. Since $\mathsf{E}[XY] = S_x$, the projection of Y onto X is $\hat{Y}(X) = (\mathsf{E}[XY]/\mathsf{E}[X^2])X = X$, and the projection of X onto Y is $\hat{X}(Y) = (\mathsf{E}[XY]/\mathsf{E}[Y^2])Y = \alpha Y$. Then $E = X - \hat{X}(Y) = X - \alpha Y$ is orthogonal to Y.

The inner right triangle in Figure 4(a) with sides $(\hat{X}(Y), E, X)$ is similar, so since $S_x = \alpha S_y$ the squared lengths of its sides are $(S_{\hat{x}} = \alpha S_x, S_e = \alpha S_n, S_x = \alpha S_y)$, respectively.

Figures 4(b)-(c) show that right triangles whose sides have squared lengths $(1, \mathrm{SNR}, 1 + \mathrm{SNR})$ or $(\alpha, 1-\alpha, 1)$ are also similar.

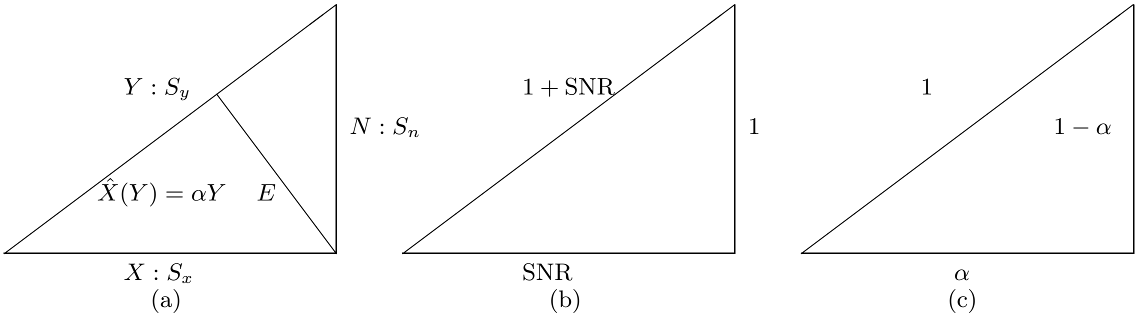


Figure 4. (a) "Pythagorean" right triangle with sides (X, N, Y), with similar inner right triangle with sides $(\hat{X}(Y) = \alpha Y, E = X - \hat{X}(Y), X)$. (b), (c): Similar right trangles.

[6] This is analogous to the tactic used by Elias [10] to prove that binary linear block codes can achieve the capacity of a binary input-symmetric channel, namely the introduction of a random translate $\mathcal{C} + \mathbf{U}$ of a binary linear block code $\mathcal{C}$ of length N, where $\mathbf{U}$ is a random binary N-tuple which is uniform over $(\mathbb{F}_2)^N$.

vector known to the transmitter. Then let the channel input be $\mathbf{X} = \mathbf{V} + \mathbf{U} - \alpha\mathbf{S} \bmod \Lambda$. The channel input is still uniform and independent of $\mathbf{V}$, by the crypto lemma, while the effect of the interference $\mathbf{S}$ is entirely cancelled in $\mathbf{Z} = \alpha\mathbf{Y} - \mathbf{U} = \mathbf{V} + \mathbf{E}_f \bmod \Lambda$. Thus the receiver needs to know nothing about the interference, the equivalent channel model is the same, and $C(\Lambda, f)$ is unaffected. □

An immediately corollary of Theorem 1 is that the capacity $C = \frac{1}{2}\log_2(1 + \mathrm{SNR})$ can be approached if $\log 2\pi e G(\Lambda) \approx 0$ and $f(\mathbf{Y})$ is taken as the linear MMSE estimator $\hat{\mathbf{X}}(\mathbf{Y}) = \alpha\mathbf{Y}$:

Corollary 3 (Mod-Λ transmission schemes can approach C) [Erez-Zamir [13]] *For any $\varepsilon > 0$, there exists a lattice Λ such that the capacity of the mod-Λ channel of Figure 1 satisfies*

$$C(\Lambda, f) \geq C - \varepsilon \quad b/d,$$

where $C = \frac{1}{2}\log_2(1 + \mathrm{SNR})$ is the capacity of the underlying AWGN channel, provided that the function $f(\mathbf{Y})$ is the linear MMSE estimator $\hat{\mathbf{X}}(\mathbf{Y}) = \alpha\mathbf{Y}$.

Remark 4 (linear MMSE estimation suffices). A true MMSE estimator would achieve an error variance at least as small as the linear MMSE estimator, and therefore would also achieve capacity. However, Corollary 3 shows that the linear MMSE estimator is good enough. □

We now show that the conditions $\log_2 2\pi e G(\Lambda) \to 0$ and $S_{e,f} = S_e$ are not only sufficient but also necessary to reach capacity. Briefly, the arguments are as follows:

1. The differential entropy per dimension of $\mathbf{X}$ and $\mathbf{Z}$, namely

$$\frac{1}{N}h(\mathbf{X}) = \frac{1}{N}h(\mathbf{Z}) = \frac{1}{2}\log_2 V(\Lambda)^{2/N} = \frac{1}{2}\log_2 2\pi e S_x - \frac{1}{2}\log_2 2\pi e G(\Lambda)$$

goes to $\frac{1}{2}\log_2 2\pi e S_x$ if and only if $\log_2 2\pi e G(\Lambda) \to 0$. This condition is necessary because the capacity of an AWGN channel with input power constraint S_x can be approached arbitrarily closely only if the differential entropy of the input distribution approaches $\frac{1}{2}\log_2 2\pi e S_x$.

Remark 5 (Gaussian approximation principle). The differential entropy of any random vector $\mathbf{X}$ with average energy per dimension S_x is less than or equal to $\frac{1}{2}\log_2 2\pi e S_x$, with equality if and only if $\mathbf{X}$ is iid Gaussian. Therefore if $\mathbf{X}_n$ is a sequence of random vectors of dimension $N(n) \to \infty$ and average energy per dimension S_x such that $h(\mathbf{X}_n)/N(n) \to \frac{1}{2}\log_2 2\pi e S_x$, we say that the sequence $\mathbf{X}_n$ is *Gaussian in the limit.* Restating the above argument, if $\mathbf{X}_n$ is uniform over $\mathcal{R}_V(\Lambda_n)$, then $\mathbf{X}_n$ is Gaussian in the limit if and only if $\log 2\pi e G(\Lambda_n) \to 0$.[7] □

2. The channel output $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ is then also Gaussian in the limit, so the linear MMSE estimator $\hat{\mathbf{X}}(\mathbf{Y}) = \alpha\mathbf{Y}$ becomes a true MMSE estimator in the limit. The MMSE estimation error $\mathbf{E} = -(1-\alpha)\mathbf{X} + \alpha\mathbf{N}$ becomes Gaussian in the limit with symbol variance $S_e = \alpha S_n$, and becomes independent of $\mathbf{Y}$. In order that $C(\Lambda, f) \to C$, it is then necessary that $S_{e,f} = S_e$, which by definition implies that $f(\mathbf{Y})$ is an MMSE estimator.[8]

In summary, these two conditions are necessary as well as sufficient:

Theorem 4 (Necessary conditions to approach C) *The capacity of the mod-Λ channel of Figure 1 approaches C only if $\log 2\pi e G(\Lambda) \to 0$ and $f(\mathbf{Y})$ is an MMSE estimator of $\mathbf{X}$ given $\mathbf{Y}$.*

[7] Zamir and Feder [30] show that if $\mathbf{X}_n$ is uniform over an $N(n)$-dimensional region $\mathcal{R}_n$ of average energy S_x and $G(\mathcal{R}_n) \to 1/(2\pi e)$, then the normalized divergence $\frac{1}{N(n)}D(\mathbf{X}_n||\mathbf{N}_n) \to 0$, where $\mathbf{N}_n$ is an iid Gaussian random vector with zero mean and variance S_x. They go on to show that this implies that any finite-dimensional projection of $\mathbf{X}_n$ converges in distribution to an iid Gaussian vector.

[8] Since $\mathbf{E}_f = \mathbf{E} + (f(\mathbf{Y}) - \hat{\mathbf{X}}(\mathbf{Y}))$ and $\mathbf{Y}$ and $\mathbf{E}$ are independent, $S_{e,f} = S_e + \frac{1}{N}\mathsf{E}[||f(\mathbf{Y}) - \hat{\mathbf{X}}(\mathbf{Y})||^2]$. Thus $f(\mathbf{Y})$ is an MMSE estimator if and only if $\mathsf{E}[||f(\mathbf{Y}) - \hat{\mathbf{X}}(\mathbf{Y})||^2] = 0$.

Remark 6 (MMSE estimation and lattice decoding). One interpretation of the Erez-Zamir result is that the scaling introduced by the MMSE estimator is somehow essential for lattice decoding of a fine-grained coding lattice Λ_c. Theorem 4 shows however that in the mod-Λ channel an MMSE estimator is necessary to achieve capacity, quite apart from any particular coding and decoding scheme. □

Remark 7 (aliasing becomes negligible). Under these conditions, Theorem 1 says that $C(\Lambda, f) \geq C$. Since $C(\Lambda, f)$ cannot exceed C, this implies that all inequalities in the proof of Theorem 1 must tend to equality, and in particular that

$$\frac{h(\mathbf{E}')}{N} \to \frac{h(\mathbf{E})}{N} \to \frac{1}{2} \log_2 2\pi e S_e,$$

where $\mathbf{E}' = \mathbf{E} \bmod \Lambda$ is the Λ-aliased version of the estimation error $\mathbf{E}$. So not only must $\mathbf{E}$ become Gaussian in the limit, *i.e.,* $h(\mathbf{E})/N \to \frac{1}{2} \log_2 2\pi e S_e$, but also $\mathbf{E}'$ must tend to $\mathbf{E}$, which means that the effect of the mod-Λ aliasing must become negligible. This is as expected, since $\mathbf{E}$ is Gaussian in the limit with symbol variance S_e and $\mathcal{R}_V(\Lambda)$ is quasi-spherical with average energy per dimension $S_x > S_e$. □

2.4 Voronoi codes

A *Voronoi code* $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda) = (\Lambda_c + \mathbf{u}) \cap \mathcal{R}_V(\Lambda)$ is the set of points in a translate $\Lambda_c + \mathbf{u}$ of an N-dimensional "coding lattice" Λ_c that lie in the Voronoi region $\mathcal{R}_V(\Lambda)$ of a "shaping" sublattice $\Lambda \subset \Lambda_c$. (Such codes were called "Voronoi codes" in [7], "Voronoi constellations" in [18], and "nested lattice codes" in [31, 12, 13, 14, 32, 1]. Here we will use the original term.)

A Voronoi code has $|\Lambda_c/\Lambda| = V(\Lambda)/V(\Lambda_c)$ code points, and thus rate

$$R(\Lambda_c/\Lambda) = \frac{1}{N} \log_2 \frac{V(\Lambda)}{V(\Lambda_c)} = \frac{1}{2} \left(\log_2 \frac{V(\Lambda)^{2/N}}{2\pi e} - \log_2 \frac{V(\Lambda_c)^{2/N}}{2\pi e} \right) \quad \text{b/d.}$$

Erez and Zamir [11, 13, 14, 32] have shown rigorously (not employing the Gaussian approximation principle) that there exists a random ensemble $\mathcal{C}((\Lambda_c + \mathbf{U})/\Lambda)$ of dithered Voronoi codes that can approach the capacity $C(\Lambda)$ of the mod-Λ transmission system of Figure 1 arbitrarily closely, if $f(\mathbf{Y}) = \hat{\mathbf{X}}(\mathbf{Y}) = \alpha\mathbf{Y}$. The decoder may be the usual minimum-Euclidean-distance decoder, even though the effective noise $\mathbf{E} = -(1-\alpha)\mathbf{X} + \alpha\mathbf{N}$ is not Gaussian.

If $C(\Lambda) \approx C$ and $P(\Lambda) = S_x$, this implies that $2\pi e G(\Lambda) \approx 1$; *i.e.,* Λ is "good for shaping." Furthermore, since the effective noise has variance S_e, if the error probability is arbitrarily small and $R(\Lambda_c/\Lambda) \approx C = \frac{1}{2} \log_2 S_x/S_e$, then

$$\log_2 S_e \approx \log_2 \frac{V(\Lambda_c)^{2/N}}{2\pi e};$$

i.e., Λ_c is "good for AWGN channel coding," or "sphere-bound-achieving."

The ensemble $\mathcal{C}((\Lambda_c + \mathbf{U})/\Lambda)$ is an ensemble of fixed-dither Voronoi codes $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$. The average probability of decoding error $\Pr_{\mathbf{U}}(E) = \mathsf{E}_{\mathbf{U}}[\Pr_{\mathbf{u}}(E)]$ is arbitrarily small over this ensemble, using a decoder that is appropriate for random dither (*i.e.,* minimum-distance decoding). This implies not only that there exists at least one fixed-dither code $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$ such that $\Pr_{\mathbf{u}}(E) \leq \Pr_{\mathbf{U}}(E)$, using the same decoder, but also that at least a fraction $1 - \varepsilon$ of the fixed-dither codes have $\Pr_{\mathbf{u}}(E) \leq \frac{1}{\varepsilon} \Pr_{\mathbf{U}}(E)$; *i.e.,* almost all fixed-dither codes have low $\Pr_{\mathbf{u}}(E)$.

This result is somewhat counterintuitive, since for fixed dither $\mathbf{u}$, $\mathbf{X}$ is not independent of $\mathbf{V}$; indeed, there is a one-to-one correspondence given by $\mathbf{X} = \mathbf{V} + \mathbf{u} \bmod \Lambda$. Therefore, the error

$$\mathbf{E} = -(1-\alpha)\mathbf{X} + \alpha\mathbf{N} = -(1-\alpha)(\mathbf{V} + \mathbf{u} \bmod \Lambda) + \alpha\mathbf{N}$$

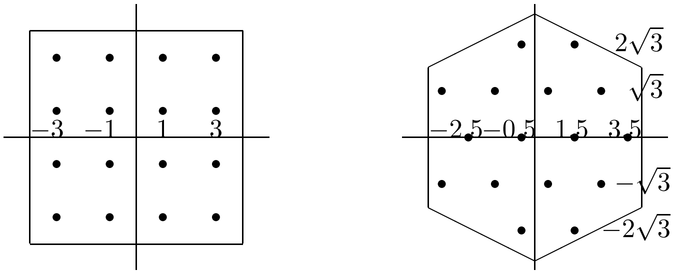
is not independent of $\mathbf{V}$; *i.e.,* there is bias in the equivalent channel output $\mathbf{Z} = \mathbf{V} + \mathbf{E} \bmod \Lambda$. Even so, we see that capacity can be achieved by a suboptimum decoder which ignores bias.

Since almost all fixed-dither codes achieve capacity, we may as well use the code $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$ that has minimum average energy $S_{\min} \le P(\Lambda) = S_x$ per dimension. But if $S_{\min} < S_x$, then we could achieve a rate greater than the capacity of an AWGN channel with signal-to-noise ratio $S_{\min}/S_n < S_x/S_n$. We conclude that the average energy per dimension of $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$ cannot be materially less than $S_x = P(\Lambda)$ for any $\mathbf{u}$, and thus must be approximately S_x for almost all values of the dither $\mathbf{u}$, in order for the average over $\mathbf{U}$ to be S_x. In summary:

Theorem 5 *(Average energy of Voronoi codes) If $\mathcal{C}((\Lambda_c+\mathbf{u})/\Lambda)$ is a capacity-achieving Voronoi code, then Λ_c is good for AWGN channel coding, Λ is good for shaping, the decoder may ignore bias, and the average energy per dimension of $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$ is $\approx P(\Lambda)$.*

Remark 8 (Average energy of Voronoi codes). Theorem 5 shows that the hope of [19] that one could find particular Voronoi codes with average energy $S_x - S_e$ was misguided. For Voronoi codes, the original "continuous approximation" of [17] holds, not the "improved continuous approximation" of [19], and the "discretization factor" of [22] does not apply.[9]

Remark 9 (observations on output scaling). It is surprising that a decoder for Voronoi codes which first scales the received signal by α and then does lattice decoding should perform better than one that just does lattice decoding. Optimum (ML) decoding on this channel is minimum-distance (MD) decoding, and ordinary lattice decoding is equivalent to minimum-distance decoding except on the boundary of the support region.



(a) (b)

Figure 5. (a) 16-point square Voronoi code $\mathcal{C}(2\mathbb{Z}^2/8\mathbb{Z}^2)$;
(b) 16-point hexagonal Voronoi code $\mathcal{C}(2A_2 - (\frac{1}{2}, 0)/8A_2)$. □

[9] Indeed, the fact that the best dithered Voronoi code does not have average energy significantly less than that of the average code is already apparent in two dimensions. Consider the two 16-point two-dimensional signal sets shown in Figure 5; these are the minimum-energy Voronoi codes based on the lattice partitions $2\mathbb{Z}^2/8\mathbb{Z}^2$ and $2A_2/8A_2$, where $\mathbb{Z}^2$ is the two-dimensional integer lattice and A_2 is the two-dimensional hexagonal lattice [18]. The average energy per dimension of the square code is 5, which is indeed less than the average energy $P(8\mathbb{Z}^2) = 16/3$ of the bounding region $\mathcal{R}_V(8\mathbb{Z}^2)$ by the average energy $P(2\mathbb{Z}^2) = 1/3$ of the basic cell $\mathcal{R}_V(2\mathbb{Z}^2)$. This is because $2\mathbb{Z}^2/8\mathbb{Z}^2$ is a Cartesian product of the one-dimensional partition $2\mathbb{Z}/8\mathbb{Z}$, and $\mathcal{C}(2\mathbb{Z}/8\mathbb{Z})$ is a perfect one-dimensional sphere packing. However, the average energy per dimension of the hexagonal code is $35/8 = 4.375$, compared to the average energy $P(8A_2) = 9/2 = 4.5$ of the bounding region $\mathcal{R}_V(8A_2)$; the difference is already noticeably less than the average energy $P(2A_2) = 9/32 = 0.28125$ of the basic cell $\mathcal{R}_V(2A_2)$.

Scaling by α seems excessive. The variance of the channel output Y is S_y. Scaling the output by α reduces the received variance to $S_{\hat{x}} = \alpha^2 S_y = \alpha S_x$, less than the input variance. This means that the scaled output αY is almost surely going to lie in a spherical shell of average energy per dimension $\approx \alpha S_x$, whereas the code vectors in the Voronoi code $\mathcal{C}(\Lambda_c/\Lambda)$ almost all lie on a spherical shell of average energy $\approx S_x$. Yet the subsequent lattice decoding to $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$ works, even though it seems that the decoder should decode to $\alpha\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$.

These questions about scaling may be resolved if as $N \to \infty$ it suffices to decode Voronoi codes based on angles, ignoring magnitudes. Then whether the decoder uses $Y, \alpha Y$ or $\sqrt{\alpha} Y$ as input, the optimum minimum-angle decoder would be the same. Indeed, Urbanke and Rimoldi [28], following Linder *et al.* [24], have shown that as $N \to \infty$ a suboptimum decoder for spherical lattice codes that does minimum-angle decoding to the subset of codewords in a spherical shell of average energy $\approx S_x$ suffices to approach capacity.

Of course, lattice decoding does depend on scale, so it seems that analyzing the performance of a properly scaled lattice decoder is just an analytical trick for finding the optimal minimum-angle decoder performance, and incidentally showing that lattice decoding of Voronoi codes suffices to reach capacity.

Finally, notice that with a fixed code and scaling by α, as $N \to \infty$ the output $\alpha\mathbf{Y}$ almost surely lies in a sphere of average energy $\approx \alpha S_x < S_x$, inside the Voronoi region $\mathcal{R}_V(\Lambda)$, so the mod-Λ operation in the receiver has negligible effect and may be omitted. □

2.5 Shannon codes, spherical lattice codes and Voronoi codes

The following remarks point out some fundamental differences between capacity-achieving spherical lattice codes and capacity-achieving Voronoi codes.

Remark 10 (Shannon codes, spherical lattice codes, and Voronoi codes). In Shannon's random code ensemble for the AWGN channel, the code point $\mathbf{X}$ asymptotically lies almost surely in a spherical shell of average energy per dimension $S_x \pm \varepsilon$, the received vector $\mathbf{Y}$ lies almost surely in a spherical shell of average energy per dimension $S_y \pm \varepsilon$, and the noise vector $\mathbf{N}$ lies almost surely in a spherical shell of average energy per dimension $S_n \pm \varepsilon$. Thus we obtain a geometrical picture in which a "output sphere" of average energy $\approx S_y$ is partitioned into $\approx (S_y/S_n)^{N/2}$ probabilistically disjoint "noise spheres" of squared radius $\approx S_n$. Curiously, the centers of the noise spheres are at average energy $\approx S_x$, even though practically all of the volumes of the noise spheres are at average energy $\approx S_y$.

Urbanke and Rimoldi [28] have shown that spherical lattice codes (the set of all points in a lattice Λ_c that lie within a sphere of average energy S_x) can achieve the channel capacity $C = \frac{1}{2}\log_2 S_y/S_n$ b/d with minimum-distance decoding. Since again $\mathbf{Y}$ and $\mathbf{N}$ must lie almost surely in spheres of average energy S_y and S_n, respectively, we again have a picture in which the output sphere must be partitioned into $\approx (S_y/S_n)^{N/2}$ effectively disjoint noise spheres whose centers are the points in the spherical lattice code, which have average energy $\approx S_x$.

Voronoi codes evidently work differently. The Voronoi region $\mathcal{R}_V(\Lambda)$ has average energy S_x, and so does any good Voronoi code $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$. Moreover, $\mathcal{R}_V(\Lambda)$ is the disjoint union (mod Λ) of $V(\Lambda)/V(\Lambda_c) \approx (S_x/S_e)^{N/2}$ small Voronoi regions, whose centers are the points in $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$. So *the centers have the same average energy as the bounding region*, in contrast to the spherical case.

By the sphere bound [19, 26] $\log_2 V(\Lambda_c)^{2/N}/(2\pi e) \geq \log_2 S_c$, where S_c is the channel noise variance, so the capacity of the mod-Λ channel is limited to $\frac{1}{2}\log_2 S_x/S_c$. If the channel noise has variance $S_c = S_n$, then the capacity is limited to $\overline{C} = \frac{1}{2}\log_2 S_x/S_n = \frac{1}{2}\log_2 \mathrm{SNR}$, which is the best that de Buda and others [9, 25] were able to achieve with Voronoi codes prior to [13]. However, the MMSE estimator reduces the effective channel noise variance to $S_c = S_e = \alpha S_n$, which allows the capacity to approach $C = \frac{1}{2}\log_2 S_x/S_e = \frac{1}{2}\log_2(1+\mathrm{SNR})$. So in the mod-$\Lambda$ setting the MMSE estimator is the crucial element that precisely compensates for the Voronoi code capacity loss from C to the "lattice capacity" $\overline{C}$.

Finally, consider a "backward-channel" view of the Shannon ensemble. The jointly Gaussian pair (X, Y) is equally well modeled by the forward-channel model $Y = X + N$ or the backward-channel model $X = \alpha Y + E$. From the latter perspective, the transmitted codeword $\mathbf{X}$ lies almost surely in a spherical shell of average energy $\approx S_e$ about the scaled received word αY, which lies almost surely in a spherical shell of average energy $\approx \alpha^2 S_y = \alpha S_x$. Thus we obtain a geometrical picture in which an "input sphere" of average energy $\approx S_x$ is partitioned into $\approx (S_x/S_e)^{N/2}$ probabilistically disjoint "decision spheres" of squared radius $\approx S_e$. The centers of the decision spheres are codewords of average energy $\approx S_x$.

Capacity-achieving Voronoi codes thus appear to be designed and decoded in accord with the backward-channel view of the Shannon ensemble, whereas capacity-achieving spherical lattice codes appear to be designed and decoded in accord with the forward-channel view.

Remark 11 ("nubbly spheres" and "nubbly Voronoi regions"). The regions formed by the unions of the decision regions of capacity-achieving spherical lattice codes and Voronoi codes further illustrate the differences between these two types of lattice codes.

Let $\mathcal{C}$ be a capacity-achieving spherical lattice code $\mathcal{C} = \mathcal{S} \cap \Lambda_c$ such that $|\mathcal{C}| \approx (1 + S_x/S_n)^{N/2}$, where $\mathcal{S}$ is a sphere of average energy per dimension $\approx S_x$ and thus volume $V(\mathcal{S}) \approx (2\pi e S_x)^{2/N}$, and Λ_c is a sphere-bound-achieving lattice for noise with variance S_n, so $S_n \approx V(\Lambda_c)^{2/N}/(2\pi e)$. The average energy of $\mathcal{R}_V(\Lambda_c)$ per dimension is then $P(\Lambda_c) \approx 2\pi e G(\Lambda_c) S_n$.

Consider the "nubbly sphere" $\mathcal{R} = \mathcal{C} + \mathcal{R}_V(\Lambda_c)$, namely the disjoint union of the Voronoi regions of Λ_c whose centers are at the code points. The intersection of $\mathcal{R}$ with Λ_c is then $\mathcal{C}$, the normalized volume of $\mathcal{R}$ is

$$V(\mathcal{R})^{2/N} = |\mathcal{C}|^{2/N} V(\Lambda_c)^{2/N} \approx 2\pi e(S_x + S_n),$$

and the average energy per dimension of $\mathcal{R}$ is $P(\mathcal{R}) = S_x + P(\Lambda_c)$, because we can write a uniform random variable U over $\mathcal{R}$ as the sum $U = X + N$ of a uniform discrete random variable X over $\mathcal{C}$ and an independent uniform continuous random variable N over $\mathcal{R}_V(\Lambda_c)$, whose average energies per dimension are S_x and $P(\Lambda_c)$, respectively. Thus the normalized second moment of the nubbly sphere $\mathcal{R}$ is

$$G(\mathcal{R}) = \frac{P(\mathcal{R})}{V(\mathcal{R})^{2/N}} \approx \frac{S_x + P(\Lambda_c)}{2\pi e(S_x + S_n)} \approx \frac{S_x + 2\pi e G(\Lambda_c) S_n}{2\pi e(S_x + S_n)}.$$

It follows that if Λ_c is "good for shaping" in the sense that $2\pi e G(\Lambda_c) \approx 1$, then so is $\mathcal{R}$. In this sense the nubbly sphere $\mathcal{R}$ is a good approximation to a sphere of average energy $S_y = S_x + S_n$.

In comparison, given a Voronoi code $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$, consider the "nubbly Voronoi region" $\mathcal{R}_v = \mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda) + \mathcal{R}_V(\Lambda_c)$, namely the disjoint union of the Voronoi regions of $\Lambda_c + \mathbf{u}$ whose centers are at the code points. The region $\mathcal{R}_v$ appears at first glance to be a close approximation to $\mathcal{R}_V(\Lambda)$; its volume is the same, and its intersection with $\Lambda_c + \mathbf{u}$ is $\mathcal{C}((\Lambda_c + \mathbf{u})/\Lambda)$. However, its average energy per dimension is $P(\mathcal{R}_v) = P(\Lambda) + P(\Lambda_c)$, not $P(\Lambda)$, by an argument similar to that above. Thus, defining $\alpha = P(\Lambda)/(P(\Lambda) + P(\Lambda_c))$, we have

$$G(\mathcal{R}_v) = \frac{P(\mathcal{R}_v)}{V(\mathcal{R}_v)^{2/N}} = \frac{P(\Lambda)}{\alpha V(\Lambda)^{2/N}} = \frac{G(\Lambda)}{\alpha},$$

so, unlike the sphere and the nubbly sphere, the nubbly Voronoi region $\mathcal{R}_v$ cannot possibly be "good for shaping," and as $\alpha \to \frac{1}{2}$ it becomes quite bad. Thus both in a second moment sense and in a shape sense, the nubbly Voronoi region $\mathcal{R}_v$ is not a good approximation to the Voronoi region $\mathcal{R}_V(\Lambda)$. □

2.6 A Voronoi quantizer

We now present a quantizer structure based on an MMSE estimator and a randomly dithered Voronoi codebook $\mathcal{C}((\Lambda_q + \mathbf{U})/\Lambda)$ which is dual to that of the mod-Λ transmission system of Figure 1. This structure is equivalent to the Wyner-Ziv source coders of [32] or [1, Appx. III] when the side information is zero. The differences between this quantizer and previous lattice quantizers based on spherical or Voronoi-bounded codebooks (see, *e.g.,* [16]) further illustrate the distinctions upon which we remarked in the previous subsection. In particular, this Voronoi quantizer approaches the rate-distortion bound at all signal-to-quantization-noise ratios.

The source will be taken as an N-dimensional iid Gaussian variable $\mathbf{Y}$ with symbol variance S_y. Our objective will be to quantize the source with mean squared distortion S_n per dimension at a rate which approximates the rate-distortion limit $R = \frac{1}{2}\log_2 \mathrm{SQNR}$ b/d, where we define $\mathrm{SQNR} = S_y/S_n$. We also define

$$\alpha = \frac{\mathrm{SQNR} - 1}{\mathrm{SQNR}}.$$

The quantizer will be based on an N-dimensional lattice partition Λ_q/Λ, where Λ_q is "good for quantizing" in the sense that $2\pi e G(\Lambda_q) \approx 1$, and $\mathcal{R}_V(\Lambda_q)$ has average energy per dimension $P(\Lambda) = S_e = \alpha S_n$, while Λ is "good for AWGN channel coding" in the sense that the probability that an iid Gaussian variable with symbol variance $S_x = \alpha S_y \approx V(\Lambda)^{2/N}/(2\pi e)$ falls outside $\mathcal{R}_V(\Lambda)$ is arbitrarily small. A codebook based on this lattice partition then has rate

$$R = \frac{1}{N}\log_2 |\Lambda_q/\Lambda| = \frac{1}{2}\left(\log_2 V(\Lambda)^{2/N} - \log_2 V(\Lambda_q)^{2/N}\right) \approx \frac{1}{2}\log_2 \frac{\alpha S_y}{\alpha S_n} = \frac{1}{2}\log_2 \mathrm{SQNR} \quad \text{b/d}.$$

To quantize the source, we first form the "MMSE estimate" $\hat{\mathbf{X}}(\mathbf{Y}) = \alpha\mathbf{Y}$. We then quantize $\hat{\mathbf{X}}(\mathbf{Y})$ to the closest element $\hat{\mathbf{Z}}$ of $\Lambda_q + \mathbf{u}$, where $\mathbf{u}$ is a realization of a random dither variable $\mathbf{U}$ uniformly distributed over $\mathcal{R}_V(\Lambda_q)$. The quantization error $\mathbf{E} = \hat{\mathbf{Z}} - \hat{\mathbf{X}}(\mathbf{Y})$ is then uniformly distributed over $\mathcal{R}_V(\Lambda_q)$, with average energy per dimension S_e, and independent of $\mathbf{Y}$. Moreover, by the Gaussian approximation principle, $\mathbf{E}$ may be regarded as approximately iid Gaussian.

The quantized value $\hat{\mathbf{Z}} \in \Lambda_q + \mathbf{u}$ is then reduced modulo Λ. This yields an element $\hat{\mathbf{Y}}$ of the codebook $\mathcal{C}(\Lambda_q + \mathbf{u}/\Lambda)$, which has rate $R \approx \frac{1}{2}\log_2 \mathrm{SQNR}$ b/d. The index of $\hat{\mathbf{Y}}$ is the quantizer output, and the reconstructed estimate of $\mathbf{Y}$ is taken as $\hat{\mathbf{Y}}$. The quantization distortion is therefore $\mathbf{N} = \mathbf{Y} - \hat{\mathbf{Y}}$.

To analyze the performance of this quantizer, we first note that $\hat{\mathbf{Z}} = \hat{\mathbf{X}}(\mathbf{Y}) + \mathbf{E}$ is the sum of two independent variables with symbol variances $\alpha^2 S_y = \alpha S_x$ and $S_e = (1-\alpha)S_x$, respectively, so $\hat{\mathbf{Z}}$ has symbol variance S_x. Moreover, since one of these variables is Gaussian and the other is approximately Gaussian, $\hat{\mathbf{Z}}$ is approximately Gaussian also. Therefore the probability that $\hat{\mathbf{Z}}$ falls outside $\mathcal{R}_V(\Lambda)$, which we may define as the event of *overload*, is arbitrarily small.

If there is no overload, then $\hat{\mathbf{Y}} = \hat{\mathbf{Z}} \bmod \Lambda = \hat{\mathbf{Z}}$. In this case

$$\mathbf{N} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \hat{\mathbf{Z}} = (1-\alpha)\mathbf{Y} - \mathbf{E},$$

so $\mathbf{N}$ is approximately Gaussian with symbol variance

$$(1-\alpha)^2 S_y + S_e = (1-\alpha)S_n + \alpha S_n = S_n.$$

Therefore the mean squared distortion is S_n in the absence of overload, and furthermore the probability of overload is arbitrarily small. Since the squared distortion in case of overload is of the order of $d^2_{\min}(\Lambda)$, the minimum energy of any nonzero point in Λ, we can assert that the mean squared distortion is arbitrarily close to S_n, as desired.

3 Generalizations

We believe that similar techniques involving Voronoi codes and dither (for proofs) may be used generally wherever MMSE principles have been found to be applicable. In general these techniques will yield constructive techniques that attain information-theoretic limits at arbitrary SNRs, if the lattices (or more general Euclidean-space codes) that are used are "good for shaping" or "good for AWGN channel coding," respectively.

For example, Cioffi *et al.* showed in [4, 5, 6] that the minimum-mean-squared-error (MMSE) decision feedback equalizer (DFE) structure is a capacity-achieving receiver structure for intersymbol interference (ISI) channels with Gaussian noise, and furthermore that the same lattice-type codes that are useful for memoryless AWGN channels can be used for ISI channels; for example, the precoding schemes of [15, 23].

The application of MMSE-DFE principles to vector channels with the linear Gaussian model $\mathbf{Y} = H\mathbf{X} + \mathbf{N}$ was made in [6] under the rubric of "generalized decision-feedback equalization" (GDFE). In this setting parallel codes operating at different SNRs must be used in general, and in some cases a kind of successive cancellation precoding must be used in order to remove the effects of "earlier" subchannels from "later" subchannels. Cioffi *et al.* [29, 3], [2, Chaps. 5, 14] have subsequently extended the GDFE approach to various multi-user linear Gaussian channels such as broadcast, multiple-access, and interference channels. In particular, they have solved the long-standing problem of finding the entire capacity region of the Gaussian broadcast channel.

Guess and Varanasi have given an information-theoretic justification of the MMSE-DFE equalizer structure in [20]; see also [27]. They show that this structure converts the original ISI channel to a *memoryless* channel in which the symbol-by-symbol mutual information is preserved. They further note that their result can be extended to other linear Gaussian channels such as the finite-dimensional ISI channel, the multivariate ISI channel, and the synchronous Gaussian multiple-access channel. In a subsequent paper [21], they propose a capacity-achieving structure that combines MMSE-DFE with interleaving so that capacity-approaching codes suitable for a memoryless channel (*e.g.,* lattice-type codes and decoding) may be used.

In their comprehensive survey paper [32], Zamir, Shamai and Erez consider the application of nested lattice (Voronoi) codes with dither to various side information problems, multiresolution and multiterminal source coding, and multiterminal channel coding, as well as to the point-to-point problems considered here. In particular, they propose a precoding version of the Guess-Varanasi [21] structure to achieve the capacity of the ISI channel at all SNRs, while avoiding the "ideal DFE assumption." They also provide an exhaustive annotated bibliography.

In summary, MMSE estimation and lattice-type coding and decoding seem to be fundamental components of constructive techniques that can approach information-theoretic limits in many different linear Gaussian channel scenarios.

Acknowledgments

I am grateful to J. M. Cioffi, U. Erez, R. Fischer and R. Zamir for many helpful comments.

References

[1] R. J. Barron, B. Chen and G. W. Wornell, "The duality between information embedding and source coding with side information, and some applications," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1159–1180, May 2003.

[2] J. M. Cioffi, *Digital Transmission Theory.* Stanford, CA: Stanford EE379c and EE479 course notes, 2003.

[3] J. M. Cioffi, "Dynamic spectrum management," Chap. 11 of *DSL Advances* by T. Starr, M. Sorbara, J. M. Cioffi and P. J. Silverman. Upper Saddle River, NJ: Prentice-Hall, 2003.

[4] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding— Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, pp. 2581–2594, Oct. 1995.

[5] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding— Part II: Coding results," *IEEE Trans. Commun.*, vol. 43, pp. 2595–2604, Oct. 1995.

[6] J. M. Cioffi and G. D. Forney, Jr., "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Communications, Computation, Control and Signal Processing* (A. Paulraj *et al.*, eds.), pp. 79–127. Boston: Kluwer, 1997.

[7] J. H. Conway and N. J. A. Sloane, "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 820–824, 1983.

[8] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.

[9] R. de Buda, "Some optimal codes have structure," *IEEE J. Select. Areas Commun.,* vol. 7, pp. 893-899, Aug. 1989.

[10] P. Elias, "Coding for noisy channels," in *IRE Conv. Rec.*, vol. 3, pp. 37–46, March 1955.

[11] U. Erez, "Coding with known interference and some results on lattices for digital communication," Ph.D. thesis, Dept. Elec. Engg. Sys., Tel-Aviv U., Israel, Dec. 2002.

[12] U. Erez, S. Shamai and R. Zamir, "Capacity and lattice strategies for cancelling known interference," in *Proc. ISITA 2000* (Honolulu, HI), pp. 681–684, Nov. 2000.

[13] U. Erez and R. Zamir, "Lattice decoding can achieve $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel," in *Proc. Int. Symp. Inform. Theory* (Washington, DC), p. 300, June 2001.

[14] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," preprint, July 2003.

[15] M. V. Eyuboglu and G. D. Forney, Jr., "Trellis precoding: Combined coding, shaping and precoding for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. 38, pp. 301–314, Mar. 1992.

[16] M. V. Eyuboglu and G. D. Forney, Jr., "Lattice and trellis quantization with lattice- and trellis-bounded codebooks— High-rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 46–59, Jan. 1993.

[17] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations— Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 877–892, Aug. 1989.

[18] G. D. Forney, Jr., "Multidimensional constellations— Part II: Voronoi constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 941–958, Aug. 1989.

[19] G. D. Forney, Jr., M. D. Trott and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 820–850, May 2000.

[20] T. Guess and M. K. Varanasi, "An information-theoretic derivation of the MMSE decision-feedback equalizer," *Proc. 1998 Allerton Conf.* (Monticello, IL), Sept. 1998.

[21] T. Guess and M. K. Varanasi, "A new successively decodable coding technique for intersymbol interference channels," in *Proc. Int. Symp. Inform. Theory* (Sorrento, Italy), p. 102, June 2000.

[22] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 913–929, May 1993.

[23] R. Laroia, "Coding for intersymbol interference channels— Combined coding and precoding," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1053–1061, July 1996.

[24] T. Linder, C. Schlegel and K. Zeger, "Corrected proof of de Buda's theorem," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1735–1737, Sept. 1993.

[25] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1767–1773, Nov. 1997.

[26] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 409–417, Mar. 1994.

[27] S. Shamai (Shitz) and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and channel precoding loss," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1388–1404, Sept. 1996.

[28] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 273–278, Jan. 1998.

[29] W. Yu and J. M. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," submitted to *IEEE Trans. Inform. Theory*, Nov. 2001.

[30] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1152–1159, July 1996.

[31] R. Zamir and S. Shamai (Shitz), "Nested linear/lattice codes for Wyner-Ziv encoding," *Proc. Inform. Theory Workshop* (Killarney, Ireland), pp. 92–93, June 1998.

[32] R. Zamir, S. Shamai (Shitz) and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250–1276, June 2002.