Reflecting on the AWGN Error Exponent

Charles Swannack[†], Uri Erez[⋆], and Gregory W. Wornell[†]

[†]*Signals, Information and Algorithms Laboratory, MIT, Cambridge MA 02139, USA*
[⋆]*Dept. of Electrical Engineering-Systems, Tel Aviv University, Ramat Aviv 69978, Israel*

Abstract

Recently it was shown that a lattice code with lattice decoding can achieve the capacity of the additive white Gaussian noise (AWGN) channel. This was achieved by using a minimum mean-square error (MMSE) scaling and dithering to transform the AWGN channel into a modulo-lattice additive noise (mod-Λ) channel. Further, Liu et. al. have shown that lattice decoding can achieve the error exponent of the AWGN channel using a scaling other than the MMSE scaling at rates above the critical rate of the channel. We present a simple geometric explanation for this result.

1 Introduction

The additive white Gaussian noise (AWGN) channel was introduced by Shannon in 1948 in his foundational work [1]. In 1959 Shannon further studied lower and upper bounds on the error exponent achieved by codes for the AWGN channel. These bounds, while quite tedious to derive, relied on simple geometric arguments. An alternative derivation of these results, which uses methods developed for general discrete memoryless channels (DMC), was later provided by Gallager in 1965 [2]. This derivation, while much simpler from an analytic standpoint lacked much of the geometry that was contained in Shannon's original work. Further work by Shannon, Gallager and Berlekamp in 1967 [3, 4] provided a tighter upper bound on the reliability function for low rates, which was recently improved upon by Ashikhmin et. al. [5]. The lower and upper bounds coincide for rates above the critical rate R_{cr} of the channel and therefore the error exponent is known for rates $R_{cr} < R < C$.

These works further show that the sphere packing exponent can be achieved for $R \geq R_{cr}$ by random spherical ensembles, *i.e.* by a code all of whose codewords are drawn uniformly over the surface of a sphere. The obtained error exponent assumes optimal maximal likelihood (ML) decoding. The ML decoding region of a codeword consists of all points that lie closer to that codeword than to any other codeword.

A different line of work aimed at developing structured codes for the AWGN channel using lattice codes was initiated by de Buda [6]. Recently, it was shown [7] that the use of lattice codes in conjunction with lattice decoding can achieve capacity on the AWGN channel. One of the key elements in the transmission scheme involves transforming the AWGN channel into an unconstrained modulo-lattice (mod-Λ) additive noise channel, having (asymptotically in the dimension) the same capacity as the original channel. For the resulting channel, if one uses a lattice code Λ_c such that $\Lambda \subset \Lambda_c$, then ML decoding amounts to lattice decoding of Λ_c.

This transformation (described in Section 6) involves pre-subtraction of a random dither known to both the transmitter and receiver. A second key ingredient of the transformation

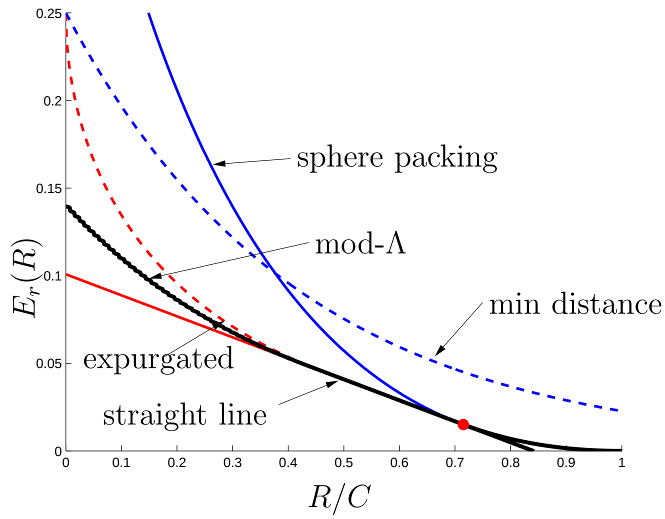

Figure 1: A few known bounds on the error exponent for the AWGN channel (SNR=10 dB)

is the use of a scaling, *i.e. linear estimator* at the receiver. The decoder first scales the received vector by the parameter α and then subtracts off the known dither. It was observed in [7] that using MMSE scaling minimizes the variance of the noise in the resulting mod-Λ channel and results in a channel having the same capacity as the AWGN channel. Thus, MMSE scaling is a natural choice and indeed is unique if one aims for capacity, see [8].

It was further conjectured by the authors of [7] that the mod-Λ transformation, while not losing in capacity, does lose in error exponent. Recently, however, Liu et. al [9] have shown that while MMSE scaling is not sufficient to obtain the error exponent for the mod-Λ channel, a *different scaling* is nonetheless sufficient to obtain the random coding exponent. Through some quite rigorous computation the authors of [9] showed that using a rate dependent scaling, α, the sphere packing exponent can be achieved for $R > R_{cr}$.

The goal of the present work is to provide a simple explanation for this result. We start by analyzing random spherical codes and observe that when working above the critical rate, the mod-Λ channel appears in a natural way when considering typical error events. We provide a simple diagram that provides a relationship between the typical error events in the mod-Λ and AWGN channels via identification of transmitted codewords and a reflection rule leading to a direct interpretation of the optimal parameters from the derivation of Gallager [10] in terms of the mod-Λ channel scaling.

2 Preliminaries and Notation

The discrete-time power-constrained additive white Gaussian noise (AWGN) channel is given by

$$Y = X + N \tag{1}$$

where the input satisfies the power constraint $\frac{1}{n}\|\mathbf{x}\|^2 \leq P$ (n being the blocklength) and where the noise N is a zero mean Gaussian random variable with variance σ^2. The signal-to-noise (SNR) ratio of the channel is $\mathrm{SNR} = P/\sigma^2$. The capacity of the channel is

$$C = \frac{1}{2}\log(1 + \mathrm{SNR}). \tag{2}$$

There are many natural ways to generate random codebook ensembles that asymptotically achieve capacity. Possible choices are: an i.i.d. Gaussian codebook, a codebook drawn uniformly over a sphere, as well as a codebook drawn uniformly over the Voronoi region of a lattice that is "good for quantization" (see [7]). In essence, the codebook distribution should tend to Gaussianity in an entropy sense. That is, its entropy (for a given power) should be close to maximal. The error exponent on the other hand is more sensitive to the input distribution. The error exponent of the AWGN channel for rates above the critical rate is:

$$\begin{aligned} E_{\mathrm{sp}}(R) \triangleq \frac{\mathrm{SNR}}{4e^{2R}} & \left(1 + e^{2R} - \left(e^{2R}-1\right)\sqrt{1+\frac{4e^{2R}}{\mathrm{SNR}(e^{2R}-1)}}\right. \\ & \left. + \frac{1}{2}\log\left(e^{2R} - \frac{\mathrm{SNR}(e^{2R}-1)}{2}\left(\sqrt{1+\frac{4e^{2R}}{\mathrm{SNR}(e^{2R}-1)}}-1\right)\right)\right) \end{aligned} \tag{3}$$

As shown in [2,11] this exponent may be achieved by a codebook drawn uniformly over the surface of a sphere[1]. We note that a Gaussian codebook does not achieve the channel's error exponent. We see below that the distinction between Gaussian and spherical distributions shows up in a dual way in the mod-Λ channel.

We denote the codebook by $\mathcal{C}$, the transmitted codeword as $\mathbf{c}$ and any other codeword as $\mathbf{c}_e$. The received vector is $\mathbf{y} = \mathbf{c} + \mathbf{z}$. Under maximum likelihood decoding we have that an error occurs when

$$\|\mathbf{y} - \mathbf{c}_e\| \leq \|\mathbf{y} - \mathbf{c}\| \tag{4}$$

for some other codeword $\mathbf{c}_e$. That is, the error probability given that the message $\mathbf{c}$ is transmitted, $\mathrm{P}_e(\mathbf{c})$, is simply

$$\mathrm{P}_e(\mathbf{c}) = \Pr\left\{\|\mathbf{y} - \mathbf{c}_e\| \leq \|\mathbf{y} - \mathbf{c}\|, \quad \text{for some } \mathbf{c}_e \in \mathcal{C}\right\}. \tag{5}$$

We denote by $\bar{P}_e$, the average error probability, averaged over all codewords as well as the codebook ensemble.

We use the following method due to Gallager [12] to bound the probability of decoding error. Let $\mathcal{R}$ be any region in $\mathbb{R}^n$. Then for any transmitted codeword $\mathbf{c}$ we may consider separately the probability of error when the received vector is in $\mathcal{R}$ or in its compliment. When the received vector, $\mathbf{y}$ is in $\mathcal{R}$, we upper bound the probability of error by using a union bound over all codewords in the codebook. When $\mathbf{y}$ is not in $\mathcal{R}$, we upper bound the probability of error by 1. More formally, we may in general write

$$\begin{aligned} \mathrm{P}_e(\mathbf{c}) &= \Pr\left(\text{error}, \mathbf{c}+\mathbf{z} \in \mathcal{R}\right) + \Pr\left(\text{error}, \mathbf{c}+\mathbf{z} \notin \mathcal{R}\right) && (6) \\ &\leq \Pr\left(\text{error}, \mathbf{c}+\mathbf{z} \in \mathcal{R}\right) + \Pr\left(\mathbf{c}+\mathbf{z} \notin \mathcal{R}\right) && (7) \\ &\leq \sum_{\mathbf{c}_e \neq \mathbf{c}} \Pr\left(\|\mathbf{y}-\mathbf{c}_e\| \leq \|\mathbf{y}-\mathbf{c}\|, \mathbf{c}+\mathbf{z} \in \mathcal{R}\right) + \Pr\left(\mathbf{c}+\mathbf{z} \notin \mathcal{R}\right) && (8) \\ &\triangleq \mathrm{P}_{\text{union}}(\mathbf{c}) + \mathrm{P}_{\text{region}}(\mathbf{c}) && (9) \end{aligned}$$

where $\mathrm{P}_{\text{region}}(\mathbf{c})$ is the probability that the received vector is not in the region $\mathcal{R}$ and $\mathrm{P}_{\text{union}}(\mathbf{c})$ is the sum appearing in (8). Averaging over the code and the ensemble of codes we have $\bar{P}_e \leq \overline{\mathrm{P}_{\text{union}}(\mathbf{c})} + \mathrm{P}_{\text{region}}(\mathbf{c}) = \overline{\mathrm{P}_{\text{union}}} + \mathrm{P}_{\text{region}}$ (since by averaging the probability of error is independent of the codeword).

[1] In [2] Gallager starts with a Gaussian distribution but applies expurgation to the same effect.

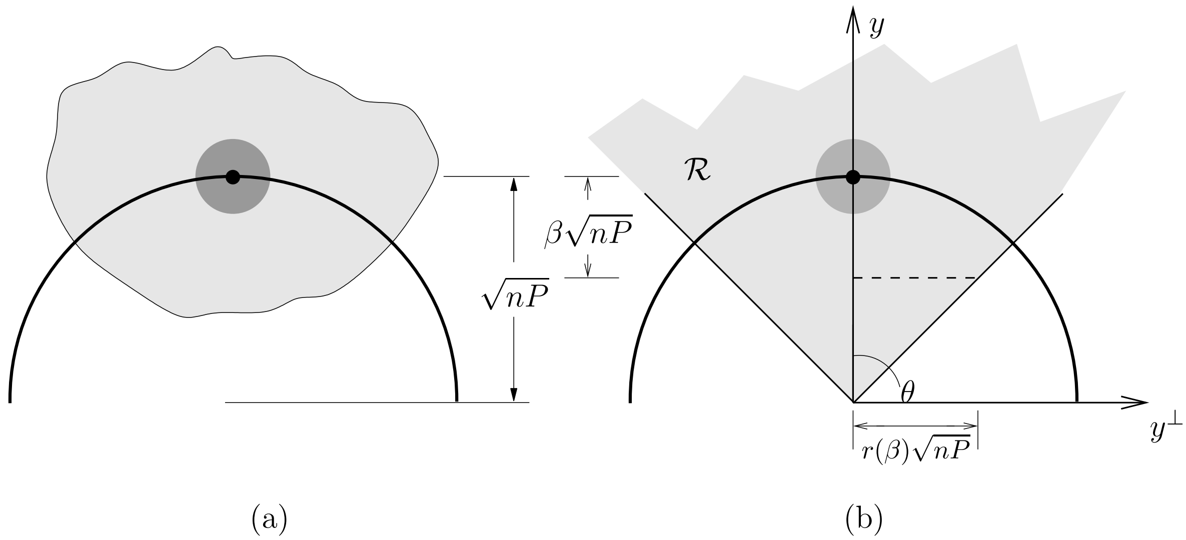

Figure 2: A depiction of the general bounding technique (9). (a) The general bound for an arbitrary region $\mathcal{R}$ and (b) the bounds when specialized to a cone of half angle θ

It turns out that with a proper choice of the region $\mathcal{R}$, which is described in the next section, this bound is tight enough to obtain the best known lower bounds on the error exponent. The region $\mathcal{R}$ can be arbitrary for the above bound to hold. However, for a random spherical ensemble there is no loss in restricting $\mathcal{R}$ to be rotationally symmetric about the axis that passes through the origin and the codeword. For the remainder of this paper we will assume that $\mathcal{R}$ is rotationally symmetric.

3 Geometric Derivation of Sphere Packing Exponent

In the case of the constrained AWGN channel we take, owing much to Shannon [11], the region $\mathcal{R}$ to be the cone[2] of half angle θ with apex at the origin and whose axis passes through $\mathbf{c}$, where $\sin\theta = \exp(-R)$. We denote this region as $\mathcal{R}_c(\theta)$. With this choice of angle, the cone has the same (in an exponential sense) solid angle as the average solid angle of an ML decoding region. It is easy to show [11] that the probability of a noise vector leaving the ML decoding region is greater than that of leaving a cone with the same solid angle. It therefore follows that the average probability of error satisfies,

$$\bar{P}_e \dot{\geq} \Pr\left(\mathbf{c} + \mathbf{z} \notin \mathcal{R}_c(\theta)\right).$$

Thus, in order to show that the sphere packing bound is tight it is sufficient to show that

1. $\Pr\left(\mathbf{c} + \mathbf{z} \notin \mathcal{R}_c(\theta)\right) \doteq e^{-nE_{sp}(R)}$.

2. $\overline{P_{\text{union}}} \dot{\leq} P_{\text{region}}$ for $R > R_{cr}$.

Property 2 is easy to prove using the distribution of a random spherical code ensemble, see [13]. We next compute $\Pr\left(\mathbf{c} + \mathbf{z} \notin \mathcal{R}_c(\theta)\right)$ and show that Property 1 holds.

[2]This form of Gallager's bound corresponds to Poltyrev's tangential sphere bound.

Let us decompose the noise into a component normal to the the surface of the sphere at $\mathbf{c}$, say z_y, and its orthogonal complement $\mathbf{z}_{y^\perp}$ (see Figure 2). Then,

$$\mathbf{z} = z_y \cdot \mathbf{e}_y + \mathbf{z}_{y^\perp}$$

where $\mathbf{e}_y$ is the unit vector normal to the sphere at $\mathbf{c}$. Since the components of the noise are independent we may condition on the component z_y and integrate over the distribution of that component. Let $y = \beta\sqrt{nP}$ be the magnitude of the normal component of the noise and let $r(\beta)\sqrt{nP}$ be the radius of the corresponding spherical cross section of the cone, as shown in Figure 2. By simple geometry, we have $r(\beta) = (1+\beta)\tan\theta$. Then,

$$\Pr(\mathbf{c}+\mathbf{z} \notin \mathcal{R}_c) = \int_{-1}^{\infty} f_{z_y}(\beta\sqrt{nP}) \cdot \Pr\left(\|\mathbf{z}_{y^\perp}\| \geq r(\beta)\sqrt{nP}\right) d\beta$$

Using a Chernoff bound on the norm of a Gaussian vector whose components each have a variance σ^2 we have,

$$\Pr\left(\|\mathbf{z}\| \geq r(\beta)\sqrt{nP}\right) \doteq \exp\left(-n\, E_h(r(\beta)^2 P/\sigma^2)\right) \tag{10}$$

where $E_h(\mu) \stackrel{\triangle}{=} \frac{1}{2}(\mu - 1 - \log\mu)$ if $\mu \geq 1$ and zero otherwise[3] Additionally, we have for a one dimensional Gaussian

$$\Pr\left(z \geq \beta\sqrt{nP}\right) \doteq \exp\left(-n\, E_v(\beta^2 P/\sigma^2)\right) \quad \text{where} \quad E_v(\mu) = \mu/2$$

Thus, the probability that the received vector is outside the cone satisfies

$$\Pr\left(\mathbf{c}+\mathbf{z} \notin \mathcal{R}_c\right) \doteq \exp\left(-n \min_{\beta} \left[\, E_v(\beta^2 \mathrm{SNR}) + E_h(r(\beta)^2\, \mathrm{SNR})\right]\right) \tag{11}$$

Finding β that minimizes the exponent of (11) we find that for $R > R_{cr}$,

$$1 + \beta^*(\mathrm{SNR}, \theta) = \frac{\cos^2\theta}{2} + \frac{\cos\theta}{2}\sqrt{\cos^2\theta + \frac{4}{\mathrm{SNR}}} \tag{12}$$

Further, with some simple arithmetic one can show that indeed

$$E_{\mathrm{sp}}(R) = E_v((\beta^*)^2 \mathrm{SNR}) + E_h(r(\beta^*)^2\, \mathrm{SNR})$$

so that Property 1 holds and the sphere packing exponent is the proper exponent for $R > R_{cr}$. Additionally, for all rates above the critical rate the error event is dominated by the event of leaving a cone at a height of $\beta^*\sqrt{nP}$. However, the probability for leaving the cone at any other height is in general *smaller* than at β^*. It is natural to ask what other regions one can use in order to derive the sphere packing bound using Gallager's bounding technique (9).

[3] $E_h(\mu)$ is Poltyrev's exponent in the sphere packing region.

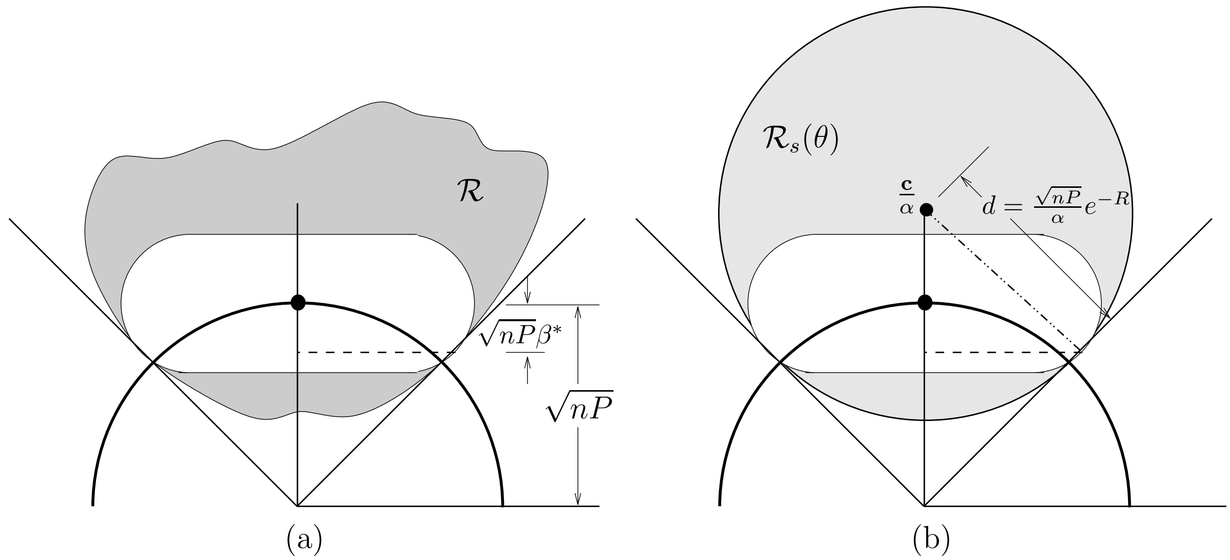

Figure 3: A depiction of "valid" regions for the general bounding technique (9) (a) The general bound for an arbitrary "valid" region. A general region $\mathcal{R}$ and a depiction of the smallest valid region. (b) the bounds when specialized to a sphere that contains the smallest region.

4 Valid Regions for Geometric Derivation

The tightness of the bound we obtained using the cone $\mathcal{R}_c(\theta)$ means that for rates above the critical rate, the error probability of leaving an ML decoding region is exponentially the same as that of leaving the cone $\mathcal{R}_c(\theta)$. Thus, the cone approximates[4] the ML decoding region in the error probability analysis for $R > R_{cr}$. We now investigate how much freedom we have in choosing a region that has this property.

We will refer to any region $\mathcal{R}$ that yields the sphere packing bound as valid. We would like to find the "smallest" valid region. It is clear that any region satisfying the following is valid:

1. $\mathcal{R} \subset \mathcal{R}_c(\theta)$.
2. $\Pr\left(\mathbf{c} + \mathbf{z} \notin \mathcal{R}\right) \doteq \Pr\left(\mathbf{c} + \mathbf{z} \notin \mathcal{R}_c(\theta)\right)$.

The first condition ensures that $\overline{P_{\text{union}}}$ will be no greater for $\mathcal{R}$ than for $\mathcal{R}_c(\theta)$. The second condition guarantees that P_{region} remains exponentially the same for $\mathcal{R}$ as it was for $\mathcal{R}_c(\theta)$.

As previously noted, the cross section of the cone that dominates the error event is that corresponding to β^*. This means that at other heights we should be able to choose a "narrower" geometrical body. Nonetheless, at a height β^* the radius of $\mathcal{R}$ has to coincide with that of $\mathcal{R}_c(\theta)$ since we cannot hope to improve the sphere packing bound. Thus, $\mathcal{R}$ must be tangent to $\mathcal{R}_c(\theta)$ at β^*. Further, it is necessary that for any valid region and for any β we have

$$f_{z_y}(\beta) \Pr\left(\mathbf{c} + \mathbf{z} \notin \mathcal{R} \mid z_y = \beta\sqrt{nP}\right) \dot{\leq} f_{z_y}(\beta^*) \Pr\left(\mathbf{c} + \mathbf{z} \notin \mathcal{R}_c \mid z_y = \beta^*\sqrt{nP}\right) \tag{13}$$

[4]It is important to note that the tightness of the sphere packing bound does not imply the existence of good cone packings, and in fact these are known not to exist.

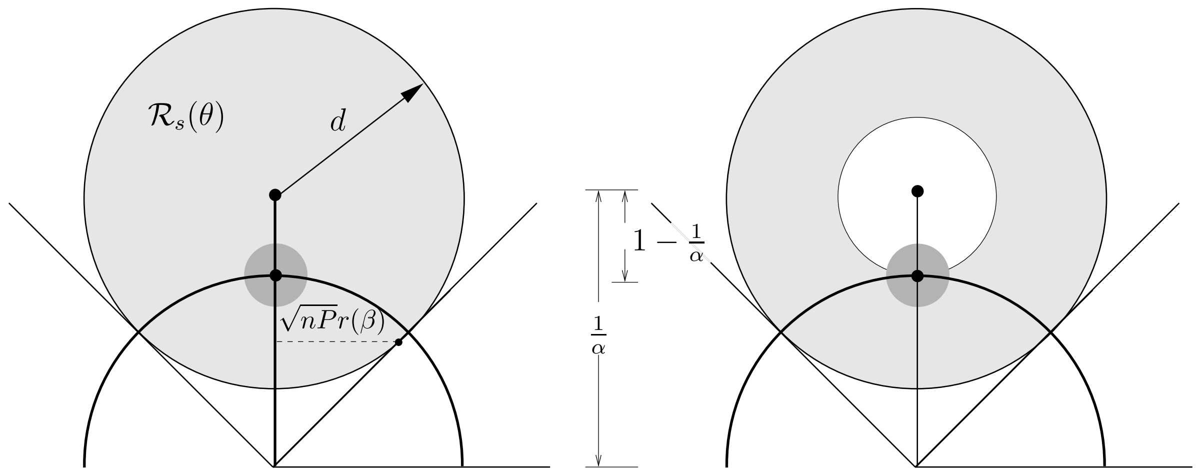

Figure 4: The derivation of the sphere packing exponent using a spherical region. (a) The probability of a Gaussian leaving a sphere whose center is located at $1/\alpha$ and tangent to a cone of half angle θ. (b) The equivalence between this and the probability that a spherical noise plus a Gaussian leaves a larger sphere

Now, consider the region that exactly meets (13) for every β. That is, the region that is parameterized by

$$E_v(\beta^2\mathrm{SNR}) + E_h(r(\beta)^2\,\mathrm{SNR}) = E_{\mathrm{sp}}(R,\mathrm{SNR})$$

This region is the smallest valid region, since the probability of leaving any other valid region is exponentially larger. A depiction of the smallest region can be seen contained in a general region $\mathcal{R}$ in Figure 3 (a) (all are tangent to the cone at β^*).

5 Geometric Derivation Using Spherical Regions

We now consider the possibility of taking $\mathcal{R}$ to be a sphere. From the previous section we know that a valid sphere must be tangent to the cone of half angle θ at β^* and also contain the smallest region. In order to make the sphere tangent at β^* we can simply draw a line perpendicular to the cone at β^* and find the point where this line intersects the line passing through the origin and the transmitted codeword (see Figure 3). We will denote this point as $\mathbf{c}/\alpha^*$. Using some simple geometry we find that

$$\frac{1}{\alpha^*} = \frac{1+\beta^*}{\cos^2\theta} = \frac{1}{2}\left(1+\sqrt{1+\frac{4}{\mathrm{SNR}\cos^2\theta}}\right) \tag{14}$$

and that the radius of this sphere is $d = \sqrt{nP}/\alpha^*\sin\theta$. Denote the sphere which is a distance l from the origin and has a radius d by $\mathcal{R}_s(l,d)$. Then, if a valid sphere exists it must be the sphere just defined

$$\mathcal{R}_s(\theta) \stackrel{\triangle}{=} \mathcal{R}_s(\sqrt{nP}/\alpha^*, \sqrt{nP}/\alpha^*\sin\theta).$$

It remains to show that $\mathcal{R}_s(\theta)$ satisfies Property 2, *i.e.* that it contains the smallest valid region. This is verified in [13] by using the same method as used for $\mathcal{R}_c(\theta)$.

It is important to note that α^* is simply Gallager's $1+\rho$ scaled by e^{-R}/SNR. Furthermore, α^* is the optimal scaling found in [9] for the mod-Λ channel. Before explaining the connection

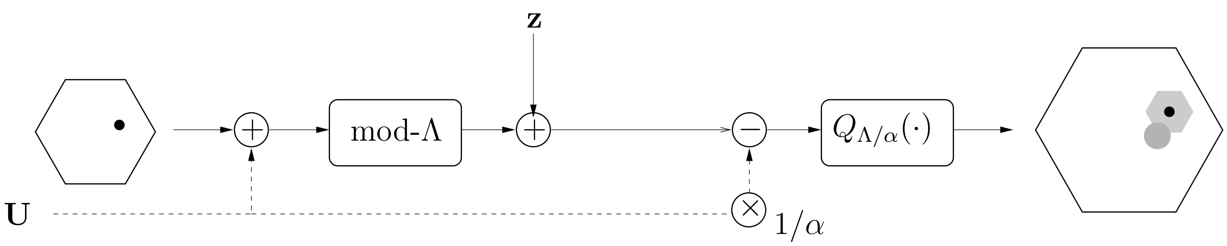

Figure 5: The mod-Λ channel diagram. A code word is chosen at random from the code book $\mathcal{C}$ and a random dither $\mathbf{u}$ is added and the result (mod-Λ) is transmitted through the additive noise channel. An estimate of the received signal is then formed, the dither subtracted and quantized (decoded) to the nearest point.

to the mod-Λ channel we make the following observation. Note that we may rewrite the received vector as

$$\mathbf{y} = \mathbf{c} + \mathbf{z} = \frac{\mathbf{c}}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\mathbf{c} + \mathbf{z} = \frac{\mathbf{c}}{\alpha} + \mathbf{w}$$

where $\mathbf{w}$ is the Gaussian vector with mean $(1 - 1/\alpha)\,\mathbf{c}$. We can think of transmitting our inflated codeword by adding a deterministic vector that enables us to meet our power constraint and then adding the Gaussian noise. Thus, the sphere packing exponent has a natural interpretation as the exponent of the probability that the scaled codeword plus a deterministic translation and a Gaussian noise remains in a sphere about the scaled codeword. Alternatively by spherical symmetry this is equivalent to the probability that a random variable that is uniform over the surface of a sphere plus a Gaussian remains in a sphere about the scaled codeword. That is,

$$\exp(-nE_{sp}(R)) \doteq \Pr\left(\frac{\mathbf{c}}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\mathbf{c} + \mathbf{z} \notin \mathcal{R}_s(\theta)\right) \doteq \Pr\left(\frac{\mathbf{c}}{\alpha} + \left[\frac{1-\alpha}{\alpha}\mathbf{b} + \mathbf{z}\right] \notin \mathcal{R}_s(\theta)\right) \tag{15}$$

where $\mathbf{b}$ is the random variable that is uniform over the ball of radius $\sqrt{nP}$. This equivalence is depicted in Figure 4. We make the final connection to the error probability in the mod-Λ channel after first briefly reviewing this transmission approach.

Remark: *This derivation has a natural interpretation as the error probability for the AWGN channel using bounded distance decoding of an inflated codebook. That is, the decoder which decodes to the codeword $\mathbf{c}$ if it is the nearest codeword and less than a distance d from the inflated codeword $\mathbf{c}/\alpha$, where $0 < \alpha \leq 1$. That is, the received vector is decoded to $\mathbf{c}$ if and only if for all $\mathbf{c}_e \in \mathcal{C}$*

$$\|\mathbf{y} - \mathbf{c}\| > \|\mathbf{y} - \mathbf{c}_e\| \quad \textit{and} \quad \left\|\mathbf{y} - \frac{\mathbf{c}}{\alpha}\right\| \leq d \tag{16}$$

Such a decoder corresponds to this method of bounding since an error is declared if the noise is not contained in a spherical region about the scaled codebook. This result means that there is no penalty in the error exponent for using such a region for $R > R_{cr}$ if one uses the optimal scaling.

6 Modulo Lattice Additive Noise Channel

We briefly review the lattice transmission approach proposed in [7]. Let Λ be an n-dimensional lattice whose fundamental Voronoi region $\mathcal{V}$ has second moment P (normalized per dimension). Let $\mathbf{u}$ be the random variable (dither) uniformly distributed over $\mathcal{V}$. The mod-Λ transmission scheme is given by,

- *Transmitter:* The input alphabet is restricted to $\mathcal{V}$. For any $\mathbf{v} \in \mathcal{V}$, the encoder sends:
$$\mathbf{x} = [\mathbf{v} + \mathbf{u}] \bmod \Lambda. \tag{17}$$
- *Receiver:* The receiver computes
$$\mathbf{y}' = \left[\mathbf{y} - \frac{1}{\alpha} \cdot \mathbf{u}\right] \quad \bmod \Lambda/\alpha. \tag{18}$$

The resulting channel is a modulo lattice additive noise channel described by the following lemma [14] :

Lemma 1 *The channel from* $\mathbf{v}$ *to* $\mathbf{y}'$ *defined by (1),(17) and (18) is equivalent in distribution to the channel*
$$\mathbf{y}' = \left[\frac{1}{\alpha} \cdot \mathbf{v} + \mathbf{z}'\right] \quad \bmod \Lambda/\alpha \quad \textit{with} \quad \mathbf{z}' = \frac{1-\alpha}{\alpha} \cdot \mathbf{u} + \mathbf{z} \tag{19}$$

Taking an MMSE scaling $\alpha = \frac{SNR}{1+SNR}$ results in a mutual information satisfying
$$I(\mathbf{v};\mathbf{y}) \geq \frac{1}{2}\log(1+\mathrm{SNR}) - \frac{1}{2}\log 2\pi e G(\Lambda) \quad \text{where} \quad G(\Lambda) = \frac{1}{n}\frac{\int_{\mathcal{V}} \|\mathbf{x}\|^2 d\mathbf{x}}{|V|^{1+2/n}} \tag{20}$$

is the normalized second moment of Λ. Thus, the gap to capacity may be made arbitrarily small by taking a lattice Λ such that $G(\Lambda)$ is sufficiently close to $1/(2\pi e)$.

We observe that the noise $\mathbf{z}$ looks very much like the effective noise appearing in (15). Note that the random vectors $\mathbf{b}$ and $\mathbf{u}$ have the same second moment but while $\mathbf{b}$ is spherical, $\mathbf{u}$ is uniform over $\mathcal{V}$. Thus, one could hope that if $\mathcal{V}$ is "spherical enough" then it could be approximated well by $\mathbf{b}$. That is, define the channel
$$\mathbf{y}'' = \left[\frac{1}{\alpha} \cdot \mathbf{v} + \mathbf{z}''\right] \quad \bmod \Lambda/\alpha \quad \text{with} \quad \mathbf{z}'' = \frac{1-\alpha}{\alpha} \cdot \mathbf{b} + \mathbf{z} \tag{21}$$

We would like to have Λ such that the error exponents of the channels (19) and (21) are the same. It is shown in [7] that such lattices do indeed exist and following [7] we refer to such a lattice as "Rogers' good". Such lattices have Voronoi regions that are close to spherical in a much stronger sense than just having a good normalized second moment. This necessity is similar to the fact that a spherical codebook is superior to a Gaussian codebook in the power constrained AWGN channel.

We may now bound the error exponent of the channel (21) by using Gallager's technique (9) as before. It is easy to see that P_{region} is precisely (15). Further, it is easy to show that, as before, the union bound P_{union} is exponentially no greater than P_{region} for $R > R_{cr}$ [13]. Thus, it follows that the mod-Λ scheme may achieve the sphere packing error exponent for these rates.

Returning to the last section we can easily see by identifying transmitted codewords of the AWGN and mod-Λ channels that the error exponents for $R > R_{cr}$ are the same. Examining Figure 4 (b) we can interpret the center of $\mathcal{R}_s(\theta)$, $\mathbf{c}/\alpha$, as the selected codeword in the mod-Λ channel and $\mathcal{R}_s(\theta)$ as the Voronoi of the codeword. Thus, the dither simply translates $\mathbf{c}/\alpha$ to a corresponding codeword of a spherical code. Further, to achieve the error exponent in the mod-Λ channel it is essential that the self noise be virtually spherical and not Gaussian, much like the distribution of codewords in the AWGN channel. This analysis also can be extended to low rates [13] giving a similar interpretation to the results of Liu [9].

References

[1] C. Shannon, "A mathematical theory of communication," *BSTJ*, pp. 379–423,623–656, 1948.

[2] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. IT-11, pp. 3–18, January 1965.

[3] C. Shannon, R. G. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inform. Contr.*, vol. 10, pp. 65–103, January 1967.

[4] ——, "Lower bounds to error probability for coding on discrete memoryless channels. II," *Inform. Contr.*, vol. 10, pp. 522–552, May 1967.

[5] A. Ashikhmin, A. Barg, and S. N. Litsyn, "A new upper bound on the reliability function of the Gaussian channel," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 1945–1961, September 2000.

[6] R. de Buda, "Some optimal codes have structure," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 6, August 1989.

[7] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, October 2004.

[8] G. Forney, Jr., "On the role of mmse estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. 41nd Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, Illinois, 2003.

[9] T. Liu, P. Moulin, and R. Koetter, "On error exponents of nested lattice codes for the AWGN channel," *IEEE Transactions on Information Theory*, 2004, submitted.

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

[11] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, May 1959.

[12] R. G. Gallager, "Low density parity check codes," Ph.D. dissertation, MIT, Cambridge, Massachusetts, 1962.

[13] C. Swannack, U. Erez, and G. W. Wornell, "Reflecting on the AWGN error exponent," *IEEE Transactions on Information Theory*, 2005, in preparation.

[14] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," in *Proc. ISITA 2000*, November 2000, pp. 681–684.