MMSE estimation and lattice encoding/decoding for linear Gaussian channels

Todd P. Coleman

6.454 9/22/02

Background: the AWGN Channel

$$Y = X + N$$

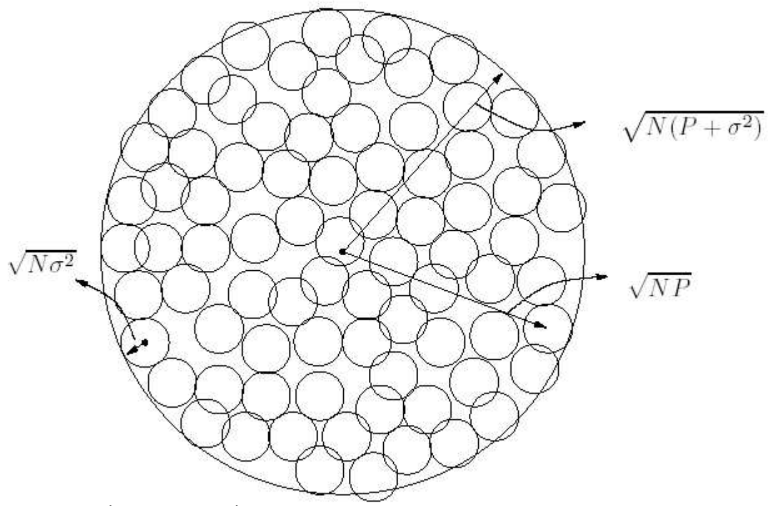
where $N \sim \mathcal{N}\left(0, \sigma_N^2\right), \frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P_X$.

- Shannon: capacity is

$$C = \frac{1}{2}\log_2(1 + SNR), \ SNR = \frac{P_X}{\sigma_N^2}$$

- Random coding argument: generate 2^{nC} i.i.d. $\mathcal{N}(0, P_X)$ codewords. Averaging across all codebooks: under ML decoding $P(e) \to 0$ as $n \to \infty$.

Geometrically Achieving Capacity



- LLN: $\mathcal{N}\left(0,\sigma^2\right)$ i.i.d. n-vector lies in sphere of radius $\sqrt{n\sigma^2}$.
- $X \sim \mathcal{N}(0,P_X), N \sim \mathcal{N}\left(0,\sigma_N^2\right) \Rightarrow Y \sim \mathcal{N}\left(0,P_X+\sigma_N^2\right)$. $\Rightarrow Y$ lies in sphere of radius $\sqrt{n(P_X+\sigma_N^2)}$.
- Codewords chosen as centers of non-overlapping spheres w/ radius $\sqrt{n\sigma_N^2}$
- Volume of n n-sphere w/ radius r is $A_n r^n$.
- $\Rightarrow$ max no. of non-overlapping decoding spheres:

$$\frac{A_n\left[n(P_X+\sigma_N^2)\right]^{\frac{n}{2}}}{A_n\left[n\sigma_N^2\right]^{\frac{n}{2}}} = 2^{\frac{n}{2}\log_2\left(1+\frac{P_X}{\sigma_N^2}\right)} = 2^{nC}$$

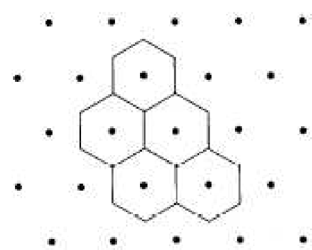
Structured Coding for AWGN channels

- Researchers for decades interested in *structured* codes, encoding mechanisms, and decoding mechanisms
- Desire: achieve capacity on the AWGN channel for *arbitrary* SNRs.
- Devote our attention to lattices: algebraic in nature.
- **Basis of today's talk:** How Uri Erez and Ram Zamir solved the decades-old problem of achieving the AWGN channel capacity at all SNRs, using lattice codes and lattice decoding.
- Surprisingly and non-so-intuitive at first glance:
  - using a *biased* MMSE estimator at the decoder is essential to achieve capacity.
  - related to deep connection between mutual information and MMSE estimation (Baris's talk in a couple weeks).

Lattices

- Lattice: a discrete group which is a subset of $\mathbb{R}^n$. Described in terms of a generator matrix:

$$\Lambda = \{\lambda = Gx : x \in \mathbb{Z}^n\}, \; G \in \mathbb{R}^{n\times n}$$



- Fundamental Voronoi region of Λ:

$$\mathcal{V} = \{x \in \mathbb{R}^n | \|x - 0\| \leq \|x - \lambda\| \;\; \forall \, \lambda \in \Lambda\}.$$

- Any $x \in \mathbb{R}^n$ uniquely expressed as

$$\begin{aligned} x &= \lambda + r, \text{ where } \lambda \in \Lambda, r \in \mathcal{V} \\ &= Q_{\mathcal{V}}(x) + x \bmod_{\mathcal{V}} \Lambda \,. \end{aligned}$$

  $\mathcal{V}$ analogous to remainder in modular arithmetic.

- Generally: any fundamental region Ω satisfies $x \in \mathbb{R}^n$ uniquely expressed as

$$\begin{aligned} x &= \lambda + r, \text{ where } \lambda \in \Lambda, r \in \Omega \\ &= Q_{\Omega}(x) + x \bmod_{\Omega} \Lambda \,. \end{aligned}$$

Desired Properties of Good Lattices

- Denote volume of any $\mathcal{R} \subset \mathbb{R}^n$ as $V(\mathcal{R})$.
- 2nd moment per dim. of $\mathcal{R}$:

$$P(\mathcal{R}) = \frac{1}{n} \frac{\int_{\mathcal{R}} \|x\|^2 dx}{V(\mathcal{R})}$$

  - Avg energy per dim. of $U \sim \text{unif}(\mathcal{R})$.
- Normalized 2nd moment of $\mathcal{R}$:

$$G(\mathcal{R}) = \frac{P(\mathcal{R})}{V(\mathcal{R})^{\frac{2}{n}}}$$

- S_{n,σ^2}: the n-sphere with radius $\sqrt{n\sigma^2}$.

$$\begin{aligned} a) \quad & V(S_{n,\sigma^2})^{\frac{2}{n}} \to 2\pi e \sigma^2,\ P(S_{n,\sigma^2}) \to \sigma^2, \\ & \Rightarrow G(S_{n,\sigma^2}) \to \frac{1}{2\pi e} \\ b) \quad & \sigma_N^2 < \sigma^2 \Rightarrow P\left([X \sim \mathcal{N}\left(0, \sigma_N^2\right)] \notin S_{n,\sigma^2} \right) \to 0. \end{aligned}$$

- Λ_S 'good for shaping' if a):

$$G(\mathcal{V}_S) \to \frac{1}{2\pi e}.$$

- Λ_C 'good for channel coding' if b):

$$\sigma_N^2 < \frac{V(\mathcal{V}_C)^{\frac{2}{n}}}{2\pi e} \Rightarrow P\left([X \sim \mathcal{N}\left(0, \sigma_N^2\right)] \notin \mathcal{V}_C \right) \to 0.$$

Lattice Codes

- *Lattice code* $\mathcal{C}$:

$$C = \Lambda_C \cap \mathcal{S}.$$

Shaping region $\mathcal{S}$ imposes signaling constraint (such as power constraint for AWGN channel).

- *Lattice decoder* $\mathcal{C}$: simply a quantizer $Q_{\Omega_C}(x)$ for Λ_C. Performs the operation

$$\lambda = Q_{\Omega_C}(y) \in \Lambda_C.$$

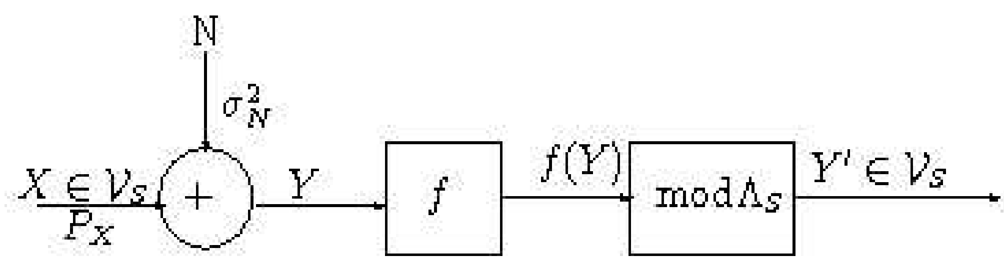
Note the decoder does not take into account the shaping region $\mathcal{S}$ associated with the lattice code, which simplifies the decoding process.

Previous Work on Lattice Codes

- De Buda considered a spherical lattice code where $\mathcal{S}$ is a sphere and is Λ_C 'good for channel coding'
- Numerous authors: $\mathcal{S}$ should be a thin spherical shell. Under ML decoding, the capacity is achieved.
  - **But** ML decoding requires finding the lattice point closest to the received signal *inside the shell* .
  - Decoding regions lose structure, have no relation to true lattice decoding.
- A spherical lattice code with a Euclidean minimum-distance decoder can achieve $\frac{1}{2}\log_2(SNR)$.
  - At high SNR, this essentially achieves capacity.
  - At low SNR, significant performance loss. We will discuss why 1 is missing here later.

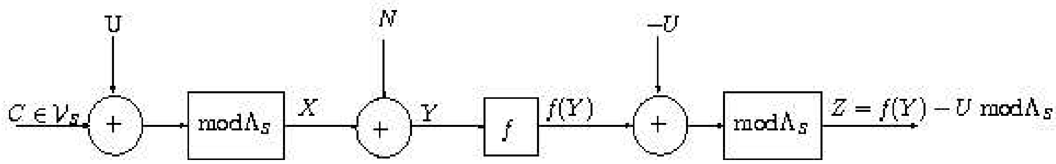
Mod-Lattice Transmission and Lattice Decoding

- Now temporarily step away from 'good for channel coding' codes Λ_C and consider Λ_S that is 'good for shaping'.
- **Desire**: $\mathcal{V}_S$ will serve as $\mathcal{S}$ and allow more structured encoding/decoding.



If Λ_S is 'good for shaping' ($G(\mathcal{V}_S) \to \frac{1}{2\pi e}$), $X \sim$ unif$(\mathcal{V}_S)$, and f an MMSE estimator of X, then $\frac{1}{2}\log_2(1 + SNR)$ is achievable.

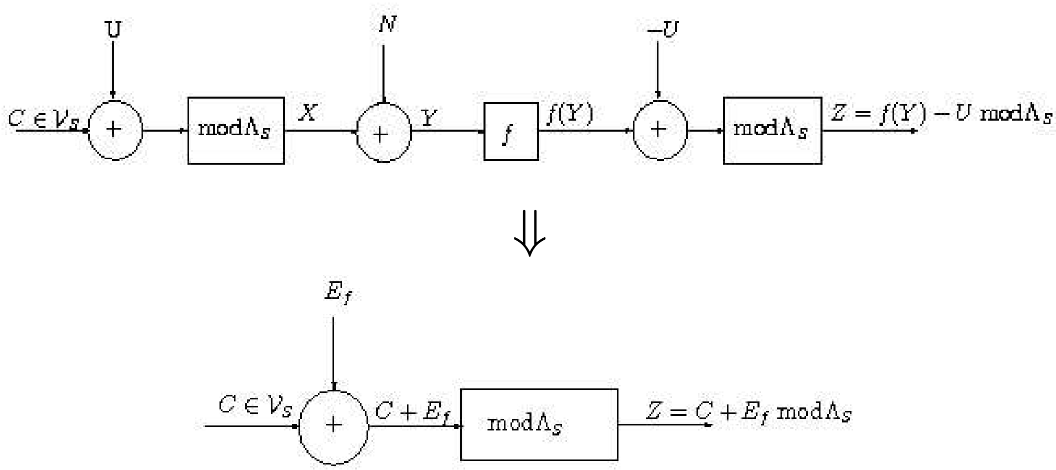
Mod-Lattice Transmission and Lattice Decoding (Cont'd)



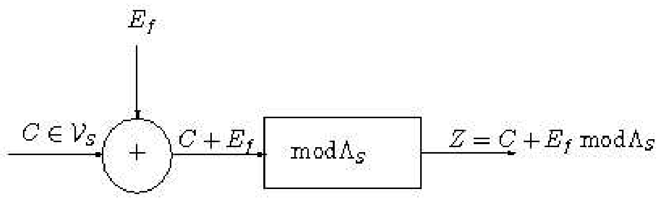
- Introduce *dither* $U \sim$ unif$(\mathcal{V}_S)$, known to both the encoder and decoder.
- Given any $C \in \mathcal{V}_S$, the channel input is $X = C + U$ modΛ_S .
- $\Rightarrow$ $X \sim$ unif$(\mathcal{V}_S)$ *and* $X \perp C$.
  Why: $P_U(u)$ constant $\forall\, u \in \mathcal{V}_S$.
  As $x \nearrow \mathcal{V}_S$, $x - c$ modΛ_S $\nearrow \mathcal{V}_S$.
  $\Rightarrow P_{X|C}(x|c) = P_U(x - c$ modΛ_S $)$,
  constant $\forall\, x \in \mathcal{V}_S, c \in \mathcal{V}_S$.

Mod-Lattice Transmission and Lattice Decoding (Cont'd)

- Dither contributes 2 nice things:
  - $X \sim \text{unif}\,(\mathcal{V}_S)$,
    $\Rightarrow$ power constraint met with equality.
  - $X \perp C$; $C \leftrightarrow X \leftrightarrow Y \Rightarrow (Y, X) \perp C$.
- $\Rightarrow E_f = f(Y) - X \;\perp C$.
- $Z = C + E_f \bmod \Lambda_S$.
  $\Rightarrow$ now an additive noise channel:



Equivalent Channel Model



$$
\begin{aligned}
C &\sim \text{unif}\,(\mathcal{V}_S) \ \text{ optimal } \ \Rightarrow Z \sim \text{unif}\,(\mathcal{V}_S)\,. \\
E'_f &\triangleq E_f \ \text{mod}\Lambda_S\ . \\
\mathbf{C} &\geq \mathbf{C}(\Lambda_S, f) = \frac{1}{N}\left[h(Z) - h(Z|C)\right] \\
&= \frac{1}{N}\left[\log_2 V(\Lambda_S) - h(E'_f)\right] \\
&= \frac{1}{2}\log_2 2\pi e P_X - \frac{1}{2}\log_2 2\pi e G(\mathcal{V}_S) - \frac{1}{N}h(E'_f) \\
&\geq \frac{1}{2}\log_2 2\pi e P_X - \frac{1}{2}\log_2 2\pi e G(\mathcal{V}_S) - \frac{1}{N}h(E_f)
\end{aligned}
$$

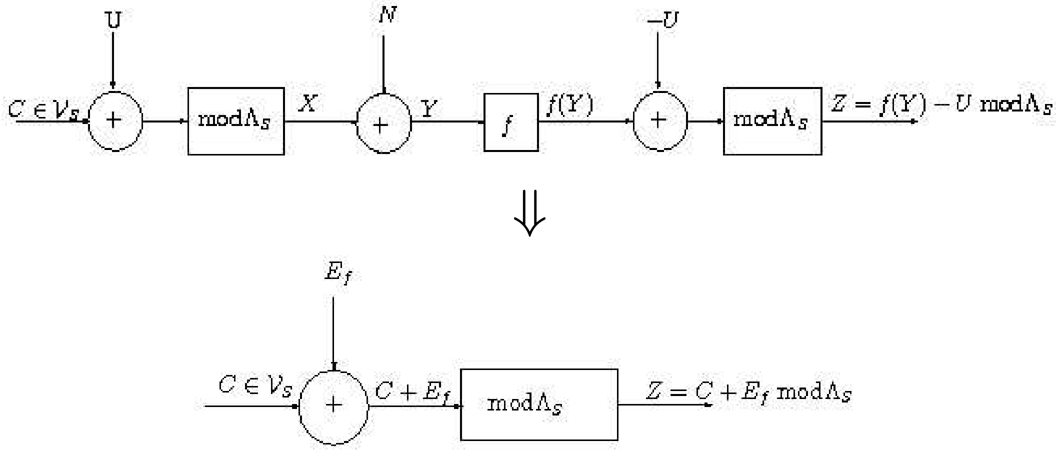
- EPI: $\frac{1}{N}h(E_f) \leq \log_2 2\pi e P_{E_f}$.

$$\Rightarrow \mathbf{C}(\Lambda_S, f) \ \geq \ \frac{1}{2}\log_2 \frac{P_X}{P_{E_f}} - \frac{1}{2}\log_2 2\pi e G(\mathcal{V}_S).$$

- Λ_S 'good for shaping': $G(\mathcal{V}_S) \to \frac{1}{2\pi e}$.

$$\Rightarrow \mathbf{C} \geq \mathbf{C}(\Lambda_S, f) \ \geq \ \frac{1}{2}\log_2 \frac{P_X}{P_{E_f}}.$$

MMSE Estimation



$$\mathbf{C} \geq \mathbf{C}(\Lambda_S, f) \geq \frac{1}{2}\log_2 \frac{P_X}{P_{E_f}}$$

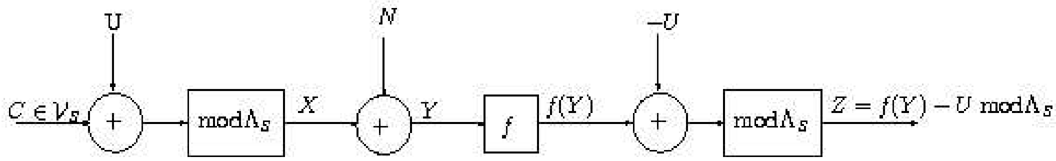
- Let $f(Y) = \widehat{X}(Y) = \alpha Y$:

$$\begin{aligned} E_f &= \alpha Y - X = \alpha N - (1-\alpha)X \\ \Rightarrow P_{E_f} &= \alpha^2 \sigma_N^2 + (1-\alpha)^2 P_X \end{aligned}$$

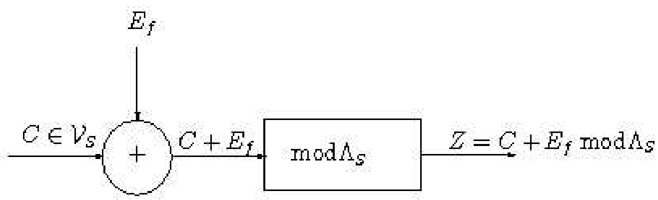
- minimize P_{E_f} $\Leftrightarrow$ choose α^* to be linear MMSE estimate:

$$\begin{aligned} \alpha^* &= \frac{P_X}{P_X + \sigma_N^2} = \frac{SNR}{1+SNR} \\ \Rightarrow P_{E_f}^* &= \frac{P_X \sigma_N^2}{P_X + \sigma_N^2} \\ \Rightarrow C(\Lambda_S, f^*) &= \frac{1}{2}\log_2(1+SNR). \end{aligned}$$

Comments on Dither, MMSE scaling



$\Downarrow$

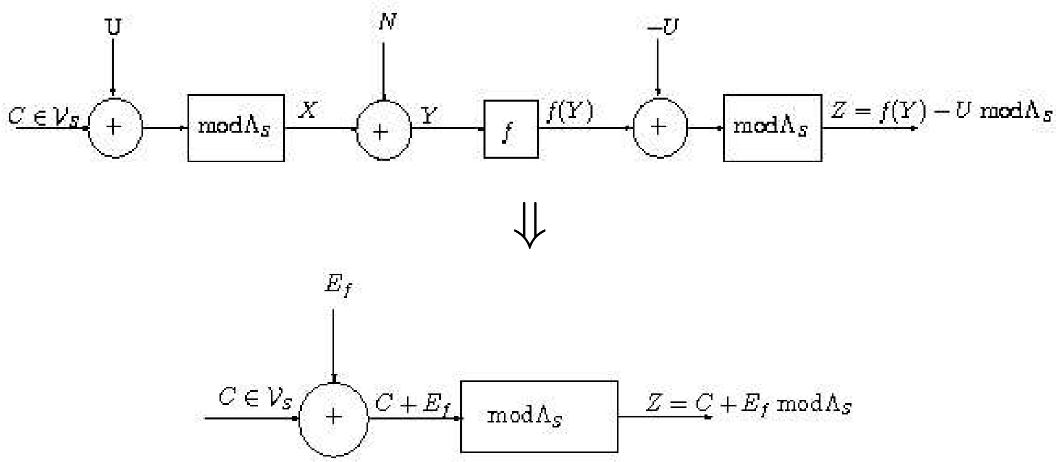


$$\mathbf{C} \;\; \geq \;\; \mathbf{C}(\Lambda_S, f) \geq \frac{1}{2}\log_2 \frac{P_X}{P_{E_f}}$$

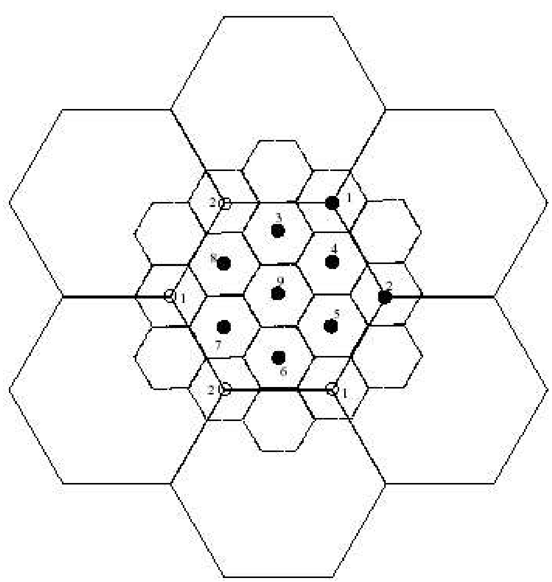
$$f(Y) = \alpha Y \;\; \Rightarrow \;\; P_{E_f} = \alpha^2 \sigma_N^2 + (1-\alpha)^2 P_X$$

- Dither U used in non-symmetric way:
  - At encoder, simply added to codeword, followed by $\bmod \Lambda_S$
  - At decoder, Y is scaled followed by dither subtraction and $\bmod \Lambda_S$ operation
- Prev. ways of using $\bmod \Lambda_S$: no scaling $\Leftrightarrow \alpha = 1 \Rightarrow \mathbf{C}(\Lambda_S, f) = \frac{1}{2}\log_2(SNR)$
- $\alpha^* \neq 1$: estimator is *biased*.
- MMSE scaling minimizes var(E_f) and increases ‘effective’ SNR by factor $\frac{SNR+1}{SNR}$.

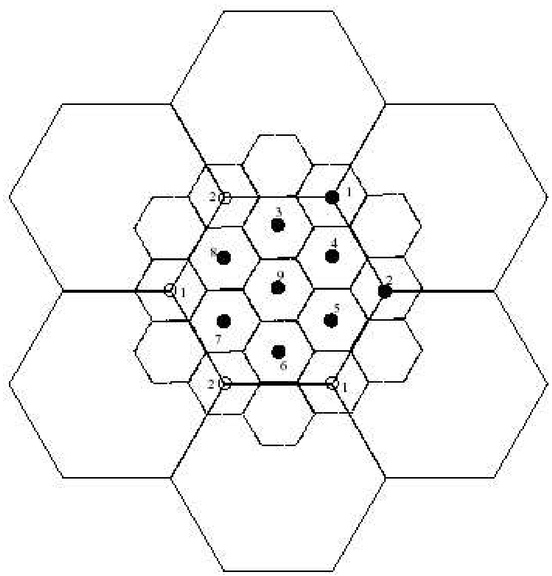
Nested Lattice Codes



- Desire: use structured coding scheme to signal $C \in \mathcal{V}_S$. Consider lattice codes.
- Fine Λ_C: 'good for channel coding'.
- Shape with $\mathcal{V}_S$, Λ_S 'good for shaping'.
- Nested lattice code: $\Lambda_S \subset \Lambda_C$.



Nested Lattice Codes (cont'd)



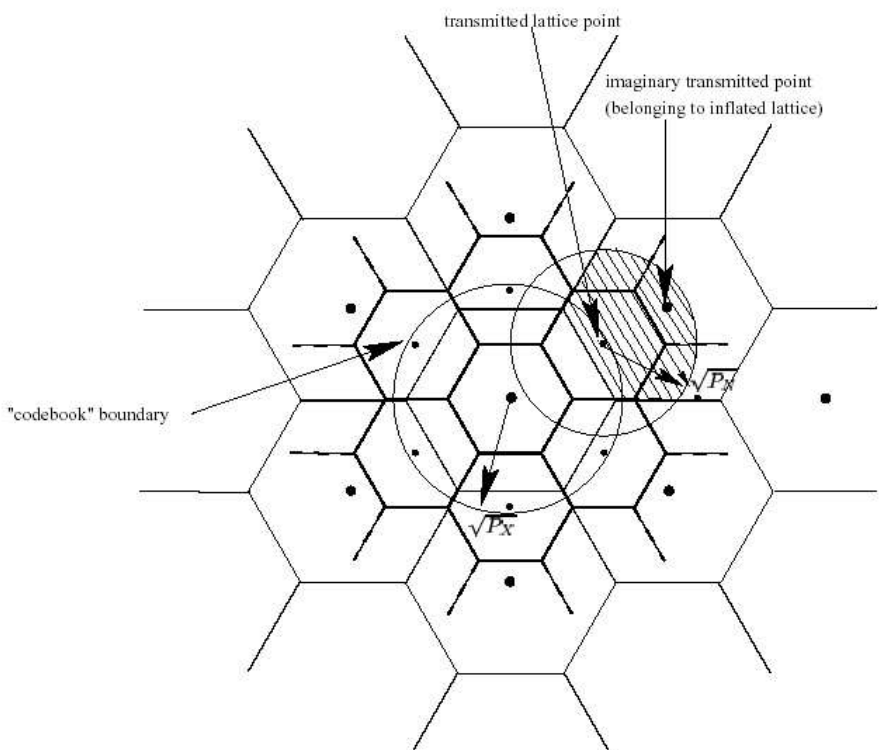
$$\mathcal{C} = \{\Lambda_C \bmod \Lambda_S\} = \{\Lambda_C \cap \mathcal{V}_S\}$$
$$R = \frac{1}{n}\log_2|\mathcal{C}| = \frac{1}{n}\log_2\frac{V(\mathcal{V}_S)}{V(\mathcal{V}_C)}$$

- Erez, Zamir show that nested lattice codes with desired properties exist *for all SNRs*.
- ML decoding with nested lattices is equivalent to lattice decoding.
  - ML decoder's quantizer:

    $$\Omega_C^* = \{e : f_{E_f}(e) \geq f_{E_f}(e - c \bmod \Lambda_S)\ \forall\, c \in \mathcal{C}\}$$
  - Note that $\Omega_C^* \neq \mathcal{V}_S$.
- Using $\mathcal{V}_S$ instead suffices and can achieve capacity.

Discussion

- Inflated lattice



- Geometry
  - Force αY to lie in same sphere as X: $\tilde{\alpha} = \sqrt{\frac{SNR}{SNR+1}}$. $\Rightarrow$ not the right intuition
  - But since $\alpha^* = \frac{SNR}{SNR+1} < \tilde{\alpha}$, with high prob. from LLN, no information loss in $\alpha^* Y \to \alpha^* Y \bmod \Lambda_S$ transformation.

Other Coding problems with Gaussian Distributions

- **Costa's 'Dirty paper coding'**

$$Y = S + X + N$$

  S known to encoder, not to decoder.

  - *Constructively* and *trivially* addressed with Erez/Zamir technique: add $\alpha^* S$ to channel input

- **Wyner-Ziv**: rate-distortion bound achieved with these codes.
- **Error exponents**: α^* only optimal as $R \to \mathbf{C}$.
  - Lower rates: α^* suboptimal.
  - Random coding error exponent can be achieved at all rates with proper choice of α.
- **MIMO flat fading channels**: generalization of these codes achieves diversity-multiplexing tradeoff.