Approaching the Capacity of the Linear Gaussian Channel

G. David Forney, Jr.
Motorola, Inc.
20 Cabot Boulevard
Mansfield, MA 02048

and

Bernard M. Gordon Adjunct Professor
Laboratory for Information and Decision Systems
M.I.T.

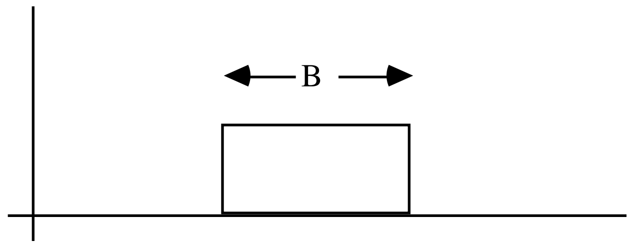
Shannon Day
Murray Hill, NJ

May 18, 1998

The ideal band-limited AWGN channel

Continuous-time channel model

Flat (brick-wall) response over a band B of width $|B| = W$ Hz



Additive white Gaussian noise, one-sided p.s.d. N_0

Input signal

- power-limited to P
- band-limited to B

Two-parameter characterization:

W: **bandwidth**

$SNR = P/N_0W$: **signal-to-noise ratio**

Equivalent discrete-time ideal AWGN channel

QAM modulation:

Two-dimensional (complex) QAM symbols x_k

Symbol rate $1/T = W$ symbols per second (2W dim/sec)

Orthonormal complex pulses $\{p(t - kT), k \in \mathbf{Z}\}$

$|P(f)|^2$ = brick-wall spectrum over B

Matched-filter receiver

Equivalent discrete-time ideal AWGN channel model:

$$z_k = x_k + n_k$$

Signal x_k, average energy $E_x = PT$ per two dimensions

Noise n_k, additive white Gaussian, variance N_0 per 2D

Single-parameter characterization (given $1/T = W$):

Signal-to-noise ratio

$$\mathrm{SNR} = E_x/N_0 = P/N_0W$$

Traditional figure of merit E_b/N_0:

$$E_b/N_0 = (E_x/R)/N_0 = \mathrm{SNR}/R.$$

Channel capacity

On an ideal AWGN channel :

$$C = \log_2 (1 + SNR) \ \text{b/2D}$$

(bits per two dimensions = b/s/Hz = spectral efficiency)

Equivalently,

$$SNR = 2^{C} - 1.$$

Reliable ($\Pr(E) \cong 0$) transmission is possible iff data rate $R < C$.

—Shannon, 1948

Normalized signal-to-noise ratio

Let

bit rate = R (bits per two dimensions) = spectral efficiency;

signal-to-noise ratio = SNR;

Define

normalized signal-to-noise ratio

$$\mathrm{SNR}_{norm} = \mathrm{SNR}/(2^R - 1).$$

Shannon limit:

Since $\mathrm{SNR}_{norm} = (2^C - 1)/(2^R - 1)$,

and $R < C$,

it follows that

$$\mathrm{SNR}_{norm} > 1.$$

Thus SNR_{norm} measures gap to capacity.

Relation to E_b/N_0 = SNR/R:

$$E_b/N_0 = ((2^R - 1)/R)\ \mathrm{SNR}_{norm}$$

Shannon limit on E_b/N_0 at a given rate R (b/2D):

$$E_b/N_0 > (2^R - 1)/R$$

As $R \to 0$, ultimate Shannon limit on E_b/N_0:

$$E_b/N_0 > \ln 2\ (-1.59\ \mathrm{dB})$$

High-SNR and low-SNR regimes

Exact expression:

$$SNR_{norm} = SNR/(2^R - 1)$$

High-SNR approximation (R >> 1):

$$SNR_{norm} \cong SNR/2^R$$

Shannon limit:

$$R < C \cong \log_2 SNR$$

Capacity goes as the logarithm of SNR

Low-SNR approximation (R << 1):

$$SNR_{norm} \cong SNR/(R \ln 2) = (\log_2 e)\, E_b/N_0$$

Shannon limit:

$$R < C \cong (\log_2 e)\, SNR$$

Capacity goes linearly with SNR

Bandwidth-limited vs. power-limited channels

Power-limited

$R < 2$ b/s/Hz

Binary transmission

Low-rate binary codes

Hamming space

Figure of merit:

$E_b/N_0 = SNR/R$

Single-pulse transmission

Matched filters

Spread spectrum

1960s -1970s

(But now turbo codes!)

Bandwidth-limited

$R > 2$ b/s/Hz

Multilevel (PAM, QAM)

High-rate nonbinary codes

Euclidean space

Figure of merit:

$SNR_{norm} = SNR/(2^R - 1)$

Sequence transmission

Intersymbol interference

Equalization

1980s - 1990s

Intermediate case

$R \cong 2\text{-}4$ b/s/Hz

Domain of PSK

Overview - Band-limited (high-SNR) regime

Linear Gaussian channel model

- Channel is a linear filter

  Intersymbol interference (ISI) is significant

- Additive Gaussian noise

Approaching channel capacity: **ideal (no ISI) channels**

- Lattice codes
- Trellis-coded modulation
- Capacity-approaching codes
- Shaping

Approaching channel capacity: **ISI channels**

Multicarrier modulation - suggested by "water pouring"

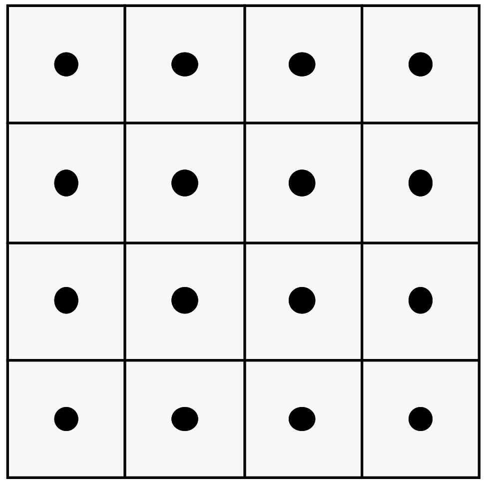
Single-carrier modulation - used in telephone modems

- Optimization of transmit band via line probing
- Combined equalization and coding via precoding

Uncoded QAM versus capacity

Uncoded QAM performance:

Square MxM constellation:



Bit rate:

$$R = \log_2 M^2 \quad \text{(bits per two dimensions)}$$

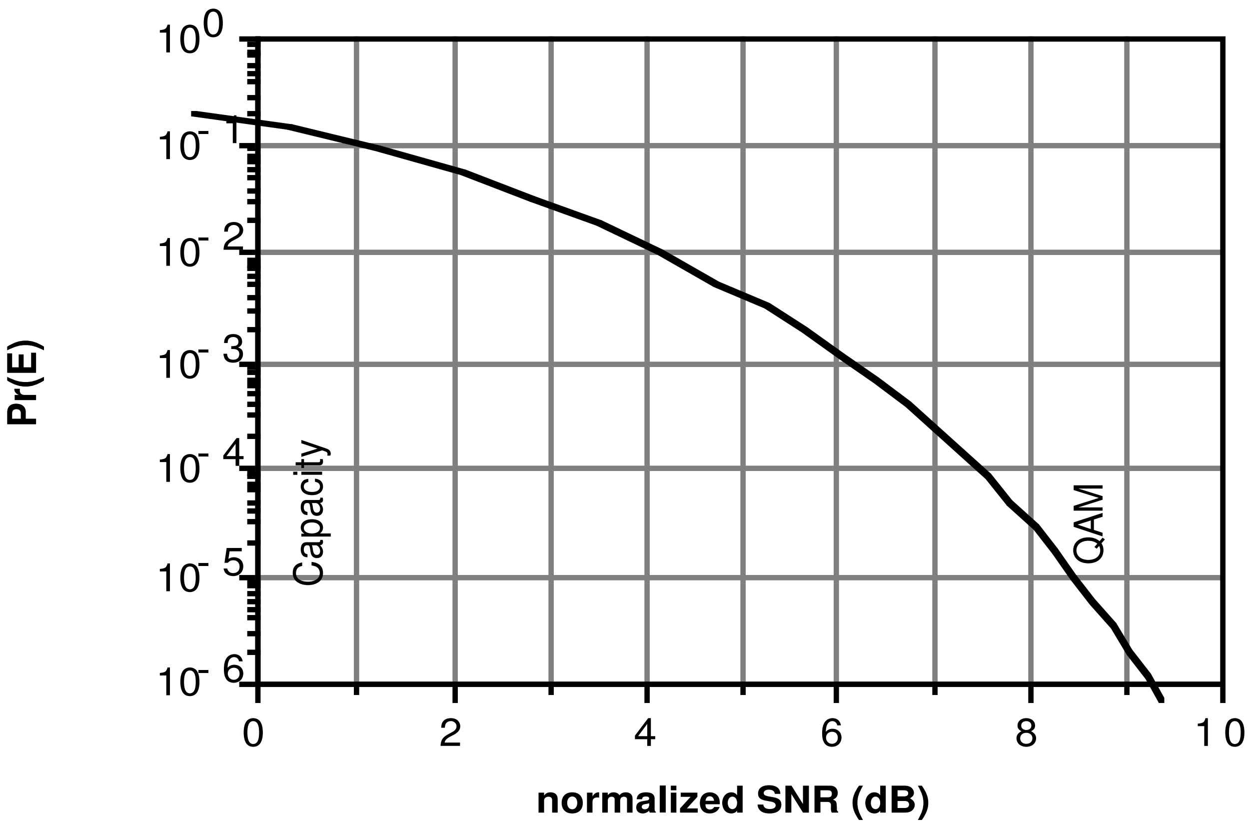
Symbol error probability:

$$\Pr(E) \cong 4 \bullet Q(3 \bullet SNR_{norm})^{1/2}.$$

$\therefore$ At high SNR, Pr(E) depends only on SNR_{norm}.

QAM vs. capacity (ideal channels)

Universal (high-SNR) performance curve :



Gap (between QAM and capacity):

6 dB @ Pr(E) $\cong 10^{-3}$

9 dB @ Pr(E) $\cong 2 \bullet 10^{-6}$

Lattice codes

Shannon : To approach capacity:

need dense sphere packings in high-dimensional space

Densest known sphere packings are usually **lattices**

N-dimensional **lattice** Λ:

Algebraically, a discrete subgroup of $\mathbf{R}^N$

Geometrically, a regular array of points in N-space

Performance of lattice constellations

Lattice constellation: $\Lambda \cap R$

In words: the set of all lattice points in a bounding region R

Symbol error probability

$$\Pr(E) \cong K_{min}(\Lambda)\cdot Q(3\cdot\gamma_c(\Lambda)\cdot\gamma_s(R)\cdot SNR_{norm})^{1/2}.$$

* Two separable gains

Coding gain : $\gamma_c(\Lambda)$

(Improvement over uncoded QAM)

Shaping gain : $\gamma_s(R)$

(Improvement over N-cube)

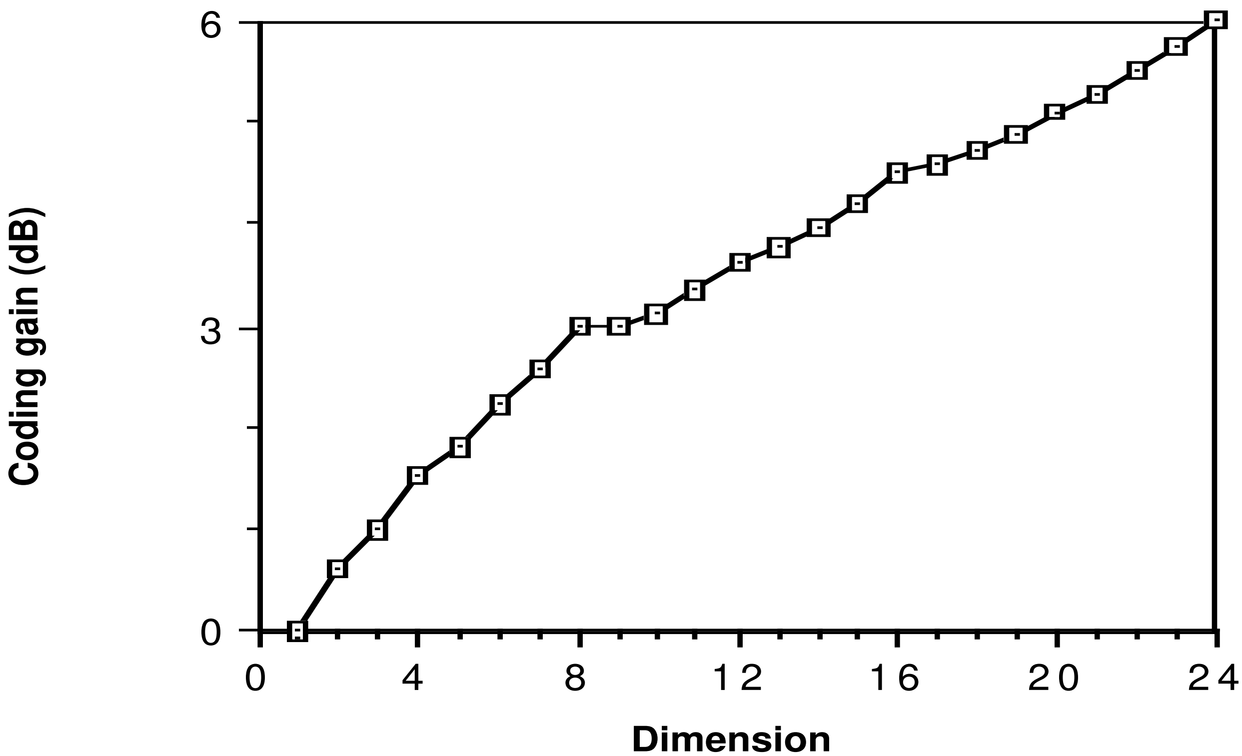
* **Effective coding gain** is reduced by the multiplicity $K_{min}(\Lambda)$

Rule of thumb : factor of two costs 0.2 dB

(@ $\Pr(E) \cong 10^{-6}$)

Coding gains of densest lattices

Nominal coding gains of the densest lattices in up to 24 dimensions:



—Conway and Sloane, 1988

- Nominal coding gain increases without limit as $N \to \infty$
- Not real, because of effect of multiplicity $K_{min}(\Lambda)$
- Effective coding gain of 24-D Leech lattice only about 4 dB

Lattice-type trellis codes

<u>Trellis code</u> **C**(Λ/Λ'; C):

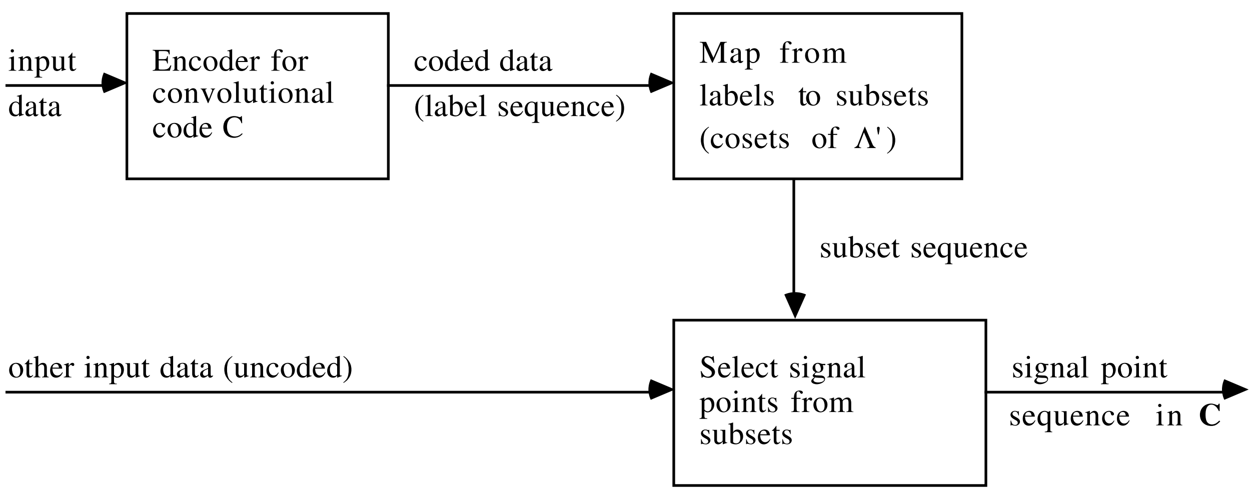
Based on:

N-dimensional **lattice partition** Λ/Λ';

A **label map** m: A → Λ/Λ' from a **label alphabet** A to the subsets (cosets of Λ') in the partition Λ/Λ';

A **convolutional code** C over A.

—Calderbank and Sloane, 1987
—Forney, 1988



Symbol error probability (per N dimensions):

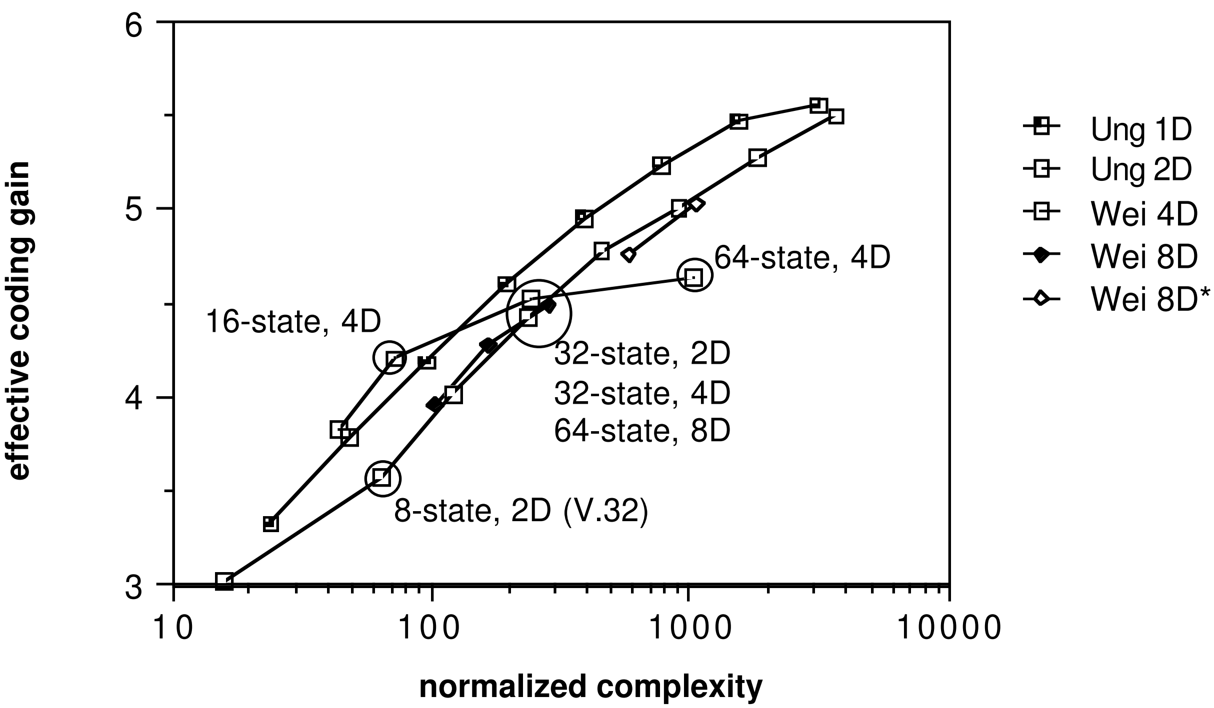
$$\Pr(E) \cong K_{min}(\mathbf{C}) \cdot Q(3 \cdot \gamma_c(\mathbf{C}) \cdot \gamma_s(R) \cdot \mathrm{SNR}_{norm})^{1/2}.$$

Progress in coded modulation for ideal channels

Trellis-coded modulation:

Ungerboeck (1982):
- 1D and 2D codes
- nominal coding gains of 3 to 6 dB
- practical : 4 to 512 states

Multidimensional codes
- Wei (1987)
- Calderbank and Sloane (1987)
- Similar coding gains
- Reduced constellation expansion
- Rotational invariance



Easy to achieve 0.6 dB gain over V.32 code

Diminishing returns
- Doubling complexity gains about 0.3 dB

Motivates reduced-complexity decoding algorithms

Capacity-approaching codes

Binary codes for the low-SNR regime:

Sequential decoding of convolutional codes (Wozencraft, 1961)

Operate at R_0: about 3 dB from the Shannon limit

Bidirectional sequential decoding: about 1.5 to 2 dB

Turbo codes (Berrou et al., 1993)

Can approach within about 0.35 dB of the Shannon limit

Low-density parity-check codes (Gallager, 1962 (!))

Can approach capacity in principle
—Gallager, 1962; MacKay and Neal, 1996

Nonbinary LDPC codes can outperform turbo codes
—Davey and MacKay, April 1998

Nonbinary codes for the high-SNR regime:

Sequential decoding of trellis codes

Operate at R_0: about 1.7 dB from the Shannon limit

Multilevel coding and multistage decoding with binary codes
—Leech and Sloane, 1971; Imai and Hayakawa, 1977

Can approach capacity in principle
—Kofman, Shamai and Zehavi, 1990

Can approach capacity in practice with turbo codes
—Wachsmann and Huber, 1995

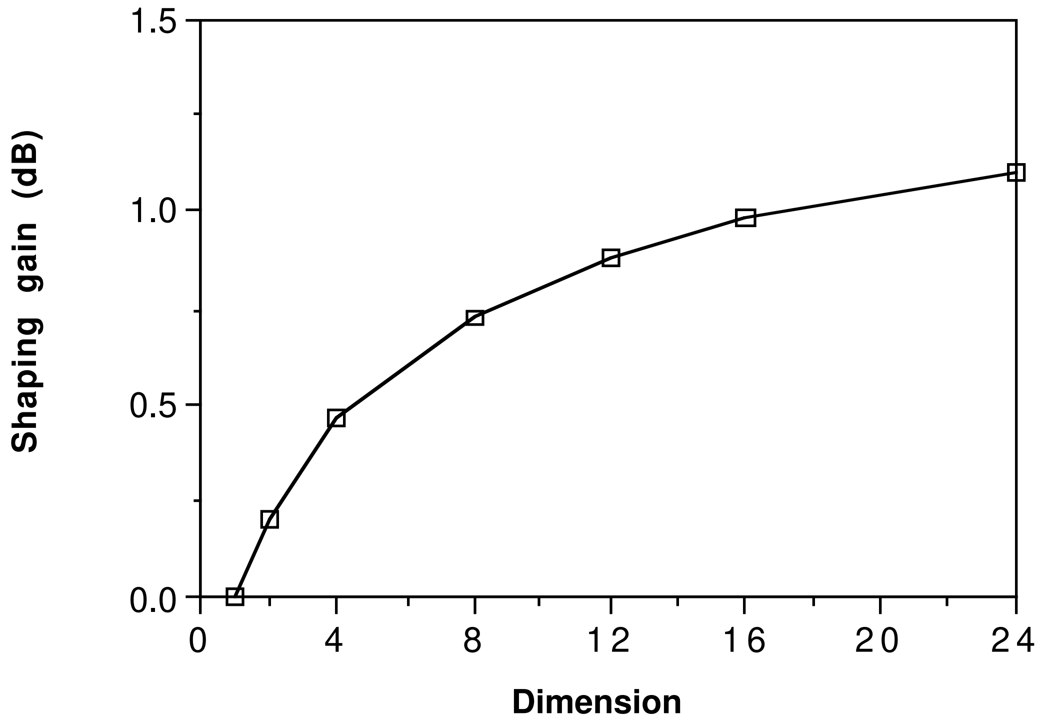
Shaping gains: Spherical shaping

Optimum bounding region R in N dimensions is an **N-sphere**

Ultimate shaping gain ($N \to \infty$) is

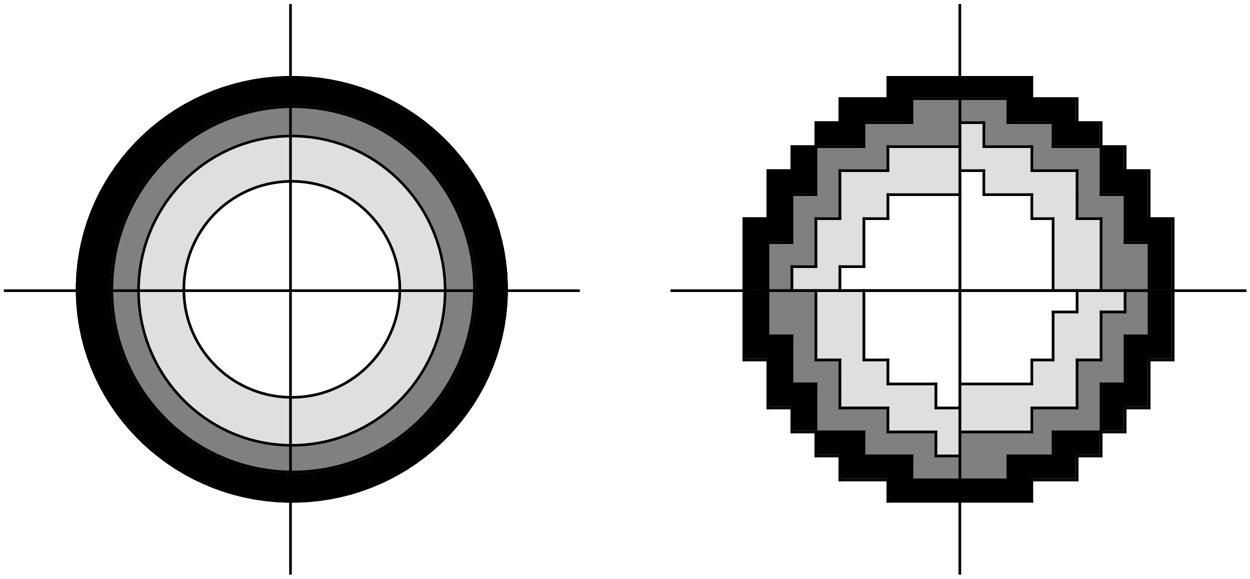
$\pi e/6 = 1.53$ dB

Shaping gains of spheres:



Shaping methods

Low-dimensional constellation partitioned into subregions:



Shaping on regions

A binary code selects a sequence of subregions

Short ($N \leq 24$) block codes can achieve 0.8 dB gain
with only 4 subregions

—Calderbank and Ozarow, 1990

Trellis shaping

Dual to trellis coding

Easily obtains shaping gains of 1 dB or more

—Forney, 1992

Shell mapping

Generating function techniques for efficient block coding

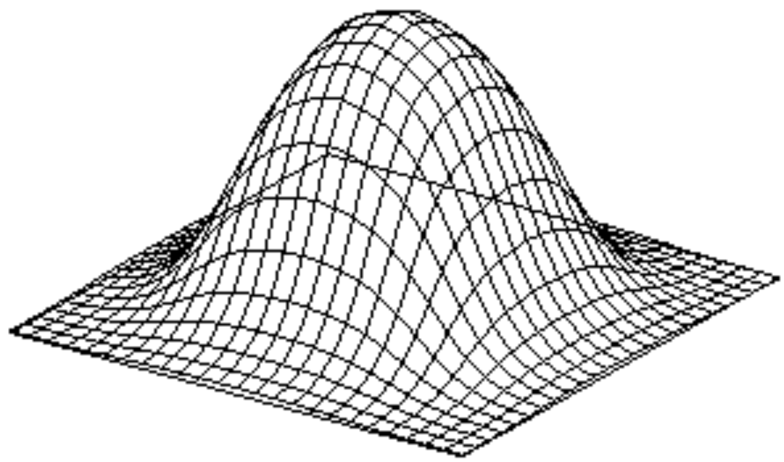
About 0.8 dB shaping gain in V.34 (16-D, 25% expansion)

—Lang and Longstaff, 1985
—Eyuboglu; Fortier; Khandani; Kschischang; Laroia, c. 1991-92

Effects of shaping

Effect/objective is to achieve nonuniform distribution over a

low-dimensional constellation:



Distribution approaches truncated Gaussian

Trade-off between constellation expansion and shaping gain

Ultimate shaping gain of $\pi e/6$ (1.53 dB):

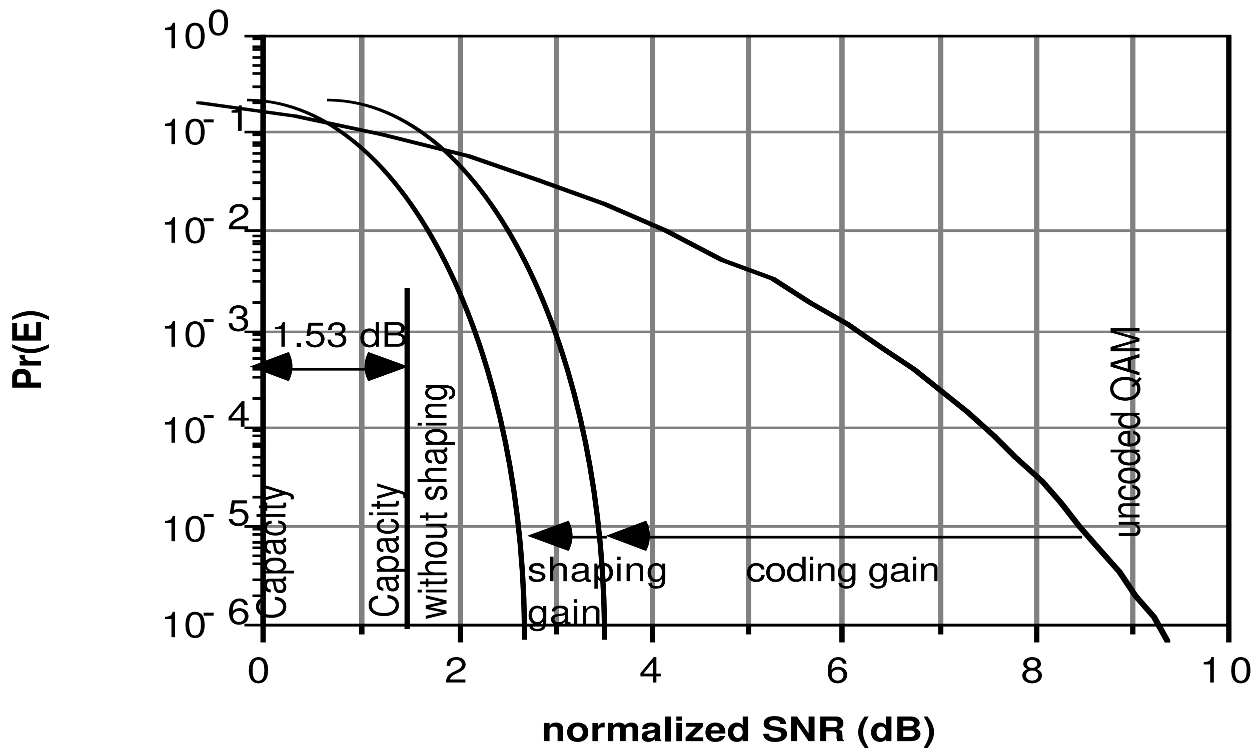
Difference between the average energies of

uniform vs. Gaussian distribution

for a given differential entropy

—Forney and Wei, 1989

Progress in coded modulation for ideal channels (cont.)



Gap at Pr(E) $\cong 2 \cdot 10^{-6}$

Coding : can achieve about 6 dB
maximum possible: 7.5 dB

Shaping: can achieve about 1 dB
maximum possible: 1.5 dB

Total : can achieve about 7 dB
maximum possible: 9 dB

(Gap is greater at lower Pr(E))

DeBuda's result

There exist lattice codes that can approach capacity as $N \rightarrow \infty$

High-SNR result

Spherical shaping (1.53 dB shaping gain)

— DeBuda (1975)

Therefore,

there exist lattices with

effective coding gain = gap minus 1.53 dB

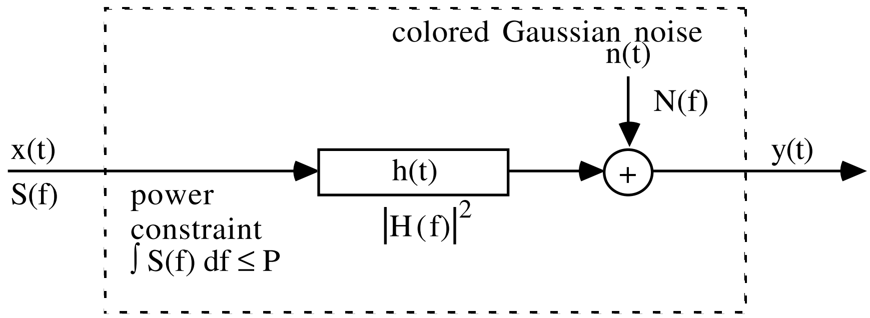
(exponential error bounds: Poltyrev, 1994)

and

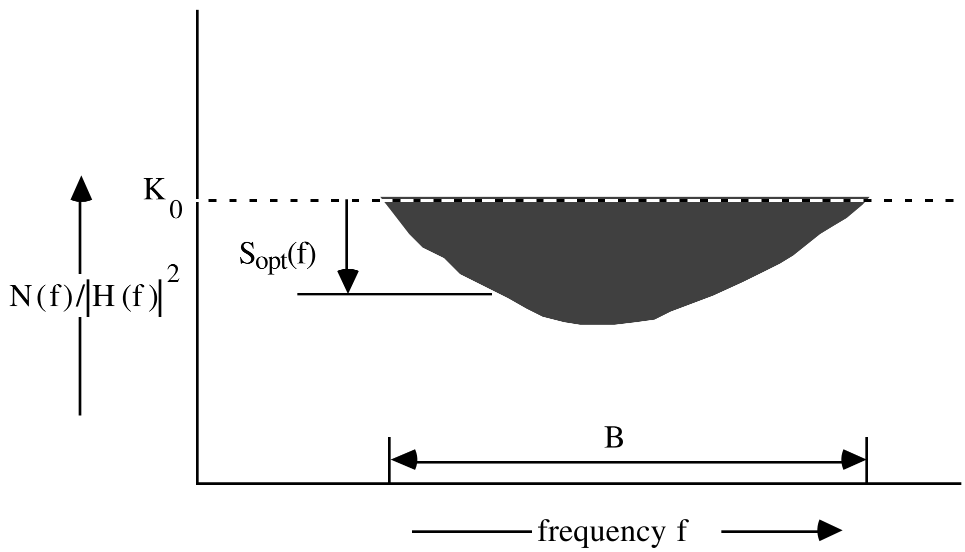
the gap can be completely closed with lattice or trellis codes

Capacity of non-ideal linear Gaussian channels

Given: linear Gaussian channel



Shannon: water pouring



Lessons :

- Spectrum : optimum spectrum nearly flat in practice
- Location of transmit band B : important to optimize
- Statistics : transmitted signal should be Gaussian-like

Effective SNR

Capacity using water-pouring spectrum $S_{opt}(f)$:

$$C = \int_B \log_2 (1 + S_{opt}(f)|H(f)|^2/N(f))\, df/W \quad (b/2D)$$

Define the **effective signal-to-noise ratio**:

$$SNR_{eff} = -1 + \exp (\int_B \ln (1 + S_{opt}(f)|H(f)|^2/N(f))\, df/W)$$

(Note:

$1 + SNR_{eff}$ = geometric mean of $1 + S_{opt}(f)|H(f)|^2/N(f)$ over B)

Capacity in terms of SNR_{eff} (as on ideal AWGN channel)

$$C = \log_2 (1 + SNR_{eff}) \quad (b/2D)$$

SNR_{norm} and Shannon limit:

$$SNR_{norm} = SNR_{eff}/(2^R - 1) > 1.$$

Two-parameter characterization of any single-band channel:

Bandwidth W

Effective signal-to-noise ratio SNR_{eff}

Equalization approaches

Problem: as Tx band widens, severe intersymbol interference (ISI)

Multicarrier solution (Shannon-Holsinger-Gallager):

Divide channel into subbands of width $\Delta f \to 0$

Each subband becomes an ideal AWGN channel

Bandwidth = Δf

No equalization required

Power allocation: $S_{opt}(f)\,\Delta f$

$$SNR(f) = S_{opt}(f)|H(f)|^2/N(f)$$

Aggregate power $\to P = \int_B S_{opt}(f)\,df$

Rate allocation:

$$R(f) \cong C(f) = \log_2 (1 + SNR(f)) \quad (b/2D)$$

Aggregate rate $\to C_{b/s} = \int_B C(f)\,df \quad (b/s)$

Note:

Requires channel measurement, known to Tx

Requires powerful coding in each subband

Practical implementations in 1990s

—Ruiz, Cioffi, Kasturia, 1992

Equalization approaches: Single-carrier

Single-carrier (QAM) methods:

1. Linear equalization

    Excessive noise enhancement with deep nulls

2. Maximum-likelihood sequence detection (VA)

    Too complex

3. Decision-feedback equalization (DFE)

    Asymptotically optimal (!)

        MLSD is not necessary

    Can't be combined with coding

4. "Precoding"

    "DFE in the transmitter"

    Asymptotically optimal performance

    Combines nicely with coding— no glue

    Usable with any shaping with minimal "dither loss"

    Goldilocks solution: "Just right!"

Price's result (general linear Gaussian channels)

Price's result:

On any linear Gaussian channel, as SNR $\rightarrow \infty$,

the gap between capacity and QAM performance

with ideal ZF-DFE (tail-cancelling equalization)

is **independent** of channel and noise spectra.

—Price, 1972
—Eyuboglu, 1988

Improved result using MMSE-type equalization:

On any linear Gaussian channel, at any SNR,

the equivalent SNR using

ideal MMSE-optimized tail-cancelling equalization

is equal to the effective SNR of the channel:

$$\mathrm{SNR}_{\mathrm{MMSE\text{-}DFE}} = \mathrm{SNR}_{\mathrm{eff}}.$$

—Cioffi, Dudevoir, Eyuboglu and Forney, 1995

Implication:

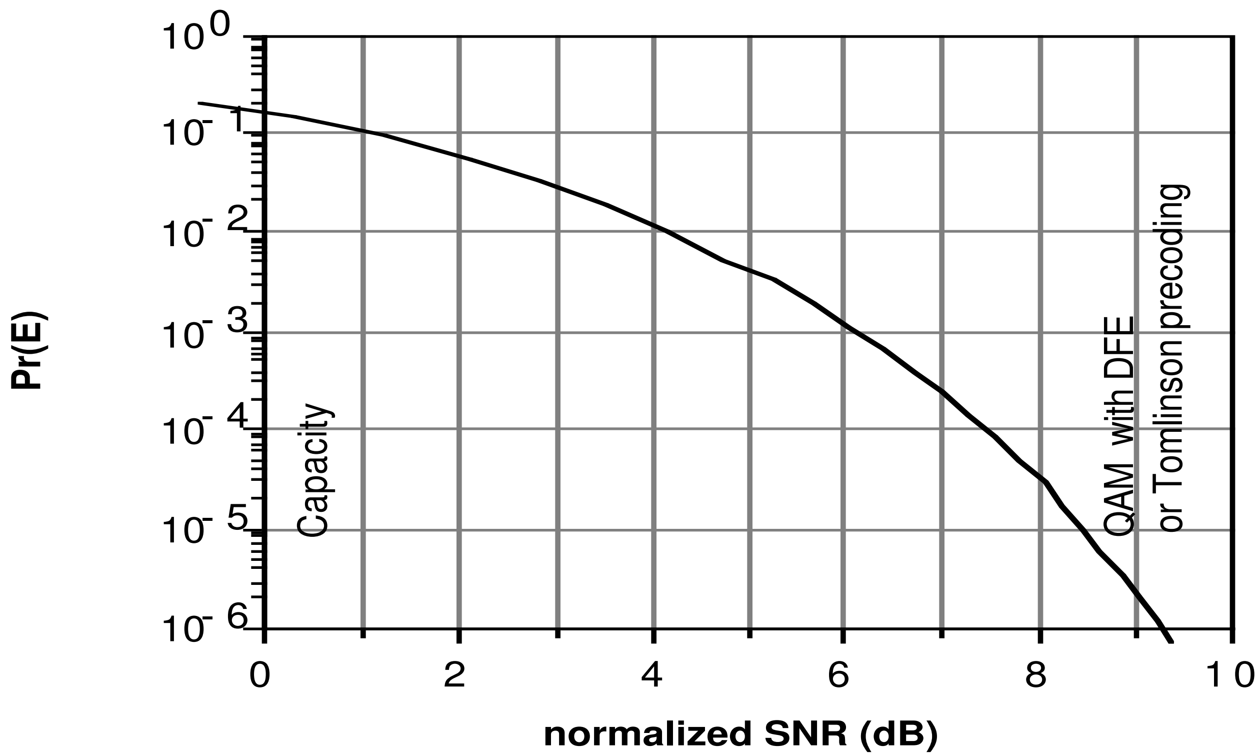
Ideal MMSE-optimized tail-cancelling equalization +

Capacity-approaching ideal-AWGN-channel coding =

Approach to capacity of any linear Gaussian channel

Universal performance curve: No coding

For any linear Gaussian channel:



$$SNR_{norm} = SNR_{eff}/(2^R - 1)$$

Assumes ideal tail-cancelling equalization

Decision-feedback equalization (DFE)

Tomlinson-Harashima precoding
—Tomlinson, 1971; Harashima and Miyakawa, 1972

Approaching capacity on non-ideal channels

"Precoding"

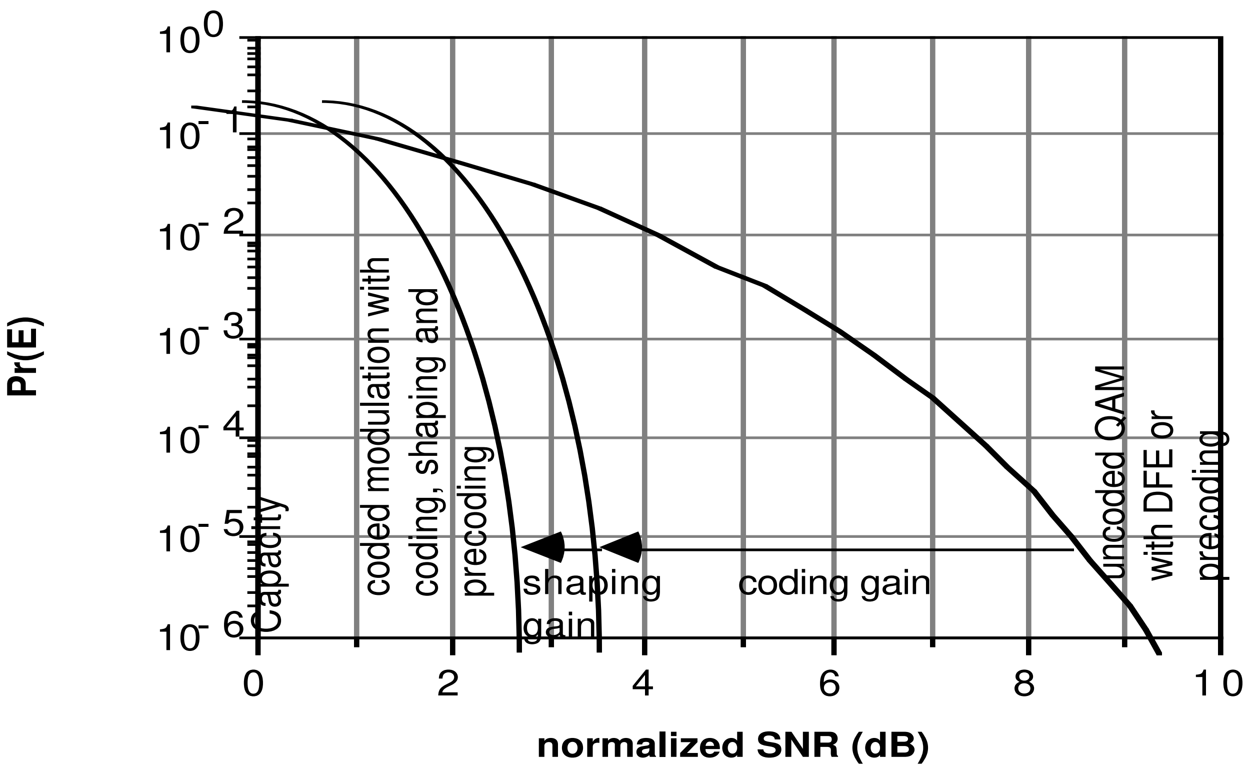
- can be combined with coding with "no glue"

—Eyuboglu and Forney, 1991-2

- can also be combined with any shaping technique

—Eyuboglu, Cole, Laroia (during V.34), 1992-93

Universal curve (any linear Gaussian channel):



Equalization gain of DFE

Coding gain $\gamma_c(\mathbf{C})$ of ideal-channel code **C**

Shaping gain γ_s of ideal-channel shaping method

Conclusions

- Practical techniques can approach channel capacity:

  Known codes can approach maximum possible **coding gain**

  Simple **shaping** methods can approach ultimate **shaping gain**

- Can approach capacity as closely on non-ideal channels

  as on ideal channels:

  Multicarrier modulation with powerful coding and shaping

  or

  Single-carrier modulation with precoding achieves

  Coding gain of known codes

  Shaping gain approaching 1.53 dB

  Equalization gain of ideal MMSE-DFE

- In principle, on <u>any</u> band-limited linear Gaussian channel

  one can <u>approach capacity as closely as desired</u>

  —(Combine results of Price, DeBuda, Cioffi *et al.*)