Applications of LDPC Codes: Hybrid ARQ schemes and coding for "Dirty-Paper"

Giuseppe Caire

Mobile Communications Department

Institut Eurécom, Sophia Antipolis

giuseppe.caire@eurecom.fr

Outline of this talk

- Hybrid ARQ schemes (joint work with Stefania Sesia and Guillaume Vivier, MOTOROLA LABS, Gif-sur-Yvette)
  1. Model and assumptions
  2. Performance limits of binary codes
  3. Performance limits of LDPC codes via Density Evolution
  4. Performances of finite-length ensembles
- Coding for “Dirty-Paper” channels (joint work with Shlomo Shamai, Technion, Israel)
  1. Information theoretic background
  2. Applications to broadcasting, data-hiding and ISI channels
  3. An LDPC design approach for the Gaussian case

Part I: Hybrid ARQ schemes

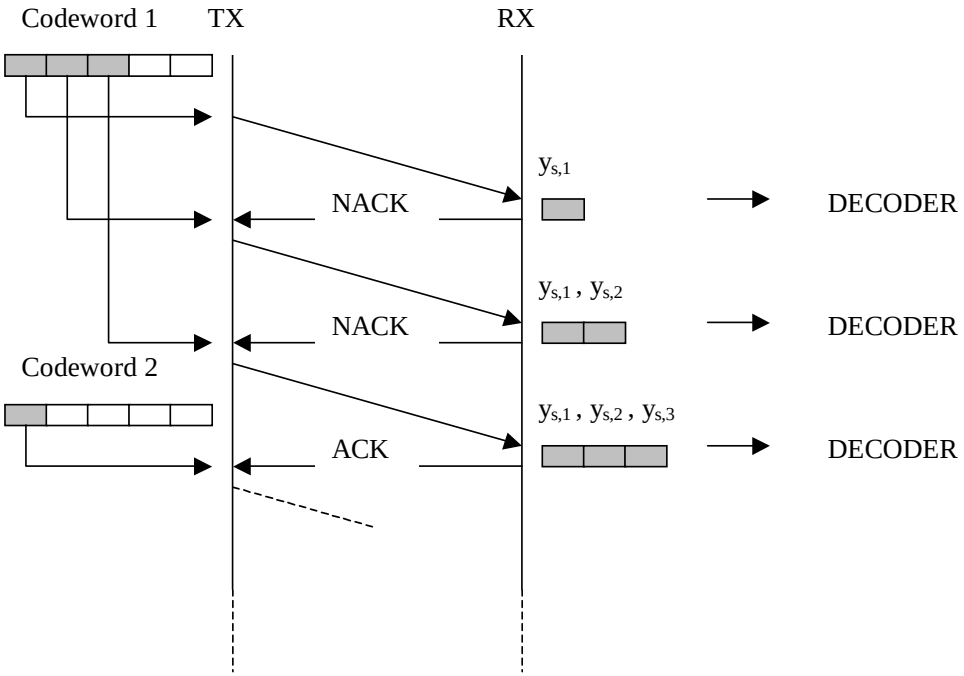
System model

- Single user transmission, block-fading Gaussian channel.
- Information is encoded by using a rate $R = \frac{b}{LM} = \frac{r}{M}$ code.
- The codeword of length LM is divided into M sub-blocks each transmitted in a time slot s

Received signal in slot s: $\mathbf{y}_s = c_s \mathbf{x}_s + \boldsymbol{\nu}_s$

- Constant energy per symbol $E = \|x_{s,l}\|^2$.
- Slowly time varying fading: constant block-fading on each slot, c_s (Rayleigh fading).

Incremental redundancy HARQ scheme



Decoding occurs at mth burst $\Longrightarrow$ effective coding rate $\frac{r}{m} = \frac{b}{Lm}$.

Background results

Achievability: for all $\epsilon > 0$ there exist L and a code $\mathcal{C} \in \mathbb{C}^{LM}$ of size 2^{rL} with $\Pr(\text{error}|\mathcal{P}, \mathcal{C}) < \epsilon$ for all $m = 1, \ldots, M$ and channel sequences $\mathcal{P}$ such that

$$I_m = \sum_{s=1}^{m} I(q(x), p_s(y|x)) > r$$

Converse: for all channel sequences $\mathcal{P}$ such that

$$I_m = \sum_{s=1}^{m} I(q(x), p_s(y|x)) < r = \frac{b}{L}$$

$\Pr(\text{error}|\mathcal{P}, \mathcal{C}) \to 1$ for any code $\mathcal{C}_m \in \mathbb{C}^{Lm}$ of size 2^{rL} as $L \to \infty$.

Error Detection: for all $\epsilon > 0$ and channel sequences $\mathcal{P}$ there exists L such that any code $\mathcal{C} \in \mathbb{C}^{LM}$ satisfies $\Pr(\text{undetected error}|\mathcal{P}, \mathcal{C}) < \epsilon$.

Throughput Analysis

Assumptions:

- The sender has an infinite buffer of messages.
- The ACK/NACK channel is delay and error free.
- $\alpha_s = |c_s|^2$ are i.i.d random variables.

Renewal-Reward Theory:

$\mathcal{E} = \{\text{Stop tx current codeword}\}$ is a recurrent event $\Rightarrow \quad \eta = \frac{\mathrm{E}[\mathcal{R}]}{\mathrm{E}[\tau]}$

- $\mathcal{R}$: Random Reward, τ: Inter-Renewal time.

Defining: $p(m) = \Pr\left(\overline{\mathcal{A}}_1, \overline{\mathcal{A}}_2, \ldots, \overline{\mathcal{A}}_m\right) = \Pr(I_1 \leq r, \ldots, I_m \leq r)$

$$\eta = R\,M\,\frac{1 - p(M)}{1 + \sum_{m=1}^{M-1} p(m)}$$

Random Binary Codes

I_m is the mutual information of the channel over the first m slots:

$$I_m = \sum_{i=1}^{m} I(\beta_i), \qquad \beta_i = \frac{E|c_i|^2}{N_0} = \alpha_i \gamma$$

$I(\beta_i)$: mutual information of the BIAWGN channel:

- $p(m)$: exact computation via convolution

$$p(m) = \Pr(I_m \leq r) = \Pr\left(\sum_{i=1}^{m} I(\beta_i) \leq r\right)$$

m-fold convolution of the pdf of $I(\beta_i)$ and subsequent integration.

- $p(m)$: Gaussian Approximation
- $p(m)$: Chernoff Bound

Random Binary Codes: Results

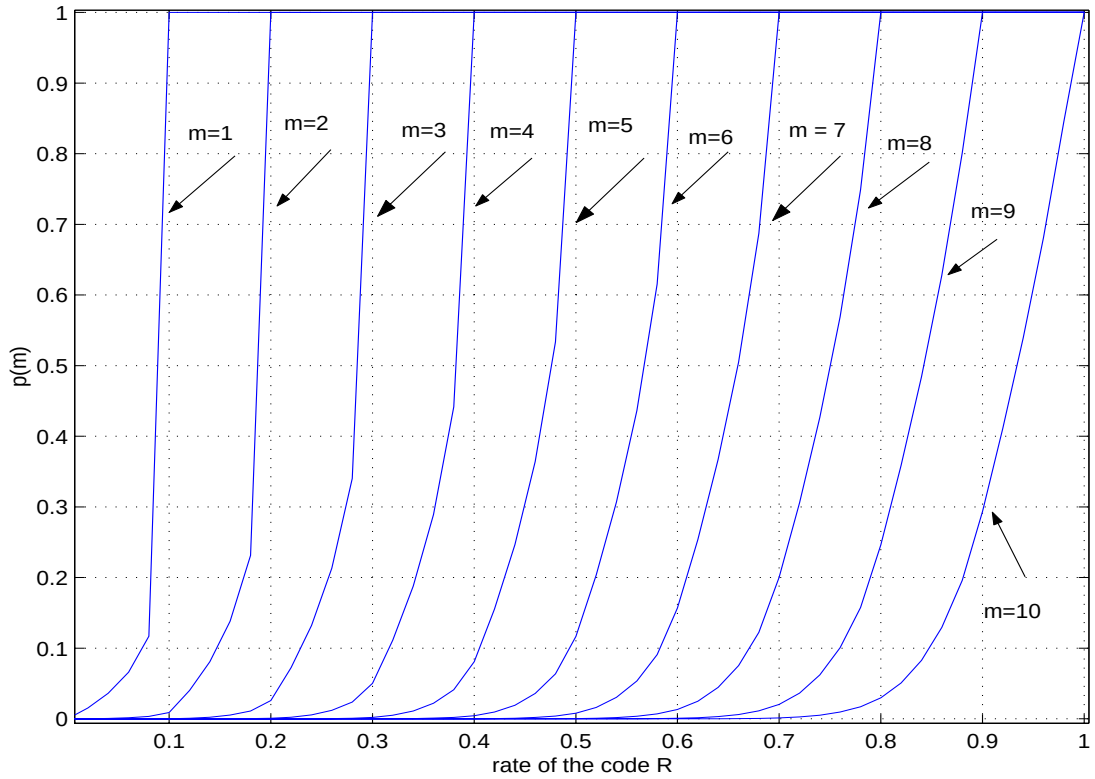


Figure 1: $p(m)$ for $M = 10$ and $\gamma = 10dB$ (calculated via convolution).

Random Binary Codes: Results

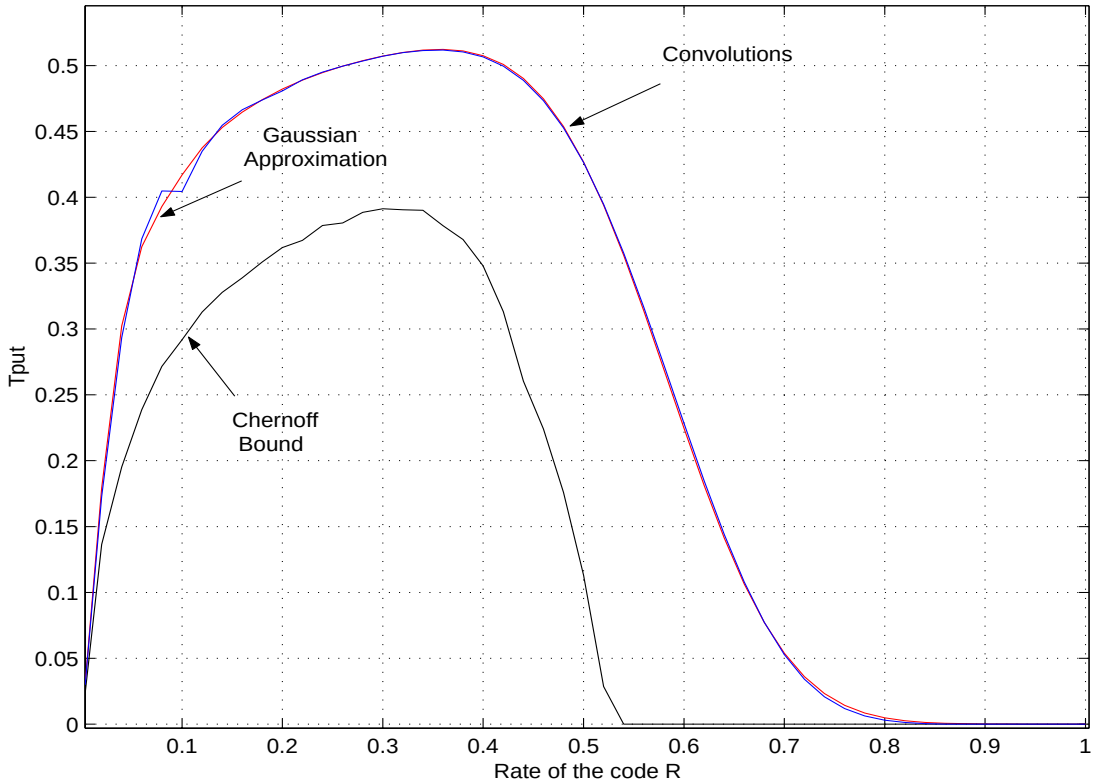


Figure 2: Throughput for $\gamma = 0$ dB.

Random Binary Codes: Results

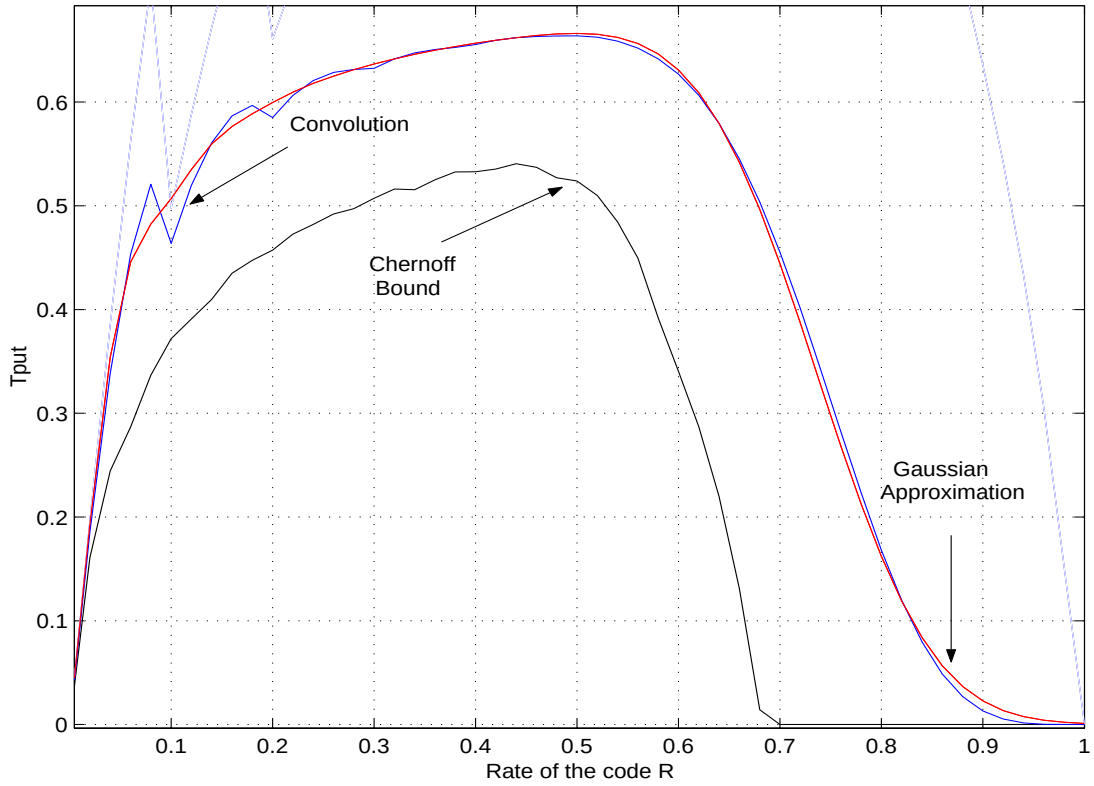


Figure 3: Throughput for $\gamma = 3$ dB.

Random Binary Codes: Results

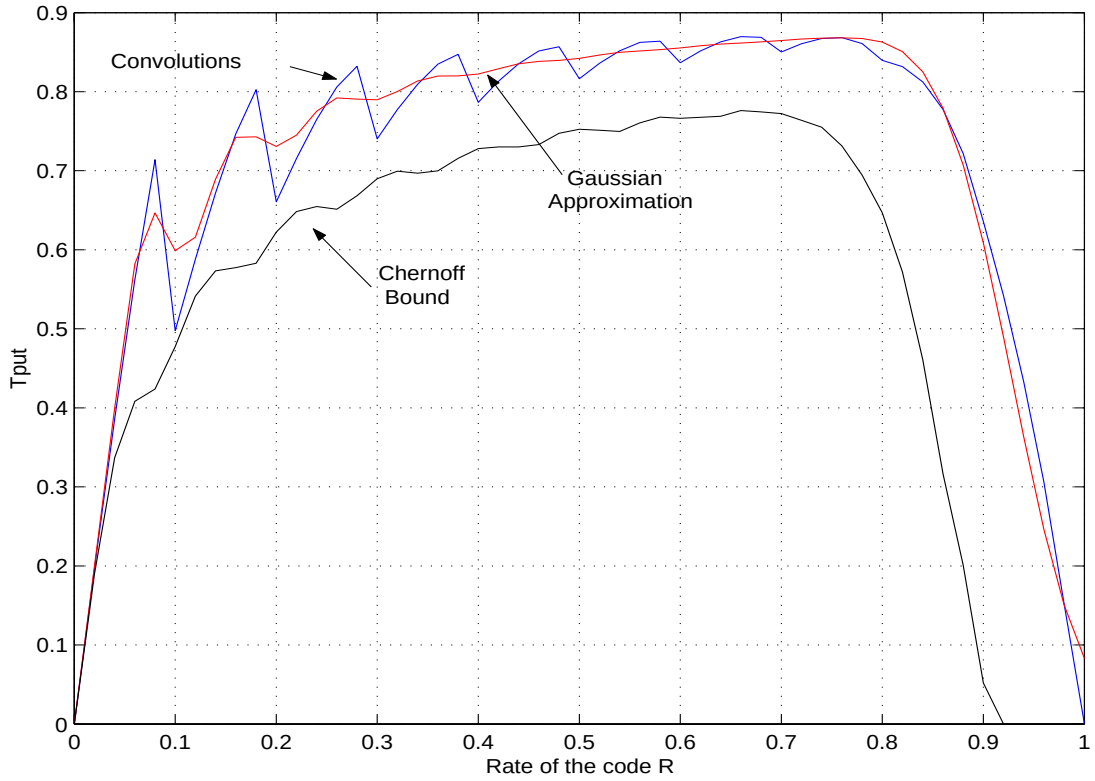


Figure 4: Throughput for $\gamma = 10$ dB.

LPDC Codes

Sparse parity-check equation $\mathbf{H}^T\mathbf{x} = \mathbf{0}$

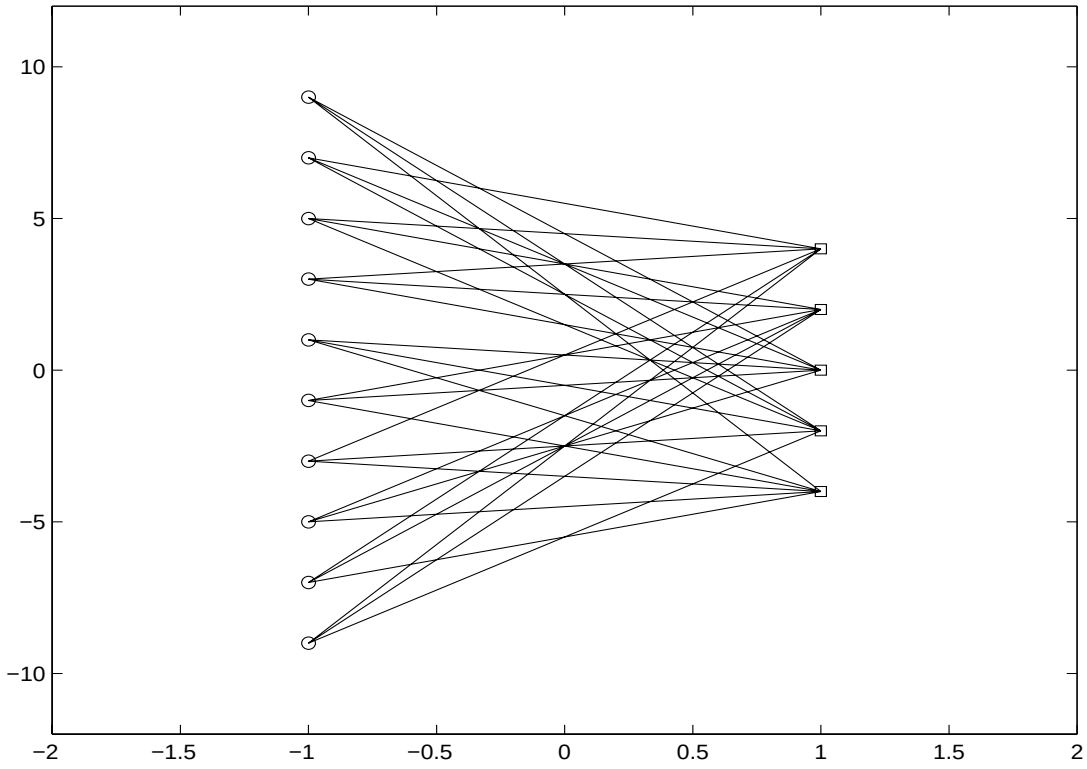


Figure 5: A regular LDPC $(3, 6)$ of length 10.

LPDC Codes

- Left and right degree sequences $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$, where λ_i (resp. ρ_i) is the fraction of edges in the graph with left (resp. right) degree i.
- Decoding algorithm: Belief-Propagation message passing.
- Performance analysis: Density Evolution.
- Significance of DE analysis: it yields the iterative decoding threshold, i.e., the maximum channel noise parameter below which the BER vanish with the number of iterations for a randomly selected graph of increasing blocklength, with exponentially large (in the blocklength) probability.

LDPC Codes: Density Evolution

- We define the event $\mathcal{A}_m = \{\mathrm{DE}_m(\boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{\alpha}_m)$ converges $\}$: the density evolution converges to the zero-BER fixed point.
- We use the so-called EXIT transfer function to approximate Density Evolution as a one dimensional dynamic system

$$X^{\ell} = \frac{1}{M} \sum_{m=1}^{M} F_{\lambda}\left(1 - F_{\rho}\left(1 - X^{\ell-1}, 0\right), \alpha_m \gamma\right)$$

where for a degree sequence $\mathbf{a} = \{a_2, \ldots, a_d\}$ we define

$$F_{\mathbf{a}}(x, y) = \sum_{k=2}^{d} a_k J((k-1)J^{-1}(x) + y)$$

$J(z)$ is the Mutual Information of BIAWGN channel.

LDPC Codes: Results

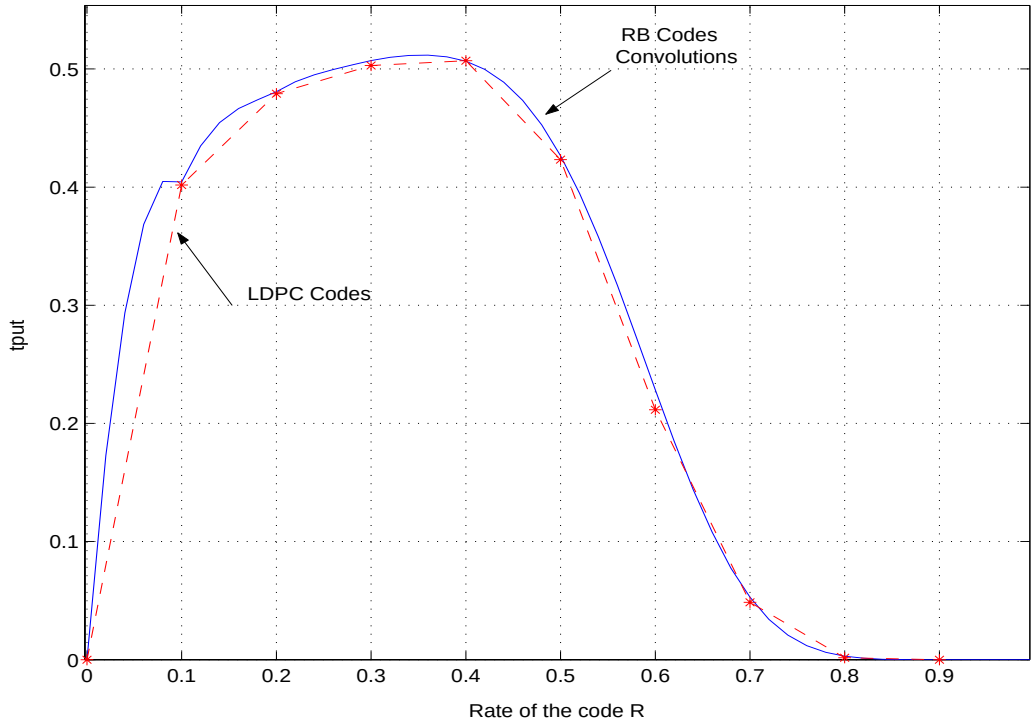


Figure 6: Throughput for $\gamma = 0dB$.

LDPC Codes: Results

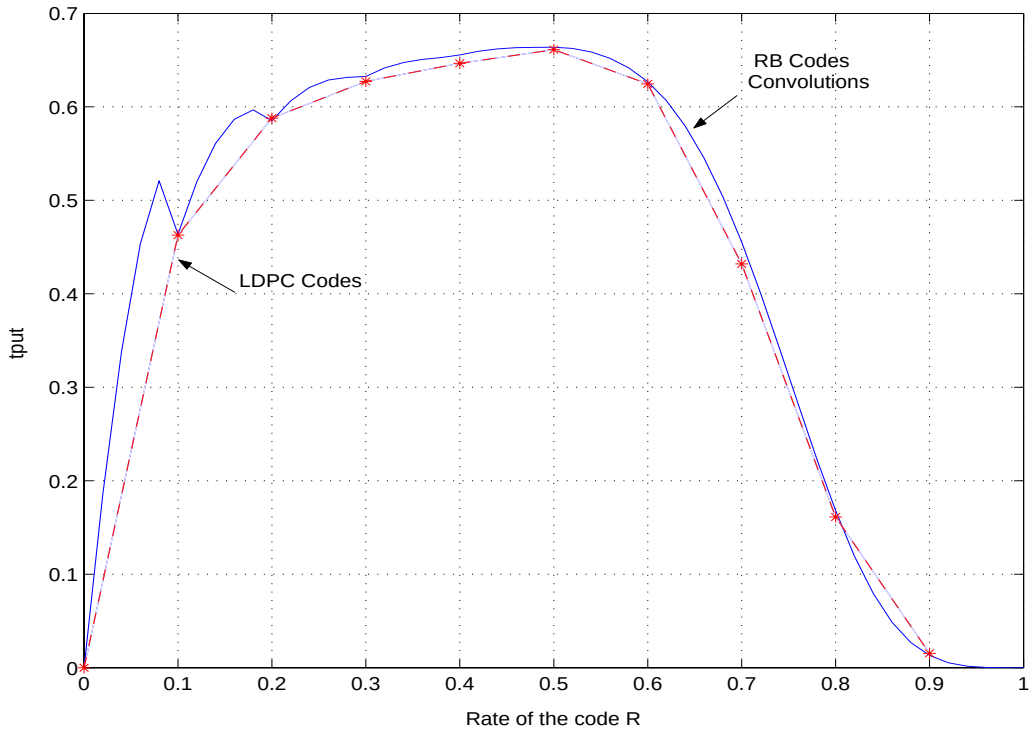


Figure 7: Throughput for $\gamma = 3dB$.

LDPC Codes: Results

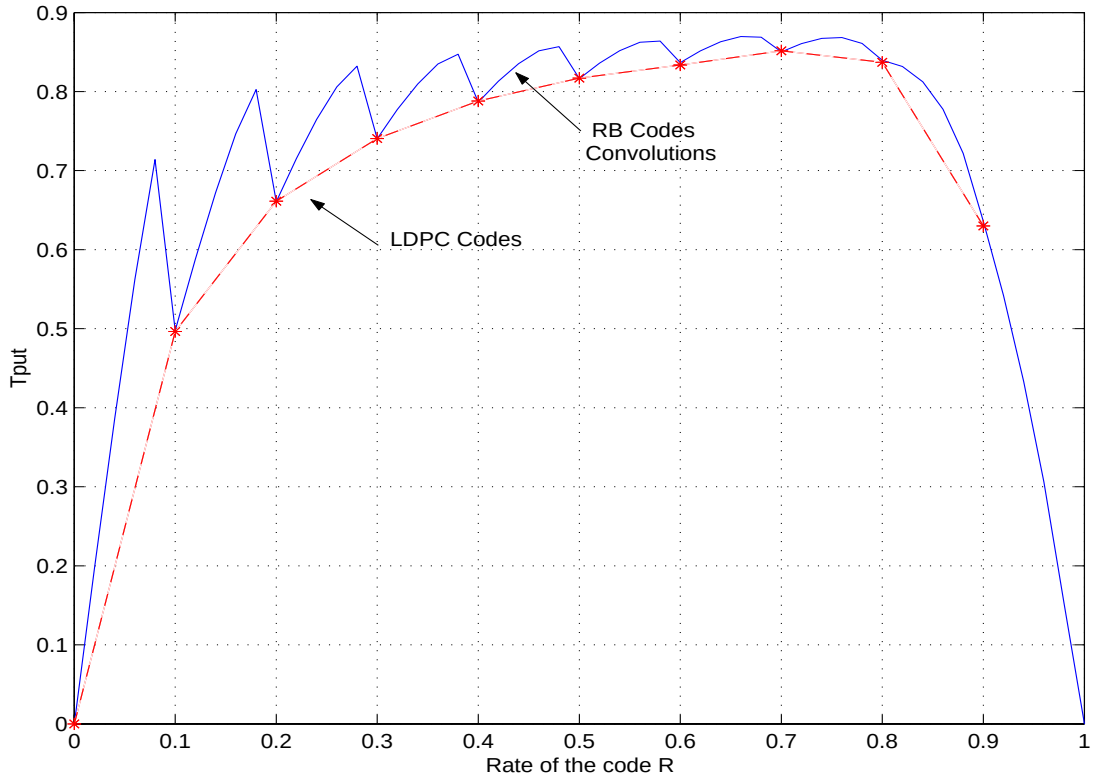


Figure 8: Throughput for $\gamma = 10dB$.

LDPC Codes: Results

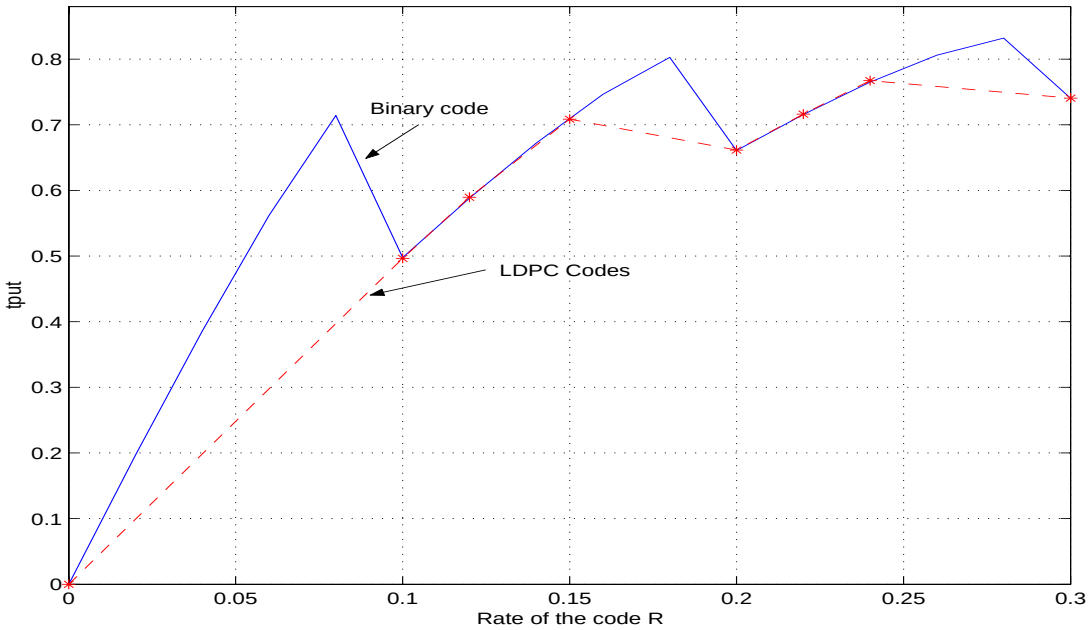


Figure 9: Throughput for $\gamma = 10dB$, zoom in the interval $R = (0.1, 0.3)$ b/s/Hz.

Finite-length codes

Small FER codes or small BER codes?

- In the first case, the retransmission unit is the codeword (as considered so far). The drawback is that we need codes with very good BLOCK error rate (FER).
- An alternative strategy is to split the information message into small packets and, after decoding, perform *selective repeat* of the packets in error only. The drawback is that we need to perform error detection on each of the packets.

Finite length performance: good FER codes

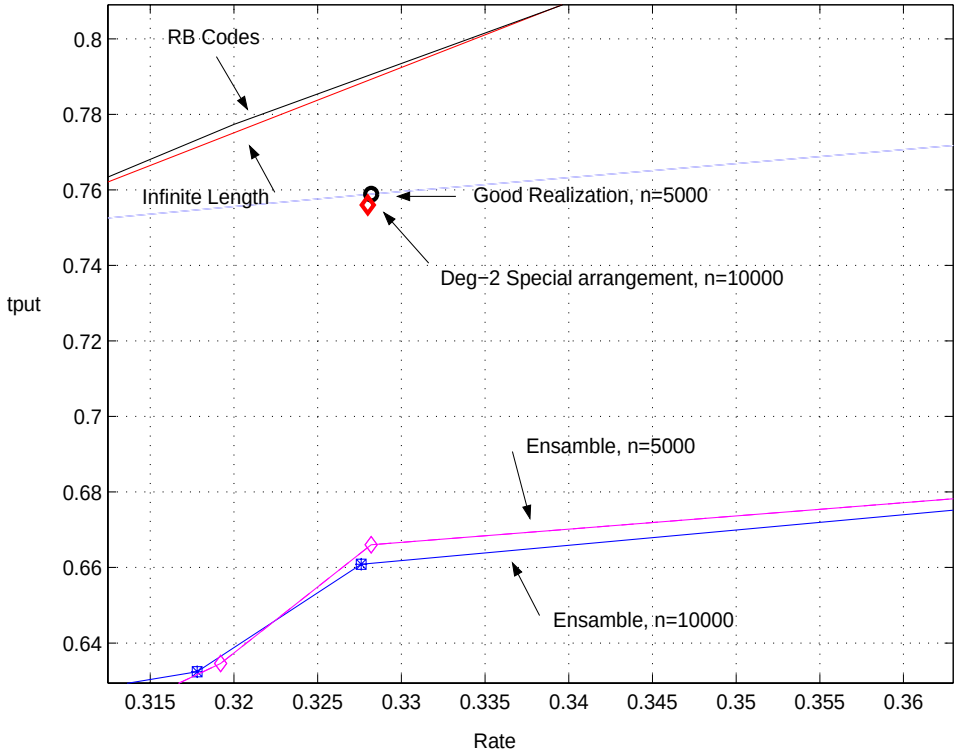


Figure 10: $\gamma = 10$dB.

External Selective Repeat

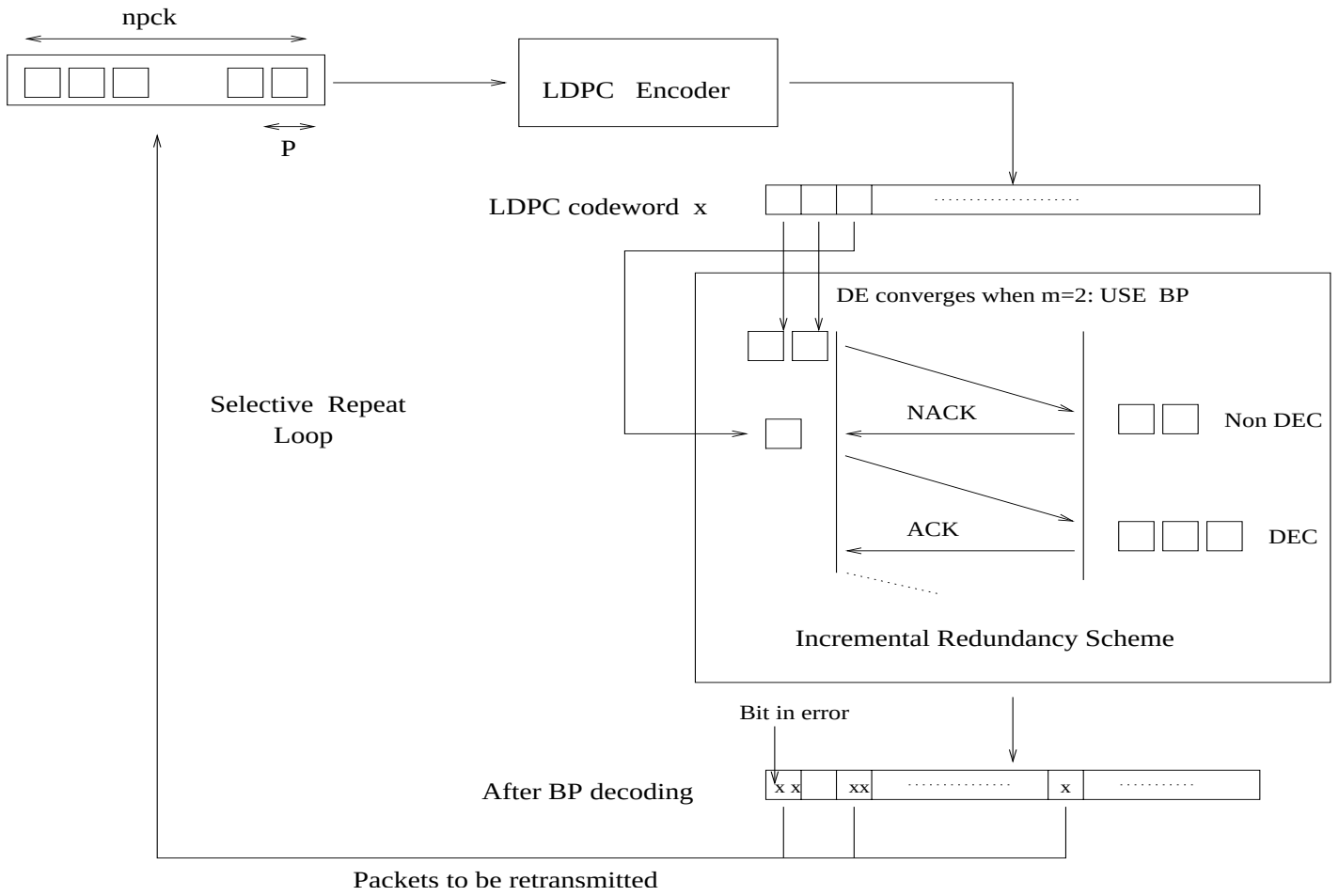
- In order to reduce the gap between infinite and finite length LDPCs, we consider the concatenation with an outer selective-repeat protocol.
- The information block is partitioned into n_p packets of length B bits. Each packet contains its own CRC and error detection can be performed per-packet.
- We re-define the recurrent event $\mathcal{E}_m$ and the random reward $\mathcal{R}_m$: let e_i be the number of packets in error after decoding at step i, let $\mathcal{B}_i = \{\mathrm{DE}_i \text{ converges}, e_i < \delta\}$, then

$$\mathcal{E}_m = \left(\overline{\mathcal{B}}_1, \overline{\mathcal{B}}_2, \ldots, \overline{\mathcal{B}}_{m-1}, \mathcal{B}_m\right)$$

and

$$\mathcal{R}_m = (n_p - e_m)B$$

External Selective Repeat



Renewal-reward throughput analysis

We define the probabilities

$$\begin{cases} \Pr(\mathcal{E}_m) \,=\, \hat{q}(m) & \text{for } m \leq M-1, \\ \Pr(\mathcal{E}_M) \,=\, 1 - \sum_{m=1}^{M-1} \hat{q}(m) & \text{for } m = M. \end{cases}$$

and $\hat{p}(m) \,=\, \Pr\big(\overline{\mathcal{B}}_1, \overline{\mathcal{B}}_2, \ldots, \overline{\mathcal{B}}_{m-1}, \overline{\mathcal{B}}_m\big)$.

The average reward is given by

$$\mathrm{E}\left[\mathcal{R}\right] \,=\, \sum_{m=1}^{M} \sum_{e=0}^{n_p} B(n_p - e) \Pr(e_m = e \,|\, \mathcal{E}_m) \Pr(\mathcal{E}_m)$$

Renewal-reward throughput analysis

After some algebra we obtain

$$\eta = RM \frac{1 - \sum_{m=1}^{M-1} r_m \hat{q}(m) - r_M \hat{p}(M-1)}{1 + \sum_{m=1}^{M-1} \hat{p}(m)}$$

where r_m is the expected fraction of packets in error given the renewal event $\mathcal{E}_m$, i.e.,

$$r_m = \frac{1}{n_p} \sum_{e=0}^{n_p} e \Pr(e_m = e | \mathcal{E}_m)$$

Performance with Outer Selective-Repeat, $R = 0.3$

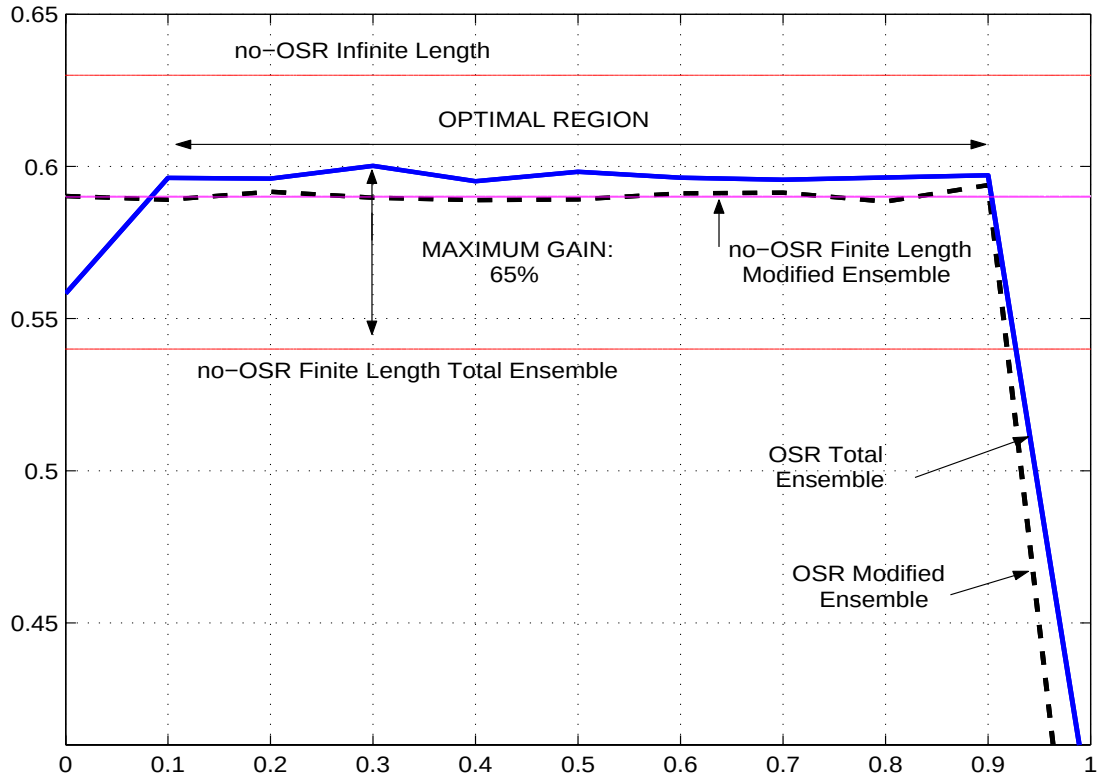


Figure 11: $\gamma = 3$dB.

Performance with Outer Selective-Repeat, $R = 0.3$

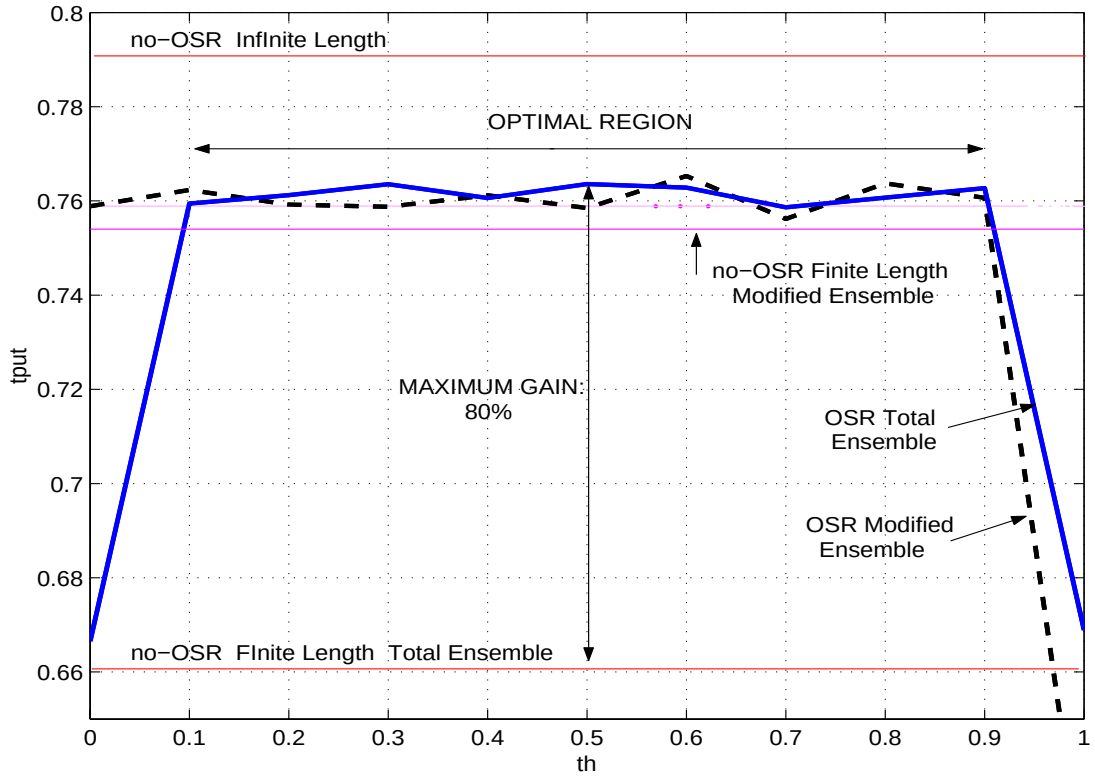
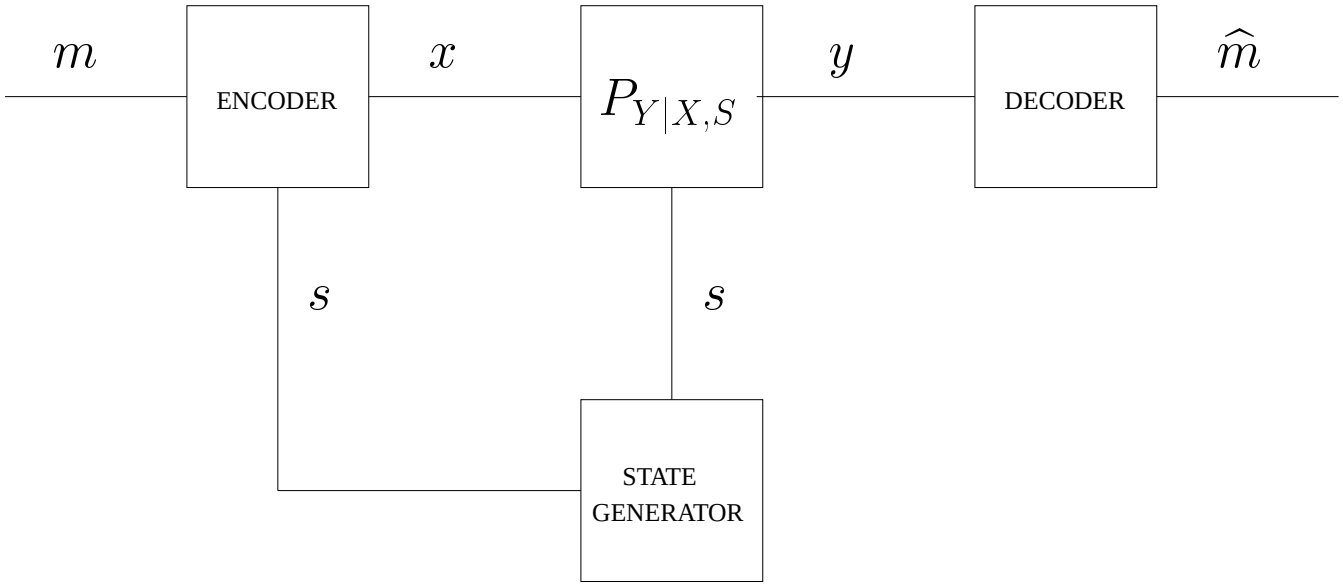


Figure 12: $\gamma = 10$dB.

Part II: writing on Dirty-Paper

Channels with states known to the transmitter



Some history (1)

- The model with **CAUSAL** state knowledge and i.i.d. state sequence was studied by Shannon (1958).
- The model with **NON-CAUSAL** state knowledge and i.i.d. state sequence was studied by Kusnetsov and Tsybakov (1974) and successively generalized by Gel'fand and Pinsker (1980).
- Ahlswede generalized Gel'fand and Pinsker result to the case of arbitrary interference signal (1986).
- Costa considered the special case of an AWGN channel with additive (Gaussian i.i.d.) interference: he nicknamed this problem "Writing on Dirty Paper".

Some history (2)

- The early “Dirty-Paper” works were motivated by storage of information in defective computer memories.
- Recently, “Dirty-Paper” coding gained renewed attention because it arises as the main tool in several important settings such as: Broadcast MIMO channels; Precoding for ISI channels; Data hiding and watermarking.
- In particular, Costa’s result has been generalized in various ways by Lapidoth and Cohen, Shamai, Erez and Zamir, Draper and Wornell and by Cover and Chiang (2000-2002).

Writing on "Dirty Tape" (the causal problem)

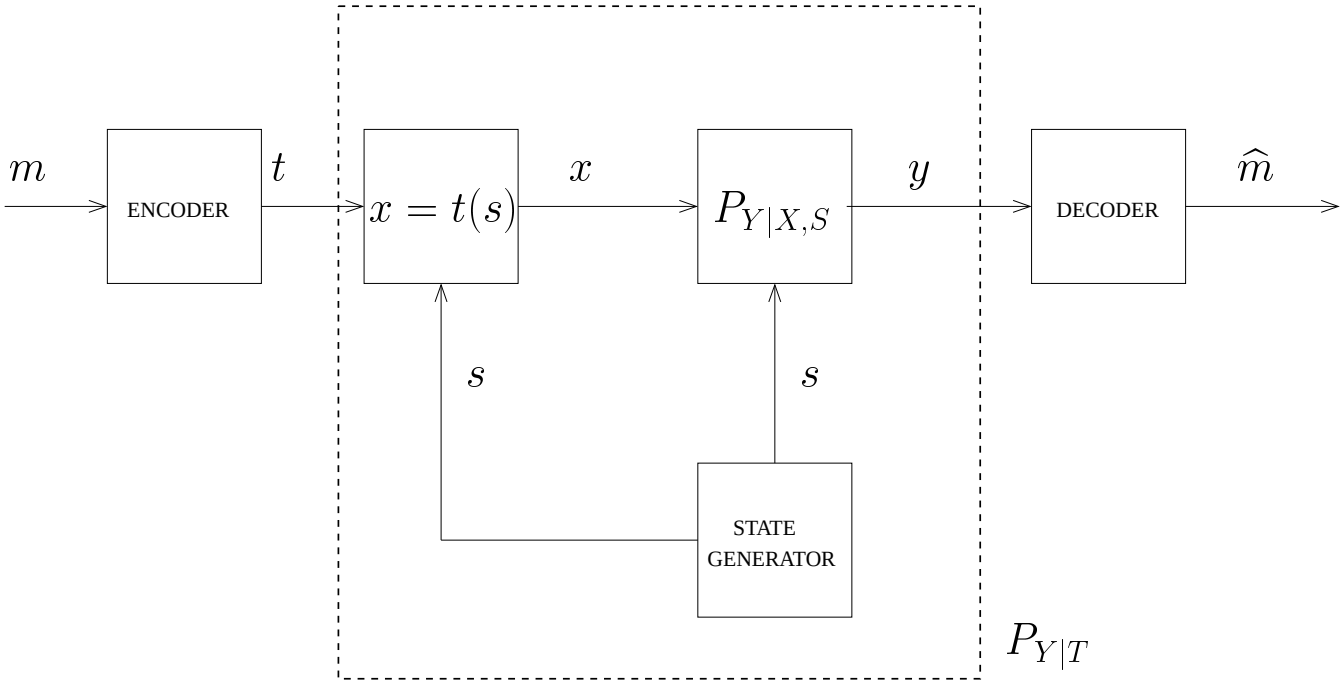
In this case, the encoder is defined by a sequence of functions $\{f_i\}_{i=1}^{n}$ such that $x_i = f_i(m, s_1^i)$. Shannon proved the capacity formula:

$$C = \max_{P_T} I(T;Y)$$

where the input T takes values over the set of memoryless functions ("strategies") $\mathcal{S} \to \mathcal{X}$.

For a power-constrained channel, maximization is over $\mathcal{T}(\mathcal{E})$, the set of probability distributions P_T such that $E[T(S)^2] \leq \mathcal{E}$.

The "Dirty-Tape" associated channel



Writing on "Dirty Paper" (the non-causal problem)

In this case, the encoder is defined by a sequence of functions $\{f_i\}_{i=1}^n$ such that $x_i = f_i(m, s_1^n)$. Gel'fand and Pinsker proved the capacity formula:

$$C = \max_{P_{T|S}} \{I(T;Y) - I(T;S)\}$$

Maximization is over the joint probability assignment

$$P_{S,T,X,Y} = P_S P_{T|S} 1\{X = T(S)\} P_{Y|X,S}$$

with given P_S and $P_{Y|X,S}$.

Remarkably, the non-causal problem includes the causal problem by restricting the maximization over $P_{T,S} = P_T P_S$.

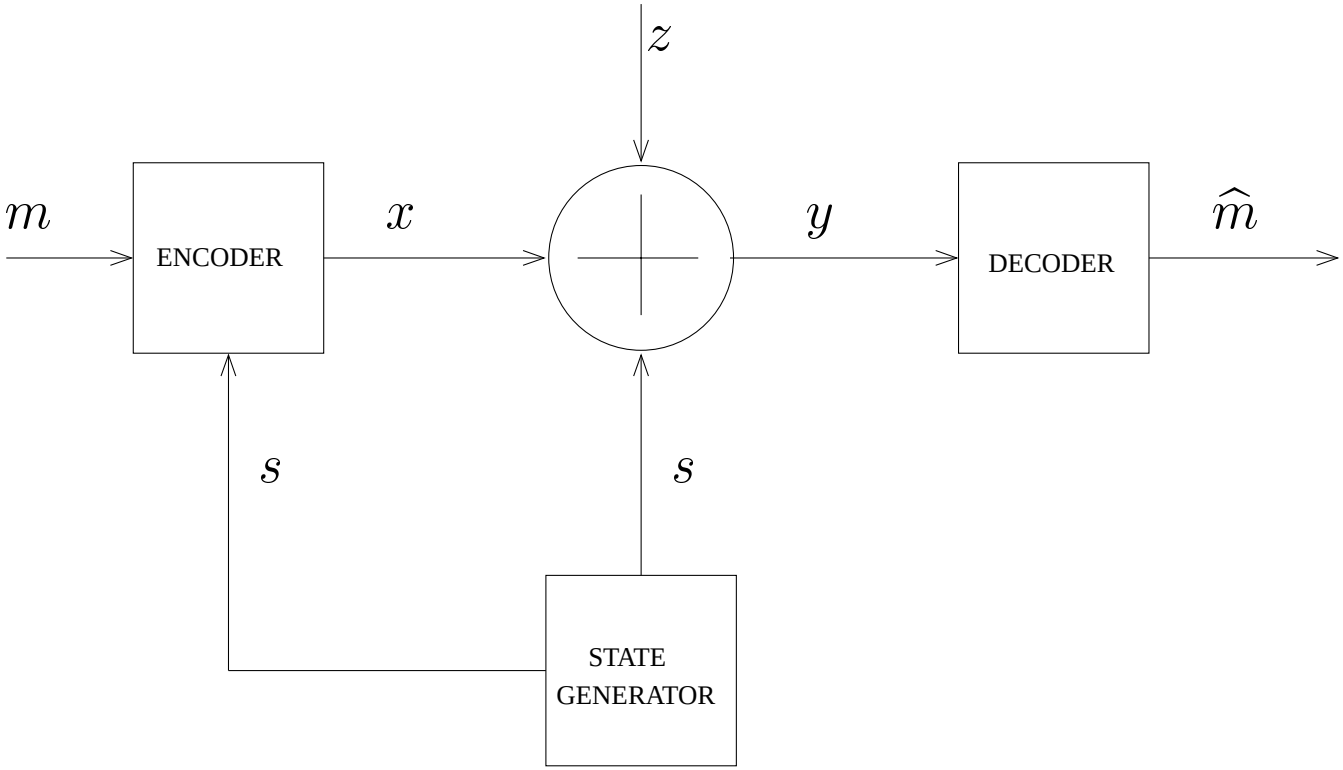
Random binning, quantization and coding

ACHIEVABILITY: Fix R, $P_{T|S}$ and a mapping $\mu : \mathcal{T}^n \times \mathcal{S}^n \to \mathcal{X}^n$; Generate a random codebook $\mathcal{C} \subseteq \mathcal{T}^n$ of size $2^{n(I(T;Y)-\epsilon)}$; Assign at random the codewords of $\mathcal{C}$ to 2^{nR} (disjoint) subsets $\mathcal{C}_m$.

Let $\mathbf{s}$ be the (known) state sequence. In order to send m, find $\mathbf{t} \in \mathcal{C}_m$ such that $(\mathbf{t}, \mathbf{s}) \in \mathcal{A}_\epsilon(n)$ and send $\mathbf{x} = \mu(\mathbf{t}, \mathbf{s})$. At the receiver, after observing the channel output $\mathbf{y}$, find the unique $\widehat{\mathbf{t}}$ such that $(\widehat{\mathbf{t}}, \mathbf{y}) \in \mathcal{A}_\epsilon(n)$ and output $\widehat{m}$ such that $\widehat{\mathbf{t}} \in \mathcal{C}_{\widehat{m}}$.

If $R < I(T;Y) - I(T;S) - 2\epsilon$, then $P_e < \epsilon$ for sufficiently large n.

Additive noise and known interference



Costa's result

Let $S \sim \mathcal{N}(0, Q)$ and $Z \sim \mathcal{N}(0, \sigma^2)$ be i.i.d., and choose

1. $P_{T|S} = \mathcal{N}(\alpha S, \mathcal{E})$;
2. $\mu(\mathbf{t}, \mathbf{s}) = \mathbf{t} - \alpha \mathbf{s}$;

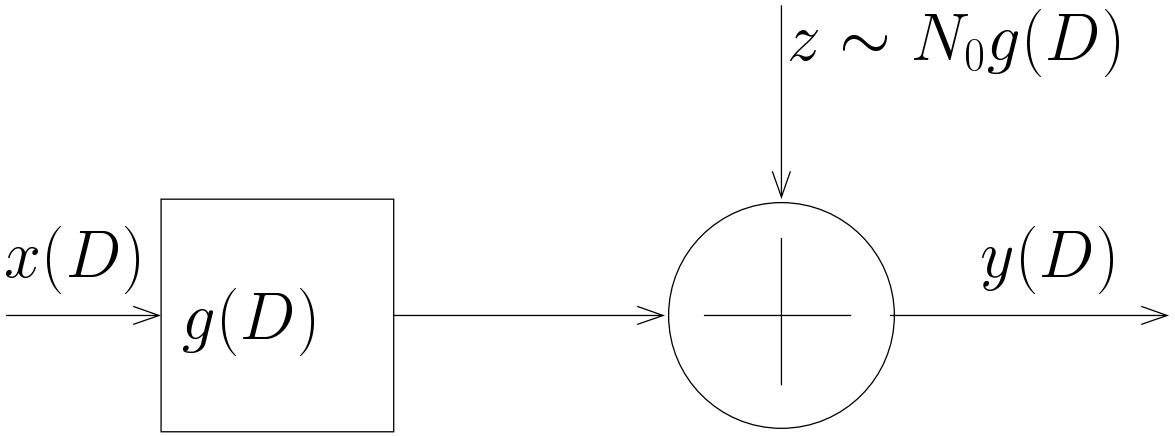
(Notice: $E[T(S)^2] = E[(T - \alpha S)^2] = \mathcal{E}$).

By letting $\alpha = \text{SNR}/(1 + \text{SNR})$, where $\text{SNR} \stackrel{\Delta}{=} \mathcal{E}/\sigma^2$, we have

$$I(T;Y) - I(T;S) = \frac{1}{2} \log_2(1 + \text{SNR})$$

Application to precoding for ISI channels (1)

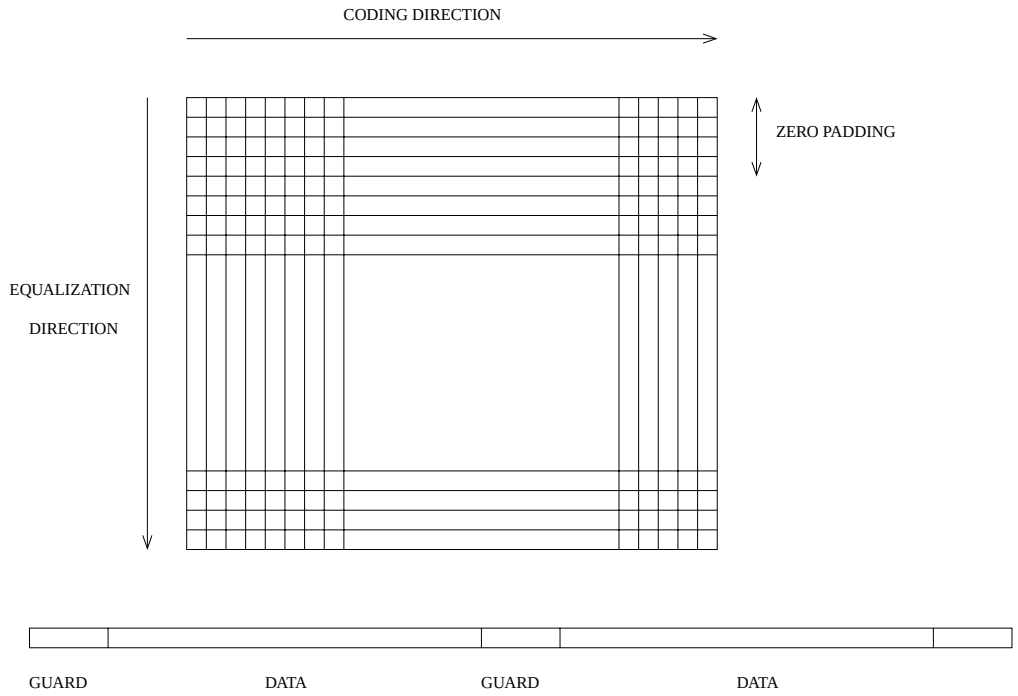
Consider the canonical ISI channel model originated by matched filtering and sampling at the symbol rate



$$g(D) = h(D)h(1/D^*)^*$$

Application to precoding for ISI channels (2)

A capacity-achieving scheme is obtained by coding, zero-padding, interleaving, and MMSE-DFE combined with successive decoding at the receiver (Varanasi and Guess, 2000):



Application to precoding for ISI channels (3)

Equivalently, we can use Costa precoding at the transmitter successive encoding at the transmitter (Zamir, Shamai and Erez, 2002).

In both cases, the achevable rate is

$$R = \log_2(1 + \mathsf{SINR}) = \int_{-1/2}^{1/2} \log_2 \left(1 + \frac{|H(\omega)|^2 \mathcal{E}}{N_0}\right) d\omega$$

where $H(\omega)$ is the total channel response (including the transmit filter).

Application to Gaussian broadcast channels (1)

The capacity region of the degraded Gaussian BC $(X, Y1, Y2)$ is given by

$$0 \leq R_1 \leq \frac{1}{2}\log_2(1+\alpha|h_1|^2\mathcal{E}/\sigma^2); \quad 0 \leq R_2 \leq \frac{1}{2}\log_2\left(1+\frac{(1-\alpha)|h_2|^2\mathcal{E}}{\sigma^2+\alpha|h_2|^2\mathcal{E}}\right)$$

It is achieved by superposition coding and successive decoding, by letting $X = W + U$, with $W \sim \mathcal{N}(0, (1-\alpha)\mathcal{E})$ and $U \sim \mathcal{N}(0, \alpha\mathcal{E})$.

Application to Gaussian broadcast channels (2)

This is a special case of Marton achievable region:

$$\begin{aligned} R_1 &\leq I(W,U;Y_1) \\ R_2 &\leq I(W,V;Y_2) \\ R_1+R_2 &\leq \min\{I(W;Y_1), I(W,Y_2)\} \\ &\quad +I(U;Y_1|W)+I(V;Y_2|W)-I(U;V|W) \end{aligned}$$

Application to Gaussian broadcast channels (3)

A different choice of Marton's auxiliary variables, yielding the same optimal result is $W =$const., $V \sim \mathcal{N}(0, (1-\alpha)\mathcal{E})$ and U achieving Costa's "dirty-paper" capacity for the induced known-interference channel with input X_1 such that $E[X_1^2] \leq \alpha\mathcal{E}$, output Y_1, Gaussian noise Z_1 and interference V.

Clearly, $\max_{U|V}\{I(U;Y_1) - I(U;V)\} = \frac{1}{2}\log_2(1 + \alpha|h_1|^2/\sigma^2)$.

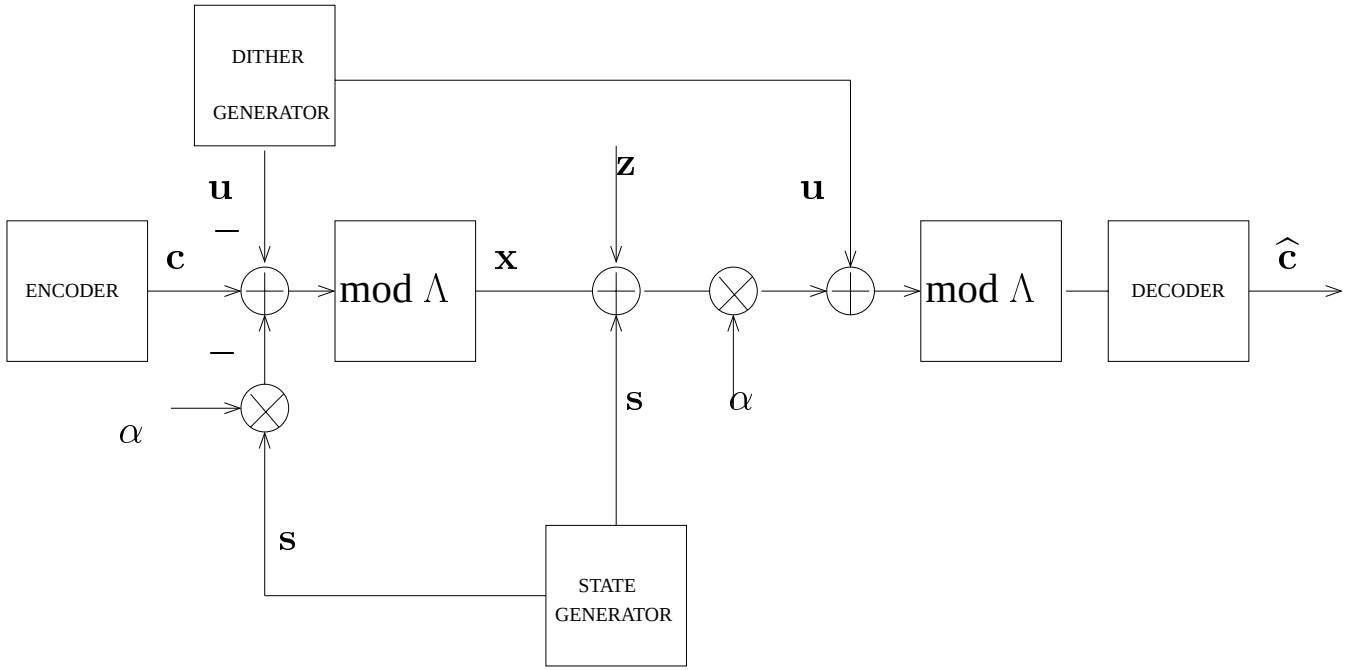
Inflated lattice strategy (1)

Let Λ be a k-dimensional lattice, with fundamental Voronoi cell $\mathcal{V}$ with mean zero and normalized second moment $\mathcal{E}$. Let $\mathbf{U} \sim$Uniform$(\mathcal{V})$ be a random signal known to both Tx and Rx (dither). The codebook $\mathcal{C}$ is chosen such that

$$\mathcal{C} \subseteq \underbrace{\mathcal{V} \times \mathcal{V} \times \cdots \times \mathcal{V}}_{n/k \text{ times}}$$

- Tx: for $\mathbf{v} \in \mathcal{C}$, send $\mathbf{x} = [\mathbf{v} - \alpha\mathbf{s} - \mathbf{u}] \bmod \Lambda$.
- Rx: given the received signal $\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{z}$, compute $\mathbf{y}' = [\alpha\mathbf{y} + \mathbf{u}] \bmod \Lambda$. Then, find $\widehat{\mathbf{v}} = \arg \max_{\mathbf{v} \in \mathcal{C}} p(\mathbf{y}'|\mathbf{v})$.

Inflated lattice strategy (2)



Associated modulo-Λ additive noise channel

Lemma: (Erez, Shamai and Zamir), The channel from $\mathbf{v}$ to $\mathbf{y}'$ induced by the inflated lattice strategy is

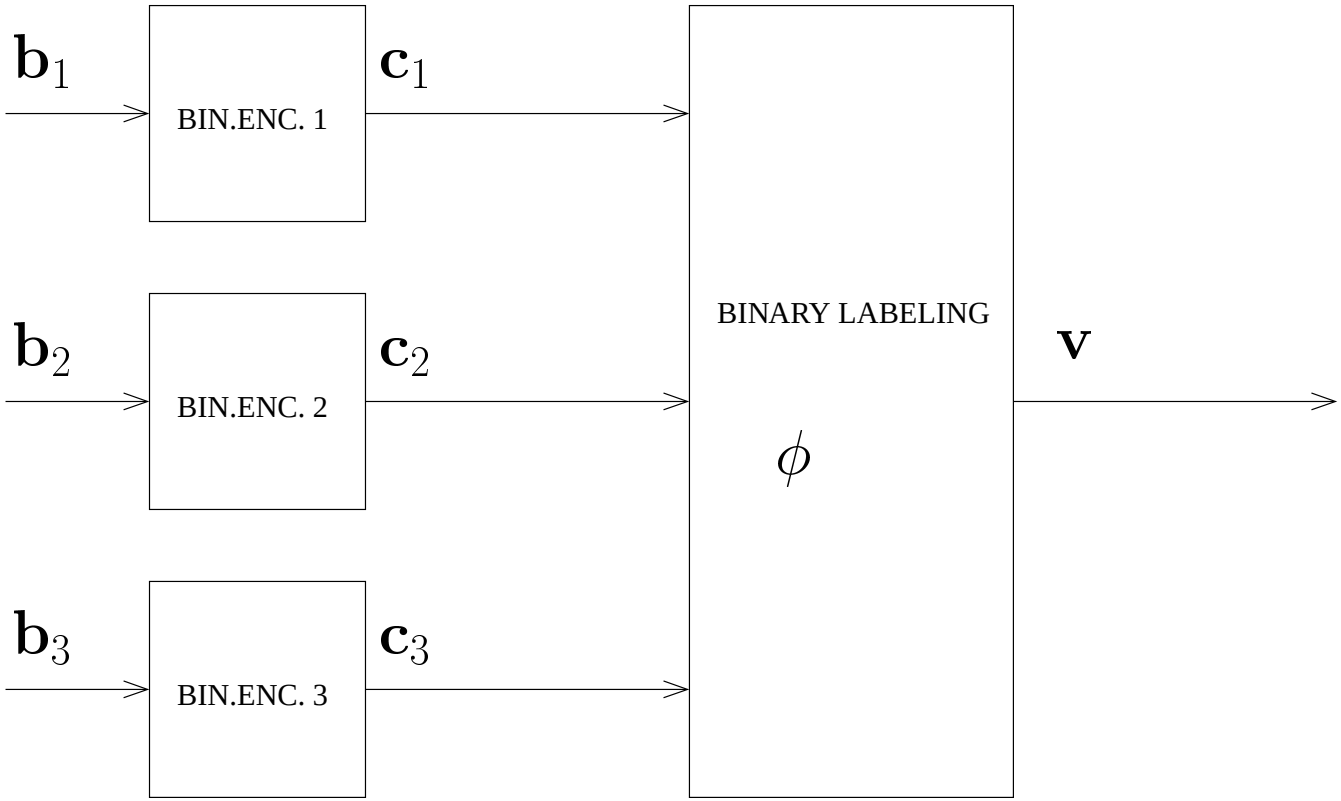
$$\mathbf{Y}' = \mathbf{v} + \mathbf{Z}' \mod \Lambda$$

where $\mathbf{Z}' \sim [(1-\alpha)\mathbf{U} + \alpha\mathbf{Z}] \mod \Lambda$.

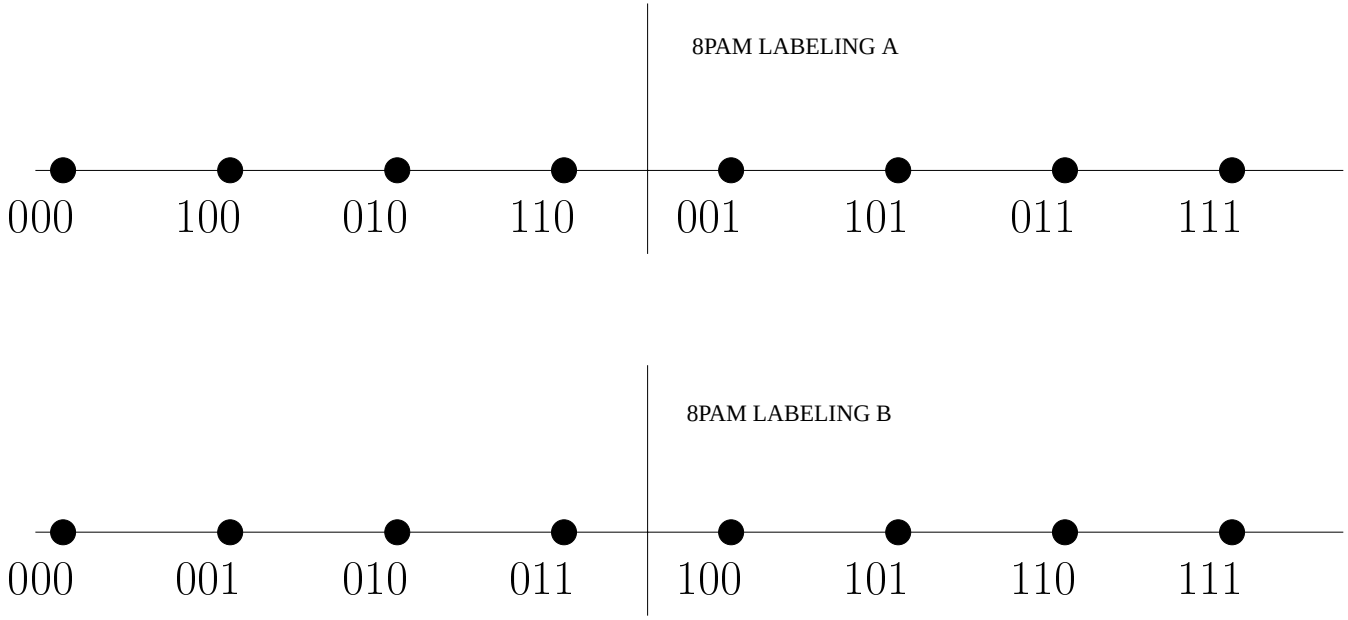
Code construction for the dirty tape

- In the case $k = 1$, $\Lambda = \Delta\mathbb{Z}$, where $\Delta = \sqrt{12\mathcal{E}}$.
- This is equivalent to the case of arbitrary k with the suboptimal choice $\Lambda = \Delta\mathbb{Z}^k$ (cubic lattice).
- As a consequence, the code $\mathcal{C}$ must be constructed over the real interval $[-\Delta/2, \Delta/2]$ (equivalently, it must be a subset of $[-\Delta/2, \Delta/2]^n$).
- A natural choice is to construct $\mathcal{C}$ over an equally spaced M-PAM constellation.

Approach 1: Multilevel Coded Modulation



Binary labeling of M-PAM



Component code rates determination

For a given binary labeling $\phi : \mathbb{F}_2^m \to \mathcal{A}$, we can compute the component code rates by using the mutual information chain rule:

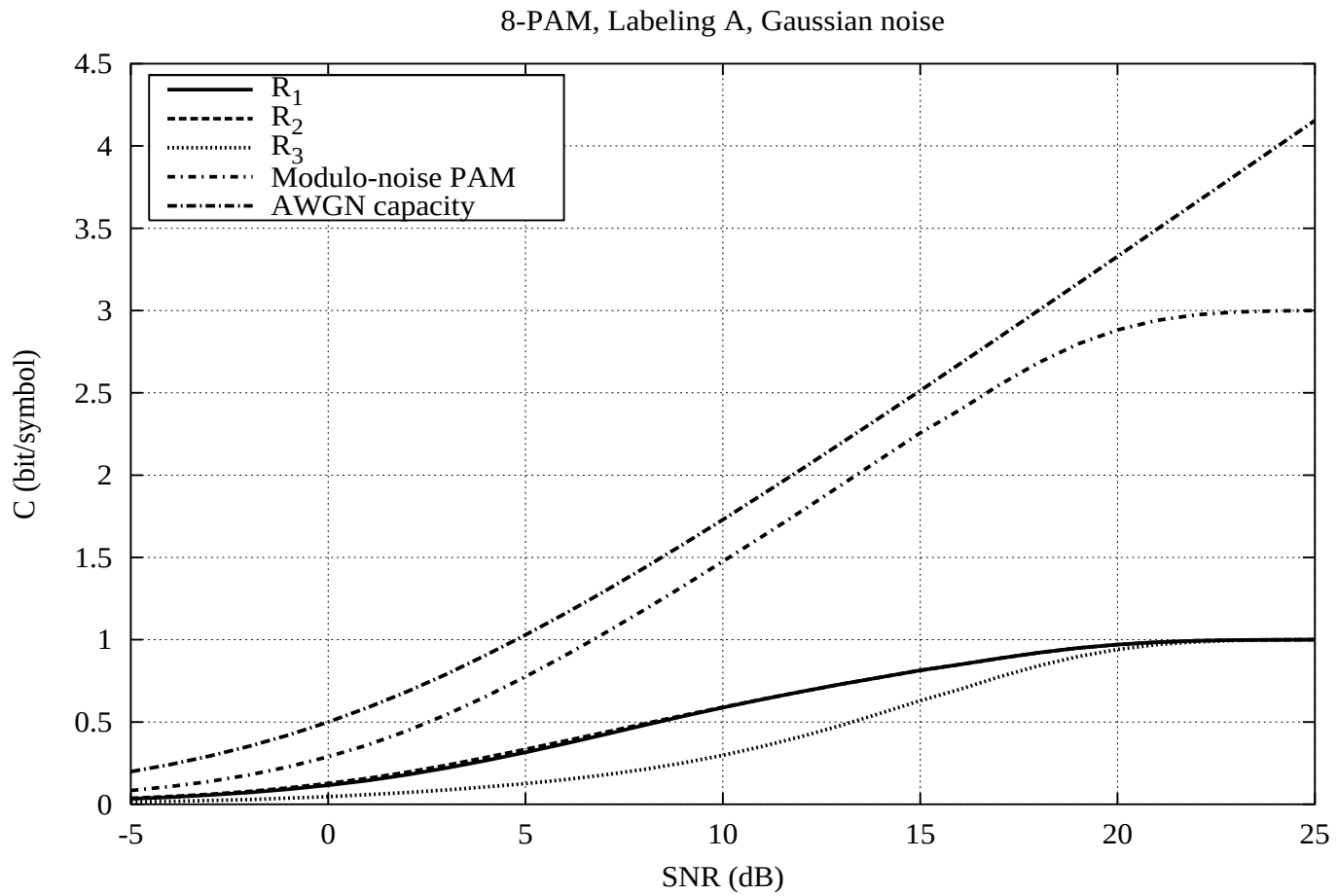
$$I(Y';V) = \sum_{i=1}^{m} R_i = \sum_{i=1}^{m} I(Y; b_i | b_1, \ldots, b_{i-1})$$

where

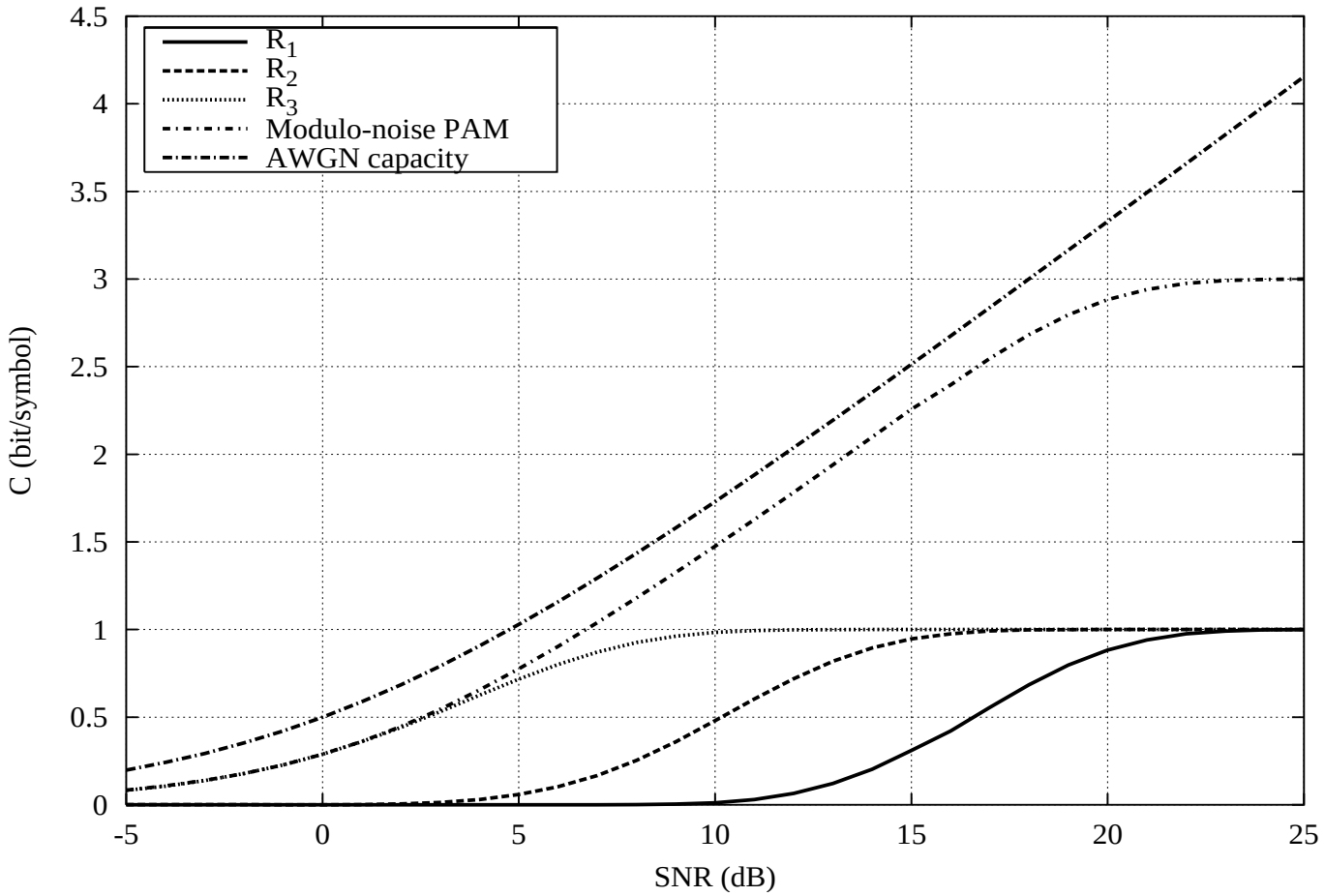
$$R_i = E\left[\log_2 \frac{\sum_{v \in \mathcal{A}(b_1,\ldots,b_i)} p_{Z'}(Y'-v)}{\sum_{u \in \mathcal{A}(b_1,\ldots,b_{i-1})} p_{Z'}(Y'-u)}\right]$$

Notice that the total rate does not depend on the binary labeling.

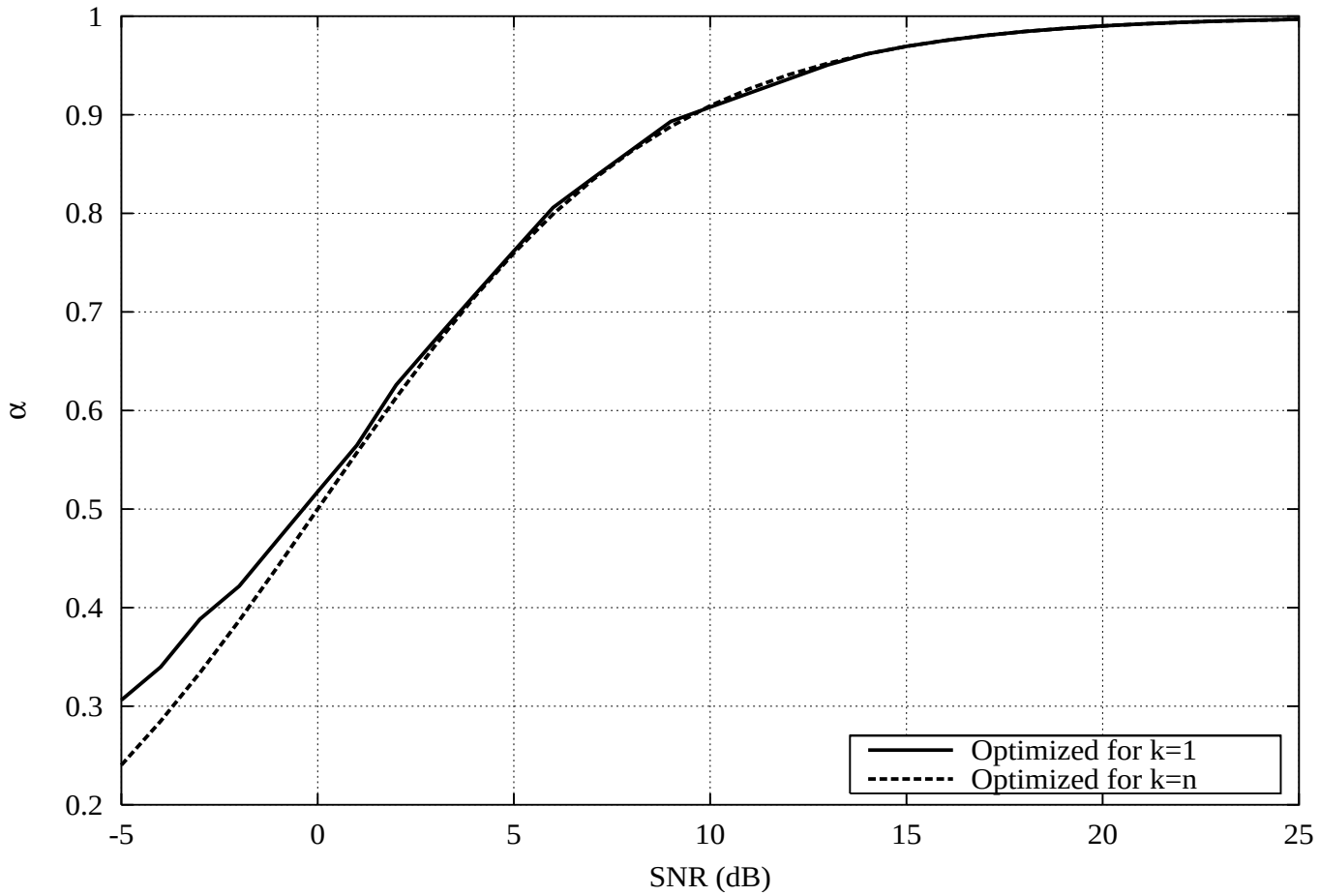
Mutual information (labeling A)



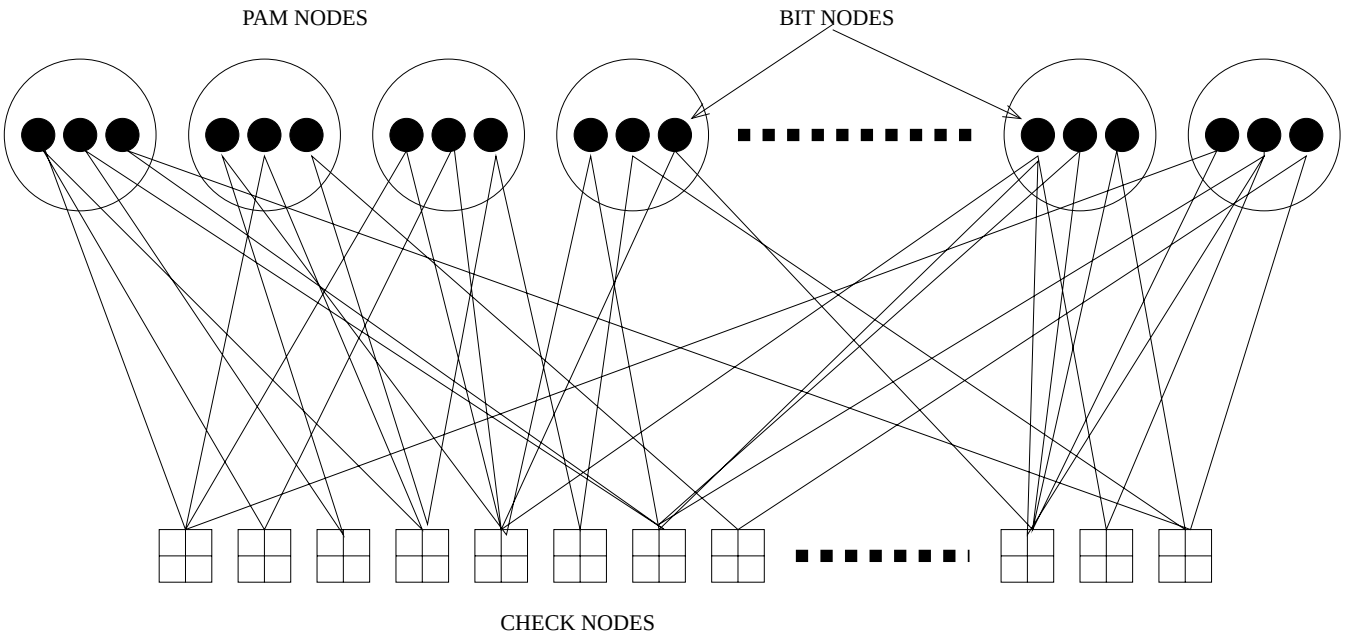
Mutual information (labeling B)



Optimized inflation factor α



Approach 2: LDPC Coded Modulation



LDPC-PAM direct optimization (1)

- We say that a PAM node has type $(d_m, \ldots, d_1)$ if its i-th label bit has degree d_i.
- We enumerate the PAM node types in lexicographic order, and let $d_{t,i}$ be the degree of the i-th bit in type t.
- We let $\lambda_{t,i}$ denote the fraction of edges connected with PAM nodes of type t in label position i.
- For a graph with e edges, the number of PAM nodes of type t is given by $n_t = e\lambda_{t,i}/d_{t,i}$, therefore $\lambda_{t,i}/d_{t,i}$ must not depend on i.

LDPC-PAM direct optimization (2)

- From $\sum_{t,i} \lambda_{t,i} = 1$ we obtain the constraint

$$\sum_t \lambda_{t,1} \sum_{i=1}^{m} \frac{d_{t,i}}{d_{t,1}} = 1$$

- The design coding rate is given by

$$R = m - \frac{\sum_j \rho_j / j}{\sum_t \lambda_{t,1} / d_{t,1}}$$

- We obtain an optimization problem for fixed right-deg sequence, in the variables $\{\lambda_{t,1}\}$

Achievable rate for $M \to \infty$

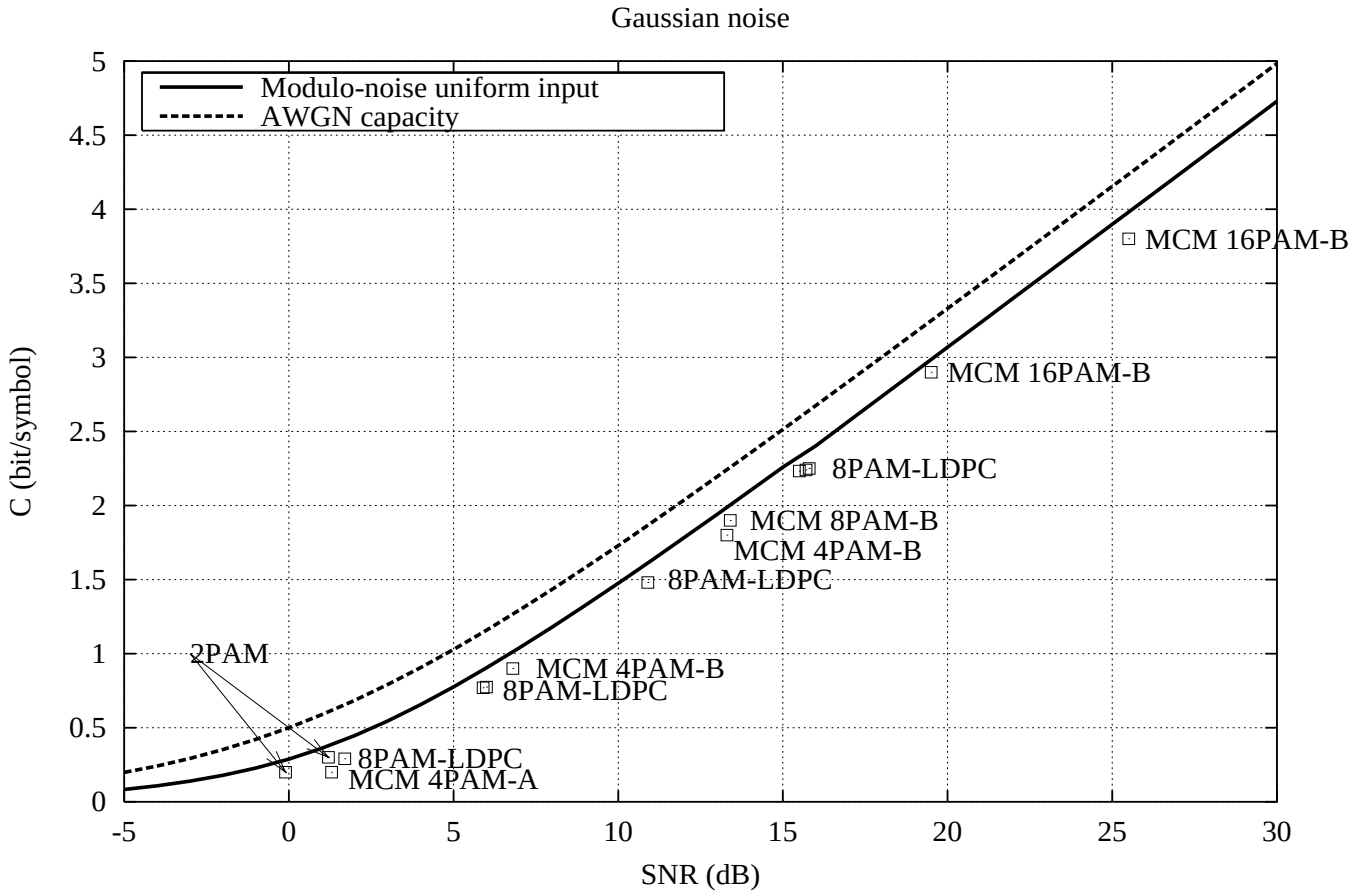
The achievable rate of a code obtained by concatenating a binary linear code (with a one-to-one binary labeling ϕ) with a very large M-PAM signal set is given by

$$I(V;Y') = \log_2 \Delta - h(Z')$$

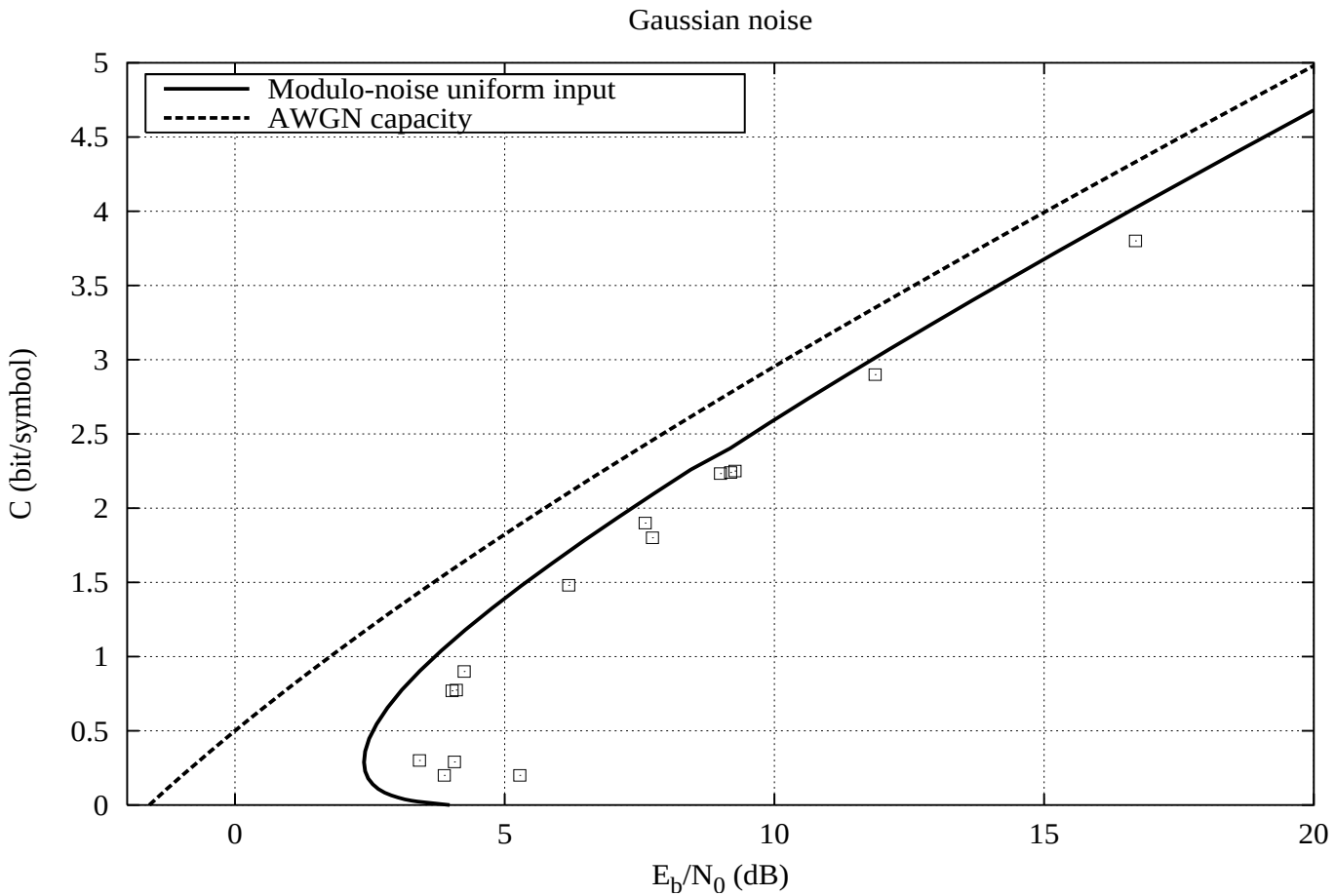
where the modulo-$[-\Delta/2, \Delta/2]$ noise has pdf

$$p_{Z'}(z) = \sum_{k \in \mathbb{Z}} \frac{P_Z\left(\frac{z+k\Delta+(1-\alpha)\Delta/2}{\alpha}\right) - P_Z\left(\frac{z+k\Delta-(1-\alpha)\Delta/2}{\alpha}\right)}{(1-\alpha)\Delta}$$

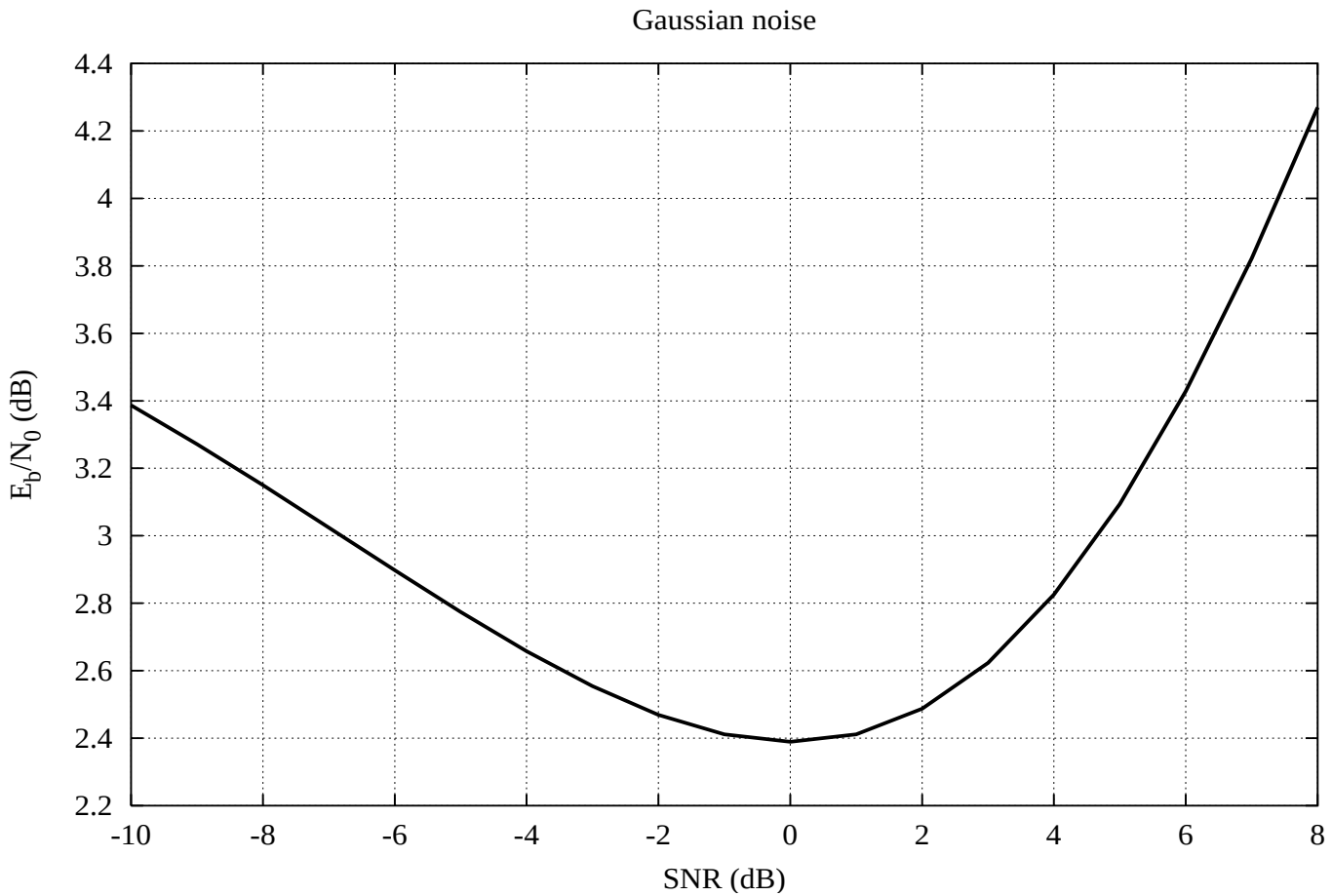
LDPC-coded PAM, $n = 20000$***, at BER*** $\leq 10^{-4}$



LDPC-coded PAM, $n = 20000$***, at BER*** $\leq 10^{-4}$



E_b/N_0 *vs. SNR for* $k = 1$*, infinite PAM*



Preliminary conclusions on code design

- For low SNR, binary coding is sufficient to approach the capacity of the modulo-noise channel.
- There exists an intermediate region of SNR (from 5 to 15 dB) for which designing specific LDPC-coded M-PAM is worthwhile.
- For high SNR, large M-PAM with multilevel coding yield good performance. Only the first (or the first two) levels need coding!

Future work: schemes for anticipation $k > 1$

- In principles, the same direct-optimization approach could be used for LDPC coded-modulation over k-dimensional signal sets, obtained as $\mathbf{Gx}$, where $\mathbf{x} \in \mathcal{A}^k$ and $\mathbf{G}$ is the generator matrix of some good k-dimensional lattice.
- The bit-wise metric computer might be based on soft-output versions of the sphere decoder for general lattices, or on some smarter scheme exploiting the structure of $\mathbf{G}$.
- For large k, the modulo-noise ML metric (the pdf $p_{Z'}$) is hard to compute, especially for lattices with a large kissing number.