Upper Bounds of Channel Capacity for Bipolar Transmission Over Gauss-Markov Fading Channel

Prepared by Ming-Chun Chiang

Advisory by Prof. Po-Ning Chen

Institute of Communications Engineering

National Chiao-Tung University

Hsinchu, Taiwan 300, R.O.C.

May 24, 2005

Outline

- **Introduction**
- The Definition of Channel Capacity
- System Model
- Upper Bounds of the Blind-CSI Capacity for Gauss-Markov Fading Channel
- Conclusions

Introduction

- In this thesis, we focus on the capacity for the time-varying Gauss-Markov fading channels.
- We remark on four different definitions of channel capacities according to whether the transmitter and the receiver have or have not the channel state information (CSI).
- As the true capacity formula for blind-CSI in both transmitter and receiver is hard to obtain, we derive its independent upper bound instead, and establish a close-form expression of the independent bound for any memory order M.
- Discussions are finally given by numerical evaluation of the independent lower bounds of error probability.

- Introduction
- **The Definition of Channel Capacity**
- System Model
- Upper Bounds of the Blind-CSI Capacity for Gauss-Markov Fading Channel
- Conclusions

Capacity definition for memoryless additive channel

- Let $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ denote the input and output sequences of the channel. Denote the noise by $N_1, N_2, \cdots$. Then, a memoryless additive channel could be defined by:

$$Y_i = X_i + N_i, \quad i = 1, 2, \cdots ,$$

where $\{X_i\}_{i=1}^{\infty}$ and $\{N_i\}_{i=1}^{\infty}$ are independent random variables.

- If $\{N_i\}_{i=1}^{\infty}$ is assumed to be Gaussian distributed with white power spectral density (PSD), then a memoryless additive white Gaussian noise (AWGN) channel is established.

- For discrete input X and discrete output Y, the mutual information is written as:

$$
\begin{aligned}
I(X;Y) &\triangleq \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{X,Y}(x,y)\cdot\log\left(\frac{P_{X,Y}(x,y)}{P_X(x)\cdot P_Y(y)}\right) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)P_{Y|X}(y|x)\cdot\log\left(\frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_X(x')P_{Y|X}(y|x')}\right) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)P_N(y-x)\cdot\log\left(\frac{P_N(y-x)}{\displaystyle\sum_{x'\in\mathcal{X}} P_X(x')P_N(y-x')}\right)
\end{aligned}
$$

- Based on this definition, the capacity-cost function for memoryless additive channel subject to input average power constraint $E[X^2]\le S$ is given by:

$$
C(S) = \max_{\{P_X \,:\, E[X^2]\le S\}} I(X;Y).
$$

Capacity for time-invariant flat fading additive channel

- The channel model for time-invariant flat fading additive channel is defined as:

$$Y_i = H \cdot X_i + N_i, \quad i = 1, 2, \cdots$$

where H is an independent time-invariant random variable.

- The capacity-cost function for time-invariant flat fading channel given $[H = h]$ is equal to:

$$C_h(S) = \max_{\{P_X \,:\, E[X^2] \le S\}} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_X(x) P_N(y - hx) \log \frac{P_N(y - hx)}{\sum_{x' \in \mathcal{X}} P_X(x') P_N(y - hx')}$$

where $P_{X|H}$ is replaced by P_X since X is independent of H.

- For continuous channel output alphabet, same derivations can give that:

$$C_h(S) = \max_{\{P_X \,:\, E[X^2]\leq S\}} \sum_{x\in\mathcal{X}} P_X(x) \int_{\mathcal{Y}} p_N(y-Hx) \log \frac{p_N(y-Hx)}{\sum_{x'\in\mathcal{X}} P_X(x')p_N(y-Hx')} dy$$

- The average signal-to-noise ratio (SNR) is usually given by:

$$\text{SNR} = \frac{E[E[H^2X^2|H]]}{E[N^2]} = E[H^2]\frac{E[X^2]}{E[N^2]}.$$

Researchers will sometimes fix $E[H^2]=1$, and varies $E[N^2]$ to examine the system performance of their coding scheme over such a channel.

Definition of average capacity $\overline{C(S)}$

- Some researchers focus on the average capacity for fading channel, such as

$$\begin{aligned}\overline{C(S)} &\triangleq \int_{\mathcal{H}} p_H(h)\cdot C_h(S)dh. \\ &= \int_{\mathcal{H}} p_H(h)\cdot\left[\max_{\{P_X\ :\ E[X^2]\le S\}} I(X;Y|h)\right]dh.\end{aligned}$$

- The operational meaning of $\overline{C(S)}$:
  - It is the underlying limit for a system in which both the transmitter and the receiver have perfect information about the channel state H.
  - The transmitter and receiver will always employ the best encoder and decoder corresponding to the perfectly estimated $H = h$ to achieve $C_h(S)$.

Definition of capacity $C(S)$ for time-invariant flat fading additive channel

- In situation where both the transmitter and the receiver are unknown of the channel state, the capacity-cost function should be given by:

$$
\begin{aligned}
& C(S) \\
\triangleq\ & \max_{\{P_X\ :\ E[X^2]\leq S\}} I\left(X;Y\right) \\
=\ & \max_{\{P_X\ :\ E[X^2]\leq S\}} \left[h\left(Y\right) - h\left(Y|X\right)\right] \\
=\ & \max_{\{P_X\ :\ E[X^2]\leq S\}} \sum_{x\in\mathcal{X}} \int_{\mathcal{Y}} p_{X,Y}\left(x,y\right) \log\left(\frac{p_{X,Y}\left(x,y\right)}{P_X\left(x\right)\cdot p_Y\left(y\right)}\right) dy \\
=\ & \max_{\{P_X\ :\ E[X^2]\leq S\}} \int_{\mathcal{H}} p_H(h) \sum_{x\in\mathcal{X}} \int_{\mathcal{Y}} p_{X,Y|H}(x,y|h) \log\left(\frac{p_{Y|X}\left(y|x\right)}{p_Y\left(y\right)}\right) dydh
\end{aligned}
$$

Remarks on four definitions of channel capacities for time-invariant flat fading additive channel

- If only the receiver knows the channel state, the transmitter cannot vary its encoding rule according the channel state; hence, there can be only one maximization input statistics in the channel capacity formula:

$$C^{(R)}(S) \triangleq \max_{\{P_X \,:\, E[X^2]\le S\}} \int_{\mathcal{H}} p_H(h) \sum_{x\in\mathcal{X}} P_X(x) \int_{\mathcal{Y}} p_{Y|X,H}(y|x,h) \cdot \log\left(\frac{p_{Y|X,H}(y|x,h)}{p_{Y|H}(y|h)}\right) dydh.$$

- On the other hand, if only the transmitter is aware of the CSI, the capacity formula will become:

$$\overline{C^{(T)}(S)} \triangleq \int_{\mathcal{H}} p_H(h) \max_{\{P_X \,:\, E[X^2]\le S\}} \sum_{x\in\mathcal{X}} P_X(x) \int_{\mathcal{Y}} p_{Y|X,H}(y|x,h) \cdot \log\left(\frac{p_{Y|X}(y|x)}{p_Y(y)}\right) dydh$$

- The operational characteristics of four definitions of capacities.

Capacity-cost function	TX Knows CSI?	RX Knows CSI?
$C(S)$	No	No
$\overline{C^{(T)}(S)}$	Yes	No
$C^{(R)}(S)$	No	Yes
$\overline{C(S)}$	Yes	Yes

- Comments on four definitions of channel capacities for time-invariant flat fading additive channel:
  - If a perfect estimate of H is available to the receiver, then the receiver can surely take advantage of $p_{X,Y|H}$ and $p_{Y|H}$ at the decoding process.

- If the receiver knows nothing about H, it can only use the average of $p_{X,Y|H}$ and $p_{Y|H}$, i.e. $p_{X,Y}$ and p_Y , in its decoding process.
- If the transmitter has full knowledge of CSI, then the encoding rule can vary according to H, such that the maximization operation shall be inside the integral with respect p_H.
- If the transmitter has no control of CSI, then the transmitter can only fix the encoding rule regardless of the CSI; hence, the maximization operation shall be placed outside the integral with respect to p_H.

• In our research, we assume antipodal transmission with input alphabet $\{-s, +s\}$ such that the power constraint on the input becomes $E[X^2] = s^2 \leq S$:

- Introduction
- The Definition of Channel Capacity
- **System Model**
- Upper Bounds of the Blind-CSI Capacity for Gauss-Markov Fading Channel
- Conclusions

Data Model

- In our system, we assume that binary phase shift keying (BPSK) signaling is transmitted as channel input.

Channel Model

- A frequency-selective fast fading channel can be modelled by:

$$y_k = \boldsymbol{x}_k^T \boldsymbol{h}_k + z_k = \begin{bmatrix} x_k & x_{k-1} & \cdots & x_{k-M+1} \end{bmatrix} \begin{bmatrix} h_{k,1} \\ h_{k,2} \\ \vdots \\ h_{k,M} \end{bmatrix} + z_k, \quad k = 1, \cdots, n$$

  where $\boldsymbol{x}_k$ and $\boldsymbol{h}_k$ denote the channel input vector and complex column vector, respectively, M is the time spread or temporal channel memory, and z_k is the complex memoryless Gaussian noise at time k with variance $E[z_k z_k^*] = \sigma_z^2$.

- The pdf of the received vector $\boldsymbol{y}$ given $\boldsymbol{x}$ and $\boldsymbol{H}$ is equal to:

$$f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{H}) = \frac{1}{(\pi\sigma_z^2)^n} \prod_{k=1}^{n} \exp\left(-\frac{\left| y_k - \boldsymbol{x}_k^T \boldsymbol{h}_k \right|^2}{\sigma_z^2} \right).$$

- We focus on the channel capacity at the situation that both the transmitter and the receiver are unaware of (and hence do not need to estimate) the channel state.
- We therefore need to compute $f\left(\boldsymbol{y}|\boldsymbol{x}\right)$

$$f\left(\boldsymbol{y}|\boldsymbol{x}\right) = \int_{\boldsymbol{\mathcal{H}}} f\left(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{H}\right) f\left(\boldsymbol{H}\right) d\boldsymbol{H}.$$

The statistics of channel state

- It remains to define $f(\boldsymbol{H})$ to establish $f(\boldsymbol{y}|\boldsymbol{x})$.
- A frequently used fading statistics is the **Guass-Markov** , which defines the statistics of the fading through a recursive first-order Markovian equation as:

$$\boldsymbol{h}_k = \alpha \boldsymbol{h}_{k-1} + \boldsymbol{v}_k,$$

  - $\boldsymbol{v}_k$ is complex, white, Gaussian distributed with mean $\boldsymbol{d}$ and covariance matrix $\boldsymbol{C}$.
  - The complex-valued constant α is a first-order Markov factor usually chosen according to $|\alpha| = e^{-\omega T}$, where T is the system sampling period and ω/π is the Doppler spread.
  - The initial channel coefficient $\boldsymbol{h}_0$ is assumed to be perfectly estimated such that $\boldsymbol{h}_0$ is treated as a known constant.

- Since $f(\boldsymbol{h}_k|\boldsymbol{h}_{k-1})$ is complex Gaussian distributed with mean $\alpha\boldsymbol{h}_{k-1}+\boldsymbol{d}$ and covariance matrix $\boldsymbol{C}$,

$$
\begin{aligned}
f(\boldsymbol{H}) &= f(\boldsymbol{h}_1)\prod_{k=2}^{n} f(\boldsymbol{h}_k|\boldsymbol{h}_{k-1}) \\
&= \frac{1}{|\pi\boldsymbol{C}|^n}\prod_{k=1}^{n}\exp\left\{-\left(\boldsymbol{h}_k-\alpha\boldsymbol{h}_{k-1}-\boldsymbol{d}\right)^H\boldsymbol{C}^{-1}\left(\boldsymbol{h}_k-\alpha\boldsymbol{h}_{k-1}-\boldsymbol{d}\right)\right\}
\end{aligned}
$$

The channel transition probability of Gauss-Markov fading channel

- As a result,

$$
\begin{aligned}
f\left(\boldsymbol{y}|\boldsymbol{x}\right) &= \frac{1}{\left(\pi\sigma_z^2\right)^n \left|\pi\boldsymbol{C}\right|^n} \int_{\boldsymbol{\mathcal{H}}} \prod_{k=1}^{n} \left[\exp\left(-\frac{\left|y_k - \boldsymbol{x}_k^T \boldsymbol{h}_k\right|^2}{\sigma_z^2}\right)\right. \\
&\quad \left.\cdot \exp\left(-\left(\boldsymbol{h}_k - \alpha\boldsymbol{h}_{k-1} - \boldsymbol{d}\right)^H \boldsymbol{C}^{-1} \left(\boldsymbol{h}_k - \alpha\boldsymbol{h}_{k-1} - \boldsymbol{d}\right)\right)\right] d\boldsymbol{H} \\
&= \frac{1}{\left(\pi\sigma_z^2\right)^n \left|\pi\boldsymbol{C}\right|^n} \left[\prod_{k=1}^{n} \left|\pi\boldsymbol{G}_k\right| \exp\left(-\frac{\left|y_k\right|^2}{\sigma_z^2}\right)\right] \\
&\quad \cdot \prod_{k=1}^{n-1} \exp\left[\left(\boldsymbol{q}_k - \alpha^*\boldsymbol{C}^{-1}\boldsymbol{d}\right)^H \boldsymbol{G}_k \left(\boldsymbol{q}_k - \alpha^*\boldsymbol{C}^{-1}\boldsymbol{d}\right) - \boldsymbol{d}^H\boldsymbol{C}^{-1}\boldsymbol{d}\right] \\
&\quad \cdot \exp\left[\boldsymbol{q}_n^H \boldsymbol{G}_n \boldsymbol{q}_n - \left(\alpha\boldsymbol{h}_0 + \boldsymbol{d}\right)^H \boldsymbol{C}^{-1} \left(\alpha\boldsymbol{h}_0 + \boldsymbol{d}\right)\right],
\end{aligned}
$$

where

$$\boldsymbol{G}_k = \begin{cases} \left(\frac{\boldsymbol{x}_1^*\boldsymbol{x}_1^T}{\sigma_z^2} + (1+|\alpha|^2)\boldsymbol{C}^{-1}\right)^{-1}, & \text{if } k=1 \\ \left(\frac{\boldsymbol{x}_k^*\boldsymbol{x}_k^T}{\sigma_z^2} + (1+|\alpha|^2)\boldsymbol{C}^{-1} - |\alpha|^2\boldsymbol{C}^{-1}\boldsymbol{G}_{k-1}\boldsymbol{C}^{-1}\right)^{-1}, & \text{if } 1<k<n \\ \left(\frac{\boldsymbol{x}_n^*\boldsymbol{x}_n^T}{\sigma_z^2} + \boldsymbol{C}^{-1} - |\alpha|^2\boldsymbol{C}^{-1}\boldsymbol{G}_{n-1}\boldsymbol{C}^{-1}\right)^{-1}, & \text{if } k=n \end{cases}$$

and

$$\boldsymbol{q}_k = \begin{cases} \frac{y_1\boldsymbol{x}_1^*}{\sigma_z^2} + \boldsymbol{C}^{-1}\boldsymbol{d} + \alpha\boldsymbol{C}^{-1}\boldsymbol{h}_0, & \text{if } k=1 \\ \frac{y_k\boldsymbol{x}_k^*}{\sigma_z^2} + \boldsymbol{C}^{-1}\boldsymbol{d} + \alpha\boldsymbol{C}^{-1}\boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1} - |\alpha|^2\boldsymbol{C}^{-1}\boldsymbol{G}_{k-1}\boldsymbol{C}^{-1}\boldsymbol{d}, & \text{if } 1<k\le n \end{cases}$$

The detail to derive the above result is described in the Appendix.

- Introduction
- The Definition of Channel Capacity
- System Model
- **Upper Bounds of the Blind-CSI Capacity for Gauss-Markov Fading Channel**
- Conclusions

Independent bound

- We focus on the independent bound of the channel capacity without knowing the channel fading both at the transmitter and at the receiver.
- First, assuming that the channel is reset every n symbols, we obtain:

$$C(S) \triangleq \frac{1}{n} \max_{\{P_{\boldsymbol{x}} \,:\, \frac{1}{n}\mathrm{tr}(E[\boldsymbol{x}^H \boldsymbol{x}]) \leq S\}} I(\boldsymbol{x}; \boldsymbol{y}),$$

  – $\boldsymbol{x}$ is formerly defined as $[x_{2-M}, \cdots, x_n]^T$.

  – Since $x_{2-M}, \cdots, x_0$ are nothing to do with the information transmission, we will abuse notation $\boldsymbol{x}$ as $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$ without ambiguity. Thus, we can equivalently replace $\boldsymbol{x}$ by X^n to result in:

$$C(S) \triangleq \frac{1}{n} \max_{\{P_{X^n} \,:\, \frac{1}{n}\mathrm{tr}(\boldsymbol{\Lambda}_X) \leq S\}} I(X^n; Y^n). \qquad (1)$$

– By elementary information theory operation,

$$\mathrm{I}\left(Y^n; X^n\right) = h(Y^n|X^n) - h(X^n)$$

where $h(\cdot)$ represents the differential entropy operation.

- Although the close-form expression for $f\left(\boldsymbol{y}|\boldsymbol{x}\right)$ is established, it is still hard to evaluate the capacity in (1). However, an upper bound based on the information-theoretical independent bound can be obtained as follows:

$$I\left(X^n; Y^n\right) \leq I\left(X_1; Y_1\right) + \cdots + I\left(X_n; Y_n\right).$$

- We then derive:

$$\begin{aligned}
&\max_{\{P_{\mathbf{X}}:\frac{1}{n}\sum_{i=1}^{n}E[X_i^2]\le S\}} I\left(X^n;Y^n\right) \\
\le &\max_{\{P_{\mathbf{X}}:\frac{1}{n}\sum_{i=1}^{n}E[X_i^2]\le S\}} \left[I\left(X_1;Y_1\right)+\cdots+I\left(X_n;Y_n\right)\right] \\
= &\max_{\{P_{\mathbf{X}}:(\forall\ i)E[X_i^2]\le S\}} \left[I\left(X_1;Y_1\right)+\cdots+I\left(X_n;Y_n\right)\right] \qquad (2) \\
\le &\max_{\{P_{\mathbf{X}}:(\forall\ i)E[X_i^2]\le S\}} I\left(X_1;Y_1\right)+\cdots+\max_{\{P_{\mathbf{X}}:(\forall\ i)E[X_i^2]\le S\}} I\left(X_n;Y_n\right),
\end{aligned}$$

where (2) holds since in our system setting, every $E[X_i^2]$ is equal to s^2 due to $x_i \in \{-s,+s\}$ for every i. Note that $\sum_{i=1}^{n} E[X_i^2] = \mathrm{tr}\left(\mathbf{\Lambda}_X\right)$.

- Let $C_k(S)$ denote the maximual mutual information of $I(X_k;Y_k)$ under input power constraints $E[X_k^2] \le S$. Then,

$$C(S) \le \frac{1}{n}\left[C_1(S)+\cdots+C_n(S)\right].$$

Derivation of $f(y_k|\boldsymbol{x}_k)$ and $f(y_k)$

In order to evaluate the independent bound of channel capacity, $f(y_k|\boldsymbol{x}_k)$ and $f(y_k)$ have to be obtained first.

- First of all, we derive:

$$\begin{aligned}
& f\left(y_k|\boldsymbol{x},\boldsymbol{h}_1,\cdots,\boldsymbol{h}_n\right) \\
\triangleq\ & \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} f\left(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}_1,\cdots,\boldsymbol{h}_n\right)dy_1\cdots dy_{k-1}dy_{k+1}\cdots dy_n \\
=\ & \frac{1}{(\pi\sigma_z^2)^n}\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}\prod_{m=1}^{n}\exp\left(-\frac{\left|y_m-\boldsymbol{x}_m^T\boldsymbol{h}_m\right|^2}{\sigma_z^2}\right)dy_1\cdot\cdot dy_{k-1}dy_{k+1}\cdot\cdot dy_n \\
=\ & \frac{1}{\pi\sigma_z^2}\exp\left(-\frac{\left|y_k-\boldsymbol{x}_k^T\boldsymbol{h}_k\right|^2}{\sigma_z^2}\right) \\
=\ & f\left(y_k|\boldsymbol{x}_k,\boldsymbol{h}_k\right),
\end{aligned}$$

where $\mathcal{C}$ denote the set of all complex numbers.

- Then, we notice that:

$$
\begin{aligned}
& f\left(y_k|\boldsymbol{x}\right) \\
= & \int_{\mathcal{C}^M} \cdot\cdot \int_{\mathcal{C}^M} f\left(y_k|\boldsymbol{x}, \boldsymbol{h}_1, \cdot\cdot, \boldsymbol{h}_n\right) f\left(\boldsymbol{h}_n|\boldsymbol{h}_{n-1}\right) \cdot\cdot f\left(\boldsymbol{h}_2|\boldsymbol{h}_1\right) f\left(\boldsymbol{h}_1\right) d\boldsymbol{h}_1 \cdot\cdot d\boldsymbol{h}_n \\
= & \int_{\mathcal{C}^M} \cdots \int_{\mathcal{C}^M} f\left(y_k|\boldsymbol{x}_k, \boldsymbol{h}_k\right) f\left(\boldsymbol{h}_n|\boldsymbol{h}_{n-1}\right) \cdots f\left(\boldsymbol{h}_2|\boldsymbol{h}_1\right) f\left(\boldsymbol{h}_1\right) d\boldsymbol{h}_1 \cdots d\boldsymbol{h}_n \\
= & \int_{\mathcal{C}^M} f\left(y_k|\boldsymbol{x}_k, \boldsymbol{h}_k\right) f\left(\boldsymbol{h}_k\right) d\boldsymbol{h}_k \\
= & f\left(y_k|\boldsymbol{x}_k\right) \\
= & \frac{1}{\pi\left(\sigma_z^2 + \boldsymbol{x}_k^T \boldsymbol{D}_k \boldsymbol{x}_k^*\right)} \exp\left(-\frac{\left|y_k - \left(\boldsymbol{x}_k^T \boldsymbol{d}_k\right)\right|^2}{\sigma_z^2 + \boldsymbol{x}_k^T \boldsymbol{D}_k \boldsymbol{x}_k^*}\right),
\end{aligned}
$$

where $\boldsymbol{D}_k \triangleq \left(\frac{1-|\alpha|^{2k}}{1-|\alpha|^2}\right) \boldsymbol{C}$ and $\boldsymbol{d}_k \triangleq \alpha^k \boldsymbol{h}_0 + \frac{1-\alpha^k}{1-\alpha} \boldsymbol{d}$.

- Accordingly, the probability distribution of $f(y_k)$ is

$$
f\left(y_k\right) = \sum_{\boldsymbol{x}_k \in \mathcal{X}^M} P_{\boldsymbol{X}_k}(\boldsymbol{x}_k) \frac{1}{\pi\left(\sigma_z^2 + \boldsymbol{x}_k^T \boldsymbol{D}_k \boldsymbol{x}_k^*\right)} \exp\left(-\frac{\left|y_k - \left(\boldsymbol{x}_k^T \boldsymbol{d}_k\right)\right|^2}{\sigma_z^2 + \boldsymbol{x}_k^T \boldsymbol{D}_k \boldsymbol{x}_k^*}\right).
$$

Independent bound for $M = 1$

- For $M = 1$,

$$
\begin{aligned}
f\left(y_k|\boldsymbol{x}_k\right) &= \frac{1}{\pi\left(\sigma_z^2 + |x_k|^2 D_k\right)} \exp\left(-\frac{\left|y_k - x_k d_k\right|^2}{\sigma_z^2 + |x_k|^2 D_k}\right) \\
&= \frac{1}{\pi\left(\sigma_z^2 + |x_k|^2 D_k\right)} \exp\left(-\frac{|d_k|^2\left|y_k/d_k - x_k\right|^2}{\sigma_z^2 + |x_k|^2 D_k}\right).
\end{aligned}
$$

- Since $I(X_k; Y_k) = I(X_k; \tilde{Y}_k)$ for invertible transformation $\tilde{Y}_k \triangleq Y_k/d_k$, we can transform $f(y_k|x_k)$ to obtain $f(\tilde{y}_k|x_k)$ as:

$$
f\left(\tilde{y}_k|x_k\right) = \frac{1}{\pi\left(\sigma_z^2 + |x_k|^2 D_k\right)/|d_k|^2} \exp\left(-\frac{\left|\tilde{y}_k - x_k\right|^2}{(\sigma_z^2 + |x_k|^2 D_k)/|d_k|^2}\right).
$$

- In our system, the transmitted symbol x_k is assumed antipodally modulated. In other words, $x_k \in \{-s, +s\}$ for some real s. We can reduce the complex channel to the real channel as:

$$f\left(\tilde{y}_{k,r}|x_k\right) = \frac{1}{\sqrt{\pi(\sigma_z^2 + s^2 D_k)/|d_k|^2}} \exp\left(-\frac{(\tilde{y}_{k,r} - x_k)^2}{(\sigma_z^2 + s^2 D_k)/|d_k|^2}\right),$$

where $\tilde{y}_{k,r}$ is the real part of $\tilde{y}_k$.

- For this real-valued symmetric additive Gaussian channel, its capacity-cost function is achieved by uniform input with $s^2 = S$:

$$\begin{aligned} C_k(S) &= I(X_k; \tilde{Y}_{k,r}) \\ &= h(\tilde{Y}_{k,r}) - \sum_{x_k \in \mathcal{X}} \frac{1}{2} \int_{\Re} f\left(\tilde{y}_{k,r}|x_k\right) \log\left[\frac{1}{f\left(\tilde{y}_{k,r}|x_k\right)}\right] d\tilde{y}_{k,r} \\ &= h(\tilde{Y}_{k,r}) - \frac{1}{2}\log\left[2\pi e \sigma_N^2\right]. \end{aligned}$$

where $\sigma_N^2 \triangleq (\sigma_z^2 + s^2 D_k)/(2|d_k|^2)$.

- Now, for uniform channel input,

$$
\begin{aligned}
& h(\tilde{Y}_{k,r}) \\
= & -\int_{\Re} f(\tilde{y}_{k,r}) \log(f(\tilde{y}_{k,r})) d\tilde{y}_{k,r} \\
= & \int_{\Re} f(\tilde{y}_{k,r}) \left[\frac{1}{2} + \frac{\tilde{y}_{k,r}^2 + s^2}{2\sigma_N^2} - \log\left(\frac{e^{\tilde{y}_{k,r} s/\sigma_N^2} + e^{-\tilde{y}_{k,r} s/\sigma_N^2}}{2} \right) \right] d\tilde{y}_{k,r} \\
= & \frac{\log(2\pi e \sigma_N^2)}{2} + \frac{s^2}{\sigma_N^2} - \int_{\Re} \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-(\tilde{y}_{k,r}-s)^2/2\sigma_N^2} \cdot \log\left(\cosh\left(\frac{\tilde{y}_{k,r} s}{\sigma_N^2}\right) \right) d\tilde{y}_{k,r} \\
= & \frac{\log(2\pi e \sigma_N^2)}{2} + \frac{S}{\sigma_N^2} - \int_{\Re} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \log\left(\cosh\left(\frac{S}{\sigma_N^2} + t\sqrt{\frac{S}{\sigma_N^2}} \right) \right) dt,
\end{aligned}
$$

- It immediately gives that:

$$
C_k(S) = \frac{2|d_k|^2 S}{\sigma_z^2 + SD_k} - \int_{\Re} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \log\left(\cosh\left(\frac{2|d_k|^2 S}{\sigma_z^2 + SD_k} + t\sqrt{\frac{2|d_k|^2 S}{\sigma_z^2 + SD_k}} \right) \right) dt.
$$

- By Cesáro-mean theorem , when taking n to infinity, we have:

$$C(S) \leq \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} C_k(S) = \lim_{n\to\infty} C_n(S) = C_\infty(S),$$

where

$$C_\infty(S) = \frac{2|d_\infty|^2 S}{\sigma_z^2 + SD_\infty} - \int_{\Re} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \log\left(\cosh\left(\frac{2|d_\infty|^2 S}{\sigma_z^2 + SD_\infty} + t\sqrt{\frac{2|d_\infty|^2 S}{\sigma_z^2 + SD_\infty}}\right)\right) dt,$$

$d_\infty = d/(1-\alpha)$ and $D_\infty = C/(1-|\alpha|^2)$, provided that $|\alpha| < 1$.

- We will demonstrate in the next three figures how channel model parameters affect the independent bound of channel capacity.

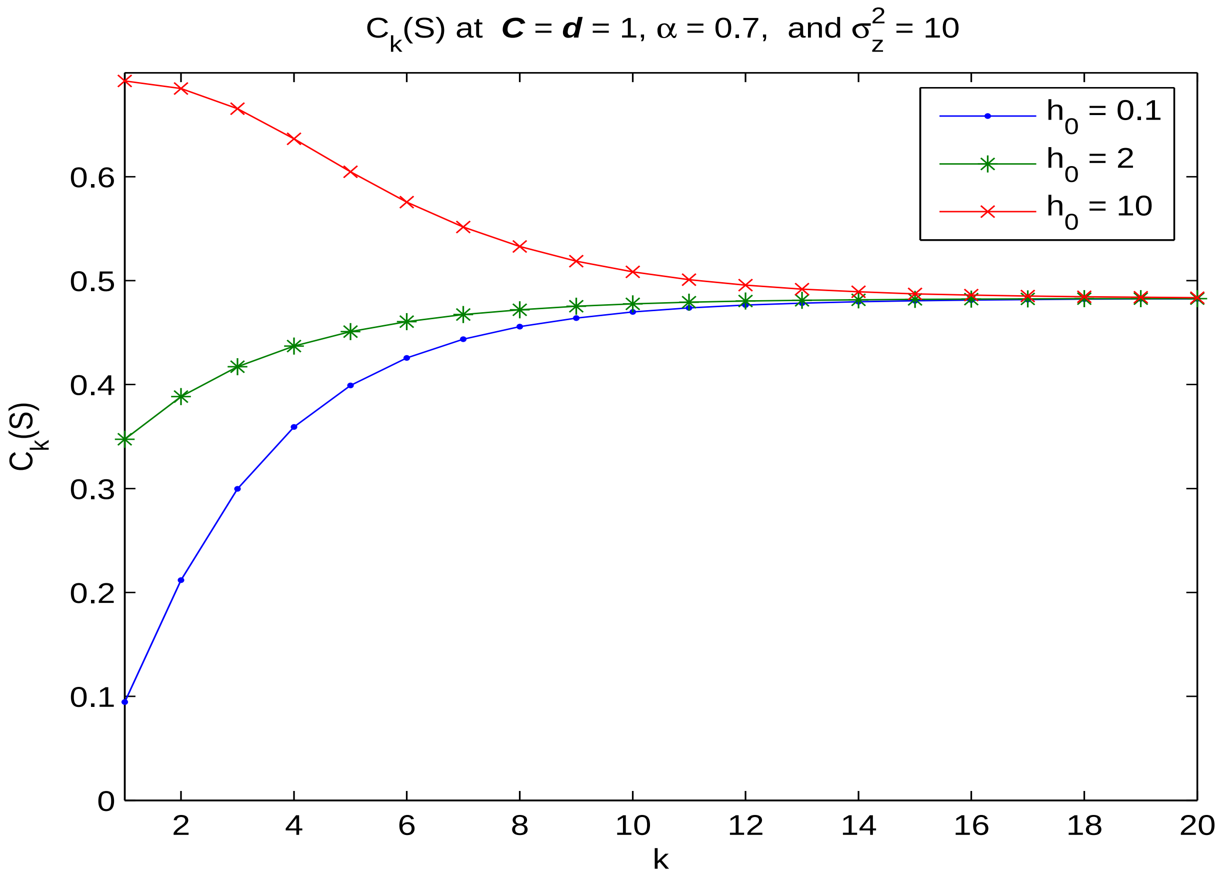

Figure 1: Illustration of $C_k(S)$ for different initial fading coefficients $\boldsymbol{h}_0 = 0.1, 2$ and 10. Other parameters for Gauss-Markov channels are $S = 1$, $\boldsymbol{C} = 1$, $\boldsymbol{d} = 1$, $\alpha = 0.7$ and $\sigma_z^2 = 10$. Obviously, $\boldsymbol{h}_0$ affects $C_k(S)$ only at small k. As k grows, $\boldsymbol{d}_k$ will be dominated by α and $\boldsymbol{d}_0$, and $C_k(S)$ will converge to the same limit $C_\infty(S)$.

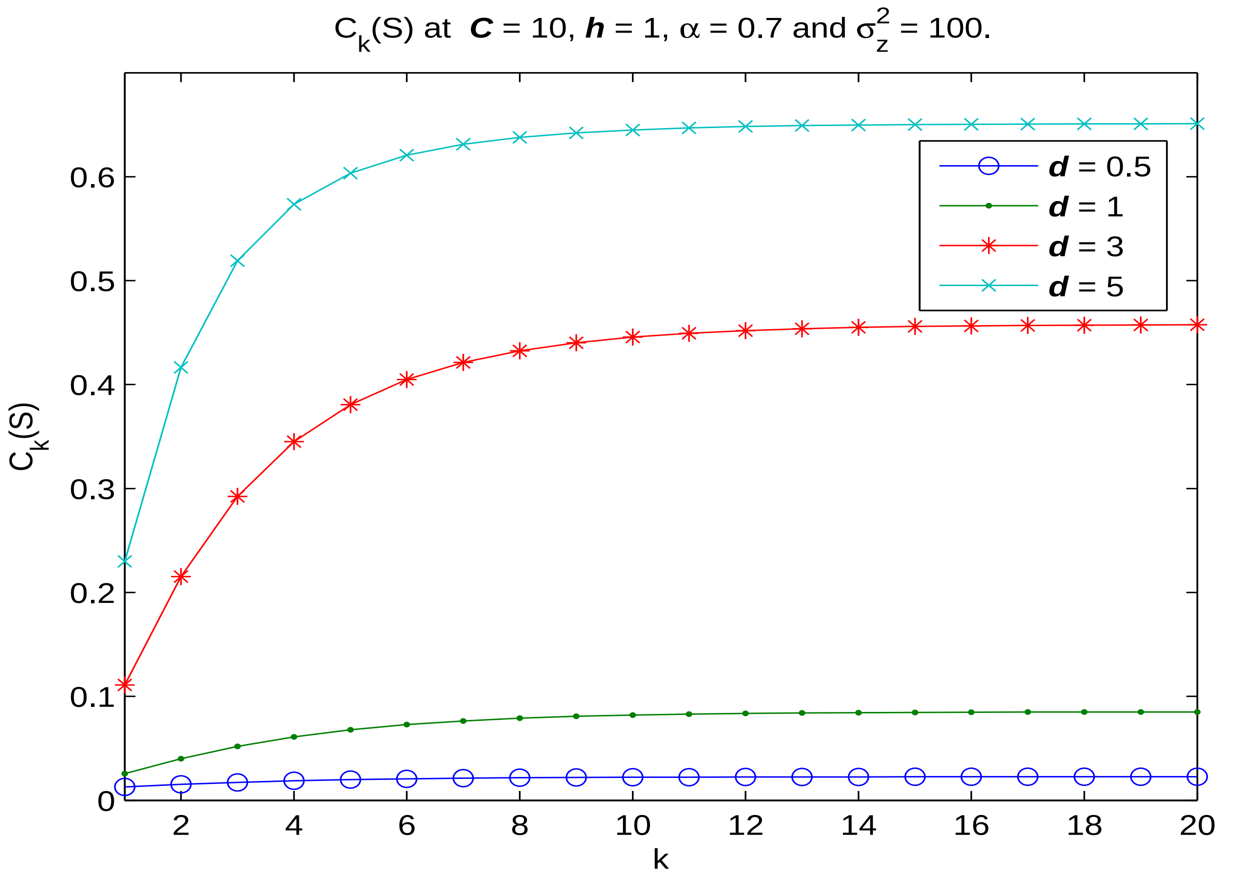

Figure 2: Illustration of $C_k(S)$ for different initial Gauss-Markov fading mean $\boldsymbol{d} = 0.5, 1, 3$ and 5. Other parameters for Gauss-Markov channels are $S = 1$, $\boldsymbol{C} = 10$, $\boldsymbol{h}_0 = 1$, $\alpha = 0.7$ and $\sigma_z^2 = 100$. The value of $\boldsymbol{d}$ determines the strength of line-of-sight (LOS) propagated signal, and a stronger $\boldsymbol{d}$ can result in larger capacity bound $C_\infty(S)$.

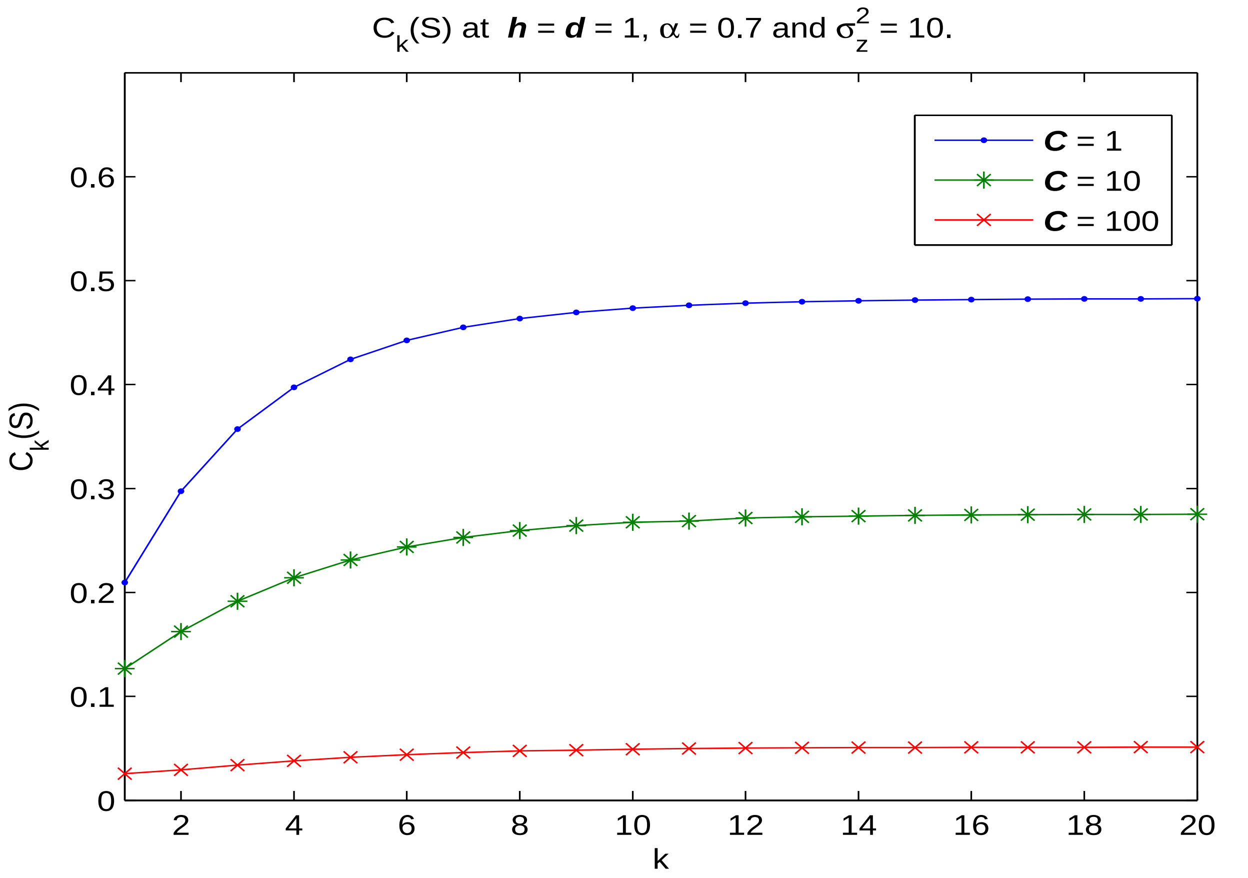

Figure 3: Illustration of $C_k(S)$ for different initial fading covariance matrix $\boldsymbol{C} = 1, 10$ and 100. Other parameters for Gauss-Markov channels are $S = 1$, $\boldsymbol{h}_0 = 1$, $\boldsymbol{d} = 1$, $\alpha = 0.7$ and $\sigma_z^2 = 10$. The figure indicates that a larger fading variance $\boldsymbol{C}$ makes a lower capacity bound.

- An interesting observation that $C_\infty(S)$ is equal to zero once $\boldsymbol{d} = \mathbf{0}$. This means that the channel capacity $C(S) = 0$ if there exists no LOS signals in the communications via Gauss-Markov channels.
- The Gauss-Markov channel model can be reduced to the additive white Gaussian noise (AWGN) channel model by letting $\alpha = 0$, $\boldsymbol{d} = \mathbf{1}$ and $\boldsymbol{C} = \mathbf{0}$. As a result,

$$C_\infty^{\text{AWGN}}(S) = C_k^{\text{AWGN}}(S) = \frac{S}{\sigma_z^2/2} - \int_{\Re} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \log\left(\cosh\left(\frac{S}{\sigma_z^2/2} + t\sqrt{\frac{S}{\sigma_z^2/2}}\right)\right) dt.$$

Notably, this is no longer an upper bound, but the **exact channel capacity formula** for binary-input AWGN channels.

Independent bound for $M = 2$

- At $M = 2$,

$$\begin{aligned} C_k(S) &= \max_{\{P_{\mathbf{X}} \,:\, (\forall\ i) E[X_i^2] \le S\}} I(X_k; Y_k) \\ &= \max_{\{P_{X_{k-1}^k} \,:\, E[X_k]^2 \le S \text{ and } E[X_{k-1}^2] \le S\}} I(X_k; Y_k). \end{aligned}$$

It is in general hard to find $C_k(S)$ for the case of $M > 1$. Hence, we made an assumption on the channel statistics below.

- **Assumption 1** *$d_{k,1} = \rho_1 d_k$ and $d_{k,2} = \rho_2 d_k$ for some real numbers ρ_1 and ρ_2, where $\boldsymbol{d}_k = \begin{bmatrix} d_{k,1} & d_{k,2} \end{bmatrix}$. Also, $\boldsymbol{C}$ is diagonal; hence, there exists $D_{k,1}$ and $D_{k,2}$ such that*

$$\boldsymbol{D}_k = \begin{bmatrix} D_{k,1} & 0 \\ 0 & D_{k,2} \end{bmatrix} = \frac{1 - |\alpha|^{2k}}{1 - |\alpha|^2} \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix}.$$

- Recall that

$$
\begin{aligned}
& f\left(y_k|\boldsymbol{x}_k\right) \\
\triangleq \quad & \frac{1}{\pi\left(\sigma_z^2+\boldsymbol{x}_k^T\boldsymbol{D}_k\boldsymbol{x}_k^*\right)}\exp\left(-\frac{\left|y_k-\left(\boldsymbol{x}_k^T\boldsymbol{d}_k\right)\right|^2}{\sigma_z^2+\boldsymbol{x}_k^T\boldsymbol{D}_k\boldsymbol{x}_k^*}\right) \\
= \quad & \frac{1}{\pi\left(\sigma_z^2+|x_k|^2D_{k,1}+|x_{k-1}|^2D_{k,2}\right)}\exp\left(-\frac{|y_k-x_kd_{k,1}-x_{k-1}d_{k,2}|^2}{\sigma_z^2+|x_k|^2D_{k,1}+|x_{k-1}|^2D_{k,2}}\right) \\
= \quad & \frac{1}{\pi\left(\sigma_z^2+s^2D_{k,1}+s^2D_{k,2}\right)}\exp\left(-\frac{|y_k/d_k-(\rho_1x_k+\rho_2x_{k-1})|^2}{(\sigma_z^2+s^2D_{k,1}+s^2D_{k,2})/|d_k|^2}\right).
\end{aligned}
$$

- Then, by letting $\tilde{y}_k=y_k/d_k$, we obtain:

$$
f\left(\tilde{y}_k|x_k,x_{k-1}\right) = \frac{1}{\pi\sigma^2}\exp\left(-\frac{|\tilde{y}_k-(\rho_1x_k+\rho_2x_{k-1})|^2}{\sigma^2}\right),
$$

where $\sigma^2\triangleq(\sigma_z^2+s^2D_{k,1}+s^2D_{k,2})/|d_k|^2$.

- By following the same reasoning as in the case of $M = 1$,

$$I(X_k; Y_k) = I(X_k; \tilde{Y}_k) = I(X_k; \tilde{Y}_{k,r}),$$

where $\tilde{Y}_k = \tilde{Y}_{k,r} + j\tilde{Y}_{k,i}$.

- Then, we derive:

$$I(X_k; \tilde{Y}_{k,r}) = \sum_{x_{k-1} \in \mathcal{X}} \sum_{x_k \in \mathcal{X}} \int_{\Re} P_{X_k, X_{k-1}}(x_k, x_{k-1}) f(\tilde{y}_{k,r} | x_k, x_{k-1}) \cdot$$

$$\log \frac{\sum_{\bar{x}_{k-1} \in \mathcal{X}} P_{X_k, X_{k-1}}(x_k, \bar{x}_{k-1}) f(\tilde{y}_{k,r} | x_k, \bar{x}_{k-1})}{\left(\sum_{\hat{x}_{k-1} \in \mathcal{X}} P_{X_k, X_{k-1}}(x_k, \hat{x}_{k-1}) \right) \left(\sum_{x'_{k-1} \in \mathcal{X}} \sum_{x'_k \in \mathcal{X}} P_{X_k, X_{k-1}}(x'_k, x'_{k-1}) f(\tilde{y}_{k,r} | x'_k, x'_{k-1}) \right)} d\tilde{y}_{k,r};$$

- By taking the derivative of

$$I(X_k;\tilde{Y}_{k,r})+\lambda\left(\sum_{x_{k-1}\in\mathcal{X}}\sum_{x_k\in\mathcal{X}}P_{X_k,X_{k-1}}(x_k,x_{k-1})-1\right)$$

with respect to $P_{X_k,X_{k-1}}(x''_k,x''_{k-1})$, we obtain:

$$\begin{aligned}&\frac{\partial\left[I(X_k;\tilde{Y}_{k,r})+\lambda\left(\sum_{x_{k-1}\in\mathcal{X}}\sum_{x_k\in\mathcal{X}}P_{X_k,X_{k-1}}(x_k,x_{k-1})-1\right)\right]}{\partial P_{X_k,X_{k-1}}(x''_k,x''_{k-1})}\\ =\;& I(x''_k,x''_{k-1};\tilde{Y}_{k,r})-1+\lambda,\end{aligned}$$

where

$$I(x''_k,x''_{k-1};\tilde{Y}_{k,r})\triangleq\int_{\Re}f(\tilde{y}_{k,r}|x''_k,x''_{k-1})\times\log\frac{\sum_{\bar{x}_{k-1}\in\mathcal{X}}P_{X_k,X_{k-1}}(x''_k,\bar{x}_{k-1})f(\tilde{y}_{k,r}|x''_k,\bar{x}_{k-1})}{\left(\sum_{\hat{x}_{k-1}\in\mathcal{X}}P_{X_k,X_{k-1}}(x''_k,\hat{x}_{k-1})\right)\left(\sum_{x'_{k-1}\in\mathcal{X}}\sum_{x'_k\in\mathcal{X}}P_{X_k,X_{k-1}}(x'_k,x'_{k-1})f(\tilde{y}_{k,r}|x'_k,x'_{k-1})\right)}d\tilde{y}_{k,r}$$

- Hence, the maximum capacity occurs at:

$$I(x_k, x_{k-1}; \tilde{Y}_{k,r}) \begin{cases} = C_k(S), & \text{if } P_{X_k,X_{k-1}}(x_k, x_{k-1}) > 0 \\ \leq C_k(S), & \text{if } P_{X_k,X_{k-1}}(x_k, x_{k-1}) = 0. \end{cases}$$

- Numerically evaluation shows that for positive ρ_1 and ρ_2, the largest $C_k(S)$ occurs at $p_{1,1} = p_{-1,-1} = 1/2$ and $p_{1,-1} = p_{-1,1} = 0$, in which case

$$C_k(S) \begin{cases} = I_k(1,1) = I_k(-1,-1) \\ \geq I_k(1,-1) \\ \geq I_k(-1,1), \end{cases}$$

where $p_{a,b}$ denotes $P_{X_k,X_{k-1}}(x_k = as, x_{k-1} = bs)$.

- An interpretation of the numerical result is that all of the four possible inputs for (x_k, x_{k-1}) share the same power, and $f(y_k|x_k, x_{k-1})$ for different (x_k, x_{k-1}) has common variance but is with aligned means (cf. Fig. 4); hence, it is advantageous to use the two inputs that are farthest to each other.

- For general ρ_1 and ρ_2, the "two inputs" become $(s \cdot \text{sgn}(\rho_1), s \cdot \text{sgn}(\rho_2))$ and $(-s \cdot \text{sgn}(\rho_1), -s \cdot \text{sgn}(\rho_2))$, where $\text{sgn}(\cdot)$ represents the sign function.

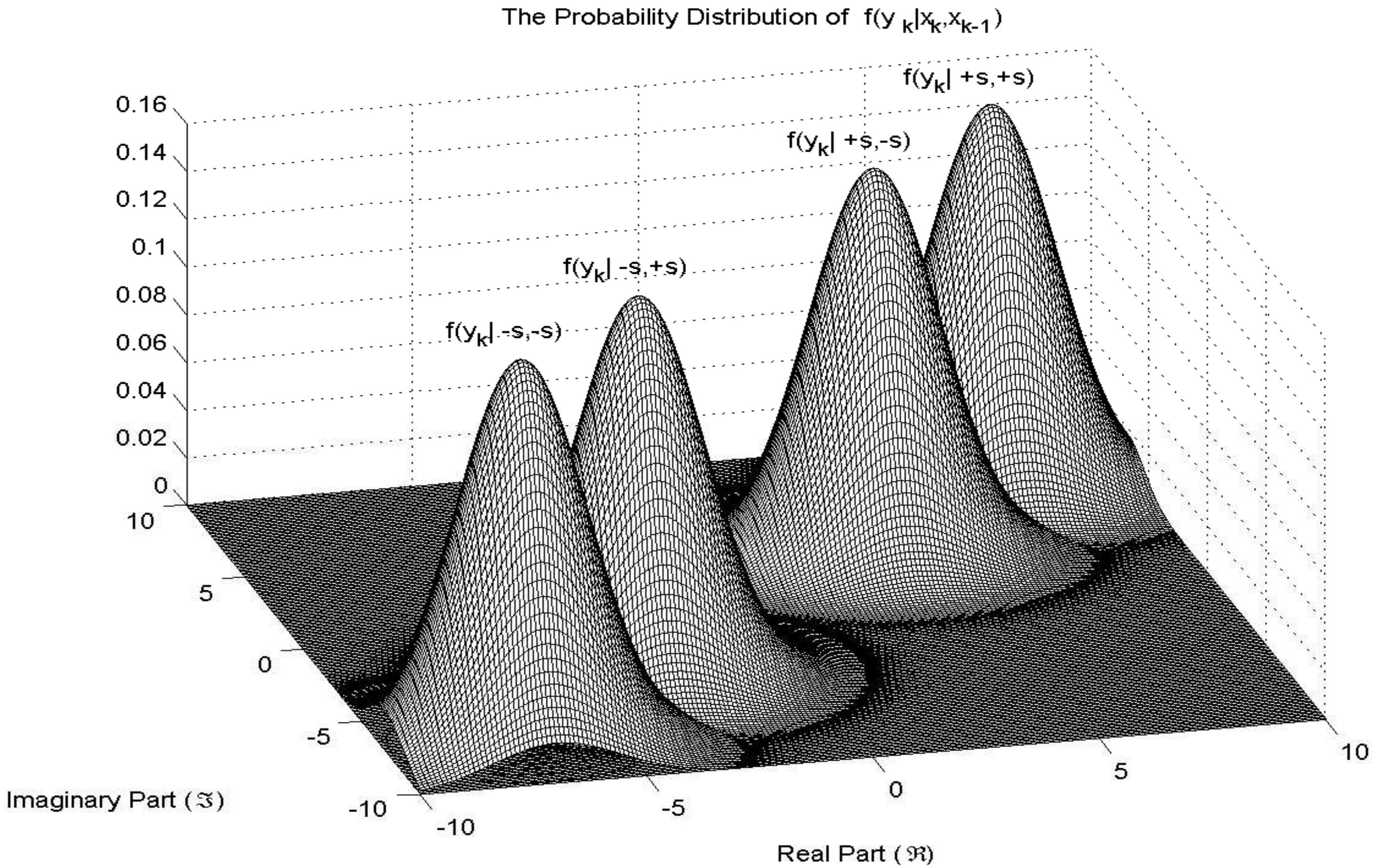

Figure 4: $f(y_k|x_k, x_{k-1})$ for four different (x_k, x_{k-1}). The parameters used in this figure are $s = 1$, $d_{k,1} = 5.25 + j5.25$, $d_{k,2} = 1.75 + j1.75$ and $D_{k,1} + D_{k,2} + \sigma_z^2 = 2$.

- In summary, for $d_{k,1} = \rho_1 d_k$ and $d_{k,2} = \rho_2 d_k$, the equivalent real channel of the original complex channel is $f(\tilde{y}_{k,r}|x_k, x_{k-1})$, which is Gaussian distributed with mean $\sum_{i=1}^{M} \rho_i x_{k-i+1}$ and variance $(1/2)\sigma^2 = (\sigma_z^2 + S\sum_{i=1}^{M} D_{k,i})/(2|d_k|^2)$. The equal probability on $(x_k, x_{k-1}) = (s \cdot \mathrm{sgn}(\rho_1), s \cdot \mathrm{sgn}(\rho_2))$ and $(x_k, x_{k-1}) = (-s \cdot \mathrm{sgn}(\rho_1), -s \cdot \mathrm{sgn}(\rho_2))$ maximizes $I(X_k, \tilde{Y}_{k,r})$, and

$$C_k(S) = \frac{2S}{\sigma^2/\rho^2} - \int_{\Re} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \left[\log \left(\cosh \left(\frac{\sqrt{2S}}{\sigma/\rho} t + \frac{2S}{\sigma^2/\rho^2} \right) \right) \right] dt,$$

where $\rho = \sum_{i=1}^{M} |\rho_i|$, and

$$\frac{\sigma^2}{\rho^2} = \frac{\sigma_z^2 + S\sum_{i=1}^{M} D_{k,i}}{\left(\sum_{i=1}^{M} |d_{k,i}| \right)^2}.$$

- Finally, by replacing

$$\frac{\sigma^2}{\rho^2} = \frac{\sigma_z^2 + \frac{S}{1-|\alpha|^2}\sum_{i=1}^{M} C_{i,i}}{\left(\frac{1}{|1-\alpha|}\sum_{i=1}^{M} |d_i|\right)^2},$$

where $\boldsymbol{C}$ is diagonal with diagonal elements $\{C_{i,i}\}_{i=1}^{M}$ and $\boldsymbol{d} = \begin{bmatrix} d_1 & d_2 & \cdots & d_M \end{bmatrix}^T$, we yield $C_\infty(S)$ for the case of $M = 2$.

Independent bound for general M

- **Theorem 1** *Let M be the true memory order of the channel. Assume that there exists a complex number d_k such that $d_{k,i} = \rho_i d_k$ for some real number ρ_i for every $1 \leq i \leq M$, where $\boldsymbol{d}_k = \begin{bmatrix} d_{k,1} & d_{k,2} & \cdots & d_{k,M} \end{bmatrix}$. Also, $\boldsymbol{C}$ is diagonal. Then, the component independent bound $C_k(S)$ is given by:*

$$C_k(S) = \frac{2S}{\delta^2} - \int_{\Re} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \left[\log \left(\cosh \left(\frac{\sqrt{2S}}{|\delta|} t + \frac{2S}{\delta^2} \right) \right) \right] dt,$$

where

$$\delta^2 = \frac{\sigma_z^2 + S \sum_{i=1}^{M} D_{k,i}}{\left(\sum_{i=1}^{M} |d_{k,i}| \right)^2}.$$

- Furthermore, the ultimate independent bound $C_\infty(S)$ has the same form as (1) with

$$\delta^2 = \frac{\sigma_z^2 + \frac{S}{1-|\alpha|^2}\sum_{i=1}^{M} C_{i,i}}{\left(\frac{1}{|1-\alpha|}\sum_{i=1}^{M}|d_i|\right)^2}.$$

- Figures 5 and 6 show the independent bounds $C_\infty(S)$ for Gauss-Markov channels of different memory orders.

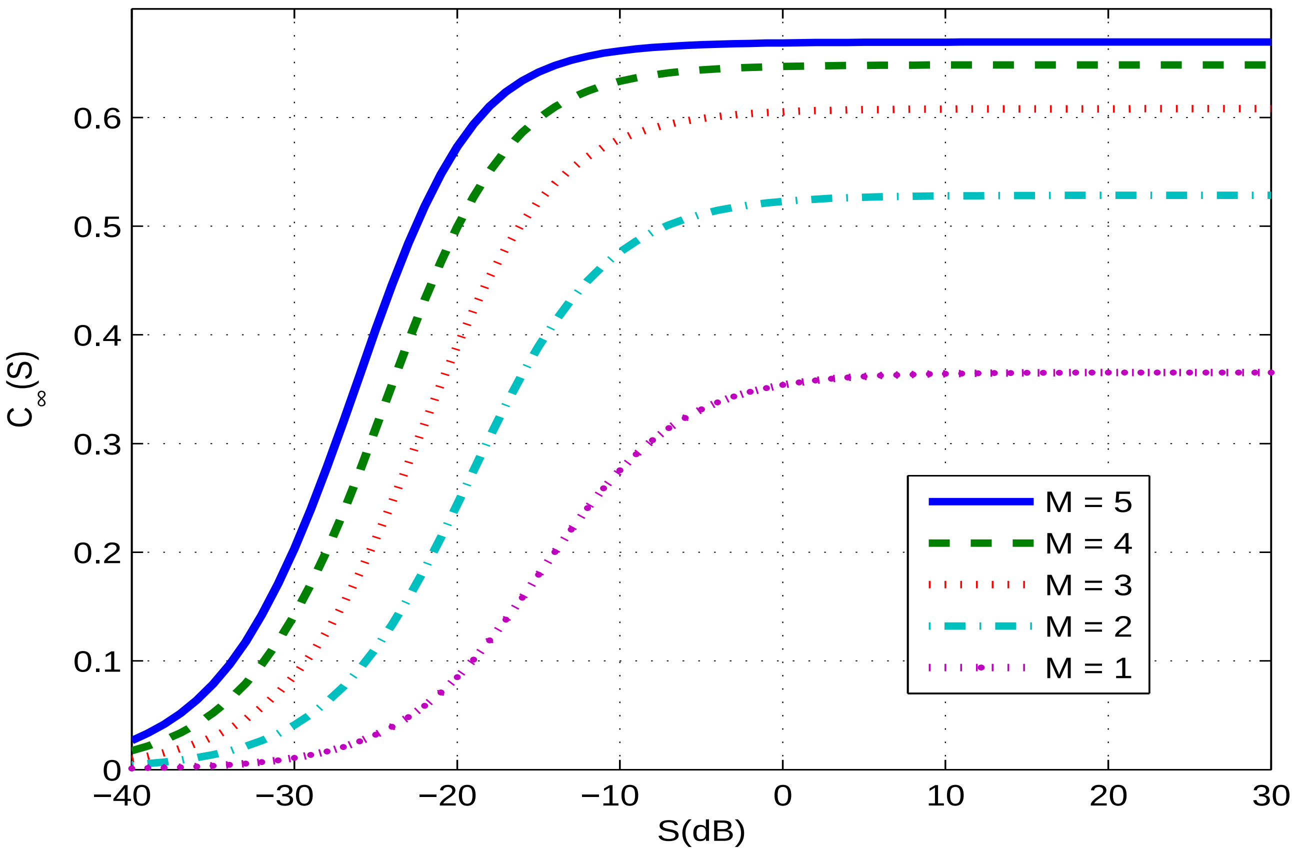

Figure 5: Illustration of $C_\infty(S)$. Other parameters for Gauss-Markov channels are $C_{1,1} = C_{2,2} = C_{3,3} = C_{4,4} = C_{5,5} = 10$, $d_1 = d_2 = d_3 = d_4 = d_5 = 1$, $\alpha = 0.7$ and $\sigma_z^2 = 1$.

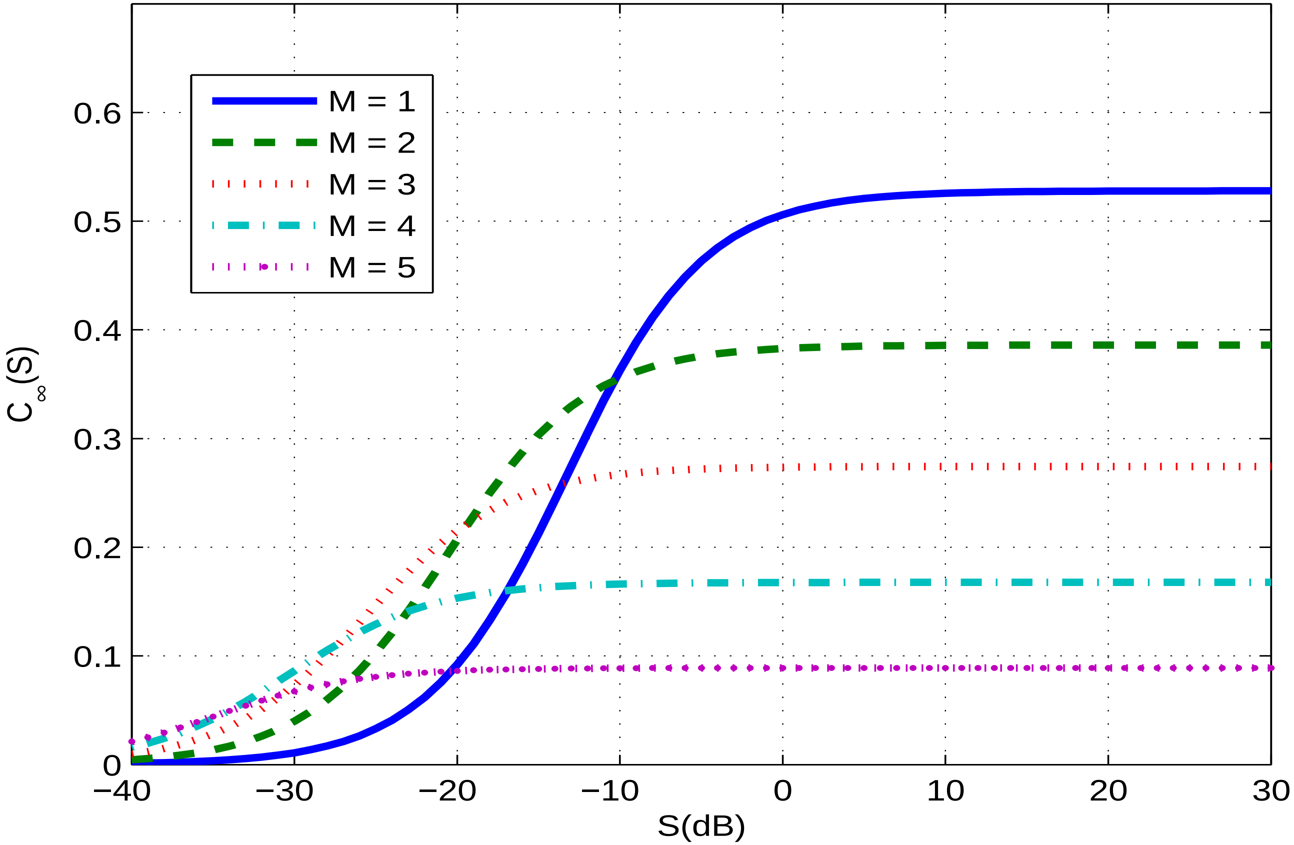

Figure 6: Illustration of $C_\infty(S)$. Other parameters for Gauss-Markov channels are $C_{1,1} = 10^{0.7}$, $C_{2,2} = 10^{1.5}$, $C_{3,3} = 10^2$, $C_{4,4} = 10^{2.5}$, $C_{5,5} = 10^3$, $d_1 = d_2 = d_3 = d_4 = d_5 = 1$, $\alpha = 0.7$ and $\sigma_z^2 = 1$.

- Figure 5: For fixed $C_{i,i}$ (and d_i), the higher the channel memory order, the more involved in received vector $\boldsymbol{y}$ at the receiver end. Thus, it is reasonable to expect a lower capacity for larger M; based on this "intuitive conjecture", the independent bound could be looser for higher M.

- Figure 6: Since $C_{i,i}$ gets larger as i increases, the influence to the current output y_k by the distant input x_{k-M+1} grows . Following the intuition, the independent bounds decrease as M increases for SNR beyond -10 dB. Nevertheless, when SNR is below -10 dB, the independent bounds become messy in channel memory order M.

The lower bound of bit error probability

- The bit error rate (P_b) (for information bits) is used as the typical performance measure in practical communication system. By means of the rate-distortion function, we can obtain a lower bound of this typical performance measure.

- First, we need to derive the average E_b/N_0 for the Gauss-Markov fading channel at channel code rate R. The average SNR is equal to:

$$\begin{aligned}\overline{SNR} &= \lim_{n\to\infty} \frac{\sum_{i=1}^{n} E[|\boldsymbol{x}_k^T \boldsymbol{h}_k|^2]}{\sum_{i=1}^{n} E[z_k^2]} \\ &= \frac{S}{\sigma_z^2}\left(\frac{1}{|1-\alpha|^2}\boldsymbol{d}^H\boldsymbol{d} + \frac{1}{1-|\alpha|^2}\mathrm{tr}(\boldsymbol{C})\right).\end{aligned}$$

Hence,

$$\gamma_b \triangleq \frac{E_b}{N_0} = \frac{\overline{SNR}}{R} = \frac{1}{R}\frac{S}{\sigma_z^2}\left(\frac{1}{|1-\alpha|^2}\boldsymbol{d}^H\boldsymbol{d} + \frac{1}{1-|\alpha|^2}\mathrm{tr}(\boldsymbol{C})\right).$$

- The rate distortion for binary input and Hamming additive distortion measure is

$$R(D) = \begin{cases} \log(2) - H_b(D), & \text{for } 0 \leq D \leq 0.5 \\ 0, & \text{for } D > 0.5 \end{cases}$$

- According to the joint source-channel coding theorem, good codes exists when

$$R(P_b) < \frac{C(S)}{R}.$$

- Therefore, we obtain a lower bound for P_b as:

$$H_b(P_b) > \log(2) - \frac{1}{R} C_\infty(S).$$

The lower bounds of bit error rate P_b corresponding to those in Figs. 5 and 6 are summarized in Figs. 7 and 8.

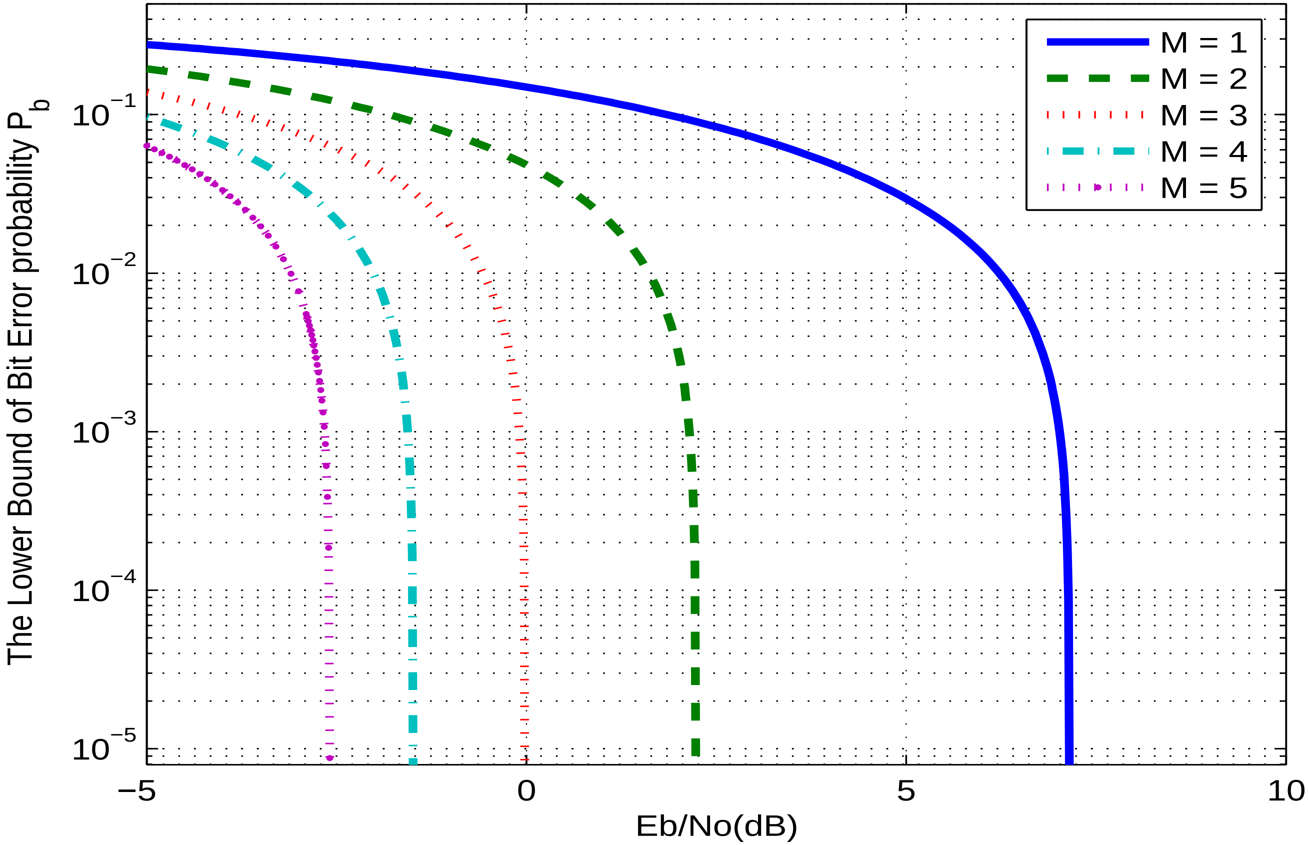

Figure 7: Illustration of lower bounds for P_b. The code rate adopted is $R = 1/3$. Other parameters used in this figure are the same as those in Fig. 5.

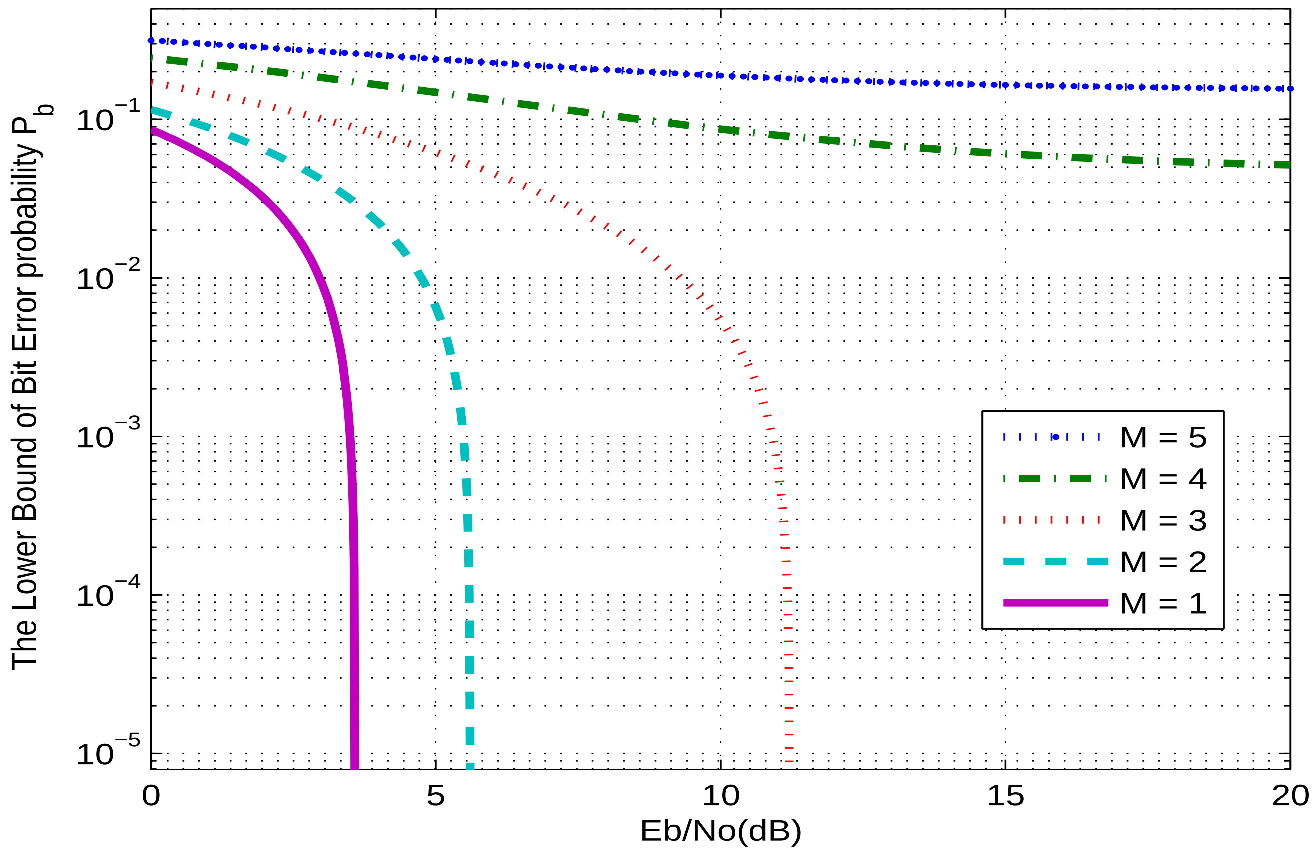

Figure 8: Illustration of lower bounds for P_b. The code rate adopted is $R = 1/3$. Other parameters used in this figure are the same as those in Fig. 6.

Conclusion

- In this thesis, we have remarked on four different definitions of channel capacities according to the transmiter/receiver with/without channel state information.
- We then turn to the derivation of the independent bounds for the channel capacity without CSI in both transmitter and receiver.
- We then found that if there is no LOS signal existing, the capacity of the blind-CSI system is reduced to zero.