Capacity of Multi-antenna Gaussian Channels (I. E. Telatar, 1999)

Mattia Lecci[1][2]

[1] Università degli Studi di Padova
Department of Information Engineering
(DEI-UniPD)

[2] Universitat Politècnica de Catalunya
Escola Tècnica Superior d'Engenyeria de Telecomunicació de Barcelona
(UPC-ETSETB)

January 2018

Overview

1. Introduction
   - Abstract
   - Notation and Assumptions
   - Properties and Lemmas
2. Gaussian Channel with Deterministic H
   - Capacity Derivation
   - Examples
   - Formal Derivation of Capacity
3. Gaussian Channel with Ergodic $\mathbf{H}$
   - Introduction to the problem
   - Capacity-achieving Distribution
   - Evaluation of the Capacity
   - Examples

Table of Contents

1. Introduction
   - Abstract
   - Notation and Assumptions
   - Properties and Lemmas
2. Gaussian Channel with Deterministic H
   - Capacity Derivation
   - Examples
   - Formal Derivation of Capacity
3. Gaussian Channel with Ergodic $\mathbf{H}$
   - Introduction to the problem
   - Capacity-achieving Distribution
   - Evaluation of the Capacity
   - Examples

Abstract

In this presentation I will talk about the capacity of a single-user Gaussian channel with multiple receiving and/or transmitting antennas [1] (also known as MIMO channel).

I will talk about 3 cases:

- Deterministic channel
- Random i.i.d. ergodic channel
- *Bonus*: multi-user case

Notation I

The notation adopted is halfway between the one used during the course and the one from the original paper.

Let's denote with t the number of transmitting and with r the number of receiving antennas.

We will consider the classical linear model, where $\mathbf{x} \in \mathbb{C}^t$ is the transmitted vector and $\mathbf{y} \in \mathbb{C}^r$ is the received vector, $H \in \mathbb{C}^{r \times t}$ is the complex channel matrix and $\mathbf{n} \in \mathbb{C}^r$ is the noise

$$\mathbf{y} = H\mathbf{x} + \mathbf{n}$$

Notation II

We assume noise at different receivers to be independent and normalized, i.e. $E\left[\mathbf{x}\mathbf{x}^{\dagger}\right] = I_r$.
The dag ($\dagger$) notation is used for the conjugate-transpose operation.

We have the power constraint

$$E\left[\mathbf{x}^{\dagger}\mathbf{x}\right] = E\left[\mathrm{tr}\left(\mathbf{x}\mathbf{x}^{\dagger}\right)\right] = \mathrm{tr}\left(E\left[\mathbf{x}\mathbf{x}^{\dagger}\right]\right) \leq P$$

A complex random vector (r.ve.) $\mathbf{x} \in \mathbb{C}^n$ is said to be Complex Gaussian if its real extension $\hat{\mathbf{x}} \triangleq \left[\begin{array}{c} \mathrm{Re}\{\mathbf{x}\} \\ \mathrm{Im}\{\mathbf{x}\} \end{array}\right] \in \mathbb{R}^{2n}$ is Gaussian.

The r.ve. $\mathbf{x}$ will have *mean* and *covariance* respectively $\mu = E[\mathbf{x}]$ and $Q = \mathrm{Cov}(Q) = E\left[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^{\dagger}\right]$

Notation III

Defining for a complex matrix A

$$\hat{A} = \begin{bmatrix} \mathrm{Re}(A) & -\mathrm{Im}(A) \\ \mathrm{Im}(A) & \mathrm{Re}(A) \end{bmatrix}$$

we say that a Complex Gaussian r.ve. is *circularly symmetric* if $\mathrm{Cov}(\hat{\mathbf{x}}) = E\left[(\hat{\mathbf{x}} - \hat{\mu})(\hat{\mathbf{x}} - \hat{\mu})^T\right] = \frac{1}{2}\hat{Q}$

pdf of circularly symmetric Complex Gaussian

The pdf of a circularly symmetric Complex Gaussian is

$$\begin{aligned}\gamma_{\mu,Q} &= \det\left(\pi\hat{Q}\right)^{-\frac{1}{2}} e^{-(\hat{x}-\hat{\mu})^T\hat{Q}(\hat{x}-\hat{\mu})} \\ &= \det(\pi Q)^{-1} e^{-(x-\mu)^\dagger Q(x-\mu)}\end{aligned}$$

Lemma 1

The following properties hold:

$$C = AB \iff \hat{C} = \hat{A}\hat{B} \tag{1a}$$

$$C = A + B \iff \hat{C} = \hat{A} + \hat{B} \tag{1b}$$

$$C = A^\dagger \iff \hat{C} = \hat{A}^T \tag{1c}$$

$$C = A^{-1} \iff \hat{C} = \hat{A}^{-1} \tag{1d}$$

$$\det\left(\hat{A}\right) = |\det(A)|^2 = \det\left(AA^\dagger\right) \tag{1e}$$

$$z = x + y \iff \hat{z} = \hat{x} + \hat{y} \tag{1f}$$

$$y = Ax \iff \hat{y} = \hat{A}\hat{x} \tag{1g}$$

$$\mathrm{Re}\left\{x^\dagger y\right\} = \hat{a}^T \hat{y} \tag{1h}$$

Properties and Lemmas II

Corollary 1

A matrix $U \in \mathbb{C}^{n\times n}$ is unitary if and only if $\hat{U} \in \mathbb{R}^{2n\times 2n}$ is orthonormal.

Corollary 2

If $Q \in \mathbb{C}^{n\times n}$ is positive semi-definite, then so is $\hat{Q} \in \mathbb{R}^{2n\times 2n}$.

Lemma 2

Suppose the complex r.ve. $\mathbf{x} \in \mathbb{C}^n$ is zero-mean and satisfies $E\left[xx^\dagger\right] = Q$. Then the differential entropy of $\mathbf{x}$ satisfies $h(\mathbf{x}) \leq \log\det(\pi e Q)$ if and only if $\mathbf{x}$ is circularly symmetric Complex Gaussian with

$$E\left[xx^\dagger\right] = Q$$

In other words, circularly symmetric Complex Gaussian r.ve. are entropy maximizers for the class of complex random vectors.

Properties and Lemmas III

Lemma 3

If $\mathbf{x} \in \mathbb{C}^n$ is a circularly symmetric Complex Gaussian then so is $y = Ax$ for any $A \in \mathbb{C}^{m \times n}$.

Lemma 4

f $\mathbf{x}$ and $\mathbf{y}$ are independent circularly symmetric Complex Gaussians, then $\mathbf{z} = \mathbf{x} + \mathbf{y}$ is also circularly symmetric Complex Gaussian.

Table of Contents

1. Introduction
   - Abstract
   - Notation and Assumptions
   - Properties and Lemmas
2. Gaussian Channel with Deterministic H
   - Capacity Derivation
   - Examples
   - Formal Derivation of Capacity
3. Gaussian Channel with Ergodic $\mathbf{H}$
   - Introduction to the problem
   - Capacity-achieving Distribution
   - Evaluation of the Capacity
   - Examples

Capacity Derivation I

We first consider the case of a deterministic (fixed) transfer function H. Note that the capacity is a function of both the transfer function and the power constraint, i.e. $C(H, P)$. We consider the case where both receiver and transmitter know the matrix H.

Singular Value Decomposition (SVD)

Any matrix $H \in \mathbb{C}^{r \times t}$ can be decomposed as

$$H = UDV^\dagger$$

where $U \in \mathbb{C}^{r \times r}$ and $V \in \mathbb{C}^{t \times t}$ are unitary, $D \in \mathbb{C}^{r \times t}$ is diagonal with non-negative real entries.
In fact, the columns of U are the eigenvectors of $HH^\dagger$, the columns of V are the eigenvectors of $H^\dagger H$ and the diagonal entries of D, called *singular values* of H, are the eigenvalues, which coincide for the two cases and given the hermitianity of such matrices, they are real and non-negative.

Capacity Derivation II

Thus, the problem can be seen as follows

$$\mathbf{y} = UDV^{\dagger}\mathbf{x} + \mathbf{n}$$

Therefore, preprocessing the transmitted symbol as $\mathbf{x} = V\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}} = U^{\dagger}\mathbf{y}$ and defining $\tilde{\mathbf{n}} = U^{\dagger}\mathbf{n}$, we get

$$\tilde{\mathbf{y}} = D\tilde{\mathbf{x}} + \tilde{\mathbf{n}}$$

as an equivalent channel.

Calling $\sigma_i = \lambda_i^{\frac{1}{2}}$ the singular values for $i = 1, \ldots, \min(r, t)$, we have

$$\begin{cases} \tilde{\mathbf{y}}_i = \sigma_i \tilde{\mathbf{x}}_i + \tilde{\mathbf{n}}_i & 1 \leq i \leq \min(r, t) \\ \tilde{\mathbf{y}}_i = \tilde{\mathbf{n}}_i & i > \min(r, t) \end{cases}$$

Capacity Derivation III

It's clear that in order to maximize the capacity we need to choose $\tilde{\mathbf{x}}_i$ for $1 \leq i \leq \min(r, t)$ (actually only up to the number of non-zero singular values) to be independent, zero-mean, circularly symmetric Complex Gaussian.

It can be shown that the optimal variances must follow once again the "water-filling" equations

$$\begin{cases} E[\mathrm{Re}(\tilde{\mathbf{x}}_i)] = E[\mathrm{Im}(\tilde{\mathbf{x}}_i)] = \frac{1}{2}\left(\mu - \lambda_i^{-1}\right)^+ \\ P = \sum_i \left(\mu - \lambda_i^{-1}\right)^+ \\ C = \sum_i \left(\log(\mu\lambda_i)\right)^+ \end{cases}$$

Example

Example 1 MAYBE??????

Formal Derivation of Capacity I

I will now proceed in the derivation of the capacity in a more formal way. This also will be used with some slight differences in the derivation of the capacity when $\mathbf{H}$ is not deterministic.

Recall, the mutual information that we are interested in is

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) = h(\mathbf{y}) - h(\mathbf{n})$$

Note that if $\mathbf{x}$ satisfies $E\left[\mathbf{x}^\dagger \mathbf{x}\right] \leq P$, so does $\mathbf{x} - E[\mathbf{x}]$, thus we will only focus on zero-mean $\mathbf{x}$.

Furthermore, if $\mathbf{x}$ is zero-mean with covariance $E\left[\mathbf{x}\mathbf{x}^\dagger\right] = Q$, then $\mathbf{y}$ is zero-mean with covariance $E\left[\mathbf{y}\mathbf{y}^\dagger\right] = HQH^\dagger + I_r$, and by Lemma 2 among such $\mathbf{y}$ the entropy is largest when $\mathbf{y}$ is circularly symmetric Complex Gaussian, which is the case when $\mathbf{x}$ is circularly symmetric Complex Gaussian (Lemmas 3 and 4).

Formal Derivation of Capacity II

Considering, then, $\mathbf{x}$ circularly symmetric Complex Gaussian, we have

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{n}) &= \log\det\Big(\pi e(I_r + HQH^\dagger)\Big) - \log\det(\pi e I_r) \\ &= \log\det\Big(I_r + HQH^\dagger\Big) = \log\det\Big(I_t + QH^\dagger H\Big) \end{aligned}$$

where the last equality is given by *Sylvester's determinant identity*, that states $\det(I + AB) = \det(I + BA)$.

Since it occurs many times during the paper, we define

$$\Psi(Q, H) = \log\det\Big(I_r + HQH^\dagger\Big)$$

Formal Derivation of Capacity III

Since we're dealing with zero-mean circularly symmetric Complex Gaussian, we only need to find a positive semi-definite matrix Q to fully determine the distribution, such that $\mathrm{tr}(Q) \leq P$.

Notice that $H^\dagger H$ is hermitian, thus it can be decomposed as $H^\dagger H = V^\dagger \Lambda V$ with unitary V and diagonal positive semi-definite Λ. Thus, using Sylvester's identity again we obtain

$$\begin{aligned}
\log\det\Big(I_t + QH^\dagger H\Big) &= \log\det\Big(I_t + \Lambda^{\frac{1}{2}} V Q V^\dagger \Lambda^{\frac{1}{2}}\Big) \\
&= \log\det\Big(I_t + \Lambda^{\frac{1}{2}} \tilde{Q} \Lambda^{\frac{1}{2}}\Big)
\end{aligned}$$

where we noticed that Q and $\tilde{Q}$ are equivalent for both definiteness and power constraint, so we can as well optimize over $\tilde{Q}$.

Formal Derivation of Capacity IV

A special case of *Hadamard's inequality* for positive semi-definite matrices, states that

$$I(\mathbf{x};\mathbf{y}) = \log\det\left(I_t + \Lambda^{\frac{1}{2}}\tilde{Q}\Lambda^{\frac{1}{2}}\right) \leq \prod_i (1+\tilde{Q}_{ii}\lambda_i)$$

with equality for diagonal matrices.

Thus, we see that the maximizing $\tilde{Q} = VQV^\dagger$ is diagonal, and the optimal entries follow the water-filling equations, i.e.

$$\begin{cases} \tilde{Q}_{ii} = (\mu - \lambda_i^{-1})^+ \\ P = \sum_i \tilde{Q}_{ii} \\ C = \sum_i (\log(\mu\lambda_i))^+ \end{cases}$$

Table of Contents

1. Introduction
   - Abstract
   - Notation and Assumptions
   - Properties and Lemmas
2. Gaussian Channel with Deterministic H
   - Capacity Derivation
   - Examples
   - Formal Derivation of Capacity
3. Gaussian Channel with Ergodic $\mathbf{H}$
   - Introduction to the problem
   - Capacity-achieving Distribution
   - Evaluation of the Capacity
   - Examples

Introduction

We now consider the case when the matrix $\mathbf{H}$ is i.i.d. random and ergodic. Ergodicity comes into play to ensure that the capacity is the actual bound as as matrix is randomly changing over time.

We assume that $\mathbf{H}$ is independent of both $\mathbf{x}$ and $\mathbf{n}$, and that the specific realization H is know at the receiver (while the transmitter only knows it's distribution). Thus, we can equivalently say that the channel yields the couple $(\mathbf{y}, \mathbf{H})$.

We will further assume that the entry of $\mathbf{H}$ are i.i.d. zero-mean Complex Gaussians, with independent real and imaginary parts, each with variance $\frac{1}{2}$.

In other words we are considering a Rayleigh fading channel for each component and enough physical separation within transmitting and receiving antennas to achieve independence between the entries of $\mathbf{H}$.

Capacity-achieving Distribution I

This time, we can compute the mutual information as follows

$$\begin{aligned} I(\mathbf{x}; (\mathbf{y}, \mathbf{H})) &= I(\mathbf{x}; \mathbf{H}) + I(\mathbf{x}; \mathbf{y}|\mathbf{H}) \\ &= I(\mathbf{x}; \mathbf{y}|\mathbf{H}) \\ &= E_{\mathbf{H}}[I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H)] \end{aligned}$$

We know that for a deterministic H, the maximizer $\mathbf{x}$ takes the form of a zero-mean circularly symmetric Complex Gaussian with covariance Q, and $\Psi(Q, H) = \log\det\big(I_r + HQH^\dagger\big)$ is the corresponding maximal mutual information.

We thus need to maximize

$$\Psi(Q) = E[\Psi(Q, \mathbf{H})] = E\left[\log\det\left(I_r + \mathbf{H}Q\mathbf{H}^\dagger\right)\right]$$

over the choice of positive semi-definite Q, such that $\mathrm{tr}(Q) \leq P$.

Capacity-achieving Distribution II

Since Q is positive semi-definite, we can write it as $Q = W\Delta W^\dagger$, where W is unitary and Δ is non-negative diagonal. Thus

$$\Psi(Q) = E\Big[\log\det\Big(I_r + (\mathbf{H}W)\Delta(\mathbf{H}W)^\dagger\Big)\Big]$$

It can be shown that $\mathbf{H}W$ has the same distribution as $\mathbf{H}$ (since W is unitary and $\mathbf{H}$ has i.i.d. entries), and thus $\Psi(Q) = \Psi(\Delta)$. It follows that we can focus on non-negative diagonal Q.

Consider the permutation matrix Π (which is orthogonal, thus unitary) and define $Q^\Pi \triangleq \Pi Q \Pi^T$. Again, we note that $\Psi(Q^\Pi) = \Psi(Q)$. Also, note that if Q is diagonal, also Q^Π is diagonal and they share the same determinant (the main diagonal is only shuffled).

Capacity-achieving Distribution III

Note that the transformation $Q \to I_r + HQH^\dagger$ is linear and preserves positive definiteness, and know that the $\log\det$ function is concave over the closed cone of positive semi-definite matrices.
Thus $Q \to \Psi(Q)$ is concave.

Let's define

$$\tilde{Q} = \frac{1}{t!} \sum_{\Pi} Q^{\Pi}$$

Note that $\tilde{Q} = \alpha I$ for some α for any Q, that $\mathrm{tr}\left(\tilde{Q}\right) = \mathrm{tr}(Q)$ and $\Psi(\tilde{Q}) \geq \Psi(Q)$.

Capacity-achieving Distribution IV

Theorem

The capacity of the channel is achieved when $\mathbf{x}$ *is a circularly symmetric Complex Gaussian with zero-mean and covariance* $Q = \frac{P}{t} I_t$.
The capacity is given by

$$E\left[\log\det\left(I_r + \frac{P}{t}\mathbf{H}\mathbf{H}^\dagger\right)\right]$$

.

Note that for fixed r, by the Law of Large Numbers $\frac{1}{t}\mathbf{H}\mathbf{H}^\dagger = I_r$, almost surely as t gets large.
Thus, the capacity in the limit of large t equals

$$r\log(1+P)$$

which scales linearly with the number of receivers!

Evaluation of the Capacity I

First of all, from the Sylvester's determinant identity we know that

$$\det\left(I_r + \frac{P}{t}\mathbf{H}\mathbf{H}^\dagger\right) = \det\left(I_t + \frac{P}{t}\mathbf{H}^\dagger\mathbf{H}\right)$$

We can then define

$$\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^\dagger & r < t \\ \mathbf{H}^\dagger\mathbf{H} & r \geq t \end{cases}$$

and also $m = \min(r, t)$, $M = \max(r, t)$.

Clearly, $\mathbf{W} \in \mathbb{C}^{m \times m}$ is a random positive semi-definite matrix, thus it has random, real, non-negative eigenvalues $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m$.

Evaluation of the Capacity II

From the literature, we know that the entries of $\mathbf{W}$ follow a *Wishart distribution with parameters* m, M and the joint density of the *ordered* eigenvalues is known to be [2]

$$p_{\lambda,\mathsf{ord}} = \frac{1}{K_{m,M}} \prod_i e^{-\lambda_i} \lambda_i^{M-m} \prod_{j>i} (\lambda_i - \lambda_j)^2, \quad \lambda_1 \geq \ldots \geq \lambda_m \geq 0$$

where $K_{m,M}$ is a normalizing factor.

After many complex calculations and transformations we reach...

Evaluation of the Capacity III

Theorem

The capacity of the Gaussian Channel with ergodic matrix $\mathbf{H}$, with t transmitters and r receivers under power constraint P equals

$$\int_0^{\infty} \log\left(1 + \frac{P}{t}\lambda\right) \sum_{k=0}^{m-1} \frac{k!}{(k+M-m)!} L_k^{M-m}(\lambda)^2 \lambda^{M-m} e^{-\lambda}\, \mathrm{d}\lambda \tag{2}$$

where $m = \min(r, t)$, $M = \max(r, t)$, and

$$L_k^{M-m}(x) = \frac{1}{k!} e^x x^{m-M} \frac{\mathrm{d}^k}{\mathrm{d}x^k}\left(e^{-x} x^{M-m+k}\right)$$

are the associated Laguerre polynomials.

Examples I

Example ($t = 1$)

In this case $m = 1$ and $M = r$. Note that $L_0^{r-1}(\lambda) = 1$, thus from Eq. (2) we get

$$C = \frac{1}{(r-1)!} \int_0^\infty \log(1 + P\lambda)\lambda^{r-1} e^{-\lambda} \, \mathrm{d}\lambda$$

As r gets large, so does the capacity.
It can be shown that as $r \to \infty$, $C \sim \log(1 + rP)$, meaning that $C - \log(1 + rP) \xrightarrow{r \to \infty} 0$.

Example

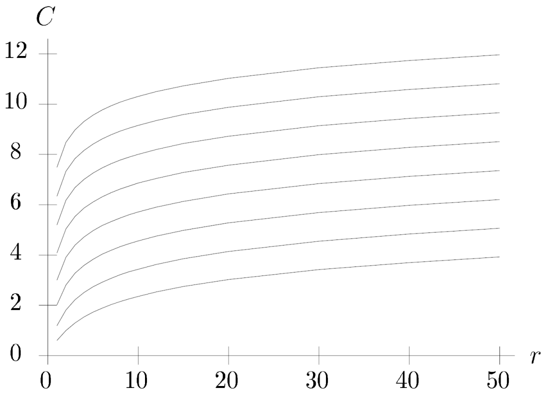

Figure: The value of the capacity (in *nats*) for $0\text{dB} \leq P \leq 35\text{dB}$ in 5dB increments

Examples III

Example ($r = 1$)

In this case $m = 1$ and $M = t$. Differently from before, though, the capacity is

$$C = \frac{1}{(t-1)!} \int_0^{\infty} \log\left(1 + \frac{P}{t}\lambda\right) \lambda^{t-1} e^{-\lambda} \,\mathrm{d}\lambda$$

Quite differently from before, the asymptotic capacity is
$C \xrightarrow{t \to \infty} \log(1 + P)$.
Thus, just increasing the number of transmitting antennas will not help getting more capacity.

Examples IV

Example ($r = 1$)

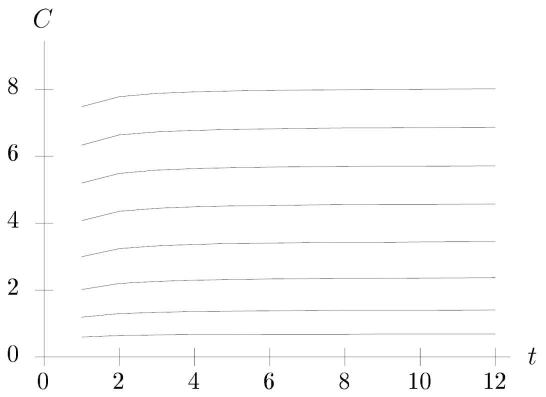

Figure: The value of the capacity (in *nats*) for $0\text{dB} \leq P \leq 35\text{dB}$ in 5dB increments

Examples V

Example ($r = t$)

In this case $m = M = r = t$. From Eq. (2), then, we get

$$C = \int_0^\infty \log\left(1 + \frac{P}{m}\lambda\right) \sum_{k=0}^{m-1} L_k(\lambda)^2 e^{-\lambda} \, \mathrm{d}\lambda$$

There actually exists a close form formula for this integral, which is

$$\lim_{m \to \infty} \frac{C}{m} = \log P - 1 + \frac{\sqrt{1+4P} - 1}{2P} + 2 \tanh^{-1}\left(\frac{1}{\sqrt{1+4P}}\right) \qquad (3)$$

which clearly means that the capacity is a linear function of the number of transmitting/receiving antennas.

Examples VI

Example ($r = t$)

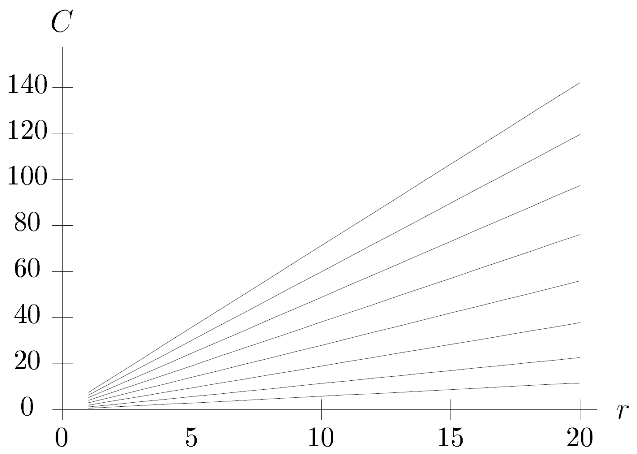

Figure: The value of the capacity (in *nats*) for $0\text{dB} \leq P \leq 35\text{dB}$ in 5dB increments

References

E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, pp. 475–501, 06 1964.