The Jacobi MIMO Channel

Ronen Dar∗, Meir Feder† and Mark Shtaif‡
School of Electrical Engineering
Tel Aviv University
Tel Aviv 69978, Israel
Email: ∗ronendar@post.tau.ac.il, †meir@eng.tau.ac.il, ‡shtaif@tauex.tau.ac.il

Abstract

A Multi-Input Multi-Output (MIMO) channel is defined by a transfer matrix $H$ which couples $m_t$ inputs into $m_r$ outputs. In the Jacobi channel $H$ is an $m_r \times m_t$ sub-matrix of an $m \times m$ Haar-distributed unitary matrix ($m \geq m_t, m_r$). The (squared) singular values of $H$ follow the law of the Jacobi ensemble, which form together with the Gaussian and Wishart ensembles the classical random matrix ensembles; hence the name of the channel. A motivation to define such a channel comes from multimode/multicore optical fiber communication. It turns out that this model is qualitatively different than the Rayleigh model, leading to interesting practical and theoretical results. This work first evaluates the ergodic capacity of the channel. In the non-ergodic case, it analyzes the outage probability and the diversity-multiplexing tradeoff. In the case where $k = m_r + m_t - m > 0$ at least $k$ degrees of freedom are guaranteed not to fade for any channel realization enabling a zero outage probability or infinite diversity order at the corresponding rates. A simple scheme that uses channel state feedback to attain the no-outage guarantee is provided. Finally, we discuss the applications in other communication scenarios.

I. INTRODUCTION

In Multi-Input Multi-Output (MIMO) channels a vector $\underline{x}$ of $m_t$ signals is transmitted, a vector $\underline{y}$ of $m_r$ signals is received, and an $m_r \times m_t$ random matrix $H$ represents the coupling of the input into the output so that the received vector is $\underline{y} = H \underline{x} + \underline{z}$ where $\underline{z}$ is a noise vector. In this paper we consider a channel matrix $H$ which is a sub-matrix of a Haar-distributed unitary matrix, i.e., drawn uniformly from the ensemble of all $m \times m$ unitary matrices matricies, and which form together with the Gaussian and Wishart the classical random matrix ensembles.

The three classical and most well-studied random matrix ensembles are the Gaussian, Wishart and Jacobi (also known as MANOVA) ensembles [1]–[3]. The Gaussian ensemble is a common model for the channel matrix $H$ in fading wireless communication (also known as the Rayleigh model). In that case, $H H^\dagger$ is the Wishart ensemble. For the model assumed in this paper, $H H^\dagger$ follows the Jacobi ensemble. It turns out that this model is both practically useful and it is qualitatively different than other MIMO models such as the Rayleigh [4]–[6], Rician [7]–[9] and Nakagami [9]–[12].

An important motivation to introduce such channels comes from recent developments in optical fiber communication. The expected capacity crunch in long haul optical fibers [13], [14] led to proposals for “space-division multiplexing” (SDM) [15], [16], that is to have several links at the same fiber, by either multiple single-mode fiber strands within a fiber cable, multiple cores within a multi-core fiber, or multiple modes within a multi-mode waveguide. An SDM system with $m$ parallel transmission paths per wavelength can potentially multiply the throughput of a certain link by a factor of $m$. Since $m$ can potentially be chosen very large, SDM technology is highly scalable. Now, a significant crosstalk between the optical paths raises the need for MIMO signal processing techniques. Unfortunately, for large size MIMO (large $m$) this is unfeasible...
Currently in the optical rates. Assuming that faster computation will be available in the future and having in mind that replacing optical fibers to support SDM is a long and expensive procedure, a long term design is sought after. To that end and more, it was proposed to design an optical system that can support relatively large number of paths for future use, but at start to address only some of the paths. In this scenario the channel can be modeled as a sub-matrix of a larger unitary matrix, i.e., the Jacobi MIMO model is applicable.

This under-addressed channel is discussed in [17] where simulations of the capacities and outage probabilities were presented. In this paper we further analyze the channel in the ergodic and non-ergodic setting, where we provide analytical expression for the capacity, outage probability and the diversity-multiplexing tradeoff. It should be noted that in optical systems the outage probability is an important measure, required to be very low. Evidently, since the entire channel matrix is unitary, when all paths are addressed a zero outage probability can be attained for any transmission rate. An interesting result that comes out of this work is that there are situations, where a partial number of paths are addressed, yet a number of streams are guaranteed to experience zero outage. Thus, choosing the number of addressed paths and the corresponding rate is a very critical design element that highly reflects on the system outage and performance.

A possibly practical outcome of this work is a simple communication scheme, with channel state feedback, that achieves the highest rate possible with no outage. The scheme works even when the feedback is “outdated”, and it allows simple decoding with no complicated MIMO signal processing, making it plausible for optical communication. The theoretical findings indicate that the no-outage promise can be attained with no feedback, yet the quest for such simple schemes is open.

While the motivation for this work comes from optical fiber communication, it should be noted that in other cases, such as in-line communication and even wireless communication, this model and the insights that follow from it can be relevant. For example, in wireless communication, it is plausible to imagine that if there were enough receive antennas capturing most, if not all, transmitted energy, the unitary assumption can be justified, and so with the smaller number of antennas the channel can be modeled as a sub-matrix of a large unitary matrix. Indeed, as will be shown, when \( m \) is large in comparison to \( m_t, m_r \), the Jacobi model (up to a normalizing constant) approaches the Rayleigh model.

The paper is organized as follows. We start by defining the system model and presenting the channel statistics in Section II. An interesting transition threshold is revealed: when the number of addressed paths is large enough, so that \( k = m_t + m_r - m > \), the statistics of the problem changes. Using this observation we give analytic expressions for the ergodic capacity in Section III. In Section IV we analyze the outage probabilities in the non-ergodic channel and show that for \( k > 0 \) a strictly zero outage probability is obtainable for \( k \) degrees of freedom. Following this finding, we present in Section V a new communication scheme which exploits a channel state feedback to achieve zero outage probability. Section VI discuss the diversity-multiplexing tradeoff of the channel where we show an absorbing difference in the maximum diversity gain between the Rayleigh fading and Jacobi channels. Section VIII summarizes and discuss the results.

II. SYSTEM MODEL AND CHANNEL STATISTICS

We consider a space-division multiplexing system that supports \( m \) spatial orthogonal propagation modes. We assume a unitary coupling among all transmission modes, allowing us to describe the transfer matrix as \( m \times m \) unitary matrix, denoted \( H \), where each entry \( h_{ij} \) represents the
complex path gain from transmitted mode \( i \) to received mode \( j \). We further assume a uniformly distributed unitary coupling, that is, the channel matrix \( H \) is assumed to be drawn uniformly from the ensemble of all \( m \times m \) unitary matrices. However, we consider the case where not all modes are being addressed, that is, the transmitter can excite \( m_t \leq m \) modes and the receiver can coherently extract \( m_r \leq m \) modes. Hence, neglecting waveguide nonlinearities and ignoring differential modal delays and mode dependent losses (MDL), the channel can be written as:

\[
y = \sqrt{\text{SNR}} H_{11} x + z,
\]

where \( x \in \mathbb{C}^{m_t} \) is the transmitted signal; \( y \in \mathbb{C}^{m_r} \) is the received signal; the additive noise \( z \) has i.i.d circularly symmetric complex Gaussian entries \( z_i \sim \mathcal{CN}(0,1) \), \( i = 1, \ldots, m_r \); \( H_{11} \) is the \( m_r \times m_t \) sub-matrix of the unitary matrix

\[
H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix};
\]

Following [17], SNR is a constant defined so that it equals the average signal-to-noise ratio at each received mode when all channel modes are excited (that is, when \( m_t = m \)). In this paper we assume an equal power constraint on each transmit mode (which is the common power constraint in optical communication systems), that is, the average signal energy transmitted at each symbol period and from each transmit mode is constrained to be not greater than 1:

\[
\mathbb{E}[x_i^* x_i] \leq 1 \quad \forall \ i = 1, \ldots, m_t.
\]

Now, to be able to analytically analyze this channel we need to understand the statistics of the channel matrix \( H_{11} \). To that end we recall the three classical random matrix ensembles [1]–[3], the Gaussian (Hermite), Wishart (Laguerre) and Jacobi (MANOVA - multivariate analysis of variance) ensembles. We limit the discussion to complex ensembles.

**Definition 1 (Gaussian ensemble).** \( \mathcal{G}(m,n) \) is \( m \times n \) matrix of i.i.d complex entries distributed as \( \mathcal{CN}(0,1) \).

The following ensembles are constructed from the Gaussian ensemble as follows.

**Definition 2 (Wishart ensemble).** \( \mathcal{W}(m,n) \), where \( m \geq n \), is \( n \times n \) Hermitian matrix which can be constructed as \( A^H A \), where \( A \in \mathcal{G}(m,n) \).

**Definition 3 (Jacobi ensemble).** \( \mathcal{J}(m_1,m_2,n) \), where \( m_1, m_2 \geq n \), is \( n \times n \) Hermitian matrix which can be constructed as \( A (A + B)^{-1} \), where \( A \) and \( B \) are \( \mathcal{W}(m_1,n) \) and \( \mathcal{W}(m_2,n) \), respectively.

The first two ensembles relate to wireless communication [6]. The Gaussian ensemble is the most common statistical model for the wireless MIMO channel. The singularity statistics of this channel follow the law of the Wishart ensemble. We claim here that the third classical ensemble, the Jacobi ensemble, is relevant to the discussed channel model.

There is a deep connection between Jacobi matrices and truncated Haar-distributed unitary matrices. More precisely, let

\[
U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}
\]

be \( m \times m \) Haar-distributed unitary matrix (that is, drawn uniformly from the ensemble of all
distribution with the eigenvalues of the Jacobi ensemble $J$.

For $m_t + m_r \leq m$, we can say that the statistics of the eigenvalues of $H$ are the same as those of $H_1 H_1^\dagger$, where $H_1$ is a Haar-distributed $m \times m$ unitary matrix, and $H$ is a $m \times m$ matrix with eigenvalues $\lambda$. The case of $m_t + m_r > m$ follows (4). By Lemma 1, the singular values of $H_1 H_1^\dagger$ follow the Jacobi ensemble (3) is well known [1]:

\[ f_\lambda(m_t, m_r, m; \lambda_1, \ldots, \lambda_{\min\{m_t, m_r\}}) = K_{m_t, m_r, m}^{-1} \prod_{i=1}^{\min\{m_t, m_r\}} \lambda_i^{m_t - m_r + (1 - \lambda_i)^{m_m - m_r - m}} \prod_{i<j} (\lambda_i - \lambda_j)^2, \]

where $K_{m_t, m_r, m}$ is a normalizing constant. Thus, the joint pdf of the ordered non-zero eigenvalues of $H_1^\dagger H_1$, for the case of $m_t + m_r \leq m$, follows (4).

B. The case of $m_t + m_r > m$

When the sum of transmitted and received addressed modes, $m_t + m_r$, is larger than the total available modes, $m$, the statistics of the singular values change. Having in mind that the columns of $H$ are orthonormal, one can think of $m_t + m_r > m$ as a transition threshold in which the sub-matrix $H_1 H_1^\dagger$ is large enough with respect to $H$ to change the singularity statistics. The following Lemma provides the joint pdf of the singular values of $H_1$, showing that for any realization of $H_1$ there are $m_t + m_r - m$ singular values which are 1.

**Lemma 1.** Suppose

\[ H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \]

is Haar distributed $m \times m$ unitary matrix, where $H_{11}$ is $m_r \times m_t$ matrix and $m_t + m_r > m$. By denoting $m_{\max} = \max\{m_t, m_r\}$ and $m_{\min} = \min\{m_t, m_r\}$, $H_1^\dagger H_{11}$ has:

- $m_t + m_r - m$ eigenvalues which are 1.
- $m_t - m_{\min}$ zero eigenvalues.
• \(m - m_{\max}\) eigenvalues which are equal to the non-zeros eigenvalues of \(H_{22}^\dagger H_{22}\), thereby are distributed according to the Jacobi ensemble

\[ \mathcal{J}(m - m_{\min}, m_{\min}, m - m_{\max}) . \]

**Proof:** We denote by \(\lambda_{1}^{(kj)}, \ldots, \lambda_{m_{t}}^{(kj)}\) the eigenvalues of \(H_{kj}^\dagger H_{kj}\) for \((k, j) = (1, 1), (2, 1)\). We further let \(\tilde{\lambda}_{1}^{(kj)}, \ldots, \tilde{\lambda}_{m - m_{r}}^{(kj)}\) be the eigenvalues of \(H_{kj}^\dagger H_{kj}\) for \((k, j) = (2, 1), (2, 2)\). Since \(H\) is unitary we can write

\[ H_{11}^\dagger H_{11} + H_{21}^\dagger H_{21} = I_{m_{r}} , \]  

(5)

where here and throughout the rest of the paper we denote by \(I_{n}\) the \(n \times n\) identity matrix. Thus, we get

\[ \lambda_{i}^{(11)} = 1 - \lambda_{i}^{(21)} \quad \forall \ i = 1, \ldots, m_{t} . \]  

(6)

In the same manner we have \(H_{21}^\dagger H_{21} + H_{22}^\dagger H_{22} = I_{m - m_{r}},\) thus

\[ \tilde{\lambda}_{i}^{(21)} = 1 - \tilde{\lambda}_{i}^{(22)} \quad \forall \ i = 1, \ldots, m - m_{r} . \]  

(7)

Now, \(H_{22}\) is \((m - m_{r}) \times m_{t}\) matrix. Since \(m - m_{r} < m_{t}, H_{22}^\dagger H_{22}\) has (at least) \(m_{t} + m_{r} - m\) zero eigenvalues; thus, by applying (6), \(H_{11}^\dagger H_{11}\) has (at least) \(m_{t} + m_{r} - m\) eigenvalues which are 1. Since \(H_{kj}^\dagger H_{kj}\) and \(H_{kj}^\dagger H_{kj}\) share the same non-zero eigenvalues we can combine (6) and (7) to conclude that the additional \(m - m_{\max}\) non-zeros eigenvalues of \(H_{11}^\dagger H_{11}\) are equal to the \(m - m_{\max}\) non-zeros eigenvalues of \(H_{22}^\dagger H_{22}\).

Above arguments hold true for any unitary matrix, in particular for any realization of the transfer matrix \(H\). Noting that \(H_{22}\) is \((m - m_{r}) \times (m - m_{t})\) matrix and therefore applies to the first case \(((m - m_{r}) + (m - m_{t}) < m),\) completes the proof.

\[ \blacksquare \]

**III. Ergodic Channel**

In this section we assume that the channel is rapidly changing or the signal samples the entire channel statistics. The channel is assumed to be known to the receiver but not to the transmitter. In this case, the mutual information between the input and output of the channel is

\[ E[I(x; y | H_{11} = H_{11})] , \]  

(8)

where the expectation is over \(H_{11}\). Since the channel is fast fading, i.e. the signal samples the entire channel statistics, we average over the channel matrix distribution.

It is well known that a circularly symmetric zero-mean Gaussian input distribution achieves the capacity of this channel, which is given by

\[ C(m_{t}, m_{r}, m; \text{SNR}) = \max_{Q, Q \succeq 0, Q_{ii} \leq 1 \forall i = 1, \ldots, m_{t}} E[\log \det(I_{m_{r}} + \text{SNR} \cdot H_{11} Q H_{11}^\dagger)] , \]  

(9)

where \(Q\) is the covariance matrix of the transmitted signal and is chosen to maximize the average mutual information. The following theorem shows that the identity matrix achieves capacity. We note that because of the equal power constraint per-mode, the capacity is symmetric in \(m_{t}\) and \(m_{r}\).

**Theorem 1.** The ergodic capacity of the channel is achieved when the transmitted signal is
circularly symmetric zero-mean Gaussian with covariance $I_m$ and is given by

$$C(m_t, m_r, m; \text{SNR}) = \mathbb{E}[\log \det(I_{m_t} + \text{SNR} \cdot H^\dagger_{11} H_{11})].$$

(10)

Proof: The capacity in (9) satisfies:

$$C(m_t, m_r, m; \text{SNR}) = \max_{Q: Q \succeq 0, \text{trace}(Q) \leq m_t} \mathbb{E}[\log \det(I_{m_r} + \text{SNR} \cdot H^\dagger_{11} Q H_{11})].$$

(11)

$$\leq \max_{Q: Q \succeq 0} \mathbb{E}[\log \det(I_{m_r} + \text{SNR} \cdot H^\dagger_{11} H_{11})].$$

(12)

In [5, Theorem 1] it was shown that $Q = I_{m_t}$ maximizes (12) for any distribution of $H_{11}$ that is invariant under unitary permutation. Since $Q = I_{m_t}$ satisfies also $Q_{ii} \leq 1$ for $i = 1, \ldots, m_t$ we can write

$$C(m_t, m_r, m; \text{SNR}) = \mathbb{E}[\log \det(I_{m_r} + \text{SNR} \cdot H^\dagger_{11} H_{11})].$$

(13)

where it can be easily shown that the distribution of $H_{11}$ is invariant under unitary permutation since $H$ is Haar-distributed, that is, invariant under unitary permutation.

To complete the proof we use

$$\det(I_{m_r} + \text{SNR} \cdot H^\dagger_{11} H_{11}) = \log \det(I_{m_t} + \text{SNR} \cdot H^\dagger_{11} H_{11}).$$

A. The case of $m_t + m_r \leq m$

The following theorem gives an analytical expression to the ergodic capacity for $m_t + m_r \leq m$.

Using the joint pdf of the eigenvalues of the Jacobi ensemble we associate the ergodic capacity with the Jacobi polynomials [19, 8.96].

Theorem 2. The ergodic capacity, for $m_t + m_r \leq m$, satisfies

$$C(m_t, m_r, m; \text{SNR}) = \int_0^1 \log(1 + \lambda \cdot \text{SNR}) \cdot \lambda^\alpha (1 - \lambda)^\beta \sum_{k=0}^{m_{\min}-1} b_{k,\alpha,\beta} [P_k^{(\alpha,\beta)}(1 - 2\lambda)]^2 d\lambda,$$

(14)

where we denote $m_{\min} = \min\{m_t, m_r\}$, $\alpha = |m_r - m_t|$, $\beta = m - m_t - m_r$,

$$b_{k,\alpha,\beta} = \frac{1}{2k + \alpha + \beta + 1} \binom{2k + \alpha + \beta}{k} \binom{2k + \alpha + \beta}{k + \alpha}^{-1}.$$

and $P_k^{(\alpha,\beta)}(x)$ are the Jacobi polynomials

$$P_k^{(\alpha,\beta)}(x) = \frac{(-1)^k}{2^k k!} (1 - x)^{-\alpha} (1 + x)^{-\beta} \frac{d^k}{dx^k} [(1 - x)^{k+\alpha} (1 + x)^{k+\beta}].$$

Proof: See Appendix A.

B. The case of $m_t + m_r > m$

Here we use Lemma 1 to compute the ergodic capacity:

Theorem 3. The ergodic channel capacity, in case $m_t + m_r > m$, is given by

$$C(m_t, m_r, m; \text{SNR}) = (m_t + m_r - m) \cdot C(1, 1; \text{SNR}) + C(m - m_r, m - m_t, m; \text{SNR}),$$

where

$$C(m_t, m_r, m; \text{SNR}) = \int_0^1 \log(1 + \lambda \cdot \text{SNR}) \cdot \lambda^\alpha (1 - \lambda)^\beta \sum_{k=0}^{m_{\min}-1} b_{k,\alpha,\beta} [P_k^{(\alpha,\beta)}(1 - 2\lambda)]^2 d\lambda,$$

and $P_k^{(\alpha,\beta)}(x)$ are the Jacobi polynomials

$$P_k^{(\alpha,\beta)}(x) = \frac{(-1)^k}{2^k k!} (1 - x)^{-\alpha} (1 + x)^{-\beta} \frac{d^k}{dx^k} [(1 - x)^{k+\alpha} (1 + x)^{k+\beta}].$$
Fig. 1. The ergodic capacity normalized by \( C(1, 1, 1; \text{SNR}) = \log(1 + \text{SNR}) \) vs. SNR (which is defined by the channel model (1)). Note that the ergodic capacity is symmetric in \( m_t, m_r \).

(a) Fixed number of supported modes \( m = 4 \), various transmit \( \times \) receive addressed modes.
(b) Fixed number of addressed modes \( m_t = m_r = 2 \), various values of supported modes \( m \).

where \( C(1, 1, 1; \text{SNR}) \) is the SISO channel capacity
\[
C(1, 1, 1; \text{SNR}) = \log(1 + \text{SNR})
\]

and \( C(m - m_r, m - m_t, m; \text{SNR}) \) is given by Theorem 2

Proof: Let
\[
H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}
\]
be the \( m \times m \) unitary coupling matrix of the channel, where \( H_{11} \) is the \( m_r \times m_t \) coupling matrix between the addressed modes. By Theorem 1 the ergodic capacity is
\[
C(m_t, m_r, m; \text{SNR}) = \mathbb{E}\left[ \log \det(I_{m_t} + \text{SNR} \cdot H_{11}^\dagger H_{11}) \right] = \mathbb{E}\left[ \sum_{i=1}^{\min\{m_t, m_r\}} \log(1 + \text{SNR} \cdot \lambda_i) \right].
\]

where \( \lambda_1, \ldots, \lambda_{\min\{m_t, m_r\}} \) are the non-zeros eigenvalues of \( H_{11}^\dagger H_{11} \). According to Lemma 1, \( m_t + m_r - m \) eigenvalues are 1 where the others are equal to the non-zeros eigenvalues of \( H_{22}^\dagger H_{22} \). Thus, we can write:
\[
C(m_t, m_r, m; \text{SNR}) = (m - m_r - m_t) \log(1 + \text{SNR}) + \mathbb{E}[\log \det(I_{m-m_t} + \text{SNR} \cdot H_{22}^\dagger H_{22})].
\]

We finish by reminding that the capacity of the discussed channel is symmetric in the number of transmitted and received modes, i.e., \( C(m - m_t, m - m_r, m; \text{SNR}) = C(m - m_r, m - m_t, m; \text{SNR}) \).

Fig. 1(a) depict the ergodic capacity (normalized by \( \log(1 + \text{SNR}) \)) as a function of SNR.

\(^1\)We define \( C(0, m - m_t, m; \text{SNR}) = C(m - m_r, 0, m; \text{SNR}) = 0.\)
for $4 \times 4$ unitary channel, for various transmit $\times$ receive systems. Theorem 3 suggests that the ergodic capacity, for the case of $m_t + m_r > m$, is as if the channel was composed of $m_t + m_r - m$ parallel SISO channels and a single MIMO channel of $m - m_r$ transmit modes and $m - m_t$ received modes. Note, that an $(m - m_t) \times (m - m_r)$ system satisfies $(m - m_r) + (m - m_t) \leq m$, therefore its capacity is given by Theorem 2. Thus, the ergodic capacity can be viewed as the sum of $\min(m_t + m_r - m, 0)$ SISO channel capacities and a residual MIMO capacity. In Fig. 1(a) one can observe that the “residual” capacity is equal for $1 \times 1$ and $3 \times 3$ systems and for $1 \times 2$ and $2 \times 3$ systems. In Fig. 1(b) we fix the number of addressed modes $m_t = m_r = 2$ and compute the ergodic capacity for different sizes of unitary matrices.

IV. NON-ERGODIC CHANNEL

In this section we assume that the channel is flat fading, i.e., the channel matrix is drawn randomly but is held fixed for the entire transmission period. The channel is assumed to be known to the receiver but not to the transmitter. In this case, the mutual information between the input and output of the channel is

$$I(x; y| H_{11} = H_{11}) .$$

(18)

Since $H_{11}$ is unknown to the transmitter, there may be a non-zero probability that the transmission rate $R$ (bits/symbol) is not supported by the channel instantiation, that is, a non-zero probability that the mutual information $I(x; y| H_{11} = H_{11})$ is smaller than $R$ and reliable communication in not feasible at this rate. This probability is termed outage probability and is given by

$$Pr[I(x; y| H_{11}) < R] .$$

For a circularly symmetric zero-mean Gaussian input distribution (which maximize the above mutual information) the outage probability satisfies

$$I(x; y| H_{11} = H_{11}) = \log \det(I_{m_r} + \text{SNR} \cdot H_{11} Q H_{11}^\dagger) ,$$

where $Q$ is the covariance matrix of the transmitted signal. We can choose $Q$ to minimize the outage probability, that is

$$Q = \arg \inf_{Q \succeq 0} \left\{ Pr[\log \det(I_{m_r} + \text{SNR} \cdot H_{11} Q H_{11}^\dagger) < R] \right\} .$$

(19)

Note that we employed the per-mode power constraint. In this work we simplify calculations by taking the covariance matrix $Q$ to be the identity matrix $I_{m_t}$. In this case the outage probability $P_{out}(m_t, m_r, m; R)$ satisfies

$$P_{out}(m_t, m_r, m; R) = Pr[\log \det(I_{m_r} + \text{SNR} \cdot H_{11} H_{11}^\dagger) < R] .$$

(20)

We next give a simple analysis of the outage probability for the case of $m_t + m_r \leq m$ while for $m_t + m_r > m$ we show that a zero outage probability is achievable for certain rates.
Fig. 2. $l(R, \text{SNR})$ as a function of $m_r$ for $m - m_r = 1, 10, 100, 1000$. Curves are drawn for outage probabilities $10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$.

A. The case of $m_t + m_r \leq m$

The outage probability for the case of $m_t + m_r \leq m$, where the covariance matrix of the transmitted signal is $I_m$, satisfies

$$P_{\text{out}}(m_t, m_r, m; R) = P \left[ \prod_{i=1}^{\min\{m_t, m_r\}} (1 + \text{SNR} \cdot \lambda_i) < 2^R \right],$$

where $\lambda_1 \leq \ldots \leq \lambda_{\min\{m_t, m_r\}}$ are the ordered non-zeros eigenvalues of $H_{11}^H H_{11}$. We can now apply the joint pdf of the relevant Jacobi ensemble to compute the outage probability:

$$P_{\text{out}}(m_t, m_r, m; R) = K_{m_t, m_r, m}^{-1} \int_{\mathcal{B}} \prod_{i=1}^{\min\{m_t, m_r\}} \lambda_i^{m_r - m_t} (1 - \lambda_i)^{m - m_r - m_t} \prod_{i<j} (\lambda_i - \lambda_j)^2 d\lambda,$$

where $K_{m_t, m_r, m}$ is a normalizing factor and

$$\mathcal{B} = \left\{ \lambda : 0 \leq \lambda_1 \leq \ldots \leq \lambda_{\min\{m_t, m_r\}} \leq 1, \quad \prod_{i=1}^{\min\{m_t, m_r\}} (1 + \text{SNR} \cdot \lambda_i) < 2^R \right\}$$

is the set that describes the outage event.

This gives an analytical expression to the outage probability for the case of $Q = I_m$. See Fig. 3.

Example 1 ($\min\{m_t, m_r\} = 1$). Suppose $\min\{m_t, m_r\} = 1$ and $m \geq 1 + \max\{m_t, m_r\}$. By denoting $m_{\max} = \max\{m_t, m_r\}$ the outage probability satisfies

$$P_{\text{out}}(m_{\max}, 1, m; R) = P_{\text{out}}(1, m_{\max}, m; R) = K_{1, m_{\max}, m}^{-1} \int_{l(R, \text{SNR})} \lambda^{m_{\max} - 1} (1 - \lambda)^{m - m_{\max} - 1} d\lambda,$$

where $l(R, \text{SNR})$ satisfies

$$l(R, \text{SNR}) = \sup\{\lambda : \log(1 + \text{SNR} \cdot \lambda) < R\} = \frac{2^R - 1}{\text{SNR}}.$$
Thus, we can write

\[ P_{\text{out}}(m_{\text{max}}, 1, m; R) = P_{\text{out}}(1, m_{\text{max}}, m; R) = \frac{B(l(R, \text{SNR}); m_{\text{max}}, m - m_{\text{max}})}{B(1; m_{\text{max}}, m - m_{\text{max}})}, \tag{24} \]

where \( B(x; a, b) \) is the incomplete beta function.

Given a desired outage probability one can compute \( l(R, \text{SNR}) \). As \( l(R, \text{SNR}) \) is higher one can afford higher data rate or lower SNR (lower transmission power). In Fig. 2 we take \( m_t = 1 \) and plot \( l(R, \text{SNR}) \) as a function of \( m_r \) for \( m - m_r = 1, 10, 100, 1000 \) and outage probabilities \( 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5} \). As the desired outage probability is smaller, \( l(R, \text{SNR}) \) is lower. This is also true as \( m \) is larger with respect to \( m_r \) (since more power is lost in the un-addressed receive modes). As \( m_r \) is larger, \( l(R, \text{SNR}) \) is higher (since the diversity at the receiver is higher, see Section VI for more information about the diversity gain in the Jacobi channel).

B. The case of \( m_t + m_r > m \)

By applying Lemma 1 we can analyze the outage probability for the case of \( m_t + m_r > m \):

**Theorem 4.** Suppose the transmission rate is

\[ R = r \log(1 + \text{SNR}) \text{ bps/Hz}, \]

with \( 0 \leq r \leq \min\{m_t, m_r\} \) in a channel that satisfies \( m_t + m_t > m \). Then, for

- \( r < (m_t + m_r - m) \) the outage probability is strictly zero.
- \( r \geq (m_t + m_r - m) \) the outage probability satisfies

\[ P_{\text{out}}(m_t, m_r, m; r \log(1 + \text{SNR})) = P_{\text{out}}(m - m_r, m - m_t, m; \tilde{r} \log(1 + \text{SNR})) , \tag{25} \]

where \( \tilde{r} = r - (m_t + m_r - m) \).

**Proof:** Let

\[ H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \]

be the \( m \times m \) unitary coupling matrix of the channel, where \( H_{11} \) is the \( m_r \times m_t \) coupling matrix between the addressed modes. By (24) the outage probability is

\[ P_{\text{out}}(m_t, m_r, m; r \log(1 + \text{SNR})) = \text{Pr} \left[ \prod_{i=1}^{\min\{m_t, m_r\}} \left( 1 + \text{SNR} \cdot \lambda_i \right) < (1 + \text{SNR})^{r} \right], \tag{26} \]

where \( \lambda_1 \leq \ldots \leq \lambda_{\min\{m_t, m_r\}} \) are the ordered non-zeros eigenvalues of \( H_{11}^\dagger H_{11} \). Applying Lemma 1 we get:

\[ P_{\text{out}}(m_t, m_r, m; r \log(1 + \text{SNR})) = \text{Pr} \left[ \prod_{i=1}^{m - \max\{m_t, m_r\}} \left( 1 + \text{SNR} \cdot \tilde{\lambda}_i \right) < (1 + \text{SNR})^{r - (m_t + m_r - m)} \right], \tag{27} \]

where \( \tilde{\lambda}_1 \leq \ldots \leq \tilde{\lambda}_{m - \max\{m_t, m_r\}} \) are the ordered non-zeros eigenvalues of \( H_{22}^\dagger H_{22} \). If \( r < m_t + m_r - m \), this probability is zero. Otherwise, the right hand is the outage probability in a
Fig. 3. Outage probability for a 20dB SNR vs. normalized rate. SNR is defined by the channel model (1).

(a) The number of supported modes is fixed $m = 4$, various transmit × receive addressed modes (note that the outage probability is symmetric in $m_t$, $m_r$).

(b) The number of addressed modes is fixed $m_t = m_r = 2$, for different number of supported modes.

system with $m - m_t$ and $m - m_r$ addressed modes at the transmitter and receiver, correspondingly, with a transfer matrix $H_{22}$.

We finish by reminding that the outage probability is symmetric in the number of transmitted and received modes, i.e., $P_{\text{out}}(m_t - m_r, m_r - m; R) = P_{\text{out}}(m_r - m_t, m_t - m; R)$.

Note that the right hand side of (25) is the outage probability of an $(m - m_t) \times (m - m_r)$ channel which applies to the first case $(m - m_r) + (m - m_t) \leq m$. Thus Theorem 4 states that one can transmit at rates below $(m_t + m_r - m) \cdot \log(1 + \text{SNR})$ with no outage. Intuitively, with probability 1, $m_t + m_r - m$ singular values of $H_{11}$ are 1, that is, the channel consists a non-fading $(m_t + m_r - m)$-dimensional subspace for any realization of $H_{11}$. One would like to transmit $(m_t + m_r - m)$ streams over this subspace to achieve zero outage probability. In the residual $(m - \max\{m_t, m_r\})$-dimensional subspace there is a non-zero probability for fading.

Fig. 3(a) depict the outage probability in a 20dB SNR for $4 \times 4$ unitary channel, for different number of addressed modes at the receiver and transmitter. This figure was also obtained in [17] using simulations. Here, these curves were computed analytically using (22) and Theorem 4. In Fig. 3(b) we keep the number of transmit and receive modes fixed, presenting the outage curves for various values of supported modes $m$.

V. ACHIEVING ZERO OUTAGE PROBABILITY

We now present a new communication scheme for $m_r \times m_t$ MIMO channel, where $m_t + m_r > m$. According to Lemma 1, $m_t + m_r - m$ singular values of $H_{11}$ are 1. The scheme, using simple manipulations, exploits a (delayed) feedback system to complete also the other $m - \max\{m_t, m_r\}$ singular values to 1. Thus the channel is transformed into $m_t + m_r - m$ independent SISO channels, supporting $m_t + m_r - m$ streams (degrees of freedom) with zero outage probability. Furthermore, our scheme removes the need for MIMO processing and allows the use of simple SISO channel decoders.
We first describe the principles of the scheme using the following simple example: Let the transmitter and receiver address 3 out of 4 available modes, i.e., the transform matrix is a 3 × 3 sub-matrix of a 4 × 4 unitary matrix (which is drawn uniformly from the manifold of all 4 × 4 unitary matrices). According to Theorem 4, two degrees of freedom can be communicated to the receiver with zero outage probability. Now, suppose only for the simplicity of the example that the channel instantiation changes independently at each channel use and let
\[ H^{(i)} = \begin{bmatrix} H_{11}^{(i)} & H_{12}^{(i)} \\ H_{21}^{(i)} & H_{22}^{(i)} \end{bmatrix} \]
be the unitary matrix realization at channel use \( i \). In addition, suppose that at each channel use \( i \) the transmitter has perfect knowledge of \( H_{21}^{(i-1)} \), the \((m - m_r) \times m_t\) sub-matrix realization of \( H^{(i-1)} \). Let the transmitter excite the following three entries vector at each channel use \( i = 1, \ldots, n \):
\[ \mathbf{x}^{(i)} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ H_{21}^{(i-1)} \mathbf{x}^{(i-1)} \end{bmatrix} \]
where we define \( \mathbf{x}^{(0)} \) to be a vector of zeros. Hence, in each channel use the transmitter communicates two new information bearing symbols and a linear combination of the previous signal.

The received signal at each channel use \( i = 1, \ldots, n \) is
\[ \mathbf{y}^{(i)} = \sqrt{\text{SNR}} H_{11}^{(i)} \mathbf{x}^{(i)} + \mathbf{z}^{(i)} . \]

Now, since \( H_{11}^{(i)} \) is assumed to be known to the receiver, \( H_{21}^{(i)} \) can be also computed using the orthonormality of \( H^{(i)} \)'s columns. We further assume that the receiver has as a side information the following noisy measure of \( \mathbf{x}^{(n)} \)
\[ \mathbf{y}^{(n)}_{si} = \sqrt{\text{SNR}} H_{21}^{(n)} \mathbf{x}^{(n)} + \mathbf{z}^{(n)}_{si} , \]
where \( \mathbf{z}^{(n)}_{si} \sim \mathcal{CN}(0, 1) \) is independent of \( \mathbf{z}^{(n)} \). Thus, the receiver can construct the following vector
\[ \mathbf{\hat{y}}^{(n)} = \begin{bmatrix} H_{11}^{(n)} & H_{21}^{(n)} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{y}^{(n)} \\ \mathbf{y}^{(n)}_{si} \end{bmatrix} \]
which satisfies
\[ \mathbf{\hat{y}}^{(n)} = \begin{bmatrix} \mathbf{y}_1^{(n)} \\ \mathbf{y}_2^{(n)} \\ \mathbf{y}_3^{(n)} \end{bmatrix} = \sqrt{\text{SNR}} \mathbf{x}^{(n)} + \mathbf{\tilde{z}}^{(n)} \]
\[ = \sqrt{\text{SNR}} \begin{bmatrix} \mathbf{x}_1^{(n)} \\ \mathbf{x}_2^{(n)} \\ H_{21}^{(n-1)} \mathbf{x}^{(n-1)} \end{bmatrix} + \mathbf{\tilde{z}}^{(n)} , \]
where \( \mathbf{\tilde{z}}^{(n)} \) has i.i.d. \( \mathcal{CN}(0, 1) \) entries. Letting \( \mathbf{y}_3^{(n)} \) be \( \mathbf{y}^{(n-1)}_{si} \), the side information for channel
use \( n - 1 \) and repeating this procedure for \( i = n - 1 \) to 1 results in two streams of measures
\[
\begin{bmatrix}
\tilde{y}_1^{(1)} \\
\tilde{y}_2^{(1)} \\
\vdots \\
\tilde{y}_1^{(n)} \\
\tilde{y}_2^{(n)}
\end{bmatrix},
\]
Thus, we get two information streams that as if were communicated through two independent AWGN SISO channels, each with signal-to-noise ratio SNR (and therefore with zero outage probability).

Note that the scheme is feasible if the side information after channel use \( n \) is being conveyed by the transmitter through a neglcajble number of channel uses (with respect to \( n \)) and if a feedback system is employed to communicate \( H_{21}^{(i)} \) to the transmitter after each channel use \( i \).

We now formalize the scheme for any \( m_r \times m_t \) Jacobi channel that satisfies \( m_t + m_r - m > 1 \). Let
\[
H^{(i)} = \begin{bmatrix}
H_{11}^{(i)} & H_{12}^{(i)} \\
H_{21}^{(i)} & H_{22}^{(i)}
\end{bmatrix}
\]
be the unitary matrix realization at channel use \( i \). We assume perfect knowledge of \( H_{11}^{(i)} \) at the receiver and a noiseless feedback communication with a delay of \( k \) channel uses. Since \( H^{(i)} \) unitary, \( H_{21}^{(i)} \) can be computed from \( H_{11}^{(i)} \) and we assume that the receiver noiselessly communicates \( H_{21}^{(i)} \) to the transmitter. Note that \( H_{21}^{(i)} \) completes \( H_{11}^{(i)} \)'s columns into orthonormal columns, thus for \( m_t + m_r - m > 1 \) and certain matrix instantiations, the computed \( H_{21}^{(i)} \) is not unique and can be chosen wisely (see Remark 4).

Now, the transmitter excites the following signal from the addressed modes at each channel use \( i = 1, \ldots, nk \):
\[
\mathbf{x}^{(i)} = \begin{bmatrix}
\mathbf{x}_1^{(i)} \\
\vdots \\
\mathbf{x}_{m_t + m_r - m}^{(i)} \\
\mathbf{H}_{21}^{(i-k)} \mathbf{x}^{(i-k)}
\end{bmatrix},
\]
where \( \mathbf{x}^{(i)} \) for \( i = -(k-1), \ldots, 0 \), is a vector of zeros. That is, the transmitter conveys \( m_t + m_r - m \) new information bearing symbols and \( \mathbf{H}_{21}^{(i-k)} \mathbf{x}^{(i-k)} \), a linear combination of the signal that was transmitted \( k \) channel uses before. Note that since \( H \) unitary, the power constraint is left satisfied.

Now, assume the transmitter communicates to the receiver the following measures
\[
\mathbf{y}_z^{(i)} = \sqrt{\text{SNR}} \mathbf{H}_{21}^{(i)} \mathbf{x}^{(i)} + \mathbf{z}_z^{(i)} \quad \forall \ i = (n-1)k + 1, \ldots, nk,
\]
that is, noisy measures of the last \( k \) transmitted signals, where \( \mathbf{z}_z^{(i)} \) are independent with i.i.d.
$\mathcal{CN}(0,1)$ entries. As was shown above, the receiver can use the side information to get

$$
\begin{bmatrix}
\tilde{y}_1^{(i)} \\
\vdots \\
\tilde{y}_{m_t}^{(i)}
\end{bmatrix} = \sqrt{\text{SNR}} \begin{bmatrix}
x_1^{(i)} \\
\vdots \\
x_{m_t+m_r-m}^{(i)}
\end{bmatrix} + \tilde{z}^{(i)}.
$$

for all $i = (n-1)k+1, \ldots, nk$, where $\tilde{z}^{(i)}$ are independent with i.i.d $\mathcal{CN}(0,1)$ entries. Letting

$$
\begin{bmatrix}
\tilde{y}_{m_t+m_r-m+1}^{(i)} \\
\vdots \\
\tilde{y}_{m_t}^{(i)}
\end{bmatrix} = \sqrt{\text{SNR}} H_{21}^{(i-k)} x^{(i-k)} + \begin{bmatrix}
\tilde{z}_{m_t+m_r-m+1}^{(i)} \\
\vdots \\
\tilde{z}_{m_t}^{(i)}
\end{bmatrix}
$$

be the side information $y_{si}^{(i-k)}$ measures for channel use $i-k$, for all $i = (n-1)k+1, \ldots, nk$, and repeating this procedure for $i = (n-1)k$ till $i = 1$ results in $m_t + m_r - m$ independent streams of measures

$$
\begin{bmatrix}
\tilde{y}_1^{(1)} \\
\vdots \\
\tilde{y}_{m_t+m_r-m}^{(1)}
\end{bmatrix}, \ldots, \begin{bmatrix}
\tilde{y}_1^{(nk)} \\
\vdots \\
\tilde{y}_{m_t+m_r-m}^{(nk)}
\end{bmatrix}.
$$

Thus, having noisy measures of the last $k$ symbols, $y_1^{(n-1)k+1}, \ldots, y_{nk}^{(nk)}$, the receiver can construct $m_t + m_r - m$ SISO channels, each with a signal-to-noise ratio SNR. Assuming the transmitter can convey these measures using a neglectable number of channel uses (with respect to $n$, see Remark 3), the scheme allows approaching the rate $(m_t + m_r - m) \log(1 + \text{SNR})$ with zero outage probability.

**Remark 1** (Delayed feedback). The scheme exploits a noiseless feedback system to communicate a (possibly) outdated information - the channel realization in previous channel uses. Thus, the feedback is not required to be fast, that is, no limitations on the delay time $k$. However, for non-ergodic systems with a short delay time, the feedback may carry information about the current channel realization. Thus, the transmitter can exploit the up-to-date feedback to use more efficient schemes (e.g., water filling). Nevertheless, for systems with a long delay time (e.g., relatively long distance optical fibers), the channel can be regarded as non-ergodic however with an outdated feedback. In these cases our scheme efficiently achieves zero outage probability.

**Remark 2** (Simple decoding). The scheme constructs $m_t + m_r - m$ independent streams of measures, each with signal-to-noise SNR. This allows the decoding stage to be simple, where a SISO channel decoder can be used, removing the need for further MIMO signal processing.

**Remark 3** (Side information measures). Given noisy measures of the last $k$ transmitted signals,

$$
\tilde{y}_{si}^{(i)} = \sqrt{\text{SNR}} H_{21}^{(i)} x^{(i)} + \tilde{z}_{si}^{(i)}, \forall i = nk - (k-1), \ldots, nk,
$$

where $\tilde{z}_{si}^{(i)}$ are independent with i.i.d $\mathcal{CN}(0,1)$ entries, the scheme can construct $m_t + m_r - m$
independent streams of measures. Thus, the transmitter has to convey \( H_{21}(i) \chi_{i}^{(i)} \), for each \( i = nk - (k - 1), \ldots, nk \), such that the receiver can extracted a vector of noisy measures with signal-to-noise ratio that is not smaller than SNR. This is feasible with a finite number of channel uses. For example, the repetition scheme\(^2\) can be used to convey these measures, each with a signal-to-noise ratio that is at least SNR (according to Lemma \( \ref{T1} \) see Section \( \ref{VI} \) Example \( \ref{Ex3} \)). Suppose each \( H_{21}(i) \chi_{i}^{(i)} \) is conveyed to the receiver within \( N_{si} \) channel uses (e.g., for the repetition scheme \( N_{si} = m_{t}(m - m_{r}) \)). By taking large enough \( n \) (with respect to \( N_{si} \)) one can approach the rate \( (m_{t} + m_{r} - m) \log(1 + \text{SNR}) \).

**Remark 4** (The uniqueness of \( H_{21} \)). The scheme can be further improved to support even an higher data rate with outage probability zero. For example, the last \( m - m_{r} \) entries of the transmitted signal at the first \( k \) channel uses can be used to excite information bearing symbols instead of the zeros symbols. Furthermore, as was mentioned above, when \( m_{t} + m_{r} - m > 1 \), \( H_{21}^{(i)} \) is not unique; there are many \( (m - m_{r}) \times m_{t} \) matrices that complete the columns of \( H_{11}^{(i)} \) into orthonormal columns. Thus, the receiver can choose \( H_{21}^{(i)} \) to be the one with the largest number of zeros rows. Now, at time \( i + k \) the transmitter excites \( m_{t} + m_{r} - m \) new symbols and \( H_{21}^{(i)} \chi_{i}^{(i)} \), a retransmission of \( \chi_{i}^{(i)} \), the transmitted signal at time \( i \). With an appropriate choice of \( H_{21}^{(i)}, H_{21}^{(i)} \chi_{i}^{(i)} \) contains entries that are zero. Instead, these entries can contain additional new information bearing symbols. An open question is how to further enhance the data rate. One would like to exploit the feedback to approach the empirical capacity for any realization of \( H_{11} \).

Note that this rate is achievable with an up-to-date feedback. Further approaching this rate with an outdated feedback system (and with zero outage probability) is left for future research.

**VI. DIVERSITY MULTIPLEXING TRADEOFF**

In this section we want to analyze the tradeoff between diversity and multiplexing in the Jacobi channel. We first examine the error probability of two simple examples - an uncoded transmission in an \( 1 \times m_{r} \) system and a repetition scheme in an \( m_{t} \times m_{r} \) system. Note that the latter can be viewed as a generalization of the first.

In this section we use \( \doteq \) to denote exponential equality, i.e., \( f(\text{SNR}) \doteq \text{SNR}^{d} \) denotes

\[
\lim_{\text{SNR} \to \infty} \frac{\log f(\text{SNR})}{\log \text{SNR}} = d.
\]

**Example 2** \( (m_{t} = 1) \). Consider a transmission of an uncoded signal through a single mode. When all available modes are being addressed at the receiver, that is \( m_{r} = m \), the entire signal power is extracted and the channel corresponds to a SISO unfading channel with an exponentially decaying (with SNR) error probability. However, for \( m_{r} < m \) some power is lost in the unaddressed modes resulting an higher error probability. Suppose the transmitter excites an uncoded QPSK signal (Similar results can be obtained for higher constellations). Using the sphere bound we can upper bound the error probability for a given channel realization:

\[
\Pr(\text{error} \mid H_{11} = H_{11}) \leq \Pr(\|z\|^{2} \geq \frac{\text{SNR}}{2}\|H_{11}\|_{F}^{2}),
\]

where \( z \sim \mathcal{CN}(0, 1) \) and \( \|\cdot\|_{F} \) is the Frobenius norm of a matrix: \( \|A\|_{F}^{2} \doteq \sum_{ij} \|A_{ij}\|^{2} = \sum_{i} \lambda_{i} \).

\( ^{2}\)The repetition scheme can convey the \( m - m_{r} \) entries of \( H_{21}^{(i)} \chi_{i}^{(i)} \) with \( N_{si} = m_{t}(m - m_{r}) \) channel uses. In each channel use a single entry is transmitted through a single mode (while all other modes are zero) in a way that all entries are transmitted through all modes.
where $\lambda_i$ are the singular values of $A$. Since for high SNR this bound is tight, we can write
\[
\Pr(\text{error} \mid H_{11} = H_{11}) \doteq \exp\left(-\frac{\text{SNR}}{2} \|H_{11}\|^2_F\right),
\]
where we further applied the cdf. of a chi-squared random variable with 2 degrees of freedom. Thus, by letting $\lambda$ be the square singular value of $H_{11}$, we can write:
\[
\Pr(\text{error} \mid \|H_{11}\|^2_F = \lambda) \doteq \exp\left(-\frac{\text{SNR}}{2} \lambda\right).
\]
By taking the expectation over (31) w.r.t $\lambda$ we get the error probability
\[
P_e(\text{SNR}) \doteq \mathbb{E}[e^{-\frac{\text{SNR}}{2} \lambda}].
\]
Now, for $m_r = m$ we always have $\lambda = 1$, thus the error probability satisfies
\[
P_e(\text{SNR}) \doteq e^{-\frac{\text{SNR}}{2}}.
\]
For $m_r < m$, we can use (4), the pdf of the eigenvalue of a Jacobi matrix $J(m_r, m - m_r, 1)$, to calculate the right hand side of (32):
\[
\mathbb{E}[e^{-\frac{\text{SNR}}{2} \lambda}] = \int_0^1 \Pr(\lambda) e^{-\frac{\text{SNR}}{2} \lambda} d\lambda
\]
\[
= K_{1,m_r,m}^{-1} \sum_{i=0}^{m-m_r-1} (-1)^i \binom{m-m_r-1}{i} \int_0^1 \lambda^{m_r+i-1}(1-\lambda)^{m-m_r-1} e^{-\frac{\text{SNR}}{2} \lambda} d\lambda.
\]
We use the Taylor expansion of $(1-x)^a$ to have
\[
\mathbb{E}[e^{-\frac{\text{SNR}}{2} \lambda}] = K_{1,m_r,m}^{-1} \sum_{i=0}^{m-m_r-1} (-1)^i \binom{m-m_r-1}{i} \left[ (m_r+i-1)!(\frac{\text{SNR}}{2})^{m_r+i-1} - e^{-\frac{\text{SNR}}{2} \lambda} \sum_{j=0}^{m_r+i-1} \binom{m_r+i-1}{j} j!(\frac{\text{SNR}}{2})^{-j-1} \right].
\]
In high SNR (37) is dominated by the term $K_{1,m_r,m}^{-1} (m_r - 1)! (\text{SNR}/2)^{-m_r}$ and since the bound is tight at this regime we can write
\[
P_e(\text{SNR}) \doteq \begin{cases} 
\text{SNR}^{-m_r} & , m_r \neq m \\
\exp\left(-\frac{\text{SNR}}{2}\right) & , m_r = m.
\end{cases}
\]
Thus, the number of received modes dictates the decaying order of the error probability at high SNR. Having in mind that the $m_r \times 1$ channel matrix can be viewed as a sub-vector of $m \times 1$ vector that was constructed by normalizing $m$ i.i.d. complex Gaussian rv’s, it is not surprising that the error probability in the analogue Rayleigh channel has a similar behavior at high SNR. But is this true also for $m_t \neq 1$? To that end we want to examine the error probability of the repetition scheme in an $m_r \times m_t$ system.

Example 3 (Repetition scheme). Suppose the transmitter excites the following ($m_t$ entries)
signals in each $m_t$ consecutive channel uses:

\[
\begin{bmatrix}
  x \\
  0 \\
  0 \\
  \vdots \\
  0 \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  x \\
  0 \\
  \vdots \\
  0 \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  0 \\
  \vdots \\
  \vdots \\
  x \\
\end{bmatrix}
\]

where $x$ is an uncoded QPSK symbol (Similar results can be obtained also for higher constellations). Let us assume that the channel is constant within the $m_t$ channel uses and, for simplicity of notations, we further assume $m_r \geq m_t$ (similar results can be obtained for $m_t > m_r$). Thus, with the same considerations as before, the error probability satisfies

\[
P_e(SNR) = \mathbb{E}[\exp \left( -\frac{SNR}{2} \sum_{i=1}^{m_t} \lambda_i \right) ],
\]

where the expectation is over $\lambda_1 \leq \cdots \leq \lambda_{m_t}$, the ordered non-zero eigenvalues of $H_{11}^\dagger H_{11}$. Now, using the joint pdf of the ordered eigenvalues of a Jacobi matrix $J(m_r, m - m_r, m_t)$ we can analyze \((39)\) for $m_t + m_r \leq m$:

\[
\mathbb{E}[\exp \left( -\frac{SNR}{2} \sum_{i=1}^{m_t} \lambda_i \right) ] = \int \cdots \int Pr(\lambda_1, \cdots, \lambda_{m_t}) e^{-\frac{SNR}{2} \sum_{i=1}^{m_t} \lambda_i} \prod_{i=1}^{m_t} d\lambda_i
\]

\[
= \frac{K_{m_t,m_r,m}}{m_t!} \int_0^1 \cdots \int_0^1 \prod_{i=1}^{m_t} \lambda_i^{m_r-m_t}(1-\lambda_i)^{m-m_r-m_t} e^{-\frac{SNR}{2} \sum_{i<j} \lambda_i(\lambda_j - \lambda_i)^2} \prod_{i=1}^{m_t} d\lambda_i,
\]

where we used the joint pdf of the non-ordered eigenvalues. Note that the term

\[
\prod_{1 \leq i < j \leq m_t} (\lambda_j - \lambda_i)
\]

is the determinant of the Vandermonde matrix

\[
\begin{bmatrix}
  1 & \cdots & 1 \\
  \lambda_1 & \cdots & \lambda_{m_t} \\
  \vdots & \vdots & \vdots \\
  \lambda_1^{m_t-1} & \cdots & \lambda_{m_t}^{m_t-1}
\end{bmatrix}.
\]

Thus we can write

\[
\prod_{1 \leq i < j \leq m_t} (\lambda_j - \lambda_i)^2 = \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} \prod_{i=1}^{m_t} \lambda_i^{\sigma_1(i) + \sigma_2(i) - 2},
\]

where $S_{m_t}$ is the set of permutations of $1, \cdots, m_t$ and $\text{sgn}(\sigma)$ denotes the signature of the permutation $\sigma$. Applying \((42)\) into \((41)\) results

\[
\mathbb{E}[\exp \left( -\frac{SNR}{2} \sum_{i=1}^{m_t} \lambda_i \right) ] = \frac{K_{m_t,m_r,m}}{m_t!} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} \prod_{i=1}^{m_t} \int_0^1 \lambda_i^{m_r-m_t+\sigma_1(i) + \sigma_2(i) - 2}(1-\lambda_i)^{m-(m_r+m_t)} e^{-\frac{SNR}{2} \lambda_i} d\lambda_i.
\]
Note that the error probability of the repetition scheme is symmetric in the scheme rate is different).

For the first case, that is, $m_t = m_r = m$ systems. Since there is a per-mode power constraint (rather than total power constraint), the error probability is symmetric in $m_t$ and $m_r$.

With the same techniques used before (to get equation (37)) we get that the error probability at high SNR is dominated by the term

$$
\frac{K^{-1}}{m_t m_r} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{\text{sgn}(\sigma_1) + \text{sgn}(\sigma_2)} \prod_{i=1}^{m_t} (m_r - m_t + \sigma_1(i) + \sigma_2(i) - 2)!(\frac{\text{SNR}}{2})^{-(m_r - m_t + \sigma_1(i) + \sigma_2(i) - 1)}.
$$

(44)

Thus, for $m_t + m_r \leq m$, the error probability satisfies

$$
P_e(\text{SNR}) \approx C_{m_t, m_r, m} \text{SNR}^{-m_t} \sum_{i=1}^{m} (m_r - m_t + 2i - 1),
$$

(45)

and conclude that the error probability of the repetition scheme satisfies

$$
P_e(\text{SNR}) \approx e^{-\frac{\text{SNR}(m_t + m_r - m)}{2}} \cdot \mathbb{E}[\exp \left(-\frac{m - m_r}{2} \sum_{i=1}^{m} \hat{\lambda}_i \right)],
$$

where $\hat{\lambda}_1 \leq \cdots \leq \hat{\lambda}_{m-m_r}$ are the non-zero ordered eigenvalues of $H_{22}^+ H_{22}$. Since $H_{22}$ applies to the first case, that is, $(m - m_t) + (m - m_r) < m$, we can use (46) and conclude that the error probability of the repetition scheme satisfies

$$
P_e(\text{SNR}) \approx \begin{cases} 
\frac{\text{SNR}^{-m_t}}{e^{\frac{\text{SNR}(m_t + m_r - m)}{2}}} \cdot \text{SNR}^{-m_t} & , m_t + m_r \leq m \\
\text{SNR}^{-m_r} & , m_t + m_r > m.
\end{cases}
$$

(47)

Fig. 4 depict the error probability vs. SNR for $m = 4$, for different $m_t \times m_r$ systems. Indeed, the curves turn exponentially decaying for systems that apply to the second case, $m_t + m_r > m$. Note that the error probability of the repetition scheme is symmetric in $m_t$ and $m_r$ (however, the scheme rate is different).

Equation (47) implies that, for $m_t + m_r \leq m$, the exponent of the dominant term in the error probability is $m_r \cdot m_t$. Thus the performance gain of an $m_t \times m_r$ system compared to a system
with a single transmit and receive mode is dictated by the SNR exponent of the error probability. This SNR exponent is termed the diversity gain. Intuitively, the total transmitted power is spread over all $m$ available modes, thus addressing only some modes results in a power loss. As the number of addressed modes at the receiver is higher and as the transmitter excites more modes, the probability for a substantial power loss is smaller. Analogously, in wireless systems, as the signal passes through more (independent) paths, the probability for a fading is smaller. However, in the Jacobi channel it turns out that there is a transition threshold in which enough modes are being addressed to ensure a certain received power. This results an exponentially decaying error probability for certain rates.

Now, in Section III we have analyzed the ergodic capacity and showed that as more modes are being addressed, the capacity increases. Thus, increasing the number of addressed modes has another potential gain - higher data rate. This gain is termed spatial multiplexing gain. In MIMO systems there is a fundamental tradeoff between the diversity and multiplexing gains. The optimal tradeoff for the Rayleigh channel was presented in [20]. We now turn to analyze this tradeoff in the Jacobi channel. To that end, we formalize the concepts of diversity gain and multiplexing gain by quoting some definitions from [20]:

**Definition 4.** Let a scheme be a family of codes $\{C(SNR)\}$ of block length $l$, one at each SNR level. Let $R(SNR)$ (bits/symbols) be the rate of the code $C(SNR)$. A scheme $\{C(SNR)\}$ is said to achieve spatial multiplexing gain $r$ and diversity gain $d$ if the data rate satisfies

$$\lim_{SNR \to \infty} \frac{R(SNR)}{\log SNR} = r$$

and the average error probability satisfies

$$\lim_{SNR \to \infty} \frac{P_e(SNR)}{\log SNR} = -d.$$  

For each $r$, define $d^*(r)$ to be the supremum of the diversity advantage achieved over all schemes.

For example, in the uncoded repetition scheme (for the case of $m_t + m_r \leq m$), the diversity gain is $m_r \cdot m_t$ when transmitting a signal from a fixed constellation. E.g., for QPSK modulation the data rate is fixed, $R(SNR) = 1/m_t$ (bps/Hz) for any SNR. Thus, for diversity gain of $m_r \cdot m_t$ the scheme achieves a multiplexing gain of 0. By increasing the constellation size with SNR to achieve an higher multiplexing gain, i.e., to support a data rate of $R(SNR) = r \log SNR$ (bps/Hz) (for some $0 < r < 1/m_t$), the minimum distance between the constellation points decreases with SNR. This results in an error probability with a smaller decaying order, that is, a lower diversity gain. See further discussion in Example 4.

We next find the optimal tradeoff, $d^*(r)$, for the Jacobi channel.

### A. The case of $m_t + m_r \leq m$

**Theorem 5.** Let the block length satisfy $l \geq m_t + m_r - 1$\(^4\). The optimal diversity-multiplexing tradeoff curve $d^*(r)$ for $m_t + m_r \leq m$, is given by the piecewise linear function that connects

\(^3\)Note that in [20] SNR is defined to be the average signal-to-noise ratio at each receive antenna. In this work, SNR as defined by (1) equals the average signal-to-noise ratio at each receive antenna when all transmit modes are excited, that is, when $m_t = m$. However, the two definitions are equivalent up to a constant, thus Definition 4 stays valid.

\(^4\)for $l < m_t + m_r - 1$, bounds on $d^*(r)$ can be obtained in the exact same manner as in [20] using results from Appendix B.
Fig. 5. Optimal DMT curve, $d^*(r)$, for $m_t = m_r = 4$ and $l \geq 7$, for different numbers of supported modes $m$. 

the points $(k, d^*(k))$ for $k = 0, 1, \cdots, \min\{m_t, m_r\}$, where

$$d^*(k) = (m_t - k)(m_r - k).$$

(48)

Proof: See Appendix B.
Hence, for \( m_t + m_r \leq m \), the optimal tradeoff curve does not depend on \( m \) and is equivalent to the optimal curve in the analogue Rayleigh channel. Since \( \mathbf{H} \)'s columns/rows are orthogonal, the extent of independence of \( \mathbf{H}_{11} \)'s elements is determined by \( m \) - as \( m \) is larger, the columns/rows are more independent. At high SNR and for \( m_t + m_r \leq m \), the impact of the orthogonality of the channel columns/rows on the error probability is negligible and indeed the decaying order of the error probability behaves as in the Rayleigh channel, where the columns are i.i.d vectors.

**B. The case of \( m_t + m_r > m \)**

According to Theorem 4 we can write

\[
P_{\text{out}}(m_t, m_r, m; r \log(1 + \text{SNR})) = P_{\text{out}}(m - m_r, m - m_t, m; (r - (m_t + m_r - m)) \log(1 + \text{SNR}))
\]

for \( r \geq m_t + m_r - m \). Thus, for rates above \( (m_t + m_r - m) \log(1 + \text{SNR}) \), the optimal diversity-multiplexing tradeoff can be found from Theorem 5. Theorem 4 further states that the outage probability for rates below \( (m_t + m_r - m) \log(1 + \text{SNR}) \) is strictly zero. Hence, for multiplexing gains below \( m_t + m_r - m \) there is a scheme that can convey unfading signals to the receiver, thereby achieving an exponentially decaying error probability. In this case the discussion about diversity is no longer relevant. Nonetheless, one can think of the gain as infinite. This reveals an interesting difference between the Jacobi and Rayleigh channels - the maximum diversity gain is “unbounded” vs. \( m_r \cdot m_t \).

The following Theorem states above.

**Theorem 6.** The optimal diversity-multiplexing tradeoff curve \( d^*(r) \), for \( m_t + m_r > m \), is given by

\[
d^*(r) = \begin{cases} 
  d_{\text{risidual}}^*(r - (m_t + m_r - m)) & , r \geq m_t + m_r - m \\
  \infty & , r < m_t + m_r - m 
\end{cases}
\]

where \( d_{\text{risidual}}^*(r) \) is the optimal curve of an \( (m - m_r) \times (m - m_t) \) system. For a block length \( l \geq m_t + m_r - 1 \), the optimal curve \( d_{\text{risidual}}^*(r) \) is the piecewise linear function that connects the points \( (k, d_{\text{risidual}}^*(k)) \) for \( k = 0, 1, \cdots, \min\{m - m_r, m - m_t\} \) where

\[
d_{\text{risidual}}^*(k) = (m - m_r - k)(m - m_t - k) .
\]

**Proof:** At high SNR, in terms of minimal outage probability, we can take the covariance matrix of the transmitted signal to be \( Q = I_{m_t} \), see Appendix B. Thus Theorem 4 can be applied: for \( r < m_t + m_r - m \) the minimal outage probability is zero hence the error probability turns exponentially decaying with SNR; for \( r \geq m_t + m_r - m \) the outage probability equals the outage probability for \( \tilde{r} = r - (m_t + m_r - m) \) in a system with \( m - m_r \) transmit and \( m - m_t \) receive modes. Noting that at high SNR the error probability is dominated by the outage probability (see Appendix B) completes the proof.

Fig. 5 depict the optimal DMT curve for \( 4 \times 4 \) system with code block length \( l \geq 7 \) for different numbers of supported modes \( m \).

In the following example we try to illuminate the concept of infinite diversity gain.

**Example 4** (\( m_t = m_r = 2 \)). We consider the \( 2 \times 2 \) Alamouti scheme [21]. Assuming a code block of length \( l \geq 3 \) and rate \( R = r \log \text{SNR} \) (bps/Hz), the transmitter excites in each two
Fig. 6. Comparison between Alamouti and the repetition scheme: $m_t = m_r = 2$, various values of $m$ and $l \geq 3$. consecutive channel uses two information bearing symbols in the following manner:

$$
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}, \begin{bmatrix}
x_2^\dagger \\
x_1^\dagger
\end{bmatrix}.
$$

ML decoding linearly combines the received measures and yields the following equivalent scalar channels:

$$
y_i = \sqrt{\|H_{11}\|^2 \cdot SNR} \cdot x_i + z_i, \quad \forall \ i = 1, 2
$$

where each $z_i$ is distributed $CN(0,1)$ independent of $x_i$ and $H_{11}$. The probability for an outage event is given by

$$
P_{out}(2, 2, m; R) = Pr \left( \log(1 + \|H_{11}\|^2 \cdot SNR) < r \log SNR \right)
$$

$$
\geq Pr \left( \|H_{11}\|^2 < SNR^{-r} \right). \quad (54)
$$

Now, for the Rayleigh fading channel $\|H_{11}\|^2$ is chi-square distributed with $2m_t m_r$ degrees of freedom. In this case, as was shown in [20], the $2 \times 2$ Alamouti scheme can achieve maximum diversity gain of 4. However, in the Jacobi channel:

- for $m = 2$ we have $\|H_{11}\|^2 = 2$ ($H_{11} = H$ unitary).
- for $m = 3$ we have $\|H_{11}\|^2 \geq 1$ (by Lemma 1 and the fact that the Frobenius norm equals the sum of the eigenvalues of $H_{11}^\dagger H_{11}$).
\* for \( m \geq 4 \) there is always a non-zero probability for an outage event.

Therefore, for \( m = 2 \) and \( m = 3 \), for any \( r \leq 1 \), we get equivalent unfading scalar channels with strictly zero outage probability and one can think of the maximum diversity gain as infinite.

For \( m \geq 4 \) it can be shown that the maximum diversity gain is 4 and the optimal tradeoff curve linearly connects the points \((1,0)\) and \((0,4)\). See Fig. 6 where the tradeoff curve of the repetition scheme is further plotted. One can note that for \( m = 3 \) the Alamouti scheme achieves the optimal DMT for \( r = 1 \).

VII. Relation To The Rayleigh Channel

The Jacobi MIMO channel is defined by the transfer matrix \( H_{11} \), a truncated \( m_r \times m_t \) version of an Haar distributed \( m \times m \) unitary matrix. The statistics of the singular values of the channel follow the law of the Jacobi ensemble \( \mathcal{J}(m_r, m - m_r, m_t) \), for \( m_r \geq m_t \) and \( m - m_r \geq m_t \). We shall now examine the case where \( m_t \) is very large with respect to \( m_r \) and \( m_r \).

Let \( G_1 \) and \( G_2 \) be \( m_r \times m_t \) and \( (m - m_r) \times m_t \) independent Gaussian matrices, each with i.i.d. \( \mathcal{CN}(0,1) \) entries. For \( m_r \geq m_t \) the Jacobi ensemble \( \mathcal{J}(m_r, m - m_r, m_t) \) can be constructed from the Gaussian ensemble as

\[
G_1^\dagger G_1 (G_1^\dagger G_1 + G_2^\dagger G_2)^{-1}.
\]  

Thus, the squared singular values of \( H_{11} \) share the same distribution with the eigenvalues of \((55)\). Intuitively, in terms of the singularity statistics, the Jacobi channel can be viewed as an \( m_r \times m_t \) sub-channel of an \( m \times m_t \) normalized Gaussian channel. Furthermore, for \( m \gg m_r \) we have

\[
G_1^\dagger G_1 (G_1^\dagger G_1 + G_2^\dagger G_2)^{-1} \approx G_1^\dagger G_1 (G_2^\dagger G_2)^{-1}
\]

\[
\approx G_1^\dagger G_1 (\mathbb{E} [g \cdot g^\dagger])^{-1}
\]

\[
= \frac{1}{m-m_r} G_1^\dagger G_1,
\]

where in the first approximation we applied \( m \gg m_r \) and in the second the Low of Large Numbers \((g)\) is a vector of \( m_t \) independent complex Gaussian rv’s \( \mathcal{CN}(0,1) \).

For \( m_t \gg m_r \) the squared singular values of \( H_{11} \) follow the law of the Jacobi ensemble \( \mathcal{J}(m_t, m - m_t, m_r) \). This ensemble can be constructed as

\[
G_1^\dagger (G_1^\dagger G_2^\dagger + G_2^\dagger G_2)^{-1}
\]

where \( G_1 \) and \( G_2 \) are \( m_r \times m_t \) and \( m_r \times (m - m_t) \) Gaussian matrices. For \( m \gg m_t \) we have

\[
G_1^\dagger (G_1^\dagger + G_2^\dagger)^{-1} \approx \frac{1}{m-m_r} G_1^\dagger
\]

Thus, we can conclude that up to a normalizing factor the Jacobi channel approaches to the Rayleigh channel for \( m \gg m_t, m_r \).

In Fig. 7 we compare the ergodic capacities and outage probabilities in the Rayleigh and Jacobi channels for \( m_t = m_r = 2 \). One can note that as \( m \) is larger, the Jacobi channel resemble the Rayleigh channel. We note that to have a proper comparison we need to impose an equal signal-to-noise ratio at the receiver in each channel scenario. SNR as defined by \((1)\) is the average signal-to-noise ratio at each receive antenna when all transmit modes are excited, that is, when \( m_t = m \). One can verify that the average signal-to-noise ratio at each receive antenna for any
$m_t \leq m$, denoted $\overline{\text{SNR}}$, is
\[
\overline{\text{SNR}} = \text{SNR} \cdot \frac{\mathbb{E}\|H_{11}\|_F^2}{m_r},
\]
where $\|H_{11}\|_F$ is the Frobenius norm of $H_{11}$. For the Jacobi channel,
\[
\mathbb{E}\|H_{11}\|_F^2 = \mathbb{E} \sum_{i=1}^{\min\{m_t, m_r\}} \lambda_i,
\]
can be computed by using the marginal distribution $P_{\lambda_i}(\lambda)$ which is given in Appendix A. We note that for $m \gg m_t, m_r$, by Equations (58) and (60), we also have
\[
\mathbb{E}\|H_{11}\|_F^2 \approx \frac{m_t \cdot m_r}{m - m_t}.
\]
For the Rayleigh channel, where the transfer matrix is Gaussian (with unit variance), one can verify that the average signal-to-noise ratio at each receive antenna is
\[
\overline{\text{SNR}} = \text{SNR} \cdot m_t.
\]

VIII. Conclusions

The Jacobi channel is defined by three parameters: $m_t$, $m_r$, and $m$. For fixed $m_t$ and $m_r$, the parameter $m$ defines the "independence measure" of the channel. Since the columns and rows of $H$ are orthonormal, $m$ determines the extent of which $H_{11}$’s elements are independent. For example, for $m = \max\{m_t, m_r\}$ the columns/rows of $H_{11}$ are orthonormal and the matrix elements are highly dependent. As $m$ is larger, the elements of $H_{11}$ are more independent. Indeed, for large $m$, the Rayleigh (i.i.d. elements) and Jacobi channels are equivalent in terms of the singular values statistics (up to a normalizing factor). The size of the unitary matrix, $m$, can be viewed as the number of orthogonal propagation paths in space whereas $m_t$ and $m_r$ are the number of addressed paths at the transmitter and receiver. Thus, the Jacobi channel model
introduces a new concept in MIMO channel modeling: defining the independence scale of the propagation paths. For example, when \( m_t = m_r \), the parameter \( m \) scales the channel from a unitary channel (\( m = m_t = m_r \)) up to the Rayleigh channel (\( m \gg m_t, m_r \)).

IX. ACKNOWLEDGEMENTS

We gratefully acknowledge Amir Dembo for the proof of Lemma [1].

APPENDIX A

PROOF OF THEOREM [2]

By Theorem [1] the ergodic capacity satisfies

\[
C(m_t, m_r, m; \text{SNR}) = \mathbb{E}[\log \det(I_{m_t} + \text{SNR} \cdot H_{11}^\dagger H_{11})]
\]

(64)

\[
= \mathbb{E}\left[\sum_{i=1}^{m_t} \log(1 + \text{SNR} \cdot \lambda_i)\right]
\]

(65)

where we denote by \( \lambda = \{\lambda_1, \ldots, \lambda_{m_t}\} \) the non-zeros eigenvalues of \( H_{11}^\dagger H_{11} \). Note that for simplicity of notations we assume \( m_r \geq m_t \) (all results hold true also for \( m_t > m_r \) by switching \( m_t \) and \( m_r \)). Thus, we can write the ergodic capacity as the expectation over only one of the unordered eigenvalues:

\[
C(m_t, m_r, m; \text{SNR}) = \mathbb{E}[\log(1 + \text{SNR} \cdot \lambda_1)].
\]

(66)

Now, the joint pdf of the ordered eigenvalues, \( f_{\lambda}(\lambda_1, \ldots, \lambda_{m_t}) \), is given by (4). The joint pdf of the unordered eigenvalues equals

\[
\frac{1}{m_t!} f_{\lambda}(\lambda_1, \ldots, \lambda_{m_t}),
\]

thus we can compute the density of \( \lambda_1 \) by integrating out \( \lambda_2, \ldots, \lambda_{m_t} \):

\[
f_{\lambda_1}(\lambda_1) = \int_0^1 \cdots \int_0^1 \frac{1}{m_t!} f_{\lambda}(\lambda_1, \ldots, \lambda_{m_t}) \prod_{i=2}^{m_t} d\lambda_i.
\]

(67)

By taking

\[
\lambda_i = \frac{1}{2}(1 - \tilde{\lambda}_i)
\]

(68)

we can write

\[
f_{\tilde{\lambda}_1}(\tilde{\lambda}_1) = \int_{-1}^1 \cdots \int_{-1}^1 f_{\tilde{\lambda}}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{m_t}) \prod_{i=2}^{m_t} d\tilde{\lambda}_i,
\]

(69)

where

\[
f_{\tilde{\lambda}}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{m_t}) = \tilde{K}^{-1}_{m_t, m_r, m} \prod_{i=1}^{m_t} (1 - \tilde{\lambda}_i)^{\alpha} (1 + \tilde{\lambda}_i)^{\beta} \prod_{i<j} (\tilde{\lambda}_i - \tilde{\lambda}_j)^2,
\]

(70)

and we denote \( \alpha = m_r - m_t \) and \( \beta = m - m_r - m_t \). Now, the term

\[
\prod_{1 \leq i < j \leq m_t} (\tilde{\lambda}_i - \tilde{\lambda}_j)
\]
is the determinant of the Vandermonde matrix

$$\begin{bmatrix}
1 & \cdots & 1 \\
\lambda_1 & \cdots & \lambda_{m_t} \\
\vdots & \ddots & \vdots \\
\lambda_{m_t-1} & \cdots & \lambda_{m_t-1}
\end{bmatrix} \quad \text{(71)}$$

With row operations we can transform (71) into the following matrix

$$\begin{bmatrix}
P_0^{(\alpha,\beta)}(\lambda_1) & \cdots & P_0^{(\alpha,\beta)}(\lambda_{m_t}) \\
\vdots & \ddots & \vdots \\
P_{m_t-1}^{(\alpha,\beta)}(\lambda_1) & \cdots & P_{m_t-1}^{(\alpha,\beta)}(\lambda_{m_t})
\end{bmatrix} \quad \text{(72)}$$

where $P_n^{(\alpha,\beta)}(x)$ are the Jacobi polynomials \[19\] 8.96 which form a complete orthogonal system in the interval $[-1, 1]$ with respect to the weighting function $w(x) = (1 - x)^\alpha (1 + x)^\beta$:

$$\int_{-1}^{1} w(x) P_n^{(\alpha,\beta)}(x) P_k^{(\alpha,\beta)}(x) dx = a_{k,\alpha,\beta} \cdot \delta_{kn} , \quad \text{(73)}$$

where for integers $\alpha$ and $\beta$

$$a_{k,\alpha,\beta} = \frac{2^{\alpha+\beta+1}}{2k + \alpha + \beta + 1} \binom{2k + \alpha + \beta}{k} \binom{2k + \alpha + \beta}{k + \alpha}^{-1} . \quad \text{(74)}$$

By the definition of determinant we have

$$\prod_{1 \leq i < j \leq m_t} (\tilde{\lambda}_i - \tilde{\lambda}_j) = C_{m_t, m_t, m} \sum_{\sigma \in S_m} (-1)^{sgn(\sigma)} \prod_{i=1}^{m_t} P_{\sigma(i)-1}^{(\alpha,\beta)}(\tilde{\lambda}_i) , \quad \text{(75)}$$

where $S_m$ is the set of permutations of $1, \ldots, m_t$, $sgn(\sigma)$ denotes the signature of the permutation $\sigma$ and $C_{m_t, m_t, m}$ is a constant picked up from transformation of the Vandermonde matrix (71) into (72). By applying (75) into (70) we get:

$$f_{\tilde{\lambda}}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{m_t}) = \tilde{C}_{m_t, m_t, m}^{-1} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{sgn(\sigma_1) + sgn(\sigma_2)} \prod_{i=1}^{m_t} (1 - \tilde{\lambda}_i)^\alpha (1 + \tilde{\lambda}_i)^\beta P_{\sigma_1(i)-1}^{(\alpha,\beta)}(\tilde{\lambda}_i) P_{\sigma_2(i)-1}^{(\alpha,\beta)}(\tilde{\lambda}_i) . \quad \text{(76)}$$

Further integrating over $\tilde{\lambda}_2, \ldots, \tilde{\lambda}_{m_t}$ results

$$f_{\tilde{\lambda}_1}(\tilde{\lambda}_1) = \tilde{C}_{m_t, m_t, m}^{-1} \sum_{\sigma_1, \sigma_2 \in S_{m_t}} (-1)^{sgn(\sigma_1) + sgn(\sigma_2)} (1 - \tilde{\lambda}_1)^\alpha (1 + \tilde{\lambda}_1)^\beta P_{\sigma_1(1)-1}^{(\alpha,\beta)}(\tilde{\lambda}_1) P_{\sigma_2(1)-1}^{(\alpha,\beta)}(\tilde{\lambda}_1) \prod_{i=2}^{m_t} a_{(\sigma_1(i)-1),\alpha,\beta} \cdot \delta_{\sigma_1(i),\sigma_2(i)} \quad \text{(77)}$$

$$= \tilde{C}_{m_t, m_t, m}^{-1} (m_t - 1)! \sum_{k=0}^{m_t-1} (1 - \tilde{\lambda}_1)^\alpha (1 + \tilde{\lambda}_1)^\beta P_k^{(\alpha,\beta)}(\tilde{\lambda}_1)^2 \prod_{i \neq k} a_{i,\alpha,\beta} \quad \text{(78)}$$

$$= \frac{1}{m_t} \sum_{k=0}^{m_t-1} \tilde{a}_{k,\alpha,\beta} P_k^{(\alpha,\beta)}(\tilde{\lambda}_1)^2 (1 - \tilde{\lambda}_1)^\alpha (1 + \tilde{\lambda}_1)^\beta , \quad \text{(79)}$$

where the first equality follows from (73) and thus implies that $\sigma_1(i) = \sigma_2(i)$ for all $i$. This results in the second equation while the third equality follows from (73) and the fact that $f_{\tilde{\lambda}_1}(\tilde{\lambda}_1)$
must integrates to unity. 

Turning back to $\lambda$ and by applying the joint pdf of the ordered eigenvalues of

$$Q \lambda \quad \text{scale of interest, we can take}$$

for $i$ and a lower bound by taking $Q$ is the set that describes the outage event. Note that we assumed $P$ for high SNR we can take $Q = I_{m_t}$ since

$$P_{out}(m_t, m_r, m; R) = \inf_{Q: \lambda \geq 0, \lambda_i \leq 1} Pr \left[ \log \det(\text{SNR} \cdot H_{11} Q H_{11}^\dagger) < R \right] ,$$

where $Q$ is the covariance matrix of the transmitted signal (with per-mode power constraint). However, for high SNR we can take $Q = I_{m_t}$ since

$$P_{out}(m_t, m_r, m; R) = \inf_{Q: \lambda \geq 0, \lambda_i \leq 1} Pr \left[ \log \det(\text{SNR} \cdot H_{11} Q H_{11}^\dagger) < R \right] .$$

This is shown in [20] - an upper bound on the outage probability is derived by picking $Q = I_{m_t}$ and a lower bound by taking $Q = m_t I_{m_t}$. These bounds are exponentially equal hence, in the scale of interest, we can take $Q = I_{m_t}$.

Now, let the transmission rate be $R = r \log(1 + \text{SNR})$ (bits/symbols). Since

$$\log \det(I_{m_t} + \text{SNR} \cdot H_{11} H_{11}^\dagger) = \log \det(I_{m_t} + \text{SNR} \cdot H_{11}^\dagger H_{11})$$

and by applying the joint pdf of the ordered eigenvalues of $H_{11}^\dagger H_{11}$ we can write

$$P_{out}(m_t, m_r, m; r \log(1 + \text{SNR})) \doteq K_{m_t, m_r, m}^{-1} \prod_{i=1}^{m_t} \lambda_i^{m_r - m_t} (1 - \lambda_i)^{m_t - m_r} \prod_{i < j} (\lambda_i - \lambda_j)^2 d\lambda ,$$

where $K_{m_t, m_r, m}$ is a normalizing factor and

$$B = \{ \lambda : 0 \leq \lambda_1 \leq \ldots \leq \lambda_{m_t} \leq 1, \prod_{i=1}^{m_t} (1 + \text{SNR} \cdot \lambda_i) < (1 + \text{SNR})^r \}$$

is the set that describes the outage event. Note that we assumed $m_t \leq m_r$ (without loss of generality, since the outage probability is symmetric in $m_t$ and $m_r$). Letting

$$\lambda_i = \text{SNR}^{-\alpha_i}$$

for $i = 1, \ldots, m_t$ allows us to write

$$P_{out}(m_t, m_r, m; r \log(1 + \text{SNR})) \doteq \log(\text{SNR})^{m_t} K_{m_t, m_r, m}^{-1} \prod_{i=1}^{m_t} \text{SNR}^{-\alpha_i (m_r - m_t + 1)} \cdot (1 - \text{SNR}^{-\alpha_i})^{m_t - m_r} \prod_{i < j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 d\alpha .$$

(87)
Since \(1 + \text{SNR}^{1-\alpha_i} \equiv \text{SNR}^{(1-\alpha_i)^+}\), where \((x)^+ = \max\{0, x\}\), we can describe the set of outage events by

\[
\mathcal{B} = \{\alpha : \alpha_1 \geq \ldots \geq \alpha_m \geq 0, \sum_{i=1}^{m_t} (1 - \alpha_i)^+ < r\}.
\]

Since the term \(\log(\text{SNR})^{m_t K_{m_t,m_r,m}^{-1}}\) has no effect on the SNR exponent, i.e., satisfies

\[
\lim_{\text{SNR} \to \infty} \frac{\log(\log(\text{SNR})^{m_t K_{m_t,m_r,m}^{-1}})}{\log \text{SNR}} = 0,
\]

we get

\[
P_{out}(m_t, m_r, m; r \log(1 + \text{SNR})) \equiv \int_{\mathcal{B}} \prod_{i=1}^{m_t} \text{SNR}^{-\alpha_i(m_r-m_t+1)} \cdot (1 - \text{SNR}^{-\alpha_i})^{m_r-m_t} \prod_{i<j}(\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^{2} \, d\alpha.
\]

Now, we note that

\[
P_{out}(m_t, m_r, m; r \log(1 + \text{SNR})) \leq \int_{\mathcal{B}} \prod_{i=1}^{m_t} \text{SNR}^{-\alpha_i(m_r-m_t+1)} \prod_{i<j}(\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^{2} \, d\alpha.
\]

In [20, Theorem 4] it was proven that the right hand side of above satisfies:

\[
\int_{\mathcal{B}} \prod_{i=1}^{m_t} \text{SNR}^{-\alpha_i(m_r-m_t+1)} \prod_{i<j}(\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^{2} \, d\alpha = \text{SNR}^{-f(\alpha^*)},
\]

where

\[
f(\alpha) = \sum_{i=1}^{m_t} (2i - 1 + m_r - m_t)\alpha_i
\]

and

\[
\alpha^* = \arg \inf_{\alpha \in \mathcal{B}} f(\alpha).
\]

By defining \(S_\delta = \{\alpha : \alpha_i > \delta \, \forall \, i = 1, \ldots, m_t\}\) for any \(\delta > 0\), we can write

\[
P_{out}(m_t, m_r, m; r \log(1 + \text{SNR})) \geq \int_{\mathcal{B} \cap S_\delta} \prod_{i=1}^{m_t} \text{SNR}^{-\alpha_i(m_r-m_t+1)} \cdot (1 - \text{SNR}^{-\alpha_i})^{m_r-m_t} \prod_{i<j}(\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^{2} \, d\alpha
\]

\[
\geq (1 - \text{SNR}^{-\delta})^{m_r(m_r-m_t)} \int_{\mathcal{B} \cap S_\delta} \prod_{i=1}^{m_t} \text{SNR}^{-\alpha_i(m_r-m_t+1)} \prod_{i<j}(\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^{2} \, d\alpha
\]

\[
\equiv \text{SNR}^{-f(\alpha^*_\delta)},
\]
where

\[ \alpha^*_\delta = \arg \inf_{\alpha \in B \cap S_\delta} f(\alpha). \]  

(100)

Using the continuity of \( f \), \( \alpha^*_\delta \) approaches \( \alpha^* \) as \( \delta \) goes to zero and we can conclude that

\[ P_{out}(m_t, m_r, m; r \log(1 + \text{SNR})) \approx \text{SNR}^{-f(\alpha^*)}. \]  

(101)

This result was obtained in [20] for the outage probability in Rayleigh channel. From here one can continue as was presented in [20], showing that the error probability is dominated by the outage probability at high SNR ( [20, Theorem 2]) for \( l \geq m_t + m_r - 1 \) (these proofs rely on (101) without making any assumptions on the channel statistics, therefore are true also for the Jacobi channel).

REFERENCES