Channel Capacity of an Asymmetric Constellation in Rayleigh Fading With Noncoherent Energy Detection

Ranjan K. Mallik, *Fellow, IEEE*, and Ross Murch, *Fellow, IEEE*

***Abstract*— The channel capacity of noncoherent reception of multi-level one-sided amplitude-shift keying (ASK), which is an asymmetric constellation, in Rayleigh fading with receive diversity and energy detection is considered. The asymmetries result in capacity achieving input probability distributions, that is, a priori probability distributions, that deviate from uniformity. An analytical expression for the mutual information in terms of a single integral is derived, and from it the set of equations, which can be solved to obtain the optimum or capacity achieving input probabilities, is obtained. High and low signal-to-noise ratio (SNR) approximations of the optimum input probabilities and the capacity are derived next. Furthermore, a logarithmic upper bound on the mutual information is obtained. Numerical results confirm that the uniform distribution of input probabilities is not capacity achieving. For example, with average SNR per symbol per branch of 6 dB, the relative deviation of the mutual information (with uniform input distribution) from the capacity is nearly 20% for 4-level ASK with one transmit diversity branch and two receive diversity branches. Furthermore, the derived high and low SNR approximations to the capacity are shown to be reasonably accurate.**

***Index Terms*— Amplitude-shift keying (ASK), asymmetric constellation, channel capacity, energy detection, input probability, mutual information, noncoherent, Rayleigh fading, receive diversity.**

I. Introduction

RECENT interest in noncoherent communication has focused on its application to the internet of things (IoT) [1] and future wireless communication systems [2]. The interest is due to its potential for low cost, low power, and low latency systems that are straightforward to implement [1]–[5]. It is particularly useful in IoT applications when integrated energy detection and low complexity receivers are essential [6]. In addition, more recently, there has been an upsurge in interest related to massive multiple-input multiple-output (MIMO) noncoherent communication as it is considered a potentially useful approach when there is fast fading [2], [7].

Noncoherent systems do not make use of channel state information (CSI) at the receiver and there are two distinct approaches in dealing with this issue. The first approach assumes that there is correlation between adjacent received symbols so that the phase and magnitude differences between symbols is known. This has led to differential detection [8], [9] and maximum likelihood sequence estimation (MLSE) techniques [10]. The advantage of this approach over coherent systems is that it provides a reduction in training overhead without significant performance loss in capacity and is one reason it is considered an option for future wireless communication in some scenarios [2], [3], [7].

The second approach in noncoherent systems is to use straightforward energy detection with the result that all knowledge of phase is lost between symbols [11]. It has also been referred to as "one-shot" noncoherent system where the receiver decodes information at the end of each symbol time [12]. To perform energy detection, envelope detection is often utilized and this can significantly reduce the complexity of the receiver architecture. This approach is particularly suitable for IoT systems that employ energy harvesting where an integrated energy harvester and receiver can be developed based around an envelope detector. An example is simultaneous wireless information and power transfer (SWIPT), where energy harvesters and information decoders can be integrated together to achieve a better rate-energy tradeoff than would be achieved with a coherent receiver [13], [14].

The capacity of noncoherent channels has been studied. For low signal-to-noise ratio (SNR) noncoherent channels it has been shown that the asymptotic minimum bit energy per channel use is the same as that for coherent systems [15]. Furthermore, results for the capacity of block fading noncoherent systems have been reported. It has been found that gains in capacity are not possible if the number of antennas is greater than the coherence time of the channels [16]. Under this condition, it has also been shown that capacity is linearly proportional to the product of the number of receive antennas and SNR in low SNR regimes, and noncoherent systems can

Manuscript received August 24, 2020; revised December 19, 2020 and March 6, 2021; accepted May 18, 2021. Date of publication June 2, 2021; date of current version November 11, 2021. The work of Ranjan K. Mallik was supported in part by the Science and Engineering Research Board, a Statutory Body of the Department of Science and Technology, Government of India, under the J. C. Bose Fellowship. The work of Ross Murch was supported in part by the Hong Kong Research Council collaborative research fund C6012-20G. The associate editor coordinating the review of this article and approving it for publication was M. C. Gursoy. *(Corresponding author: Ranjan K. Mallik.)*

Ranjan K. Mallik is with the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: rkmallik@ee.iitd.ernet.in).

Ross Murch is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: eermurch@ust.hk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3083497.

Digital Object Identifier 10.1109/TWC.2021.3083497

have near coherent performance [17]. Various other important results for capacity with channel memory have also been developed [18].

Few results for capacity are available for noncoherent systems when energy detection is utilized and block fading is not assumed [12]. One reason for this is that without any CSI, it is impossible to separate out the MIMO channels [19]–[21]. Therefore, previous work has focused on energy detection MIMO when partial channel knowledge is available [22], for which fundamental limits on system performance including capacity and diversity have been obtained [23], [24]. These reveal that capacity and diversity order are approximately halved compared to coherent MIMO systems. For single-input multiple-output (SIMO) systems using energy detection without CSI, various results including scaling laws are available [12], [25]. It has been shown that for channel magnitude based energy detection SIMO systems, the diversity performance is the same as that for full coherent systems [26].

Noncoherent systems with energy detection (one-shot noncoherent detection) lead to the use of asymmetric constellations such as one-sided amplitude-shift keying (ASK) since phase information is not available. Multi-level one-sided ASK can be conveniently applied to noncoherent SIMO systems and it has been shown that minimization of symbol error probability (SEP) in SIMO Rayleigh fading channels can be performed by encoding the ASK amplitudes with a near geometric progression [11]. Furthermore, earlier work on noncoherent detectors considering energy detection with single-input single-output [27]–[30] and SIMO [11] in additive white Gaussian noise (AWGN) channels [27], [28], lognormal multipath fading channels [29], and Rayleigh channels with receive diversity [11] are also available. The technique for obtaining the optimal constellation signal amplitude levels has also been considered in [11], [28], [29]. Results for frequency-shift keying (FSK) are also available [31], where continuous phase FSK has been utilized and comparisons between coherent and noncoherent detection in AWGN are given. Extensions to MIMO are also possible but they cannot provide multiplexing gain due to the lack of phase information [21].

In this paper, we consider one-shot noncoherent reception of multi-level one-sided ASK, which is an *asymmetric* constellation, in Rayleigh fading with receive diversity and energy detection. The asymmetries result in channel capacity achieving input probability distributions, that is, a priori probability distributions, that deviate from uniformity. An analytical expression for the mutual information in terms of a single integral is derived, and from it the set of equations, which can be solved to obtain the optimum or capacity achieving input probabilities, is obtained. High and low SNR approximations of the optimum input probabilities and the capacity are derived next. Although not supported by a rigorous mathematical analysis, the quality of the approximations are validated by means of numerical examples. Furthermore, a logarithmic upper bound on the mutual information is obtained. Numerical results showing the variation of the input probabilities and the capacity with respect to the system parameters and the accuracy of the approximations are presented. For example, with average SNR per symbol per branch of 6 dB, the relative deviation of the mutual information (with uniform input distribution) from the capacity is nearly 20% for 4-level ASK with one transmit diversity branch and two receive diversity branches.

The paper is organized as follows. In Section II, an analytical expression for the mutual information in terms of a single integral is derived and the set of equations whose solution gives the capacity achieving input probabilities is obtained. High and low SNR approximations of the capacity achieving input probabilities and the capacity are derived in Sections III and IV, respectively. Section V presents a logarithmic upper bound on the mutual information. Numerical results are presented in Section VI. Section VII gives some concluding remarks.

II. Mutual Information and Channel Capacity

Consider a digital communication system in flat Rayleigh fading with N receive diversity branches and noncoherent reception. The modulation scheme used is multi-level one-sided ASK with L levels, and symbol-by-symbol detection is performed. The $N\times 1$ complex baseband sampled signal vector received at the N diversity branches in a symbol interval is expressed as [11]

$$\mathbf{r} = \mathbf{h}s + \mathbf{n}, \tag{1}$$

where s is the information-bearing ASK symbol belonging to the L-ary constellation $\mathcal{S}$, $\mathbf{h}$ the random complex fading gain vector, and $\mathbf{n}$ the AWGN vector, such that

$$\mathcal{S} = \left\{\sqrt{E_1}, \ldots, \sqrt{E}_L\right\}, \tag{2a}$$

$$E_1 \geq 0, \;\; E_{i-1} < E_i, \; i = 2, \ldots, L, \tag{2b}$$

$$s \in \mathcal{S}, \;\; \mathbf{h} \sim \mathcal{CN}\left(\mathbf{0}_N, \sigma_{\mathrm{h}}^2 \mathbf{I}_N\right), \;\; \mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}_N, \sigma_{\mathrm{n}}^2 \mathbf{I}_N\right), \tag{2c}$$

with $\mathbf{0}_N$ denoting the $N \times 1$ vector of zeros and $\mathbf{I}_N$ the $N \times N$ identity matrix. Thus $\mathbf{h}$, which is independent of $\mathbf{n}$, is a zero-mean complex circular Gaussian random vector with covariance matrix $\sigma_{\mathrm{h}}^2 \mathbf{I}_N$, implying independent and identically distributed (i.i.d.) Rayleigh fading, while $\mathbf{n}$ is a zero-mean complex circular Gaussian random vector with covariance matrix $\sigma_{\mathrm{n}}^2 \mathbf{I}_N$. Due to noncoherent reception, the CSI $\mathbf{h}$ is not known at the receiver.

The conditional probability density function (p.d.f.) of $\mathbf{r}$, conditioned on s, is given by

$$f(\mathbf{r}|s) = \frac{\exp\left\{-\dfrac{\mathbf{r}^H \mathbf{r}}{(|s|^2\sigma_{\mathrm{h}}^2 + \sigma_{\mathrm{n}}^2)}\right\}}{\pi^N \left(|s|^2\sigma_{\mathrm{h}}^2 + \sigma_{\mathrm{n}}^2\right)^N}, \tag{3}$$

where $(\cdot)^H$ denotes the Hermitian (conjugate transpose) operator. The energy detector makes its decision on the basis of $\mathbf{r}^H\mathbf{r}$. Note that in the absence of CSI, $\mathbf{r}^H\mathbf{r}$ is a sufficient statistic for the detection of s. Let s_i denote the ith symbol, given by

$$s_i = \sqrt{E_i}, \quad i = 1, \ldots, L. \tag{4}$$

If $p(s_i) = p_i$ denotes the a priori probability of symbol s_i such that

$$\sum_{i=1}^{L} p_i = 1, \tag{5}$$

then the mutual information $\mathcal{I}(s;\mathbf{r})$ between the transmitted ASK symbol s and the received signal vector $\mathbf{r}$ is expressed as

$$\mathcal{I}(s;\mathbf{r}) = -\int\left(\sum_{s\in\mathcal{S}} p(s)f(\mathbf{r}|s)\right)\log_2\left(\sum_{s'\in\mathcal{S}} p(s')f(\mathbf{r}|s')\right)d\mathbf{r} - N\sum_{s\in\mathcal{S}} p(s)\log_2\left(\pi e\left(|s|^2\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)\right),$$

that is

$$\mathcal{I}(s;\mathbf{r}) = -\int\left(\sum_{i=1}^{L} p_i f(\mathbf{r}|s_i)\right)\log_2\left(\sum_{j=1}^{L} p_j f(\mathbf{r}|s_j)\right)d\mathbf{r} - N\sum_{i=1}^{L} p_i\log_2\left(\pi e\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)\right). \quad (6)$$

Note that the term $N\sum_{i=1}^{L} p_i\log_2\left(\pi e\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)\right)$ in (6) is the conditional entropy of $\mathbf{r}|s$. The channel capacity is the maximum of the mutual information between the transmitted ASK symbol s and the received signal vector $\mathbf{r}$ over the input probability distribution, that is, the a priori probability distribution, $\{p_1,\ldots,p_L\}$.

Substituting (3) in (6) we get

$$\begin{aligned}
\mathcal{I}(s;\mathbf{r}) &= -N\log_2 e - N\sum_{i=1}^{L} p_i\log_2\left(\pi\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)\right) \\
&\quad -\int\sum_{i=1}^{L}\frac{p_i\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N} \\
&\quad\times\log_2\left(\sum_{j=1}^{L}\frac{p_j\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}\right)d\mathbf{r} \\
&= -N\log_2 e - \sum_{i=1}^{L} p_i\log_2\left(\pi^N\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N\right) \\
&\quad\times\int\frac{\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}d\mathbf{r} \\
&\quad -\sum_{i=1}^{L} p_i\int\frac{\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N} \\
&\quad\times\log_2\left(\sum_{j=1}^{L}\frac{p_j\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}\right)d\mathbf{r} \\
&= -N\log_2 e \\
&\quad -\sum_{i=1}^{L} p_i\int\frac{\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N} \\
&\quad\times\left[\begin{array}{l}\log_2\left(\displaystyle\sum_{j=1}^{L}\frac{p_j\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}\right) \\ +\log_2\left(\pi^N\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N\right)\end{array}\right]d\mathbf{r} \\
&= -N\log_2 e \\
&\quad -\sum_{i=1}^{L} p_i\int\frac{\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\pi^N\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N} \\
&\quad\times\log_2\left(\sum_{j=1}^{L}\frac{p_j\exp\left\{-\dfrac{\mathbf{r}^H\mathbf{r}}{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\right\}}{\left[\dfrac{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}\right]}\right)d\mathbf{r}. \quad (7)
\end{aligned}$$

Putting $\mathbf{r} = \left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^{1/2}\mathbf{v}$ for each i in (7), we obtain

$$\begin{aligned}
\mathcal{I}(s;\mathbf{r}) &= -N\log_2 e \\
&\quad -\sum_{i=1}^{L} p_i\int\frac{\exp\left\{-\mathbf{v}^H\mathbf{v}\right\}}{\pi^N} \\
&\quad\times\log_2\left(\sum_{j=1}^{L}\frac{p_j\exp\left\{-\dfrac{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\mathbf{v}^H\mathbf{v}\right\}}{\left[\dfrac{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}\right]}\right)d\mathbf{v}. \quad (8)
\end{aligned}$$

Viewing $\mathbf{v} = [V_1,\cdots,V_N]^T$, where $(\cdot)^T$ denotes the transpose operator, as a $\mathcal{CN}(\mathbf{0}_N,\mathbf{I}_N)$ random vector, (8) can be rewritten as

$$\begin{aligned}
\mathcal{I}(s;\mathbf{r}) &= -N\log_2 e \\
&\quad -\sum_{i=1}^{L} p_i\mathbf{E}_{\mathbf{v}}\left[\log_2\left(\sum_{j=1}^{L}\frac{p_j\exp\left\{-\dfrac{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)}\mathbf{v}^H\mathbf{v}\right\}}{\left[\dfrac{\left(E_j\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}{\left(E_i\sigma_{\mathrm{h}}^2+\sigma_{\mathrm{n}}^2\right)^N}\right]}\right)\right], \quad (9)
\end{aligned}$$

where $\mathbf{E}_{\mathbf{v}}[\cdot]$ denotes expectation over the statistics of $\mathbf{v}$.

Denote the *average SNR per branch of the ith symbol* as Γ_i, which is given by

$$\Gamma_i = \frac{E_i\sigma_{\mathrm{h}}^2}{\sigma_{\mathrm{n}}^2},\quad i=1,\ldots,L. \quad (10)$$

From (2b), we get

$$\Gamma_1 \geq 0,\ \ \Gamma_{i-1} < \Gamma_i,\ i=2,\ldots,L. \quad (11)$$

Put $X_i = |V_i|^2$, $i=1,\ldots,N$, where $X_1,\ldots,X_N$ are N i.i.d. exponentially distributed random variables (r.v.s), each

with mean 1. The mutual information expression (9) can be simplified in terms of $\Gamma_1, \ldots, \Gamma_L$ as

$$\begin{aligned}\mathcal{I}(s;\mathbf{r}) = &- N\log_2 e \\ &- \sum_{i=1}^{L} p_i \mathbf{E}_{X_1,\ldots,X_N} \\ &\left[\log_2\left(\sum_{j=1}^{L} \frac{p_j \exp\left\{-\dfrac{(\Gamma_i+1)}{(\Gamma_j+1)}\displaystyle\sum_{k=1}^{N} X_k\right\}}{\left[\dfrac{(\Gamma_j+1)^N}{(\Gamma_i+1)^N}\right]}\right)\right]. \quad (12)\end{aligned}$$

This can be rewritten in the form of an expectation over the statistics of r.v. X, which is the sum of $X_1, \ldots, X_N$, as

$$\begin{aligned}&\mathcal{I}(s;\mathbf{r}) \\ &= - N\log_2 e \\ &\quad - \sum_{i=1}^{L} p_i \mathbf{E}_X\left[\log_2\left(\sum_{j=1}^{L} \frac{p_j \exp\left\{-\dfrac{(\Gamma_i+1)}{(\Gamma_j+1)}X\right\}}{\left[\dfrac{(\Gamma_j+1)^N}{(\Gamma_i+1)^N}\right]}\right)\right]. \quad (13)\end{aligned}$$

It is clear that X has a gamma distribution with mean N and shape parameter N, implying that (13) can be expressed as

$$\begin{aligned}\mathcal{I}(s;\mathbf{r}) = &- \frac{N}{\ln 2} \\ &- \sum_{i=1}^{L} p_i \int_0^{\infty} \log_2\left(\sum_{j=1}^{L} \frac{p_j \exp\left\{-\dfrac{(\Gamma_i+1)}{(\Gamma_j+1)}x\right\}}{\left[\dfrac{(\Gamma_j+1)^N}{(\Gamma_i+1)^N}\right]}\right) \\ &\times \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx, \quad (14)\end{aligned}$$

which is an *analytical expression for the mutual information in terms of a single integral*. The expression (14) can be rewritten as

$$\begin{aligned}\mathcal{I}(s;\mathbf{r}) = &- \frac{N}{\ln 2} \\ &- \frac{1}{\ln 2}\sum_{i=1}^{L} p_i \int_0^{\infty} \ln\left(\sum_{j=1}^{L} p_j R_{j,i}^{N} \exp\{-xR_{j,i}\}\right) \\ &\times \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx, \quad (15)\end{aligned}$$

where the quantity $R_{j,i}$ is defined as

$$R_{j,i} \triangleq \frac{(\Gamma_i+1)}{(\Gamma_j+1)}, \quad i = 1,\ldots,L, \;\; j = 1,\ldots,L. \quad (16)$$

It is clear from (16) and (11) that

$$\begin{aligned}R_{j,j} &= 1, \quad j = 1,\ldots,L, \\ R_{j,j+1} &> 1, \quad j = 1,\ldots,L-1, \\ R_{j,i} &= R_{i,j}^{-1} > 1, \;\; j = 1,\ldots,i-1, \;\; i = 2,\ldots,L. \quad (17)\end{aligned}$$

To obtain the capacity, we need to maximize the mutual information $\mathcal{I}(s;\mathbf{r})$ over the input probability distribution $\{p_1, \ldots, p_L\}$ subject to the constraints

$$\sum_{i=1}^{L} p_i = 1, \quad 0 \le p_i \le 1, \;\; i = 1,\ldots,L. \quad (18)$$

We construct the Lagrangian function $\mathcal{L}$ as

$$\begin{aligned}\mathcal{L} &= \left(\frac{\ln 2}{N}\right)\mathcal{I}(s;\mathbf{r}) - \lambda\left(\sum_{i=1}^{L} p_i - 1\right) + \sum_{i=1}^{L} v_i p_i \\ &= -1 - \frac{1}{N}\sum_{i=1}^{L} p_i \int_0^{\infty} \ln\left(\sum_{j=1}^{L} p_j R_{j,i}^{N} \exp\{-xR_{j,i}\}\right) \\ &\quad \times \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\ &\quad - \left(\sum_{i=1}^{L} p_i(\lambda - v_i) - \lambda\right), \quad (19)\end{aligned}$$

where λ is a Lagrangian multiplier and $v_1, \ldots, v_L$ are nonnegative slack variables which are chosen such that $p_1, \ldots, p_L \ge 0$ and $v_\ell p_\ell = 0$ for $\ell = 1, \ldots, L$. We obtain the capacity achieving $p_1, \ldots, p_L$ from (19) by solving the set of equations

$$\left(\frac{\ln 2}{N}\right)\frac{\partial \mathcal{I}(s;\mathbf{r})}{\partial p_\ell} = \lambda - v_\ell, \quad \ell = 1,\ldots,L, \quad (20a)$$

$$\sum_{i=1}^{L} p_i = 1, \quad (20b)$$

$$v_\ell p_\ell = 0, \quad \ell = 1,\ldots,L, \quad (20c)$$

where $v_1, \ldots, v_L \ge 0$ and $p_1, \ldots, p_L \ge 0$.

By following the derivation presented in Appendix A, we obtain from (20) the set of equations

$$\begin{aligned}&-\frac{1}{N}\int_0^{\infty} \ln\left(\sum_{j=1}^{L} p_j R_{j,\ell}^{N} \exp\{-xR_{j,\ell}\}\right)\frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\ &\quad = \lambda - v_\ell + \frac{1}{N}, \quad \ell = 1,\ldots,L, \quad (21a)\end{aligned}$$

$$\sum_{i=1}^{L} p_i = 1, \quad (21b)$$

$$v_\ell p_\ell = 0, \quad \ell = 1,\ldots,L, \quad (21c)$$

where $v_1, \ldots, v_L \ge 0$ and $p_1, \ldots, p_L \ge 0$.

By solving the set of equations given by (21), we obtain the optimum or capacity achieving $p_1, \ldots, p_L, \lambda, v_1, \ldots, v_L$, which we denote as $p_{1,\mathrm{opt}}, \ldots, p_{L,\mathrm{opt}}, \lambda_{\mathrm{opt}}, v_{1,\mathrm{opt}}, \ldots, v_{L,\mathrm{opt}}$, respectively. Note that (21) can be conveniently solved using a mathematical package such as MATLAB, without any restriction on the $R_{j,i}$. By applying (21) to (15), we obtain the capacity $\mathcal{C}$ (in *bits per channel use*) as

$$\mathcal{C} = \frac{(N\lambda_{\mathrm{opt}} + 1 - N)}{\ln 2}. \quad (22)$$

III. Approximation for High SNR

Consider the case of high SNR with

$$\frac{(\Gamma_{j+1}+1)}{(\Gamma_j+1)} = R_{j,j+1} = R_{j+1,j}^{-1} \gg 1, \quad j = 1, \ldots, L-1. \tag{23}$$

The logarithm in (15) can be expressed as

$$\begin{aligned}
&\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right) \\
&= \ln\left(\sum_{j=1}^{L} \exp\{\ln p_j + N\ln R_{j,i} - xR_{j,i}\}\right) \\
&= \ln p_i - x \\
&\quad + \ln\left(1 + \sum_{\substack{j=1 \\ j\neq i}}^{L} \exp\left\{\ln\frac{p_j}{p_i} + N\ln R_{j,i} - x(R_{j,i}-1)\right\}\right) \\
&\approx \ln p_i - x \\
&\quad + \sum_{\substack{j=1 \\ j\neq i}}^{L} \exp\left\{\ln\frac{p_j}{p_i} + N\ln R_{j,i} - x(R_{j,i}-1)\right\}
\end{aligned} \tag{24}$$

under the necessary condition

$$\ln\frac{p_j}{p_i} + N\ln R_{j,i} - x(R_{j,i}-1) < 0, \quad j \neq i, \;\; j,i = 1,\ldots,L. \tag{25}$$

The condition (25) implies

$$x > \frac{\left(N\ln R_{j,i} + \ln\dfrac{p_j}{p_i}\right)}{(R_{j,i}-1)}, \quad R_{j,i} > 1, \;\; j = 1,\ldots,i-1, \;\; i = 2,\ldots,L, \tag{26a}$$

$$x < \frac{\left(N\ln R_{i,j} + \ln\dfrac{p_i}{p_j}\right)}{\left(1-R_{i,j}^{-1}\right)}, \quad R_{i,j} > 1, \;\; j = i+1,\ldots,L, \;\; i = 1,\ldots,L-1. \tag{26b}$$

Note that

$$\min_{\substack{j \\ j<i}} R_{j,i} = R_{i-1,i}, \quad i = 2,\ldots,L, \tag{27a}$$

$$\min_{\substack{j \\ j>i}} R_{i,j} = R_{i,i+1}, \quad i = 1,\ldots,L-1. \tag{27b}$$

Retaining only the significant terms of the summation over j in (24) under the conditions (23), (25), and (27), we get

$$\begin{aligned}
&\ln\left(\sum_{j=1}^{L} p_j R_{j,1}^N \exp\{-xR_{j,1}\}\right) \\
&\approx \ln p_1 - x \\
&\quad + \exp\left\{\ln\frac{p_2}{p_1} - N\ln R_{1,2} + x\left(1-R_{1,2}^{-1}\right)\right\}, \quad i = 1,
\end{aligned} \tag{28a}$$

$$\begin{aligned}
&\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right) \\
&\approx \ln p_i - x \\
&\quad + \exp\left\{\ln\frac{p_{i-1}}{p_i} + N\ln R_{i-1,i} - x(R_{i-1,i}-1)\right\} \\
&\quad + \exp\left\{\ln\frac{p_{i+1}}{p_i} - N\ln R_{i,i+1} + x\left(1-R_{i,i+1}^{-1}\right)\right\}, \\
&\quad i = 2,\ldots,L-1,
\end{aligned} \tag{28b}$$

$$\begin{aligned}
&\ln\left(\sum_{j=1}^{L} p_j R_{j,L}^N \exp\{-xR_{j,L}\}\right) \\
&\approx \ln p_L - x \\
&\quad + \exp\left\{\ln\frac{p_{L-1}}{p_L} + N\ln R_{L-1,L} - x(R_{L-1,L}-1)\right\}, \\
&\quad i = L.
\end{aligned} \tag{28c}$$

Taking the limit as $R_{j,j+1}$ goes to infinity for $j = 1,\ldots,L-1$, we obtain the limiting mutual information under the infinite SNR condition

$$\frac{(\Gamma_{j+1}+1)}{(\Gamma_j+1)} = R_{j,j+1} = R_{j+1,j}^{-1} \to \infty, \quad j = 1,\ldots,L-1. \tag{29}$$

Applying (29) to (28) we get the approximation

$$\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right) \approx \ln p_i - x, \quad i = 1,\ldots,L;$$

substituting this in (15) and noting that

$$\int_0^\infty \frac{x^N \exp\{-x\}}{(N-1)!}dx = N,$$

we obtain the limiting mutual information under the infinite SNR condition (29) as

$$\mathcal{I}_{\mathrm{SNR}\to\infty}(s;\mathbf{r}) = -\frac{N}{\ln 2} - \frac{1}{\ln 2}\sum_{i=1}^{L} p_i \ln p_i + \frac{N}{\ln 2}\sum_{i=1}^{L} p_i,$$

that is,

$$\mathcal{I}_{\mathrm{SNR}\to\infty}(s;\mathbf{r}) = -\sum_{i=1}^{L} p_i \log_2 p_i, \tag{30}$$

which is maximized by the *uniform input probability distribution* given by

$$p_i = \frac{1}{L}, \quad i = 1,\ldots,L. \tag{31}$$

Note that under the infinite SNR condition, the mutual information is the same as the source entropy, since the signal is noise free at the output. Substituting (31) in (30), we find that the limiting capacity (in bits per channel use) under the infinite SNR condition (29) is

$$\mathcal{C}_{\mathrm{SNR}\to\infty} = \log_2 L. \tag{32}$$

It is clear from the infinite SNR behavior of the mutual information that the capacity achieving $p_1,\ldots,p_L$ under the high SNR condition (23) will not deviate largely from the

uniform distribution. As a result, we can rewrite (26) approximately as

$$x > \frac{N \ln R_{j,i}}{(R_{j,i}-1)}, \quad R_{j,i} > 1, \;\; j = 1, \ldots, i-1, \;\; i = 2, \ldots, L, \tag{33a}$$

$$x < \frac{N \ln R_{i,j}}{\left(1 - R_{i,j}^{-1}\right)}, \quad R_{i,j} > 1, \;\; j = i+1, \ldots, L, \;\; i = 1, \ldots, L-1. \tag{33b}$$

Application of (27) to (33) gives

$$x > \frac{N \ln R_{i-1,i}}{(R_{i-1,i}-1)}, \quad i = 2, \ldots, L, \tag{34a}$$

$$x < \frac{N \ln R_{i,i+1}}{\left(1 - R_{i,i+1}^{-1}\right)}, \quad i = 1, \ldots, L-1. \tag{34b}$$

Let the quantity β_i be given by

$$\beta_i = \frac{N \ln R_{i,i+1}}{(R_{i,i+1}-1)}, \quad i = 1, \ldots, L-1. \tag{35}$$

The conditions (34a) and (34b) can be expressed in terms of β_i as

$$x > \beta_{i-1}, \quad i = 2, \ldots, L, \tag{36a}$$

$$x < R_{i,i+1}\beta_i, \quad i = 1, \ldots, L-1. \tag{36b}$$

Substituting (28) in (15) and applying (36), the mutual information under the high SNR condition (23) approximates as

$$\begin{aligned}
&\mathcal{I}(s;\mathbf{r}) \\
&\approx -\frac{1}{\ln 2}\sum_{i=1}^{L} p_i \ln p_i \\
&\quad - \frac{1}{\ln 2}\sum_{i=2}^{L} p_{i-1} \int_{\beta_{i-1}}^{\infty} \frac{R_{i-1,i}^{N} x^{N-1} \exp\{-xR_{i-1,i}\}}{(N-1)!}dx \\
&\quad - \frac{1}{\ln 2}\sum_{i=1}^{L-1} p_{i+1} \int_{0}^{R_{i,i+1}\beta_i} \frac{R_{i,i+1}^{-N} x^{N-1} \exp\left\{-xR_{i,i+1}^{-1}\right\}}{(N-1)!}dx \\
&= -\frac{1}{\ln 2}\sum_{i=1}^{L} p_i \ln p_i \\
&\quad - \frac{1}{\ln 2}\sum_{i=2}^{L} p_{i-1} \int_{R_{i-1,i}\beta_{i-1}}^{\infty} \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\
&\quad - \frac{1}{\ln 2}\sum_{i=1}^{L-1} p_{i+1} \int_{0}^{\beta_i} \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx.
\end{aligned} \tag{37}$$

Now

$$\begin{aligned}
&\int_{R_{i-1,i}\beta_{i-1}}^{\infty} \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\
&\quad = \exp\{-R_{i-1,i}\beta_{i-1}\}\sum_{k=0}^{N-1}\frac{(R_{i-1,i}\beta_{i-1})^k}{k!}, \\
&\quad R_{i-1,i}\beta_{i-1} \gg 1, \quad i = 2, \ldots, L, \\
&\int_{0}^{\beta_i} \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\
&\quad = 1 - \exp\{-\beta_i\}\sum_{k=0}^{N-1}\frac{\beta_i^k}{k!} \\
&\quad = 1 - \exp\{-\beta_i\}\left[\exp\{\beta_i\} - \sum_{k=N}^{\infty}\frac{\beta_i^k}{k!}\right] \\
&\quad = \exp\{-\beta_i\}\sum_{k=N}^{\infty}\frac{\beta_i^k}{k!}, \\
&\quad \beta_i \ll 1, \quad i = 1, \ldots, L-1,
\end{aligned}$$

which can be approximated using (35) as

$$\begin{aligned}
&\int_{R_{i-1,i}\beta_{i-1}}^{\infty} \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\
&\quad \approx \exp\{-R_{i-1,i}\beta_{i-1}\}\frac{(R_{i-1,i}\beta_{i-1})^{N-1}}{(N-1)!} \\
&\quad = \frac{(N\ln R_{i-1,i})^{N-1}}{(N-1)!\left(1-R_{i-1,i}^{-1}\right)^{N-1} R_{i-1,i}^{\frac{N}{\left(1-R_{i-1,i}^{-1}\right)}}}, \\
&\quad i = 2, \ldots, L,
\end{aligned} \tag{38a}$$

$$\begin{aligned}
&\int_{0}^{\beta_i} \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\
&\quad \approx \frac{\beta_i^N}{N!} \\
&\quad = \frac{(N\ln R_{i,i+1})^N}{N!(R_{i,i+1}-1)^N}, \quad i = 1, \ldots, L-1.
\end{aligned} \tag{38b}$$

Substitution of (38) in (37) results in $\mathcal{I}(s;\mathbf{r}) \approx \mathcal{I}_{\mathrm{hiSNR}}(s;\mathbf{r})$, where

$$\begin{aligned}
&\mathcal{I}_{\mathrm{hiSNR}}(s;\mathbf{r}) \\
&\quad = -\frac{1}{\ln 2}\sum_{i=1}^{L} p_i \ln p_i \\
&\qquad - \frac{1}{\ln 2}\sum_{i=2}^{L} p_{i-1}\frac{(N\ln R_{i-1,i})^{N-1}}{(N-1)!\left(1-R_{i-1,i}^{-1}\right)^{N-1} R_{i-1,i}^{\frac{N}{\left(1-R_{i-1,i}^{-1}\right)}}} \\
&\qquad - \frac{1}{\ln 2}\sum_{i=1}^{L-1} p_{i+1}\frac{(N\ln R_{i,i+1})^N}{N!(R_{i,i+1}-1)^N}.
\end{aligned} \tag{39}$$

By following the derivation presented in Appendix B, we obtain the *optimum or high SNR capacity achieving* $p_1, \ldots, p_L$, which are denoted as

$$p_{1,\mathrm{hiSNR,opt}}, \ldots, p_{L,\mathrm{hiSNR,opt}},$$

respectively, and are given in *closed form* by

$$p_{\ell,\mathrm{hiSNR,opt}} = \frac{\exp\{-\nu_\ell\}}{\sum_{k=1}^{L}\exp\{-\nu_k\}}, \quad \ell = 1, \ldots, L, \tag{40a}$$

where

$$\nu_1 = \frac{(N\ln R_{1,2})^{N-1}}{(N-1)!\left(1-R_{1,2}^{-1}\right)^{N-1} R_{1,2}^{\frac{N}{\left(1-R_{1,2}^{-1}\right)}}}, \tag{40b}$$

$$\nu_\ell = \frac{(N\ln R_{\ell,\ell+1})^{N-1}}{(N-1)!\left(1-R_{\ell,\ell+1}^{-1}\right)^{N-1} R_{\ell,\ell+1}^{\frac{N}{\left(1-R_{\ell,\ell+1}^{-1}\right)}}} + \frac{(N\ln R_{\ell-1,\ell})^N}{N!(R_{\ell-1,\ell}-1)^N}, \quad \ell = 2,\ldots,L-1, \quad (40c)$$

$$\nu_L = \frac{(N\ln R_{L-1,L})^N}{N!(R_{L-1,L}-1)^N}. \quad (40d)$$

Substituting (40) in (39), the *capacity* (in bits per channel use) *under the high SNR condition (23)* is obtained as

$$\mathcal{C}_{\text{hiSNR}} = \log_2\left(\sum_{k=1}^{L}\exp\{-\nu_k\}\right), \quad (41)$$

where $\nu_1,\ldots,\nu_L$ are given by (40b), (40c), and (40d).

Note that under the infinite SNR condition (29), $\nu_1,\ldots,\nu_L \to 0$, and we obtain from (41)

$$\mathcal{C}_{\text{hiSNR}} \to \mathcal{C}_{\text{SNR}\to\infty} = \log_2 L,$$

where $\mathcal{C}_{\text{SNR}\to\infty}$ is given by (32).

IV. Approximation for Low SNR

Consider the case of low SNR with

$$\frac{(\Gamma_{j+1}+1)}{(\Gamma_j+1)} - 1 = R_{j,j+1} - 1 = \left(R_{j+1,j}^{-1} - 1\right) \ll 1, \quad j = 1,\ldots,L-1, \quad (42)$$

which implies

$$|R_{j,i} - 1| \ll 1 \text{ for } j \neq i. \quad (43)$$

Under the low SNR condition (42) (when $R_{j,i}$ is close to 1 for $j \neq i$), the logarithm term in (15) with the constant $-N/\ln 2$ absorbed inside the logarithm, the argument of which is a convex combination of

$$R_{1,i}^N\exp\{-x(R_{1,i}-1)\},\ldots,R_{L,i}^N\exp\{-x(R_{L,i}-1)\},$$

can be expressed as

$$-\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-x(R_{j,i}-1)\}\right) \approx -\mathcal{K}(x)\sum_{j=1}^{L} p_j \ln\left(R_{j,i}^N\exp\{-x(R_{j,i}-1)\}\right), \quad (44)$$

where $\mathcal{K}(x)$ is a scale factor which depends on x. The range of integration over x in (15) is from 0 to ∞, and it is numerically observed that applying (44) to (15) and replacing $\mathcal{K}(x)$ by $1/2$ in the integrand gives a good approximation. As a result, application of (44) to (15) and replacement of $\mathcal{K}(x)$ by $1/2$ results in $\mathcal{I}(s;\mathbf{r}) \approx \mathcal{I}_{\text{loSNR}}(s;\mathbf{r})$, where

$$\mathcal{I}_{\text{loSNR}}(s;\mathbf{r}) = -\frac{N}{2\ln 2} - \frac{1}{2\ln 2}\sum_{i=1}^{L} p_i \sum_{j=1}^{L} p_j \int_0^\infty \ln\left(R_{j,i}^N\exp\{-xR_{j,i}\}\right) \times \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx. \quad (45)$$

The integral in (45) simplifies to

$$\begin{aligned}\int_0^\infty \ln\left(R_{j,i}^N\exp\{-xR_{j,i}\}\right)\frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx &= N\ln(R_{i,j})\int_0^\infty \frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\ &\quad - R_{j,i}\int_0^\infty \frac{x^N\exp\{-x\}}{(N-1)!}dx \\ &= N\ln(R_{j,i}) - NR_{j,i}. \end{aligned} \quad (46)$$

Substituting (46) in (45), we get

$$\begin{aligned}\mathcal{I}_{\text{loSNR}}(s;\mathbf{r}) &= -\frac{N}{2\ln 2} + \frac{N}{2\ln 2}\sum_{i=1}^{L} p_i \sum_{j=1}^{L} p_j \left[R_{j,i} - \ln(R_{j,i})\right] \\ &= -\frac{N}{2\ln 2} + \frac{N}{4\ln 2}\sum_{i=1}^{L}\sum_{j=1}^{L} p_i p_j \left[R_{j,i} + R_{i,j} - \ln(R_{j,i}R_{i,j})\right], \end{aligned}$$

which simplifies to (since $R_{j,i}R_{i,j} = 1$)

$$\mathcal{I}_{\text{loSNR}}(s;\mathbf{r}) = -\frac{N}{2\ln 2} + \frac{N}{4\ln 2}\sum_{i=1}^{L}\sum_{j=1}^{L} p_i p_j \left[R_{j,i} + R_{i,j}\right]; \quad (47)$$

this is a quadratic function of $p_1,\ldots,p_L$. Let $R_{i,j}$ denote the element in the ith row and jth column of the $L\times L$ matrix $\mathbf{R}$, that is,

$$\mathbf{R} = [R_{i,j}]_{i,j=1}^{L}, \quad (48)$$

having ones as its diagonal elements, and let the $L \times 1$ probability vector $\mathbf{p}$ be given by

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_L \end{bmatrix}. \quad (49)$$

Using (48) and (49), we can express (47) as

$$\mathcal{I}_{\text{loSNR}}(s;\mathbf{r}) = -\frac{N}{2\ln 2} + \frac{N}{4\ln 2}\mathbf{p}^T\left[\mathbf{R}+\mathbf{R}^T\right]\mathbf{p}. \quad (50)$$

We conclude from (16) and (17) that $R_{j,i} + R_{i,j}$ increases with decrease of i or increase of j for $i < j$, and increases with increase of i or decrease of j for $i > j$. As a result, the maximum element of $\mathbf{R}+\mathbf{R}^T$ is in its Lth row and 1st column and its 1st row and Lth column, which is $R_{1,L} + R_{1,L}^{-1}$.

If we express $R_{j,j+1}$ as

$$R_{j,j+1} = 1 + \delta_j, \quad 0 < \delta_j \ll 1, \quad j = 1,\ldots,L-1,$$

then, from (47), we get

$$\mathcal{I}_{\text{loSNR}}(s;\mathbf{r}) = \frac{N}{4\ln 2}\sum_{i=1}^{L}\sum_{j=1}^{L} p_i p_j \left[R_{j,i} + R_{i,j} - 2\right]$$

$$= \frac{N}{2\ln 2}\sum_{j=2}^{L}\sum_{i=1}^{j-1} p_i p_j \times \left[\prod_{k=i}^{j-1}(1+\delta_k) + \prod_{k=i}^{j-1}(1+\delta_k)^{-1} - 2\right],$$

which can be simplified as (using second order approximation of $\prod_{k=i}^{j-1}(1+\delta_k)^{-1}$)

$$\mathcal{I}_{\mathrm{loSNR}}(s;\mathbf{r}) = \frac{N}{2\ln 2}\sum_{j=2}^{L}\sum_{i=1}^{j-1} p_i p_j \left(\sum_{k=i}^{j-1}\delta_k\right)^2;$$

from this, it can be shown that the convex combination $\mathbf{p}^T\left[\mathbf{R}+\mathbf{R}^T\right]\mathbf{p}$ of the elements of $\mathbf{R}+\mathbf{R}^T$ is maximized by choosing

$$p_1 = p_L = \frac{1}{2}, \qquad p_\ell = 0, \;\; \ell = 2,\ldots,L-1,$$

which implies that *in the low SNR regime, we should use binary modulation with equiprobable symbols*. The *low SNR capacity achieving* $p_1,\ldots,p_L$, which are denoted as $p_{1,\mathrm{loSNR,opt}},\ldots,p_{L,\mathrm{loSNR,opt}}$, respectively, are therefore given by

$$p_{1,\mathrm{loSNR,opt}} = p_{L,\mathrm{loSNR,opt}} = \frac{1}{2},$$
$$p_{\ell,\mathrm{loSNR,opt}} = 0, \;\; \ell = 2,\ldots,L-1. \tag{51}$$

Substituting (51) in (50), the *capacity* (in bits per channel use) *under the low SNR condition (42)* is obtained as

$$\mathcal{C}_{\mathrm{loSNR}} = \frac{N\left(R_{1,L} + R_{1,L}^{-1} - 2\right)}{8\ln 2}. \tag{52}$$

Note that when $R_{j,i} \to 1$ for $j \neq i$, we obtain from (52) $\mathcal{C}_{\mathrm{loSNR}} \to 0$, as expected.

V. A Logarithmic Upper Bound on the Mutual Information

The integral in (15) can be expressed as

$$-\int_0^\infty \ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right)\frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx = -\mathbf{E}_X\left[\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-XR_{j,i}\}\right)\right], \tag{53}$$

where X is a gamma distributed r.v. with mean N and shape parameter N. Since $-\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right)$ is a concave function of x, we apply Jensen's inequality to (53) over the p.d.f. of X, noting that $\mathbf{E}[X] = N$, and obtain

$$\begin{aligned}
&-\int_0^\infty \ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right)\frac{x^{N-1}\exp\{-x\}}{(N-1)!}dx \\
&\leq -\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-\mathbf{E}[X]R_{j,i}\}\right) \\
&= -\ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-NR_{j,i}\}\right).
\end{aligned} \tag{54}$$

By applying (54) to (15) we obtain $\mathcal{I}(s;\mathbf{r}) \leq \mathcal{I}_{\mathrm{LUB}}(s;\mathbf{r})$, where $\mathcal{I}_{\mathrm{LUB}}(s;\mathbf{r})$ is a *logarithmic upper bound* on the mutual information and is given by

$$\begin{aligned}
&\mathcal{I}_{\mathrm{LUB}}(s;\mathbf{r}) \\
&= -\frac{N}{\ln 2} \\
&\quad -\frac{1}{\ln 2}\sum_{i=1}^{L} p_i \ln\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-NR_{j,i}\}\right).
\end{aligned} \tag{55}$$

If $p_{1,\mathrm{opt}},\ldots,p_{L,\mathrm{opt}}$ denote the optimum or capacity achieving input probabilities, then a logarithmic upper bound $\mathcal{C}_{\mathrm{LUB}}$ on the capacity is obtained by putting $p_i = p_{i,\mathrm{opt}}$ for $i = 1,\ldots,L$ in (55), resulting in

$$\begin{aligned}
&\mathcal{C}_{\mathrm{LUB}} \\
&= -\frac{N}{\ln 2} \\
&\quad -\frac{1}{\ln 2}\sum_{i=1}^{L} p_{i,\mathrm{opt}} \ln\left(\sum_{j=1}^{L} p_{j,\mathrm{opt}} R_{j,i}^N \exp\{-NR_{j,i}\}\right).
\end{aligned} \tag{56}$$

It can be easily shown that for $y > 0$, the function $y^N\exp\{-Ny\}$, where N is a natural number, monotonically increases for $0 < y < 1$, attains a maximum of $\exp\{-N\}$ at $y = 1$, and monotonically decreases for $y > 1$. As a result, in (55), we have

$$\max_{i,j} R_{j,i}^N \exp\{-NR_{j,i}\} = \exp\{-N\},$$

and the maximum occurs when $j = i$. Therefore, in the case of the high SNR condition (23), we get

$$R_{j,i}^N \exp\{-NR_{j,i}\} \ll 1 \;\text{ for } j \neq i. \tag{57}$$

Furthermore, we have

$$\lim_{R_{j,i}\to\infty} R_{j,i}^N \exp\{-NR_{j,i}\} = 0 \;\text{ for } j \neq i. \tag{58}$$

Under the infinite SNR condition (29), we obtain from (55) and (58)

$$\begin{aligned}
&\lim_{R_{j,j+1}\to\infty,\; j=1,\ldots,L-1} \mathcal{I}_{\mathrm{LUB}}(s;\mathbf{r}) \\
&= -\frac{N}{\ln 2} - \frac{1}{\ln 2}\sum_{i=1}^{L} p_i \ln(p_i\exp\{-N\}) \\
&= -\frac{N}{\ln 2} + \frac{N}{\ln 2}\sum_{i=1}^{L} p_i - \frac{1}{\ln 2}\sum_{i=1}^{L} p_i \ln p_i \\
&= -\sum_{i=1}^{L} p_i \log_2 p_i.
\end{aligned} \tag{59}$$

Denoting the limiting logarithmic upper bound on the mutual information under the infinite SNR condition (29), which is given by (59), as $\mathcal{I}_{\mathrm{LUB,SNR}\to\infty}(s;\mathbf{r})$, we get

$$\mathcal{I}_{\mathrm{LUB,SNR}\to\infty}(s;\mathbf{r}) = -\sum_{i=1}^{L} p_i \log_2 p_i, \tag{60}$$

which is the same as the limiting mutual information given by (30). This is maximized by the uniform input probability

distribution given by $p_1 = \cdots = p_L = 1/L$, as in (31). Substituting (31) in (60), we find that the limiting capacity upper bound (in bits per channel use) under the infinite SNR condition (29) is $\mathcal{C}_{\mathrm{LUB,SNR}\to\infty} = \log_2 L$, which is the same as the limiting capacity given by (32).

Note further that when $R_{j,i} \to 1$ for $j \neq i$, we obtain from (56) $\mathcal{C}_{\mathrm{LUB}} \to 0$.

Computation of the bound (55) is easy since it does not require evaluation of integrals. Furthermore, this bound approximates $\mathcal{I}_{\mathrm{hiSNR}}(s;\mathbf{r})$ given by (39) under the high SNR condition (23), and approximates $\mathcal{I}_{\mathrm{loSNR}}(s;\mathbf{r})$ given by (47) under the low SNR condition (42).

VI. Numerical Results

Denote the *average SNR per symbol per branch* as Γ_{av}, which is given by

$$\Gamma_{\mathrm{av}} = \frac{1}{L}\sum_{i=1}^{L}\Gamma_i. \tag{61}$$

Note that Γ_{av} is the *arithmetic mean of* $\Gamma_1,\ldots,\Gamma_L$. The *average consumed signal power to noise power ratio per symbol per branch*, denoted as Γ_{cav}, is given by

$$\Gamma_{\mathrm{cav}} = \sum_{i=1}^{L} p_i\Gamma_i.$$

If the input probability distribution is not uniform, then $\Gamma_{\mathrm{av}} \neq \Gamma_{\mathrm{cav}}$.

We consider L-level one-sided ASK with *SNR common ratio* R_{SNR} ($R_{\mathrm{SNR}} > 1$) given by

$$R_{\mathrm{SNR}} = R_{i,i+1} = \frac{(\Gamma_{i+1}+1)}{(\Gamma_i+1)}, \quad i = 1,\ldots,L-1,$$

and $\Gamma_1 = 0$, which imply

$$\Gamma_i + 1 = R_{\mathrm{SNR}}^{i-1}, \quad i = 1,\ldots,L. \tag{62}$$

A common SNR ratio is used for noncoherent ASK just like equally spaced constellation points are used for coherent ASK. It has also been shown in earlier work [11] that an approximately common SNR ratio is optimal for high SNR in the noncoherent ASK case. Substituting (62) in (61), we obtain

$$\sum_{i=1}^{L} R_{\mathrm{SNR}}^{i-1} = L(\Gamma_{\mathrm{av}}+1). \tag{63}$$

The SNR common ratio R_{SNR} is obtained from the average SNR per symbol per branch Γ_{av} by finding the real positive solution (that is greater than 1) of the $(L-1)$th degree polynomial equation (in R_{SNR})

$$\sum_{i=2}^{L} R_{\mathrm{SNR}}^{i-1} - \Big[L(\Gamma_{\mathrm{av}}+1) - 1\Big] = 0.$$

The values of R_{SNR} obtained for $\Gamma_{\mathrm{av}} = 1, 5, 10, 20, 100$, that is, 0 dB, 7 dB, 10 dB, 13 dB, 20 dB, are

$1.4883, 2.4433, 3.1138, 3.9824, 7.0246$, respectively, when $L = 4$,

$1.1917, 1.4927, 1.6673, 1.8649, 2.4108$, respectively, when $L = 8$,

$1.0864, 1.2103, 1.2764, 1.3471, 1.5241$, respectively, when $L = 16$.

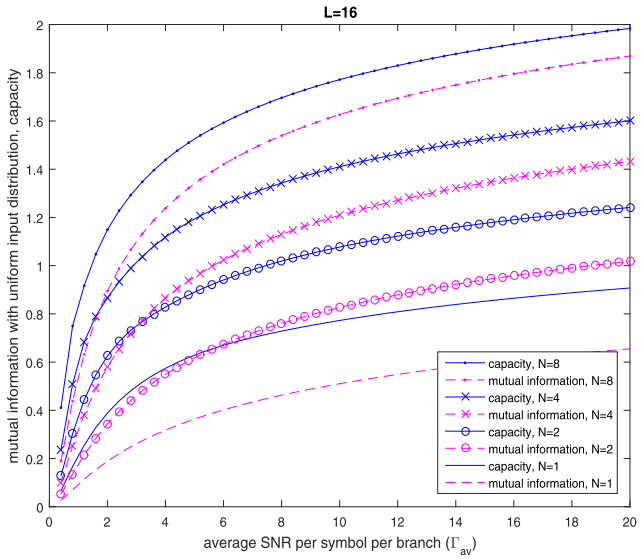

Fig. 1. Mutual information with uniform input distribution and capacity versus average SNR per symbol per branch Γ_{av} for 16-level ASK ($L = 16$).

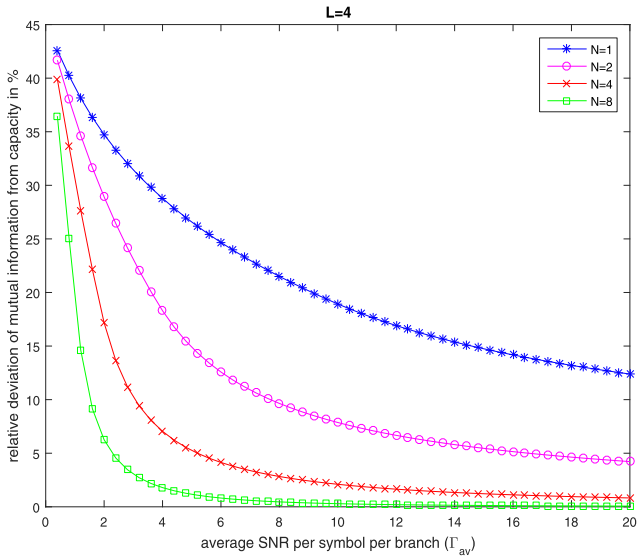

Fig. 2. Relative deviation of mutual information (with uniform input distribution) from capacity versus average SNR per symbol per branch Γ_{av} for 4-level ASK ($L = 4$).

Plots of the mutual information with uniform distribution (computed from (15)) and the capacity (obtained from numerical maximization of (15) subject to the constraints (18)) versus Γ_{av} for 16-level ASK with different values of the number of receive diversity branches N are shown in Fig. 1. We find from the figure that the capacity and the mutual information with uniform input distribution increase with increase in Γ_{av} and increase in N. Furthermore, for each N and Γ_{av}, the capacity is higher than the mutual information with uniform input distribution, as expected. The relative deviation of the mutual information (with uniform input distribution) from the

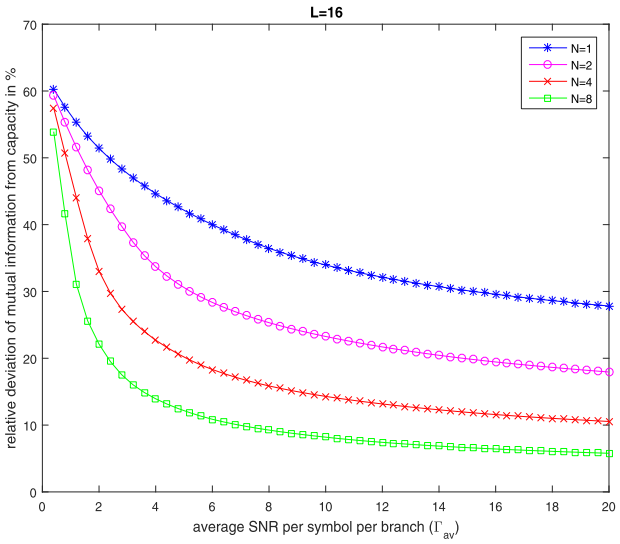

Fig. 3. Relative deviation of mutual information (with uniform input distribution) from capacity versus average SNR per symbol per branch $\Gamma_{\rm av}$ for 16-level ASK ($L = 16$).

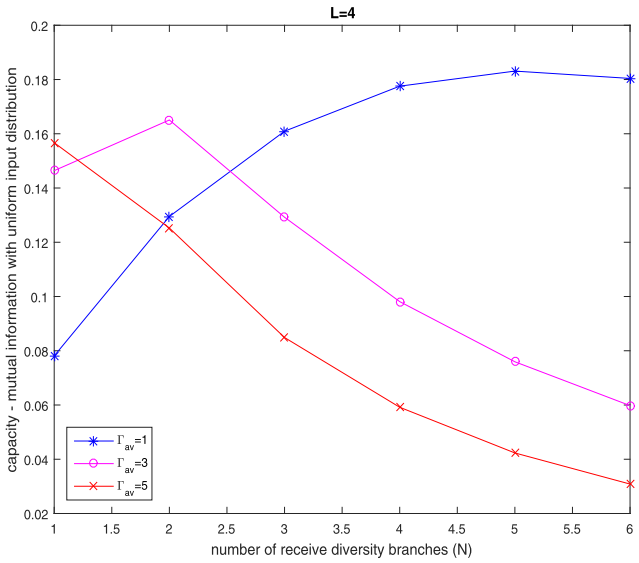

Fig. 4. Capacity minus mutual information with uniform input distribution versus number of receive diversity branches N for 4-level ASK ($L = 4$).

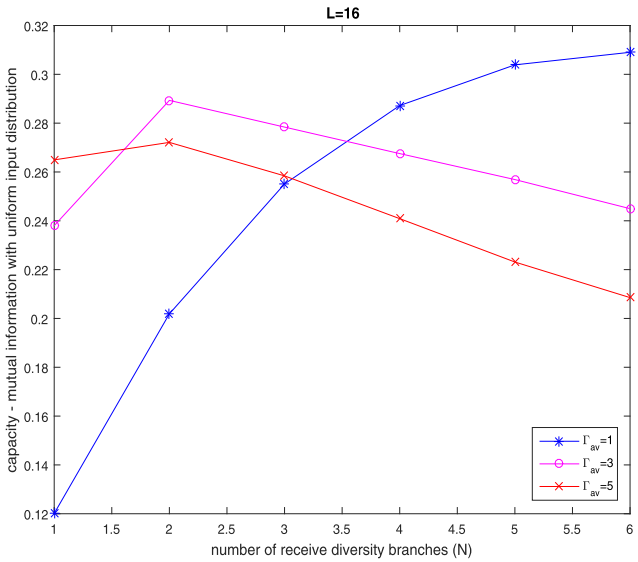

Fig. 5. Capacity minus mutual information with uniform input distribution versus number of receive diversity branches N for 16-level ASK ($L = 16$).

capacity, given by

$$\begin{aligned}&\text{relative deviation of mutual information}\\&\text{(with uniform input distribution) from capacity}\\&=\frac{\text{capacity}-\text{mutual information with uniform input distribution}}{\text{capacity}},\end{aligned}\tag{64}$$

versus $\Gamma_{\rm av}$ is plotted in Figs. 2 and 3 for $L = 4$ and $L = 16$, respectively, with different values of N. It is observed from the figures that the relative deviation decreases with increase in $\Gamma_{\rm av}$ and increase in N, and increases with increase in L. At $\Gamma_{\rm av} = 4$ (that is, 6 dB), the relative deviations for $L = 4$ (Fig. 2) are 1.8%, 7%, 18.3%, and 28.8% with $N = 8$, $N = 4$, $N = 2$, and $N = 1$, respectively, for $L = 8$ (plots are not shown) they are 9.6%, 17.8%, 28.8%, and 39.8% with $N = 8$, $N = 4$, $N = 2$, and $N = 1$, respectively, and for $L = 16$ (Fig. 3) they are 13.9%, 22.7%, 33.7%, and 44.7% with $N = 8$, $N = 4$, $N = 2$, and $N = 1$, respectively. Thus there is an advantage in using capacity achieving input probabilities instead of a uniform input distribution for signaling, with the advantage growing with decrease in $\Gamma_{\rm av}$, decrease in N, and increase in L. Plots of the capacity minus mutual information with uniform input distribution versus N are shown in Figs. 4 and 5 for $L = 4$ and $L = 16$, respectively, with different values of $\Gamma_{\rm av}$. We find from the figures that there exists a value of N for which this difference attains a maximum, and this maximizing N decreases with increase in $\Gamma_{\rm av}$. The relative deviation of the uniform distribution from the optimum or capacity achieving probabilities $p_{1,\rm opt}, \ldots, p_{L,\rm opt}$, given by

$$\begin{aligned}&\text{relative deviation of uniform distribution from optimum}\\&\text{probabilities}\\&=\frac{\sqrt{\sum_{\ell=1}^{L}\left(p_{L,\rm opt}-\frac{1}{L}\right)^2}}{\sqrt{\sum_{\ell=1}^{L}p_{L,\rm opt}^2}},\end{aligned}\tag{65}$$

versus $\Gamma_{\rm av}$ ($0.4 \le \Gamma_{\rm av} \le 20$) is plotted in Fig. 6 for $L = 8$ with different values of N. We find from the figure that the relative deviation ranges from 86.6% to 86.7% at $\Gamma_{\rm av} = 0.4$ (-4 dB) over $N = 1, 2, 4, 8$, and from 38.7% to 81.3% at $\Gamma_{\rm av} = 20$ (13 dB) over $N = 1, 2, 4, 8$. On the other hand, for $L = 4$ (plots are not shown), the relative deviation ranges from 70.8% to 70.9% at $\Gamma_{\rm av} = 0.4$ over $N = 1, 2, 4, 8$, and from 4.3% to 58.7% at $\Gamma_{\rm av} = 20$ over $N = 1, 2, 4, 8$, while for $L = 16$ (plots are not shown), the relative deviation is about 93% at $\Gamma_{\rm av} = 0.4$ over $N = 1, 2, 4, 8$, and ranges from about 72% to about 91% at $\Gamma_{\rm av} = 20$ over $N = 1, 2, 4, 8$. These indicate that the relative deviation increases to significant amounts with increase in L.

Plots of the capacity achieving probabilities (obtained by numerically maximizing (15) subject to the constraints (18))

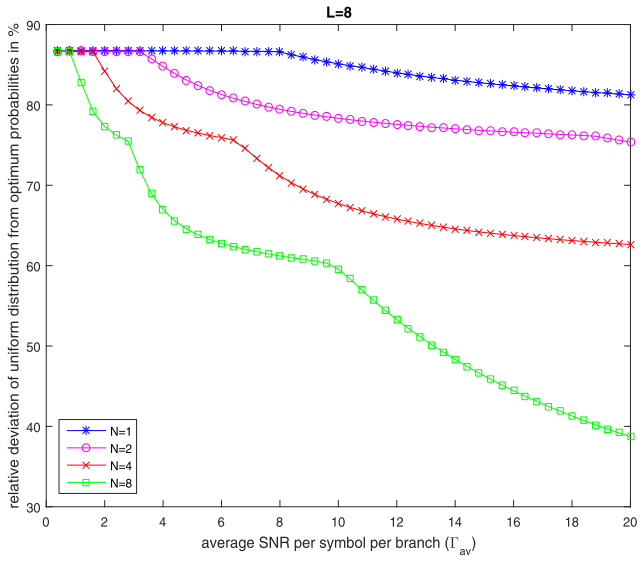

Fig. 6. Relative deviation of uniform distribution from optimum probabilities versus average SNR per symbol per branch $\Gamma_{\rm av}$ for 8-level ASK ($L = 8$).

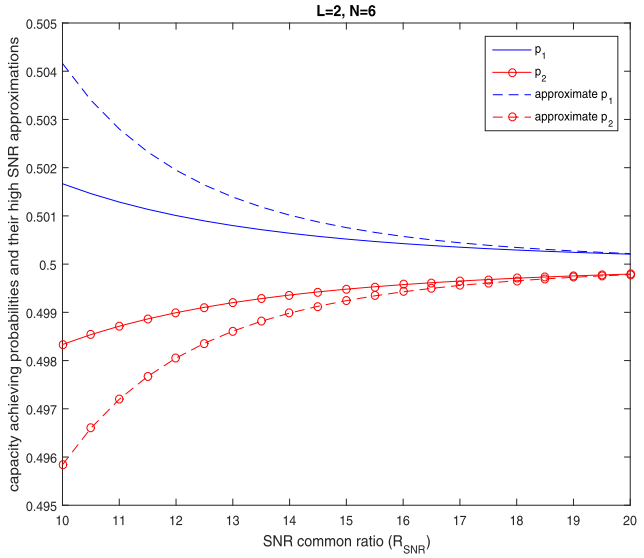

Fig. 7. Variation of capacity achieving probabilities and their high SNR approximations with SNR common ratio $R_{\rm SNR}$ for binary ASK ($L = 2$) with number of receive diversity branches $N = 6$ (relative approximation error $\in [2.4458 \times 10^{-3}\%, 0.5\%]$, average relative approximation error over all data points $= 0.11\%$).

and their high SNR approximations (computed from (40)) versus $R_{\rm SNR}$ are shown in Fig. 7 for $L = 2$, $N = 6$ and in Fig. 8 for $L = 4$, $N = 6$. The relative approximation error in terms of the capacity achieving probabilities $p_{1,\rm opt}, \ldots, p_{L,\rm opt}$ and their high SNR approximations $p_{1,\rm hiSNR,opt}, \ldots, p_{L,\rm hiSNR,opt}$ is given by

$$\text{relative approximation error in probability} = \frac{\sqrt{\sum_{\ell=1}^{L} \left(p_{L,\rm opt} - p_{L,\rm hiSNR,opt}\right)^2}}{\sqrt{\sum_{\ell=1}^{L} p_{L,\rm opt}^2}}. \quad (66)$$

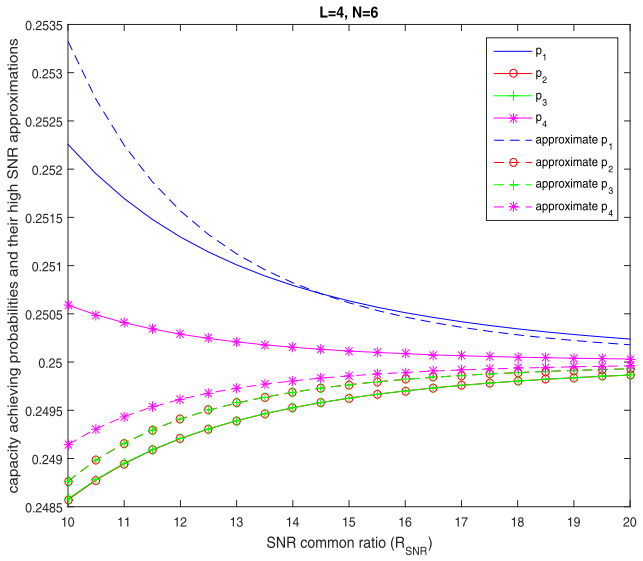

Fig. 8. Variation of capacity achieving probabilities and their high SNR approximations with SNR common ratio $R_{\rm SNR}$ for 4-level ASK ($L = 4$) with number of receive diversity branches $N = 6$ (relative approximation error $\in [0.03\%, 0.36\%]$, average relative approximation error over all data points $= 0.1\%$).

We find that for $L = 2$, $N = 6$ (Fig. 7), the relative approximation error is in the range $[2.4458 \times 10^{-3}\%, 0.5\%]$, and the average relative approximation error over all data points is 0.11%, while for $L = 2$, $N = 2$, using the same values of $R_{\rm SNR}$ as in Fig. 7 (plots are not shown), the relative approximation error is in the range $[0.01\%, 0.76\%]$ and the average relative approximation error over all data points is 0.5%. On the other hand, for $L = 4$, $N = 6$ (Fig. 8), the relative approximation error is in the range $[0.03\%, 0.36\%]$, and the average relative approximation error over all data points is 0.1%, while for $L = 4$, $N = 2$, using the same values of $R_{\rm SNR}$ as in Fig. 8 (plots are not shown), the relative approximation error is in the range $[2.89\%, 6.58\%]$ and the average relative approximation error over all data points is 4.29%. Therefore the high SNR approximation (40) is reasonably accurate, with the average relative approximation error over all data points decreasing with increase in N.

Plots of the capacity (obtained from numerical maximization of (15) subject to the constraints (18)), its high SNR approximation (computed from (41)), its low SNR approximation (computed from (52)), and its logarithmic upper bound (computed from (56)) versus $R_{\rm SNR}$ are shown in Fig. 9 for $L = 2$, $N = 2$ and in Fig. 10 for $L = 8$, $N = 2$. We find from the figures that the capacity increases with increase in $R_{\rm SNR}$, and, to achieve a high capacity close to $\log_2 L$ bits per channel use, such as a capacity of $0.8 \log_2 L$ bits per channel use, a large $R_{\rm SNR}$ is needed for a small N. However, the value of $R_{\rm SNR}$ needed reduces with increase in N. For example, for $L = 2$, $N = 2$ (Fig. 9), an $R_{\rm SNR}$ of about 12 is needed to achieve a capacity of 0.8 bits per channel use, whereas for $L = 2$, $N = 6$ (plots are not shown), an $R_{\rm SNR}$ of only about 3.9 is needed for the same. Similarly, for $L = 8$, $N = 2$ (Fig. 10), an $R_{\rm SNR}$ of about 7.2 is needed to achieve a capacity of

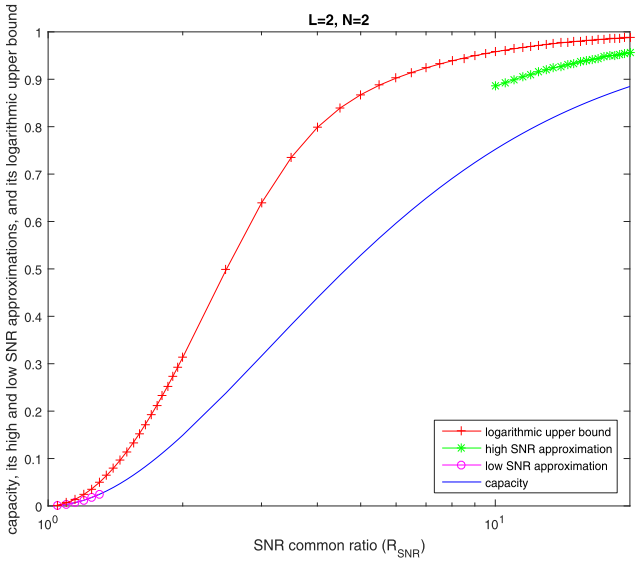

Fig. 9. Variation of capacity, its high and low SNR approximations, and its logarithmic upper bound with SNR common ratio R_{SNR} for binary ASK ($L = 2$) with number of receive diversity branches $N = 2$.

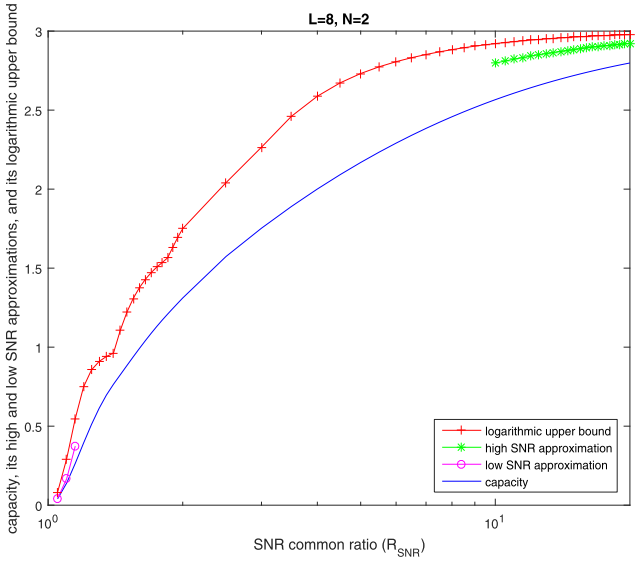

Fig. 10. Variation of capacity, its high and low SNR approximations, and its logarithmic upper bound with SNR common ratio R_{SNR} for 8-level ASK ($L = 8$) with number of receive diversity branches $N = 2$.

2.4 bits per channel use, whereas for $L = 8$, $N = 6$ (plots are not shown), an R_{SNR} of only about 2.9 is needed for the same. We also find that the high and low SNR approximations of the capacity are reasonably accurate. The accuracy of the high SNR approximation increases with increase in N and the accuracy of the low SNR approximation increases with decrease in N.

VII. Conclusion

Analytical and numerical results for the capacity of noncoherent reception of multi-level one-sided ASK, which is an asymmetric constellation, in Rayleigh fading with receive diversity and energy detection have been provided. An analytical expression for the mutual information in terms of a single integral is derived, and high and low SNR approximations of the optimum or capacity achieving input probabilities and the capacity are provided. The approximations are found to be reasonably accurate. Although not supported by a rigorous mathematical analysis, the quality of the approximations is validated by means of numerical examples. The numerical results also confirm that the uniform distribution of input probabilities is not capacity achieving. For example, with average SNR per symbol per branch of 6 dB, the relative deviation of the mutual information (with uniform input distribution) from the capacity is nearly 20% for 4-level ASK with two receive diversity branches. The relative deviation increases with increase in the number of ASK levels; for 16-ASK with two receive diversity branches, the relative deviation increases to over 30%. On the other hand, the relative deviation decreases with increase in average SNR per symbol per diversity branch and number of receive diversity branches. These findings imply that in the design of noncoherent energy detection systems with one-sided ASK, optimization of the input probabilities is useful for increasing channel capacity.

Appendix A
Derivation of (21)

The partial derivatives of $\mathcal{I}(s;\mathbf{r})$ in (20a) are expressed using (15) as

$$\begin{aligned}\frac{\partial \mathcal{I}(s;\mathbf{r})}{\partial p_\ell} = &-\frac{1}{\ln 2}\sum_{i=1}^{L} p_i \int_0^\infty \frac{R_{\ell,i}^N \exp\{-xR_{\ell,i}\}}{\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right)} \\ &\times \left[\frac{x^{N-1}\exp\{-x\}}{(N-1)!}\right] dx \\ &-\frac{1}{\ln 2}\int_0^\infty \ln\left(\sum_{j=1}^{L} p_j R_{j,\ell}^N \exp\{-xR_{j,\ell}\}\right) \\ &\times \frac{x^{N-1}\exp\{-x\}}{(N-1)!} dx, \\ &\ell = 1,\ldots,L. \end{aligned} \tag{67}$$

Changing the variable of integration from x to $x' = xR_{\ell,i}$ in each term in the summation over i in (67), we obtain

$$\begin{aligned}&\int_0^\infty \frac{R_{\ell,i}^N \exp\{-xR_{\ell,i}\}}{\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-xR_{j,i}\}\right)} \\ &\qquad \times \left[\frac{x^{N-1}\exp\{-x\}}{(N-1)!}\right] dx \\ &= \int_0^\infty \frac{\exp\{-x'\}}{\left(\sum_{j=1}^{L} p_j R_{j,i}^N \exp\{-x' R_{\ell,i}^{-1} R_{j,i}\}\right)} \\ &\qquad \times \left[\frac{(x')^{N-1}\exp\{-x' R_{\ell,i}^{-1}\}}{(N-1)!}\right] dx'. \end{aligned} \tag{68}$$

Observe from (16) and (17) that

$$R_{\ell,i}^{-1}R_{j,i} = R_{i,\ell}R_{j,i} = R_{j,\ell}, \quad R_{\ell,i}^{-1} = R_{i,\ell},$$
$$R_{j,i}^{N} = R_{\ell,i}^{N}R_{j,\ell}^{N} = R_{i,\ell}^{-N}R_{j,\ell}^{N}. \tag{69}$$

Substituting (69) in (68), we get

$$\begin{aligned}
&\sum_{i=1}^{L} p_i \int_0^{\infty} \frac{R_{\ell,i}^{N}\exp\{-xR_{\ell,i}\}}{\left(\sum_{j=1}^{L} p_j R_{j,i}^{N}\exp\{-xR_{j,i}\}\right)} \\
&\qquad \times \left[\frac{x^{N-1}\exp\{-x\}}{(N-1)!}\right] dx \\
&= \int_0^{\infty} \frac{\left(\sum_{i=1}^{L} p_i R_{i,\ell}^{N}\exp\{-x'R_{i,\ell}\}\right)}{\left(\sum_{j=1}^{L} p_j R_{j,\ell}^{N}\exp\{-x'R_{j,\ell}\}\right)} \\
&\qquad \times \left[\frac{(x')^{N-1}\exp\{-x'\}}{(N-1)!}\right] dx' \\
&= \int_0^{\infty} \frac{(x')^{N-1}\exp\{-x'\}}{(N-1)!} dx' \\
&= 1.
\end{aligned} \tag{70}$$

Substitution of (70) in (67) results in

$$\begin{aligned}
&\frac{\partial \mathcal{I}(s;\mathbf{r})}{\partial p_\ell} \\
&\quad = -\frac{1}{\ln 2} \\
&\qquad - \frac{1}{\ln 2}\int_0^{\infty} \ln\left(\sum_{j=1}^{L} p_j R_{j,\ell}^{N}\exp\{-xR_{j,\ell}\}\right) \\
&\qquad \times \frac{x^{N-1}\exp\{-x\}}{(N-1)!} dx, \\
&\qquad \ell = 1,\ldots,L.
\end{aligned} \tag{71}$$

Further substitution of (71) in (20) results in the set of equations (21).

APPENDIX B
DERIVATION OF (40)

To obtain the capacity under the high SNR condition (23), we need to maximize $\mathcal{I}_{\mathrm{hiSNR}}(s;\mathbf{r})$ over $\{p_1,\ldots,p_L\}$ subject to the constraint $\sum_{i=1}^{L} p_i = 1$. We construct the Lagrangian function $\mathcal{L}_{\mathrm{h}}$ as

$$\mathcal{L}_{\mathrm{h}} = (\ln 2)\mathcal{I}_{\mathrm{hiSNR}}(s;\mathbf{r}) - \lambda_{\mathrm{h}}\left(\sum_{i=1}^{L} p_i - 1\right), \tag{72}$$

where λ_{h} is a Lagrangian multiplier. We obtain the high SNR capacity achieving $p_1,\ldots,p_L$ from (72) by solving the set of equations

$$(\ln 2)\frac{\partial \mathcal{I}_{\mathrm{hiSNR}}(s;\mathbf{r})}{\partial p_\ell} = \lambda_{\mathrm{h}}, \quad \ell = 1,\ldots,L, \tag{73a}$$

$$\sum_{i=1}^{L} p_i = 1. \tag{73b}$$

Substituting (39) in (73a), we get

$$\begin{aligned}
\ln p_1 =& -1-\lambda_{\mathrm{h}} \\
&- \frac{(N\ln R_{1,2})^{N-1}}{(N-1)!\left(1-R_{1,2}^{-1}\right)^{N-1} R_{1,2}^{\frac{N}{\left(1-R_{1,2}^{-1}\right)}}}, \qquad (74a) \\
\ln p_\ell =& -1-\lambda_{\mathrm{h}} \\
&- \frac{(N\ln R_{\ell,\ell+1})^{N-1}}{(N-1)!\left(1-R_{\ell,\ell+1}^{-1}\right)^{N-1} R_{\ell,\ell+1}^{\frac{N}{\left(1-R_{\ell,\ell+1}^{-1}\right)}}} \\
&- \frac{(N\ln R_{\ell-1,\ell})^{N}}{N!(R_{\ell-1,\ell}-1)^{N}}, \\
&\ell = 2,\ldots,L-1, \qquad (74b) \\
\ln p_L =& -1-\lambda_{\mathrm{h}} - \frac{(N\ln R_{L-1,L})^{N}}{N!(R_{L-1,L}-1)^{N}}. \qquad (74c)
\end{aligned}$$

Applying the constraint (73b) to (74) gives

$$\begin{aligned}
&\exp\{1+\lambda_{\mathrm{h}}\} \\
&\quad = \exp\left\{-\frac{(N\ln R_{1,2})^{N-1}}{(N-1)!\left(1-R_{1,2}^{-1}\right)^{N-1} R_{1,2}^{\frac{N}{\left(1-R_{1,2}^{-1}\right)}}}\right\} \\
&\qquad + \exp\left\{-\frac{(N\ln R_{L-1,L})^{N}}{N!(R_{L-1,L}-1)^{N}}\right\} \\
&\qquad + \sum_{\ell=2}^{L-1}\exp\left\{-\frac{(N\ln R_{\ell,\ell+1})^{N-1}}{(N-1)!\left(1-R_{\ell,\ell+1}^{-1}\right)^{N-1} R_{\ell,\ell+1}^{\frac{N}{\left(1-R_{\ell,\ell+1}^{-1}\right)}}}\right\} \\
&\qquad \times \exp\left\{-\frac{(N\ln R_{\ell-1,\ell})^{N}}{N!(R_{\ell-1,\ell}-1)^{N}}\right\}.
\end{aligned} \tag{75}$$

Substitution of (75) in (74) results in the optimum or high SNR capacity achieving $p_1,\ldots,p_L$, which are denoted as $p_{1,\mathrm{hiSNR,opt}},\ldots,p_{L,\mathrm{hiSNR,opt}}$, respectively, and are given by (40).

REFERENCES

[1] X.-C. Gao, J.-K. Zhang, H. Chen, Z. Dong, and B. Vucetic, "Energy-efficient and low-latency massive SIMO using noncoherent ML detection for industrial IoT communications," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6247–6261, Aug. 2019.

[2] C. Xu *et al.*, "Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures," *IEEE Access*, vol. 7, pp. 178246–178299, 2019.

[3] H. Xie, W. Xu, H. Q. Ngo, and B. Li, "Non-coherent massive MIMO systems: A constellation design approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3812–3825, Jun. 2020.

[4] D. Feng, C. Jiang, G. Lim, L. J. Cimini, Jr., G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, 1st Quart., 2013.

[5] J. K. Devineni and H. S. Dhillon, "Non-coherent detection and bit error rate for an ambient backscatter link in time-selective fading," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 602–618, Jan. 2021.

[6] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 117–125, Apr. 2015.

[7] D. Yu, G. Yue, A. Liu, and L. Yang, "Absolute amplitude differential phase spatial modulation and its non-coherent detection under fast fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2742–2755, Apr. 2020.

[8] J. G. Lawton, "Theoretical error rates of 'differentially coherent' binary and 'Kineplex' data transmission systems," *Proc. IRE*, vol. 47, no. 2, pp. 333–334, 1959.

[9] C. Cahn, "Performance of digital phase-modulation communication systems," *IRE Trans. Commun. Syst.*, vol. 7, no. 1, pp. 3–6, May 1959.

[10] T. Kailath, "Correlation detection of signals perturbed by a random channel," *IEEE Trans. Inf. Theory*, vol. IT-6, no. 3, pp. 361–366, Jun. 1960.
[11] R. K. Mallik and R. D. Murch, "Noncoherent reception of multi-level ASK in Rayleigh fading with receive diversity," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 135–143, Jan. 2014.
[12] M. Chowdhury, A. Manolakos, and A. Goldsmith, "Scaling laws for noncoherent energy-based communications in the SIMO MAC," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1980–1992, Apr. 2016.
[13] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.
[14] D. I. Kim, J. H. Moon, and J. J. Park, "New SWIPT using PAPR: How it works," *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, pp. 672–675, Dec. 2016.
[15] I. Jacobs, "The asymptotic behavior of incoherent M-ary communication systems," *Proc. IEEE*, vol. 51, no. 1, pp. 251–252, Jan. 1963.
[16] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[17] S. Ray, M. Medard, and L. Zheng, "On noncoherent MIMO channels in the wideband regime: Capacity and reliability," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 1983–2009, Jun. 2007.
[18] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.
[19] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
[20] A. Schenk and R. F. Fischer, "Noncoherent detection in massive MIMO systems," in *Proc. 17th Int. ITG Workshop Smart Antennas (WSA)*, Stuttgart, Germany, Mar. 2013, pp. 1–8.
[21] R. K. Mallik, S. Singh, R. D. Murch, and S. Mehra, "Signal design for multiple antenna systems with spatial multiplexing and noncoherent reception," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1245–1258, Apr. 2015.
[22] Y. Li, R. K. Mallik, and R. Murch, "Channel magnitude-based MIMO with energy detection for Internet of Things applications," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9893–9907, Dec. 2019.
[23] G. Psaltopoulos and A. Wittneben, "Nonlinear MIMO: Affordable MIMO technology for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 824–832, Feb. 2010.
[24] G. K. Psaltopoulos and A. Wittneben, "Diversity and spatial multiplexing of MIMO amplitude detection receivers," in *Proc. IEEE 20th Int. Symp. Pers., Indoor Mobile Radio Commun.*, Tokyo, Japan, Sep. 2009, pp. 202–206.
[25] M. Chowdhury, A. Manolakos, F. Gomez-Cuba, E. Erkip, and A. J. Goldsmith, "Capacity scaling in noncoherent wideband massive SIMO systems," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jerusalem, Israel, Apr./May 2015, pp. 1–5.
[26] R. K. Mallik, R. D. Murch, and Y. Li, "Channel magnitude based energy detection with receive diversity for multi-level amplitude-shift keying in Rayleigh fading," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3079–3094, Jul. 2017.
[27] A. Al-Dweik and F. Xiong, "Frequency-hopped multiple-access communications with noncoherent M-ary OFDM-ASK," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 33–36, Jan. 2003.
[28] A. Anttonen, A. Kotelba, and A. Mämmelä, "Energy detection of multilevel PAM signals with systematic threshold mismatch," *Res. Lett. Commun.*, vol. 2009, pp. 1–4, Jan. 2009.
[29] A. Anttonen, A. Kotelba, and A. Mämmelä, "Error performance of PAM systems using energy detection with optimal and suboptimal decision thresholds," *Phys. Commun.*, vol. 4, no. 2, pp. 111–122, Jun. 2011.
[30] Y. Maghsoodi and A. Al-Dweik, "Error-rate analysis of FHSS networks using exact envelope characteristic functions of sums of stochastic signals," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 974–985, Mar. 2008.
[31] W. Osborne and M. Luntz, "Coherent and noncoherent detection CPFSK," *IEEE Trans. Commun.*, vol. COM-22, no. 8, pp. 1023–1036, Aug. 1974.

Ranjan K. Mallik (Fellow, IEEE) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Kanpur, Kanpur, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1988 and 1992, respectively.

From August 1992 to November 1994, he was a scientist with the Defence Electronics Research Laboratory, Hyderabad, India, working on missile and EW projects. From November 1994 to January 1996, he was a faculty member with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, Kharagpur. From January 1996 to December 1998, he was with the faculty of the Department of Electronics and Communication Engineering, Indian Institute of Technology Guwahati, Guwahati. Since December 1998, he has been with the Faculty of the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, where he is currently an Institute Chair Professor. His research interests are in diversity combining and channel modeling for wireless communications, space-time systems, cooperative communications, multiple-access systems, power line communications, molecular communications, difference equations, and linear algebra.

Dr. Mallik is a member of Eta Kappa Nu, the IEEE Communications, Information Theory and Vehicular Technology Societies, the American Mathematical Society, and the International Linear Algebra Society; a fellow of the Indian National Academy of Engineering, the Indian National Science Academy, The National Academy of Sciences, India, Prayagraj, the Indian Academy of Sciences, Bengaluru, The World Academy of Sciences-for the advancement of science in developing countries (TWAS), The Institution of Engineering and Technology, U.K., The Institution of Electronics and Telecommunication Engineers, India, and The Institution of Engineers (India); and a life member of the Indian Society for Technical Education. He is a recipient of the Hari Om Ashram Prerit Dr. Vikram Sarabhai Research Award in the field of electronics, telematics, informatics, and automation, the Shanti Swarup Bhatnagar Prize in engineering sciences, the Khosla National Award, and the J. C. Bose Fellowship. He has served as an Area Editor and an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.

Ross Murch (Fellow, IEEE) received the Bachelor's and Ph.D. degrees in electrical and electronic engineering from the University of Canterbury, Christchurch, New Zealand.

He is currently a Chair Professor with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, where he was the Department Head from 2009 to 2015. He has taken sabbaticals at Imperial College London, London, U.K.; MIT, Cambridge, MA, USA; Allgon, Akersberga, Sweden; and AT&T, Newman Springs, NJ, USA. His unique expertise lies in his combination of knowledge from both wireless communication systems and electromagnetic areas. He has over 200 publications and 20 patents on wireless communication systems and antennas with over 17000 citations. His current research interests include multiport antennas, RF energy harvesting, the Internet of Things, acoustics, and RF imaging.

Dr. Murch is a recipient of several awards, including the Computer Simulation Technology (CST) University Publication Award in 2015 and two teaching awards. He has served the IEEE in various positions, including an Area Editor, the Technical Program Chair, a Distinguished Lecturer, and a member of the Fellow Evaluation Committee.