# Polar Codes for Decomposed Multi-Input Multi-Output Gaussian Broadcast Channels

M. Yusuf Şener *, Gerhard Kramer, Shlomo Shamai (Shitz), Ronald Böhnke, Wen Xu

Posted Date: 19 May 2026

doi: 10.20944/preprints202605.1186.v1

Keywords: broadcast channel; capacity; dirty paper coding; multi-input multi-output

 doi:10.20944/preprints202605.1186.v1



*Article*

# Polar Codes for Decomposed Multi-Input Multi-Output Gaussian Broadcast Channels

**M. Yusuf Şener [1,2,*], Gerhard Kramer [1], Shlomo Shamai (Shitz) [3], Ronald Böhnke [2] and Wen Xu [2]**

[1] Institute for Communications Engineering, School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany
[2] Huawei Technologies Duesseldorf GmbH, 80992 Munich, Germany
[3] Department of Electrical and Computer Engineering, Technion—Israel Institute of Technology, Haifa 3200003, Israel
* Correspondence: yusuf.sener@tum.de
† This article is a revised and expanded version of the paper entitled "Achieving Gaussian Vector Broacast Channel Capacity with Scalar Lattices", which was presented at the IEEE International Symposium on Information Theory 2024, Athens, 7–12 July 2024.

**Abstract**

Dirty paper coding (DPC) is applied to linear multi-input multi-output (MIMO) broadcast channels with additive white Gaussian noise and one message per receiver. The method decomposes each receiver channel into parallel scalar channels with known interference, then applies modulo operators, amplitude-shift keying (ASK), and probabilistic shaping. The achievable rate tuples include all points inside the capacity region by choosing truncated Gaussian shaping, large ASK alphabets, and large modulo intervals. Simulations with short polar codes show significant rate gains from DPC compared to conventional linear precoding, while maintaining similar encoder and decoder complexities.

**Keywords:** broadcast channel; capacity; dirty paper coding; multi-input multi-output

## 1. Introduction

This paper is dedicated to Professor H. Vincent Poor on the occasion of his 75th birthday. The authors would like to express their appreciation for his outstanding contributions to engineering research and his unwavering commitment to the information theory and communications communities.

The paper studies the downlink of wireless systems, focusing on how to achieve higher data rates than those achievable with linear precoding (LP) without increasing complexity. Professor Poor authored numerous influential papers on LP and nonlinear processing, covering topics such as multiuser detection [1–3], code-division multiple access (CDMA) [4–6], orthogonal and non-orthogonal multiple access (OMA and NOMA) [7,8,8–10], rate-splitting multiple access (RSMA) [11–13], and dirty paper coding (DPC) [14].

### *1.1. Linear Precoding*

LP is a linear version of superposition coding that combines two independently encoded signals. It involves modulating "beams" or vectors and has two basic variants. The first variant, which we focus on, selects one beam per message and maximizes a weighted sum rate while treating interference as noise. This approach is referred to as space division multiple access (SDMA) in [11], but we will refer to it simply as LP. The second variant avoids interference by using orthogonal beams, such as time- or frequency-division multiplexing, and is often called OMA. It is worth noting that SDMA, OMA, and other variants of LP can achieve higher rates by utilizing multiple transmission modes with optimized resource allocation across time, frequency, and power [15]. Different transmission modes can be represented by an auxiliary random variable [16] [p. 278], which is generally preferable to using a convex hull operator; see [16] [p. 288]. However, adding modes introduces latency.

 doi:10.20944/preprints202605.1186.v1



SDMA and OMA are commonly used in practice, but they yield suboptimal results. An LP variation that achieves the capacity of broadcast channels (BCs) with single-antenna transceivers utilizes successive interference cancellation (SIC), as discussed by Cover [17] [Sec . VII], Bergmans [18,19], Gallager [20], and others [21,22]. In this approach, each UE first decodes a set of signals intended for other UEs, then cancels the interference from those signals before decoding its own signal. This method is commonly referred to as NOMA. However, SIC increases the complexity and latency for the UEs.

A variation of NOMA involves splitting each message into multiple parts, encoding them separately, and decoding them individually. This approach effectively increases the total number of messages being transmitted. Message splitting was introduced by Carleial for interference channels [23] and was applied to broadcast channels around the same time; see [24] [Sec. III]. This general method is sometimes referred to as RSMA [11,12,25]. RSMA increases both the complexity and latency of base stations (BSs) and user equipments (UEs), but it can offer performance advantages over DPC when channel state information at the transmitter (CSIT) is imperfect [26,27].

The most general approach combines superposition coding with DPC, and we refer to this method as Marton coding [28]; see also [16] [p. 391]. Marton coding can be implemented by targeting specific corner points of the rate region using DPC, and then applying time-sharing. We remark that time-sharing latency can be significantly reduced by utilizing fine time-shifts and alternating DPC [29].

### *1.2. Dirty Paper Coding*

DPC arises in the context of channels with non-causal state information at the transmitter, for which the capacity was characterized by Gelfand and Pinsker [30]. Costa applied the theory to additive white Gaussian noise (AWGN) channels with additive Gaussian interference known at the transmitter. Remarkably, the capacity is the same as without interference [31], a result similar to Wyner-Ziv coding for Gaussian sources and mean square error distortion [32]. The result was later extended to general interference by using high-dimensional lattices [33,34].

DPC achieves all points in the capacity region of a linear, multi-input multi-output (MIMO), AWGN BC with independent messages, each destined for a different receiver [35–41]. The idea is to encode the messages successively, treating the signals of previously encoded messages as interference. One thereby achieves higher rates than LP. However, DPC is rarely used in practice because it is considered too complex. We argue here that DPC can be implemented with a similar level of complexity to coded LP, albeit with higher latency due to successive encoding.

A useful insight for understanding DPC is that successive encoding is the dual of successive decoding at the receiver of an appropriate (dual) multiple-access channel (MAC) [39]. Remarkably, Gaussian signaling, linear minimum mean-square error (LMMSE) estimation, and convex optimization yield the best MAC rates, and the best MAC beamforming vectors can be transformed into the best BC precoding vectors. A refined approach decomposes the BC into scalar channels and shows that the best BC precoding and MAC beamforming vectors are co-linear [42].

### *1.3. Code Constructions*

Explicit DPC constructions are described in [43–46]. These papers use low-density parity-check (LDPC) or turbo codes for reliability and trellis codes for shaping. The schemes require long block lengths, and the complexity is high. Similarly, the superposition coding scheme in [47] employs long LDPC and trellis codes. Lattice constructions with integer-forcing were proposed in [48–51]; these perform well, but typically do not achieve capacity. Non-asymptotic analyses of random coding are described in [52–60].

Polar codes [61] are an attractive alternative because they permit practical nested coding for reliability, quantization, and shaping. The paper [62] analyzes binary DPC, the paper [63] shows how to perform shaping, the paper [64] introduces polar coded modulation, and the thesis [65] designs polar lattices for DPC; see also [29,66–76] and [77–82]. We also use polar codes, but with scalar modulo operations, dithering, and probabilistic shaping [57,59]. A modulo operator reduces the transmitter

 doi:10.20944/preprints202605.1186.v1



dynamic range, and dithering decouples the constellation and interference statistics, which provides useful independencies.

### 1.4. Organization and Contributions

This paper is organized as follows. Section 2 reviews notation and results on symmetric unimodal functions. Section 3 describes the vector BC and shows that noise whitening and the singular value decomposition (SVD) give parallel scalar channels with known interference. One may thus apply scalar DPC to achieve all rate tuples in the capacity region [83]. Section 4 and Section 5 prove that the scalar coding scheme in [59] achieves these rate tuples. Section 7 designs short polar codes and provides frame error rate (FER) curves based on numerical simulations. Section 8 concludes the paper.

## 2. Preliminaries

### 2.1. Notation

Column vectors are denoted as $\underline{x} = [x_1, \ldots, x_n]^T$, where $\underline{x}^T$ is the transpose of $\underline{x}$. Bold letters such as $\mathbf{Q}$ denote matrices; $\mathbf{Q}^\dagger$ is the complex-conjugate transpose of $\mathbf{Q}$; trace$(\mathbf{Q})$ is the trace of $\mathbf{Q}$; det$(\mathbf{Q})$ is the determinant of $\mathbf{Q}$; $\mathbf{I}$ is an identity matrix. We write $j = \sqrt{-1}$, $[x]^+ = \max(0, x)$, and

$$x \bmod A = x - k'A \tag{1}$$

where $k'$ is the integer so that $x - k'A$ lies in $[-A/2, A/2)$.

Upper- and lowercase letters denote random variables (RVs) and their realizations; e.g., $\underline{X}$ is a random vector and $\underline{x}$ is its realization. $p_X$ denotes the probability density function of the random variable $X$. We remove subscripts if the argument is a lowercase version of the RV, e.g., $p(x) = p_X(x)$. Define the expectation and entropy with respect to a density $q(.)$ as

$$\mathrm{E}_q[f(X)] := \int_{\mathbb{R}} q(x)\, f(x)\, dx \tag{2}$$

$$h_q(X) := \mathrm{E}_q[-\log q(X)]. \tag{3}$$

We write $\mathrm{E}[X]$ and $h(X)$ if the density is clear from the context. $I(X;Y)$ denotes the mutual information of $X$ and $Y$. $Q(\cdot)$ denotes the $Q$-function.

### 2.2. Decomposition of a Covariance Matrix

Let $\underline{m} = \mathrm{E}[\underline{X}]$ and $\mathbf{K} = \mathrm{E}[(\underline{X} - \underline{m})(\underline{X} - \underline{m})^\dagger]$ be the mean and covariance matrix of $\underline{X}$. The matrix $\mathbf{K}$ is positive semi-definite and has the eigenvalue decomposition

$$\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger \tag{4}$$

where $\mathbf{U}$ is unitary and $\mathbf{\Lambda}$ is diagonal with non-negative entries. Suppose there are $n_+$ positive eigenvalues. Let $\bar{\mathbf{\Lambda}}$ be the $n_+ \times n_+$ matrix with these eigenvalues, and let $\bar{\mathbf{U}}$ be the $n \times n_+$ matrix with the corresponding columns of $\mathbf{U}$ removed. Then a Cholesky decomposition is

$$\mathbf{K} = \mathbf{P}\mathbf{P}^\dagger \text{ where } \mathbf{P} = \bar{\mathbf{U}}\bar{\mathbf{\Lambda}}^{1/2}\bar{\mathbf{V}}^\dagger \tag{5}$$

for any $n_+ \times n_+$ unitary matrix $\bar{\mathbf{V}}$. Observe that $\mathbf{P}$ is an $n_t \times n_+$ matrix with the left inverse

$$\mathbf{P}^+ = \bar{\mathbf{V}}\bar{\mathbf{\Lambda}}^{-1/2}\bar{\mathbf{U}}^\dagger. \tag{6}$$

For example, if $n_+ = n$ and $\bar{\mathbf{V}} = \mathbf{U}$ then $\mathbf{K} = \mathbf{P}^2$. In this case, one usually writes $\mathbf{P} = \mathbf{K}^{1/2}$ and $\mathbf{P}^+ = \mathbf{P}^{-1} = \mathbf{K}^{-1/2}$.

 doi:10.20944/preprints202605.1186.v1



*2.3. Symmetric and Unimodal Functions*

The real-valued function $f(.)$ is symmetric if $f(x) = f(-x)$ for all $x \in \mathbb{R}$. The following (known) lemma is proved in Appendix A.

**Lemma 1.** *The convolution $f * g(.)$ of two symmetric functions $f(.)$ and $g(.)$ is symmetric.*

We study symmetric and unimodal probability density functions (p.d.f.s) $p(.)$, i.e., $p(x)$ is non-increasing for $x \geq 0$. The following lemma was proved in [84, pp. 30-32] and [85, Thm. 2.1]; Appendix A provides an alternative proof.

**Lemma 2.** *The convolution $p * q(.)$ of two symmetric unimodal p.d.f.s $p(.)$ and $q(.)$ is symmetric unimodal.*

Finally, we derive bounds on the sum of uniformly spaced samples of a symmetric unimodal function $f(.)$. Appendix A proves the following lemma for the spacing $\kappa$.

**Lemma 3.** *Consider a symmetric unimodal $f(.)$. We have*

$$\left| \int_{-\frac{A}{2}}^{\frac{A}{2}} f(y)dy - \sum_{k=-\lfloor M/2 \rfloor}^{\lceil M/2 \rceil - 1} \kappa f(x + k\kappa) \right| \leq \kappa f(0) \tag{7}$$

*for $x$ satisfying $0 \leq x < \kappa/2$. Similarly, if $f(.)$ has finite area then for any $x$ we have*

$$\left| \int_{\mathbb{R}} f(y)dy - \sum_{k \in \mathbb{Z}} \kappa f(x + k\kappa) \right| \leq \kappa f(0). \tag{8}$$

*2.4. Scalar DPC Scheme*

We review the scalar DPC scheme from [59]. Consider the real-alphabet additive noise channel

$$Y = X + S + Z \tag{9}$$

where $X$ satisfies the power constraint $\mathrm{E}[X^2] \leq P_X$, $S$ is interference known at the transmitter, and $Z$ (not necessarily Gaussian) is independent of $(X, S)$ and has variance $P_Z$. The encoder computes [59] [eq. (1)]

$$S' = (\alpha S + D) \bmod A \tag{10}$$

where the dither $D$ is continuously uniform over the interval $[-A/2, A/2)$. Consider a shaping density $q(x)$. A symbol $U$ is selected from the $M$-ASK constellation

$$\mathcal{U} = \left\{ -\frac{A}{2} + (k + 1/2)\kappa \right\}_{k=0}^{M-1}, \quad \kappa = \frac{A}{M} \tag{11}$$

by using the discrete shaping distribution

$$P(u|s') = \frac{q((u - s') \bmod A)}{\sum_{v \in \mathcal{U}} q((v - s') \bmod A)}, \quad u \in \mathcal{U}. \tag{12}$$

Note that $\kappa$ is the ASK-spacing. The transmitted signal is (see [59, eq. (2)])

$$X = (U - S') \bmod A \tag{13}$$

 doi:10.20944/preprints202605.1186.v1



where $\alpha = P_X/(P_X + P_Z)$. The shaping (12) induces the input distribution

$$p(x) = q(x)/d(x), \quad x \in [-A/2, A/2) \tag{14}$$

where

$$d(x) = \sum_{k=0}^{M-1} \kappa\, q((x + k\kappa) \bmod A). \tag{15}$$

The received signal is (see [59] [eq. (3)-(4)])

$$\begin{aligned} Y' &= (\alpha X + S' + \alpha Z) \bmod A \\ &= (U + Z') \bmod A. \end{aligned} \tag{16}$$

where the effective noise is

$$Z' = (\alpha - 1)X + \alpha Z \tag{17}$$

with variance $\sigma_{Z'}^2 = \alpha P_Z$.

We mostly consider truncated Gaussian shaping with parameters $A$ and $\sigma_q$, namely (see [59] [eq. (12)])

$$q(x) = \frac{e^{-x^2/\left(2\sigma_q^2\right)}}{c\sqrt{2\pi\sigma_q^2}}, \quad x \in [-A/2, A/2) \tag{18}$$

where $c = 1 - 2Q(A/(2\sigma_q))$. We have $\mathrm{E}_q[X] = 0$, and compute the variance and entropy [59] [eq. (16) and (17)]

$$\mathrm{E}_q[X^2] = \sigma_q^2 \left[1 - \frac{A\, e^{-A^2/(8\sigma_q^2)}}{c\sqrt{2\pi\sigma_q^2}}\right] \tag{19}$$

$$h_q(X) = \frac{1}{2}\log(2\pi e\, \sigma_q^2 c^2) - \frac{1}{2}\left(1 - \frac{\mathrm{E}_q[X^2]}{\sigma_q^2}\right)\log e. \tag{20}$$

## 3. Gaussian Vector Broadcast Channels

Consider a complex-alphabet BC with the $n_t$-dimensional input $\underline{X}$ and $n_k$-dimensional outputs

$$\underline{Y}_k = \mathbf{H}_k \underline{X} + \underline{Z}_k, \quad k = 1, \ldots, K. \tag{21}$$

where $\mathbf{H}_k$ is an $n_k \times n_t$ complex matrix and $\underline{Z}_k$ is a circularly-symmetric, complex, Gaussian (CSCG) noise vector with invertible covariance matrix $\mathbf{Q}_k$. We require $\underline{X}$ to satisfy the power constraint $\mathrm{E}[\|\underline{X}\|^2] \le P_X$.

Consider one message per UE with rate $R_k$ for all $k$. Let $\mathbf{K}_k$ be an $n_t \times n_t$ covariance matrix associated with message $k$. The following rate tuples $(R_1, \ldots, R_K)$ are achievable by considering the UE ordering $1, \ldots, K$ and treating the messages $l$ with $l < k$ as being known at UE $k$:

$$R_k \le \log\det\left(\mathbf{I} + \frac{\mathbf{H}_k \mathbf{K}_k \mathbf{H}_k^\dagger}{\mathbf{I} + \sum_{l>k} \mathbf{H}_k \mathbf{K}_l \mathbf{H}_k^\dagger}\right) \tag{22}$$

subject to $\sum_{k=1}^K \mathrm{trace}(\mathbf{K}_k) \le P_X$. The capacity region is the union of all such rate tuples, taken over all $\mathbf{K}_1, \ldots, \mathbf{K}_K$ and UE orderings; see [28,40].

 doi:10.20944/preprints202605.1186.v1



### 3.1. Capacity-Achieving Scheme

The BS sends $\underline{X} = \sum_{k=1}^{K} \underline{X}_k$ where the $\underline{X}_k$ are statistically independent [39] and each $\underline{X}_k$ has an $n_{\mathrm{t}} \times n_{\mathrm{t}}$ covariance matrix $\mathbf{K}_k$ optimized for a particular rate tuple. The optimal $\underline{X}_k$ are CSCG. Using (5), we have

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{P}_k^{\dagger} \text{ where } \mathbf{P}_k = \bar{\mathbf{U}}_k \bar{\mathbf{\Lambda}}_k^{1/2} \bar{\mathbf{V}}_k^{\dagger} \tag{23}$$

where $\bar{\mathbf{U}}_k$, $\bar{\mathbf{\Lambda}}_k$, and $\bar{\mathbf{V}}_k$ have dimensions $n_{\mathrm{t}} \times n_{k+}$, $n_{k+} \times n_{k+}$, and $n_{k+} \times n_{k+}$, respectively. It suffices to consider $n_{k+} \leq n_k$ for all $k$.

Consider the ordering $1, \ldots, K$. UE $k$ sees

$$\underline{Y}_k = \mathbf{H}_k \underline{X}_k + \underbrace{\left( \sum_{l<k} \mathbf{H}_k \underline{X}_l \right)}_{:= \underline{S}_k} + \underbrace{\left( \sum_{l>k} \mathbf{H}_k \underline{X}_l \right) + \underline{Z}_k}_{:= \check{\underline{Z}}_k} \tag{24}$$

and treats $\underline{S}_k$ as interference and $\check{\underline{Z}}_k$ as noise. Let $\check{\mathbf{Q}}_k$ be the $n_k \times n_k$ covariance matrix of $\check{\underline{Z}}_k$. We use the SVD to write

$$\check{\mathbf{Q}}_k^{-1/2} \mathbf{H}_k \bar{\mathbf{U}}_k \bar{\mathbf{\Lambda}}_k^{1/2} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^{\dagger} \tag{25}$$

where $\mathbf{U}_k$ and $\mathbf{V}_k$ are unitary and $\mathbf{\Sigma}_k$ is an $n_k \times n_{k+}$ diagonal matrix with the singular values of $\check{\mathbf{Q}}_k^{-1/2} \mathbf{H}_k \bar{\mathbf{U}}_k \bar{\mathbf{\Lambda}}_k^{1/2}$. Now choose $\bar{\mathbf{V}}_k = \mathbf{V}_k^{\dagger}$ in (23). UE $k$ left-multiplies $\underline{Y}_k$ with the $n_k \times n_k$ dimensional noise-whitening filter

$$\mathbf{F}_k = \mathbf{U}_k^{\dagger} \check{\mathbf{Q}}_k^{-1/2} \tag{26}$$

to obtain the parallel channel

$$\tilde{\underline{Y}}_k = \mathbf{\Sigma}_k \tilde{\underline{X}}_k + \tilde{\underline{S}}_k + \tilde{\underline{Z}}_k \tag{27}$$

where $\tilde{\underline{S}}_k$ is known at the BS and

$$\tilde{\underline{X}}_k = \mathbf{P}_k^{+} \underline{X}_k, \quad \tilde{\underline{S}}_k = \mathbf{F}_k \underline{S}_k, \quad \tilde{\underline{Z}}_k = \mathbf{F}_k \check{\underline{Z}}_k. \tag{28}$$

Note that only the first $n_{k+}$ entries of $\tilde{\underline{Y}}_k$ are useful channel outputs, i.e., we have $n_{\mathrm{ch}} := \sum_{k=1}^{K} n_{k+}$ useful sub-channels.

The covariance matrices of $\tilde{\underline{X}}_k$ and $\tilde{\underline{Z}}_k$ are identity matrices, i.e., the entries $\tilde{X}_{k,i}$, $i = 1, \ldots, n_{k+}$, and $\tilde{Z}_{k,i}$, $i = 1, \ldots, n_k$, are independent and identically distributed (i.i.d.) CSCG with unit variance. The power constraint is $\sum_{k=1}^{K} P_k \leq P_X$ where $P_k = \mathrm{trace}\, \mathbf{K}_k$. The BS can use $\mathbf{P}_k$ as an $n_{\mathrm{t}} \times n_{k+}$ precoding matrix to compute $\underline{X}_k = \mathbf{P}_k \tilde{\underline{X}}_k$.

### 3.2. Parallel DPC with Scalar Modulo Operations

The $\tilde{\underline{X}}_k$ are independent since the $\underline{X}_k$ are independent. Moreover, the decomposition (27) specifies $n_{\mathrm{ch}}$ sub-channels where the $i$th entry of $\tilde{\underline{Y}}_k$ is

$$\tilde{Y}_{k,i} = \sigma_{k,i} \tilde{X}_{k,i} + \tilde{S}_{k,i} + \tilde{Z}_{k,i} \tag{29}$$

for $k = 1, \ldots, K$ and $i = 1, \ldots, n_{k+}$. The parameter $\sigma_{k,i}$ is a singular value, $\tilde{S}_{k,i}$ is known at the BS, and $\tilde{Z}_{k,i}$ is CSCG and independent of $\tilde{X}_{k,i}, \tilde{S}_{k,i}$. Thus, one achieves capacity if one achieves the capacity $\log(1 + \sigma_{k,i}^2)$ for each scalar channel (29) in (27), and for the covariance matrices $\mathbf{K}_1, \ldots, \mathbf{K}_K$ of all required rate tuples.

 doi:10.20944/preprints202605.1186.v1



One may write $\underline{Z}_k = \mathbf{Q}_k^{1/2}\underline{W}_k$ where the entries $W_{k,i}$ of $\underline{W}_k$ are i.i.d. CSCG with unit variance. Thus, using (24) and (28), we have

$$\underline{\tilde{Z}}_k = \mathbf{F}_k\left[\left(\sum_{l>k}\mathbf{H}_k\mathbf{P}_l\underline{\tilde{X}}_l\right) + \mathbf{Q}_k^{1/2}\underline{W}_k\right]. \tag{30}$$

Alternatively, there are constants $a_{k,i,l,h}$ and $b_{k,i,h}$ for which

$$\tilde{Z}_{k,i} = \left(\sum_{l>k}\sum_{h=1}^{n_t} a_{k,i,l,h}\tilde{X}_{l,h}\right) + \sum_{h=1}^{n_k} b_{k,i,h}W_{k,h}. \tag{31}$$

We apply the DPC in [57] to each channel (29) by treating the real and imaginary parts of $\tilde{Y}_{k,i}$ as independent with the same channel gain $\sigma_{k,i}$ and noise variance $1/2$. We focus on the real part and abuse notation by using the same symbols as for the complex alphabet channels, i.e., all random variables are now real-valued. Also, the $\tilde{X}_{k,i}$, $\tilde{S}_{k,i}$, and $\tilde{Z}_{k,i}$ are no longer Gaussian, with the exception of the $\tilde{Z}_{K,i}$.

More precisely, consider the real part of sub-channel $k,i$ and define

$$Y := \frac{\tilde{Y}_{k,i}}{\sigma_{k,i}} = X + S + Z \tag{32}$$

where

$$X := \tilde{X}_{k,i}, \quad S := \frac{\tilde{S}_{k,i}}{\sigma_{k,i}}, \quad Z := \frac{\tilde{Z}_{k,i}}{\sigma_{k,i}}. \tag{33}$$

Now code as in Section 2.4, where we identify $D := D_{k,i}$ and the parameters

$$M := M_{k,i}, \quad A := A_{k,i}, \quad \kappa := \kappa_{k,i}, \quad \alpha := \alpha_{k,i}. \tag{34}$$

Note that $Z$ is not necessarily Gaussian. We choose the $D_{k,i}$ across the sub-channels as independent, which makes the $\tilde{X}_{k,i}$ independent for all $k,i$; see Appendix B. Thus, all RVs on the right-hand side of (31) are independent and the density of $\tilde{Z}_{k,i}$ is the convolution of the densities of the $a_{k,i,l,h}\tilde{X}_{l,h}$ and $b_{k,i,h}W_{k,h}$ (more generally, if a $\tilde{X}_{l,h}$ does not have a density, the density of $\tilde{Z}_{k,i}$ is the derivative of the convolution of the distribution functions of the $a_{k,i,l,h}\tilde{X}_{l,h}$ and $b_{k,i,h}W_{k,h}$).

We prove the following results in Section 4 and Section 5.

**Lemma 4** (cf. [59, Lemma 2]). *Consider fixed $A_{k,i}$. In the limit of large $M_{k,i}$, we have $p_{k,i}(x) = q_{k,i}(x)$ for all $x$.*

**Theorem 1** (cf. [59, Theorem 1]). *Consider the DPC scheme in Section 2.4 applied to all $n_{ch}$ sub-channels. For each sub-channel, consider the variables (33)–(34) and the shaping density (18). Given $A$, the parameter $\sigma_q^2$ is chosen so that (19) takes on the value $E_q[X^2] = 1/2$. Let*

$$\underline{R} = (R_{1,1}, \ldots, R_{K,n_{K+}})$$

*be an $n_{ch}$-tuple of achievable rates for the specified parameters*

$$\underline{A} = (A_{1,1}, \ldots, A_{K,n_{K+}}) \text{ and } \underline{M} = (M_{1,1}, \ldots, M_{K,n_{K+}}).$$

 doi:10.20944/preprints202605.1186.v1



*Given $\underline{A}$, there is a sequence of achievable rates $\underline{R}$ satisfying*

$$\lim_{\underline{M}\to\underline{\infty}} R_{k,i} \geq \frac{1}{2}\log\left(2\sigma_q^2 c^2(1+\sigma_{k,i}^2)\right) - \frac{1}{2}\left(1-\frac{1}{2\sigma_q^2}\right)\log e \tag{35}$$

*for all $k,i$. Moreover, for each sub-channel, we have $c\to 1$ and $\sigma_q^2 \to 1/2$ as $\underline{A}\to\underline{\infty}$, yielding*

$$\lim_{\underline{A}\to\underline{\infty}}\lim_{\underline{M}\to\underline{\infty}} R_{k,i} = \frac{1}{2}\log\left(1+\sigma_{k,i}^2\right) \text{ for all } k,i. \tag{36}$$

Note that (35) specifies achievable rates for finite $\underline{A}$. This bound can thus be used for peak power constraints.

## 4. Proof of Lemma 4

Consider $d(x)$ in (15), and recall that $x \bmod \kappa = x - k'\kappa$ for a unique integer $k'$. We have

$$d(x) \overset{(a)}{=} \sum_{k=0}^{M-1} \kappa\, q((x \bmod \kappa + k\kappa) \bmod A) = d(x \bmod \kappa) \tag{37}$$

where step $(a)$ follows because the modulo operator permits summing over any $M$ successive integers. Thus, $d(.)$ is periodic with period $\kappa$. For example, Figure 1 shows $d(x)$ for $A=6$, $M=7,8$, and the truncated Gaussian $q(x)$ in [59] [eq. (11)] with $\sigma_q = 1.8$. Note that $d(.)$ satisfies (37) for general $q(.)$ and not just a truncated Gaussian $q(x)$.

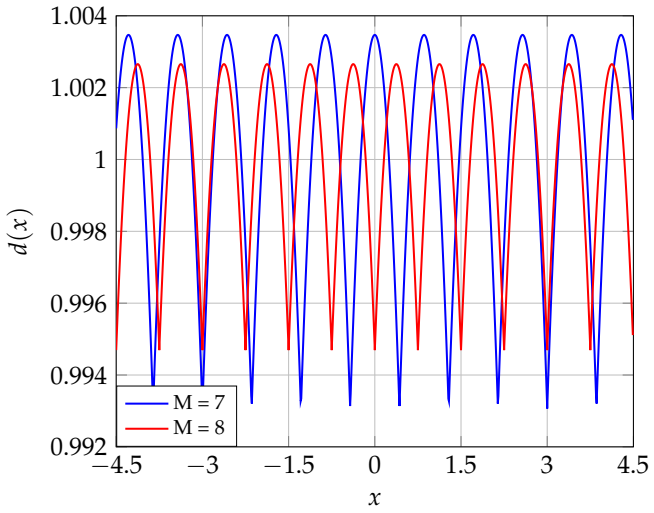


**Figure 1.** $d(x)$ for the shaping density $q(x)$ in (18) with $A=6$ and $\sigma_q = 1.8$, and for $M=7$ ($\kappa \approx 0.857$), and $M=8$ ($\kappa = 0.75$). Observe that $d(x)$ is periodic with period $\kappa$ and $d(x)\approx 1$.

Suppose $q(x)$ is symmetric, which means that $d(x)$ and $p(x) = q(x)/d(x)$ are symmetric. For $x\in[0,\kappa/2)$, we have

$$d(x) = \sum_{k=-\lfloor M/2\rfloor}^{k=\lceil M/2\rceil-1} \kappa\, q(x+k\kappa). \tag{38}$$

Now suppose $q(.)$ is symmetric and unimodal. Applying (7) in Lemma 3, we have $d_{\min} \leq d(x) \leq d_{\max}$ for all $x$, where

$$d_{\min} = 1-\kappa q(0), \quad d_{\max} = 1+\kappa q(0). \tag{39}$$

 doi:10.20944/preprints202605.1186.v1



We thus have $d_{\min}, d_{\max} \to 1$ for $M \to \infty$ and

$$\frac{q(x)}{d_{\max}} \le p(x) \le \frac{q(x)}{d_{\min}} \tag{40}$$

where we assume $M$ is sufficiently large so $d_{\min} > 0$. Figure 2 illustrates the bounds (40) for $A = 6$, $M = 8$, and the truncated Gaussian $q(.)$ in [59, Equation (12)] with $\sigma_q = 1.8$. We compute $d_{\min} \approx 0.82$ and $d_{\max} \approx 1.18$. The bounds are loose because $M$ is relatively small.

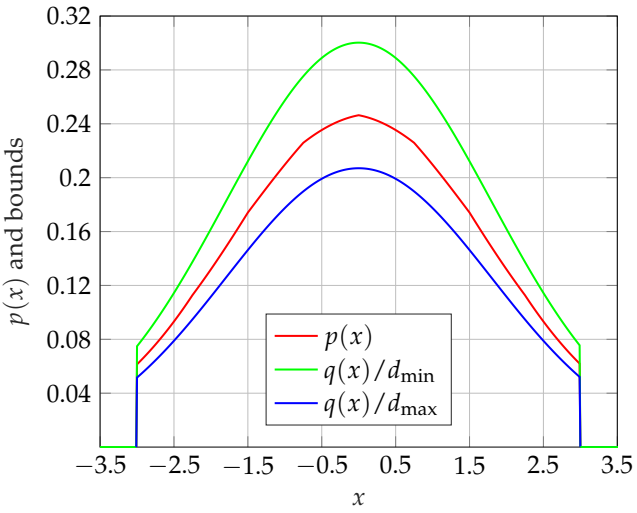


**Figure 2.** Curves of (40) for $A = 6$, $M = 8$, and truncated Gaussian shaping with $\sigma_q = 1.8$.

We derive more properties of $d(.)$ for the truncated Gaussian $q(.)$. The derivative of (38) for $x > 0$ is

$$d'(x) = \sum_{k=-\lfloor M/2 \rfloor}^{\lceil M/2 \rceil - 1} \frac{-(x + k\kappa)\kappa}{\sqrt{2\pi} c \sigma_q^3} e^{-(x+k\kappa)^2/(2\sigma_q^2)} \tag{41}$$

and, by symmetry, we have $d'(x) = -d'(-x)$. Consider first odd $M$ and $x = 0$; the positive and negative summands cancel and $d'(0) = 0$; see Figure 1. For even $M$, there is a summand at $k = -M/2$ but not at $k = M/2$. Thus, $d'(x) > 0$ for small positive $x$. By symmetry, we thus have $d'(x) < 0$ for small negative $x$; see Figure 1.

## 5. Proof of Theorem 1

We use the definitions (33)-(34). Observe that we have

$$P_X = \frac{1}{2}, \quad P_Z = \frac{1}{2\sigma_{k,i}^2}, \quad \alpha = \frac{\sigma_{k,i}^2}{1 + \sigma_{k,i}^2}. \tag{42}$$

The achievable rate of sub-channel $k, i$ is [59, eq. (5)]

$$\begin{aligned} R_{k,i} &= I(U; Y') - I(U; S') \\ &= h(Y') - h(Z' \bmod A) - h(S') + h(X). \end{aligned} \tag{43}$$

We anticipate the result $R_{k,i} \approx \frac{1}{2}\log(1 + \sigma_{k,i}^2)$ as follows. Section 5.2 shows that $h(Y') \approx \log A$, which cancels $h(S') = \log A$ for large $M$. Section 5.1 shows that $h(X) \approx \frac{1}{2}\log(\pi e)$ if $M$ is large and

 doi:10.20944/preprints202605.1186.v1



$\sigma_q^2 \approx 1/2$ and $c \approx 1$. Section 5.3 shows that the latter two approximations are valid for large $A$. Finally, we have the upper bound

$$\begin{aligned} h(Z' \bmod A) &\overset{(a)}{\leq} \frac{1}{2}\log\Big(2\pi e\, \mathrm{E}\Big[(Z' \bmod A)^2\Big]\Big) \\ &\leq \frac{1}{2}\log\Big(2\pi e\, \mathrm{E}\Big[(Z')^2\Big]\Big) \\ &\overset{(b)}{=} \frac{1}{2}\log\left(\frac{\pi e}{1+\sigma_{k,i}^2}\right). \end{aligned} \tag{44}$$

Step $(a)$ follows because a zero-mean Gaussian maximizes entropy under a second-moment constraint, and step $(b)$ follows from $\sigma_{Z'}^2 = \alpha P_Z$; see (17) and (42).

The following two sections first study $P_X$ and $h(X)$ and then $p(y')$ and $h(Y')$.

### 5.1. Power and Differential Entropy of $X$

The power $P_X = \mathrm{E}[X^2]$ is based on an expectation with respect to $p(.)$. Using (2) and (40), for symmetric unimodal $q(.)$, we have

$$\frac{\mathrm{E}_q[X^2]}{d_{\max}} \leq P_X \leq \frac{\mathrm{E}_q[X^2]}{d_{\min}} \tag{45}$$

and thus $P_X \to \mathrm{E}_q[X^2]$ for $M \to \infty$. Using (40), we have (see Appendix C)

$$h(X) \geq \frac{h_q(X)}{d_{\max}} + \log d_{\min} - \left(\frac{1}{d_{\min}} - \frac{1}{d_{\max}}\right)[q(0)\log q(0)]^+ \tag{46a}$$

$$h(X) \leq \frac{h_q(X)}{d_{\min}} + \log d_{\max} + \left(\frac{1}{d_{\min}} - \frac{1}{d_{\max}}\right)[q(0)\log q(0)]^+ \tag{46b}$$

and for $M \to \infty$, we have (see (20))

$$h(X) \to h_q(X) = \frac{1}{2}\log(2\pi e\,\sigma_q^2 c^2) - \frac{1}{2}\left(1 - \frac{1}{2\sigma_q^2}\right)\log e. \tag{47}$$

### 5.2. Density and Differential Entropy of $Y'$

Consider $Y' = (U + Z') \bmod A$ in (16) where $U$ and $Z' = (\alpha - 1)X + \alpha Z$ in (17) are independent. Note that $Z$ is the sum of one Gaussian and several shaped random variables; see (31) and (33).

Suppose $M$ is even; the case where $M$ is odd can be treated similarly. Using symmetric shaping for each sub-channel and Lemma 1, it follows that $p(z') = p(-z')$ and

$$\begin{aligned} p(y') &= \sum_{\ell\in\mathbb{Z}}\sum_{v\in\mathcal{U}} \frac{1}{M} p_{Z'}(y' - v - \ell A) \\ &= \sum_{k\in\mathbb{Z}} \frac{1}{M} p_{Z'}\Big((y' \bmod \kappa) - \Big(\frac{\kappa}{2} - k\kappa\Big)\Big) \qquad (48a) \\ &= p_{Y'}(y' \bmod \kappa) \qquad (48b) \end{aligned}$$

for $y' \in [-A/2, A/2)$ and $p(y') = 0$ otherwise. Hence, $p(y')$ is symmetric and periodic with period $\kappa$ in $[-A/2, A/2)$. Figure 3 illustrates these results for a point-to-point channel with $A = 6$, $M = 4$, $\sigma_q = 1.8$, $\sigma = 1$, $\tilde{Z}$ Gaussian, and $P_{\tilde{Z}} = P_X$. We show next that $p(y') \approx 1/A$ in $[-A/2, A/2)$ if $M$ is large. In fact, increasing $M$ to 8 in Figure 3 already gives $|p(y') - 1/6| \leq 10^{-10}$.

 doi:10.20944/preprints202605.1186.v1



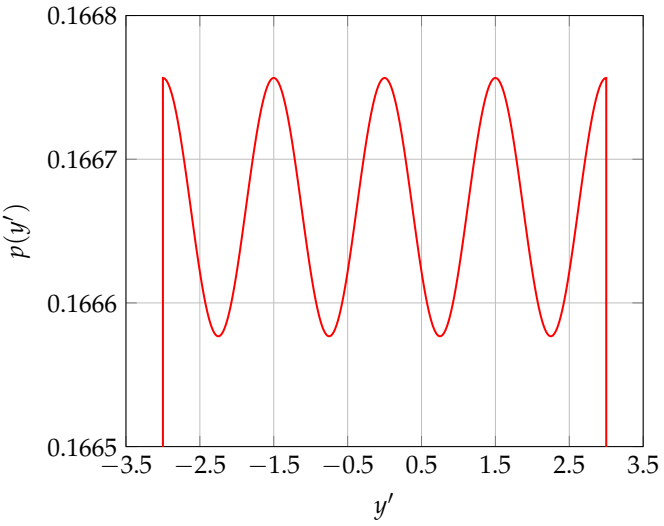


**Figure 3.** $p(y')$ for $A = 6$, $M = 4$, $\sigma_q = 1.8$, $\sigma = 1$, $\tilde{Z}$ Gaussian and $P_{\tilde{Z}} = P_X$. Observe that $p(y')$ is symmetric and periodic with period $\kappa = 1.5$ and $p(y') \approx 1/A$ in $[-A/2, A/2)$.

The density $p(z')$ is the convolution of three classes of densities, namely those of:

- $(\alpha_{k,i} - 1)X_{k,i}$ ,
- the $\alpha_{k,i}\, a_{k,i,l,h} X_{l,h}$ for $l > k$ and all $h$, and
- the $\alpha_{k,i}\, b_{k,i,h} W_{k,h}$ for all $h$.

The densities $q_{k,i}(.)$ and $q_{l,h}(.)$ are chosen to be symmetric unimodal, but the densities of $X_{k,i}$ and $X_{l,h}$ are not necessarily unimodal. However, the latter densities can each be lower-bounded by appropriate $q_{k,i}(.)/d_{\max}$ and $q_{l,h}(.)/d_{\max}$ as in (40), and these bounds are symmetric unimodal. Moreover, the Gaussian densities of the $W_{k,h}$ are symmetric unimodal. Thus, by Lemma 2 and using (39)–(40), we can lower bound $p_{Z'}(.)$ by a symmetric unimodal $\underline{p}(.)$ that approaches $p_{Z'}(.)$ for large $M$. Similarly, we can upper bound $p_{Z'}(.)$ by a symmetric unimodal $\overline{p}(.)$ that approaches $p_{Z'}(.)$ for large $M$. That is, we obtain

$$\underline{p}(z') \le p(z') \le \overline{p}(z') \tag{49}$$

for symmetric unimodal $\underline{p}(z')$ and $\overline{p}(z')$ that both converge to $p(z')$ for large $M$.

For example, point-to-point DPC has the effective noise $Z' = (\alpha - 1)X + \alpha Z$ where $X$ and $Z$ are independent; see (17). We hence have the convolution

$$p_{Z'}(.) = p_{(\alpha-1)X} * p_{\alpha Z}(.). \tag{50}$$

Note that $p(x)$ is symmetric but not necessarily unimodal, and hence $p(z')$ is symmetric but not necessarily unimodal. However, we may bound

$$\begin{aligned} p(z') &= \int_{-A/2}^{A/2} p(x)\, p(z'|x)\, dx \\ &\overset{(a)}{\ge} \int_{-A/2}^{A/2} \frac{q(x)}{d_{\max}}\, p(z'|x)\, dx \\ &\overset{(b)}{=} \frac{1}{d_{\max}} q_{(\alpha-1)X} * p_{\alpha Z}(z') := \underline{p}(z') \end{aligned} \tag{51}$$

where step $(a)$ follows by (40) and step $(b)$ uses the density

$$q_{(\alpha-1)X}(x) = \frac{q(x/(\alpha-1))}{1-\alpha} \tag{52}$$

 doi:10.20944/preprints202605.1186.v1



which is symmetric unimodal in $x$. The function $\underline{p}(z')$ defined in (51) is a convolution of two symmetric unimodal functions. Thus, $\underline{p}(z')$ is symmetric unimodal by Lemma 2. Moreover, one can iterate the above argument for successive DPC with $K > 1$, and thereby obtain a symmetric unimodal $\underline{p}(z')$ with $\underline{p}(z') \leq p(z')$.

Let $\overline{p}(z')$ be the same as $\underline{p}(z')$ but with $d_{\min}$ replacing $d_{\max}$ in (51). We have the bounds (49) where both $\underline{p}(z')$ and $\overline{p}(z')$ are symmetric unimodal. Figure 4 shows the curves for $A = 6$, $M = 8$, the truncated Gaussian $q(.)$ with $\sigma_q = 1.8$, $\sigma = 1$, $P_X = P_Z$, and $Z$ is Gaussian.

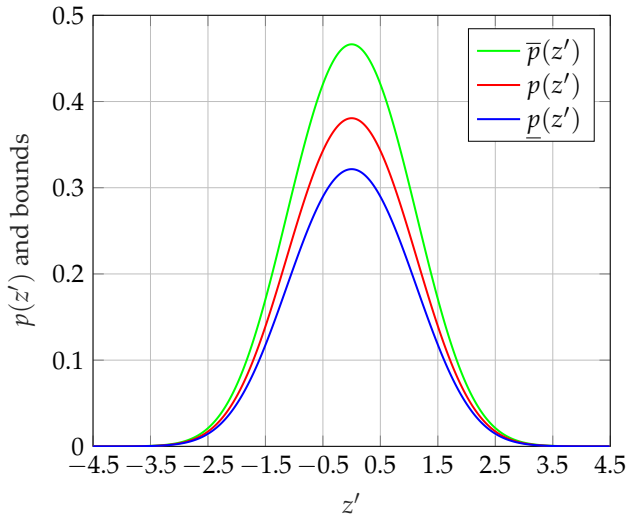


**Figure 4.** $p(z')$ and bounds for $A = 6$, $M = 8$, and truncated Gaussian shaping.

Now substitute $\underline{p}_{Z'}(.)$ and $\overline{p}_{Z'}(.)$ for $f(.)$ in (8) of Lemma 3 to write

$$-\kappa\, \underline{p}_{Z'}(0) \leq \frac{1}{d_{\max}} - \sum_{k\in\mathbb{Z}} \kappa\, \underline{p}_{Z'}(y' + k\kappa) \leq \kappa\, \underline{p}_{Z'}(0) \tag{53a}$$

$$-\kappa\, \overline{p}_{Z'}(0) \leq \frac{1}{d_{\min}} - \sum_{k\in\mathbb{Z}} \kappa\, \overline{p}_{Z'}(y' + k\kappa) \leq \kappa\, \overline{p}_{Z'}(0) \tag{53b}$$

for any $y'$. Define the expressions

$$p_{\min} = \frac{1}{A}\left(\frac{1}{d_{\max}} - \kappa\, \underline{p}_{Z'}(0)\right) \tag{54a}$$

$$p_{\max} = \frac{1}{A}\left(\frac{1}{d_{\min}} + \kappa\, \overline{p}_{Z'}(0)\right). \tag{54b}$$

Applying (53) gives

$$p_{\min} \leq \sum_{k\in\mathbb{Z}} \frac{1}{M} \underline{p}_{Z'}(y' + k\kappa) \overset{(a)}{\leq} p(y') \overset{(b)}{\leq} \sum_{k\in\mathbb{Z}} \frac{1}{M} \overline{p}_{Z'}(y' + k\kappa) \leq p_{\max} \tag{55}$$

for $y' \in [-A/2, A/2)$, where steps $(a)$ and $(b)$ follow by (48a) and (49). For example, for the parameters in Figure 3, we have $p_{\min} \approx 0.1$ and $p_{\max} \approx 0.26$. The bounds are again loose because $M$ is small. However, as $M \to \infty$, we have $p_{\min}, p_{\max} \to 1/A$, and hence $p(y')$ converges to the uniform distribution over $[-A/2, A/2)$. Finally, assuming $M$ is sufficiently large such that $p_{\min} > 0$, we obtain

$$-\log p_{\max} \leq h(Y') \leq -\log p_{\min} \tag{56}$$

which implies $h(Y') \to \log A$ for $M \to \infty$.

 doi:10.20944/preprints202605.1186.v1



*5.3. Achieving Capacity*

For fixed $A$, $\sigma_q$ satisfying $\mathrm{E}_q[X^2] = 1/2$, and $M \to \infty$, we obtain the bound (35) by using (44), (47), and (56). Finally, consider three possible scalings as $A \to \infty$:

- $A/\sigma_q \to \infty$: We have $c = 1 - 2Q\big(A/(2\sigma_q)\big) \to 1$ and $\sigma_q^2 \to 1/2$. We thus obtain the desired (36).
- $A/\sigma_q \to b$ for a constant $b$: This is impossible because $\sigma_q^2 \to \infty$ as $A \to \infty$, which forces $P_X \to \infty$, contradicting the power constraint.
- $A/\sigma_q \to 0$: We have $c = 1 - 2Q\big(A/(2\sigma_q)\big) \to 0$. Consider the Taylor expansions around $x = 0$:

$$Q(x/2) = \frac{1}{2} - \frac{x}{2\sqrt{2\pi}} + \frac{x^3}{48\sqrt{2\pi}} + O(x^5) \tag{57}$$

$$e^{-x^2/8} = 1 - x^2/8 + x^4/128 + O(x^6). \tag{58}$$

From (57), we obtain

$$c = \frac{1}{\sqrt{2\pi}}\frac{A}{\sigma_q} - \frac{1}{24\sqrt{2\pi}}\left(\frac{A}{\sigma_q}\right)^3 + \mathcal{O}\left(\left(\frac{A}{\sigma_q}\right)^5\right). \tag{59}$$

Applying (57)–(59) to (2), we have

$$\begin{aligned}\mathrm{E}[X^2] &= \lim_{A/\sigma_q \to 0} \sigma_q^2 \left[1 - \frac{1 - \frac{1}{8}\left(\frac{A}{\sigma_q}\right)^2 + \frac{1}{128}\left(\frac{A}{\sigma_q}\right)^4}{1 - \frac{1}{24}\left(\frac{A}{\sigma_q}\right)^2}\right] \\ &= \frac{A^2}{12} + \lim_{A/\sigma_q \to 0} \sigma_q^2\, \mathcal{O}\left(\left(\frac{A}{\sigma_q}\right)^4\right).\end{aligned} \tag{60}$$

Thus, we have $\mathrm{E}[X^2] \to \infty$ for $A \to \infty$, contradicting the power constraint.

## 6. Simulation Models and Capacity

Consider a BS serving $K = 2$ UEs, each with $n_k = 1$ antenna. We have $\underline{X}_k = \mathbf{P}_k \tilde{X}_k$ and

$$Y_1 = \mathbf{H}_1 \underline{X}_1 + \mathbf{H}_1 \underline{X}_2 + Z_1 \tag{61a}$$

$$Y_2 = \mathbf{H}_2 \underline{X}_1 + \mathbf{H}_2 \underline{X}_2 + Z_2 \tag{61b}$$

where each $\mathbf{H}_k$ is a column vector of dimension $n_t$, and $\|\mathbf{P}_1\|^2 + \|\mathbf{P}_2\|^2 \le P_X$.

*6.1. DPC and Duality*

The DPC rate bounds (22) simplify to

$$R_1 \le \log\left(1 + \frac{|\mathbf{H}_1\mathbf{P}_1|^2}{1 + |\mathbf{H}_1\mathbf{P}_2|^2}\right) \tag{62a}$$

$$R_2 \le \log\left(1 + |\mathbf{H}_2\mathbf{P}_2|^2\right). \tag{62b}$$

The optimal precoders $\mathbf{P}_k$ can be found by using the dual MAC with output

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{H}_1^\dagger X_1 + \mathbf{H}_2^\dagger X_2 + \underline{Z} \tag{63}$$

 doi:10.20944/preprints202605.1186.v1



and independent inputs $X_1$, $X_2$ with variances $P_1$, $P_2$ that satisfy $P_1 + P_2 \leq P_X$; see [39]. The noise $\underline{Z} = [Z_1 \; Z_2]^T$ has i.i.d. entries with unit variance. Let $\overline{\mathbf{P}}_1^\dagger$ and $\overline{\mathbf{P}}_2^\dagger$ be MAC receiver beamforming vectors that create the scalar outputs

$$\overline{Y}_1 = \overline{\mathbf{P}}_1^\dagger \underline{Y}, \quad \overline{Y}_2 = \overline{\mathbf{P}}_2^\dagger \underline{Y} \tag{64}$$

for UEs 1 and 2, respectively. The optimal $\overline{\mathbf{P}}_k$ are specified by LMMSE estimation; see Appendix D. It remains to determine the powers $P_1$ and $P_2$, which turns out to be a concave optimization problem in general.

Next, following [42], we choose the BC precoding vectors as co-linear with the $\overline{\mathbf{P}}_k$:

$$\mathbf{P}_1 = a_1 \overline{\mathbf{P}}_1, \quad \mathbf{P}_2 = a_2 \overline{\mathbf{P}}_2 \tag{65}$$

for real scalars $a_1, a_2$. Suppose the BC first encodes for UE 1, and the MAC receiver first decodes the UE 2 signal. For the desired rate equivalences, we require the signal-to-interference-and-noise ratios (SINRs) to satisfy

$$\frac{|\mathbf{H}_1\mathbf{P}_1|^2}{1+|\mathbf{H}_1\mathbf{P}_2|^2} = \frac{\left|\mathbf{P}_1^\dagger \mathbf{H}_1^\dagger\right|^2 P_1/a_1^2}{\|\mathbf{P}_1\|^2/a_1^2} \tag{66a}$$

$$|\mathbf{H}_2\mathbf{P}_2|^2 = \frac{\left|\mathbf{P}_2^\dagger \mathbf{H}_2^\dagger\right|^2 P_2/a_2^2}{\|\mathbf{P}_2\|^2/a_2^2 + \left|\mathbf{P}_2^\dagger \mathbf{H}_1^\dagger\right|^2 P_1/a_2^2} \tag{66b}$$

where the left- and right-hand sides are BC and MAC SINRs, respectively. Using $|\mathbf{H}_k\mathbf{P}_l|^2 = |\mathbf{P}_l^\dagger \mathbf{H}_k^\dagger|^2$, we obtain

$$P_1 = \frac{\|\mathbf{P}_1\|^2}{1+|\mathbf{H}_1\mathbf{P}_2|^2} \tag{67a}$$

$$P_2 = \|\mathbf{P}_2\|^2 + \frac{|\mathbf{H}_1\mathbf{P}_2|^2}{1+|\mathbf{H}_1\mathbf{P}_2|^2}\|\mathbf{P}_1\|^2 \tag{67b}$$

and thus

$$P_1 + P_2 = \|\mathbf{P}_1\|^2 + \|\mathbf{P}_2\|^2. \tag{68}$$

For example, consider the channel vectors

$$\mathbf{H}_1 = h_1 \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \mathbf{H}_2 = h_2 \begin{bmatrix} 1 & 1 \end{bmatrix} \tag{69}$$

with $h_1 = 1$, $h_2 = 0.5$, and $P_X = 25$. We use the dual MAC to compute the capacity region shown in Figure 5. This region is the union of two regions corresponding to the two encoding orders, which are shown as dashed lines.

 doi:10.20944/preprints202605.1186.v1



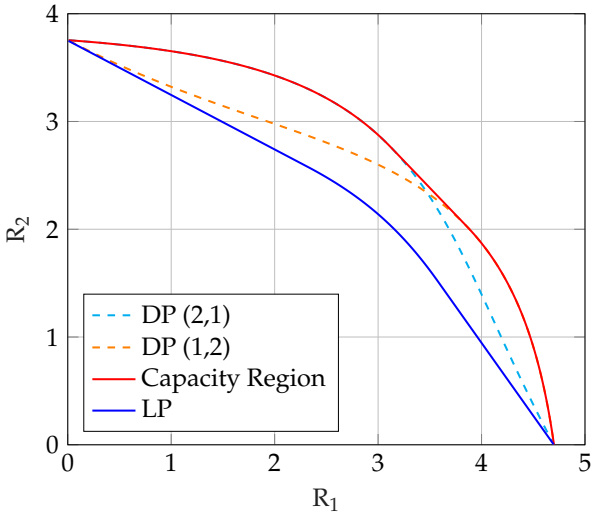


**Figure 5.** Capacity and LP rate regions for the channels (69) with $h_1 = 1$, $h_2 = 0.5$ and $P_X = 25$.

*6.2. Coded LP*

We compare with linear precoding (LP) where the transmit vector is $\mathbf{X} = \sum_{k=1}^{K} \mathbf{X}_k$ with independent $\mathbf{X}_k$. The bound (22) becomes more restrictive:

$$R_k \leq \log\det\left(\mathbf{I} + \frac{\mathbf{H}_k\mathbf{K}_k\mathbf{H}_k^\dagger}{\mathbf{I} + \sum_{l\neq k}\mathbf{H}_k\mathbf{K}_l\mathbf{H}_k^\dagger}\right) \tag{70}$$

where the denominator sum is now over $l \neq k$. For example, for $K = 2$ and $n_k = 1$, the bounds are (see (62))

$$R_1 \leq \log\left(1 + \frac{|\mathbf{H}_1\mathbf{P}_1|^2}{1 + |\mathbf{H}_1\mathbf{P}_2|^2}\right) \tag{71a}$$

$$R_2 \leq \log\left(1 + \frac{|\mathbf{H}_2\mathbf{P}_2|^2}{1 + |\mathbf{H}_2\mathbf{P}_1|^2}\right). \tag{71b}$$

Appendix E reviews a BC-MAC duality for coded LP. The optimal beamforming vectors $\overline{\mathbf{P}}_k$ in (64) for the dual MAC are again specified by LMMSE estimation, and we may again use the relations (65); see [42]. Unfortunately, unlike DPC, the power allocation problem is no longer concave. We thus perform an exhaustive search over a fine quantization of power allocations, which is possible for small numbers of UEs and antennas. The LP region for the channels (69) with $h_1 = 1$, $h_2 = 0.5$, and $P_X = 25$ is shown in Figure 5.

We remark that coded LP can be improved, see Section 1.1. For example, one may use multiple transmission modes, as well as RSMA and SIC; see Appendixes F-G. However, SIC increases UE complexity and latency.

*6.3. Computing Rate Regions*

We study parallel channels for the polar code simulations, e.g., corresponding to tones in an orthogonal frequency-division multiplexing (OFDM) system. A BC with two transmit antennas, one antenna per UE, and two tones may be written as having $n_t = 4$ and $n_1 = n_2 = 2$, where the rows of each $\mathbf{H}_k$ are orthogonal.

 doi:10.20944/preprints202605.1186.v1



For example, consider

$$\mathbf{H}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \end{bmatrix} \tag{72a}$$

$$\mathbf{H}_2 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.6 \end{bmatrix}. \tag{72b}$$

For DPC, we use the dual MAC and optimize weighted sum rates $w_1 R_1 + w_2 R_2$ over

- the power allocation across the parallel MACs;
- the users' transmit powers for the two MACs;
- the MAC decoding orders.

One can apply any convex solver, e.g., gradient descent [86]. For coded LP, one may consider either the BC or dual MAC and optimize the weighted sum rate over

- the power allocation across the parallel MACs;
- the users' transmit powers for the two MACs.

We use the dual MAC and LMMSE estimation, see Section 6.2, and perform an exhaustive search over a fine quantization of power allocations. For problems with more UEs and antennas, one must resort to other methods, e.g., gradient-based methods as in [87].

The capacity and LP rate regions are plotted in Figure 6. The circles show points $(R_1, R_2)$ that maximize the sum rates for DPC and coded LP, namely

$$(R_1, R_2) \approx (7.01, 4.63) \text{ and } (R_1, R_2) \approx (5.62, 4.56). \tag{73}$$

The asterisks show points targeted for the polar code simulations in Section 7, namely

$$(R_1, R_2) \approx (5.61, 3.27) \text{ and } (R_1, R_2) \approx (4.22, 3.19). \tag{74}$$

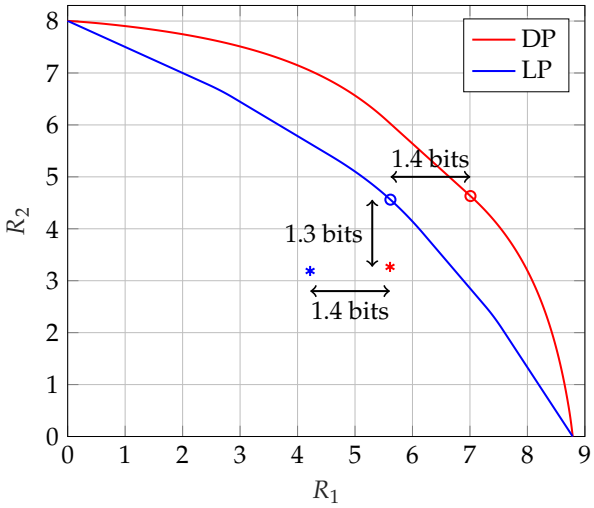


**Figure 6.** Capacity and LP rate regions for the channels (72). Circles show points that maximize the sum rate and are targeted for the polar code simulations. Asterisks show the rates achieved by polar codes of length 256 bits when the target FER for each user is $10^{-3}$.

 doi:10.20944/preprints202605.1186.v1



*6.4. Polar Code Simulations*

We use the channel (72) with $P_X = 50$ to perform polar code simulations described in Section 7. We target the maximum sum rate via the DPC order 1,2, and apply the steps in Section 3.1. The resulting singular values for the sub-channels (29) are

$$\text{UE 1: } \sigma_{1,1} \approx 3.6675, \quad \sigma_{1,2} \approx 2.8131 \tag{75a}$$

$$\text{UE 2: } \sigma_{2,1} \approx 2.2552, \quad \sigma_{2,2} \approx 1.7519. \tag{75b}$$

For coded LP, the optimized precoders $\mathbf{P}_k$ are used to compute the covariance matrices $\mathbf{\Sigma}_k$, and then the steps in Section 3.1 are used to compute the singular values with no interference $\underline{S}_1 = \underline{S}_2 = \underline{0}$ and enhanced noise

$$\check{\underline{Z}}_1 = \mathbf{H}_1 \underline{X}_2 + \underline{Z}_1, \quad \check{\underline{Z}}_2 = \mathbf{H}_2 \underline{X}_1 + \underline{Z}_2. \tag{76}$$

The resulting singular values for the best sum rate are

$$\text{UE 1: } \sigma_{1,1} \approx 2.8635, \quad \sigma_{1,2} \approx 2.0786 \tag{77a}$$

$$\text{UE 2: } \sigma_{2,1} \approx 2.2722, \quad \sigma_{2,2} \approx 1.6843. \tag{77b}$$

## 7. Short Polar Codes

Let $N = 2^n$ for a positive integer $n$. The $N \times N$ polar encoding matrix is

$$\mathbf{G}_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n}, \quad n = \log_2 N. \tag{78}$$

where $\otimes n$ denotes an $n$-fold Kronecker product. An $N$-dimensional bit vector $\underline{b}$ is encoded as $\underline{c}^T = \underline{b}^T \mathbf{G}_N$. The recursive block structure of $\mathbf{G}_N$ yields

$$\begin{aligned} \underline{c}^T &= \underline{b}^T \begin{bmatrix} \mathbf{G}_{N/2} & 0 \\ \mathbf{G}_{N/2} & \mathbf{G}_{N/2} \end{bmatrix} \\ &= [\underbrace{\underline{b}_{1:N/2}^T \mathbf{G}_{N/2} \oplus \underline{b}_{N/2+1:N}^T \mathbf{G}_{N/2}}_{\underline{c}_1^T}, \underbrace{\underline{b}_{N/2+1:N}^T \mathbf{G}_{N/2}}_{\underline{c}_2^T}] \end{aligned} \tag{79}$$

and thus $\underline{b}_{1:N/2}^T \mathbf{G}_{N/2} = \underline{c}_1^T \oplus \underline{c}_2^T$ and $\underline{c}_2^T$ are the codewords obtained by encoding the first and second halves of $\underline{b}^T$, respectively. Thus, $\underline{c}^T$ is constructed by first encoding the subblocks $\underline{b}_{1:N/2}^T$ and $\underline{b}_{N/2+1:N}^T$, and then applying a polarization transform to combine these subcodes. For example, Figure 7 shows that the encoder can first map $\underline{b}_{1:4}$ and $\underline{b}_{5:8}$ to $\underline{c}_{1:4} \oplus \underline{c}_{5:8}$ and $\underline{c}_{5:8}$, respectively, and then apply one level of the polarization transform ($n = 1$) to the corresponding sub-codeword elements. We exploit this structure to construct a long polar code across two sub-channels of a UE by using two separate polar sub-codes.

 doi:10.20944/preprints202605.1186.v1



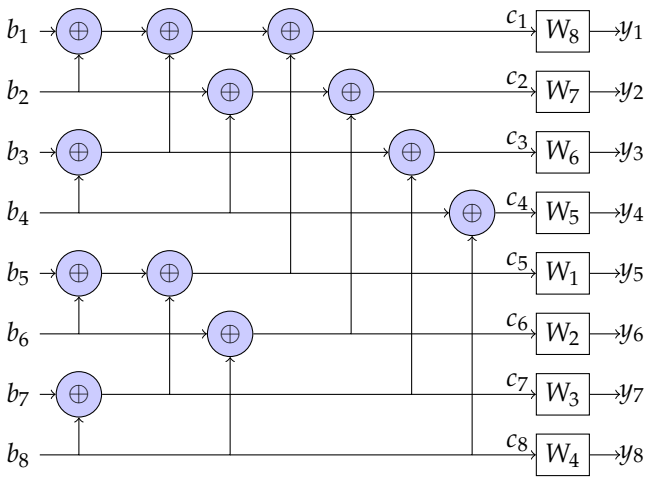


**Figure 7.** Polar encoding for $N = 8$.

### *7.1. Shaped Multi-Level Coding*

There are two variants of shaping with polar codes: all coded bits are assigned the same or different *a priori* distributions. The former approach applies to time-invariant coding, and the latter to time-varying coding. An example of a time-varying coding is DPC, because the interference varies with each coded bit.

We use shaped multi-level coding (MLC) to map message bits to ASK symbols. A simple approach uses separate polar codes for each sub-channel. However, at short block lengths, performance improves by using a single, long multi-level polar code across all sub-channels of a user.

Another effective approach leverages the different singular values in (29), to enhance polarization [60,72,88]. Let $I(W_i)$, $i \in \{1, \ldots, N\}$, denote the mutual information of the polar code bit channels, ordered such that $I(W_i) \leq I(W_j)$ for $i \leq j$. Further, let $\underline{W}^-$ and $\underline{W}^+$ denote the indices of the least reliable $N/2$ and most reliable $N/2$ code bit channels, respectively, with $\underline{W}^-$ in ascending order and $\underline{W}^+$ in descending order. By matching the corresponding elements of $\underline{W}^+$ and $\underline{W}^-$ in the last polarization stage, the FER performance of the polar code can be improved. For example, in Figure 7, the code bits transmitted on $W_1, W_2, W_3, W_4$ are matched in the last polarization stage with the code bits transmitted on $W_8, W_7, W_6, W_5$, respectively.

### *7.2. Polar Coded Modulation*

The BS sends 64 symbols $\underline{x}(t)$, $t = 1, \ldots, 64$. Alternatively, for our example channel (72), it sends 64 symbols over each of the $n_{\text{ch}} = 4$ parallel channels with inputs

$$\tilde{x}_{1,1}(t),\ \tilde{x}_{1,2}(t) \text{ for UE 1}$$
$$\tilde{x}_{2,1}(t),\ \tilde{x}_{2,2}(t) \text{ for UE 2.}$$

We choose the $\tilde{x}_{1,1}(t)$ and $\tilde{x}_{1,2}(t)$ in DPC to have a 256-QAM alphabet, while all other $\tilde{x}_{k,i}(t)$ have a 64-QAM alphabet. These alphabets are split into two real alphabets: 16-ASK and 8-ASK. There are, therefore, 128 real symbols $\tilde{x}_{k,i,\mathbb{R}}(t')$ for fixed $i, k$ and $t' = 1, \ldots, 128$. We map these 128 symbols to the complex QAM symbols via

$$\tilde{x}_{k,i}(t) = \tilde{x}_{k,i,\mathbb{R}}(2t-1) + j\,\tilde{x}_{k,i,\mathbb{R}}(2t) \tag{80}$$

for $t = 1, \ldots, 64$. The precoded transmit vectors are

$$\underline{x}_k(t) = \mathbf{P}_k \begin{bmatrix} \tilde{x}_{k,1}(t) \\ \tilde{x}_{k,2}(t) \end{bmatrix} \tag{81}$$

 doi:10.20944/preprints202605.1186.v1



for $k = 1, 2$ and $t = 1, \ldots, 64$. The BS in DPC thus sends $64 \cdot \log_2(256) = 512$ and $64 \cdot \log_2(64) = 384$ coded bits over the sub-channels of UE 1 and UE 2, respectively, corresponding to a total of 1024 and 768 coded bits for UE 1 and UE 2, respectively. The BS uses MLC with 4 or 3 levels across sub-channel pairs, and each level has 256 coded bits.

Consider DPC with the precoding order $k = 1, 2$. The message bits of UE 1 are encoded directly into the 256 real 16-ASK symbols in the order

$$\tilde{x}_{1,1,\mathbb{R}}(1), \ldots, \tilde{x}_{1,1,\mathbb{R}}(128), \tilde{x}_{1,2,\mathbb{R}}(1), \ldots, \tilde{x}_{1,2,\mathbb{R}}(128)$$

by using MLC with 4 levels and a length-256 polar code for each level. From (80) and (81), the first and second halves of the polar code bits protecting the bit levels of $\underline{\tilde{x}}_{1,\mathbb{R}}$ are transmitted over sub-channels with channel coefficients $\sigma_{1,1}$ and $\sigma_{1,2}$, respectively; see (27).

Let $\mathbf{1}$ be the 128-dimensional vector with 128 ones. The encoding parameters for the two sub-channels of UE 2 can be interpreted as the 256-dimensional vectors

$$\underline{\alpha}_2 = [\alpha_{2,1}, \alpha_{2,2}]^T \otimes \mathbf{1} \tag{82}$$

$$\underline{A}_2 = [A_{2,1}, A_{2,2}]^T \otimes \mathbf{1} \tag{83}$$

$$\underline{\sigma}^2_{\tilde{X}_2} = [\sigma^2_{\tilde{X}_{2,1}}, \sigma^2_{\tilde{X}_{2,2}}]^T \otimes \mathbf{1} \tag{84}$$

$$\underline{\sigma}^2_{Z'_2} = [\sigma^2_{Z'_{2,1}}, \sigma^2_{Z'_{2,2}}]^T \otimes \mathbf{1}. \tag{85}$$

For UE 2, the BS treats $\underline{\tilde{s}}_2 = \mathbf{F}_2\mathbf{H}_2\underline{X}_1$ as interference; see (30). To encode to 256 real 8-ASK symbols $\underline{u}_2$, the BS applies

$$s'_{2,i,\mathbb{R}}(t') = (\alpha_{2,i}\tilde{s}_{2,i,\mathbb{R}}(t') + d_{2,i}(t')) \bmod A_{2,i} \tag{86}$$

for $i = 1, 2$ and $t' = 1, \ldots, 128$. To simplify notation, write the 256 coded bit indexes as $i = 1, \ldots, 256$ and define

$$\tilde{s}_{2,\mathbb{R}}(1), \ldots, \tilde{s}_{2,\mathbb{R}}(256) := \tilde{s}_{2,1,\mathbb{R}}(1), \ldots, \tilde{s}_{2,1,\mathbb{R}}(128), \tilde{s}_{2,2,\mathbb{R}}(1), \ldots, \tilde{s}_{2,2,\mathbb{R}}(128). \tag{87}$$

Note that we used the same ordering as above for $\tilde{x}_{1,i,\mathbb{R}}(t')$. Similarly write

$$s'_{2,\mathbb{R}}(1), \ldots, s'_{2,\mathbb{R}}(256) := s'_{2,1,\mathbb{R}}(1), \ldots, s'_{2,1,\mathbb{R}}(128), s'_{2,2,\mathbb{R}}(1), \ldots, s'_{2,2,\mathbb{R}}(128). \tag{88}$$

The BS performs shaping by computing the log-likelihood ratio (LLR)

$$L\Big(c^l(i)|c^1(i), \ldots, c^{l-1}(i), \tilde{s}_{2,\mathbb{R}}(i)\Big) = \frac{\sum\limits_{u_2 \in \mathcal{V}_2\left(c^l(i)=0|c^1(i),\ldots,c^{l-1}(i)\right)} \exp\left(\frac{-\left(\left(u_2 - s'_{2,\mathbb{R}}(i)\right) \bmod A_2(i)\right)^2}{2\sigma_{\tilde{X}_2}(i)^2}\right)}{\sum\limits_{u_2 \in \mathcal{V}_2\left(c^l(i)=1|c^1(i),\ldots,c^{l-1}(i)\right)} \exp\left(\frac{-\left(\left(u_2 - s'_{2,\mathbb{R}}(i)\right) \bmod A_2(i)\right)^2}{2\sigma_{\tilde{X}_2}(i)^2}\right)} \tag{89}$$

of the coded bit $c^l(i)$ at the $l$th bit-level, $i = 1, \ldots, 256$, $l = 1, 2, 3$, where we used the shaping density (18) and similar notation as (82)-(85). A polar decoder treats both the frozen and message bits as frozen to compute the shaping bits. The set $\mathcal{V}_2\left(c^l(i)|c^1(i), \ldots, c^{l-1}(i)\right)$ in (89) is the subset of 8-ASK symbols for which the first $l$ labeling bits are $c^1(i), \ldots, c^l(i)$.

The encoding output is 256 real 8-ASK symbols in the order

$$u_{2,1,\mathbb{R}}(1), \ldots, u_{2,1,\mathbb{R}}(128), u_{2,2,\mathbb{R}}(1), \ldots, u_{2,2,\mathbb{R}}(128).$$

 doi:10.20944/preprints202605.1186.v1



Using these symbols, the BS computes

$$\tilde{x}_{2,i,\mathbb{R}}(t') = \left(u_{2,i,\mathbb{R}}(t') - s'_{2,i,\mathbb{R}}(t')\right) \bmod A_{2,i} \tag{90}$$

for $i = 1, 2$ and $t' = 1, \ldots, 128$. These symbols are then mapped to the $\tilde{x}_{2,i}(t)$ according to (80), and to the $\underline{x}_2(t)$ according to (81).

To decode, the UEs compute the LLRs

$$L(c^l(i)|\hat{c}^1(i), \ldots, \hat{c}^{l-1}(i), y'(i)) = \frac{\sum\limits_{u_2 \in \mathcal{V}_2\left(c^l(i)=0|\hat{c}^1(i),\ldots,\hat{c}^{l-1}(i)\right)} \sum\limits_{m=-\infty}^{\infty} \exp\left(\frac{-(y'(i)-u_2+mA_2(i))^2}{2\sigma_{Z'_2}(i)^2}\right)}{\sum\limits_{u_2 \in \mathcal{V}_2\left(c^l(i)=1|\hat{c}^1(i),\ldots,\hat{c}^{l-1}(i)\right)} \sum\limits_{m=-\infty}^{\infty} \exp\left(\frac{-(y'(i)-u_2+mA_2(i))^2}{2\sigma_{Z'_2}(i)^2}\right)} \tag{91}$$

for their bits $c^l(i)$, where the inner sum in (91) is due to the modulo operator in (13).

### 7.3. Simulation Results

The modulo operator limits the dynamic range at the UEs and BS, especially in the presence of interference. However, the dispersion increases at low rates, and this increases the FER at short block length [59, Figure 5 and Figure 6]. We therefore use a modulo operator for DPC only, i.e., we do not use a modulo interval for UE 1 under DPC, or for either UE in the LP scheme. Instead, we shaped with a sampled Gaussian distribution proportional to $e^{-\nu x^2}$ with $\nu = 0.494$. The ASK spacings were $\kappa = 0.93$ and 0.415 for $M = 8$ and 16, respectively. For UE 2 with DPC, we cancel interference by shaping according to (12), and use $A = 6.64$ and $\sigma_q = 1$ for both sub-channels.

We allocate an asymptotically optimal fraction of sub-channels to shaping at each bit level; see [57, Sec. IV]. For the two polar subcodes corresponding to bit level $j = 1, 2, 3, 4$, the mutual information is computed from this quantity and translated proportionally into the fraction of shaping bits assigned to each subcode. Within each bit level, the most reliable sub-channels are allocated to the shaping bits, the next most reliable sub-channels to the message bits, and the remaining sub-channels to the frozen bits. The sub-channel reliability order specified in 5G NR [89] is used. We employ 6 CRC bits, which are transmitted on the reliable subcode of the last bit level. At this level, after allocating the shaping bits, the most reliable remaining sub-channels are assigned to the CRC bits. Natural labeling maps bits to modulation symbols.

After allocating the shaping and CRC bits, the assignment of message bits to the bit levels and their subcodes is optimized to maximize the achievable information rate (AIR) at a target FER of $\epsilon = 10^{-3}$ for a total transmit power $P_X = 50$. The resulting parameters are given in Table 1, where $s^i$ and $b^i$ are the numbers of shaping and information bits at bit level $i$, respectively, and "sub" indicates the subcode index. The parameters in Table 1 would make the variances of the $\tilde{X}_{k,i}$ slightly larger than one due to finite block length shaping, so we normalize so that the $\tilde{X}_{k,i}$ have unit variance.

**Table 1.** Simulation Parameters.

| | | sub | $s^1$ | $s^2$ | $s^3$ | $s^4$ | $b^1$ | $b^2$ | $b^3$ | $b^4$ |
|---|---|---|---|---|---|---|---|---|---|---|
| DP | UE 1 | 1 | 0 | 0 | 0 | 55 | 0 | 5 | 97 | 73 |
| | no modulo | 2 | 0 | 0 | 10 | 110 | 3 | 52 | 117 | 12 |
| | UE 2 | 1 | 0 | 0 | 55 | - | 0 | 35 | 68 | - |
| | modulo | 2 | 0 | 10 | 110 | - | 6 | 88 | 12 | - |
| LP | UE 1 | 1 | 0 | 0 | 55 | - | 0 | 61 | 72 | - |
| | no modulo | 2 | 0 | 10 | 110 | - | 18 | 107 | 12 | - |
| | UE 2 | 1 | 0 | 0 | 55 | - | 0 | 34 | 67 | - |
| | no modulo | 2 | 0 | 10 | 110 | - | 5 | 86 | 12 | - |

 doi:10.20944/preprints202605.1186.v1



The BS uses successive cancellation (SC) decoding for shaping, and the UEs use SC list decoding with list passing [57] across subcodes and bit levels for channel decoding.

Figure 8 plots the FER versus SNR $= P_X/4$. The UE codes are designed at $P_X = 50$, or SNR $\approx 11$ dB, and achieve the target error rate $\epsilon = 10^{-3}$. The precoders are optimized for each SNR value. Overall, DPC transmits a sum of 568 bits to both UEs over 64 channel uses, with a sum rate of $R_{\mathrm{DP}} \approx 8.88$ bits per channel use (bpcu); see Figure 6. Coded LP transmits 474 bits, yielding a total rate of $R_{\mathrm{LP}} \approx 7.41$ bpcu, so DPC gains 1.47 bpcu. The asymptotic rates at $P_X = 50$ for DPC and coded LP are approximately 11.64 bpcu and 10.18 bpcu, respectively, so DPC gains 1.46 bpcu.

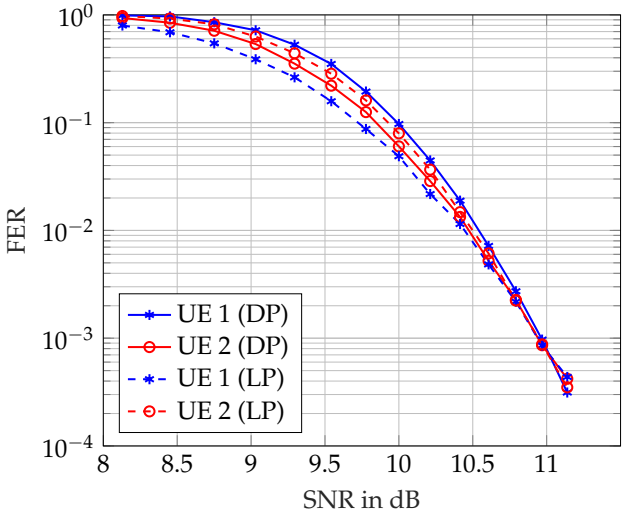


**Figure 8.** FER vs. SNR for DPC and coded LP with multi-level polar codes. The code parameters are optimized at 11 dB.

## 8. Conclusions

We showed that scalar DPC based on modulo operations, probabilistic shaping, and dithering can approach any rate tuple in the capacity region of a complex-alphabet MIMO BC with CSCG noise and one message per receiver. The encoder and decoder complexities are similar to those of coded LP. Short polar code simulations demonstrate the rate gains achieved through DPC relative to coded LP. Future work may derive better bounds on the achievable rates, e.g., by using informational divergence rather than bounds based on symmetric unimodal functions.

**Author Contributions:** Conceptualization, M.Y.Ş. and G.K.; methodology, M.Y.Ş., G.K., and R.B.; software, M.Y.Ş.; validation, M.Y.Ş., G.K., S.S., R.B., and W.X.; formal analysis, M.Y.Ş. and G.K.; investigation, M.Y.Ş.; resources, R.B. and W.X.; data curation, M.Y.Ş.; writing—original draft preparation, M.Y.Ş. and G.K.; writing—review and editing, M.Y.Ş., G.K., S.S., R.B., and W.X.; visualization, M.Y.Ş. and G.K.; supervision, G.K., R.B., and W.X.; project administration, M.Y.Ş. and G.K.; funding acquisition, G.K., S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the German Research Foundation (DFG) via the German-Israeli Project Cooperation (DIP) under Projects KR 3517/13-1 and SH 1937/1-1.

**Conflicts of Interest:** The authors declare no conflict of interest.

 doi:10.20944/preprints202605.1186.v1



## Appendix A. Proof of Lemmas

### *Appendix A.1. Proof of Lemma 1*

If $f(.)$ and $g(.)$ are both symmetric then

$$\begin{aligned} f * g(-x) &= \int_{\mathbb{R}} f(y)\, g(-x-y)\, dy \\ &\overset{(a)}{=} \int_{\mathbb{R}} f(-y)\, g(x+y)\, dy \\ &\overset{(b)}{=} \int_{\mathbb{R}} f(\tilde{y})\, g(x-\tilde{y})\, d\tilde{y} = f * g(x) \end{aligned} \tag{A1}$$

where step $(a)$ follows by symmetry and step $(b)$ follows by substituting $\tilde{y} = -y$.

### *Appendix A.2. Proof of Lemma 2*

Unimodality implies that $p(.)$ can have a Dirac-delta component $\delta(.)$ at $x = 0$ only, i.e., one may write

$$p(.) = c_p \delta(.) + f(.) \tag{A2}$$

for a constant $c_p$ satisfying $0 \leq c_p \leq 1$, and for a non-negative, symmetric, unimodal $f(.)$ without $\delta(.)$ components. We may assume the derivative $f'(.)$ exists almost everywhere [85, Prop. 2.1]. The $x$ where $f'(x)$ does not exist include "jumps" in $f(.)$. For example, a negative "jump" at $x \geq 0$ from $f(x) = a$ to $f(x+\epsilon) = b$, where $\epsilon$ is a vanishing positive number and $a > b \geq 0$, becomes a $(b-a)\delta(.-x)$ component in $f'(.)$.

As in (A2), consider $q(.) = c_q \delta(.) + g(.)$. We compute

$$p * q(x) = c_p c_q \delta(x) + c_p g(x) + c_q f(x) + f * g(x) \tag{A3}$$

which is symmetric; see Lemma 1. The first three summands in (A3) are unimodal, so it remains to show that $f * g(.)$ is unimodal. Taking the derivative for $x \geq 0$, we have

$$\begin{aligned} (f * g)'(x) &= \int_{-\infty}^{\infty} f(\tilde{y})\, g'(x-\tilde{y})\, d\tilde{y} \\ &\overset{(a)}{=} \int_{0}^{\infty} [f(|x-y|) - f(x+y)]\, g'(y)\, dy \end{aligned} \tag{A4}$$

where step $(a)$ follows by substituting $y = x - \tilde{y}$ and because $f(.)$ and $g(.)$ are symmetric. But the term in square brackets is non-negative because $|x-y| \leq x+y$ and $f(.)$ is symmetric unimodal. We also have $g'(y) \leq 0$ for $y \geq 0$ because $g(.)$ is symmetric unimodal.

### *Appendix A.3. Proof of Lemma 3*

The sum in (7) is a Riemann sum, so we use the left and right rules of Riemann summation. Figure A1 shows a symmetric unimodal $f(.)$ for $\kappa = 1$ (here $f(.)$ is a truncated Gaussian density). The sample points for $x = 0.2$ are located at the top-left corner of each of the six bars. The area of each bar is one of the summands in (7); the area of the green bar is at most $\kappa f(0)$. Figure A2 shifts the red bars of positive sampling points to the left by $\kappa$, and we see that the area of the five red bars is less than the integral in (7). This can be done for any $x$ with $0 \leq x < \kappa/2$, so the sum in (7) is at most the integral plus $\kappa f(0)$.

 doi:10.20944/preprints202605.1186.v1



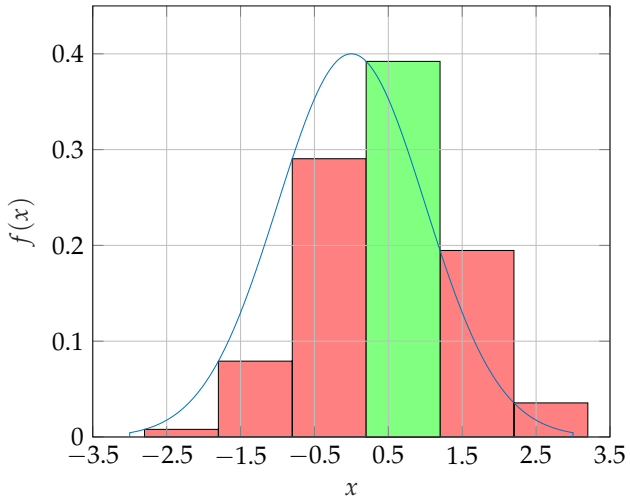


**Figure A1.** $f(x)$ for $A = M = 6$. The sampling points are shifted by $x = 0.2$ and are located at the top left corner of each of the six bars.

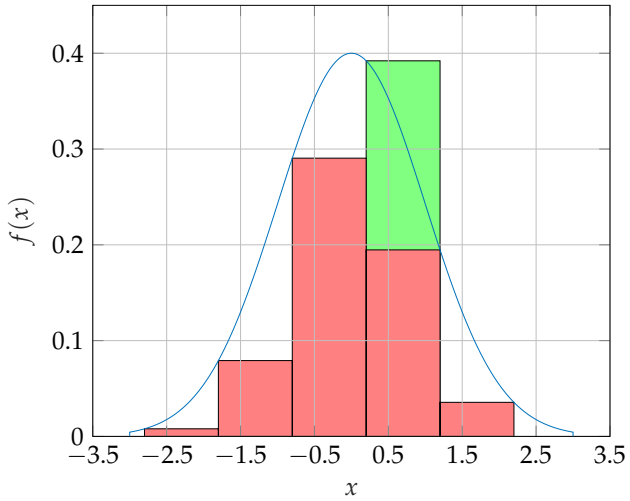


**Figure A2.** $f(x)$ for $A = M = 6$. The area of the red bars is less than the area of $f(x)$, which is here 1.

Similarly, Figure A3 shifts the red bars of negative sampling points to the left by $\kappa$ so they lie above $f(x)$. If we add one (blue) bar of area $\kappa f(0)$, then the sum of the areas of the seven bars is greater than the integral in (7). This proves (7). To prove (8), we perform similar steps with a countable number of bars and appropriate left and right shifts. Note that we may restrict attention $0 \leq x < \kappa/2$ by symmetry.

 doi:10.20944/preprints202605.1186.v1



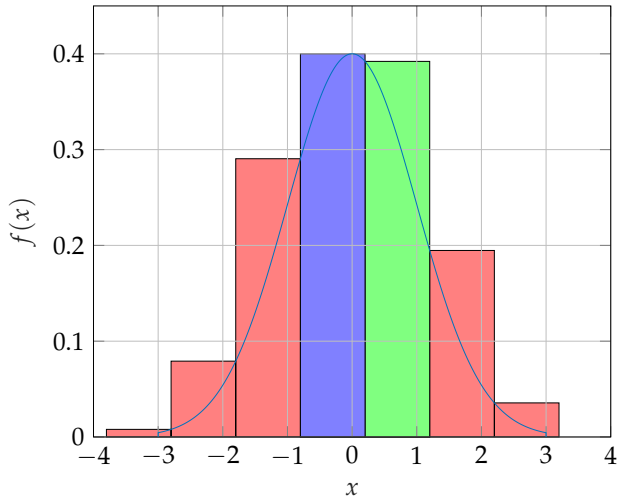


**Figure A3.** $f(x)$ for $A = M = 6$. The area of the red, green, and blue bars is greater than the area of $f(x)$, which is here 1.

## Appendix B. Independence of the Channel Inputs $\tilde{X}_{k,i}$

Let $\underline{\tilde{X}}^c_{k,i}$ be the vector of all $\tilde{X}_{l,m}$ except $\tilde{X}_{k,i}$. Note from (24) that $\tilde{S}_{k,i}$ is a function of $\underline{\tilde{X}}^c_{k,i}$. Moreover, the chain $\underline{\tilde{X}}^c_{k,i} - \tilde{S}'_{k,i} - U_{k,i}$ is Markov because $U_{k,i}$ is chosen using $S'_{k,i}$ as shown in (12). Suppose the dithers $D_{k,i}$ are mutually independent, and consider the identities

$$\begin{aligned} p(\tilde{x}_{k,i}|\underline{\tilde{x}}^c_{k,i}) &= \int_{-A_{k,i}/2}^{A_{k,i}/2} p(d_{k,i})\, p(\tilde{x}_{k,i}|\tilde{s}_{k,i}, d_{k,i})\, dd_{k,i} \\ &\overset{(a)}{=} \sum_{u\in\mathcal{U}_{k,i}} \frac{1}{A_{k,i}} \underbrace{P_{U_{k,i}|S'_{k,i}}\big(u\big|(u-\tilde{x}_{k,i}) \bmod A\big)}_{\overset{(b)}{=}\, \kappa_{k,i} q(\tilde{x}_{k,i})/d(\tilde{x}_{k,i})} \\ &\overset{(c)}{=} p(\tilde{x}_{k,i}) \end{aligned} \tag{A5}$$

where step $(a)$ follows because $\tilde{X}_{k,i}$ is a discrete RV given $S'_{k,i} = s'_{k,i}$ (see (13)), step $(b)$ follows by (12), (14), and (15) and step $(c)$ follows because $\kappa_{k,i} = A_{k,i}/M_{k,i}$ and $p(\tilde{x}_{k,i}) = q(\tilde{x}_{k,i})/d(\tilde{x}_{k,i})$ (see (14)). Thus, the channel inputs $\tilde{X}_{k,i}$ in (29) are mutually independent.

## Appendix C. Entropy Bounds

To prove (46a), let $\mathcal{I}_1 = \{x : q(x) < 1\}$ and $\mathcal{I}_2 = \{x : q(x) \geq 1\}$. Observe that the length of $\mathcal{I}_2$ is at most one since $q(.)$ is a density. We bound

$$\begin{aligned} h(X) &\geq \int_{\mathbb{R}} -p(x) \log \frac{q(x)}{d_{\min}}\, dx \\ &= \log d_{\min} - \int_{\mathcal{I}_1} p(x) \log q(x)\, dx - \int_{\mathcal{I}_2} p(x) \log q(x)\, dx \\ &\geq \log d_{\min} - \int_{\mathcal{I}_1} \frac{q(x)}{d_{\max}} \log q(x)\, dx - \int_{\mathcal{I}_2} \frac{q(x)}{d_{\min}} \log q(x)\, dx \\ &\quad \pm \int_{\mathcal{I}_2} \frac{q(x)}{d_{\max}} \log q(x)\, dx \\ &= \log d_{\min} + \frac{h_q(x)}{d_{\max}} - \left(\frac{1}{d_{\min}} - \frac{1}{d_{\max}}\right) \int_{\mathcal{I}_2} q(x) \log q(x)\, dx \\ &\geq \log d_{\min} + \frac{h_q(x)}{d_{\max}} - \left(\frac{1}{d_{\min}} - \frac{1}{d_{\max}}\right) [q(0) \log q(0)]^+ \end{aligned} \tag{A6}$$

 doi:10.20944/preprints202605.1186.v1



where the final step follows because $1 \le q(x) \le q(0)$ for $\mathcal{I}_2 \neq \emptyset$ and $x \in \mathcal{I}_2$, and because $\int_{\mathcal{I}_2} dx \le 1$. Similar steps prove (46b) and (56).

## Appendix D. MAC Beamformers

The best MAC receiver beamformers are LMMSE filters. Let $\mathbf{K}_{\underline{Y}}$ be the covariance matrix of $\underline{Y}$. The LMMSE operation is the conditional expectation $\mathrm{E}\left[X_k|\underline{Y}\right] = \overline{\mathbf{P}}_k^{\dagger}\,\underline{Y}$ with the LMMSE filters

$$\overline{\mathbf{P}}_k^{\dagger} = \mathrm{E}\left[X_k\,\underline{Y}^{\dagger}\right]\mathbf{K}_{\underline{Y}}^{-1}. \tag{A7}$$

For example, for the channels (69), if the receiver decodes UE 1's signal first, we have

$$\begin{aligned}\overline{\mathbf{P}}_1^{\dagger} &= [P_1 h_1\ 0]\begin{bmatrix} P_1h_1^2 + P_2h_2^2 + 1 & P_2h_2^2 \\ P_2h_2^2 & P_2h_2^2 + 1\end{bmatrix}^{-1} \\ &\propto \begin{bmatrix} P_2h_2^2 + 1 & -P_2h_2^2\end{bmatrix}.\end{aligned} \tag{A8}$$

Similarly, decoding UE 2's message first, we have

$$\begin{aligned}\overline{\mathbf{P}}_2^{\dagger} &= h_2P_2\,[1\ 1]\begin{bmatrix} 1 + P_1h_1^2 + P_2h_2^2 & P_2h_2^2 \\ P_2h_2^2 & 1 + P_2h_2^2\end{bmatrix}^{-1} \\ &\propto \begin{bmatrix} 1 & 1 + P_1h_1^2\end{bmatrix}.\end{aligned} \tag{A9}$$

Observe that $\overline{\mathbf{P}}_1$ is matched to $\mathbf{H}_1$ for small $P_2$, and is orthogonal to $\mathbf{H}_2$ for large $P_2$. Moreover, the LMMSE filter for transmitter 2, after removing the interference of transmitter 1, is matched to $\mathbf{H}_2$. Similarly, $\overline{\mathbf{P}}_2$ is matched to $\mathbf{H}_2$ for small $P_2$, and is orthogonal to $\mathbf{H}_1$ for large $P_2$. The LMMSE filter for transmitter 1, after removing the interference of transmitter 2, is matched to $\mathbf{H}_1$.

## Appendix E. Dual MAC for coded LP

Consider the dual MAC, beamforming, and scaling in (63)–(65). The new SINR equivalences are (cf. (66) and [42])

$$\frac{|\mathbf{H}_1\mathbf{P}_1|^2}{1+|\mathbf{H}_1\mathbf{P}_2|^2} = \frac{\left|\mathbf{P}_1^{\dagger}\mathbf{H}_1^{\dagger}\right|^2 P_1/a_1^2}{\|\mathbf{P}_1\|^2/a_1^2 + \left|\mathbf{P}_1^{\dagger}\mathbf{H}_2^{\dagger}\right|^2 P_2/a_1^2} \tag{A10a}$$

$$\frac{|\mathbf{H}_2\mathbf{P}_2|^2}{1+|\mathbf{H}_2\mathbf{P}_1|^2} = \frac{\left|\mathbf{P}_2^{\dagger}\mathbf{H}_2^{\dagger}\right|^2 P_2/a_2^2}{\|\mathbf{P}_2\|^2/a_2^2 + \left|\mathbf{P}_2^{\dagger}\mathbf{H}_1^{\dagger}\right|^2 P_1/a_2^2}. \tag{A10b}$$

These equations simplify to

$$\frac{1}{1+|\mathbf{H}_1\mathbf{P}_2|^2} = \frac{P_1}{\|\mathbf{P}_1\|^2 + \left|\mathbf{H}_2\mathbf{P}_1\right|^2 P_2} \tag{A11a}$$

$$\frac{1}{1+|\mathbf{H}_2\mathbf{P}_1|^2} = \frac{P_2}{\|\mathbf{P}_2\|^2 + \left|\mathbf{H}_1\mathbf{P}_2\right|^2 P_1}. \tag{A11b}$$

Rearranging, we obtain

$$\begin{bmatrix}\|\mathbf{P}_1\|^2 \\ \|\mathbf{P}_2\|^2\end{bmatrix} = \underbrace{\begin{bmatrix}1+|\mathbf{H}_1\mathbf{P}_2|^2 & -|\mathbf{H}_2\mathbf{P}_1|^2 \\ -|\mathbf{H}_1\mathbf{P}_2|^2 & 1+|\mathbf{H}_2\mathbf{P}_1|^2\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}P_1 \\ P_2\end{bmatrix} \tag{A12}$$

 doi:10.20944/preprints202605.1186.v1



and directly see the power equivalence (cf. (68))

$$P_1 + P_2 = \|\mathbf{P}_1\|^2 + \|\mathbf{P}_2\|^2. \tag{A13}$$

The matrix $\mathbf{A}$ in (A12) has an inverse with positive entries. Non-negative $\|\mathbf{P}_1\|^2, \|\mathbf{P}_2\|^2$ thus give non-negative $P_1, P_2$. The reverse may be seen by writing $\mathbf{P}_k = \|\mathbf{P}_k\| \, \underline{u}_k$ where $\|\underline{u}_k\| = 1$ and inserting into (A11). We obtain

$$\|\mathbf{P}_1\|^2(1 + a_{21}P_2) = P_1\Big(1 + a_{12}\|\mathbf{P}_2\|^2\Big) \tag{A14a}$$

$$\|\mathbf{P}_2\|^2(1 + a_{12}P_1) = P_2\Big(1 + a_{21}\|\mathbf{P}_1\|^2\Big). \tag{A14b}$$

where $a_{kl} = |\mathbf{H}_k \, \underline{u}_l|^2$. Note that swapping $(P_1, P_2)$ with $(\|\mathbf{P}_2\|^2, \|\mathbf{P}_1\|^2)$ does not change the equations. Thus, non-negative $P_1, P_2$ give non-negative $\|\mathbf{P}_1\|^2, \|\mathbf{P}_2\|^2$; see (A12).

## Appendix F. Information Rates of LP with Two Modes

Consider LP with $K = 2$ UEs and two transmission modes that are given fractions $\alpha$ and $1 - \alpha$ of the time/frequency resources. One may optimize over $\alpha$, and the covariance matrices $\mathbf{K}_{k,t}$ of each mode $t = 1, 2$. An achievable weighted sum-rate is characterized by the problem:

$$\begin{aligned} &\max \; w_1(\alpha R_{1,1} + (1-\alpha)R_{1,2}) + w_2(\alpha R_{2,1} + (1-\alpha)R_{2,2}) \\ &\text{s.t. } \alpha \mathrm{Tr}(\mathbf{K}_{1,1} + \mathbf{K}_{2,1}) + (1-\alpha)\mathrm{Tr}(\mathbf{K}_{1,2} + \mathbf{K}_{2,2}) \le P_X \end{aligned} \tag{A15}$$

where

$$R_{k,t} = \log\left|\mathbf{I} + \mathbf{H}_k\mathbf{K}_{k,t}\mathbf{H}_k^\dagger\Big(\mathbf{I} + \mathbf{H}_k\mathbf{K}_{3-k,t}\mathbf{H}_k^\dagger\Big)^{-1}\right|. \tag{A16}$$

For example, for the channels (69) with $n_k = 1$, one should optimize $\alpha$ and the angles and powers of four precoding vectors (two precoding vectors for each UE).

## Appendix G. Rate Splitting

Consider $2^K - 1$ independent vectors $\mathbf{X}_\mathcal{S}$ with $\emptyset \subsetneq \mathcal{S} \subseteq \{1, \ldots, K\}$. The BS transmits

$$\mathbf{X} = \textstyle\sum_\mathcal{S} \mathbf{X}_\mathcal{S}. \tag{A17}$$

The idea is to split each message $k$ into up to $2^{K-1}$ different messages, each carried by exactly one of the $\mathbf{X}_\mathcal{S}$ with $k \in \mathcal{S}$. UE $k$ decodes these $\mathbf{X}_\mathcal{S}$, either using joint decoding or SIC. Clearly, the number of parameters and bounds grows rapidly with $K$. For $K = 2$, we have three messages and (70) is replaced with four decoding bounds:

$$R_{\{1,2\}} \le \log\det\left(\mathbf{I} + \frac{\mathbf{H}_k\mathbf{K}_{\{1,2\}}\mathbf{H}_k^\dagger}{\mathbf{I} + \sum_{l=1,2}\mathbf{H}_k\mathbf{K}_{\{l\}}\mathbf{H}_k^\dagger}\right) \tag{A18a}$$

$$R_{\{k\}} \le \log\det\left(\mathbf{I} + \frac{\mathbf{H}_k\mathbf{K}_{\{k\}}\mathbf{H}_k^\dagger}{\mathbf{I} + \sum_{l\neq k}\mathbf{H}_k\mathbf{K}_{\{l\}}\mathbf{H}_k^\dagger}\right) \tag{A18b}$$

for $k = 1, 2$. We also have three rate-splitting bounds

$$R_1 \le R_{\{1\}} + R_{\{1,2\}} \tag{A19a}$$

$$R_2 \le R_{\{2\}} + R_{\{1,2\}} \tag{A19b}$$

$$R_1 + R_2 \le R_{\{1\}} + R_{\{2\}} + R_{\{1,2\}}. \tag{A19c}$$

 doi:10.20944/preprints202605.1186.v1



For example, setting $\mathbf{K}_{\{1,2\}} = \mathbf{0}$, we have $R_{\{1,2\}} = 0$ and obtain the two LP bounds (70). A more interesting case is $\mathbf{K}_{\{1\}} = \mathbf{0}$ so that $R_{\{1\}} = 0$ and one chooses

$$(R_1, R_2) = \left(R_{\{1,2\}}, R_{\{2\}}\right). \tag{A20}$$

In this case, UE 2 recovers both messages. As for coded LP, time- or frequency-sharing across multiple transmission modes can increase coded LP rates; see [15].

## References


1. Poor, H.; Verdu, S. Single-user detectors for multiuser channels. *IEEE Trans. Commun.* **1988**, *36*, 50–60. https://doi.org/10.1109/26.2728.
2. Poor, H.; Verdu, S. Probability of error in MMSE multiuser detection. *IEEE Trans. Inf. Theory* **1997**, *43*, 858–871. https://doi.org/10.1109/18.568697.
3. Wang, X.; Poor, H. Blind multiuser detection: a subspace approach. *IEEE Trans. Inf. Theory* **1998**, *44*, 677–690. https://doi.org/10.1109/18.661512.
4. Mitra, U.; Poor, H. Adaptive receiver algorithms for near-far resistant CDMA. *IEEE Trans. Commun.* **1995**, *43*, 1713–1724. https://doi.org/10.1109/26.380222.
5. Nelson, L.; Poor, H. Iterative multiuser receivers for CDMA channels: an EM-based approach. *IEEE Trans. Commun.* **1996**, *44*, 1700–1710. https://doi.org/10.1109/26.545900.
6. Wang, X.; Poor, H. Iterative (turbo) soft interference cancellation and decoding for coded CDMA. *IEEE Trans. Commun.* **1999**, *47*, 1046–1061. https://doi.org/10.1109/26.774855.
7. Ding, Z.; Yang, Z.; Fan, P.; Poor, H.V. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Proc. Lett.* **2014**, *21*, 1501–1505. https://doi.org/10.1109/LSP.2014.2343971.
8. Ding, Z.; Adachi, F.; Poor, H.V. The application of MIMO to non-orthogonal multiple access. *IEEE Trans. Wireless Commun.* **2015**, *15*, 537–552. https://doi.org/10.1109/TWC.2015.2475746.
9. Ding, Z.; Liu, Y.; Choi, J.; Sun, Q.; Elkashlan, M.; Chih-Lin, I.; Poor, H.V. Application of non-orthogonal multiple access in LTE and 5G networks. *IEEE Commun. Mag.* **2017**, *55*, 185–191. https://doi.org/10.1109/MCOM.2017.1500657CM.
10. Vaezi, M.; Schober, R.; Ding, Z.; Poor, H.V. Non-orthogonal multiple access: common myths and critical questions. *IEEE Wireless Commun.* **2019**, *26*, 174–180. https://doi.org/10.1109/MWC.2019.1800598.
11. Clerckx, B.; Mao, Y.; Schober, R.; Poor, H.V. Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: a simple two-user rate analysis. *IEEE Wireless Commun. Lett.* **2020**, *9*, 349–353. https://doi.org/10.1109/LWC.2019.2954518.
12. Park, J.; Choi, J.; Lee, N.; Shin, W.; Poor, H.V. Rate-splitting multiple access for downlink MIMO: a generalized power iteration approach. *IEEE Trans. Wireless Commun.* **2023**, *22*, 1588–1603. https://doi.org/10.1109/TWC.2022.3205480.
13. Wang, C.X.; You, X.; Gao, X.; Zhu, X.; Li, Z.; Zhang, C.; Wang, H.; Huang, Y.; Chen, Y.; Haas, H.; et al. On the road to 6G: visions, requirements, key technologies, and testbeds. *IEEE Commun. Surveys & Tutor.* **2023**, *25*, 905–974. https://doi.org/10.1109/COMST.2023.3249835.
14. Katz, A.; Peleg, M.; Vincent Poor, H.; Shamai Shitz, S. The Gaussian primitive discrete time and filtered diamond channel: correlated noise and dirty-paper coding. *IEEE Trans. Commun.* **2025**, *73*, 5495–5508. https://doi.org/10.1109/TCOMM.2024.3523975.
15. Bergmans, P.; Cover, T. Cooperative broadcasting. *IEEE Trans. Inf. Theory* **1974**, *20*, 317–324. https://doi.org/10.1109/TIT.1974.1055232.
16. I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Channels*; Akadémiai Kiadó: Budapest, 1981.
17. Cover, T. Broadcast channels. *IEEE Trans. Inf. Theory* **1972**, *18*, 2–14. https://doi.org/10.1109/TIT.1972.1054727.
18. Bergmans, P. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inf. Theory* **1973**, *19*, 197–207. https://doi.org/10.1109/TIT.1973.1054980.
19. Bergmans, P. A simple converse for broadcast channels with additive white Gaussian noise. *IEEE Trans. Inf. Theory* **1974**, *20*, 279–280. https://doi.org/10.1109/TIT.1974.1055184.
20. Gallager, R.G. Capacity and coding for degraded broadcast channels. *Probl. Peredachi Inf.* **1974**, *10*, 3–14.

 doi:10.20944/preprints202605.1186.v1




21. Cover, T. An achievable rate region for the broadcast channel. *IEEE Trans. Inf. Theory* **1975**, *21*, 399–404. https://doi.org/10.1109/TIT.1975.1055418.
22. van der Meulen, E. Random coding theorems for the general discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory* **1975**, *21*, 180–190. https://doi.org/10.1109/TIT.1975.1055347.
23. Carleial, A. Interference channels. *IEEE Trans. Inf. Theory* **1978**, *24*, 60–70. https://doi.org/10.1109/TIT.1978.1055812.
24. Gamal, A. The capacity of a class of broadcast channels. *IEEE Trans. Inf. Theory* **1979**, *25*, 166–169. https://doi.org/10.1109/TIT.1979.1056029.
25. Joudeh, H.; Clerckx, B. Sum-rate maximization for linearly precoded downlink multiuser MISO systems With partial CSIT: A rate-splitting approach. *IEEE Trans. Commun.* **2016**, *64*, 4847–4861. https://doi.org/10.1109/TCOMM.2016.2603991.
26. Mao, Y.; Clerckx, B.; Li, V.O. Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming SDMA and NOMA. *EURASIP J. Wireless Commun. Netw.* **2018**, *2018*, 133.
27. Mao, Y.; Clerckx, B. Beyond dirty paper coding for multi-antenna broadcast channel With partial CSIT: a rate-splitting approach. *IEEE Trans. Commun.* **2020**, *68*, 6775–6791. https://doi.org/10.1109/TCOMM.2020.3014153.
28. Marton, K. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory* **1979**, *25*, 306–311. https://doi.org/10.1109/TIT.1979.1056046.
29. Runge, C.; Kramer, G. Time-shifted alternating Gelfand-Pinsker coding for broadcast channels. In Proceedings of the IEEE Int. Symp. Inf. Theory, Athens, Greece, 2024; pp. 1700–1705. https://doi.org/10.1109/ISIT57864.2024.10619498.
30. Gelfand, S. Coding for channel with random parameters. *Probl. Contr. Inf. Theory* **1980**, *9*, 19–31.
31. Costa, M. Writing on dirty paper. *IEEE Trans. Inf. Theory* **1983**, *29*, 439–441. https://doi.org/10.1109/TIT.1983.1056659.
32. Wyner, A.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10. https://doi.org/10.1109/TIT.1976.1055508.
33. Zamir, R.; Shamai, S.; Erez, U. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Inf. Theory* **2002**, *48*, 1250–1276. https://doi.org/10.1109/TIT.2002.1003821.
34. Erez, U.; Shamai, S.; Zamir, R. Capacity and lattice strategies for canceling known interference. *IEEE Trans. Inf. Theory* **2005**, *51*, 3820–3833. https://doi.org/10.1109/TIT.2005.856935.
35. Yu, W.; Sutivong, A.; Julian, D.; Cover, T.; Chiang, M. Writing on colored paper. In Proceedings of the IEEE Int. Symp. Inf. Theory, 2001, pp. 302–. https://doi.org/10.1109/ISIT.2001.936165.
36. Yu, W.; Cioffi, J.M. Sum capacity of Gaussian vector broadcast channels. *IEEE Trans. Inf. Theory* **2004**, *50*, 1875–1892. https://doi.org/10.1109/TIT.2004.833336.
37. Cohen, A.S.; Lapidoth, A. The Gaussian watermarking game. *IEEE Trans. Inf. Theory* **2002**, *48*, 1639–1667. https://doi.org/10.1109/TIT.2002.1003844.
38. Caire, G.; Shamai, S. On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Trans. Inf. Theory* **2003**, *49*, 1691–1706. https://doi.org/10.1109/TIT.2003.813523.
39. Vishwanath, S.; Jindal, N.; Goldsmith, A. Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Trans. Inf. Theory* **2003**, *49*, 2658–2668. https://doi.org/10.1109/TIT.2003.817421.
40. Weingarten, H.; Steinberg, Y.; Shamai, S.S. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inf. Theory* **2006**, *52*, 3936–3964. https://doi.org/10.1109/TIT.2006.880064.
41. Geng, Y.; Nair, C. The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *IEEE Trans. Inf. Theory* **2014**, *60*, 2087–2104. https://doi.org/10.1109/TIT.2014.2304457.
42. Hunger, R.; Joham, M. A general rate duality of the MIMO multiple access channel and the MIMO broadcast channel. In Proceedings of the IEEE Global Telecommun. Conf., 2008, pp. 1–5. https://doi.org/10.1109/GLOCOM.2008.ECP.178.
43. Erez, U.; ten Brink, S. A close-to-capacity dirty paper coding scheme. *IEEE Trans. Inf. Theory* **Oct. 2005**, *51*, 3417–3432. https://doi.org/10.1109/TIT.2005.855586.
44. Sun, Y.; Uppal, M.; Liveris, A.D.; Cheng, S.; Stankovic, V.; Xiong, Z. Nested turbo codes for the Costa problem. *IEEE Trans. Commun.* **Mar. 2008**, *56*, 388–399. https://doi.org/10.1109/TCOMM.2008.050101.
45. Sun, Y.; Yang, Y.; Liveris, A.D.; Stankovic, V.; Xiong, Z. Near-capacity dirty-paper code design: a source-channel coding approach. *IEEE Trans. Inf. Theory* **2009**, *55*, 3013–3031. https://doi.org/10.1109/TIT.2009.2021319.

 doi:10.20944/preprints202605.1186.v1




46. Shilpa, G.; Thangaraj, A.; Bhashyam, S. Dirty paper coding using sign-bit shaping and LDPC codes. In Proceedings of the IEEE Int. Symp. Inf. Theory, Austin, TX, USA, Jun. 2010; pp. 923–927. https://doi.org/10.1109/ISIT.2010.5513493.
47. Bennatan, A.; Burshtein, D.; Caire, G.; Shamai, S. Superposition coding for side-information channels. *IEEE Trans. Inf. Theory* **2006**, *52*, 1872–1889. https://doi.org/10.1109/TIT.2006.872985.
48. Gariby, T.; Erez, U.; Shamai, S. Dirty paper coding for PAM signaling. In Proceedings of the IEEE Int. Symp. Inf. Theory, Nice, France, 2007; pp. 376–380. https://doi.org/10.1109/ISIT.2007.4557091.
49. Silva, D.; Pivaro, G.; Fraidenraich, G.; Aazhang, B. On integer-forcing precoding for the Gaussian MIMO broadcast channel. *IEEE Trans. Wireless Commun.* **2017**, *16*, 4476–4488. https://doi.org/10.1109/TWC.2017.2699178.
50. He, W.; Nazer, B.; Shamai Shitz, S. Uplink-downlink duality for integer-forcing. *IEEE Trans. Inf. Theory* **2018**, *64*, 1992–2011. https://doi.org/10.1109/TIT.2018.2791589.
51. Venturelli, R.B.; Silva, D. Optimization of integer-forcing precoding for multi-user MIMO downlink. *IEEE Wireless Commun. Lett.* **2020**, *9*, 1860–1864. https://doi.org/10.1109/LWC.2020.3006393.
52. Liu, T.; Moulin, P.; Koetter, R. On error exponents of modulo lattice additive noise channels. *IEEE Trans. Inf. Theory* **2006**, *52*, 454–471. https://doi.org/10.1109/TIT.2005.862077.
53. Verdú, S. Non-asymptotic achievability bounds in multiuser information theory. In Proceedings of the Allerton Conf. Commun., Control, and Computing, Allerton, IL, 2012; pp. 1–8. https://doi.org/10.1109/Allerton.2012.6483192.
54. Watanabe, S.; Kuzuoka, S.; Tan, V.Y.F. Nonasymptotic and second-order achievability bounds for coding With side-information. *IEEE Trans. Inf. Theory* **2015**, *61*, 1574–1605. https://doi.org/10.1109/TIT.2015.2400994.
55. Scarlett, J. On the dispersions of the Gel'fand–Pinsker channel and dirty paper coding. *IEEE Trans. Inf. Theory* **2015**, *61*, 4569–4586. https://doi.org/10.1109/TIT.2015.2449844.
56. Tamir, R.; Merhav, N. Error exponents of the dirty-paper and Gel'fand–Pinsker channels. *IEEE Trans. Inf. Theory* **2023**, *69*, 7479–7498. https://doi.org/10.1109/TIT.2023.3314210.
57. Şener, M.Y.; Boehnke, R.; Xu, W.; Kramer, G. Dirty paper coding based on polar codes and probabilistic shaping. *IEEE Commun. Lett.* **2021**, *25*, 3810–3813. https://doi.org/10.1109/LCOMM.2021.3113722.
58. Şener, M.Y.; Boehnke, R.; Xu, W. A practical dirty paper coding scheme for MISO broadcast channels. In Proceedings of the IEEE Global Commun. Conf., Rio de Janeiro, Brazil, Dec. 2022; pp. 215–220. https://doi.org/10.1109/GLOBECOM48099.2022.10001479.
59. Şener, M.Y.; Boehnke, R.; Xu, W.; Kramer, G. Achieving the dirty paper channel capacity with scalar lattices and probabilistic shaping. *IEEE Commun. Lett.* **2024**, *28*, 29–33. https://doi.org/10.1109/LCOMM.2023.3339252.
60. Şener, M.Y.; Boehnke, R.; Xu, W. Polar coding for parallel dirty paper channels. In Proceedings of the IEEE Global Commun. Conf. Workshops, 2024, pp. 1–6. https://doi.org/10.1109/GCWkshp64532.2024.11100314.
61. Arikan, E. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theory* **2009**, *55*, 3051–3073.
62. Korada, S.B.; Urbanke, R.L. Polar codes are optimal for lossy source coding. *IEEE Trans. Inf. Theory* **2010**, *56*, 1751–1768. https://doi.org/10.1109/TIT.2010.2040961.
63. Honda, J.; Yamamoto, H. Polar coding without alphabet extension for asymmetric models. *IEEE Trans. Inf. Theory* **2013**, *59*, 7829–7838. https://doi.org/10.1109/TIT.2013.2282305.
64. Seidl, M.; Schenk, A.; Stierstorfer, C.; Huber, J.B. Polar-coded modulation. *IEEE Trans. Commun.* **2013**, *61*, 4108–4119. https://doi.org/10.1109/TCOMM.2013.090513.130433.
65. Liu, L. Polar codes and polar lattices for efficient communication and source quantization. PhD thesis, Dept. Electrical and Electronics Engineering, Imperial College London, 2016.
66. Sutter, D.; Renes, J.M.; Dupuis, F.; Renner, R. Achieving the capacity of any DMC using only polar codes. In Proceedings of the IEEE Inf. Theory Workshop, Lausanne, Switzerland, 2012; pp. 114–118. https://doi.org/10.1109/ITW.2012.6404638.
67. Böcherer, G.; Prinz, T.; Yuan, P.; Steiner, F. Efficient polar code construction for higher-order modulation. In Proceedings of the IEEE Wireless Commun. Networking Conf. Workshops, San Francisco, CA, USA, 2017; pp. 1–6. https://doi.org/10.1109/WCNCW.2017.7919039.
68. Prinz, T.; Yuan, P.; Böcherer, G.; Steiner, F.; İşcan, O.; Boehnke, R.; Xu, W. Polar coded probabilistic amplitude shaping for short packets. In Proceedings of the IEEE Int. Workshop Signal Proc. Adv. Wireless Commun., Sapporo, Japan, 2017; pp. 1–5. https://doi.org/10.1109/SPAWC.2017.8227653.

 doi:10.20944/preprints202605.1186.v1




69. Mondelli, M.; Hassani, S.H.; Urbanke, R.L. How to achieve the capacity of asymmetric channels. *IEEE Trans. Inf. Theory* **2018**, *64*, 3371–3393. https://doi.org/10.1109/TIT.2018.2789885.
70. İşcan, O.; Boehnke, R.; Xu, W. Shaped polar codes for higher order modulation. *IEEE Commun. Lett.* **2018**, *22*, 252–255. https://doi.org/10.1109/LCOMM.2017.2766621.
71. Wiegart, T.; Steiner, F.; Schulte, P.; Yuan, P. Shaped on-off keying using polar codes. *IEEE Commun. Lett.* **2019**, *23*, 1922–1926. https://doi.org/10.1109/LCOMM.2019.2930511.
72. Wiegart, T.; Prinz, T.; Steiner, F.; Yuan, P. Design of polar codes for parallel channels with an average power constraint. In Proceedings of the IEEE Int. Symp. Inf. Theory, Paris, France, 2019; pp. 1942–1946. https://doi.org/10.1109/ISIT.2019.8849335.
73. İşcan, O.; Boehnke, R.; Xu, W. Probabilistic shaping using 5G New Radio polar codes. *IEEE Access* **2019**, *7*, 22579–22587. https://doi.org/10.1109/ACCESS.2019.2898103.
74. İşcan, O.; Boehnke, R.; Xu, W. Sign-bit shaping using polar codes. *Trans. Emerging Telecommun. Technol.* **2020**, *31*, e4058. https://doi.org/https://doi.org/10.1002/ett.4058.
75. Boehnke, R.; İşcan, O.; Xu, W. Multi-level distribution matching. *IEEE Commun. Lett.* **2020**, *24*, 2015–2019. https://doi.org/10.1109/LCOMM.2020.2993929.
76. Runge, C.; Wiegart, T.; Lentner, D.; Prinz, T. Multilevel binary polar-coded modulation achieving the capacity of asymmetric channels. In Proceedings of the IEEE Int. Symp. Inf. Theory, Espoo, Finland, 2022; pp. 2595–2600. https://doi.org/10.1109/ISIT50566.2022.9834740.
77. Eghbalian-Arani, S.; Behroozi, H. Polar codes for a quadratic-Gaussian Wyner-Ziv problem. In Proceedings of the Int. Symp. Wireless Commun. Systems, Ilmenau, Germany, 2013; pp. 1–5.
78. Liu, L.; Ling, C. Polar codes and polar lattices for independent fading channels. *IEEE Trans. Commun.* **2016**, *64*, 4923–4935. https://doi.org/10.1109/TCOMM.2016.2613109.
79. Liu, L.; Yan, Y.; Ling, C.; Wu, X. Construction of capacity-achieving lattice codes: polar lattices. *IEEE Trans. Commun.* **2019**, *67*, 915–928. https://doi.org/10.1109/TCOMM.2018.2876113.
80. Liu, L.; Shi, J.; Ling, C. Polar lattices for lossy compression. *IEEE Trans. Inf. Theory* **2021**, *67*, 6140–6163. https://doi.org/10.1109/TIT.2021.3097965.
81. Liu, L.; Lyu, S.; Ling, C.; Bai, B. On the equivalence between probabilistic shaping and geometric shaping: a polar lattice perspective. In Proceedings of the IEEE Int. Symp. Inf. Theory, Athens, Greece, 2024; pp. 2174–2179. https://doi.org/10.1109/ISIT57864.2024.10619275.
82. Jha, S. Universal Gaussian quantization with side-information using polar lattices. *IEEE J. Sel. Areas Inf. Theory* **2022**, *3*, 639–650. https://doi.org/10.1109/JSAIT.2023.3247864.
83. Şener, M.Y.; Kramer, G.; Shamai Shitz, S.; Böhnke, R.; Xu, W. Achieving Gaussian Vector Broadcast Channel Capacity with Scalar Lattices. In Proceedings of the IEEE Int. Symp. Inf. Theory, 2024, pp. 1706–1711. https://doi.org/10.1109/ISIT57864.2024.10619389.
84. Wintner, A. *Asymptotic distributions and infinite convolutions*; Edwards Brothers: Ann Arbor, Michigan, 1938.
85. Purkayastha, S. Simple proofs of two results on convolutions of unimodal distributions. *Stat. Prob. Lett.* **1998**, *39*, 97–100. https://doi.org/https://doi.org/10.1016/S0167-7152(98)00013-3.
86. Boehnke, R.; Kammeyer, K.D. Weighted sum rate maximization for the MIMO-downlink using a projected conjugate gradient algorithm. In Proceedings of the Int. Workshop on Cross Layer Design, 2007, pp. 82–85. https://doi.org/10.1109/IWCLD.2007.4379043.
87. Christensen, S.S.; Agarwal, R.; De Carvalho, E.; Cioffi, J.M. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. *IEEE Trans. Wireless Commun.* **2008**, *7*, 4792–4799. https://doi.org/10.1109/ICC.2009.5199574.
88. Liu, S.; Hong, Y.; Viterbo, E. Polar codes for block fading channels. In Proceedings of the IEEE Wireless Commun. Network. Conf. Workshops, San Francisco, CA, USA, Mar. 2017; pp. 1–6. https://doi.org/10.1109/WCNCW.2017.7919041.
89. Lin, X.; Li, J.; Baldemair, R.; Cheng, J.F.T.; Parkvall, S.; Larsson, D.C.; Koorapaty, H.; Frenne, M.; Falahati, S.; Grovlen, A.; et al. 5G New Radio: unveiling the essentials of the next generation wireless access technology. *IEEE Commun. Stand. Mag.* **2019**, *3*, 30–37. https://doi.org/10.1109/MCOMSTD.001.1800036.