

Capacity Approaching Low Density Spreading in Uplink NOMA via Asymptotic Analysis

Hossein Asgharimoghaddam *Student Member, IEEE*, Jarkko Kaleva, *Member, IEEE*, and Antti Tölli, *Senior Member, IEEE*

Abstract

Low-density spreading non-orthogonal multiple-access (LDS-NOMA) is considered where K single-antenna user-equipments (UEs) communicate with a base-station (BS) over F fading sub-carriers. Each UE k spreads its data symbols over $d_k < F$ sub-carriers. We aim to identify the LDS-code allocations that maximize the ergodic mutual information (EMI). The BS assigns resources solely based on pathlosses. Conducting analysis in the regime where F , K , and $d_k, \forall k$ converge to $+\infty$ at the same rate, we present EMI as a deterministic equivalent plus a residual term. The deterministic equivalent is a function of pathlosses and spreading codes, and the small residual term scales as $\mathcal{O}(\frac{1}{\min(d_k^2)})$. We formulate an optimization problem to get the set of all spreading codes, irrespective of sparsity constraints, which maximize the deterministic EMI. This yields a simple resource allocation rule that facilitates the construction of desired LDS-codes via an efficient partitioning algorithm. The acquired LDS-codes additionally harness the small incremental gain inherent in the residual term, and thus, attain near-optimal values of EMI in the finite regime. While regular LDS-NOMA is found to be asymptotically optimal in symmetric models, an irregular spreading arises in generic asymmetric cases. The spectral efficiency enhancement relative to regular and random spreading is validated numerically.

I. INTRODUCTION

The objective in 5G and beyond network is to move on from merely connecting people to fully realizing the internet of things (IoT) and the fourth industrial revolution [1]. This has resulted in a shift towards techniques that both increase the spectral efficiency and support massive connectivity. Non-orthogonal multiple access (NOMA) as such a technique allows multiple user equipments (UEs) to share the same signal dimension via power domain (PD) or code domain

H. Asgharimoghaddam and A. Tölli are with the Centre for Wireless Communications, University of Oulu, Oulu, Finland (Firstname.Lastname@oulu.fi). J. Kaleva (jarkko.kaleva@solmutech.com) is with the Solmu Technologies Oy.

This work has been supported in part by the Academy of Finland 6Genesis Flagship (grant no. 318927).

A preliminary version of this paper has been presented at the IEEE International Symposium on Information Theory (ISIT) 2020 Los Angeles, California, USA.

(CD) multiplexing. PD-NOMA distinguishes UE's signals via superposition decoding principle and by exploiting the signal to interference plus noise ratio (SINR) difference among UEs. In CD-NOMA, including low-density spreading (LDS) multiple access [2], sparse code multiple access (SCMA) [3], multi-user shared access (MUSA) [4], distinctive codes are assigned to UEs as in code division multiple access (CDMA) system [5], [6]. The focus of this work is on LDS-NOMA, which employs an LDS code comprising a small number of d_k non-zero elements for spreading UE k 's symbol over a number of F shared radio resources [2], [7]–[9]. Link level aspects of LDS-NOMA, e.g. bit-error-rate performance [2], [3], [8], [9] and envelop fluctuation [10], have been well-studied in the literature. However, the theoretical results about the boundaries of the achievable rates in LDS-NOMA is rather limited [11]. Here, a multi-carrier LDS-NOMA scheme with spreading in frequency domain [8], [12] is considered. The objective is to characterize the boundaries of the ergodic sum rates at which UEs can jointly transmit reliably, and identify the LDS-code allocation policies that closely attain these boundaries. Such analysis provides insight for system design [5], [12], and hence, is instrumental for the optimal utilization of scarce radio resources.

A. Prior related works

The spectral efficiency (SE) analysis of LDS-NOMA with spreading in time domain has been considered in [11], [13]–[17] under a symmetric AWGN channel model. These works compare the SE limits in the structured regular LDS codes and in the randomly generated irregular ones. The sparse mapping between numbers of K UEs and F resources in LDS-NOMA is called regular when each UE occupies a number of $d_k = d, \forall k$ resources and each resource is used by a number of $\frac{K}{F}d$ UEs; or irregular otherwise [13]. The irregular schemes with d_k being randomly Poissonian distributed with fixed mean [16], [17], and randomly uniformly distributed [15] are studied using replica method [18] and random matrix framework developed in [19], respectively. The regular scheme is considered in [13], [14], where in [14] a closed-form approximation is given for the SE limit. These analyses indicate that the regular codes, in symmetric AWGN channel, yield superior SE as compared to the irregular and the dense spreading (the case with $d_k = F, \forall k$) schemes. The aforementioned works rely on the analysis of random matrices in the large system regime [18], [20], where F grows large with a fixed ratio of K/F . Such analysis yields rather accurate approximations in the finite regime that become arbitrarily tight as F grows large. However, since the mathematical literature studying the limiting behavior of sparse random matrices is distinctly smaller than that for non-sparse ones [21], the extension of such

SE analysis to more generic settings is rather difficult. The only large system analysis on the SE limit in presence of fading is considered in [22] in a special setting with $d_k = 1, \forall k$. An information theoretic analysis of LDS-NOMA with fading also appears in [23], wherein the ergodic sum-rate of random spreading has been evaluated numerically. The application of large system analysis for studying the SE limits of spread spectrum system has been motivated in the pioneer works such as [19], [24]–[29]. These works characterize the spectral efficiency of random dense-spreading ($d_k = F, \forall k$) CDMA with linear [25], [26] and/or optimal [19], [24], [27]–[29] receivers. A common conclusion in a number of the aforementioned works is that the random dense-spreading CDMA incurs negligible spectral efficiency loss relative to the optimum if an optimal receiver is used and the number of UEs per chip is sufficiently large [19], [27].

B. Contributions

In this paper, a multi-carrier LDS-NOMA scheme with spreading in frequency domain [8], [12] is considered. Different from the aforementioned works, the UEs are allowed to have distinct pathloss values, and frequency and time selective fading is assumed on the sub-carriers. Also, instead of assuming a particular sparse mapping, we consider $d_k, \forall k$ as design parameters, and identify the LDS-code allocation policies that closely attain the maximum of the ergodic mutual information (EMI). A key feature of these policies is that they assign the codes only based on the pathloss values. Conducting analysis in the large system limits where F , K , and $d_k, \forall k$ converge to $+\infty$ at the same rate, we present EMI as a deterministic equivalent plus a residual term. The deterministic equivalent is given as a function of pathloss values and LDS-codes, and the small residual term is shown to quickly vanish inversely proportional to d^2 where $d = \min\{d_k, \forall k\}$. First, we formulate an optimization problem to get the set of all spreading codes, irrespective of the sparsity constraints, which maximize the deterministic equivalent of EMI. This yields a simple resource allocation rule that facilitates the construction of the desired sparse spreading codes via an efficient partitioning algorithm. The analysis in the finite regime shows that the acquired sparse solutions additionally harness the small incremental gain inherent in the residual term, and thus, attain near-optimal values of the EMI in the finite regime. It is observed that the regular spreading matrices are asymptotically optimal for the symmetric scenarios with the same pathlosses and power constraints for all UEs. However, in the generic asymmetric scenarios, an irregular structure might arise. Numerical simulations validate the attainable spectral efficiency enhancement as compared to the random and the regular spreading schemes.

Parts of this paper have been published in the conference publication [30]. Specifically, the large system analysis and the optimization approach are sketched therein, while the precise proofs along with the derivation details are presented in the current work. In addition, the numerical analysis is extended to visualize the resource allocation policy proposed herein.

The remainder of this work is organized as follows. In Section II, the network model and the problem formulation are given. The proposed optimization approach based on large system analysis is presented in Section III. The maximization of the deterministic EMI is considered in Section IV-A, and the gap to the optimum is characterized in Section IV-B. The algorithmic solution for the LDS codes assignment is presented in Section V. The numerical results are given in Section VI. Conclusions are drawn in Section VII while all the proofs are presented in the Appendices.

II. PROBLEM STATEMENT

A. General Notations

The following notations are used throughout the manuscript. All boldface letters indicate vectors (lower case) or matrices (upper case). Superscripts $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, $(\cdot)^{-1}$, $(\cdot)^{1/2}$ stand for transpose, Hermitian transpose, conjugate operator, matrix inversion and positive semidefinite square root, respectively. We use $\mathbb{C}^{m \times n}$ and $\mathbb{R}^{m \times n}$ to denote the set of $m \times n$ complex and the real valued matrices, respectively. $\text{diag}(\cdots)$ denotes a diagonal matrix with elements (\cdots) on the main diagonal. The $(i, j)^{\text{th}}$ element of a matrix \mathbf{A} is denoted by $[\mathbf{A}]_{i,j}$ or $a_{i,j}$. The sets are indicated by calligraphic letters. The cardinality of a set \mathcal{A} is denoted by $|\mathcal{A}|$, and $\mathcal{A} \setminus k$ is used to exclude the index k from the set. $\mathbb{E}\{\cdot\}$, $\text{Tr}\{\cdot\}$ denote statistical expectation and trace operator, respectively. Euclidean (spectral) norm for vectors (matrices) are denoted by $\|\cdot\|$. The notation $|\cdot|$ is used to denote both the absolute value for a complex scalar, and the determinant for a square matrix. Notation $x_F = \mathcal{O}(\alpha_F)$ represents inequality $|x_F| \leq C\alpha_F$ as $F \rightarrow \infty$ with C being a generic constant independent of system size F .

B. System Model

Consider an uplink multi-carrier system with K single-antenna UEs transmitting to a base station (BS) on a common frequency band. The set of UEs' indices is denoted by \mathcal{K} hereafter. Exploiting the OFDM technique, the total frequency band is divided into a set of narrow band sub-channels $\mathcal{F} = \{1, \dots, F\}$. Then, each UE k spreads its data symbol in frequency direction using a low density spreading code $\mathbf{w}_k \in \mathbb{R}^F$. The code is a sparse vector consisting of F chips

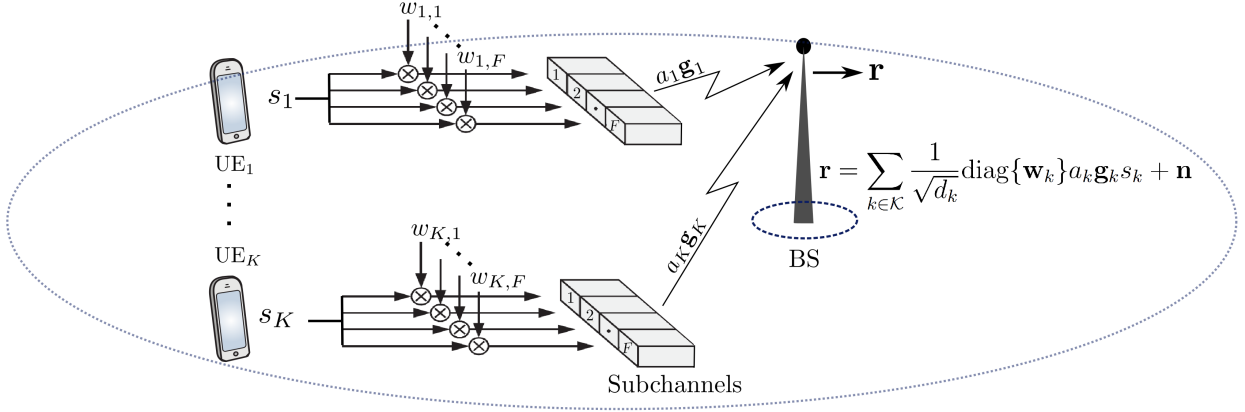


Fig. 1: The system model illustration.

with d_k , a small number, of non-zero values. The UEs' codes are not restricted to be orthogonal, and thus, the signals of UEs transmitting on the same sub-channel will be superimposed. The received signal vector $\mathbf{r} \in \mathbb{C}^F$ is given as

$$\mathbf{r} = \sum_{k \in \mathcal{K}} \frac{1}{\sqrt{d_k}} \text{diag}\{\mathbf{w}_k\} a_k \mathbf{g}_k s_k + \mathbf{n} \quad (1)$$

where the f^{th} element of the vector \mathbf{r} corresponds to the signal received within f^{th} sub-channel. The noise vector is given by $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_F)$. The unit variance symbol of k^{th} UE is denoted by s_k . The channel vector for UE k is denoted by $a_k \mathbf{g}_k \in \mathbb{C}^F$, defined in more detail in the sequel. The vector $\mathbf{w}_k \in \mathbb{R}^F$ denotes the UE k 's spreading code, and is assumed to satisfy the transmit power condition

$$\frac{1}{d_k} \sum_{f \in \mathcal{F}} |w_{f,k}|^2 \leq P_k, \forall k \in \mathcal{K} \quad (2)$$

where P_k is the total power available at UE k . The normalization by d_k values in (1) and (2) is considered to make the sparsity assumption in the system model explicit.

C. Channel Model

The channel matrix entries are generated based on the uncorrelated fading channel model utilized in the context of multi-carrier systems [31, Chapter 1]. This channel model is based on the assumption that the fading on the adjacent data symbols after inverse OFDM and de-interleaving can be considered as uncorrelated [31]. This assumption holds when, for example, time and frequency interleavers with sufficient interleaving depth are applied. Thus, the resulting complex-valued channel fading coefficient is generated independently for each sub-carrier and

OFDM symbol. For a propagation scenario without line of sight, the fading amplitude is generated by a Rayleigh distribution and the channel model is referred to as uncorrelated Rayleigh fading channel [31]. In particular, we present the channel vector for UE k as $a_k \mathbf{g}_k$ where a_k^2 includes the pathloss due to large scale fading, and the matrix $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_K]$ represents the small-scale fading. The entries of \mathbf{G} are independent complex Gaussian random variables. Each entry has zero-mean independent real and imaginary parts with variance of $\frac{1}{2}$. In the following, we use $\mathbf{h}_k = \frac{a_k}{\sqrt{d_k}} \text{diag}\{\mathbf{w}_k\} \mathbf{g}_k$ and $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]$ to denote the equivalent channel vectors and matrix, respectively, including the spreading vectors.

D. Ergodic Capacity of the Channel

Let $\mathbf{W} \in \mathbb{R}^{F \times K}$ to be the spreading matrix that contains all the spreading vectors $\{\mathbf{w}_k\}$ in its columns. We define the set of feasible sparse spreading matrices as $\mathcal{C}_1 = \{\mathbf{W} | w_{f,k} \in \mathbb{R}, \frac{1}{d_k} \sum_f |w_{f,k}|^2 \leq P_k, \|\mathbf{w}_k\|_0 = d_k, \forall k \in \mathcal{K}\}$ where $\|\mathbf{w}_k\|_0 = d_k$ restricts the number of non-zero elements in \mathbf{w}_k to be equal to d_k . Our interest is in the scenario where the BS assigns the spreading codes based on the UEs' pathlosses, and the assigned spreading codes are known to the UEs. Implicit in this model is the assumption that the channel statistics vary much more slowly than the small-scale fading coefficients, so that the statistical properties of the channel can be assumed constant for a long period of communication [32]. Given a perfect knowledge of fading coefficients at the BS side, the ergodic mutual information (EMI) between the transmitters and the receiver, for a spreading matrix $\mathbf{W} \in \mathcal{C}_1$, is

$$J(\mathbf{W}, \sigma^2) = \frac{1}{F} \mathbb{E} \log |\mathbf{I}_F + \frac{1}{\sigma^2} \sum_{k \in \mathcal{K}} \frac{a_k^2}{d_k} \text{diag}\{\mathbf{w}_k\} \mathbf{g}_k \mathbf{g}_k^H \text{diag}\{\mathbf{w}_k\}| \quad (3)$$

where the expectation is over the small-scale fading only. We are particularly interested in the ergodic capacity of the channel, which is equal to the maximum of $J(\mathbf{W}, \sigma^2)$ over the set of all the sparse spreading matrices in \mathcal{C}_1 , i.e.,

$$C_E(\sigma^2) = \max_{\mathbf{W} \in \mathcal{C}_1} J(\mathbf{W}, \sigma^2). \quad (4)$$

The corresponding conventional problem without sparsity constraints is considered in [33] for symmetric additive white Gaussian noise (AWGN) channel model. It is shown in [33] that Welch-Bound-Equality (WBE) signature sequences achieve the sum-capacity of symmetric AWGN channel. The optimal spreading codes in asymmetric AWGN model is studied in [34] via the concept of Majorization and Schur-concavity [35] of the sum-capacity with respect to eigenvalues

of $\mathbf{H}\mathbf{H}^H$. While AWGN channel capacity depends on the spreading codes through their cross-correlations [33], [34], the transmission of the signal over the i.i.d Rayleigh fading channel destroys the orthogonality of the spreading codes [36]. This, together with the hypotheses on the availability of only channel statistics, allow the spreading codes to be taken from the set of vectors $\{\mathbf{w}_k\}$ with positive values and independent of the small-scale fading gains, as stated in the following proposition,

Proposition 1. *The expectation in (3) is invariant with respect to the signs of the entries in real valued spreading vectors \mathbf{w}_k , $\forall k \in \mathcal{K}$.*

Proof. The proof follows by showing that a column of \mathbf{H} , given as $\mathbf{h}_k = \frac{a_k}{\sqrt{d_k}} \text{diag}\{\mathbf{w}_k\} \mathbf{g}_k$, has the same distribution as $\frac{a_k}{\sqrt{d_k}} \text{diag}\{|w_{1,k}|, \dots, |w_{F,k}|\} \mathbf{g}_k$. The claim follows directly from the invariance of i.i.d complex Gaussian vectors in distribution under unitary transformation [37]. \square

Even though the cross-correlation properties of the spreading codes is not a determining parameter in (4), the optimal pairing of UEs and the power loading on each sub-channel need to be studied. The sparsity requirement for spreading codes impose binary constraints in (4) that makes the problem non-convex. Also, the convexity of the objective function in (4) cannot be verified due to the expectation operator. However, the objective function without the expectation operator can be shown to be non-convex. Moreover, the expectation in (4) needs to be evaluated in a concise form and in terms of the spreading codes and the pathloss values.

III. AN OPTIMIZATION APPROACH BASED ON ASYMPTOTIC ANALYSIS

In the following, we use theory of large random matrices [20] to characterize the EMI in terms of the spreading codes and the pathloss values. The large system analysis of the problem is carried out in the asymptotic regime where $F \rightarrow \infty$ with $KF^{-1} \in (0, \infty)$ and $d_k F^{-1} \in (0, 1], \forall k$. The limiting results yield rather accurate approximations for the finite-size scenarios [20]. In deriving the large system analysis, we use subscript F to denote the dependency of the entities on the system size.

Theorem 1. *Consider the channel matrices $\mathbf{H}_F = [\mathbf{h}_1, \dots, \mathbf{h}_K]$ with $\mathbf{h}_k = \frac{a_k}{\sqrt{d_k}} \text{diag}\{\mathbf{w}_k\} \mathbf{g}_k, \forall k \in \mathcal{K}$. The entries of $\mathbf{G}_F = [\mathbf{g}_1, \dots, \mathbf{g}_K]$ are i.i.d standard complex Gaussian random variables. The deterministic vectors $\mathbf{w}_k, \forall k \in \mathcal{K}$ are the columns of $\mathbf{W}_F \in \mathcal{C}_1$, each with d_k non-zero values.*

The scalars $\{a_k\}_{1 \leq k \leq K}$ are bounded real values. Then, as $F \rightarrow \infty$ with $KF^{-1} \in (0, \infty)$ and $d_k F^{-1} \in (0, 1], \forall k$, the EMI converges to a deterministic equivalent such that

$$J_F(\mathbf{W}_F, \sigma^2) = \bar{J}_F(\mathbf{W}_F, \sigma^2) + \epsilon_F \quad (5)$$

where $\epsilon_F = \mathcal{O}(\frac{1}{d^2})$, $d = \min\{d_k\}_{1 \leq k \leq K}$, and

$$\begin{aligned} \bar{J}_F(\mathbf{W}_F, \sigma^2) &= \frac{1}{F} \sum_{k \in \mathcal{K}} \log(1 + \frac{a_k^2}{\sigma^2 d_k} \sum_{f \in \mathcal{F}} w_{f,k}^2 r_f) \\ &+ \frac{1}{F} \sum_{f \in \mathcal{F}} \log(1 + \frac{1}{\sigma^2} \sum_{k \in \mathcal{K}} \frac{1}{d_k} w_{f,k}^2 a_k^2 \tilde{r}_k) - \frac{1}{\sigma^2 F} \sum_{f \in \mathcal{F}} \sum_{k \in \mathcal{K}} \frac{w_{f,k}^2}{d_k} a_k^2 r_f \tilde{r}_k \end{aligned} \quad (6)$$

where $r_f(\mathbf{W}_F, \sigma^2)$ and $\tilde{r}_k(\mathbf{W}_F, \sigma^2)$ are the solutions of

$$\begin{aligned} r_f &= (1 + \frac{1}{\sigma^2} \sum_{k \in \mathcal{K}} \frac{1}{d_k} w_{f,k}^2 a_k^2 \tilde{r}_k)^{-1}, \forall f \in \mathcal{F}, \\ \tilde{r}_k &= (1 + \frac{a_k^2}{\sigma^2 d_k} \sum_{f \in \mathcal{F}} w_{f,k}^2 r_f)^{-1}, \forall k \in \mathcal{K}. \end{aligned} \quad (7)$$

Proof. The proof of the theorem is given in Appendix I wherein we use an Integration by parts formula [38] to derive an expression for the expectation of the mutual information as in (5). Then, we derive an upper-bound for ϵ_F using Nash-Poincaré inequality [38] where the convergence rate of $\mathcal{O}(\frac{1}{d^2})$ is claimed accordingly. The convergence $J_F(\mathbf{W}_F, \sigma^2) - \bar{J}_F(\mathbf{W}_F, \sigma^2) \rightarrow 0$ can be also claimed relying on Girko's law [39, Section 3.2.3] [20, Theorem 6.10]. An alternative proof based on Replica method is also given in [40]. Regarding the convergence rate, it is shown in [41] that in the case where $\mathbf{h}_k, \forall k \in \mathcal{K}$ are Gaussian vectors with given variance profiles, the convergence rate is $\mathcal{O}(\frac{1}{F^2})$. One might be able to obtain the convergence rate declared in the theorem by properly scaling the variances in [41] while ensuring that the assumptions therein remain valid. However, for convenience of the reader and to avoid ambiguity, a straightforward proof of the theorem is presented in Appendix I based on an alternative technique, known as the Gaussian method [20], [38], which is particularly suited to random matrix models with Gaussian entries. \square

According to Theorem 1, the EMI $J_F(\mathbf{W}_F, \sigma^2)$ converges asymptotically to the deterministic equivalent $\bar{J}_F(\mathbf{W}_F, \sigma^2)$ with a convergence rate of $\mathcal{O}(\frac{1}{d^2})$. In the finite scenarios of interest with a moderate number of sub-channels, d_k values can be small relative to F .¹ In such finite cases,

¹Note that while the limiting results are obtained in the asymptotic regime, those can be applied as approximations for the finite scenarios with dimensions as small as 8 and even 4 or 2 [20, Section 2.2.1].

the analysis in Section IV-B shows that the residual term ϵ_F appears as a small incremental gain in the EMI of the sparse spreading scheme, which is dictated mainly by the number of non-zero elements in the codes. Keeping this in mind, we propose an optimization approach as in the following. In Section IV-A, we first formulate an optimization problem to get the set $\bar{\mathcal{C}}^*$ of all power constrained spreading codes, irrespective of the sparsity constraints, which maximize the deterministic EMI $\bar{J}_F(\mathbf{W}_F, \sigma^2)$. This yields a simple resource allocation rule that facilitates the construction of the desired sparse spreading codes via an efficient partitioning algorithm. The details about the partitioning algorithms is delegated to Section V. In Section IV-B, we show that the sparse solutions in $\bar{\mathcal{C}}^*$ additionally harness the small incremental gain inherent in the residual term ϵ_F in the finite regime. The analysis in Section IV eventually yields an upper-bound on the gap to the optimum, which is shown to be close to zero for the sparse solutions in $\bar{\mathcal{C}}^*$.

IV. MAXIMIZING THE ERGODIC MUTUAL INFORMATION

In the sequel, we omit the subscript F denoting the dependency on system size. Also, observe that $\bar{J}(\mathbf{W}, \sigma^2)$ in (6) depends only on squares of $w_{f,k}$ values. Therefore, with a change of variable $v_{f,k} = \frac{1}{d_k} w_{f,k}^2$, hereafter, we express the EMI and the related entities as a function of matrix $\mathbf{V} = [v_{f,k}]_{f \in \mathcal{F}, k \in \mathcal{K}}$. Given a matrix \mathbf{V} , the corresponding spreading vectors $\{\mathbf{w}_k\}$ can be obtained up to an uncertainty in the signs of the entries in the spreading vectors. It is shown in Proposition 1 that the objective function under the considered i.i.d channel model is indifferent to the signs of the spreading code entries. Thus, hereafter, we refer to \mathbf{V} and \mathbf{W} interchangeably as the spreading matrix.

A. The optimal spreading in the asymptotic regime

We first neglect the sparsity constraints, and define $\mathcal{C}_2 \triangleq \{\mathbf{V} | v_{f,k} \in \mathbb{R}^+, \sum_{f \in \mathcal{F}} v_{f,k} \leq P_k, \forall k \in \mathcal{K}\}$ to be the set of all the power constrained spreading matrices \mathbf{V} . Then, we formulate the problem of maximizing the deterministic EMI $\bar{J}(\mathbf{V}, \sigma^2)$ as follows

$$\max_{\mathbf{V} \in \mathcal{C}_2} \bar{J}(\mathbf{V}, \sigma^2) \quad (8a)$$

$$\text{s.t.} \quad \sum_{f \in \mathcal{F}} v_{f,k} \leq P_k, \forall k \in \mathcal{K}, \quad (8b)$$

$$v_{f,k} \geq 0, \forall k \in \mathcal{K}, f \in \mathcal{F}. \quad (8c)$$

This optimization problem yields the set of all spreading codes that maximize the deterministic EMI $\bar{J}(\mathbf{V}, \sigma^2)$ subject to the power constraints and irrespective of the sparsity constraints.

The Karush-Kuhn-Tucker (KKT) conditions [42] are necessary conditions for a matrix \mathbf{V} to be a local optimal solution of the problem in (8). However, the KKT conditions are not necessarily the sufficient conditions. The sufficiency and the globally optimality of the solutions are discussed later. The Lagrangian associated with (8) is given as

$$L(v_{i,j}, \lambda_{i,j}, \delta_j) = -\bar{J}(\mathbf{V}, \sigma^2) - \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{K}} \lambda_{i,j} v_{i,j} + \sum_{j \in \mathcal{K}} \delta_j \left(\sum_{i \in \mathcal{F}} v_{i,j} - P_j \right) \quad (9)$$

where the Lagrangian variables δ_j and $\lambda_{i,j}$ are associated with constraints (8b) and (8c), respectively. The gradient of the Lagrangian can then be evaluated as

$$\nabla_{f,k} L(v_{i,j}, \lambda_{i,j}, \delta_j) = -\frac{\partial \bar{J}(\mathbf{V}, \sigma^2)}{v_{f,k}} - \lambda_{f,k} + \delta_k, \quad \forall f \in \mathcal{F}, k \in \mathcal{K}. \quad (10)$$

Note that $\bar{J}(\mathbf{V}, \sigma^2)$ depends on the entries of \mathbf{V} via $r_f(\mathbf{V}, \sigma^2)$ and $\tilde{r}_k(\mathbf{V}, \sigma^2)$ as in (7) and (6). Since those are the solutions to the saddle point equations, the partial derivatives $\frac{\partial \bar{J}(\mathbf{V}, \sigma^2)}{\partial r_f}$ and $\frac{\partial \bar{J}(\mathbf{V}, \sigma^2)}{\partial \tilde{r}_k}$ are zero at any point given by $(\mathbf{V}, r_f(\mathbf{V}, \sigma^2), \tilde{r}_k(\mathbf{V}, \sigma^2))$.² Therefore, the chain rules of derivatives [43] allow the partial derivative $\frac{\partial \bar{J}(\mathbf{V}, \sigma^2)}{v_{f,k}}$ to be evaluated by assuming r_f and \tilde{r}_k as constants. This, in particular, gives $\frac{\partial \bar{J}(\mathbf{V}, \sigma^2)}{v_{f,k}} = \frac{1}{\sigma^2 F} a_k^2 \tilde{r}_k r_f$, which yields the KKT conditions as

$$\begin{aligned} \lambda_{f,k}^* \geq 0, \quad \lambda_{f,k}^* v_{f,k} = 0, \quad \delta_k^* \geq 0, \quad \delta_k^* \left(\sum_{i \in \mathcal{F}} v_{i,k} - P_k \right) = 0, \quad \forall f \in \mathcal{F}, k \in \mathcal{K}, \\ -\frac{1}{\sigma^2 F} a_k^2 \tilde{r}_k r_f - \lambda_{f,k}^* + \delta_k^* = 0, \quad \forall f \in \mathcal{F}, k \in \mathcal{K} \end{aligned} \quad (11)$$

where $\lambda_{f,k}^*$ and δ_k^* denote the optimal values of the Lagrangian variables. Since $\lambda_{f,k}$ can be solved from the last equation, the KKT conditions can be simplified as

$$\delta_k^* \left(\sum_{i \in \mathcal{F}} v_{i,k} - P_k \right) = 0, \quad \forall k \in \mathcal{K}, \quad (12a)$$

$$\left(\delta_k^* - \frac{1}{\sigma^2 F} a_k^2 \tilde{r}_k r_f \right) v_{f,k} = 0, \quad \forall f \in \mathcal{F}, k \in \mathcal{K}, \quad (12b)$$

$$\delta_k^* \geq \frac{1}{\sigma^2 F} a_k^2 \tilde{r}_k r_f, \quad \forall f \in \mathcal{F}, k \in \mathcal{K}. \quad (12c)$$

According to the KKT conditions, some properties for the optimal solutions to (8) can be summarized as in the following proposition.

Proposition 2. *The spreading matrices that maximize the ergodic mutual information in (8) have the following properties:*

- *The power constraints in (8b) are satisfied with equality for all UEs, i.e., all UEs are active, and transmit with full power.*

²One can verify this by evaluating the partial derivatives using (7) and (6).

- The parameters r_f are equal to r^* , $\forall f \in \mathcal{F}$ where r^* is the solution of the following fixed point iterations

$$r^* = \left(1 + \frac{1}{F} \sum_{k \in \mathcal{K}} \frac{P_k a_k^2}{\sigma^2 + P_k a_k^2 r^*}\right)^{-1}. \quad (13)$$

- The parameters \tilde{r}_k , $\forall k$ are equal to \tilde{r}_k^* , $\forall k$ where

$$\tilde{r}_k^* = \frac{\sigma^2}{\sigma^2 + P_k a_k^2 r^*}, \forall k \in \mathcal{K}. \quad (14)$$

Proof. The proof is given in Appendix II. \square

As a result of Proposition 2, the solutions satisfying the KKT conditions must give $r_f = r^*$, $\forall f \in \mathcal{F}$ and $\tilde{r}_k = \tilde{r}_k^*$, $\forall k \in \mathcal{K}$. Observe that the values of r^* and \tilde{r}_k^* are given independently from the values of $v_{f,k}$. Let $\bar{J}(\mathbf{V}, \tilde{r}_k^*, r^*, \sigma^2)$ denotes a function obtained by plugging the r^* and \tilde{r}_k^* values into (6), i.e.,

$$\begin{aligned} \bar{J}(\mathbf{V}, \tilde{r}_k^*, r^*, \sigma^2) &= -\frac{1}{F} \sum_{k \in \mathcal{K}} \log(\tilde{r}_k^*) - \log(r^*) - \frac{r^*}{\sigma^2 F} \sum_{k \in \mathcal{K}} a_k^2 \tilde{r}_k^* \sum_{f \in \mathcal{F}} v_{f,k}, \\ &= -\frac{1}{F} \sum_{k \in \mathcal{K}} \log(\tilde{r}_k^*) - \log(r^*) - \frac{r^*}{\sigma^2 F} \sum_{k \in \mathcal{K}} P_k a_k^2 \tilde{r}_k^* \end{aligned} \quad (15)$$

where the arguments of the logarithms are replaced with their equivalents from (7), and in the last equality we used the first property from Proposition 2. Observe that the values of $\bar{J}(\mathbf{V}, \tilde{r}_k^*, r^*, \sigma^2)$ depend only on \tilde{r}_k^* and r^* values. Thus, all the solutions of (8), satisfying the KKT conditions, attain the same value of the objective function, i.e., its global maximum value. This proves the sufficiency of the KKT conditions for a spreading matrix \mathbf{V} to be the optimal solution of (8). As a result of the KKT conditions, the spreading codes \mathbf{v}_k , which maximize $\bar{J}(\mathbf{V}, \sigma^2)$ in (8), can be evaluated as the positive solutions of the following indeterminate system of equations

$$\sum_{f \in \mathcal{F}} v_{f,k} = P_k, \forall k \in \mathcal{K}, \quad (16a)$$

$$\sum_{k \in \mathcal{K}} \beta_k v_{f,k} = \frac{1}{r^*} - 1, \forall f \in \mathcal{F} \quad (16b)$$

where $\beta_k = \frac{a_k^2}{\sigma^2 + P_k a_k^2 r^*}$. In these equations, r^* is a fixed scalar, which is evaluated from (13). The equalities in (16b) are obtained by setting $r_f = r^*$, $\forall f$ in (7). These equalities follow since any spreading matrix that gives $r_f = r^*$, $\forall f$, equivalently, satisfies the second and third KKT conditions in (12) as well. The equalities in (16a) are given as a result of the first KKT condition in (12).

The system of equations in (16) has a simple implication. The first line of equalities in (16a) indicates that the UEs need to transmit with full power. The second line implies that the entries $v_{f,k}$ should be assigned such that r_f values become the same across all sub-channels. One can verify that the dense spreading, i.e., $v_{f,k} = \frac{P_k}{F}, \forall f, k$ is a solution of (16). This has been also observed in [24] where the authors show that in a scenario with randomly assigned dense spreading sequences, the frequency-dependency of r_f values vanishes asymptotically. Aside from the dense spreading matrix, sparse spreading matrices can be also designed to satisfy (16). To this end, the non-zero elements $v_{f,k}$ should be assigned to the sub-channels such that the weighted sum $\sum_{k \in \mathcal{K}} \beta_k v_{f,k}$ becomes the same, i.e., equal to $\frac{1}{r^*} - 1$, across all sub-channels. In the symmetric case with $a_k = a$, $P_k = P$ and $d_k = d, \forall k$, the second line of equalities in (16b) becomes $\frac{a^2}{\sigma^2 + Pa^2r^*} \sum_{k \in \mathcal{K}} v_{f,k} = \frac{1}{r^*} - 1, \forall f \in \mathcal{F}$. Given dK/F to be integer, one can verify that any regular spreading matrix with non-zero elements being $\frac{P}{d}$ is a solution to (16). However, in the generic non-symmetric case, an irregular assignment of the non-zero values might arise to ensure the conditions in (16) to hold. Such an assignment for the generic case is done in Section V via a simple partitioning algorithm.

B. On the optimality of the asymptotic sparse spreading codes

Based on the analysis in Section IV-A, we know that the power-constrained spreading codes in \mathcal{C}_2 that maximize the deterministic equivalent $\bar{J}(\mathbf{V}, \sigma^2)$ in (8) are given as the solutions of the system of equations in (16). This set of solutions has been defined as $\bar{\mathcal{C}}^*$. Hereafter, we use $\bar{\mathbf{V}}^*$ to refer to a member of the set $\bar{\mathcal{C}}^*$. The analysis in Theorem 1 shows that the residual term ϵ vanishes with a rate inversely proportional to the square of the number of non-zero elements in the codes. Thus, the solutions in $\bar{\mathcal{C}}^*$ attain the maximum of EMI $J(\mathbf{V}, \sigma^2)$ in the asymptotic regime. However, in the finite regime, ϵ term appears as a small incremental gain in the EMI formulation, which needs to be considered. Let \mathbf{V}_d^* to be the unknown optimal spreading matrix that maximizes $J(\mathbf{V}, \sigma^2)$ subject to the sparsity constraints. Also, let $J_d^* \triangleq J(\mathbf{V}_d^*, \sigma^2)$ to be the maximum of $J(\mathbf{V}, \sigma^2)$ attained by \mathbf{V}_d^* . Now, the penalty when using any spreading matrix $\bar{\mathbf{V}}^* \in \bar{\mathcal{C}}^*$, given as a solution of the system of equations in (16), instead of the optimal one \mathbf{V}_d^* can be written as

$$\Delta_d = \underbrace{J(\mathbf{V}_d^*, \sigma^2)}_{J_d^*} - J(\bar{\mathbf{V}}^*, \sigma^2). \quad (17)$$

Writing $J(\mathbf{V}_d^*, \sigma^2)$ and $J(\bar{\mathbf{V}}^*, \sigma^2)$ as the deterministic equivalents plus the residual terms, we get

$$J(\mathbf{V}_d^*, \sigma^2) - \bar{J}(\mathbf{V}_d^*, \sigma^2) = \epsilon^{(1)} \quad (18a)$$

$$J(\bar{\mathbf{V}}^*, \sigma^2) - \bar{J}(\bar{\mathbf{V}}^*, \sigma^2) = \epsilon^{(2)} \quad (18b)$$

where the additional index i in $\epsilon^{(i)}$ is added to distinguish the above differences. Subtracting the sides of the above equalities, and rearranging the terms, we get

$$\underbrace{(J(\mathbf{V}_d^*, \sigma^2) - J(\bar{\mathbf{V}}^*, \sigma^2))}_{\Delta_d} + (\bar{J}(\bar{\mathbf{V}}^*, \sigma^2) - \bar{J}(\mathbf{V}_d^*, \sigma^2)) = \Delta_\epsilon$$

where $\Delta_\epsilon = \epsilon^{(1)} - \epsilon^{(2)}$. Since the subtraction $(\bar{J}(\bar{\mathbf{V}}^*, \sigma^2) - \bar{J}(\mathbf{V}_d^*, \sigma^2))$ in the left-hand is positive³, it can be claimed that the gap to the optimum Δ_d is bounded as

$$0 \leq \Delta_d \leq \Delta_\epsilon^+ \quad (19)$$

where Δ_ϵ^+ denotes the positive values of Δ_ϵ . In general, Δ_d may attain negative values. However, $J_d^* < J(\bar{\mathbf{V}}^*, \sigma^2)$ can happen only when $\bar{\mathbf{V}}^*$ violates the sparsity constraints, which is not the case of interest. Therefore, Δ_d is lower bounded by zero. Next, we look into the properties of the residual terms in the finite regime to characterize the gap to the optimum as given in (19).

Generally speaking, the residual term can be roughly associated with the concentration of UEs' powers in a fewer number of the elements in the spreading codes. This gives rise to the variance of the random channel entries, which eventually appears as ϵ in the EMI formulation. In order to get further insight into the structure of ϵ term, a numerical example is illustrated in Fig. 2. The results are generated based on the simulation assumptions given in Section VI-A with $F = 50$, $K = 100$, and $d_K = d, \forall k$. For generating the results in Fig. 2, a randomly selected drop of UEs is taken, and the mean and the variance of the residual term $\epsilon = J(\mathbf{V}_d, \sigma^2) - \bar{J}(\mathbf{V}_d, \sigma^2)$ are evaluated over 1000 randomly selected spreading matrices $\mathbf{V}_d \in \mathcal{C}_1$. The subscript d is used to emphasize that the spreading codes in the columns of \mathbf{V}_d have d non-zero values. Fig. 2 shows the mean and the variance of ϵ values along with $J(\mathbf{V}_d, \sigma^2)$ and $\bar{J}(\mathbf{V}_d, \sigma^2)$ versus the number of non-zero elements in the codes. The variance of ϵ , depicted by the bars in the figure, is magnified 20 times for better illustration. The first observation is that the mean values of the ϵ term are distinct for different d values. Moreover, the small variance of the residual term indicates that the values of ϵ do not vary abruptly among the power constrained spreading vectors with the

³Note that $\bar{J}(\bar{\mathbf{V}}^*, \sigma^2)$ is the maximum of the objective function in (8).

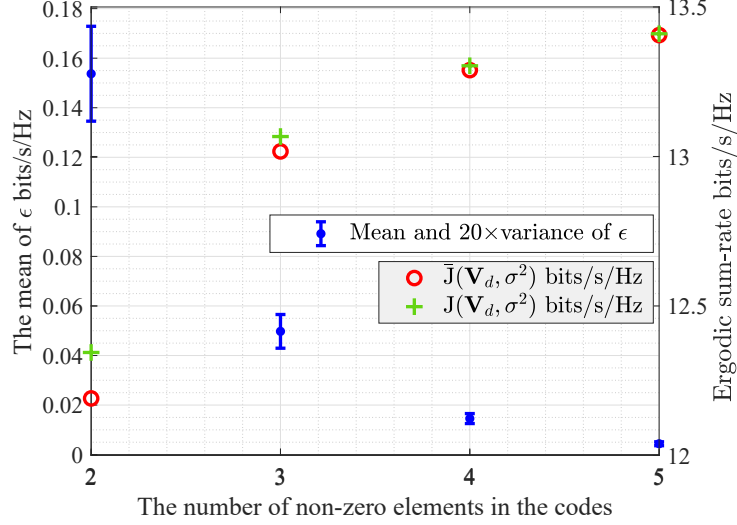


Fig. 2: Mean&Var of ϵ (left axis), and ergodic sum-rate (right axis) vs. d for random spreading.

same number of non-zero elements. Thus, one might infer that the values of the residual term is dictated mainly by the number of non-zero elements in the spreading codes, which implies a decline in ϵ values as d increases. The other observation is that the residual term is small as compared to $\bar{J}(\mathbf{V}_d, \sigma^2)$. Generally speaking, in the scenarios with a moderate number of sub-channels F as in Fig. 2, the residual term is larger for the sparse spreading schemes with $d \ll F$. While it almost disappears in the dense spreading case with $d = F$. This motivates the small incremental gain in ϵ for sparse spreading to be interpreted as sparsity gain. Note that there is a small loss associated with spreading in the considered system model.⁴ Thus, the sparsity gain is coined here to reflect the increment in EMI due to sparse spreading as compared to dense spreading. Recall the symmetric scenario mentioned in Section IV-A wherein both the sparse-regular and the dense spreading matrices were among the solutions maximizing the deterministic $\bar{J}(\mathbf{V}, \sigma^2)$ in (8). According to the above discussion, we expect that the exact EMI $J(\mathbf{V}, \sigma^2)$ for the sparse-regular codes to be better of that in the dense spreading scheme by an amount of ϵ . Note that ϵ for the dense spreading scheme is negligible. This has been also observed in [13] under the symmetric AWGN channel model. It is shown in [13] that the sparse-regular spreading codes yield slightly higher spectral efficiency as compared to the dense spreading scheme in the symmetric scenario considered therein.

To conclude this section, let us recall the Δ_d formulation in (19), which gives the gap to the optimum J_d^* when using a solution $\bar{\mathbf{V}}^* \in \bar{\mathcal{C}}^*$ in the finite regime. Relying on the above analysis, we expect the $\epsilon^{(i)}, i = 1, 2$ terms to be small as compared to the corresponding deterministic parts.

⁴ See [44] for the spreading coding trade-off.

Also, a solution $\bar{\mathbf{V}}^*$ that satisfies the sparsity constraints is expected to harness an incremental $\epsilon^{(2)}$ value close to the $\epsilon^{(1)}$ value. Therefore, the anticipated performance gap Δ_d is close to zero. Next, we find the desired sparse solutions in $\bar{\mathcal{C}}^*$ via an algorithmic solution.

V. AN ALGORITHM FOR CONSTRUCTING THE SPARSE SPREADING MATRICES

The problem of finding the subset of solutions in $\bar{\mathcal{C}}^*$ with the desired sparsity includes a zero-norm. The discrete and discontinuous nature of the zero-norm impedes the application of standard convex analysis [45]. Fortunately, the system of equations in (16) unveils a simple rule for the allocation of the spreading codes. This allows the desired sparse codes to be obtained using an alternative algorithmic solution. We are interested in determining the sparse spreading codes that satisfy (16). Let the elements $v_{f,k}$ for each UE k to be taken from $\{0, \frac{P_k}{d_k}\}$ and subject to the power constraint enforced by equalities in (16a). Based on (16b), the problem is to allocate $\frac{P_k}{d_k}\beta_k$ values to sub-channels such that the sums of $\frac{P_k}{d_k}\beta_k$ values on each sub-channel become the same, i.e., equal to $(\frac{1}{r^*} - 1)$, across all the sub-channels. This problem falls within a class of partitioning problems that arises in number theory and computer science [46]. Although the partitioning problem is NP-complete, there are heuristics that solve the problem in many instances, either optimally or approximately [47]. One such approach is the greedy algorithm, which iterates through $\frac{P_k}{d_k}\beta_k$ values in descending order, assigning each of them to whichever sub-channel has the smallest sum [48]. These steps are summarized in Alg. 1.

Algorithm 1 Partitioning solution

- 1: Divide the total power of each UE k into d_k power fragments.
 - 2: Set $v_{f,k} = 0, \forall f, k$, and $\mathcal{J} = \{1, \dots, K\}$.
 - 3: **while** \mathcal{J} is non-empty **do**
 - 4: Set $k = \arg \max_{j \in \mathcal{J}} \frac{P_j}{d_j} \beta_j$.
 - 5: Set $f = \arg \min_{i \in \mathcal{F}} \eta_i$ with $\eta_i = \sum_{j \in \mathcal{K}} \beta_j v_{i,j}$.
 - 6: Set $v_{f,k} = P_k/d_k$
 - 7: **if** $\sum_{i \in \mathcal{F}} v_{i,k} = P_k$, i.e., UE k has assigned all of its power-fragments **then**
 - 8: Remove index k from \mathcal{J} .
 - 9: **end if**
 - 10: **end while**
-

In Alg. 1, we try to make the sum terms η_f across the sub-channels as equal as possible. Let η_{\max}^* denotes the maximum of $\eta_f, \forall f \in \mathcal{F}$ in an optimal partitioning solution. Alg. 1 yields η_f values such that $\frac{\max(\eta_f)}{\eta_{\max}^*} \leq \frac{4}{3} - \frac{1}{3F}$ [48]. One can always improve the homogeneity of η_f

values by offloading the power from the sub-channels with largest η_f values into those with corresponding smallest values until (16b) holds up to the desired accuracy. However, numerical analysis shows that Alg. 1 yields satisfactory results in most cases, and thus, further fine-tuning steps are redundant. Alg. 1 has a running time of $\mathcal{O}(2^F(d_{\max}K)^2)$ [48].

VI. NUMERICAL RESULTS

A. Simulation assumptions

The simulation results are generated in a scenario where a single-antenna BS serves single-antenna UEs in uplink. Transmit power of each UE is 1 Watt, and the noise power is set to -120dB . The pathloss values are taken randomly and uniformly from the range of -150dB to -60dB to account for diverse received SNRs at the BS. The final channel gains are given by the product of the pathlosses and the small-scale fading entries as in (1). To keep the results comparable, the numbers of non-zero elements in the spreading codes are assumed to be the same for all UEs, i.e., $d_k = d, \forall k$. The number of sub-channels is $F = 50$ unless mentioned otherwise.

In addition to the method proposed in Section V, two alternative spreading schemes from the literature are considered. The first one is the random spreading method that allocates the power constrained sparse spreading codes to UEs randomly. A practically useful property of the random spreading scheme is that no coordination overhead is imposed. The other scheme is the coordinated regular spreading, which assigns the non-zero elements in the codes in a way that each UE occupies a number of d resources and each resource is used by a number of $\frac{K}{F}d$ UEs. The ratio $\frac{K}{F}$ is chosen such that $\frac{K}{F}d$ be an integer. In evaluating the performance of the considered methods, we average the corresponding attainable rates over 1000 UE drops, where, in each drop, the expectation involved in $\text{EMI } J(\mathbf{V}, \sigma^2)$ is evaluated over 1000 random realizations of small-scale fading. The deterministic equivalent values $\bar{J}(\mathbf{V}, \sigma^2)$ are evaluated from (6).

B. Evaluation of the performance of the proposed method

The assessment of an optimal solution to the problem in (4) requires exhaustive search over all the power constrained spreading matrices with the desired sparsity. This impedes the comparison of the results to the optimum. However, still we can evaluate and compare the performance in Alg. 1 and in the aforementioned uncoordinated random and the coordinated regular spreading schemes. Note that while the uncoordinated assignment of the spreading codes is a useful property, the random spreading causes a significant performance degradation. On the other hand, the coordinated allocation of spreading codes in the regular scheme does not consider the

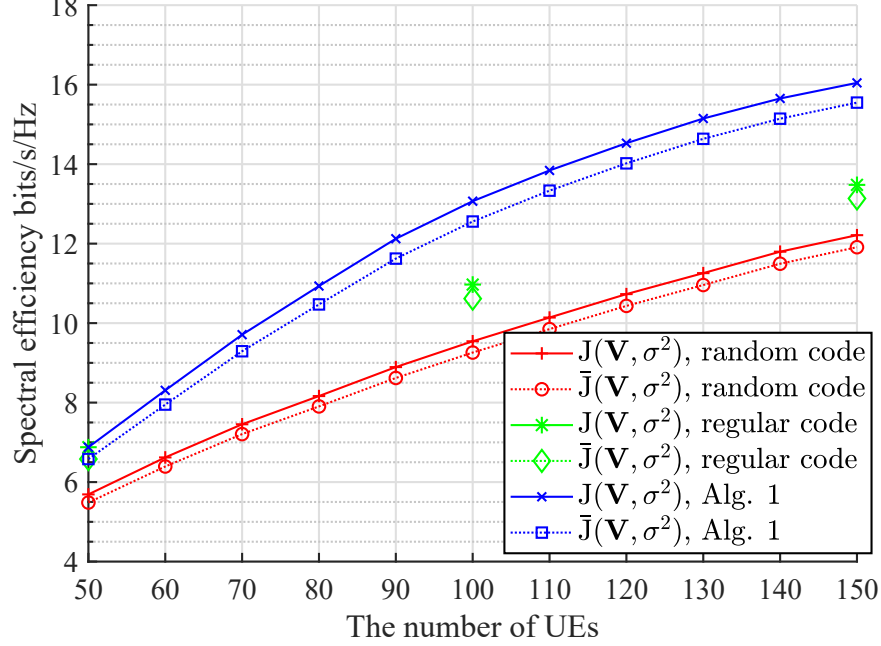


Fig. 3: The spectral efficiency $J(\mathbf{V}, \sigma^2)$ vs. the number of UEs, $d = 1$, $F = 50$.

asymmetry in the system model due to the diverse pathloss values and power constraints at the UEs. On the contrary, the spreading codes in Alg. 1 are allocated according to the UEs' pathloss values. This results in the minimal coordination requirement since the statistical properties of the channel matrix can be assumed to remain constant for a sufficiently large number of reception phases [32].

Figures 3 and 4 show the attainable rates (bits/s/Hz) for the aforementioned methods as a function of the number of UEs. In Fig. 3, we apply Alg.1 to the special non-spreading case with $d = 1$ as well. Note that Alg.1 allocates the sparse codes to UEs such that the deterministic equivalent of EMI is maximized. The motivation therein is that the residual term is small relative to the deterministic equivalent part, and the small gain in ϵ is harnessed inherently due to the sparsity of the allocated codes. While we expect ϵ to be relatively small for the cases with $d > 1$, due to the fast convergence rate of $\mathcal{O}(\frac{1}{d^2})$, the analysis in the non-spreading case with $d = 1$ may be considered as a heuristic attempt. Interestingly, the difference between $J(\mathbf{V}, \sigma^2)$ and $\bar{J}(\mathbf{V}, \sigma^2)$ is relatively small even in such a case, and the coordinated allocation of resources gives 20% and 35% enhancement in the spectral efficiency at 100% and 300% system load, respectively. The system load is defined as the ratio of $\frac{K}{F}$ in percentage. Fig. 4 shows the rates for the case with LDS codes having $d = 2$ non-zero values. In this case, the gain in the coordinated assignment of spreading codes is about 6.5% and 11% at 100% and 300% load, respectively, which is less than

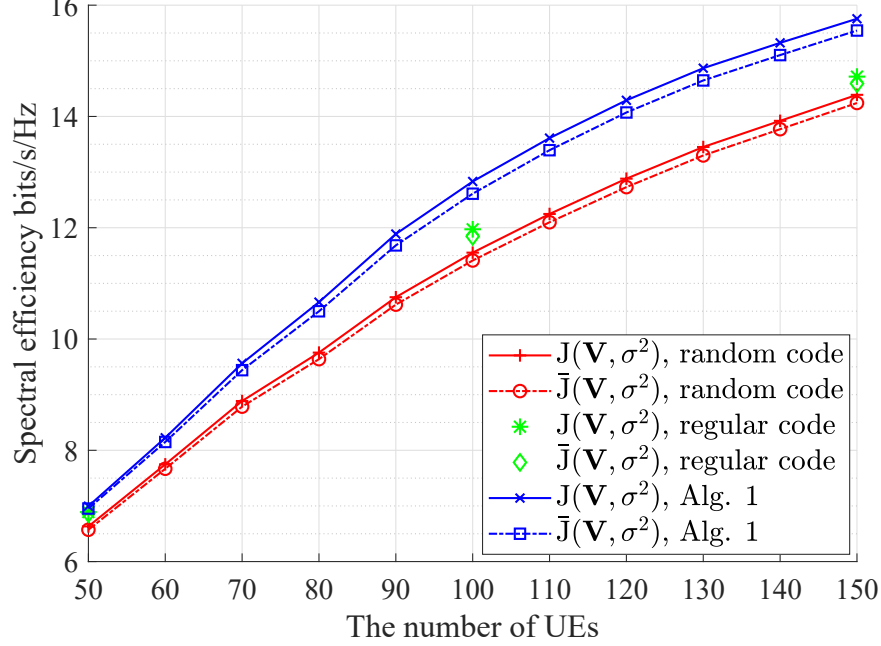


Fig. 4: The spectral efficiency $J(\mathbf{V}, \sigma^2)$ vs. the number of UEs, $d = 2$, $F = 50$.

that in Fig. 3. The other observation is that the regular spreading method in both of Figs. 3 and 4 gives slightly better spectral efficiency as compared to the random spreading scheme, however, its performance is inferior to that of Alg. 1.

In Fig. 5, the attainable SEs are presented versus the number of non-zero elements d , for a system load of 300%. It can be seen that the performance of the random spreading method improves as d grows larger. Spreading on more sub-channels allows UEs to attain interference diversity. This, in general, reduces the loss imposed by the uncoordinated resource allocation. Note that, even though one can enhance the spectral efficiency of the uncoordinated method by spreading on further sub-channels, the number of UEs overlapping on the same sub-channel increases correspondingly. In a system with 300% load, the average number of overlapping UEs in the case with $d = 2$ and $d = 6$ is equal to 6 and 18, respectively. Thus, the detection complexity is greatly increased with larger values of d . In Fig. 5, the spectral efficiency of the dense spreading scheme is also depicted. As mentioned in Section IV-A, the dense spreading matrix is a solution of the optimization problem in (8), and thus, the values of deterministic $\bar{J}(\mathbf{V}, \sigma^2)$ for both the dense spreading scheme and the sparse spreading in Alg. 1 are the same. However, the values of the exact EMI $J(\mathbf{V}, \sigma^2)$ for the sparse spreading in Alg. 1 are better of that in the dense spreading scheme by an amount of ϵ . This can be seen from Fig. 5 where the curves of $\bar{J}(\mathbf{V}, \sigma^2)$ and $J(\mathbf{V}, \sigma^2)$ are almost overlapping for the dense spreading case, while

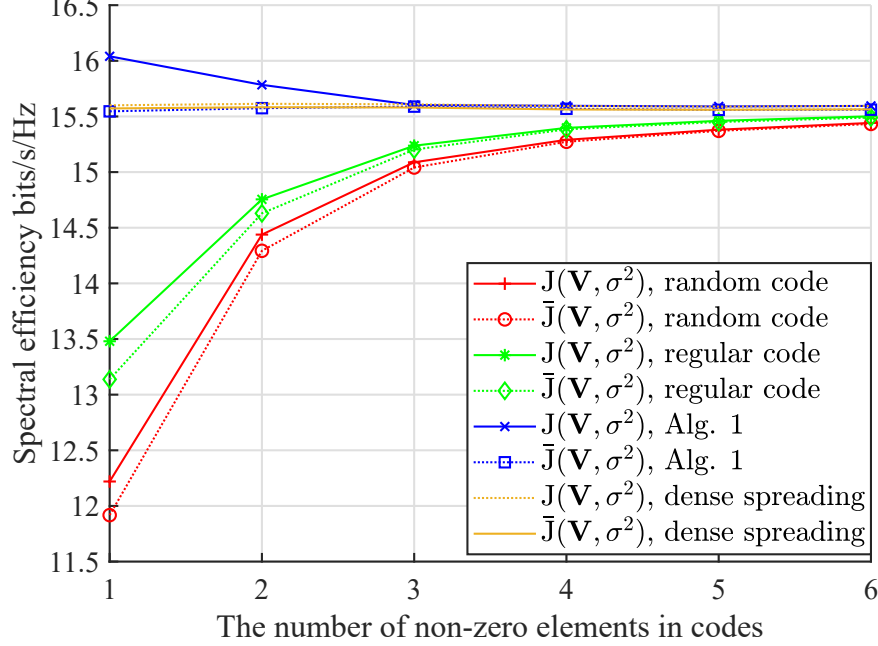


Fig. 5: The spectral efficiency $J(\mathbf{V}, \sigma^2)$ vs. d , $F = 50$, $\frac{K}{F} = 3$.

the value of $J(\mathbf{V}, \sigma^2)$ for Alg. 1 at $d = 2$ is nearly 0.2 bits/s/Hz higher than the deterministic EMI. This additional gain in $J(\mathbf{V}, \sigma^2)$ for sparse spreading as compared to dense spreading was referred to as sparsity gain in Section IV-B. Observe that the sparsity gain decreases as the number of non-zero elements in the codes increases. Finally, we observe that, in contrast to the symmetric model⁵, the regular spreading method is inferior to Alg.1 and even to the dense spreading scheme in the considered asymmetric scenario. As mentioned in Section IV-B, the regular spreading matrix is an optimal solution to the optimization problem in (8) in the symmetric scenarios.

C. Visualization of the resource allocation strategy in the proposed method

In Fig. 6, the resource allocation strategy in Alg. 1 is visualized in a scenario with $F = 30$, $K = 90$, and $d = 2$. This figure illustrates the allocation of the spreading codes in Alg. 1 for a particular drop of UEs. The colorbar represents the unitless β_k values, introduced in (16). It can be seen that the power fragments are allocated to the sub-channels such that the sum terms $\eta_f = \sum_{k \in \mathcal{K}} \beta_k v_{f,k}$ become equal. Observe also that the UEs with large β_k values are distributed across the sub-carriers. Then, those with smaller β_k values are placed such that the sum terms $\eta_f = \sum_{k \in \mathcal{K}} \beta_k v_{f,k}$ become equalized. It can be seen from Fig. 6 that the UEs overlapping on each sub-carriers have diverse β_k values.

⁵See Section IV-B for definition of symmetric system model

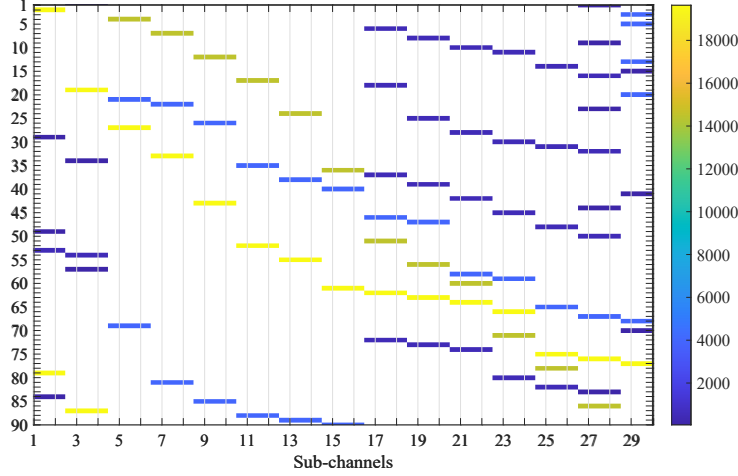


Fig. 6: Visualization of the resource allocation strategy in Alg. 1, $F = 30$, $K = 90$, $d = 2$.

VII. CONCLUSIONS

In this paper, a simple and efficient rule for close-to-optimal allocation of sparse spreading codes was derived based on rigorous analysis. The analysis reduced the dilemma of maximizing the ergodic mutual information to a partitioning problem, which was solved via an efficient algorithmic solution. The proposed algorithm allocates the spreading codes based on the system load, the sparsity constraints and pathloss values. The simulation results showed that the proposed algorithm with minimal coordination provides a superior performance as compared to the uncoordinated random spreading and the coordinated regular spreading schemes. It was shown that the regular spreading matrices are asymptotically optimal only in the symmetric system model, while in the asymmetric case the performance of the regular spreading method is even inferior to the dense spreading scheme. As the future work, we are interested in extending the results to the multi-antenna BS scenario with correlation introduced among the antenna elements and among the sub-carriers. Following the same optimization approach as in here, we expect the optimal low density spreading policies to be characterized based on the spatial and the spectral correlation properties of the channel matrix.

APPENDIX I

PROOF OF THEOREM 1

The ergodic mutual information $J_F(\mathbf{W}_F, \sigma^2)$ is related to $\frac{1}{F} \mathbb{E} \text{Tr}(\frac{1}{\sigma^2} \mathbf{H}_F \mathbf{H}_F^H + \mathbf{I}_F)^{-1}$ as follows

$$\frac{\partial J_F(\mathbf{W}_F, \sigma^2)}{\partial \sigma^2} = \frac{1}{\sigma^2 F} \mathbb{E} \text{Tr}(\frac{1}{\sigma^2} \mathbf{H}_F \mathbf{H}_F^H + \mathbf{I}_F)^{-1} - \frac{1}{\sigma^2}. \quad (20)$$

Denoting $\mathbf{Q}_F(\sigma^2) = (\frac{1}{\sigma^2}\mathbf{H}_F\mathbf{H}_F^H + \mathbf{I}_F)^{-1}$, equivalently we have

$$J_F(\mathbf{W}_F, \sigma^2) = \int_{\sigma^2}^{+\infty} \left(\frac{1}{z} - \frac{1}{zF} \mathbb{E} \text{Tr} \mathbf{Q}_F(z) \right) dz. \quad (21)$$

Thus, given an expression for the trace term $\mathbb{E} \text{Tr} \mathbf{Q}_F(z)$, $\forall z \in \mathbb{R}^+$, one can equivalently derive an expression for EMI based on (21). In derivation of the results, we utilize so-called Gaussian method [20], [38]. Let $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta})$ be a circularly symmetric Gaussian random vector with covariance matrix $\mathbf{\Theta} \in \mathbb{C}^{N \times N}$. Also, let the function $f(\mathbf{x}, \mathbf{x}^*) \in \mathbb{C}$ together with its derivatives be polynomial bounded. The Gaussian method consists of two ingredients [38]:

- Integration by parts formula: $\mathbb{E} x_j f(\mathbf{x}) = \sum_{i=1}^N [\mathbf{\Theta}]_{j,i} \mathbb{E} \frac{\partial f(\mathbf{x})}{\partial x_i^*}$,
- Nash-Poincaré inequality: $\text{var}(f(\mathbf{x})) \leq \mathbb{E} \nabla_{\mathbf{x}} f(\mathbf{x})^T \mathbf{\Theta} (\nabla_{\mathbf{x}} f(\mathbf{x}))^* + \mathbb{E} (\nabla_{\mathbf{x}^*} f(\mathbf{x}))^H \mathbf{\Theta} \nabla_{\mathbf{x}^*} f(\mathbf{x})$.

In the sequel, we show that a diagonal element $q_{p,p}$ of matrix \mathbf{Q}_F can be written as the sum of $h_{i,j} q_{p,i} h_{p,j}^*$ terms. Defining the function $f(\mathbf{H}_F) \triangleq q_{p,i} h_{p,j}^*$, then, we expand the terms of type $\mathbb{E} h_{i,j} f(\mathbf{H}_F)$ using the Gaussian integration by parts formula. This retrieves an implicit but deterministic expression for $\mathbb{E} \text{Tr}(\mathbf{Q}_F)$ up to a small residual term. Then, Nash-Poincaré inequality allows us to derive an upper bound on the residual term and declares the convergence of the trace term under the realms of Theorem 1 as

$$\frac{1}{F} \mathbb{E} \text{Tr} \mathbf{Q}_F(\sigma^2) = \frac{1}{F} \text{Tr} \mathbf{R}_F(\mathbf{W}_F, \sigma^2) + \zeta_F(\sigma^2) \quad (22)$$

where $\mathbf{R}_F(\mathbf{W}_F, \sigma^2)$ is as defined in the theorem, and $\zeta_F(\sigma^2) = \mathcal{O}(\frac{1}{d^2})$ is a fast diminishing term. Given the deterministic equivalent for $\frac{1}{F} \mathbb{E} \text{Tr} \mathbf{Q}_F(\sigma^2)$ as in (22), the convergence of EMI $J_F(\mathbf{W}_F, \sigma^2)$ to the deterministic equivalent $\bar{J}_F(\mathbf{W}_F, \sigma^2)$, as declared in Theorem 1, follows from the relation in (21). In the following, we proof the convergence of the trace term in (22). The proof of the convergence of EMI is straightforward, and thus, is omitted. We invite the reader to refer to [38, Theorem 1] for the details. In the sequel, we frequently omit the subscript F denoting the dependency of the entities on the system size. In proving the convergence of the trace term, we frequently use the following elementary results,

$$\frac{\partial q_{p,d}}{\partial h_{i,j}} = -[\mathbf{Q}(\frac{\partial \mathbf{Q}^{-1}}{\partial h_{i,j}}) \mathbf{Q}]_{p,d} = -\frac{1}{\sigma^2} [\mathbf{Q}(\frac{\partial \mathbf{H} \mathbf{H}^H}{\partial h_{i,j}}) \mathbf{Q}]_{p,d} \quad (23)$$

$$= -\frac{1}{\sigma^2} q_{p,i} [\mathbf{h}_j^H \mathbf{Q}]_d, \quad (24)$$

and similarly,

$$\frac{\partial q_{p,d}}{\partial h_{i,j}^*} = -\frac{1}{\sigma^2} [\mathbf{Q} \mathbf{h}_j]_p q_{i,d}. \quad (25)$$

We start by noticing that $\mathbf{Q} = \mathbf{I}_F - \frac{1}{\sigma^2} \mathbf{Q} \mathbf{H} \mathbf{H}^H$, a relation often referred to as the resolvent identity. This allows one to write $\mathbb{E}q_{p,p}$ as a function of $\mathbb{E}[\mathbf{Q} \mathbf{H} \mathbf{H}^H]_{p,p}$, i.e.,

$$\mathbb{E}q_{p,p} = 1 - \frac{1}{\sigma^2} \sum_{j=1}^K \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^*, \quad p \in \mathcal{F}. \quad (26)$$

Now, we work on term $\mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* = \sum_{i=1}^F \mathbb{E}q_{p,i} h_{i,j} h_{p,j}^*$ to expand it using the Gaussian integration by parts formula. Let the function $f(\mathbf{H})$ to be defined as $f(\mathbf{H}) \triangleq q_{p,i} h_{p,j}^*$. Then from the integration by parts formula, we get

$$\mathbb{E}h_{i,j} q_{p,i} h_{p,j}^* = \mathbb{E}h_{i,j} f(\mathbf{H}) = a_j^2 v_{i,j} \mathbb{E} \frac{\partial f(\mathbf{H})}{\partial h_{i,j}^*} \quad (27a)$$

$$= a_j^2 v_{i,j} \mathbb{E} \left(\frac{\partial q_{p,i}}{\partial h_{i,j}^*} h_{p,j}^* + \frac{\partial h_{p,j}^*}{\partial h_{i,j}^*} q_{p,i} \right) \quad (27b)$$

$$= -\frac{1}{\sigma^2} a_j^2 v_{i,j} \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p q_{i,i} h_{p,j}^* + a_j^2 v_{i,j} \delta(i-p) \mathbb{E}q_{p,i} \quad (27c)$$

where $v_{i,j} \triangleq \frac{1}{d_j} w_{i,j}^2$. Summing the sides of above equality over index i , we get

$$\mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* = -\frac{1}{\sigma^2} \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* a_j^2 \text{Tr}(\mathbf{Q} \mathbf{V}_j) + a_j^2 v_{p,j} \mathbb{E}q_{p,p} \quad (28)$$

where $\mathbf{V}_j = \text{diag}\{\mathbf{v}_j\}$ with \mathbf{v}_l defined as $\mathbf{v}_l \triangleq [v_{1,l}, \dots, v_{F,l}]$, $\forall l$. Let us define $\beta_j = a_j^2 \text{Tr}(\mathbf{Q} \mathbf{V}_j)$, $\alpha_j = \mathbb{E}\beta_j$ and $\overset{\circ}{\beta}_j = \beta_j - \alpha_j$. Then, the terms in the above equation can be separated as

$$\mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* = -\frac{1}{\sigma^2} \alpha_j \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* + a_j^2 v_{p,j} \mathbb{E}q_{p,p} - \frac{1}{\sigma^2} \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* \overset{\circ}{\beta}_j. \quad (29)$$

Solving the above equality for $\mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^*$, we get

$$\mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* = \tilde{e}_j a_j^2 v_{p,j} \mathbb{E}q_{p,p} - \tilde{e}_j \frac{1}{\sigma^2} \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* \overset{\circ}{\beta}_j \quad (30)$$

where $\tilde{e}_j \triangleq \frac{\sigma^2}{\sigma^2 + \alpha_j}$. Summing the sides of above equality over index j , we get

$$\mathbb{E}[\mathbf{Q} \mathbf{H} \mathbf{H}^H]_{p,p} = \text{Tr}(\mathbf{\Lambda}_p \mathbf{A} \tilde{\mathbf{E}}) \mathbb{E}q_{p,p} - \frac{1}{\sigma^2} \sum_{j=1}^K \tilde{e}_j \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* \overset{\circ}{\beta}_j \quad (31)$$

where $\mathbf{\Lambda}_p = \text{diag}\{v_{p,1}, \dots, v_{p,K}\}$, $\mathbf{A} = \text{diag}\{a_1^2, \dots, a_K^2\}$, $\tilde{\mathbf{E}} = \text{diag}\{\tilde{e}_1, \dots, \tilde{e}_K\}$, and $\overset{\circ}{\boldsymbol{\beta}} = \text{diag}\{\overset{\circ}{\beta}_1, \dots, \overset{\circ}{\beta}_K\}$.

Utilizing the resolvent identity, (26) and (31) yield

$$\sigma^2 - \sigma^2 \mathbb{E}q_{p,p} = \tilde{\alpha}_p \mathbb{E}q_{p,p} - \frac{1}{\sigma^2} \sum_{j=1}^K \tilde{e}_j \mathbb{E}[\mathbf{Q} \mathbf{h}_j]_p h_{p,j}^* \overset{\circ}{\beta}_j \quad (32)$$

where $\tilde{\alpha}_p = \text{Tr}(\mathbf{\Lambda}_p \mathbf{A} \tilde{\mathbf{E}})$. Now, one can solve (32) to obtain $\mathbb{E}q_{p,p}$. Let \mathbf{T} be a $F \times F$ diagonal non-negative matrix with bounded spectral norm. Multiplying the acquired $\mathbb{E}q_{p,p}$ from (32) by elements of \mathbf{T} , and summing over p yields

$$\frac{1}{F} \text{Tr}(\mathbf{TQ}) = \frac{1}{F} \text{Tr}(\mathbf{TE}) + \frac{1}{\sigma^4 F} \sum_{j=1}^K \tilde{e}_j \mathbb{E} \mathbf{h}_j^H \mathbf{TEQ} \mathbf{h}_j \overset{o}{\beta}_j \quad (33)$$

where $\mathbf{E} = \text{diag}\{e_1, \dots, e_K\}$ with $e_p \triangleq \frac{\sigma^2}{\sigma^2 + \tilde{\alpha}_p}$. Next in Appendix I-A, we prove that the last term in the above equation, hereafter denoted by $\zeta(\sigma^2)$, vanishes with $\mathcal{O}(\frac{1}{d^2})$ rate. As the result, we get the convergence $\frac{1}{F} \text{Tr}(\mathbf{TQ}) - \frac{1}{F} \text{Tr}(\mathbf{TE}) \rightarrow 0$. However, notice that the term $\text{Tr}(\mathbf{TE})$ still depends on the unknown parameters $\alpha_j = a_j^2 \mathbb{E} \text{Tr}(\mathbf{V}_j \mathbf{Q})$. Therefore, in the last step of the proof given in Appendix I-B, we need to show that the matrix \mathbf{E} can be replaced by the deterministic matrix \mathbf{R} , as defined in the theorem.

A. The upper-bound on $\zeta(\sigma^2)$ term

In proving the convergence rate, we first derive an upper-bound for $|\zeta(\sigma^2)|$ in terms of the variance of $\text{Tr}(\mathbf{Q})$. Then, we show that the upper-bound vanishes with $\mathcal{O}(\frac{1}{d^2})$ rate. observing that $\beta_j = a_j^2 \text{Tr}(\mathbf{V}_j \mathbf{Q})$, the $\zeta(\sigma^2)$ term, given as the last term in (33), can be presented as in the following

$$|\zeta(\sigma^2)| = \frac{1}{\sigma^4 F} \left| \sum_{i=1}^F \mathbb{E} \overset{o}{q}_{i,i} \sum_{j=1}^K v_{i,j} a_j^4 \tilde{e}_j \mathbf{h}_j^H \mathbf{TEQ} \mathbf{h}_j \right| \quad (34a)$$

$$\leq \frac{1}{\sigma^4 F} \sum_{i=1}^F \left| \mathbb{E} \overset{o}{q}_{i,i} \sum_{j=1}^K v_{i,j} a_j^4 \tilde{e}_j \mathbf{h}_j^H \mathbf{TEQ} \mathbf{h}_j \right| \quad (34b)$$

$$\leq \frac{1}{\sigma^4 d F} \sum_{i=1}^F \left| \mathbb{E} \|\tilde{\mathbf{E}}\| \|\mathbf{A}\|^2 \overset{o}{q}_{i,i} \sum_{j=1}^K \mathbf{h}_j^H \mathbf{TEQ} \mathbf{h}_j \right| \quad (34c)$$

where $\overset{o}{q}_{i,i} = q_{i,i} - \mathbb{E}q_{i,i}$, $\mathbf{A} = \text{diag}\{a_1^2, \dots, a_K^2\}$, and (34b) is due to the triangle inequality. Let $\phi \triangleq \text{Tr}(\mathbf{Q})$, then, using the resolvent identity we get

$$|\zeta(\sigma^2)| \leq \frac{C}{dF} \sum_{i=1}^F \left| \mathbb{E}(\overset{o}{q}_{i,i} \text{Tr}(\mathbf{TE}(\mathbf{I}_F - \mathbf{Q}))) \right| \quad (35a)$$

$$\leq \frac{C}{dF} \text{var}(\phi) \quad (35b)$$

where C is a generic constant independent of the system size, and (35a) follows since the matrices $\tilde{\mathbf{E}}, \mathbf{A}, \mathbf{T}, \mathbf{E}$ have bounded spectral norms. In (35b) we used the fact that $\overset{o}{q}_{i,i}$ terms are zero mean, where then, the positive correlations among the entries yields the last inequality.

Next, we use Nash-Poincaré inequality to find an upper-bound for the variance of $\phi = \text{Tr}(\mathbf{Q})$.

In particular, observing that $|\frac{\partial \phi}{\partial h_{i,j}}| = |\frac{\partial \phi}{\partial h_{i,j}^*}|$, Nash-Poincaré inequality yields,

$$\text{var}(\phi) \leq 2 \sum_{i=1}^F \sum_{j=1}^K a_j^2 v_{i,j} \mathbb{E} \left| \frac{\partial \text{Tr}(\mathbf{Q})}{\partial h_{i,j}} \right|^2 \quad (36a)$$

$$= 2 \sum_{i=1}^F \sum_{j=1}^K a_j^2 v_{i,j} \mathbb{E} \left| \sum_{p=1}^F \frac{\partial q_{p,p}}{\partial h_{i,j}} \right|^2 \quad (36b)$$

$$= 2 \sum_{i=1}^F \sum_{j=1}^K a_j^2 v_{i,j} \mathbb{E} \left| \frac{1}{\sigma^2} [\mathbf{h}_j^H \mathbf{Q} \mathbf{Q}]_i \right|^2 \quad (36c)$$

$$= \frac{2}{\sigma^4} \sum_{j=1}^K a_j^2 \mathbb{E} \mathbf{h}_j^H \mathbf{Q} \mathbf{Q} \mathbf{V}_j \mathbf{Q} \mathbf{Q} \mathbf{h}_j \quad (36d)$$

$$\leq \mathbb{E} \frac{2 \|\mathbf{A}\|^2 \|\mathbf{Q}\|^4}{\sigma^4 d} \text{Tr}(\mathbf{H} \mathbf{H}^H) \quad (36e)$$

$$\leq C \frac{K}{d}. \quad (36f)$$

The inequality in (36f) follows since from the resolvent identity one can verify that $\|\mathbf{Q}\| \leq 1$. Also, it is can be verified that $\text{Tr}(\mathbf{H} \mathbf{H}^H) \leq K$. Finally, putting the results from (35) and (36) together, we get $|\zeta(\sigma^2)| \leq \frac{C}{d^2}$.

B. Replacing the matrix \mathbf{E}_F by the deterministic matrix \mathbf{R}_F

In the last step of the proof, we need to show that the matrix $\mathbf{E}_F(\mathbf{V}_F, \sigma^2)$ can be replaced by the deterministic matrix $\mathbf{R}_F(\mathbf{V}_F, \sigma^2) = \text{diag}\{r_p(\mathbf{V}_F, \sigma^2), \forall p \in [1, \dots, F]\}$ where the r_p values are given as the unique positive solution of the following system of equations

$$r_p = \frac{\sigma^2}{\sigma^2 + \tilde{\delta}_p}, \quad \tilde{\delta}_p = \text{Tr}(\Lambda_p^{(F)} \mathbf{A}_F \tilde{\mathbf{R}}_F), \quad p = 1, \dots, F, \quad (37)$$

$$\tilde{r}_j = \frac{\sigma^2}{\sigma^2 + \delta_j}, \quad \delta_j = a_j^2 \text{Tr}(\mathbf{V}_j^{(F)} \mathbf{R}_F), \quad j = 1, \dots, K \quad (38)$$

where the matrices $\Lambda_p^{(F)}$, \mathbf{A}_F , and $\mathbf{V}_j^{(F)}$ are as defined in (31). The superscript and subscript F denote the dependency of the entities on the system size. The matrix $\tilde{\mathbf{R}}_F$ is defined as $\tilde{\mathbf{R}}_F(\mathbf{V}_F, \sigma^2) = \text{diag}\{\tilde{r}_j(\mathbf{V}_F, \sigma^2), \forall j \in [1, \dots, K]\}$. Let us define \mathbf{T}_F and \mathbf{B}_F to be $F \times F$ and $K \times K$ diagonal deterministic matrices with uniformly bounded spectral norm. Then, we show that under the assumptions in the theorem, the following holds for every $\sigma^2 \in \mathbb{R}^+$

$$\begin{aligned} \frac{1}{F} \text{Tr}(\mathbf{T}_F \mathbf{E}_F(\mathbf{V}_F, \sigma^2)) &= \frac{1}{F} \text{Tr}(\mathbf{T}_F \mathbf{R}_F(\mathbf{V}_F, \sigma^2)) + \mathcal{O}\left(\frac{1}{d^2}\right), \\ \frac{1}{F} \text{Tr}(\mathbf{B}_F \tilde{\mathbf{E}}_F(\mathbf{V}_F, \sigma^2)) &= \frac{1}{F} \text{Tr}(\mathbf{B}_F \tilde{\mathbf{R}}_F(\mathbf{V}_F, \sigma^2)) + \mathcal{O}\left(\frac{1}{d^2}\right). \end{aligned} \quad (39)$$

In the following, the dimension-superscript and subscript F are omitted. In proving the results we need to develop a well-quantified bound on the difference of the trace term $\frac{1}{F}\text{Tr}(\mathbf{T}(\mathbf{E} - \mathbf{R}))$ as the dimensions grow large. First, by a mere development, we have

$$\begin{aligned} \frac{1}{F}|\text{Tr}(\mathbf{T}(\mathbf{E} - \mathbf{R}))| &= \frac{1}{F}|\text{Tr}(\mathbf{T}\mathbf{E}(\mathbf{R}^{-1} - \mathbf{E}^{-1})\mathbf{R})| \\ &= \frac{1}{\sigma^2 F} \left| \sum_{p=1}^F t_p e_p r_p (\tilde{\delta}_p - \tilde{\alpha}_p) \right| \\ &\leq \frac{\|\mathbf{T}\|}{\sigma^2 F} \sum_{p=1}^F |\tilde{\delta}_p - \tilde{\alpha}_p| \end{aligned} \quad (40)$$

where the last equality follows from the upper bounds $\|\mathbf{R}\| \leq 1$ and $\|\mathbf{E}\| \leq 1$, implied by the definitions of \mathbf{R} and \mathbf{E} in (37) and (33), respectively. Under the same arguments, the term $|\tilde{\delta}_p - \tilde{\alpha}_p|$ can be bounded as

$$\begin{aligned} |\tilde{\delta}_p - \tilde{\alpha}_p| &= \frac{1}{\sigma^2} \left| \sum_{j=1}^K a_j^2 v_{p,j} (\alpha_j - \delta_j) \tilde{e}_j \tilde{r}_j \right| \\ &\leq \frac{1}{\sigma^2} \sum_{j=1}^K a_j^2 v_{p,j} |\alpha_j - \delta_j|. \end{aligned} \quad (41)$$

On the other hand, expanding $|\alpha_j - \delta_j|$, we have

$$\begin{aligned} |\alpha_j - \delta_j| &= a_j^2 \left| \mathbb{E} \text{Tr}(\mathbf{V}_j \mathbf{Q}) - \sum_{p=1}^F \frac{v_{p,j}}{1 + \tilde{\delta}_p / \sigma^2} \right| \\ &\stackrel{(a)}{=} a_j^2 \left| \sum_{p=1}^F v_{p,j} \left(\frac{1}{1 + \tilde{\alpha}_p / \sigma^2} - \frac{1}{1 + \tilde{\delta}_p / \sigma^2} \right) \right| + \mathcal{O}\left(\frac{1}{d^2}\right) \\ &\stackrel{(b)}{\leq} \frac{a_j^2}{\sigma^2} \sum_{p=1}^F v_{p,j} |\tilde{\delta}_p - \tilde{\alpha}_p| + \mathcal{O}\left(\frac{1}{d^2}\right) \end{aligned} \quad (42)$$

where equality (a) follows from (33), and (b) is given due to the bounds on the spectral norm of \mathbf{R} and \mathbf{E} . These inequalities together yield,

$$\begin{aligned} \frac{1}{K} \sum_{j=1}^K |\alpha_j - \delta_j| &\stackrel{(42)}{\leq} \frac{\|\mathbf{A}\|}{\sigma^2 d} \sum_{p=1}^F |\tilde{\delta}_p - \tilde{\alpha}_p| + \mathcal{O}\left(\frac{1}{d^2}\right) \\ &\stackrel{(41)}{\leq} \frac{\|\mathbf{A}\|^2}{\sigma^4 d} \sum_{p=1}^F \sum_{j=1}^K v_{p,j} |\alpha_j - \delta_j| + \mathcal{O}\left(\frac{1}{d^2}\right) \\ &= \frac{P \|\mathbf{A}\|^2}{\sigma^4 d} \sum_{j=1}^K |\alpha_j - \delta_j| + \mathcal{O}\left(\frac{1}{d^2}\right) \end{aligned} \quad (43)$$

where the last equality follows since $\sum_{p=1}^F v_{p,j} \leq P$ where $P = \max\{P_k, \forall k \in \mathcal{K}\}$. Let us do a change of variable as $z = 1/\sigma^2$ and express all the related functions in terms of z . Then, from the inequality in (43), we get

$$(1 - \frac{K}{d} z^2 P \|\mathbf{A}\|^2) \frac{1}{K} \sum_{j=1}^K |\alpha_j - \delta_j| = \mathcal{O}(\frac{1}{d^2}). \quad (44)$$

This inequality ensures that there exist a z_0 value such that $(1 - \frac{K}{d} z^2 P \|\mathbf{A}\|^2) > 0$. This further implies that $\frac{1}{K} \sum_{j=1}^K |\alpha_j - \delta_j| = \mathcal{O}(\frac{1}{d^2})$ holds for $z \leq z_0$. Thus, once we prove that the above inequality holds for all $z \in \mathbb{R}^+$, the convergence of the trace term to the deterministic equivalent can be claimed. To do so, we consider $\alpha_j = a_j^2 \mathbb{E} \text{Tr}(\mathbf{V}_j \mathbf{Q})$ as a function in z with extended domain from $z \in \mathbb{R}^+$ to $z \in \mathbb{C} \setminus \mathbb{R}^-$. It can be shown that the following integral representation for $\alpha_j(z)$ holds (see [49, Appendix A]),

$$\alpha_j(z) = \int_0^{+\infty} \frac{\mu_j(d\lambda)}{1 + z\lambda} \quad (45)$$

where μ_j is a uniquely defined positive measure on \mathbb{R}^+ such that $\mu_j(\mathbb{R}^+) = a_j^2 \text{Tr} \mathbf{V}_j$. Based on the properties of Stieltjes transform [20, Theorem 3.2], $\alpha_j(z)$ can be upper-bounded as

$$\begin{aligned} \alpha_j(z) &\leq a_j^2 \text{Tr} \mathbf{V}_j \frac{1}{|z| \text{dist}(-\frac{1}{z}, \mathbb{R}^+)} \\ &\leq \|\mathbf{A}\| \frac{1}{|z| \text{dist}(-\frac{1}{z}, \mathbb{R}^+)}. \end{aligned} \quad (46)$$

Similarly, a bound on $\delta_k(z)$ can be developed using integral representation. This analysis shows that the functions $\delta_k(z)$ and $\alpha_k(z)$ belongs to the class of Stieltjes transforms of finite positive measures carried by \mathbb{R}^+ . Thus, $|\alpha_k(z) - \delta_k(z)|$ belongs to a family of analytic functions, which are bounded on any compact subset of $\mathbb{C} \setminus \mathbb{R}^-$. As the result, the Vitali's convergence theorem [20] ensures that $\frac{1}{K} \sum_{j=1}^K |\alpha_j - \delta_j|$ goes to zero for any $z \in \mathbb{R}^+$ as $F \rightarrow \infty$. What remains is to show that the derived convergence rate $\mathcal{O}(\frac{1}{d^2})$ holds for all $z \in \mathbb{R}^+$, which completes the proof. The results follows from straightforward calculus where the reader is invited to refer to [49, Appendix C] for the details about the derivation steps.

APPENDIX II

PROOF OF PROPOSITION 2

The first property in the proposition follows directly since the mutual information is a strictly increasing function of the UEs' powers. One can also drive the same conclusion based on (12c) and (12a). The former implies that $\delta_k \geq \frac{1}{\sigma^2 F} a_k^2 \tilde{r}_k r_f > 0$, and thus, the latter gives $\sum_{i \in \mathcal{F}} v_{i,k} = P_k$

for all UEs. The second property in the proposition can be justified by noticing that (12c) is satisfied for a UE k only if we set $\delta_k^* = \frac{1}{\sigma^2 F} a_k^2 \tilde{r}_k \max\{r_f\}$. Hence, from (12b), we observe that UE k transmits only on the sub-channels with largest r_f values. Since, all other UEs also have the same preference, the condition in (12b) and (12c) are satisfied only if the UEs assign $v_{f,k}$ values such that $r_f = r, \forall f$. The value r can be shown that is unique, i.e., $r = r^*$ for any solution to (8). Assume the matrix \mathbf{V} to be a solution to (8) that results in r_f values to be equalized across sub-channels, i.e., $r_f = r, \forall f$. Plugging the \mathbf{V} entries into (7), we get

$$\tilde{r}_k = \frac{1}{1 + \frac{a_k^2}{\sigma^2} \sum_{f \in \mathcal{F}} v_{f,k}^2 r}, \forall k \in \mathcal{K} \quad (47)$$

where, from the first property in the proposition, we have $\sum_{i \in \mathcal{F}} v_{i,k} = P_k, \forall k \in \mathcal{K}$, and thus,

$$\tilde{r}_k = \frac{\sigma^2}{\sigma^2 + P_k a_k^2 r}, \forall k \in \mathcal{K}. \quad (48)$$

Given the values of $\tilde{r}_k, \forall k \in \mathcal{K}$ as above, the value of r can be evaluated from the following system of equations

$$r = \left(1 + \sum_{k \in \mathcal{K}} \frac{a_k^2 v_{f,k}}{\sigma^2 + P_k a_k^2 r}\right)^{-1}, \quad \forall f \in \mathcal{F} \quad (49)$$

that holds for all $f \in \mathcal{F}$. Inverting the sides of (49) and summing over all $f \in \mathcal{F}$ results in

$$\frac{F}{r} = F + \sum_{k \in \mathcal{K}} \frac{a_k^2}{\sigma^2 + P_k a_k^2 r} \sum_{f \in \mathcal{F}} v_{f,k} \quad (50)$$

where from the first property we have $\sum_{i \in \mathcal{F}} v_{i,k} = P_k, \forall k \in \mathcal{K}$, which gives $r = r^*$ as in (13).

REFERENCES

- [1] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127 639–127 651, 2019, doi:10.1109/ACCESS.2019.2939938.
- [2] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1616–1626, 2008.
- [3] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Vehicular Technology Conference (VTC2014-Fall)*, Sep. 2014, pp. 1–5.
- [4] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.
- [5] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, Oct 2017.
- [6] L. Dai, B. Wang, Y. Yuan, S. Han, C.-I. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, pp. 74–81, 09 2015.
- [7] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5G wireless networks," in *11th International Symposium on Wireless Communications Systems (ISWCS)*, Aug 2014, pp. 781–785.

- [8] R. Hoshyar, R. Razavi, and M. Al-Imari, "LDS-OFDM an efficient multiple access technique," in *Vehicular Technology Conference*, May 2010, pp. 1–5.
- [9] R. Razavi, M. AL-Imari, M. A. Imran, R. Hoshyar, and D. Chen, "On receiver design for uplink low density signature OFDM (LDS-OFDM)," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3499–3508, November 2012.
- [10] M. Al-Imari and R. Hoshyar, "Reducing the peak to average power ratio of LDS-OFDM signals," in *International Symposium on Wireless Communication Systems*, Sep. 2010, pp. 922–926.
- [11] M. T. P. Le, G. C. Ferrante, G. Caso, L. De Nardis, and M. Di Benedetto, "On information-theoretic limits of code-domain NOMA for 5G," *IET Communications*, vol. 12, no. 15, pp. 1864–1871, 2018.
- [12] M. Al-Imari, M. Imran, and R. Tafazolli, "Low density spreading multiple access," *Information Technology Software Engineering*, vol. 2, 09 2012.
- [13] O. Shental, B. M. Zaidel, and S. S. Shitz, "Low-density code-domain NOMA: Better be regular," in *IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2628–2632.
- [14] B. M. Zaidel, O. Shental, and S. S. Shitz, "Sparse NOMA: A closed-form characterization," in *IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1106–1110.
- [15] G. C. Ferrante and M. D. Benedetto, "Spectral efficiency of random time-hopping CDMA," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6643–6662, Dec 2015.
- [16] M. Yoshida and T. Tanaka, "Analysis of sparsely-spread CDMA via statistical mechanics," in *IEEE International Symposium on Information Theory*, July 2006, pp. 2378–2382.
- [17] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove tanaka's formula)," in *IEEE Information Theory Workshop - ITW '06 Punta del Este*, March 2006, pp. 160–164.
- [18] R. R. Miller, "Random matrices, free probability and the replica method," in *12th European Signal Processing Conference*, Sep. 2004, pp. 189–196.
- [19] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information theory*, vol. 45, no. 2, pp. 622–640, 1999.
- [20] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.
- [21] P. M. Wood, "Universality and the circular law for sparse random matrices," *Ann. Appl. Probab.*, vol. 22, no. 3, pp. 1266–1300, 06 2012.
- [22] M. T. P. Le, G. C. Ferrante, T. Q. S. Quek, and M. Di Benedetto, "Fundamental limits of low-density spreading NOMA with fading," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4648–4659, July 2018.
- [23] R. Razavi, R. Hoshyar, M. A. Imran, and Y. Wang, "Information theoretic analysis of lds scheme," *IEEE Communications Letters*, vol. 15, no. 8, pp. 798–800, 2011.
- [24] R. Miller and A. M. Tulino, "Minimum bit error probability of large randomly spread MC-CDMA systems in multipath rayleigh fading," in *Eighth IEEE International Symposium on Spread Spectrum Techniques and Applications - Programme and Book of Abstracts (IEEE Cat. No.04TH8738)*, Aug 2004, pp. 560–564, doi:10.1109/ISSSTA.2004.1371762.
- [25] D. N. C. Tse and O. Zeitouni, "Performance of linear multiuser receivers in random environments," in *IEEE Communications Theory Mini-Conference (Cat. No.99EX352)*, June 1999, pp. 163–167, doi=10.1109/CTMC.1999.790257.
- [26] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: effective interference, effective bandwidth and user capacity," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 641–657, March 1999, doi=10.1109/18.749008.
- [27] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2832–2836, Nov 1998, 10.1109/18.737515.
- [28] —, "Randomly selected spreading sequences for coded CDMA," in *Proceedings of ISSSTA'95 International Symposium on Spread Spectrum Techniques and Applications*, vol. 1, Sep. 1996, pp. 54–57 vol.1, doi=10.1109/ISSSTA.1996.563742.

- [29] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001, doi=10.1109/18.923717.
- [30] H. Asgharimoghaddam and A. Tili, "Resource allocation in low density spreading uplink NOMA via asymptotic analysis," in *IEEE International Symposium on Information Theory (ISIT)*, Los Angeles, California, USA, June 2020.
- [31] K. Fazel and S. Kaiser, *MultiCarrier and Spread Spectrum Systems: From OFDM and MCCDMA to LTE and WiMAX, Second Edition*. John Wiley & Sons, 2008, vol. 2.
- [32] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 974–978.
- [33] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1261–1266, July 1994.
- [34] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1984–1991, Sept 1999.
- [35] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its applications*, 01 1979, vol. 143.
- [36] M. Reinhardt and J. Lindner, "Transformation of a Rayleigh fading channel into a set of parallel AWGN channels and its advantage for coded transmission," *Electronics Letters*, vol. 31, no. 25, pp. 2154–2155, Dec 1995.
- [37] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.4460100604>
- [38] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for capacity analysis of large dimensional multi-antenna channels," *IEEE Transactions on Information Theory*, vol. 54, 01 2007.
- [39] R. R. Müller, G. Alfano, B. M. Zaidel, and R. de Miguel, "Applications of large random matrices in communications engineering," *ArXiv*, vol. abs/1310.5479, 2013.
- [40] R. R. Müller, "The replica method in multiuser communications," in *Random Matrix Theory and Its Applications*, July 2009, pp. 139–165, doi = 10.1142/97898142731210005.
- [41] W. Hachem, P. Loubaton, and J. Najim, "A CLT for information-theoretic statistics of Gram random matrices with a given variance profile," *Ann. Appl. Probab.*, vol. 18, no. 6, pp. 2071–2130, 12 2008, doi:10.1214/08-AAP515.
- [42] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [43] H. Anton, *Calculus with Analytic Geometry*. John Wiley and Sons Inc., 1995, vol. 10.
- [44] V. V. Veeravalli and A. Mantravadi, "The coding-spreading tradeoff in CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 2, pp. 396–408, Feb 2002, doi=10.1109/49.983362.
- [45] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.
- [46] R. E. Korf, "A complete anytime algorithm for number partitioning," *Artificial Intelligence*, vol. 106, no. 2, pp. 181 – 203, 1998.
- [47] B. Hayes, "Computing science: The easiest hard problem," *American Scientist*, vol. 90, no. 2, pp. 113–117, 2002.
- [48] R. L. Graham, "Bounds on multiprocessing timing anomalies," *SIAM Journal of Applied Mathematics*, vol. 17, pp. 416–429, 1969.
- [49] F. Dupuy and P. Loubaton, "On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 5737–5753, Sept 2011.