Sparse NOMA: A Closed-Form Characterization

Benjamin M. Zaidel
Faculty of Engineering
Bar-Ilan University
Ramat-Gan 52900, Israel
Email: benjamin.zaidel@gmail.com

Ori Shental
Communications Theory Department
Bell Labs
Holmdel, New Jersey 07733, USA
Email: ori.shental@nokia-bell-labs.com

Shlomo Shamai (Shitz)
Department of Electrical Engineering
Technion
Haifa 32000, Israel
Email: sshlomo@ee.technion.ac.il

Abstract—**Understanding fundamental limits of the various technologies suggested for future 5G and beyond cellular systems is crucial for developing efficient state-of-the-art designs. A leading technology of major interest is non-orthogonal multiple-access (NOMA). In this paper, we derive an explicit rigorous *closed-form* analytical expression for the optimum spectral efficiency in the large-system limit of *regular* sparse NOMA, where only a *fixed and finite* number of orthogonal resources are allocated to any designated user, and vice versa. The basic Verdú-Shamai formula for (dense) randomly-spread code-division multiple-access (RS-CDMA) turns out to coincide with the limit of the derived expression, when the number of orthogonal resources per user grows large. Furthermore, regular sparse NOMA is rigorously shown to be spectrally more efficient than RS-CDMA across the entire system load range. It may therefore serve as an efficient means for reducing the throughput gap to orthogonal transmission in the underloaded regime, and to the ultimate Cover-Wyner bound in overloaded systems. The results analytically reinforce preliminary conclusions in [1], which mostly relied on heuristics and numerical observations. The spectral efficiency is also derived in closed form for the sub-optimal linear minimum-mean-square-error (LMMSE) receiver, which again extends the corresponding Verdú-Shamai LMMSE formula to regular sparse NOMA.**

I. INTRODUCTION

Orthogonal transmission is the holy grail of multiple-access communications. Orthogonal multiple-access (OMA) allows for a feasible transceiver design, which also achieves the ultimate total throughput in a fully-loaded Gaussian channel, where the number of users equals the total number of available orthogonal resources. However, non-orthogonal multiple-access (NOMA) schemes with tractable near-optimal receivers open up the practicality of operation in the overloaded regime, where the number of designated users *exceeds* the number of available resources. Note that even for pragmatic underloaded systems it is typically hard to maintain orthogonality, and OMA leads to severe throughput penalties. Thus, NOMA exhibits significant advantages over OMA by either supporting more concurrent users or, alternatively, facilitating higher user throughput when orthogonality breaks down. In a nutshell, these are the main incentives driving NOMA as a key enabler for the Internet-of-Things (IoT) and 5G and beyond networks. An overview of the ample literature on NOMA can be found, e.g., in [2] and references therein.

Sparse, or low-density code-domain (LDCD) NOMA is a prominent sub-category, which conceptually relies on multiplexing low-density signatures (LDS) [3]. Sparse spreading codes comprising only a small number of non-zero elements

are employed for linearly modulating each user's symbols over shared physical orthogonal resources. The sparse mapping between users and resources is dubbed: *regular* when each user occupies a fixed number of resources, and each resource is used by a fixed number of users; *irregular* when the respective numbers are random, and only fixed on average; *partly-regular* when each user occupies a fixed number of resources, and each resource is used by a random, yet fixed on average, number of users (or vice versa). The main attractiveness of this central class of NOMA schemes is in its inherent receiver complexity reduction, achieved by utilizing message-passing algorithms (MPAs) [2], similar to low-density parity-check (LDPC) codes. Different variants of sparse NOMA, e.g., sparse-code multiple-access (SCMA) [4], have recently gained much attention in 5G standardization. Naturally, information-theoretic analysis of sparse NOMA has been a fruitful grounds for research, providing a great deal of insight into its workings, e.g., [5]–[8]. However, it is important to note that neither of these contributions provide an explicit rigorous closed-form analytical characterization of sparse NOMA throughput. This holds even in the large-system limit, and comes in sheer contrast to the insightful work by Verdú and Shamai on (dense) randomly-spread code-division multiple-access (RS-CDMA) [9].

In a recent paper [1], the authors have presented a closed-form analytical expression for the *limiting empirical squared singular value density* of a spreading (signature) matrix, corresponding to *regular* sparse NOMA and feasible, e.g., via regular LDPC design. The result was rigorously derived only for a repetition-based sparse spreading scheme, while for binary sparse random spreading the same result was claimed to hold based on the *heuristic* cavity method of statistical physics. The limiting density was then used to plot the optimum asymptotic spectral efficiency, but only by means of numerical integration.

In this contribution, using a recent fundamental result by Bordenave and Lelarge [10], we first re-establish the aforementioned limiting density result in full rigor, and in fact extend it to the more general case of sparse signatures, of which non-zero entries reside on the unit-circle in the complex plane. Then, a rigorous *closed-form explicit analytical expression* for the spectral efficiency of regular sparse NOMA with optimum decoding is derived. The formula is obtained for Gaussian signaling and non-fading channels in the asymptotic large-system limit. The Verdú-Shamai optimum spectral efficiency formula [9] turns out to coincide

with the limit of the derived expression as the spreading codes become dense. A closed-form expression for the asymptotic spectral efficiency of the linear minimum-mean-square-error (LMMSE) receiver is also provided. An extreme-SNR characterization leading to useful insights is given as well. The results allow for easy comparison to the spectral efficiency of RS-CDMA [9] (representing *dense* NOMA), and other sparse NOMA variants, e.g., [5], [8]. Assuming optimal decoding, the remarkable superiority of the feasible regular sparse NOMA over not only its irregular, or partly-regular, counterparts, but also over the intractable RS-CDMA, as originally observed (only) numerically in [1], is here being analytically verified.

II. SPARSE NOMA SYSTEM MODEL

Consider a system where the signals of K users are multiplexed over N shared orthogonal resources. These resources can designate, e.g., orthogonal frequencies in orthogonal frequency-division multiplexing (OFDM) based systems, different time-slots in time-hopping (TH) multiple-access systems [8], or combinations thereof. Multiplexing is performed while employing randomly chosen sparse spreading signatures of length N (namely, N-dimensional vectors). Each sparse signature is assumed to contain a small number of non-zero entries (typically much smaller than N), while the remaining entries are set to zero. The N-dimensional received signal, at some arbitrary time instance, at the output of a generic complex Gaussian vector channel, adheres to

$$\mathbf{y} = \sqrt{\tfrac{\mathrm{snr}}{d}} \mathbf{A}\mathbf{x} + \mathbf{n} \ , \tag{1}$$

where $\mathbf{x}$ is a K-dimensional complex vector comprising the coded symbols of the users. Assuming Gaussian signaling, full symmetry, fixed powers, and no cooperation between encoders corresponding to different users, the input vector $\mathbf{x}$ is distributed as $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$; i.e., a unit energy per symbol per user is assumed. $\mathbf{A}$ denotes the $N \times K$ sparse signature matrix, whose kth column represents the spreading signature of user k, and its non-zero entries designate the user-resource mapping (user k occupies resource n if $A_{nk} \neq 0$, where A_{nk} denotes the (n, k)'th entry of $\mathbf{A}$). Finally, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ denotes the N-dimensional circularly-symmetric complex additive white Gaussian noise (AWGN) vector at the receiving end. The normalization factor $\frac{1}{\sqrt{d}}$ in (1), $d \in \mathbb{N}^+$, controls the signature norms (as explained in the sequel). The parameter snr thus designates the received signal-to-noise ratio (SNR) of each user, and we further use $\beta \triangleq \frac{K}{N}$ to denote the *system load* (users per resource).

In sparse NOMA, due to the sparsity of $\mathbf{A}$, typically only a few of the users' signals collide over any given orthogonal resource. The regularity assumption dictates that each column of $\mathbf{A}$ (respectively, row) has *exactly* d (respectively, βd) non-zero entries. β is therefore chosen here so that $\beta d \in \mathbb{N}^+$. The non-zero entries of $\mathbf{A}$ are assumed to arbitrarily reside on the unit-circle in the complex plane. Repetition-based spreading and random binary spreading [1] thus constitute special cases. The model may also account for phase-fading scenarios in conjunction with sparse spreading. The normalization in (1)

ensures that the columns of $\frac{1}{\sqrt{d}}\mathbf{A}$ have unit norm. Finally, we also assume here that $\mathbf{A}$ is perfectly known at the receiving end, and uniformly chosen randomly and independently per each channel use from the set of $(\beta d, d)$-regular matrices.

Note at this point that a key tool in the derivations to follow is the observation that the signature matrix $\mathbf{A}$ can be associated with the *adjacency matrix* of a random $(\beta d, d)$-semiregular bipartite (factor) graph $\mathcal{A}$, where a user node k and a resource node n are connected if and only if $A_{nk} \neq 0$. This factor graph is assumed henceforth to be *locally tree-like*[1], and to converge in the large-system limit (as $N \to \infty$) to a *bipartite Galton-Watson tree* (BGWT), as specified in Section III. This assumption (valid for regular LDPC codes, cf. [1], [10]) essentially implies that for large dimensions short cycles are rare, and it allows for reduced complexity iterative near-optimal multiuser detection by applying MPAs over the underlying factor graph.

III. ASYMPTOTIC SPECTRAL DENSITY

Our first step is to characterize the limiting empirical distribution of the squared singular values of the normalized signature matrix $\frac{1}{\sqrt{d}}\mathbf{A}$, which is a cornerstone in the derivation of our main analytical results. To this end, we first review a useful result by Bordenave and Lelarge [10] on properties of random weighted bipartite graphs, whose random weak limit is the probability measure of a BGWT. Due to space limitations, we avoid a fully formal representation, while referring the reader to [10] for a full account.

Consider a sequence of random bipartite graphs $\{\mathcal{G}_N\}$, converging in law to a BGWT with degree distribution $(\delta_{\beta d}, \delta_d)$, and parameter $\frac{1}{1+\beta}$. Let $\mathbf{W}$ be an $N \times K$ complex random matrix independent of $\mathcal{G}_N$, with i.i.d. entries having finite absolute second moments. With some abuse of notation, let the weighted adjacency matrix of $\mathcal{G}_N$ read

$$\tilde{\mathbf{A}}_N = \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^\dagger & \mathbf{0} \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{0} & \mathbf{W} \\ \mathbf{W}^\dagger & \mathbf{0} \end{pmatrix} \circ \begin{pmatrix} \mathbf{0} & \bar{\mathbf{A}} \\ \bar{\mathbf{A}}^\dagger & \mathbf{0} \end{pmatrix} \ , \tag{2}$$

where $\circ$ denotes the Hadamard product, $\bar{\mathbf{A}}$ is an $N \times K$ matrix such that $\bar{A}_{ij} = 1$ if vertex i is connected by an edge to vertex $N + j$ on $\mathcal{G}_N$, and $\bar{A}_{ij} = 0$ otherwise, where $i \in \{1, \dots, N\}$, and $j \in \{1, \dots, K\}$. Let $\mathcal{H}$ denote the set of holomorphic functions $f : \mathbb{C}^+ \to \mathbb{C}^+$ such that $|f(z)| \leq \frac{1}{\mathrm{Im}(z)}$. Finally, let $\mathcal{P}(\mathcal{H})$ denote the space of probability measures on $\mathcal{H}$.

Theorem 1 ([10, Theorems 4 & 5]). *Let the above assumptions on the structure of $\tilde{\mathbf{A}}_N$ hold. Then*

1) *There exists a unique pair of probability measures $(\mu_a, \mu_b) \in \mathcal{P}(\mathcal{H}) \times \mathcal{P}(\mathcal{H})$ such that for all $z \in \mathbb{C}^+$*

$$Y^a(z) \overset{d}{=} -(z + \textstyle\sum_{i=1}^{\beta d-1} |W_i^b|^2 Y_i^b(z))^{-1} \ , \tag{3}$$

$$Y^b(z) \overset{d}{=} -(z + \textstyle\sum_{i=1}^{d-1} |W_i^a|^2 Y_i^a(z))^{-1} \ , \tag{4}$$

where Y^a, Y_i^a (respectively, Y^b, Y_i^b) are i.i.d. random variables with law μ_a (respectively, μ_b), and W_i^a, W_i^b are i.i.d. random variables distributed as W_{11}^2.

---

[1] A precise mathematical definition can be found, e.g., in [10].

[2] The equality $\overset{d}{=}$ in (3) and (4) is in the sense that the *distribution* of the random variables on both sides of the equation is the same.

2) *For all $z \in \mathbb{C}^+$, the Stieltjes transform[3] $m_{\tilde{\mathbf{A}}_N}(z)$ of the empirical eigenvalue distribution of $\tilde{\mathbf{A}}_N$ converges as $N \to \infty$ in L^1 to $m_{\tilde{\mathbf{A}}}(z) = \frac{1}{1+\beta}\mathbb{E}\{X^a(z)\} + \frac{\beta}{1+\beta}\mathbb{E}\{X^b(z)\}$, where for all $z \in \mathbb{C}^+$*

$$X^a(z) \stackrel{d}{=} -(z + \textstyle\sum_{i=1}^{\beta d} |W_i^b|^2 Y_i^b(z))^{-1} , \quad (5)$$

$$X^b(z) \stackrel{d}{=} -(z + \textstyle\sum_{i=1}^{d} |W_i^a|^2 Y_i^a(z))^{-1} , \quad (6)$$

where $Y_i^a, Y_i^b, W_i^a,$ and W_i^b are i.i.d. copies with laws as in Part (1).

Turning to the sparse regular NOMA setting specified in Section II, we immediately observe that the graph $\mathcal{A}$ associated with the signature matrix $\mathbf{A}$ (cf. (1)) falls exactly within the framework of Theorem 1, and its weighted adjacency matrix can be expressed as in (2). Furthermore, the underlying assumption on the spreading signatures dictates that all entries of the corresponding weight matrix $\mathbf{W}$ have surely a unit absolute value. We thus get the following result.

Theorem 2. *Let $\mathbf{A}$ be a sparse random $N \times K$ matrix with exactly $2 \le d \in \mathbb{N}^+ < \infty$ (respectively, $2 \le \beta d \in \mathbb{N}^+ < \infty$) non-zero entries in each column (respectively, row), arbitrarily distributed over the unit-circle in $\mathbb{C}$. Assume that the $(\beta d, d)$-semiregular bipartite graph $\mathcal{A}$ associated with $\mathbf{A}$ is locally tree-like, with a BGWT having degree distribution $(\delta_{\beta d}, \delta_d)$ and parameter $\frac{1}{1+\beta}$ as a weak limit. Let $\alpha \triangleq \frac{d-1}{d}$ and $\gamma \triangleq \frac{\beta d-1}{d}$. Then*

1) *For all $z \in \mathbb{C}^+$, the Stieltjes transform of the empirical eigenvalue distribution of $\frac{1}{d}\mathbf{A}\mathbf{A}^\dagger$ converges as $N \to \infty$ in L^1 to*

$$m_{\frac{1}{d}\mathbf{A}\mathbf{A}^\dagger}(z) = -\left(z - \frac{\beta}{1+\alpha m(z)}\right)^{-1} , \quad (7)$$

where $m(z)$ solves the following deterministic equation:

$$m(z) = -\left(z - \frac{\gamma}{1+\alpha m(z)}\right)^{-1} . \quad (8)$$

2) *Subject to the convergence of the Stieltjes transform, the weak limit of the empirical eigenvalue distribution of $\frac{1}{d}\mathbf{A}\mathbf{A}^\dagger$ as $N \to \infty$ is a distribution with density*

$$\rho(\lambda, \beta, d) = [1-\beta]^+ \delta(\lambda) + \frac{\beta d}{2\pi} \frac{\sqrt{[\lambda-\lambda^-]^+[\lambda^+-\lambda]^+}}{\lambda(\beta d-\lambda)} , \quad (9)$$

where $\lambda^\pm = (\sqrt{\alpha} \pm \sqrt{\gamma})^2$, $\delta(\lambda)$ is a unit point mass at $\lambda = 0$, and $[z]^+ \triangleq \max\{0, z\}$.

Proof Outline: The weights $\{W_i^a, W_i^b\}$ in (3) and (4) have unit absolute values. The equations hence admit a unique *deterministic* solution. Next, note that the eigenvalues of

$$\tilde{\mathbf{A}}_N^2 = \begin{pmatrix} \mathbf{A}\mathbf{A}^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^\dagger \mathbf{A} \end{pmatrix} \quad (10)$$

are simply the eigenvalues of $\mathbf{A}\mathbf{A}^\dagger$ together with those of $\mathbf{A}^\dagger \mathbf{A}$ (which are in fact the same up to $|K - N|$ additional

[3]The Stieltjes transform of a probability measure μ on $\mathbb{R}$ reads $m(z) \triangleq \int_{\mathbb{R}} \frac{1}{x-z} \, d\mu(x)$, $z \in \mathbb{C}^+$. The measure μ can be recovered from $m(z)$ via the Stieltjes inversion formula $d\mu(\lambda) = \frac{1}{\pi}\lim_{\epsilon \to 0^+} \mathrm{Im}(m(z))|_{z=\lambda+j\epsilon} \, d\lambda$, where the limit is in the sense of weak convergence of measures (e.g., [11]).

zero eigenvalues). Furthermore, the limiting Stieltjes transform of the empirical eigenvalue distribution of $\tilde{\mathbf{A}}_N^2$ admits the following relation $z m_{\tilde{\mathbf{A}}^2}(z^2) = m_{\tilde{\mathbf{A}}}(z)$. This lets us conclude that $m_{\frac{1}{d}\mathbf{A}\mathbf{A}^\dagger}(z) = \sqrt{\frac{d}{z}}X^a(\sqrt{dz})$, where $X^a(z)$ is obtained from (5). Eqs. (7) and (8) then follow after some algebra. Finally, we get (9) using the Stieltjes inversion formula. ∎

IV. Optimum Receiver

The fundamental figure of merit for system performance is taken here as the normalized spectral efficiency (total throughput) in bits/sec/Hz per dimension. For optimum processing this quantity corresponds to the ergodic sum-capacity, given by the normalized conditional input-output mutual information [9]:

$$C_N^{\mathrm{opt}}(\mathsf{snr}, \beta, d) \triangleq \tfrac{1}{N}\mathrm{I}(\mathbf{x}; \mathbf{y}|\mathbf{A})$$
$$= \tfrac{1}{N} \mathbb{E}\left\{\log_2 \det\left(\mathbf{I}_N + \tfrac{\mathsf{snr}}{d}\mathbf{A}\mathbf{A}^\dagger\right)\right\} . \quad (11)$$

Focusing on the large-system limit, the asymptotic spectral efficiency of the optimum receiver corresponds to $C^{\mathrm{opt}}(\mathsf{snr}, \beta, d) \triangleq \lim_{N \to \infty} C_N^{\mathrm{opt}}(\mathsf{snr}, \beta, d)$, where the existence of the limit is established next. The following result is one of the main contributions of this paper.

Theorem 3. *Let d, β, α and γ be as in Theorem 2. Further let $\tilde{\beta} \triangleq \frac{\alpha}{\gamma}$ and $\zeta \triangleq \frac{\beta d}{\gamma}$. Then, the optimum spectral efficiency (11) converges as $N \to \infty$ to*

$$C^{\mathrm{opt}}(\mathsf{snr}, \beta, d) = (\beta - 1)\log_2\left(1 + \alpha\mathsf{snr} - \tfrac{1}{4}\mathcal{F}(\gamma\mathsf{snr}, \tilde{\beta})\right)$$
$$+ \tfrac{\beta(d-1)+1}{2}\log_2\left(1 + (\gamma + \alpha)\mathsf{snr} - \tfrac{1}{4}\mathcal{F}(\gamma\mathsf{snr}, \tilde{\beta})\right)$$
$$- \tfrac{\beta(d-1)-1}{2}\log_2\left(\frac{(1+\beta d\,\mathsf{snr})^2}{\mathcal{G}(\gamma\mathsf{snr}, \zeta, \tilde{\beta})}\right) , \quad (12)$$

where (cf. [9]) $\mathcal{F}(x, z) \triangleq \left(\left(x\left(1 + \sqrt{z}\right)^2 + 1\right)^{1/2} - \left(x\left(1 - \sqrt{z}\right)^2 + 1\right)^{1/2}\right)^2$ and

$$\mathcal{G}(x, y, z) \triangleq$$
$$\left(\frac{\sqrt{\left(y - (1-\sqrt{z})^2\right)\left(x(1+\sqrt{z})^2+1\right)} - \sqrt{\left(y - (1+\sqrt{z})^2\right)\left(x(1-\sqrt{z})^2+1\right)}}{\sqrt{y - (1-\sqrt{z})^2} - \sqrt{y - (1+\sqrt{z})^2}}\right)^2$$
$$x, y, z \in \mathbb{R}^+, \ y \ge (1 + \sqrt{z})^2 . \quad (13)$$

Proof Outline: The space $\mathcal{H}$ specified in Section III, equipped with an appropriate topology, is a complete separable metrizable compact space (see [10]). Considering the sequence of Stieltjes transforms converging to (7) by Theorem 2, recall that the *ergodic* normalized sum-capacity (11) is determined by the *distribution* of the signature matrices. Therefore, by the Skorokhod representation theorem (e.g., [10, Theorem 7]), we can assume that the sequence of Stieltjes transforms and its limit (here a deterministic function) are defined on a common probability space, and that the convergence to the limit is in the almost sure sense. We next observe that by Hadamard's inequality, along with the unit absolute value of the non-zero entries of $\mathbf{A}$, $\frac{1}{N}\log_2 \det\left(\mathbf{I}_N + \frac{\mathsf{snr}}{d}\mathbf{A}\mathbf{A}^\dagger\right) \le \log_2(1 + \beta\mathsf{snr}) < \infty$. This implies, by uniform integrability and the weak convergence stated in Theorem 2 (cf. (9)), that the sequence $\frac{1}{N}\log_2 \det\left(\mathbf{I}_N + \frac{\mathsf{snr}}{d}\mathbf{A}\mathbf{A}^\dagger\right)$ converges to $\int_0^\infty \log_2(1 + \mathsf{snr}\lambda)\rho(\lambda, \beta, d) \, d\lambda$, and finally, by the bounded convergence theorem, that $C_N^{\mathrm{opt}}(\mathsf{snr}, \beta, d)$ converges to the

same limit as well. Explicit calculation of the integral finally yields (12). ∎

Remark 4. Although (12) applies to $2 \leq d, \beta d \in \mathbb{N}^+ < \infty$, total throughputs on the convex closure of the respective rates are achievable by means of time-sharing between different $(d, \beta d)$ points in the admissible set, allowing for *effective* average non-integer values for $(d, \beta d)$ combinations.

We complete the asymptotic analysis of the optimum receiver by means of extreme-SNR characterization. Recall that a spectral efficiency R is approximated in the low-SNR regime as $R \approx \frac{S_0}{3\text{dB}}(\frac{E_b}{N_0}|_{\text{dB}} - \frac{E_b}{N_0}{}_{\min}|_{\text{dB}})$, where S_0 denotes the low-SNR slope, $\frac{E_b}{N_0}{}_{\min}$ is the minimum $\frac{E_b}{N_0}$ that enables reliable communications, and $3\text{dB} \triangleq 10 \log_{10} 2$ [12]. The SNR and $\frac{E_b}{N_0}$ are related via $\beta\text{snr} = R\frac{E_b}{N_0}$. In the high-SNR regime the spectral efficiency is approximated as $R \approx S_\infty (\log_2 \text{snr} - \mathcal{L}_\infty)$, where S_∞ denotes the high-SNR slope (multiplexing gain), and $\mathcal{L}_\infty$ denotes the high-SNR power offset [12]. The results are summarized in the following proposition (the proof is omitted due to space limitations).

Proposition 5. *Let d and β be as in Theorem 2. Then, the low-SNR parameters of the optimum receiver read:* $\left(\frac{E_b}{N_0}\right)^{\text{opt}}_{\min} = \ln 2$, *and* $S_0^{\text{opt}} = \frac{2\beta d}{d(\beta+1)-1}$. *The high-SNR slope of the optimum receiver is given by* $S_\infty^{\text{opt}} = \min(\beta, 1)$, *while the high-SNR power offset satisfies*

$$\mathcal{L}_\infty^{\text{opt}} =$$
$$\begin{cases} \left(\frac{1}{\beta} - 1\right)\log_2(1-\beta) - (d-1)\log_2\left(1 - \frac{1}{d}\right), & \beta < 1 \\ -(d-1)\log_2\left(1 - \frac{1}{d}\right), & \beta = 1 \\ (\beta - 1)\log_2(\beta - 1) - \beta\log_2\beta \\ \qquad - (\beta d - 1)\log_2\left(1 - \frac{1}{\beta d}\right), & \beta > 1. \end{cases} \quad (14)$$

Comparison to RS-CDMA [12] reveals that $\frac{E_b}{N_0}{}_{\min}$ and S_∞ are identical in both settings. However, the low-SNR slope of regular sparse NOMA is strictly *higher*, while the high-SNR power offset is strictly *lower*, compared to RS-CDMA. Regular sparse NOMA thus exhibits superior performance in extreme-SNR regimes. The spectral efficiency coincides with that of RS-CDMA as $d \to \infty$ in *all* SNR regimes[4], which is also evident from the underlying expectation and (9) (cf. [1]).

V. LMMSE RECEIVER

In this section we turn to investigate the ergodic spectral efficiency of the LMMSE receiver. Recall that the corresponding error covariance matrix is given by $\mathbf{M} = (\mathbf{I}_K + \text{snr}\,\mathbf{R})^{-1}$, where $\mathbf{R} \triangleq \frac{1}{d}\mathbf{A}^\dagger\mathbf{A}$ is the signature crosscorrelation matrix [9]. The signal-to-interference-plus-noise ratio (SINR) at the output of the receiver for user k is $\frac{1}{M_{kk}} - 1$, and the spectral efficiency of the LMMSE receiver thus reads

$$C_N^{\text{mmse}}(\text{snr}, \beta, d) = \beta\,\mathbb{E}\left\{\frac{1}{K}\sum_{k=1}^K \log_2\left(\frac{1}{M_{kk}}\right)\right\}. \quad (15)$$

The following theorem characterizes the spectral efficiency of the LMMSE receiver in the large-system limit.

---

[4]Note here that first $N, K \to \infty$, while d remains *fixed*, and only then d is increased. The locally tree-like property thus remains valid.

Theorem 6. *Let the definitions and assumptions of Theorem 3 hold. Then, the spectral efficiency of the LMMSE receiver $C_N^{\text{mmse}}(\text{snr}, \beta, d)$ converges as $N \to \infty$ to*

$$C^{\text{mmse}}(\text{snr}, \beta, d) = \beta\log_2\left(\frac{1+\beta d\,\text{snr}}{1+d\gamma\,\text{snr} - \frac{d\mathcal{F}(\gamma\text{snr}, \tilde{\beta})}{4}}\right). \quad (16)$$

Proof Outline: Let $\mathsf{R_R}(z) \triangleq (\mathbf{R} - z\mathbf{I}_K)^{-1}$, $z \in \mathbb{C}^+$, denote the resolvent of $\mathbf{R}$. Following [10], it can be shown that the diagonal entries of $\mathsf{R_R}(z)$ converge in distribution to $\sqrt{\frac{d}{z}}X^b(\sqrt{dz})$ (cf. (6)), which gives the Stieltjes transform of the limiting empirical eigenvalue distribution of $\mathbf{R}$, $m_\mathbf{R}(z)$. Applying analytic continuation, we thus conclude that $M_{kk} \xrightarrow{d} \frac{1}{\text{snr}}m_\mathbf{R}(-\frac{1}{\text{snr}}) \triangleq M_1$. Since the random variables $\{M_{kk}\}$ have a bounded strictly positive support, we may conclude by the continuity of $\log_2(\frac{1}{x})$ and uniform integrability of $\log_2(\frac{1}{M_{kk}})$, that $C^{\text{mmse}}(\text{snr}, \beta, d) = \beta\log_2(M_1^{-1})$, leading to (16). ∎

We note here that the time-sharing argument in Remark 4 holds for the LMMSE receiver as well. The extreme-SNR characterization of this receiver is given next.

Proposition 7. *Let d and β be as in Theorem 2. Then, the low-SNR parameters of the LMMSE receiver read:* $\left(\frac{E_b}{N_0}\right)^{\text{mmse}}_{\min} = \ln 2$, *and* $S_0^{\text{mmse}} = \frac{2\beta d}{(2\beta+1)d-2}$. *The high-SNR slope of the LMMSE receiver is given by*

$$S_\infty^{\text{mmse}} = \begin{cases} \beta, & \beta < 1 \\ \frac{1}{2}, & \beta = 1 \\ 0, & \beta > 1 \end{cases} \quad (17)$$

while the high-SNR power offset satisfies

$$\mathcal{L}_\infty^{\text{mmse}} = \begin{cases} \log_2\left(\frac{1}{1-\beta}\right) + \log_2\left(\frac{d-1}{d}\right), & \beta < 1 \\ \log_2\left(\frac{d-1}{d}\right), & \beta = 1. \end{cases} \quad (18)$$

Comparison to [12] leads again to the same conclusions as drawn for the optimum receiver.

VI. NUMERICAL RESULTS

Some numerical results for the limiting spectral efficiency of regular sparse NOMA, following (12) and (16), are shown in Fig. 1, and plotted as a function of the system load β for fixed $\frac{E_b}{N_0} = 10\text{dB}$. The results were calculated for $d = 2$ ($\beta \geq 1$), 3 and 10. The markers designate the points for which $\beta d \in \mathbb{N}^+$. The piecewise linear (dotted) lines represent the achievable throughputs by *exercising time-sharing* as in Remark 4. Fig. 1 clearly indicates that regular sparse NOMA outperforms RS-CDMA [9, Eqs. (9), (12)]. In the underloaded regime ($\beta < 1$), the gap to the spectral efficiency of orthogonal transmissions [9, Eq. (8)] is reduced. In the overloaded regime the scheme reduces the gap to the ultimate performance limit, as given by the Cover-Wyner bound [9, Eq. (4)] corresponding to the absence of spreading. Note also that the spectral efficiency of regular sparse NOMA approaches that of RS-CDMA *from above* with the increase of the sparsity parameter d, as already indicated in [1]. For the sake of comparison we also included in Fig. 1 the spectral efficiencies of irregular sparse NOMA [5, Eqs. (28)–(32)], and partly-regular sparse NOMA when the transmit
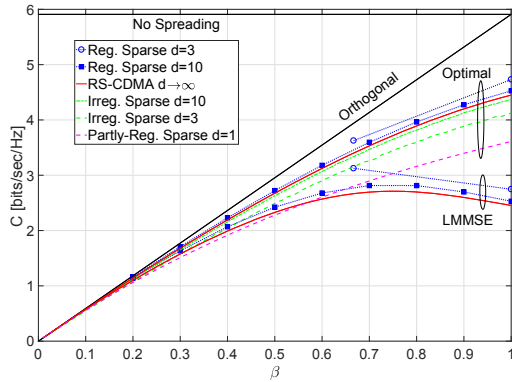
(a) Underloaded regime
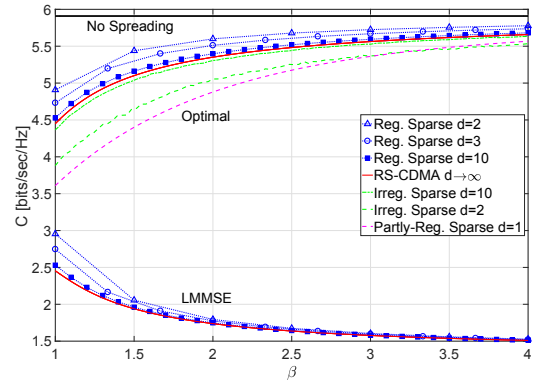
(b) Overloaded regime

Fig. 1: Limiting spectral efficiency vs. the system load β for $E_b/N_0 = 10$ dB.

energy is concentrated in a single orthogonal resource (namely, $d = 1$) [8, Eq. (13)]. The results reinforce initial observations in [1], indicating that irregular sparse NOMA schemes exhibit degraded performance due to the fact that some resources may be left unused (and in the fully irregular case some users may end up without any designated resources). Irregular sparse NOMA schemes are also observed to approach the spectral efficiency of RS-CDMA with the increase of d, however the approach to the limit is from *below* (cf. [8]).

VII. CONCLUDING REMARKS

Regular sparse NOMA has been investigated in this paper in the large-system limit. Considering the optimum receiver (whose performance can be approached using *practical* LDPC-based constructions and MPAs even in *overloaded* regimes), and the simple LMMSE receiver, the respective spectral efficiencies were expressed in *closed explicit form*, and shown to outperform the achievable throughputs of both RS-CDMA and irregular sparse NOMA. A paper with detailed derivations of the results is currently in preparation and deferred elsewhere. It is crucial to emphasize that the underlying NOMA system model is markedly different here from previously analyzed settings. Firstly, the random matrices are sparse, as opposed to standard RS-CDMA [9], where performance is governed by the *Marčenko-Pastur law*. Secondly, the entries of the signature matrix are not i.i.d., as opposed to Poissonian irregular sparse NOMA (e.g., [5]). The setting also differs from the recently analyzed partly-regular scheme [8] (time-hopping CDMA), where sparse random spreading sequences with a fixed number of non-zero entries per time frame are employed. This is since in the current setting the number of non-zero entries is identical and fixed also in each *row* of the signature matrix, in contrast to [8]. Finally, the setting differs also from the sparse models considered, e.g., in [7], where the limiting average sparsity amounts to a fixed (small) *fraction* of the dimensions (implying linear scaling); and [6], where the number of non-zero signature entries amounts to a vanishing fraction of its dimension, but is still *infinite* in the large-system limit. We

conclude by noting that our observations facilitate the understanding of the potential performance gains of sparse NOMA, and advocate employing *regular* schemes as a key practical tool for enhancing performance of future highly loaded cellular systems. Extensions to fading channels and non-Gaussian signaling are thus of major importance, and are currently pursued.

REFERENCES

[1] O. Shental, B. M. Zaidel, and S. Shamai (Shitz), "Low-density code-domain NOMA: Better be regular," in *Proc. 2017 IEEE Int. Symp. Inf. Theory (ISIT'17)*, Aachen, Germany, Jun. 25–30, 2017.

[2] Z. Ding, X. Lei, G. J. 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 J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[3] R. Hoshyar, R. Razavi, and M. Al-Imari, "LDS-OFDM an efficient multiple access technique," in *Proc. 2010 IEEE 71st Veh. Technol. Conf.*, Taipei, Taiwan, May 2010, pp. 1–5.

[4] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. 2013 IEEE 24th Int. Symp. Personal Indoor and Mobile Radio Commun. (PIMRC)*,, London, UK, Sep. 2013, pp. 332–336.

[5] M. Yoshida and T. Tanaka, "Analysis of sparsely-spread CDMA via statistical mechanics," in *Proc. 2006 IEEE Int. Symp. Inf. Theory (ISIT'06)*, Seatle, WA, Jul. 2006, pp. 2378–2382.

[6] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.

[7] D. Guo, D. Baron, and S. Shamai (Shitz), "A single-letter characterization of optimal noisy compressed sensing," in *Proc. 47th Ann. Allerton Conf. Commun. Control Comp.*, Monticello, IL, Sep./Oct. 2009.

[8] G. C. Ferrante and M.-G. Di Benedetto, "Spectral efficiency of random time-hopping CDMA," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6643–6662, Dec. 2015.

[9] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[10] C. Bordenave and M. Lelarge, "Resolvent of large random graphs," *Random Struct. Alg.*, vol. 37, pp. 332–352, 2010.

[11] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[12] S. Shamai (Shitz) and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.