Title	Energy-Efficient Precoder Design for Downlink Multi-User MISO Networks with Finite Blocklength Codes
Authors(s)	Singh, Keshav, Ku, Meng-Lin, Flanagan, Mark F.
Publication date	2021-03
Publication information	Singh, Keshav, Meng-Lin Ku, and Mark F. Flanagan. "Energy-Efficient Precoder Design for Downlink Multi-User MISO Networks with Finite Blocklength Codes." IEEE, March 2021. https://doi.org/10.1109/TGCN.2020.3045687.
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/26550
Publisher's statement	© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/TGCN.2020.3045687

Energy-Efficient Precoder Design for Downlink Multi-User MISO Networks With Finite Blocklength Codes

Keshav Singh, *Member, IEEE,* Meng-Lin Ku, *Senior Member, IEEE,* and Mark F. Flanagan, *Senior Member, IEEE*

Abstract

One of the key applications in the fifth generation (5G) cellular networks is to support ultra-reliable and low-latency communication (URLLC) which requires extremely high reliability ($\sim 99.999\%$) and low latency (< 1 ms). In this paper, the energy efficiency (EE) is maximized for downlink multi-user multiple-input single-output (MISO) networks under short packet transmission. We jointly optimize the precoders at the base station (BS) for serving multiple downlink users and the decoding error probability (DEP) with finite blocklength (FBL) codes, subject to the constraints on the BS transmit power and DEP per URLLC user. This non-convex optimization problem is then accurately approximated by applying the Dinkelbach method and approximating the channel dispersion in the high signal-to-interference-plus-noise ratio (SINR) regime and the entire (practical) SINR regime, respectively. The resulting problems are convex in the design variables of the precoders and DEPs when the remaining variables are held fixed, and using this fact we provide iterative algorithms to efficiently find locally optimal solutions. Furthermore, the convergence of the proposed algorithms and the closed-form expressions of the precoders are also derived. Simulation results validate the effectiveness of the proposed algorithms

Keshav Singh was with the School of Electrical and Electronic Engineering, University College Dublin, Ireland. He is now with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung 80424, Taiwan. (E-mail: keshav.singh@mail.nsysu.edu.tw).

Mark F. Flanagan is with the School of Electrical and Electronic Engineering, University College Dublin, Ireland. (E-mail: mark.flanagan@ieee.org).

Meng-Lin Ku is with the Department of Communication Engineering, National Central University, Taoyuan, Taiwan. (E-mail: mlku@ce.ncu.edu.tw).

that support energy-efficient URLLC as well as the impact of various network parameters such as the blocklength, number of users, distance between the BS and URLLC users and the threshold of the DEP on the EE and spectral efficiency (SE) performance.

Index Terms

Precoder design, finite blocklength (FBL) codes, base station, multiple-input single-output (MISO), ultra-reliable low-latency communication (URLLC), multi-user, energy efficiency, optimization.

I. Introduction

Ultra-reliable low-latency communication (URLLC) is one of the most important components in the fifth generation (5G) cellular networks and enables mission-critical applications such as enhanced vehicle to everything (eV2XC), e-health, tactile internet, haptic communications, and machine-type communications (MTC) [1]–[5]. The quality-of-service requirements of URLLC applications are very rigorous. For instance, mission-critical applications for industrial automation require very low latency (e.g., < 1 ms) and packet error probabilities (PEP) (e.g., 10^{-9}), and the PEP and user plane latency for eV2X applications are 10^{-5} and $3-10$ ms, respectively [6]. However, the existing wireless systems cannot meet such critical requirements. On the other side, energy efficiency (EE) of the networks is another crucial performance metric in realizing 5G radio access solutions. Therefore, it is essential to revisit the designs of the existing wireless communication networks and the associated transmission strategies to meet the requirements of URLLC applications [7], [8] as well to enhance the network's EE.

The EE depends on the achievable rate which is derived using Shannon's capacity formula in existing literature [9]-[15]. In [9], the beamforming design techniques have been investigated to maximize the EE of multi-user multiple-input single-output (MISO) downlink networks, while energy-efficient beamforming methods have been investigated in [10] for MISO heterogeneous cellular networks. The authors in [11] have proposed a pricing-based distributed beamforming algorithm for MISO interference channels. In [12], the beamforming design problems have been considered for multi-cell multi-user downlink systems along with simultaneous wireless information and power transfer (SWIPT), for which centralized and distributed beamforming algorithms have been studied to maximize the energy harvesting efficiency. Furthermore, the joint subcarrier and power allocation strategies have been investigated in [13], [14] for dual-hop two-way decode-and-forward (DF) and amplify-and-forward (AF) relay networks, while

a generalized resource allocation scheme has been studied for DF relaying networks in [15]. However, the proposed methods in [9]-[15] cannot be directly extended to URLLC systems, since much lower sum-millisecond latency and transmission error rates are required to be fulfilled while the maximal EE of the system is computed.

Polyanskiy *et al.* have derived the expression for computing the achievable rate under finite blocklength (FBL) codes and analyzed the performance of a URLLC communication system [16], [17]. By using the achievable rate derived in [17], the physical-layer and link-layer performance of a relay system for quasi-static Rayleigh fading channels and the throughput of a spectrum sharing network under finite-blocklength codewords have been investigated in [18] and [19], respectively, while the resource allocation and transmission strategies for URLLC systems have been studied in [20]. Dynamic resource allocation schemes have been investigated in [21] for vehicular networks under optimized latency and reliability, while a two-phase transmission protocol with URLLC has been proposed in [22] for device-to-device networks. Furthermore, energy-efficient resource allocation for URLLC systems has been studied in [23]-[27]. The authors in [23] have studied the energy-efficient packet scheduling algorithms, while the fundamental trade-off between energy and latency has been explored in [24] for URLLC systems. Furthermore, the authors in [26] have investigated identical content transmission for correlated sources over non-orthogonal multiple access channels. The effectiveness of successive interference cancellation (SIC) has also been studied in [26], which pointed to the nature of SIC as the main cause for spectral efficiency (SE) degradation. To the best of the authors' knowledge, the joint optimization of precoders and decoding error probability (DEP) for downlink MISO networks to support URLLC for all signal-to-interference-plus-noise ratio (SINR) regimes has not yet been fully explored in the literature. Although a precoder design method was proposed in [27] for the URLLC downlink MISO system under FBL coding, the solution in [27] is limited to the high-SINR regime, and no closed-form expression for the optimal precoder was derived or exploited for this case.

Motivated by the aforementioned discussion, in this paper, we aim at maximizing the EE of a downlink multi-user MISO network with the FBL codes, where a multi-antenna base station (BS) serves multiple single-antenna URLLC users by jointly optimizing the precoders and DEP.

The main contributions of this work are highlighted below:

- We formulate an optimization problem for downlink multi-user MISO URLLC[1] networks to maximize the network's EE under constraints of the DEP per URLLC user and on the total transmit power at the BS.
- The formulated problem is non-convex due to the non-concave EE expression, and thus it is difficult to solve. Therefore, we analyze the structure of the objective function of EE and derive an approximate but concave EE expression. For this, we firstly apply the Dinkelbach method and then approximate the channel dispersion under high-SINR regimes and also approximate the channel dispersion along with applying a successive convex approximation (SCA) method in all SINR regimes so that the non-convex problem can be successfully converted into one that is provably jointly convex in the design variables of the precoders and DEPs for fixed values of the remaining auxiliary variables and penalty factor.
- For each scenario, an energy-efficient iterative algorithm is proposed which finds a locally optimal solution for maximizing the EE of the network. Furthermore, we prove the convergence of the algorithms and derive closed-form expressions for the resulting precoders.
- Through simulation examples, we validate the effectiveness of the proposed algorithms that support energy-efficient URLLC. In addition, we simulate the SE/throughput maximization (SEM) algorithm under the URLLC consideration, enabling us to uncover the impact of various system parameters such as the blocklength, number of URLLC users, distance between the BS and URLLC users and DEP constraints on the network's EE and SE performance.

Organization: The remainder of the paper is structured as follows. In Section II, we illustrate the URLLC system model and preliminary. Next, we formulate the optimization problem for URLLC with the FBL codes in Section III, while the problem transformation and the iterative algorithms are depicted in Section IV. In Section V, we describe simulation results. Finally,

[1]Note that the end-to-end (E2E) delay of data packet transmission contains several components including the queueing delay, transmission delay, propagation delay, processing/decoding delay, and routing delay in the backhaul and core networks. For tractability, in this paper, we focus on the transmission delay [28]-[29], which is independent of the other components of the E2E delay.

conclusions and future works are presented in Section VI.

Notation: We use following notations throughout the paper. The lowercase and uppercase boldface letters (e.g., $\mathbf{a}$ and $\mathbf{A}$) are utilized to denote a vector and a matrix, respectively, while the operations of transpose, conjugate transpose, matrix inversion and element-wise conjugate are represented by $(\cdot)^T$, $(\cdot)^\dagger$, $(\cdot)^{-1}$, and $(\cdot)^*$, respectively. The symbol $\mathbb{R}_+$ denotes the set of non-negative real numbers, whereas the matrix $\mathbf{I}_M$ denotes an $M \times M$ identity matrix. The notation $E(x)$ is the expectation of x, while the real and imaginary part of the argument are denoted by $\mathfrak{R}(\cdot)$ and $\mathfrak{I}(\cdot)$, respectively. Moreover, $[\cdot]^+$ indicates $\max(0, \cdot)$.

II. System Model and Preliminaries

A. System Model

Consider a downlink multi-user MISO network in Fig. 1, consisting of K single-antenna URLLC users and one BS. The BS has M antennas which are used to transmit a different message to each of the K users. We focus on URLLC applications in low-mobility environments in which the channel coherence time is much larger than the required time for calculating the solutions with the proposed algorithms and given channel state information (CSI). Further, the channel from the BS to the kth URLLC user is assumed to be quasi-static Rayleigh fading and denoted by $\mathbf{h}_k \in \mathbb{C}^M$, $\forall k$. Suppose L_k information bits are dedicated to the k-th URLLC user and thus, the BS first encodes these information bits into a block code of length m_d symbols, expressed by $x_k[n], n = 1, 2, \ldots, m_d$. Thus, the received signal at the k-th user is expressed as

$$y_k[n] = \mathbf{h}_k^\dagger \mathbf{w}_k x_k[n] + \sum_{l=1, l\neq k}^{K} \mathbf{h}_k^\dagger \mathbf{w}_l x_l[n] + z_k[n], \quad k = 1, 2, \ldots, K, n = 1, 2, \ldots, m_d, \tag{1}$$

where $\mathbf{w}_k \in \mathbb{C}^M$ is the precoding vector at the BS for the k-th user, while $z_k[i] \sim \mathcal{CN}(0, \sigma_k^2)$ is the complex additive white Gaussian noise (AWGN). Using (1), the SINR of the user k can be computed as

$$\gamma_k = \frac{|\, \mathbf{h}_k^\dagger \mathbf{w}_k \,|^2}{\sum_{l=1, l\neq k}^{K} |\, \mathbf{h}_k^\dagger \mathbf{w}_l \,|^2 + \sigma_k^2}, \; k = 1, 2, \ldots, K. \tag{2}$$

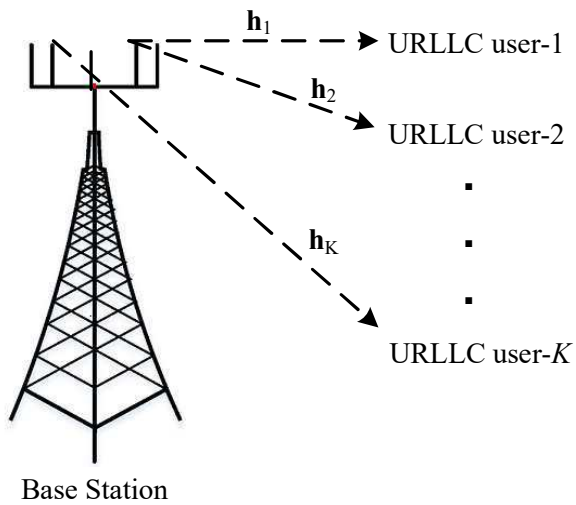


Fig. 1: Multi-user downlink MISO URLLC system.

B. Achievable Rate and Error Probability Analysis for a Point-to-Point Transmission System with FBL

Before deriving the achievable rate for the considered framework, the derivation of the achievable rate for a point-to-point communication system is discussed. For the real AWGN channel with FBL codes, an approximation to the achievable rate is given by [16, Theorem 54]

$$r \simeq C(\gamma) - \sqrt{\frac{1}{m}\left(1 - \frac{1}{(1+\gamma)^2}\right)} Q^{-1}(\epsilon) \log_2 e\,, \tag{3}$$

where γ denotes the signal-to-noise ratio (SNR), and $C(\gamma) = \log_2(1+\gamma)$, while ϵ and m indicate the DEP and blocklength, respectively. Here, $Q^{-1}(\cdot)$ in (3) represents the inverse of Gaussian $Q-$function, defined as $Q(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\omega}^{\infty} e^{-t^2/2} dt$.

Further, the authors in [17] and [18] have extended the above result with a real AWGN channel model to a complex quasi-static Rayleigh fading channel model and derived the achievable rate as follows:

$$\begin{aligned} r &= \mathcal{R}(\gamma, \epsilon, m) \\ &\simeq \mathcal{C}(\gamma) - \sqrt{\left(1 - 2^{-2\mathcal{C}(\gamma)}\right)/m}\; Q^{-1}(\epsilon) \log_2 e\,. \end{aligned} \tag{4}$$

From (4), the DEP for single-hop transmission at the receiver can be obtained by

$$\epsilon \approx Q\left(\frac{\mathcal{C}(\gamma) - r}{\sqrt{\left(1 - 2^{-2\mathcal{C}(\gamma)}\right)/m}\; \log_2 e}\right). \tag{5}$$

Remark 1: Note that when $m \geq 100$, the approximated achievable rate in (4) and DEP in (5) give more accurate results. Moreover, the achievable rate r in (4) approaches to the Shannon capacity given by $r = \log_2(1+\gamma)$ when m is sufficiently large.

III. Problem Formulation for URLLC with FBL

A. System Performance Metrics

Using (2) and (4), the achievable rate of the k-th URLLC user is computed as [16]

$$\mathcal{R}_k\left(\mathbf{w}_k, \epsilon_k, m_d\right) \triangleq \mathcal{C}\left(\gamma_k\right) - \sqrt{\frac{1}{m_d}\underbrace{\left(1-2^{-2\mathcal{C}(\gamma_k)}\right)}_{V_k}}\, Q^{-1}\left(\epsilon_k\right)\log_2 e\,, \tag{6}$$

where ϵ_k is the DEP at user k. From (6), the throughput of the network is given by

$$\mathcal{F}\left(\mathbf{w}, \boldsymbol{\epsilon}\right) = \sum_{k=1}^{K} \mathcal{R}_k\left(\mathbf{w}_k, \epsilon_k, m_d\right)\,. \tag{7}$$

where $\mathbf{w} = \{\mathbf{w}_k\}$ and $\boldsymbol{\epsilon} = \{\epsilon_k\}$. As a result, this formula provides an important insight into quantifying the influence of the BS precoders and the DEP and blocklength of each URLLC user on the network throughput.

B. Power Consumption Model

We consider transmit power, processing power, and circuit power in computing the EE of the network. The transmit power depends on channel conditions, cell coverage areas, etc., while the processing and circuit power are also known as static power consumption which depends on the number of antennas. Accordingly, the power dissipation model is defined as follows [15]:

$$P_T\left(\mathbf{w}\right) = \underbrace{\sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2}_{\substack{\text{Dynamic Power Dissipation}\\ \leqslant P_{max}}} + \underbrace{(M+K)P_{ST}}_{\substack{\text{Static Power Dissipation}\\ Q^C \geqslant 0}}\,, \tag{8}$$

where P_{ST} indicates the static power consumption per antenna at each node, while P_{max} is the total transmit power budget of the BS.

C. Problem Formulation

Using (7) and (8), a joint design problem of the precoders and DEP for maximizing the EE of the network is formulated as [15]

$$\begin{aligned} \max_{\mathbf{w},\boldsymbol{\epsilon}} \quad & \frac{\mathcal{F}\left(\mathbf{w}, \boldsymbol{\epsilon}\right)}{P_T\left(\mathbf{w}\right)} \\ \text{subject to} \quad & (C.1)\ \epsilon_k \leq \epsilon_{k,max}\,, \quad k = 1, \ldots, K\,; \\ & (C.2) \sum\nolimits_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 \leq P_{max}\,, \end{aligned} \tag{9}$$

where the constraint $(C.1)$ ensures the reliability of each URLLC user by limiting the error rate to $\epsilon_{k,max}$, while the constraint $(C.2)$ guarantees that the total BS transmit power is bounded by the maximum power budget P_{max}.

IV. Proposed EE Maximization Algorithm

The joint design problem (9) for the precoding weights and DEP is non-convex due to the non-concavity of the objective function and thus, in general, it is difficult to solve [30]. In this section, we convert this non-convex problem into one that is jointly convex in the design variables of the precoders and DEPs when the remaining variables are held fixed, and we then use this fact to propose iterative algorithms which can find locally optimal solutions. Firstly, we introduce the expansion of $\sqrt{1-z}$ and then the approximation of the inverse error function as follows.

Using [31], we get

$$\sqrt{V_k} = \sqrt{1 - \frac{1}{(1+\gamma_k)^2}} = 1 - \sum_{i=1}^{\infty} \frac{a_i}{(1+\gamma_k)^{2i}}, \tag{10}$$

with $a_i = \left| \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}-1\right), \ldots, \left(\frac{1}{2}-i+1\right)}{i!} \right|$.

Fig. 2 shows the series in (10), together with its approximation using zero terms and one term. Note that $\sqrt{V_k}$ is approximated to 1 if all terms of the infinite series in (10) are ignored, which is accurate in the high-SNR regime ($\gamma_k > 10$ dB). In Fig. 2, we can observe that when $\gamma_k > 0$ dB, the approximation of $\sqrt{V_k}$ to one term is very accurate. To accurately approximate $\sqrt{V_k}$ at lower SINR, it is necessary to include more terms of the infinite series (10). In this work, since low-SINR operation is of less practical interest, we will consider only the approximations which use zero term and one term.

We introduce the following lemma for the approximation of $Q^{-1}(z)$.

Lemma 1: For $0 < z < 1$, the approximation of $Q^{-1}(z)$ is given as

$$Q^{-1}(z) \approx \sqrt{\frac{\pi}{2}}\,(A - Bz). \tag{11}$$

where A and B are defined as

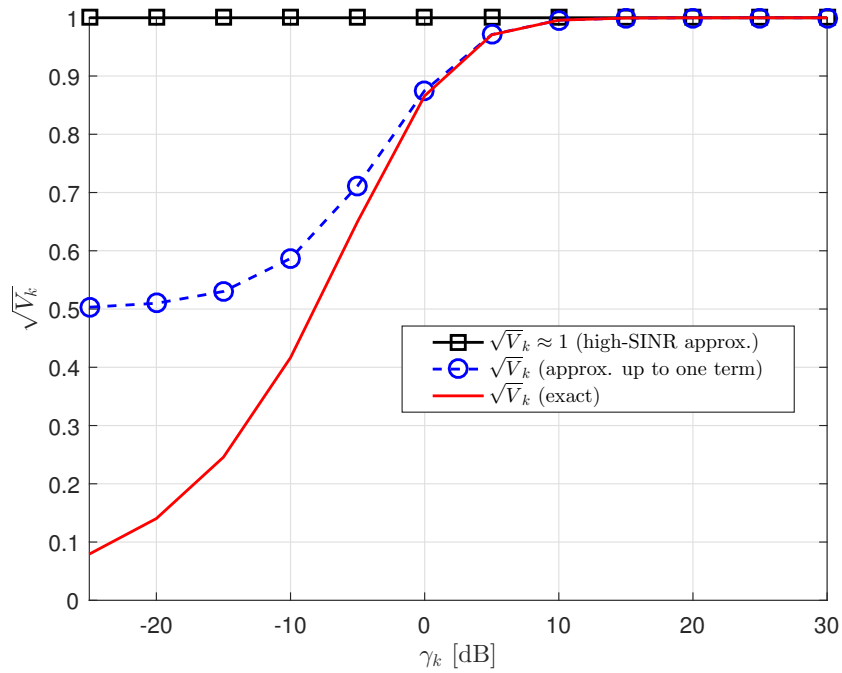


Fig. 2: Approximation of channel dispersion V_k with only zero and one term.

$$A = \left(1 + \frac{\pi}{12} + \frac{7\pi^2}{480} + \frac{127\pi^3}{40320} + \cdots\right); \tag{12}$$

$$B = \left(2 + \frac{\pi}{2} + \frac{7\pi^2}{48} + \frac{127\pi^3}{2880} + \cdots\right). \tag{13}$$

Proof: Please refer to Appendix A. ■

We will utilize (10) and Lemma 1 in the following subsections in order to transform the non-convex problem (9) into one that is jointly convex in the design variables of the precoders $\mathbf{w}$ and DEPs $\boldsymbol{\epsilon}$.

A. *Transformation of Problem* (9)

We utilize the Dinkelbach method [32] to transform the fractional objective function into an equivalent subtractive form. For this, we introduce the following proposition:

Proposition 1: Let $(\mathbf{w}^\star, \boldsymbol{\epsilon}^\star)$ be the optimal strategy which attains the maximum $q^\star$ such that

$$q^\star = \frac{\mathcal{F}(\mathbf{w}^\star, \boldsymbol{\epsilon}^\star)}{P_T(\mathbf{w}^\star)} = \max_{\mathbf{w},\boldsymbol{\epsilon}} \frac{\mathcal{F}(\mathbf{w}, \boldsymbol{\epsilon})}{P_T(\mathbf{w})}. \tag{14}$$

This is true if and only if

$$\max_{\mathbf{w},\boldsymbol{\epsilon}} \mathcal{F}(\mathbf{w}, \boldsymbol{\epsilon}) - q^\star P_T(\mathbf{w}) = \mathcal{F}(\mathbf{w}^\star, \boldsymbol{\epsilon}^\star) - q^\star P_T(\mathbf{w}^\star) = 0. \tag{15}$$

Proof: The detailed proof may be found in [32]. ■

By using Proposition 1, the problem (9) can be equivalently written as

$$\begin{aligned} \max_{\mathbf{w},\boldsymbol{\epsilon}} \quad & \mathcal{F}\left(\mathbf{w},\boldsymbol{\epsilon}\right) - qP_T\left(\mathbf{w}\right) \\ \text{subject to} \quad & (C.1)\ \epsilon_k \leq \epsilon_{k,max}\,, \quad k=1,\ldots,K\,; \\ & (C.2) \sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 \leq P_{max}\,, \end{aligned} \tag{16}$$

where the variable q can be regarded as a penalty factor. The problem (16) is still non-convex due to the presence of the channel dispersion in the objective function. Since the channel dispersion V_k in (6) depends on γ_k, it is very difficult to find the solution of the problem (9). Therefore, we consider the high-SINR and one-term approximation of V_k and propose energy-efficient iterative algorithms in the following subsections.

B. Approximation of V_k under high-SINR regime

At the high-SINR regime, the value of the term $\left(1-\frac{1}{\left(1+\gamma_k\right)^2}\right)$ in V_k becomes close to zero and thus, the channel dispersion V_k can be approximated as

$$V_k = 1 - \frac{1}{\left(1+\gamma_k\right)^2} \approx 1\,. \tag{17}$$

Also we have

$$1+\gamma_k \approx \gamma_k. \tag{18}$$

It is noticed that with the approximations in (17) and (18), the approximated throughput in the objective function can serve as a lower bound for $\mathcal{F}\left(\mathbf{w},\boldsymbol{\epsilon}\right)$ in (7). That is, from (A.4), (17), and (18), the optimization problem (16) can be reformulated by maximizing the corresponding lower bound on the EE as follows:

$$\begin{aligned} \max_{\mathbf{w},\boldsymbol{\epsilon}} \quad & \sum_{k=1}^{K}\left(\log_2\left(\gamma_k\right) - \sqrt{\frac{\pi}{2m_d}}\left(A-B\epsilon_k\right)\log_2 e\right) - q\times P_T\left(\mathbf{w}\right) \\ \text{subject to} \quad & (C.1)\ \epsilon_k \leq \epsilon_{k,max}\,, \quad k=1,\ldots,K\,; \\ & (C.2) \sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 \leq P_{max}\,. \end{aligned} \tag{19}$$

Without loss of generality, we assume that the inner product of the channel vector $\mathbf{h}_k$ and the precoding weight $\mathbf{w}_k$ is a positive real number, i.e., producing an amplitude gain. By substituting

the value of γ_k in (19) and introducing auxiliary variables $\{\Omega_k\}$, the problem (19) can be equivalently written as

$$\begin{aligned}
\max_{\mathbf{w},\boldsymbol{\epsilon},\boldsymbol{\Omega}} \quad & \sum_{k=1}^{K}\left(2\log_2\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right)-\log_2\left(\Omega_k\right)-\sqrt{\frac{\pi}{2m_d}}\left(A-B\epsilon_k\right)\log_2 e\right)-q\times P_T\left(\mathbf{w}\right) \\
\text{subject to} \quad & (C.1)\ \epsilon_k \leq \epsilon_{k,max}\,,\quad k=1,\ldots,K\,; \\
& (C.2)\sum_{k=1}^{K}\parallel \mathbf{w}_k \parallel_2^2 \leq P_{max}\,; \\
& (C.3)\ \sum_{l=1,l\neq k}^{K}\mid \mathbf{h}_k^{\dagger}\mathbf{w}_l \mid^2 + \sigma_k^2 \leq \Omega_k\,,\ k=1,\ldots,K\,; \\
& (C.4)\ \Re\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right)\geq 0\,,\quad k=1,\ldots,K\,; \\
& (C.5)\ \Im\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right)=0\,,\quad k=1,\ldots,K\,,
\end{aligned} \tag{20}$$

where $\boldsymbol{\Omega}=\{\Omega_k\}$.

Lemma 2: The optimization problem (20) is jointly convex in variables $\mathbf{w}$ and $\boldsymbol{\epsilon}$ for given $\boldsymbol{\Omega}$ and q.

Proof: Please refer to Appendix B. ■

1) Iterative EE maximization (EEM) algorithm (Algorithm 1): In this subsection, we illustrate an iterative EE maximization (EEM) algorithm that provides a solution to the optimization problem (20). Note that the approximated problem under high-SINR conditions given by (20) is not jointly convex in the variables $\mathbf{w}$, $\boldsymbol{\epsilon}$, $\boldsymbol{\Omega}$ and q due to non-concavity (coupling of variables) of the objective function. However, for given $\boldsymbol{\Omega}$ and q, the approximated optimization problem (20) becomes jointly convex in variables $\mathbf{w}$ and $\boldsymbol{\epsilon}$. Therefore, we can apply standard numerical techniques, *but in an iterative manner*, to solve the problem (20). For example, we can use the CVX tool [23] to solve this problem for given $\boldsymbol{\Omega}$ and q in the first step and then update $\boldsymbol{\Omega}$ and q with the obtained solutions of $\mathbf{w}$ and $\boldsymbol{\epsilon}$ in the second step. However, in this paper we propose an iterative algorithm via primal-dual decomposition to solve the problem (20) in the first step. As compared with the CVX tool, this yields a very efficient splitting optimization scheme, according to which a solution to the original problem is iteratively computed through solving a sequence of simpler subproblems, each one involving only one of the terms appearing in the objective function.

The Lagrange function of the problem (20) is expressed by

$$\begin{aligned}
&\mathcal{L}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\lambda_k\}, \mu, \{\nu_k, \upsilon_k, \bar{\upsilon}_k\}\right) = \\
&\sum_{k=1}^{K}\left(2\log_2\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right) - \log_2\left(\Omega_k\right) - \sqrt{\frac{\pi}{2m_d}}\left(A - B\epsilon_k\right)\log_2 e\right) \\
&- q \times \left(\sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 + Q^C\right) - \sum_{k=1}^{K}\lambda_k\left(\epsilon_k - \epsilon_{k,max}\right) - \mu\left(\sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 - P_{max}\right) \\
&- \sum_{k=1}^{K}\nu_k\left(\sum_{l=1, l\neq k}^{K}\frac{1}{\sigma_k^2} \mid \mathbf{h}_k^{\dagger}\mathbf{w}_l \mid^2 + 1 - \frac{\Omega_k}{\sigma_k^2}\right) + \sum_{k=1}^{K}\upsilon_k\Re\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right) + \sum_{k=1}^{K}\bar{\upsilon}_k\Im\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right), \qquad (21)
\end{aligned}$$

where $\{\lambda_k\}$, $\mu \geq 0$ and $\nu_k, \upsilon_k, \bar{\upsilon}_k \geq 0, \forall k$ are the Lagrange multipliers associated with the constraints $(C.1)$, $(C.2)$, $(C.3)$, $(C.4)$, and $(C.5)$ of the problem (20), respectively. Since the problem in (20) satisfies Slater's condition [30], we can find the optimal solution of (20) by solving its dual as follows:

$$\min_{\{\lambda_k\}, \mu, \{\nu_k, \upsilon_k, \bar{\upsilon}_k\}} \max_{\mathbf{w}, \boldsymbol{\epsilon}} \mathcal{L}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\lambda_k\}, \mu, \{\nu_k, \upsilon_k, \bar{\upsilon}_k\}\right) . \qquad (22)$$

The dual problem (22) can be solved iteratively. We firstly update the precoding vector $\mathbf{w}$ and the DEP vector $\boldsymbol{\epsilon}$ by fixing $(\lambda, \mu, \{\nu_k\})$ and then, we compute the Lagrangian multipliers $(\lambda, \mu, \{\nu_k\})$ with obtained $\mathbf{w}$ and $\boldsymbol{\epsilon}$. The proposed iterative EEM algorithm under the V_k approximation for the high-SINR regime is summarized in Algorithm 1.

Theorem 1: The closed-form expression for the precoding vector $\mathbf{w}_k$ and the update equation for the DEP ϵ_k are given, respectively, by

$$\mathbf{w}_k^{\star} = \left(\frac{1}{2}\upsilon_k - \frac{1}{2j}\bar{\upsilon}_k\right) \times \left((-q+\mu)\,\mathbf{I}_M + \sum_{l=1,\, l\neq k}^{K}\frac{\nu_l}{\sigma_l^2}\mathbf{h}_l\mathbf{h}_l^{\dagger}\right)^{-1}\mathbf{h}_k\,,\ k = 1, \ldots, K\,; \qquad (23)$$

$$\epsilon_k(l+1) = \left[\epsilon_k(l) + \alpha_{\epsilon_k(l)}\left(\sqrt{\frac{\pi}{2m_d}}B\log_2 e - \lambda_k\right)\right]^{+}\,,\ k = 1, \ldots, K\,, \qquad (24)$$

where $\alpha_{\epsilon_k}(l)$ is the step size at iteration l.

Proof: Please refer to Appendix C. ■

By applying the gradient method [30], [34], the Lagrange multipliers $(\lambda, \mu, \{\nu_k, \upsilon_k, \bar{\upsilon}_k\})$ are updated with the obtained $\mathbf{w}^{\star}$ and $\boldsymbol{\epsilon}$, while we adopt the method defined in [35, Section V.C] to update $\boldsymbol{\Omega}$.

Algorithm 1 An iterative EEM algorithm under V_k approximation for high-SINR regimes

Input: • Define the maximum number of iterations L_{max} and convergence tolerance ϵ_{out}

• Initialize the iteration counter $l = 0$ and penalty factor $q[0] \leftarrow 0$ and Ω_k , $\forall k$.

Repeat

Solve the problem (20) for given $\mathbf{\Omega}$ and q to achieve $\mathbf{w}^\star$ and $\boldsymbol{\epsilon}^\star$.

Update Ω_k as in [35] for obtained $\mathbf{w}^\star$, $\boldsymbol{\epsilon}^\star$ and given q.

Update $q(l+1) \leftarrow \dfrac{\mathcal{F}\left(\mathbf{w}^\star(l), \boldsymbol{\epsilon}^\star(l)\right)}{P_T\left(\mathbf{w}^\star(l)\right)}$.

Set $l \leftarrow l+1$

Update $\mathbf{w}_k(l+1) \leftarrow \mathbf{w}_k(l)$ and $\epsilon_k(l+1) \leftarrow \epsilon_k(l)$.

Until convergence or $l = L_{max}$

Return $(\mathbf{w}_k^\star, \epsilon_k^\star)$.

2) Updating q: The update procedure of q which works as a penalty factor for allocating the transmit power at the BS is described as follows. Let $(\mathbf{w}^\star, \boldsymbol{\epsilon}^\star, \mathbf{\Omega}^\star)$ be the optimal solution of the problem (20) with respect to $q^\star$ at the iteration l, and the penalty factor q is updated as

$$q(l+1) = \frac{\mathcal{F}\left(\mathbf{w}^\star(l), \boldsymbol{\epsilon}^\star(l)\right)}{P_T\left(\mathbf{w}^\star(l)\right)} . \tag{25}$$

Theorem 2: The penalty factor q increases monotonically with the iteration number l until convergence.

Proof: Please refer to Appendix D. ■

C. Approximation of V_k up to one term

Using (10), $\sqrt{V_k}$ can be approximated with one more term as follows:

$$\sqrt{V_k} = \sqrt{1 - \frac{1}{\left(1+\gamma_k\right)^2}} \approx 1 - \frac{1}{2\left(1+\gamma_k\right)^2} . \tag{26}$$

Using (26) and Lemma 1, we can reformulate $\mathcal{F}(\mathbf{w},\boldsymbol{\epsilon})$ as

$$\mathcal{F}(\mathbf{w},\boldsymbol{\epsilon}) = \sum_{k=1}^{K} \mathcal{R}_k(\mathbf{w}_k, \epsilon_k, m_d) \ . \tag{27}$$

$$\triangleq \sum_{k=1}^{K} \left(\log_2(1+\gamma_k) - \sqrt{\frac{\pi}{2m_d}} \left(1 - \frac{1}{2(1+\gamma_k)^2}\right) (A - B\epsilon_k) \log_2 e \right) . \tag{28}$$

Notice that the term $1 - \dfrac{1}{2(1+\gamma_k)^2}$ works as a dynamic penalty factor for the DEP requirement and varies with the instantaneous SINR value γ_k.

From (27), the optimization problem in (16) is rewritten as

$$\max_{\mathbf{w},\boldsymbol{\epsilon}} \quad \sum_{k=1}^{K} \left(\log_2(1+\gamma_k) - \sqrt{\frac{\pi}{2m_d}} \left(1 - \frac{1}{2(1+\gamma_k)^2}\right) (A - B\epsilon_k) \log_2 e \right) - q \times P_T(\mathbf{w})$$

$$\text{subject to} \quad (C.1) \ \text{and} \ (C.2) \ \text{in (16)} . \tag{29}$$

Since the problem (29) is still non-convex due to the non-concavity of the objective function, we introduce auxiliary variables $\{\Xi_k\}$ and $\{\Psi_k\}$ as follows:

$$\sum_{l=1, l\neq k}^{K} \mid \mathbf{h}_k^{\dagger} \mathbf{w}_l \mid^2 + \sigma_k^2 \leq \Xi_k \ , \ k = 1 \ldots, , K \ . \tag{30}$$

$$1 - \frac{1}{2(1+\gamma_k)^2} \leq \Psi_k \ , \ k = 1 \ldots, , K. \tag{31}$$

From (2) and (30), we can re-express (31) as

$$\mid \mathbf{h}_k^{\dagger} \mathbf{w}_k \mid^2 \leq \Xi_k \left(\sqrt{\frac{1}{2(1-\Psi_k)}} - 1 \right) \ , \ k = 1 \ldots, , K \ . \tag{32}$$

By using (30) and (32), the optimization problem in (29) can be reformulated as

$$\max_{\mathbf{w},\boldsymbol{\epsilon},\boldsymbol{\Psi}} \quad \sum_{k=1}^{K} \left(\log_2 \left(1 + \frac{\mid \mathbf{h}_k^{\dagger} \mathbf{w}_k \mid^2}{\Xi_k} \right) - \Psi_k \sqrt{\frac{\pi}{2m_d}} \ (A - B\epsilon_k) \log_2 e \right) - q \times P_T(\mathbf{w})$$

$$\begin{aligned} \text{subject to} \quad & (C.1) \ \epsilon_k \leq \epsilon_{k,max} \ , \quad k = 1, \ldots, K \ ; \\ & (C.2) \sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 \leq P_{max} \ ; \\ & (C.3) \sum_{l=1, l\neq k}^{K} \mid \mathbf{h}_k^{\dagger} \mathbf{w}_l \mid^2 + \sigma_k^2 \leq \Xi_k \ , \ k = 1, \ldots, K \ ; \\ & (C.4) \mid \mathbf{h}_k^{\dagger} \mathbf{w}_k \mid^2 \leq \Xi_k \left(\sqrt{\frac{1}{2(1-\Psi_k)}} - 1 \right) \ , \ k = 1, \ldots, K \ . \end{aligned} \tag{33}$$

Due to the non-concavity of the objective functions in (33), the optimization problem (33) is still non-convex. Now, we consider the lower bound on $\sum_{k=1}^{K}\log_2\left(1+\Upsilon_k\right)$ as [13, Lemma 2]

$$\sum_{k=1}^{K}\log_2\left(1+\Upsilon_k\right)\geq\sum_{k=1}^{K}\rho_k\log_2\left(\Upsilon_k\right)+\varrho_k\,, \tag{34}$$

where $\Upsilon_k=\dfrac{\mid\mathbf{h}_k^{\dagger}\mathbf{w}_k\mid^2}{\Xi_k}$, and both the coefficients ρ_k and ϱ_k are selected for any given $\tilde{\Upsilon}_k\geq 0$ as

$$\rho_k=\frac{\tilde{\Upsilon}_k}{1+\tilde{\Upsilon}_k}\,; \tag{35}$$

$$\varrho_k=\log_2\left(1+\tilde{\Upsilon}_k\right)-\rho_k\log_2\left(\tilde{\Upsilon}_k\right)\,. \tag{36}$$

With the chosen coefficients in (35) and (36), the equality of (34) holds when $\Upsilon_k=\tilde{\Upsilon}_k$ and $(\varrho_k,\rho_k)=(1,0)$ and this bound becomes tight as $\Upsilon_k\rightarrow\infty$.

Using (34) and similar to Section IV.B, we assume that the inner product of $\mathbf{h}_k$ and $\mathbf{w}_k$ is a positive real number. Thus, by using (34) the problem (33) can be equivalently rewritten as

$$\max_{\mathbf{w},\boldsymbol{\epsilon},\boldsymbol{\Xi},\boldsymbol{\Psi}}\ \sum_{k=1}^{K}\left(\rho_k\left(2\log_2\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right)-\log_2\left(\Xi_k\right)\right)+\varrho_k-\Psi_k\sqrt{\frac{\pi}{2m_d}}\left(A-B\epsilon_k\right)\log_2 e\right)-q\times P_T\left(\mathbf{w}\right)$$

$$\begin{aligned}
\text{subject to}\ &(C.1)\ \ \epsilon_k\leq\epsilon_{k,max}\,,\quad k=1,\ldots,K\,;\\
&(C.2)\sum_{k=1}^{K}\parallel\mathbf{w}_k\parallel_2^2\leq P_{max}\,;\\
&(C.3)\sum_{l=1,l\neq k}^{K}\mid\mathbf{h}_k^{\dagger}\mathbf{w}_l\mid^2+\sigma_k^2\leq\Xi_k\,,\ k=1,\ldots,K\,;\\
&(C.4)\ \mid\mathbf{h}_k^{\dagger}\mathbf{w}_k\mid^2\leq\Xi_k\left(\sqrt{\frac{1}{2\left(1-\Psi_k\right)}}-1\right)\,,\ k=1,\ldots,K\,;\\
&(C.5)\ \ \Re\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right)\geq 0\,,\quad k=1,\ldots,K\,;\\
&(C.6)\ \ \Im\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right)=0\,,\quad k=1,\ldots,K\,,
\end{aligned} \tag{37}$$

where $\boldsymbol{\Xi}=\{\Xi_k\}$.

Lemma 3: For given $\boldsymbol{\Xi}$, $\boldsymbol{\Psi}$ and q, the optimization problem in (37) is convex in the variables $\mathbf{w}$ and $\boldsymbol{\epsilon}$.

Proof: Please refer to Appendix E. ■

1) Iterative EEM algorithm (Algorithm 2): In this subsection, an iterative EEM algorithm is proposed to find a solution to the optimization problem (37). The Lagrange function of the problem (37) is written as

$$\begin{aligned}
&\tilde{\mathcal{L}}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\tilde{\lambda}_k\}, \tilde{\mu}, \{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\}\right) = \\
&\sum_{k=1}^{K}\left(\rho_k\left(2\log_2\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right) - \log_2\left(\Xi_k\right)\right) + \varrho_k - \Psi_k\sqrt{\frac{\pi}{2m_d}}\left(A - B\epsilon_k\right)\log_2 e\right) \\
&- q \times \left(\sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 + Q^C\right) - \sum_{k=1}^{K}\tilde{\lambda}_k\left(\epsilon_k - \epsilon_{k,max}\right) - \tilde{\mu}\left(\sum_{k=1}^{K} \parallel \mathbf{w}_k \parallel_2^2 - P_{max}\right) \\
&- \sum_{k=1}^{K}\varsigma_k\left(\sum_{l=1, l\neq k}^{K} \mid \mathbf{h}_k^{\dagger}\mathbf{w}_l \mid^2 + \sigma_k^2 - \Xi_k\right) - \sum_{k=1}^{K}\tilde{\nu}_k\left(\mid \mathbf{h}_k^{\dagger}\mathbf{w}_k \mid^2 - \Xi_k\left(\sqrt{\frac{1}{2\left(1-\Psi_k\right)}} - 1\right)\right) \\
&+ \sum_{k=1}^{K}\psi_k\mathfrak{R}\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right) + \sum_{k=1}^{K}\bar{\psi}_k\mathfrak{I}\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right) , \qquad (38)
\end{aligned}$$

where $\{\tilde{\lambda}_k\}, \tilde{\mu} \geq 0$ and $\{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\} \geq 0$, $\forall k$ are the Lagrange multipliers corresponding to the constraints $(C.1)$, $(C.2)$, $(C.3)$, $(C.4)$, $(C.5)$, and $(C.6)$ of the problem (37), respectively. Similar to the dual function of (21), we can define the dual problem of (37) as

$$\min_{\{\tilde{\lambda}_k\}, \tilde{\mu}, \{\varsigma_k \tilde{\nu}_k, \rho_k, \varrho_k\}} \max_{\mathbf{w}, \boldsymbol{\epsilon}} \tilde{\mathcal{L}}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\tilde{\lambda}_k\}, \tilde{\mu}, \{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\}\right) , \qquad (39)$$

and obtain the optimal solution of the problem (37). We summarize the proposed iterative EEM algorithm under V_k approximation with one term in Algorithm 2.

Theorem 3: The structure of the optimal precoding vector $\mathbf{w}_k$ is given by

$$\mathbf{w}_k^{\star} = \left(\frac{1}{2}\rho_k - \frac{1}{2j}\varrho_k\right) \times \left(\left(q + \tilde{\mu}\right)\mathbf{I}_M + \tilde{\nu}_k\mathbf{h}_k\mathbf{h}_k^{\dagger} + \sum_{l=l\neq k}^{K}\varsigma_l\mathbf{h}_l\mathbf{h}_l^{\dagger}\right)^{-1}\mathbf{h}_k , \quad k = 1, \ldots, K . \qquad (40)$$

Proof: Please refer to Appendix F. ■

Lemma 4: The update expression for the DEP ϵ_k is given by

$$\epsilon_k(l+1) = \left[\epsilon_k(l) + \bar{\alpha}_{\epsilon_k(l)}\left(\sqrt{\frac{\pi}{2m_d}}\Psi_k B\log_2 e - \lambda_k\right)\right]^{+} , \; k = 1, \ldots, K , \qquad (41)$$

where $\bar{\alpha}_{\epsilon_k}$ is the step size at iteration l.

Proof: Please refer to Appendix G. ■

Algorithm 2 An iterative EEM algorithm under V_k approximation with one term

Input: • Define the maximum number of iterations L_{max} and convergence tolerance ϵ_{out}

• Initialize the iteration counter $l = 0$ and penalty factor $q(0) \leftarrow 0$ and $\Omega_k, \forall k$.

Repeat

Solve the problem (37) for given $\mathbf{\Psi}$ and q to achieve $\mathbf{w}^\star$ and $\boldsymbol{\epsilon}^\star$.

Update $\mathbf{\Xi}$ as in [35] for obtained $\mathbf{w}^\star$ and $\boldsymbol{\epsilon}^\star$ and given q.

Update Ψ_k as $\Psi_k(t+1) \leftarrow \left[1 - \frac{1}{2\left(1+\gamma_k^\star(t)\right)^2}\right]$ for obtained $\mathbf{w}^\star$ and given q.

Update $q(l+1) \leftarrow \frac{\mathcal{F}\left(\mathbf{w}^\star(l), \boldsymbol{\epsilon}^\star(l)\right)}{P_T\left(\mathbf{w}^\star(l)\right)}$.

Set $l \leftarrow l+1$

Update $\mathbf{w}_k(l+1) \leftarrow \mathbf{w}_k(l)$ and $\epsilon_k(l+1) \leftarrow \epsilon_k(l)$.

Until convergence or $l = L_{max}$

Return $(\mathbf{w}_k^\star, \epsilon_k^\star)$.

2) Update of Ψ_k: Let $\mathbf{w}, \boldsymbol{\epsilon}$ be the optimal solution of the problem (37) for fixed q in the l-th outer iteration. Then, the auxiliary variable Ψ_k can be updated in the $(t+1)$-th inner iteration with the obtained solution $\mathbf{w}(t)$ in the t-th inner iteration as

$$\Psi_k(t+1) = 1 - \frac{1}{2\left(1+\gamma_k^\star(t)\right)^2}, \tag{42}$$

where $\gamma_k^\star(t)$ is the SINR with respect to the obtained solution $\mathbf{w}(t)$. By using the gradient method [30], [34], we update the Lagrange multipliers $\{\tilde{\lambda}_k\}$, $\tilde{\mu}$ and $\left(\{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\}\right)$ with the obtained $\mathbf{w}^\star$ and $\boldsymbol{\epsilon}^\star$. To update $\mathbf{\Xi}$, we adopt the method given in [35, Section V.C].

Theorem 4: The penalty factor q is monotonically increasing with the iteration number l and converges

Proof: The proof of this result is similar to that of Theorem 2. ■

D. Design Relation of Algorithm 1 and Algorithm 2

The main design flow of our work is illustrated in Fig. 3 and can be briefly summarized as follows. First, the Dinkelbach method is used to convert the fractional EE objective function

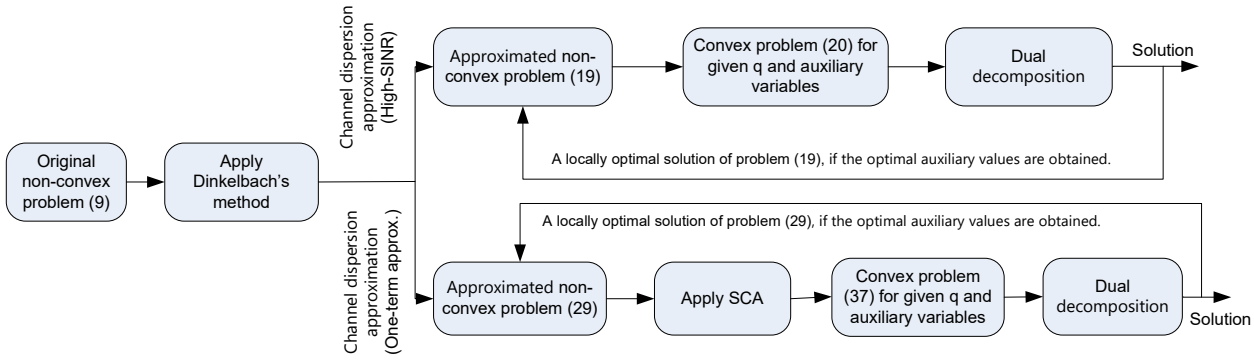


Fig. 3: Relation between the solutions of the transformed problems 1 and 2 and the corresponding original non-convex problems.

to a subtractive form. Second, two channel dispersion approximation schemes are adopted to transform the original non-convex problem (9) into two approximated non-convex problems (19) and (29). After this, the successive convex approximation (SCA) is further applied to the second approximated non-convex problem (29), and finally, these two approximated non-convex problems become convex for given auxiliary variables and penalty factor in (20) and (37), respectively. This lower bound is adopted in the SCA in the second approximated non-convex problem (29) to deal with the original non-concave objective function. Instead of considering a fixed lower bound, the lower bound is adaptively adjusted in successive iterations along with the updated coefficients ρ_k and ϱ_k, as described in (35) and (36). This approach is called SCA [33], which enables the monotonic improvement of the EE performance and finally reaches to a Karush-Kuhn-Tucker (KKT) point of the approximated non-convex problem under given auxiliary variables. As such, we say that the policies obtained by solving (20) and (37) are locally optimal solutions for the approximated non-convex problems (19) and (29) under the optimal values of auxiliary variables and penalty factor, respectively.

E. Complexity Analysis of Algorithm 1 and Algorithm 2

To get a better insight into the complexity of the proposed algorithms, in the following we provide a complexity analysis for the proposed Algorithms 1 and 2, assuming that the network penalty factor q converges in U iterations. We require to solve one sub-problem for each of the K users, thus the complexity is $\mathcal{O}(K)$. If ϵ_k takes V discrete values, then the complexity to solve the problem is $\mathcal{O}(V^3)$. If Ω_k takes Z discrete values, then the complexity is $\mathcal{O}(Z^3)$. The

TABLE I: Complexity analysis for the proposed algorithms

Complexity Analysis	
Algorithm	**Complexity**
EEM Algorithm 1	$\mathcal{O}\left(PQRK5(2K)^{\rho}(3M+5)(V^3+Z^3)\right)$
EEM Algorithm 2	$\mathcal{O}\left(PQURK6(2K)^{\rho}(3M+6)(V^3+Z^3)\right)$

computational complexity due to constraint $(C.2)$ is $\mathcal{O}(3M)$ while the other constraints impose a complexity of $\mathcal{O}(1)$. Further, updating the dual variable has a complexity of $\mathcal{O}((2K)^{\rho})$ (for example, we set $\rho = 2$ if the ellipsoid method is used), thus the total complexity to update all the dual variables and obtain $\mathbf{w}$ becomes $\mathcal{O}(5(2K)^{\rho})$. Let us assume that the dual objective function and the update of the auxiliary variables Ω and q take P, Q, R iterations to converge, respectively. Then the total complexity of Algorithm 1 is given by $\mathcal{O}\left(PQRK5(2K)^{\rho}(3M+5)(V^3+Z^3)\right)$. In a similar fashion, let us consider that the update of the auxiliary variable Ξ takes U iterations to converge in the case of the optimization problem (48), and thus there is an additional dual variable, adding a complexity of $\mathcal{O}((2K)^{\rho})$. Therefore, the total complexity for Algorithm 2 can be given by $\mathcal{O}\left(PQURK6(2K)^{\rho}(3M+6)(V^3+Z^3)\right)$.

F. Maximize EE with Infinite blocklength (IFBL)

1) IFBL under high-SINR regime: When $m_d \to \infty$, the EE maximization problem in (20) becomes that of EE maximization with infinite blocklength (IFBL):

$$\begin{aligned} \max_{\mathbf{w},\mathbf{\Omega}} \quad & \sum\nolimits_{k=1}^{K} \left(2\log_2\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right) - \log_2\left(\Omega_k\right)\right) - q \times P_T\left(\mathbf{w}\right) \\ \text{subject to} \quad & (C.2)\ ,(C.3)\ ,(C.4)\ \&\ (C.5) \text{ in (20).} \end{aligned} \tag{43}$$

The proposed Algorithm 1 can also be applied to the problem (43) to find the optimal precoding vectors $\mathbf{w}^{\star}$ for this case.

2) IFBL under one-term approximation of dispersion:: Similar to (43), the EE maximization problem with IFBL can be formulated by letting $m_d \to \infty$ in the problem (37) as

$$\begin{aligned} \max_{\mathbf{w},\mathbf{\Xi}} \quad & \sum\nolimits_{k=1}^{K} \left(\rho_k\left(2\log_2\left(\mathbf{h}_k^{\dagger}\mathbf{w}_k\right) - \log_2\left(\Xi_k\right)\right) + \varrho_k\right) - q \times P_T\left(\mathbf{w}\right) \\ \text{subject to} \quad & (C.2)\ ,(C.3)\ ,(C.5)\ \&\ (C.6) \text{ in (37).} \end{aligned} \tag{44}$$

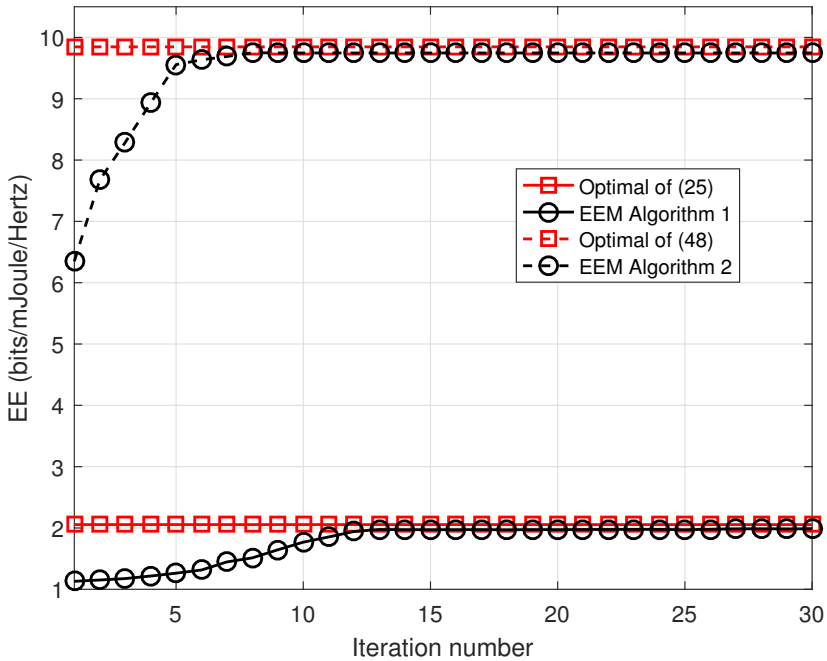


Fig. 4: Convergence of the proposed EEM algorithms.

and Algorithm 2 can be applied to this problem to obtain the optimal $\mathbf{w}^{\star}$ which maximizes the EE for this case. In fact, the obtained solution in this case is a local maximizer for the problem (9) with IFBL:

$$\begin{aligned} \max_{\mathbf{w},\boldsymbol{\epsilon}} \quad & \frac{\mathcal{F}\left(\mathbf{w},\boldsymbol{\epsilon}\right)}{P_T\left(\mathbf{w}\right)} \\ \text{subject to} \quad & (C.2) \text{ in (9).} \end{aligned} \tag{45}$$

V. Numerical Results

In this section, we demonstrate the effectiveness of the proposed EEM algorithms for downlink multiuser MISO networks under FBL codes by computer simulations.

A. Simulation Setup

A single cell with a radius of 500 meters is considered, where the BS is placed at the center of the cell and all the users are assumed to be located at the cell edge. The maximum number of iterations is set as $L_{max} = 30$. The transmit power budget at the BS is set as $P_{max} = 45$ dBm and the convergence tolerance is set to 10^{-5}. Furthermore, we model the path loss as $35.3 + 37.6 \times \log_{10}(d_k)$ dB, (where d_k denotes the distance from the BS to URLLC user k in

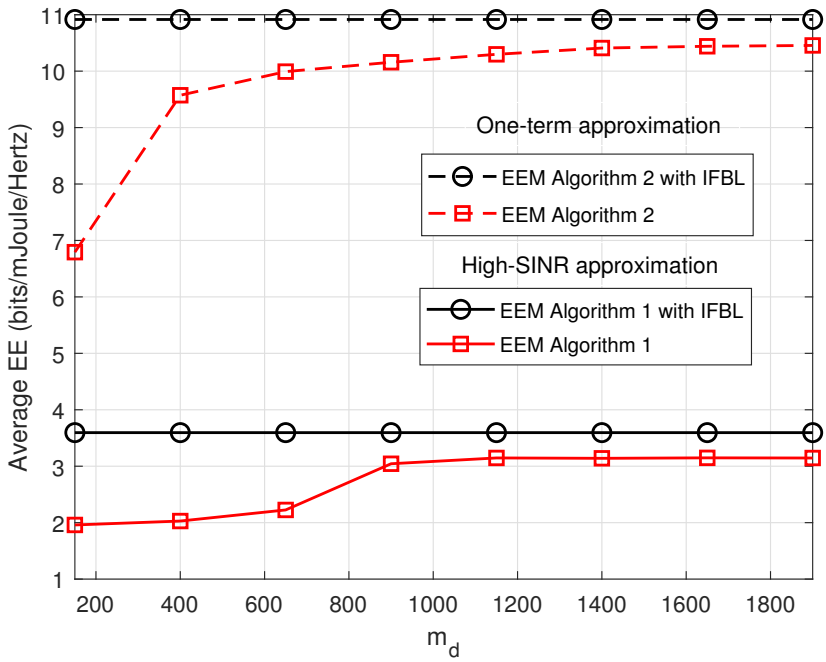


Fig. 5: Impact of the blocklength m_d on the average EE.

meters) [36], while the bandwidth is set to be 15 MHz. We set $d_k = 500$ meters, $P_{ST} = 10$ dBm and the thermal noise density is -174 dBm/Hz. The value of the maximum tolerable DEP for each user is set as $\epsilon_{k,max} = 10^{-7}$. The maximization of the EE with IFBL codes, which works as an upper bound on the EE attainable by the proposed algorithms, is simulated for performance comparison.

B. Convergence of the proposed algorithm

Fig. 4 shows the convergence behavior of the proposed EEM Algorithm 1 and Algorithm 2 for the considered network. An exhaustive search algorithm that gives the *optimal* solution is also implemented in both scenarios i.e., high-SINR and one-term approximation, for performance comparison. The parameter settings are $K = 2$, $M = 12$, $m_d = 250$, and $d_k = 500$ meters. In Fig. 4, we can observe that in both scenarios, the EE of the network increases monotonically and the proposed EEM Algorithm 1 and Algorithm 2 take fewer than 15 iterations to converge and to approach the performance of the optimal solution.

C. Impact of the blocklength

Fig. 5 demonstrates the impact of the blocklength on the achievable EE for $\epsilon_{k,max} = 10^{-7}$ and $d_k = 500$ meters. For both algorithms, the values of K and M are set as 2 and 12, respectively.

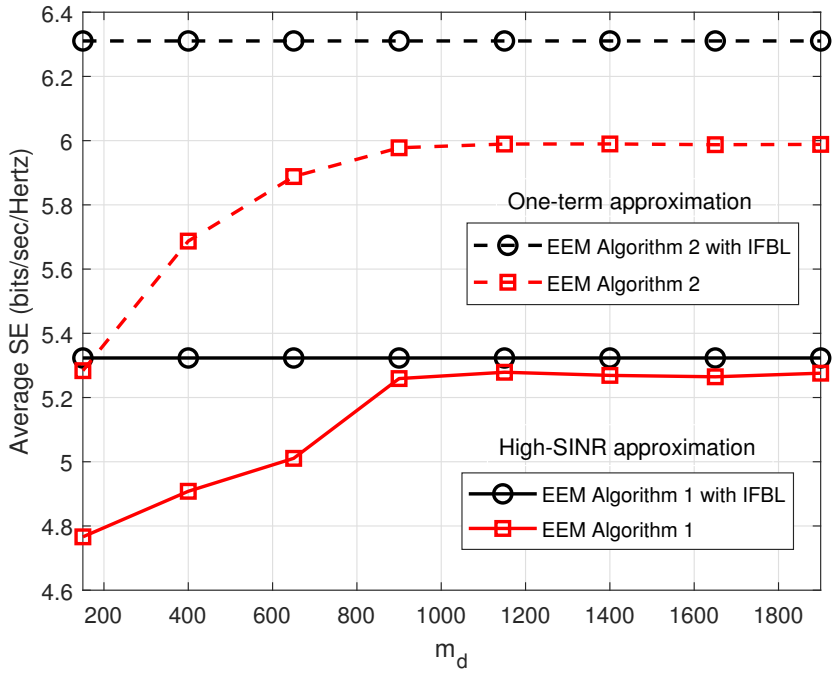


Fig. 6: Impact of the blocklength m_d on the average SE.

It can be seen in Fig. 5 that the average EE of the considered system with Algorithm 1 and Algorithm 2 improves as the blocklength size m_d increases from 150 to 1000 and it begins to saturate after $m_d > 1000$. In addition, we can also notice that Algorithm 2 performs much better than Algorithm 1 when $m_d > 1000$, while the EE performance of the both algorithms becomes closer to that with IFBL.

The impact of the blocklength on the achievable SE performance is shown in Fig. 6 for $\epsilon_{k,max} = 10^{-7}$, $K = 2$, $M = 12$ and $d_k = 500$ meters. Similar to Fig. 5, the average SE performance of the proposed algorithms increases when the blocklength size m_d increases from 150 to 1000 and it becomes approximately constant after $m_d > 1000$. Furthermore, when $m_d > 1000$, the average SE performance gap between the proposed algorithm and the corresponding IFBL becomes very small. Furthermore, the average SE performance of Algorithm 2 is always better than that of the Algorithm 1 because the approximation approach used in Algorithm 2 is much tighter.

D. Impact of the maximum tolerable DEP $\epsilon_{k,max}$

Fig. 7 shows the impact of the DEP constraint on the achievable average EE performance for $m_d = 250$, $K = 2$, $M = 12$ and $d_k = 500$ meters. It can be observed that the average EE of the

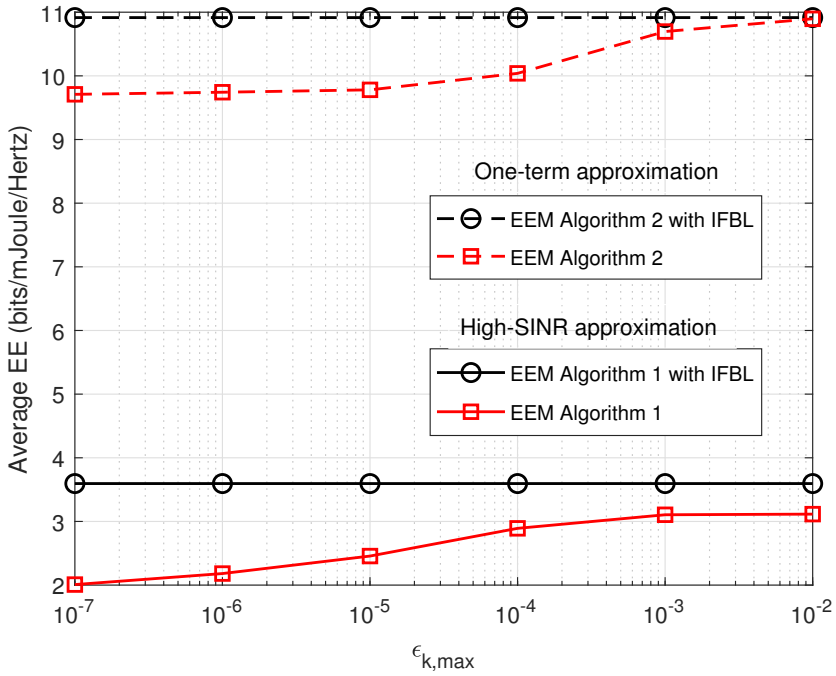


Fig. 7: Impact of the constraint on the DEP on the average EE.

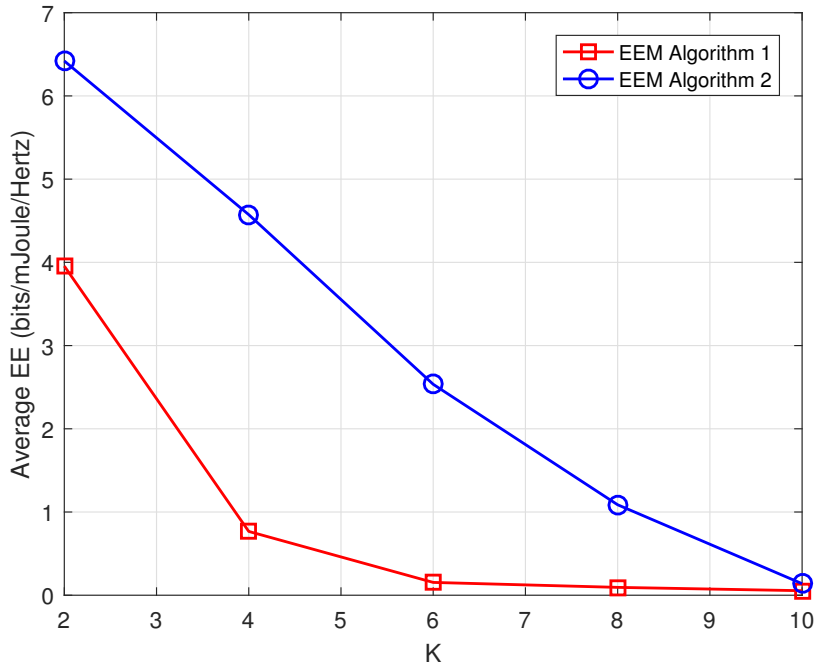


Fig. 8: Impact of the number of URLLC users K on the average EE.

network with Algorithm 1 and Algorithm 2 increases and is closer to the performance of the respective IFBL case as the maximum tolerable value on the DEP increases. The main reason is that when the maximum tolerable value increases in both scenarios, the rate loss in (6) due to the DEP constraint becomes smaller, and thus the EE is improved.

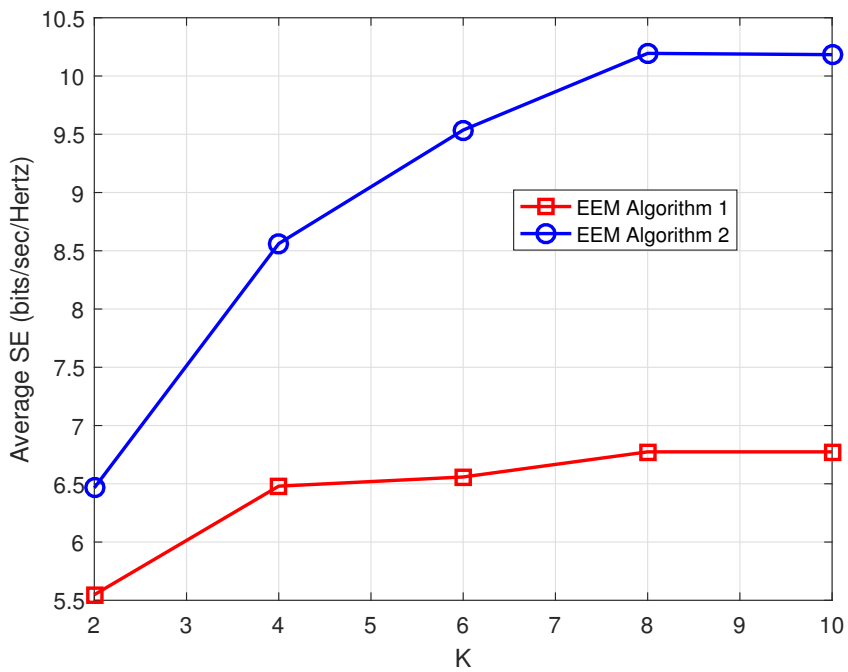


Fig. 9: Impact of the number of URLLC users K on the average SE.

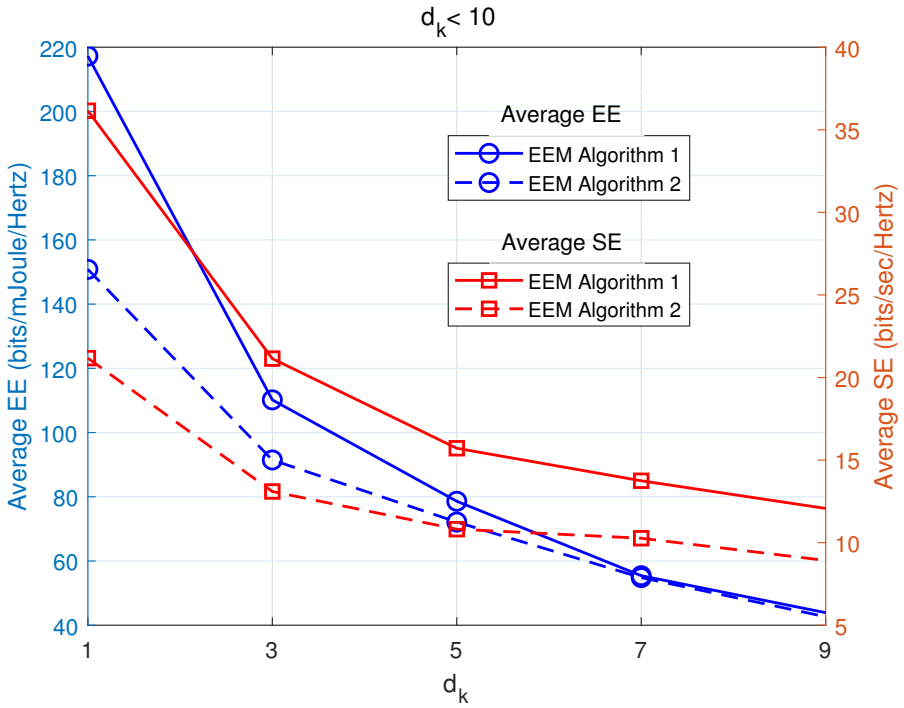
E. Impact of the number of URLLC users K

We shows the impact of increasing the number of URLLC users K on the average EE and SE performance under $\epsilon_{k,max} = 10^{-7}$ and $d_k = 500$ meters. For both algorithms, we set $m_d = 250$ and $M = 20$. In Fig. 8, it can be observed that the average EE for both the algorithms decreases significantly with increasing K and it approaches to zero for both algorithms when $K = 10$. This is due to fact that the overall power consumption increases with K. In addition, Fig. 9 shows the average SE performance for both algorithms enhances when we increase the number of URLLC users K and it becomes constant for $K \geq 8$ due to the penalty factor q. The main reason behind the improvement in the SE performance is multi-user diversity gain achieved by increasing K.

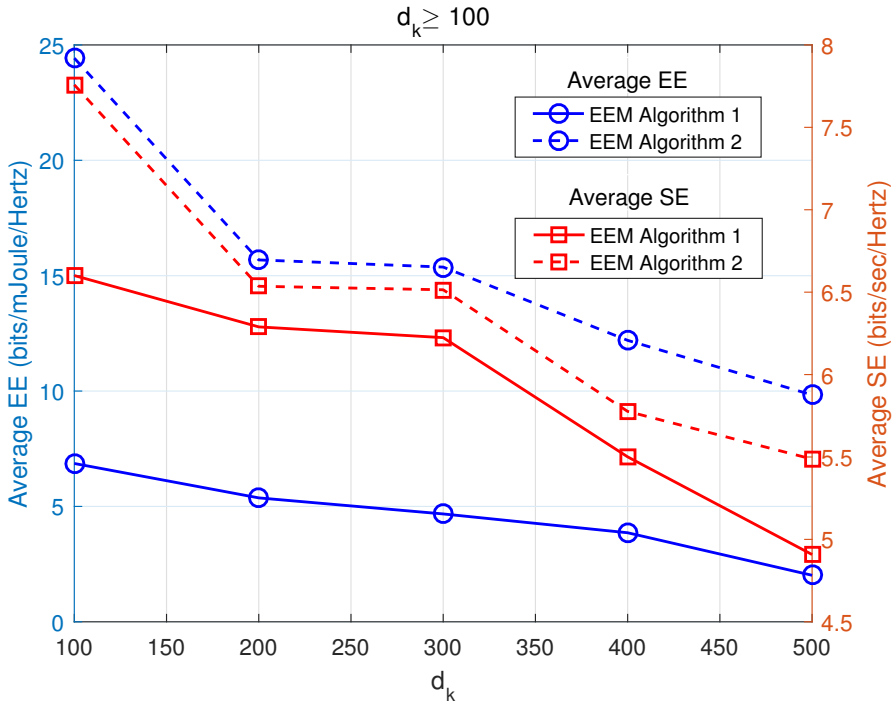
F. Impact of distance d_k

Fig. 10 shows the impact of distance between the BS and the URLLC users on the average EE and SE performance for $\epsilon_{k,max} = 10^{-7}$, $K = 2$ and $M = 12$. In Fig. 10(a), it can be seen that when the URLLC users are very close to the BS i.e., $d_k \leq 10$ meters, the system operates in high-SINR regimes and thus Algorithm 1 performs better than Algorithm 2 in terms

of average EE and SE. Moreover, the performance gap between the two algorithms decreases with increasing d_k.



(a) Average EE and SE versus d_k



(b) Average EE and SE versus d_k

Fig. 10: Impact of distance from the BS to URLLC users on the average EE and SE performance.

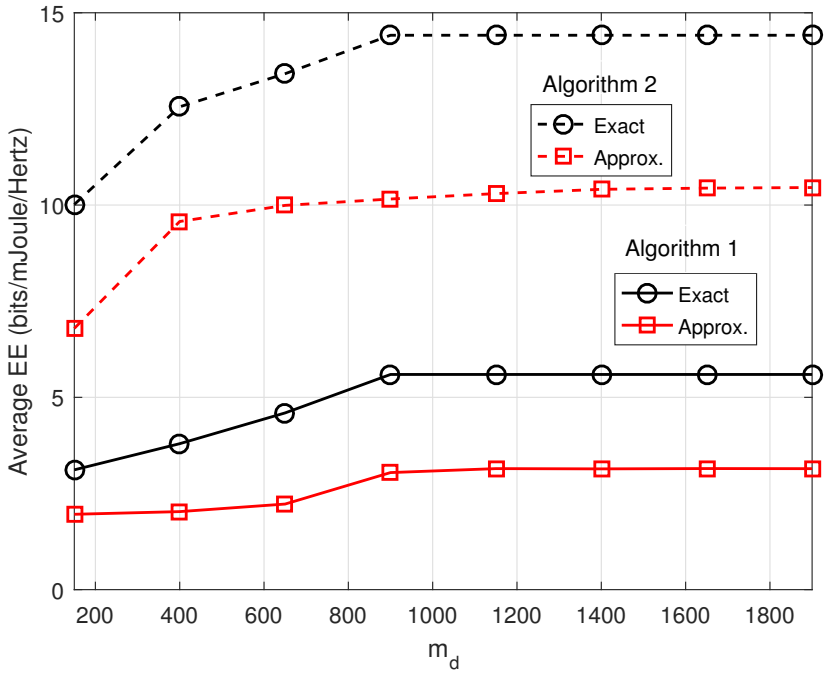


Fig. 11: Average EE performance of the Algorithm 1 and Algorithm 2 with approximate and exact objective functions.

Fig. 10(b) shows the performance of both algorithms when $d_k \geq 100$ meters. Similar to Fig. 10(a), the average EE and SE decreases when we increase the value of d_k. The system switching from high-SINR regime to the low-SINR regime with increasing d_k and thus for $d_k \geq 100$ meters, Algorithm 2 starts performing better than Algorithm 1.

G. Exact versus approximate objective

We plot the optimal values of the objective function in (20) and (37) obtained by using Algorithm 1 and Algorithm 2, corresponding to the *approximate* average EE (curves marked "approx."), while we also plot the *exact* average EE (curves marked "exact") by substituting the optimal solutions obtained by Algorithm 1 and Algorithm 2 into the objective function of (9). In Fig. 11 we demonstrate the average EE performance of Algorithm 1 and Algorithm 2 with the approximate and the exact objective functions. The values of K and M are set as 2 and 12, respectively, while $\epsilon_k = 10^{-7}$ and $d_k = 500$ meters. It can be observed that the approximated objective represents a lower bound on the exact objective, as expected.

H. Impact of blockages

Fig. 12 shows the impact of blockages on the average EE performance of Algorithm 1 and Algorithm 2 with $K = 2$ and $M = 12$, $\epsilon_k = 10^{-7}$ and $d_k = 100$ meters. In order to

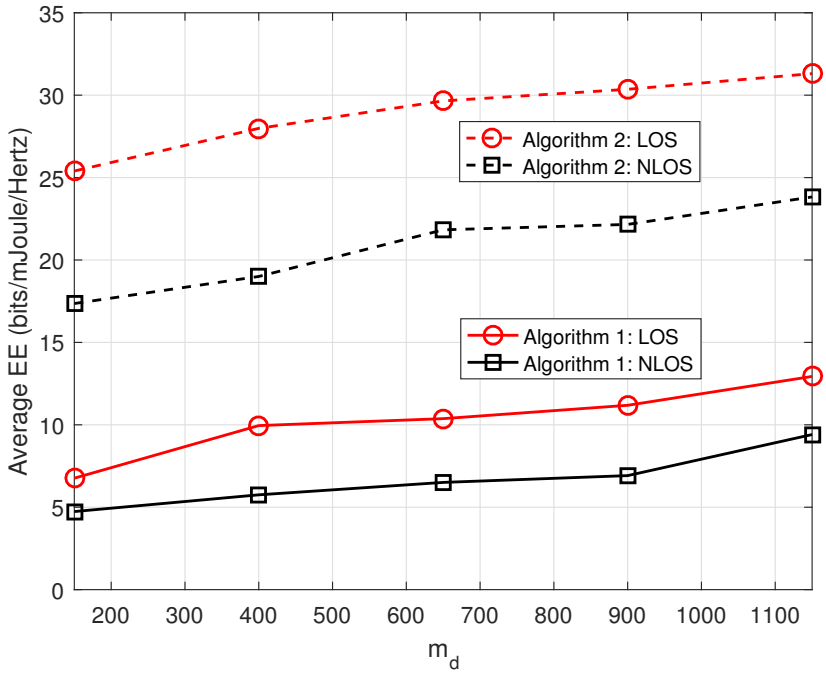


Fig. 12: Impact of blockages on the performance of the Algorithm 1 and Algorithm 2.

show the impact of blockages (non-line-of-sight (NLOS)), we include in the simulation the effect of shadowing, which follows a lognormal distribution with zero mean and 8 dB standard deviation [37]. In Fig. 12, we can observe that the blockages degrade the average EE performance of both algorithms and thus, both algorithms performance better in case of line-of-sight (LOS) propagation.

VI. Conclusions and Future Work

In this paper, we focused on designing the optimal energy-efficient precoder weights and DEP for downlink MISO URLLC networks. For this, we formulated an optimization problem under the constraints of DEP per URLLC user and BS transmit power. Since the formulated problem was non-convex, we analyzed the structure of the problem and derived two approximate versions of the problem, one of which is valid for high SINR and the other is for all practical values of SINR. For fixed values of the auxiliary variables and penalty factor, the approximate problems are jointly convex in the design variables of the precoders and DEPs. Two iterative EEM algorithms were then proposed to obtain locally optimal solutions with a zero-term (high SINR) and a one-term (general SINR) approximation to the channel dispersion. Moreover, we proved the convergence of the algorithms and derived closed-form expressions for the resulting precoders.

By simulation examples, we illustrated the impact of the blocklength, the number of URLLC users, distance between the BS and users and the threshold of DEP on the network's EE and SE performance. Future research directions include extension of the proposed framework to multiple-input and multiple-output (MIMO) multi-hop communications and imperfect CSI knowledge, and the investigation of deep learning approaches for fast varying channel environments.

APPENDIX A: PROOF OF LEMMA 1

We utilize the relation $Q^{-1}(z) = \sqrt{2}\,\mathrm{erfc}^{-1}(2z) = \sqrt{2}\,\mathrm{erf}^{-1}(1-2z)$ to get the approximation of $Q^{-1}(z)$. For $|z| < 1$, the $\mathrm{erf}^{-1}(z)$ is expanded as [31]

$$\mathrm{erf}^{-1}(z) = \sum_{k=0}^{\infty} \frac{c_k}{2k+1} \left(\frac{\sqrt{\pi}}{2} z\right)^{2k+1}, \tag{A.1}$$

where c_k is defined as

$$c_k = \begin{cases} 1 & \text{for } k = 0\,; \\ \sum\limits_{m=0}^{k-1} \dfrac{c_m c_{k-1-m}}{(m+1)(2m+1)} & \text{for } k \geq 1\,. \end{cases} \tag{A.2}$$

Using (A.1), the expansion of $Q^{-1}(z)$ is given by

$$\begin{aligned} Q^{-1}(z) = &\sqrt{\frac{\pi}{2}} \Bigg((1-2z) + \frac{\pi}{12}(1-2z)^3 + \frac{7\pi^2}{480}(1-2z)^5 \\ &+ \frac{127\pi^3}{40320}(1-2z)^7 + \cdots \Bigg), \; 0 < z < 1\,. \end{aligned} \tag{A.3}$$

It can be observed that the values of the second and higher power terms of z in (A.3) are close to zero, and hence we consider only linear terms to approximate $Q^{-1}(z)$ as

$$Q^{-1}(z) \approx \sqrt{\frac{\pi}{2}}\,(A - Bz)\,, \tag{A.4}$$

where A and B are given by (12) and (13), respectively. This completes the proof.

APPENDIX B: PROOF OF LEMMA 2

For any given values of $\mathbf{\Omega}$ and q, the objective function in (20) is concave, while the constraint $(C.1)$ is linear. The constraints $(C.2)$ and $(C.3)$ in (20) are the summation of the non-negative weighted quadratic functions and thus, they are convex. Due to the linear form, the constraints $(C.4)$ and $(C.5)$ are also convex. Therefore the optimization problem (20) is convex in $\mathbf{w}$ and $\boldsymbol{\epsilon}$ for given $\mathbf{\Omega}$ and q, and thus the solution can be obtained by solving this problem and using standard convex optimization tools [30].

APPENDIX C: PROOF OF THEOREM 1

By taking the partial derivative of the Lagrangian dual function defined in (21) with respect to $\mathbf{w}_k^\dagger$ and ϵ_k, we obtain

$$\frac{\partial \mathcal{L}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\lambda_k\}, \mu, \{\nu_k, \upsilon_k, \bar{\upsilon}_k\}\right)}{\partial \mathbf{w}_k^\dagger} = -\, q \times \mathbf{w}_k - \mu \mathbf{w}_k - \sum_{l=1,\, l\neq k}^{K} \frac{\nu_l}{\sigma_l^2} \mathbf{h}_l \mathbf{h}_l^\dagger \mathbf{w}_k + \frac{1}{2} \upsilon_k \mathbf{h}_k - \frac{1}{2j} \bar{\upsilon}_k \mathbf{h}_k \,; \tag{C.1}$$

$$\frac{\partial \mathcal{L}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\lambda_k\}, \mu, \{\nu_k, \upsilon_k, \bar{\upsilon}_k\}\right)}{\partial \epsilon_k} = \sqrt{\frac{\pi}{2m_d}} B \log_2 e - \lambda_k \,. \tag{C.2}$$

By equating the obtained partial derivative to zero, we get

$$\mathbf{w}_k^\star = \left(\frac{1}{2}\upsilon_k - \frac{1}{2j}\bar{\upsilon}_k\right) \times \left((-q+\mu)\,\mathbf{I}_M + \sum_{l=1,\, l\neq k}^{K} \frac{\nu_l}{\sigma_l^2} \mathbf{h}_l \mathbf{h}_l^\dagger\right)^{-1} \mathbf{h}_k \,. \tag{C.3}$$

Using (C.2), the DEP ϵ_k can be updated as

$$\epsilon_k(l+1) = \left[\epsilon_k(l) + \alpha_{\epsilon_k(l)} \left(\sqrt{\frac{\pi}{m_d}} B \log_2 e - \lambda_k\right)\right]^+ . \tag{C.4}$$

Hence, this completes the proof.

APPENDIX D: PROOF OF THEOREM 2

Suppose $(\mathbf{w}^\star, \boldsymbol{\epsilon}^\star, \boldsymbol{\Omega}^\star)$ is the optimal solution at the l-th iteration, we have

$$\begin{aligned} \mathcal{G}\left(q(l)\right) &= \mathcal{F}\left(\mathbf{w}^\star(l), \boldsymbol{\epsilon}^\star(l)\right) - q(l) P_T\left(\mathbf{w}^\star(l)\right); \\ &\geq \mathcal{F}\left(\mathbf{w}^\star(l-1), \boldsymbol{\epsilon}^\star(l-1)\right) - q(l) P_T\left(\mathbf{w}^\star(l-1)\right) = 0\,.. \end{aligned} \tag{D.1}$$

By applying (25) into (D.1), it gives

$$P_T\left(\mathbf{w}^\star(l)\right) \times \left(q(l+1) - q(l)\right) \geq 0\,. \tag{D.2}$$

As we know that $P_T\left(\mathbf{w}^\star(l)\right) \geq 0$, this implies that $q(l+1) \geq q(l)$. Hence, the penalty factor q increases monotonically with the iteration number l until convergence.

APPENDIX E: PROOF OF LEMMA 3

For any fixed values of $\boldsymbol{\Xi}$, $\boldsymbol{\Psi}$ and q, the objective function in (37) is concave, while the constraint $(C.1)$ is linear. The constraints $(C.2)$, $(C.3)$ and $(C.4)$ in (20) are the summation of the non-negative weighted quadratic functions and thus, they are convex. The constraints $(C.5)$

and $(C.6)$ are also convex because of the linear forms. Therefore the optimization problem (37) is convex in $\mathbf{w}$ and $\boldsymbol{\epsilon}$ for given $\boldsymbol{\Xi}$, $\boldsymbol{\Psi}$ and q, and thus the solution can be obtained by solving (37) using standard convex optimization tools [30].

APPENDIX F: PROOF OF THEOREM 3

Taking the partial derivative of (38) with respect to $\mathbf{w}_k^\dagger$ yields

$$\frac{\partial \tilde{\mathcal{L}}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\tilde{\lambda}_k\}, \tilde{\mu}, \{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\}\right)}{\partial \mathbf{w}_k^\dagger} = - q \times \mathbf{w}_k - \tilde{\mu} \times \mathbf{w}_k - \sum_{l=l\neq k}^{K} \varsigma_l \mathbf{h}_l \mathbf{h}_l^\dagger \mathbf{w}_k - \tilde{\nu}_k \mathbf{h}_k \mathbf{h}_k^\dagger \mathbf{w}_k + \frac{1}{2}\rho_k \mathbf{h}_k - \frac{1}{2j}\varrho_k \mathbf{h}_k \,. \tag{F.1}$$

By letting $\frac{\partial \tilde{\mathcal{L}}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\tilde{\lambda}_k\}, \tilde{\mu}, \{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\}\right)}{\partial \mathbf{w}_k^\dagger} = 0$, we get

$$- q \times \mathbf{w}_k - \tilde{\mu} \times \mathbf{w}_k - \sum\nolimits_{l=l\neq k}^{K} \varsigma_l \mathbf{h}_l \mathbf{h}_l^\dagger \mathbf{w}_k - \tilde{\nu}_k \mathbf{h}_k \mathbf{h}_k^\dagger \mathbf{w}_k + \frac{1}{2}\rho_k \mathbf{h}_k - \frac{1}{2j}\varrho_k \mathbf{h}_k = 0 \,; \tag{F.2}$$

$$\Rightarrow \left((q + \tilde{\mu})\, \mathbf{I}_M + \tilde{\nu}_k \mathbf{h}_k \mathbf{h}_k^\dagger + \sum_{l=l\neq k}^{K} \varsigma_l \mathbf{h}_l \mathbf{h}_l^\dagger \right) \mathbf{w}_k = \left(\frac{1}{2}\rho_k - \frac{1}{2j}\varrho_k \right) \mathbf{h}_k \,; \tag{F.3}$$

$$\Rightarrow \mathbf{w}_k^\star = \left(\frac{1}{2}\rho_k - \frac{1}{2j}\varrho_k \right) \times \left((q + \tilde{\mu})\, \mathbf{I}_M + \tilde{\nu}_k \mathbf{h}_k \mathbf{h}_k^\dagger + + \sum_{l=l\neq k}^{K} \varsigma_l \mathbf{h}_l \mathbf{h}_l^\dagger \right)^{-1} \mathbf{h}_k \,, \quad k = 1, \ldots, K \,. \tag{F.4}$$

This completes the proof of the theorem.

APPENDIX G: PROOF OF LEMMA 4

By taking the partial derivative of $\tilde{\mathcal{L}}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\tilde{\lambda}_k\}, \tilde{\mu}, \{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\}\right)$ with respect to ϵ_k, we get

$$\frac{\partial \tilde{\mathcal{L}}\left(\mathbf{w}, \boldsymbol{\epsilon}, \{\tilde{\lambda}_k\}, \tilde{\mu}, \{\varsigma_k, \tilde{\nu}_k, \psi_k, \bar{\psi}_k\}\right)}{\partial \epsilon_k} = \sqrt{\frac{\pi}{2m_d}} \Psi_k B \log_2 e - \lambda_k \,. \tag{G.1}$$

By using (G.1), the update equation for ϵ_k is given by

$$\epsilon_k(l+1) = \left[\epsilon_k(l) + \alpha_{\epsilon_k(l)} \left(\sqrt{\frac{\pi}{2m_d}} \Psi_k B \log_2 e - \lambda_k \right) \right]^+ \,. \tag{G.2}$$

This completes the proof of Lemma 4.

REFERENCES

[1] P. Popovski, Č. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783-5801, Aug. 2019.

[2] G. Ghatak, "Cooperative relaying for URLLC in V2X networks," *IEEE Wireless Commun. Lett.*, in press, 2020.

[3] N. Bonjorn, F. Foukalas, F. Cañellas, and P. Pop, "Cooperative resource allocation and scheduling for 5G eV2X services," *IEEE Access*, vol. 7, pp. 58212-58220, Jan. 2019.

[4] J. Yang *et al.*, "Ultra-reliable communications for industrial internet of things: Design considerations and channel modeling," *IEEE Netw.*, vol. 33, no. 4, pp. 104-111, Jul. 2019.

[5] M. Khoshnevisan, V. Joseph, P. Gupta, F. Meshkati, R. Prakash, and P. Tinnakornsrisuphap, "5G industrial networks with CoMP for URLLC and time sensitive network architecture," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 947-959, Apr. 2019.

[6] 3GPP, "Study on scenarios and requirements for next generation access technologies," *Tech. Specification Group Radio Access Netw.*, 3GPP, Valbonne, France, Tech. Rep. 38.913, Release 14, Jun. 2017.

[7] P. Schulz et al., "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70-78, Feb. 2017.

[8] B. Makki, T. Svensson and M. Zorzi, "Finite block-length analysis of the incremental redundancy HARQ," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 529-532, Oct. 2014.

[9] O. Tervo, L. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574-5588, Oct. 2015.

[10] M. Sheng, L. Wang, X. Wang, Y. Zhang, C. Xu, and J. Li, "Energy efficient beamforming in MISO heterogeneous cellular networks with wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 954-968, Apr. 2016.

[11] C. Pan, W. Xu, J. Wang, H. Ren, W. Zhang, N. Huang, and M. Chen, "Pricing-based distributed energy-efficient beamforming for MISO interference channels," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 710-722, Apr. 2016.

[12] S. Jang, H. Lee, S. Kang, T. Oh, and I. Lee, "Energy efficient SWIPT systems in multi-cell MISO networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8180-8194, Dec. 2018.

[13] K. Singh, A. Gupta, and T. Ratnarajah, "A utility-based joint subcarrier and power allocation for green communications in multi-user two-way regenerative relay networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3705-3722, Sept. 2017.

[14] K. Singh, A. Gupta, and T. Ratnarajah, "QoS-driven energy-efficient resource allocation in multiuser amplify-and-forward relay networks," *IEEE Trans. Sig. Inf. Process. Netw.*, vol. 3, no. 4, pp. 771-786, Dec. 2017.

[15] K. Singh, A. Gupta, T. Ratnarajah, and M.-L. Ku, "A general approach toward green resource allocation in relay-assisted multiuser communication networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 848-862, Feb. 2018.

[16] Y. Polyanskiy, H. V. Poor, and S. Verdu, , "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307-2359, 2010.

[17] W. Yang , G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232-4265, Jul. 2014.

[18] Y. Hu, A. Schmeink and J. Gross, "Blocklength-limited performance of relaying under quasi-static rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4548-4558, Jul. 2016.

[19] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of spectrum sharing networks using rate adaptation," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2823-2835, Aug. 2015.

[20] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimal power control at finite blocklength," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 4598-4615, Sept. 2015.

[21] M. I. Ashraf, C. Liu, M. Bennis, W. Saad, and C. S. Hong, "Dynamic resource allocation for optimized latency and reliability in vehicular networks," *IEEE Access*, vol. 6, pp. 63843-63858, Oct. 2018.

[22] L. Liu and W. Yu, "A D2D-based protocol for ultra-reliable wireless communications for industrial automation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5045-5058, Aug. 2018.

[23] S. Xu, T. Chang, S. Lin, C. Shen and G. Zhu, "Energy-efficient packet scheduling with finite blocklength codes: Convexity analysis and efficient algorithms," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5527-5540, Aug. 2016.

[24] A. Avranas, M. Kountouris, and P. Ciblat, "Energy-latency tradeoff in ultra-reliable low-latency communication with retransmissions," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2475-2485, Nov. 2018.

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

[26] N. A. K. Beigi and M. R. Soleymani, "Ultra-reliable energy-efficient cooperative scheme in asynchronous NOMA with correlated sources," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7849-7863, Oct. 2019.

[27] K. Singh, M.-L. Ku, and M. F. Flanagan, "Energy-efficient precoder design for URLLC-enabled downlink multi-user MISO networks using finite blocklength codes," in *Proc. IEEE VTC Spring*, 2020.

[28] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink MISO OFDMA-URLLC systems," *IEEE Trans. Commun.*, Aug. 2020.

[29] A. Avranas, M. Kountouris, and P. Ciblat, "The influence of CSI in ultra-reliable low-Latency communications with IR-HARQ," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1-6.

[30] S. P. Boyd and L. Vandenberghe, "Convex optimization," *Cambridge Univ. Press*, Cambridge, U.K., 2004.

[31] I. S. Gradshteyn and I. M. Ryzhik, "Table of Integrals, Series, and Products," *Elseview*, 8th ed., Sept. 2014.

[32] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492-498, March 1967.

[33] K. Singh and M. Ku, "Toward green power allocation in relay-assisted multiuser networks: A pricing-based approach," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2470-2486, May 2015.

[34] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439-1451, Aug. 2006.

[35] M.-L. Ku, L.-C. Wang, and Y.-T. Su, "Toward optimal multiuser antenna beamforming for hierarchical cognitive radio systems," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2872-2885, Oct. 2012.

[36] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 127-141, Jan. 2018.

[37] 3GPP TR 36.814, "Evolved universal terrestrial radio access (EUTRA); further advancements for E-UTRA physical layer aspects," Tech. rep. Release 9. 3GPP, 2010.