

Opportunistic Spectrum Sharing for D2D-Based URLLC

Zheng Chu, *Member, IEEE*, Wenjuan Yu, *Member, IEEE*, Pei Xiao, *Senior Member, IEEE*, Fuhui Zhou, *Member, IEEE*, Naofal Al-Dhahir, *Fellow, IEEE*, Atta ul Quddus, and Rahim Tafazolli, *Senior Member, IEEE*

Abstract—A device-to-device (D2D) ultra reliable low latency communications (URLLC) network is investigated in this paper. Specifically, a D2D transmitter opportunistically accesses the radio resource provided by a cellular network and directly transmits short packets to its destination. A novel performance metric is adopted for finite block-length code. We quantify the maximum achievable rate for the D2D network, subject to a probabilistic interference power constraint based on imperfect channel state information (CSI). First, we perform a convexity analysis which reveals that the finite block-length rate for the D2D pair in short-packet transmission is not always concave. To address this issue, we propose two effective resource allocation schemes using the successive convex approximation (SCA)-based iterative algorithm. To gain more insights, we exploit the monotonicity of the average finite block-length rate. By capitalizing on this property, an optimal power control policy is proposed, followed by closed-form expressions and approximations for the optimal average power and the maximum achievable average rate in the finite block-length regime. Numerical results are provided to confirm the effectiveness of the proposed resource allocation schemes and validate the accuracy of the derived theoretical results.

Index Terms—Finite block-length codes, URLLC, average achievable rate, performance analysis, optimal power allocation.

I. INTRODUCTION

Due to the explosive growth of data traffic in wireless services, the evolution of the fifth-generation (5G) networks is necessary to secure higher data rates while guaranteeing a highly reliable transmission [1]. In addition, wireless communication mechanisms are employed in emerging mission-

This work was supported by the U.K. Engineering and Physical Sciences Research Council under Grant EP/R001588/1. The work of W. Yu and A. Quddus was supported by grant N62909-17-1-2114 from the US Office of Naval Research Global. The work of F. Zhou was supported by the Natural Science Foundation of China under Grant 61701214, The Excellent Youth Foundation of Jiangxi Province under Grant 2018ACB21012 and The Open Foundation of The State Key Laboratory of Integrated Services Networks under Grant ISN19-08. The work of N. Al-Dhahir was supported by the Qatar National Research Fund (a member of the Qatar Foundation) under Grant NPRP 8-627-2-260. The authors also would like to acknowledge the support of the University of Surrey 5GIC (<http://www.surrey.ac.uk/5gic>) members for this work. The associate editor coordinating the review of this paper and approving it for publication was Jingon Joung. (Corresponding author: Wenjuan Yu.)

Z. Chu, W. Yu, P. Xiao, A. Quddus, and R. Tafazolli are with the 5G Innovation Centre, Institute for Communication Systems, University of Surrey, Guildford, GU2 7XH, United Kingdom. (Emails: {zheng.chu; w.yu; p.xiao; a.quddus; r.tafazolli}@surrey.ac.uk).

F. Zhou is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210000, P. R. China. F. Zhou is also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xian 710071, China. (Email: zhoufuhui@ieee.org).

N. Al-Dhahir is with the Department of Electrical and Computer Engineering, University of Texas at Dallas, Dallas, TX 75080 USA (Email: aldhahir@utdallas.edu).

critical applications and services (such as remote surgery, unmanned aerial vehicle (UAV) deliveries, smarter transportation, and augmented/virtual reality (AR/VR)), which necessitates ultra low-latency (time delay required for information exchange via wireless networks) transmissions [2]. Thus, the scenario of ultra reliable low latency communications (URLLC) has been considered as one of the major use cases in 5G networks.

Unlike the human-centric 4G long-term evolution (LTE) networks, the URLLC system is required to achieve no less than 99.999% reliability (i.e., 10^{-5} packet error probability) and no longer than 1 ms latency for a small packet size (e.g., 32 bytes) in 5G [3]. *Ultra-high reliability* and *low-latency* are considered to be the most difficult technical challenges in URLLC networks [4], and there is a tradeoff between them. Specifically, the reliability can be improved by introducing an increase in latency, which is due to the fact that some network resources, i.e., signalling, re-transmission, etc., may be employed [5]. On the other hand, the existing wireless networks have been designed to focus on long-packet transmission scenarios to maximize the system throughput/energy efficiency, where high reliability and low latency are generally challenging to be guaranteed simultaneously. This highlights the need for developing novel architectures and transmission schemes to satisfy the URLLC requirements which emphasizes both high-reliability and ultra low-latency.

According to [6]–[10], the main challenge in URLLC applications is to support short packets transmission, which is the generic form of data traffic in mission-critical communications. As a traditional information-theoretical metric, Shannon's capacity has been widely used to characterize the maximal achievable rate. Specifically, ergodic capacity and outage capacity are typically considered for sufficiently large packet lengths. However, these conventional performance metrics are not accurate performance indicators in URLLC networks. This is due to the fact that the block-length of channel coding is short due to the low-latency requirement and small packet size. Under this scenario, the impact of finite block-length and decoding error is more prominent and cannot be ignored. There has been some recent work to support short packet transmissions. In [11], a novel performance metric was exploited for the maximum achievable rate, by taking into account a finite packet length and a given packet error probability. Ultra high-reliability and stringent low-latency requirements can be satisfied by sacrificing the achievable throughput/energy efficiency with a finite transmitted packet size [11]. Recently, finite packet-length transmission has been introduced in some typical scenarios, i.e., downlink broadcast

channel, and uplink random access channel [7]. Specifically, new theories focusing on the design of wireless protocols for finite packet-length transmission have been recently developed which have a direct impact on the system design of URLLC networks [7]. In addition, short packet transmission was applied to the relay-enabled URLLC network in [8], [9], where accurate performance modeling and optimization were discussed. These references highlighted the advantage of applying relaying in 5G URLLC transmissions in comparison to the direct transmission scenario (without relaying). Very recently, a novel access technique, namely, non-orthogonal multiple access (NOMA), was introduced and analyzed to support low-latency communications, by applying the coding rate in the finite block-length regime [12]–[15], to improve the fairness and spectral efficiency.

Device-to-device (D2D) communications have recently attracted increasing research attentions as one of the promising techniques in 5G networks, which was standardized by the 3GPP release 12 [16], [17]. A generic D2D pair consists of one D2D transmitter (DT) and one D2D receiver (DR) in close proximity to communicate directly rather than via a cellular base station (BS). D2D communication provides new mobile service opportunities and potential capability to reduce the traffic load in cellular networks [18]. In addition, D2D communications can be applied to mission-critical applications such as providing public safety and disaster monitoring and relief services due to the low-cost deployment and low-complexity configuration of D2D devices [19]–[22]. Furthermore, D2D communications can improve the spectral utilization via enabling direct communication between the neighbouring mobile devices without the involvement of BS, which is well suited for high data rate, delay constrained and quality of service (QoS)-specific communication [23]. This implies that the latency between two neighbouring devices can be greatly reduced by transmitting with the D2D proximity link while guaranteeing the data rate. In [24], a D2D network was designed to satisfy stringent delay requirements and high reliability, which is suitable for factory automation configurations. The presented results revealed that the introduction of D2D pairs or groups helps to support stringent URLLC requirements. The above-mentioned works only focus on the conventional performance metric for long packets to support stringent URLLC requirements, which is not accurate for short packet transmission. Moreover, there are only few papers in the existing literature which exploit the performance metric of D2D networks in support of URLLC services, which motivates this paper.

In this paper, we study a generic D2D cellular network, in which the D2D transmitter reuses the cellular link to directly send short packets to its destination. Since short packets are essential for low-latency mission-critical communications, we adopt the maximum achievable rate in the finite block-length regime as a suitable performance metric [11]. Our aim in this paper is to maximize the achievable rate of the D2D pair for low-latency communications, subject to a received-power outage constraint at the cellular BS and the imperfect channel state information (CSI) between the D2D user and the cellular BS. The probabilistic received-power constraint implies that the QoS of the cellular network can be guaranteed while only

imperfect channel knowledge may be available in a practical environment, when the D2D users are relatively close to each other and far away from the cellular BS. *We summarize the contributions of this paper as follows:*

- 1) *Optimal Resource Allocation:* The convexity analysis is first presented which reveals that the achievable rate in short-packet transmission is not always concave/convex. This makes the rate-maximization problem for the D2D pair in the finite block-length regime intractable and difficult to solve. In order to circumvent this issue, we resolve the instantaneous achievable rate maximization problem via convex optimization reformulations. Accordingly, the successive convex approximation (SCA) method is adopted to linearize the objective function so that the finite block-length rate-maximization problem can be iteratively solved to determine the optimal resource allocation.
- 2) *Performance Analysis:* To gain more insights, we prove that the average achievable rate in the finite block-length regime is a monotonically increasing function in the high signal-to-noise ratio (SNR) region. By utilizing this property, an optimal power control policy is then proposed, which is achieved at the maximum power limit and maximizes the average achievable rate for the D2D pair for short-packet transmissions. Finally, closed-form expressions and approximations are derived for the optimal average transmit power and the maximum achievable average rate, respectively.

The rest of this paper is organized as follows. The D2D network which supports short packets transmission is first described in Section II, followed by the formulation of the achievable finite block-length rate-maximization problem. The optimal resource allocation and performance analysis are investigated in Section III and Section IV, respectively. Numerical results are presented in Section V, and we conclude this paper in Section VI.

II. SYSTEM MODEL

In this section, we investigate a classic D2D cellular system which consists of a D2D pair, a cellular user (CU) and a cellular BS. In this model, the DT opportunistically accesses and reuses the frequency resources of the cellular network, and reliably transmits mission-critical short packets to the DR. Moreover, the interference power at the cellular BS will be limited to guarantee the QoS in the cellular network. The discrete-time block-fading channel model is considered in this paper, which means that the channel gains remain fixed within one coherence interval and vary independently at the following interval.¹ The complex channel coefficient between the DT and DR is denoted by $h_d[i]$ where i indicates the i -th interval, and $h_c[i]$ represents the channel coefficient between the DT and the cellular BS. Specifically, it is assumed that $h_d[i]$ and $h_c[i]$ are independent and identically distributed (i.i.d.) zero mean circularly symmetric complex Gaussian (ZMCSCG) random

¹Channel coherence time is larger than the frame duration including control signaling and data transmission. Hence, the channel code of finite block-length can be transmitted within one coherence duration [25].

variables, i.e., $\mathcal{CN}(0, 1)$. Furthermore, we assume that the D2D pair has perfect CSI of $h_d[i]$, whereas only imperfect CSI of $h_c[i]$ is available between the DT and the cellular BS. This may happen when the two D2D users are relatively close to each other and far away from the cellular BS. Thus, the actual channel coefficient $h_c[i]$ is given by

$$h_c[i] = \hat{h}_c[i] + \Delta h_c[i], \quad (1)$$

where $\hat{h}_c[i]$ is the minimum mean square error (MMSE) estimate of $h_c[i]$ and $\Delta h_c[i]$ denotes the estimation error. It is assumed that $\Delta h_c[i] \sim \mathcal{CN}(0, \sigma_e^2)$ and the channel estimate $\hat{h}_c[i]$ is ZMCSG distributed with variance $1 - \sigma_e^2$. Note that perfect CSI is obtained when $\sigma_e^2 = 0$.

Furthermore, it is noted that for short-packet transmissions, finite error probability is inevitable [7]. We assume that ϵ denotes the block error probability, i.e., $0 < \epsilon < 1$, m represents the finite block-length, and γ denotes SNR. Then, the instantaneous achievable rate at the DR with finite block-length codes in the i -th time interval is approximately given by [11]

$$\begin{aligned} \hat{R}_i &\approx \log_2(1 + \gamma|h_d[i]|^2) \\ &\quad - \sqrt{\frac{1}{m} \left(1 - \frac{1}{(1 + \gamma|h_d[i]|^2)^2}\right) \frac{Q^{-1}(\epsilon)}{\ln 2}}, \end{aligned} \quad (2)$$

where $Q^{-1}(x)$ is the inverse of the Gaussian Q-function with $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. We model the transmission process as a service process. Note that the transmissions may fail with a probability ϵ . If the transmitted message cannot be decoded correctly, we claim that the effective service rate is zero. Hence, the effective service rate (in bits per m block-length) is given by [26]²

$$\bar{R}_i = \begin{cases} 0, & \text{with probability } \epsilon, \\ m\hat{R}_i, & \text{with probability } 1 - \epsilon. \end{cases} \quad (3)$$

Thus, the achievable normalized average rate at the DR with finite block-length codes is written as

$$R = (1 - \epsilon)\mathbb{E}\left[\frac{\bar{R}_i}{m}\right] = (1 - \epsilon)\mathbb{E}[\hat{R}_i]. \quad (4)$$

Remark 1: In ultra reliability and low latency scenarios, we expect that the users transmit short packets to achieve low latency. Under such a scenario, finite block-length rate becomes a more accurate performance metric for the D2D network. Hence, we adopt finite block-length rate to measure the low-latency feature, which is typical for event-driven packets in URLLC [25]. This shows that we have taken the latency requirement of the D2D network into consideration. On the other hand, ultra high reliability is considered in terms of successful packet decoding probability, which may be as high as $(1 - \epsilon)$ where ϵ is the block error probability. From the definition of finite block-length coding rate given in (2), it can be seen that the impacts of the finite block-length m and the block error probability ϵ have been taken into account. Specifically, when m decreases, a larger penalty is placed on the instantaneous achievable rate. This demonstrates the

²The transmission failure with probability ϵ can be affected by finite block length and the CSI imperfection. When the decoding error occurs, the DR can adopt a simple automatic repeat request (ARQ) mechanism which will transmit a negative acknowledgement (NACK) requesting the re-transmission.

tradeoff between the achievable finite block-length rate and low-latency requirements.

A. Problem Formulation

In this subsection, we formulate the average achievable rate maximization problem for the D2D pair, subject to an outage probability constraint on the received power at the cellular BS and the CSI imperfectness of $h_c[i]$. In this paper, we assume that the interference caused by the DT due to spectrum sharing may exceed the peak limit, but only for a very small percentage of time. First, we make the definitions $g_d = |h_d|^2$, $\hat{g}_c = |h_c|^2$, and $\Delta g_c = |\Delta h_c|^2$.³ The probability density functions (PDFs) of g_d , \hat{g}_c , and Δg_c are respectively given as

$$f(g_d) = e^{-g_d}, \quad f(\hat{g}_c) = \frac{1}{1 - \sigma_e^2} e^{-\frac{\hat{g}_c}{1 - \sigma_e^2}}, \quad f(\Delta g_c) = \frac{1}{\sigma_e^2} e^{-\frac{\Delta g_c}{\sigma_e^2}}. \quad (5)$$

Then, we perform optimal power control over time to maximize the average achievable rate for the D2D pair, which is given as

$$\begin{aligned} R &= \max_{P_d} (1 - \epsilon)\mathbb{E} \left[\log_2 \left(1 + \frac{P_d g_d}{N_0 B} \right) \right. \\ &\quad \left. - \frac{Q^{-1}(\epsilon)}{\ln 2} \sqrt{\frac{1}{m} \left(1 - \frac{1}{\left(1 + \frac{P_d g_d}{N_0 B} \right)^2} \right)} \right], \end{aligned} \quad (6a)$$

$$s.t. \Pr\{P_d(\hat{g}_c + \Delta g_c) \geq I_{th}\} \leq \varrho, \quad (6b)$$

where P_d is the transmit power at the DT, which depends upon the channel condition of the link between D2D pair as well as the link between the DT and the cellular BS. $\mathbb{E}\{\cdot\}$ indicates the expectation over the joint PDF of g_d and g_c . B and N_0 are the channel bandwidth and the single-sided noise power spectral density. Furthermore, I_{th} is the interference power threshold. The constraint (6b) is defined as the interference outage probability constraint imposed at the cellular BS with imperfect CSI to guarantee that this interference does not exceed a threshold, which is probabilistically constrained by a given outage parameter ϱ . Constraint (6b) is the functional requirement for D2D transmission on the cellular frequency band. Problem (6) is a packet delivery mechanism for event-driven packets in URLLC, and the block error probability is employed to measure the reliability of URLLC [25]. For example, the VR/AR system provides high-quality multimedia content which is delivered to the specific devices over the air and enables new business verticals in media and entertainment. This scenario not only requires ultra reliable and low latency transmissions, but also high data rates to guarantee the quality of service (QoS) of the device. In addition, the remote live production is another 5G use case that requires the high-quality signal transmission without sacrificing the ultra reliability and low latency requirements. In order to solve the formulated problem (6), we first analyze the interference constraint (6b)

³Here, the time index i is omitted for simplicity.

under the condition of $\frac{I_{th}}{P_d} \geq \hat{g}_c$, as

$$\begin{aligned} \Pr\{P_d(\hat{g}_c + \Delta g_c) \geq I_{th}\} &\leq \varrho \Rightarrow \Pr\left\{\Delta g_c \geq \frac{I_{th}}{P_d} - \hat{g}_c\right\} \leq \varrho \\ &\Rightarrow \int_0^\infty f(g_d)dg_d \int_0^\infty f(\hat{g}_c)d\hat{g}_c \int_{\frac{I_{th}}{P_d} - \hat{g}_c}^\infty f(\Delta g_c)d\Delta g_c \leq \varrho \\ &\Rightarrow \int_0^\infty e^{-g_d} \int_0^\infty e^{-\frac{I_{th}}{P_d\sigma_e^2} + \frac{\hat{g}_c}{\sigma_e^2}} \frac{1}{1 - \sigma_e^2} e^{-\frac{\hat{g}_c}{1 - \sigma_e^2}} d\hat{g}_c dg_d \leq \varrho. \quad (7) \end{aligned}$$

Apparently, it is challenging to derive a closed-form power allocation which maximizes the average achievable rate while satisfying the interference outage constraint (7). However, one sufficient instantaneous power constraint can be obtained which guarantees that the interference constraint (7) will be always satisfied, which has been derived in [27]. In order to obtain the optimal power allocation satisfying the interference outage constraint (7), we consider the following upper-bound

$$\begin{aligned} \int_0^\infty e^{-g_d} \int_0^\infty e^{-\frac{I_{th}}{P_d\sigma_e^2} + \frac{\hat{g}_c}{\sigma_e^2}} \frac{1}{1 - \sigma_e^2} e^{-\frac{\hat{g}_c}{1 - \sigma_e^2}} d\hat{g}_c dg_d \\ \leq \int_0^\infty e^{-g_d} \int_0^\infty \varrho \frac{1}{1 - \sigma_e^2} e^{-\frac{\hat{g}_c}{1 - \sigma_e^2}} d\hat{g}_c dg_d. \end{aligned}$$

Hence, the following sufficient interference power constraint holds

$$\begin{aligned} e^{-\frac{I_{th}}{P_d\sigma_e^2} + \frac{\hat{g}_c}{\sigma_e^2}} \leq \varrho, \Rightarrow -\frac{I_{th}}{P_d\sigma_e^2} + \frac{\hat{g}_c}{\sigma_e^2} \leq \ln \varrho, \\ \Rightarrow \frac{I_{th}}{P_d\sigma_e^2} \geq \frac{\hat{g}_c}{\sigma_e^2} - \ln \varrho, \Rightarrow P_d \leq \frac{I_{th}}{\hat{g}_c - \sigma_e^2 \ln \varrho}. \quad (8) \end{aligned}$$

If the proposed power allocation satisfies the constraint in (8), then it is guaranteed that the outage constraint (7) is also met [27]. Note that (8) is a sufficient condition of (7). In other words, the solution which satisfies the constraint (8) definitely guarantees the constraint (7). Thus, we substitute (8) into (6b), and problem (6) can be reformulated as

$$\begin{aligned} R = \max_{P_d} (1 - \epsilon) \mathbb{E} \left[\log_2 \left(1 + \frac{P_d g_d}{N_0 B} \right) \right. \\ \left. - \frac{Q^{-1}(\epsilon)}{\ln 2} \sqrt{\frac{1}{m} \left(1 - \frac{1}{\left(1 + \frac{P_d g_d}{N_0 B} \right)^2} \right)} \right], \\ \text{s.t. } P_d \leq \frac{I_{th}}{\hat{g}_c - \sigma_e^2 \ln(\varrho)}. \quad (9) \end{aligned}$$

In order to solve the problem (9), we first analyze the convexity of the instantaneous achievable rate in the following section, and then solve it via iterative SCA.

III. OPTIMAL RESOURCE ALLOCATION

In this section, we exploit the convexity of the instantaneous achievable rate in problem (9) for a given finite block-length and a block error probability. By removing the expectation notation, we focus on the instantaneous achievable rate maximization problem which can be written as

$$\max_{b \geq 0} \hat{R} = \log_2(1 + b) - c \frac{\sqrt{b(b+2)}}{b+1} \quad (10a)$$

$$\text{s.t. } b \leq \frac{I_{th} g_d}{[\hat{g}_c - \sigma_e^2 \ln(\varrho)] N_0 B}, \quad (10b)$$

where $b = \frac{P_d g_d}{N_0 B}$, $c = \frac{Q^{-1}(\epsilon)}{\sqrt{m} \ln 2}$. Note that problem (10) is intractable due to the unknown properties of its objective function (10a). Hence, in the following, we first provide the convexity analysis of the achievable rate in the finite block-length regime, which can help us to design effective and low-complexity resource allocation algorithms.

First, let us define $f_0(b)$ as

$$f_0(b) = \log_2(1 + b) - c \frac{\sqrt{b(b+2)}}{b+1}. \quad (11)$$

Then, the following *lemma* is proposed to characterize the convexity of problem (10).

Lemma 1: The convexity of the achievable rate in the finite block-length regime, i.e., $f_0(b)$, is analyzed as follows:

- 1) $f_0(b)$ is a concave function for $b > f_1^{-1}(c)$ if $c > d$,
- 2) $f_0(b)$ is a concave function for $b > f_2^{-1}(c)$ if $c \leq d$,
- 3) $f_0(b)$ is a convex function for $f_1^{-1}(c) \leq b \leq f_2^{-1}(c)$ if $c \leq d$,

where $d = f_1(b_0) = f_2(b_0)$, and $b_0 = 1.1481$ is the solution of the equation $f_1(b) = f_2(b)$ satisfying $b_0 \geq 0$.

$$f_1(b) = \frac{(1+b) \log_2(1+b)}{\sqrt{b(b+2)}},$$

$$f_2(b) = \frac{(b+1)[b(b+2)]^{\frac{3}{2}}}{(3b^2 + 6b + 1)}.$$

Proof: Please refer to Appendix A. ■

From the analysis in *Lemma 1*, one can see that $f_0(b)$ is not always concave/convex, which means that the optimization problem (10) cannot be directly solved using convex optimization techniques. However, the optimization problem (10) belongs to the class of difference of convex (DC) programming problems, since its objective function (10a) can be written as the difference of two concave functions and the constraint (10b) is linear. Thus, a first-order Taylor approximation will be used to convexify the non-convex objective function to reformulate problem (10) into a convex one [28]. Its basic mathematical manipulation at the approximated value $x^{(n)}$ with the n -th iteration is given by $f(x) \leq f(x^{(n)}) + \nabla f(x^{(n)})(x - x^{(n)})$. The resulting problem can be solved using standard convex optimization solvers iteratively [29]. In the following, we introduce two transformation schemes to linearize the non-convex objective function (10a) such that the original problem can be reformulated into a convex optimization framework and finally tackled via the SCA method or the interior-point algorithm.

A. Scheme 1

In this subsection, we employ the first-order Taylor approximation to convexify the second term of (10a), i.e.,

$$\begin{aligned} c \frac{\sqrt{b(b+2)}}{b+1} &\approx c \frac{\sqrt{b^{(n)}(b^{(n)}+2)}}{b^{(n)}+1} \\ &+ \frac{c}{(1+b^{(n)})^2 \sqrt{b^{(n)}(b^{(n)}+2)}} (b - b^{(n)}), \end{aligned}$$

where $b(n)$ is the updated value of the variable b at the n -th iteration. Hence, problem (10) can be reformulated as

$$\begin{aligned} \max_{b \geq 0} \log_2(1+b) - c \frac{\sqrt{b(n)(b(n)+2)}}{b(n)+1} \\ - \frac{c}{(1+b(n))^2 \sqrt{b(n)(b(n)+2)}} (b-b(n)) \\ \text{s.t. } b \leq \frac{I_{th} g_d}{[\hat{g}_c - \sigma_e^2 \ln(\varrho)] N_0 B}. \end{aligned} \quad (12)$$

After applying the first-order Taylor approximation, the transformed problem (12) at the n -th iteration is a concave problem which can be easily solved. Specifically, for a given $b(n)$, we solve problem (12) by using the interior-point method [29], and then update $b(n+1) = b(n)$ until convergence. The procedure is summarized in **Algorithm 1**. We note that the obtained optimal solution at each iteration is guaranteed to be a feasible solution for the next iteration. Thus, the iteratively updated finite block-length achievable rate obtained in the $(n+1)$ -th iteration is larger than or equal to that in the n -th iteration. In other words, **Algorithm 1** generates a non-decreasing sequence of objective values. Moreover, problem (12) is bounded due to its interference power constraint. Hence, **Algorithm 1** converges to a locally optimum solution [30], [31].

Algorithm 1: Algorithm to solve problem (10)

- 1) **Initialization:** $b(0)$ is a feasible solution for problem (10).
 - 2) Solve problem (12) with a given $b(n)$ to obtain b at n -th iteration.
 - 3) Update $b(n) = b$.
 - 4) $n := n + 1$.
 - 5) **Until convergence**
-

B. Scheme 2

In this subsection, we propose an approximation scheme to solve problem (10) in the high interference threshold region. Particularly, we apply $\sqrt{1-a} \approx 1 - \frac{1}{2}a$, for $|a| < 1$, to approximate the objective function in problem (10) and reformulate the problem as follows

$$\begin{aligned} \max_{b \geq 0} \hat{R} \approx \log_2(1+b) + \frac{c}{2(1+b)^2} - c \\ \text{s.t. } b \leq \frac{I_{th} g_d}{[\hat{g}_c - \sigma_e^2 \ln(\varrho)] N_0 B}. \end{aligned} \quad (13)$$

After applying the approximation step, problem (13) is still non-concave because of its objective function. To tackle this problem, we consider the first-order Taylor approximation to reformulate the problem, which is given by

$$\begin{aligned} \max_{b \geq 0} \log_2(1+b) + \frac{c}{2(1+b(n))^2} - \frac{c}{(1+b(n))^3} (b-b(n)) \\ \text{s.t. } b \leq \frac{I_{th} g_d}{[\hat{g}_c - \sigma_e^2 \ln(\varrho)] N_0 B}. \end{aligned} \quad (14)$$

The reformulated problem (14) at the n -th iteration can be easily proved to be a concave problem. Hence, it can be iteratively solved using the procedure similar to **Algorithm 1**, which is omitted here. These two proposed schemes are solved iteratively by employing **Algorithm 1** and the standard

interior-point methods, thus, it follows that the computational complexity of the SCA approach is on the order of $\mathcal{O}(N)$, where N is the maximum iteration time [32, Lecture 6].

IV. PERFORMANCE ANALYSIS

In the previous section, we solved the instantaneous achievable rate maximization problem (10) approximately via the iterative SCA method. To further investigate problem (9), in this section we first analyze the monotonicity of the average achievable finite block-length rate in the high SNR region. Then, based on the theoretical analysis, we propose an optimal power control at high SNRs, followed by the closed-form expressions and approximations for the average optimal transmit power and the average achievable rate.

First, we analyze the monotonicity of the objective function in problem (9). Note that the average achievable rate R at the DR with finite block-length codes is rewritten as $R = (1 - \epsilon) \mathbb{E}\{\hat{R}_i\}$, where the term $\mathbb{E}\{\hat{R}_i\}$ is simplified and expanded as follows:

$$\hat{R} = \mathbb{E} \left[\log_2(1 + \rho g_d) - \frac{Q^{-1}(\epsilon)}{\ln 2} \sqrt{\frac{1}{m} \left(1 - \frac{1}{(1 + \rho g_d)^2} \right)} \right], \quad (15)$$

where $\rho = \frac{P_d}{N_0 B}$ is the transmit SNR. The following is required to characterize the monotonicity of (15).

Lemma 2: For the AWGN channel, its achievable rate monotonically increases with the transmit SNR $\bar{\rho}$, under the following assumptions, $m \geq 10^2$, $\epsilon \geq 10^{-6}$, and $\bar{\rho} \geq -10$ dB⁴.

Proof: Please refer to Appendix B. ■

Note that *Lemma 2* proves the monotonicity for the AWGN channel without the consideration of fading. For a fading channel, in the following we will argue that the average finite block-length rate \hat{R} is a monotonically increasing function when the transmit SNR is large. Firstly, the average finite block-length rate \hat{R} can be approximated at high SNRs as follows:

$$\hat{R}(\rho) \approx E[\log(1 + \rho g_d) - \frac{Q^{-1}(\epsilon)}{\ln(2)} \sqrt{1/m}]. \quad (16)$$

By taking the first-derivative of (16), it can be easily verified that the average finite block-length rate \hat{R} is a monotonically increasing function with ρ in the high SNR region. Since the transmit power P_d is equal to ρ multiplied by the constant noise power, we conclude that the achievable finite block-length rate \hat{R} monotonically increases with P_d , for sufficiently large values of ρ . To relax the condition of high SNR, we plot Fig. 1 which shows that although the monotonicity is proved for the high SNR region, it holds true for SNR level as small as -6 dB. Based on the above analysis and Fig. 1, the optimal power allocation P_d^* should be achieved at the maximum power limit such that the maximum value of the average finite block-length rate is achieved. Thus, we have

$$P_d^* = \frac{I_{th}}{\hat{g}_c - \sigma_e^2 \ln(\varrho)}. \quad (17)$$

By substituting (17) into the objective function in (9), we

⁴The assumptions here are reasonable for URLLC applications.

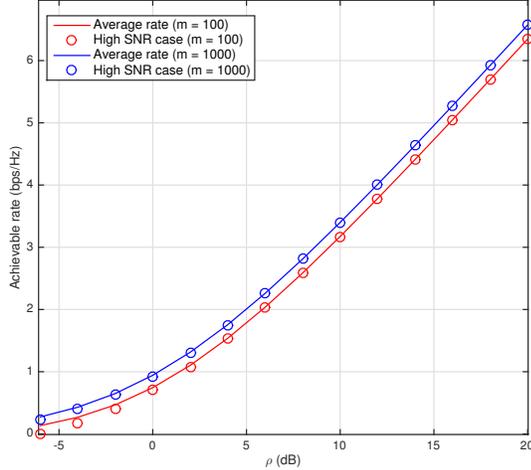


Fig. 1: Comparison between high SNR finite block-length rate and achievable finite block-length rate.

have

$$R = (1 - \epsilon) \mathbb{E} \left\{ \log_2 \left(1 + \frac{I_{th} g_d}{N_0 B (\hat{g}_c - \sigma_e \ln(\rho))} \right) - \sqrt{\frac{1}{m}} \frac{Q^{-1}(\epsilon)}{\ln 2} \sqrt{1 - \frac{1}{\left(1 + \frac{I_{th} g_d}{N_0 B (\hat{g}_c - \sigma_e \ln(\rho))} \right)^2}} \right\}, \quad (18)$$

and the optimal average transmit power is given by

$$\bar{P}_d^* = \mathbb{E}[P_d^*] = \mathbb{E} \left[\frac{I_{th}}{\hat{g}_c - \sigma_e^2 \ln(\rho)} \right]. \quad (19)$$

To proceed, in the following *theorem*, we derive the average achievable rate and the optimal average power allocation in closed-form.

Theorem 1: A closed-form approximation can be derived for the maximum average achievable rate, i.e., R in (18), which is given in (20) on the top of the next page, where $\eta_1 = \frac{N_0 B}{I_{th}}$, $\eta_2 = \frac{N_0 B \sigma_e^2 \ln(\rho)}{I_{th}}$, $e_n(x) = \sum_{k=0}^n \frac{x^k}{k!}$, $\beta = \sqrt{\frac{1}{m}} Q^{-1}(\epsilon) \log_2(\epsilon)$, and $E_1(\cdot)$ is the exponential integral function [33]. In addition, the optimal average transmit power \bar{P}_d^* for the DR can be derived in closed-form

$$\bar{P}_d^* = \frac{I_{th}}{1 - \sigma_e^2} e^{-\frac{\sigma_e^2 \ln(\rho)}{1 - \sigma_e^2}} E_1 \left[-\frac{\sigma_e^2 \ln(\rho)}{1 - \sigma_e^2} \right]. \quad (21)$$

Proof: Please refer to Appendix C. ■

Theorem 1 presents the closed-form expression for the maximum average achievable rate and the optimal average power of the D2D pair. Then, we provide the asymptotic analysis for the maximum achievable average rate at high SNRs. Firstly, it can be easily shown that in the high SNR regime,

$$\left[1 - \frac{1}{\left(1 + \frac{I_{th} g_d}{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))} \right)^2} \right] \rightarrow 1. \quad (22)$$

Accordingly, the maximum achievable average rate obtained at

TABLE I: Notations

Parameters	Notation	Typical Values
Channel estimation error variance	σ_e^2	10^{-2}
Interference outage probability	ϱ	10^{-2}
Block error probability	ϵ	10^{-2}
Interference threshold	I_{th}	10 dBW
Block-length	m	500

high SNRs, denoted by R_h , can be approximated as follows:

$$R_h \approx (1 - \epsilon) \left\{ \mathbb{E} \left[\log_2 \left(1 + \frac{I_{th} g_d}{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))} \right) \right] - \beta \right\}. \quad (23)$$

Following similar proof steps as in *Theorem 1*, a closed-form expression can be derived for (23), which is omitted here. The gap between them depends on the values of the error probability and the block-length. This phenomenon will be clearly shown in Section V, which confirms the above analysis and indicates that the gap between the achievable finite block-length rate and ergodic capacity can be reduced by carefully selecting the parameter values.

Remark 2: In the high SNR region, the average finite block-length rate is expressed as the difference between the ergodic capacity and the penalty term, i.e., β in (23), due to the fact that the term $\mathbb{E} \left\{ \beta \sqrt{1 - \frac{1}{(1 + \rho g_d)^2}} \right\}$ will approach β as the SNR ρ increases. In addition, Fig. 1 demonstrates that the average finite block-length rate versus SNR ρ , and as indicated by the figure, the average rate in the high SNR case can achieve the same performance when $\rho \geq 0$ dB at a given block error probability ϵ . Moreover, the optimal power allocation can improve the average finite block-length rate. This is because the transmit power is optimized to satisfy the upper bound of the interference power constraint. Thus, the average finite block-length rate can be maximized with the optimal power allocation in the high SNR region. The block length m can improve the average rate, which represents the low latency feature.

V. NUMERICAL RESULTS

In this section, numerical results are presented to validate the proposed iterative SCA schemes and the accuracy of the theoretical analysis will be numerically confirmed by comparing with Monte Carlo simulations. The summary of the simulation parameter configuration is presented in Table I unless otherwise specified.

First, we evaluate the proposed SCA-based resource allocation schemes given in Section III. Fig. 2 shows the convergence of **Algorithm 1** for Scheme 1 and Scheme 2. It is observed that the proposed resource allocation schemes can converge to the same optimal value of the finite block-length achievable rate. This confirms the accuracy of the approximation $\sqrt{1-x} \approx 1 - \frac{1}{2}x$ and the effectiveness of these two schemes. In addition, the interference threshold I_{th} and the block-length value m can produce a positive effect to the finite block-length achievable rate, while the larger channel estimation errors σ_e^2 can lead to a degradation of the finite block-length achievable rate.

In Fig. 3, it is observed that the finite block-length achievable rate increases with the interference threshold I_{th} with

$$\begin{aligned}
R \approx (1 - \epsilon) & \left\{ \frac{I_{th}}{I_{th} \ln 2 - N_0 B \ln 2(1 - \sigma_e^2)} \left[-e^{-\frac{\sigma_e^2 \ln(\varrho)}{1 - \sigma_e^2}} E_1 \left[-\frac{\sigma_e^2 \ln(\varrho)}{1 - \sigma_e^2} \right] + e^{-\eta_2} E_1[-\eta_2] \right] \right. \\
& - \beta \left[1 - \frac{\eta_1}{2}(1 - \sigma_e^2) + \frac{\eta_2}{2} - \frac{e^{-\frac{\eta_2}{(1 - \sigma_e^2)\eta_1}}}{\eta_1(1 - \sigma_e^2) \left(\frac{1}{(1 - \sigma_e^2)\eta_1} - 1 \right)^3} \left(E_1 \left(-\frac{\eta_2}{(1 - \sigma_e^2)\eta_1} \right) \right. \right. \\
& \left. \left. - e_2 \left[-\eta_2 \left(\frac{1}{(1 - \sigma_e^2)} - 1 \right) \right] e^{\eta_2 \left(\frac{1}{(1 - \sigma_e^2)} - 1 \right)} E_1(-\eta_2) + e^{\frac{\eta_2}{(1 - \sigma_e^2)\eta_1}} \sum_{k=1}^2 \frac{e^{k-1} \left(-\frac{\eta_2}{\eta_1(1 - \sigma_e^2)} \right)}{k \left(1 + \frac{1}{\eta_1(1 - \sigma_e^2)} \right)^k} \right) \right] \right\} \quad (20)
\end{aligned}$$

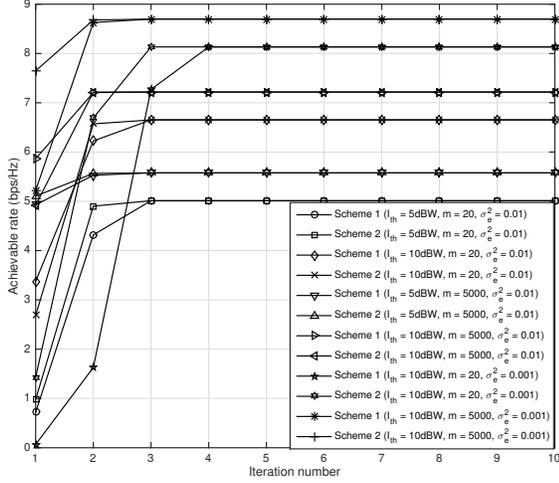


Fig. 2: Convergence of **Algorithm 1**.

$m = 20$ or $m = 6000$ for the URLLC context. Both proposed SCA-based iterative schemes converge to the same performance in terms of the finite block-length achievable rate, which validates the proposed **Algorithm 1**. Moreover, in order to validate the proposed resource allocation schemes, we compare our proposed schemes with the optimal solutions obtained from Monte Carlo simulations. It can be observed that the Monte Carlo simulation has a slightly better performance than the proposed SCA-based iterative schemes in terms of the finite block-length achievable rate. This is owing to the fact that Monte Carlo simulations can guarantee a globally optimal solution for the average achievable rate maximization problem, whereas our proposed schemes consider the instantaneous achievable rate maximization with some approximations, leading to sub-optimal solutions. However, the performance difference between the proposed schemes and the Monte Carlo simulations is not significant, which confirms the accuracy and effectiveness of the proposed schemes.

Fig. 4 plots the instantaneous achievable rate versus the finite block-length m , obtained using the proposed resource allocation schemes in Section III. As one can see from this figure, both proposed schemes first monotonically increase with the block-length m and become flat in the large block-length region. Moreover, both of them have the same performance in

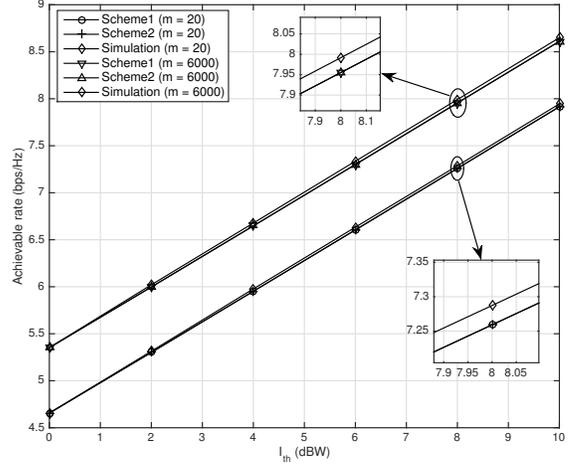


Fig. 3: The achievable rate versus the interference power threshold I_{th} .

terms of the achievable finite block-length rate, which validates the proposed approximation schemes. In addition, a larger interference threshold increases the achievable finite block-length rate, while a larger block error probability leads to a smaller achievable finite block-length rate. Interestingly, it can be observed that in Fig. 6, similar trends and arguments hold for the average achievable rate performance as in the closed-form expressions derived in Section IV. Further details will be given in the following discussions.

Next, we evaluate the average achievable rate performance to confirm the accuracy of the proposed closed-form expressions given in Section IV. Fig. 5 shows the finite block-length rate versus the interference threshold I_{th} . It is observed from this figure that the average achievable rate versus the interference power threshold I_{th} , with two different block-length values, namely, $m = 20$ and $m = 2000$. From this figure, it is clear that the analytical result provided by the closed-form expression of the average achievable rate R matches with the Monte Carlo simulation results for a relatively large interference power threshold, i.e., $I_{th} \geq 0$ dBW. In addition, there is a negligible difference between the theoretical and numerical results in the low interference tolerance region because of the slight inaccuracy of the approximation step, i.e., $\sqrt{1-x} \approx 1 - \frac{1}{2}x$. In addition, the average achievable rate

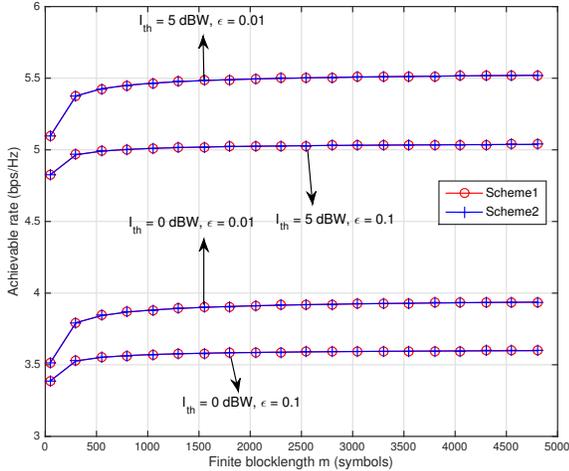


Fig. 4: The achievable rate versus block-length m .

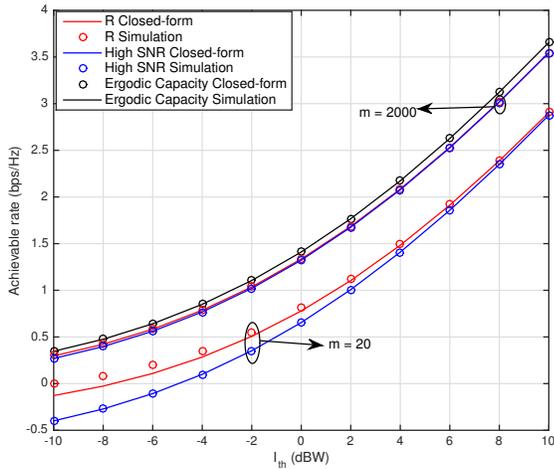


Fig. 5: The average finite block rate versus the interference power threshold I_{th} .

at high SNRs, i.e., R_h , approaches R in the high interference power region. This is due to the fact that the average finite block-length rate has a slightly increasing trend with a larger block length m and the achievable finite block-length rate. Specifically, it depends on the term

$$\mathbb{E} \left\{ \sqrt{\frac{1}{m} \frac{Q^{-1}(\epsilon)}{\ln 2}} \sqrt{1 - \frac{1}{\left(1 + \frac{I_{th} g_d}{N_0 B(\hat{g}_c - \sigma_e^2 \ln(\rho))}\right)^2}} \right\},$$

which approaches $\mathbb{E} \left\{ \sqrt{\frac{1}{m} \frac{Q^{-1}(\epsilon)}{\ln 2}} \right\}$ with a larger m and I_{th} such that R_h achieves a performance very close to the average finite block-length rate R in the high interference threshold regime.

Then, we evaluate the average achievable rate versus the block-length m in Fig. 6. It can be observed that the average achievable rate R first monotonically increases with the block-length m and then stays flat in the large block-length regime. This is due to the fact that the achievable average rate with fi-

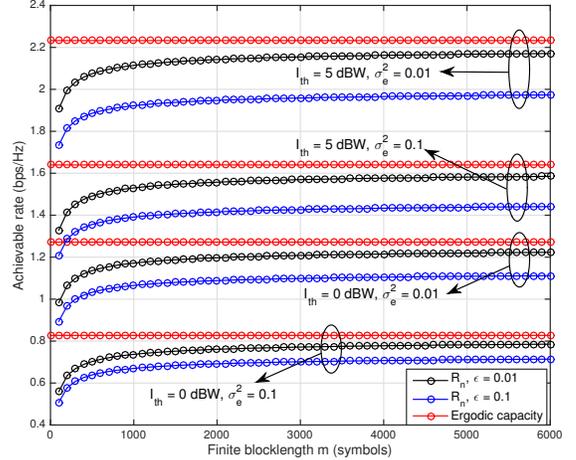


Fig. 6: The average achievable rate versus block-length m .

nite block-length codes is basically an expression calculated by incurring a penalty on ergodic capacity, in which the penalty is inversely proportional to m . Hence, when m becomes larger, the penalty added to the achievable rate in the finite block-length regime gradually reduces. Meanwhile, Fig. 6 shows that the traditional ergodic capacity is flat with the block-length m . This is due to the fact that the ergodic capacity is defined for arbitrarily small packet error probability and sufficiently large packet lengths, which implies that it is independent of the finite block-length m . In addition, the interference tolerance has a positive effect on the average achievable rate, which is easy to understand since a larger interference tolerance power results in a larger transmit power limit and thus contributes to a higher average achievable rate. Moreover, Fig. 6 also indicates that a larger channel estimation error variance σ_e^2 and a larger block error probability ϵ will lead to a smaller average achievable rate. This indicates that accurate channel estimation can greatly benefit the achievable rate performance in the finite block-length regime.

Fig. 7 shows the performance of the optimal average power versus the interference outage probability ρ . From this figure, it can be observed that the derived closed-form expression of the average optimal power \bar{P}_d^* matches with the Monte Carlo simulation results, which confirms the accuracy of our derivations. Furthermore, Fig. 7 shows that a larger value of ρ leads to a higher average power allocation which is required to support the D2D transmission. In this case, the interference power at the cellular BS may exceed the interference threshold I_{th} with a higher probability. In addition, we can observe that a larger interference tolerance leads to a larger average optimal power, whereas the channel error estimation variance has an adverse effect on the average optimal power allocation.

VI. CONCLUSION

In this paper, we investigated opportunistic spectrum sharing for a D2D based URLLC system. A novel performance metric for finite block-length codes was adopted to formulate an achievable rate maximization problem subject to

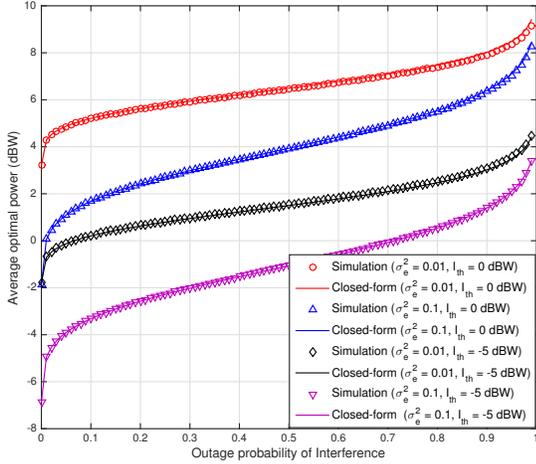


Fig. 7: The average optimal power allocation versus interference outage probability threshold ρ with finite block-length value $m = 500$.

the outage probability constraint of the interference power and the imperfect CSI. To facilitate the design of effective resource allocation algorithms, convexity analysis was first presented which showed that the objective function is not always concave/convex. Then, the original problem was transformed using some approximations and then solved by applying the iterative SCA-based schemes. In addition, we analyzed the monotonicity property of the average rate in the finite block-length regime for the AWGN channel, and we also argue that the property is true for the fading channel. Based on this property, an optimal power control policy was proposed, followed by derivations of closed-form expressions and approximations for the optimal average transmit power and the maximum achievable average rate. Numerical results confirmed the effectiveness of the proposed resource allocation schemes and validated the accuracy of the theoretical closed-form expressions.

APPENDIX

A. Proof of Lemma 1

First, it is required that the finite block-length achievable rate should be non-negative, i.e., $f_0(b) \geq 0$ such that the following inequality can be obtained

$$c \leq f_1(b) = \frac{(1+b) \log_2(1+b)}{\sqrt{b(b+2)}}. \quad (24)$$

Next, we find the second-order derivative to check the convexity of $f_0(b)$, which yields

$$\frac{\partial^2 f_0(b)}{\partial b^2} = -\frac{1}{(1+b)^2} + \frac{c[2b^2 + 2b + (1+b)^2]}{(1+b)^3[b(b+2)]^{\frac{3}{2}}}, \quad (25)$$

$$= \frac{c(3b^2 + 6b + 1) - (b+1)[b(b+2)]^{\frac{3}{2}}}{(b+1)^3[b(b+2)]^{\frac{3}{2}}}. \quad (26)$$

If $f_0(b)$ is concave, it means that $\frac{\partial^2 f_0(b)}{\partial b^2} \leq 0$. In other words, $c(3b^2 + 6b + 1) - (b+1)[b(b+2)]^{\frac{3}{2}} \leq 0$, which can be written as

$$c \leq f_2(b) = \frac{(b+1)[b(b+2)]^{\frac{3}{2}}}{(3b^2 + 6b + 1)}. \quad (27)$$

From (24) and (27), we notice that both $f_1(b)$ and $f_2(b)$ are monotonically increasing functions in terms of $b \geq 0$. Also, it can be easily verified that $f_1(b) \geq f_2(b)$ for $0 \leq b \leq b_0$ and $f_1(b) \leq f_2(b)$ for $b \geq b_0$, where $b_0 = 1.1481$ satisfying $f_1(b) = f_2(b)$. Hence, by denoting $d = f_1(b_0) = f_2(b_0)$, the concavity of $f_0(b)$ can be analyzed as follows:

- If $c > d$, i.e., $c > f_1(b_0) = f_2(b_0)$, this means that $b_0 \leq f_2^{-1}(c) \leq f_1^{-1}(c)$. From (24), we note that $b \geq f_1^{-1}(c)$ is required to guarantee a non-negative finite block-length achievable rate. Therefore this condition implies that $b \geq f_1^{-1}(c) \geq f_2^{-1}(c)$, which satisfies (27) and confirms that $f_0(b)$ is concave.
- If $c \leq d$, i.e., $c \leq f_1(b_0) = f_2(b_0)$, this means that $b_0 \geq f_2^{-1}(c) \geq f_1^{-1}(c)$. From (24), we note that $b \geq f_1^{-1}(c)$ is required to guarantee the feasibility of the calculated finite block-length rate. Therefore there are two possible cases: 1) $b \geq f_2^{-1}(c) \geq f_1^{-1}(c)$, 2) $f_2^{-1}(c) \geq b \geq f_1^{-1}(c)$. If b belongs to case 1), then it satisfies (27) and $f_0(b)$ is concave. On the other hand, if b belongs to case 2), then $c \geq f_2(b)$, $\frac{\partial^2 f_0(b)}{\partial b^2} \geq 0$ and hence $f_0(b)$ is convex.

B. Proof of Lemma 2

Let us focus on the AWGN channel with a transmit SNR $\bar{\rho}$, and we take the first derivative of the instantaneous achievable rate \hat{R} , given by

$$\frac{\partial \hat{R}}{\partial \bar{\rho}} = \frac{1}{(1+\bar{\rho}) \ln 2} \left[1 - \sqrt{\frac{1}{m}} Q^{-1}(\epsilon) \left(1 - \frac{1}{(1+\bar{\rho})^2} \right)^{-\frac{1}{2}} \frac{1}{(1+\bar{\rho})^2} \right]. \quad (28)$$

The remaining part is to show that (28) is greater than zero under some practical assumptions. Specifically, for $m \geq 10^2$ and $\epsilon \geq 10^{-6}$, we have $m^{-\frac{1}{2}} \leq 10^{-1}$ and $Q^{-1}(\epsilon) \leq Q^{-1}(10^{-6})$.⁵ In addition, we define $g(\bar{\rho}) = \sqrt{\frac{1}{m}} Q^{-1}(\epsilon) \left(1 - \frac{1}{(1+\bar{\rho})^2} \right)^{-\frac{1}{2}} \frac{1}{(1+\bar{\rho})^2}$, which is easily shown to be a monotonically decreasing function with $\bar{\rho}$ and $g(\bar{\rho}) < 1$ when $\bar{\rho}$ is not extremely low, e.g., $\bar{\rho} \geq -10$ dB. We completed the proof of Lemma 2.

C. Proof of Theorem 1

First, we rewrite the average achievable rate R as

$$R = (1-\epsilon) \left[\underbrace{\mathbb{E} \left\{ \log_2 \left(1 + \frac{I_{th} g d}{N_0 B (\hat{g}_c - \sigma_e \ln(\varrho))} \right) \right\}}_C \right] - \beta \underbrace{\mathbb{E} \left\{ \sqrt{1 - \frac{1}{\left(1 + \frac{I_{th} g d}{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\varrho))} \right)^2}} \right\}}_D, \quad (29)$$

⁵ $Q^{-1}(x)$ monotonically decreases in terms of x .

where $\beta = \sqrt{\frac{1}{m}} Q^{-1}(\epsilon) \log_2(\epsilon)$. Next, we derive the closed-form expression for the first term C , which is given by

$$\begin{aligned} C &= \int_0^\infty \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} \int_0^\infty \frac{1}{\ln 2} \ln \left(1 + \frac{I_{th} g_d}{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))} \right) e^{-g_d} d g_d d \hat{g}_c \\ &= \frac{1}{\ln 2} \int_0^\infty \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} e^{\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}}} E_1 \left(\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}} \right) d \hat{g}_c. \end{aligned} \quad (30)$$

Let us denote $x = \eta_1 \hat{g}_c - \eta_2$, where $\eta_1 = \frac{N_0 B}{I_{th}}$ and $\eta_2 = \frac{N_0 B \sigma_e^2 \ln(\rho)}{I_{th}}$. Hence, we have

$$\begin{aligned} C &= \frac{e^{-\frac{\eta_2}{(1-\sigma_e^2)\eta_1}}}{\eta_1(1-\sigma_e^2) \ln 2} \int_{-\eta_2}^\infty e^{\left(1 - \frac{1}{(1-\sigma_e^2)\eta_1}\right)x} E_1(x) dx \\ &= \frac{I_{th}}{[I_{th} - N_0 B(1-\sigma_e^2)] \ln 2} \left[-e^{-\frac{\sigma_e^2 \ln(\rho)}{1-\sigma_e^2}} E_1 \left[-\frac{\sigma_e^2 \ln(\rho)}{1-\sigma_e^2} \right] \right. \\ &\quad \left. + e^{-\eta_2} E_1[-\eta_2] \right]. \end{aligned} \quad (31)$$

Then, we focus on the derivation of closed-form expression for D , which can be expanded as

$$D = \int_0^\infty \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} \int_0^\infty \sqrt{1 - \frac{1}{\left(1 + \frac{I_{th} g_d}{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}\right)^2}} e^{-g_d} d g_d d \hat{g}_c. \quad (32)$$

Note that $\sqrt{1-a}$ can be approximated as $1 - \frac{1}{2}a$, for $|a| < 1$. Thus, we have that

$$\begin{aligned} D &\approx \int_0^\infty \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} \int_0^\infty \left[1 - \frac{1}{2} \frac{1}{\left(1 + \frac{I_{th} g_d}{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}\right)^2} \right] e^{-g_d} d g_d d \hat{g}_c \\ &= \int_0^\infty \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} d \hat{g}_c \\ &\quad - \frac{1}{2} \int_0^\infty \frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}} \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} \left[\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}} \right. \\ &\quad \left. \times e^{\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}}} E_i \left(-\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}} \right) + 1 \right] d \hat{g}_c \\ &= 1 - \frac{1}{2} \int_0^\infty \frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}} \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} d \hat{g}_c \\ &\quad - \frac{1}{2} \int_0^\infty \frac{e^{-\frac{\hat{g}_c}{1-\sigma_e^2}}}{1-\sigma_e^2} \left[\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}} \right]^2 e^{\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}}} \\ &\quad \times E_i \left(-\frac{N_0 B (\hat{g}_c - \sigma_e^2 \ln(\rho))}{I_{th}} \right) d \hat{g}_c \\ &= 1 - \frac{\eta_1}{2} (1 - \sigma_e^2) + \frac{\eta_2}{2} \\ &\quad + \underbrace{\frac{1}{2\eta_1(1-\sigma_e^2)} e^{-\frac{\eta_2}{(1-\sigma_e^2)\eta_1}} \int_{-\eta_2}^\infty x^2 e^{\left(1 - \frac{1}{(1-\sigma_e^2)\eta_1}\right)x} E_1(x) dx}_{\Gamma}. \end{aligned} \quad (33)$$

To obtain the closed-form expression for Γ , we consider [34, Eq. 4.2.16], given as

$$\begin{aligned} \int x^n e^{ax} E_1(bx) dx &= \frac{n!}{a^{n+1}} E_1[(a+b)x] \\ &\quad - \frac{n!}{a^{n+1}} e_n(ax) e^{-ax} E_1(bx) + \frac{n!}{a^{n+1}} E_1(bx) \\ &\quad + \frac{n!}{a^{n+1}} e^{-(a+b)x} \sum_{k=1}^n \frac{e_{k-1}[(a+b)x]}{k \left(1 + \frac{b}{a}\right)^k}, \end{aligned} \quad (34)$$

where $e_n(x) = \sum_{k=0}^n \frac{x^k}{k!}$. Thus, Γ is derived in closed-form as follows:

$$\begin{aligned} \Gamma &= -\frac{e^{-\frac{\eta_2}{(1-\sigma_e^2)\eta_1}}}{\eta_1(1-\sigma_e^2) \left(\frac{1}{(1-\sigma_e^2)^2 \eta_1} - 1 \right)^3} \left\{ E_1 \left(-\frac{\eta_2}{(1-\sigma_e^2)\eta_1} \right) \right. \\ &\quad \left. - e_2 \left[-\eta_2 \left(\frac{1}{(1-\sigma_e^2)} - 1 \right) \right] e^{\eta_2 \left(\frac{1}{(1-\sigma_e^2)\eta_1} - 1 \right)} E_1(-\eta_2) \right. \\ &\quad \left. + e^{\frac{\eta_2}{(1-\sigma_e^2)\eta_1}} \sum_{k=1}^2 \frac{e_{k-1} \left(-\frac{\eta_2}{\eta_1(1-\sigma_e^2)} \right)}{k \left(1 + \frac{1}{\eta_1(1-\sigma_e^2)} \right)^k} \right\}. \end{aligned} \quad (35)$$

By exploiting (31), as well as the approximation of D in (33) with Γ in (35), the approximated closed-form expression of R can be derived and is given in (20). Similarly, the optimal average transmit power \bar{P}_d^* can be derived in closed-form, given in (21), and its proof is omitted here for simplicity.

REFERENCES

- [1] W. Yu, L. Musavian, and Q. Ni, "Tradeoff analysis and joint optimization of link-layer energy efficiency and effective capacity toward green communications," *IEEE Trans. Wirel. Commun.*, vol. 15, pp. 3339–3353, Jan. 2016.
- [2] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches." available online: <http://arxiv.org/abs/1709.00560>, Sept. 2017.
- [3] "3GPP, Summary of email discussion on the link level evaluation for LTE URLLC." available online: http://www.3gpp.org/ftp/Meeting_3GPP_SYNC/RAN1/Docs/, TSG RAB WG1 Meeting 92, R1-1801385, Mar. 2018.
- [4] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, Jun. 2018.
- [5] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, "Fundamental tradeoffs among reliability, latency and throughput in cellular networks," in *Proc. GC Workshops, Austin, TX, USA*, pp. 1391–1396, Dec. 2014.
- [6] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. International Conference on 5G for Ubiquitous Connectivity*, pp. 146–151, Nov. 2014.
- [7] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultra-reliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sept. 2016.
- [8] Y. Li, M. C. Gursoy, and S. Velipasalar, "Throughput of two-hop wireless channels with queueing constraints and finite blocklength codes," in *Proc. ISIT, Barcelona, Spain*, pp. 2599–2603, Jul. 2016.
- [9] Y. Hu, M. C. Gursoy, and A. Schmeink, "Relaying-enabled ultra-reliable low-latency communications in 5G," *IEEE Netw.*, vol. 32, no. 2, pp. 62–68, Mar. 2018.
- [10] W. Yu, A. Qudus, S. Vahid, and R. Tafazolli, "Opportunistic spectrum access in support of ultra-reliable and low-latency communications," in *Proc. IEEE GLOBECOM Workshop, Abu Dhabi, UAE*, Dec. 2018.
- [11] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [12] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, Jul. 2018.
- [13] Y. Xu, C. Shen, T. Chang, S. Lin, Y. Zhao, and G. Zhu, "Energy-efficient non-orthogonal transmission under reliability and finite blocklength constraints," in *Proc. IEEE Globecom Workshops*, pp. 1–6, Dec. 2017.
- [14] Y. Xu, C. Shen, T.-H. Chang, S.-C. Lin, Y. Zhao, and G. Zhu, "On energy-efficient NOMA designs for heterogeneous low-latency downlink transmissions." available online: <https://arxiv.org/abs/1806.09144>, 2018.

- [15] W. Yu, L. Musavian, and Q. Ni, "Link-layer capacity of NOMA under statistical delay QoS guarantees," *IEEE Trans. Commun.*, in press.
- [16] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, Apr. 2014.
- [17] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 40–48, Apr. 2014.
- [18] Z. Chu, H. X. Nguyen, T. A. Le, M. Karamanoglu, E. Ever, and A. Yazici, "Secure wireless powered and cooperative jamming D2D communications," *IEEE Trans. Green Commun. Net.*, vol. 2, no. 1, pp. 1–13, Mar. 2018.
- [19] K. Ali, H. X. Nguyen, Q. T. Vien, and P. Shah, "Disaster management communication networks: Challenges and architecture design," in *Proc. Pervasive Computing and Communication Workshops (PerCom Workshops)*, pp. 537–542, Mar. 2015.
- [20] K. Ali, H. X. Nguyen, P. Shah, Q. T. Vien, and N. Bhuvanandaram, "Architecture for public safety network using D2D communication," in *Proc. IEEE WCNC, Doha, Qatar*, pp. 1–6, Apr. 2016.
- [21] Z. Chu, T. A. Le, H. X. Nguyen, A. Nallanathan, and M. Karamanoglu, "A stackelberg-game approach for disaster-recovery communications utilizing cooperative D2D," *IEEE Access*, vol. 6, pp. 10733–10742, Mar. 2018.
- [22] K. Ali, H. X. Nguyen, Q. T. Vien, P. Shah, and Z. Chu, "Disaster management using D2D communication with power transfer and clustering techniques," *IEEE Access*, vol. 6, pp. 14643–14654, 2018.
- [23] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, "A survey of device-to-device communications: Research issues and challenges," *IEEE Commun. Survey Tuts.*, vol. 20, pp. 2133–2168, thirdquarter 2018.
- [24] 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.
- [25] 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.
- [26] J. G. S. Schiessl and H. Al-Zubaidy, "Delay analysis for wireless fading channels with finite blocklength channel coding," in *Proc. ACM MSWiM*, 2015.
- [27] L. Musavian and S. Aissa, "Fundamental capacity limits of cognitive radio in fading environments with imperfect channel information," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3472–3480, Nov. 2009.
- [28] H. Tuy, *Convex Analysis and Global Optimization*. Springer Publishing Company, Incorporated, 2nd ed., 2016.
- [29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [30] L. N. Tran, M. F. Hanif, A. Tolli, and M. Juntti, "Fast converging algorithm for weighted sum rate maximization in multicell MISO downlink," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 872–875, Dec. 2012.
- [31] L. N. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, Jan. 2014.
- [32] A. Ben-Tal and A. S. Nemirovskiaei, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2001.
- [33] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions*. New York: Dover, 1965.
- [34] M. Geller and E. Ng, "Table of integrals of the exponential integral," *J. Res. Nat. Bur. Stand.*, no. 73.