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-critical applications and services (such as remote surgery, unmanned aerial vehicle (UAV) deliveries, smarter transportations, 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. 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 (ZMCSG) random 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

\[
h_c[i] = \hat{h}_c[i] + \Delta h_c[i], \tag{1}
\]

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

Furthermore, it is noted that for short-packet transmissions, finite error probability is inevitable [7]. We assume that \( \epsilon \)

\( ^1 \)Channel 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].
denotes the block error probability, i.e., $0 < \epsilon < 1$, $m$ represents the finite block-length, and $\gamma$ 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]
\[
\hat{R}_i \approx \log_2 \left(1 + \gamma |h_d[i]|^2\right) - \sqrt{\frac{1}{m} \left(1 - \frac{1}{1 + \gamma |h_d[i]|^2}\right) Q^{-1}(\epsilon) / \ln 2},
\]
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 transmission may fail with a probability $\epsilon$. 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]\(^2\)
\[
\hat{R}_i = \begin{cases} 
0, & \text{with probability } \epsilon, \\
m\hat{R}_i, & \text{with probability } 1 - \epsilon.
\end{cases}
\]
Thus, the achievable normalized average rate at the DR with finite block-length codes is written as
\[
R = (1 - \epsilon)\mathbb{E}\left[\frac{\hat{R}_i}{m}\right] = (1 - \epsilon)\mathbb{E}[\hat{R}_i].
\]

Remark 1: In ultra-reliable 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 $\epsilon$ 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 $\epsilon$ have been taken into account. Specifically, when $m$ decreases, a larger penalty is placed on the instantaneous achievable rate. This demonstrates the 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 imperfection 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 $\Delta g_c = |\Delta h_c|^2$. The probability density functions (PDFs) of $g_d$, $\hat{g}_c$, and $\Delta g_c$ are respectively given as
\[
f(g_d) = e^{-g_d}, \; f(\hat{g}_c) = \frac{1}{1 - \sigma_c^2} e^{-\frac{\hat{g}_c}{1 - \sigma_c^2}}, \; f(\Delta g_c) = \frac{1}{\sigma_c^2} e^{-\frac{\Delta g_c}{\sigma_c^2}^2}.
\]

Then, we perform optimal power control over time to maximize the average achievable rate for the D2D pair, which is given as
\[
R = \max_{P_d} (1 - \epsilon)\mathbb{E}\left[\log_2 \left(1 + \frac{P_d g_d}{N_0 B}\right) - \frac{Q^{-1}(\epsilon)}{\ln 2} \sqrt{\frac{1}{m} \left(1 - \frac{1}{1 + \frac{P_d g_d}{N_0 B}}\right)^2}\right],
\]
s.t. $\Pr\{P_d(\hat{g}_c + \Delta g_c) \geq I_{th}\} \leq \varrho$,
\[
\Rightarrow \int_{0}^{\infty} f(g_d)dg_d \int_{0}^{\infty} f(\hat{g}_c)dg_c \int_{0}^{\frac{I_{th}}{g_d}} f(\Delta g_c)d\Delta g_c \leq \varrho
\Rightarrow \int_{0}^{\infty} e^{-g_d} \int_{0}^{\infty} e^{-\frac{\hat{g}_c}{1 - \sigma_c^2} + \frac{\Delta g_c}{\sigma_c^2}} \frac{1}{1 - \sigma_c^2} e^{-\frac{\Delta g_c}{\sigma_c^2}^2} dg_c dg_d \leq \varrho.
\]

2The transmission failure with probability $\epsilon$ 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.

3Here, the time index $i$ is omitted for simplicity.
Hence, the following sufficient interference power constraint holds
\[
\int_0^\infty e^{-ya} \int_0^\infty e^{-\frac{yb}{a^2} + \frac{yb}{a^2}^2} \frac{1}{1 - \sigma_e^2} e^{-\frac{yb}{a^2}} dg_idgd.
\]
Furthermore, the following sufficient interference power constraint holds
\[
\int_0^\infty e^{-ya} \int_0^\infty a^{-1} - \frac{yb}{a^2} e^{-\frac{yb}{a^2}} dg_idgd.
\]
Hence, the following sufficient interference power constraint holds
\[
e^{-\frac{yb}{a^2} + \frac{yb}{a^2}^2} \leq q \Rightarrow \frac{I_{th}}{P_d\sigma_e^2} + \frac{\hat{g}_c}{\sigma_e^2} \leq \ln q,
\]
\[
\Rightarrow \frac{I_{th}}{P_d\sigma_e^2} \geq \frac{\hat{g}_c}{\sigma_e} - \ln q \Rightarrow P_d \leq \frac{I_{th}}{\hat{g}_c - \sigma_e^2 \ln q}.
\]
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
\[
R = \max_{P_d} (1 - e)E \left[ \log_2 \left( 1 + \frac{P_d g_d}{N_0 B} \right) \right]
\]
\[
- \frac{Q^{-1}(c)}{\ln 2} \left( 1 - \frac{1}{1 + \frac{P_d g_d}{N_0 B}} \right)^2, \]
s.t. \( P_d \leq \frac{I_{th}}{\hat{g}_c - \sigma_e^2 \ln q}. \)

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

### III. Optimal Resource Allocation

In this section, we exploit the convexity of the instantaneous achievable rate 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
\[
\max_{b \geq 0} \hat{R} = \log_2(1 + b) - c \sqrt{b + 2}\left(\frac{b + 2}{b + 1}\right)^2 \leq \frac{I_{th, g_d}}{\hat{g}_c - \sigma_e^2 \ln q} \left(\frac{b + 2}{b + 1}\right), \quad \text{s.t. } b \geq \frac{P_d g_d}{N_0 B}, \quad c = \frac{Q^{-1}(c)}{\sqrt{2} \ln 2}.
\]
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
\[
f_0(b) = \log_2(1 + b) - c \sqrt{b + 2}\left(\frac{b + 2}{b + 1}\right).
\]
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 \( b_0 \geq 0 \).
\[
f_1(b) = \frac{(1 + b) \log_2(1 + b)}{\sqrt{b(b + 2)}}, \quad f_2(b) = \frac{(b + 1)[b(b + 2)]^{\frac{3}{2}}}{(3b^2 + 6b + 1)}.
\]

**Proof:** Please refer to Appendix A.

From the analysis in Lemma 1, one can see that if \( f_0(b) \) 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.,
\[
c \sqrt{b(n) + 2} \left(\frac{b(n) + 2}{b(n) + 1}\right) \approx c \sqrt{b(n)(b(n) + 2)} \left(\frac{b(n) + 2}{b(n) + 1}\right)
\]
\[
+ \left(\frac{1 + b(n))}{2} \sqrt{b(n)(b(n) + 2)}(b - b(n)),
\]
where \( b(n) \) is the updated value of the variable \( b \) at the \( n \)-th iteration. Hence, problem (10) can be reformulated as
\[
\max_{b \geq 0} \log_2(1 + b) - c \sqrt{b(n)(b(n) + 2)} \left(\frac{b(n) + 2}{b(n) + 1}\right)
\]
\[
- \left(2 + (b(n)) \sqrt{b(n)(b(n) + 2)}(b - b(n))
\]
s.t. \( b \leq \frac{I_{th, g_d}}{\hat{g}_c - \sigma_e^2 \ln q} \left(\frac{b(n) + 2}{b(n) + 1}\right), \quad (12)\)

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 \(\sqrt{1 - \alpha} \approx 1 - \frac{1}{2}\alpha\), for \(|\alpha| < 1\), to approximate the objective function in problem (10) and reformulate the problem as follows

\[
\max_{b \geq 0} \hat{R} \approx \log_2(1 + b) + \frac{c}{2(1 + b)^2} - c
\]

\[
s.t. \ b \leq \frac{I_{th}g_d}{[g_c - \sigma^2 \ln(\rho)]N_0B}.
\]

(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

\[
\max_{b \geq 0} \log_2(1 + b) + \frac{c}{2(1 + b(n))^2} - \frac{c}{(1 + b(n))^3} (b - b(n))\]

\[
s.t. \ b \leq \frac{I_{th}g_d}{[g_c - \sigma^2 \ln(\rho)]N_0B}.
\]

(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 \(O(N)\), 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:

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

(15)

where \(\rho = \frac{P_d}{N_0B}\) 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 \(\rho\), under the following assumptions, \(m \geq 10^2, \epsilon \geq 10^{-6}\), and \(\rho \geq -10\) dB.

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:

\[
\hat{R}(\rho) \approx \mathbb{E}[\log_2(1 + \rho g_d) - \frac{Q^{-1}(\epsilon)}{\ln 2} \sqrt{1/m}].
\]

(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 \(\rho\) in the high SNR region. Since the transmit power \(P_d\) is equal to \(\rho\) 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 \(\rho\). 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

\[
P_d^* = \frac{I_{th}}{g_c - \sigma^2 \ln(\rho)}.
\]

(17)

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

\[
R = (1 - \epsilon)\mathbb{E} \left[ \log_2 \left(1 + \frac{I_{th}g_d}{N_0B(g_c - \sigma \ln(\rho))}\right) \right]
\]

\[
\frac{1}{m} \ln 2 \frac{1}{\left(1 + \frac{I_{th}g_d}{N_0B(g_c - \sigma^2 \ln(\rho))}\right)^2},
\]

(18)

and the optimal average transmit power is given by

\[
\hat{P}_d^* = \mathbb{E}[P_d^*] = \mathbb{E} \left[ \frac{I_{th}}{g_c - \sigma^2 \ln(\rho)} \right].
\]

(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 \(\eta_1 = \frac{N_0B}{I_{th}}, \eta_2 = \frac{N_0B^2 \sigma^2 \ln(\rho)}{I_{th}}, e_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}, \beta = -3\).

The assumptions here are reasonable for URLLC applications.
average finite block-length rate versus SNR

\[ \frac{1}{m} Q^{-1}(\epsilon) \log_2(\epsilon), \] and \( E_1(\cdot) \) is the exponential integral function [33]. In addition, the optimal average transmit power \( P_d^* \) for the DR can be derived in closed-form

\[ P_d^* = \frac{I_{th} \ln(\epsilon)}{1 - \sigma_x^2} e^{\frac{\sigma_x^2 \ln(\epsilon)}{1 - \sigma_x^2}} E_1 \left( \frac{\sigma_x^2 \ln(\phi)}{1 - \sigma_x^2} \right). \]  

(21)

**Proof:** Please refer to Appendix C.

**Theorem 1** presents the closed-form expression for the maximum achievable average 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,

\[ 1 - \frac{1}{1 + \frac{I_{th} \sigma_x^2}{N_0 B (\hat{g}_c - \sigma_x^2 \ln(\phi))}} \approx 1. \]  

(22)

Accordingly, the maximum achievable average rate obtained at high SNRs, denoted by \( R_h \), can be approximated as follows:

\[ R_h \approx (1 - \epsilon) \left\{ \mathbb{E} \left[ \ln \left( 1 + \frac{I_{th} \sigma_x^2}{N_0 B (\hat{g}_c - \sigma_x^2 \ln(\phi))} \right) \right] - \beta \right\}. \]  

(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., \( \beta \) in (23), due to the fact that the term \( \mathbb{E} \left( \beta \sqrt{1 - \frac{1}{(1 + \rho d^2)}} \right) \) will approach \( \beta \) as the SNR \( \rho \) increases. In addition, Fig. 1 demonstrates that the average finite block-length rate versus SNR \( \rho \), 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 \( \epsilon \). 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 \( \sigma_x^2 \) 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 \( m = 20 \) 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

### TABLE I: Notations

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Notation</th>
<th>Typical Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel estimation error variance</td>
<td>( \sigma_x^2 )</td>
<td>( 10^{-2} )</td>
</tr>
<tr>
<td>Interference outage probability</td>
<td>( \theta )</td>
<td>( 10^{-2} )</td>
</tr>
<tr>
<td>Block error probability</td>
<td>( \epsilon )</td>
<td>( 10^{-2} )</td>
</tr>
<tr>
<td>Interference threshold</td>
<td>( I_{th} )</td>
<td>( 10 ) dBW</td>
</tr>
<tr>
<td>Block-length</td>
<td>( m )</td>
<td>( 500 )</td>
</tr>
</tbody>
</table>

Fig. 1: Comparison between high SNR finite block-length rate and achievable finite block-length rate.
\[ R \approx (1 - e) \left\{ \frac{I_{th}}{I_{th} \ln 2 - N_0 B \ln 2(1 - \sigma_e^2)} \left[ -e^{-\frac{\sigma^2 \ln(\rho)}{1 - \sigma_e^2}} E_1 \left[-\frac{\sigma^2 \ln(\rho)}{1 - \sigma_e^2}\right] + e^{-\eta_2} E_1[-\eta_2] \right] - \beta \left[ 1 - \frac{\eta_1}{2} (1 - \sigma_e^2) + \frac{\eta_2}{2} - \frac{e^{-\frac{\eta_2}{\eta_1(1 - \sigma_e^2)}}}{\eta_1(1 - \sigma_e^2)} \right]^{-\frac{\eta_2}{\eta_1(1 - \sigma_e^2)}} E_1(-\eta_2) + e^{\frac{\eta_2}{\eta_1(1 - \sigma_e^2)}} \sum_{k=1}^{\infty} \frac{e^{k-1} - \eta_2}{\eta_1(1 - \sigma_e^2) - 1} \right\} \] (20)

Fig. 2: Convergence of Algorithm 1.

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
Achievable rate (bps/Hz)

-0.5

average achievable rate. This indicates that accurate channel estimation results in a larger transmit power limit and thus contributes to a higher average achievable rate. Moreover, Fig. 6 also shows that a larger interference tolerance power has a positive effect on the average achievable rate, which is finite 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 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 $\sigma_e^2$ and a larger block error probability $\epsilon$ 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. 5: The average finite block rate versus the interference power threshold $I_{th}$.

at high SNRs, i.e., $R_{th}$, 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

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

which approaches $\mathbb{E} \{ \frac{1}{m} Q^{-1}(\epsilon) \}$ with a larger m and $I_{th}$ such that $R_{th}$ 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 finite 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 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 $\sigma_e^2$ and a larger block error probability $\epsilon$ 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. 6: The average achievable rate versus 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 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 $\sigma_e^2$ and a larger block error probability $\epsilon$ 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.

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 formul-ate an achievable rate maximization problem subject to 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

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

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

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

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

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

$$c \leq f_2(b) = \frac{(b + 1)[b(b + 2)]^{3/2}}{(3b^2 + 6b + 1)}.$$ (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 $b \geq f_2^{-1}(c)$ 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 $\dot{R}$, given by

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

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 $Q^{-1}(\epsilon) \leq Q^{-1}(10^{-6})$. 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

$$R = (1 - \epsilon) \left[ \frac{\log_2 \left( 1 + \frac{I_{th} g_d}{N_0 B(\gamma_c - \sigma_e \ln(\bar{\rho}))} \right)}{c} \right],$$

$$-\beta \left[ \frac{1}{1 + \frac{I_{th} g_d}{N_0 B(\gamma_c - \sigma_e \ln(\bar{\rho}))}} \right] \right] \right] \right] \right],$$

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

$$C = \int_0^{\infty} \frac{e^{-\frac{\eta_1}{\eta_2}}}{1 - \eta_2} \int_0^{\infty} \frac{1}{\ln 2} \left[ 1 + \frac{I_{th} g_d}{N_0 B(\gamma_c - \sigma_e \ln(\bar{\rho}))} \right] e^{-\eta_4 g_d g_d g_c} \frac{I_{th} g_d}{N_0 B(\gamma_c - \sigma_e \ln(\bar{\rho}))} \left( I_{th} g_d / N_0 B(\gamma_c - \sigma_e \ln(\bar{\rho})) \right) d g_c.$$ (29)

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

$$C = e^{-\frac{\eta_1}{\eta_2}} \int_{-\eta_2}^{\infty} \left( 1 - \frac{\eta_1}{\eta_2} \right) E_1(x) dx$$

$$= \frac{I_{th} - N_0 B (1 - \sigma_e^2)}{2} \left[ e^{-\frac{\eta_1}{\eta_2} \sigma_e^2 \ln(\bar{\rho})} \right] E_1 \left[ \frac{\sigma_e^2 \ln(\bar{\rho})}{1 - \sigma_e^2} \right]$$

$$+ e^{-\eta_2} E_1[-\eta_2].$$ (31)

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

5
Then, we focus on the derivation of closed-form expression for $D$, which can be expanded as

$$D = \int_0^\infty \frac{e^{\frac{R}{2}}}{1 - \sigma^2 e^\sqrt{\frac{R}{2}}} \left(1 - \frac{1}{\left(1 + \frac{I_{th}g_0}{N_0 B(\sigma^2 \ln(\varrho))}\right)^2}\right) e^{-2d} \sigma d\gamma. \tag{32}$$

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

$$D \approx \int_0^\infty \frac{e^{\frac{R}{2}}}{1 - \sqrt{\frac{R}{2}}} \left[1 - \frac{1}{2} \left(1 + \frac{I_{th}g_0}{N_0 B(\sigma^2 \ln(\varrho))}\right)^2\right] e^{-2d} \sigma d\gamma \tag{33}$$

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

$$\int x^n e^{-ax} E_1(bx) dx = \frac{n!}{a^{n+1}} E_1\left((a+b)x\right) - \frac{n!}{a^{n+1}} e_{n+1}(ax) e^{-ax} E_1(bx) + \frac{n!}{a^{n+1}} E_1(bx) + \frac{n!}{a^{n+1}} e^{-a(b+x)} \sum_{k=1}^{n} e_{k-1}\left((a+b)x\right) \frac{1}{k!} \left(1 + \frac{b}{a}\right)^k, \tag{34}$$

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

$$\Gamma = -\frac{e^{-\frac{\eta_2}{1 - \sigma^2 \ln(\varrho)}}}{\eta_1 (1 - \sigma^2 \ln(\varrho))} \sum_{k=1}^{2} \left[\eta_2 \left(1 - \frac{1}{1 - \sigma^2 \ln(\varrho)}\right) - \frac{1}{(1 - \sigma^2 \ln(\varrho))}\right] E_1(-\eta_2) + \frac{e^{\frac{\eta_2}{1 - \sigma^2 \ln(\varrho)}}}{\eta_1 (1 - \sigma^2 \ln(\varrho))} \sum_{k=1}^{2} e_{k-1}\left(\frac{\eta_2}{1 - \sigma^2 \ln(\varrho)}\right) \frac{1}{k!} \left(1 + \frac{1}{1 - \sigma^2 \ln(\varrho)}\right)^k. \tag{35}$$

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

**References**


