Low-Latency Short-Packet Transmission over a Large Spatial Scale

Lei Huang [1], Xiaoyu Zhao [1], Wei Chen [1] and H. Vincent Poor [2,*]

1 Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China; huang-l17@tsinghua.org.cn (L.H.); xy-zhao16@mails.tsinghua.edu.cn (X.Z.); wchen@tsinghua.edu.cn (W.C.)

2 Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA

* Correspondence: poor@princeton.edu

Abstract: Short-packet transmission has attracted considerable attention due to its potential to achieve ultralow latency in automated driving, telesurgery, the Industrial Internet of Things (IIoT), and other applications emerging in the coming era of the Six-Generation (6G) wireless networks. In 6G systems, a paradigm-shifting infrastructure is anticipated to provide seamless coverage by integrating low-Earth orbit (LEO) satellite networks, which enable long-distance wireless relaying. However, how to efficiently transmit short packets over a sizeable spatial scale remains open. In this paper, we are interested in low-latency short-packet transmissions between two distant nodes, in which neither propagation delay, nor propagation loss can be ignored. Decode-and-forward (DF) relays can be deployed to regenerate packets reliably during their delivery over a long distance, thereby reducing the signal-to-noise ratio (SNR) loss. However, they also cause decoding delay in each hop, the sum of which may become large and cannot be ignored given the stringent latency constraints. This paper presents an optimal relay deployment to minimize the error probability while meeting both the latency and transmission power constraints. Based on an asymptotic analysis, a theoretical performance bound for distant short-packet transmission is also characterized by the optimal distance–latency–reliability tradeoff, which is expected to provide insights into designing integrated LEO satellite communications in 6G.

Keywords: 6G; short-packet transmission; URLLC; finite-blocklength coding; large spatial scale; relaying; propagation delay; end-to-end delay; decoding delay; asymptotic analysis

1. Introduction

5G wireless communication systems and a series of perspectives on conceptualized 6G systems have emerged as powerful platforms for multiple use cases such as augmented/virtual reality (AR/VR), teleoperated surgery, and automatic driving [1,2]. An enhanced ultrareliable and low-latency communication (URLLC) has been proposed to support more stringent requirements on reliability and latency to meet the demands of those services. Specifically, the envisioned 6G wireless network may require URLLC based on a 10^{-10} packet loss probability and an over-the-air latency on the order of 0.1 ms [3,4].

With latency at a submillisecond level, short-packet transmissions are necessary [5]. Therefore, we should consider finite-blocklength channel coding to ensure reliability in URLLC. For the finite-blocklength (FBL) regime, the authors of [6] formulated the maximal coding rate of FBC over additive white Gaussian noise (AWGN) channels. The coding rate analysis has been extended to other practical scenarios, e.g., multiple-antenna fading channels [7], multiple-antenna Rayleigh-fading channels [8], block-fading channels [9], and hybrid automatic repeat request (HARQ) [10]. Latency is not only due to the physical layer transmission discussed in the literature above, but is also caused by the queuing of data packets in the network layer [11]. With a cross-layer design of variable-length coding for a single link, we have also achieved extremely low-latency communications in [12]. The

cross-layer design for URLLC was also discussed in [13,14]. The author of [13] minimized the power consumption with a reliability constraint by adopting a packet-dropping policy. In [14], the violation probabilities of the maximal delay and peak-age of information were given for URLLC.

Nonterrestrial communications are also at the forefront of 6G technology research [15], in order to provide three-dimensional (3D) coverage by complementing on-the-ground infrastructures with aerial platforms, including satellites. For large-scale communication systems, the signal-to-noise ratio, which decays rapidly with distance, is a crucial factor that affects system performance. For this reason, relaying is essential to gain higher reliability. Relay communication began to attract researchers' attention from the classical work of [16]. Various relaying schemes, including fixed relaying, selection relaying, and incremental relaying, were first discussed by Laneman, Tse, and Wornell in [17]. Different cooperative schemes between the direct link and the relay link were discussed under both amplify-and-forward (AF) or decode-and-forward (DF) modes in [18]. In [19], we extended the decode-and-cancel protocol proposed in our previous work [20] to cancel inter-relay interference (IRI) in the scheme CAO-SIR. In CAO-SIR, we achieved an equivalent parallel relay model for DF-based successive relaying since the interference is entirely canceled by delicately adjusting the transmission order of relay nodes.

Recently, researchers have also investigated relaying under the finite-blocklength regime (FBL). The authors in [21] derived closed-form expressions for the coding rates of relay communications under the Nakagami-m fading channel in the finite-blocklength regime. In [22], the researchers studied the relaying throughput performance of the quasistatic Rayleigh channels under the finite-blocklength regime, optimized the optimal distance and blocklength for URLLC, and compared that to the infinite-blocklength regime (IBL). However, relaying communications over a long distance suffer from technical difficulties such as the deterioration of the signal-to-noise ratio caused by the long distance, the increased latency caused by the increased number of relays, and the unclear theoretical upper limit of the optimal delay caused by the speed of light. In [21,22], large-scale communication was not taken into consideration, and the authors' model was based on a fixed resource allocation and relay deployment scheme. Though [23] considered resource allocation for blocklength and power, this did not include a latency constraint because it was based on infinite-blocklength coding and a Poisson field of interferers. Thus, achieving low-latency short-packet communications over a large spatial scale is still an open problem.

In this paper, we focus on long-distance and short-packet communication. We aim to minimize the overall error probability while limiting the power cost under a time constraint. We prove the convexity of the overall error probability with respect to the blocklength and power under a particular condition when the relays are placed equidistantly. We also find a blocklength power allocating scheme to minimize the error probability, where more power should be allocated to the relays that have fewer encoded symbols. By scaling the problem conditions, we simplify the problem and obtain an analytical expression of the optimal error probability. Thereby, we find that the optimal relaying strategy is to allocate the blocklength and power equally to each relay. Then, we analyze the theoretical limits of the optimal solution. The discussion includes two cases, where the number of relays is fixed and the number of relays is variable. When there is a fixed number of relays in the communication system, we find that the error probability decreases superexponentially with the blocklength and the logarithm of the signal-to-noise ratio. When the number of relays is not fixed, we find there is an optimal number of relays in the relaying communication system. Therefore, the performance of relay communication has a theoretical upper limit. Under certain circumstances, no matter how many relays are added, the performance of the communication system cannot be improved. Based on a numerical simulation, we solve the optimization problem and compare it to our theoretical result.

The rest of this paper is organized as follows. Section 2 presents the system model. In Section 3, we form an optimization problem for the overall error probability in terms of

the distance, blocklength, and signal-to-noise ratio. Then, we give a theoretical solution for the optimal bit-error rate. In Section 4, we analyze the theoretical asymptotic limit and approximate results of the optimal solution. In Section 5, we verify the correctness of our conclusions by a numerical simulation. Finally, conclusions and future suggestions are provided in Section 6.

2. System Model

We focus on an orbital wireless system with the requirement of extremely low latency, which is demonstrated in Figure 1. The satellites communicate with another planet, e.g., Mars, via the space station. Each of the satellites is considered as a DF relay. For simplicity, we only consider the transmission of one packet from the source to destination, as shown in Figure 2, where a source node R_0 transmits a k-bit packet in each transmission to its destination node R_L with the help of $L - 1$ decode-and-forward (DF) relays, called $R_1, R_2, \ldots, R_{L-1}$. The distance between node $R_{\iota-1}$ and R_ι is d_ι. The k-bit packet is encoded into symbols to be transmitted from R_0 to R_L. The modulated waveforms of those symbols are reliably sent by node $R_{\iota-1}$. Node R_ι receives and detects it without distortion. We denote by τ the average time consumed in sending one symbol. As a result, $\frac{1}{\tau}$ is the symbol rate. Let n_ι be the number of symbols transferred between node $R_{\iota-1}$ and R_ι, $n_\iota \in \mathbb{N}_+$.

We assume node $R_{\iota-1}$ encodes the k-bit packet into n_ι symbols. $R_{\iota-1}$ sends a symbol vector $\boldsymbol{x_\iota}$ to R_ι. In particular, $\boldsymbol{x_\iota} = (x_{\iota 1}, x_{\iota 2}, \cdots, x_{\iota n_\iota})$ and $|x_\iota|_2^2 = 1$. $x_{\iota j}$ is the jth symbol. Let h_ι denote the channel coefficient between node $R_{\iota-1}$ and R_ι. Let g_ι denote the channel gain, namely $g_\iota = |h_\iota|^2$. We denote by $\boldsymbol{y_\iota}$ the received symbol vector at node R_ι. Specifically, $\boldsymbol{y_\iota} = (y_{\iota 1}, y_{\iota 2}, \cdots, y_{\iota n_\iota})$. For $1 \leq j \leq n_\iota$, $y_{\iota j}$ can be presented by:

$$y_{\iota j} = h_\iota x_{\iota j} + z_\iota, \tag{1}$$

where z_ι is the additive Gaussian white noise. Let d_ι denote the distance between node $R_{\iota-1}$ and R_ι. The path loss of the transmission from $R_{\iota-1}$ to R_ι is normalized as $d_\iota^{-\alpha}$, where α is the loss exponent. Due to the large scale of our model, the shadow-fading component can be ignored, and the channel gain is presented as $g_\iota = d_\iota^{-\alpha}$ [24]. We assume the signal-to-noise ratio (SNR) of the signal emitted by $R_{\iota-1}$ is γ_ι. Thus, node R_ι receives the signal with the SNR equal to $\gamma_\iota d^{-\alpha}$ and sends it to the next node $R_{\iota+1}$ with the SNR equal to $\gamma_{\iota+1}$ if R_ι detects no error.

We then present the overall reliability from the source to the destination. To achieve extremely low latency and ultrareliability, a finite-blocklength coding scheme is adopted. For the real AWGN channel, a tight bound on the coding rate was given by [6]. Let ϵ_ι denote the error probability from node $R_{\iota-1}$ to R_ι. Since $R_{\iota-1}$ transmits k bits on n_ι symbols, the coding rate is $\frac{k}{n_\iota}$. Using the normal approximation in this paper, we have the formula for the channel coding rate in the transmission from $R_{\iota-1}$ and R_ι, i.e.,

$$\frac{k}{n_\iota} = C_\iota - \sqrt{\frac{V_\iota}{n_\iota}} Q^{-1}(\epsilon_\iota), \tag{2}$$

in which $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt$ is the Gaussian Q-function. Q^{-1} is the inverse function of $Q(x)$. We denote by C_ι and V_ι the channel capacity and the channel diversity, respectively. Under our configuration, the channel capacity is represented by:

$$C_\iota = \log(1 + \gamma_\iota d_\iota^{-\alpha}). \tag{3}$$

The channel dispersion is represented by:

$$V_\iota = \frac{(\gamma_\iota d_\iota^{-\alpha} + 1)^2}{\gamma_\iota d_\iota^{-\alpha}(\gamma_\iota d_\iota^{-\alpha} + 2)} \log_2 e. \tag{4}$$
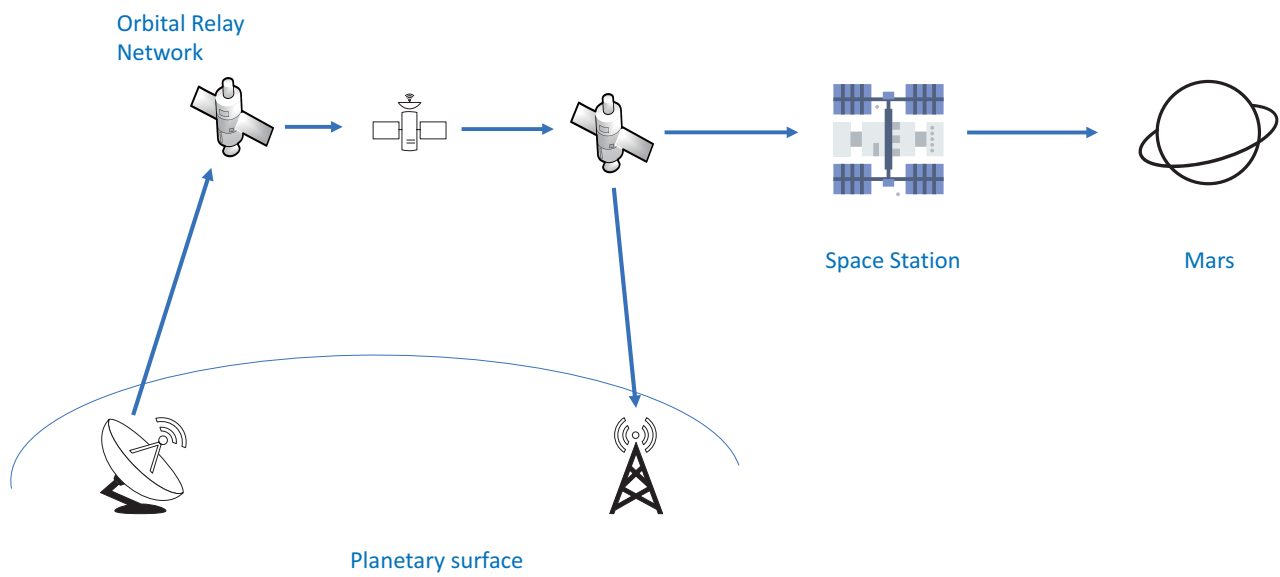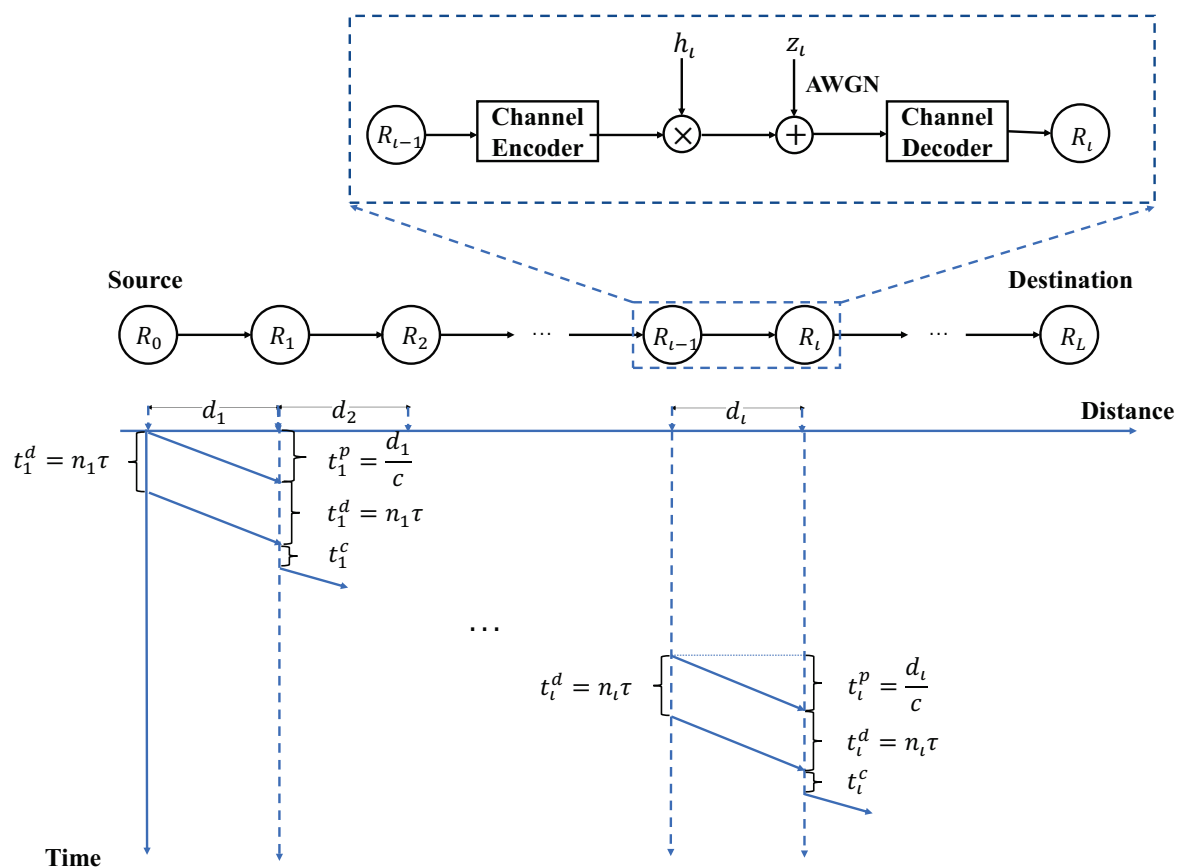
Figure 1. An orbital relaying system.

Figure 2. System model: a communication system with $L-1$ relays.

Substituting Equation (3) into Equation (2), we present the error probability ϵ_ι:

$$\epsilon_\iota = Q\left(\left(\ln(1 + \gamma_\iota d_\iota^{-\alpha}) - \frac{k \ln 2}{n_\iota}\right)\frac{n_\iota}{V_\iota}\right). \tag{5}$$

We assume all relay nodes and the destination node can decode the data packets and detect the error reliably. The relay nodes will forward the messages if and only if no error occurs. As a result, the overall error probability is given by:

$$
\begin{aligned}
\epsilon &= 1 - \prod_{\iota=1}^{L}(1 - \epsilon_\iota) \\
&\approx \sum_{\iota=1}^{L} \epsilon_{\iota,}.
\end{aligned}
\tag{6}
$$

The first-order approximation in Equation (6) gives us a reliable result when $\epsilon_\iota \approx 0$.

Then, we focus on the overall latency from the source to the destination. Since the data packet is transmitted serially, the overall latency T is the sum of the latency from $R_{\iota-1}$ to R_ι, named t_ι.

$$
T = \sum_{\iota=1}^{L} t_\iota.
\tag{7}
$$

In the URLLC communication without retransmission, the physical layer latency can be divided into the following components [11]:

$$
t_\iota = t_\iota^c + t_\iota^p + t_\iota^d,
\tag{8}
$$

where t_ι^c is the time consumed by channel coding, t_ι^p is the time consumed by electronic wave propagation, and t_ι^d is the time consumed by decoding messages. Thus, we have:

$$
t_\iota^d = n_\iota \tau.
\tag{9}
$$

The propagation delay t_ι^p is caused by the signal travel time with the speed of light c, which is presented by:

$$
t_\iota^p = \frac{d_\iota}{c}.
\tag{10}
$$

We assume that each relay is equipped with a strong calculation force so that the time consumed by encoding messages is negligible for t_ι^d and t_ι^p. Thus, the time to transfer for the data packet from $R_{\iota-1}$ to R_ι is:

$$
t_\iota = n_\iota \tau + \frac{d_\iota}{c}.
\tag{11}
$$

To meet the latency constraint, the overall delay $\sum_{i=1}^{L} t_\iota$ is no greater than a constant T^{th}, which gives:

$$
\sum_{\iota=1}^{L} n_\iota \tau + \sum_{\iota=1}^{L} \frac{d_\iota}{c} \leq T^{th}.
\tag{12}
$$

T^{th} represents the maximal tolerable latency from the source to the destination. Since $d_1, d_2, \cdots, d_L$ are constants when the positions of the relays are fixed, Equation (12) can be simplified as:

$$
\sum_{\iota=1}^{L} n_\iota \leq N^{th}(T^{th}),
\tag{13}
$$

where $N^{th}(T^{th})$ is a linear function of Tth, i.e., $N^{th}(T^{th}) = \frac{T^{th} - \sum_{\iota=1}^{L} d_\iota}{\tau}$. By this means, we rewrite the latency constraint in Equation (12) as Equation (13), which is given as a constraint on the overall blocklength.

We consider an overall energy constraint assumption. The energy to transmit the packet from $R_{\iota-1}$ to R_ι equals $n_\iota \tau \gamma_\iota N_0$, for $1 \leq \iota \leq L$. N_0 denotes the single-side noise density. Since τ and N_0 are constants, we have the normalized energy constraint:

$$\sum_{\iota=1}^{L} n_\iota \gamma_\iota \leq E^{th}, \tag{14}$$

where E^{th} is the overall normalized energy of all source nodes and relay nodes.

The error probability ϵ_ι is a function of blocklength n_ι and signal-to-noise ratio γ_ι, as shown in Equation (5).

We next minimize the error probability ϵ subject to the energy and latency constraints.

3. Optimal Relay Deployment and Resource Allocation

In this section, we formulate the optimal tradeoff among the latency, normalized energy, and error probability. For this purpose, we first present the overall error probability in terms of the blocklength n_ι and signal-to-noise ratio γ_ι of each node R_ι. Then, the optimization problem is formulated to minimize the error probability with constraints on the overall blocklength and normalized energy.

3.1. Problem Formulation

We let $Q(\mathcal{Y}(\gamma_\iota, d_\iota, n_\iota))$ be the error probability of the link from $R_{\iota-1}$ to R_ι. $\mathcal{Y}$ is given by:

$$\mathcal{Y}(\gamma, d, n) = \left(\ln\left(1 + \gamma d^{-\alpha}\right) - \frac{k \ln 2}{n} \right) \sqrt{\frac{n(\gamma d^{-\alpha} + 1)^2}{\gamma d^{-\alpha}(\gamma d^{-\alpha} + 2)}}. \tag{15}$$

We denote by D the summation of d_ι. Then, $\sum_{\iota=1}^{L} d_\iota = D$. In Section 2, we have the latency (blocklength) constraint in Equation (13) and the energy constraint in Equation (14). From the previous discussion, we establish the optimization problem (16) to minimize the error probability by allocating blocklength and power to each relay:

$$\min_{\gamma_\iota, d_\iota, n_\iota} \epsilon = \sum_{\iota=1}^{L} Q(\mathcal{Y}(\gamma_\iota, d_\iota, n_\iota)) \tag{16a}$$

$$\text{s.t. } \sum_{\iota=1}^{L} n_\iota \gamma_\iota \leq E^{th}, \tag{16b}$$

$$\sum_{\iota=1}^{L} d_\iota = D, \tag{16c}$$

$$\sum_{\iota=1}^{L} n_\iota \leq N^{th}, \tag{16d}$$

$$\gamma_\iota \geq 0, \tag{16e}$$

$$n_\iota \in \mathbb{N}_+. \tag{16f}$$

The optimization problem (16) is subject to the relay placement in Equation (16c) and the finite-blocklength coding in Equations (16b) and (16d). We consider a case where the blocklength of a finite-blocklength code tends to infinity, i.e., $N^{th} \to \infty$. As the overall blocklength N^{th} increases, the effect of finite-blocklength coding is no longer apparent. There is no loss of generality in assuming the same coding process for each relay. Therefore, in this case, a uniform distribution of relays minimizes the error probability. Thus, when we consider the general FBC case, unless otherwise stated, we assume that the relays are deployed equidistantly. We then only consider resource allocation in the finite-blocklength coding. Furthermore, this deployment strategy can also be verified as the optimal one in the simulation results in Section 6.

Then, we simplify Problem (16) under the assumption that relays are placed equidistantly. The error probability ϵ given by (16a) is a decreasing function of n_ι and γ_ι. Thus, we can certainly rewrite the inequations in Equations (16b) and (16d) as equations. Moreover, we expand the domain of n_ι to real numbers in the interval $[1, \infty)$ to avoid the complex integer programming.

In the case where relays are placed equidistantly, we denote by $Q(\mathcal{G}(\gamma, n))$ the error probability to transfer k bits from $R_{\iota-1}$ to R_ι, where:

$$\mathcal{G}(\gamma, n) = \left(\ln(1 + \gamma) - \frac{k \ln 2}{n} \right) \sqrt{\frac{n(\gamma + 1)^2}{\gamma(\gamma + 2)}}. \tag{17}$$

With the discussion above, we have the optimization problem (18).

$$\min_{\gamma_\iota, n_\iota} \epsilon = \sum_{\iota=1}^{L} Q(\mathcal{G}(\gamma_\iota, n_\iota)) \tag{18a}$$

$$\text{s.t. } \sum_{\iota=1}^{L} n_\iota \gamma_\iota \le E^{th}, \tag{18b}$$

$$\sum_{\iota=1}^{L} n_\iota \le N^{th}, \tag{18c}$$

$$\gamma_\iota \ge 0, \tag{18d}$$

$$n_\iota \ge 1, \tag{18e}$$

The γ_ι in Problem (18) denotes the received SNR with a path loss of $d^{-\alpha}$ from γ_ι in Problem (16). This is the same situation for the constant E^{th}.

3.2. Properties of the Optimal Resource Allocation

Problem (18) is challenging due to its nonlinear constraint Equation (18b). Therefore, we relax the problem's constraints. In particular, we present the following theorem.

Theorem 1. *If* $\ln(1 + \gamma_\iota) - \frac{k \ln 2}{n_\iota} > \max\left\{ \frac{1}{k \ln 2}, \frac{1}{n_\iota} \right\}$, *the* ϵ *in Problem (18) achieves its minimum when all the* γ_ι *are greater than a constant* γ_0 *and the sequence* n_ι *and* γ_ι *is in reverse order, i.e., for any* $n_i \ge n_j$, *we have* $\gamma_i \le \gamma_j$.

Proof of Theorem 1. We use the notation $\varphi(\gamma, n)$ to denote $Q(\mathcal{G}(\gamma, n))$. For any n and any γ_1, γ_2 satisfying the constraints of Problem (18) and $\gamma_1 > \gamma_2$, let $\chi(n) = \varphi(n, \gamma_2) - \varphi(n, \gamma_1)$. We first prove that $\chi(n)$ is a decreasing function in its domain. We present the derivative of the function $\chi(n)$ as:

$$\chi'(n) = \frac{1}{\sqrt{2\pi}} \left(-\exp\left(-\frac{\mathcal{G}_2^2}{2} \right) \frac{\partial \mathcal{G}_2}{\partial n} + \exp\left(-\frac{\mathcal{G}_1^2}{2} \right) \frac{\partial \mathcal{G}_1}{\partial n} \right). \tag{19}$$

We next show $\chi'(n) < 0$, which is equivalent to the following inequation by substituting $\mathcal{G}_n'$ and $\mathcal{G}_\gamma'$ in Appendix A.

$$\exp\left(\frac{\mathcal{G}_2^2 - \mathcal{G}_1^2}{2} \right) < \frac{\ln(1 + \gamma_2) + \frac{k \ln 2}{n}}{\ln(1 + \gamma_1) + \frac{k \ln 2}{n}} \frac{\gamma_2 + 1}{\gamma_1 + 1} \sqrt{\frac{\gamma_1}{\gamma_2} \frac{\gamma_1 + 2}{\gamma_2 + 2}}. \tag{20}$$

Because $\gamma_1 > \gamma_2$, we derive Equation (21) and Equation (22).

$$\frac{\ln(1 + \gamma_2)}{\ln(1 + \gamma_1)} < \frac{\ln(1 + \gamma_2) + \frac{k \ln 2}{n}}{\ln(1 + \gamma_1) + \frac{k \ln 2}{n}}, \tag{21}$$

$$\frac{\gamma_2 + 1}{\gamma_1 + 1} \sqrt{\frac{\gamma_1}{\gamma_2} \frac{\gamma_1 + 2}{\gamma_2 + 2}} > 1. \tag{22}$$

To prove Equation (20), we let $\zeta(\gamma) = \exp\left(-\frac{\mathcal{G}^2(\gamma, n)}{2}\right) \ln(1 + \gamma)$. By calculating its derivative, we obtain:

$$\zeta'(\gamma) = \exp\left(-\frac{\mathcal{G}^2}{2}\right) \left(\frac{1}{1+\gamma} - \mathcal{G}\mathcal{G}'_\gamma \ln(1+\gamma)\right). \tag{23}$$

Because $\ln(1 + \gamma) - \frac{k \ln 2}{n} > \frac{1}{n}$, we have:

$$\begin{aligned}
\mathcal{G}\mathcal{G}'_\gamma \ln(1+\gamma) &> \left(\ln(1+\gamma) - \frac{k\ln 2}{n}\right) n \left(\frac{(\gamma+1)^2}{\gamma(\gamma+2)}\right)^{\frac{1}{2}} \frac{\ln(1+\gamma)}{\sqrt{\gamma(\gamma+2)}} \\
&> \frac{\gamma+1}{\gamma(\gamma+2)} \ln(1+\gamma) \\
&> \frac{1}{\gamma+1}.
\end{aligned} \tag{24}$$

The last step holds when $\gamma > \gamma_0$, where γ_0 is the solution of the function $\ln(1 + \gamma) = \frac{\gamma(\gamma+2)}{(\gamma+1)^2}$ and $\gamma_0 \approx 1.22$.

Now, we proved that $\zeta(\gamma)$ decreases when $\gamma > \gamma_0$. Thus, $\zeta(\gamma_1) < \zeta(\gamma_2)$, which leads to:

$$\exp\left(\frac{\mathcal{G}_2^2 - \mathcal{G}_1^2}{2}\right) < \frac{\ln(1+\gamma_2)}{\ln(1+\gamma_1)}. \tag{25}$$

Combining Equations (21), (22), and (25), we can assert that Equation (20) holds, which gives us $\chi'(n) < 0$. Now, let $n_1 > n_2$. We have $\chi(n_1) < \chi(n_2)$, i.e.,

$$\varphi(n_1, \gamma_2) + \varphi(n_2, \gamma_1) < \varphi(n_1, \gamma_1) + \varphi(n_1, \gamma_1), \tag{26}$$

in which $\gamma_1 > \gamma_2$. According to Equation (26), it is sufficient to show that when n_ι and γ_ι are in reverse order, the error probability is smaller, which completes the proof. $\square$

We consider the condition $\ln(1 + \gamma_\iota) - \frac{k \ln 2}{n_\iota} > \max\left\{\frac{1}{k \ln 2}, \frac{1}{n_\iota}\right\}$ in Theorem 1 to ensure the error probability is within a reasonable range. Noticing Equation (6) holds when $\epsilon_\iota \approx 0$, we force channel capacity C_ι to be relatively larger than the coding rate $\frac{k}{n_\iota}$. If $C_\iota - \frac{k}{n_\iota} \approx 0$, we obtain $\epsilon_\iota \approx 0.5$ from Equation (5). For those ϵ_ι being rather large, we can simply remove relay R_ι, which does not affect the overall error probability. (Thus, we consider the condition $\ln(1 + \gamma_\iota) - \frac{k \ln 2}{n_\iota} > \max\left\{\frac{1}{k \ln 2}, \frac{1}{n_\iota}\right\}$ in the rest of the paper.)

By this means, we further shed some light on allocating the blocklength and power to the relays in Theorem 1. When the overall error probability is minimized, the blocklength allocated to a relay should be a decreasing function of its transmission power. According to Theorem 1, we present the following corollary.

Corollary 1. *Problem (18) is equivalent to the following optimization problem:*

$$\min_{\gamma_\iota, n_\iota} \epsilon = \sum_{\iota=1}^{L} Q(\mathcal{G}(\gamma_\iota, n_\iota)) \tag{27a}$$

$$s.t. \sum_{\iota=1}^{L} n_\iota \gamma_\iota = E^{th}, \tag{27b}$$

$$\sum_{\iota=1}^{L} n_\iota = N^{th}, \tag{27c}$$

$$n_1 \geq n_2 \geq \cdots \geq n_L \geq 1, \tag{27d}$$

$$0 \leq \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_L. \tag{27e}$$

Proof of Corollary 1. We start the proof by observing that each pair (n_ι, γ_ι) is independent in Problem (18) of the others. If we exchange the subscript of (n_i, γ_i) and (n_j, γ_j) for $i \neq j$, the objective ϵ does not change. Thus, we can assume $n_1 \geq n_2 \geq \cdots \geq n_L$. Then, from Theorem 1, we know that Problem (18) achieves the minimum value when $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_L$.

We denote by ϵ_1^* and ϵ_2^* the optimal value of Problem (18) and Problem (27), respectively. Because the domain of Problem (27) is a part of the domain of Problem (18), $\epsilon_1^* \leq \epsilon_2^*$. Moreover, ϵ_1^* is achieved when $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_L$ from Theorem 1, which satisfies (27d) and (27e). Therefore, the optimal solution of Problem (18) falls into the domain of Problem (27). This gives $\epsilon_2^* \leq \epsilon_1^*$. Thus, $\epsilon_1^* = \epsilon_2^*$, which completes the proof. □

3.3. An Approximate, but Analytical Solution

Finally, we formulate an optimization problem with the overall SNR and blocklength constraints. To obtain an analytical solution to the optimal error probability, we first handle the nonlinear constraint in Equation (27b) by approximation. Then, we add a few conditions to simplify the proof.

We present the following lemma to approximate the overall power in Equation (29).

Lemma 1 (Rearrangement inequality [25]). *For $L \in \mathbb{N}_+$ and two sequences satisfying $a_1 \geq a_2 \geq \cdots a_L$ and $b_1 \geq b_2 \geq \cdots b_L$,*

$$\sum_{\iota=1}^{L} a_\iota b_{L+1-\iota} \leq \frac{1}{L} \left(\sum_{\iota=1}^{L} a_\iota \right) \left(\sum_{\iota=1}^{L} b_\iota \right) \leq \sum_{\iota=1}^{L} a_\iota b_\iota. \tag{28}$$

Equations (27b) and (27c) suggest that the sum of γ_ι can not exceed a certain value. Applying Lemma 1, we present the approximation for the overall SNR (Actually, the SNR is proportional to the transmission power in our model. Therefore, we also consider Equation (28) as the power constraint in the following text.) in Equation (29).

$$\sum_{\iota=1}^{L} \gamma_\iota \approx \frac{LE^{th}}{N^{th}}. \tag{29}$$

Thus, we estimate the overall power by Equation (29). Finally, we obtain the optimization problem (30).

$$\min_{\gamma_\iota, n_\iota} \epsilon = \sum_{\iota=1}^{L} Q(\mathcal{G}(\gamma_\iota, n_\iota)) \tag{30a}$$

$$\text{s.t.} \ \sum_{\iota=1}^{L} \gamma_\iota = \frac{LE^{th}}{N^{th}}, \tag{30b}$$

$$\sum_{\iota=1}^{L} n_\iota = N^{th}, \tag{30c}$$

$$\max\left\{ \frac{1}{k \ln 2}, \frac{1}{n_\iota} \right\} < \ln(1 + \gamma_\iota) - \frac{k \ln 2}{n_\iota}, \tag{30d}$$

$$n_1 \geq n_2 \geq \cdots \geq n_L \geq 0, \tag{30e}$$

$$0 \leq \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_L. \tag{30f}$$

In order to solve Problem (30), we have Theorem 2 for the convexity of $Q(\mathcal{G}(\gamma, n))$.

Theorem 2. *The function $Q(\mathcal{G}(\gamma, n))$ is a binary convex function for the pair (γ, n) in the domain of Problem (30).*

Proof of Theorem 2. For a function $f(x, y)$, we use the abbreviations $f'_x = \frac{\partial f}{\partial x}$, $f'_y = \frac{\partial f}{\partial y}$, $f''_{xy} = \frac{\partial^2 f}{\partial x \partial y}$, $f''_x = \frac{\partial^2 f}{\partial x^2}$, and $f''_y = \frac{\partial^2 f}{\partial y^2}$.

We focus on the Hessian matrix of the function $\mathcal{H} = Q(\mathcal{G}(\gamma, n))$, which is given by $\nabla^2 \mathcal{H} = \begin{bmatrix} \mathcal{H}''_\gamma & \mathcal{H}''_{\gamma n} \\ \mathcal{H}''_{n\gamma} & \mathcal{H}''_n \end{bmatrix}$. The main idea of the proof is to show that the Hessian matrix is positive definite. Actually, we calculate the partial derivatives of $\mathcal{H}$ as follows:

$$\mathcal{H}''_n = (\mathcal{G}(\mathcal{G}'_n)^2 - \mathcal{G}''_n) \exp\left(-\frac{\mathcal{G}^2}{2}\right). \tag{31}$$

Similarly, we obtain:

$$\mathcal{H}''_\gamma = (\mathcal{G}(\mathcal{G}'_\gamma)^2 - \mathcal{G}''_\gamma) \exp\left(-\frac{\mathcal{G}^2}{2}\right), \tag{32}$$

and:

$$\mathcal{H}''_{n\gamma} = \mathcal{H}''_{\gamma n} = (\mathcal{G}\mathcal{G}'_n \mathcal{G}'_\gamma - \mathcal{G}''_{n\gamma}) \exp\left(-\frac{\mathcal{G}^2}{2}\right). \tag{33}$$

From Appendix A, we have $\mathcal{G}''_n < 0$ and $\mathcal{G}''_\gamma < 0$. Substituting the negative condition property into Equation (32), we have:

$$\begin{aligned} \mathcal{H}''_\gamma &= \left((\mathcal{G}'_\gamma)^2 \mathcal{G} - \mathcal{G}''_\gamma\right) \exp\left(-\frac{\mathcal{G}^2}{2}\right) \\ &> (\mathcal{G}'_\gamma)^2 \mathcal{G} \exp\left(-\frac{\mathcal{G}^2}{2}\right) \\ &> 0. \end{aligned} \tag{34}$$

Similarly, $\mathcal{H}''_n > 0$.

To deal with the cross partial derivative $\mathcal{H}''_{n\gamma}$, we calculate the factor in the first round brackets in Equation (33) as:

$$\mathcal{G}\mathcal{G}'_n\mathcal{G}'_\gamma - \mathcal{G}''_{n\gamma} = n^{-\frac{1}{2}} [\gamma(\gamma+2)]^{-\frac{3}{2}} \left\{ \left[\frac{\gamma(\gamma+2) - \left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right)}{2} \right] \times \right.$$
$$\left. \left[n \left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right) \left(\ln(1+\gamma) + \frac{k\ln 2}{n} \right) - 1 \right] - \frac{k\ln 2}{n} \right\}. \tag{35}$$

Since $\ln(1+\gamma) > \frac{k\ln 2}{n}$, we have:

$$\left(n \left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right) \left(\ln(1+\gamma) + \frac{k\ln 2}{n} \right) - 1 \right) > \left(2n \left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right) \frac{k\ln 2}{n} - 1 \right)$$
$$= 2k\ln 2 \left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right) - 1$$
$$> 1, \tag{36}$$

which applies the condition of (30d). Thus,

$$\mathcal{G}\mathcal{G}'_n\mathcal{G}'_\gamma - \mathcal{G}''_{n\gamma} > n^{-\frac{1}{2}} [\gamma(\gamma+2)]^{-\frac{3}{2}} \left[\frac{\gamma(\gamma+2) - \ln(1+\gamma) - \frac{k\ln 2}{n}}{2} \right]$$
$$> n^{-\frac{1}{2}} (\gamma(\gamma+2))^{-\frac{3}{2}} \left[\frac{\gamma(\gamma+2) - 2\ln(1+\gamma)}{2} \right] \tag{37}$$
$$> 0.$$

Since $\mathcal{G}\mathcal{G}'_n\mathcal{G}'_\gamma - \mathcal{G}''_{n\gamma} > 0$ and $\mathcal{G}''_{n\gamma} < 0$, from Appendix A, we obtain:

$$\mathcal{H}''_{n\gamma} \mathcal{H}''_{\gamma n} = (\mathcal{G}\mathcal{G}'_n\mathcal{G}'_\gamma - \mathcal{G}''_{n\gamma})^2 \exp\left(-\mathcal{G}^2 \right)$$
$$< (\mathcal{G}\mathcal{G}'_n\mathcal{G}'_\gamma)^2 \exp\left(-\mathcal{G}^2 \right) \tag{38}$$
$$< \mathcal{H}''_n \mathcal{H}''_\gamma.$$

This yields that $\nabla^2 \mathcal{H}$ is positive definite. The proof is complete. $\square$

Theorem 2 proves the convexity of the objective function mathematically. Furthermore, we have the optimal solution of Problem (30) as follows.

Theorem 3. *The optimal solution of Problem (30) is* $LQ\left(\mathcal{G}\left(\frac{E^{th}}{N^{th}}, \frac{N^{th}}{L} \right) \right)$*, when we set* $n_\iota = \frac{N^{th}}{L}$ *and* $\gamma_\iota = \frac{E^{th}}{N^{th}}$ *for each* $\iota = 1, \dots, L$.

Proof of Theorem 3. From Theorem 2, we have that $Q(\mathcal{G}(\gamma, n))$ is a convex function. Applying Jensen's inequality, we have:

$$\epsilon \geq LQ\left(\mathcal{G}\left(\frac{\sum_{\iota=1}^L \gamma_\iota}{L}, \frac{\sum_{\iota=1}^L n_\iota}{L} \right) \right) = LQ\left(\mathcal{G}\left(\frac{E^{th}}{N^{th}}, \frac{N^{th}}{L} \right) \right). \tag{39}$$

The proof is complete. $\square$

Theorem 3 states that the system achieves the highest reliability when the transmission power and blocklength of the relays are the same. Furthermore, we obtain an analytical expression of the optimal error probability. This relay resource allocation strategy is intuitively reasonable, corresponding to the equal distance between each successive relay pair. Finally, from the discussion above, we generate a suboptimal result of the initial problem.

4. Asymptotic Performance Analysis for Long-Distance Short-Packet Transmission

In this section, we focus on the solution given by Theorem 3, i.e., $\epsilon = LQ\left(\mathcal{G}\left(\frac{E^{th}}{N^{th}}, \frac{N^{th}}{L}\right)\right)$. Let $\bar{\gamma}$ and $\bar{n}$ denote $\frac{E^{th}}{N^{th}}$ and $\frac{N^{th}}{L}$, respectively. To give more insight, we develop an asymptotic analysis of the error probability for large $\bar{n}$ or $\bar{\gamma}$.

To start with, we present Theorem 4 to approximate the overall error probability ϵ when $\mathcal{G}\left(\frac{E^{th}}{N^{th}}, \frac{N^{th}}{L}\right) \to \infty$.

Theorem 4. *If E^{th} or N^{th} is large enough, the overall error probability is approximated by:*

$$\epsilon \approx \frac{L^{\frac{3}{2}}\sqrt{\frac{E^{th}}{N^{th}}\left(\frac{E^{th}}{N^{th}}+2\right)}\exp\left(-\frac{N^{th}\left(\ln\left(1+\frac{E^{th}}{N^{th}}\right)-\frac{Lk\ln 2}{N^{th}}\right)^2\left(1+\frac{E^{th}}{N^{th}}\right)^2}{2L\frac{E^{th}}{N^{th}}\left(\frac{E^{th}}{N^{th}}+2\right)}\right)}{\sqrt{2\pi N^{th}}\left(\ln\left(1+\frac{E^{th}}{N^{th}}\right)-\frac{Lk\ln 2}{N^{th}}\right)\left(1+\frac{E^{th}}{N^{th}}\right)}. \tag{40}$$

Proof. We notice that for any $x > 0$, the following inequation holds:

$$\frac{x}{1+x^2}\exp\left(-\frac{x^2}{2}\right) < \int_x^{+\infty}\exp\left(-\frac{u^2}{2}\right)du < \frac{1}{x}\exp\left(-\frac{x^2}{2}\right). \tag{41}$$

Substitute $x = \mathcal{G}\left(\frac{E^{th}}{N^{th}}, \frac{N^{th}}{L}\right)$ into Equation (41), and the proof is completed. $\square$

From Theorem 4, we also express ϵ with $\bar{\gamma}$ and $\bar{n}$ as:

$$\epsilon \approx \frac{L\sqrt{\bar{\gamma}(\bar{\gamma}+2)}\exp\left(-\frac{\bar{n}\left(\ln(1+\bar{\gamma})-\frac{k\ln 2}{\bar{n}}\right)^2(1+\bar{\gamma})^2}{2\bar{\gamma}(\bar{\gamma}+2)}\right)}{\sqrt{2\pi\bar{n}}\left(\ln(1+\bar{\gamma})-\frac{k\ln 2}{\bar{n}}\right)(1+\bar{\gamma})}, \tag{42}$$

in which we recall that $\bar{n} = \frac{N^{th}}{L}$ and $\bar{\gamma} = \frac{E^{th}}{N^{th}}$.

4.1. Asymptotic Analysis with a Fixed Number of Relays

In this part, we assume that the number of relays is fixed as a constant L. When $\bar{n}$ or $\bar{\gamma}$ is large, we have $\left(\ln(1+\bar{\gamma})-\frac{k\ln 2}{n}\right) \approx (\ln(1+\bar{\gamma})$ and $\frac{(1+\bar{\gamma})^2}{\bar{\gamma}(\bar{\gamma}+1)} \approx 1$. Thus, we approximate ϵ in Equation (43) by applying Theorem 4.

$$\epsilon \approx \frac{L\exp\left(-\frac{\bar{n}\ln^2(1+\bar{\gamma})}{2}\right)}{\sqrt{2\pi\bar{n}}\ln(1+\bar{\gamma})}. \tag{43}$$

For a large $\bar{n}$, we present ϵ as a function of $\bar{n}$, i.e.,

$$\epsilon \approx \frac{\mu_1\exp(-\theta_1 n)}{\sqrt{n}}, \tag{44}$$

where $\mu_1 = \frac{L}{\sqrt{2\pi}\ln(1+\bar{\gamma})}$ and $\theta_1 = -\frac{\ln^2(1+\bar{\gamma})}{2}$. In this case, as:

$$\begin{aligned}\bar{n} &= \frac{N^{th}}{L} \\ &= \frac{T-\frac{D}{c}}{L\tau} \to \infty,\end{aligned} \tag{45}$$

the whole problem degenerates into an infinite-blocklength problem. The well-known Shannon formula can be applied to assure $\epsilon \to 0$ with a given data rate. Moreover, from (44), we can see that ϵ decreases nearly exponentially with n.

For a large $\bar{\gamma}$, we present ϵ as a function of $\bar{\gamma}$, i.e.,

$$\epsilon \approx \frac{\mu_2 \bar{\gamma}^{-\theta_2 \ln(\bar{\gamma})}}{\ln(\bar{\gamma})}, \tag{46}$$

where $\mu_2 = \frac{L}{\sqrt{2\pi\bar{n}}}$ and $\theta_2 = -\frac{\bar{n}}{2}$. From Equation (46), we can see that ϵ decreases nearly exponentially with $\log(\bar{\gamma})$, which is an infinitesimal of higher order compared to any fractional polynomials, but a low-order infinitesimal compared to $\exp(-\bar{\gamma})$.

4.2. Analysis of Relay Gains

In this part, we express error probability ϵ as a function of the number of relays L. In this case, the maximal normalized energy E^{th} and the overall blocklength N^{th} are a function of L. For instance, in practice, a longer distance means a more severe path loss, and in those cases, we should allocate more energy. We denote by $E^{th}(L)$ the maximal normalized energy and by $N^{th}(L)$ the overall blocklength, respectively. Certain types of $E^{th}(L)$ and $N^{th}(L)$ are now discussed. Define $E = E^{th}(1)$ and $N = N^{th}(1)$. We still use the same assumption in the last part to approximate ϵ.

First, we let $N^{th}(L) = N$ and $E^{th}(L) = E$. The constraints do not change regardless of the number of relays. From Equation (40), we express ϵ in terms of the number of relays L, i.e.,

$$\epsilon \approx \mu_3 L^{\frac{3}{2}} \exp\left(-\frac{\theta_3}{L}\right), \tag{47}$$

where $\mu_3 = \frac{1}{\sqrt{2\pi N} \ln\left(1+\frac{E}{N}\right)}$ and $\theta_3 = \frac{N \ln\left(1+\frac{E}{N}\right)}{2}$.

From Equation (47), the error probability increases with L. This is because both communication resources, i.e., the blocklength and transmission power, are fixed. The resources allocated to each relay decreases with L. Thus, the error probability of each hop increases with L.

Next, we let $E^{th}(L)$ increase linearly with L, i.e., $E^{th}(L) = LE$. Moreover, we let $N^{th}(L) = N$ be a fixed value. Under this assumption, we have:

$$\epsilon \approx \frac{L^{\frac{3}{2}} \exp\left(-N \frac{\ln^2\left(1+\frac{LE}{N}\right)}{2L}\right)}{\sqrt{2\pi N} \ln\left(1+\frac{LE}{N}\right)}. \tag{48}$$

We denote by $\epsilon = \mathcal{F}(L)$ the function of ϵ and L in Equation (48). By solving $\mathcal{F}'(L) = 0$, there is a local minimum point L_0. (Actually, L_0 may not be an integer. However, we do not use this fact in any essential way.) $\mathcal{F}(L) \geq \mathcal{F}(L_0)$ holds for all L. Thus, there are an optimal number of relays $L = L_0$ to minimize the error probability ϵ. When the number of relays L increases substantially, the error probability goes to infinity because there is a limited blocklength to be allocated. The discussion enlightens us that there is an optimal number of relays to choose in practice.

Then, we let $E^{th}(L) = LE$ and $N^{th}(L) = LN$. In this case, we assume that the maximal tolerable latency is proportional to L. It is a practical scenario when the total distance increases linearly with L. Under this consumption, we have Equation (49), i.e.,

$$\epsilon \approx \frac{L \exp\left(-N \frac{\ln^2\left(1+\frac{E}{N}\right)}{2}\right)}{\sqrt{2\pi N} \ln\left(1+\frac{E}{N}\right)}, \tag{49}$$

In this case, ϵ increases linearly with L.

Last but not least, we let $E^{th}(L) = L^{\alpha}E$ and $N^{th}(L) = N$. In this case, we increase the number of relays in a fixed total distance D from the source to the destination. The distance between each node is given by $\frac{D}{L}$. Thus, the received SNR of each relay is proportional to L^{α}, i.e., $E^{th}(L) = L^{\alpha}E$. The error probability is given by:

$$\epsilon \approx \frac{L^{\frac{3}{2}} \exp\left(-N\frac{\ln^2(1+\frac{L^{\alpha}E}{N})}{2L}\right)}{\sqrt{2\pi N}\ln(1+\frac{L^{\alpha}E}{N})}. \tag{50}$$

In Equation (50), there is also an optimal L to minimize ϵ.

To summarize, we investigate four cases in this part as L changes in Equations (47)–(50). More generally, we let $E^{th}(L)$, $N^{th}(L)$ be any function of L. We have Equation (51) for the general case.

$$\epsilon \approx \frac{L^{\frac{3}{2}}\sqrt{\frac{E^{th}(L)}{N^{th}(L)}\left(\frac{E^{th}(L)}{N^{th}(L)}+2\right)}\exp\left(-\frac{N^{th}(L)\left(\ln\left(1+\frac{E^{th}(L)}{N^{th}(L)}\right)-\frac{Lk\ln 2}{N^{th}(L)}\right)^2\left(1+\frac{E^{th}(L)}{N^{th}(L)}\right)^2}{2L\frac{E^{th}(L)}{N^{th}(L)}\left(\frac{E^{th}(L)}{N^{th}(L)}+2\right)}\right)}{\sqrt{2\pi N^{th}(L)}\left(\ln\left(1+\frac{E^{th}(L)}{N^{th}(L)}\right)-\frac{Lk\ln 2}{N^{th}(L)}\right)\left(1+\frac{E^{th}(L)}{N^{th}(L)}\right)}. \tag{51}$$

4.3. The Blocklength–Power Tradeoff

In this part, we determine the blocklength and power tradeoff with a given maximal error probability $\epsilon^{\max}$.

Applying Theorem 3 and $\frac{(\bar{\gamma}+1)^2}{\bar{\gamma}(\bar{\gamma}+2)} \approx 1$, we have:

$$Q^{-1}\left(\frac{\epsilon^{\max}}{L}\right) \approx \sqrt{\bar{n}}\left(\ln(1+\bar{\gamma}) - \frac{k\ln 2}{\bar{n}}\right). \tag{52}$$

When the number of relays L is fixed, we can see that the average blocklength $\bar{n}$ should decrease with the average transmission power $\bar{\gamma}$. This result corresponds to the resource allocation property in Theorem 1.

When the number of relays L is not fixed, we can substitute the functions $E^{th}(L)$ and $N^{th}(L)$ into Equation (52). For example, since:

$$\bar{\gamma} = \frac{E^{th}(L)}{N^{th}(L)}$$
$$= \frac{E^{th}(L)}{L\bar{n}}, \tag{53}$$

we have:

$$Q^{-1}\left(\frac{\epsilon^{\max}}{L}\right) \approx \sqrt{\bar{n}}\left(\ln\left(1+\frac{E^{th}(L)}{L\bar{n}}\right) - \frac{k\ln 2}{\bar{n}}\right). \tag{54}$$

Similarly, we have:

$$Q^{-1}\left(\frac{\epsilon^{\max}}{L}\right) \approx \sqrt{\frac{E^{th}(L)}{L\bar{\gamma}}}\left(\ln(1+\bar{\gamma}) - \frac{Lk\bar{\gamma}\ln 2}{E^{th}(L)}\right). \tag{55}$$

Equations (54) and (55) indicate a tradeoff between $\bar{n}$ and $\bar{\gamma}$. Other types of $E^{th}(L)$ and $N^{th}(L)$ are also applicable. However, this topic exceeds the scope of this paper.

More importantly, it is worth pointing out that Equation (52) reveals whether a transmission is possible by increasing the number of relays with the given error probability, overall energy, and latency requirements.

5. Simulation Results

In this section, numerical and simulation results are presented to validate the theoretical analysis and demonstrate the advantages of our work.

First, Figures 3 and 4 present the error probability with respect to the overall blocklength and the overall energy, respectively. The two curves are drawn for the loss exponent $\alpha = 2$ and $\alpha = 2.2$. We consider the transmission of a one-hundred-bit data packet with the help of two relays, i.e., $L = 3$. The distance from the source to the destination is $D = 6$. The normalized unit distance is $d_0 = 1$. In Figure 3, the overall normalized energy E^{th} is set as 4200 and the overall blocklength N^{th} ranges from 200 to 500. In Figure 4, the overall blocklength N^{th} is set as 450 and the overall normalized energy E^{th} ranges from 2400 to 4800. In the subfigure of Figure 3, the relative error is less than 10^{-5}, which is mainly caused by the rounding function of $\bar{n} = [\frac{N^{th}}{L}]$. In Figure 4, the relative error is less than 10^{-4}, which is mainly caused by the approximation of ϵ in Equation (6). It is easy to see that the theoretical results obtained via Theorem 3 match well with their corresponding simulation results. Furthermore, we can see that a larger α leads to a higher error probability.

Next, we present the numerical result of the optimal solution in Figures 5 and 6. The number of relays was set as $L = 5$. Figure 5 presents the overall blocklength versus the error probability curves for three average transmission powers, namely $\bar{\gamma} = 1.5$, $\bar{\gamma} = 2$, and $\bar{\gamma} = 2.5$. Figure 6 presents the overall normalized energy versus error probability curves for three average blocklengths, namely $\bar{n} = 60$, $\bar{n} = 80$, and $\bar{n} = 100$.

In Figure 7, we present the optimal ϵ curves with respect to the number of relays L when different kinds of $E^{th}(L)$ and $N^{th}(L)$ are given. In most cases, the error probability increases with the number of relays L. However, under the circumstance where $E^{th}(L) = L^2E$, $N^{th}(L) = N$, we have an optimal $L = 2$.

Finally, Figure 8 presents the optimal distance–latency–reliability tradeoff. The parameters were set as follows: the number of relays $L = 3$, the overall normalized energy $E^{th} = 4150$, and the symbol duration $\tau = 2 \times 10^{-4}$.

We chose the total distance $D \in [300 \text{ km}, 600 \text{ km}]$. The normalized distance was unitized by a factor $c\tau$, where c is the speed of light. Figure 8a shows the reliability with respect to the distance and latency. Figure 8b–d presents the projection of the surface when the distance, latency, or reliability is fixed. Specifically, Figure 8b shows the reliability–latency tradeoff when the total distance is fixed as 360 km. Figure 8c presents the reliability–distance tradeoff when the latency is set as a constant of 10.14 ms. Figure 8d shows the latency–distance tradeoff when the error probability is 1×10^{-12}.
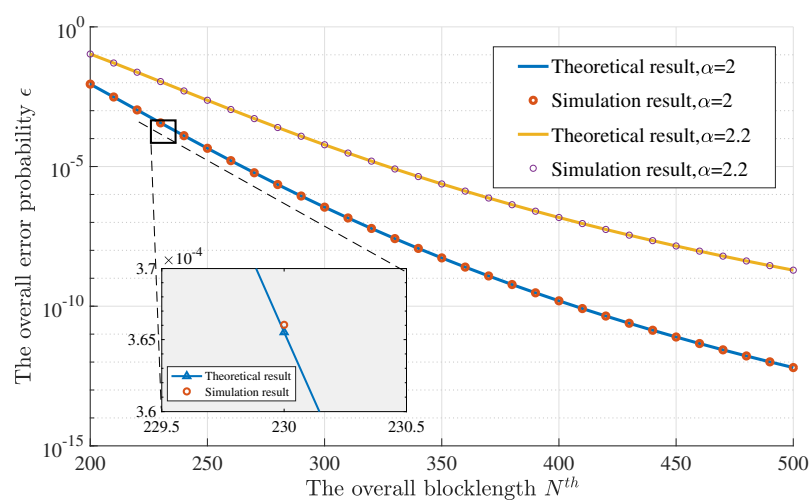

**Figure 3.** The numerical result of the optimal error probability–blocklength curve.
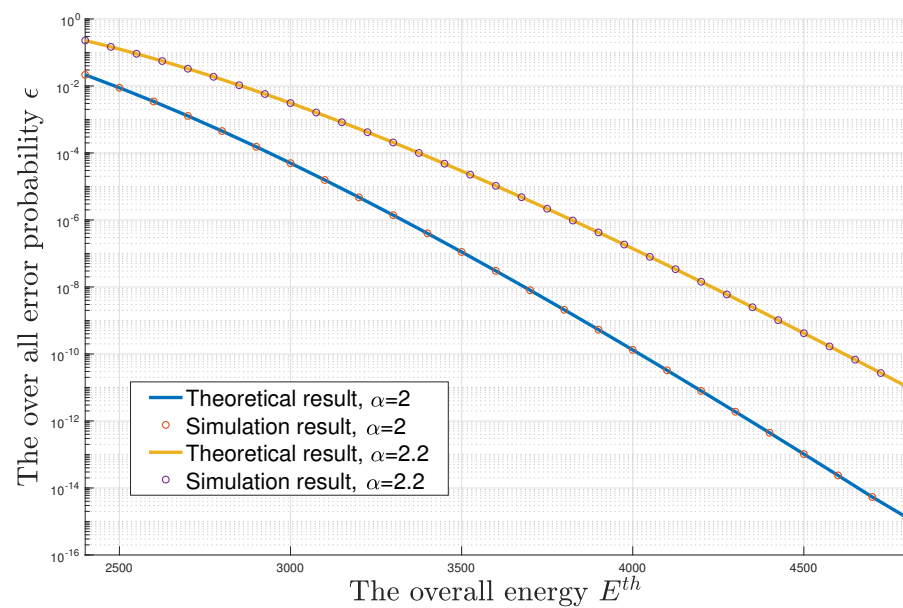
Figure 4. The numerical result of the optimal error probability–energy curve.

From the simulation results, we can draw some conclusions about the advantages of our model as follows. First, we consider both propagation delay and decoding delay. Our model was established to assess the performance of large-scale low-latency communication systems, which has not been included in any other previous works. We obtained a low-complexity solution for the error probability by giving an analytical solution of the relaxed problem. Furthermore, we attained the best resource allocation policy. Last but not least, we determined the theoretical limit of our model, under the condition of whether the number of relays is fixed or not. In particular, our result determines the minimum error probability of relaying communication given the latency, power allocation scheme, and the relay nodes' positions.
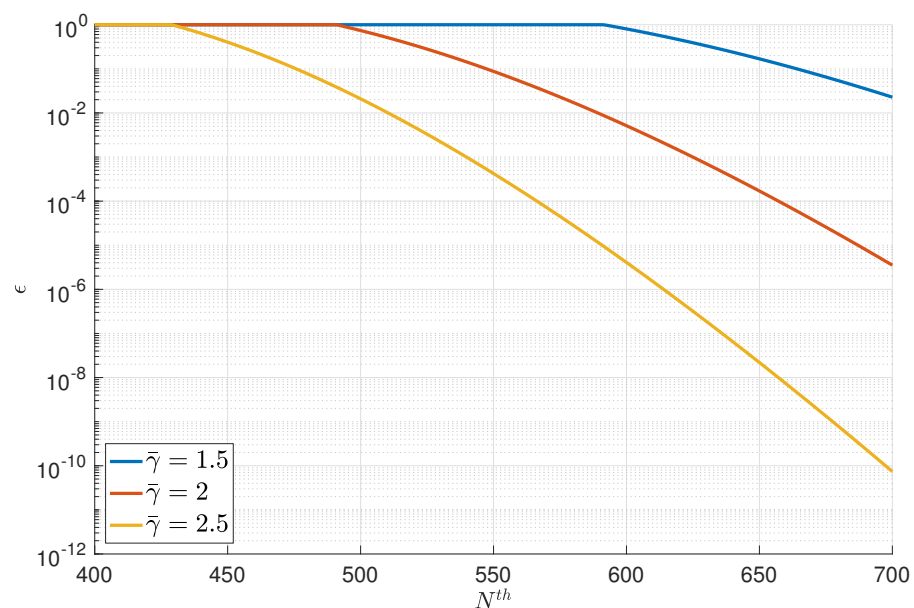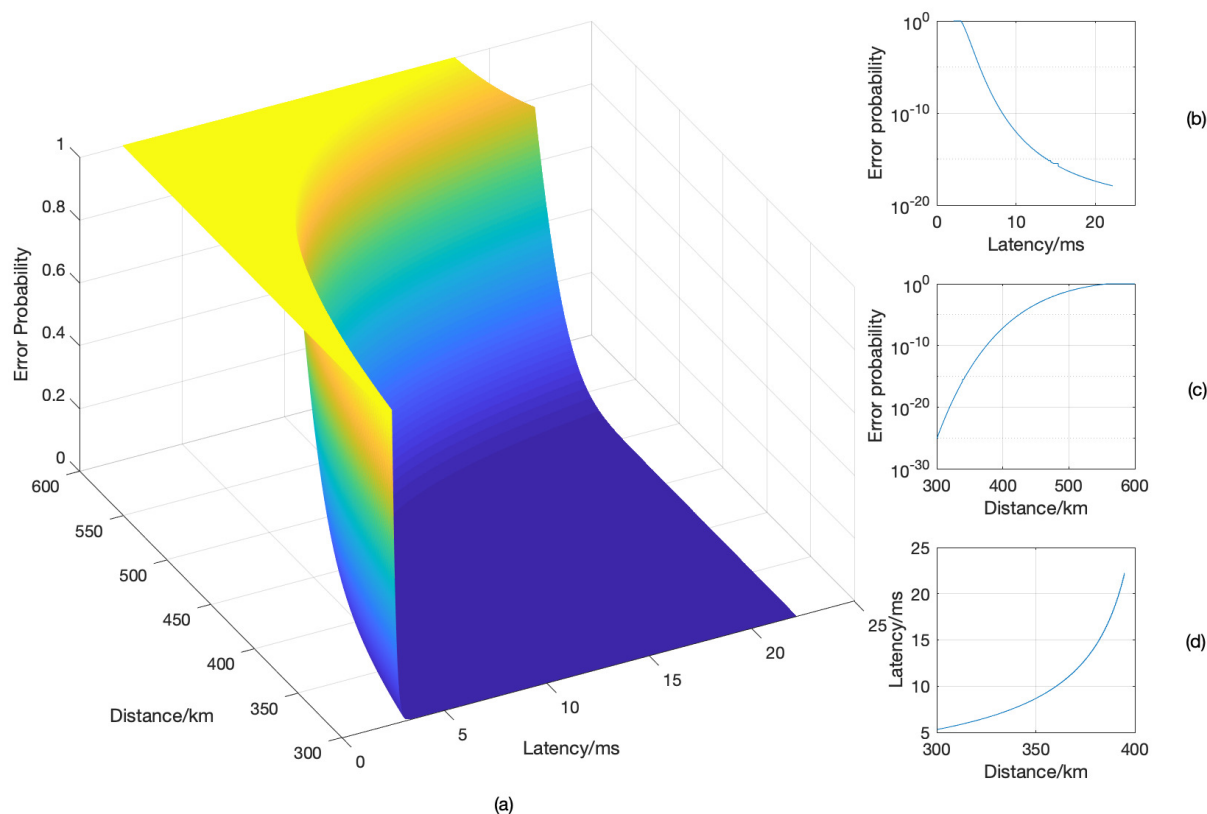

Figure 5. The overall blocklength versus error probability curves.
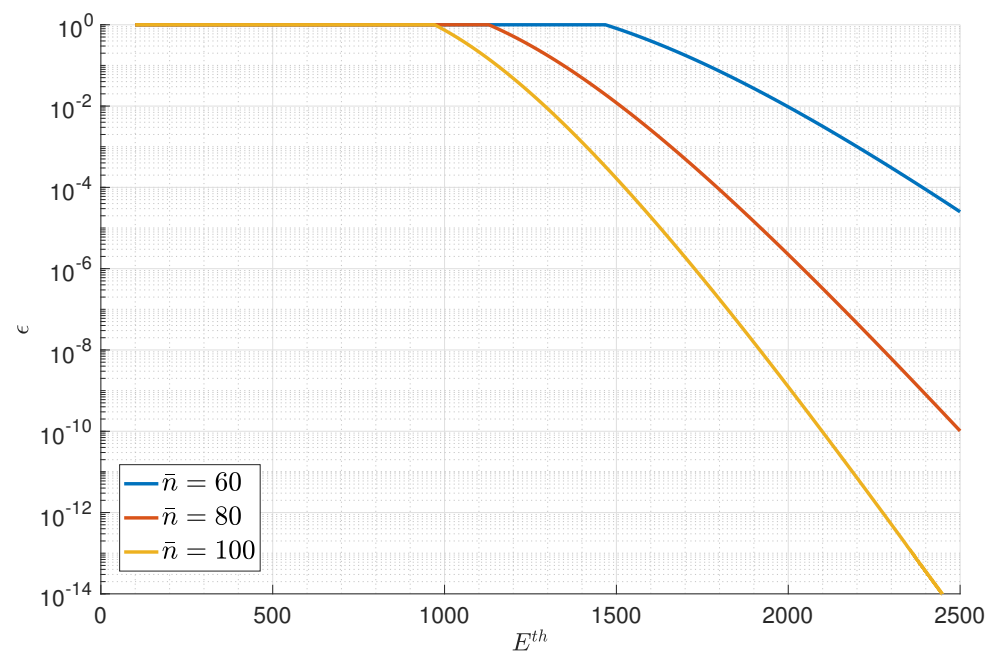
Figure 6. The overall normalized energy versus error probability curves.
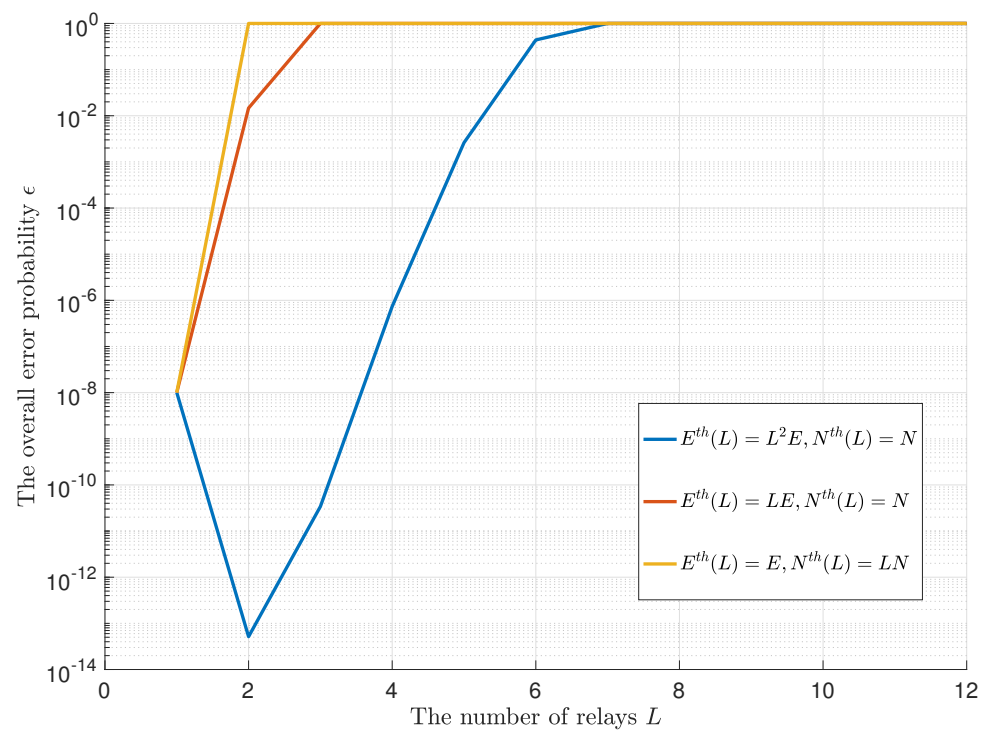
Figure 7. The optimal ϵ curves w.r.t the number of relays L.

Figure 8. The optimal distance–latency–reliability tradeoff.

6. Conclusions

In this paper, we studied the optimal policies and performance bounds for low-latency short-packet transmission over a large spatial scale, in which both the propagation loss of the SNR and the propagation delay should be taken into consideration. Although DF relays can be deployed to regenerate packets reliably and mitigate their SNR loss due to propagation, they induce a severe decoding delay. To address these issues, we optimized the relay deployment, power allocation, and blocklength for each relay. Moreover, we investigated the performance limits of the distant transmission of short packets via asymptotic analysis. Given the overall transmission power, we found the optimal distance–latency–reliability tradeoff. Our results may provide some engineering insights into the design of integrated LEO satellite communications in 6G. Important future work includes the study of amplify-and-forward short-packet relaying, multi-user short-packet transmission, and two-way short-packet relaying over a large spatial scale. Furthermore, we will consider the situation where the relay nodes are not at fixed positions. We will investigate whether it is possible to allocate the resources in real time based on the positions of the nodes.

Author Contributions: Conceptualization, W.C. and H.V.P.; formal analysis, L.H.; methodology, L.H. and X.Z.; software, L.H.; writing—original draft, L.H.; writing—review and editing, X.Z., W.C., and H.V.P. All authors read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Appendix A. Derivatives of Function $\mathcal{G}$

The partial derivatives of $\mathcal{G}(n, \gamma)$ are listed in this Appendix. Furthermore, we prove whether each of them is positive or negative.

The partial derivative of $\mathcal{G}$ w.r.t. n is:

$$\mathcal{G}'_n = \sqrt{\frac{(\gamma+1)^2}{\gamma(\gamma+2)}} \left(\frac{k\ln 2}{2n^{\frac{3}{2}}} + \frac{\ln(1+\gamma)}{2n^{\frac{1}{2}}} \right) \tag{A1}$$
$$> 0,$$

and the partial derivative of $\mathcal{G}$ w.r.t. γ is:

$$\mathcal{G}'_\gamma = \sqrt{n} \left[\gamma(\gamma+2) \right]^{-\frac{3}{2}} \left[\gamma(\gamma+2) - \left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right) \right] \tag{A2}$$
$$> 0.$$

This is because $\gamma(\gamma+2) - \left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right) > \gamma - \ln(1+\gamma) > 0$.

The second-order partial derivative of $\mathcal{G}$ w.r.t. n is given by:

$$\mathcal{G}''_n = -\sqrt{\frac{(\gamma+1)^2}{\gamma(\gamma+2)}} \left(\frac{3k\ln 2}{4n^{\frac{5}{2}}} + \frac{\ln(1+\gamma)}{4n^{\frac{3}{2}}} \right) \tag{A3}$$
$$< 0.$$

The second-order partial derivative of $\mathcal{G}$ w.r.t. γ is given by:

$$\mathcal{G}''_\gamma = -\sqrt{n} \left[\gamma(\gamma+2) \right]^{-\frac{3}{2}} \left[\gamma+1 + \frac{1}{\gamma+1} - \frac{3\left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right)(\gamma+1)}{\gamma(\gamma+2)} \right]$$
$$< -\sqrt{n} \left[\gamma(\gamma+2) \right]^{-\frac{3}{2}} \left[\gamma+1 - \frac{3\gamma(\gamma+1)}{\gamma(\gamma+2)} \right] \tag{A4}$$
$$< 0.$$

The first inequation holds because:

$$\gamma+1 + \frac{1}{\gamma+1} - \frac{3\left(\ln(1+\gamma) - \frac{k\ln 2}{n} \right)(\gamma+1)}{\gamma(\gamma+2)}$$
$$> \gamma+1 + \frac{1}{\gamma+1} - \frac{3\ln(1+\gamma)(\gamma+1)}{\gamma(\gamma+2)}$$
$$> \gamma+1 + \frac{1}{\gamma+1} - \frac{3\gamma+1}{\gamma+2} \tag{A5}$$
$$> \gamma+2 + \frac{4}{\gamma+2} - 4$$
$$> 0,$$

which applies the inequality of the arithmetic and geometric means (AM-GM inequality) in the last step.

Last but not least, we have:

$$
\begin{aligned}
\mathcal{G}''_{n\gamma} &= [n\gamma(\gamma+2)]^{-\frac{3}{2}} \left[\frac{n\gamma(\gamma+2)}{2} + k - \frac{n\ln(1+\gamma) - k\ln 2}{2} \right] \\
&> [n\gamma(\gamma+2)]^{-\frac{3}{2}} \left[\frac{n\gamma(\gamma+2)}{2} - \frac{n\ln(1+\gamma)}{2} \right] \\
&> [n\gamma(\gamma+2)]^{-\frac{3}{2}} \left[\frac{n\gamma(\gamma+2)}{2} - \frac{n\gamma}{2} \right] \\
&> 0.
\end{aligned}
\tag{A6}
$$

In conclusion, we have the expressions of the derivatives of $\mathcal{G}$ and proved whether each of the derivatives is positive or negative.

References

1. Letaief, K.B.; Chen, W.; Shi, Y.; Zhang, J.; Zhang, Y.J.A. The Roadmap to 6G: AI Empowered Wireless Networks. *IEEE Commun. Mag.* **2019**, *57*, 84–90. [CrossRef]
2. Andrews, J.G.; Buzzi, S.; Choi, W.; Hanly, S.V.; Lozano, A.; Soong, A.C.K.; Zhang, J.C. What Will 5G Be? *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1065–1082. [CrossRef]
3. Zhang, Z.; Xiao, Y.; Ma, Z.; Xiao, M.; Ding, Z.; Lei, X.; Karagiannidis, G.K.; Fan, P. 6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies. *IEEE Veh. Technol. Mag.* **2019**, *14*, 28–41. [CrossRef]
4. David, K.; Berndt, H. 6G Vision and Requirements: Is There Any Need for Beyond 5G? *IEEE Veh. Technol. Mag.* **2018**, *13*, 72–80. [CrossRef]
5. Giordani, M.; Polese, M.; Mezzavilla, M.; Rangan, S.; Zorzi, M. Toward 6G Networks: Use Cases and Technologies. *IEEE Commun. Mag.* **2020**, *58*, 55–61. [CrossRef]
6. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel Coding Rate in the finite-blocklength Regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [CrossRef]
7. Yang, W.; Durisi, G.; Koch, T.; Polyanskiy, Y. Quasi-Static Multiple-Antenna Fading Channels at finite-blocklength. *IEEE Trans. Inf. Theory* **2014**, *60*, 4232–4265. [CrossRef]
8. Durisi, G.; Koch, T.; Östman, J.; Polyanskiy, Y.; Yang, W. Short-Packet Communications Over Multiple-Antenna Rayleigh-Fading Channels. *IEEE Trans. Commun.* **2016**, *64*, 618–629. [CrossRef]
9. Yang, W.; Durisi, G.; Koch, T.; Polyanskiy, Y. Block-Fading Channels at finite-blocklength. In Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS), Ilmenau, Germany, 27–30 August 2013; pp. 1–4.
10. Östman, J.; Devassy, R.; Ferrante, G.C.; Durisi, G. Low-Latency Short-Packet Transmissions: Fixed Length or HARQ? In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–6. [CrossRef]
11. Ji, H.; Park, S.; Yeo, J.; Kim, Y.; Lee, J.; Shim, B. Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects. *IEEE Wirel. Commun.* **2018**, *25*, 124–130. [CrossRef]
12. Zhao, X.; Chen, W. Queue-Aware Variable-Length Coding for Ultra-Reliable and Low-Latency Communications. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6. [CrossRef]
13. She, C.; Yang, C.; Quek, T.Q.S. Cross-Layer Optimization for Ultra-Reliable and Low-Latency Radio Access Networks. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 127–141. [CrossRef]
14. Devassy, R.; Durisi, G.; Ferrante, G.C.; Simeone, O.; Uysal, E. Reliable Transmission of Short Packets Through Queues and Noisy Channels Under Latency and Peak-Age Violation Guarantees. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 721–734. [CrossRef]
15. Gui, G.; Liu, M.; Tang, F.; Kato, N.; Adachi, F. 6G: Opening New Horizons for Integration of Comfort, Security, and Intelligence. *IEEE Wirel. Commun.* **2020**, *27*, 126–132. [CrossRef]
16. Cover, T.; Gamal, A. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory* **1979**, *25*, 572–584. [CrossRef]
17. Laneman, J.; Tse, D.; Wornell, G. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Inf. Theory* **2004**, *50*, 3062–3080. [CrossRef]
18. Nabar, R.; Bolcskei, H.; Kneubuhler, F. Fading relay channels: Performance limits and space-time signal design. *IEEE J. Sel. Areas Commun.* **2004**, *22*, 1099–1109. [CrossRef]
19. Chen, W. CAO-SIR: Channel Aware Ordered Successive Relaying. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 6513–6527. [CrossRef]
20. Chen, W.; Dai, L.; Letaief, K.B.; Cao, Z. A Unified Cross-Layer Framework for Resource Allocation in Cooperative Networks. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 3000–3012. [CrossRef]
21. Zhang, X.; Zhu, Q.; Poor, H.V. Finite-Blocklength Performance of Relay-Networks over Nakagami-m Channels. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6. [CrossRef]
22. Du, F.; Hu, Y.; Qiu, L.; Schmeink, A. Finite-blocklength performance of multi-hop relaying networks. In Proceedings of the International Symposium on Wireless Communication Systems (ISWCS), Poznań, Poland, 20–23 September 2016; pp. 466–470. [CrossRef]

23. Lee, J.; Shin, H.; Lee, I.; Heo, J. Optimal Linear Multihop System for DF Relaying in a Poisson Field of Interferers. *IEEE Commun. Lett.* **2013**, *17*, 2029–2032. [CrossRef]
24. Erceg, V.; Greenstein, L.; Tjandra, S.; Parkoff, S.; Gupta, A.; Kulic, B.; Julius, A.; Bianchi, R. An empirically based path loss model for wireless channels in suburban environments. *IEEE J. Sel. Areas Commun.* **1999**, *17*, 1205–1211. [CrossRef]
25. Hardy, G.H.; Littlewood, J.E.; Pólya, G. *Inequalities*; Cambridge University Press: Cambridge, UK, 1952; p. 89.