Multi-Relay-Assisted Low-Latency High-Reliability Communications With Best Single Relay Selection

Yulin Hu, *Senior Member, IEEE*, Christopher Schnelling, M. Cenk Gursoy, *Senior Member, IEEE*, and Anke Schmeink, *Senior Member, IEEE*

***Abstract*—We study a multi-node Internet of Things system supporting low-latency high-reliability communication to a destination node. The rest of the nodes are potential relays in which the best single relay (BSR) is selected to assist the transmission to the destination. The system operates with finite blocklength (FBL) codes to satisfy the low-latency requirement. The scope of this work is to derive and improve the FBL performance of the considered BSR system. On the one hand, we extend Polyanskiy's FBL model of a single-hop scenario to the considered relaying system and derive the corresponding achievable reliability. On the other hand, by employing a practical FBL coding scheme, namely polar codes (PCs), an FBL performance bound attainable by a low-complexity coding scheme is presented. In particular, we provide a reliability bound of a dynamic-length PC scheme. Addressing a source-driven BSR strategy, as well as a relay-driven BSR strategy, we investigate two viable strategies for relay selection in the FBL regime, while the corresponding performance under an infinite blocklength (IBL) assumption serves as a reference. We prove that the two BSR strategies have the same performance in the IBL regime, while the relay-driven strategy is significantly more reliable than the source-driven one when considering the FBL regime. Furthermore, following the derived FBL performance model, we provide an optimal design to minimize the overall error probability via blocklength allocation. Through simulation and numerical investigations, we show the appropriateness of the proposed analytical model. Moreover, we evaluate both the achievable performance with FBLs and the performance of PCs in the considered scenarios while comparing the source-driven and relay-driven strategies.**

***Index Terms*—Decode-and-forward, finite blocklength regime, punctured polar codes, rate-compatible codes, relaying.**

I. Introduction

FUTURE wireless networks are expected to support high speed, low-latency and high reliability transmissions while connecting a massive number of smart devices, *e.g.*, enabling the Internet of Things (IoT) [2], [3]. Many envisioned IoT applications, such as industrial control applications, autonomous driving, cyber-physical systems, E-health, haptic feedback in virtual and augmented reality, smart grid, and remote surgery, will have stringent transmission latency and reliability requirements [4], [5] that cannot be met by existing wireless networks. The common features of these IoT applications are as follows: (i) the transmission reliability is usually a concern, (ii) due to low-latency constraints, the coding blocklengths for wireless transmissions are quite short, (iii) usually multiple IoT nodes are densely deployed.

It is known that cooperative relaying significantly promotes the transmission performance and greatly capitalizes on dense node packings [6]–[8]. Consequently, the performance of a network with multiple IoT nodes may be enhanced by relaying [9], [10], as each node may potentially act as a relay, *e.g.*, via BSR selection, assisting transmissions for peer nodes [11]–[13]. However, the above studies of relaying and the application of relaying in multi-node scenarios are conducted under the ideal assumption of communicating arbitrarily reliably at rates close to Shannons channel capacity. They thus implicitly assume an IBL regime, which does not allow for for the accurate assessment of the performance in latency-critical IoT scenarios operating with short blocklengths to satisfy the low-latency requirement.

In the FBL regime, the error probability in communication is not negligible due to the impact of the short blocklength. Early in 1962, Strassen has presented a normal approximation of the coding rate [14]. More recently, an achievable upper bound on the coding rate is identified in [15] for a single-hop transmission system, taking the error probability into account. The result of [15] has been extended to Gilbert-Elliott Channels [16], as well as to quasi-static fading channels [17], [18]. Recently, an achievable FBL performance for relaying under a single relay scenario was addressed analytically in [19]–[22]. Furthermore, bounds when using single relays with practical codes, *e.g.*, polar codes (PCs), are discussed in [23]. In fact, PCs have been considered in several relaying scenarios. In [24], PCs are employed to devise coding schemes for decode-and-forward (DF), as well as compress-and-forward relaying scenarios assuming relay channels with orthogonal receivers. A DF relaying scenario without the assumption of orthogonal receivers is addressed in [25], by implementing a Markov block coding scheme proposed in [26]

Manuscript received November 22, 2018; revised March 1, 2019 and May 13, 2019; accepted May 26, 2019. Date of publication June 6, 2019; date of current version August 13, 2019. This work was supported by the DFG research under Grant SCHM 2643/17. This paper was presented in part at the IEEE International Symposium on Wireless Communication Systems (ISWCS), Bologna, Italy, August 2017 [1]. The review of this paper was coordinated by Dr. X. Huang. *(Corresponding author: Yulin Hu.)*

Y. Hu, C. Schnelling, and A. Schmeink are with the ISEK Research Group, RWTH Aachen University, D-52074 Aachen, Germany (e-mail: hu@umic.rwth-aachen.de; schnelling@ti.rwth-aachen.de; schmeink@umic.rwth-aachen.de).

M. C. Gursoy is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: mcgursoy@syr.edu).

Digital Object Identifier 10.1109/TVT.2019.2921253

using PCs. Inspired by [24], PCs are presented in an opportunistic cooperative DF relaying scenario in [27]. While these works exploit structural properties of PCs, *e.g.*, the subset property of the information bit index sets in case of degraded channels as in [25], the considered relay scenarios are, to the best of our knowledge, either two-hop situations assuming a single-relay, or, as in [27] or [28], a main link assisted by a two-hop side channel based on a single relay network.

To the best of our knowledge, for multi-node scenarios with BSR selection that facilitate latency-critical IoT networks, both the achievable reliability bound as well as practical bounds, *i.e.*, reliability achieved by specific coding systems, have not been studied. In particular, the reliability-oriented design via blocklength allocation for such scenarios is still open. In our recent work [1], we have addressed the throughput of a BSR network, without explicitly considering the reliability. In this paper, we extend the work in [1] to a BSR system supporting latency-critical transmissions in IoT networks, where the system is assumed to operate with FBL codes to satisfy the low-latency requirement. Our major focus is to derive the reliability performance and provide reliability-oriented blocklength allocation designs.

The contributions of this work are summarized as follows:

- We derive the achievable reliability performance bound of the considered BSR network based on Polyanskiy's FBL model.
- By distinguishing between a source-driven BSR selection strategy and a relay-driven BSR selection strategy, we investigate the performance in the FBL regime, while the corresponding performance under an IBL assumption is provided as a reference. More importantly, we prove that the above two BSR strategies have the same reliability performance in the IBL regime, while in the FBL regime the relay-driven strategy is significantly more reliable than the source-driven one.
- To support our theoretical findings, a practical FBL performance bound based on PCs is provided. In particular, the reliability bound of a dynamic-length polar code scheme is studied.
- Following the derived FBL performance model, we investigate the achievable reliabilities by applying efficient blocklength allocations for the network under both the source-driven and relay-driven BSR selection strategies. By applying an appropriate approximation, we address the convexity of the approximated problems.
- Via simulations, we show the appropriateness of our analytical model. Moreover, we evaluate the achievable reliability performance of the network with optimal blocklength allocation. It is observed that the gaps between the practical PCs and the FBL bounds depend on the decoder employed. More importantly, the results indicate that the optimal blocklength allocation based on the analytical bounds does provide reasonable design guidelines for the BSR network operating with a practical FBL coding scheme.

The remainder of the paper is organized as follows. Section II describes the system model and briefly introduces the theoretical background of the FBL regime, as well as the relevant aspects of the rate-compatible coding scheme based on punctured PCs used in this work. In Section III, we first derive the FBL performance of the system assuming static channels, and subsequently extend the performance model to a quasi-static channel fading scenario. Efficient blocklength allocations are proposed in Section IV for the network under both the source-driven and relay-driven BSR selection strategies. Section V presents our numerical results. Finally, we conclude our work in Section VI.

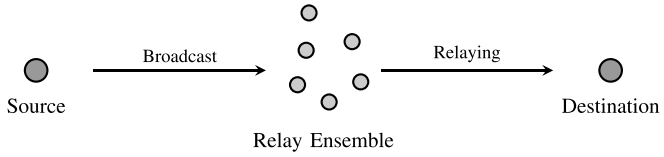


Fig. 1. Example of the considered IoT network with BSR selection.

II. Preliminaries

In this section, we first describe the system model and subsequently review the theoretical background of the FBL regime. Furthermore, we provide necessary details of PCs and the rate-compatible coding scheme based on punctured PCs employed in this work.

A. System Model

We consider an IoT network with a source node S, a destination node D, and other J IoT nodes as potential DF relays $\{j\}$, where $j \in \mathcal{J} = \{1, \ldots, J\}$. This scenario is schematically depicted in Fig. 1. In general, the distances among relays are significantly shorter than the distances between the source, the relay group and the destination. Data packets at the source are required to be transmitted to the destination via the relay group, while guaranteeing an end-to-end latency of S symbols. In addition, we denote by l the total overhead cost in symbols, *e.g.*, for the channel state information (CSI) acquisition and beacon transmission (*e.g.*, relay selection decision or acknowledgment). In particular, l is increasing in J and is assumed to be constant for a given J. Thus, each frame accommodates two phases of lengths m_1 and m_2, which are referred to as broadcasting phase and relaying phase. Hence, $m_1 + m_2 \leq M = S - l$ is necessary to satisfy the end-to-end latency requirement. In the broadcasting phase, the source sends a data block to the relays. After reception, if any relay decodes the block successfully, such relay will be able to forward the data to the destination.

We refer to such relays as *active relays*. In the subsequent relaying phase, one of the active relays, namely that with the best channel to the destination, will be selected to forward the data to the destination. We refer to this scenario as the BSR network[1].

[1] In comparison to letting all active relays forward the packet, the BSR strategy is more energy efficient, as with a given (total) transmit power constraint/limit at the relays, BSR transmits the data packet via the best channel and results in a higher signal-to-noise ratios (SNRs) than letting all active relay forward using the same transmit power. In addition, multiple relays forwarding likely makes the destination receive multiple unsynchronized signals and additional costs for the synchronization can be incurred. In this work, we focus on the scenario selecting only the BSR to assist the transmission.

Channels are assumed to be subject to block-fading with unit average gain. Hence, channels are constant within the duration of each transmission period but vary from one period to the next. We denote the instantaneous channel gains from the source to relay R_j and from R_j to the destination by $z_{\mathrm{S},j}$ and $z_{j,\mathrm{D}}$, $j \in \mathcal{J}$. In addition, the random channel gains are with unit mean, *i.e.*, $\mathbb{E}(z_{\mathrm{S},j}) = \mathbb{E}(z_{j,\mathrm{D}}) = 1$. The instantaneous SNRs from the source to Rj and from Rj to the destination are denoted by $\gamma_{\mathrm{S},j}$ and $\gamma_{j,\mathrm{D}}$. We consider a homogeneous scenario in which the average received SNRs of all the links from the source to relays are identical, denoted by $\bar{\gamma}_{\mathrm{S,R}}$, while the same applies to the average SNR of the links from relays to the destination, denoted by $\bar{\gamma}_{\mathrm{R,D}}$. Both the source as well as each of the relays is assumed to have access to perfect CSI. These assumptions allow us to study the fundamental performance of the considered multi-relay network, regardless of a specific network topology. Therefore, we have $\gamma_{\mathrm{S},j} = z_{\mathrm{S},j}\bar{\gamma}_{\mathrm{S,R}}$ and $\gamma_{j,\mathrm{D}} = z_{j,\mathrm{D}}\bar{\gamma}_{\mathrm{R,D}}$, $j \in \mathcal{J}$.

B. Blocklength-Limited Performance of a Single-Hop Transmission Scenario With Perfect CSI

For additive white Gaussian noise (AWGN) channels, the authors of [15] have derived a tight achievable bound for the coding rate of a single-hop transmission. With blocklength m, block error probability ϵ and SNR γ, the coding rate in bits per channel use is given by

$$\frac{1}{2}\log(1+\gamma) - \frac{\log e}{\gamma+1}\sqrt{\gamma\frac{\gamma+2}{2m}}Q^{-1}(\varepsilon) + \frac{O(\log m)}{m}, \quad (1)$$

where $Q^{-1}(\cdot)$ is the inverse of the Q-function given by

$$Q(w) = \frac{1}{\sqrt{2\pi}}\int_w^\infty e^{-t^2/2}dt. \quad (2)$$

In [29], [30], the above result has been extended to a complex channel model with received SNR γ, where the coding rate in bits per channel use is

$$r = \mathcal{R}(\gamma,\varepsilon,m) \approx \mathcal{C}(\gamma) - \sqrt{\frac{V}{m}}Q^{-1}(\epsilon), \quad (3)$$

where $\mathcal{C}(\gamma) = \log_2(1+\gamma)$ is the channel's Shannon capacity. Moreover,

$$V = \left(1 - \frac{1}{1+\gamma^2}\right)(\log_2 e)^2 \quad (4)$$

is the channel dispersion [15, Def.1]. Hence, for a single-hop transmission with blocklength m and coding rate r, the block error probability at the receiver is given by

$$\varepsilon = \mathcal{P}(\gamma,r,m) \approx Q\left(\sqrt{\frac{m}{V}}\big(\mathcal{C}(\gamma) - r\big)\right). \quad (5)$$

C. Polar Codes

Recently introduced by Arıkan [31], PCs answer a long-standing, open question in information theory by providing a practical coding scheme which provably achieves the symmetric capacity of any binary-input discrete memoryless channel (B-DMC). Relying on channel polarization, PCs are constructed explicitly targeting a specific design channel. To do so, we may select an index set $\mathcal{I} \subseteq \mathbb{N}$ for a PC of length $m = 2^n$, $n \in \mathbb{N}$, based on the probabilities of decision error under successive cancellation (SC) decoding

$$P_{\mathrm{e}}(\mathcal{W}_{m,i}) = \sum_{\mathbf{y}' \in \mathcal{Y}^m \times \mathcal{X}^{i-1}} \frac{1}{2}\min\left\{W_{m,i}(\mathbf{y}' \,|\, 0), W_{m,i}(\mathbf{y}' \,|\, 1)\right\}, 1 \le i \le m, \quad (6)$$

which estimates

$$\hat{u}_i = \underset{u_i \in \{0,1\}}{\arg\max}\left\{W_{m,i}\big(\mathbf{y}, \hat{\mathbf{u}}_1^{i-1} \,|\, u_i\big)\right\} \quad (7)$$

based on a received block $\mathbf{y}$ and assuming $i-1$ correctly estimated information bits $\hat{\mathbf{u}}_1^{i-1}$. In the decisions, $W_{m,i}$ denotes the channel statistic of the (virtual) channels $\mathcal{W}_{m,i}$,

$$\mathcal{W}_{m,i} : \{0,1\} \to \mathcal{Y}^m \times \{0,1\}^{i-1},\ 1 \le i \le m, \quad (8)$$

modelling these decisions, where $\mathcal{Y}$ denotes the output alphabet of the basic channel $\mathcal{W}$. As $m \to \infty$, these polarize in the sense that the fraction of decisions vanishes for which $P_{\mathrm{e}}(\mathcal{W}_{m,i})$ does not approach either 0 or $\frac{1}{2}$.

Encoding and decoding using SC decoding [31] is possible in quasi-linear complexity of $O(m \log m)$, where m denotes the blocklength. To increase the performance at short lengths, SC decoding has been extended to successive cancellation list (SCL) decoding [32], [33], which runs in $O(L \cdot m \log m)$ with list size L, *i.e.*, the number of candidate prefixes maintained in each step of the decoding process.

In their original formulation given in [31], the blocklength of a PC is restricted to integer powers of 2, and we have $m = 2^n$ for some $n \in \mathbb{N}$. However, as greater length flexibility is required in many communication scenarios, both puncturing and shortening of PCs have been considered to facilitate fine-grained length adaptions of PCs, *cf.* [34], [35], or [36].

In this work, we employ a rate-compatible coding scheme based on punctured PCs presented in [37] as an example of an FBL coding scheme. The rate-compatible code

$$\mathfrak{C} = \{\mathcal{C}_t : t \in [T]\} \quad (9)$$

is implemented as a family of T punctured PCs $\mathcal{C}_t$, such that the code $\mathcal{C}_t$ has length m_t and rate $r_t = \frac{K}{m_t}$, for a fixed and identical dimension K for all codes C_t, $t \in [T] := \{1,\ldots,T\}$. We assume $m_t < m_{t+1}$ and thus have $r_t > r_{t+1}$, $t \in [T-1]$.

Each code $\mathcal{C}_t$ is obtained by puncturing a mother PC of appropriate length. Given a mother PC $\mathcal{C}$ of length $m = 2^n$ and a puncturing pattern $\mathcal{P} \subset [m]$, a punctured PC $\mathcal{C}'$ of length $\tilde{m} = m - |\mathcal{P}|$ is given by

$$\mathcal{C}' = \{\mathbf{x}_{\mathcal{P}^{\mathrm{c}}} : \mathbf{x} \in \mathcal{C}\}, \quad (10)$$

that is, by all codewords $\mathbf{x} \in \mathcal{C}$, punctured at positions x_i, $i \in \mathcal{P}$. Here, we write $\mathcal{P}^{\mathrm{c}} := [m] \setminus \mathcal{P}$ to denote the complement of $\mathcal{P}$ with respect to $[m]$, and $\mathbf{x}_{\mathcal{S}} = (x_i)_{i \in \mathcal{S}}$ to denote a subvector of $\mathbf{x}$. As a result, the rate-compatible code $\mathfrak{C}$ may be defined by

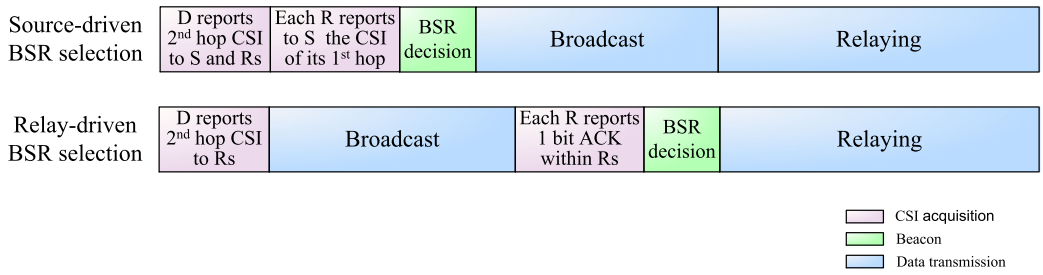


Fig. 2. Frame structure examples of BSR selection strategies: source-driven vs. relay-driven.

a sequence

$$\mathfrak{P} = \{\mathcal{P}_t \subset [m] : t \in [T]\} \tag{11}$$

of puncturing patterns and a mother PC of length m, such that $m_t = m - |\mathcal{P}_t|$. For the specific instance of $\mathfrak{C}$ considered in this work, we assume $\mathcal{P}_T = \emptyset$ and hence use the unpunctured mother code as $\mathcal{C}_T$.

Finally, we require all codes $\mathcal{C}_t$ to share the same index set $\mathcal{I}$, indexing non-frozen information positions, and we have $|\mathcal{I}| = K$. The indices forming $\mathcal{I}$ are chosen based on a density evolution (DE) [38].

III. The Achievable Reliability of Best Single Relay Selection

By selecting the best single relay from the relay candidates to assist in the transmission to the destination, we achieve selection diversity. We begin this section by discussing two different BSR selection strategies. Subsequently, we derive the achievable reliability performance of these two strategies in both the IBL and FBL regimes. Finally, we address the reliability bound of the considered BSR network when operated with the rate-compatible coding scheme based on punctured PCs as discussed above.

A. BSR Selection Strategies: Source-Driven vs. Relay-Driven

The source-driven BSR selection strategy makes the source responsible for selecting the BSR, and it has been has been widely discussed in the IBL regime. In particular, as shown in Fig. 2, under the source-driven strategy, the destination is required to feed back (broadcast) the second hop CSIs to the relays and the source, while each relay subsequently feeds back its backhaul CSI to the source. Then, the source determines the optimal relay (as well as the other operation parameters, *e.g.*, for resource allocation) based on the CSIs of the two hops. Then, the source broadcasts a beacon including the BSR decision to the relays and subsequently transmits the data packet to the selected relay. The selected relay attempts to decode the data packet and forwards it to the destination if decoding is successful.

On the other hand, the relay group (including a leader relay) may also drive the selection. In such a relay-driven strategy, the destination still needs to report the second-hop CSI to the relay group. But unlike the source-driven strategy, the source does not need any CSI and simply broadcasts the data packet to all relays. Consequently, it is possible that multiple relays decode the data packet correctly and hence may become active relays. We denote the set of such relays by $\mathcal{A}$ and refer to it as the *active relay set*. Then, each relay other than the leader relay will send a one bit acknowledgement (ACK), indicating its active state to the leader. Subsequently, the leader relay selects the best relay from all the active relays and broadcasts the decision in a beacon to all relays. Finally the selected relay forwards the data packet to the destination in the relaying phase. The destination obtains an SNR given by $\max_{j\in\mathcal{A}}\{\gamma_{j,\mathrm{D}}\}$. An example of the frame structure is provided in Fig. 2.

As of now, we have outlined example structures of typical source-driven and relay-driven strategies, which incur comparable feedback overhead. In the next subsections, we will investigate the performance of these two strategies in both the IBL regime and FBL regime, respectively. In particular, we aim at answering the following questions. Why has the relay-driven strategy received much less attention in comparison to the source-driven strategy in the IBL regime? Why is it essential to study the relay-driven strategy in the FBL regime?

B. Achievable Reliability of BSR in the IBL Regime

First of all, it should be pointed out that for a practical communication system, the blocklength is definitely not infinite. Hence, when we study the performance of such system in the IBL regime, this only indicates that the impact of the limited length of coding blocks is ignored in the analysis. In particular, the analysis in the IBL regime follows the assumption which is only true when the blocklength is infinitely long: a packet is assumed to be decoded with arbitrarily small error probability given that the coding rate is lower than the Shannon capacity.

In the considered multiple relay network, having the CSI of all links, the BSR selector (either the source or the leader of the relay group) knows exactly which relays may decode the packet successfully for a given data packet and blocklength. Hence, the selection decisions of the source-driven and the relay-driven strategies are exactly the same. Denote by μ_j the IBL throughput (in bits per frame) of the two-hop transmission via relay j. With bocklength m_1, a total of $m_1 \log_2(1+\gamma_{\mathrm{S},j})$ bits can be transmitted via the first hop of relaying, according to the Shannon capacity theory, while this value for the second hop is $m_2 \log_2(1+\gamma_{j,\mathrm{D}})$. Note that μ_j is limited by the minimum of the throughputs of the two hops of relay j. We thus have

$$\mu_j = \min\left\{m_1 \log_2\left(1+\gamma_{\mathrm{S},j}\right), m_2 \log_2\left(1+\gamma_{j,\mathrm{D}}\right)\right\}. \tag{12}$$

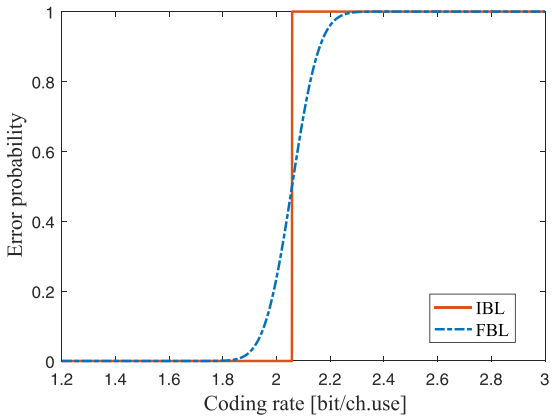


Fig. 3. The error performance of a single-hop scenario with perfect CSI. In the numerical analysis, we set an SNR$= 5\,dB$ and a blocklength $m = 300$, and thus have a Shannon capacity of about 2.06 bits per channel use.

The BSR decision (under either the source-driven or relay-driven strategy) results in choosing the relay with the highest IBL throughput. Then, the maximal IBL throughput over all relays is given by

$$\mu_{\rm IBL}^{\max} = \max_{j\in\mathcal{J}} \{\mu_j\}\,. \tag{13}$$

Finally, we use the error probability to characterize the achievable reliability of the transmission for a packet of size D via the considered BSR network, given by

$$\Pr_{\rm IBL} = \begin{cases} 1, & \text{if } \mu_{\rm IBL}^{\max} < {\rm D}\,, \\ 0, & \text{if } \mu_{\rm IBL}^{\max} \geq {\rm D}\,. \end{cases} \tag{14}$$

C. Achievable Reliability of BSR in the FBL Regime

Unlike in the IBL regime, Equation (5) provides a different error model for the FBL regime. In Fig. 3, we provide numerical results of a single-hop transmission based on (5) in comparison to the IBL regime.

We observe that in the FBL regime the (single-hop) transmission is not necessarily correct even for coding rates lower than the Shannon capacity. Therefore, in the considered network, even having perfect CSI of all links does not enable the source to know exactly which relays will decode the broadcasted packet successfully. In other words, if the source selects a relay as the forwarding relay, *i.e.*, working under the source-driven model, it is possible that this relay fails in decoding the packet and hence cannot forward the packet. As a result, the decision of BSR selection of the source-driven and relay-driven strategies are not necessarily identical, resulting in different performance levels. We note that this difference only exists in the FBL regime (recall that in the IBL regime, the two strategies have the same result).

1) Source-Driven BSR in the FBL Regime: According to (5), with coding rate $r = D/m_1$ the achievable error probability of the link from the source to Rj is given by $\varepsilon_{{\rm S},j} = \mathcal{P}(\gamma_{{\rm S},j}, D/m_1, m_1)$. Let $\varepsilon_{{\rm tot},j}$ denote the overall error probability of transmission when relay j is selected as the relay. Then, we have

$$\begin{aligned}\varepsilon_{{\rm tot},j} &= \varepsilon_{{\rm S},j} + \varepsilon_{j,{\rm D}} - \varepsilon_{{\rm S},j}\cdot\varepsilon_{j,{\rm D}} \\ &= \mathcal{P}\left(\gamma_{{\rm S},j}, \frac{D}{m_1}, m_1\right) + \mathcal{P}\left(\gamma_{j,{\rm D}}, \frac{D}{m_2}, m_2\right) \\ &\quad - \mathcal{P}\left(\gamma_{{\rm S},j}, \frac{D}{m_1}, m_1\right)\cdot\mathcal{P}\left(\gamma_{j,{\rm D}}, \frac{D}{m_2}, m_2\right).\end{aligned} \tag{15}$$

Hence, the selection decision under the source-driven strategy is implemented by choosing the relay with the lowest $\varepsilon_{{\rm tot},j}$. Finally, the achievable error probability of transmitting a packet with size D via this BSR network is given by

$$\varepsilon_{\rm SD} = \min_{j\in\mathcal{J}} \{\varepsilon_{{\rm tot},j}\}\,. \tag{16}$$

2) Relay-Driven BSR in the FBL Regime: Under the relay-driven strategy, the relay is selected from the active relay set. Note that with probability $1-\varepsilon_{{\rm S},j}$, relay j will decode the data block correctly. Hence, regardless of the coding rate $r = D/m_1$, there is a positive probability that some relays decode the data block successfully and can be active in the relaying phase. Let $n_{\rm a} = |\mathcal{A}|$ denote the cardinality of the active relay set $\mathcal{A}$, *i.e.*, $\mathcal{A}$ contains $n_{\rm a}$ active relays which have decoded the data block correctly. Then, we have $n_{\rm a} = \sum_{j=1}^{J} x_j$, where x_j is a Bernoulli random variable, that is, $x_j \sim {\rm Ber}(1-\varepsilon_{{\rm S},j})$, indicating if R$j$ is in $\mathcal{A}$ or not. Hence, the expected value of $n_{\rm a}$ is given by $\mathbb{E}[n_{\rm a}] = \sum_{j=1}^{J}(1-\varepsilon_{{\rm S},j})$. In addition to that, we present the following lemma to characterize $n_{\rm a}$, while a proof is provided in Appendix A.

Lemma 1: The expected number of active relays $\mathbb{E}[n_a]$ is increasing in m_1.

From $\mathcal{A}$, the relay with the highest SNR of the link to the destination is selected to achieve the highest reliability for the second hop transmission. Hence, the received SNR at the destination is given by $\max_{j\in\mathcal{A}}\{\gamma_{j,{\rm D}}\}$. Consequently, the error probability of the second hop is given by $\mathcal{P}(\max_{j\in\mathcal{A}}\{\gamma_{j,{\rm D}}\}, \frac{D}{m_2}, m_2)$, which represents the overall error probability conditioned on the active relay set $\mathcal{A}$. On the other hand, the event of the active relay set being $\mathcal{A}$ happens if all the relays in $\mathcal{A}$ successfully decode the packet while the remaining relays fail, i.e., the probability of this event is $\prod_{j\notin\mathcal{A}}\varepsilon_{{\rm S},j}\prod_{n\in\mathcal{A}}(1-\varepsilon_{{\rm S},n})$. Denote by $\varepsilon_{\rm RD}$ the overall error probability of transmitting a packet with size D via the relay-driven BSR network. According to the law of total probability, $\varepsilon_{\rm RD}$ can be calculated by summing up the conditional error probabilities given all possible active relay sets $\mathcal{A}$, and can be expressed as

$$\begin{aligned}\varepsilon_{\rm RD} = \sum_{\mathcal{A}\in(\mathcal{J})} \Bigg\{ &\prod_{j\notin\mathcal{A}}\varepsilon_{{\rm S},j}\prod_{n\in\mathcal{A}}(1-\varepsilon_{{\rm S},n}) \\ &\times \mathcal{P}\left(\max_{j\in\mathcal{A}}\{\gamma_{j,{\rm D}}\}, \frac{D}{m_2}, m_2\right)\Bigg\},\end{aligned} \tag{17}$$

where $\mathbb{P}(\mathcal{J})$ is the power set of the relay set $\mathcal{J}$.

We have derived the reliability performance models of the source-driven and relay-driven BSR networks in the FBL regime. The following lemma characterizes the performance difference between the two models, while a proof is provided in Appendix B.

Lemma 2: In the FBL regime, the expected error probability of the relay-driven strategy is lower than that of the source-driven strategy, *i.e.*, the relay-driven strategy is more reliable.

Note that Lemma 2 essentially answers the questions posed at the end of Section III-A affirmatively, and verifies that studying the relay-driven BSR selection is particular important in latency-critical IoT networks operating with FBL codes.

D. Reliability Bound of the BSR Network With Polar Codes

In this subsection, we address the reliability bound of the considered network when operating with PCs. While there are plenty of well-established choices for a coding system, we opt for PCs for several reasons. By design, PCs allow for a wide range of rates via fixing the number of channel uses employed to convey information to an arbitrary integer. Combined with an explicit upper bound on the probability of block error under SC decoding, but also to that of the same code under more powerful SCL decoding, this allows for great flexibility for evaluating our findings numerically.

To construct the punctured PCs used as component codes $\mathcal{C}_t$ of the rate-compatible code $\mathfrak{C}$, we perform DEs taking into account the puncturing patterns. As in [37], we obtain the puncturing patterns via the quasi-uniform puncturing (QUP) presented in [39]. We then select the index set based on the channel parameters $P_{\mathrm{e}}(\mathcal{W}_{m,i})$ as given in (6) such that the union bound of block error

$$\mathcal{P}_{\mathrm{PC}}\big(\gamma, r, m\big) \;\leq\; \sum_{i\in\mathcal{I}} P_{\mathrm{e}}(\mathcal{W}_{m,i}) \tag{18}$$

is minimized. The channel parameters, that is, the error probabilities $P_{\mathrm{e}}(\mathcal{W}_{m,i})$, are approximated by a DE taking into account the puncturing pattern. Assuming the all-zero codeword, this method tracks densities of the logarithmic likelihood-ratios (LLRs) on the factor graph (FG) representation of the decoder according to the sequential decoding schedule [38]. This results in channel parameters

$$\begin{aligned} P_{\mathrm{e}}(\mathcal{W}_{m,i}) &= \mathbb{P}[L_{m,i} \leq 0] \\ &= \lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\epsilon} \mathsf{L}(x)\,dx + \frac{1}{2}\int_{-\epsilon}^{+\epsilon} \mathsf{L}(x)\,dx\right), \end{aligned} \tag{19}$$

where L denotes the probability density of the LLR

$$L_{m,i} = \log \frac{W_{m,i}\big(\mathbf{y}, \hat{\mathbf{u}}_1^{i-1} \,|\, u_i = 0\big)}{W_{m,i}\big(\mathbf{y}, \hat{\mathbf{u}}_1^{i-1} \,|\, u_i = 1\big)}. \tag{20}$$

Note that by this approach, we obtain a theoretically justified, explicit upper bound on the performance of the component codes under SC decoding, which also holds for the punctured instances $\mathcal{C}_t$ of the mother PC. To obtain an analytical bound on the reliability of each PC instance in the BSR network, we then employ the upper bound on $\mathcal{P}_{\mathrm{PC}}(\gamma, r, m)$ as in (18) as $\mathcal{P}(\gamma, r, m)$, and obtain the achievable reliability as in the previous subsections.

In this work, we employ a rate-compatible code $\mathfrak{C}$ based on punctured PCs as described above. Specifically, we construct $\mathfrak{C}$ by puncturing a mother PC of length $m = 2^8 = 256$ offering $T = 13$ different rates via length adaption for a fixed information dimension $K = 96$. The puncturing patterns $\mathcal{P}_t$, $t \in [T]$, are obtained via the QUP approach presented in [39]. We use $\mathcal{P}_1$ of cardinality $|\mathcal{P}_1| = 96$, and thus obtain $m_1 = 256 - 96 = 160$ as the shortest code length supported by $\mathfrak{C}$, having rate $r_1 = \frac{K}{m_1} = \frac{3}{5}$. Increasing lengths in steps of 8, we obtain $m_t = 160 + 8(t-1)$, $t \in [13]$ and hence support rates

$$r_t = \frac{K}{m_t} = \frac{96}{160 + 8(t-1)} = \frac{12}{19 + t},\; t \in [13]\,. \tag{21}$$

IV. Achievable Reliability via Blocklength Allocation for the BSR Network

In this section, we study the achievable instantaneous reliability (per frame) via blocklength allocation. In particular, in this work we are interested in a reliability-optimal system to support latency-critical transmissions in IoT networks. On the one hand, the code blocklength of transmission via each link is relatively short. On the other hand, the transmitted packets are also relatively short and the transmissions are required to be ultra-reliable, *i.e.*, $\varepsilon_{\mathrm{S},j} \ll 0.1$ and $\varepsilon_{j,\mathrm{D}} \ll 0.1$, $j \in \mathcal{J}$. According to the reliability-oriented principle of the design, we study the blocklength allocations for the source-driven and relay-driven BSR selection strategies, respectively.

A. Achievable Reliability via Blocklength Allocation for Source-Driven BSR Networks

Based on the instantaneous error probability given by (16), the blocklength allocation problem under the source-driven strategy can be expressed by

$$\begin{aligned} \min_{m_1,m_2}\;\; & \varepsilon_{\mathrm{SD}} \\ \text{s.t.}:\;\; & m_1 + m_2 \leq M, \\ & m_1, m_2 \in \mathbb{Z}^+\,. \end{aligned} \tag{22}$$

To solve the problem, we first provide the following lemma, proved in Appendix C.

Lemma 3: The equality in the delay constraint of (22), i.e., $m_1 + m_2 = M$, is always required in order to minimize $\varepsilon_{\mathrm{SD}}$.

In addition, recall that we consider an ultra-reliable scenario where $\varepsilon_{\mathrm{S},j} \ll 0.1$ and $\varepsilon_{j,\mathrm{D}} \ll 0.1$. Hence, $\varepsilon_{\mathrm{S},j} + \varepsilon_{j,\mathrm{D}} \gg \varepsilon_{\mathrm{S},j} \cdot \varepsilon_{j,\mathrm{D}}$. This motivates the approximation $\varepsilon_{\mathrm{tot},j} = \varepsilon_{\mathrm{S},j} + \varepsilon_{j,\mathrm{D}} - \varepsilon_{\mathrm{S},j} \cdot \varepsilon_{j,\mathrm{D}} \approx \varepsilon_{\mathrm{S},j} + \varepsilon_{j,\mathrm{D}}$. We will validate the accuracy of this approximation by simulation. Then, based on this approximation and Lemma 3, we reformulate problem (22) and obtain

$$\begin{aligned} \min_{m_1,m_2}\;\; & \min_{j\in\mathcal{J}}\{\varepsilon_{\mathrm{S},j} + \varepsilon_{j,\mathrm{D}}\} \\ \text{s.t.}:\;\; & m_1 + m_2 = M, \\ & m_1, m_2 \in \mathbb{Z}^+\,. \end{aligned} \tag{23}$$

Note that the objective is the minimum over J functions, and it is not necessarily convex when $J > 2$. Hence, we further

reformulate the problem and obtain[2]

$$\begin{aligned} \min_{j\in\mathcal{J}} \quad & \min_{m_1,m_2} \{\varepsilon_{\mathrm{S},j}+\varepsilon_{j,\mathrm{D}}\} \\ & \text{s.t.}: m_1+m_2=M, \\ & m_1,m_2\in\mathbb{Z}^+. \end{aligned} \tag{24}$$

Then, we provide the following key lemma for solving the problem in (24), and give a proof in Appendix D.

Lemma 4: $f_j(m_1,m_2)=\varepsilon_{\mathrm{S},j}+\varepsilon_{j,\mathrm{D}}$ is convex in (m_1,m_2).

$$P_j=(1-\varepsilon_{\mathrm{S},j})\sum_{\mathcal{F}'\in\mathbb{P}(\mathcal{J}\text{-}\{j\})}\left\{\prod_{k\notin\mathcal{F}'}\varepsilon_{\mathrm{S},k} \times\prod_{n\in\mathcal{F}'}(1-\varepsilon_{\mathrm{S},n})\,\mathbb{1}\{\max_{k\in\mathcal{F}'}\{\gamma_{k,\mathrm{D}}\}\le\gamma_{j,\mathrm{D}}\}\right\}, \tag{25}$$

According to Lemma 4, the sub-problem of $\min_{m_1,m_2} f_j(m_1,m_2)=\min_{m_1,m_2}\{\varepsilon_{\mathrm{S},j}+\varepsilon_{j,\mathrm{D}}\}$ subject to the linear constraint $m_1+m_2=M$ can be efficiently solved by applying convex optimization techniques. Note that blocklengths should be non-negative integers, solving the sub-problem contains two steps: First, obtain the optimal solution (m_1°,m_2°) of the relaxation of the sub-problem. In the next step, we check this solution. If m_1° and m_2° are integers, this solution is also optimal for the original sub-problem, *i.e.*, $(m_1^*,m_2^*)=(m_1^\circ,m_2^\circ)$. Otherwise, let $m_{\text{ceil}}=\lceil m_1^*\rceil$ and $m_{\text{floor}}=\lfloor m_1^*\rfloor$, where $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ are ceil and floor functions, respectively. Then, the optimal solution of the original sub-problem is given by $(m_1^*,M-m_1^*)$, where m_1^* is given by

$$m_1^*=\operatorname*{arg\,min}_{m_1\in\{m_{\text{ceil}},m_{\text{floor}}\}} f_j(m_1,M-m_1). \tag{26}$$

In total, we need to solve at most J sub-problems and find the optimal solution for problem (24). Recall that we do not have to solve the sub-problems for relays with poor channel gains in both of the two hops. In particular, only the following three types of relays are candidates for which we need to solve the sub-problems with respect to the SNRs of the two hops: (i) The relay j^* with the highest bottleneck SNR over all relays, where the bottleneck SNR of the j^*-th relay is given by $\min\{\gamma_{\mathrm{S},j^*},\gamma_{j^*,\mathrm{D}}\}$. (ii) The relays with a first-hop SNR higher than γ_{S,j^*}. (iii) The relays with a second-hop SNR higher than $\gamma_{j^*,\mathrm{D}}$.

B. Achievable Reliability via Blocklength Allocation for Relay-Driven BSR Networks

Under the relay-driven strategy, the blocklength allocation problem becomes

$$\begin{aligned} \min_{m_1,m_2} \quad & \varepsilon_{\mathrm{RD}} \\ \text{s.t.}: \quad & m_1+m_2\le M, \\ & m_1,m_2\in\mathbb{Z}^+. \end{aligned} \tag{27}$$

[2]Note that unlike the max-min inequality, interchanging the two minimums does not alter the optimization problem.

It is challenging to solve this problem due to the fact that according to (17), $\varepsilon_{\mathrm{RD}}$ is calculated as a summation of 2^J terms, while each term is a product of $J+1$ error probability functions. Therefore, we reformulate the problem in the following way.

Recall that $\varepsilon_{\mathrm{tot},j}$ represents the overall error probability of transmission if relay j is selected. Consequently, we represent the objective of the problem by $\varepsilon_{\mathrm{RD}}=\sum_{j=1}^{J}\varepsilon_{j,\mathrm{D}}\cdot P_j$, where P_j is the probability that relay j is selected, *i.e.*, relay j is in the active relay set and has the highest second hop SNR in the set. Hence, we have the expression of P_j provided in Equation (25) on the next page, where $\mathcal{F}'=\mathcal{F}\setminus\{j\}$ is the set of relays in the active relay set excluding relay j. In addition, $\mathbb{1}\{\cdot\}$ is an indicator function. In particular, $P_j=1-\varepsilon_{\mathrm{S},j}$ holds if relay j has the highest SNR in the second hop, *i.e.*, $\max_{k\in\mathcal{J}}\{\gamma_{k,\mathrm{D}}\}=\gamma_{j,\mathrm{D}}$. In other words, the SNRs in the second hop directly influence the selection of the relay. We therefore sort the relay group in decreasing order according to the SNR values of the second hop. For instance, the relay with the largest second-hop SNR is denoted as relay 1, and the relay with the second largest second-hop SNR is indicated as relay 2. This sorted relay set is denoted by $\mathcal{J}^\circ$, where for $k\in\mathcal{J}^\circ$ and $k\le J-1$, we have $\gamma_{k,\mathrm{D}}\ge\gamma_{k+1,\mathrm{D}}$. We define $\varepsilon_{\mathrm{S},0}=1$, and have $P_j=(1-\varepsilon_{\mathrm{S},j})\prod_{k=0}^{j-1}\varepsilon_{\mathrm{S},k}$. Note that the intuitive description is that with the new indexing of the relays, relay j is selected if it is in the active relay set and the relay 1 through $j-1$ (which have higher second-hop SNRs than relay j) are not in the active set. Then, we obtain

$$\begin{aligned} \varepsilon_{\mathrm{RD}} &= \sum_{j\in\mathcal{J}^\circ}\varepsilon_{j,\mathrm{D}}(1-\varepsilon_{\mathrm{S},j})\prod_{k=0}^{j-1}\varepsilon_{\mathrm{S},k} \\ &= \sum_{j\in\mathcal{J}^\circ}(\varepsilon_{j,\mathrm{D}}-\varepsilon_{\mathrm{S},j}\varepsilon_{j,\mathrm{D}})\prod_{k=0}^{j-1}\varepsilon_{\mathrm{S},k} \\ &\approx \sum_{j\in\mathcal{J}^\circ}\varepsilon_{j,\mathrm{D}}\prod_{k=0}^{j-1}\varepsilon_{\mathrm{S},k} \\ &= \sum_{j\in\mathcal{J}^\circ}\mathcal{P}\left(\gamma_{j,\mathrm{D}},\frac{D}{m_2},m_2\right)\prod_{k=0}^{j-1}\mathcal{P}\left(\gamma_{\mathrm{S},k},\frac{D}{m_1},m_1\right), \end{aligned} \tag{28}$$

where the approximation is motivated by the fact that in the considered ultra-reliable transmission scenario $\varepsilon_{\mathrm{S},j}\ll 0.1$ and $\varepsilon_{j,\mathrm{D}}\ll 0.1$, and therefore $\varepsilon_{j,\mathrm{D}}\gg\varepsilon_{\mathrm{S},j}\varepsilon_{j,\mathrm{D}}$. In the next section, we will provide simulation results to show that the approximation is tight for ultra-reliable transmission scenarios. According to (28), the solution of the problem in (27) can be obtained approximately by solving

$$\begin{aligned} \min_{m_1,m_2} \quad & \sum_{j\in\mathcal{J}^\circ}\varepsilon_{j,\mathrm{D}}\prod_{k=0}^{j-1}\varepsilon_{\mathrm{S},k} \\ \text{s.t.}: \quad & m_1+m_2\le M, \\ & m_1,m_2\in\mathbb{Z}^+. \end{aligned} \tag{29}$$

The constraint of the problem can also be improved using the following lemma, proved in Appendix E.

Lemma 5: The equality in the delay constraint of (29), *i.e.*, $m_1 + m_2 = M$, is always required in order to minimize $\sum_{j\in\mathcal{J}^\circ} \varepsilon_{j,\mathrm{D}} \prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k}$.

According to Lemma 5, the problem in (29) is equivalent to

$$\begin{aligned} \min_{m_1,m_2} \quad & \sum_{j\in\mathcal{J}^\circ} \varepsilon_{j,\mathrm{D}} \prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k} \\ \text{s.t.}: \quad & m_1 + m_2 = M, \\ & m_1, m_2 \in \mathbb{Z}^+ . \end{aligned} \tag{30}$$

Then, we provide the following key lemma for solving the problem in (30), for which a proof is given in Appendix F.

Lemma 6: Under the condition $m_1 + m_2 = M$, the objective $\sum_{j\in\mathcal{J}^\circ} \varepsilon_{j,\mathrm{D}} \prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k}$ is convex in (m_1, m_2).

However, according to Lemma 6, the relaxed problem of (30) can be efficiently solved by applying convex optimization techniques. Then, the optimal solution of the original problem in (30) can be obtained by checking the nearest two integers (on both the left and right sides) with respect to the optimal solution of the relaxed problem. The detail process is the same as the discussion provided after Lemma 4.

In this section, we have proposed the instantaneous blocklength allocation policies for the considered BSR network under both source-driven and relay-driven strategies. It is worth mentioning that these two per-frame blocklength allocation policies can be easily extended to a constant frame structure design. That is, once the optimal structure is determined, it will be fixed for all frames over time. In this case, the aim is to minimize the expected error probability with respect to channel-fading. Note that the sum of convex functions maintains the convexity. Therefore, the above lemmas extend to the constant design problems as well. Moreover, it should be pointed out that to achieve the minimal error probability via the optimal blocklength allocation, CSI is required for determining the blocklength. In particular, the average CSI is required for the above constant frame structure design. Moreover, to apply the instantaneous blocklength allocation, the instantaneous CSI is required to be known by the BSR selector at the beginning of each frame. For instance, under the source-driven BSE selection strategy, the source could apply the instantaneous blocklength allocation after receiving the instantaneous CSI. On the other hand, under the relay-driven strategy, the source does not have instantaneous CSI. Hence, the instantaneous blocklength allocation is required to be applied at the relay group (a leader relay), and the allocation decision needs to be reported to the source (with a certain overhead) before the broadcast phase.

V. Numerical Results and Discussion

In this section, we present simulation and numerical results to support our findings. We first validate our performance models by simulations. In addition to that, we evaluate the BSR network performance under both the source-driven, as well as the relay-driven strategies. In particular, both the FBL achievable performance results and the practical PCs results are provided. For the simulation and numerical analysis, we consider a low SNR scenario, thus motivating the application of relays to enhance transmission reliability. The average received SNR at each link is assumed to be 5 dB. In addition, we set a unit average channel gain for all links, while assuming that all links experience independent and identically distributed (i.i.d.) Rayleigh quasi-static fading. As a default setup, we set the number of relays to $J = 5$, where we vary this setup in Fig. 6 and Fig. 8. Finally, the packet size is set to 96 bits, while the total blocklength for the two-hop data transmission is set to $M = 412$.

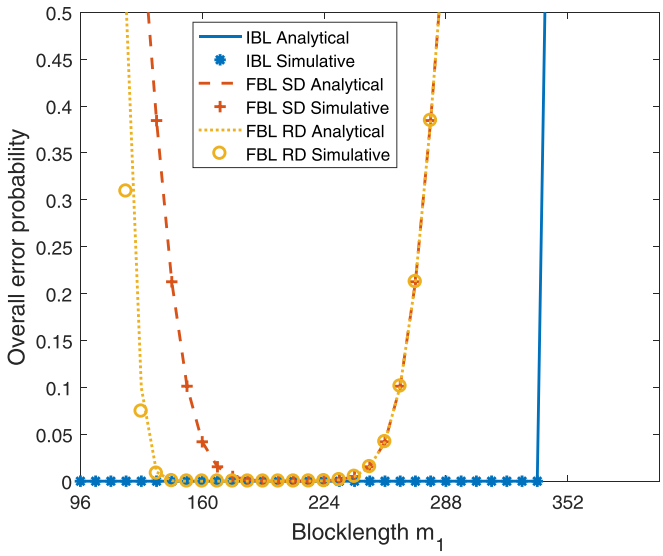


Fig. 4. The expected overall error probability of an instantaneous frame is convex in blocklength m_1 in the reliable region, *e.g.*, when the error probability is lower than 0.1. In the analysis, we set $J = 5$.

Fig. 4 illustrates the impact of blocklength allocation on the expected reliability performance within a frame. Both the source-driven (SD) and the relay-driven (RD) selection strategies are investigated. First of all, we observe a very close match between the simulations and the proposed analytical models, which confirms the accuracy of the derived reliability bounds. In addition to that, as expected, the error probability in the IBL regime is zero when the IBL throughput of the two-hop transmission is higher than the packet size, and switches to one when the condition does not hold. At the same time, all FBL reliabilities are convex in the blocklength m_1 in the reliable regime, which indicates that the employed approximations in both (23) and (28) is appropriate. Moreover, the relay-driven strategy provides a better reliability performance than the source-driven one, confirming our analytical characterization given in Lemma 2.

To illuminate the achievable reliabilities of the two strategies, we change the y-axis scale of Fig. 4 from linear to logarithmic and only provide the FBL results in Fig. 5 (where the IBL curves with value 0 cannot be shown in the logarithmic scale). From the figure, we observe more clearly that the relay-driven selection strategy introduces a (4 orders of magnitude) lower error probability in comparison to the source-driven one. More interestingly, the optimal reliabilities of the two strategies are attained at different blocklength allocations, where the relay-driven strategy prefers a relatively short blocklength for the backhaul link. Finally, recall that the analytical results of the relay-driven selection strategy are obtained based on the approximation in (28). The close match between the simulation and analytical results of

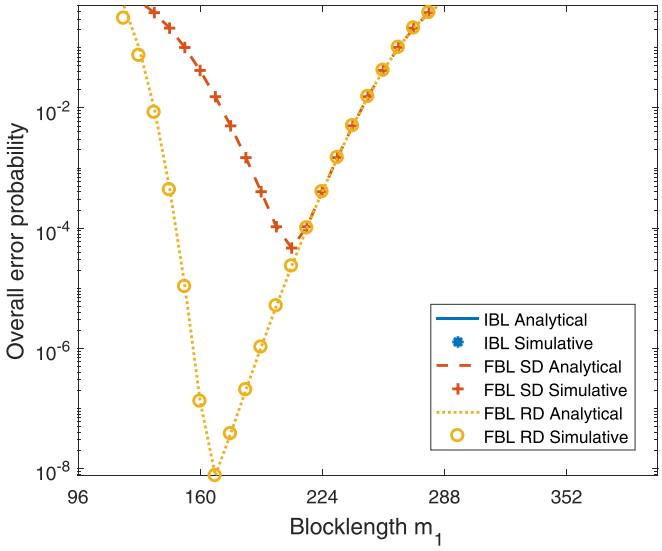


Fig. 5. The expected overall error probability of an instantaneous frame in logarithmic scales. In the analysis, we set $J = 5$.

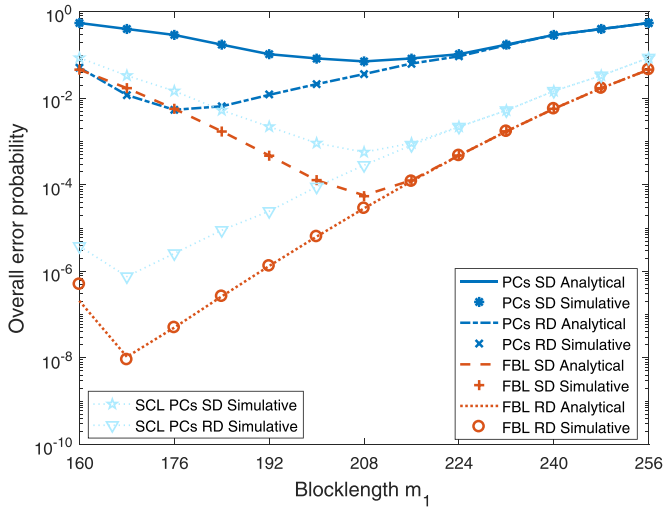


Fig. 7. The expected overall error probability in an instantaneous frame: Analytical FBL bounds and simulation results for PCs with SC and SCL decoding. In the analysis, we set $J = 5$.

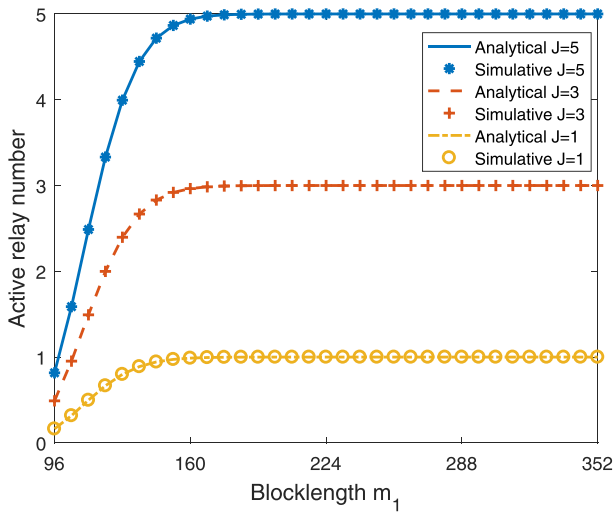


Fig. 6. The expected number of active relays in an instantaneous frame is increasing in blocklength m_1.

the relay-driven strategy confirms that the approximation in (28) is tight for a reliable transmission scenario.

Next, we study the relationship between the active relay number and the blocklength allocation.

As shown in Fig. 6, for all the setups of the total number of relays deployed in the system, the expected number of active relays in a transmission frame is increasing in the blocklength of the backhaul link, which confirms our analytical result in Lemma 1. Finally, the simulation results again match well with our analytical results for all scenarios with different number of deployed relays.

To assess the validity of these findings with respect to practical application scenarios, we employ the rate-compatible code $\mathfrak{C}$ obtained by puncturing PCs as described in Sections II-C and III-D. We obtain bounds for the performance of the codes $\mathcal{C}_t \in \mathfrak{C}$ under SC decoding via DE, and provide them alongside the FBL performance bounds for both relay selection strategies in Fig. 7. Furthermore, in the same plot, we give simulation results for the codes $\mathcal{C}_t \in \mathfrak{C}$ under both SC decoding, as well as cyclic redundancy check (CRC)-aided SCL decoding, using a CRC with generator polynomial $g(x) = x^7 + x^3 + 1$ as an outer code of length 7 and a list size of 16. Note that for a fair comparison, the inner codes which are punctured PCs are of dimension $K' = K + 7 = 103$ to accommodate for the CRC.

We observe SC decoding results, which, albeit matching the bounds obtained via DE, do not facilitate competitive system performance in terms of overall error probability with respect to the FBL bounds provided. On the other hand, the CRC-aided SCL decoder helps to achieve a system performance quite close to the analytical FBL bounds. In addition to that, we observe that for both relay selection strategies, the simulation results for the channel coding system employed here exhibit the same characteristics such as convexity as the corresponding analytical bounds. More importantly, the results indicate that optimal blocklength allocation based on the analytical bounds does provide reasonable design guidelines for systems relying on practical codes, *e.g.*, on PCs. In particular, the optimal blocklength allocation choices based on the FBL bound are also promising for PCs, especially when the practical codes have excellent reliability performance.

So far, we have validated our analytical model, where in particular the instantaneous achievable reliability can be optimized via blocklength allocation. By applying the optimal blocklength allocation per frame, the average (over fading) of the achievable reliability can be obtained. We present the results in Fig. 8, where we vary the number of relays deployed in the network.

Most prominently, we observe again a good match between the simulation and analytical results. Secondly, as expected, the source-driven and relay-driven selection strategies have the same performance when the system has only one relay. The reliability

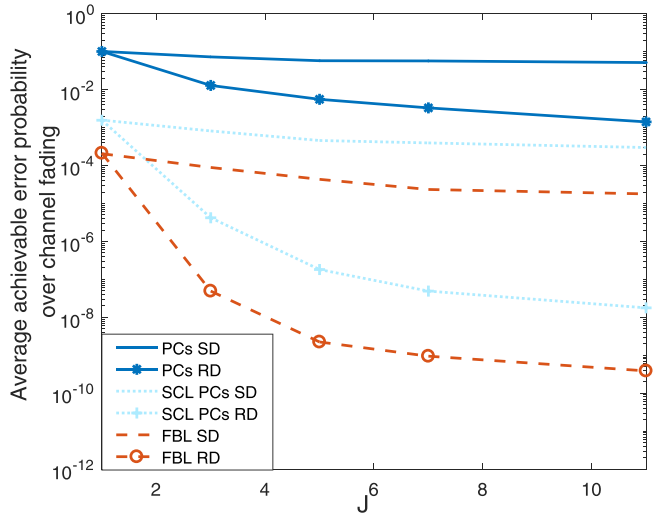


Fig. 8. The average achievable overall error probability (over channel fading). In the figure, we vary the number of relays J from 1 to 15. Note that the results in the figure are numerical, *i.e.*, obtained from our analytical model. The FBL and PC curves are obtained based on the validated analytical bound, *i.e.*, analytical results, while the SCL PCs curves (PCs under CRC-aided SCL decoding) are simulation results.

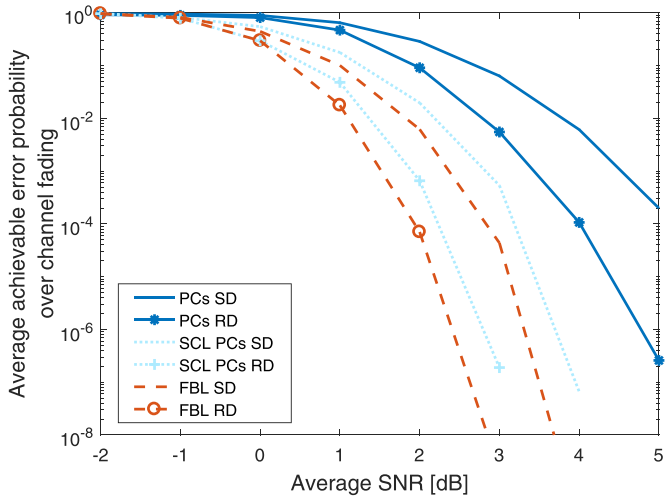


Fig. 9. The average achievable overall error probability (over channel fading) at different average SNRs. In this figure, we set $J = 5$. The FBL and PCs curves are obtained based on the validated analytical bound, *i.e.*, analytical results, while the SCL PCs curves (PCs under CRC-aided SCL decoding) are simulation results.

performance of all cases improves as more relays are deployed. In particular, the relay-driven strategy benefits more significantly than the source-driven strategy, *i.e.*, curves of the source-driven strategy are relatively flat. However, as we add more relays to the system, the reliability enhancement by deploying more relays becomes less significant. For the design of low-latency short blocklength systems, this result suggests a certain cluster size of nodes acting as relay candidates. Furthermore, we again observe that the simulation results for punctured PCs under both decoders have characteristics (*e.g.*, flatness of either the source-driven strategy, or the relay-driven strategy, and the performance difference between the two strategies) similar to the corresponding FBL bounds.

Finally, we evaluate the average achievable reliability while considering different average channel qualities. As shown in Fig. 9, the error probabilities are decreasing in the average SNR of all links. In particular, for the relay-driven strategy they decrease more rapidly compared to the source-driven one. In addition, the gap between the two strategies of either the PC or FBL case becomes more significant as the channels get better. However, when SCL decoding is employed, for both strategies performance close to the analytical FBL optimum may be achieved.

VI. Conclusion

We have characterized the achievable FBL performance bounds for a BSR network and illustrated it employing PCs as an example of a low-complexity coding system which greatly profits from BSR-induced diversity gains. Both the source-driven and the relay-driven BSR selection strategies have been studied. In particular, we have proved that the two BSR strategies have the same performance in the IBL regime, while the relay-driven strategy is significantly more reliable than the source-driven one in the FBL regime. This confirms the importance and necessity of the study presented in this work. Moreover, following the derived performance model, we have investigated the achievable reliability via an optimal efficient blocklength allocation for the network under both the source-driven and relay-driven BSR selection strategies.

The appropriateness of the proposed analytical model, *i.e.*, the FBL and PC bounds as well as the proposed lemmas, is validated by simulations. In addition, we have evaluated the BSR networks under both the relay-driven and source-driven BSR selection strategies. It was shown that the relay-driven strategy provides significantly more reliable transmissions than the source-driven one. In addition, the relay-driven strategy benefits more significantly from deploying additional relays or having better channel qualities. In our simulations, we have provided the results for general FBL codes according to Polyanskiy's model and we also applied PCs under SC and SCL decoding. We have observed that SCL decoding helps PCs to offer a competitive reliability performance in comparison to the FBL bound. More importantly, although the performances of applying practical PCs are lower than that in the general FBL model, their reliability behaviors in terms of blocklength allocation are very similar. In particular, the above relationship between the relay-driven and source-driven BSR selection strategies holds for the PC cases as well as the FBL case. Moreover, the optimal design based on the FBL model holds also for the scenario with a practical FBL coding scheme.

Appendix A
Proof of Lemma 1

According to (5), we have $\mathcal{P}(\gamma, \frac{D}{m}, m) = Q(\frac{\sqrt{m}\log_2(1+\gamma)-\frac{D}{\sqrt{m}}}{\sqrt{1-\frac{1}{(\gamma+1)^2}}})$. Let $\phi = (1-\frac{1}{(\gamma+1)^2})^{-\frac{1}{2}}$. ϕ is a positive constant with respect to m. In addition, let $t = \sqrt{m}$. Then,

we obtain

$$\mathcal{P}\left(\gamma, r, t\right) = Q\left(\phi t \log_2\left(1+\gamma\right) - \phi D t^{-1}\right), \tag{31}$$

and the corresponding first order derivative with respect to t

$$\begin{aligned}&\frac{\partial P\left(\gamma, r, t\right)}{\partial t}\\ &= -\frac{\phi}{\sqrt{2\pi}} e^{-\frac{\left[\phi t \log_2(1+\gamma) - \phi D t^{-1}\right]^2}{2}} \left[\log_2\left(1+\gamma\right) + D t^{-2}\right] \leq 0.\end{aligned} \tag{32}$$

We have $\frac{\partial P(\gamma,r,t)}{\partial t} \leq 0$ as ϕ is positive. We can further show that $\frac{\partial \mathcal{P}}{\partial m} = \frac{\partial \mathcal{P}}{\partial t}\frac{\partial t}{\partial m} = \frac{1}{2}\frac{\partial \mathcal{P}}{\partial t} m^{-\frac{1}{2}} \leq 0$. Note that $\varepsilon_{\mathrm{S},j} = \mathcal{P}(\gamma_{\mathrm{S},j}, \frac{D}{m_1}, m_1)$. Considering $\frac{\partial \mathbb{E}[n_a]}{\partial m_1} = -\sum_{j=1}^{J} \frac{\partial \varepsilon_{\mathrm{S},j}}{\partial m_1}$, we therefore conclude $\frac{\partial \mathbb{E}[n_a]}{\partial m_1} \geq 0$. In other words, $\mathbb{E}[n_a]$ is increasing in m_1.

Appendix B
Proof of Lemma 2

We prove the lemma by distinguishing the following 3 cases:

- *Case 1:* The relay selected by the source-driven strategy decodes the data packet successfully and at the same time this relay has the highest channel gain of the relay-destination link than other relays in the active relay set. In this case, the relay-driven strategy selects the same relay as the source-driven one, *i.e.*, results in the same reliability.
- *Case 2:* The relay selected by the source-driven strategy decodes the data packet successfully but this relay is not the one with the highest channel gain (of the relay-destination link). Then, the relay-driven strategy selects a relay with a stronger second hop and therefore leads to a lower overall error probability.
- *Case 3:* The relay selected by the source-driven strategy decodes the data packet incorrectly. The source-driven strategy has an error probability of 1. At the same time, the relay-driven strategy definitely has a lower error probability as long as the active relay set is not empty.

Note that the probabilities of Case 2 and Case 3 definitely have positive values. Hence, the error probability of the relay-driven strategy is lower than that of the source-driven one.

$$\begin{aligned}\frac{\partial^2 \mathcal{P}\left(\gamma, r, t\right)}{\partial t^2} &= \frac{2 D t^{-3} \phi}{\sqrt{2\pi}} e^{-\frac{\left[\phi t \log_2(1+\gamma) - \phi D t^{-1}\right]^2}{2}}\\ &\quad + \frac{\phi^3}{\sqrt{2\pi}} e^{-\frac{\left[\phi t \log_2(1+\gamma) - \phi D t^{-1}\right]^2}{2}}\\ &\quad \times \left[t \log_2\left(1+\gamma\right) - D t^{-1}\right]\left[\log_2(1+\gamma) + D t^{-2}\right]^2.\end{aligned} \tag{33}$$

Appendix C
Proof of Lemma 3

We prove Lemma 3 by contradiction. We first assume that there exists an optimal solution (m_1', m_2') satisfying the constraint with strict inequality, *i.e.*, $M - (m_1' + m_2') = n > 0$. The relay j is selected as the optimal relay. Hence, the optimal reliability with this solution is $\varepsilon_{\mathrm{SD}}\left(m'\_1, m'\_2\right) = \varepsilon_{\mathrm{tot},j}\left(m'\_1, m'\_2\right)$. On the other hand, we can find another feasible solution $(m_1'' = m_1', m_2'' = m_2' + n)$. Recall that in the proof of Lemma 1, we have shown that $\frac{\partial \mathcal{P}}{\partial m} \leq 0$. Hence, $\forall j \in \mathcal{J}$ we conclude that $\varepsilon_{\mathrm{S},j} = \mathcal{P}(\gamma_{\mathrm{S},j}, \frac{D}{m_1}, m_1)$ is decreasing in m_1 but constant in m_2 and that $\varepsilon_{j,\mathrm{D}} = \mathcal{P}(\gamma_{j,\mathrm{D}}, \frac{D}{m_2}, m_2)$ is decreasing in m_2 but constant in m_1. In comparison to the assumed optimal m_1' and m_2', the solution (m_1'', m_2'') does not change the value of $\varepsilon_{\mathrm{S},j}$ but results in a lower value of $\varepsilon_{j,\mathrm{D}}$ for each $j \in \mathcal{J}$. According to (15), the solution (m_1'', m_2'') leads to a lower value of $\varepsilon_{\mathrm{tot},j}$, *i.e.*, the assumption that (m_1', m_2') is the optimal solution of the problem in (22) is violated.

Appendix D
Proof of Lemma 4

According to (32), we have the second order derivative of $\mathcal{P}$ with respect to t, given by (33).

Recall that we consider an ultra-reliable scenario, where the error probability of a transmission is lower than 0.5. Hence, we have $\mathcal{P}\left(\gamma, r, t\right) = Q\left(\phi t \log_2\left(1+\gamma\right) - \phi D t^{-1}\right) < 0.5 \Rightarrow t\log_2\left(1+\gamma\right) - D t^{-1} > 0$. Therefore, $\frac{\partial^2 \mathcal{P}}{\partial t^2} \geq 0$ holds. In addition, as $t = m^{\frac{1}{2}}$, it is easy to show $\frac{\partial t}{\partial m} = \frac{1}{2} m^{-\frac{1}{2}}$ and $\frac{\partial^2 t}{\partial m^2} = -\frac{1}{4} m^{-\frac{3}{2}}$.

Recall that it has been shown in (32) that $\frac{\partial \mathcal{P}}{\partial t} \leq 0$. Then, we obtain

$$\frac{\partial^2 \mathcal{P}}{\partial m^2} = \underbrace{\frac{\partial^2 \mathcal{P}}{\partial t^2}}_{\geq 0}\left(\frac{\partial t}{\partial m}\right)^2 + \underbrace{\frac{\partial \mathcal{P}}{\partial t}}_{\leq 0}\underbrace{\frac{\partial^2 t}{\partial m^2}}_{\leq 0} \geq 0. \tag{34}$$

Note that $f_j(m_1, m_2) = \varepsilon_{\mathrm{S},j} + \varepsilon_{j,\mathrm{D}} = \mathcal{P}(\gamma_{\mathrm{S},j}, \frac{D}{m_1}, m_1) + \mathcal{P}(\gamma_{j,\mathrm{D}}, \frac{D}{m_2}, m_2)$. According to (34), the Hessian matrix of $f_j(m_1, m_2)$ with respect to (m_1, m_2) is positive semi-definite:

$$\begin{aligned}\mathbf{H}_j &= \begin{pmatrix} \frac{\partial^2 f_j}{\partial m_1^2} & \frac{\partial^2 f_j}{\partial m_1 \partial m_2} \\ \frac{\partial^2 f_j}{\partial m_2 \partial m_1} & \frac{\partial^2 f_j}{\partial m_2^2} \end{pmatrix}\\ &= \begin{pmatrix} \frac{\partial^2 P\left(\gamma_{\mathrm{S},j}, \frac{D}{m_1}, m_1\right)}{\partial m_1^2} & 0 \\ 0 & \frac{\partial^2 P\left(\gamma_{\mathrm{D},j}, \frac{D}{m_2}, m_2\right)}{\partial m_2^2} \end{pmatrix} \geq 0.\end{aligned} \tag{35}$$

Thus, $f_j(m_1, m_2)$ is convex in (m_1, m_2) for all $j \in \mathcal{J}$.

Appendix E
Proof of Lemma 5

Lemma 5 can also be proved by contradiction. Similar to the proof of Lemma 3, we first assume that there exists an optimal solution (m_1', m_2') satisfying the constraint with strict inequality, *i.e.*, $M - (m_1' + m_2') = n > 0$, and find another feasible solution $(m_1'' = m_1', m_2'' = m_2' + n)$. Recall that $\frac{\partial \mathcal{P}}{\partial m} \leq 0$. Hence, $\forall j \in \mathcal{J}^{\circ}$ we conclude that $\prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k} = \prod_{k=0}^{j-1} \mathcal{P}(\gamma_{\mathrm{S},k}, \frac{D}{m_1}, m_1)$ is decreasing in m_1 but constant in m_2 and that $\varepsilon_{j,\mathrm{D}} = \mathcal{P}(\gamma_{j,\mathrm{D}}, \frac{D}{m_2}, m_2)$ is decreasing in m_2 but constant in m_1.

In comparison to the assumed optimal m_1' and m_2', the solution (m_1'', m_2'') does not change the value of $\prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k}$ but results in lower values for all $\varepsilon_{j,\mathrm{D}}$, $j \in \mathcal{J}^{\circ}$. Hence, a lower value of $\sum_{j \in \mathcal{J}^{\circ}} \varepsilon_{j,\mathrm{D}} \prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k}$ is introduced by the solution (m_1'', m_2''),

i.e., the assumption that (m_1', m_2') is the optimal solution of the problem in (29) is violated. Therefore, we conclude that $m_1 + m_2 = M$ is always guaranteed in the optimal solution of the optimization problem given in (29).

APPENDIX F
PROOF OF LEMMA 6

According to (34), we have $\frac{\partial^2 \mathcal{P}}{\partial m^2} \geq 0$. Hence, $\varepsilon_{j,\mathrm{D}}$ is convex in m_2 and $\varepsilon_{\mathrm{S},k}$ is convex in m_1. According to Lemma 5, we have $m_2 = M - m_1$. Then, $\frac{\partial m_2}{\partial m_1} = -1$ and $\frac{\partial^2 m_2}{\partial {m_1}^2} = 0$ hold. Hence, we have $\frac{\partial^2 \varepsilon_{j,\mathrm{D}}}{\partial {m_1}^2} = \frac{\partial^2 \varepsilon_{j,\mathrm{D}}}{\partial {m_2}^2}(\frac{\partial m_2}{\partial m_1})^2 + \frac{\partial \varepsilon_{j,\mathrm{D}}}{\partial m_2}\frac{\partial^2 m_2}{\partial {m_1}^2} = \frac{\partial^2 \varepsilon_{j,\mathrm{D}}}{\partial {m_2}^2}$. In other words, $\varepsilon_{j,\mathrm{D}}$ is also convex in m_2. Then, $\varepsilon_{j,\mathrm{D}} \prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k}$ becomes actually a product of non-negative convex functions.

In the following, we prove that the product of non-negative convex functions $\mathcal{P}_1(m_2)$ and $\mathcal{P}_2(m_2)$, given by $\mathcal{G}(m_2) = \mathcal{P}_1(m_2)\mathcal{P}_2(m_2)$, is also a non-negative convex function. Define $\lambda \in [0,1]$ and consider two points $m_2' \geq m_2''$ in the feasible set. As $\mathcal{P}_1(m_2)$ and $\mathcal{P}_2(m_2)$ are convex functions, we have: $\mathcal{P}_1\left(\lambda m_2' + (1-\lambda)\, m_2''\right) \leq \lambda \mathcal{P}_1\left(m_2'\right) + (1-\lambda)\,\mathcal{P}_1\left(m_2''\right)$ and $\mathcal{P}_2\left(\lambda m_2' + (1-\lambda)\, m_2''\right) \leq \lambda \mathcal{P}_2\left(m_2'\right) + (1-\lambda)\,\mathcal{P}_2\left(m_2''\right)$. Therefore, we have

$$\begin{aligned}
&\mathcal{G}\left(\lambda m_2' + (1-\lambda)\, m_2''\right) \\
&= \mathcal{P}_1\left(\lambda m_2' + (1-\lambda)\, m_2''\right) \mathcal{P}_2\left(\lambda m_2' + (1-\lambda)\, m_2''\right) \\
&\leq \left[\lambda \mathcal{P}_1(m_2') + (1-\lambda)\,\mathcal{P}_1(m_2'')\right] \left[\lambda \mathcal{P}_2(m_2') + (1-\lambda)\,\mathcal{P}_2(m_2'')\right]
\end{aligned}$$

Hence, $\mathcal{G}(m_2)$ is convex in m_2 as

$$\begin{aligned}
&\lambda \mathcal{G}\left(m_2'\right) + (1-\lambda)\,\mathcal{G}\left(m_2''\right) - \mathcal{G}\left(\lambda m_2' + (1-\lambda)\, m_2''\right) \\
&\geq \lambda \mathcal{G}\left(m_2'\right) + (1-\lambda)\,\mathcal{G}\left(m_2''\right) \\
&\quad - \left[\lambda \mathcal{P}_1\left(m_2'\right) + (1-\lambda)\,\mathcal{P}_1(m_2'')\right] \left[\lambda \mathcal{P}_2(m_2') + (1-\lambda)\,\mathcal{P}_2(m_2'')\right] \\
&= \lambda\,(1-\lambda)\left(\mathcal{P}_1\left(m_2'\right) - \mathcal{P}_1\left(m_2''\right)\right)\left(\mathcal{P}_2\left(m_2'\right) - \mathcal{P}_2\left(m_2''\right)\right) \geq 0.
\end{aligned}$$

By iteratively applying this result, the convexity of $\varepsilon_{j,\mathrm{D}} \prod_{k=0}^{j-1} \varepsilon_{\mathrm{S},k}$ can be determined. Finally, the sum of convex functions is also convex, *i.e.*, the objective of the problem in (30) is convex.

REFERENCES

[1] Y. Hu, C. Schnelling, Y. Amraue, and A. Schmeink, "Finite blocklength performance of a multi-relay network with best single relay selection," in *Proc. IEEE Int. Symp. Wireless Commun. Syst.*, Bologna, Italy, Aug. 2017, pp. 122–127.

[2] K. Campbell, J. Diffley, B. Flanagan, B. Morelli, B. O'Neil, and F. Sideco, "The 5G economy: How 5G technology will contribute to the global economy," IHS Econ. IHS Technol., Tech. Rep., Jan. 2017.

[3] C. She, C. Yang, and T. Q. S. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 72–78, Jun. 2017.

[4] Y. Hu, M. C. Gursoy, and A. Schmeink, "Relaying-enabled ultra-reliable low latency communications in 5G," *IEEE Netw.*, vol. 32, no. 2, pp. 62–68, Mar./Apr. 2018.

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

[6] J. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[7] Y. Huang, F. Al-Qahtani, C. Zhong, Q. Wu, J. Wang, and H. Alnuweiri, "Performance analysis of multiuser multiple antenna relaying networks with co-channel interference and feedback delay," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 59–73, Jan. 2014.

[8] Y. Hu, J. Gross, and A. Schmeink, "QoS-constrained energy efficiency of cooperative ARQ in multiple DF relay systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 2, pp. 848–859, Feb. 2016.

[9] W. Guo, S. Zhou, Y. Chen, S. Wang, X. Chu, and Z. Niu, "Simultaneous information and energy flow for IoT relay systems with crowd harvesting," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 143–149, Nov. 2016.

[10] T. Lv, Z. Lin, P. Huang, and J. Zeng, "Optimization of the energy-efficient relay-based massive IoT network," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3043–3058, Aug. 2018.

[11] S. Atapattu, Y. Jing, H. Jiang, and C. Tellambura, "Relay selection and performance analysis in multiple-user networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1517–1529, Aug. 2013.

[12] L. Liu, C. Hua, C. Chen, and X. Guan, "Semidistributed relay selection and power allocation for outage minimization in cooperative relaying networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 295–305, Jan. 2017.

[13] J. Kim, K. Kim, and J. Lee, "Energy-efficient relay selection of cooperative HARQ based on the number of transmissions over Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 610–621, Jan. 2017.

[14] V. Strassen, "Asymptotische Abschätzungen in shannons informationstheorie," in *Proc. Trans. Third Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723.

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

[16] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Dispersion of the Gilbert–Elliott channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1829–1848, Apr. 2011.

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

[18] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Finite-blocklength channel coding rate under a long-term power constraint," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 2067–2071.

[19] Y. Hu, J. Gross, and A. Schmeink, "On the capacity of relaying with finite blocklength," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1790–1794, Mar. 2016.

[20] Y. Hu, J. Gross, and A. Schmeink, "On the performance advantage of relaying under the finite blocklength regime," *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 779–782, May 2015.

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

[22] Y. Li, M. C. Gursoy, and S. Velipasalar, "Throughput of two-hop wireless channels with queueing constraints and finite blocklength codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 2599–2603.

[23] L. Wang, "Polar coding for relay channels," in *Proc. IEEE Int. Symp. Inf. Theory*, HongKong, Jun. 2015, pp. 1532–1536.

[24] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund, "Polar codes for cooperative relaying," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3263–3273, Nov. 2012.

[25] M. Karzand, "Polar codes for degraded relay channels," in *Proc. Int. Zurich Seminar Commun.*, Zurich, Switzerland, Mar. 2012, pp. 59–62.

[26] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[27] D. S. Karas, K. N. Pappi, and G. K. Karagiannidis, "Smart decode-and-forward relaying with polar codes," *IEEE Wireless Commun. Lett.*, vol. 3, no. 1, pp. 62–65, Feb. 2014.

[28] R. Blasco-Serrano, R. Thobaben, V. Rathi, and M. Skoglund, "Polar codes for compress-and-forward in binary relay channels," in *Proc. Conf. Rec. 44th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1743–1747.

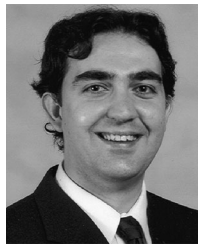
[29] G. Ozcan and M. C. Gursoy, "Throughput of cognitive radio systems with finite blocklength codes," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2541–2554, Nov. 2013.

[30] C. She, C. Yang, and T. Q. S. Quek, "Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2266–2280, May 2018.

[31] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
[32] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint-Petersburg, Russia, Jul. 2011, pp. 1–5.
[33] I. Tal and A. Vardy, "List decoding of polar codes," *CoRR*, vol. abs/1206.0050, May 2012.
[34] A. Eslami and H. Pishro-Nik, "A practical approach to polar codes," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 16–20.
[35] D.-M. Shin, S.-C. Lim, and K. Yang, "Design of length-compatible polar codes based on the reduction of polarizing matrices," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2593–2599, Jul. 2013.
[36] V. Miloslavskaya, "Shortened polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4852–4865, Sep. 2015.
[37] C. Schnelling, M. Rothe, R. Mathar, and A. Schmeink, "Rateless codes based on punctured polar codes," in *Proc. Int. Symp. Wireless Commun. Syst.*, Aug. 2018, pp. 2593–2599.
[38] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 1496–1500.
[39] K. Niu, K. Chen, and J.-R. Lin, "Beyond turbo codes: Rate-compatible punctured polar codes," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2013, pp. 3423–3427.

Yulin Hu (S'11–M'15–SM'18) received the M.Sc.E.E. degree from USTC, China, in 2011. He received the Ph.D. degree under the supervision of Prof. Anke Schmeink at RWTH Aachen University, Aachen, Germany and Prof. James Gross at KTH Royal Institute of Technology, Stockholm, Sweden, in Dec. 2015 and received the Ph.D.E.E. degree Hons. from RWTH Aachen University where he was a Postdoctoral Research Fellow since January to December 2016. From May to July in 2017, he was a Visiting Scholar with Prof. M. Cenk Gursoy at Syracuse University, USA. Since 2017, he has been working as a Senior Researcher and Project Lead with ISEK Research Group, RWTH Aachen University. He has been invited to contribute submissions to multiple conferences. His research interests include information theory and optimal design of wireless communication systems. Prof. Hu was a recipient of the IFIP/IEEE Wireless Days Student Travel awards in 2012. He received two best paper awards at IEEE International Symposium on Wireless Communication Systems (ISWCS) 2017 and IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2017, respectively. He served as a TPC member for many conferences and was an Organizer of a special session in IEEE ISWCS 2018. He is currently serving as an Editor for *Physical Communication* (Elsevier), *EURASIP Journal on Wireless Communications and Networking,* and the Lead Editor of the Urllc-LoPIoT Special Issue in *Physical Communication*.

Christopher Schnelling received the diploma in computer engineering in 2014, and the master's degree in management, business, and economics in 2016, both from RWTH Aachen University, Aachen, Germany. He is currently pursuing a Ph.D. degree in electrical engineering. His research interests include channel coding and polar codes.



M. Cenk Gursoy (S'01–M'04–SM'16) received the B.S. degree with high distinction in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 1999 and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2004. Between 2004 and 2011, he was a Faculty Member with the Department of Electrical Engineering, University of Nebraska-Lincoln (UNL), Lincoln, NE, USA. He is currently a Professor with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA. His research interests include the general areas of wireless communications, information theory, communication networks, signal processing, and machine learning. Prof. Gursoy was a recipient of the Gordon Wu Graduate Fellowship from Princeton University between 1999 and 2003. He is currently a member of the editorial boards of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and he is an Area Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He also served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS between 2010 and 2015, IEEE COMMUNICATIONS LETTERS between 2012 and 2014, IEEE Journal on Selected Areas in Communications–Series on Green Communications and Networking (JSAC-SGCN) between 2015 and 2016, Physical Communication (Elsevier) between 2010 and 2017, and IEEE TRANSACTIONS ON COMMUNICATIONS between 2013 and 2018. He has been the Co-Chair of the 2017 International Conference on Computing, Networking and Communications (ICNC)–Communication QoS and System Modeling Symposium, the Co-Chair of 2019 IEEE Global Communications Conference (Globecom)–Wireless Communications Symposium, and the Co-Chair of 2019 IEEE Vehicular Technology Conference Fall–Green Communications and Networks Track. He received an NSF CAREER Award in 2006. More recently, he received the EURASIP Journal of Wireless Communications and Networking Best Paper Award, 2017 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications Best Paper Award, 2017 IEEE Green Communications & Computing Technical Committee Best Journal Paper Award, UNL College Distinguished Teaching Award, and the Maude Hammond Fling Faculty Research Fellowship. He is the Aerospace/Communications/Signal Processing Chapter Co-Chair of IEEE Syracuse Section.

Anke Schmeink (M'11–SM'18) received the diploma degree in mathematics with a minor in medicine and the Ph.D. degree in electrical engineering and information technology from RWTH Aachen University, Aachen, Germany, in 2002 and 2006, respectively. She was a Research Scientist with Philips Research before joining RWTH Aachen University in 2008, where she has been an Associate Professor since 2012. She spent several research visits with the University of Melbourne and with the University of York. Her research interests include information theory, systematic design of communication systems and bioinspired signal processing. Prof. Schmeink is a member of the Young Academy, North Rhine-Westphalia Academy of Science.