Permutation-Based Transmissions in Finite Blocklength Regime: Efficient and Effective Resource Utilisation

Wenyao Li, Yuli Yang, Bingli Jiao

***Abstract*—In this paper, the concept of permutation-based transmissions is developed at the transport layer of short-packet communications, where the application-layer data is divided into two parts: one is conveyed by the permutation of various lengths in a group of packets, and the other is encapsulated into these various-length packets. The former part, referred to as permutation-conveyed data unit (PCDU), improves the goodput and reduces the latency. From a finite blocklength perspective, the maximal payload rate of permutation-based transmissions is formulated, based on which the goodput and latency are derived and obtained in analytical forms. Moreover, the spectral efficiency and energy efficiency are analysed. Illustrative numerical results on the performance and resource utilisation comparisons not only substantiate the advantages of our permutation-based transmissions over conventional transport-layer encapsulation schemes, but also provide useful tools and specifications for the PCDU design in short-packet communications.**

***Index Terms*—Permutation-conveyed data unit (PCDU), short-packet communications, finite blocklength, maximal payload rate, goodput, latency, spectral efficiency, energy efficiency.**

I. Introduction

As a major driver for the development of future wireless communication infrastructures, the Internet of Everything (IoE) promises to fundamentally change how we live and work. The foundation of IoE-enabled services is constituted by ultra-reliable and low-latency communications (URLLCs), where the target delay for individual interfaces is expected to be lower than 1ms [1], [2]. For supporting mission-critical services, e.g., in real-time detection and control, short packets are utilised to meet the delay constraints [3], [4].

Since Shannon's classical framework established for infinite blocklength is not applicable to URLLCs using short packets, the information-theoretic advances in finite blocklength regime have created a powerful framework for the design, analysis, and optimisation of short-packet communications [5], [6]. The seminal works [7], [8] investigated maximal coding rate of short packets in the finite blocklength regime and obtained non-asymptotic bounds on it. Inspired by these works, a series of asymptotic approximations of converse and achievability bounds on the maximal coding rates have been developed in the finite-blocklength information theory [9], [10]. Particularly for the short-packet communications over wireless channels of practical interest, various transmission protocols and short channel codes have been designed and optimised to achieve URLLCs at low cost and low computational complexity [11], [12].

W. Li and B. Jiao are with the Department of Electronics, Peking University, Beijing 100871, China (e-mail: liwenyao@stu.pku.edu.cn, jiaobl@pku.edu.cn).

Y. Yang is with the School of Engineering, University of Lincoln, Lincoln LN6 7TS, U.K. (e-mail: yyang@lincoln.ac.uk).

On the other hand, a huge number of connected devices, giving rise to the IoE, have been challenging how to efficiently utilise the scarce resources of spectrum and energy. A large volume of attempts have been committed to the enhancement of spectral efficiency and energy efficiency for wireless communications in the IoE [13], [14]. Among them, the philosophy of permutation modulation at the physical layer is one of the promising solutions, which exploits the physical resources in a smart way [15]. It utilises the activation of physical-layer resource units, e.g., in space, frequency, time or code domain, to convey a part of information bits, rather than solely relies on classical modulation, i.e., amplitude and/or phase modulation, to map and physically transmit them. The physical-layer community has been encouraged by this philosophy to conceive the ideas of spatial modulation [16]–[18], frequency index modulation [19], [20], and time index modulation [21], [22]. The permutation philosophy has been used at the physical layer in a wide range of applications, such as relay channels [23], [24], cognitive radio networks [25], cooperative communications [26], physical-layer security [27]–[29], aeronautical *ad-hoc* networking [30], and reconfigurable intelligent surfaces [31], [32].

As reliable end-to-end connectivity in the IoE relies on transport-layer protocols that operate above physical channels, the philosophy of permutation modulation has been developed at the transport layer to increase the goodput and reduce the latency, where a portion of application-layer data is conveyed by the permutation of activated transport-layer resource units, e.g., the transmission order or the packet length. In [33] and [34], a portion of information bits in the data unit (DU) of a packet are referred to as opportunistic bits (OBs) and conveyed by the index of the time slot (TS) when the packet is transmitted. In [35] and [36], the permutation-based encapsulation was proposed, where a portion of application-layer data is conveyed by the permutation with repetition of various DU lengths in a group of packets, rather than encapsulated into the packets.

In comparison to their long-packet counterparts, short-packet communications need to address two associated challenges with the aid of finite-blocklength information theory [37], [38]. Firstly, the transceiver's buffer size is limited by the IoE device configuration and the service latency requirement. Secondly, the header length of a packet cannot be neglected, as it is comparable to the DU length. Motivated by this, we will optimise the packet structure for the permutation-based encapsulation in short-packet communications based on the advances of finite-blocklength information theory.

Our main objective is to maximise the performance and resource utilisation efficiency for the application of permutation-

TABLE I
CONTRASTING THE NOVELTY OF THIS WORK TO THE LITERATURE.

Novelty	**This Work**	[16]–[32]	[33]–[36]
Permutation Philosophy	✓	✓	✓
Transport-Layer Encapsulation	✓		✓
Goodput	✓		✓
Latency	✓		✓
Resource Utilisation	✓		✓
Optimal Design in Finite Blocklength Regime	✓		

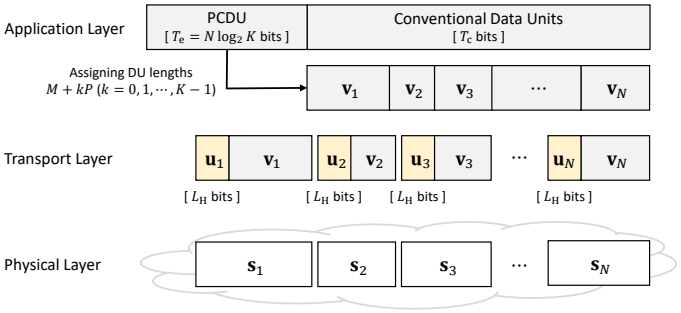


Fig. 1. Data encapsulation in a permutation-based transmission.

based transmissions into practice. The novelties of this work are boldly and explicitly contrasted to the state-of-the-art in Table I at a glance. More specifically, our contributions in this work are three-fold as below.

- *Payload Rate Optimisation:* With the aid of finite-blocklength information theory, the maximal payload rate of permutation-based transmissions is formulated and maximised through optimising the short-packet structure design. Specifically, increasing the number of alternative DU lengths brings extra bits conveyed in a permutation-based transmission. However, the average DU length will be decreased under the constraint of maximum DU length, which leads to the payload loss due to the influence of header in the finite blocklength regime. Closed-form solutions are provided to achieve the optimal balance.
- *Performance Evaluation:* On the basis of the maximised maximal payload rate, the performance of our permutation-based encapsulation within short-packet communications is evaluated and compared with that of conventional transmissions, in the metrics of goodput and latency.
- *Resource Utilisation Analysis:* To investigate our permutation-based encapsulation in short-packet communications from the perspective of resource utilisation efficiency, its spectral and energy efficiency gains are formulated and maximised in closed-form expressions through optimising the packet structure design.

The remainder of this paper is organized as follows. In Section II, the system model of permutation-based transmissions is presented, followed by the introduction to information-theoretic analysis in finite blocklength regime. Section III formulates the maximal payload rate of permutation-based transmissions through optimising the structure of short packets, based on which Section IV analyses the goodput and latency improvements offered by our permutation-based transmissions. Further, the resource utilisation efficiency is addressed in Section V. Finally, the paper is concluded in Section VI.

II. SYSTEM MODEL

The permutation-based transmission has been proposed to reduce the header redundancy at the transport layer and improve the spectral efficiency at the physical layer. In this section, the permutation-based transmission is developed in finite blocklength regime, for further balancing the payload length and the header length.

A. Permutation-Based Transmissions

In a permutation-based transmission, a group of N various-length packets, denoted by $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N$, are delivered over the network interface at physical layer. Each packet $\mathbf{s}_n$ consists of a header $\mathbf{u}_n$ of length L_{H} bits and a DU $\mathbf{v}_n$ of a various length, $n = 1, 2, \cdots, N$. This is the structure of a conventional packet, where the DU contains application-layer information bits, and the packet length is recorded in a formatted metadata field of the header. Different from the conventional packet structure, the permutation-based transmission chooses the DU lengths in a group from a set of cardinality K, denoted by $\{L_0, L_1, \cdots, L_{K-1}\}$. Herein, these lengths are set to an arithmetic sequence [35], [36], i.e.,

$$L_k = M + kP, \quad k = 0, 1, \cdots, K-1, \tag{1}$$

where the initial term M is the shortest DU length, and P is the common difference of successive lengths, $M, P \in \mathbb{N}$.

The encapsulation of application-layer data into a permutation-based transmission is shown in Fig. 1, where the application-layer information bits are divided into two portions: the permutation-conveyed DU (PCDU), composed of T_e bits, and the conventional DUs, composed of T_c bits. The T_e-bit PCDU is mapped onto the permutation with repetition in allocating the K available lengths into the N DUs, while the other T_c bits are encapsulated into these N DUs in a conventional way. As the length of each DU is selected from K candidates, the lengths of N DUs form K^N permutation patterns in total, and each pattern is determined by a PCDU. In this way, each PCDU has $T_e = N \log_2 K$ bits, which are the extra bits conveyed though a permutation-based transmission. For example, the 16 permutations in the case of $N = 4$ and $K = 2$ are listed in Table II, where the PCDU has $T_e = 4$ bits. To convey the same number of application-layer information bits, a permutation-based transmission requires less resource units at the physical layer than a conventional transmission, as the PCDU is not physically transmitted but conveyed by the DU length permutation, which does not consume any physical resources.

The probability that the DU length L_k occurs in an encapsulation is denoted by γ_k, with $0 < \gamma_k < 1$ and $\sum_{k=0}^{K-1} \gamma_k = 1$.

TABLE II
16 PERMUTATIONS IN THE CASE OF $N = 4$ AND $K = 2$.

Permutation ♯	PCDU	DU Length Allocation	
1	0000	DU 1 $\leftarrow M$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M$
2	0001	DU 1 $\leftarrow M$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M + P$
3	0010	DU 1 $\leftarrow M$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M$
4	0011	DU 1 $\leftarrow M$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M + P$
5	0100	DU 1 $\leftarrow M$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M$
6	0101	DU 1 $\leftarrow M$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M + P$
7	0110	DU 1 $\leftarrow M$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M$
8	0111	DU 1 $\leftarrow M$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M + P$
9	1000	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M$
10	1001	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M + P$
11	1010	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M$
12	1011	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M + P$
13	1100	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M$
14	1101	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M$	DU 4 $\leftarrow M + P$
15	1110	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M$
16	1111	DU 1 $\leftarrow M + P$	DU 2 $\leftarrow M + P$
		DU 3 $\leftarrow M + P$	DU 4 $\leftarrow M + P$

Thus, the average length of a DU in permutation-based transmissions is given by

$$\bar{L} = M + \sum_{k=0}^{K-1} \gamma_k P k. \tag{2}$$

In practice, the maximum DU length, denoted by $L_{\max}$, is determined by specific transport-layer protocols, e.g, Transmission Control Protocol (TCP) or User Datagram Protocol (UDP). Explicitly, we have

$$M + P(K-1) \leqslant L_{\max}. \tag{3}$$

Moreover, the amount of application-layer data is limited by the transceiver's buffer size, denoted by F_s. As N packets, each containing a header of length L_{H} and a DU of the maximum length $L_{\max}$, are loaded into the buffer, we have

$$N[L_{\mathrm{H}} + M + P(K-1)] \leqslant N(L_{\mathrm{H}} + L_{\max}) = F_s. \tag{4}$$

Given these constraints, we will optimise the packet structure design for permutation-based transmissions, based on the information-theoretic analysis in finite blocklength regime.

B. Information-Theoretic Analysis

In Shannon's classical analysis framework [39], the channel capacity is used to define the maximal coding rate in the case of infinite code length with an arbitrarily small error probability. However, the code length in practical transmissions is finite and the error probability cannot be made arbitrarily small. Therefore, the advances in finite blocklength regime, e.g., [8], are preferred in the information-theoretic analysis of practical transmissions. Given the blocklength B, the block error probability ϵ, and the signal-to-noise power ratio (SNR) ρ, the maximum number of information bits conveyed over the block is given by [8, Eq. (1)]

$$D(B, \epsilon, \rho) \approx BC(\rho) - \sqrt{BV(\rho)}Q^{-1}(\epsilon), \tag{5}$$

where the channel capacity $C(\rho) = (1/2)\log_2(1+\rho)$. The channel dispersion $V(\rho) = (\rho/2)[(\rho+2)/(\rho+1)^2](\log_2 e)^2$, and $Q^{-1}(\cdot)$ is the inverse function of $Q(x) = \int_x^\infty (1/\sqrt{2\pi})\exp(-t^2/2)dt$.

At the physical layer, although the header $\mathbf{u}_n$ and the DU $\mathbf{v}_n$ are usually encoded with different-rate channel codes for a better protection of the header, the average coding rate is used to scale the transmission rate from the information-theoretic point of view, where the encapsulation process is deemed to be a consistent mapping from the transport-layer bits to the physical-layer symbols.

On the basis of finite blocklength regime, we investigate the optimal packet structures in permutation-based transmissions for two scenarios, concerning the minimum values of the shortest DU length M and the common difference of successive lengths P. Firstly, the practical scenario with popular transport-layer protocols, e.g., TCP or UDP, is considered. In this scenario, a formatted metadata field, of length l in the header $\mathbf{u}_n$, is used to specify the packet length in the unit of bytes, which equals to the header length L_{H} plus the DU length. Therefore, the minimum values of M and P, denoted by $M_{\min}$ and $P_{\min}$, are both 8 bits. In addition, a loose constraint on the maximum DU length $L_{\max}$ is

$$L_{\max} \leqslant (2^l - 1) \times 8 - L_{\mathrm{H}} < (2^{L_{\mathrm{H}}} - 1) \times 8 - L_{\mathrm{H}}. \tag{6}$$

Subsequently, to demonstrate full benefits achieved by our permutation-based transmissions, the ideal scenario with $M_{\min} = P_{\min} = 1$ bit is taken into account, which will provide effectual specifications for permutation-based packet design in advanced protocols.

III. MAXIMAL PAYLOAD RATE

In this section, the maximal payload rate is formulated to scale the effective data rate in permutation-based transmissions, excluding the contribution from metadata in headers.

At the physical layer, the average amount of payload data within a block is $T_e/N + \bar{L}$, which is conveyed by a permutation-based packet of average length $L_{\mathrm{H}} + \bar{L}$. Therefore, the payload rate of permutation-based transmissions is given by $(T_e/N + \bar{L})/B$, where B is the blocklength, defined as the number of symbols used to convey a packet of average length $L_{\mathrm{H}} + \bar{L}$ over the network interface.

According to (5), the minimum blocklength required to convey $L_{\mathrm{H}}+\bar{L}$ bits, with the target packet error probability ϵ at the SNR ρ, is expressed as

$$B^* = D^{-1}(L_{\mathrm{H}}+\bar{L},\epsilon,\rho), \tag{7}$$

where $D^{-1}(\kappa,\epsilon,\rho)$ denotes the inverse of the function $B \mapsto D(B,\epsilon,\rho)$, given by

$$D^{-1}(\kappa,\epsilon,\rho) = \left(\frac{\sqrt{V(\rho)}Q^{-1}(\epsilon)+\sqrt{V(\rho)(Q^{-1}(\epsilon))^2+4C(\rho)\kappa}}{2C(\rho)}\right)^2. \tag{8}$$

Subsequently, the maximal payload rate of a permutation-based transmission is obtained by

$$R = \frac{T_e/N+\bar{L}}{B^*} = \frac{\log_2 K+\bar{L}}{D^{-1}\left(L_{\mathrm{H}}+\bar{L},\epsilon,\rho\right)}. \tag{9}$$

Without loss of generality, allocating a fair DU length follows the discrete uniform distribution, i.e., $\gamma_k = 1/K$ for $k = 0, 1, \cdots, K-1$. Thus, the average length of a DU, given by (2), is rewritten as

$$\bar{L} = M + \frac{K-1}{2}P. \tag{10}$$

Substituting (10) into (9), we rewrite the maximal payload rate as

$$R(M,K,P) = \frac{\log_2 K+M+P(K-1)/2}{D^{-1}\left(L_{\mathrm{H}}+M+P(K-1)/2,\epsilon,\rho\right)}. \tag{11}$$

Obviously, there are six parameters in (11). Among them, L_{H}, ϵ, and ρ are determined by the transport-layer protocol, the reliability requirement, and the channel condition, respectively. Then, the packet structure design will be optimised with respect to the shortest DU length M, the common difference of successive lengths P, and the number of available DU lengths K for maximising the maximal payload rate.

Now that ρ and ϵ are constants during the maximisation of R, we will replace the notations $C(\rho)$, $V(\rho)$, and $Q^{-1}(\epsilon)$ in (8) with C, V, and Q^{-1}, respectively, for brevity's sake.

Concerning the constraints introduced in Section II-A, the maximisation of the maximal payload rate R is formulated by an optimisation problem as

$$\mathcal{P}1: R^* = \max_{M,K,P} R(M,K,P) \tag{12}$$

$$\text{s.t.} \quad M+P(K-1) \leqslant L_{\max}, \tag{12a}$$

$$N[L_{\mathrm{H}}+M+P(K-1)] \leqslant F_s, \tag{12b}$$

$$M,P,K \in \mathbb{N}, \tag{12c}$$

where the objective function $R(M,K,P)$ is given by (11). Since $R(M,K,P)$ is independent of N, the constraint (12b) can be attained by arbitrarily adjusting N. Therefore, the constraint (12b) will be removed and the constraint (12a) will be the primary concern in the optimisation process hereafter.

Next, we will solve $\mathcal{P}1$ for two scenarios, concerning the minimum values of M and P, i.e., the practical scenario with $M_{\min} = P_{\min} = 8$ bits and the ideal scenario with $M_{\min} = P_{\min} = 1$ bit.

A. Practical Scenario

Herein, the optimisation problem $\mathcal{P}1$ is solved in three steps: *1)* converting the inequality constraint (12a) into an equality one by introducing the constraint on the longest DU length $\alpha = M+P(K-1)$; *2)* optimising P and K under the converted equality constraint, given that the shortest DU length M is deemed to be a constant; *3)* optimising M based on the optimised P^* and K^*.

1) Conversion of (12a): Given the longest DU length $\alpha = M+P(K-1) \leqslant L_{\max}$, the shortest DU length M in (1) can be expressed as $M = \alpha - P(K-1)$. Then, the optimisation problem $\mathcal{P}1$ in (12) is rewritten as

$$\mathcal{P}2: R^* = \max_{\alpha,K,P} R(\alpha,K,P) \tag{13}$$

$$\text{s.t.} \quad \alpha \leqslant L_{\max}, \tag{13a}$$

$$P(K-1) < \alpha, \tag{13b}$$

$$\alpha,P,K \in \mathbb{N}, \tag{13c}$$

where the objective function $R(\alpha,K,P)$ is

$$R(\alpha,K,P) = \frac{\log_2 K+\alpha-P(K-1)/2}{D^{-1}\left(L_{\mathrm{H}}+\alpha-P(K-1)/2,\epsilon,\rho\right)}. \tag{14}$$

Relaxing the constraint $\alpha \in \mathbb{N}$ and allowing α to take real values, i.e., $\alpha \in \mathbb{R}$, the partial derivative of the function $R(\alpha,K,P)$ with respect to α is given by

$$\frac{\partial R}{\partial \alpha} = \frac{V(Q^{-1})^2+\sqrt{V}Q^{-1}\Psi(\alpha)+4C\left(L_{\mathrm{H}}-\log_2 K\right)}{\Psi(\alpha)[\sqrt{V}Q^{-1}+\Psi(\alpha)]^3/(2C)^2}, \tag{15}$$

where

$$\Psi(\alpha) \triangleq \sqrt{V(Q^{-1})^2+4C(L_{\mathrm{H}}+\alpha-P(K-1)/2)} > 0.$$

According to the constraints (13a), (13b), and (6), we have

$$P(K-1) < \alpha \leqslant L_{\max} < (2^{L_{\mathrm{H}}}-1)\times 8 - L_{\mathrm{H}}. \tag{16}$$

That is,

$$K < \frac{(2^{L_{\mathrm{H}}}-1)\times 8-L_{\mathrm{H}}}{P}+1 < \frac{(2^{L_{\mathrm{H}}}-1)\times 8}{P_{\min}}+1, \tag{17}$$

where $P_{\min} = 8$ bits in the practical scenario.

Therefore, the inequality (17) is equivalent to

$$K < 2^{L_{\mathrm{H}}}. \tag{18}$$

Substituting (18) into (15), it is obvious that $\partial R/\partial\alpha > 0$ always holds. Hence, the maximal payload rate R is a monotonically increasing function of α, which is also confirmed by Fig. 2.

In other words, to maximise R, the longest DU length α should be set at its maximum value, i.e. $\alpha^* = L_{\max}$, which is a constant independent of K and P. As such, the inequality constraint (12a) can be converted into the equality constraint

$$M+P(K-1) = L_{\max}. \tag{19}$$

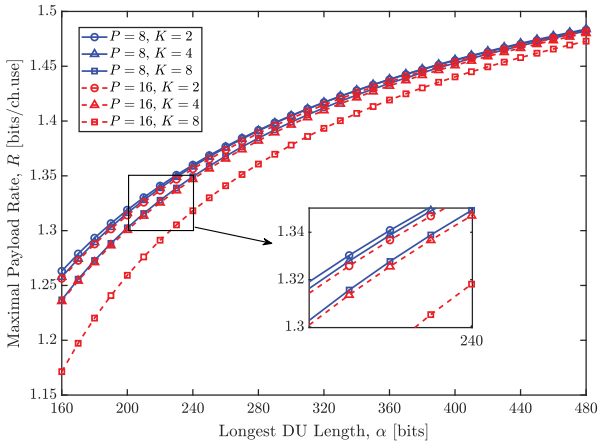


Fig. 2. Maximal payload rates versus α, given $P = 8, 16$ bits and $K = 2, 4, 8$, for $L_{\mathrm{H}} = 24$ bits, at $\epsilon = 10^{-3}$ and $\rho = 10$ dB.

2) Optimisations of P and K: In this step, the tradeoff between P and K is analysed on the premise that the shortest DU length M is a constant.

Based on the equality constraint on the longest DU length, i.e., (19), the common difference of successive lengths and the number of available DU lengths can be expressed as

$$P = (L_{\max} - M)/(K-1), \tag{20a}$$

$$K = (L_{\max} - M)/P + 1. \tag{20b}$$

The maximal payload rate is expressed as a function of K, i.e.,

$$R(K) = \frac{\log_2 K + (L_{\max} + M)/2}{[\sqrt{V}Q^{-1} + \Phi(M)]^2/(2C)^2}, \tag{21a}$$

and a function of P, i.e.,

$$R(P) = \frac{\log_2\left((L_{\max} - M)/P + 1\right) + (L_{\max} + M)/2}{[\sqrt{V}Q^{-1} + \Phi(M)]^2/(2C)^2}, \tag{21b}$$

by substituting (20a) and (20b) into (11), respectively, where

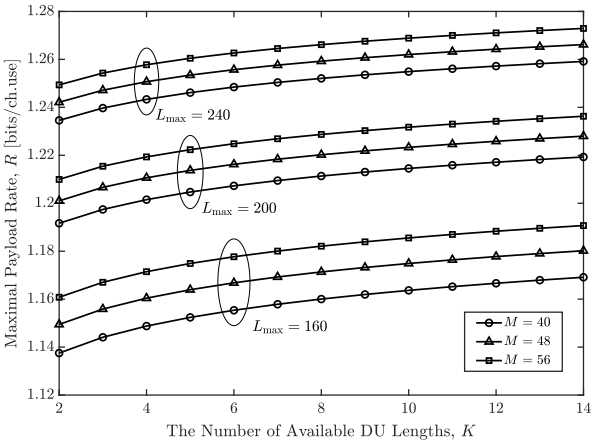
$$\Phi(M) \triangleq \sqrt{V(Q^{-1})^2 + 4C[L_{\mathrm{H}} + (L_{\max} + M)/2]} \tag{22}$$

is deemed to be a positive constant given M.

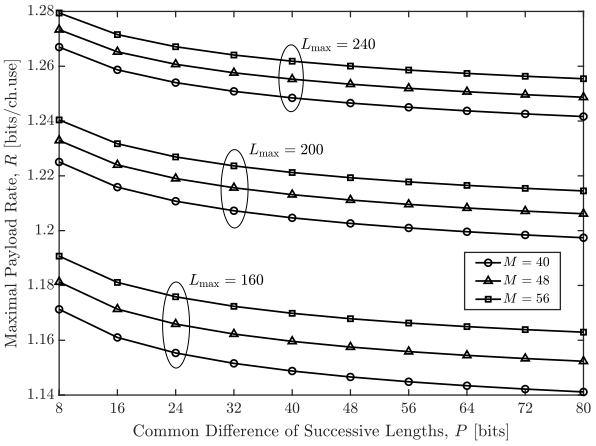
It is obvious from (21a) and (21b) that $R(K)$ is a monotonically increasing function of K and $R(P)$ is a monotonically decreasing function of P, given M. This phenomenon is confirmed by Fig. 3 as well, where the maximal payload rate R is plotted as a function of K in Fig. 3(a) and as a function of P in Fig. 3(b), for the permutation-based transmissions with the longest DU length $L_{\max} = 160, 200, 240$ bits and the shortest DU length $M = 40, 48, 56$ bits. Hence, to maximise the maximal payload rate given $L_{\max}$ and M, the optimal P is its minimum value and the optimal K is its maximum value, i.e., $P^* = 8$ and $K^* = (L_{\max} - M)/8 + 1$ in the practical scenario.

3) Optimisation of M: Given the optimised $P^* = 8$ and $K^* = (L_{\max} - M)/8 + 1$, the maximisation of R through optimising M is formulated as

$$\mathcal{P}3 : R^* = \max_{M} R_1(M) \tag{23}$$



(a) $R(K)$ versus K at $P = (L_{\max} - M)/(K-1)$



(b) $R(P)$ versus P at $K = (L_{\max} - M)/P + 1$

Fig. 3. Maximal payload rates under the constraint (19), at $\epsilon = 10^{-3}$ and $\rho = 10$ dB, for $L_{\mathrm{H}} = 24$ bits.

$$\text{s.t.} \quad L_{\max} - M \geqslant 8, \tag{23a}$$

$$M \geqslant 8, \quad M \in \mathbb{N}, \tag{23b}$$

where the objective function $R_1(M)$ is

$$R_1(M) = \frac{\log_2((L_{\max} - M)/8 + 1) + (L_{\max} + M)/2}{D^{-1}\left(L_{\mathrm{H}} + (L_{\max} + M)/2, \epsilon, \rho\right)}. \tag{24}$$

The constraint (23a) indicates that at least $K = 2$ lengths are available in a permutation-based transmission, and (23b) specifies that the shortest DU length M is one byte in the practical scenario.

Relaxing the constraint $M \in \mathbb{N}$ and allowing M to take real values, i.e., $M \in \mathbb{R}$, we have the following lemma.

Lemma 1: The first-order derivative of the objective function $R_1(M)$ is a monotonically decreasing function of M.

Proof: The first-order derivative of $R_1(M)$ with respect to M is derived as

$$R_1'(M) = \frac{\sqrt{V}Q^{-1}/2 - \Phi(M)/[(L_{\max} - M + 8)\ln 2]}{\Phi(M)[\sqrt{V}Q^{-1} + \Phi(M)]^2/(2C)^2} + \frac{(2C)^3[L_{\mathrm{H}} - \log_2((L_{\max} - M)/8 + 1)]}{\Phi(M)[\sqrt{V}Q^{-1} + \Phi(M)]^3}, \tag{25}$$

where the function $\Phi(M) > 0$ is given by (22).

The second-order derivative of $R_1(M)$ is calculated using

$$R_1''(M) = \frac{-(2C)^2 \cdot [\Theta_1(M) + \Theta_2(M) + \Theta_3(M)]}{\Phi^2(M)[\sqrt{V}Q^{-1} + \Phi(M)]^4}, \quad (26)$$

where

$$\Theta_1(M) \triangleq \frac{[\sqrt{V}Q^{-1} + \Phi(M)][\sqrt{V}Q^{-1} + 3\Phi(M)]}{2\Phi(M)/(C\sqrt{V}Q^{-1})}, \quad (26a)$$

$$\Theta_2(M) \triangleq \frac{2C^2\,[L_{\rm H} - \log_2((L_{\max} - M)/8 + 1)]}{\Phi(M)/[\sqrt{V}Q^{-1} + 4\Phi(M)]}, \quad (26b)$$

$$\Theta_3(M) \triangleq \frac{1}{\ln 2}\left(\frac{\Phi(M)[\sqrt{V}Q^{-1} + \Phi(M)]}{L_{\max} - M + 8} - 2C\right)^2 - \frac{4C^2}{\ln 2}. \quad (26c)$$

It is obvious from (26a) and (26c) that

$$\Theta_1(M) > 0 \quad \text{and} \quad \Theta_3(M) > -\frac{4C^2}{\ln 2}. \quad (27)$$

From (26b), we have

$$\Theta_2(M) > 8C^2[L_{\rm H} - \log_2((L_{\max} - M)/8 + 1)]. \quad (28)$$

Given that $l \leqslant L_{\rm H} - 1$, the constraint (6) leads to

$$L_{\max} \leqslant (2^l - 1) \times 8 - L_{\rm H} \leqslant (2^{L_{\rm H}-1} - 1) \times 8 - L_{\rm H}. \quad (29)$$

Substituting the minimum value of M given by (23b) and the maximum value of $L_{\max}$ given by (29) into the right-hand side of (28), we have

$$\begin{aligned}\Theta_2(M) &> 8C^2\left[L_{\rm H} - \log_2\left(2^{L_{\rm H}-1} - L_{\rm H}/8 - 1\right)\right] \\ &> 8C^2\left[L_{\rm H} - \log_2\left(2^{L_{\rm H}-1}\right)\right] = 8C^2\end{aligned} \quad (30)$$

Therefore, $\Theta_2(M) + \Theta_3(M) > (8 - 4/\ln 2)C^2 > 0$.

As a result, the second-order derivative $R_1''(M) < 0$ always holds and, consequently, the first-order derivative $R_1'(M)$ in (25) is a monotonically decreasing function of M, which completes the proof. ∎

Based on *Lemma 1*, the optimisation of M is discussed concerning the range of $L_{\max}$.

When M is set at its minimum value, i.e., $M_{\min} = 8$, (25) is written as

$$R_1'(8) = \frac{2C\,[L_{\rm H} - \log_2(L_{\max}/8)] + [\sqrt{V}Q^{-1} + \Phi(8)]\theta}{\Phi(8)[\sqrt{V}Q^{-1} + \Phi(8)]^3/(4C^2)}, \quad (31)$$

where $\Phi(8) = \Phi(M)|_{M=8}$, and the constant

$$\theta \triangleq \frac{\sqrt{V}Q^{-1}}{2} - \frac{\Phi(8)}{L_{\max}\ln 2}. \quad (32)$$

Apparently, $R_1'(8) > 0$ holds if $\theta > 0$, which is satisfied given that

$$L_{\max} > \frac{2}{\ln 2} + \frac{8C}{V(Q^{-1})^2(\ln 2)^2} + \frac{8\sqrt{C}}{\sqrt{V}Q^{-1}\ln 2} + \frac{4\sqrt{CL_{\rm H}}}{\sqrt{V}Q^{-1}\ln 2}. \quad (33)$$

In practice, it is easy to meet (33), as we usually have $\sqrt{C}/(\sqrt{V}Q^{-1}) \in (0, 1)$. Hence, $R_1'(8) > 0$ agrees with the reality of packet transmissions under study.

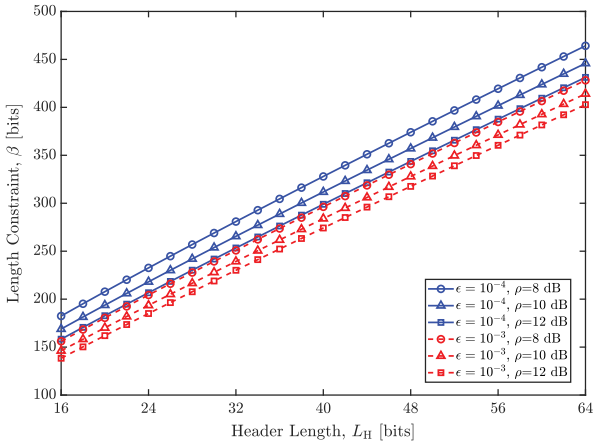


Fig. 4. The constraint β in (35) versus the header length $L_{\rm H}$. Given $L_{\max} \geqslant \beta$, R is maximised on $M \in [8, L_{\max} - 8]$.

When M is set at its maximum value, i.e., $M_{\max} = L_{\max} - 8$, (25) is obtained by

$$R_1'(L_{\max} - 8) = \frac{(4C^2)[\sqrt{V}Q^{-1}/2 - \Phi(L_{\max} - 8)/(16\ln 2)]}{\Phi(L_{\max} - 8)[\sqrt{V}Q^{-1} + \Phi(L_{\max} - 8)]^2} + \frac{8C^3(L_{\rm H} - 1)}{\Phi(L_{\max} - 8)[\sqrt{V}Q^{-1} + \Phi(L_{\max} - 8)]^3}, \quad (34)$$

where $\Phi(L_{\max} - 8) = \Phi(M)|_{M=L_{\max}-8}$. To meet $R_1'(L_{\max} - 8) < 0$, we have

$$\begin{aligned}L_{\max} &> [32(\ln 2)^2 - 1/2]V(Q^{-1})^2/(4C) + (8\ln 2 - 1)L_{\rm H} \\ &\quad - (8\ln 2 - 4) + (8\ln 2 - 1)\sqrt{V}Q^{-1}/(4C) \\ &\quad \times \sqrt{(4\ln 2 + 1/2)^2 V(Q^{-1})^2 + 32(\ln 2)C(L_{\rm H} - 1)} \\ &\triangleq \beta,\end{aligned} \quad (35)$$

More explicitly, the length β in (35) is plotted in Fig. 4.

In the case that $L_{\max} \geqslant \beta$, the first-order derivative $R_1'(M)$ in (25) is a monotonically decreasing function of M, ranging from positive to negative values, and the maximal payload rate $R_1(M)$ in (24) is concave with respect to M.

In the case that $L_{\max} < \beta$, $R_1'(M) > 0$ always holds and, thus, $R_1(M)$ is a monotonically increasing function of M on $[8, L_{\max} - 8]$. That is, the maximal payload rate $R_1(M)$ is maximised at $M = L_{\max} - 8$, where $K = 2$ DU lengths are available in a permutation-based transmission.

As a consequence, the optimal value of the shortest DU length M in the practical scenario with $P^* = 8$, denoted by M_1^*, is achieved at

$$M_1^* = \begin{cases} M^\star, & L_{\max} \geqslant \beta, \\ L_{\max} - 8, & L_{\max} < \beta, \end{cases} \quad (36)$$

where $M^\star$ is the root of the equation

$$R_1'(M) = 0 \quad (37)$$

and can be obtained through an iterative algorithm, e.g., Newton's method.

In Fig. 5, the maximal payload rate R is plotted versus M, for various values of the maximum DU length $L_{\max}$, with

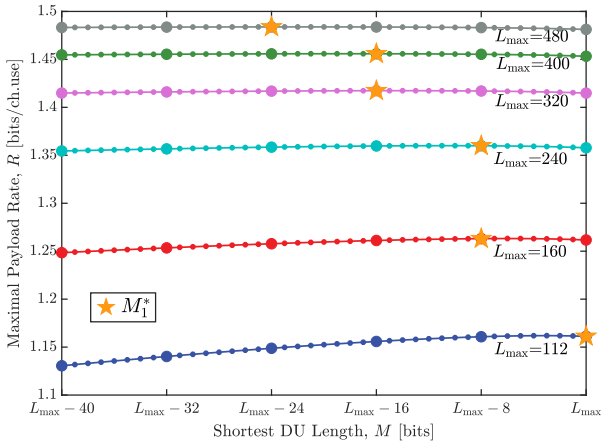


Fig. 5. Maximal payload rates versus M in the practical scenario with $P^*=8$, at $\epsilon=10^{-3}$ and $\rho=10$ dB, for $L_{\mathrm{H}}=24$ bits.

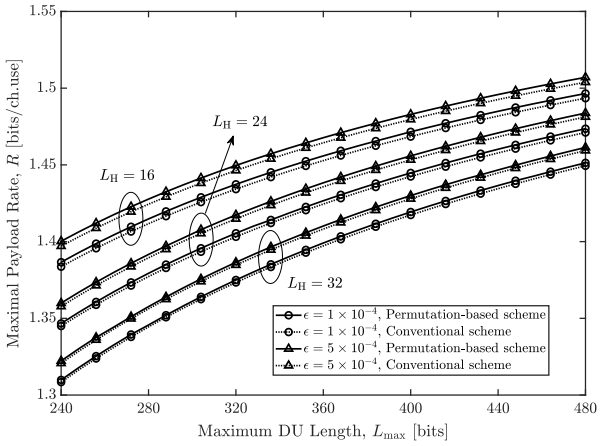


Fig. 6. The maximal payload rate R^* maximised at $\rho=10$ dB.

$P^*=8$ and $K^*=(L_{\max}-M)/8+1$. Given $\epsilon=10^{-3}$, $\rho=10$ dB, $L_{\mathrm{H}}=24$ bits, we have $\beta=193$ bits based on (35). As shown in this figure, increasing $L_{\max}$ improves the maximal payload rate of a permutation-based transmission, where more DU lengths are available. However, if $L_{\max}<\beta$, e.g. when $L_{\max}=160$ bits, only $K=2$ DU lengths are available. More specifically, if $L_{\max}$ further decreases, e.g. given that $L_{\max}=112$ bits, the performance gain achieved by permutation-based transmissions over conventional approaches will disappear. The main reason behind this is that, in shorter packets, a larger portion of the resource is consumed by the headers, which however does not contribute to the payload at all. As such, if $L_{\max}$ is very small, the payload gain contributed by the PCDU cannot compensate the payload loss caused by the DU reduction.

In summary, if $L_{\max}\geqslant\beta$, our permutation-based transmission outperforms conventional approaches, and its maximal payload rate is maximised at $P^*=8$, $K^*=(L_{\max}-M_1^*)/8+1$, and M_1^* as

$$R_1^*(L_{\max},L_{\mathrm{H}},\epsilon,\rho)=\frac{\log_2((L_{\max}-M_1^*)/8+1)+(L_{\max}+M_1^*)/2}{D^{-1}\left(L_{\mathrm{H}}+(L_{\max}+M_1^*)/2,\epsilon,\rho\right)}. \tag{38}$$

In Fig. 6, the maximised R_1^* is plotted as a function of $L_{\max}$. For comparison, the maximal payload rate of conventional scheme is also plotted in terms of the most advantageous case, where the DU length of all packets is set to $L_{\max}$. In this case, the packet length $L_{\mathrm{H}}+L_{\max}$ in conventional scheme is larger than the average packet length $L_{\mathrm{H}}+\bar{L}$ in the permutation-based scheme, consuming more communication resources. In spite of this, the maximal payload rate of the permutation-based transmission still outperforms that of the conventional transmission, thanks to the extra application-layer data conveyed by the PCDU. The advantage becomes more obvious as the header length gets shorter, because the DU reduction has smaller influence on the payload rate gain in this case. Besides, the maximal payload rate is improved upon increasing $L_{\max}$ or ϵ.

B. Ideal Scenario

Herein, given $M_{\min}=P_{\min}=1$ bit, the optimisation problem $\mathcal{P}1$ is solved in three steps as well. The first two steps are exactly the same as Steps *1)* and *2)* in Section III-A, i.e., under the equality constraint $M+P(K-1)=L_{\max}$, the optimal common difference of successive lengths P is its minimum value, while the optimal number of available DU lengths K is its maximum value. More specifically, $P^*=1$ and $K^*=L_{\max}-M+1$ in the ideal scenario.

As shown in (21b), the maximal payload rate R is improved upon decreasing P, from which it is readily inferred that permutation-based transmissions achieve higher maximal payload rate in the ideal scenario with $P^*=1$ than in the practical scenario with $P^*=8$.

In the last step, we will optimise the shortest DU length M based on the optimised $P^*=1$ and $K^*=L_{\max}-M+1$. For the ideal scenario, the maximisation of R through optimising M is formulated as

$$\mathcal{P}4: R^*=\max_{M} R_2(M) \tag{39}$$
$$\text{s.t.}\quad L_{\max}-M\geqslant 1, \tag{39a}$$
$$M\geqslant 1,\quad M\in\mathbb{N}, \tag{39b}$$

where the objective function $R_2(M)$ is

$$R_2(M)=\frac{\log_2(L_{\max}-M+1)+(L_{\max}+M)/2}{D^{-1}\left(L_{\mathrm{H}}+(L_{\max}+M)/2,\epsilon,\rho\right)}. \tag{40}$$

Relaxing the constraint $M\in\mathbb{N}$ and allowing M to take real values, i.e., $M\in\mathbb{R}$, we have the following lemma.

Lemma 2: The first-order derivative of the objective function $R_2(M)$ is a monotonically decreasing function of M.

Proof: The first-order derivative of $R_2(M)$ with respect to M is derived as

$$R_2'(M)=\frac{\sqrt{V}Q^{-1}/2-\Phi(M)/[(L_{\max}-M+1)\ln 2]}{\Phi(M)[\sqrt{V}Q^{-1}+\Phi(M)]^2/(2C)^2}+\frac{(2C)^3\left(L_{\mathrm{H}}-\log_2(L_{\max}-M+1)\right)}{\Phi(M)[\sqrt{V}Q^{-1}+\Phi(M)]^3}, \tag{41}$$

where the function $\Phi(M)>0$ is given by (22).

The second-order derivative of $R_2(M)$ is calculated using

$$R_2''(M) = \frac{-(2C)^2 \cdot [\Omega_1(M) + \Omega_2(M) + \Omega_3(M)]}{\Phi^2(M)[\sqrt{V}Q^{-1} + \Phi(M)]^4}, \quad (42)$$

where

$$\Omega_1(M) \triangleq \frac{[\sqrt{V}Q^{-1} + \Phi(M)][\sqrt{V}Q^{-1} + 3\Phi(M)]}{2\Phi(M)/(C\sqrt{V}Q^{-1})}, \quad (42a)$$

$$\Omega_2(M) \triangleq \frac{2C^2\,[L_{\rm H} - \log_2(L_{\max} - M + 1)]}{\Phi(M)/[\sqrt{V}Q^{-1} + 4\Phi(M)]}, \quad (42b)$$

$$\Omega_3(M) \triangleq \frac{1}{\ln 2}\left(\frac{\Phi(M)[\sqrt{V}Q^{-1}+\Phi(M)]}{L_{\max} - M + 1} - 2C\right)^2 - \frac{4C^2}{\ln 2}. \quad (42c)$$

Through the derivations similar with (27)–(30), we can prove that the second derivative $R_2''(M) < 0$ in the case of $P^* = 1$. Therefore, the first-order derivative $R_2'(M)$ in (41) is a monotonically decreasing function of M, which completes the proof. ■

Based on *Lemma 2*, the optimisation of M is discussed concerning $L_{\max} \in [1, L_{\max} - 1]$.

When M is set at its minimum value $M_{\min} = 1$, (41) is written as

$$R_2'(1) = \frac{2C\,[L_{\rm H} - \log_2 L_{\max}] + [\sqrt{V}Q^{-1} + \Phi(1)]\omega}{\Phi(1)[\sqrt{V}Q^{-1} + \Phi(1)]^3/(4C^2)}, \quad (43)$$

where $\Phi(1) = \Phi(M)|_{M=1}$, and the constant

$$\omega \triangleq \frac{\sqrt{V}Q^{-1}}{2} - \frac{\Phi(1)}{L_{\max}\ln 2}. \quad (44)$$

Apparently, $R_2'(1) > 0$ holds if $\omega > 0$, which is satisfied given that

$$L_{\max} > \frac{2}{\ln 2} + \frac{8C}{V(Q^{-1})^2(\ln 2)^2} + \frac{2\sqrt{2C}}{\sqrt{V}Q^{-1}\ln 2} + \frac{4\sqrt{CL_{\rm H}}}{\sqrt{V}Q^{-1}\ln 2}. \quad (45)$$

In comparison with the condition (33) in the case of $P^* = 8$, (45) is easier to meet. Hence, $R_2'(1) > 0$ agrees with the reality of packet transmissions under study.

When M is set at its maximum value, i.e., $M_{\max} = L_{\max} - 1$, (41) is achieved at

$$R_2'(L_{\max}-1) = -\frac{2C^2\sqrt{V}Q^{-1}(1/\ln 2 - 1)}{\Phi(L_{\max}-1)[\sqrt{V}Q^{-1}+\Phi(L_{\max}-1)]^2} - \frac{(8C^3/\ln 2)(L_{\max}+\ln 2 - 1/2) + 8C^3L_{\rm H}(1/\ln 2 - 1)}{\Phi(L_{\max}-1)[\sqrt{V}Q^{-1}+\Phi(L_{\max}-1)]^3} < 0. \quad (46)$$

That is, $R_2'(M)$ monotonically decreases from a positive value to a negative one, with the increase in M. Hence, $R_2(M)$ is a concave function of $M \in [1, L_{\max} - 1]$.

As a result, the optimal value of the shortest DU length M in the ideal scenario with $P^* = 1$, for maximising the maximal payload rate $R_2(M)$, is the root of the equation $R_2'(M) = 0$, denoted by M_2^*.

Now that $M_2^* < L_{\max} - 1$, the maximal payload rate $R_2(M)$ is maximised with $K^* > 2$ DU lengths available in a permutation-based transmission. In other words, our permutation-based encapsulation always outperforms conventional transmissions.

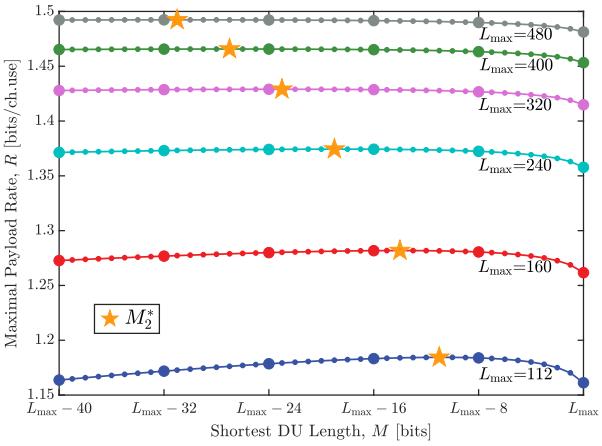


Fig. 7. Maximal payload rates versus M in the ideal scenario with $P^* = 1$, at $\epsilon = 10^{-3}$ and $\rho = 10$ dB, for $L_{\rm H} = 24$ bits.

Given $P^* = 1$ and $K^* = L_{\max} - M + 1$, the maximal payload rate R_2^* maximised at $M = M_2^*$ is plotted in Fig. 7. Different from M_1^* in the practical scenario with $P^* = 8$, M_2^* is not constrained to be a multiple of 8, i.e., it can be an arbitrary positive integer on $[1, L_{\max} - 1]$.

Comparing Fig. 7 with Fig. 5, we may find that $R_2(M) > R_1(M)$ and $M_2^* < M_1^*$. More specifically, given the same conditions of $L_{\max}$, $L_{\rm H}$, ϵ, ρ, the maximal payload rate in the ideal scenario with $P^* = 1$ is higher than that in the practical scenario with $P^* = 8$, and the shortest DU length for maximising $R_2(M)$ is shorter than than for maximising $R_1(M)$. The main reason behind this is that, as the common difference of successive lengths, P^*, decreases, the payload gain contributed by the PCDU is larger than the payload loss caused by the DU reduction.

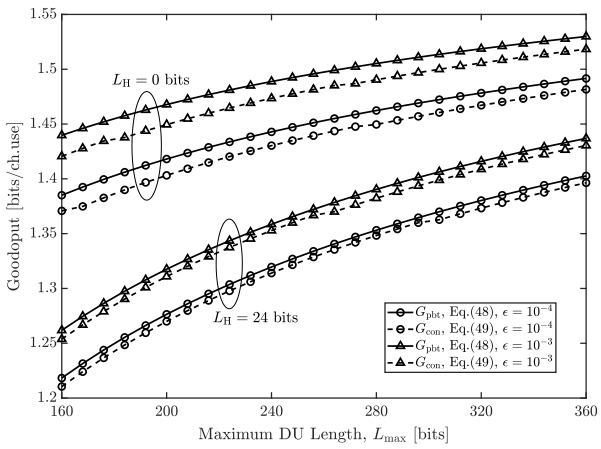
In addition, we may notice that $R_2(L_{\max}) = R_1(L_{\max})$. That is, when $M = L_{\max}$, the permutation-based transmissions in both scenarios achieve the same maximal payload rate, which is equal to that of conventional transmissions, i.e., with $K = 1$ DU length available.

IV. Performance Analysis

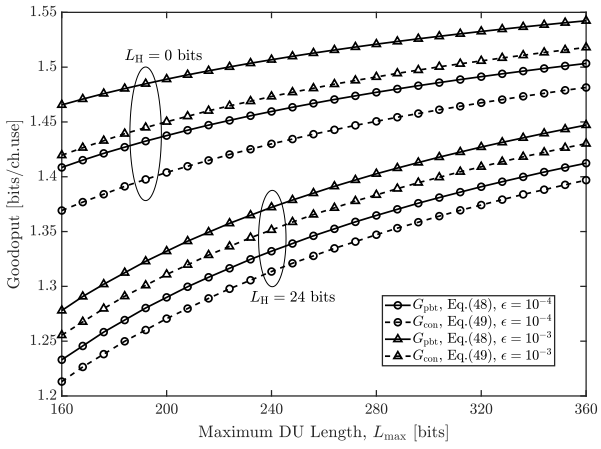
In this section, the performance of permutation-based transmissions is evaluated in the metrics of goodput and latency, based on our formulation of the maximal payload rate in finite blocklength regime, denoted by $R_{\rm pbt} = R(M^*, K^*, P^*)$, where $R(M, K, P)$ is given by (11).

For the purpose of performance comparison between the permutation-based transmissions and conventional approaches, the maximal payload rate of conventional transmissions, denoted by $R_{\rm con}$, is investigated for the average length of a DU, $\bar{L} = M^* + P^*(K^* - 1)/2$, where

$$R_{\rm con} = \frac{\bar{L}}{D^{-1}(L_{\rm H} + \bar{L}, \epsilon, \rho)} = \frac{M^* + P^*(K^* - 1)/2}{D^{-1}(L_{\rm H} + M^* + P^*(K^* - 1)/2, \epsilon, \rho)}. \quad (47)$$



(a) $P^* = 8$



(b) $P^* = 1$

Fig. 8. Goodput comparison between permutation-based and conventional transmissions at $\rho = 10$ dB.

A. Goodput

The metric of goodput is used to measure the effective delivery of application-layer data, i.e., excluding the metadata arising from headers. Concerning the block error probability ϵ, the goodput of our permutation-based transmissions is defined as

$$\begin{aligned} G_{\rm pbt} &= (1-\epsilon)R_{\rm pbt} \\ &= (1-\epsilon)\frac{\log_2 K^* + M^* + P^*(K^*-1)/2}{D^{-1}(L_{\rm H}+M^*+P^*(K^*-1)/2,\epsilon,\rho)}, \end{aligned} \tag{48}$$

and the goodput of conventional transmissions is obtained by

$$\begin{aligned} G_{\rm con} &= (1-\epsilon)R_{\rm con} \\ &= (1-\epsilon)\frac{M^* + P^*(K^*-1)/2}{D^{-1}(L_{\rm H}+M^*+P^*(K^*-1)/2,\epsilon,\rho)}. \end{aligned} \tag{49}$$

In Figs. 8(a) and 8(b), the goodput of our permutation-based transmissions is compared with that of conventional approaches at $\rho = 10$ dB for the practical scenario with $P^* = 8$ and the ideal scenario with $P^* = 1$, respectively. Note that, given the average length of a conventional DU as

$$\bar{L} = M^* + P^*(K^*-1)/2 = (L_{\max} + M^*)/2, \tag{50}$$

the conventional goodput $G_{\rm con}$ remains the same in both scenarios, since it is independent of P^*. From both figures, we may find that the permutation-based goodput is considerably improved in comparison with the conventional goodput, specifically for the ideal scenario with $P^* = 1$. Moreover, the goodput gap between permutation-based and conventional transmissions is increased as the header length $L_{\rm H}$ decreases. The case of $L_{\rm H} = 0$ demonstrates the maximum goodput gain achieved by permutation-based transmissions over the conventional, where the overhead of headers is disregarded.

B. Latency

In comparison to conventional transmissions, our permutation-based encapsulation reduces the payload for delivering the same amount of application-layer data, since the PCDU is not physically encapsulated into the payload and does not consume any resource.

To investigate the latency reduction capability of permutation-based transmissions, the time resource consumed in delivering a single bit of the application-layer data is used to scale the latency, which is referred to as latency per bit and calculated using

$$\begin{aligned} T_{\rm pbt} &= T_s/R_{\rm pbt} \\ &= \frac{T_s \cdot D^{-1}(L_{\rm H}+M^*+P^*(K^*-1)/2,\epsilon,\rho)}{\log_2 K^* + M^* + P^*(K^*-1)/2}, \end{aligned} \tag{51}$$

where T_s denotes the time resource consumed per physical-layer channel use, including the time consumption of all necessary operations at the physical layer, e.g., transmission through the physical channel, modulation and demodulation, coding and decoding, etc.

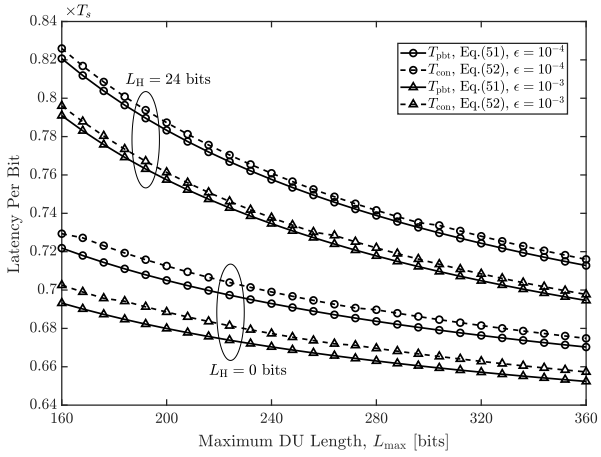
Similarly, the latency per bit in conventional transmissions is obtained by

$$\begin{aligned} T_{\rm con} &= T_s/R_{\rm con} \\ &= \frac{T_s \cdot D^{-1}(L_{\rm H}+M^*+P^*(K^*-1)/2,\epsilon,\rho)}{M^* + P^*(K^*-1)/2}. \end{aligned} \tag{52}$$

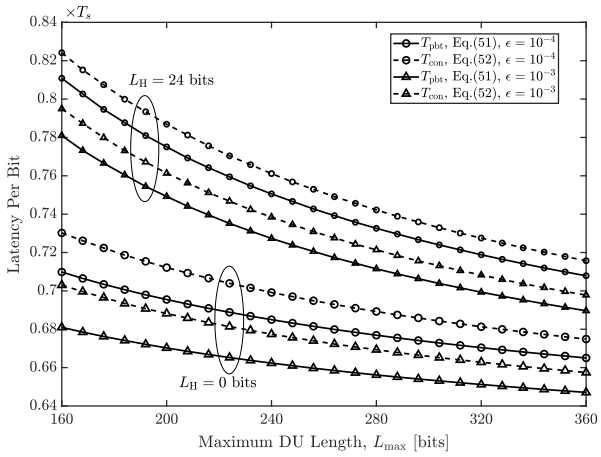
In Figs. 9(a) and 9(b), the latency per bit is compared between permutation-based and conventional transmissions at $\rho = 10$ dB for the practical scenario with $P^* = 8$ and the ideal scenario with $P^* = 1$, respectively. As shown in these figures, the conventional latency per bit $T_{\rm con}$ remains the same in both scenarios. This phenomenon is a counterpart of that in Fig. 8, and the reasons behind them are the same. Moreover, the permutation-based latency is always lower than the conventional one, and the gap between them is increased as P^* or $L_{\rm H}$ decreases.

V. Resource Utilisation Efficiency

An outstanding advantage of our permutation-based encapsulation is using the permutation of DU lengths to map the PCDU, rather than consuming physical resources. In this section, the spectral and energy efficiency gains of the permutation-based encapsulation are formulated to demonstrate its resource utilisation efficiency.



(a) $P^* = 8$



(b) $P^* = 1$

Fig. 9. Comparisons of latency per bit between permutation-based and conventional transmissions at $\rho = 10$ dB.

A. Spectral Efficiency Gain

In comparison with conventional transmissions, the spectral efficiency gain (%) of our permutation-based encapsulation, denoted by Δ_{SE}, is defined as the percentage of the payload rate improvement of permutation-based transmissions against the conventional payload rate, i.e.,

$$\begin{aligned}\Delta_{\text{SE}}(M,P,K) &= \frac{R(M,K,P) - R_c}{R_c} \\ &= \frac{\log_2 K}{M + P(K-1)/2},\end{aligned} \tag{53}$$

where the permutation-based payload rate $R(M,K,P)$ is given by (11), and the conventional payload rate

$$R_c = \frac{M + P(K-1)/2}{D^{-1}(L_{\text{H}} + M + P(K-1)/2, \epsilon, \rho)}. \tag{54}$$

Obvious, $\Delta_{\text{SE}}(M,K,P)$ is a monotonically decreasing function of M or P, and the partial derivative of the function $\Delta_{\text{SE}}(M,K,P)$ with respect to K is

$$\frac{\partial \Delta_{\text{SE}}}{\partial K} = \frac{4M - 2P + 2PK(1-\ln K)}{K[2M + P(K-1)]^2 \ln 2}. \tag{55}$$

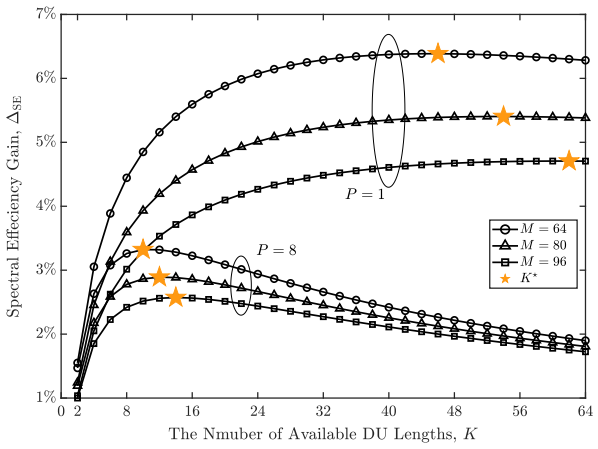


Fig. 10. Spectral efficiency gain Δ_{SE} versus the number of available DU lengths, K.

Given M and P, the spectral efficiency gain Δ_{SE} is maximized when $\partial \Delta_{\text{SE}}/\partial K = 0$, namely

$$K(\ln K - 1) = 2M/P - 1. \tag{56}$$

The solution to (56) is obained by

$$K^\star = \frac{2M/P - 1}{\mathcal{W}(0, (2M/P-1)/e)}, \tag{57}$$

where e is the Euler's number, and $\mathcal{W}(0,x)$ is the 0^{th} branch of the Lambert's W function [40].

More specifically, the spectral efficiency gain Δ_{SE} is plotted as a function of K in Fig. 10, where P is set to its minimum value, i.e., $P = 8$ bits in the practical scenario and $P = 1$ bit in the ideal scenario. As shown in this figure, the spectral efficiency gain of permutation-based transmissions in the ideal scenario is higher than that in the practical scenario. In addition, the spectral efficiency gain is improved as the shortest DU length M decreases, which implies that our permutation-based encapsulation achieves higher spectral efficiency gain in shorter packet communications.

B. Energy Efficiency Gain

In comparison with conventional transmissions, the energy efficiency gain (%) of our permutation-based encapsulation, denoted by Δ_{EE}, is defined as the percentage of the energy saved by permutation-based transmissions against the conventional energy consumption, i.e.,

$$\begin{aligned}\Delta_{\text{EE}}(M,P,K) &= \frac{E_s/R_c - E_s/R(M,P,K)}{E_s/R_c} \\ &= \frac{\log_2 K}{\log_2 K + M + P(K-1)/2},\end{aligned} \tag{58}$$

where E_s denotes the energy per symbol and, thus, E_s/R_c and $E_s/R(M,P,K)$ are the energy consumed per bit in conventional and permutation-based transmissions, respectively.

Apparently, $\Delta_{\text{EE}}(M,K,P)$ is also a monotonically decreasing function of M or P. The partial derivative of $\Delta_{\text{EE}}(M,K,P)$ with respect to K is given by

$$\frac{\partial \Delta_{\text{EE}}}{\partial K} = \frac{4M - 2P + 2PK(1-\ln K)}{K[2\log_2 K + 2M + P(K-1)]^2 \ln 2}. \tag{59}$$

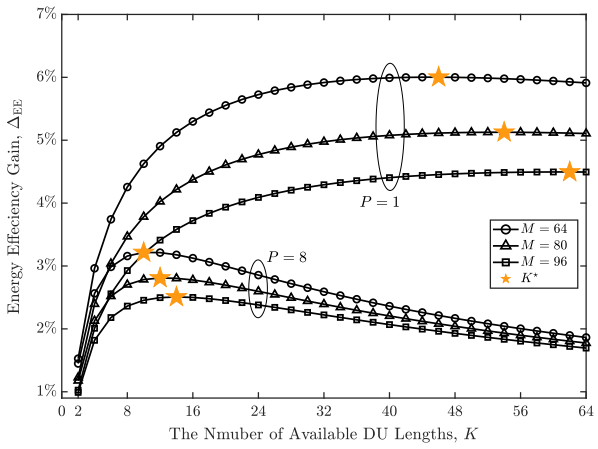
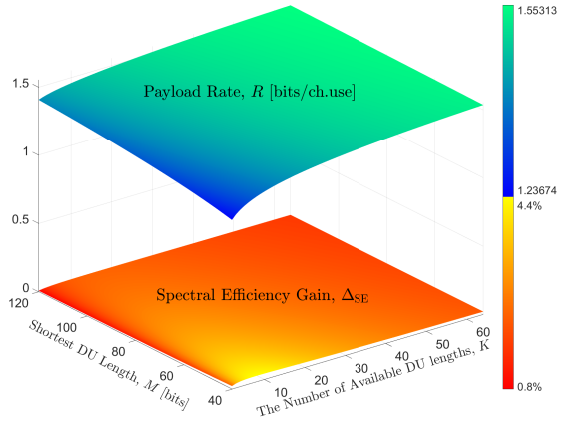


Fig. 11. Energy efficiency gain Δ_{SE} versus the number of available DU lengths, K.

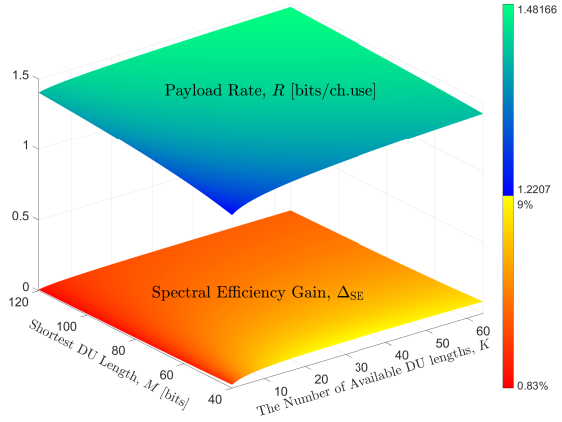
Given M and P, the energy efficiency gain Δ_{EE} is maximized when $\partial\Delta_{\mathrm{EE}}/\partial K = 0$, whose solution is the same as (57). The energy efficiency gain Δ_{EE} versus K is plotted in Fig. 11. Comparing this figure with Fig. 10, we may find that Δ_{EE} and Δ_{SE} have the same variations with respect to the number of DU lengths K, the shortest DU length M, and the common difference of successive lengths P. In brief, using shorter packets leads to higher resource utilisation efficiency of permutation-based transmissions.

C. Tradeoff between Performance and Resource Utilisation

In Fig. 12, the tradeoff between the maximal payload rate $R(M, K, P)$ and the spectral efficiency gain $\Delta_{\mathrm{SE}}(M, K, P)$ of permutation-based transmissions is investigated at $\rho = 10$ dB and $\epsilon = 10^{-3}$, for $L_{\mathrm{H}} = 0$. As shown in this figure, the permutation-based encapsulation using longer packets achieves better performance, i.e., higher goodput and lower latency per bit pertaining to higher maximal payload rate. The main reason behind this lies in (5), which indicates that longer blocklength results in higher coding rate. On the other hand, the permutation-based encapsulation using shorter packets achieves better resource utilisation, i.e., higher spectral or energy efficiency gain, because the efficiency gain is defined in comparison to conventional systems and depends on the contribution from the PCDU. Moreover, the ideal scenario with $P = 1$ can more efficiently balance the tradeoff between the performance and resource utilisation, compared with the practical scenario with $P = 8$.



(a) $P = 8$



(b) $P = 1$

Fig. 12. Tradeoff between the maximal payload rate $R(M, K, P)$ and the spectral efficiency gain $\Delta_{\mathrm{SE}}(M, K, P)$ of permutation-based transmissions at $\rho = 10$ dB and $\epsilon = 10^{-3}$, for $L_{\mathrm{H}} = 0$.

VI. Conclusion

To improve the goodput and reduce the latency in short-packet communications, permutation-based encapsulation has been developed at the transport layer and investigated in this paper, where the PCDU is not transmitted via physical resources but mapped onto the permutation of various DU lengths. Based on the finite-blocklength information theory, the maximal payload rate of permutation-based transmissions was formulated in an analytical form, and the optimal packet structure was obtained for maximising the maximal payload rate in both practical and ideal scenarios. As per the optimal packet structure, the performance of permutation-based transmissions was evaluated in terms of goodput and latency, as well as compared with the conventional performance. Moreover, the resource utilisation efficiency of permutation-based transmissions was analysed in the metrics of spectral and energy efficiency gains. Illustrative numerical results on the performance and resource utilisation comparisons substantiated the advantage achieved by our permutation-based encapsulation over conventional short-packet communications, specifically in the ideal scenario. In particular, an important insight was reached to facilitate the system design of permutation-based transmissions. That is, in contrast to the practical scenario with $P = 8$, the ideal scenario with $P = 1$ is more beneficial to permutation-based transmissions, as it leads to better performance and higher resource utilisation efficiency as well as more efficiently balances the tradeoff between the performance and resource utilisation.

ACKNOWLEDGMENT

The authors would like to thank the editor, Dr. Ahmed El Shafie, and the anonymous reviewers for their valuable comments to improve the presentation of this paper.

REFERENCES

[1] H. Ji, *et. al.*, "Ultra-reliable and low-latency communications in 5G downlink: physical layer aspects", *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124-130, Jun. 2018.

[2] H. Chen, *et. al.*, "Ultra-reliable low latency cellular networks: use cases, challenges and approaches", *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119-125, Dec. 2018.

[3] D. Zheng, Y. Yang, L. Wei, and B. Jiao, "Decode-and-forward short-packet relaying in the internet of things: Timely status updates", *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8423-8437, Dec. 2021.

[4] H. Ren, C. Pan, Y. Deng, M. Elkashlan and A. Nallanathan, "Joint power and blocklength optimization for URLLC in a factory automation scenario", *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1786-1801, Mar. 2020.

[5] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets", *Proc. IEEE*, vol. 104, no. 9, pp. 1711-1726, Aug. 2016.

[6] D. Yoo, W. E. Stark, K. Yar and S. Oh, "Coding and modulation for short packet transmission", *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 2104-2109, May 2010.

[7] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels", *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962-1990, May 2008.

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

[9] J. Hoydis, R. Couillet, and P. Piantanida, "The second-order coding rate of the MIMO quasi-static Rayleigh fading channel", *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6591-6622, Dec. 2015.

[10] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels", *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7041-7051, Nov. 2013.

[11] M. Shirvanimoghaddam, *et al.*, "Short block-length codes for ultra-reliable low latency communications", *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 130-137, Feb. 2019.

[12] T. Erseghe, "Coding in the finite-blocklength regime: bounds based on Laplace integrals and their asymptotic approximations", *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6854-6883, Dec. 2016.

[13] J. E. Siegel, S. Kumar and S. E. Sarma, "The future internet of things: secure, efficient, and model-based", *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2386-2398, Aug. 2018.

[14] Q. Wu, W. Chen, D. W. K. Ng and R. Schober, "Spectral and energy-efficient wireless powered IoT networks: NOMA or TDMA?", *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6663-6667, Jul. 2018.

[15] N. Ishikawa, S. Sugiura, and L. Hanzo, "50 years of permutation, spatial and index modulation: From classic RF to visible light communications and data storage", *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1905-1938, 3rd Quart., 2018.

[16] Y. Yang and B. Jiao, "Information-guided channel-hopping for high data rate wireless communication", *IEEE Commun. Lett.*, vol. 12, no. 4, pp. 225-227, Apr. 2008.

[17] Y. Yang and S. Aissa, "Bit-padding information guided channel hopping", *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 163–165, Feb. 2011.

[18] Y. Yang and S. Aissa, "Information guided channel hopping with an arbitrary number of transmit antennas", *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1552-1555, Oct. 2012.

[19] E. Basar, U. Aygolu, E. Panayirci, and H. Poor, "Orthogonal frequency division multiplexing with index modulation", *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5536-5549, Nov. 2013.

[20] E. Basar, "Index modulation techniques for 5G wireless networks", *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 168-175, Jul. 2016.

[21] M. I. Kadir, "Generalized space–time–frequency index modulation", *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 250-253, Feb. 2019.

[22] S. Sugiura, T. Ishihara and M. Nakao, "State-of-the-art design of index modulation in the space, time, and frequency domains: benefits and fundamental limitations", *IEEE Access*, vol. 5, pp. 21774-21790, 2017.

[23] Y. Yang, "Information-guided relay selection for high throughput in half-duplex relay channels", in *Proc. IEEE Global Commun. Conf.*, Dec. 2009.

[24] Y. Yang and S. Aissa, "Information-guided transmission in decode-and-forward relaying systems: Spatial exploitation and throughput enhancement", *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2341–2351, Jul. 2011.

[25] Y. Yang and S. Aissa, "Cross-layer combining of information-guided transmission with network coding relaying for multiuser cognitive radio systems", *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 26–29, Feb. 2013.

[26] Y. Yang, "Spatial modulation exploited in non-reciprocal two-way relay channels: Efficient protocols and capacity analysis", *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2821–2834, Jul. 2016.

[27] Y. Yang and M. Guizani, "Mapping-varied spatial modulation for physical layer security: Transmission strategy and secrecy rate", *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 877–889, Apr. 2018.

[28] Y. Yang, M. Ma, S. Aissa, and L. Hanzo, "Physical-layer secret key generation via CQI-mapped spatial modulation in multi-hop wiretap *ad-hoc* networks", *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1322–1334, 2021.

[29] M. Yin, Y. Yang, and B. Jiao, "Security-oriented trellis code design for spatial modulation", *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1875–1888, Mar. 2021.

[30] Y. Liu, Y. Yang, L. Yang, and L. Hanzo, "Physical layer security of spatially modulated sparse-code multiple access in aeronautical *ad-hoc* networking", *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2436–2447, Mar. 2021.

[31] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G", *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 3187-3196, May 2020.

[32] Q. Li, M. Wen and M. Di Renzo, "Single-RF MIMO: From spatial modulation to metasurface-based modulation", *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 88-95, Aug. 2021.

[33] B. Jiao, "A high spectral efficiency method enabled by opportunistic-bit", *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12368-12372, Dec. 2018.

[34] M. Yin, Y. Yang, J. Wu and B. Jiao, "Opportunistic bits in short-packet communications: a finite blocklength perspective", *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8085-8099, Dec. 2021.

[35] Y. Yang, "Permutation-based transmissions in ultra-reliable and low-latency communications", *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 1024-1028, March 2021.

[36] Y. Yang and L. Hanzo, "Permutation-based TCP and UDP transmissions to improve goodput and latency in the internet-of-things", *IEEE Internet Things J.*, vol. 8, no. 18, pp. 14276–14286, Sep. 2021.

[37] B. Lee, S. Park, D. J. Love, H. Ji and B. Shim, "Packet structure and receiver design for low latency wireless communications with ultra-short packets", *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 796-807, Feb. 2018.

[38] L. Zhang and Y. Liang, "Average throughput analysis and optimization in cooperative IoT networks with short packet communication", *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11549-11562, Dec. 2018.

[39] C. E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379-423, Jul. 1948.

[40] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function", *Adv. in Comput. Math.*, vol. 5, pp. 329-359, Dec. 1996.