A Throughput-Enhanced HARQ Scheme for 5G System via Partial Superposition

Qianfan Wang, *Graduate Student Member, IEEE*, Suihua Cai, *Member, IEEE*,
Li Chen, *Senior Member, IEEE*, and Xiao Ma, *Member, IEEE*

***Abstract*— In this letter, we propose partial superposition for 5G hybrid automatic retransmission request (HARQ), resulting in a throughput-enhanced HARQ scheme for the 5G system. The novelty of the proposed scheme is to retransmit an erroneous packet by superimposing (XORing) its partial bits onto a fresh packet, costing no extra bandwidth and transmission power. The proposed HARQ schemes can be implemented with the basis of the hardware components of the 5G standard. Numerical results show that the proposed HARQ schemes can yield a throughput improvement up to 10% over the original 5G HARQ scheme.**

***Index Terms*— 5G, HARQ, LDPC codes, partial superposition.**

I. INTRODUCTION

MANY wireless channels are time-varying and the channel state information (CSI) is unknown to the transmitter. In this case, it is difficult to achieve sufficiently low error rates by just error-correcting codes. Hybrid automatic retransmission request (HARQ), which combines both forward error correction (FEC) and automatic retransmission request (ARQ), is a commonly used technique to ensure the reliability of data transmission over wireless links. 3GPP has adopted low-density parity-check (LDPC) codes for data channels in 5G enhanced Mobile Broad Band (eMBB) scenario [1] and formulated the corresponding HARQ schemes [2].

Throughput is one of the foremost performance metrics to evaluate a HARQ scheme. One way to improve the throughput is based on single-packet design, where the transmitter employs fixed-rate codes with repetition redundancy (RR) [3] and Chase combining [4] or rate-compatible (RC) codes [5] with incremental redundancy (IR) [6]. Practically, RC-LDPC codes have been adopted by the 5G standard. Another way is based on multiple-packet design, where redundancies can be generated from multiple data packets. For example, cross-packet coding as investigated in [7]–[9] encodes the previous erroneous packet together with the current packet. The cross-packet coding can improve the performance in terms of throughput over the single-packet scheme but usually has a higher implementation complexity due to the requirement of multiple encoders/decoders with increasing code rates.

In this letter, we propose a new HARQ scheme based on 5G LDPC codes, which integrates the aforementioned two approaches. This scheme is inspired by the work of [10] and is referred to as *partial superposition* (PS)-RR-HARQ (or PS-IR-HARQ). In the case when a packet is unsuccessfully received, the first retransmission in the proposed scheme is generated by superimposing its partial coded bits on a fresh packet, which is a simple way to implement the cross-packet coding. At the receiver, the decoder performs an iterative belief propagation (BP) algorithm to recover the erroneous (target) packet. If it is still unsuccessful, the transmitter switches into the single-packet mode by transmitting subsequently RRs (or IRs). Different from the existing cross-packet coding HARQ schemes [7]–[9], which usually require multiple encoders/decoders with increasing code rates, the proposed PS-HARQ scheme requires only one encoder/decoder and keeps the basic coding structure almost unchanged. This resembles the backtrack retransmission (BRQ) scheme [11] and the layer-coded HARQ (L-HARQ) scheme [12]. Indeed, the L-HARQ is implemented on the basis of a single off-the-shelf code. The main difference lies in the following. At each retransmission, the L-HARQ transmits a codeword that is obtained by encoding a "punctured" erroneous packet along with a few fresh information bits, while the PS-HARQ transmits either a superposition of a punctured coded packet with a fresh codeword or an RR/IR packet as in the conventional HARQ. As a result, the data sequence is packetized into sub-blocks with a *constant* length k for the PS-HARQ but *delayed-CSI-dependent* lengths for the L-HARQ. In addition, we need to point out that the partial superposition is introduced between an erroneous packet and a fresh packet and is performed over the binary field. This is different from the superposition-coding-aided HARQ scheme [13], where the superposition is performed over the real field. This is also different from the network-coded HARQ scheme [14], where the redundancy packets are generated by XORing two erroneous packets.

Moreover, the proposed scheme distinguishes by its simplicity and flexibility. It can be implemented in conjunction with any existing single-packet HARQ schemes. Practically, it can

---

Manuscript received April 27, 2020; revised May 27, 2020 and June 15, 2020; accepted June 16, 2020. Date of publication June 23, 2020; date of current version October 9, 2020. This work was supported by the NSF of China (No. 61971454 and No. 61671486), the Science and Technology Planning Project of Guangdong Province (No. 2018B010114001) and the Basic Research Project of Guangdong Provincial NSF (No. 2016A030308008). The associate editor coordinating the review of this letter and approving it for publication was A. Balatsoukas-Stimming. *(Corresponding authors: Li Chen; Xiao Ma.)*

Qianfan Wang is with the School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510006, China (e-mail: wangqf6@mail2.sysu.edu.cn).

Suihua Cai and Xiao Ma are with the Guangdong Key Laboratory of Information Security Technology, Sun Yat-sen University, Guangzhou 510006, China, and also with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China (e-mail: caish23@mail.sysu.edu.cn; maxiao@mail.sysu.edu.cn).

Li Chen is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China (e-mail: chenli55@mail.sysu.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2020.3004407

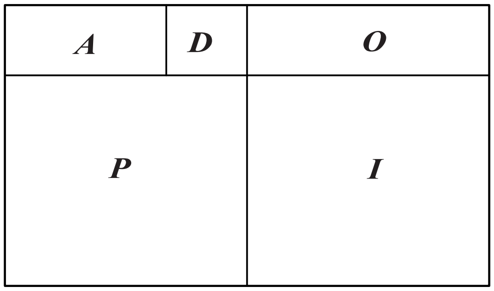


Fig. 1. The structure diagram of the parity-check matrix for the 5G LDPC codes.

be easily implemented based on the hardware components of the 5G standard. The increase in implementation complexity is not too much while the improvement in throughput can be up to 10% over the 5G HARQ as suggested by simulations over independent block Rayleigh fading channels.

II. Background Knowledge

A. System Model

We consider an end-to-end transmission scenario, which consists of a noisy channel for data transmission and a noiseless feedback channel. At time slot $t \geqslant 0$, a coded packet of length n is modulated using binary phase-shift keying (BPSK) signalling and transmitted over a block fading channel, resulting in a received signal $\boldsymbol{y}^{(t)}$, given by

$$\boldsymbol{y}^{(t)} = h_t \boldsymbol{x}^{(t)} + \boldsymbol{w}^{(t)}, \tag{1}$$

where $\boldsymbol{x}^{(t)} \in \{+1, -1\}^n$ is the transmitted signal vector, $\boldsymbol{w}^{(t)}$ is an additive white Gaussian noise (AWGN) vector with zero mean and variance σ^2, and h_t is a fading coefficient which follows the Rayleigh distribution. We assume that the fading coefficient, known at the receiver, is constant over one block and changes independently at different blocks. We also assume that the feedback channel is delay-free to simplify the notation. Our proposal can be easily adapted to 5G systems with delayed feedback and multiple HARQ processes.

B. Review of 5G HARQ Scheme

The 5G LDPC code is a class of rate-compatible quasi-cyclic (QC) LDPC codes with raptor-like construction. Two base matrices (equivalently, two base graphs BG1 and BG2) have been designed for different range of block lengths and code rates [2], from which the parity-check matrices can be constructed by lifting with a lifting factor Z. The structure diagram of the parity-check matrix is illustrated in Fig. 1, where the sub-matrices $\boldsymbol{A}$ and $\boldsymbol{P}$ consist of circulant permutation matrices and zero matrices, the sub-matrix $\boldsymbol{D}$ is a dual-diagonal matrix, $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{O}$ is a zero matrix.

The information bits are encoded systematically using the lowest rate, where the resulting coded bits (except those punctured[1]) are stored in a circular buffer as illustrated in Fig. 2. A codeword in the buffer can be grouped (with possible overlaps which result in repeated bits) to generate four redundancy versions (RVs) as required by the HARQ, denoted by $\boldsymbol{v}_{1\mathrm{T}}$, $\boldsymbol{v}_{2\mathrm{T}}$, $\boldsymbol{v}_{3\mathrm{T}}$ and $\boldsymbol{v}_{4\mathrm{T}}$, see Fig. 2 for reference. For codes based on BG1, the starting positions of these four RVs are specified by $\{\mathrm{B}_0, \mathrm{B}_1, \mathrm{B}_2, \mathrm{B}_3\} = \{0, 17, 33, 56\} \times Z$. For codes based on BG2, the starting positions are specified by $\{\mathrm{B}_0, \mathrm{B}_1, \mathrm{B}_2, \mathrm{B}_3\} = \{0, 13, 25, 43\} \times Z$. At every transmission needed in the 5G HARQ, the transmitter sends different RVs using an LTE-like order [2].

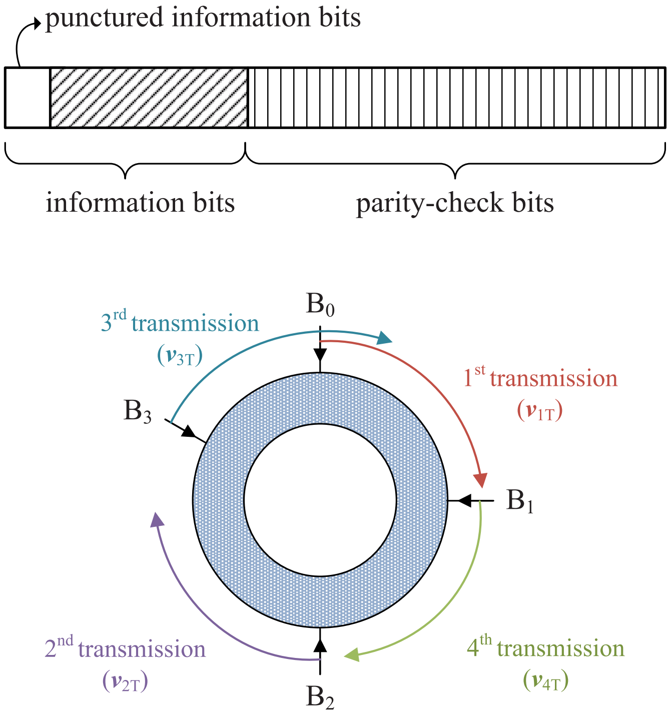


Fig. 2. The 5G HARQ codeword circular buffer.

III. The PS-HARQ Schemes

A. PS-RR-HARQ Scheme

Let $\mathbb{F}_2$ denote the binary field, and $\mathbb{F}_2^k$ and $\mathbb{F}_2^n$ denote the k-dimensional and n-dimensional vectors, respectively. $\mathscr{C}[n, k]$ further denotes a binary block code of dimension k and length n. Assume that the transmitter attempts to send a sequence of information blocks $\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(1)}, \cdots, \boldsymbol{u}^{(L)}$, where $\boldsymbol{u}^{(t)} \in \mathbb{F}_2^k$ and is encoded into a codeword $\boldsymbol{v}^{(t)} \in \mathbb{F}_2^n$ by performing the encoding algorithm of the $\mathscr{C}[n, k]$. Here, we consider the case that $\mathscr{C}[n, k]$ is a 5G LDPC code and $\boldsymbol{v}^{(t)} = \boldsymbol{v}_{1\mathrm{T}}^{(t)}$ for comparison. We only need to show how to successfully transmit $\boldsymbol{v}^{(0)}$. We assume that the first transmission of $\boldsymbol{v}^{(0)}$ is unsuccessful and a NACK is fed back. For the conventional truncated RR-HARQ, the transmitter retransmits $\boldsymbol{v}^{(0)}$ until an ACK is fed back or the maximum number $T_{\max}$ of transmissions is reached. In contrast, for the PS-RR-HARQ, the transmitter employs cross-packet coding for the first retransmission of $\boldsymbol{v}^{(0)}$ and retransmits partial bits of $\boldsymbol{v}^{(0)}$ along with $\boldsymbol{v}^{(1)}$ in a more compact way via superposition. To be precise, the first retransmission of $\boldsymbol{v}^{(0)}$

---

[1] According to the standard, the first $2Z$ information bits are punctured for improving the threshold performance.

TABLE I

COMPARISON OF PS-HARQ AND CONVENTIONAL HARQ SCHEMES IN THE WORST-CASE SCENARIO

	RR-HARQ	PS-RR-HARQ	5G-HARQ	PS-IR-HARQ
$t=0$	$\{\boldsymbol{v}^{(0)},\boldsymbol{v}^{(0)},\boldsymbol{v}^{(0)},\boldsymbol{v}^{(0)}\}$	$\{\boldsymbol{v}^{(0)},\boldsymbol{c}^{(1)},\boldsymbol{v}^{(0)},\boldsymbol{v}^{(0)},\boldsymbol{v}^{(0)}\}$	$\{\boldsymbol{v}_{1\mathrm{T}}^{(0)},\boldsymbol{v}_{2\mathrm{T}}^{(0)},\boldsymbol{v}_{3\mathrm{T}}^{(0)},\boldsymbol{v}_{4\mathrm{T}}^{(0)}\}$	$\{\boldsymbol{v}_{1\mathrm{T}}^{(0)},\boldsymbol{c}^{(1)},\boldsymbol{v}_{2\mathrm{T}}^{(0)},\boldsymbol{v}_{3\mathrm{T}}^{(0)},\boldsymbol{v}_{4\mathrm{T}}^{(0)}\}$
$t=1$	$\{\boldsymbol{v}^{(1)},\boldsymbol{v}^{(1)},\boldsymbol{v}^{(1)},\boldsymbol{v}^{(1)}\}$	$\{\boldsymbol{c}^{(2)},\boldsymbol{v}^{(1)},\boldsymbol{v}^{(1)},\boldsymbol{v}^{(1)}\}$	$\{\boldsymbol{v}_{1\mathrm{T}}^{(1)},\boldsymbol{v}_{2\mathrm{T}}^{(1)},\boldsymbol{v}_{3\mathrm{T}}^{(1)},\boldsymbol{v}_{4\mathrm{T}}^{(1)}\}$	$\{\boldsymbol{c}^{(2)},\boldsymbol{v}_{2\mathrm{T}}^{(1)},\boldsymbol{v}_{3\mathrm{T}}^{(1)},\boldsymbol{v}_{4\mathrm{T}}^{(1)}\}$
$t=2$	$\{\boldsymbol{v}^{(2)},\boldsymbol{v}^{(2)},\boldsymbol{v}^{(2)},\boldsymbol{v}^{(2)}\}$	$\{\boldsymbol{c}^{(3)},\boldsymbol{v}^{(2)},\boldsymbol{v}^{(2)},\boldsymbol{v}^{(2)}\}$	$\{\boldsymbol{v}_{1\mathrm{T}}^{(2)},\boldsymbol{v}_{2\mathrm{T}}^{(2)},\boldsymbol{v}_{3\mathrm{T}}^{(2)},\boldsymbol{v}_{4\mathrm{T}}^{(2)}\}$	$\{\boldsymbol{c}^{(3)},\boldsymbol{v}_{2\mathrm{T}}^{(2)},\boldsymbol{v}_{3\mathrm{T}}^{(2)},\boldsymbol{v}_{4\mathrm{T}}^{(2)}\}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t=L$	$\{\boldsymbol{v}^{(L)},\boldsymbol{v}^{(L)},\boldsymbol{v}^{(L)},\boldsymbol{v}^{(L)}\}$	$\{\boldsymbol{v}^{(L)},\boldsymbol{v}^{(L)},\boldsymbol{v}^{(L)}\}$	$\{\boldsymbol{v}_{1\mathrm{T}}^{(L)},\boldsymbol{v}_{2\mathrm{T}}^{(L)},\boldsymbol{v}_{3\mathrm{T}}^{(L)},\boldsymbol{v}_{4\mathrm{T}}^{(L)}\}$	$\{\boldsymbol{v}_{2\mathrm{T}}^{(L)},\boldsymbol{v}_{3\mathrm{T}}^{(L)},\boldsymbol{v}_{4\mathrm{T}}^{(L)}\}$

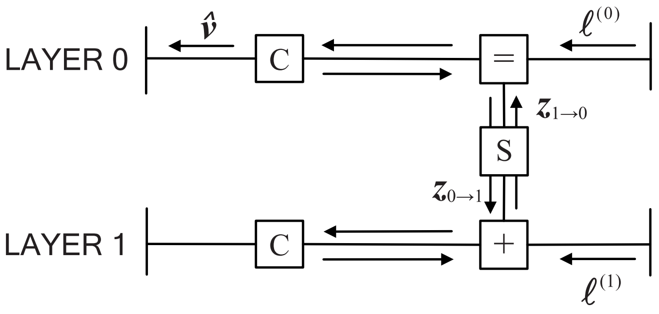


Fig. 3. Normal graph for decoding with two layers.

can be expressed as

$$\boldsymbol{c}^{(1)} = \boldsymbol{v}^{(1)} + \boldsymbol{v}^{(0)}\boldsymbol{S}, \tag{2}$$

where $\boldsymbol{S}$ is a square matrix of order n, which can be obtained by nulling the first $n(1-\alpha)$ columns of a randomly generated but fixed permutation matrix of order n. The parameter α $(0 < \alpha \leq 1)$ is referred to as *superposition fraction*, which can be optimized numerically by one-dimensional search. Notice that when $\alpha = 0$, it is reduced to the conventional RR-HARQ scheme, suggesting that an optimized α enables a higher throughput. At the receiver, since both $\boldsymbol{v}^{(0)}$ and $\boldsymbol{c}^{(1)}$ carry information related to $\boldsymbol{v}^{(0)}$, their noisy versions can be employed to recover $\boldsymbol{v}^{(0)}$ by an iterative message-passing algorithm over a normal graph as illustrated in Fig. 3, where the notation is the same as that in [10]. If the recovery is still unsuccessful, the transmitter will retransmit $\boldsymbol{v}^{(0)}$ until an ACK is received or the number of transmissions for $\boldsymbol{v}^{(0)}$ reaches $T_{\max}$. The key point is that the transmission of $\boldsymbol{c}^{(1)}$ will not be counted as a transmission of $\boldsymbol{v}^{(0)}$ but as that of $\boldsymbol{v}^{(1)}$. This is because, once $\boldsymbol{v}^{(0)}$ is recovered or estimated at the receiver and its effect on $\boldsymbol{c}^{(1)}$ is removed, the received version of $\boldsymbol{c}^{(1)}$ will be reduced to a noisy version of $\boldsymbol{v}^{(1)}$. Taking $T_{\max} = 4$ as an example and considering the worst-case scenario, we compare the PS-RR-HARQ scheme with the conventional RR-HARQ scheme by tabulating the transmitted packets in Table I. Notice that the number of transmitted packets is 5 for $\boldsymbol{v}^{(0)}$ and 3 for $\boldsymbol{v}^{(L)}$ in the PS-RR-HARQ scheme.

In the following, we present a formal description of the PS-RR-HARQ scheme. We assume that the soft-input soft-output (SISO) decoding takes the log-likelihood ratios (LLRs) associated with each coded bits as input and delivers extrinsic LLRs as output for possible use in iterative processing. The LLR associated with the i-th coded bit is initially calculated from the received signal and the channel fading coefficient as $2hy_i/\sigma^2$ for $0 \leqslant i \leqslant n-1$. We use ℓ_i to stand for the i-th initial LLR and $\boldsymbol{\ell}$ for the vector of LLRs. In the case when the first transmission of $\boldsymbol{v}^{(0)}$ is not successful, the cross-packet coded packet $\boldsymbol{c}^{(1)}$ will be transmitted and $\boldsymbol{v}^{(0)}$ can be re-decoded from $\boldsymbol{\ell}^{(0)}$ and $\boldsymbol{\ell}^{(1)}$ by performing an iterative message-passing algorithm, resulting in $\hat{\boldsymbol{v}}^{(0)} = D(\boldsymbol{\ell}^{(0)}, \boldsymbol{\ell}^{(1)})$, which is similar to the decoding algorithm in [10] and is outlined in Algorithm 1, where $\boldsymbol{z}_{0\to 1}$ and $\boldsymbol{z}_{1\to 0}$, initialized by zero vectors, are messages exchanged between the two layers (see Fig. 3 for reference). For completeness, the PS-RR-HARQ scheme is summarized in Algorithm 2.

Algorithm 1 Message-Passing Algorithm for Decoding

```
Input: $\boldsymbol{\ell}^{(0)}, \boldsymbol{\ell}^{(1)}$
repeat
    Taking into account the constraints specified by $\boxed{\mathrm{C}}$ and $\boxed{=}$, the processor at layer 0 takes $\boldsymbol{\ell}^{(0)}$ and $\boldsymbol{z}_{1\to 0}$ as input, and delivers the extrinsic LLRs $\boldsymbol{z}_{0\to 1}$ to layer 1;
    Taking into account the constraints specified by $\boxed{\mathrm{C}}$ and $\boxed{+}$, the processor at layer 1 takes $\boldsymbol{\ell}^{(1)}$ and $\boldsymbol{z}_{0\to 1}$ as input, and delivers the extrinsic LLRs $\boldsymbol{z}_{1\to 0}$ to layer 0;
    Make hard decisions based on the first layer output, resulting in $\hat{\boldsymbol{v}}$;
until $\hat{\boldsymbol{v}}$ is a valid codeword or a preset maximum global iteration number $J_{\max}$ is reached;
Output: $\hat{\boldsymbol{v}} = D(\boldsymbol{\ell}^{(0)}, \boldsymbol{\ell}^{(1)})$
```

B. PS-IR-HARQ Scheme

The basic idea of the PS-IR-HARQ scheme is the same as the PS-RR-HARQ scheme. The difference lies in the retransmissions in the case when the cross-packet coding is not sufficient, where IRs instead of RRs are transmitted. The comparison of the PS-IR-HARQ scheme with the 5G HARQ scheme is also shown in Table I, where $T_{\max} = 4$ and $\boldsymbol{v}_{i\mathrm{T}}$ $(1 \leq i \leq 4)$ are specified by the 5G standard [2].

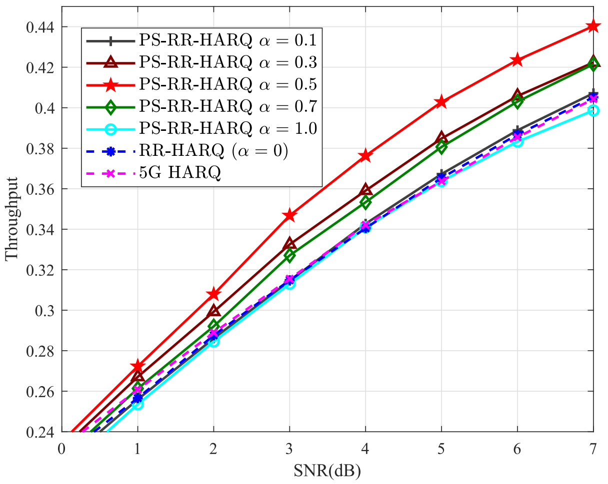
At the receiver, we need modify the decoding algorithms for PS-IR-HARQ as follows. First, if repeated bits exist, their LLRs are updated by Chase combining. Then the SISO algorithm for the first layer in Algorithm 1 is adapted to the lower-rate LDPC code with IRs. Finally, Algorithm 2 is modified accordingly.

Algorithm 2 PS-RR-HARQ Scheme

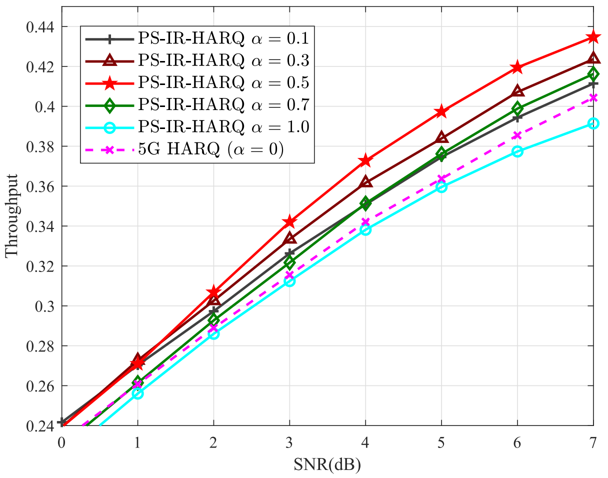
```
Initialization: The transmitter transmits $\boldsymbol{v}^{(0)}$ and the receiver computes LLRs $\boldsymbol{\ell}^{(0)}$;
for $t = 0, 1, \ldots, L$ do
    The receiver decodes estimation $\hat{\boldsymbol{v}}^{(t)}$ of $\boldsymbol{v}^{(t)}$ from $\boldsymbol{\ell}^{(t)}$;
    if $\hat{\boldsymbol{v}}^{(t)}$ is a valid codeword then
        The receiver feeds back an ACK;
        The transmitter transmits $\boldsymbol{v}^{(t+1)}$ and the receiver computes LLRs $\boldsymbol{\ell}^{(t+1)}$;
    else
        The receiver feeds back a NACK;
        The transmitter transmits $\boldsymbol{c}^{(t+1)} = \boldsymbol{v}^{(t+1)} + \boldsymbol{v}^{(t)}\boldsymbol{S}$ and the receiver computes LLRs $\boldsymbol{\ell}^{(t+1)}$;
        The receiver performs the message-passing decoding (Algorithm 1), resulting in $\hat{\boldsymbol{v}}^{(t)} = D(\boldsymbol{\ell}^{(t)}, \boldsymbol{\ell}^{(t+1)})$;
        while $\hat{\boldsymbol{v}}^{(t)}$ is not a valid codeword and the number of transmissions $\boldsymbol{v}^{(t)}$ is less than $T_{\max}$ do
            The receiver feeds back a NACK;
            The transmitter retransmits $\boldsymbol{v}^{(t)}$ and the receiver computes LLRs $\boldsymbol{r}$;
            The receiver updates $\boldsymbol{\ell}^{(t)}$ by Chase combining, i.e., $\boldsymbol{\ell}^{(t)} \leftarrow \boldsymbol{\ell}^{(t)} + \boldsymbol{r}$, and performs the message-passing decoding (Algorithm 1), resulting in $\hat{\boldsymbol{v}}^{(t)} = D(\boldsymbol{\ell}^{(t)}, \boldsymbol{\ell}^{(t+1)})$;
        The receiver feeds back an ACK or a NACK according to the decoding results and updates $\boldsymbol{\ell}^{(t+1)}$ by removing the effect of $\boldsymbol{v}^{(t)}$, i.e., $\boldsymbol{\ell}^{(t+1)} \leftarrow \boldsymbol{\ell}^{(t+1)} \cdot (-1)^{\hat{\boldsymbol{v}}^{(t)}\boldsymbol{S}}$;
```

C. Complexity Analysis

We take the transmission of $\boldsymbol{v}^{(0)}$ as an example to analyze the complexity increase in comparison with the conventional schemes. When $\boldsymbol{v}^{(0)}$ is successfully decoded at the first transmission, no extra computational loads are required. Otherwise, $n\alpha$ binary additions are required at the transmitter and an iterative message-passing algorithm is performed over a spatially coupled graph that specifies the superpositions. Since the $n\alpha$ extra nodes are of degree three, the main extra computational loads are caused by the enlarged graph, which is roughly twice as large as the original graph. This increase in complexity can be far less than one-fold over the conventional schemes in the high signal-to-noise ratio (SNR) region of interest where the first transmission is likely to be successful.



(a) PS-RR-HARQ scheme



(b) PS-IR-HARQ scheme

Fig. 4. Throughput comparison between the PS-HARQ schemes and the conventional HARQ schemes. Each information block of length $k = 720$ is protected by a 5G LDPC code, and each transmitted packet has a length of $n = 1440$. Notice that when $\alpha = 0$, PS-RR-HARQ is reduced to RR-HARQ and PS-IR-HARQ is reduced to 5G HARQ.

IV. Numerical Results

In this section, we compare the PS-HARQ schemes with the conventional HARQ schemes over independent block fading channels. In the case of cross-packet coding, the maximum global iteration number $J_{\max} = 2$, and the embedded LDPC codes are decoded by the sum-product algorithm (SPA) with a maximum iteration number of 25. For the original HARQ scheme, the maximum iteration number is 100. A binary sequence is segmented into consecutive blocks of length $k = 720$, each of which is encoded by a 5G LDPC code (constructed based on BG2 [2] with lifting factor $Z = 72$) of length 3600. The resulting coded bits except those punctured are stored in a circular buffer for future use. For the 5G HARQ scheme, the first and successive packets (upon request) are formed as $\boldsymbol{v}_{1\mathrm{T}}$, $\boldsymbol{v}_{2\mathrm{T}}$, $\boldsymbol{v}_{3\mathrm{T}}$ and $\boldsymbol{v}_{4\mathrm{T}}$ by specifying the starting positions (see Sec. II-B), each having length of $n = 1440$. The transmitted packets (upon request) for the other three schemes are scheduled in order as demonstrated in Table I. Notice that all the retransmissions $\boldsymbol{v}$ in the RR-HARQ schemes are identical to $\boldsymbol{v}_{1\mathrm{T}}$. The long-term average throughput (LTAT) is defined as the average number of successfully decoded information bits per transmitted symbol, which can be expressed as [15],

$$\rho = \lim_{t \to \infty} \frac{K(t)}{N(t)}, \tag{3}$$

where $K(t)$ is the total number of successfully decoded information bits up to time slot t and $N(t)$ is the total number of transmitted symbols up to time slot t. In this letter, we use Monte Carlo simulations with a sufficiently large number of time slots to estimate the LTAT performance.

Using the one-dimensional search with a step size of 0.1 for the superposition fraction α, some simulation results for PS-RR-HARQ schemes with different α are shown in Fig. 4(a), where we observe that the proposed PS-RR-HARQ scheme with $\alpha = 0.5$ can yield a throughput improvement of up to 10% over both the original 5G HARQ scheme and the RR-HARQ scheme. The simulation results also confirm that the throughput performance can be benefited from partial superposition with an optimized superposition fraction α. We have made similar observations for the proposed PS-IR-HARQ scheme as shown in Fig. 4(b). Compared with the 5G HARQ scheme, introducing partial superposition transmission brings a slightly increased implementation complexity but a throughput gain of up to 10%.

V. CONCLUSION

We have proposed throughput-enhanced HARQ schemes based on 5G LDPC codes, which introduce cross-packet coding using partial superposition into the original single-packet HARQ schemes. The proposed HARQ schemes are simple and flexible in the sense that they can be easily implemented with the basis of the existing single-packet HARQ schemes. Numerical results have shown that the proposed PS-HARQ schemes can have a throughput improvement of up to 10% over the original 5G HARQ scheme.

REFERENCES

[1] *Final Report of 3GPP TSG RAN WG1 #87 V1.0.0*, document R1-1701552, 3GPP, Feb. 2017. [Online]. Available: https://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_87/Report/Final_Minutes_report_RAN1%2387_v100.zip

[2] *Multiplexing and Channel Coding (Release 15)*, document TS 38.212 V15.1.1, 3GPP, Apr. 2018.

[3] G. Benelli, "An ARQ scheme with memory and soft error detectors," *IEEE Trans. Commun.*, vol. 33, no. 3, pp. 285–288, Mar. 1985.

[4] D. Chase, "Code combining—A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. 33, no. 5, pp. 385–393, May 1985.

[5] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[6] D. Mandelbaum, "An adaptive-feedback coding scheme using incremental redundancy (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 388–389, May 1974.

[7] C. Hausl and A. Chindapol, "Hybrid ARQ with cross-packet channel coding," *IEEE Commun. Lett.*, vol. 11, no. 5, pp. 434–436, May 2007.

[8] M. Jabi, A. Benyouss, M. Le Treust, E. Pierre-Doray, and L. Szczecinski, "Adaptive cross-packet HARQ," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2022–2035, May 2017.

[9] G. Cocco and D. Floreano, "Cross-packet coding for delay-constrained streaming applications," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1962–1966, Nov. 2019.

[10] Q. Wang, S. Cai, W. Lin, L. Chen, and X. Ma, "Spatially coupled LDPC codes via partial superposition," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019, pp. 2614–2618.

[11] P. Popovski, "Delayed channel state information: Incremental redundancy with backtrack retransmission," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 2045–2051.

[12] M. Jabi, E. Pierre-Doray, L. Szczecinski, and M. Benjillali, "How to boost the throughput of HARQ with off-the-shelf codes," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2319–2331, Jun. 2017.

[13] R. Zhang and L. Hanzo, "Superposition-coding-aided multiplexed hybrid ARQ scheme for improved end-to-end transmission efficiency," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4681–4686, Oct. 2009.

[14] Y. Lang, D. Wübben, A. Dekorsy, V. Braun, and U. Doetsch, "Improved HARQ based on network coding and its application in LTE," in *Proc. IEEE Wireless Commun. Netw. Conf*, Shanghai, China, Apr. 2012, pp. 1958–1963.

[15] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001.