

# Superposition-Coding Aided Multiplexed Hybrid ARQ Scheme for Improved End-to-End Transmission Efficiency

Rong Zhang and Lajos Hanzo

School of ECS., Univ. of Southampton, SO17 1BJ, UK.

Tel: +44-23-80-593 125, Fax: +44-23-80-593 045

Email: {rz05r,lh}@ecs.soton.ac.uk, <http://www-mobile.ecs.soton.ac.uk>

**Abstract**—In this paper, we propose a novel superposition coding aided multiplexed Hybrid Automatic Repeat reQuest (HARQ) scheme for the sake of improving both the link-layer and Transport Control Protocol (TCP) layer efficiency. The detailed system design is presented and the transmission efficiency metrics of both layers are discussed. Furthermore, the achievable link-layer packet error rate as well as a range of other transmission efficiency metrics of both layers are quantified. It is demonstrated that our scheme substantially improves the attainable transmission efficiency of both layers and it is particularly suitable for delay-sensitive services.

## I. INTRODUCTION

The Transport Control Protocol (TCP) supports reliable end-to-end data transmission and facilitates congestion control, where the transmission frame loss due to link errors is often assumed to be negligible in wired networks. However, directly transplanting the TCP into wireless applications suffers from link impairments, such as radio link attenuation, fading, handover, mobility and co-channel interference. For the sake of avoiding congestion due to physical retransmissions induced by channel errors, we may either conceal the effects of the wireless link from the TCP-enabled transmitter or we make it aware of the wireless channel effects [1]–[3].

From a cross-layer point of view [4], link-layer approaches such as Hybrid Automatic Repeat reQuest (HARQ) [5], which is proposed for the sake of further improving the robustness against link adaptation inaccuracy due to various implementation impairments, attempt to conceal the channel-induced packet loss events from the TCP-enabled transmitter by reducing the effects of wireless link errors with the aid of channel coding combined with retransmissions on a prompt packet-based timescale. This solution is appealing as it does not incur the typical overhead associated with TCP-awareness and yet obeys the TCP semantics. However, this HARQ aided approach introduces extra delay due to local link layer retransmissions, which may potentially lead to a timeout and hence may trigger the slow-start phase of the TCP transmission.

**Acknowledgements:** The work reported in this paper has formed part of the Core 4 Research Programme of the Virtual Center of Excellence in Mobile and Personal Communications, Mobile VCE, [www.mobilevce.com](http://www.mobilevce.com), whose funding support, including that of EPSRC, is gratefully acknowledged. Fully detailed technical reports on this research are available to Industrial Members of Mobile VCE.

Against this backdrop, in this paper we aim at improving the overall *end-to-end* TCP transmission efficiency by reducing the link layer's *hop-by-hop* HARQ retransmission delay with the aid of our proposed superposition coding aided Multiplexed HARQ (M-HARQ) scheme, which jointly encodes the current new packet to be transmitted and any packets that are about to be retransmitted. In other words, the link-layer retransmissions are embedded in the next new packet's transmission, which avoids any potential throughput reduction imposed by retransmissions although naturally, they do impose additional interference. A similar idea was proposed in [6], which requires a specifically designed channel code and its application is limited to twin-packet joint transmissions. As a benefit, our proposed scheme is capable of jointly and simultaneously transmitting multiple packets and it is equally applicable to both Type I and II HARQ techniques. Hence, the advocated technique can be seamlessly integrated with diverse existing and future systems.

In a nutshell, the contribution of this paper is that we propose a novel *M-HARQ* scheme, which improves both the link layer as well as the *end-to-end* TCP layer transmission efficiency at the cost of a marginal link layer Packet Error Ratio (PER) performance degradation.

The rest of the paper is organized as follows. In Section II, we provide a general description of the classic HARQ approach. Furthermore, the structure of our proposed M-HARQ arrangement is described, followed by the associated encoding and iterative decoding algorithms. In Section III, the link layer and TCP layer transmission efficiency metrics used are introduced. In Section IV, both the link layer PER performance and the transmission efficiency of both the conventional HARQ and the proposed M-HARQ scheme are evaluated and discussed. Finally, we conclude our discourse in Section V.

## II. MULTIPLEXED HYBRID ARQ

### A. Conventional Approach

Being a physical-layer-aware ARQ scheme, HARQ combines the Cyclic Redundancy Check (CRC) encoding function of the link layer with channel coding in the physical layer. In HARQ, the receiver asks for a packet's retransmission using the reverse-direction channel with the aid of a single-bit Negative-ACKnowledgement (NACK) flag, whenever its currently decoded packet is deemed to be erroneous based on the

decision of the CRC scheme. In general, the retransmissions in ARQ-aided systems can be carried out in different manners, for example using a Type-I Packet Combining (PC) scheme and a Type-II Incremental Redundancy (IR) scheme [7]. In this paper, we elaborate on Type-I HARQ although our proposed scheme is equally applicable to both types. In Type-I HARQ, the same coded packet is used in consecutive retransmissions, allowing the receiver having a sufficiently large memory to perform soft combining of the various replicas of the packets before decoding. Naturally, each packet is also individually decodable for a receiver without sufficient memory to decode each replica of retransmitted packets, although typically this results in a residual PER penalty or in an increased number of retransmissions.

Following the above conceptual introduction, let us now describe it mathematically. The information arriving from the upper layer, which is referred to here as a *frame*, is partitioned into  $M$  packets of equal length  $N_i$ ,  $\mathbf{u}_m \in \{0, 1\}^{N_i}$ ,  $m \in [1, M]$ , where  $N_i$  denotes the information packet length. These packets are protected by the channel coding function  $f_{c,v} \in \Omega = \{f_1, \dots, f_V\}$  of rate  $r_{c,v} \in \mathbb{R} = \{r_1, \dots, r_V\}$ , where  $\Omega$  and  $R$  represent a set of predefined discrete rate-compatible codes and their corresponding rates. The selection of a particular code-rate is based on the Channel Quality Indicator (CQI) controlling the link adaptation procedure. The maximum number of retransmissions is  $L < M$ , i.e. had a total of  $(L + 1)$  transmission attempts. For Type-I HARQ, the same coded packet is repeated  $L$  times, i.e. we have  $f_{c,v}^0(\mathbf{u}_m) = f_{c,v}^l(\mathbf{u}_m)$ ,  $l \in [1, L]$ , where the superscript '0' stands for the initial transmission. After successfully decoding the  $m$ th packet during the  $(L + 1)$ st transmission attempt, the transmission of the  $(m + 1)$ st packet is activated as illustrated in Fig. 1.

### B. Proposed Approach

The strategy of transmitting the next new packet only when the successful reception of the current one was confirmed is highly inefficient, which is analogous to the widely recognized drawback of conventional Stop-and-Wait ARQ [7]. However, if the receiver is capable of tolerating a modest amount of additional interference, the next new packet can be simultaneously transmitted with the retransmissions of the previous  $K \in [1, L]$  erroneous packets, as seen in Fig. 1. In other words,  $M$  new packets are continuously transmitted, while the  $K$  erroneous packets are transmitted on a virtual channel, appropriately combined with the new packets.

1) *Structure*: In general, different packets require different number of retransmissions, depending on the instantaneous channel conditions. We consider the worst-case scenario, where each packet exploited the maximum number of retransmissions  $L$ , so that we can evaluate the maximum of the PER after  $L$  retransmissions. In the worst-case scenario considered and when employing the superposition coding scheme to be introduced shortly, the resultant interference of our M-HARQ arrangement becomes similar to that of the Inter-Symbol-Interference (ISI) effects experienced for transmission over a dispersive channel in the absence of HARQ transmissions.

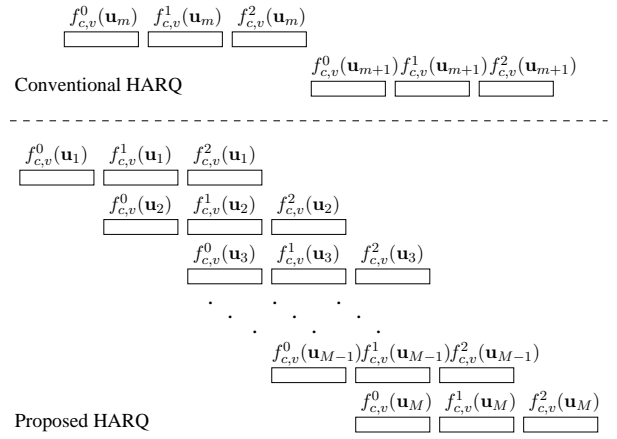


Fig. 1. Classic HARQ and the proposed multiplexed HARQ in conjunction with  $L = 2$ .

Analogously, our scheme may be interpreted as generating Inter-Packet-Interference (IPI) and it may be inferred from Fig. 1 that the conventional scheme requires a total of  $M_r = M(L + 1)$  packet transmissions, while our scheme necessitates only  $M_r = M + L$  transmissions.

2) *Encoding and Decoding*: Generally speaking, the joint encoding function  $F$  of the  $m$ th transmission can be represented as  $F(\mathbf{u}_{a_1}, \dots, \mathbf{u}_{a_2})$ , where we have:

$$\begin{cases} (a_1, a_2) = (m, 1) & 1 \leq m < L, \\ (a_1, a_2) = (m, m - L) & L \leq m \leq M, \\ (a_1, a_2) = (M, m - L) & M < m \leq M + L. \end{cases} \quad (1)$$

Although in principle specifically designed coding functions may be created, we opt for the powerful superposition coding concept in this paper:

$$F(\cdot) = \sum_{i=a_2}^{a_1} \rho_i e^{j\theta_i} f [f_{c,v}^{m-i}(\mathbf{u}_i)], \quad (2)$$

where each superimposed packet is referred to as a layer, while  $\rho_i$  and  $\theta_i \in [0, \pi)$  denote the layer-specific amplitude- and phase-rotation, respectively. In this paper, an identical amplitude allocation and uniform phase rotations are employed for the individual superimposed layers. The benefit of choosing this particular superposition coding technique is that by opting for this simple linear operation, the specific modulation function  $f(\cdot)$  and channel coding function  $f_{c,v}(\cdot)$  of the individual layers may be retained.

Our M-HARQ scheme employs iterative Multiple Packets Detection (MPD) and Channel Decoding (DEC) exchanging extrinsic information between these two receiver components. The choice of the DEC algorithm depends on the specific channel code employed, however, a host of MPD schemes may be invoked, including the powerful but high-complexity Maximum Likelihood (ML) detection scheme or we may opt for employing a low-complexity soft interference cancellation scheme [8] in this paper.

*Remarks*: Instead of superposition coding, multiple packets may be orthogonally multiplexed within a specific transmission attempt without imposing any IPI. However, maintaining orthogonality amongst the packets requires additional Direct

Sequence (DS)-spreading of the original channel coded packet, hence resulting in a rate-loss. Since orthogonal channel codes are hard to design, we may exploit the multiplexing capability inherently provided by channel codes having a rate less than unity by differentiating the layers with the aid of their unique, layer-specific channel codes. Naturally, this is achieved at the cost of an increased complexity and marginal PER performance degradation.

### III. TRANSMISSION EFFICIENCY METRIC

A general packet-based wireless network is constituted by a wired link spanning from a server to an Access Point (AP) and a wireless link from the AP to the Mobile Terminal (MT). In this section, we discuss the transmission efficiency metrics to be applied for the link layer and for the TCP layer, respectively, which are the effective throughput  $\eta$  and the mean frame arrival rate  $\lambda$ .

The following simplifying assumptions are stipulated. Firstly, when considering the TCP layer's transmission efficiency, we ignore any propagation delay over the wireless channel and the feedback delay encountered during the transmission of the link layer's error-free ACK and NACK indication. Furthermore, for transmission over the block fading channel considered in this paper, the packet error events are independently and identically distributed (i.i.d.), resulting in a low probability of TCP timeouts due to consecutive packet loss events. Hence, we ignore the so-called slow-start TCP phase and only consider the congestion-avoidance phase of the widely used TCP Reno [9]. We also assume that the M/G/1 queue [9] has a Poissonian arrival process having an arrival rate of  $\lambda$ , a general i.i.d. service time duration  $T$  and a single server.

#### A. Effective Throughput $\eta$

Let us define the normalised effective throughput  $\eta$  as the product of the throughput per packet  $\eta_0$  and the total number of packets  $M$  divided by the total number  $M_r$  of transmissions required, yielding  $\eta = \eta_0 M / M_r$ , where the per-packet throughput is given by:  $\eta_0(\gamma_b) = r \cdot b [1 - p_e(\gamma_b)]$ , where  $r$  and  $b$  are the channel coding rate and the number of bits per symbol determined by the modulation scheme employed. Furthermore,  $p_e$  denotes the link layer's PER as a function of the Signal-to-Noise-Ratio per-bit denoted by  $\gamma_b$ . This metric assumes that each packet exhausts all the  $L$  retransmissions for the sake of simplified comparisons.

#### B. Mean Frame Arrival Rate $\lambda$

The normalised effective throughput achieved at the TCP layer may be measured by the so-called mean transmission frame arrival rate  $\lambda$  encountered, which is determined by the average number of TCP frames  $N$  successfully transmitted within the average Round Trip Time (RTT)  $\tau$  [10]. More explicitly, we have  $\lambda = N / \tau$ , where  $N = 3w/4$  for the Stop-and-Wait ARQ regime assumed in the TCP layer [11]. Below, we will elaborate on the effects of two parameters used in the calculation of  $\lambda$ , namely those of the congestion window size  $w$  and of the average RTT  $\tau$ .

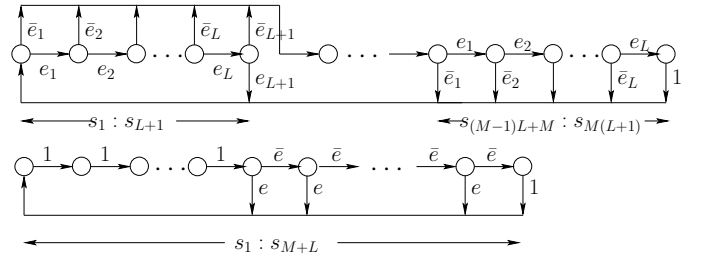


Fig. 2. The Markov model of both the conventional scheme (top) and of the M-HARQ scheme (bottom).

1) *Average RTT*: The average RTT is given by  $\tau = D + D_q + E[T]$ , where  $D$  is the propagation delay in the wired link spanning from the server to the AP,  $E[T]$  is the expected transmission duration of a TCP frame and  $D_q$  is the expected queueing delay, which is given by the Pollaczek-Khinchine formula of [10], [11]:  $D_q = \lambda E[T^2] / [2(1 - \lambda E[T])]$ . Let us now derive the first and second moments of the transmission duration  $T$  of a TCP frame, i.e.  $E[T]$  and  $E[T^2]$  in order to get the average RTT  $\tau$ .

A TCP frame is deemed to be successfully received only when all of its  $M$  link-layer packets are correctly received. If any packet of a TCP frame fails the CRC-test after  $(L + 1)$  transmissions, this frame will be dropped and the next frame is to be sent. This procedure may be modelled with the aid of a Markov chain as follows [11], [12].

*Markov chain.* Denote the set of states in the Markov chain by  $\mathcal{S}$ , where  $|\mathcal{S}| = M_r$  stands for the maximum number of link layer packet transmissions within a single TCP frame. Let us consider the conventional HARQ first as seen in the upper part of Fig. 2. There are a total of  $M_r = M(L + 1)$  transmission states and each link layer packet of a TCP frame has a maximum of  $(L + 1)$  transmissions. Every erroneous packet transmission occurring with a probability of  $e_j, j \in [1, L]$  leads from the current state represented by a circle to a state transition within the current packet. Following the final erroneous transmission of the link layer, the entire current TCP frame is dropped and the corresponding state transition leads directly to the next TCP frame represented by  $s_1$  with a probability of  $e_{L+1}$ . By contrast, a successful transmission having a probability of  $\bar{e}_j, j \in [1, L + 1]$  leads to a state transition by-passing the retransmissions and initiating a new packet's transmission.

However, the proposed M-HARQ scheme exhibits different characteristics, as seen in the lower part of Fig. 2. There are a total of  $M_r = (M + L)$  transmission states and each packet has a maximum of  $(L + 1)$  transmissions. The first  $L$  states are associated with a unity transmission probability, since the maximum number of  $(L + 1)$  transmissions of the first packet has not been exhausted. Once the  $(L + 1)$  packet transmissions of the first packet have been exhausted, the state transition diagram of Fig. 2 traverses to the next state corresponding to the next new packet's transmission in the current TCP frame with a probability of  $\bar{e}$ . By contrast, if none of the  $(L + 1)$  transmissions were successful, then the state transition evolves to the next new frame with a probability of  $e$ .

*Transition probability.* The transition probability  $p_c(s, s')$  of

$p_c(s, s')$	$\{s, s'\}$
$e_j$	$\mathcal{S}_1 = \{i, j, 1\}$ $i \in [1, M-1], j = L+1$
$e_j$	$\mathcal{S}_2 = \{(i-1)(L+1) + j, s+1\}$ $i \in [1, M], j \in [1, L]$
$\bar{e}_j$	$\mathcal{S}_3 = \{(M-1)(L+1) + j, 1\}$ $j \in [1, L]$
$\bar{e}_j$	$\mathcal{S}_4 = \{(i-1)(L+1) + j, i(L+1) + 1\}$ $i \in [1, M-1], j \in [1, L+1]$
1	$\mathcal{S}_5 = \{M(L+1), 1\}$
0	$s, s' \in \mathcal{S} - \{\mathcal{S}_i\}_{i=1}^5$
$p_p(s, s')$	$\{s, s'\}$
$e$	$\mathcal{S}_1 = \{i, 1\}$ $i \in [L+1, M+L-1]$
$\bar{e}$	$\mathcal{S}_2 = \{i, i+1\}$ $i \in [L+1, M+L-1]$
1	$\mathcal{S}_3 = \{M+L, 1\}; \mathcal{S}_4 = \{i, i+1\}$ $i \in [1, L]$
0	$s, s' \in \mathcal{S} - \{\mathcal{S}_i\}_{i=1}^4$

TABLE I

THE TRANSITION PROBABILITY OF THE CONVENTIONAL SCHEME  $p_c(s, s')$  FROM STATE  $s \in \mathcal{S}$  TO  $s' \in \mathcal{S}$  AND OF THE PROPOSED SCHEME  $p_p(s, s')$ .

the conventional HARQ scheme's Markov chain from state  $s \in \mathcal{S}$  to  $s' \in \mathcal{S}$  and that of the proposed M-HARQ scheme denoted by  $p_p(s, s')$  are given in Table I, where  $e_j$  and  $\bar{e}_j = 1 - e_j$  denotes the probability of erroneous and correct link layer packet reception during the  $j$ th transmission, respectively. Likewise,  $e = \prod_{j=1}^{L+1} e_j$  and  $\bar{e} = 1 - e$  denotes the probability of erroneous and correct link layer packet reception after  $(L+1)$  transmissions, respectively. These transition probabilities are then incorporated in the transition probability matrix  $\mathbf{P}$  of size  $|\mathcal{S}| \times |\mathcal{S}|$ .

*State probability.* The state probability vector  $\mathbf{p}$  hosts the probability of each state  $s \in \mathcal{S}$ . The initial state probability  $\mathbf{p}^0$  is modelled by the entries  $p^0(s=1) = 1$  and  $p^0(s > 1) = 0$ . The transition state probability  $\mathbf{p}^q$  at the  $q$ th transition phase is given by  $\mathbf{p}^q = \mathbf{p}^0 \hat{\mathbf{P}}^q$ , where we have  $q \in [\min(L+1, M), |\mathcal{S}|]$  and  $\hat{\mathbf{P}}$  equals to the transition probability matrix  $\mathbf{P}$ , except that the transition to state  $s = 1$  corresponding to the first column of  $\mathbf{P}$ , namely to  $\mathbf{z} = \mathbf{P}(:, 1)$  is omitted from  $\hat{\mathbf{P}}$ . Finally, the steady state probability  $\mathbf{p}^s$  is obtained by solving the linear equations  $p^s(s) = \sum_{s' \in \mathcal{S}} p^s(s') p(s', s)$  subject to the constraint  $\sum_{s \in \mathcal{S}} p^s(s) = 1$ .

Hence, the first and second moments of  $T$  are expressed as  $E[T] = \sum_{q=a_2}^{a_1} \mathbf{p}^q \mathbf{z} t_q$  and  $E[T^2] = \sum_{q=a_2}^{a_1} \mathbf{p}^q \mathbf{z} t_q^2$ , respectively, where we have  $a_1 = |\mathcal{S}|$  and  $a_2 = \min(L+1, M)$ . Furthermore,  $t_q = q t_0$  stands for the time required for the transmission of  $q$  link-layer packets with each packet's transmission duration  $t_0$  set to unity in this paper.

2) *Congestion Window Size:* The congestion window size  $w$  in the TCP mechanism is employed to control the frame flow by additively increasing its window size every RTT, until either the maximum allowable buffer-size is reached or the congestion is resolved. If so, then the window size is halved. Hence the maximum congestion window size is determined by two design parameters, namely the affordable buffer-size and the acceptable Frame Error Rate (FER).

We refer to a buffer-size limited scenario, when potentially no frame error events are experienced, because the window

size of  $w_B$  is increased until the buffer is filled. The subsequently arriving TCP frames are then dropped and the window size of  $w_B$  is halved in the interest of getting ready for future window expansion to prevent future frame dropping events. The window size  $w_B$  is given by  $w_B = B + \tau/E[T] + 1$  [11], where  $1/E[T]$  represents the average TCP frame transmission rate. On the other hand, if residual frame error events persist after link layer HARQ retransmissions, the window size  $w_e$  is determined by the residual FER  $p_e^f$  experienced and  $w_e$  is halved for the same reason as argued above. The FER  $p_e^f$  corresponds to the maximum of  $(L+1)$  link layer HARQ transmissions, which is given by  $p_e^f = \mathbf{p}_s \mathbf{z}_e / \mathbf{p}_s \mathbf{z}$ , where  $\mathbf{z}_e$  equals to the vector of  $\mathbf{z}$  hosting the transition probabilities leading to state  $s = 1$  due to failure. Hence the average number of TCP frames transmitted between two consecutive frame loss events is roughly  $1/p_e^f$ , which is also approximately given by  $3w_e(w_e/2 + 1)/4$ , hence we arrive at the window size of  $w_e \approx \sqrt{8/3p_e^f}$ .

When considering both the buffer-size limited and the FER limited behavior, the ultimate window size  $w$  is recommended to be set to  $w = \min(w_B, w_e)$ . Based on the value of both the average RTT  $\tau$  and the congestion window size  $w$ , the mean frame arrival rate may be calculated as  $\lambda = 3w/4\tau$ , where we obtain the value of  $\lambda$  by fixed point iterations.

## IV. PERFORMANCE EVALUATION

### A. PER Investigation

Let us now evaluate the link layer PER performance of our proposed M-HARQ scheme. Fig. 3 shows the PER performance of the proposed arrangement against that of the conventional scheme for a total of  $L+1 = 3$  transmissions employing Type-I HARQ. In practice, a total of two or three transmissions are sufficient, since the HARQ scheme acts like a 'safety net' in support of the link adaptation procedure, which is capable of preventing most of the potential packet loss events. In our simulations, each packet of length  $N_i = 256$  bits is QPSK modulated and channel coded by a rate-1/3 irregular systematic Repeat Accumulate (RA) code [13]. A Rayleigh distributed block-fading channel is used and the feedback channel conveying the NACK indicator is assumed to be error-free. Again, we consider the worst case scenario, where each of the  $M$  packets employs the maximum affordable number of  $L = 2$  retransmissions.

We now investigate the PER of all the  $(L+1)$  transmissions for each of the first  $(L+1)$  packets, since they correspond to different typical interference patterns  $\Omega$ . For instance, when  $L = 2$  is considered, the number of layers for each of the  $L+1 = 3$  transmissions of the  $L+1 = 3$  first packets is given by  $\Omega_{pck1} = [1, 2, 3]$ ,  $\Omega_{pck2} = [2, 3, 3]$  and  $\Omega_{pck3} = [3, 3, 3]$ . Fig. 3 suggests that during the first transmission the PER performance of our proposed scheme is the same as that of the conventional scheme. By contrast, for two and three transmissions, there is an observable but marginal PER degradation for our proposed scheme compared to that of the conventional one. Apart from this slight difference, all packets experience a near-identical PER performance.

### B. Efficiency Evaluation

Let us now investigate the link layer's effective throughput  $\eta$  for both our proposed M-HARQ scheme and for the conventional scheme. Observe in Fig. 4 the significantly improved effective throughput of our proposed M-HARQ arrangement as compared to that of the conventional one for both  $M = 4$  and  $M = 12$ . When the total number of transmitted packets  $M$  is significantly higher than  $L$ , the effective throughput  $\eta$  of our proposed scheme approaches that of the single-transmission scenario, which can be verified by comparing the results of both the  $L = 1$  and  $L = 2$  scenarios corresponding to  $M = 12$  in Fig. 4, where the  $L = 0$  curve is printed using the continuous line. This implies that there is only a marginal retransmission delay penalty for our proposed M-HARQ scheme for  $M \rightarrow \infty$ .

Fig. 5 compares the mean frame arrival rate  $\lambda$  recorded for a Poissonian source frame generation process for both the conventional HARQ scheme and for our proposed arrangement at the TCP layer. Two different phenomena may be observed in Fig. 5, namely the 'buffer-size limited' and the 'FER limited' situations seen at the right and left of the figure. More explicitly, at low  $E_b/N_0$  values,  $\lambda$  is limited by the high FER imposed by the channel, which activates more retransmissions. On the other hand,  $\lambda$  is limited by the finite buffer-size of  $B$ , when the FER is low, such as  $p_e \leq 0.02$  for our proposed scheme and  $p_e \leq 0.003$  for the conventional one. This is clearly seen in Fig. 5 upon increasing the buffer-size from  $B = 1$  to 20. It can also be seen in Fig. 5 that our proposed scheme substantially improves the mean successful TCP frame arrival rate at low  $\gamma_b$  values, before the conventional scheme reaches its buffer-limited maximum. Although both schemes reach the same maximum value of  $\lambda$  bounded by the buffer-size  $B$ , the proposed M-HARQ arrangement requires lower  $E_b/N_0$  values than the conventional HARQ, when comparing points A and C in Fig. 5. This implies that our scheme is more tolerant to frame error events and hence has a higher end-to-end throughput as a benefit of employing the proposed M-HARQ scheme.

### C. Discussion

Our proposed scheme is based on the superposition coding approach and hence the resultant composite packet of multiple superimposed layers becomes effectively 'interference-limited'. Therefore, the per-layer throughput rate should not be excessive in order to ensure that the decoded PER remains low and approaches the single-layer best-case performance, as illustrated in Fig. 3. More explicitly, this requirement discourages the employment of high-throughput, but interference-sensitive, high-order modulation schemes.

Furthermore, relatively low-rate channel codes are preferred for the sake of supporting the low PER transmission of multiple superimposed layers at a near-single-layer PER performance. Since the number of retransmissions  $L$  is typically low in practice, so is the number of superimposed layers. Additionally, only a Type-I HARQ scheme was investigated based on the argument that Type-II HARQ provides a limited extra gain over Type-I HARQ for low-order modulation and

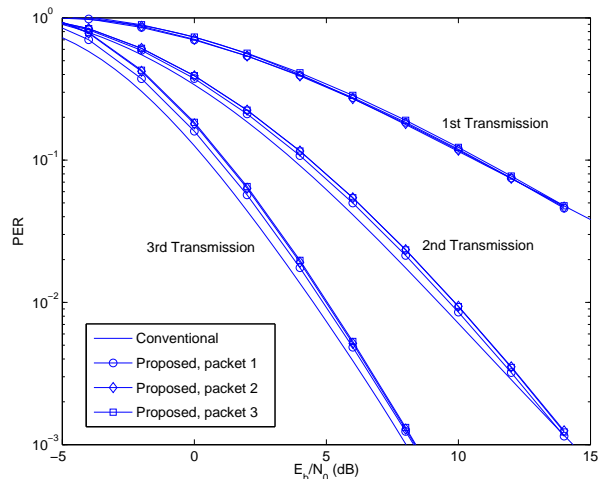


Fig. 3. The PER performance of all  $L + 1 = 3$  transmissions for both the conventional HARQ and for the first  $L + 1 = 3$  packets of the proposed M-HARQ scheme.

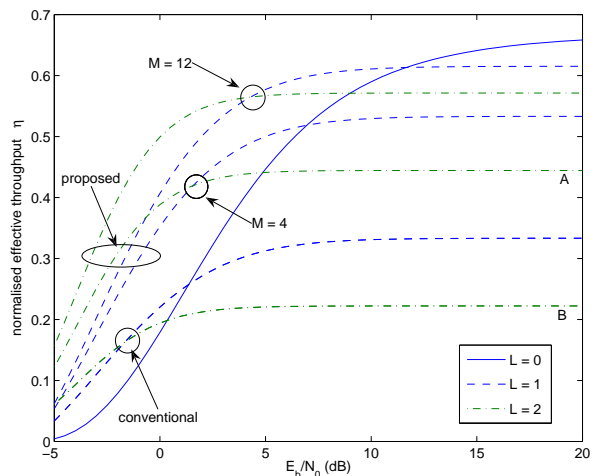


Fig. 4. The effective throughput of the first and second retransmissions for both the conventional HARQ and the proposed M-HARQ scheme.

low-rate channel coding [14], although the employment of Type-II HARQ is also straightforward.

## V. CONCLUSION

In this paper, a novel superposition-aided multiplexed HARQ scheme was proposed, which is capable of substantially improving the link layer's effective throughput for all transmitted packets at a marginal PER performance degradation. This improved link layer transmission efficiency also contributed towards an improved overall end-to-end transmission efficiency. Our superposition coding aided arrangement may be readily integrated with existing systems without substantially modifying the current design. It is particularly suitable for delay-sensitive low-rate services and for providing cell-edge users with an improved end-to-end throughput and/or transmission integrity.

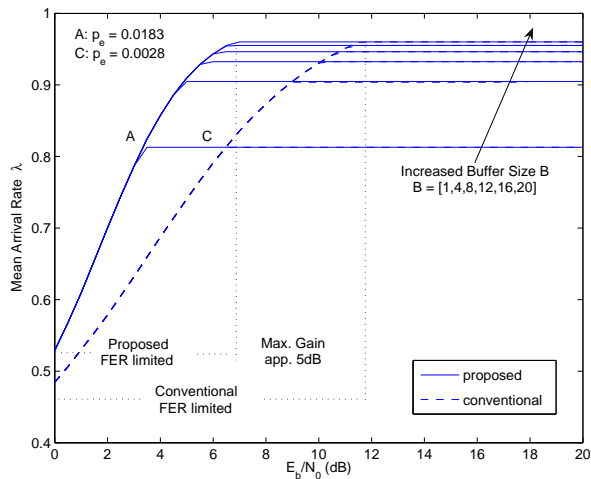


Fig. 5. The mean arrival rate for both the conventional HARQ and for the proposed M-HARQ scheme.

## REFERENCES

- [1] A. Bakre and B. Badrinath, "I-TCP: indirect TCP for mobile hosts," in *Proceedings of IEEE 15th International Conference on Distributed Computing Systems*, Vancouver, Canada, May28-31 1995, pp. 136–143.
- [2] H. Balakrishnan, S. Seshan, and R. Katz, "Improving reliable transport and handoff performance in cellular wireless networks," *ACM Wireless Networks*, vol. 4, pp. 129–136, Dec. 1995.
- [3] R. Caceres and L. Iftode, "Improving the performance of reliable transport protocols in mobile computing environments," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 850–857, June 1995.
- [4] S. Shakkottai, T. Rappaport, and P. Karlsson, "Cross-layer design for wireless networks," *IEEE Communications Magazine*, vol. 41, pp. 74–80, Oct. 2003.
- [5] L. Hanzo, J. Blogh, and S. Ni, *3G, HSDPA, HSUPA and Intelligent FDD versus TDD Networking: Smart Antennas and Adaptive Modulation*. IEEE Press - John Wiley, 2008.
- [6] C. Hausl and A. Chindapol, "Hybrid ARQ with cross-packet channel coding," *IEEE Communications Letters*, vol. 11, pp. 434–436, May 2007.
- [7] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, 2nd ed. New York: Prentice-Hall, Inc., 2005.
- [8] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Transactions on Communications*, vol. 47, pp. 1046–1061, July 1999.
- [9] W. Stevens, *TCP/IP Illustrated, Volume I: The Protocols*. MA: Addison-Wesley, 1994.
- [10] L. Kleinrock, *Queueing Systems, Volume I and II*. New York: Wiley, 1976.
- [11] Q. Qian Huang, S. Chan, L. Ping, and M. Zukerman, "Improving wireless TCP throughput by a novel TCM-based hybrid ARQ," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 2476–2485, 2007.
- [12] A. Bakre and B. Badrinath, "Improving TCP over wireless through adaptive link layer setting," in *Proceedings of IEEE GLOBECOM '01*, San Antonio, USA, Nov.25-29 2001, pp. 1766–17703.
- [13] J. Jin, A. Khandekar, and R. J. McEliece, "Irregular repeat-accumulate codes," in *Proc. of the 2nd International Conference on Turbo Codes*, Brest, France, Sept.4 – 7, 2000, pp. 125–127.
- [14] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA," in *Proceedings of IEEE VTC '01 Fall*, Atlantic City, New Jersey, USA, Oct.7-11, 2001, pp. 1829–1833.