This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Grouped Packet Coding: A Method for Reliable Communication Over Fading Channels With Long Delays

Rameez Ahmed and Milica Stojanovic, *Fellow, IEEE*

***Abstract*—In this paper, we investigate an automatic repeat request (ARQ) for reliable transmission over half-duplex links. We design a method based on grouped packet coding (GPC) that combines a stop-and-wait (S&W) ARQ procedure with random linear packet coding and selective acknowledgments applied to groups of coded packets. Our goal in doing so is to boost the throughput efficiency on poor-quality links with long delay. Such links are notably encountered in underwater acoustic channels, where the bit error rate may be as high as 10^{-3} and round-trip delays can be measured in thousands of bits. To quantify the benefits of the proposed S&W-GPC method, we evaluate its throughput efficiency analytically and compare it with the throughput efficiency of standard S&W methods, as well as the benchmark efficiency of full-duplex methods. Our results show that S&W-GPC outperforms all other techniques on half-duplex links with long delay, as well as rateless packet coding on full-duplex links with long delay. We present results for a point-to-point link, as well as for a multicast network. In addition to the performance analysis, we offer guidelines for an optimal system design, which involves a judicious choice of the packet size, packet coding, and grouping parameters.**

***Index Terms*—Underwater communication, packet coding, ARQ, half-duple.**

I. Introduction

In this paper, we investigate random linear packet coding for channels that experience long delays and time-varying propagation conditions that contribute to the high bit (packet) error rates and latency. Our work is motivated by the problems in underwater acoustic communications, but applies to other wireless systems that experience either high latency or poor quality of the physical link. We note that an arbitrary channel cannot be replaced by an acoustic one when it comes to physical-layer aspects of propagation modeling and signal processing, but data-link-layer aspects of our analysis are applicable to all channels that exhibits poor bit error rate (BER) and long propagation/processing delays.

Propagation delay is a significant challenge for underwater acoustic communication, which is used in a variety of systems, such as those involving multiple underwater autonomous vehicles, deep-sea oil and gas field maintenance, climate recording, and biological ecosystem monitoring. Acoustic signals propagate well underwater, but bring a number of challenges for communication [1]. Since the speed of sound in water is low (nominally 1500 m/s), acoustic communication suffers from long propagation delays, which contribute to link latency similar to that of satellite systems. Underwater acoustic channels also suffer from pronounced multipath and Doppler distortion, which contributes to high packet loss rates similar to that of some mobile terrestrial systems, e.g., in-flight systems and cellular services in high-speed trains. Combined with the half-duplex nature of most acoustic modems [2], these challenges call for a dedicated link design to provide reliable communication.

Traditionally, automatic repeat request (ARQ) protocols, such as stop-and-wait (S&W), go-back-N, and selective repeat, are used to make a link reliable. Coupled with long propagation delays, these techniques become inefficient as they rely on waiting for feedback from the receiver.

Optimization of ARQ techniques for underwater acoustic channels was discussed in [3]. As acoustic modems are currently constrained to the half-duplex operation, the choice of an ARQ protocol is limited to the S&W family. Two modifications of the basic S&W techniques that draw on the concepts used in satellite communications were studied there, and their performance was compared in terms of the throughput efficiency. It was shown that the grouping of packets and the use of selective acknowledgments improve the throughput efficiency. Reliable data transfer from one half-duplex node to another was also addressed in [4], where the long propagation delay was exploited by allowing the two nodes to transmit simultaneously in a juggling fashion. A comparative performance analysis of ARQ protocols for underwater acoustic networks was presented in [5]. A variation of the selective-repeat ARQ and two hybrid ARQ techniques were studied there for a multiuser underwater network, and their performances were compared in terms of the throughput efficiency and the average packet delay. Another variation of the selective-repeat ARQ was proposed in [6], where the authors used the long propagation delay to set up an interlaced time-division duplexed link. A transmission scheme for a continuous ARQ protocol over underwater acoustic channels was proposed in [7], in which the authors used an idle period after the transmission of every packet to time the reception of acknowledgments for continuous transmission, similar to juggling.

Manuscript received September 18, 2017; revised April 18, 2018; accepted July 6, 2018. This work was supported in part by the National Science Foundation under Grant NSF CNS-1726512 and Grant NSF 1428567, and in part by the Office of Naval Research under Grant ONR N00014-15-1-2550. *(Corresponding author: Rameez Ahmed.)*

Associate Editor: M. Chitre.

The authors are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: rarameez@gmail.com; millitsa@ece.neu.edu).

Digital Object Identifier 10.1109/JOE.2018.2855498

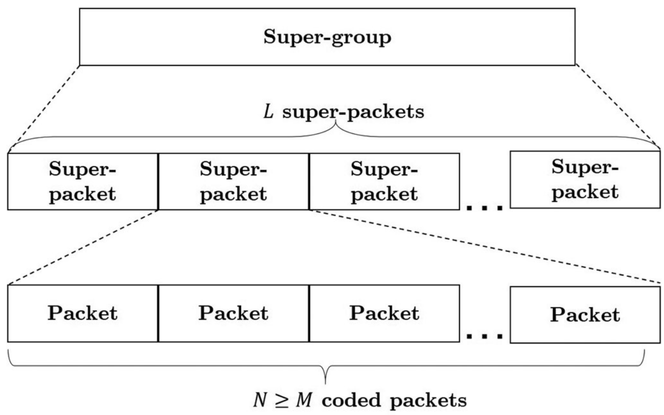


Fig. 1. Hierarchy of the proposed GPC technique. M information-bearing packets are encoded into$N \geq M$ packets for transmission. N coded packets form one superpacket. The superpackets are grouped into groups of L to form supergroups. A supergroup is subject to a selective acknowledgment procedure.

In this paper, we explore random linear packet coding as an attractive addition for achieving link reliability. In a packet-coded system, a group of M information-bearing packets is encoded into $N \geq M$ coded packets for transmission [8]. The receiver can decode the original information-bearing packets from a subset of any M out of the N coded packets. It should be noted that the packet coding does not replace channel coding, but can be used in addition to it. Since packet coding is performed at the packet level (as opposed to the bit level), it is readily applicable to any existing physical-layer technique.

Packet coding for channels with long propagation delay was studied in [9]–[13]. In [9], rateless codes were considered for reliable data transfer in underwater acoustic networks. It was shown there that the throughput efficiency improved since the feedback was used less often. In [10], optimal schedules were investigated for packet coding on a half-duplex link, and showed that an optimal number of coded packets exist, which minimizes the time (or energy) required to complete the transmission of a group of packets. Optimal strategies for information broadcasting using random linear packet coding were addressed in [11] showing performance improvements over traditional ARQ techniques. Multihop reliable data transfer for an underwater acoustic network using fountain codes was proposed in [12], in which, under the assumption of half-duplex operation, the block size of each hop was adapted so as to optimize the end-to-end delay. Joint power and rate control for an acoustic link employing random linear packet coding was considered in [13]. It was shown there that a small additional redundancy suffices to maintain a prespecified reliability at the receiver.

In this paper, we design a fully reliable link using ARQ in conjunction with packet coding. In particular, we regard a group of N coded packets as one unit, which we refer to as a superpacket. We group L such superpackets together, forming a supergroup which is then transmitted. The receiver sends a selective acknowledgment for each supergroup. If a certain superpacket within a supergroup is negatively acknowledged, it is retransmitted according to an ARQ technique. The packet grouping hierarchy is illustrated in Fig. 1.

We compare the throughput efficiency of the proposed technique with conventional ARQ techniques and benchmark it against a full-duplex link. We present numerical results for a point-to-point link as well as for a multicast network. We also offer guidelines to determine the optimal packet length and supergroup and superpacket size, such that the overall throughput efficiency is maximized.

The rest of this paper is organized as follows. We present the system model in Section II. Section III is devoted to performance analysis using numerical examples and simulation. The results for a multicast network are presented in Section IV, and the conclusions are summarized in Section V.

II. Throughput Efficiency

A. Basic ARQ

We begin with the basic S&W ARQ technique since we consider a half-duplex channel. In this technique, the transmitter sends a packet and waits for an acknowledgment (ACK) from the receiver before sending the next packet. The basic S&W has throughput efficiency [15]

$$\eta_{\text{S\&W}} = (1 - P_E)\frac{N_b}{K_b + N_{\text{rt}} + N_a} \approx (1 - P_E)\frac{N_b}{K_b + N_{\text{rt}}} \quad (1)$$

where the following notation is used:

- P_E probability of packet loss;
- N_b number of information-bearing bits per packet;
- K_b total number of bits per packet (including the N_{oh} overhead bits used for synchronization, control, cyclic redundancy check, etc., there are $C_b = N_b + N_{\text{oh}}$ bits per packet; after channel coding at a rate ρ_c, there are $K_b = \lceil C_b/\rho_c \rceil$ bits per packet);
- N_{rt} number of bits corresponding to the round-trip delay (round-trip delay over a distance d is $T_{\text{rt}} = 2d/c$, where c is the speed of signal propagation, and $N_{\text{rt}} = \lceil T_{\text{rt}}/R_b \rceil$, where R_b is the bit rate in the channel);
- N_a number of bits needed to acknowledge one packet (negligible compared to K_b).

For simplicity, we ignore the possibility of acknowledgment packets being lost (in practice, acknowledgment packets can be protected by a strong channel code).

Throughput efficiency of S&W is severely limited on links with high latency. In comparison, throughput efficiency of a full-duplex link employing selective repeat procedure is insensitive to the round-trip delay. Assuming negligible ACK duration, it is given by [15]

$$\eta_{\text{FD}} = (1 - P_E)\frac{N_b}{K_b}. \quad (2)$$

The above-mentioned expression corresponds to an ideal case with unlimited memory space. In practice, a limit on the transmission window (number of packets that have been transmitted but whose ACKs are yet to arrive) is imposed by two factors: storage space (unacknowledged packets need to be buffered) and numbering overhead (all packets have to be numbered because retransmissions are delivered out of order; if one must reckon with a possibly infinite delivery delay, the numbering overhead grows accordingly). A tolerable average delivery delay also plays a role. This problem is solved by combining selective repeat with either "stuttering" or Go-Back-N, whereby only a finite number of selective repeats are attempted before resorting to persistent retransmission of an erroneous packet [15], [16]. The buffer sizes are now finite, but the corresponding efficiency is somewhat decreased. Specifically, if the preset maximum number of selective repeats is ν and Go-Back-N is used with L_{rt} packets fitting into the round-trip delay, the number of packets to be buffered is νL_{rt}, and the throughput efficiency is [16]

$$\eta'_{\text{FD}} = \frac{1-P_E}{1+L_{\text{rt}}P_E^{\nu+1}} \cdot \frac{N_b}{K_b}. \tag{3}$$

This throughput efficiency is lower than that of the ideal full-duplex (2), but approaches it as ν grows. With that in mind, in what follows we will not concern ourselves with buffering issues, and will use the ideal full-duplex as a benchmark upper bound.

When full-duplex is not available, as is the case in acoustic channels, a classical improvement to S&W is sought through packet grouping and selective repeat ARQ [3]. In this method, which we call S&W with grouping (S&W-G), L packets are sent together and a selective acknowledgment is received for each packet in a group. Packets that are negatively acknowledged are retransmitted in the next group of L packets, along with any new packets. The resulting throughput efficiency is [3]

$$\eta_{\text{S\&W,G}} = (1-P_E)\frac{LN_b}{LK_b+N_{\text{rt}}+LN_a} \approx (1-P_E)\frac{LN_b}{LK_b+N_{\text{rt}}}. \tag{4}$$

Clearly, as L increases, so does the throughput efficiency of S&W-G. A practical limit on the group size L is again imposed by the finite memory.

Packet grouping improves the performance of S&W by filling the idle time with new packets, thus alleviating the issue of long delay. However, propagation delay is not the only factor that limits the ARQ performance. The other factor is the probability of packet loss P_E. On poor-quality links where P_E is high, such as acoustic links, the performance of S&W-G suffers even if large group sizes L are used. Packet coding offers a remedy for this problem.

B. Packet Coding

In a packet-coded system, M information-bearing packets, each carrying N_b information bits, are encoded into N packets, and such a group of N coded packets is regarded as one superpacket. The process of random linear packet coding is explained in detail with examples in [8]. With random linear packet coding performed in the Galois field $GF(2^q)$, each coded packet contains $C_b = N_b + Mq + N_{\text{oh}}$ bits, where the Mq extra bits are used to represent packet coding coefficients. After channel coding at a rate ρ_c, each packet contains $K_b = \lceil C_b/\rho_c \rceil$ bits. The structure of a coded packet is shown in Fig. 2.

The M information-bearing packets will be decoded successfully at the receiver so long as any M out of the N transmitted packets are received correctly. The probability of successfully decoding the original M packets is thus given by[1]

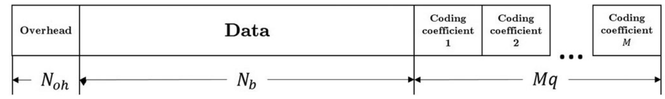


Fig. 2. Structure of a coded packet.

$$P_S = \sum_{m=M}^{N} \binom{N}{m} (1-P_E)^m P_E^{N-m}. \tag{5}$$

The number of coded packets N can now be determined as the smallest value that ensures a desired success rate for a given M and P_E [13]. In other words, if we set the target success rate to some P_S^*, then N is chosen as the smallest value for which $P_S \geq P_S^*$.

Packet error rate P_E depends on the BER P_b and the packet size. Specifically, $P_E = 1-(1-P_b)^{C_b}$. Hence, for a given M, the number of coded packets N depends on P_S^*, P_b, and N_b. This dependence is illustrated in Fig. 3.

The number of coded packets N can also be chosen greater than the minimum indicated by the desired P_S^*. The incentive in doing so is to allow for a greater P_E, which in turn allows for a greater P_b and, thus, a lower transmit power. However, with an increase in N, the energy per successfully transmitted information bit increases. The two trends (lower transmit power and longer packets) yield an optimal spot where the energy per bit is minimized (see [13] for details). While this spot can be used to identify the corresponding N, we keep with the original choice of the smallest N, as this choice will inform us about the maximum achievable throughput efficiency.

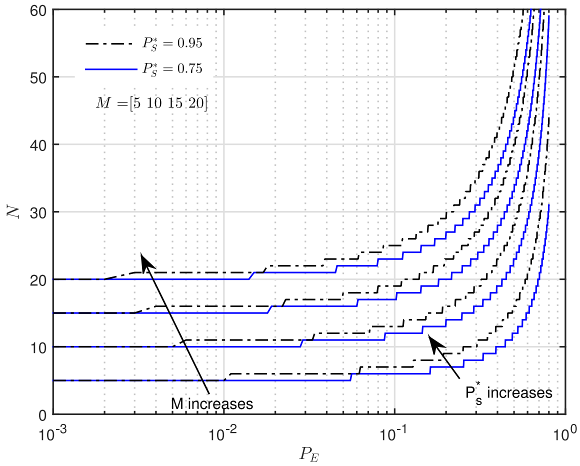
C. Packet Coding and ARQ

Applied in the ordinary S&W fashion, a packet-coded system, which we refer as S&W with packet coding (S&W-PC), will have the throughput efficiency

$$\eta_{\text{S\&W,PC}} = P_S\frac{MN_b}{NK_b+N_{\text{rt}}+N_a}. \tag{6}$$

This expression is analogous to (1), except that the role of a packet is now played by a superpacket, and the role of packet error rate is played by $(1-P_S)$. The latter fact enables us to gain control over situations when P_E is high. Namely, for a given P_E, S&W-PC has an additional degree of freedom over ordinary S&W in that it can choose M and N (or equivalently, M and P_S^*). Note that once the values of M and P_S^* are chosen,

[1]We are assuming that coded packets represent linearly independent combinations of the original information-bearing packets. Strictly speaking, this assumption is not true with random coding; however, the probability of generating a linearly dependent combination decreases with the Galois field size [10]. In addition, we are not considering an infinite-length rateless packet coding, but only a finite number of coded packets N (possibly only slightly greater than M), in which case provisions can be made to ensure linear independence. One way of making such a provision would be to test for linear dependence at the transmitter. If a particular set of coefficients that define a superpacket is found to exhibit linear dependence, a new set can be generated. Strictly speaking, the code would then no longer be random, but that would not affect the system throughput.



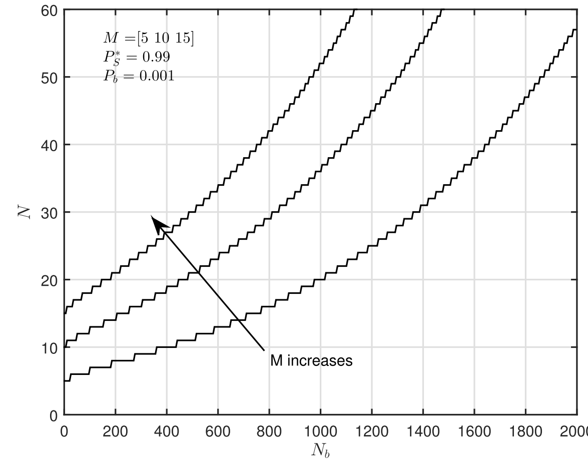


Fig. 3. (Left) Number of coded packets N as a function of the packet error rate P_E. The four solid curves correspond to the target success rate $P_s^* = 0.95$, while the four dashed curves correspond to $P_s^* = 0.75$. Each set of four curves corresponds to four values of M (5, 10, 15, and 20 information-bearing packets). (Right) Number of coded packets N as a function of the number of bits per packet N_b. BER P_b is indicated in the figure along with P_s^* and M.

they imply the value of N and P_S.[2] In other words, M and P_S^* are design parameters that can be adjusted to maximize the performance. As a result, we expect the throughput efficiency (6) to be higher than its counterpart (1).

Given the benefits of packet coding on half-duplex links, the question arises as to whether packet coding also has benefits on full-duplex links. There are several ways in which packet coding can be employed on full-duplex links. One obvious way is to use selective repeat full-duplex with ordinary packets replaced by superpackets. This method, which we refer as FD-PC1, will yield throughput efficiency analogous to (2) as

$$\eta_{\text{FD,PC1}} = P_S \frac{MN_b}{NK_b}. \tag{7}$$

Another possibility is to use packet coding in a rateless fashion, i.e., take a set of M information-bearing packets and keep transmitting their coded combinations until an ACK is received; then move on to the next set. This method, which we refer as FD-PC2, takes on average $\frac{MK_b}{(1-P_E)} + N_{\text{rt}} + N_a$ bits to transmit the set of M packets successfully [10]. We define the corresponding efficiency as follows:

$$\eta_{\text{FD,PC2}} = \frac{MN_b}{MK_b/(1-P_E) + N_{\text{rt}} + N_a}. \tag{8}$$

This efficiency is approached by the time-division duplexing method of [10], which can be used on half-duplex links.

Returning to the issue of half-duplex links and the problem of long delay, we finally investigate the possibility of combining packet coding with grouped S&W. Specifically, we group superpackets into groups of L, and refer to one such group as supergroup. The resulting method, which we call S&W with grouped packet coding (S&W-GPC), has throughput efficiency

$$\eta_{\text{S\&W,GPC}} = P_S \frac{LMN_b}{LNK_b + N_{\text{rt}} + LN_a}. \tag{9}$$

[2] P_S can differ slightly from P_S^*, but that is a detail of no concern at the moment.

This expression is analogous to (4), except that the role of packets and groups is now played by superpackets and supergroups, respectively, while the superpacket error rate is $1 - P_S$. L and P_S^* are the system design parameters that can be adjusted to control the overall throughput efficiency. While $L > 1$ is used to counteract latency, P_S^* is used to counteract packet loss.

In Section III, we will quantify the performance of various ARQ strategies through numerical examples. Before we do so, a few words are in order regarding time-varying channels.

D. Time-Varying Channels

On a fading channel, transmit power is typically determined so as to satisfy an outage criterion, i.e., to ensure that the BER stays below a prespecified level P_b^* with some probability $(1 - P_{\text{out}})$. This requirement can be met by using a fixed fading margin, but such an approach is wasteful as the link remains active when conditions are not favorable. An alternative is to adjust the transmit power so as to ensure that the signal-to-noise ratio (SNR) stays at $\gamma = \gamma^*$, the value corresponding to the target BER P_b^*. Such power control can be exercised so long as the required power is within some maximum value $P_{T,\max}$. When more than $P_{T,\max}$ is needed, transmission is shut off. Hence, when the BER is kept at P_b^*, the system is active (not in outage).

Power control is implemented based on the channel gain G, which is measured at the receiver and fed back to the transmitter. The gain G pertains to the large-scale fading, which is slow enough that it can withstand the feedback delay. Small-scale fading is embodied into the functional dependence between the BER and the SNR $\gamma = P_T G/P_N$, where P_T and P_N denote the transmit power and the noise power, respectively. Specifically, a power control policy is defined as follows [13]:

$$P_T = \begin{cases} P_{T,\max}\frac{G_{\text{out}}}{G} = \gamma^* \frac{P_N}{G}, G \geq G_{\text{out}} \\ 0, \text{ otherwise} \end{cases} \tag{10}$$

where the value of G_{out} is determined from the condition $P_{\text{out}} = P\{P_b > P_b^*\} = P\{\gamma < \gamma^*\} = P\{G < G_{\text{out}}\}$. For example, with log-normally distributed gain, which can be assumed in

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

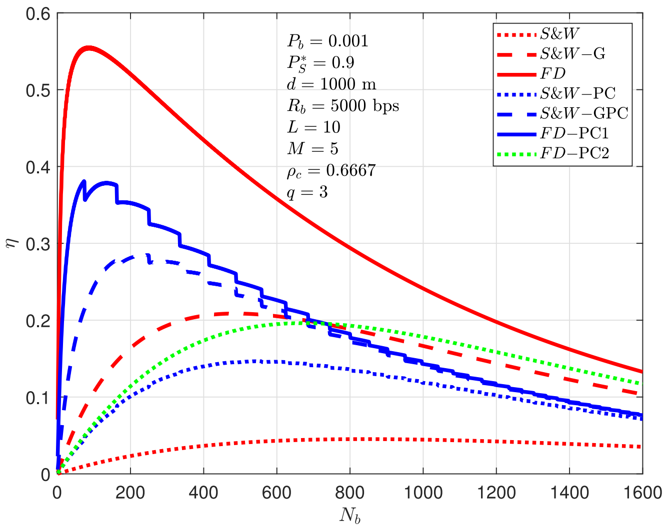


Fig. 4. Throughput efficiency as a function of N_b for various ARQ methods. Specific system parameters are listed in the figure. (Spikes occur because the number of coded packets N changes with N_b.)

certain acoustic channels [14], i.e., $g = 10\log_{10} G \sim \mathcal{N}(\bar{g}, \sigma_g)$, we have that $P_{\text{out}} = Q(\frac{g_{\text{out}}-\bar{g}}{\sigma_g})$ and $g_{\text{out}} = 10\log_{10} G_{\text{out}}$.

In the limiting case of infinite maximum power, there is no outage ($P_{\text{out}} = 0$, $G_{\text{out}} = 0$), and the BER is constantly kept at $P_b = P_b^*$. Although hypothetical, we will consider this case when we illustrate the results numerically. In practice, P_{out} will of course have a finite value, and all the throughput efficiency results will simply be scaled by $(1 - P_{\text{out}})$.

III. Performance Analysis

We begin by looking at the throughput efficiency as a function of the packet size N_b. Fig. 4 illustrates the results for various techniques. These results pertain to the specific system parameters that are listed in the figure. The total round-trip delay is $T_{\text{rt}} = 1.3$ s or 6667 b in this example, in which we assume a transmission distance $d = 1$ km, the speed of propagation $c = 1500$ m/s, and bit rate $R_b = 5$ kb/s. Channel coding rate is $\rho_c = 2/3$ and packet coding is performed in $GF(8)$, i.e., $q = 3$. The number of fixed overhead bits is $N_{\text{oh}} = 8$, and an acknowledgment takes $N_a = 1$ b. These parameters are fixed for all the examples presented.

The throughput efficiencies of systems with and without packet coding are shown in Fig. 4. Without packet coding, we have the ordinary S&W (1), S&W-G (4), and the full-duplex benchmark FD (2). With packet coding, we have S&W-PC (6), S&W-GPC (9), and the two full-duplex benchmarks FD-PC1 (7) and FD-PC2 (8). Clearly, S&W-GPC outperforms all other half-duplex techniques. The fact that it outperforms S&W-G speaks in favor of packet coding, whereas the benefits of grouping are evident from the fact that S&W-GPC outperforms S&W-PC (its $L = 1$ counterpart). Grouping is beneficial for systems without packet coding as well.

An important observation to make is that there exists an optimal packet size for each technique, given a specific M, L, and P_S^*. The explanation of this phenomenon is rather intuitive: When packets are too short, waiting time is not used efficiently; when packets are too long, chances of one or more bits in a packet being erroneous are greater, hence a retransmission is more likely. Deviating significantly from the optimal packet size can have a detrimental effect on throughput efficiency, regardless of the ARQ method used. For example, S&W-GPC achieves its maximum throughput of about 0.3 with 237 b per packet. Choosing 1000 b per packet will instead yield only half of the maximum throughput. The situation is similar with ordinary S&W-G, although its maximum throughput efficiency of about 0.2 is achieved with 484 b per packet. We, thus, see that GPC offers about 50% improvement in maximum throughput efficiency in this example.

Comparing the full-duplex benchmarks, we note that the ordinary full-duplex has the best performance. This should not come as a surprise: When full-duplex is available, there is no point in coding the packets ahead because redundant packets waste transmission time, and when the round-trip time is long, there will be many redundant packets filling it. It is also in the best interest of throughput efficiency that the packets be kept short, thus ensuring that retransmissions are minimized. The FD performance shown in Fig. 4 is that of an ideal full-duplex with unlimited storage, and should in that sense be regarded as an unattainable bound used for benchmarking purposes only. Rateless transmission over a full-duplex link (FD-PC2), which can be approached by time-division duplexing [10] on half-duplex links, does not have the buffering issues, but its performance is markedly worse. In fact, S&W-GPC outperforms FD-PC2. This is explained by the fact that when the round-trip delay is long, FD-PC2 will keep transmitting (long) after the M packets have been successfully decoded, and will learn of successful decoding only once the ACK arrives. In contrast, GPC transmits a new supergroup without waiting for the ACK on the current one, thus filling the round-trip time with new information. Its efficiency thus becomes somewhat remarkable, amounting to more than half the maximum efficiency of ideal FD. Also note that this result includes full overhead of random linear packet coding. If that overhead is removed, throughput efficiency increases further, as illustrated in Fig. 5.

While grouping clearly helps to counteract latency, the cycle between successive ACK packets grows with the superpacket size L. Specifically, the duration of a cycle is $T_{\text{cyc}} = T_{\text{sg}} + T_{\text{rt}}$, where $T_{\text{sg}} = LNK_b/R_b$ is the duration of a supergroup. When the ACK feedback link is also used to convey the channel state information, one must make sure that this feedback is frequent enough, i.e., that the channel state does not change much during one cycle. Fig. 6 shows the duration of a supergroup as a function of the packet size N_b. As a sanity check, we verify that the cycle duration corresponding to the optimal packet size is well within the coherence time of the large-scale fading.

Let us now focus on the performance of S&W-GPC, and ask how can this performance, shown in Fig. 4 for an ad hoc selection of the key parameters (M, L, P_S^*), be maximized. To answer this question, Fig. 7 zooms in on the throughput efficiency of GPC, showing it for different values of M, the number of information-bearing packets in a superpacket. As before, the efficiency is shown versus the number of bits per packet. Clearly, the performance of S&W-GPC changes with M. Initially, as M grows, the throughput efficiency improves,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

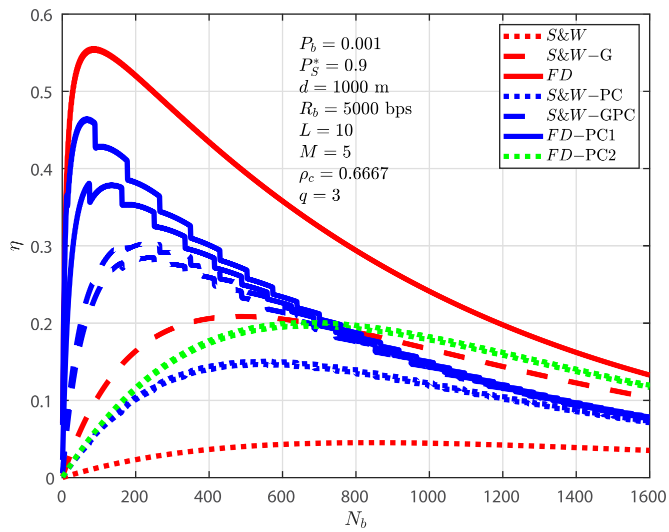


Fig. 5. Same as Fig. 4, but with an additional set of curves for packet-coded systems. The additional curves represent throughput efficiency without the Mq bits of packet coding overhead. These curves are labeled the same as their original counterpart, but lie above them. Packet coding overhead can be removed if the transmitter and receiver agree up front on the coding coefficients.

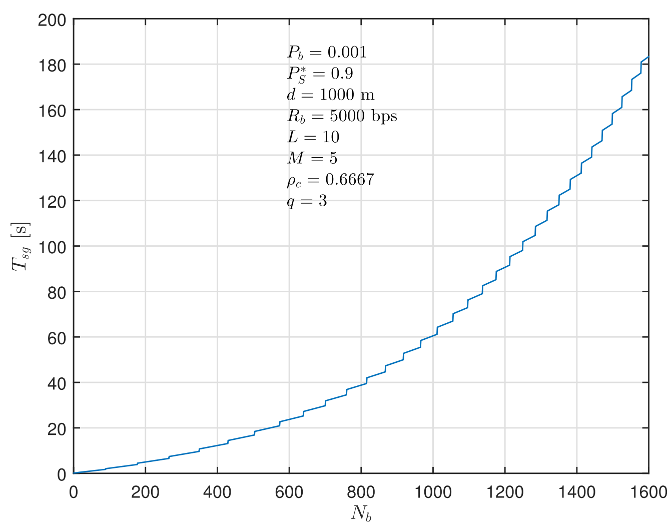


Fig. 6. Supergroup duration, $T_{\text{sg}} = LNK_b/R_b$, as a function of N_b. At the optimally chosen packet size (relatively small N_b), the supergroup duration is 6 s in this example. The corresponding cycle duration is $T_{\text{cyc}} = T_{\text{sg}} + T_{\text{rt}} = 7.3$ s.

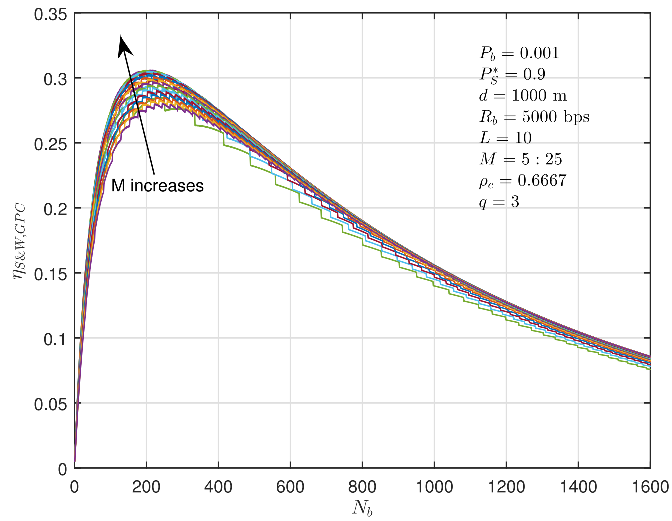


Fig. 7. Throughput efficiency of S&W-GPC changes with the group size M.

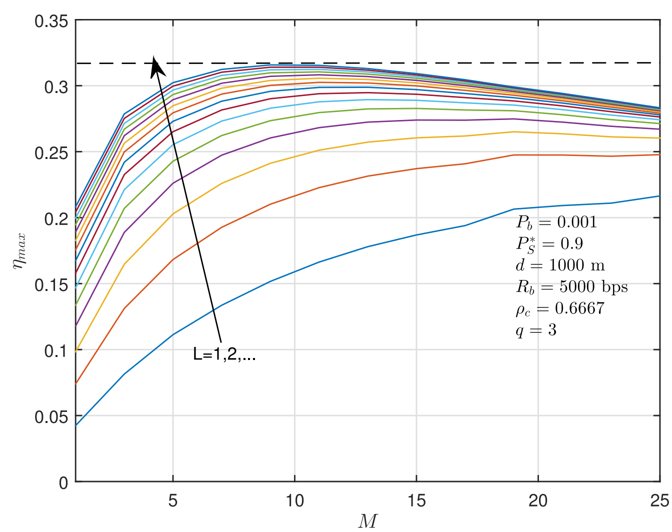


Fig. 8. Maximum throughput efficiency of S&W-GPC. For every point shown, the packet size N_b is chosen as that for which the efficiency is maximized (see Fig. 7).

but apparently it reaches a limit. The question thus becomes: Is there a best value of M?

To address the question of optimal selection of M, Fig. 8 shows the maximum throughput efficiency of S&W-GPC as a function of M. The result is shown for various values of the supergroup size L. Clearly, maximum throughput efficiency increases with L. For every M and L, the packet size is chosen as that which corresponds to the maximum throughput efficiency; in other words, the points along each curve may correspond to different values of N_b. The existence of an optimal point (best value of M) is evident.

An equivalent result can be shown in the more usual framework of throughput efficiency versus packet size, i.e., in the style of Fig. 4. This is done in Fig. 9, which shows the throughput efficiency of S&W-GPC versus N_b, where M is now chosen optimally for each point on the curves. More precisely, this figure shows the maximum throughput efficiency as a function of N_b for various values of L. Looking across different supergroup sizes, it is again obvious that the maximum throughput efficiency increases with L. The peak efficiency for a given L is the maximum throughput efficiency corresponding to an optimally chosen packet size. This maximum is shown as a function of L in the figure on the right. (One may be tempted to label it $\eta_{\max,\max}$, but we refrain from this at the risk of some abuse of notation.) If one were to impose a limit on L, be it because of limited storage or channel coherence considerations for power control, that limit would represent the optimal choice for L. In this example, we have set the limit to $L = 15$, and indicated the corresponding efficiency by the flat dashed line in Fig. 9 (left). The line corresponds to the rightmost point on the graph in Fig. 9 (right). It is very interesting to note that the performance limit,

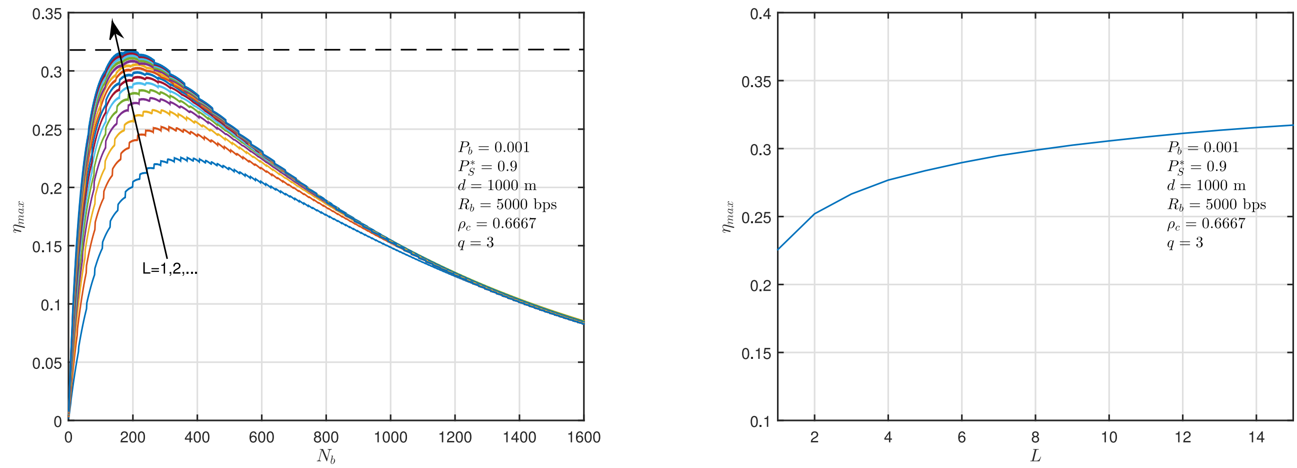


Fig. 9. (Left) Maximum throughput efficiency of half-duplex GPC. For every N_b, the group size M is chosen as that for which the efficiency is maximized. (Right) Maximum achievable throughput efficiency as a function of the group size L. Each point of this graph corresponds to a maximum from the graph on the left.

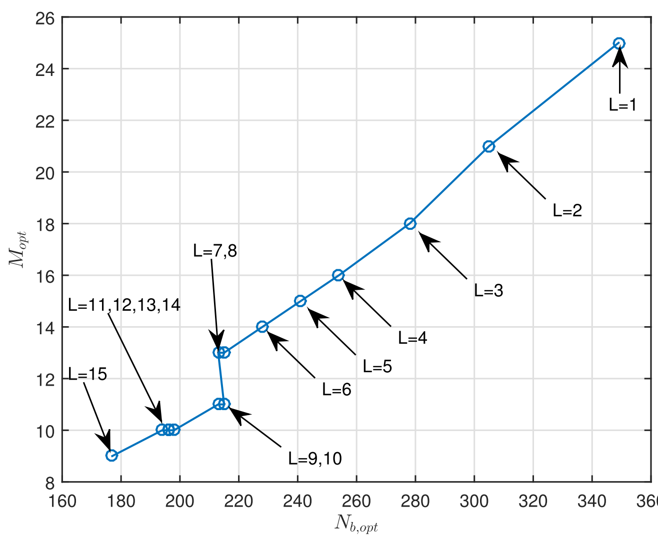


Fig. 10. Pairs of optimum (N_b, M) that yield maximum throughput efficiency for various values of the group size L. System parameters are as indicated in Fig. 9.

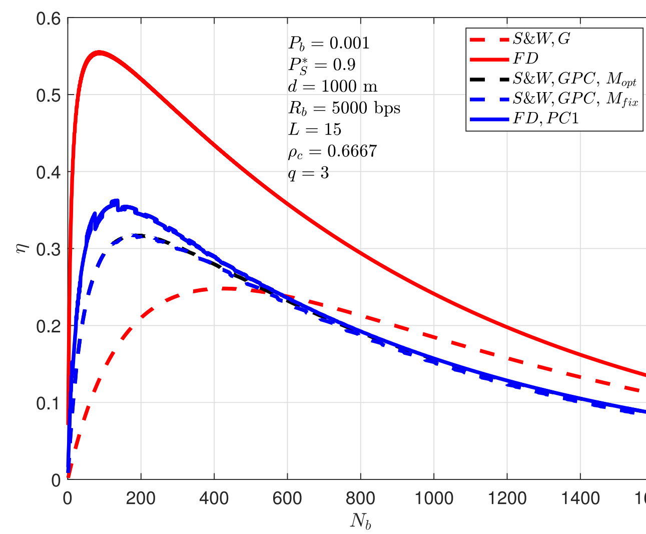


Fig. 11. Throughput efficiency of S&W-GPC with optimally chosen M is shown in black dashed line; that of S&W-GPC with fixed M is shown in blue dashed line. The fixed value of M is 9, which is optimal for $L = 15$ (see Fig. 10). FD-PC1 operates with the same M_{opt} as S&W-GPC. The black curve is less choppy than the blue curve indicating that the effect of choosing M optimally for every N_b is to straighten the occasional small dips in efficiency.

$\lim_{L\to\infty} \max_{N_b,M} \{\eta_{\text{SW,GPC}}\}$, is reached quickly as L increases. This is a remarkable feature, as it implies that there is no need to choose overly large values of L, which in turn implies that practical buffer sizes may be well-within reach of the optimal performance.

Fig. 10 summarizes the key design points. In this figure, optimal values of N_b and M are shown for various choices of L. As L increases, both N_b and M decrease. At the same time, the efficiency increases. The optimal design can thus be summarized as follows: Choose the supergroup size L as large as the system constraints allow, then use the result of Fig. 10 to find the best pair (N_b, M) for the so-chosen L. Keep in mind that this figure corresponds to the specific values of the round-trip delay, BER, and the target success rate P_S^*. While the round-trip delay is dictated by the physics of propagation, the BER and P_S^* can be controlled. We will address their impact on efficiency shortly, after a brief final comment on optimization.

The performance of an optimally designed S&W-GPC is contrasted to that of S&W-G (no packet coding) in Fig. 11. The supergroup size is $L = 15$, and M is chosen optimally for each value of N_b. The full-duplex benchmark is also included, as is the FD-PC1 performance obtained with the same parameters as the optimized S&W-GPC. Also shown in the figure are the performance of S&W-GPC with a fixed value of M, chosen as $M = 9$, which corresponds to the maximum point (optimal N_b) of the optimized S&W-GPC. We note that the optimal choice of M yields a smoother efficiency curve without the occasional small dips in performance that accompany the fixed choice of M. However, the difference is very small. These effects are an artifact of dealing with integers, and should be of no concern.

So far, we have been looking at the system performance for a given BER and an ad hoc selection of the target success rate

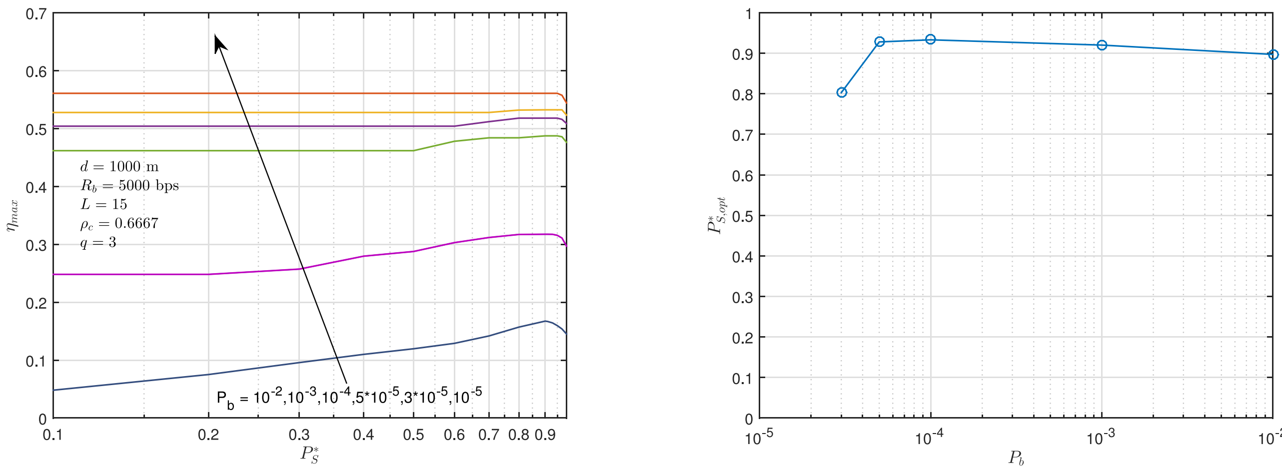


Fig. 12. (Left) Maximum throughput efficiency as a function of the target success rate P_s^* for various values of the BER. Note that maximum efficiency depends heavily on P_b, and less so on P_s^*. For $P_b = 10^{-5}$, there is essentially no dependence on P_s^* indicating that packet coding is not necessary on very good channels (very low BER). (Right) Value of P_s^* at which the maximum throughput efficiency peaks.

P_S^*. The question we now want to address is how does the performance change with the BER, and whether there exists a good choice for the target success rate P_S^*. Fig. 12 offers the answer. This figure shows the maximum throughput efficiency as a function of P_S^* for various values of the BER. As one might expect, higher throughput efficiency is achievable on better links with lower BER. In contrast, performance remains rather insensitive to the exact choice of P_S^*, so long as it is not excessively low or extremely close to 1, but within some meaningful range, say between 0.75 and 0.95. While this observation may be somewhat surprising, an explanation is found in Fig. 3, in which we see that two rather different values of P_S^* yield similar (M, N) pairs for all except the very high packet error rates P_E. On links with poor BER performance (say $P_b = 10^{-3}$), P_E is controlled by choosing short packets. Consequently, while one does not expect P_E to be so low that packet coding is not needed ($N = M$), one does not expect it to be so high as to require excessive packet coding ($N \gg M$) either. In the gray zone between these two extremes (the far left and the far right of the graph in Fig. 3), we note that $P_S^* = 0.95$ requires only a few more coded packets than $P_S^* = 0.75$. What little difference there is, averages out in the throughput efficiency (which increases with P_S^*, but decreases with N).

Fig. 13 provides a closer look at the maximum throughput efficiency in the range of meaningful P_S^* values. We note that there is very little sensitivity to the choice of P_S^* in this range. This is good news as it simplifies our design and, more importantly, indicates that the maximum throughput efficiency is robust to the choice of P_S^*. However, note that performance metrics other than the throughput efficiency may be sensitive to the choice of P_S^*. This is particularly true for the cycle duration, which increases linearly with the number of coded packets N. Fig. 14 provides a closer look into this effect.

Fig. 14 shows how various system parameters change with the BER. Included in this figure, underneath each other for easy reference, are the maximum throughput efficiency, optimal packet size N_b, optimal number of information-bearing packets M in a superpacket, corresponding packet coding ratio N/M, and cycle duration $T_{\rm cyc}$. The set of plots on the left corresponds to the

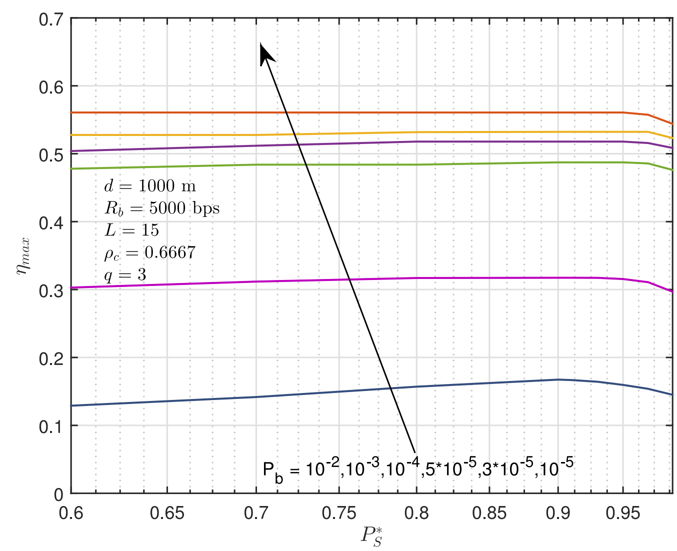


Fig. 13. Zoomed portion of Fig. 12. Maximum throughput efficiency remains fairly constant in the range of P_S^* values shown indicating that small deviations from $P^*_{S,\rm opt}$ do not cause any significant degradation in performance, i.e., that the system design is robust with respect to choosing the target success rate P_S^*.

optimal choice of the target success rate P_S^*. The set of plots on the right corresponds to $P_S^* = 0.7$ for all values of the BER. We note that while there is little change in the maximum throughput efficiency between these two designs (optimal P_S^* and fixed P_S^*), there is a considerable difference in the corresponding cycle times. Hence, if shorter cycle times are of interest because they yield shorter decoding delay and more frequent feedback of the channel state (which is in turn used to keep the BER at a given value), setting P_S^* below the optimum is preferable.

A. *Additional Results: Simulation*

In this section, we report on the results of numerical simulation of S&W-GPC. The goal of simulation was twofold: First, to assess the impact of simplifying assumptions (neglecting the possibility of ACK packets being lost and the possibility of

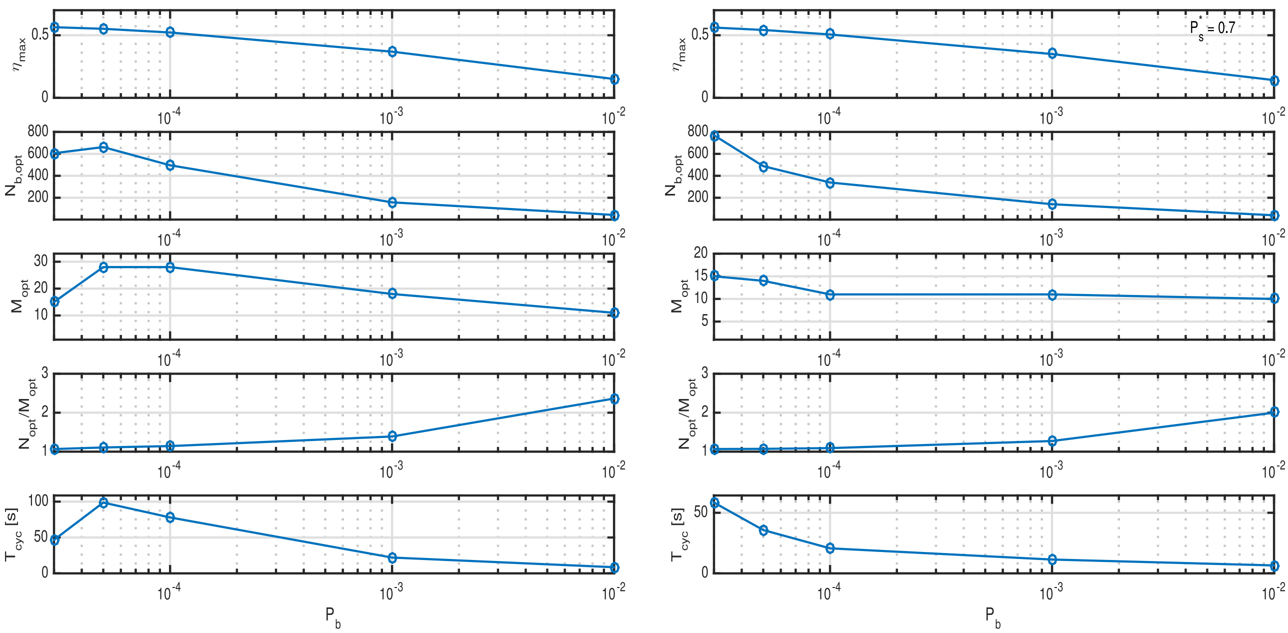


Fig. 14. Salient system parameters. (Left) $P_S^* = P_{S,\text{opt}}^*$. (Right) $P_S^* = 0.7$. While throughput efficiency remains practically unchanged with P_S^*, note the difference in cycle time.

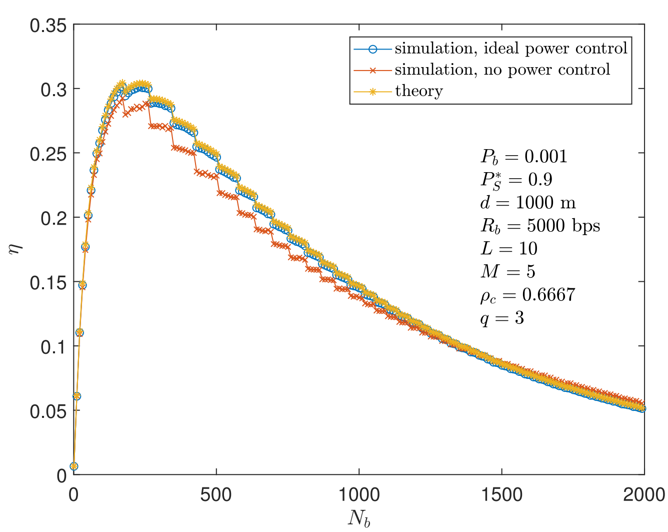


Fig. 15. Throughput efficiency of S&W-GPC: Results of numerical simulation using experimentally validated statistical channel model [14]. Large-scale channel gain is log-normally distributed with standard deviation of 5 dB, and 3 dB Doppler bandwidth of 0.05 Hz.

decoding failure due to linear dependence of packets in a superpacket), and second, to assess the impact of feedback delay, i.e., the assumption of perfect channel state feedback. Our simulation is based on the experimentally validated statistical channel models [14], which present the large-scale channel gain as a first-order autoregressive Gauss–Markov process on the logarithmic scale.

Fig. 15 illustrates the simulation results, along with the theoretical throughput efficiency (9). Ideal power control corresponds to adjusting the transmit power such that the BER is kept constant at the design value P_b^*. Simulated performance of ideal power control includes the possibility of ACK packets being lost, and shows very little difference from the theoretical result. The more interesting question of feedback delay is addressed by the simulated performance of a system that uses no power control (throughput efficiency curve labeled by x-marks). No power control can be thought of as the worst-case scenario to which one would resort because of excessive feedback delay, unreliable channel state estimation, or hardware limitations. For a fair comparison, we set the transmit power to a constant level equal to the average power consumed under ideal power control policy. The resulting large-scale SNR is now time-varying, $\gamma = \gamma^* G E\{\frac{1}{G}\}$, and so is the BER. Theoretical analysis of throughput efficiency becomes complicated in this case, but simulation offers the much needed answer. In Fig. 15, simulated throughput efficiency of an S&W-GPC system operating with no power control over a fading channel is labeled by o-marks. As one might expect, this throughput efficiency is lower than the ideal. However, and perhaps somewhat surprisingly, the loss in not excessive at all, amounting to about 1% reduction at the optimal point. This is very good news from the standpoint of a practical system design, where implementation of power control may be fraught with difficulties.

IV. Multicast

Random linear packet coding is known to show full benefits in multicast scenarios, where coding ahead in a one-size-fits-all fashion caters to all the receivers at once. This is true for full-duplex rateless coding over links with negligible delay, where transmission stops as soon as all the receivers have successfully decoded the information-bearing packets. The multicast situation on half-duplex links with long delay is different.

On half-duplex links, we distinguish two boundary cases, one in which the link quality (BER) varies drastically between the receivers, and another in which all the links are of the same quality. In the first case, system performance will be dominated by the worst link, i.e., all the nodes will experience (more or less) the throughput efficiency of the link with the highest BER. This is explained simply by noting that if there is one receiver whose BER is disproportionately higher than the BER of other receivers, that receiver will keep requesting retransmissions, and the other receivers will have to wait until the requesting receiver is satisfied. The case in which all the links have the same BER is the most interesting one. What is the throughput efficiency in this case? Clearly, a retransmission will be more likely when there are more receivers: Although a particular receiver may have received all the information it needs, it will keep receiving that same information until the last of the remaining receivers has received all it needs.

To arrive at the throughput efficiency of a multicast system, let us assume that there are R receivers, and let us denote by T_R the expected time it takes to convey a superpacket to all the R receivers. In addition, let T denote the time needed for one transmission round (superpacket, round-trip, and acknowledgment). Without any grouping, we have that $T = NK_b/R_b + T_{\text{rt}} + N_a/R_b$. We already know that T_1, i.e., the value corresponding to the throughput efficiency (6), is

$$T_1 = \frac{T}{P_S}. \tag{11}$$

Consider now a simple case with three receivers for example. Upon completion of the first round, which takes the time T, the following possibilities exist: all three receivers have received the information; two out of three have received the information; only one has received it, or none has received it. The expected remaining time is $T_0 = 0, T_1, T_2$, and T_3, respectively. Hence, we can write

$$T_r = T + \sum_{i=0}^{r} P_{r,i} T_i \tag{12}$$

where $P_{r,i}$ is the probability that i out of r receivers are requesting retransmission. Given that all the receivers have the same BER, and hence the same P_S, this probability is simply

$$P_{r,i} = \binom{r}{i} (1 - P_S)^i P_S^{r-i}. \tag{13}$$

From (12), we have that

$$T_r = \frac{1}{1 - P_{r,r}} \left(T + \sum_{i=0}^{r-1} P_{r,i} T_i \right). \tag{14}$$

To obtain T_R, the above-mentioned expression is evaluated recursively for $r = 1,2,\ldots,R$. As a check, note that T_1 reduces to (11).

The above-mentioned result corresponds to transmission without grouping. Grouping of size L effectively yields L nongrouped ARQ links operating in parallel, with each link having the one-round time $T = LNK_b/R_b + T_{\text{rt}} + LN_a/R_b$. Thus, the throughput efficiency of multicast S&W-GPC with R equal-BER receivers is

$$\eta_{\text{S\&W,GPC}}(R) = \frac{LMN_b/R_b}{T_R} \tag{15}$$

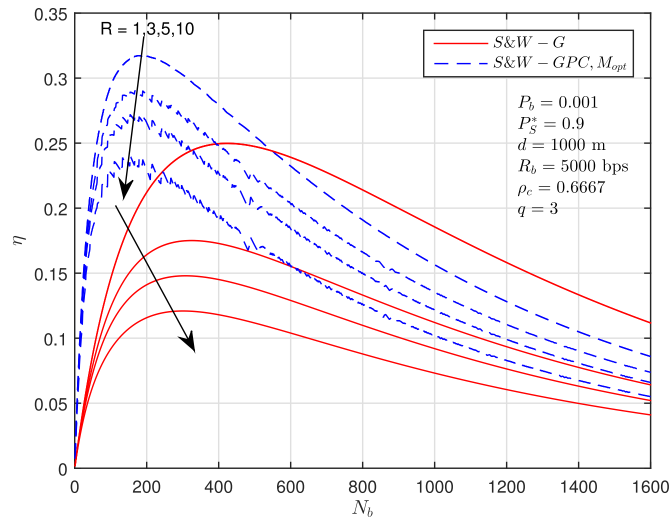


Fig. 16. Throughput efficiency of S&W-GPC and ordinary S&W-G in multicast scenarios with $R = 1, 3, 5$, and 10 receivers. (Spikiness is an artifact of N being the smallest integer that yields $P_S \geq P_S^*$.)

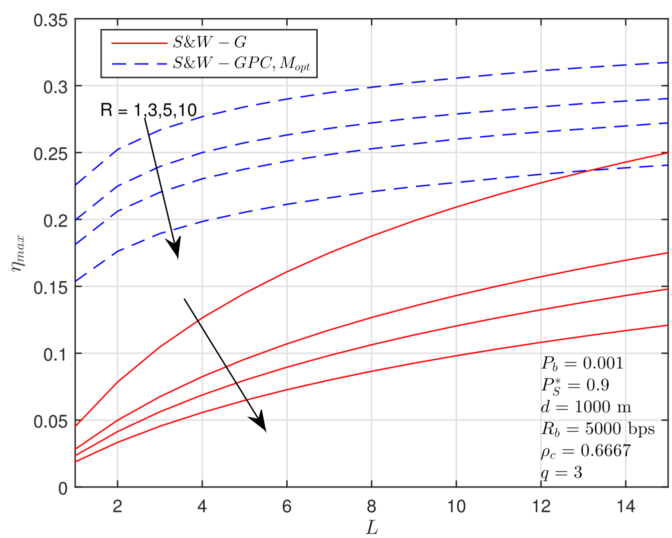


Fig. 17. Maximum throughput efficiency for varying number of multicast receivers R.

where T_R is defined by (14) and $TR_b = LNK_b + N_{\text{rt}} + LN_a$.

Throughput efficiency without packet coding, $\eta_{\text{S\&W,G}}(R)$, is the special case of the above-mentioned result, obtained for $M = N = 1$, $P_S = 1 - P_E$, and an appropriate packet size.[3]

Figs. 16 and 17 illustrate the results, quantifying the decrease in throughput efficiency that occurs with an increasing number of receivers in a multicast system. The advantage of GPC is evident: not only does the S&W-GPC outperform the ordinary S&W-G for a given number of receivers, but also the throughput efficiency it offers in a multicast configuration can surpass that of the ordinary S&W-G in a unicast configuration.

[3]Recall that $K_b = \lceil C_b/\rho_c \rceil$, with $C_b = N_b + qM + N_{\text{oh}}$ when packet coding is used, and $C_b = N_b + N_{\text{oh}}$ when packet coding is not used. The value of C_b also determines the relevant P_E.

V. Conclusion

We investigated an ARQ method based on an S&W procedure with GPC, and found that it offers a very good practical solution for reliable communication over half-duplex links where long delay and poor physical link quality challenge the performance of conventional ARQ methods. S&W-GPC is based on two levels of hierarchy: On the first level, a superpacket is formed by applying random linear packet coding to a group of information-bearing packets; on the second level, several superpackets are combined to form a supergroup. The strength of S&W-GPC lies in its ability to control the probability of superpacket retransmission (which is done through packet coding, by choosing the number of coded packets so as to yield a desired success rate), and the supergroup size (transmitting a group of superpackets and waiting for a single selective acknowledgment is more effective than waiting for an acknowledgment after each individual superpacket transmission). In this manner, it outperforms the conventional S&W procedure that uses packet grouping, but no packet coding.

When the round-trip delay is long compared to the superpacket duration, GPC outperforms even the full-duplex rateless packet coding, whose performance can be approached on half-duplex links by the time-division duplexing method [10]. GPC gains this advantage by avoiding unnecessary redundant transmissions that occur with full-duplex rateless coding between the time of successful decoding at the receiver and the time the transmitter becomes aware of success (arrival of the acknowledgment). The corresponding power expenditure will clearly be in favor of GPC as well.

The advantages of S&W-GPC are also retained in the multicast scenario, where it consistently outperforms the ordinary S&W-G. Finally, GPC is simple to implement: It can operate on top of any physical layer, it does not replace conventional channel coding but works together with it, and it does not impose any synchronization requirements beyond the ones that normally exist in a packetized digital system. The parameters that control its throughput efficiency are few—in addition to the packet size (which needs to be optimized in conventional ARQ as well), GPC requires specification of the supergroup size (number of information-bearing packets per supergroup) and the target success rate (or equivalently, number of coded packets per supergroup). This paper offers guidelines for choosing the system parameters in an optimal manner, as well as a detailed analysis of their impact on throughput efficiency. Future development would benefit from implementing S&W-GPC in a readily available acoustic modem and evaluating its performance in the field.

References

[1] J. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater sensor networks: Applications, advances and challenges," *Philos. Trans. Roy. Soc.,* vol. 370, no. 1958, pp. 158–175, Nov. 2011.

[2] S. Sendra, J. Lloret, J. M. Jimenez, and L. Parra, "Underwater acoustic modems," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4063–4071, Jun. 2016.

[3] M. Stojanovic, "Optimization of a data link protocol for an underwater acoustic channel," in *Proc. Eur. Oceans*, Jun. 2005, vol. 1, pp. 68–73.

[4] M. Chitre and W. S. Soh, "Reliable point-to-point underwater acoustic data transfer: To juggle or not to juggle?" *IEEE J. Ocean. Eng.*, vol. 40, no. 1, pp. 93–103, Jan. 2015.

[5] J. Yu, H. Chen, L. Xie, and J. H. Cui, "Performance analysis of hybrid ARQ schemes in underwater acoustic networks," in *Proc. Oceans—St. John's*, Sep. 2014, doi: 10.1109/OCEANS.2014.7003000.

[6] S. Azad, P. Casari, and M. Zorzi, "The underwater selective repeat error control protocol for multiuser acoustic networks: Design and parameter optimization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4866–4877, Oct. 2013.

[7] M. Gao, W. S. Soh, and M. Tao, "A transmission scheme for continuous ARQ protocols over underwater acoustic channels," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009, doi: 10.1109/ICC.2009.5198684.

[8] D. J. C. MacKay, "Fountain codes," *Proc. Inst. Elect. Eng. Commun.*, vol. 152, no. 6, pp. 1062–1068, Dec. 2005.

[9] M. Chitre and M. Motani, "On the use of rate-less codes in underwater acoustic file transfers," in *Proc. OCEANS 2007—Europe*, Jun. 2007, doi: 10.1109/OCEANSE.2007.4302275.

[10] D. Lucani, M. Medard, and M. Stojanovic, "On coding for delay—Network coding for time-division duplexing," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2330–2348, Apr. 2012.

[11] P. Casari, M. Rossi, and M. Zorzi, "Towards optimal broadcasting policies for HARQ based on fountain codes in underwater networks," in *Proc. 5th Annu. Conf. Wireless Demand Netw. Syst. Serv.*, Jan. 2008, pp. 11–19.

[12] Z. Zhou, H. Mo, Y. Zhu, Z. Peng, J. Huang, and J. H. Cui, "Fountain code based adaptive multi-hop reliable data transfer for underwater acoustic networks," in *Proc. IEEE Int. Conf.Commun.*, Jun. 2012, pp. 6396–6400.

[13] R. Ahmed and M. Stojanovic, "Joint power and rate control for packet coding over fading channels," *IEEE J. Ocean. Eng.*, vol. 42, no. 3, pp. 697–710, Jul. 2017.

[14] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE J. Ocean. Eng.*, vol. 38, no. 4, pp. 701–717, Oct. 2013.

[15] S. Lin, D. Costello, and M. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Mag.*, vol. 22, no. 12, pp. 5–17, Dec. 1984.

[16] M. Miller and S. Lin, "The analysis of some selective-repeat ARQ schemes with finite receiver buffer," *IEEE Trans. Commun.*, vol. 29, no. 9, pp. 1307–1315, Sep. 1981.

Rameez Ahmed received the Bachelors degree in telecommunication from the Vellore Institute of Technology University, Vellore, India, in 2008, and the M.S. and Ph.D. degrees in electrical engineering from Northeastern University, Boston, MA, USA, in 2011 and 2016, respectively.

He is currently a Senior Development Engineer with The MathWorks, Inc., Natick, MA, USA. His research interests include digital communications, statistical signal processing, wireless networks, network coding, and their applications to satellite systems and underwater acoustic systems.

Milica Stojanovic (SM'08–F'10) received the Dipl. Ing. degree from the University of Belgrade, Belgrade, Serbia, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from Northeastern University, Boston, MA, USA, in 1991 and 1993, respectively.

She was the Principal Scientist with the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. In 2008, she joined Northeastern University, where she is currently a Professor of electrical and computer engineering. She is also a Guest Investigator with the Woods Hole Oceanographic Institution and a Visiting Scientist with MIT. Her research interests include digital communications theory, statistical signal processing, and wireless networks and their applications to underwater acoustic systems.

Dr. Stojanovic is the recipient of the 2015 IEEE/OES Distinguished Technical Achievement Award. She is an Associate Editor for the IEEE Journal of Oceanic Engineering and a Past Associate Editor for the IEEE Transactions on Signal Processing and the IEEE Transactions on Vehicular Technology. She also serves on the Advisory Board of the IEEE Communication Letters, and Chairs the IEEE Ocean Engineering Society's Technical Committee for underwater communication, navigation, and positioning.