A New Outage Probability Bound for IR-HARQ and Its Application to Power Adaptation

Wenyu Wang and Kaiming Shen
School of Science and Engineering
The Chinese University of Hong Kong, Shenzhen, China
E-mail: wenyuwang@link.cuhk.edu.cn, shenkaiming@cuhk.edu.cn

***Abstract*—This work proposes a novel upper bound on outage probability for the incremental redundancy hybrid automatic repeat request (IR-HARQ) protocol over block fading channels. The new bound is much tighter than the classical bound due to Laneman, especially when the transmit power is limited as is often the case in the Internet of Things (IoT). We demonstrate the practical use of this new bound by showing that it enables a geometric programming approach to power adaptation for IR-HARQ. Because the new bound approximates the outage probability much more exactly, the resulting new power adaptation scheme significantly outperforms those existing methods based on the classical bound.**

I. Introduction

There has been a revival of research interests in hybrid automatic repeat request (HARQ) due to the key role it plays in enabling ultrareliable communications for the Internet of Things (IoT). It is an urgent need to improve the conventional HARQ protocol to account for the low-energy requirement for the IoT devices. Aiming to enhance the energy efficiency of ultrareliable communications, this paper proposes a power optimization strategy for incremental redundancy HARQ, i.e., IR-HARQ, by means of a novel outage probability bounding technique that is much tighter than the existing ones.

More specifically, we seek the optimal power allocation across the transmission rounds of IR-HARQ over single-input multiple-output (SIMO) channels given the channel distribution, in order to minimize the expected energy consumption under the target rate and the target outage probability. This power adaptation problem stems from the energy-reliability tradeoff: we would reduce transmit power for the energy saving purpose, but on the other hand higher transmit power gives higher reliability of transmission (i.e., lower outage probability). It is numerically difficult to address the power adaptation problem directly due to the complicated form of the outage probability—which consists of successive convolutions with the power variables.

Power adaptation for HARQ, either of the incremental redundancy type or of the Chase combining type, is extensively studied in the literature. Regardless of the specific problem setting, a common central issue is how to cope with the outage probability function. An early work [1] considers the relay channel. Its main idea is to discretize the outage probability function by assuming discrete power values so that dynamic programming (DP) starts to apply. For the relay channel with continuous power, [2] suggests using a classical bound on outage probability from [3] to approximate the original problem as a geometric programming (GP) problem. The outage probability bound in [3] also forms the building block of many other works in this area. For example, [4] optimizes power across the different transmission rounds of IR-HARQ by the classical bound, showing that its power strategy outperforms the conventional equal power allocation significantly. This performance gain is also verified in the power-limited regime in [5]. A further step is taken in [6], [7] to find the closed-form solution of the GP problem. Moreover, [8] extends the classical bound in [3] to the multi-antenna case, while [9] extends to the multi-bit feedback case. Differing from the above works, [10] approximates the outage probability by Chernoff bound. However, this method works only when all the transmission rounds of HARQ use equal power level.

Roughly speaking, the bound in [3] is obtained by letting the power variable be infinitely large in part of the outage probability function. As a consequence, the classical bound would significantly overestimate outage probability when the transmit power is limited in practice, so that its solution can incur much higher energy consumption than is necessary to attain the target outage probability. To remedy this error, the recent work [11] proposes incorporating the practical power constraint into the outage probability bound. We explore this direction further. The paper shows that the proposed outage probability bound encompasses the existing bounds in [3], [11] as two special cases, and is strictly tighter than them. In aid of the new bound, we convert the power adaptation problem to a GP formulation. In contrast to the existing GP approach based on the classical bound in [3], the new method based on the improved bound not only yields remarkably lower energy consumption, it can also accommodate much stricter requirements for ultrareliable communications.

II. Channel Model

Suppose the transmitter has a single antenna and the receiver has $M \geq 1$ antennas. The SIMO channel $\boldsymbol{h}_n \in \mathbb{C}^M$ has bandwidth W and varies from block to block in an i.i.d. fashion, where $n = 1, \ldots, N$ is the block index, namely *block fading channel*. We model each $\boldsymbol{h}_n$ as

$$\boldsymbol{h}_n = \sqrt{\beta}\boldsymbol{g}_n, \tag{1}$$

where the pathloss $\beta > 0$ is fixed and the Rayleigh fading $\boldsymbol{g}_n \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M)$ is an i.i.d. Gaussian random vector. Let p_n

be the transmit power in block n and let σ^2 be the background noise power. By coherent combining at the receiver side, we can achieve the following data rate r_n in block n:

$$r_n = W \log_2 \left(1 + \frac{\|\boldsymbol{h}_n\|^2 p_n}{\sigma^2}\right). \tag{2}$$

If the block duration equals L, then at most $r_n L$ message bits can be conveyed in the nth block.

The transmitter aims to reliably send a t-bit message to the receiver by IR-HARQ. In principle, IR-HARQ allows the mutual information to be accumulated over time, i.e., the total number of decodable bits after n blocks is given by $\sum_{i=1}^n r_i L$. An outage takes place in block n if the accumulated mutual information $\sum_{i=1}^n r_i L$ is below the target message size t, so the outage probability of block n can be computed as

$$Q_n = \Pr\left[\sum_{i=1}^n r_i L < t\right]. \tag{3}$$

The retransmission process of IR-HARQ works as follows. After each round of transmission, the receiver sends back `NACK`/`ACK` signal indicating whether it has successfully received the message. If the feedback is `NACK`, then the transmitter continues to the next block; otherwise, the retransmission process finishes. In particular, the process must terminate after the final block N, so the ultimate outage probability is determined by Q_N. Consequently, the expected value of the total energy consumption across the N blocks is given by

$$E = p_1 L + \sum_{n=2}^N p_n L Q_{n-1}, \tag{4}$$

where the first term does not contain the outage probability because the initial block is always used. Likewise, the expected overall latency is given by

$$D = L + \sum_{n=2}^N L Q_{n-1}. \tag{5}$$

Moreover, we use ϵ, δ, and P to denote the constraints on the ultimate outage probability, the overall latency, and the transmit power, respectively. We seek the optimal power choice $(p_1, \ldots, p_N)$ that minimizes the expected energy consumption under the above constraints, i.e.,

$$\begin{aligned}
&\underset{(p_1,\ldots,p_N)}{\text{minimize}} \quad E && \text{(6a)}\\
&\text{subject to} \quad Q_N \le \epsilon, && \text{(6b)}\\
&\qquad\qquad\quad D \le \delta, && \text{(6c)}\\
&\qquad\qquad\quad 0 \le p_n \le P. && \text{(6d)}
\end{aligned}$$

The full channel information $\{h_n\}$ is unknown to the transmitter. We assume that only the channel distribution is available.

III. Power Adaptation by Bounding Q_n

A. Actual Outage Probability

Recognize $2\|\boldsymbol{g}_n\|^2$ as a chi-squared random variable with M degrees of freedom, and thus the probability density function (PDF) of $\|\boldsymbol{h}_n\|^2 = \beta/2 \cdot 2\|\boldsymbol{g}_n\|^2$ can be obtained as

$$f_{\|\boldsymbol{h}_n\|^2}(x) = \frac{x^{M-1} e^{-x/\beta}}{(M-1)!\,\beta^M}. \tag{7}$$

Extending the above PDF to (2) gives the PDF of r_n:

$$f_{r_n}(x) = \left(\frac{\sigma^2}{\beta}\right)^M \frac{2^{x/W}(2^{x/W}-1)^{M-1}\ln 2}{(M-1)!\,W p_n^M} \cdot J(p_n), \tag{8}$$

where

$$J(p_n) = \exp\left(-\frac{\sigma^2(2^{x/W}-1)}{\beta p_n}\right). \tag{9}$$

We now rewrite the outage probability Q_n in (3) in terms of the PDF $f_{r_n}(x)$, i.e.,

$$\begin{aligned}
Q_n &= \Pr\left[\sum_{i=1}^n r_i < t/L\right] && \text{(10a)}\\
&= \int_0^{t/L} \big(f_{r_1}(x) * f_{r_2}(x) * \cdots * f_{r_n}(x)\big)(\tau)d\tau && \text{(10b)}\\
&= \left(\left(\int_0^x f_{r_1}(z)dz\right) * f_{r_2}(x) * \cdots * f_{r_n}(x)\right)(t/L), && \text{(10c)}
\end{aligned}$$

where (10b) follows by the fact that the PDF of the sum of independent random variables equals the successive convolutions of their respective PDFs, while (10c) follows by the identity $\frac{d}{dz}\big(u(x) * v(x)\big)(z) = \big(\frac{d}{dx}u(x) * v(x)\big)(z)$.

Substituting (10c) into (6), we rewrite the problem explicitly in terms of the power variables $(p_1, \ldots, p_N)$. It is difficult to optimize $(p_1, \ldots, p_N)$ directly because they are nested in the successive convolutions inside Q_n. We propose to remove the power variables from the successive convolutions by relaxing Q_n properly, as discussed in the next subsection.

B. Proposed Outage Probability Bound

We start with a new upper bound on the PDF $f_{r_n}(x)$.

Proposition 1: For any auxiliary variable $\alpha \in \mathbb{R}$, the PDF $f_{r_n}(x)$ in (8) is upper bounded by

$$\hat{f}_{r_n}(x) = \left(\frac{\sigma^2}{\beta}\right)^M \frac{2^{x/W}(2^{x/W}-1)^{M-1}\ln 2}{(M-1)!\,W p_n^M} \cdot \hat{J}(p_n|\alpha), \tag{11}$$

where

$$\hat{J}(p_n|\alpha) = p_n^\alpha \cdot \max_{0 \le y_n \le P}\left\{\frac{1}{y_n^\alpha}\exp\left(-\frac{(2^{x/W}-1)\sigma^2}{\beta y_n}\right)\right\}. \tag{12}$$

Proof: Notice that $\hat{J}(p_n|\alpha)$ reduces to $J(p_n)$ when y_n is suboptimally set to p_n, regardless of the choice of α. Then it immediately follows that $f_{r_n}(x) \le \hat{f}_{r_n}(x)$. ■

Remark 1: The main idea behind Proposition 1 is to use a power function $\gamma \cdot p_n^\alpha$ ($\gamma > 0$) to bound $J(p_n)$ from above.

Remark 2: The optimal y_n in (12) is given by

$$y_n^\star = \min\left\{\frac{(2^{x/W}-1)\sigma^2}{\alpha\beta},\, P\right\}. \tag{13}$$

Moreover, we will discuss how to tune the auxiliary variable α in Remark 3 later on. Because the outage probability Q_n

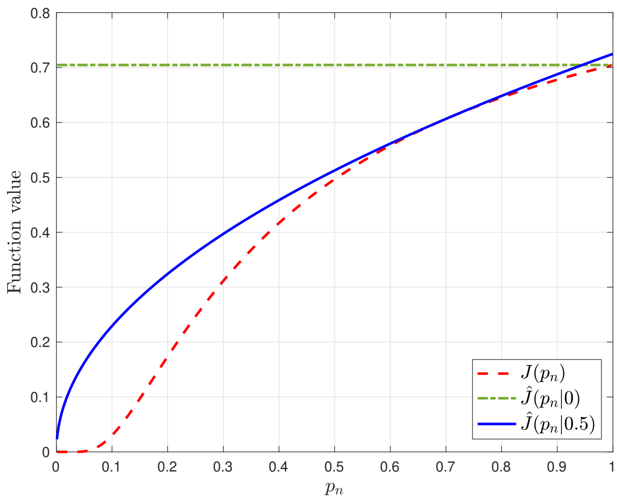

Fig. 1. The exponential function $J(p_n)$ in (9) and its upper bound $\hat{J}(p_n|\alpha)$ in (12) with the different values of the auxiliary variable α.

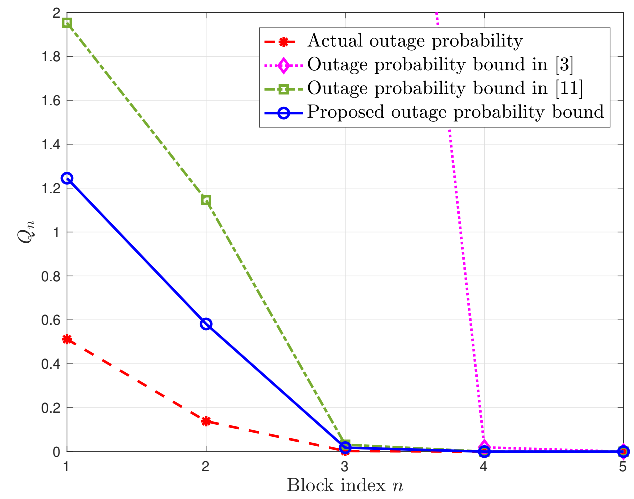

Fig. 2. Various upper bounds on outage probability Q_n with optimal α when $M = 4$, $N = 5$, $\beta = 1$, $\sigma = 1$, and $p_n = 0.8$ for each $n = 1, \ldots, N$.

increases with $f_{r_n}(x)$ monotonically, we can further construct an upper bound on Q_n by replacing $f_{r_n}(x)$ with $\hat{f}_{r_n}(x)$ in (10), as stated in the following proposition.

Proposition 2: For any auxiliary variable $\alpha \in \mathbb{R}$, the outage probability Q_n in (3) is upper bounded by

$$\begin{aligned}\hat{Q}_n &= \left(\left(\int_0^x \hat{f}_{r_1}(z)dz\right) * \hat{f}_{r_2}(x) * \cdots * \hat{f}_{r_n}(x)\right)(t/L)\\ &= A_n \prod_{i=1}^{n} p_i^{\alpha-M},\end{aligned} \tag{14}$$

where

$$A_n = \left(\frac{\sigma^{2M}\ln 2}{(M-1)!W\beta^M}\right)^n \big(\varphi(x) * \underbrace{\eta(x) * \cdots * \eta(x)}_{n-1}\big)(t/L) \tag{15}$$

along with

$$\begin{aligned}\eta(x) = {}& 2^{x/W}(2^{x/W}-1)^{M-1}\\ &\cdot \max_{0\le y_n\le P}\left\{\frac{1}{y_n^\alpha}\exp\left(-\frac{\sigma^2(2^{x/W}-1)}{\beta y_n}\right)\right\}\end{aligned} \tag{16}$$

and

$$\varphi(x) = \int_0^x \eta(z)dz. \tag{17}$$

Note that the parameter A_n is independent of $(p_1, \ldots, p_N)$.

Remark 3: The choice of α in (14) can be restricted to the interval $[0, M)$. We let $\alpha < M$ in order to render $\hat{Q}_n$ a decreasing function of the power variables (to mimic the behavior of the actual outage probability); we let $\alpha \ge 0$ because $\hat{Q}_n$ with $\alpha \ge 0$ is strictly tighter than that with $\alpha < 0$.

C. New Bound vs. Existing Bounds

We further explore the relationship between the proposed outage probability bound and the existing bounds in [3], [11], showing that the new bound $\hat{Q}_n$ in (14) is strictly tighter in general.

First, note that $y_n^\star$ in (13) equals P when α tends to zero, and hence $\hat{J}(p_n|0) = \exp\big(-\frac{(2^{x/W}-1)\sigma^2}{\beta P}\big)$; it is worth mentioning that $\hat{J}(p_n|0)$ is a constant function independent of p_n. Substituting the PDF bound $\hat{J}(p_n|0)$ in (11) and (14) yields a suboptimal version of $\hat{Q}_n$, which is exactly the upper bound from [11]. As compared to the constant function $\hat{J}(p_n|0)$, the proposed PDF bound $\hat{J}(p_n|\alpha)$ is an exponential function with higher flexibility, so it can approximate $J(p_n)$ more exactly, as illustrated in Fig. 1.

Moreover, we could obtain $\hat{J}(p_n|0) = 1$ by assuming that $\beta \to \infty$. The resulting $\hat{Q}_n$ in (14) then boils down to the classical outage probability bound in [3]. Because the value of $\hat{J}(p_n|0)$ increases with β, the classical bound [3] is looser.

D. Application to Power Adaptation

Using the new upper bound $\hat{Q}_n$ to approximate the outage probability Q_n, we convert (6) to the following GP problem

$$\underset{(p_1,\ldots,p_N)}{\text{minimize}} \quad p_1 + \sum_{n=2}^{N}\left(A_{n-1}p_n\prod_{i=1}^{n-1}p_i^{\alpha-M}\right) \tag{18a}$$

$$\text{subject to} \quad A_N\prod_{i=1}^{N}p_i^{\alpha-M} \le \epsilon \tag{18b}$$

$$\sum_{n=2}^{N}\left(A_{n-1}p_n\prod_{i=1}^{n-1}p_i^{\alpha-M}\right) \le \frac{\delta-L}{L} \tag{18c}$$

$$0 \le p_n \le P, \tag{18d}$$

which can be solved efficiently by the standard convex optimization technique. Furthermore, in light of Remark 3, we suggest searching through $[0, M)$ for the optimal α; we solve the GP problem (18) with respect to every possible value of α and choose the best. It is worthwhile to point out that the power solution $(p_1^\star, \ldots, p_N^\star)$ of the above GP problem is guaranteed to satisfy the constraints in the original problem (6) since the outage probability is bounded from above.

It remains to compute the parameters A_n in the above GP problem according to (15). In practice, we would partition the range $(0, t/L]$ into many slots, say K slots, so that the continuous convolutions in (15) are cast to numerically tractable discrete convolutions. A naive idea is to compute these discrete convolutions successively. By fast Fourier transform, each discrete convolution requires a computational complexity of $O(K \log K)$, and it takes the complexity $O(NK \log K)$ in total to obtain $(A_1, \ldots, A_N)$.

It turns out that A_n can be acquired in a more efficient way. As a key observation, the successive convolutions in (15), i.e., $\varphi(x) * \eta(x) * \cdots * \eta(x)$, have the last $n-1$ functions be the same; this special structure can be utilized to reduce complexity. Recall that the essence of successive convolutions is to try out all possible allocations of the K slots among the n functions and then add up the corresponding function products $\varphi(x)\eta(x)\cdots\eta(x)$. Since the last $n-1$ functions are identical, some of these allocations would yield the same function product. For instance, if we use the tuple (K_1, K_2, K_3, K_4) to denote the allocation of 100 slots among 4 functions, then the following allocations $(94, 1, 2, 3)$, $(94, 1, 3, 2)$, $(94, 2, 1, 3)$, $(94, 2, 3, 1)$, $(94, 3, 1, 2)$, and $(94, 3, 2, 1)$ yield the same function product. The aforementioned naive approach however does not take this repetition into account, and just calculates the same function product 6 times. Clearly, it suffices to consider only one of these allocations; this is the rationale behind the proposed fast way of computing $(A_1, \ldots, A_N)$.

We now illustrate our method through a toy example, assuming that $K = 5$ and $N = 3$. Consider such binary tree: each internal node is denoted as $(K_1, K_2, K_3 \,|\, k_0, m)$, where K_1, K_2, K_3, k_0, m are all nonnegative integers and satisfy $K_1 + K_2 + K_3 = K - k_0$; each leaf node is denoted as (K_1, K_2, K_3) with $K_1 + K_2 + K_3 = K$. This binary tree is recursively built according to the following three rules:

1) The root is $(0, 0, 0 \,|\, K, N)$;
2) Each internal node $(K_1, K_2, K_3 \,|\, k_0, m)$ has left child $(K_1, K_2, K_3 \,|\, k_0, m-1)$ if $m > 1$;
3) Each internal node $(K_1, K_2, K_3 \,|\, k_0, m)$ has right child if $k_0 \geq m$. The right child is $(K_1', K_2', K_3' \,|\, k_0 - m, m)$ if $k_0 > m > 1$, is (K_1', K_2', K_3') if $k_0 = m$, and is $(K_1 + k_0, K_2, K_3)$ if $m = 1$, where $K_i' = K_i + 1$ for $i \leq m$ and $K_i' = K_i$ otherwise.

Intuitively, the left child assumes that the remaining k_0 slots are not allocated to the mth function, while the right child assumes that at least one slot is allocated to the first m functions each. Fig. 3 displays the resulting binary tree.

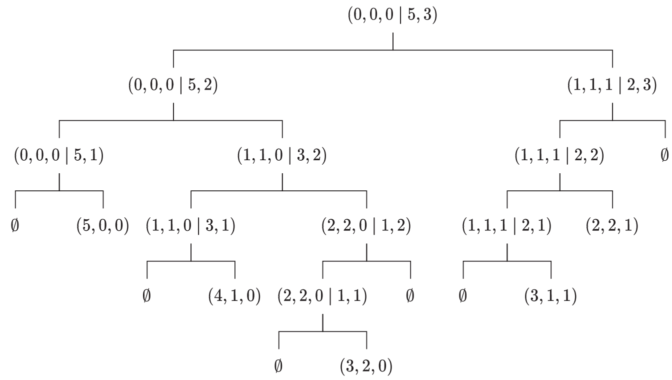

Fig. 3. The binary tree for computing A_N when $N = 3$ and $K = 5$, where $\emptyset$ represents the nil nodes.

Each leaf node shows a particular allocation that represents a group of allocations of the same combination, e.g., the rightmost leaf node $(2, 2, 1)$ in Fig. 3 represents the group $\{(2, 2, 1), (2, 1, 2), (1, 2, 2)\}$. At each leaf node, we try out all possible values for K_1, then compute the corresponding function product $\varphi(K_1\tau)\eta(K_2\tau)\cdots\eta(K_n\tau)$ with respect to each possible value of K_1 in the group, where $\tau = \frac{t}{LK}$. For example, the leaf node $(2, 2, 1)$ corresponds to the group $\{(2, 2, 1), (2, 1, 2), (1, 2, 2)\}$, so the possible values of K_1 are $\{1, 2\}$. We then compute the function product respectively for $(2, 2, 1)$ and $(1, 2, 2)$; because two allocations have $K_1 = 2$ and one allocation has $K_1 = 1$ in the group, the sum of function products at the leaf node $(2, 2, 1)$ is given by $2\varphi(2\tau)\eta(2\tau)\eta(\tau) + \varphi(\tau)\eta(2\tau)\eta(2\tau)$. We further add up function products across all the leaf nodes to obtain the overall successive convolutions. It can be shown that the overall complexity of finding $(A_1, \ldots, A_N)$ is $O(N^3 K)$. Because N is usually a small number in practice (e.g., $N = 4$ in 5G NR [12]), the complexity is dominated by K. Thus, the above way of computing $(A_1, \ldots, A_N)$ runs in linear time.

IV. Simulation Results

We now validate the performance gain of the proposed power adaptation algorithm via simulations. Assume that the size of message $t = 4$ bits, the number of receive antennas $M = 4$, the maximum number of transmission rounds $N = 5$, the outage probability constraint $\epsilon = 10^{-5}$. Without loss of generality, we normalize the background noise σ^2 and the bandwidth W to 1. Moreover, the latency in our case is measured in terms of block duration.

Fig. 4 shows the energy consumption vs. the pathloss β for the different power adaptation algorithms. It can be seen that the GP method based on the proposed bound provides more energy-efficient transmission. For instance, when $\beta = 6$, the energy consumption of the proposed method is less than half of that of the classical bound based method, and is about 60% lower than that of the maximum power scheme. Fig. 5 shows how the optimal α changes with β. It suggests that we need to raise α when the channel condition becomes worse.

Fig. 6 shows the energy-latency tradeoff. Observe that the proposed algorithm requires the lowest energy E under any latency constraint δ. The new bound allows more than 50% energy cutoff as compared to the classical bound in [3]. Observe also that the new bound can reach smaller δ values than the classical bound, so it can accommodate stricter requirements on delay. Finally, the power allocations displayed in Fig. 7 show that the new algorithm achieves more energy-efficient transmission by putting higher power in the last few transmissions rather than the initial transmission.

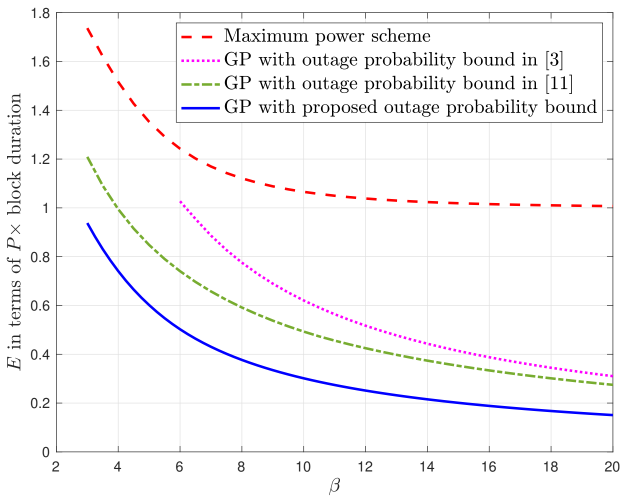

Fig. 4. Energy E vs. pathloss β when the latency constraint $\delta = 2$.

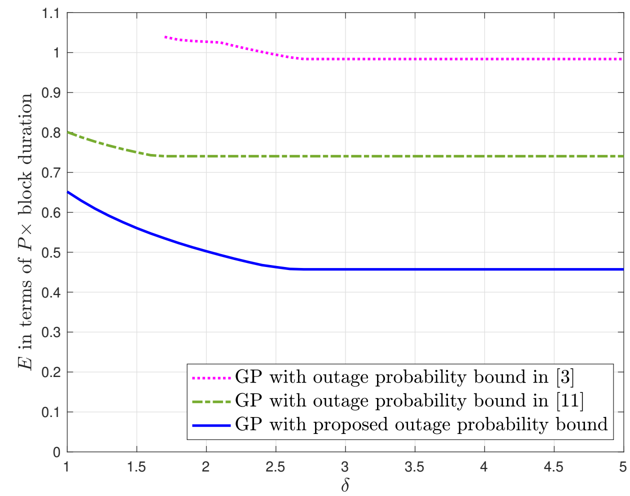

Fig. 6. Tradeoff between energy E and latency δ when $\beta = 6$.

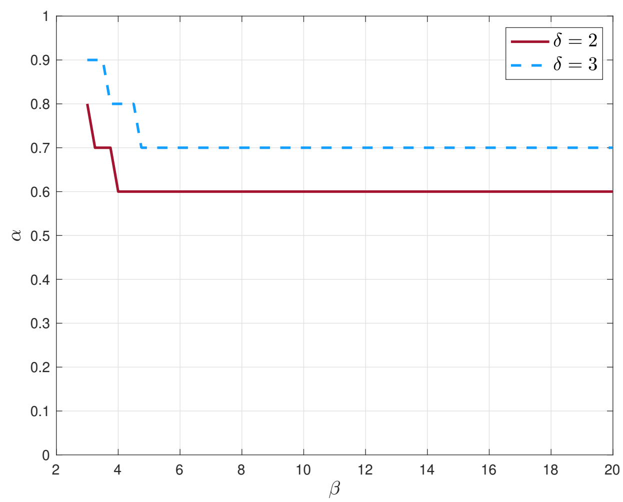

Fig. 5. Optimal α vs. pathloss β under different latency constraints.

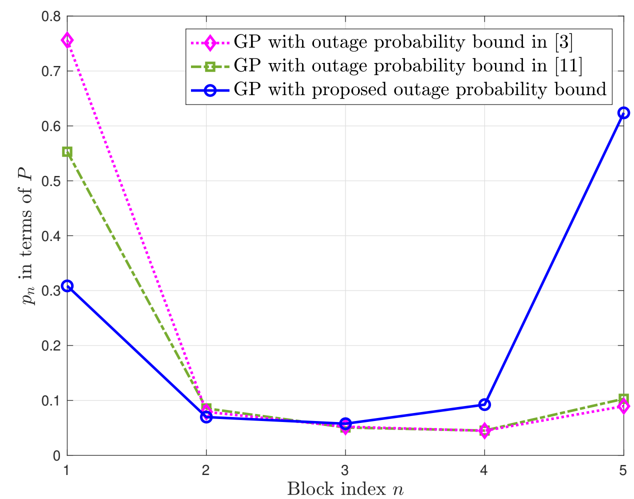

Fig. 7. Power allocations of the different algorithms when $\beta = 8$.

V. Conclusion

This work proposes a novel upper bound on outage probability for IR-HARQ over SIMO block fading channels, which is much tighter than the existing bounds. We further use the new bound to facilitate power adaptation. The optimized power allocation by GP and the new bound attains much higher energy efficiency in ultrareliable communications.

References

[1] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of harq with chase combining and incremental redundancy for HSDPA," in *IEEE Veh. Technol. Conf. (VTC)*, Oct. 2001.

[2] S. Lee, W. Su, S. Batalama, and J. D. Matyjas, "Cooperative decode-and-forward ARQ relaying: Performance analysis and power optimization," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2632–2642, Aug. 2010.

[3] J. N. Laneman, "Limiting analysis of outage probabilities for diversity schemes in fading channels," in *IEEE GLOBECOM*, Dec. 2003.

[4] W. Su, S. Lee, D. A. Pados, and J. D. Matyjas, "Optimal power assignment for minimizing the average total transmission power in hybrid-ARQ Rayleigh fading links," *IEEE Trans. Wireless Commun.*, vol. 59, no. 7, pp. 1867–1877, Jul. 2011.

[5] T. V. K. Chaitanya and E. G. Larsson, "Outage-optimal power allocation for hybrid ARQ with incremental redundancy," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2069–2074, Jul. 2011.

[6] S. Ge, Y. Xi, S. Huang, and J. Wei, "Packet error rate analysis and power allocation for CC-HARQ over Rayleigh fading channels," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1467–1470, Aug. 2014.

[7] E. Dosti, M. Shehab, H. Alves, and M. Latva-Aho, "Ultra reliable communication via optimum power allocation for HARQ retransmission schemes," *IEEE Access*, vol. 8, pp. 89 768–89 781, May 2020.

[8] S. Ge, Y. Xi, S. Huang, Y. Ma, and J. Wei, "Approximate closed-form power allocation scheme for multiple-input–multiple-output hybrid automatic repeat request protocols over Rayleigh block fading channels," *IET Commun.*, vol. 9, no. 16, pp. 2023–2032, Nov. 2015.

[9] M. Jabi, L. Szczecinski, M. Benjillali, and F. Labeau, "Outage minimization via power adaptation and allocation in truncated hybrid ARQ," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 711–723, Mar. 2015.

[10] J. Choi, "On HARQ-IR for downlink NOMA systems," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3576–3584, Aug. 2016.

[11] K. Shen, X. Chen, W. Yu, and S. R. Khosravirad, "Power adaptive HARQ for ultrareliability via a novel outage probability bound," in *IEEE ICC Workshops*, Jun. 2021.

[12] E. Dahlman, S. Parkvall, and J. Sköld, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2018.