Power Adaptive HARQ for Ultrareliability via a Novel Outage Probability Bound

Kaiming Shen[†], Xihan Chen[‡], Wei Yu[♯], and Saeed R. Khosravirad[§]
[†]School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), China
[‡]College of Information Science and Electronic Engineering, Zhejiang University, China
[♯]Electrical and Computer Engineering Dept., University of Toronto, ON M5S 3G4, Canada
[§]Nokia - Bell Labs, Murray Hill, NJ, U.S.A.
E-mail of corresponding author: shenkaiming@cuhk.edu.cn

***Abstract*—Hybrid automatic repeat request (HARQ) is a key enabler for the extremely strict reliability requirements as imposed in future networks. This paper studies power control for HARQ in order to minimize the expected energy consumption of wireless transmission given the outage probability constraint and the maximum average number of retransmissions. The main difficulty in solving the above power control problem is due to the fact that the outage probability cannot be expressed analytically in terms of the power variables. Prior works have suggested using a classic upper bound on the outage probability by assuming unbounded transmit powers throughout HARQ, thus approximating the power control problem in a geometric programming (GP) form. In contrast, this work proposes a novel and much tighter upper bound by taking practical power constraints into account. Numerical examples show that the GP method with the new upper bound outperforms one with the classic upper bound significantly. Analytical insights are also gained into the proposed GP method. Moreover, the paper extends the above results to the multi-antenna channels in which antenna diversity is harnessed to further enhance reliability.**

I. Introduction

Ultrareliable communications refer to transmitting data wirelessly with a target outage probability lower than 10^{-5}, as compared to the traditional cellular systems typically with the outage probability 10^{-2} [1]. The above stringent performance standard is driven by a multitude of evolving applications of the Internet of Things (IoT), ranging from self-driving vehicles to remote surgery [1]–[3]. This paper seeks an energy-efficient retransmission protocol to accommodate the ultrareliability requirements in these application scenarios.

More specifically, under the constraints on the outage probability and the average number of retransmissions, we consider minimizing the expected energy consumption for the hybrid automatic repeat request (HARQ) scheme. The corresponding optimization problem is difficult to tackle directly because the outage probability cannot be expressed as a closed-form function of powers. Our approach relies on a new upper bound of the outage probability to approximate the original problem in a geometric programming (GP) form. In contrast, the existing GP methods in [4]–[7] are all based on a classic upper bound [8] with high-power assumptions and are prone to overestimation of the outage probability when the transmit powers are constrained. The paper further extends the proposed upper bound and the GP-based power control method to the multi-antenna channels with antenna diversity.

In the existing literature, ultrareliable communications are examined from various perspectives. Short packet coding has attracted extensive research interests for the joint optimization of ultrareliability and low latency. Many works [9]–[11] in this area are empirically based, aimed at determining what types of codes (e.g., LDPC and polar codes) are most suited for the short blocklength regime. In order to tame the tail of the outage distribution with small deviations, [1], [12] propose using the extreme value theory. To optimize resource allocation for ultrareliable communications, [13] devises a multiplexing queueing method and [14] considers the network slicing approach. Moreover, spatial diversity is exploited in [15] to improve the ultrareliability of wireless communications. For the relay broadcast channel, [16] uses rate splitting to facilitate ultrareliable communications.

This paper is most closely related to a line of works [4]–[7] about the power control in HARQ by means of GP. Nevertheless, this work differs from [4]–[7] by developing a new outage probability bound that is much tighter than the existing one in [8]. We extend the problem formulation in [4]–[7] to include a constraint on the average number of retransmissions. We numerically demonstrate the advantage of the new upper bound and the GP method for power control as compared to the benchmark.

II. System Model

Consider a block fading channel with a sequence of N fading blocks over which the transmitter wishes to send a t-bit message toward the receiver. With the pathloss $\beta > 0$ fixed and the Rayleigh fading $z_n \sim \mathcal{CN}(0,1)$ drawn i.i.d. across the N blocks, the channel $h_n \in \mathbb{C}$ in each block n is given by

$$h_n = \sqrt{\beta} z_n,\ \forall n = 1, \ldots, N. \tag{1}$$

Let p_n be the transmit power in block n and let σ^2 be the background noise level. Assuming the normalized spectrum bandwidth and the normalized block duration without loss of generality, the achievable rate of the message transmission in block n is computed as

$$r_n = \log_2\left(1 + \frac{|h_n|^2 p_n}{\sigma^2}\right),\ \forall n = 1, \ldots, N. \tag{2}$$

The pathloss-to-noise ratio is defined by

$$s = \frac{\beta}{\sigma^2}. \tag{3}$$

We can then rewrite the achievable rate r_n as

$$r_n = \log_2\left(1 + s|z_n|^2 p_n\right),\ \forall n = 1, \ldots, N. \tag{4}$$

Moreover, by using *incremental redundancy* transmission, e.g., IR-LDPC codes [17], the mutual information accumulates in the sense that the overall achievable rate of data transmission across the first n blocks can be shown to be

$$R_n = \sum_{m=1}^{n} r_m,\ \forall n = 1, \ldots, N. \tag{5}$$

The randomness of the Rayleigh fadings $(z_1, \ldots, z_n)$ results in the following outage probability after n blocks:

$$Q_n = \Pr\big[R_n < t\big],\ \forall n = 1, \ldots, N. \tag{6}$$

The HARQ scheme handles the outage as follows. After each block n, the receiver gives a feedback ACK/NACK signal depending on whether or not the t-bit message has been successfully received, i.e., whether or not $R_n \geq t$, thereby either terminating HARQ or continuing to the next block $n+1$. The HARQ procedure finishes after the final block N, so the *ultimate* outage probability is determined by Q_N.

Since block n is used only if the previous $n-1$ blocks are insufficient, the expected value of the total energy consumption throughout the N blocks amounts to

$$E = \sum_{n=1}^{N} p_n \Pr\big[R_{n-1} < t\big] \tag{7a}$$

$$= p_1 + \sum_{n=2}^{N} p_n Q_{n-1}, \tag{7b}$$

where $R_0 = 0$. Similarly, the average number of retransmissions can be computed as

$$D = \sum_{n=1}^{N-1} \Pr\big[R_n < t\big] \tag{8a}$$

$$= \sum_{n=1}^{N-1} Q_n. \tag{8b}$$

We can now formulate an optimization problem of minimizing E under the power constraint P, given the target outage probability ϵ and the retransmission constraint δ, i.e.,

$$\underset{\mathbf{p}}{\text{minimize}} \quad E \tag{9a}$$

$$\text{subject to} \quad Q_N \leq \epsilon, \tag{9b}$$

$$D \leq \delta, \tag{9c}$$

$$0 \leq p_n \leq P, \tag{9d}$$

where $\mathbf{p} = (p_1, \ldots, p_N)$. The main obstacle posed in (9) is that none of (Q_N, E, D) could be written analytically. Because E and D are comprised of an array of Q_n's, the heart of the problem is the intractable expression (6) of the outage probability. This work overcomes the obstacle by using a new upper bound, which is tighter than the existing one in [8], to approximate Q_N and those Q_n terms contained in E and D. The resulting approximation of (9) turns out to be a GP problem readily solvable by the convex optimization methods.

In addition, it is worth mentioning that the three metrics (Q_N, E, D) can be put together in other ways, e.g., we could have minimized D under the E and Q_N constraints, and our approach is amenable to most of these formulations as well.

III. Upper Bound on Outage Probability

We begin with a single block. If the receiver decodes the message by using block n alone without incremental redundancy, the cumulative distribution function (CDF) for the achievable rate r_n is

$$F_n(r) = \Pr\big[r_n < r\big] \tag{10a}$$

$$= \Pr\left[|z_n|^2 < \frac{2^r - 1}{sp_n}\right] \tag{10b}$$

$$= 1 - \exp\left(-\frac{2^r - 1}{sp_n}\right), \tag{10c}$$

and the corresponding probability density function (PDF) of r_n is

$$f_n(r) = \frac{d}{dr} F_n(r) \tag{11a}$$

$$= \frac{2^r \ln 2}{sp_n} \cdot \exp\left(-\frac{2^r - 1}{sp_n}\right). \tag{11b}$$

Recall that the accumulated rate R_n equals to $\sum_{m=1}^{n} r_m$, so the PDF of R_n, denoted by $g_n(R)$, can be computed as the successive convolution of the respective PDFs of r_m's over the range $[0, R)$, i.e.,

$$g_n(R) = \underset{\substack{0 \leq r_m \leq R,\, \forall m \\ r_1 + \cdots + r_n = R}}{\int \cdots \int} f_1(r_1) \cdots f_n(r_n) dr_1 \cdots dr_n \tag{12a}$$

$$= (f_1 * f_2 * \cdots * f_n)(R). \tag{12b}$$

Further, the corresponding CDF of the accumulated rate R_n is given by

$$G_n(R) = \int_0^R g_n(\tau) d\tau \tag{13a}$$

$$= \left(\left(\int_0^{r_1} f_1(\tau) d\tau\right) * f_2 * \cdots * f_n\right)(R) \tag{13b}$$

$$= (F_1 * f_2 * \cdots * f_n)(R), \tag{13c}$$

in which (13b) follows by the identity $(u * v)' = u' * v$. Observe that the outage probability Q_n in (6) equals to the CDF of the accumulated rate R_n when $R_n = t$, so it can be rewritten as

$$Q_n = G_n(t). \tag{14}$$

In order to make the optimization of $\mathbf{p}$ tractable, we wish to isolate $\mathbf{p}$ from the successive convolution in (13c). Toward this

$$A_n(t) = P\left(\frac{\ln 2}{s}\right)^{n-1} \cdot \left(\left[1 - \exp\left(-\frac{2^{r_1}-1}{sP}\right)\right] * \underbrace{\left[2^{r_2}\exp\left(-\frac{2^{r_2}-1}{sP}\right)\right] * \cdots * \left[2^{r_n}\exp\left(-\frac{2^{r_n}-1}{sP}\right)\right]}_{(n-1)\text{ terms}}\right)(t) \quad (18)$$

---

end, we first relax $f_n(t)$ in (11) as

$$\begin{aligned} f_n(r) &\le \hat{f}_n(r) && (15a)\\ &= \frac{2^r \ln 2}{sp_n} \cdot \exp\left(-\frac{2^r - 1}{sP}\right), && (15b)\end{aligned}$$

which further yields an upper bound on $F_n(r)$:

$$\begin{aligned} F_n(r) &\le \hat{F}_n(r) && (16a)\\ &= \int_0^r \hat{f}_n(\tau)d\tau && (16b)\\ &= \frac{P}{p_n}\left(1 - \exp\left(-\frac{2^r-1}{sP}\right)\right). && (16c)\end{aligned}$$

With $f_n(r)$ and $F_n(r)$ respectively bounded by $\hat{f}_n(r)$ and $\hat{F}_n(r)$ in (13c), we construct an upper bound on Q_n as

$$\begin{aligned} \hat{Q}_n &= (\hat{F}_1 * \hat{f}_2 * \cdots * \hat{f}_n)(t) && (17a)\\ &= A_n(t)\left(\prod_{m=1}^{n} p_m\right)^{-1}, && (17b)\end{aligned}$$

where $A_n(t)$ is independent of p_n as shown in (18). The following proposition summarizes the above result.

Proposition 1 (Upper Bound on Outage Probability): For a block fading channel over N i.i.d. Rayleigh fading blocks in which a t-bit message is transmitted with incremental redundancy, when the pathloss-to-noise ratio s and the power constraint P are fixed, we have the following upper bound on the outage probability Q_n after n transmissions:

$$Q_n \le \hat{Q}_n, \quad \forall n = 1, \ldots, N, \tag{19}$$

where $\hat{Q}_n$ is given by (17) and (18).

Further, the upper bound in [8] is a special case of the proposed bound $\hat{Q}_n$, as shown in the subsequent proposition.

Proposition 2 (Connection to Classic Upper Bound): As $P \to \infty$, the parameter $A_n(t)$ in (18) tends to

$$A_{n,\infty}(t) = \frac{(\ln 2)^{n-1}}{s^n} \cdot \big((2^{r_1} - 1) * \underbrace{2^{r_2} * \cdots * 2^{r_n}}_{(n-1)\text{ terms}}\big)(t) \tag{20}$$

and accordingly $\hat{Q}_n$ tends to

$$\hat{\hat{Q}}_n = A_{n,\infty}(t)\left(\prod_{m=1}^{n} p_m\right)^{-1}, \tag{21}$$

which is the upper bound proposed in [8]. Moreover, given any $t > 0$, we have

$$Q_n \le \hat{Q}_n \le \hat{\hat{Q}}_n, \quad \forall n = 1, \ldots, N, \tag{22}$$

where the two inequalities are tight simultaneously as $s \to \infty$.

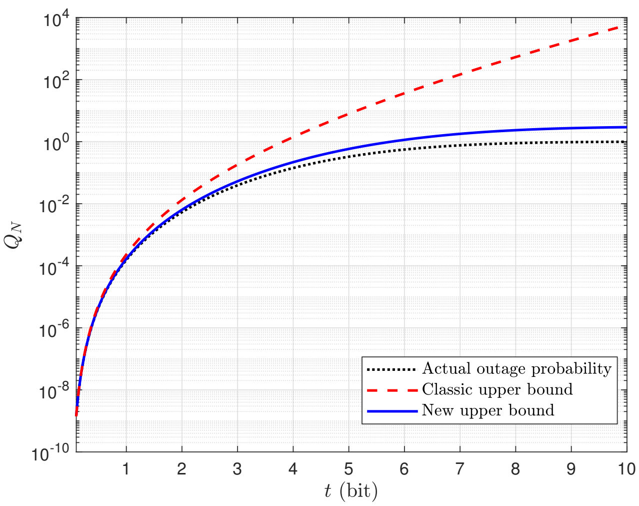

Fig. 1. Actual outage probability vs. classic upper bound vs. proposed upper bound when $N = 5$, $s = 2$, $P = 1$, and $p_n/P = 0.8$ for any n.

Proof: Observe that each $\hat{f}_n(r)$ in (15b) is monotonically increasing with P when r is fixed; observe also that $\hat{Q}_n$ in (17) equals to $\int_0^t (\hat{f}_1 * \hat{f}_2 * \cdots * \hat{f}_n)(\tau)d\tau$, so $\hat{Q}_n$ is monotonically increasing with P and thus $\hat{Q}_n \le \hat{\hat{Q}}_n$. Regarding the equalities, it can be seen that $f_n(r) = \hat{f}_n(r)$ and $F_n(r) = \hat{F}_n(r)$ when $s \to \infty$, so $Q_n = \hat{Q}_n$ also holds. This argument works for $\hat{\hat{Q}}_n$ as well. ■

The initial derivation of $\hat{\hat{Q}}_n$ in [8] builds upon a piecewise squeezing argument, whereas we have obtained the same result by specializing the proposed bounding technique in (15)–(18). Furthermore, given moderate s and P, i.e., when the signal-to-noise ratio is not sufficiently high, the gap between the two upper bounds can be quite large especially for a large t, as illustrated in Fig. 1.

IV. Power Control by Geometric Programming

Using the new upper bound $\hat{Q}_n$ to approximate the outage probability Q_n in (9) leads us to the following GP problem:

$$\begin{aligned} \underset{\mathbf{p}}{\text{minimize}} \quad & p_1 + \sum_{n=2}^{N} A_{n-1}(t)p_n\left(\prod_{m=1}^{n-1} p_m\right)^{-1} && (23a)\\ \text{subject to} \quad & A_N(t)\left(\prod_{m=1}^{N} p_m\right)^{-1} \le \epsilon, && (23b)\\ & \sum_{n=1}^{N-1} A_n(t)\left(\prod_{m=1}^{n} p_m\right)^{-1} \le \delta, && (23c)\\ & 0 \le p_n \le P, && (23d)\end{aligned}$$

which can be efficiently solved by the standard numerical methods. We further provide insights into this GP method.

Remark 1: Using the upper bound rather than the lower bound here guarantees that (E, D) can only be overestimated, so the outage probability and the retransmission constraints in the original problem (9) can be guaranteed in spite of the approximation.

Remark 2: The proposed method basically approximates Q_N as a monomial, E as a posynomial, and D as another posynomial. Many other problems involving (Q_N, E, D) can also be converted into a GP form using this approximation.

Proposition 3: When $P \to \infty$, the solution of (23) satisfies

$$p_n \hat{Q}_{n-1} = 2p_{n+1}\hat{Q}_n + \lambda \hat{Q}_n,\ \forall n = 1, \ldots, N-1, \quad (24)$$

for some $\lambda \geq 0$.

Proof: First, incorporate the retransmission constraint (23c) into the objective function in (23a) by using a Lagrange multiplier $\lambda \geq 0$. Next, it can be shown that the outage probability constraint (23b) must be tight at the optimum; otherwise the objective function can be further reduced by decreasing p_N properly; this tightness allows p_N to be expressed in $(p_1, \ldots, p_{N-1})$. Further, introduce a set of new variables $y_n \in \mathbb{R}$ and substitute $p_n = e^{y_n}$ in the GP problem; note that the constraint $p_n \geq 0$ is now satisfied automatically. We then arrive at an unconstrained convex problem of $(y_1, \ldots, y_{N-1})$. Solving the first-order condition gives

$$e^{2y_n} A_{n-1}(t) = (2e^{y_{n+1}} + \lambda) A_n(t),\ \forall n = 1, \ldots, N, \quad (25)$$

which recovers (24) when $y_n = \ln p_n$ is substituted in. ■

Remark 3: We might be tempted to remove the power constraint (9d) by postulating that the minimization of the expected energy consumption E would suppress each p_n already. However, this is not true. For ease of discussion, we relax the retransmission constraint (23c) by setting $\lambda = 0$, then obtain from (24) that

$$p_N = \frac{p_1}{2^{N-1}\hat{Q}_{N-1}}. \quad (26)$$

It turns out that the denominator $2^{N-1}\hat{Q}_{N-1}$ could be quite close to zero in practice. For a typical ultrareliable communication system, the approximated outage probability $\hat{Q}_{N-1}$ lies in the interval $[10^{-9}, 10^{-5}]$ while 2^{N-1} is around 100 at most. As a consequence, there is an impractical 70 dB gap between p_1 and p_N. The intuition is that minimizing E can only suppress the expected value $p_n Q_{n-1}$, but the actual transmit power p_n may still spike as Q_{n-1} becomes arbitrarily small.

V. Multi-Antenna Channels

This section generalizes the proposed upper bound $\hat{Q}_n(t)$ and the GP-based power control method to a multiple-input single-output (MISO) channel. In statistics terms, the generalized upper bound is an extension from 2 degrees of freedom to higher degrees of freedom for the chi-square distribution; in wireless communication terms, the generalized upper bound takes antenna diversity into consideration.

A. Antenna Diversity

Assume that the transmitter is now deployed with $L > 1$ antennas. The channel $\mathbf{h}_n \in \mathbb{C}^L$ in block n is

$$\mathbf{h}_n = \big(h_{1n}, h_{2n}, \ldots, h_{Ln}\big), \quad (27)$$

wherein each entry h_{jn} is modeled as

$$h_{jn} = \sqrt{\beta} z_{jn} \quad (28)$$

with $\beta > 0$ fixed and z_{jn} drawn i.i.d. from $\mathcal{CN}(0, 1)$ for each (j, n) pair. By *space-time coding*, the following rate is achievable in block n:

$$r_n = \log_2\left(1 + \sum_{j=1}^{L} \frac{|h_{jn}|^2 p_n}{\sigma^2}\right) \quad (29a)$$

$$= \log_2\left(1 + s p_n \sum_{i=1}^{L} |z_{jn}|^2\right) \quad (29b)$$

$$= \log_2\left(1 + \frac{1}{2} s p_n \nu_n\right), \quad (29c)$$

where the chi-square random variable

$$\nu_n \sim \chi^2(d) \quad \text{with } d = 2L \quad (30)$$

has d degrees of freedom. The rest of setting follows that in Section II. Please note that the above model works for a single-input multiple-output (SIMO) channel as well. In this case, the receiver has L antennas and achieves the data rate r_n in (29) via maximum ratio combining (MRC). Furthermore, these results can be extended to the Nakagami fading channels.

B. Generalized Upper Bound on Q_n

The Rayleigh distribution in Section III can be recognized as a special case of the chi-square distribution with 2 degrees of freedom. We seek a generalization of the upper bound $\hat{Q}_n$ that works for an arbitrary number of degrees of freedom.

Following the steps in Section III, we first compute the CDF of the rate r_n in (29) as

$$F_n(r) = \Pr\left[\nu_n < \frac{2(2^r - 1)}{s p_n}\right] \quad (31a)$$

$$= \frac{1}{(d-1)!} \cdot \gamma\left(d, \frac{2^r - 1}{s p_n}\right), \quad (31b)$$

where $\gamma(\cdot, \cdot)$ refers to the *lower incomplete gamma function*

$$\gamma(a, b) = \int_0^b x^{a-1} \exp(-x) dx. \quad (32)$$

The corresponding PDF of r_n is

$$f_n(r) = \frac{(2^r - 1)^{d-1}}{s^d (d-1)!} \cdot \exp\left(-\frac{2^r - 1}{s p_n}\right) \cdot \frac{2^r \ln 2}{p_n^d}. \quad (33)$$

Following (15b), we make use of the power constraint P to construct an upper bound on $f_n(r)$ as

$$\hat{f}_n(r) = \frac{(2^r - 1)^{d-1}}{s^d (d-1)!} \cdot \exp\left(-\frac{2^r - 1}{sP}\right) \cdot \frac{2^r \ln 2}{p_n^d}, \quad (34)$$

$$\tilde{A}_n(t) = \left(\frac{P^d (\ln 2)^{n-1} \gamma\left(d, \frac{2^{r_1}-1}{sP}\right)}{s^{d(n-1)}\left((d-1)!\right)^n} * \underbrace{\left[2^{r_2}(2^{r_2}-1)^{d-1} \mathsf{exp}\left(-\frac{2^{r_2}-1}{sP}\right)\right] * \cdots * \left[2^{r_n}(2^{r_n}-1)^{d-1} \mathsf{exp}\left(-\frac{2^{r_n}-1}{sP}\right)\right]}_{(n-1) \text{ terms}} \right)(t) \tag{37}$$

which in return gives an upper bound on $F_n(r)$ according to (16):

$$\hat{F}_n(r) = \left(\frac{P}{p_n}\right)^d \cdot \frac{1}{(d-1)!} \cdot \gamma\left(d, \frac{2^r - 1}{sP}\right). \tag{35}$$

Substituting $\hat{f}_n(r)$ and $\hat{F}_n(r)$ into (17a) further leads us to an upper bound on Q_n:

$$\hat{Q}_n = \tilde{A}_n(t) \left(\prod_{m=1}^{n} p_m\right)^{-d}, \tag{36}$$

where $\tilde{A}_n(t)$ is given by (37). In particular, as $P \to \infty$, the proposed upper bound $\hat{Q}_n$ reduces to

$$\hat{\hat{Q}}_n = \tilde{A}_{n,\infty}(t) \left(\prod_{m=1}^{n} p_m\right)^{-d} \tag{38}$$

along with

$$\tilde{A}_{n,\infty}(t) = \left(\frac{\ln 2}{s^d (d-1)!}\right)^n \frac{d}{\ln 2} \cdot \left((2^{r_1}-1)^d * [2^{r_2}(2^{r_2}-1)^{d-1}] * \cdots * [2^{r_n}(2^{r_n}-1)^{d-1}]\right)(t), \tag{39}$$

which is an extension of the classic bound [8] as proposed in [7]. Moreover, the optimality analysis in Proposition 2 carries over to the multi-antenna channels, i.e., given any $t > 0$, we have

$$Q_n \le \hat{Q}_n \le \hat{\hat{Q}}_n, \quad \forall n = 1, \ldots, N, \tag{40}$$

where the two inequalities are tight simultaneously as $s \to \infty$.

Most importantly, each $\hat{Q}_n$ remains a monomial in the multi-antenna channels, so (E, D) can still be approximated as polynomials and hence the GP method readily extends.

VI. Numerical Examples

This section shows numerically that the new upper bound $\hat{Q}_n$ is superior to the classic upper bound $\hat{\hat{Q}}_n$ for solving the power control problem (9). We set the number of blocks $N = 5$, the message size $t = 4$ bits, and the outage probability constraint $\epsilon = 10^{-5}$. For ease of illustration, the power variables p_n are normalized by the max power P.

Fig. 2 compares GP with the classic upper bound $\hat{\hat{Q}}_n$ and GP with the new upper bound $\hat{Q}_n$. As opposed to the maximum power scheme with $p_n = P$, the GP methods reduce the expected energy consumption E significantly, e.g., E is reduced by nearly 50% when $s = 30$. The two GP methods have close performance when s is sufficiently large. Nonetheless, when the number of transmit antennas raises to 3, as shown in Fig. 3, GP with new upper bound has a considerable advantage since it works in a much wider range of the pathloss-to-noise ratio s. In contrast, GP with classic upper bound fails to give valid solution because of its overestimation of the outage probability. Observe also that the GP methods can save more energy when L increases.

Fig. 4 takes a closer look at the two GP methods when the transmitter has $L = 3$ antennas and at most $\delta = 2$ retransmissions are allowed on average. As it turns out, GP with new bound can save more energy because it sets a much lower (25% less) power level in the first block when $s = 6$, even though it requires higher power than GP with classic bound in the last block. The figure also shows that the two GP methods give similar power solutions when $s = 20$. This result agrees with our analysis in Section V that the two upper bounds both converge to the actual outage probability as $s \to \infty$; the result also agrees with what we have observed in Fig. 3, i.e., GP with new bound has much better performance when s is small.

Finally, we present the E-δ tradeoffs for the two GP methods in Fig. 5. Observe that the GP methods flatten out in E when δ is sufficiently large; that is the case where power control can be carried out as if the average latency constraint does not exist. Along the E-axis, it can be seen that GP with new bound can cut the energy consumption by over 20% as compared to GP with classic bound at the same δ. Besides, along the δ-axis, the figure shows that GP with new bound can accommodate much stricter average latency constraints.

VII. Conclusion

This work aims to enhance energy efficiency for ultra-reliable communications by optimizing transmit powers with HARQ. The main contribution is a new upper bound on the outage probability, which can be used to approximate the originally intractable power optimization problem much more closely than the classic upper bound. We formulate the problem as a GP and show how to adaptively optimize the power allocation to achieve ultrareliability while satisfying a delay constraint. Further, the proposed upper bound along with the GP method for power control is extended to the multi-antenna channels in order to reap the antenna diversity gain.

References

[1] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[2] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.

[3] A. Azari, M. Ozger, and C. Cavdar, "Risk-aware resource allocation for URLLC: Challenges and strategies with machine learning," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 42–48, Mar. 2019.

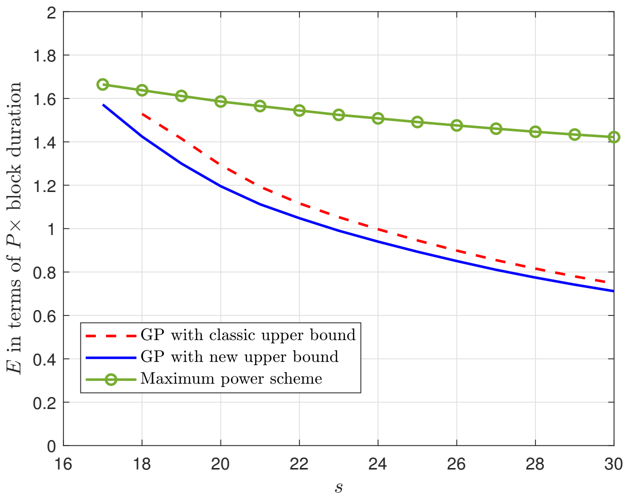

Fig. 2. Energy E vs. pathloss-to-noise ratio s when $L = 1$ and $\delta = 3$.

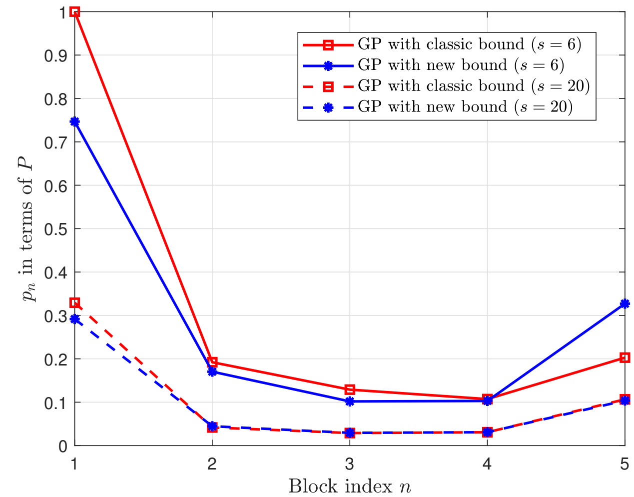

Fig. 4. Optimized power p_n for each block n when $L = 3$ and $\delta = 2$.

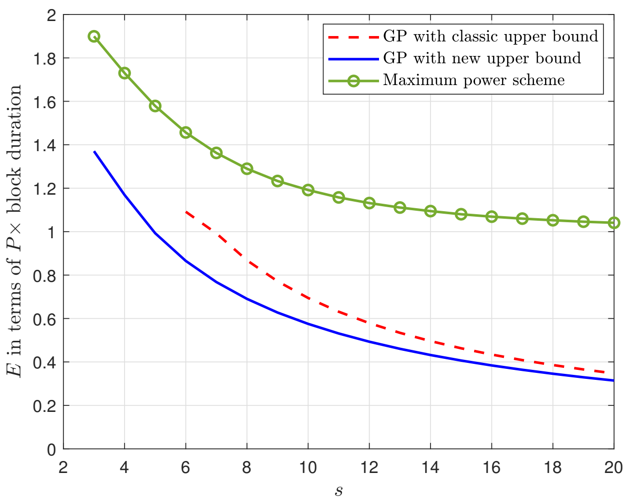

Fig. 3. Energy E vs. pathloss-to-noise ratio s when $L = 3$ and $\delta = 2$.

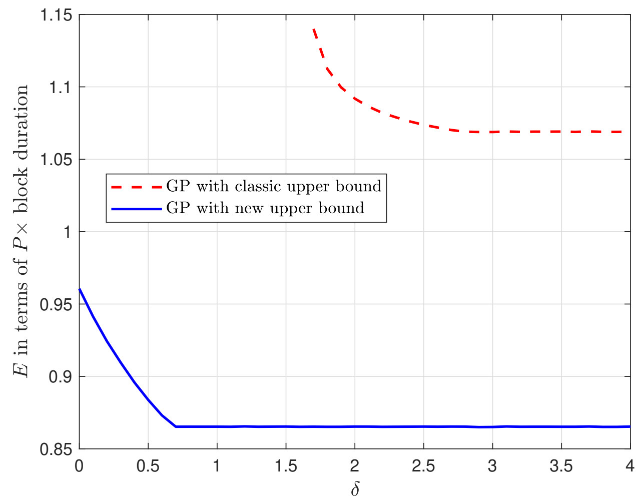

Fig. 5. Energy E vs. latency threshold δ when $s = 6$ and $L = 3$.

[4] 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.

[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, July 2011.

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

[7] 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.

[8] J. N. Laneman, "Limiting analysis of outage probabilitites for diversity schemes in fading channe," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2003, pp. 1242–1246.

[9] M. Sybis, K. Wesołowski, K. Jayasinghe, V. Venkatasubramanian, and V. Vukadinovic, "Channel coding for ultra-reliable low-latency communication in 5G systems," in *IEEE Veh. Tech. Conf. (VTC Fall)*, Sept. 2016.

[10] X. Wu, M. Jiang, C. Zhao, L. Ma, and Y. Wei, "Low-rate PBRL-LDPC codes for URLLC in 5G," vol. 7, no. 5, pp. 800–803, Oct. 2018.

[11] M. Shirvanimoghaddam, M. S. Mohammadi, R. Abbas, A. Minja, C. Yue, B. Matuz, G. Han, Z. Lin, W. Liu, Y. Li, S. Johnson, and B. Vucetic, "Short block-length codes for ultra-reliable low latency communications," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 130–137, Feb. 2019.

[12] C.-F. Liu and M. Bennis, "Ultra-reliable and low-latency vehicular transmission: An extreme value theory approach," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1292–1295, June 2018.

[13] C. She, C. Yang, and T. Q. S. Quek, "Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2266–2280, May 2018.

[14] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 881–895, Apr. 2019.

[15] V. N. Swamy, S. Suri, P. Rigge, M. Weiner, G. Ranade, A. Sahai, and B. Nikolić, "Real-time cooperative communication for automation over wireless," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7168–7183, Nov. 2017.

[16] K. Shen, W. Yu, and S. R. Khosravirad, "Ultrareliable wireless communication with message splitting," in *IEEE Int. Workshop Sig. Process. Advances Wireless Commun. (SPAWC)*, July 2019.

[17] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, Aug. 2004.