H-ARQ based Non-Orthogonal Multiple Access with Successive Interference Cancellation

Jinho Choi
Institute of Advanced Telecommunications
Swansea University, UK
Email: j.choi@swansea.ac.uk

***Abstract*—In this paper, we consider non-orthogonal multiple access (NOMA) over block fading channels where the performance is limited by interference and fading. Re-transmission and interference cancellation techniques can be employed to improve performance. Re-transmissions can provide diversity gain, while successive interference cancellation (SIC) can improve signal to interference ratio (SIR). We can show that the proposed NOMA approach with re-transmission outperforms an orthogonal multiple access (OMA) approach with re-transmission using information outage probability when the signal to noise ratio (SNR) is low. It can also be shown that the outage probability of the NOMA with SIC is lower than that of the OMA when the transmission rate is sufficiently low where SIC can be facilitated.**

Keywords: Non-orthogonal multiple access, re-transmission diversity, successive interference cancellation

I. Introduction

The superiority of non-orthogonal multiple access (NOMA) schemes (e.g., code division multiple access (CDMA)) to orthogonal multiple access (OMA) schemes (e.g., time division multiple access (TDMA)) has been well-known in terms of the channel capacity [1], [2], [3]. The use of stripping or cancellation plays a key role in achieving channel capacity for NOMA [1], [2]. In [3], the capacity of randomly spread CDMA systems is studied under various conditions. It is shown that the multiuser gain is crucial to offset the performance loss caused by multiuser interference.

In [4], [5], [6], it is shown that the linear minimum mean square error (MMSE) receiver with successive interference cancellation (SIC) can be used at the receiver for NOMA systems to achieve channel capacity. This is an important observation as a relatively simple receiver can be used to achieve channel capacity.

In this paper, we consider a NOMA scheme which is different from CDMA. It is assumed that there are multiple users in a common channel (multiple access channel) and hybrid-automatic repeat request (H-ARQ) protocol [7], [8] is used for reliable communications. Each user transmits coded packets and incremental redundancy (IR) is used for H-ARQ, which is called the H-ARQ with IR. Note that this system is also considered and compared with other systems using throughput in [9], [10]. We assume block fading [11] in this paper. Thus, the information outage probability [12] rather than the channel capacity is used as a performance index. In general, we consider the two key ingredients for NOMA systems to improve performance: H-ARQ with IR and SIC. The H-ARQ with IR is an effective means to adapt transmission rate over fading channel condition without fixing transmission rate in advance. The transmission rate can be effectively adjusted by feedback in H-ARQ: the effective transmission (code) rate decreases as more re-transmissions are carried out in the H-ARQ with IR. The SIC can effectively reduce the signal to interference-plus-noise ratio (SINR) in NOMA systems. In fading channels, since the SIC can cancel or strip strong interfering signals first, it becomes an effective means to exploit the multiuser gain and improve the performance. Note that the multiuser gain by using the SIC is different from the multiuser diversity gain in [13]. The multiuser diversity gain can be achieved by selecting the user of the highest channel gain to access channel, while the multiuser gain in the context of SIC could increase the SINR by cancelling interfering signals (but not improve diversity gain). Throughout the paper, we show the impact of H-ARQ with IR and SIC on the outage probability and consider comparisons with OMA systems.

II. System Model

In this section, we consider a NOMA scheme based on the H-ARQ with IR. In multiuser system, if multiple packets are transmitted, there is collision and re-transmission request is sent to the users in collision. Although the packets are collided, each packet can transmit a certain amount of its information. Thus, in order to improve the spectral efficiency, the H-ARQ with IR is utilized for combining all the retransmitted signals effectively. The resulting NOMA can be called the H-ARQ with IR based NOMA. However, for convenience, throughout the paper, the H-ARQ with IR based NOMA will be simply referred to as the NOMA if there is no significant risk of confusion. In addition, the SIC can be used to increase the SINR, which can also improve the spectral efficiency of the NOMA.

Suppose that there are K senders or transmitters. We assume that all the signals share the same time and frequency channel. In addition, each transmitter has its dedicated re-

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

ceiver[1]. At the receiver, the received signal is written as

$$r_m = \sum_{k=1}^{K} h_k s_{k,m} + n_m, \; m = 0, 1, \ldots, M-1, \quad (1)$$

where r_m denotes the mth received signal sample, h_k and $s_{k,m}$ denote the channel coefficient over the packet duration (block fading is assumed) and the mth data symbol from transmitter k, respectively, and n_m denotes the background zero-mean white Gaussian noise. We consider the following addition assumptions.

- A packet consists of M data symbols and can be decodable individually. That is, a packet corresponds to a codeword which is self-decodable and each re-transmitted packet can be considered the codeword which is generated by a different code from the the same message sequence for the IR. (This approach would be different from [14] where a single code is used.)
- All the packets are synchronized[2] (a slotted multiple access system is assumed).

In addition, it is assumed that the signal from transmitter 1 is the desired signal (i.e., it is assumed that receiver 1 is dedicated to transmitter 1). Fig. 1 shows an example of the system where $K = 4$. The same multiple access is considered in [9]. Throughout the paper, we assume that no power control is used and the feedback information from the receiver to its corresponding transmitter is ACK (when the packet is successfully decoded) or NACK (when the packet is not successfully decoded). In this system, we can see that the rate control is adaptively carried out by re-transmission.

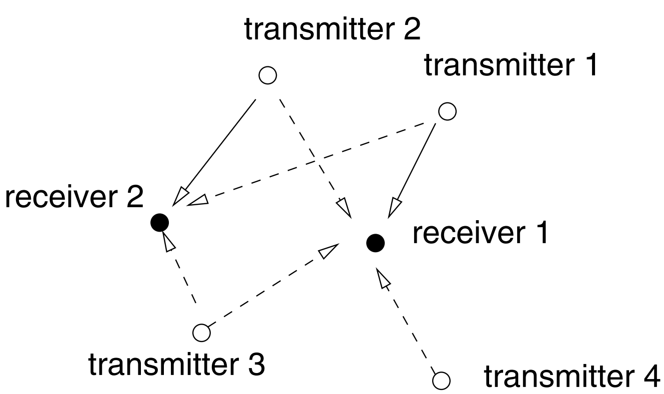


Fig. 1. An example of the system with $K = 4$ (solid lines represent the desired signals, while dashed lines represents undesired signals).

A. H-ARQ with Incremental Redundancy

Assume that a codeword is transmitted over a packet duration. If the signal is not decodable due to interfering signals and/or background noise, the re-transmission request is sent to the transmitter. The re-transmitted packet is combined with the previous packet(s) for decoding. We assume that a capacity achieving channel code is employed. Note that in [15] the throughput of H-ARQ with IR is studied with LDPC codes and it is shown that a practical design using LDPC codes can closely approach the information-theoretical limit found in [9].

For the lth packet from transmitter k, the (normalized) instantaneous channel capacity can be given as follows [9]:

$$I_{k,l} = \log_2 \left(1 + \frac{\gamma_k |h_{k,l}|^2}{1 + \sum_{q \neq k} \gamma_q |h_{q,l}|^2} \right), \quad (2)$$

where $h_{k,l}$ denotes the channel coefficient from transmitter k to the receiver over the lth data packet duration from transmitter k and γ_k is the normalized signal to noise ratio (SNR) of the signal from transmitter k. Suppose that the normalized rate of each data packet (from all the transmitters) is R_k. When the H-ARQ with IR is employed, the information outage probability for the desired signal with L re-transmissions (a transmission and $L-1$ re-transmissions) can be given as follows:

$$P_{\mathrm{no,out}}(R_1, L) = \Pr(S_{1,L} < R_1),$$

where $S_{1,L} = \sum_{l=1}^{L} I_{1,l}$ [9].

Note that if K signals have their dedicated *orthogonal* channels (i.e., in OMA systems), each signal's instantaneous mutual information can be given as

$$I_{\mathrm{orth},k,l} = \frac{1}{K} \log_2 \left(1 + \gamma_k |h_{k,l}|^2 \right), \quad (3)$$

Here, we assume that the channel is equally divided into K sub-channels. Then, for orthogonal channels, the outage probability becomes

$$P_{\mathrm{orth,out}}(R_1, L) = \Pr(S_{\mathrm{orth},1,L} < R_1),$$

where $S_{\mathrm{orth},1,L} = \sum_{l=1}^{L} I_{\mathrm{orth},1,l}$. For the sake of simplicity, we assume that the rate, R_k, is the same for all the signals, i.e., $R_k = R$ for all k (symmetric case in terms of rate). In practice, however, they can be different from one signal to another and depend on the actual data rate of each signal and coding scheme.

B. SIC Based Decoding

In (2), the other $(K-1)$ signals become the interference and lower the SINR. If some signals can be corrected decoded and cancelled or stripped, the SINR can be higher and the performance can be improved. For example, if the other $K-1$ packets are decodable, the cancellation of $K-1$ signals is possible and the resulting instantaneous channel capacity becomes

$$I_{1,l} = \log_2 \left(1 + \gamma_1 |h_{1,l}|^2 \right).$$

Thus, we can see that the probability of successful decoding can increase and the average number of re-transmissions can be smaller. This results in a higher throughput or spectral efficiency for the NOMA scheme.

[1]As an example, we can consider a cellular system with a reuse factor of 1. For uplink channels, receivers become base stations, while transmitters become users and each user has its dedicated base station. As the reuse factor is 1, there could be multiple users (from adjacent cells) who access the same uplink channel.

[2]Synchronization is not necessary unless the SIC is used. However, for convenience, we assume a synchronous NOMA system.

Throughout the paper, we assume that all the re-transmitted packets from transmitter 1 are optimally combined, while no combining is used for the signals from the other $(K-1)$ transmitters. As the receiver is dedicated to the desired signal, the receiver does not know that whether another (interfering) transmitter re-transmits a packet or not. Thus, the decoding and cancellation of the other signals should be performed using the current packet without any combining with previously transmitted packets.

For the SIC, the order of the detection/cancellation becomes important. For example, suppose that the packet from transmitter K is firstly decoded. If the decoding succeeds, that is,

$$\log_2\left(1+\frac{\gamma_K|h_{K,l}|^2}{1+\sum_{q=1}^{K-1}\gamma_q|h_{q,l}|^2}\right)\geq R,$$

the packet from transmitter K can be cancelled or stripped. Then, the packet from transmitter $K-1$ is to be decoded, and it can be decodable and cancelled if

$$\log_2\left(1+\frac{\gamma_{K-1}|h_{K-1,l}|^2}{1+\sum_{q=1}^{K-2}\gamma_q|h_{q,l}|^2}\right)\geq R.$$

To cancel more interfering signals through the SIC, the signal (among the $(K-1)$ interfering signals) that has the highest SINR should be chosen first:

$$u_1^*=\arg\max_{k\in\{2,3,\ldots,K\}}\log_2\left(1+\frac{\gamma_k|h_{k,l}|^2}{1+\sum_{q\neq k}\gamma_q|h_{q,l}|^2}\right),$$

where u_m denotes the index for the mth cancelled signal. Let

$$\mathcal{W}_m=\{2,3,\ldots,K\}\setminus\{u_1,u_2,\ldots,u_m\}.$$

Then, it follows

$$u_m^*=\arg\max_{k\in\mathcal{W}_{m-1}}\log_2\left(1+\psi_{k,l}\right),$$

where

$$\psi_{k,l}=\frac{\gamma_k|h_{k,l}|^2}{1+(\gamma_1|h_{1,l}|^2+\sum_{q\in\mathcal{W}_{m-1}\setminus k}\gamma_q|h_{q,l}|^2)}$$

and $\mathcal{W}_0=\{2,3,\ldots,K\}$. This SIC continues until the mutual information is less than R or decoding fails. If D interfering signals are cancelled, the resulting instantaneous mutual information for the desired signal is

$$\tilde{I}_{1,l}=\log_2\left(1+\frac{\gamma_1|h_{1,l}|^2}{1+\sum_{q\in\mathcal{W}_D}\gamma_q|h_{q,l}|^2}\right). \tag{4}$$

Note that D depends on the channel realizations, $h_{k,l}$, SNRs, and R. When the SIC is employed, the outage probability with L re-transmissions becomes

$$P_{\text{no}-\text{sic},\text{out}}(R,L)=\Pr(\tilde{S}_{1,L}<R),$$

where $\tilde{S}_{1,L}=\sum_{l=1}^{L}\tilde{I}_{1,l}$.

In general, we can see that

$$P_{\text{no}-\text{sic},\text{out}}(R,L)\leq P_{\text{no},\text{out}}(R,L).$$

That is, the NOMA with SIC can provide a better performance than that without SIC.

Note that the rate, R, would be sufficiently small to facilitate the SIC. For example, suppose that the amplitude of the channel coefficient is Rayleigh distributed and the probability density function (pdf) of the SNR is $f(x)=e^{-x}$, $x\geq 0$. Then, the probability that the SIC can be successfully performed without error, denoted by P_{sic}, becomes $P_{\text{sic}}=\Pr(\log_2(1+x)\geq R)=e^{-(2^R-1)}$. If $R=1$ and $R=0.25$, $P_{\text{sic}}=0.37$ and 0.83, respectively. The lower R, the higher probability the SIC can be performed. Consequently, a smaller R (i.e., $R<1$) is preferred in the NOMA with SIC.

III. Outage Probability Analysis

In this section, we consider the information outage probability to understand the performance of NOMA with/without SIC. We have some technical assumptions for analysis as follows.

A1) $\gamma_k=\gamma$ for all k (As previously assumed that $R_k=R$ for all k, we assume that the system is symmetric in terms of SNR and rate).

A2) The $h_{k,l}$'s are independent and their amplitudes are Rayleigh distributed for all k and l.

In this paper, we mainly present upper bounds on the outage probabilities without details. In the full paper, we provide detailed derivations and lower bounds.

A. Without SIC

The instantaneous mutual information is a function of multiple random variables, which is not easy to analyze. Thus, to make the analysis tractable, we approximate the instantaneous mutual information $I_{1,l}$ in (2) as

$$X_l=\log_2(1+|h_{1,l}|^2\bar{\gamma}_K)\simeq I_{1,l},$$

where $\bar{\gamma}_K=\frac{\gamma}{1+(K-1)\gamma}$, which is obtained by taking the expectation of the power of the $(K-1)$ interfering signals. This approximation would be reasonable as K becomes larger. For convenience, let

$$Y_l=\log(1+|h_{1,l}|^2\bar{\gamma}_K)=X_l\log 2.$$

Then, the outage probability after L (re-) transmissions of IR becomes

$$\begin{aligned}P_{\text{no},\text{out}}(R,L)&=\Pr(X_1+X_2+\ldots+X_L<R)\\&\leq\min_{\lambda\geq 0}e^{\lambda\bar{R}}\left(E\left[\frac{1}{(1+|h|^2\bar{\gamma}_K)^\lambda}\right]\right)^L,\end{aligned} \tag{5}$$

where $\bar{R}=R\log 2$. For the inequality in (5), we use the Chernoff inequality.

Noting that the pdf of $|h_{k,l}|^2$ (from A2)) is

$$f(x)=e^{-x},\ x\geq 0,$$

we have

$$E\left[\frac{1}{(1+|h|^2\bar{\gamma}_K)^\lambda}\right]=\frac{1}{\bar{\gamma}_K}A_\lambda\left(\frac{1}{\bar{\gamma}_K}\right), \tag{6}$$

where

$$A_n(\mu) = \int_0^\infty \frac{e^{-\mu x}}{(1+x)^n} \mathrm{d}x.$$

Then, the upper bound of the outage probability becomes

$$\Pr\left(\sum_{l=1}^{L} X_l < R\right) \le \min_{\lambda \ge 0} \xi(\lambda), \quad (7)$$

where $\xi(\lambda) = e^{\lambda \bar{R}} \left(\frac{1}{\bar{\gamma}_K} A_\lambda \left(\frac{1}{\bar{\gamma}_K}\right)\right)^L$.

Based on the results above, the outage probability is given by

$$P_{\text{no,out}}(R, L) \le e^{-L \log L}, \qquad L \gg 1.$$

Clearly, this asymptotic behavior is independent of R and $\bar{\gamma}_K$.

B. With SIC

In this subsection, we consider the outage probability when the SIC is used. The SIC is an effective means to reduce the interference and improve the performance without joint decoding.

It is possible to obtain a closed-form expresson for the outage probability for the case of K (i.e., the case of two users). However, unfortunately, a general case is not straightforward. Thus, we focus on approximations.

Among $K - 1$ interfering signals, some signals can be cancelled if their instantaneous channel capacity is greater than R. According to (4), D depends on the channel realizations, $h_{k,l}$, R, and γ. For approximation, we assume that D is an independent random variable.

Assuming that D is given, the mutual information can be approximated becomes

$$\begin{aligned} \tilde{I}_{1,l} &= \log_2\left(1 + \frac{\gamma |h_{1,l}|^2}{1 + \gamma \sum_{q=D+1}^{K-1} \alpha_q}\right) \\ &\simeq \hat{I}_{1,l}(D) \\ &= \log_2\left(1 + \frac{\gamma |h_{1,l}|^2}{1 + \gamma \sum_{q=D+1}^{K-1} \bar{\alpha}_q}\right), \end{aligned} \quad (8)$$

where α_q denotes the qth largest $|h_{k,l}|^2$, $k = 2, 3, \ldots, K$, and $\bar{\alpha}_q = E[\alpha_q]$. Taking D an independent random variable, the average mutual information can be an approximate of $\tilde{I}_{1,l}$ in (4) as

$$\tilde{I}_{1,l} \simeq \hat{I}_{1,l} = E_D[\hat{I}_{1,l}(D)]. \quad (9)$$

We can show that $\hat{I}_{1,l}$ is bounded as follows:

$$\begin{aligned} &\log_2\left(1 + \frac{\gamma |h_{1,l}|^2}{E_D[1 + \gamma \sum_{q=D+1}^{K-1} \bar{\alpha}_q]}\right) \le \hat{I}_{1,l} \\ &\le \log_2\left(1 + E_D\left[\frac{\gamma |h_{1,l}|^2}{1 + \gamma \sum_{q=D+1}^{K-1} \bar{\alpha}_q}\right]\right). \end{aligned} \quad (10)$$

For convenience, let

$$\begin{aligned} \gamma_{\mathrm{L}} &= \frac{\gamma}{E_D[1 + \gamma \sum_{q=D+1}^{K-1} \bar{\alpha}_q]} \\ \gamma_{\mathrm{U}} &= E_D\left[\frac{\gamma}{1 + \gamma \sum_{q=D+1}^{K-1} \bar{\alpha}_q}\right]. \end{aligned} \quad (11)$$

If γ_{sic} denotes the effective SNR after SIC, γ_{L} and γ_{U} become a lower and a upper bounds on γ_{sic}, respectively. Using γ_{L} and γ_{U}, we can find upper and lower bounds on the outage probability. From (10), we have the following upper bounds:

$$\begin{aligned} \Pr\left(\sum_{l=1}^{L} \hat{I}_{1,l} < R\right) &\le \Pr\left(\sum_{l=1}^{L} \log_2(1 + \gamma_{\mathrm{L}} |h_{1,l}|^2) < R\right) \\ &\le \min_{\lambda \ge 0} \xi_{\text{sic}}(\lambda), \end{aligned} \quad (12)$$

where the last inequality results from (5) and

$$\xi_{\text{sic}}(\lambda) = e^{\lambda \bar{R}} \left(\frac{1}{\gamma_{\mathrm{L}}} A_\lambda \left(\frac{1}{\gamma_{\mathrm{L}}}\right)\right)^L. \quad (13)$$

Although the derivation is not shown in the paper due to the space limitation, a lower bound of the outage probability can also be found. As shown in Section IV, the lower bound is tighter.

IV. Simulation Results

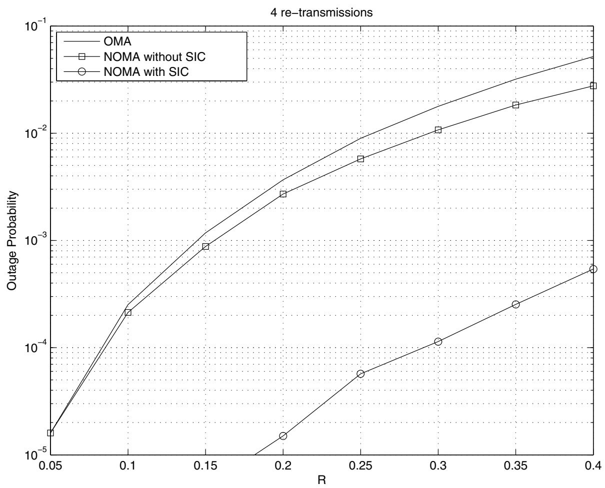
In this section, we present simulation results for the OMA and NOMA with/without SIC. Fig. 2 shows the outage probability with $K = 4$. In general, we can see that the NOMA has lower outage probability than the OMA and the SIC can improve the performance of the NOMA. In addition, we can also confirm that the re-transmission is a very effective means to reduce the outage probability.

In Fig 3, we present the outage probability from simulation results and its upper and lower bounds from the results of Section 3 with $K = 4$. In general, it is shown that the lower bound is tighter than the upper bound for the NOMA without SIC: see Fig. 3 (a). As shown previously, when the SINR is low, the approximate result is tight and it results in a tight lower bound. However, when the SIC is used, the SINR becomes higher and the lower bound is no longer tight. This is shown in Fig. 3 (b). Then, the upper bound becomes useful together with the lower bound to predict the performance of the NOMA with SIC.

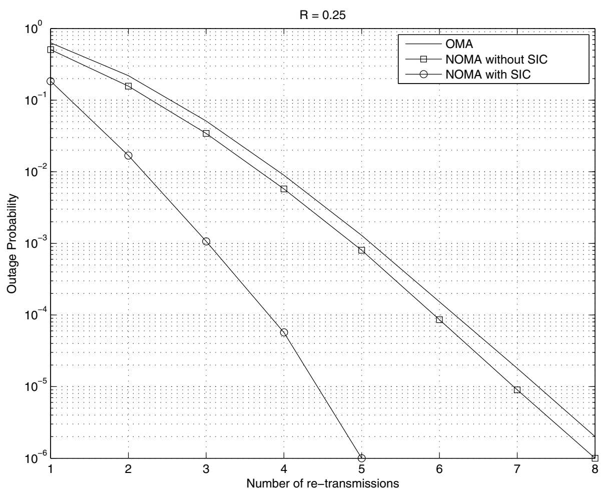
V. Concluding Remarks

In this paper, we studied the NOMA when the H-ARQ with IR is used and derived upper and lower bounds on information outage probability. It was shown that the outage probability decreases with L as $e^{-L \log L}$, which is faster than the exponential decay. In addition, this decay rate with respect to the number of re-transmissions is the same for both H-ARQ schemes with IR and RTD. From the outage probability, it was observed that the NOMA without SIC outperforms the OMA when the SNR is low (less than 0 dB). In addition, when $R < 1$, the NOMA with SIC outperforms the OMA for all SNR.

To facilitate the SIC, a low rate is preferable in the NOMA with SIC. It was shown that the outage probability increases slowly with K for a sufficiently low R. Thus, the throughput can be improved by increasing K with a sufficiently low R in the NOMA with SIC.

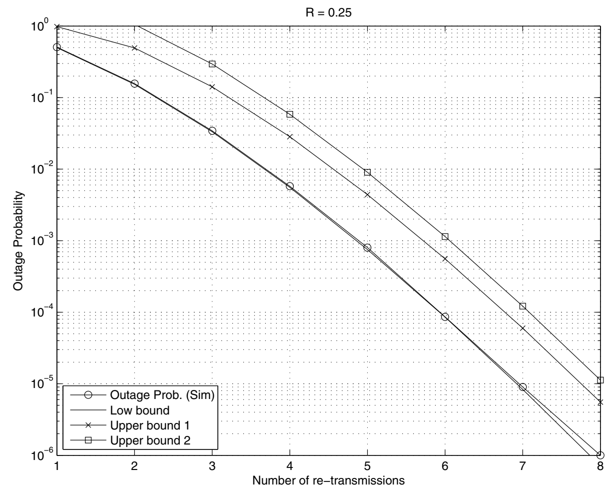


(a)

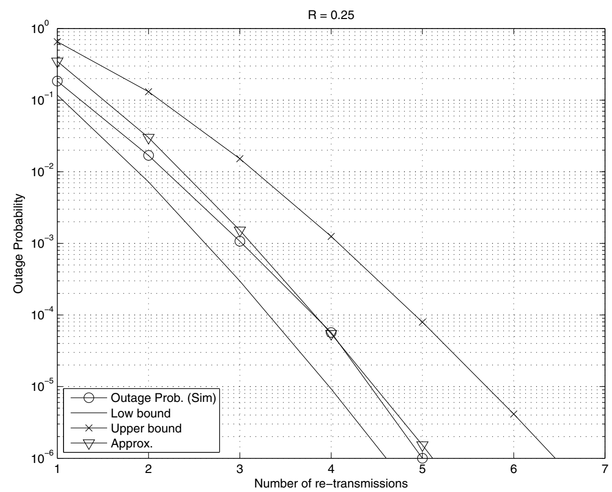


(b)

Fig. 2. Outage probability for OMA and NOMA with/without SIC when $\gamma = 0$ dB: (a) Outage probability versus R; (b) Outage probability versus L.



(a)



(b)

Fig. 3. Outage probability versus the number of re-transmissions when $\gamma = 0$ dB: (a) without SIC; (b) with SIC.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[2] R. G. Gallager, "An inequality on the capacity region of multiple access multipath channels," in *Communications and Cryptography: Two Sides and One Tapestry*, R. E. Blahut, D. J. Costello, U. Maurer, and T. Mittelholzer, Eds, Boston MA: Kluwer, 1994, pp. 129-139.

[3] S. Shamai and S. Verdu, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inform. Theory*, vol.47, pp. 1302-1327, May 2001.

[4] T. Guess and M. K. Varanasi, "An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels," *IEEE Trans. Inform. Theory*, vol.51, pp. 173-187, January 2005.

[5] G. D. Forney Jr., "Shannon meets Wiener II: On MMSE estimation in successive decoding schemes," in *Proc. 42nd Annual Allerton Conf. Communication, Control, and Computing*, Monticello, IL, USA, 2004.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[7] S. Lin and D. Costello, *Error Control Coding, 2nd Ed.*, Prentice Hall, 2004.

[8] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, Prentice Hall, 1995.

[9] G. Caire and D. Tunietti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inform. Theory*, vol.47, pp. 1971-1988, July 2001.

[10] D. Tunietti and G. Caire, "The throughput of some wireless multiaccess systems," *IEEE Trans. Inform. Theory*, vol.48, pp. 2773-2785, October 2002.

[11] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol.44, pp. 2619-2692, October 1998.

[12] L. Ozarow, S. Shamai, and A.D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Vehicular Tech.*, vol.43, pp.359-378, May 1994.

[13] P. Viswanath, D.N.C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol.48, pp.1277-1294, June 2002.

[14] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Comm.*, vol.36, pp. 389-400, April 1988.

[15] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Trans. Comm.*, vol.52, pp. 1311-1321, August 2004.