ARQ: The Gateway from NOMA to NOM

Arafat Al-Dweik $^{1,1}$, Ahmed Iraqi $^2$, and Youssef Iraqi $^2$

$^1$Khalifa University
$^2$Affiliation not available

November 8, 2023

Abstract

This work presents a new framework that utilizes power-domain (PD) nonorthogonal multiple access (NOMA) as a multiplexing scheme to improve the throughput of point-to-point (P2P), or single user, communications. The proposed framework synergizes PD-NOMA and automatic repeat request (ARQ) to enable multiplexing and transmitting multiple packets that belong to the same user simultaneously. To overcome channel estimation and feedback limitations, and to reduce the system complexity, a simple adaptation scheme is proposed to select the appropriate number packets to be transmitted within a given transmission slot. Moreover, the number of transmitted packets is limited to a maximum of two to allow the receiver to blindly identify the number of transmitted packets in a particular transmission slot. The obtained results show that the proposed NOM scheme can eventually double the system throughput at high signal-to-noise ratios (SNRs), and hence, reduce the delay by 50%. The system complexity and overhead are generally comparable to conventional ARQ systems.
ARQ: The Gateway from NOMA to NOM

Arafat Al-Dweik, Senior Member, IEEE, and Youssef Iraqi, Senior Member, IEEE

Abstract—This work presents a new framework that utilizes power-domain (PD) nonorthogonal multiple access (NOMA) as a multiplexing scheme to improve the throughput of point-to-point (P2P), or single user, communications. The proposed framework synergizes PD-NOMA and automatic repeat request (ARQ) to enable multiplexing and transmitting multiple packets that belong to the same user simultaneously. To overcome channel estimation and feedback limitations, and to reduce the system complexity, a simple adaptation scheme is proposed select the appropriate number packets to be transmitted within a given transmission slot. Moreover, the number of transmitted packets is limited to a maximum of two to allow the receiver to blindly identify the number of transmitted packets in a particular transmission slot. The obtained results show that the proposed NOM scheme can eventually double the system throughput at high signal-to-noise ratios (SNRs), and hence, reduce the delay by 50%. The system complexity and overhead are generally comparable to conventional ARQ systems.

Index Terms—Automatic repeat request (ARQ), nonorthogonal multiple access (NOMA), multiplexing, throughput.

I. INTRODUCTION

In packet switched networks, the packet is considered as the basic data unit, and the system should guarantee that each packet is received correctly, to enable the application that utilizes these packets to function properly. Therefore, each packet is generally protected by two defending lines, which are forward error correction coding (FEC) and automatic repeat request (ARQ). At the receiver, the channel decoder attempts to correct the errors in the received packet. If the decoder fails, ARQ is invoked where the transmitter will be asked to resend the erroneous packet. However, FEC and ARQ are very costly in terms of power and spectral efficiency. Therefore, the system power efficiency and throughput can be severely degraded [1]–[3]. The combined FEC and ARQ is typically called hybrid ARQ (HARQ).

A. HARQ Overview

The typical HARQ process consists of four main operations, FEC, error detection, combining, and retransmission. At the transmitter, HARQ systems generally follow the structure described in [1]:

- Each sequence is then applied to a channel encoder that appends \( l_{ec} \) bits to form a codeword.
- The channel encoder output is interleaved to mitigate the effect of burst errors.
- The interleaver output is modulated using a particular modulation scheme.
- To form a packet, one or more modulated sequences are grouped, then the packet header is appended.

At the receiver, the typical processing starts with demodulation, deinterleaving, FEC, error detection, and finally retransmission if the packet still has errors after FEC decoding. In such a case, the receiver sends a negative acknowledgment to inform the transmitter to resend the packet. This approach is called stop-and-wait ARQ. The receiver can combine the initial and retransmitted packets to improve the packet signal-to-noise ratio (SNR). An ARQ with packet combining is known as Type-I ARQ. The retransmission process is terminated when the packet is received correctly, or when the maximum number of retransmissions is reached, and in this case the packet will be dropped.

B. NOMA Overview

Power domain (PD) nonorthogonal multiple access (NOMA) has been considered extensively in the literature due to its promising spectral efficiency. The main concept of NOMA is that users who are far from the base-station require more power as compared to near users [4], [5]. Therefore, it would be possible to pair certain users based on their channel conditions, and assign them the same frequency and time resources, but different powers. Fig. 1 shows a simplified diagram of a two user NOMA system. At the receiver, the user with the maximum power can recover its own data symbols while considering the multiuser interference as unknown additive noise. For other users, successive interference cancellation (SIC) can be applied by each user to extract its own data. The power assignment for each NOMA user is essential to minimize the error probability of each user.

C. Related Work

PD-NOMA has received enormous attention by the research community due to its potential to improve the spectral efficiency of various wireless networks. The work reported in the literature covers a broad range of topics such as bit error rate (BER) [4], [5], outage and capacity analysis. NOMA is also considered for cooperative communications, cognitive radio, multiple-input multiple-output (MIMO), visible light communications (VLC), free space optics (FSO) and intelligent reflecting surfaces (IRS).
In addition to the aforementioned applications, NOMA has been considered with ARQ to satisfy certain QoS requirements. For example, Cai et al. [6] investigated the outage performance of ARQ assisted NOMA systems and studied the trade-off between the minimum number of retransmissions and the transmit power allocation. Cao et al. [7] used ARQ to assist short-packet communications for NOMA networks. The work focuses on deriving analytical expressions for the average packet error rate and effective throughput over Nakagami-$m$ channels. Moreover, the article discusses the optimal power allocation and packet length. Yu et al. [8] considered a NOMA relaying protocol with truncated ARQ. The error detection and retransmission are performed at both the relay and destination. The relay is considered as the near user and hence it is allocated less power as compared to the destination. The article discusses several retransmission strategies, derives the system throughput and investigates the impact of power allocation, target rate, and distance between terminals. The work in [8] was extended in [9], [10] to include the relay selection problem.

Secure transmission scheme using ARQ assisted cognitive NOMA networks is proposed in [11] where a security-required user is paired with QoS-sensitive user to perform NOMA. The ARQ is used to mitigate the successive interference cancellation SIC errors and improve the secrecy performance. The authors in [12] considered the combination between NOMA and ARQ for realizing ultra-reliability and high throughput system. The article presents the diversity order analysis of various NOMA-ARQ systems including Type-I ARQ, ARQ with chase combining and ARQ with incremental redundancy. Furthermore, power-efficient transmission with imperfect CSI is also considered.

D. Motivation and Contribution

As can be noted from the surveyed literature and the references listed therein, the transmission power per packet can actually be much larger what is needed to achieve a certain packet error rate (PER), particularly if the packet is retransmitted multiple times. Although the power of the first and each subsequent retransmission can be optimized, throughput reduction caused by the retransmission process will still be an issue. Inspired by NOMA, we propose in this article a protocol that combines ARQ with nonorthogonal multiplexing (NOM) such that whenever a particular transmission does not require the full power, the residual power is used to transmit another packet simultaneously using NOMA. Consequently, the throughput can be significantly increased. Moreover, to reduce the system complexity and signaling overhead, we propose a heuristic protocol that utilizes fixed power transmission. The obtained results demonstrate that the proposed protocol offers about 100% throughput improvement over conventional ARQ systems.

II. System Model

In this article we consider stop-and-wait Type-I ARQ with turbo product codes (TPC). The channel is assumed to be a slow block Rayleigh fading channel, and the transmitter has knowledge of the channel state information (CSI) for the packet to be transmitted. The CSI is typically available at the transmitter side through a feedback channel, or using the channel reciprocity concept. In block fading channels, all symbols in the transmitted packet undergo the same fading. However, the fading from one packet to the other may vary. The error detection process is assumed to be perfect, which is typical in practical ARQ systems given that sufficient cyclic redundancy check (CRC) bits are used [2]. The system quality of service (QoS) requirement is the PER, which should be less than a certain threshold $PER_{th}$. The packets to be transmitted are called $A, B, C, \ldots$, and the subscript corresponds to the transmission count. For example, $A_2$ indicates the second transmission of packet $A$. The transmitter maximum transmit power is denoted as $P_{max}$. The NOM is generally similar to NOMA, except that the multiplexed packets belong to the same user, as shown in the following discussions.

A. Proposed ARQ-NOM

The proposed ARQ-NOM can be described as follows:
Based on the channel condition, the transmitter selects the transmission power for packet $A_1$, such that \( \text{PER}[A_1] = \text{PER}_{th} \). In this case the power will be denoted as $P_{A_1}$.

If $P_{A_1} \leq P_{\text{max}}$, then another packet, $B_1$, can be transmitted using NOM, with power $P_{B_1} \leq P_{\text{max}} - P_{A_1}$. The value of $P_{B_1}$ is selected such that \( \text{PER}[B_1] = \text{PER}_{th} \). If the available power for $B_1$ is not sufficient to satisfy the PER condition, we set $P_{B_1} = P_{\text{max}} - P_{A_1}$, and then stop. Otherwise, the same process is repeated for packet $C_1$, and so forth until all the transmission power is allocated. Consequently, $P_{A_1} > P_{B_1} > \cdots$.

When multiple packets will be transmitted, the PER should be computed while considering the multiuser interference [4], [5].

At the receiver, packet $A_1$ will be demodulated while considering that the interference from other packets is unknown additive noise. Then the packet is applied to the FEC decoder. The FEC decoder output is then used by the SIC to detect packet $B_1$, then $A_1$ and $B_1$ are used by the SIC to detect $C_1$, etc. Finally, all packets are checked for errors, and the receiver sends a feedback vector that contains the Ack/Nak for each packet. For example, if the transmitter sent only packets $A_1$ and $B_1$, then the feedback vector will be one of the following set: \{Ack,Ack\}, \{Ack,Nak\}, \{Nak,Ack\} or \{Nak,Nak\} to indicate the status of the detection process of the two packets.

Based on the feedback vector, the transmitter retransmits the packets with Nak, and may include new packets if the channel conditions allow. When a particular packet is retransmitted, the retransmission power is assigned while considering the previously transmitted packets as Chase combining can reduce the power requirements for the retransmitted packets. In such scenarios we may obtain $P_{B_1} > P_{A_1}$. Moreover, the PER and power allocation should be computed while considering the Chase combining impact on the PER.

packets that did not pass the error check process can be stored at the receiver, and then Chase combining can be used [1] to combine all copies of such packets.

Various packet drop strategies can be applied. For example, a packet is dropped only when it consumes the same power that it would consume in a conventional system. For example, assume that a conventional ARQ system allows a maximum of 4 transmissions, and the power for each transmission is normalized to unity, then the total power is 4. Consequently, a packet might be transmitted using the proposed system over six transmissions each of which has a power of 2/3. In other words, a packet is dropped only when it consumes all the power budgeted to it. It is also possible to design the power allocation process while considering other packet drop criteria.

### B. Packet Combining

Chase combining of the multiple received copies of a packet is essential to improve the system throughput [1]. If the channel gains over the combined packets are independent, then the diversity order in conventional ARQ is equal to the number of the combined packets. Although Chase combining can be also adopted in the proposed ARQ-NOM, there are certain modifications that are necessary due to the fact that each transmission may have multiple packets with mutual interference, inter-packet interference (IPI). Therefore, the following procedure can be adopted for Chase combining:

1) If a packet is transmitted more than once, and it was allocated the highest power as compared to other packets, then such packets are combined using conventional Chase combining by assuming that the IPI is unknown additive noise. Therefore, maximum ratio combining (MRC) can be applied to all symbols in the combined packets.

2) After detecting packets with the largest power, we apply the SIC to extract the packets with next highest power and combine them. For example, assume that the receiver has transmitted packets $A$ and $B$ three times, so the receiver has in the buffer $[A_1,B_1],[A_2,B_2],[A_3,B_3]$. In this case, the receiver applies Chase combining to $\{A_1,A_2,A_3\}$, and detects packet $A$. Then, packet $A$ is encoded, modulated, scaled by the channel gain and transmission power, and finally subtracted from the three buffered transmissions. Then, the three buffered sequences correspond to $B_1$, $B_2$, and $B_3$, which now can be applied to the Chase combiner after the IPI has been eliminated $B$.

3) An important case to consider is when a packet is dropped. As an example, assume that packet $A$ has consumed its power budget after three transmissions and it will be dropped. Moreover, based on the channel conditions, only two packets can be multiplexed at the transmitter. Packet $B$ has also failed three times, but still did not use its power budget. Consequently, the receiver buffer would have packets $[A_1,B_1],[A_2,B_2],[A_3,B_3]$. In this case $A$ will be dropped, and the next transmission will $[B_1,C_1]$. For this case, it could be better not to combine $B_1,\ldots,B_4$ because $B_1, B_2, B_3$ are contaminated with high IPI because the SIC process was mostly unsuccessful, and $A$ can not be subtracted from the composite packet correctly.

4) In step 3, packet $B_4$ might be detected successfully, which implies that the IPI caused by $B$ on $A$ can be eliminated, which improves the chances of $A$ being detected correctly. Therefore, if the system can tolerate some delay, we propose to introduce an additional step before completely dropping a packet, which is called the hold state. In the hold state, the packet that exceeds the number of allowed transmissions is put into hold, stored at the receiver, until the decision with respect to packet $B$ is finalized. In fact, any successfully packet can help detecting previous erroneous packets. This feature is invaluable for particular emerging applications. For example, consider the case of surveillance unmanned aerial vehicles (UAVs) that are used to stream videos and photos on the move. In such cases, the number of
retransmissions should be small to allow live streaming of the captured videos and photos, and as a consequence, the quality can’t be guaranteed. Therefore, the live streaming can be used as an initial step to by the UAVs’ pilots, and higher quality videos and photos can be obtained later offline.

C. Design Tuning

Although the design presented in Sec. II-A is attractive because it reduces the transmission power per packet and improves the throughput while satisfying the PER constraint, some design tuning can be performed to reduce the system complexity and simplify its realization. By noting that NOM is generally similar to NOMA where the signal-to-interference plus noise ratio (SINR) is inversely proportional to the number of multiplexed packets, then in most cases only two packets can be multiplexed at the transmitter. Therefore, the number of packets that can be multiplexed at the transmitter will be limited to a maximum of two packets.

The availability of accurate CSI at the transmitter can be challenging, particularly in mobile systems because it requires substantial signaling overhead. Therefore, power assignment based on the channel statistical information, or even without any knowledge about the channel would be highly desirable given that it would not severely degrade the system performance. Moreover, packet combining complexity at the receiver can become highly complex in certain scenarios because packets without SIC have to be combined with packets after the SIC. Therefore, we adopt fixed power assignment and show that the system can perform generally well using fixed power assignment. Moreover, the packet with the minimum index should be given higher priority until it is successfully received, or dropped.

To reduce the buffering requirements at the receiver, and noting that if the packet with the highest power was not decoded successfully, then the SIC will mostly be unsuccessful, and combining the packets with low power might be of limited benefit. Therefore, if a packet is dropped, then all other packets multiplexed with that packet can be dropped. Consequently, the packet drop criterion can be defined such that the transmission counter is set only for the main packet.

Based on the considered design tuning, the ARQ-NOM can be summarized as:

1) Two packets are always multiplexed in each transmission regardless of the channel conditions. The primary packet, i.e. the packet with the minimum index will be assigned more power, therefore $P_{A_1} > P_{B_1}$.

2) The primary packet will be always assigned more power than the secondary packet, even when retransmission is needed. For example, if packet $A$ is to be transmitted then the power allocated to packet $A$ in the first or consequent transmissions should be more than the power allocated to any other packet.

3) All copies of the primary packet can be combined while considering the IPI as unknown additive noise.

4) The maximum number of transmissions can be set to a particular number only for the primary packet.

5) If the primary packet is dropped, all secondary packets are deleted from the receiver buffer.

Fig. 2 depicts the state diagram of the transmitter, where a state is defined as the transmitter output. The figure shows all possible scenarios for transmitting three packets \{A, B, C\}. The maximum number of transmissions for each packet is two. Each state is represented by circle, and the packet/packets in each circle are labeled with subscript that indicates the transmission number, and a superscript that indicates the number of transmissions for that packet while it is the primary packet. For example $B_1^0$ indicates that packet $B$ has been transmitted two times as a secondary packet, and zero times as a primary packet. The transmission process starts from the green state $A_1^1B_1^1$, and it is terminated at state $E$. The transitions between states are labeled with the receiver feedback where 1 indicates an Ack and 0 indicates a Nack. As can be noted from the figure, the minimum number of transmissions required to receive the three packets correctly is two, and the maximum is six. In conventional ARQ, the minimum is three and the maximum is six transmissions. Consequently, the throughput gain is expected to be higher for good channel conditions.

III. Simulation Results

Thus section presents the Monte Carlo simulation results of the proposed ARQ-NOM system. In each simulation run $10^5$ transmissions are generated. The binary bits are encoded using turbo product codes (TPC) where the component codes are extended Bose, Chaudhuri, and Hocquenghem (eBCH). Each TPC code is denoted as $\text{eBCH}(n,k,d_{\text{min}})$ where $n$ is the component codeword length, $k$ is the number of information bits per component codeword, and $d_{\text{min}}$ is the minimum
Hamming distance. The information bits are modulated using binary phase shift keying (BPSK) and then multiplexed following the NOM principle where the primary packet is allocated power $\alpha_1$ and the secondary packet is allocated power $\alpha_2 = 1 - \alpha_1$, $\alpha_1 > \alpha_2$. The maximum number of transmissions allowed is 5. The channel is modeled as a block Rayleigh fading where the fading block is equal to codeword length that is $n^2$.

At the received side, based on the SIC principle, the primary packet can be detected directly while ignoring the IPI. Therefore a zero-forcing equalized followed by a hard decision BPSK demodulator is used to detect the binary bits in the primary packet. Then the bits are fed to a hard-input hard-output turbo decoder [13]. The packet is then checked for errors, if it is error free, it is applied to the SIC receiver to extract the bits of the secondary packet following the same approach applied to the primary packet. Finally, the feedback is sent to the transmitter to report which packet was received correctly. Multiple packets are combined using Chase combining as described in Packet combining is applied as described in Sec. II.

The results are presented for the throughput, normalized delay, and dropped packets rate. The throughput is defined as the ratio of the total number of information bits received correctly to the total number of transmitted channel symbols. The normalized delay is defined as the average number of transmissions for each packet. And finally, the packet drop rate is the average ratio of packets dropped to the total number of transmitted packets.

Fig. 3. Throughput of the ARQ-NOM and conventional ARQ for TPC eBCH(32, 21, 6)\(^2\) using $\alpha_1 = 0.7, 0.8, 0.9$.

Fig. 3 shows the throughput of the ARQ-NOM using the eBCH(32, 21, 6)\(^2\) TPC, which has a code rate of about 0.43. The results are presented using three different power values. As can be noted from the figure, the proposed ARQ-NOM may offer up to 100% throughput gain at moderate and high signal to noise ratios (SNRs) as compared to conventional ARQ. At low SNRs, conventional ARQ outperforms the ARQ-NOM because the IPI and power sharing increases the packet error rate, which reduces the throughput. Such behavior can be also noted by varying the values of $\alpha_1$ where the figure shows that at low SNRs, increasing the value of $\alpha_1$ improves the throughput. Therefore, the values of $\alpha_1$ should be adaptively chosen to maximize the throughput. Nevertheless, using fixed values of $\alpha_1$ in the absence of SNR information at the transmitter is still an effective approach because the throughput loss in the low SNRs region is small while a significant gain can be achieved at moderate and high SNRs. If the average SNR is known at the receiver a simple switching between the ARQ and ARQ-NOM, denoted as adaptive ARQ-NOM, would provide the best performance given that $\alpha_1 = 0.8$.

Fig. 4 is similar to Fig. 3 except that the high code rate TPC eBCH(32, 26, 4)\(^2\) is used. The packets’ length in both cases is similar, but the code rate for this code is 0.66. As can be noted, both codes provide generally similar throughput trends, except that throughput becomes more sensitive to the values of $\alpha_1$. Moreover, the switch-over between the ARQ-NOM and ARQ is shifted to the right by about 4 dB, which implies that the error correction capability of the code may provide an advantage for the ARQ-NOM system.

Fig. 4 shows the throughput of the ARQ-NOM and conventional ARQ for TPC eBCH(32, 26, 4)\(^2\) using $\alpha_1 = 0.7, 0.8, 0.9$.
probability of retransmission. Nevertheless, at moderate and high SNRs the ARQ-NOM drops to 50% of the ARQ. Such performance is obtained because the ARQ-NOM effectively needs half of the time that the ARQ requires to transmit the same number of packets. The figure also shows the adaptive case which switches between the ARQ and ARQ-NOM.

Fig. 6. Percentage of dropped packets.

Fig. 6 shows the packet drop rate of the ARQ and ARQ-NOM. The results in the figure show that IPI and power sharing has significant effect on the packet drop rate at low SNRs, while it converges to the ARQ at moderate and high SNRs. Using a powerful code show that the drop packet can be improved at low SNRs for both systems.

IV. Conclusions and Future Work

This work presented a new approach to apply the NOMA concept to single user applications where it is denoted as NOM because it is used for multiplexing instead of multiple access. The proposed technique exploits the excess power available at the transmitter to multiplex and send additional packets simultaneously. Consequently, the system throughput can be improved while maintaining the QoS requirements satisfied. To simplify the system realization, a low complexity system was designed, which is highly efficient in the absence of CSI at the transmitter side. If the SNR is available at the transmitter, then a simple adaptation process is applied where the system may switch between the conventional ARQ and ARQ-NOM. The obtained results show how the proposed ARQ-NOM can increase the system throughput by 100% at high SNRs.

Although the proposed low complexity NOM system managed to provide substantial throughput gain, the main goal of this work was to present a proof of concept rather than presenting a holistic framework design and performance evaluation. Therefore, there are several open problems that should be considered to evaluate the performance limits of the proposed system. For example, the power optimization should be considered for the cases where full and statistical CSI are available at the transmitter. In such cases, a particular packet might be considered as the primary packet at one transmission, and a secondary in the next transmission. Consequently, an optimal packet combining rule should be derived to consider packets with perfect and imperfect SIC. The system performance should be also evaluated while considering the hold state. Evaluating the system performance using soft decision decoding and deriving the log likelihood ratios (LLRs) with successful and unsuccessful SIC for the proposed system should be also considered.

References


**Arafat Al-Dweik** (S’97–M’01–SM’04) received the M.S. degree (summa cum laude) and Ph.D. degree (magna cum laude) in electrical engineering from Cleveland State University, Cleveland, OH, USA. He served with Efficient Channel Coding, Inc., Cleveland, USA, the Department of Information Technology, Arab American University in Palestine and the University of Guelph, ON, Canada. Since 2003, he has been with the Department of Electrical Engineering, Khalifa University in UAE.

He is currently an Adjunct Research Professor with Western University and the University of Guelph in Canada, and University of Manchester in UK. Dr. Al-Dweik is a recipient of the Fulbright Scholarship, the Hijjawi Award for Applied Sciences, the Fulbright Alumni Development Grant, Dubai Award for Sustainable Transportation and the Leader-Founder Award in UAE. He serves as an Editor for the IEEE Transactions on Vehicular Technology and IET Communications. He is a registered Professional Engineer in the Province of Ontario, Canada.

**Youssef Iraqi** (M’03, SM’10) is currently an Associate Professor with the Electrical Engineering and Computer Science Department at Khalifa University, UAE. Before that, Dr. Iraqi was the chair of the Computer Science Department at Dhofar University, Oman for four years. From 2004 to 2005 he was a Research Assistant Professor in the David R. Cheriton School of Computer Science at the University of Waterloo, Canada. Dr. Iraqi has published more than 110 research papers in international journals and refereed conference proceedings. His research interests include adaptive resource management in wireless networks, trust and reputation management, cloud computing, and stylometry. In 2008, he received the IEEE Communications Society Fred W. Ellersick Paper Award in Communications Systems. Dr. Iraqi is on many technical program committees of international conferences and always approached for his expertise by international journals in his field. Dr. Iraqi is a senior member of the IEEE.