A New Adaptive Rate IR-HARQ Combining with AMC

A. Alkhoder, A. Assimi and M. Alhariri

Abstract—In this paper, we analyze a new cross layer combining scheme of adaptive modulation and coding (AMC) at the physical layer and hybrid automatic repeat request (HARQ) protocol at the data link layer. Our purpose is to improve the performance of the communication system in terms of throughput and errors. We propose a new adaptive HARQ protocol based on the estimated mutual information in addition to signal to noise ratio as channel state information (CSI), which are unavailable at the receiver. Unlike the traditional protocols, the transmitter is aware of the transmission mode only, not the instantaneous CSI, which minimize the load on the feedback channel. The transmission is supposed to be done over block fading channel. Mathematical formulation is improved for the performance criterions, in terms of throughput of the new HARQ, average number of transmissions and the spectral efficiency of the AMC/HARQ. The comparison with the traditional variable rate HARQ shows that our new adaptive HARQ protocol provides an upper bound for the throughput. In addition, the simulation results show that a small number of transmissions is enough to reach the maximum spectral efficiency while maintaining the error probabilities under certain acceptable level.

Index Terms—Adaptive modulation and coding (AMC), Adaptive Rate Incremental Redundancy Hybrid Automatic Repeat reQuest ARIR-HARQ protocol, cross-layer design, Mutual Information (MI), spectral efficiency.

I. INTRODUCTION

In modern wireless communication systems, much efforts have been made in order to overcome the effects of channel fading and enhancing system performance in term of spectral efficiency under certain services constraints such as Packet Error Rate (PER) and transmission delay, a cross-layer combining is performed between Adaptive Modulation and Coding (AMC) at the physical layer and Hybrid Automatic Repeat reQuest (HARQ) at the data link layer [1], [2] and [3].

Manuscript received October 14, 2018; revised November 17, 2018. The associate editor Prof. Đinko Begušić has been coordinating the review of this manuscript and approved it for publication.

Authors are with the Department of Telecommunication, Higher Institute of Applied Science and Technology, Damascus, Syria.

E-mails: assal.khoder@hiast.edu.sy, abdelnasser.assimi@hiast.edu.sy, hariri66@mail.ru.

Digital Object Identifier (DOI): 10.24138/jcomss.v14i4.642

The most common technique in performing AMC is to simply adapt the transport format (coding rate and modulation mode) according to the Channel State information (CSI). In HARQ protocols, additional coded bits are retransmitted upon the reception of a negative acknowledgment (NACK) from the receiver. The receiver, in turn, combines these bits with the previously received coded bits according to the used HARQ scheme in order to decode the packet [4] [5]. In AMC/HARQ combining, the adaptation of the transport format and the number of transmitted bits is performed jointly depending on CSI and the feedback from the receiver.

To enhance the throughput in the system, the transmitter tries to minimize the channel uses or the amount of transmitted bits necessary to the correct decoding of the packet at the receiver. This observation provides a basis for designing a spectrally efficient AMC/HARQ scheme [5], [28] and [29]. A big effort is made to adapt the number of retransmitted bit to CSI. An adaptive incremental redundancy (AIR) scheme is proposed in [5]. In order to guarantee correct decoding of the combined transmissions, while minimizing channel uses, authors in [5] propose a new scheme that adapts the transmission format according to feedback information. With AIR, the transport format of each retransmission is optimized based on CSI during the current and previous transmissions. In [6], the achievable throughput of incremental redundancy (IR) HARQ over block fading channels, is analyzed. The transmission rate is updated based on CSI experienced by the receiver in the past transmissions of a failed decoded packet. This work concludes that, even if CSI is fully outdated, the adaptive scheme outperforms a non-adaptive HARQ. Furthermore, a few transmissions are enough to approach closely the ergodic capacity of the channel at high SNR. In [7], the idea of using adaptive IR was analyzed within the framework of information theory and upper bounds on system performance are given.

The average SNR per packet was one of the most used measures of CSI to quantify the link quality in the system. In practical implementation this measure is subject to estimation error and uncertainty. This leads the researcher to find a more robust measure for the quality of the received packets. The estimated Mutual Information (MI) was found to be on candidate for this purpose [5] [6] [8]. In these works, the MI is mainly used as a measure for the correct decoding at the
packet.

In this paper, we propose a new AMC/HARQ combining technique based on both of SNR and MI as link quality measures to perform adaptive retransmission scheme in order to enhance spectral efficiency under the constraint of a given packet error rate. We focus in this work on the use of the AIR-HARQ scheme [6] combined with AMC. The targeted services are assumed to be constrained to PER and delay. To limit the delay in our system, we estimate the amount of transmitted information needed by the receiver in order to decode the packet correctly from the first transmission attempt. In the case of failure, we re-estimate the additional information to be sent at the next retransmission. Which means, adapting the redundancy ratio to CSI. We name our proposed technique by Adaptive Rate Incremental Redundancy (ARIR) HARQ (ARIR-HARQ). The obtained results are shown in term of: a) system throughput, b) outage probability, c) average number of transmission attempts, and d) spectral efficiency of the AMC/HARQ combining scheme.

In section 2, we present the system model and assumptions for the system, including the communication system with the channel model and the structure of the frames and the packets. The proposed scheme of combining AMC and HARQ is presented in 3, where the design requirements, at the physical layer in 3.1 and the data link layer in 3.2. In section 4, the performance of the proposed ARIR-HARQ protocol is analyzed in terms of throughput in 4.1, throughput bounds in 4.2 and the average number of transmission in 4.3. The performance of AMC/ARIR-HARQ is analyzed in section 5 in terms of spectral efficiency, errors and average number of transmission. Parameters adopted in the simulation and the numerical results of the simulation are presented in section 6, including PER, BER, spectral efficiency and the average number of transmission. Parameters adopted in the simulation and the numerical results of the simulation are presented in section 6, including PER, BER, spectral efficiency and the average number of transmission in addition to changes in average values of redundancy ratio and coding rate with the the maximum number of transmission. Finally, we summarize the conclusion of our study in section 7.

II. SYSTEM MODEL AND ASSUMPTIONS

A point-to-point SISO wireless packet communication system is considered as shown in Fig. 1. It consists of a joint AMC module at the physical layer and HARQ module at the data link layer in a cross layer combining structure. Multiple transmission modes are assumed in the system at the physical layer. Each mode consists of a specific Modulation and Coding Scheme (MCS) as in [6]. The AMC/HARQ controller determines the AMC scheme based on CSI estimated at the receiver. Where, CSI in our proposal is the SNR and MI values.

The considered flat fading channel can be modeled by means of the received SNR $\gamma$. We adopt the Nakagami-$m$ distribution to describe $\gamma$ statistically, because it applies to a large class of fading channels [9]. It includes Rayleigh channel as a special case when $m = 1$, Nakagami-$n$ (Rician) fading channel for $m > 1$. Using one-to-one mapping between the Ricean factor $n$ and the Nakagami fading parameter $m$, Nakagami-$m$ channel will approximate Nakagami-$n$ channels [9].

In addition to SNR, we also consider MI between the transmitted and received bits as a metric to describe the link quality. MI can be estimated using the log likelihood ratios (LLRs) of received symbols [10], [11], [12]. Considering the SISO channel, the relationship between the input and output messages can be modeled as follow [13]:

$$y_l = h_l x_l + z_l,$$

where $x$ is the input symbol, $y$ is the output symbol, $h$ is the channel gain which follows Nakagami distribution, $z$ is an additive white Gaussian noise and $l$ is the sub-carrier index in the OFDM configuration, where $L$ is the number of sub-carriers.

In our new scheme, the SNR range is divided to $k$ regions, to each region, a modulation mode and an initial coding rate are associated. For practical reasons, the initial coding rate is chosen from a set of quantized values. In other words, initial MCS schemes are associated with the SNR regions, and the SNR switching thresholds must be determined between these regions. The proposed method to determine these thresholds between MCSs is described in our previous publication in [14]. SNR regions and thresholds are obtained as in [14], and they will be tabulated in section 6.

The principle of combining AMC and HARQ is well described in the literature as in [1], [3], [13], [15], [16]. The maximum number of HARQ transmissions $N$ of one packet could be determined by dividing the maximum allowable system delay (afforded in practice) over the round trip delay required for each transmission [1], [2]. The coding rate used in HARQ protocol, and by consequence the amount of transmitted redundancy, is adaptive [4], [16], [28] and [29]. In this work, we propose a new method to choose the coding rate depending on the estimated values of MI and SNR.

At the physical layer, the transmission is performed on the base of frame by frame, where each frame contains a fixed number ($N_f$) of M-QAM symbols to construct an OFDM symbol. Each packet in the data link layer is encoded by a mothercode giving a codeword of $N_c$ bits $b_1, b_2, ... b_{N_c}$ and
divided into sub-codewords \(x_1, x_2, ..., x_{N_c}\) (of \(N_1, N_2, ..., N_{N_c}\) bits, respectively). In this work, we assume a perfect coding and \(x_1\) is the information bits, while each of \(x_i\) (\(i = 2, 3, ..., N_c\)) is a redundancy block which has an adaptive length depending on the estimated SNR or MI value. These blocks are chosen according to a given puncturing.

In our new proposed model, the AMC/HARQ controller sends back the MCS selection to the transmitter through an error free feedback channel. The transmission mode is then updated by the AMC/HARQ controller at the transmitter. Soft decoding is used at the receiver, and HARQ protocol with AMC is used. This means, after the first transmission of a sub-codeword consisting of 

\[N_1\] information bits with \(N_2\) redundancy bits. This means that the initial coding rate will be \(N_1/(N_1 + N_2)\). If the decoding of that packet fails, the receiver determines the amount of required additional redundancy (\(N_3\) length) for correct decoding based on the estimated MI of the previous received message, and sends it back as a NACK signaling to the transmitter. The transmitter, in turn, sends the required additional redundancy. Then, the receiver combines the original received sub-codeword with the additional redundancy, and tries to decode it. This operation continues until the correct decoding of the packet and ACK signaling is sent back to the transmitter, or the maximum number of transmissions \(N\) is achieved. Therefore, if \(n\) transmissions are performed per packet, the coding rate will be \(N_1/\sum_{i=1}^{n} N_i\). ACK/NACK signaling, in our design, is a multi-bit message carrying the indices of required redundancy blocks and the modulation index as in [29].

Each frame at the physical layer contains multiple messages (\(m\) messages) from the data link layer (variable number of sub-codewords depending on AMC scheme). It should be noted that the frame may contain empty parts, and we suppose (as in [6] and [17]) that these parts are small in comparison with the frame length, especially for a high number of packets in the frame. Therefore, the loss in the throughput caused by these empty parts could be negligible [6] [17].

III. PROPOSED SCHEME

In this section, we explain our proposed model for AMC/ARIR-HARQ. We first determine the required performance at the physical layer in term of the achieved average PER (\(P_0\)) to meet the PER target (\(P_{\text{target}}\)), given that, N- HARQ is implemented at data link layer. Then, we determine the required performance at the data link layer in term of the coding rate or redundancy ratio at each transmission attempt to meet the PER goal (\(P_{\text{out}}\)).

Assuming perfect channel estimation, we adopt the information outage probability as the basis of performance criteria as in [18], [19]:

\[P_{\text{out}} = \Pr(\text{NACMI}(\gamma) < 1), \tag{2}\]

where \(\text{NACMI}(\gamma)\) is the normalized accumulated mutual information conditioned on \(\gamma\). M-ary symbols are assumed to have equal probabilities. The \(\text{NACMI}\) can hence be calculated, for IR-HARQ, at the \(N\)-th transmission attempt as [20]:

\[\text{NACMI}(\gamma) = \sum_{n=1}^{N} \frac{N_n}{N} M I(\gamma_n), \tag{3}\]

where \(\gamma_n\) is the estimated SNR at the \(n\)-th transmission attempt. It’s clear from (3) that \(\text{NACMI}\), in the IR scheme, is accumulated directly from the MI values of the \(N\) transmission attempts, where each one of MI values depends on the estimated SNR at the corresponding transmission attempt. Therefore, \(\text{NACMI}\) in the case of IR varies depending on the distribution of MI, and it is necessary to use the entire SNR history to estimate it accurately.

A. Physical Layer

Our problem is to design AMC (finding switching thresholds) to maximize the spectral efficiency while ensuring that an average PER (\(P_0\)) is satisfied. Suppose that a packet is dropped if it is not correctly decoded after \(N\) transmissions (\(N-1\) retransmission). Hence, the packet loss probability at the data link layer can be computed as \(P_0^N\). To satisfy the PER goal (\(P_{\text{out}}\)) at the data link layer, that is, the probability of losing the packet at the end of the HARQ process, we need to impose: \(P_0^N \leq P_{\text{out}}\). Thus, the packet loss will be upper bounded by a target PER (\(P_{\text{target}}\)) at the physical layer, i.e:

\[P_0 \leq P_{\text{out}}^{1/N} := P_{\text{target}}, \tag{4}\]

As in [21], we assume constant transmission power, and partition the total SNR range into \(K+1\) non-overlapping consecutive regions, with boundary points (SNR switching thresholds) denoted as \(\{\gamma^{(k)}\}_{k=0}^{K+1}\), where \(K\) is the total number of initial MCSs (modulation mode and initial coding rate at the first transmission of a packet). Specifically, the \(k\)-th MCS is used when \(\gamma \in [\gamma^{(k)}, \gamma^{(k+1)}]\). The procedure of finding \(\{\gamma^{(k)}\}_{k=0}^{K+1}\) is described in our previous work in [14]. The same procedure is used here to get the SNR switching thresholds values for different MCSs as mentioned above.

B. Data Link Layer

In order to determine the coding rate for a new transmission, the estimated SNR value is used at the first transmission of a packet. From the normalized accumulated mutual information \(\text{NACMI}\) in (3), the coding rate at each transmission attempt can be found, when the condition (2) is satisfied, for incremental redundancy as:

\[\frac{N_1}{\sum_{i=1}^{n} N_i} = \frac{1}{1+\sum_{i=1}^{n} P_i} \leq \frac{1}{\sum_{i=1}^{n} N_i} \sum_{i=1}^{n} M I_i N_i, \tag{5}\]

where \(P_i = N_i/N_1\) is the redundancy ratio at the \(i\)-th transmission attempt. For successful decoding at the \(n\)-th transmission attempt, \(\text{NACMI}\) should be:

\[\text{NACMI}_n = \sum_{i=1}^{n} M I_i P_i \geq 1. \tag{6}\]
Then, the required $\rho_n$ for correct decoding at the $n$-th transmission attempt (supposing that the decoding has failed up to the $(n - 1)$-th transmission attempt) can be evaluated from (3) after replacing $N_t/N_i$ by $\rho_i$, as:

$$\rho_n = \frac{1}{M_{in}^n} - \frac{1}{M_{in}^{n-1}} \sum_{i=1}^{n-1} M_i \rho_i, \quad (7)$$

where $M_i = MI(\gamma_i)$ is the mutual information per symbol that the receiver collects after the $i$-th transmission attempt with SNR $\gamma_i$.

Equation (7) shows that the redundancy $\rho_n$ depends on all the previous $\rho_i$: $i = 1, 2, ..., n - 1$ in addition to $M_i$: $i = 1, 2, ..., n - 1$, which are collected at the receiver, and used to find the redundancy ratio in the next transmission attempt. The mentioned redundancy ratio is known at the transmitter thanks to the feedback channel. Therefore, the transmitter updates the redundancy using the adaptation function:

$$\rho_n = \rho_n(M_i), \quad i = 1, 2, ..., n. \quad (8)$$

However, from (6), the decoding in the $n$-th transmission attempt will be successful if $NACMI_n = NACMI_{n-1} + \rho_n M_{in} \geq 1$. Thus, the decoding success or failure depends on $NACMI_{n-1}$ which is already known, and on $M_{in}$, which is random in its nature, and it needs to be predicted from previous CSI in general. Therefore, the adaptation function of the redundancy is:

$$\rho_n = \rho_n(NACMI_{n-1}, M_{in}), \quad (9)$$

where from (6)

$$\rho_n = (1 - NACKMI_{n-1}) \frac{1}{M_{in}}. \quad (10)$$

And for block fading channel the redundancy adaptation function is:

$$\rho_n = (1 - NACKMI_{n-1}) \frac{1}{M_{in}}, \quad (11)$$

where $M_{i2}$ is the mutual information estimated for the first received subcodeword.

Note that: $NACMI_{n-1}$ is still less than 1, where the $n$-th transmission is needed only if the decoding at the $(n - 1)$-th attempt was failed. Therefore, the coding rate is adapted with the adaptation of the redundancy; depending on the estimated $NACMI$ value in addition to the predicted value of MI using (10) (or (11) for slow fading channel), or equivalently depending on $\rho_1, \rho_2, ..., \rho_{n-1}$ in addition to $M_{i1}, M_{i2}, ..., M_{in}$ (or $M_{i2}$ for block fading channel) using (7).

Tell now, the initial transmission rate associated to SNR region $[\gamma^{(k)}, \gamma^{(k+1)}]$ is then:

$$R_{initial} = R_m \frac{N_1}{N_1 + N_2} = R_m R_c = R_m \frac{1}{1 + \rho_2} = R_m \frac{1}{1 + \rho_{initial}} \quad (12)$$

where $\rho_{initial} = \rho_2$ is the initial anticipated redundancy ratio to be used at the first transmission of a packet, $R_m$ is the modulation rate (bit per QAM symbol) and $R_c$ is the coding rate at the first transmission (the initial anticipated coding rate). The best spectral efficiency is obtained when minimum redundancy is chosen, this is the case where $\rho_i$ is chosen as minimum as possible.

Note that, in practical systems, to avoid the additional usage of feedback channel, the acknowledgement feedback shouldn’t be large. While the acknowledgement must express the amount of redundancy, and the word “just enough redundancy” may be between one bit and a codeword containing a large number of bits. Therefore we need to quantize the required redundancy $\rho_i$ to some prefixed values $\{\rho_j\}$ via our explained procedure in [14].

- Determine the quantized redundancies $\{\rho_j\}$ via the above-explained procedure.
- The operating stages in the proposed cross-layer design follow the flow chart shown in Fig. 2.

**Fig. 2.** The operating stages in the proposed ARIR/HARQ

In our proposed technique, the additional computational complexity comes from mutual information estimation.
Supposing now (in the worst case), that N HARQ attempts are reached, the length of the code-word will be \( N_s \) symbols (containing the systematic sub-codeword and redundancies). Where \( N_s = N \times N_{s1} = constant \times N_{s1} \) (\( N_{s1} \) is the number of symbols obtained from the systematic sub-codeword after modulation). Therefore, the input vector length of MI estimation procedure is \( N_s \). As a mutual information estimation procedure, we use the nearest-neighbor distances in estimating MI, where the complexity is limited by the length of the input vector length [30]. Therefore the additional complexity is limited by \( constant \times N_{s1} \times N = \alpha N_{s1} \) (where \( \alpha \) is constant). Which means that the additional computational complexity is \( O(N_{s1}) \).

IV. PERFORMANCE ANALYSIS OF ARIR-HARQ

Our objective now, is to evaluate the benefits of the proposed adaptive rate incremental redundancy (ARIR) HARQ adopting the adaptation function (9), and compare it with the Variable-Rate Incremental Redundancy (VR-IR-HARQ) proposed in [17]. The performance criterion is the throughput. If the redundancy ratios \( \rho_n \) of the transmission attempts are not the same, we talk about Variable-rate Retransmission HARQ which was discussed in [17], while if \( \rho_n \equiv \rho_1, \forall n \), we obtain the fixed-rate (FR) transmission considered in [7] or [22]. In our proposed design, we make \( \rho_n \) adaptively adjusted with the CSI, and we call this scheme adaptive rate incremental redundancy HARQ (ARIR-HARQ). Since \( \rho_n \) decreases with increasing \( n \), the sub-codewords have decrementing lengths with HARQ process proceeding, recover VR transmission, and we can deal ARIR-HARQ as VR-IR-HARQ, while the difference is that \( \rho_n \) are not predetermined.

A. Throughput for Adaptive Rate IR-HARQ (ARIR-HARQ)

According to the reward-renewal theorem [23], the throughput is the ratio between the expected number of correctly received bits \( \bar{N}_b \) (after \( N \) transmission attempts for each packet delivering) and the expected number of channel uses \( \bar{N}_e \) required by the HARQ process to deliver the packet, i.e.

\[
\eta_{\text{ARIR}} = \frac{\bar{N}_b}{\bar{N}_e}, \tag{13}
\]

Let \( \text{NACK}_n \) denote the event of decoding failure in the \( n \)-th transmission attempt and \( f_{\text{ARIR}}^{(n)} = \Pr\{\text{NACK}_1, \text{NACK}_2, ..., \text{NACK}_n\} \) denote the probability of decoding failure after \( n \) transmission attempts. Equivalently, \( f_{\text{ARIR}}^{(n)} \) has the meaning of HARQ “outage probability”, defined above in the condition (4). In ARIR-HARQ, from (2) and (3), the decoding fails in the \( n \)-th transmission attempt if the accumulated mutual information is lower than the transmission rate (\( \text{NACMI} < 1 \)), which yields the following condition

\[
f_{\text{ARIR}}^{(n)} = \Pr\{\sum_{i=1}^{n} MI(y_i) \rho_i < 1 \}, \tag{14}
\]

where \( y_i \) is the SNR during the \( i \)-th transmission attempt.

To calculate \( f_{\text{ARIR}}^{(n)} \) we may follow the same suggested procedure in [7], where the authors introduce a random variable \( g_i = MI(y_i); i = 1, ..., n \) whose PDF can be obtained by

\[
g_i(x) = MI(y_i) = \ln(2) p \left( \frac{x}{2^{\rho_i} - 1}; m \right) \frac{2^{\rho_i}}{\rho_i}, \tag{15}
\]

Alternatively, we may approximate \( g_i = MI(y_i) \) by the Gaussian variable, [22][14], i.e.

\[
g_i(x) \approx \frac{1}{\sqrt{2\pi\sigma_{mi}}} \exp \left( -\frac{(x-MI_{mi})^2}{2\sigma_{mi}^2} \right), \tag{16}
\]

and

\[
\sigma_{mi}^2 = \int_0^\infty M^2 I(y) p(y)dy - \bar{M}^2_{mi}, \tag{17}
\]

are, respectively the mean and the variance of \( MI(y) \), and they are found based on the channel model. In addition to Gaussian, other distribution for MI may be taken like gamma or beta distribution [24] and [25].

To find the throughput, we need to evaluate \( f_{\text{ARIR}}^{(n)} \) in (14) using the pdf in (15)

\[
f_{\text{ARIR}}^{(n)} = \int_0^1 \sum_{i=1}^{n} g_i(x) \rho_i dx = \sum_{i=1}^{n} \rho_i \int_0^1 g_i(x) dx, \tag{18}
\]

and

\[
g_i(x) \approx \frac{1}{\sqrt{2\pi\sigma_{mi}}} \exp \left( -\frac{(x-MI_{mi})^2}{2\sigma_{mi}^2} \right) = N[M_{mi}, \sigma_{mi}^2], \tag{19}
\]

where \( N[M_{mi}, \sigma_{mi}^2] \) is the Gaussian distribution with mean \( M_{mi} \) and variance \( \sigma_{mi}^2 \).

\[
\int_0^1 g_i(x) dx = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{M_{mi}-1}{\sqrt{2}\sigma_{mi}} \right) \right). \tag{20}
\]

Therefore:

\[
f_{\text{ARIR}}^{(n)} = \sum_{i=1}^{n} \left[ \frac{\rho_i}{2} \left( 1 + \text{erf} \left( \frac{M_{mi}-1}{\sqrt{2}\sigma_{mi}} \right) \right) \right]. \tag{21}
\]

Going back to (13), by definition: [7], [23]

\[
\bar{N}_b = N_1 \left( 1 - f_{\text{ARIR}}^{(n)} \right), \tag{22}
\]

and

\[
\bar{N}_e = \sum_{n=1}^{N} \bar{N}_n, \tag{23}
\]

where \( \bar{N}_n = \rho_n \cdot f_{\text{ARIR}}^{(n-1)} \), then,
\[ \bar{N}_e = N_1 \sum_{n=1}^{N} [\rho_n f_{\text{ARIR}}^{(n-1)}] \, . \]  

Therefore:

\[ \eta_{\text{ARIR}} = \frac{1 - f_{\text{ARIR}}^{(n)}}{\sum_{n=1}^{N} [\rho_n f_{\text{ARIR}}^{(n-1)}]} \, , \]  

where \( f_{\text{ARIR}}^{(n)} \) can be obtained from (21) for \( N \) transmission attempts. Replacing \( f_{\text{ARIR}}^{(n)} \) (the probability of decoding failure after \( N \) transmission attempts) from (21) in (22), we get:

\[ \bar{N}_b = N_1 \left( 1 - \sum_{i=1}^{\eta} \frac{\rho_i}{2} \left( 1 + erf \left( \frac{M_{\text{mi}}^{-1}}{\sqrt{2} \sigma_{\text{mi}}} \right) \right) \right) \, . \]  

Replacing (21) in (24) yields:

\[ \bar{N}_e = N_1 \sum_{n=1}^{N} \rho_n \sum_{i=1}^{\eta} \left[ \frac{\rho_i}{2} \left( 1 + erf \left( \frac{M_{\text{mi}}^{-1}}{\sqrt{2} \sigma_{\text{mi}}} \right) \right) \right] \, . \]  

Substituting now in (13) we obtain the throughput of ARIR-HARQ

\[ \eta_{\text{ARIR}} = \frac{1 - \sum_{i=1}^{\eta} \frac{\rho_i}{2} \left( 1 + erf \left( \frac{M_{\text{mi}}^{-1}}{\sqrt{2} \sigma_{\text{mi}}} \right) \right)}{\sum_{i=1}^{\eta} \frac{\rho_i}{2} \left( 1 + erf \left( \frac{M_{\text{mi}}^{-1}}{\sqrt{2} \sigma_{\text{mi}}} \right) \right)} \, . \]  

We can see from (28), knowing that \( \lim_{n \to \infty} \rho_n = 0 \) and \( \rho_n \leq 1 \), that \( \eta_{\text{ARIR}} \) increases with decreasing \( \rho_n \) in ARIR-HARQ system.

When ARIR is used, \( \rho_n^{(\text{ARIR})} \) are chosen (depending on the adaptation function in (10), or (11) for block fading channel) as minimum as possible to guarantee the condition in (6), while \( \rho_n^{(\text{VR-IR})} \) (for variable rate incremental redundancy HARQ) are pre-fixed values. Thus, \( \rho_n^{(\text{ARIR})} \) are lower bounds for \( \rho_n^{(\text{VR-IR})} \) for correct decoding. When \( \rho_n^{(\text{VR-IR})} \) is smaller than \( \rho_n^{(\text{ARIR})} \), an error has more chance to occur—because \( \rho_n^{(\text{ARIR})} \) are chosen as minimum as possible to guarantee correct decoding. Therefore, we can expect higher throughput in ARIR-HARQ comparing to VR-IR-HARQ.

\[ \text{B. Achievable Throughput for Adaptive Rate IR-HARQ (ARIR-HARQ)} \]

In practical design (quantizing \( \rho \) to fixed values \( \{\rho^{(j)}\}_{j=0}^{\eta} \)), the lower bound of \( \eta \) could be obtained from (25) (or equivalently from (28)) where \( \rho_n \) is quantized to \( \rho^{(j)} \) when the required \( \rho \in [\rho^{(j)}, \rho^{(j+1)}) \). In this case, \( f_{\text{ARIR}}^{(n)} \) reaches 0, while \( f_{\text{ARIR}}^{(n-1)} = \prod_{i=1}^{n-1} p_i \) increases, supposing \( n - 1 \) failed attempts. Therefore, the minimum of \( \eta_{\text{ARIR}} \) will be:

\[ \min_{\rho_n \in [\rho^{(j)}]} \left[ \eta_{\text{ARIR}} \right] = \frac{1}{\sum_{n=1}^{\eta} \left[ f_{\text{ARIR}}^{(j)} \right] \left[ \prod_{i=1}^{n-1} p_i \right]}. \]  

The upper bound can be reached when \( \rho_n \) is quantized to \( \rho^{(j+1)} \). \( f_{\text{ARIR}}^{(n)} \) hence reaches \( p_{\text{out}} \), while \( f_{\text{ARIR}}^{(n-1)} = \prod_{i=1}^{n-1} p_i \) decreases, supposing \( n - 1 \) failed attempt. Therefore, the maximum of \( \eta_{\text{ARIR}} \) will be:

\[ \max_{\rho_n \in [\rho^{(j)}]} \left[ \eta_{\text{ARIR}} \right] = \frac{1 - p_{\text{out}}}{\sum_{n=1}^{\eta} \left[ f_{\text{ARIR}}^{(j)} \right] \left[ \prod_{i=1}^{n-1} p_i \right]}. \]  

This relation represents the maximum achievable throughput of the proposed ARIR-HARQ system.

\[ \text{C. Average Number of Transmissions of ARIR-HARQ} \]

Supposing that \( N \) HARQ attempts are done including the first transmission attempt, therefore, given \( f_n \) is the decoding failure at the \( n \)-th transmission attempt, the average number of transmissions is given by \( \sum_{n=1}^{N} (1/f_n) \cdot f_n \). Where, \( l_n \) is the number of packet transmissions within the \( n \)-th transmission attempt (in our design: \( l_n = 1 : \forall n \)), and \( f_0 = 1 \) (the first transmission attempt is a certain event). Therefore,

\[ N_{\text{avg}} = \sum_{n=1}^{N} f_{n-1} = 1 + \sum_{n=1}^{N-1} f_n. \]  

In our design, it can be seen from (31) and (29) that, we can expect smaller average number of transmissions compared to fixed and variable rate incremental redundancy due to the minimization of \( \rho_n \) which makes \( f_n \) as lower bounds in our design.

\[ \text{V. PERFORMANCE ANALYSIS OF AMC COMBINED WITH ARIR-HARQ} \]

In this section, we derive the average PER and the spectral efficiency of our AMC/ARIR-HARQ cross-layer design. This is intended for applications where PER goal \( P_e^N \leq P_{\text{target}} \) under the constraint of a maximum number of transmissions \( N \) and the required round trip delay for each transmission (the time elapsed from sending a packet until receiving its acknowledgment), supposing that it doesn’t exceed the channel coherence time.

The average PER at the physical layer will be lower than \( P_{\text{target}} \), since \( P_{\text{target}} \) is the upper bound of the instantaneous PER in our AMC design. We need to evaluate this average PER at the physical layer.

Let \( f_{\text{ARIR}}^{(k,1)} \) be the probability of failed decoding after the first transmission when the \( k \)-th MCS (MCS) is used. Therefore, the PER (\( \text{PER}_k \)) is given by (21):

\[ \text{PER}_k = \int_{y_k}^{y_{k+1}} f_{\text{ARIR}} p_r(y) dy = f_{\text{ARIR}} \text{ Pr}(k), \]  

where \( \text{Pr}(k) \) is the probability of using the \( k \)-th modulation.
mode (the integration of the fading channel pdf between $\gamma_k$ and $\gamma_{k+1}$) [21].

The average PER of AMC can then be computed as the ratio of the average number of incorrectly received packets over the total average number of transmitted packets (cf. [21, eq. (35)])

$$\text{PER} = \frac{\sum_{k=1}^{K} R_k W_k}{\sum_{k=1}^{K} R_k P(k)} = \frac{\sum_{k=1}^{K} R_k f(k) \text{Pr}(k)}{\sum_{k=1}^{K} R_k P(k)}. \quad (33)$$

Since HARQ is implemented at the data link layer, in addition to the modulation mode $m$, redundancies for the same packet in error will be transmitted, up to $N - 1$ times. For notational brevity, let us denote $p := \text{PER}$. Therefore, the average number of transmissions per packet can be found as [(26, p. 397)] or from (31):

$$\bar{N}(p, N) = p + p^2 + p^3 + \ldots + p^{N-1} = \frac{p-p^N}{1-p}. \quad (34)$$

With the average PER in (33), knowing that $\bar{N} \geq 1$, the actual packet loss probability at the data link layer with $N$-HARQ is:

$$P_{\text{actual loss}} = p\bar{N} \leq P_{\text{target}} = P_{\text{out}}, \quad (35)$$

which verifies the requirement in (4).

When the $k$-th MCS is used ($M_k$ - QAM modulation mode and $R_k$ coding rate), each transmitted symbol will carry $R_k = R_s g_s(M_k)$ information bits. As in [21], we assume, the symbol rate is $T_s$, the bandwidth of the considerable Nyquist pulse shaping filter is $B = 1/T_s$. Therefore, at the physical layer (when packet retransmissions are not considered), the achieved average spectral efficiency (bit rate per unit bandwidth), supposing no fading, is $\eta_{\text{phy}} = R_k$ (where only physical layer AMC design is considered without fading).

According to the AMC rule (the $k$-th MCS is chosen when $\gamma \in [y_k, y_{k+1})$), the modulation mode and coding rate depend on the received SNR. Then, the spectral efficiency is [27]:

$$\eta_{\text{ARIR}} = \eta_{\text{ARIR}}(N) = \eta_{\text{ARIR}} R_k. \quad (40)$$

Replacing (39), (40) in (36), we get the spectral efficiency of our proposed AMC/ARIR-HARQ system as:

$$\bar{\eta}_{\text{AMC/ARIR-HARQ}} = \sum_{k=1}^{K} \left[ R_k \left( \frac{1}{\sum_{n=1}^{N} n \sum_{i=1}^{n-1} \rho_i \left( 1 + \text{erf} \left( \frac{\sqrt{2} \eta_{\text{MI}}(\gamma_{k+1})}{\sigma_{\text{MI}}(\gamma_k)} \right) \right) \right) \text{Pr}(k) \right], \quad (41)$$

where $\bar{\eta}_{\text{MI}}(\gamma_k)$ and $\sigma^2_{\text{MI}}(\gamma_k)$ are the mean and variance of MI when the $k$-th modulation mode is used, where $\gamma \in [y_k, y_{k+1})$.

VI. NUMERICAL RESULTS

In this section, numerical results for the new design are presented. We set the packet length to be 1056. While specific numerical values will be affected if we choose another packet length. However, because of the unifying development in Sections 2 and 3, similar observations are expected.

Let the performance constraints $P_{\text{target}} = 10^{-3}$ and the maximum number of transmissions is 4. The channel is a Nakagami-4 fading channel, over SNR range $[0 - 26]$dB. Table I gives the anticipated initial coding rates for each modulation mode, where each modulation mode and a coding rate construct an MCS. The initial coding rates adopted in the simulation are chosen from the literature where they are in convenience with the note mentioned in sec.3.2. To find SNR switching thresholds between MCSs, we resort to the same procedure followed in our previous work in [14] for the MCSs whose initial coding rates are 1s. Then, in order to find SNR switching threshold for the MCSs whose coding rate more than 1, we use the relation: [15]

$$y_{th}^k = y_{th}^{k-1} - 10 \log \left( \frac{1}{1+e^{-\gamma_{th}}(\gamma_{th})} \right), \quad (42)$$

where, $y_{th}$ is the switching threshold for the MCSs whose coding rate more than 1, $\gamma_{th}$ is the target SNR, and $e$ is the base of the natural logarithm.
where \( \gamma_{kn}^{j} \) is the switching threshold of the MCS whose the \( k \)-th modulation mode and the \( j \)-th coding rate, and \( CR^{(j)} \) is the \( j \)-th coding rate. Table 1 summarizes the SNR switching thresholds of MCSs used in the simulation.

When the system is working, the MCS (modulation mode and initial anticipated coding rate for correct decoding) is chosen based on the estimated SNR. It should be noted that the estimated SNR may exceed 26dB after the fading channel and adding the noise, this is why we use MCSs with higher rates (512-QAM with coding rate 1 or \( \frac{1}{2} \) in the last row of Table 1). If the decoding has failed, the receiver estimates MI value and predict the amount of redundancy for correct decoding and send it back to the transmitter which, in turn, sends the request to the receiver. Again, this operation will repeat until the correct decoding or the maximum number of transmission is achieved.

**TABLE I**

<table>
<thead>
<tr>
<th>MCS</th>
<th>Coding rate</th>
<th>1</th>
<th>4/5</th>
<th>2/5</th>
<th>4/7</th>
<th>1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPSK</td>
<td></td>
<td>1.7464</td>
<td>0.9546</td>
<td>0.2851</td>
<td>-0.2948</td>
<td>Less than -0.2948</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.7155</td>
<td>1.7464</td>
<td>0.9546</td>
<td>-0.2948</td>
<td>-0.2948</td>
</tr>
<tr>
<td>QPSK</td>
<td></td>
<td>5.4074</td>
<td>6.3765</td>
<td>5.4074</td>
<td>6.3765</td>
<td>5.4074</td>
</tr>
<tr>
<td>8-QAM</td>
<td></td>
<td>8.9639</td>
<td>9.9330</td>
<td>8.1721</td>
<td>9.1202</td>
<td>8.1721</td>
</tr>
<tr>
<td>64-QAM</td>
<td></td>
<td>18.6493</td>
<td>19.6184</td>
<td>17.8575</td>
<td>18.6493</td>
<td>17.8575</td>
</tr>
<tr>
<td>256-QAM</td>
<td></td>
<td>24.3564</td>
<td>25.3255</td>
<td>23.5646</td>
<td>24.3564</td>
<td>23.5646</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR [dB]</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2851</td>
<td>-0.2948</td>
<td>Less than -0.2948</td>
<td>-0.2948</td>
<td>-0.2948</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.7464</td>
<td>0.9546</td>
<td>0.2851</td>
<td>-0.2948</td>
<td>-0.2948</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.7155</td>
<td>1.7464</td>
<td>0.9546</td>
<td>-0.2948</td>
<td>-0.2948</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5.4074</td>
<td>6.3765</td>
<td>5.4074</td>
<td>6.3765</td>
<td>5.4074</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>8.9639</td>
<td>9.9330</td>
<td>8.1721</td>
<td>9.1202</td>
<td>8.1721</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3 shows the spectral efficiency of ARIR-HARQ \( K = 2, 3, 4 \) with the lower bound \( \eta_{a} \) (i.e., when \( K = 1 \)) and the upper bound \( C \) (the ergodic capacity). Observe that the throughput of ARIR-HARQ gets very close to the ergodic capacity when \( K > 2 \). We can easily observe that \( K = 2 \) is enough to reach the convergence, while diminishing returns are observed for \( M > 2 \). As we recall that the transmitter has no knowledge of the instantaneous CSI, but it seems to be aware of it. Comparing to AIR-HARQ [6], gains are notable in the spectral efficiency, for example, for \( K = 2 \), where the residual throughput (the difference between the throughput and the upper bound) is between 0.1 bit per symbol per Hertz at \( SNR = 5dB \) and 0.2 bit per symbol per Hertz at \( SNR = 25dB \) in our design, while it changes between 0.4 and 0.6 bit per symbol per Hertz at the same conditions. Same results are depicted in comparison with the case of VR-HARQ and FR-HARQ [17]. In addition, less transmission attempts to reach the maximum achievable spectral efficiency comparing to AIR-HARQ and VR-HARQ [6] [17].

![Image](324x196 to 534x369)

**Fig. 3.** Average spectral efficiency for AMC/ARIR-HARQ design versus SNR [dB].

![Image](572 to 743)

**Fig. 4.** Average packet delivering delay for AMC/ARIR-HARQ design versus SNR [dB].

In contrast, we expect more average redundancy over the whole SNR range, but still as minimum as possible to get better spectral efficiency. Unlike [6], in our design the redundancy decreases with the transmission attempts, and this is expected while we look for minimum redundancy for correct decoding after each transmission attempt. Fig. 5 shows the average value of redundancy changes with the number of transmissions over the SNR range.
The difference between our design and [6] is that the initial coding rate in [6] is always close to the ergodic capacity and the progressive redundancy ratio increases with the progress of HARQ process.

While in our design, the initial coding rate is chosen based on SNR estimation, in addition, the first redundancy doesn’t exceed the packet length, which means that, the initial coding rate may be less than or equal the ergodic capacity, the initial redundancy ratio hence will be in the domain \([0, 1]\). This is clear from the Fig. 5 for No HARQ case. Fig. 6 shows the progressive coding rate with the progress of HARQ process.

Fig. 7 and 8 respectively show the average BER and PER of the system for \(K = 1\) to \(4\). We can see from Fig. 7 that BER for \(K = 1\) (the case of AMC only) doesn’t reach the performance requirement \(P_{target}\), and we need at least 2 HARQ attempts to reach the target after \(SNR = 2.1dB\), this result is with agreement with the results abstracted from Fig. 3, where we found that 2 HARQ attempts are enough to reach the maximum throughput. In addition, we need more HARQ attempts to reach the target at low SNRs. Same results are depicted from PER curves in Fig. 8.

VII. CONCLUSION

In this paper, a new scheme for HARQ with adaptive rate incremental redundancy (ARIR-HARQ) is proposed and analyzed for transmissions over Nakagami-\(m\) block-fading channels. The proposed HARQ is combined with AMC in cross layer design to overcome the fading. The design benefits from the instantaneous estimation of SNR and MI as CSI to anticipate the convenient initial coding rate and modulation order in order to initialize the transmission with the convenient MCS. Then the coding rate (or equivalently, the redundancy) will be updated with the estimated MI after each transmission attempt. We established a general formulation for the throughput when the redundancy is updated with CSI.

We have shown that the proposed ARIR-HARQ provides gains over AIR-HARQ, VR-HARQ and FR-HARQ-IR in terms of increased throughput and decreased average number of transmissions at the expense of the computational burden spent for CSI estimation.

REFERENCES
