Adaptive Retransmission Protocol Based on Mutual Information

Assal Alkhoder¹, Abdelnasser Assimi² and Mohammad Alhariri³

¹,²,³Higher Institute of Applied Science and Technology
¹assal.alkhoder@hiast.edu.sy, ²abdelnasser.assimi@hiast.edu.sy, ³hariri66@mail.ru

Abstract

We examine a new method for cross-layer design of adaptive modulation and coding at the physical layer and an automatic repeat request protocol at the data link layer, in order to maximize spectral efficiency under prescribed error performance constraint. The new proposed method depends on the mutual information concepts at the design stage of the system. The design relies on the mutual information statistics, and then statistical study of the mutual information has done in order to demonstrate its random nature of the mutual information. The system is simulated over Nakagami-m block fading channel. Numerical results reveal an enhancement in throughput and error performance with the new design over the traditional design. In addition, more margin in error probabilities could be sacrificed to optimize the throughput in order to reach the maximum achievable throughput in the communication channel.

Keywords: Adaptive modulation and coding (AMC), automatic repeat request (ARQ) protocol, cross-layer design, Mutual Information (MI), fading channels, spectral efficiency

1. Introduction

To enhance throughput in future wireless data communication systems, adaptive modulation and coding (AMC) have been studied extensively and adopted at the physical layer, in order to match transmission rates to time-varying channel conditions; see for example. [1, 2] and references therein. For example, in recent wireless network standards, such as 3GPP/3GPP2, HIPERLAN/2 and IEEE 802.11/16 [3],[5], AMC has been advocated at the physical layer to enhance channel utilization and throughput of future wireless communication systems. However, this method cannot achieve the maximum spectral efficiency [6], especially when the number of available AMC modes is limited. To achieve high reliability at the physical layer, one has to reduce the transmission rate using either small size constellations, or powerful but low-rate error control codes [3].

An alternative way to mitigate channel fading is to rely on the Automatic Repeat Request (ARQ) protocol at the data link layer, that requests retransmissions for those packets which are received in error [3]. Hybrid ARQ (HARQ), best prescribed and used by 3GPP, 3GPP2 and IEEE 802.16 standards [2, 7], is adopted to mitigate channel fading at the data link layer.

In AMC schemes, the constellation size and the code rate are chosen based on channel state information (CSI), while in HARQ, only the code rate which is adapted based on feedback acknowledgments. In HARQ schemes, a packet is first transmitted with some parity bits for error detection and none or a few parity bits for error correction. Additional blocks of bits are transmitted upon retransmission requests. The receiver combines the transmitted and retransmitted blocks together to form a more powerful codeword in order to recover the original packet [3, 6]. Obviously, the mechanism of HARQ schemes makes
itself autonomously adaptive to the variations of instantaneous channel conditions by performing a different number of retransmissions. This makes it insensitive to errors and delays incurred in channel estimation [6]. However, during a long deep fading, a large number of retransmissions will be required for certain packets to be correctly decoded at the receiver. This, in turn, may lead to unacceptable large delay and buffer sizes.

Based on the above analysis of benefits and drawbacks of AMC and HARQ, a cross-layer design by combining AMC and HARQ is performed by using one's advantages to overcome the other's disadvantages [3, 4, 6], aiming to achieve maximum of spectral efficiency under certain QoS constraints such as Packet Error Rate (PER) and transmission delay.

The conventional way of performing AMC adaptation is to simply divide the signal to noise ratio (SNR) range into several regions and assign different AMC modes with different constellation sizes and coding rates to the different SNR regions [1, 3, 6]. Thus, there are several SNR switching thresholds between the mentioned regions. The traditional way to find these SNR thresholds is to simulate the system performance with the AMC scheme and extract them from Bit Error Rate (BER) or PER curves. We can find in [1] mathematical formulas which give SNR thresholds between modulation schemes. The same procedure is also followed to find the thresholds between coding rates in HARQ process [3, 6]. However, the gain in system performance depends on the accuracy of SNR estimation.

Looking for other link quality metrics which are more accurate than SNR in estimating BER or PER, we find that the mutual information (MI) is a good candidate for this purpose. Results in [8] indicate that mutual information is an appropriate measure of channel quality to predict error rates reasonably well for convolutional and turbo codes. In [8], several link abstraction models are compared in terms of accuracy in the Block Error Rate (BLER) estimation. The results of this work show that the Mutual Information Effective SINR Metric (MIESM) outperforms Exponential Effective SNR Mapping (EESM) in terms of the BLER prediction accuracy. However, the MIESM method has the drawback that, since there is no closed form expression for the Mutual Information (MI) between transmitted and received modulation symbols, or between transmitted and received coded bits, it must be approximated or computed numerically. Authors in [9] focus on the comparison of MIESM and EESM models for Long Term Evaluation Down Link (LTE DL) in terms of complexity and BLER prediction accuracy. Emphasis is put in [9] on finding the model parameters that are suitable to predict the LTE DL BLER for a variety of physical layer configurations, ranging from single-input-single-output (SISO) to multi-input-multi-output (MIMO) in spatial multiplexing mode. Authors in [10] propose an alternative measure called Mean MI Bit Mapping (MMIBM), as a link quality metric to be used for fast link adaptation and compare it against the MI effective SNR mapping (MIESM). In [11], we find a comparison between both measures by means of realistic link-level simulations including channel estimation and feedback delay. In addition, authors in [11] develop a methodology to obtain performance bounds for any fast link adaptation algorithm, i.e. an upper bound for throughput and a lower bound for PER, and study the tightness of the throughput performance bound. In [12], the average mutual information per code bit is used as a single parameter for the codeword error probability estimation of a Forward Error Correction (FEC) codeword that is transmitted on chunks with different modulation formats and for the estimation of Channel gain to Noise Ratios (CNR). While MI is used as a performance measure in many works in the literature [8-13], it’s also used as a measure to track the channel state variations in AMC/HARQ systems. In [14], as a means to consider the system behavior, including the channel code, regarding bit or frame error rate (BER, FER) versus normalized average mutual information (NAMI) of an Orthogonal Frequency Division Modulation (OFDM) symbol for several M-QAM alphabets, was investigated. Based on the mutual information results, two algorithms using this channel knowledge were introduced. On one hand, the
error rates can be optimized assuming a fixed rate requirement and on the other hand, the achieved transmission rate can be optimized under a BER/FER constraint. In both cases, a NAMI threshold was derived from the simulated system behavior. In [4], the proposed method uses the estimation of the mutual information between received and sent message symbols, in order to estimate the number of bits necessary to correctly decode the message at next transmission with a predefined targeted probability (say 0.5). If the message has not been successfully decoded, the transmission will be repeated after a new mutual information estimation with another predefined probability (say 0.3) of successful decoding, until the successful decoding of the message, or a maximum number of retransmission is achieved. Algorithms for rate adaptation in [15] rely on the observation that the mutual information is sufficient as single parameter for the estimation of the FEC performance in the case of fading. The work in [16] has verified the good accuracy of MI model for rate compatible Low Density Parity Check (RC LDPC) coded OFDM system. While in [17], the MI model is used to predict the retransmission length of two hybrid ARQ schemes, Partial Chase Combining (PCC) and Incremental Redundancy (IR).

Traditionally, switching thresholds between modulation schemes and the maximum number of transmissions are predetermined depending on PER(SNR) (or BER(SNR)) curves or mathematical models [3, 6]. Then, modulation mode is chosen depending on the estimated value of SNR. The transmitter sends a codeword to the receiver. If the decoding of that packet has failed, a not acknowledge signaling (NACK) will be sent back to the transmitter asking for another copy of the codeword (Chase combining concept) or a redundancy (incremental redundancy concept). The transmitter responds the not acknowledge. This operation will be repeated until acknowledge signaling (ACK) is sent back to the transmitter (successful decoding) or a maximum number of transmissions is achieved.

In this paper, we propose a new and accurate method to determine SNR switching thresholds between modulation and coding schemes (MCSs) in a cross layer combining of AMC and ARQ. The proposed method is based on MI estimation to determine SNR thresholds. In our proposed design, we supposed that the number of transmissions for successful decoding could be anticipated based on the estimated SNR. Therefore, based on the estimated SNR, a modulation mode and an anticipated number of copies of the same packet are transmitted. If the decoding has failed, NACK signaling will be sent back to the transmitter asking for a new version of the packet, and the operation continues until the correct decoding (ACK signaling is sent back to the transmitter) or the maximum number of retransmissions is achieved.

We start by statistical simulation to find SNR switching thresholds as we will describe in the following sections. Therefore, SNR range is divided to several sub-regions. To limit the delay in our system, the anticipated number of transmissions will be defined independently for each of SNR sub-regions, so it won’t be a fixed number, but a limited one. Which means that we determine the anticipated number of transmissions after estimating the effective SNR and before the first transmission of the packet.

This paper is organized as follows: In Section 2, we present the system model and the related assumptions. The principle of combining AMC and HARQ is presented in 2.1. The conventional method to determine the switching thresholds between MCSs is summarized in 2.2. In 2.3, we describe the estimation method of the mutual information which will be used in our proposed method to find the mentioned switching thresholds as described in Section 2.4. The performance analysis of the AMC/ARQ system is derived in Section 3. As we are interested in error probability and spectral efficiency as performance metrics, we describe these metrics in this section. Numerical simulations are presented in Section 4. Finally, conclusions and future works are summarized in Section 5.
2. System Model and Assumptions

We consider a single-transmit single-receive antenna system (single input single output (SISO)) as a point-to-point wireless packet communication system, consisting one source node and one destination node as shown in Figure 1. It consists of a joint AMC module at the physical layer, and an ARQ module at the data link layer. The processed unit at the data link layer is a frame of multiple packets, while the processed unit at the physical layer is a frame which contains a fixed number of transmitted symbols.

We assume that multiple transmission modes are available at the physical layer. Each mode consists of a specific Modulation and Coding Scheme (MCS) and FEC code scheme as in HIPERLAN/2, IEEE 802.11a, and 3GPP standards [2, 5, 19]. The AMC controller determines the AMC scheme based on CSI estimated at the receiver. The controller send back the estimated CSI or the MCS selection to the transmitter through an error free feedback channel. The transmission mode is then updated by the AMC controller at the transmitter. Coherent demodulation and maximum-likelihood (ML) decoding are used at the receiver. The decoded bit streams are mapped to packets and pushed upwards to the data link layer. HARQ protocol with AMC is used in the system. This means, if the receiver cannot correctly decode the received data packet, it feeds back a not acknowledge (NACK) signaling to the transmitter, otherwise, it feeds back an acknowledge (ACK) signaling. While the number of transmissions is predetermined for each modulation scheme independently in each SNR sub-region, the number of transmissions is limited by $N_{\text{max}}$. Feedback for ARQ and AMC related parameters are sent back together via the same physical feedback channel.

![Figure 1. Cross-layer Structure Combining AMC with ARQ](image-url)
At the physical layer, we deal with frame by frame transmissions, where each frame construct an OFDM symbol, and each frame contains a fixed number \( N_L \) of M-QAM symbols. Each frame at the physical layer may contain multiple packets from the data link layer (variable number of packets depending on AMC scheme). The packet and frame structures are depicted in Figure 2. Each packet contains \( N_p \) bits, \( b_1, b_2, ..., b_{N_p} \) (\( b = 0 \) or \( 1 \)), including payload, and cyclic redundancy check (CRC) bits to facilitate error detection. After modulation with mode \( k \) of rate \( R_k \) bits/symbol, each packet is mapped to a block of symbols \( (x_1, x_2, ..., x_{N_x}) \) containing \( N_x = N_p / R_k \) M-QAM symbols. Multiple such blocks, \( B_1, B_2, ..., B_{N_B} \), where \( N_B \) is the number of packets per frame, are gathered with \( N_c \) pilot symbols, in order to constitute one OFDM symbol to be transmitted at the physical layer, as in HIPERLAN/2 and IEEE 802.11a standards [19]. when mode \( k \) is used, the number of symbols per frame will be \( N_L = N_c + N_B N_p / R_k \), which means, the number of packets per frame (\( N_B \)) depends on the chosen MCS and coding rate pair.

We next list the operating assumptions adopted in this paper:

**A1:** The transmission is carried over a block fading frequency flat channel, which means that it remains invariant per frame, but is allowed to vary from frame to frame. This model is suitable for slowly-varying fading channels [20]. As a consequence, AMC is adjusted on a frame-by-frame basis.

**A2:** Perfect channel state information (CSI) is available at the receiver. The corresponding mode selection is fed back to the transmitter without error and latency, as in [1]. The assumption that the feedback channel is error free and has no latency, could be at least approximately satisfied by using a fast feedback link with powerful error control for feedback information [3]. Further considerations on system design with e.g., delayed, or, noisy CSI [21, 22], may be left for future investigation.

**A3:** Error detection based on CRC is perfect, provided that sufficiently reliable error detection CRC codes are used [1, 23]. As in [23, 3], the redundancy and serial number bits in each packet are not included in the throughput calculation.

**A4:** The fading channel coefficients corresponding to the original packet and its related retransmitted versions are the same, while they are independent from packet to another packet transmission. Therefore, they are independent and identically distributed (i.i.d.) random variables.

The round trip delay of ARQ is the time elapsed from sending a packet until receiving its acknowledgment. In general, these round trip delays doesn’t exceed the channel coherence time. Hence, the original packet and possible subsequent retransmissions experience the same channel, which justifies what is mentioned above.
For flat fading channels (A1), single parameter is enough to capture the channel quality, namely the received SNR $\gamma$. Since the channel varies from frame to frame, we adopt the general Nakagami-$m$ model to describe $\gamma$ statistically [25]. The received SNR $\gamma$ per frame is thus a random variable with a Gamma probability density function (pdf) [1]:

$$p(\gamma) = \frac{m^m \gamma^{m-1}}{\dot{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\dot{\gamma}}\right)$$

(1)

Where $\gamma$ is the received instantaneous SNR and $\dot{\gamma} := E(\gamma)$ is the average received SNR, $\Gamma(m) := \int_0^\infty t^{m-1}e^{-t}dt$ is the Gamma function, and $m$ is the Nakagami fading parameter ($m \geq 1/2$). We choose the Nakagami-$m$ channel model because it applies to a large class of fading channels. It includes Rayleigh channel as a special case when $m = 1$. And Rician fading channel for $m > 1$. One-to-one mapping between the Ricean factor $K$ and the Nakagami fading parameter $m$, Nakagami channel will approximate Ricean channels [24].

Considering the SISO channel, the conditional probability of this channel is $p(x|y)$, where $x$ is the input symbol, $y$ is the output symbol, the relationship between the input and output messages can be modeled as follow [25]:

$$y_l = h_l x_l + z_l$$

(2)

Where $h$ is the channel gain, $z$ is an additive white Gaussian noise and $l$ ($l = 1, 2, ..., L$) is the sub-carrier index in the OFDM configuration, where $L$ is the number of sub-carriers. Since OFDM turns wide-band frequency selective channel into multiple narrow-band flat fading channels [18], OFDM signaling converts frequency selective channel $h$ into $L$ parallel flat-fading channels [26].
Hard decoding is implemented on the received M-QAM symbols to recover the transmitted packets in order to find error probabilities and the spectral efficiency of the system as performance metrics.

2.1. Combining AMC and HARQ

In this section, we describe the principle of combining the AMC at PHY and HARQ at DLL. Since only finite delays could be afforded in practice, the number of HARQ retransmissions is constrained. So the maximum number of transmissions $N_{\text{max}}$ of one packet could be determined by dividing the maximum allowable system delay over the round trip delay required for each transmission [3], [6]. If a packet is still erroneous after $N_{\text{max}}$ transmissions, it will be dropped, and we declare a packet loss. As a result, the acceptable PER or BER [3], [6] at DLL should be predetermined.

Traditionally, PER (or BER) goal and $N_{\text{max}}$ are predetermined, then PER(SNR) (or BER(SNR)) curves or mathematical models [3], [6] are used to determine switching thresholds between MCSs. Some optimization on these thresholds can be performed in order to optimize the throughput or the delay [27], [28], [29].

In our proposed design, we supposed that the number of transmissions for successful decoding could be anticipated based on the estimated SNR. Therefore, based on the estimated SNR, a modulation mode $k$ and an anticipated number of copies $N_k$ of the same packet are transmitted. If the decoding is failed, NACK signaling will be sent back to the transmitter asking for a new version of the packet, and the operation continues until the correct decoding (ACK signaling is sent back to the transmitter) or the maximum number of retransmissions is achieved. That means, an initial anticipation of the coding rate (number of transmissions (copies) of the same packet) and modulation mode is predetermined, a first transmission is performed with the modulation mode $k$ and $N_k$ copies of the same packet, then minimizing the coding rate will be done (if the packet isn’t correctly decoded) up tell ACK signaling is sent back, or the minimum coding rate is achieved (the maximum number of transmission $N_{\text{max}}$ is achieved).

We propose in the next section a new method to determine thresholds between MCSs.

2.2. SNR Regions Determination

In our new design, the SNR range is divided to $k$ regions, to each region, a modulation mode is associated. Each of these regions is divided to $n$ sub-regions, to each sub-region, an anticipated number of transmissions (initial coding rate) is associated. In other words, modulation and coding schemes (MCSs) are associated with the sub-regions, and SNR switching thresholds must be determined between these sub-regions.

The proposed method to determine thresholds between MCSs is described in this section. It relies on the mutual information statistics. In our system, we must specify the number of initial transmitted copies of the same packet (limited by $N_{\text{max}}$) for each MCS. For this purpose, we first perform some numerical simulations to describe the statistical behavior of MI versus both SNR and BER.

An initial simulation is performed for a large number of packets. For a given SNR value, each packet is transmitted $N_{\text{max}}$ times independently of feedback acknowledgments. At the receiver, we record for each transmission the estimated instantaneous MI and the Bit Error Count (BEC) for the combined received copies of the same packet. This simulation is repeated for different modulation schemes (M-QAM) at each SNR value. This simulation allows to model the statistical behavior of MI with respect to other system parameters including SNR, modulation
scheme, BER and the number of transmissions. Before we present the results of these simulations, we first describe in the next sub-section the MI estimation method used in this paper.

2.3. Mutual Information

Several ESM approaches for predicting the instantaneous link performance have been proposed in the literature, including: mean instantaneous capacity [30, 31], exponential effective SINR mapping (EESM) [32, 33] and mutual information effective SINR mapping (MIESM) [8, 34]. Each of these approaches uses a different function to map the vector of SINR values to a single number. In general, any ESM physical layer abstraction method can be described via the following equation [35, 36]:

$$SNR_{eff} = \Phi^{-1}\left\{ \frac{1}{L} \sum_{i=1}^{L} \int_{LLR_i} \Phi(SNR_i) \right\}$$  \hspace{1cm} (3)

Where: $SNR_{eff}$ is the effective SNR, $SNR_i$ is the SNR in the $l$-th subcarrier, $L$ is the number of symbols in a coded block, or the number of subcarriers used in an OFDM system, and $\Phi$ is the invertible function that defines the specific ESM.

In this paper, we consider the mutual information for the invertible function $\Phi$. The mutual information can be then evaluated based on [13, 35, 36]:

$$MI = MI(SNR_{eff}) = \frac{1}{L} \sum_{i=1}^{L} MI(SNR_i)$$ \hspace{1cm} (4)

Suppose that $x \in X$ and $y \in Y$ where $X$ and $Y$ are random variables. If $x$ and $y$ are identically independent distributed (i.i.d), MI between $X$ and $Y$ is then [36]:

$$MI(X;Y) = \sum_{x \in X} \int_{y} p_{XY}(y|x)p(x) \log_2 \frac{p(y|x)}{p(y)} dy$$ \hspace{1cm} (5)

Where $X$ and $Y$ denote the sample spaces of $X$ and $Y$, respectively. While $x$ are unknown, and just LLRs of the received symbols and the original alphabet of the transmitted symbols are known. Therefore, the computation of the symbol mutual information (MI) is dependent on the symbol-level log-likelihood ratio (LLR). The symbol-level LLR given that the symbol $x_i$ is transmitted, can be computed for the Maximum Likelihood receiver as follows [36]:

$$LLR_i = \ln \left( \frac{p(y|x_i)}{\sum_{k=1,k \neq i}^{M} p(y|x_k)} \right) = \ln \left( \frac{d_i^2}{\sum_{k=1,k \neq i}^{M} e^{-d_k^2/2\sigma^2}} \right) \hspace{1cm} i = 1,2, ..., M$$ \hspace{1cm} (6)

where, $d_i$, ($i = 1,2, ..., M$) is the Euclidean distance of the symbol $x_i$ from the current received symbol. Specifically, $d_i = \| y - H x_i \| = \sqrt{(y-Hx_i) (y-Hx_i)^T}$, where $x_i$ represents the $i^{th}$ symbol.

Therefore, the computation of the mutual information per coded bit can be derived as in [13, 35, 36] leading to:

$$MI = \frac{1}{M} \sum_{i=1}^{M} \int_{LLR_i} p(LLR_i) \log_2 \frac{M}{1+ e^{-LLR_i}} dLLR_i$$ \hspace{1cm} (7)

Where LLR$_i$ are the log likelihood ratios of received symbols and which is assumed to be normally distributed $p(LLR_i) = N(AVE_i, VAR_i)$ with average $AVE_i$ and variance $VAR_i$ [13, 36]. Therefore, $p(LLR_i)$ is given by:
\[ p(\text{LLR}_i) = \frac{1}{\sqrt{2\pi\text{VAR}_i}} e^{-\frac{(\text{LLR}_i-\text{AVE}_i)^2}{2\text{VAR}_i}} \] (8)

Then:

\[ \text{MI} = \frac{1}{M} \sum_{i=1}^{M} \int_{\text{LLR}_i}^{\text{LLR}_i+\text{VAR}_i} e^{-\frac{(\text{LLR}_i-\text{AVE}_i)^2}{2\text{VAR}_i}} \log_2 \frac{M}{1+e^{-\text{LLR}_i}} d\text{LLR}_i \] (9)

To find \( \text{MI} \) value in the bit level, the right side of (9) must be divided by \( \log_2(M) \) for M-QAM modulation. Therefore, \( \text{MI} \) changes between 0 and 1.

2.4. The Proposed Method for AMC/ARQ Combining

2.4.1. Performance Requirement of AMC/ARQ: We first determine the required reliable performance at the physical layer to meet the goal (BER, PER, Delay or outage probability), given that, \( N_{\text{max}} - \text{ARQ} \) (\( N_{\text{max}} \) is the maximum number of transmission) is implemented at the data link layer.

Suppose that AMC achieves an average PER (\( P_0 \)), and a packet is dropped if it is not received correctly after \( n_k \) transmissions. Referring to the assumption A4, the packet loss probability at the data link layer can be computed as \( P_0^{n_k} \). To satisfy the PER goal (\( P_{\text{loss}} \)) at the data link layer, we need to impose: \( P_0^{n_k} \leq P_{\text{loss}} \). Thus, we obtain the target PER (\( P_{\text{target}} \)) at the physical layer:

\[ P_0 \leq P_{\text{loss}}^{1/n_k} := P_{\text{target}} \] (10)

Therefore, if we use AMC to realize \( P_0 \) bounded as in (10) at the physical layer, and implement an \( N_{\text{max}} - \text{ARQ} \) at the DLL, then the number of transmissions required per packet \( n_k \); \( n_k = 1, ..., N_{\text{max}} \) will be upper bounded by \( N_{\text{max}} \) and the probability of packet loss after \( n_k \) transmissions won’t be larger than \( P_{\text{target}} \). Our problem now is to design AMC to maximize spectral efficiency while ensuring that \( P_0 \) satisfies (10).

In other words, our objective here is to maximize the average data rate, while maintaining the required performance (\( P_{\text{target}} \)) at the physical layer. As in [1], we assume constant transmission power, and partition the total SNR range into \( K + 1 \) non-overlapping consecutive regions, with boundary points denoted as \( \{y_k\}_{k=0}^{K+1} \), where \( K \) is the total number of modulation schemes. Specifically, mode \( k \) is used when \( y \in [y_k, y_{k+1}) \).

To avoid deep channel fades, the maximum number of transmissions for payload bits is sent when \( y_0 \leq y < y_1 \) is \( N_{\text{max}} \). What remains now is to determine the boundary points \( \{y_k\}_{k=0}^{K+1} \) to meet the required PER goal (\( P_{\text{target}} \)).

The adaptive system is constrained to PER goal (\( P_{\text{target}} \)). Therefore, the corresponding BER goal \( \text{BER}_{\text{target}} \) must be defined. For a packet with \( N_p \) bits, and if each bit inside the packet has the same bit error probability and bit-errors are uncorrelated, the PER can be related to the BER through \( \text{PER} = 1 - (1 - \text{BER})^{N_p} \) [37], [3]. Therefore:

\[ \text{BER}_{\text{target}} = 1 - \left( 1 - P_{\text{target}} \right)^{N_p} \] (11)

2.4.2. Switching Thresholds Determination: In our proposed design, thresholds determination relies on MI statistics. Therefore, for each modulation mode, we need to
find the relation between BER and MI, from which we find the MI threshold $M_{I_{th}}$ that corresponds the BER goal (BER_{target}). When $MI$ exceeds $M_{I_{th}}$, we expect that BER will be less than BER_{target}.

Statistical simulation results show that MI values have a random nature at each point of SNR, for which, the simulation gives us the mean and the variance values. Figure 3 shows the histogram and the CDF of MI at $SNR = 11dB$ for 64-QAM. Figure 4 shows the mean and the variance values dependencies of SNR values for 64-QAM in the domain $[0 − 25]dB$.

While the packet length and the number of packets used in the simulation are limited, we need to fit the simulated CDF to a continuous one. Fitting the distribution of MI with a normal distribution is an obvious possible choice, as the central limit theorem ensures that PDF$(MI)$ converges to a Gaussian distribution with mean AVE$[MI]$ and variance VAR$[MI]$. Since $MI$ is non-negative, it is also worth considering the approximation of PDF$(MI)$ by a Gamma distribution. Even better, as $MI$ is upper bounded by 1, the Beta distribution seems to be another natural candidate, being defined for variables in the real interval $[0, 1]$. Of course, the Gamma and the Beta are asymptotically correct, too [39] and [40]. For simplicity, we choose the normal distribution to fit the distribution of MI. Therefore, the CDF is normally fitted after finding the variance and the average values. Figure 3 shows the fitted PDF and CDF for 64-QAM at $SNR = 11dB$.

As mentioned above, we want to ensure that $MI > M_{I_{th}}$. Since MI is a random variable, this can be satisfied only for a predetermined probability $P_g$. i.e.,

$$Pr(MI > M_{I_{th}}|k, SNR) = P_g$$  \hfill (12)

Figure 3. PDF and CDF of MI with the Normally Fitted Ones for 64-QAM over Gaussian Channel at SNR = 11dB
Which leads to:

\[ \text{CDF}_{k,\text{SNR}}(MI) = 1 - p_g \]  

Then we need to resolve this equation with respect to SNR for each modulation mode. Using the normal distribution for MI in (13) leads to:

\[ 1 - p_g = \int_{-\infty}^{M_{\text{th}}} \frac{1}{\sqrt{2\pi \cdot \text{VAR}(MI)}} e^{-\frac{(MI - \text{AVE}(MI))^2}{2 \cdot \text{VAR}(MI)}} \, dMI \]  

(14)

This can be rewritten using the error function [41] as,

\[ p_g = \frac{1}{2} - \frac{1}{2} \text{erf} \left( \frac{M_{\text{th}} - \text{AVE}(MI)}{\sqrt{2 \cdot \text{VAR}(MI)}} \right) \]  

(15)

As mentioned earlier, the mean and variance \( \text{AVE}(MI) \) and \( \text{VAR}(MI) \) are functions of the modulation mode \( k \) and the SNR value. Let's note by \( \gamma_{th}^{k} \) the SNR value that satisfy (15) for a given modulation mode. We can rewrite (15) to show this dependency as follows:

\[ p_g = \frac{1}{2} - \frac{1}{2} \text{erf} \left( \frac{M_{\text{th}}(k) - \text{AVE}(MI(k,\gamma_{th}^{k}))}{\sqrt{2 \cdot \text{VAR}(MI(k,\gamma_{th}^{k}))}} \right) \]  

(16)

Therefore, the remaining question is to resolve (16) to find \( \gamma_{th}^{k} \). While \( M_{\text{th}} \) is determined depending on \( \text{BER}_{\text{target}} \) as explained later, then \( \gamma_{th}^{k} \) for each modulation mode \( k \) is function of \( p_g \) and \( \text{BER}_{\text{target}} \). Therefore, we deduce by mean of numerical resolution.
\( y_{th}^{k} = f_k(p_g, BER_{target}) \) 

(17)

The proposed method for finding an SNR threshold \( y_{th}^{k} \) numerically is described through an example. Figure 5 shows the variation of BER versus MI, for BPSK, QPSK, 16QAM and 64-QAM modulation. The packet consists of 1056 bits. At each point of SNR, 10000 packets are transmitted; each packet is transmitted one time. MI and BER are collected for each packet in order to find these curves. Figure 5 points out that BER could be estimated for a specific value of MI, provided a fixed length of the packet is adopted. Figure 6 shows a zoom in for Figure 5 for 64-QAM in a region of MI close to 1. As shown in Figure 6, we can choose an MI threshold which guarantees BER less than \( BER_{target} \). For example, in our design we look for \( BER_{target} = 10^{-3} \), therefore, \( M_I_{th} \) should be 0.9988.

![Figure 5. BER Variation Versus Average MI for M-QAM](image)

Once \( M_I_{th} \) is determined, we need to find the average MI value of the packets who’s \( M_I > M_I_{th} \) with some probability or confidence \( p_g \). This means, between CDF curves, have already obtained from the simulation, we need the CDF curve of the packet’s MI, which have \( M_I > M_I_{th} \) with a predetermined probability \( p_g \). We choose the CDF curve at which \( 1 - p_g \) is corresponding \( M_I_{th} \) value. From this curve, we can find the average value of MI \( mean(M_I_g) \). That means, the chosen CDF will satisfy (14). In our example, between CDFs curves, we find the CDF curve at which 0.1 \( (p_g = 0.9) \) of the CDF, MI is equal \( M_I_{th} \), as shown in Figure 7 (Which means that 90% of MI values will be more than \( M_I_{th} \), or the mutual information of 90% of received packets are equal or more than \( M_I_{th} \)). From that curve, we find the mean value of MI \( (mean(M_I_g)) \), which is the MI value at 0.5 of the CDF. Which means that, to get the \( mean(M_I_g) \), we find \( AVE(M_I) \) after replacing \( 1 - p_g = 0.5 \) in (14). Therefore, from Figure 7, we can find that \( mean(M_I_g) \approx 0.99925 \).
From average MI curve (as a function of SNR) of the considered modulation mode, we find SNR value that corresponds $\text{mean}(M_I_g)$, found from the last step. Figure 8 shows the relation between the average value of MI and SNR. This curve is found for 64-QAM. At $\text{mean}(M_I_g)$, found in the previous step ($\text{mean}(M_I_g) = 0.99925$), we can find the SNR value, denoted $\gamma_{th}^{k,1}$ ($k$-th for the modulation mode and 1 for the first transmission). This SNR value is the SNR threshold ($\gamma_{th}^{k,1}$) for the $k$-th modulation scheme used with one transmission, after which, we guarantee that 90% of the received packets will have a mutual information higher than $M_{I_{th}}$, means that BER will be less than $BER_{target}$. From Figure 8 (which satisfy (17)), we can see that $\gamma_{th}^{1,1} \approx 23.51$. 

Figure 6. Zoom in of MI of Figure 5 for 64-QAM between 0.996 and 1

Figure 7. CDF Curve of MI at which 0.1 is Corresponding $M_{I_{th}}$
Repeating these steps for all modulation schemes, we obtain SNR switching thresholds between modulation schemes $\gamma_{th}^{k,1}$ in our system.

We find the SNR switching thresholds of the same modulation scheme and the $n_k$-th transmission ($n_k$-1 retransmission) $\gamma_{th}^{k,n_k}$, when Chase Combining is used, using curves like mentioned above but for the $n_k$-th transmission, or simply using the relation [38]:

$$\gamma_{th}^{k,n_k} = \gamma_{th}^{k,1} - 10 \cdot \log_{10}(n_k)$$  \hspace{1cm} (18)

Till now, SNR region is divided to $k$ regions, in each one of them, a modulation mode is used. Each SNR region is also divided to $n$ sub-regions, in each one of them; an initial anticipation of number of transmissions is associated with the corresponding modulation mode. This way, at an SNR value, we can determine the modulation mode and the anticipated number of transmissions for Maximum Ratio Combining (MRC) or Chase Combining (CC). The maximum numbers of transmissions for all modulation modes at SNR sub-regions are limited by $N_{max}$.

With $\gamma_{th}^{k,n_k}$ specified as mentioned above, one can verify that the AMC with mode $k$ and $n_k$ transmission chosen when $\gamma \in [\gamma_{th}^{k,n_k} - \gamma_{th}^{k,n_k+1}]$ guarantees the PER in (10) with confidence factor $P_g$.

Summarizing our results in the previous and this subsection, the AMC/ARQ is designed following these steps:

**Step 1:** Given the limit of maximum number of transmissions and the PER goal, determine the $P_{target}$ from (10).

**Step 2:** For this $P_{target}$, determine $\{\gamma_{th}^{k,n_k}\}$ via the explained procedure.
The proposed cross-layer design hence leads to the following operating stages in the overall system (as explained in 2.1):

**Stage 1:** Update transmission modes \((k \text{ for the modulation mode and } n_k \text{ for the number of transmission})\) for each frame by using AMC/ARQ.

**Stage 2:** Start the first transmission using the modulation mode \(k\) and the number or transmission \(n_k\).

**Stage 3:** Retransmit erroneous packets by \(N_{\text{max}} - \text{ARQ}\).

### 3. Performance Analysis of AMC/ARQ

In this section, we derive the average PER and the spectral efficiency of our cross-layer design. For applications where the assumption A4 is violated, even though, PER goal and maximum number of transmissions are still guaranteed.

Since the instantaneous PER is upper-bounded by \(P_{\text{target}}\) in our AMC design, the average PER at the physical layer will be lower than \(P_{\text{target}}\). We first evaluate this average PER at the physical layer.

The PER (\(\text{PER}_k\)) for the modulation scheme \(k\) (MCS\(^k\)) is given by [3]:

\[
\text{PER}_k = \frac{1}{\text{Pr}(k)} \Gamma(m) \left(\frac{m}{\gamma}\right)^m \times \frac{\Gamma(m, b_k \gamma_k) - \Gamma(m, b_k \gamma_k + 1)}{(b_k)^m}
\]  

(19)

Where, \(a_k\) and \(b_k\) are fitting parameters of the simulated PER [3], [6], and:

\[
\text{Pr}(k) = \int_{y_k}^{y_{k+1}} p(y)dy = \frac{\Gamma(m, \frac{m y_k}{\gamma}) - \Gamma(m, \frac{m y_k + 1}{\gamma})}{\Gamma(m)}
\]  

(20)

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. [1, eq. (35)])

\[
\text{PER} = \frac{\sum_{k=1}^{K} R_k \text{Pr}(k) \text{PER}_k}{\sum_{k=1}^{K} R_k \text{Pr}(k)}
\]  

(21)

Since ARQ is implemented at the data link layer, the packets in error during the original reception may be retransmitted, up to \(N_{\text{max}}\) times. For notational brevity, let us define \(p := \text{PER}\), and the average number of anticipated number of transmission is \(\bar{N}_a\). Therefore, the actual maximum number of retransmission will be \(N_{a,\text{max}} = N_{\text{max}} - \bar{N}_a\). By considering the assumption A4, the average number of transmissions per packet can be found as ([42, p. 397]):

\[
\bar{N}(p, N_{a,\text{max}}) = 1 + p + p^2 + \ldots + p^{N_{\text{max}} - \bar{N}_a - 1} = \frac{1 - p^{N_{\text{max}} - \bar{N}_a + 1}}{1 - p}
\]  

(22)

When \(N_{\text{max}} = 1\), then \(\bar{N}_a = 1\), we have, \(\bar{N}(p, N_{a,\text{max}} = 1) = 1\) which corresponds to the special case of AMC-only. With the average PER in (21), the actual packet loss probability at the data link layer with \(N_{\text{max}} - \text{ARQ}\) is:

\[
P_{\text{actual loss}} = p^{N_{a,\text{max}}} \leq P_{\text{target}}^{N_{a,\text{max}}} \leq P_{\text{target}}^{N_{\text{max}}} = P_{\text{loss}}
\]  

(23)
It verifies: $b_0^{N_{a,max}} \leq P_{\text{target}}$

When $b_0^{N_{a,max}} \leq P_{\text{target}}$ and the maximum number of transmissions allowed per packet is $N_{a,max}$. We are now ready to evaluate the achieved system spectral efficiency. When mode $k$ is used, each transmitted symbol will carry $R_k = R_c \log_2(M_k)$ information bits for the mode adhering to a $M_k$-QAM constellation, and a rate $R_c$ FEC code. For uncoded transmission modes, we set $R_c = 1$. As in [1], we assume a Nyquist pulse shaping filter with bandwidth $B = 1/T_s$, where $T_s$ is the symbol rate. Therefore, the average spectral efficiency (bit rate per unit bandwidth) achieved at the physical layer without considering possible packet retransmission is (similar to [1] where only physical layer AMC design is considered)

$$\bar{S}e_{\text{physical}} = \sum_{k=1}^{K} R_k \Pr(k)$$  \hspace{1cm} (24)

When ARQ is implemented, each packet, and thus each information bit, is equivalently transmitted $\bar{N}(p, N_{a,max})$ times. Hence, the overall average spectral efficiency, as a function of $N_{a,max}$, is obtained as

$$\bar{S}e(N_{a,max}) = \frac{\bar{S}e_{\text{physical}}}{\bar{N}(p, N_{a,max})} = \frac{1}{\bar{N}(p, N_{a,max})} \sum_{k=1}^{K} R_k \Pr(k)$$  \hspace{1cm} (25)

Setting $N_{max} = 1$, then $N_{a,max} = 1$. We obtain from (25) the average spectral efficiency for AMC-only as:

$$\bar{S}e(N_{max} = 1) = \sum_{k=1}^{K} R_k \Pr(k)$$  \hspace{1cm} (26)

The form in (26) is in agreement with [1] and (24), where the AMC design is considered only at the physical layer.

While we consider an anticipated number of transmissions as one transmission (the transmitter won’t wait an acknowledgement before finishing the anticipated number of transmissions), $N_{a,max}$ in our new design is less than $N_{max}$ in the traditional design. Therefore, more transmission could be performed in the new design to reach the maximum number of transmission as the traditional one. Then $\bar{N}(p, N_{a,max}) \leq \bar{N}(p, N_{max})$. Then, from (25) we deduce:

$$\bar{S}e(N_{a,max}) \geq \bar{S}e(N_{max})$$  \hspace{1cm} (27)

Where $\bar{S}e(N_{max})$ is the spectral efficiency of the traditional design [3]. Therefore, from relation (27), we can expect that the spectral efficiency of the new design will be higher than which for the traditional one. From relation (23), we can also deduce that error probabilities in our new design will be less than error probabilities in the traditional one. These results are examined in the following section.

4. Simulation and Numerical Results

In this section, we present numerical results. We set the packet length 1056. Specific numerical values will be affected if one chooses a different $N_p$. However, similar observations are expected, because of the unifying development in Section 2.

Let the performance constraints $P_{\text{target}} = 10^{-3}$ and $N_{max} = 4$. We set the Nakagami fading parameter $m = 4$, which corresponds to the existence of line of
sight component between the transmitter and the receiver. Table 1 gives the anticipated number of transmissions for each modulation mode. For comparison purposes, we have simulated a conventional adaptive modulation system with Type-1 HARQ adopting the same constraints ($P_{\text{target}}$ and $N_{\text{max}}$) as in our new design. We have adopted the same procedure followed by [3], [6] to find SNR switching thresholds, in order to simulate the conventional system.

Table 1. Switching Thresholds Between Modulation Modes

<table>
<thead>
<tr>
<th>Modulation</th>
<th>$n_k$</th>
<th>$\gamma_{th}^{k,n_k}$ (dB)</th>
<th>$N_{\text{max}}$</th>
<th>$\gamma_{th}^k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPSK</td>
<td>4</td>
<td>0.7689</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>BPSK</td>
<td>3</td>
<td>2.0183</td>
<td>4</td>
<td>6.4613</td>
</tr>
<tr>
<td>BPSK</td>
<td>2</td>
<td>3.7792</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPSK</td>
<td>1</td>
<td>6.7895</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QPSK</td>
<td>2</td>
<td>7.2414</td>
<td>4</td>
<td>9.5829</td>
</tr>
<tr>
<td>QPSK</td>
<td>1</td>
<td>10.2517</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-QAM</td>
<td>3</td>
<td>12.4702</td>
<td>4</td>
<td>16.2107</td>
</tr>
<tr>
<td>16-QAM</td>
<td>2</td>
<td>14.2311</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-QAM</td>
<td>1</td>
<td>17.2414</td>
<td>4</td>
<td>22.7821</td>
</tr>
<tr>
<td>64-QAM</td>
<td>3</td>
<td>18.7388</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>64-QAM</td>
<td>2</td>
<td>20.4997</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>64-QAM</td>
<td>1</td>
<td>23.5100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>256-QAM</td>
<td>3</td>
<td>24.6288</td>
<td>4</td>
<td>23.4488</td>
</tr>
</tbody>
</table>

Figure 9 shows the average spectral efficiency of our new designed AMC/ARQ and for the conventional design. Figure 9 shows also the Nakagami fading channel capacity for $m = 4$. Figure 9 points out that, the average spectral efficiency of our new design is closer to the Nakagami Channel capacity than the traditional design. About 2dB of gain (equivalent 0.5 bit per symbol per hertz) for our new design against the conventional one at high SNR is observed. This gain decreases with decreasing SNR to vanish at low SNR. To improve spectral efficiency over the entire SNR range, a practical system could also combine uncoded and coded transmission.
Figure 9. Average Spectral Efficiency for AMC/ARQ Design Versus SNR. Dotted Line for NFC \((m = 4)\), Continuous Line for our New Method, Dashed Line for the Conventional Method

The average packet error rates and bit error rate at the physical layer, for the new and conventional design, are depicted in Figure 10 and Figure 11. These figures point out that the two designs verify the mentioned goals (PER and BER). From Figure 10, we can see that PER for our design is less than the PER of the conventional one. The PER for the conventional design is too close to the PER goal. Since enhancing spectral efficiency will affect the PER negatively, no chance to enhance the spectral efficiency at the expense of PER. While in our design, we have more margin in PER to lose in order to enhance spectral efficiency. This point my cast a case study to enhance the spectral efficiency using some optimization methods as in [27-29]. The same conclusions can be extracted from Figure 11.
Figure 10. Average BER for AMC/ARQ Design Versus SNR. Continuous Line for our New Method, Dashed Line for the Conventional Method

Figure 11. Average PER for AMC/ARQ Design Versus SNR. Continuous Line for our New Method, Dashed Line for the Conventional Method
5. Conclusion and Future Work

In this paper, a cross-layer design, which combines adaptive modulation and coding at the physical layer with ARQ at the data link layer, is developed in order to maximize system throughput under prescribed performance constraint. We derived a closed-form expression of the average spectral efficiency for packets transmitted over Nakagami block fading channels, in addition to packet loss. Numerical results demonstrated the throughput improvement of our cross-layer design over the conventional one. More margin in packet loss is observed from simulation results, therefore, more chance to enhance the spectral efficiency. This point may cast a case study to enhance the spectral efficiency using some optimization method. Diminishing returns appear on the spectral efficiency improvement as the number of retransmissions increases, which suggests that a small number of retransmissions strikes a desirable delay-throughput tradeoff in practice. To improve spectral efficiency over the entire SNR range, a practical system could also combine uncoded and coded transmission.

Although, the assumptions that perfect CSI is available at the receiver, and that the feedback channel has zero delay and is error free, may not always hold true. While the feedback in our design is not a measured value of CSI, but a modulation and a number of retransmission selections. Therefore, the above assumption may be hold. However, one possible extension of this work is to design and analyze our cross-layer design with imperfect CSI at the transmitter. Finally, the impact of coding on our design is also worth pursuing.

References


Authors

Assal Alkhoder, was born in Damascus, Syria in 1978. He received the B.Eng. Degree in electronic engineering from Albaath university, Homs, Syria in 2002, diploma of Higher studies in communications Engineering from Damascus University in 2003, and M.S. degree in high frequency materials and components for communication systems from université de Bretagne Occidentale (UBO), Brest, France, in 2010. He is currently working towards the Ph.D. degree at the Higher Institute for Applied Sciences and Technology, Damascus, Syria. His research interests are on Enhancing wireless communication systems.

Abdelnasser Assimi, was born in Derra, Syria in 1973. He received the Eng. degree from the ENST, Paris, France in 1996 and the Ph.D. degree from the University of Cergy-Pontoise, Cergy-Pontoise, France in 2009. He is currently an Assistant Professor and Head of Digital Electronics Lab at the Higher Institute for Applied Sciences and Technology, Damascus, Syria. His research interests include Man-Machine communication, mobile communication, and Turbo-processing and its application to communication systems.

Mohammad Alhariri, was born in Damascus Syrian Arab Republic in 1966. Received the electrical engineering degree in 1991 from the Damascus University. He received his Master degree in Antennas and Microwave Device at the Bauman Moscow State Technical University in 1997. He received his Ph.D. degree in Antennas, microwave devices and their technologies at the Moscow Aviation Institute in 2007. His research interests are in Electrodynamics modeling of printed slit antennas. He work in the Higher Institute of Applied Science and Technology in Damascus (HIAST-Syria).