ARQ strategies for MIMO eigenmode transmission with adaptive modulation and coding

De Carvalho, Elisabeth; Popovski, Petar

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
ARQ strategies for MIMO eigenmode transmission with adaptive modulation and coding

Elisabeth de Carvalho and Petar Popovski
Aalborg University, Niels Jernes Vej 12
9220 Aalborg, Denmark
email: {edc,petarp}@es.aau.dk

Abstract—Packet retransmission strategies are presented for MIMO eigenmode transmission where adaptive modulation and coding (AMC) is implemented. The retransmission design is based on weighted linear MMSE. It includes the transmit and receiver filter, the power and eigenmode allocation and AMC level when new packets are transmitted. The weight matrix of WMMSE is used to appropriately weight streams with different AMC levels in order to maximize the system throughput. Simulations show that the choice of the weight factors has a major impact on the performance.

I. INTRODUCTION

Automatic Repeat Request (ARQ) design for MIMO communications [1] offer many challenges mainly due to the multiplicity of spatial channels for simultaneous transmission. ARQ design includes the selection of the spatial channel used for retransmission as well as for transmission of new packets. This paper focuses on ARQ for MIMO eigenmode transmission. The transmitter is assumed to have full channel state information (CSI) which remains approximately constant over a packet transmission. Based on the CSI, the transmit filter, the power allocation and the adaptive modulation and coding (AMC) level are adjusted.

In eigenmode transmission, independently coded streams are subject to a linear processing which depends on the singular vectors of the MIMO channel. The total transmit power is constrained to be smaller than a set value and the power assigned to each stream depends on the singular values of the channel. When all the input packets are transmitted for the first time, several criteria can be used to optimize the transmit filter and power allocation. Popular criteria are the maximization of capacity and weighted linear minimization mean squared error (WMMSE). When the weight matrix of WMMSE is appropriately selected, capacity maximization and WMMSE give the same solution [2].

When packets are decoded in error, they are retransmitted with the same data content, meaning that the AMC level for the retransmitted packets is kept unchanged. Only packets decoded with errors are retransmitted. Soft information about the erroneously decoded packets are kept in memory and combined with the retransmitted packets to increase the probability of correct detection. Retransmission optimization includes: 1) the choice of the spatial channel for retransmitted packets and newly transmitted packets, 2) the power allocation among streams, 3) the transmit filter 4) AMC level for new packets.

Let us take the example of a $2 \times 2$ MIMO system where 2 packets are newly transmitted. A criterion based on capacity maximization or (linear) WMMSE can be used to optimize the transceiver parameters. The result of the optimization is optimal if the input streams are Gaussian. In practice, the input data belongs to a discrete constellation resulting in a loss of optimality. When adaptive modulation and coding (AMC) is applied and as the number of possible AMC level increases, this suboptimality is compensated for.

Assume that the 2 transmitted packets are decoded incorrectly and have to be retransmitted. The transceiver parameters need to be optimized taking into account packet combining with previous transmission. The main difference with the first transmission optimization is that the AMC level is now fixed. The design should account for the fact that the input data belongs to a discrete constellation. This increases the computational complexity of the solution in general. For this reason, it is preferable to keep a linear criterion, such as linear WMMSE. However, because the AMC level is fixed, performance of linear WMMSE can actually be quite poor. We propose to use linear WMMSE but select appropriately the weight matrix. The weighting between 2 streams cannot be the same in the case where the 2 packets have the same AMC level and the case where they have a different one. Assume that one of the streams carry a 16-QAM and the other stream a QPSK. The first stream requires a higher post-processing SNR to get good probability of correct detection. This can be achieved by giving more weight to the first stream. After the first transmission, when only 1 packet is decoded with errors and the other one is decoded correctly, the packet in error gets retransmitted and a new packet is transmitted. The design additionally involves the choice for the AMC level for the new packet.

In [3], a retransmission design for MIMO eigenmode transmission was proposed based on linear MMSE. The design criterion in [3] is not changed to account for the fact that the retransmitted packets have a fixed constellation which can affect performance as seen in the simulation section in our paper. Furthermore, [3] does not consider to transmit new packets when only a subset of the packets is decoded with errors thus limiting the spatial efficiency.

The paper is organized as follows. Section II recalls the two main criteria used when packets are newly transmitted. Section III presents the design proposed in this paper. As last, section IV presents simulations. Only the $2 \times 2$ case
is presented. However, the principles explained here can be extended to more general cases. Furthermore, we treat the case of 2 consecutive transmission periods, where new packets are transmitted during the first period and retransmissions occur during the second period. Again, the general case can be constructed recursively based on the content of this paper.

II. DESIGN FOR FIRST TRANSMISSION

We consider the first transmission of two packets denoted $\tilde{X}_A^{t=k}$ and $\tilde{X}_B^{t=k}$ from antenna 1 and antenna 2 respectively at time $t = k$. The $m^{th}$ symbols of the packets are denoted as $\tilde{z}_A^{t=k}$ and $\tilde{z}_B^{t=k}$. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ denote respectively the conjugate, the transpose and Hermitian transpose operations. The vectorial input-output relationship for the $2 \times 2$ MIMO system is:

$$y^{t=k} = H^{t=k}x^{t=k} + n^{t=k}$$

where $y^{t=k}$ is the received signal at antenna 1 and antenna 2, $x^{t=k} = [\tilde{z}_A^{t=k}, \tilde{z}_B^{t=k}]^T$ is the vector of input data, $n^{t=k} = [n_1^{t=k}, n_2^{t=k}]^T$ is the received noise at antenna 1 and antenna 2. It is assumed to be a centered complex circularly symmetric Gaussian random variable: $n \sim CN(0, \sigma_n^2 I)$. The channel matrix $H^{t=k} = [H_1^{t=k}, H_2^{t=k}]$, where $H_i^{t=k}$ is the $i^{th}$ column of $H^{t=k}$.

Two well-known design criteria for transmission of new packets are presented below. They are used to optimize the transmit and receive filters as well as the transmit power allocation among eigenmodes. The first criterion relies on the maximization of capacity and the second criterion on weighted linear minimum mean squared error (WMSE). WMMSE becomes equivalent to capacity maximization when the weight is appropriately selected [2]. This correspondence between capacity and WMMSE motivates the design approach adopted for retransmissions in this paper where maximization of the throughput is the purpose. As explained in the introduction, retransmissions imply a decreased number of degrees of freedom as the AMC level of the retransmitted packet has been fixed in a previous transmission. Information theory or (non linear) MMSE criteria can be developed based on a fixed AMC level but result in high computational complexity. In this paper, we rely on linear WMMSE and adjust the weight matrix to adapt to the AMC level of the streams in order to maximize the throughput.

A. Capacity maximization

When the full CSI (channel and noise variance at receiver) is available at the transmitter and receiver, the system capacity has the following expression:

$$C = \max_{R_{xx}, R_{xx} \leq P} \log_2 \det \left( I + \frac{1}{\sigma_n^2} HR_{xx}H^H \right)$$

where $x = [x_A \ x_B]^T$, $R_{xx} = ExxH$. The total transmit power is constrained to be smaller than a set value $P$. As the $i^{th}$ diagonal element of $R_{xx}$ is equal to the transmit power $P_i$ from antenna $i$, the power constraint can be written as trace($R_{xx}$) $\leq P$. The optimal correlation matrix can be written as

$$H = U\Lambda V^H.$$  

where $U$ and $V$ are unitary matrices containing the left and right singular vectors. $\Lambda$ is a diagonal matrix containing the singular values of $H$ denoted as $\lambda_i$.

The optimal correlation matrix is $R_{xx} = V\Phi^2V^H$, $\Phi^2$ is a diagonal matrix with element $(i,i)$ equal to $P_i$. The optimal transmitter consists in a) independently coded streams using an AWGN encoder 2) a transmit filter $F$:

$$F = V\Phi$$

and a power assignment following the waterfilling policy:

$$\Phi = \left(\mu^{-1/2} - \sigma_n^2 \Lambda^{-1}\right)^{1/2} \Lambda$$

where $(z)_+ = z$ if $z > 0$ and $(z)_+ = 0$ otherwise. $\mu$ is adjusted to conform to the transmit power constraint $P_1 + P_2 \leq P$. Capacity maximization imply that the receiver is optimal (based on maximum likelihood).

B. Weighted MMSE

We impose a transmit and receive structure consisting of a linear filter at the transmitter denoted $F$ and a linear filter at the receiver denoted $G$. The vector input of $F$ is denoted as $x$. The output vector is $\hat{x} = Fx$. The vector output of $G$ is $G\hat{y} = G(HFx + n)$. Denoting the weight matrix as $W$ and the weighted error as $W^{1/2}e$, the weighted MMSE criterion is:

$$\min_{F,G} E\|W^{1/2}e\|^2, \ e = x - G(HFx + n).$$

The solution is of the form $F = V\Phi F$ and $G = U^H\Phi G$, with:

$$\Phi_F = \left(\mu^{-1/2} - \Lambda^{-1/2}W^{-1/2} - \Lambda^{-1}\right)^{1/2}$$

$$\Phi_G = \left(\mu^{-1/2} - \Lambda^{-1/2}W^{-1/2} - \mu\Lambda^{-1}W^{-1}\right)^{1/2} \Lambda^{-1/2}$$

For $W = \Lambda$, the transmit filter is the same as the one resulting from capacity maximization. Hence, the transmitter structure is the same for both cases.

C. AMC Level

The capacity based criterion and linear WMMSE are optimal when the input signals are Gaussian. In practice, the input signals belong to a discrete constellation and the number of possible AMC level is finite. A known approach is to use an alternate criterion aiming at maximizing the throughput sometimes under a minimum bit error rate constraint [4]. Let $PER_K(\gamma_K)$ denote the packet error rate of the $K^{th}$ stream: it depends on the selected AMC and the post-processing SNR of the stream $\gamma_K$. The throughput for stream $K$ is:

$$r_K = \frac{1}{[1 - PER_K(\gamma_K)]}.$$  

$r_K$ is the nominal throughput for stream $K$. In this paper, the AMC level is selected as the one maximizing the throughput (9). In practice, SNR intervals can be associated with each AMC level.
III. Retransmission Design

Linear WMMSE is optimal if the input data is Gaussian. In reality, the input data belongs to a finite alphabet. However, in general, the number of AMC levels is sufficiently high so that the suboptimality gets decreased. Weighted and non weighted MMSE gives reasonable performance in this case. The AMC level of the retransmitted packet is fixed however. This should be taken into account by an appropriate design.

Next, we formulate a criterion based on the maximization of the total system throughput: this is the criterion we would like to maximize but its optimization is too costly. As an alternative, we adopt WMMSE where the weight matrix will be tuned to maximize the throughput.

A. Main Assumptions

We recall the main hypotheses of the retransmission process.

- For each packet transmission, transmit and receive filters as well as transmit power of each stream and eigenmode allocation are determined.

- Packets transmitted over different eigenchannels might have a different AMC level. Packets are assumed to have the same duration: they contain the same number of symbols, but possibly a different number of bits. Retransmitted packets keep the same symbol content.

- Received signals containing contributions of erroneously decoded packets are kept in memory and are combined with received signals containing retransmitted packets. When packets are successfully decoded, their contribution is removed from the current received signal and possibly from received signals corresponding to previous transmissions.

- If 2 packets are decoded with errors at time $t$, they are retransmitted at time $t + 1$.

- If only 1 packet is successfully decoded, its contribution is removed from the received signal at time $t$ and possibly from past received signals kept in memory. The erroneously received signal is retransmitted at time $t + 1$ as well as a new packet.

- When both packets are decoded correctly at time $t$, 2 new packets are transmitted at time $t + 1$.

B. Throughput Maximization

If the packets transmitted at time $t$ are newly transmitted, the system throughput can be written as: $r_A \left(1 - PER_A(\gamma_A)\right) + r_B \left(1 - PER_B(\gamma_B)\right)$. This expression is not valid in the retransmission case, because it does not take into account that packet errors occurred in previous transmission periods.

However, the following throughput maximization criterion:

$$\max r_A \left(1 - PER_A(\gamma_A)\right) + r_B \left(1 - PER_B(\gamma_B)\right)$$

(10)

was used in [5] where it was shown to be a good approximate criterion for retransmission. There appears to be no low cost solution for determining the transmit and receive filter using this criterion. WMMSE will be used. However, we will come back to this criterion to determine the power allocation.

C. Weighted MMSE

1) Equivalent channels: Suppose that packets $X_A$ and $X_B$ are newly transmitted at time $t = 1$. Both packets are first processed by the transmit filter $F^{t=1}$. Here, we assume that $F^{t=1}$ incorporates the power allocation information as in section II-B. The output of the transmit filter is $F^{t=1}x^{t=1}$. We denote $H^{t=i} = H^{t=i}F^{t=i}$ at time $t = i$ and $H^{t=1}_j$ its $j^{th}$ column.

If both packets $X_A$ and $X_B$ are decoded with errors, they are both retransmitted through channel $H^{t=2}$.

2) Transmit and receive filters: A full length derivation of the transmit and receive filters $F^{t=2}$ and $G^{t=2}$ can be found in [3]. Here we give an alternate and brief derivation that is based on the results in [2].

The weighted MMSE criterion is:

$$\min_{F^{t=2}, G^{t=2}} E\left[\|W^{1/2}(G^{t=2}y - x^{t=2})\|^2\right].$$

(14)

$G^{t=2}$ is the receive filter at time $t = 2$ and $W$ is the diagonal weight matrix. The solution for $G^{t=2}$ is the classical receive MMSE filter based on the received signal $y$:

$$G^{t=2} = (\sigma_n^2 I + \rho^2 G^{t=1})^{-1} H^{t=2}.$$

(15)

The transmit filter $F^{t=2}$ is solution of:

$$\min_{F^{t=2}} \text{tr} \left[ W \left( D + \frac{1}{\sigma_n^2} F^{t=2} H^{t=2} H^{t=2} F^{t=2} \right)^{-1} \right].$$

(16)

where $D = I + \frac{1}{\sigma_n^2} F^{t=1} H^{t=1} H^{t=1} F^{t=1}$.

The cost function can be rewritten as: $\text{tr} \left[ D^{-1/2} W D^{-1/2} \left( I + \rho D^{1/2} F^{t=2} H^{t=2} H^{t=2} F^{t=2} D^{1/2} \right) \right]$. Denoting $\tilde{F}^{t=2} = F^{t=2} D^{1/2}$, the optimization problem for $\tilde{F}^{t=2}$ is similar to the problem in [2] with weight matrix equal to $D^{-1/2} W D^{-1/2}$. The main difference is in the
power constraint which becomes: $\text{tr}(F^{t=2}H F^{t=2}) = \text{tr}(D^{1/2} (\tilde{F}^{t=2}H \tilde{F}^{t=2}D^{1/2}) \leq \bar{P}$. Using the same kind of arguments as in [2] to find the power allocation, $F^{t=2}$ verifies:

$$F^{t=2} = V^{t=2} \Phi^{t=2}$$

$$\Phi^{t=2} = (\mu^{-1/2} [\Lambda^{t=2}]^{-1/2} W^{1/2} - [\Lambda^{t=2}]^{-1} D)^{1/2}$$

The structure of $F^{t=2}$ is similar to the case where the packets are all newly transmitted. The difference is in the power allocation between streams which depends on the previous transmission. The weight matrix $W$ only impacts the power allocation and not the transmit filter. Based on these last observations, the idea is:

(a) base the design of the transmit filter $F^{t=2}$ on the (unweighted) MMSE criterion.

(b) the power allocation is determined using an alternate criterion. A given power allocation corresponds to a given weight matrix $W$. So the alternate criterion serves also to select the weight matrix.

D. Power allocation

Plugging the optimal expression of $F^{t=2}$, the MMSE is:

$$\text{MMSE} = \left( I + \frac{1}{\sigma_n^2} \Lambda_1 \Phi(1^2) + \rho \Lambda_2 \Phi(2^2) \right)^{-1}.$$  \hspace{1cm} (19)

The post processing SNR for stream $K$ is chosen as the output of the unbiased MMSE receiver [6]:

$$\frac{1}{\text{MMSE}_{K,K}} - 1$$  \hspace{1cm} (20)

where MMSE$_{K,K}$ is $K^{th}$ diagonal element the $2 \times 2$ matrix MMSE. Denoting $p^{t=i}_A$ the power on stream A at time $t = i$, the post-processing SNR is:

$$\gamma_A = \frac{1}{\sigma_n^2} \Lambda_1^{t=1} p^{t=1}_A + \frac{1}{\sigma_n^2} \Lambda_1^{t=2} p^{t=2}_A$$  \hspace{1cm} (21)

$$\gamma_B = \frac{1}{\sigma_n^2} \Lambda_2^{t=1} p^{t=1}_B + \frac{1}{\sigma_n^2} \Lambda_2^{t=2} (\bar{P} - p^{t=2}_A).$$  \hspace{1cm} (22)

The ideal would be to choose the power allocation which minimizes the throughput (10) as:

$$\min_{p_A} r_A \text{PER} A(\gamma_A) + r_B \text{PER} B(\gamma_B).$$  \hspace{1cm} (23)

When the AMC level on stream A and B is the same, the optimization problem has an easy solution for most cases. It can be proven that maximizing (23) is equivalent to finding the power allocation that makes $\gamma_A$ equal to $\gamma_B$, in the domain where the cost function is concave, which is the case for medium to high SNR [5].

This finding motivates the treatment of the case of different AMC levels on both streams. Indeed, we propose to optimize the power allocation as the one verifying:

$$\gamma_A = \beta_{AB} \gamma_B$$  \hspace{1cm} (24)

$\beta_{AB}$ is determined through simulations for each combination of AMC levels by averaging over many channel realizations. In the simulation part, we will see that this gives performance closed to the optimal one.

1) Eigenmode allocation:

a) 2 packets in error: The eigenmode allocation of the retransmitted packets is selected as the one minimizing $|\gamma_A - \beta_{AB} \gamma_B|$ for a predetermined fixed power allocation (power equally distributed among eigenmodes). This insures a more efficient optimal power allocation. Indeed, in some cases, it is not possible to make $\gamma_A$ equal to $\beta_{AB} \gamma_B$ by playing on power allocation. But, by insuring that $|\gamma_A - \beta_{AB} \gamma_B|$ is minimal for the predetermined power allocation, we maximize the chance of satisfying the targeted criterion.

In [3], the eigenmode allocation is switched compared to the previous transmission. If one packet was transmitted from the strongest eigenmode of the channel at time $t = 1$, then it is retransmitted from the weakest eigenmode of the channel at time $t = 2$. This allocation can be detrimental especially if the streams have different AMC levels. This aspect will be highlighted in the simulation section.

b) 1 packet in error: When only 1 of the packets is decoded with errors, a new packet is transmitted for which the AMC level has to be decided. We test all the AMC levels for this new packet and select the eigenmode allocation that minimizes $|\gamma_A - \beta_{AB} \gamma_B|$. When the AMC level for the new packet is too high, an allocation such that $\gamma_A = \beta_{AB} \gamma_B$ is not possible. So the rule is to select the highest AMC level for which $\gamma_A$ can be made equal to $\beta_{AB} \gamma_B$.

IV. Simulations

The simulation part treats the case where the streams have different AMC levels. Indeed, the performance difference between the power allocation proposed in this paper and an unweighted MMSE approach as in [3] is mostly noticeable in this case.

The packets are uncoded and AMC levels can be selected among BPSK, 4-QAM, 16-QAM, 64-QAM. For first transmission, we select channel realizations that result in a 16-QAM for the first stream and 4-QAM for the second stream. The value of $\beta_{AB}$ in equation 24 is determined by simulations and is equal to 6.32. The channel follows a Kronecker model with same transmit and receive correlation equal to 0.3. For each packet transmission, a new channel realization is drawn independent of the previous one.

In figure 1, 2, 3, we show the performance measure for the second transmission. Three different scenarios are tested:

Scen 1 The first transmission results in 2 packets in error. Two packets are retransmitted. As $r_A$ and $r_B$ are the same for the 3 designs that we compare, we show the average value of $r_A \text{PER} A(\gamma_A) + r_B \text{PER} B(\gamma_B)$ for the second transmission as a function of average SNR.

Scen 2 The stream with strongest AMC level is decoded with errors and the other stream is decoded correctly. We show the average value of $r_A \text{Thr} A(\gamma_A) + r_B \text{Thr} B(\gamma_B)$ as a function of average SNR.

Scen 3 The stream with weakest AMC level is decoded with errors and the other stream is decoded correctly. We show the average value of $r_A \text{Thr} A(\gamma_A) + r_B \text{Thr} B(\gamma_B)$ as a function of average SNR.

The following designs are compared:
Design 1: Power allocation optimizes (23).
Design 2: Power allocation optimizes (24).
Design 3: Power allocation using unweighted WMMSE. Eigenmode allocation is switched.

Design 3 corresponds to [3]: in [3], when only 1 packet is decoded correctly and the other one is decoded with errors, the error free packet gets retransmitted along with the erroneous packet during the subsequent transmission period. Here, we modify this design. The contribution of the error free packet is removed from the received signal at time $t = 1$. At time $t = 2$, the erroneous packet gets retransmitted and a new packet is transmitted.

The following conclusions can be drawn from the simulations: a) Design 2 gives performance very close to the more optimal Design 1, b) Design 3 gives worse performance than Design 1: power allocation should take into account the difference of AMC levels between the streams.

V. CONCLUSION

This paper has proposed a design for retransmissions in MIMO eigenmode transmission. To achieve a low computational complexity, we have based our design on linear weighted MMSE. We have shown how important it is to appropriately weight the streams of the MIMO transmission, especially when the AMC level of the streams is different. Using unweighted MMSE results in performance that can be quite far from optimal. This problem is specific to a retransmission process where packets are retransmitted with the same symbol content. When packets are newly transmitted and provided that the number of AMC levels is not too limited, non weighted MMSE performs reasonably well. In the retransmission process, the AMC levels are fixed: the weight matrix has be to appropriately selected and depends on the AMC levels of each stream.

REFERENCES