Low Complexity Link Abstraction for Retransmission in LTE/LTE-Advanced with IR-HARQ

Imran Latif, Florian Kaltenberger, Raymond Knopp, Joan Olmos**
Eurecom, Campus SophiaTech
Les Templiers, Route des Chappes
06410 Biot France
**Escola dEnginyeria de Telecomunicació i Aeroespacial de Castelldefels (EETAC-UPC) Esteve Terradas,
7. 08860 Castelldefels, SPAIN

Phone: +33 4 93 00 81 67
Fax: +33 4 93 00 82 00
Email: {latif,kaltenbe,knopp}@eurecom.fr
Low Complexity Link Abstraction for Retransmission in LTE/LTE-Advanced with IR-HARQ

Imran Latif*, Florian Kaltenberger*, Raymond Knopp*, and Joan Olmos†
*EURECOM, Sophia Antipolis, France, Email: firstname.name@eurecom.fr
†Universitat Politecnica de Catalunya, Castelldefels, Spain, Email: olmos@tsc.upc.edu

Abstract—Recent wireless cellular systems, like 3GPP's Long Term Evolution (LTE) and LTE-Advanced (LTE-A), benefit from many enhancements of wireless techniques. These include multicarrier modulation schemes, fast link adaptation, forward error correction with HARQ, powerful turbo codes and multiple antenna techniques. System level simulators usually use link abstraction models for capturing the effects of these enhancements from link level simulators. This reduces both time and complexity of large system simulations. Also the link abstraction models are used in the mobile stations for an efficient feedback of instantaneous channel quality indicator in real systems. Therefore, it is extremely important for a link abstraction model to not only have low complexity but also very good accuracy. However modeling the effects of incremental redundancy (IR)-HARQ in LTE-based systems is a challenging and interesting task. In LTE-based systems two types of gains can be achieved with IR-HARQ at the receiver. These are coding gain, which is achieved through the transmission of new parity bits in each of the retransmission, and an SNR gain, which is achieved due to repetition of coded bits in each of the retransmission. A good link abstraction model should be able to abstracts the effects IR-HARQ and other enhancements with as much low complexity as possible and reasonable accuracy. In this paper we propose a low complexity link abstraction which abstracts the performance of LTE-based link level simulators with multiple antennas, incremental redundancy (IR) HARQ and other enhancements. The complexity is reduced by the fact that our proposed scheme only uses three reference curves corresponding to the mother code rate for each type of QAM modulation. Then we show that how the reference curves for an arbitrary code rate can be generated by using these three reference curves and an offset. This offset is based on the effective code rate after combining the transmissions from different HARQ rounds. To strengthen the case for our proposed link abstraction model we show the results for LTE-based wireless system. The results show that our method provides reasonably accurate results with very low complexity and even no calibration of adjustment factors is performed. For all the simulations we used Eurecom’s OpenAirInterface platform which implements physical layer of LTE release 8 as per the standards of 3GPP.

I. INTRODUCTION

In recent cellular systems inclusion of hybrid automatic repeat request (HARQ) and fast link adaptation (LA) at the physical layer is the key performance enabler. It helps in increasing the throughput and spectral efficiency of the system without the need of increasing the bandwidth or the transmit power [1]. Despite their benefits at the link level, their true evaluation is performed by system level simulations. System simulations normally use link abstraction models to reduce the complexity and duration. But modeling the coding gains of HARQ at the system level is not a linear process and has been an active area of research especially in the industry for 3GPP LTE and IEEE 802.16m WiMAX [2]. The main purpose of link abstraction is to provide a low complexity tool for huge system level evaluations with very good accuracy [3]. Recently many link abstraction models have been proposed for the OFDM systems but most of them do not address the requirements of the LTE [4][5][6]. While others are too much complex to be actually implemented in real systems. Since the link abstraction has to be implemented not only for system evaluations at eNodeB but also for fast link adaptation in the hand sets (UEs) so the requirement of reduced complexity becomes even more strict.

However to model the IR-HARQ in LTE link performance prediction is not a trivial job because of the structure of the retransmissions. The retransmissions can not only include new parity bits which are source of the coding gain but also include some portions of the last transmission. So in this way the link prediction model should be able to track the number of bits which is not difficult but huge complexity is the bottleneck for this approach. A very comprehensive analysis of modeling the coding performance of HARQ schemes is presented in [7] where the authors have presented the concept of accumulated mutual information and also they compared the performance gain of incremental redundancy and chase combining. In [8] the authors have also presented method of HARQ modeling but they do not consider the case for LTE and their approach requires a huge number of reference AWGN curves. In [9] the authors have presented a link prediction model which is based on MIESM and showed results with acceptable accuracy but in their approach lots of information is to be calculated before hand and stored in the UE and eNodeB to reach the accuracy. Further HARQ modeling for OFDM systems is discussed in [10] where they have presented a recursive method using exponential effective SINR mapping which requires the calibration of adjustment factors and thus limits its applicability.
In this paper we address the issues of modeling the IR-HARQ in system level evaluations for the LTE systems and propose a novel unified link abstraction method which is capable of modeling the gains offered by not only all of the retransmissions of a certain codeword but also the initial transmission with the help of only a few reference curves. For this purpose we present the semi-analytical expressions which can accommodate for the coding gain or loss in turbo codes with respect to the reference curve and we show that our scheme can map all transmissions of the certain codeword on the reference curve very accurately. Another novelty in our scheme is that it can abstract link performance even if a different code rate is used for the retransmissions of a certain codeword.

In the rest of the paper we present the system model and the process of IR-HARQ for LTE in section II. Then in section III we present the necessary factors which are crucial part of IR-HARQ and must be considered for its performance modeling in system level evaluations. In section IV we present the link abstraction for IR-HARQ. In section V we present our proposed IR-HARQ modeling scheme and show that how it can be performed for the initial transmission and then for the retransmissions. In section VI we provide the results and finally we conclude the paper in section VII with concluding remarks and future work.

II. IR-HARQ IN LTE

A. System Model

We consider the baseline configuration of LTE, i.e. LTE transmission mode 1 which represents a SISO system. OFDMA is applied on the downlink in LTE and the codeword to be transmitted is mapped onto the physical resources known as resource elements after encoding, interleaving, rate matching and puncturing, scrambling and OFDM modulation. Interleaving in the LTE is contention free and pseudo-random and makes sure that the bits are spread uniformly across all of the available resource blocks.

In LTE transmission mode 1, a single antenna eNodeB communicates with a single antenna UE. The received signal at the UE for j-th resource element (LTE acronym for subcarrier or frequency tone) is given by

\[ r_j = h_j x_j + z_j, \quad j = 1, 2, \cdots, J \]

where \( J \) is the number of resource elements, \( h_j \in \mathbb{C}^{1 \times 1} \) symbolizes the flat Rayleigh fading SISO channel from the eNodeB to the UE. It can be modeled as independent identically distributed (i.i.d) zero mean circularly symmetric complex Gaussian random variable with the variance of 0.5 per dimension. \( z_j \) is ZMCSGC white noise of variance \( N_0 \) at the UE. Complex symbol \( x_j \) is assumed to be independent and of variances \( \sigma^2 \) belonging to discrete QAM constellations \( \chi \).

If the first transmission is not successful, then UE sends back a NACK signal to eNodeB on its physical uplink control channel (PUCCH) indicating that the signal is received in error. The eNodeB transmits the fresh copy of the signal which is not exactly the same and contains some repeated bits and some new parity bits. This is possible in LTE with the help of redundancy version indexing and virtual circular buffer (VCB). The design of rate matching and puncturing in LTE is very sophisticated. It can be used for implementing an arbitrary code size and HARQ benefits from the design of different redundancy versions which basically sends some new information bits and some repeated coded bits.

After receiving the certain number of retransmission, the effective signal is calculated at the receiver by applying the maximum ratio combining on the received log-likelihood ratios during different retransmissions. The gains received after the decoding can be actually divided into two categories, the coding gain which comes from the transmission of new parity bits in each retransmission and the receive SNR gain which comes from the repetition of coded bits in each retransmission. It is very well known that for turbo codes with contention free interleavers, most of the Hamming weight in the minimum distance resides in the parity streams. Therefore, the retransmissions are designed to have more new parity bits so that the effective minimum distance of the punctured code is improved and provides good performance at high code rates [11].

III. IR-HARQ MODELING AT SYSTEM LEVEL

To model the effects of the IR-HARQ at the system level, it is required to track down the new and repeated coded bits in each of the retransmission. This puts an extra burden on the UE and eNodeB as it may involve replicating the rate matching, interleaving, scrambling and then their opposites. This is an added complexity especially on the UE side for the calculation of the effective SINR so that it can signal back to the eNodeB its required CQI. So we need to rather have a simple model which can not only easily take into account all the gains of IR-HARQ but also can provide us with the required accuracy and reduced level of complexity.

A. Circular Buffer Rate Matching

Our approach makes use of the intelligent design of the circular buffer rate matching and the fact that all of the arbitrary code rates are obtained from the mother code rate. So it is important to understand the design of circular buffer rate matching. It is shown in figure 1 [12].

![Fig. 1. Turbo Encoder and Virtual Circular Buffer (VCB)](image-url)
Based on this circular buffer, the eNodeB can transmit data with any arbitrary code rate. For example for high code rates only some portion of the circular buffer is transmitted whereas for the very low code rate circular buffer can be transmitted several times. So, in general for any required transmission code rate the circular buffer can be read out from an arbitrary starting point and if the end of the buffer is reached then the it overlaps from the beginning of the buffer, hence named as circular buffer.

For HARQ, different redundancy versions are created by just pointing out the different starting point from the circular buffer. The adjacent redundancy versions contains equal number of the repeated bits and if the same MCS is used for retransmission then each of the RV is also of the same size.

IV. ABSTRACTION FOR HARQ

For the abstraction of HARQ we shall use the concept of effective SINR mapping which at first maps the varying SINRs of a codeword to an effective SINR ($\gamma_{eff}$) value which is then used to read the equivalent BLER from the AWGN performance curves of a particular modulation and code scheme (MCS).

$$\gamma_{eff} = \beta_1 I^{-1} \left[ \frac{1}{J} \sum_{j=1}^{J} I \left( \frac{\gamma_j}{\beta_2} \right) \right]$$

$$BLER(\gamma, \text{mcs}) \approx BLER_{\text{AWGN}}(\gamma_{eff}, \text{mcs})$$

Where $\gamma$ represents the $\mathbb{R}^{J \times 1}$ vector of the post processed SINR values for each of the subcarrier ($\gamma_j$) and $J$ is the number of total subcarriers in a codeword. $I(\gamma_j)$ is a mapping function which transforms SINR of each subcarrier to some “information measure” where it is linearly averaged over the codeword. Then these averaged values are transformed back to SNR domain. $\beta_1$ and $\beta_2$ are called calibration factors and they are there to compensate for different modulation orders and code rates.

Mutual information based abstraction models have been shown to be more accurate for mapping function of (2) than other standard models [13][6][3] and also MI-based models do not require the calibration of adjustment factors for at least the case of SISO systems so we shall use MI-based ESM as the base abstraction methodology in this paper.

A. Chase Combining (CC)

In CC same copy of the previous codeword is retransmitted. Let’s say there were total of $M$ transmissions and for each of the retransmission the same RV was sent then the $\gamma_{eff}$ can be written as [2],

$$\gamma_{eff} = \delta_1 I^{-1} \left[ \frac{1}{J} \sum_{j=1}^{J} I \left( \frac{\sum_{m=1}^{M} \gamma_{m,j}}{\delta_2} \right) \right]$$

B. Incremental Redundancy Without Coded Repetition

For the case when only new parity bits are retransmitted and there is no repetition of coded bits, the $\gamma_{eff}$ can be written as,

$$\gamma_{eff} = \delta_1 I^{-1} \left[ \frac{1}{J} \sum_{m=1}^{M} \sum_{j=1}^{J} I \left( \frac{\gamma_{m,j}}{\delta_2} \right) \right]$$

i.e., by the new retransmission only coding gain is achieved in terms of the better received mutual information.

C. Incremental Redundancy with Coded Repetition

In LTE incremental redundancy is implemented with coded bit repetitions which means that in each of the retransmissions there are always some new bits and some repeated bits when compared with the previous transmission. So the $\gamma_{eff}$ must be able to reflect the effects of the new bits and the repeated coded bits. Depending on the initial transmission code rate, the number of repeated bits can be different (from 0-3 time within the initial transmission to 0-15 times for all 4 transmissions). And to keep track of these bits for each and every retransmission for a huge variety of modulation and coding schemes of LTE is extremely complex and against the basic requirements of the link abstraction.

a) Brute-Force Method: One method is the brute-force approach. In this method $\gamma_{eff}$ is calculated using (2) for all retransmissions and the effects of coding gain is captured from the reference AWGN-BLER curve which corresponds to the specific code rate after certain retransmissions. This approach seems straight forward but is very complex as it needs to calculate a huge number of reference AWGN curves for the equivalent code rates which is function of multiple factor, i.e. block size, code rate, modulation etc. So, using this method poses a great amount of complexity over-head in the training phase of the link abstraction and also requires a huge number of reference AWGN curves to be stored which becomes impractical for the case of all of the different MCSs and allowed number of transmissions. For example it requires at least 112 reference AWGN curves only for the case when the retransmission is done using the same mcs and an adjacent RV is transmitted in each of the allowed retransmission.

b) Bit-Tracking: One better method to approach this problem is to count the number of the different sets of possible combinations of the bits which are received at least once at the receiver in all of the transmissions and also to track the frequency of the occurrence of all of the bits. Then using this information mutual information at bit level can be used for an accurate modeling of IR-HARQ [14]. This approach is definitely more accurate but also more complex.

V. PROPOSED LINK ABSTRACTION MODEL

We propose a rather simple approach for which actually it is not required to store more than 3 reference curves for the link quality measurement. We also give semi analytical
solution with the help of which the gains of the receiver with IR-HARQ can be reliably represented in the system level evaluations.

Our approach stems from the fact that in the practical systems the desired channel codes are derived from a fixed mother code $r_m$ which contains $L$ coded bits. If transmission with higher code rate than the mother code rate is required then the codeword of mother code rate $r_m$ is punctured according to some predetermined criterion. Whereas for the transmission with low code rate than the mother code rate, either some portion of the codeword of mother code rate or the whole codeword is repeated depending on how low the required code rate is. This actually has significant effect on the performance of the implemented HARQ in the system. For example if the initial transmission code rate is really low then there are no added parity bits to be sent into retransmissions so there is no difference in the performance of IR and CC. Where as for the higher initial transmission code rate there usually are some additional parity bits which can be sent in the retransmissions to provide higher coding gains.

A. Initial Transmission

We denote the initial transmission rate with $r_i$. If during the first transmission $r_i$ is lower than $r_m$ then the transmission of this codeword will definitely contain more than one rounds of the circular buffer and if $r_i$ is higher than $r_m$ then only some portion of the circular buffer will be transmitted. Thus we can easily calculate these turns or the size of the circular buffer which is transmitted in the initial transmission as, $R = \frac{r_m}{r_i}$.

Then we propose a simple calculation of the $\gamma_i^{\dagger}$ using $\gamma_{eff}$ of (2) with the idea of mapping $r_i$ to the mother code rate $r_m$ with the help of $R$, i.e., $\gamma_i^{\dagger} = R \times \gamma_{eff}$. $R$ takes into account the effect of either the repeated coded bits or the punctured coded bits when compared to the $r_m$. In other words $R$ shifts of the performance curve for a specific codeword to either left or right of the performance curve of the mother code rate as is shown in figure 2 so that,

$$BLER(\gamma_i^{\dagger}, mcs) \simeq BLER_{AWGN}(\gamma_{eff}^{\dagger}, mcs(r_m))$$  \hspace{1cm} (6)

Which means that now we do not need to store a huge number of the reference BLER curves corresponding to all of the MCSs but only for the $mcs$ which corresponds to the $r_m$.

B. Retransmissions

Now we show that how we can use the same reference curves even for the case of all of the retransmissions. To this purpose we need to find the effective code rate after certain number of retransmissions, then again we can use the idea of mapping this effective code rate to the $r_m$ using all of the repeat transmissions. The effective code rate $r_{eff}$ after certain number of allowed retransmissions can be calculated as,

$$r_{eff} = \left\{ \begin{array}{ll}
\frac{r_i}{r_i + \sum_{k=1}^{2} r_{rd}^{k} + \sum_{k=1}^{2} r_{rd}^{k}} & 1^{st} \text{retransmission} \\
\frac{r_i}{r_i + \sum_{k=1}^{2} r_{rd}^{k} + \sum_{k=1}^{2} r_{rd}^{k} + r_{rd}^{3} + r_{rd}^{4} + r_{rd}^{5}} & 2^{nd} \text{retransmission} \\
\frac{r_i}{r_i + \sum_{k=1}^{2} r_{rd}^{k} + \sum_{k=1}^{2} r_{rd}^{k} + r_{rd}^{3} + r_{rd}^{4} + r_{rd}^{5} + r_{rd}^{6}} & 3^{rd} \text{retransmission} 
\end{array} \right.$$  \hspace{1cm} (7)

Important thing to note here is that the retransmissions can have either the same code rate or different than the $r_{i-1}$. Index $i$ represents the current transmission and $i \leq 4$ as in LTE maximum of 4 transmissions of a certain codeword are allowed. Now for the case of CC this is good enough as, $r_{eff} = \frac{T}{K'}$ where $T$ is the transport block size which is to be transmitted in a certain codeword and $K'$ is the total number of transmitted coded bits without repetition. But for the IR-HARQ of LTE the $r_{eff}$ is not the real effective code rate as there are some bits in the retransmissions which are sent new and others are repeated. As mentioned earlier that most of the Hamming weight of the turbo code lies in its parity bits, the redundancy versions in LTE are designed in such a way to have less repetition of the coded bits. So we need to find the $r_{eff}'$ which represents the actual code rate considering the number of coded repeated bits in retransmissions. It is clear that $r_{eff} > r_{eff}'$ because of the overlapping of adjacent redundancy versions as $r_{eff}' = \frac{T}{K'}$ where $K'$ is the total number of coded transmitted bits with repetition. Then $r_{eff}'$ can be calculated as,

$$r_{eff}' = r_{eff} + \xi r_{eff}$$  \hspace{1cm} (7)

where $\xi$ represents the portion of the redundancy version which is overlapping with the previous transmission and this factor can only be calculated from the rate matching algorithm of LTE. This factor is only important as long as the VCB at the decoder is not exhausted. In other words it tells us that what percentage of the VCB is still not received and using this we can calculate the true value of $R$. Where as for the case when after retransmissions the VCB is exhausted at the decoder then $r_{eff}$ alone is able to represent the gain due to the coded bit repetitions as the portion of the repeated coded bits becomes insignificant to the overall received bits. So it is important to mention that as soon as the full circular buffer is exhausted, the second part of (7) becomes insignificant. Please note that the number of repeated bits in all of the RVs of specific MCS are uniformly distributed in LTE so we only need to calculate this number for just any one of the retransmission. Then one can calculate $R = \frac{r_m}{r_{eff}'}$ and $\gamma_{ret}$ as,
\[ \gamma_{\text{ret}} = I^{-1} \left[ \frac{1}{J \times M} \sum_{j=1}^{J} \sum_{m=1}^{M} I \left( \gamma_{m,j} \right) \right] \] (8)

and

\[ \gamma^\dagger_{\text{ret}} = R \times \gamma_{\text{ret}} \] (9)

where \( \gamma_{\text{ret}} \) takes into account the expected coding gain and \( R \) provides the gain due to the repetitions of the coded bits. Then we can assume,

\[ BLER(\gamma',\text{mcs}) \simeq BLER_{\text{AWGN}}(\gamma^\dagger_{\text{ret}},\text{mcs}(r_m)) \] (10)

Where \( \gamma' \) represents the \( \mathbb{R}^{(J \times M) \times 1} \) vector of the post processed SINR values for each of the subcarrier \( (\gamma_{m,j}) \) for each of the transmission \( m \) of certain codeword.

VI. RESULTS

For the validation of our proposed scheme, we used Eu- recom’s OpenAirInterface\(^1\) link level simulator which implements 3GPP LTE Release 8.6 physical layer [15], [12], [16] with 5 MHz bandwidth and 25 physical resource blocks (PRB). We show the results for a highly frequency selective Rayleigh channel model with 8-taps. To test our proposed method we performed simulation for more than 50 different channel realizations and during each of the channel realization we simulated the system for 1000 packets or 500 erroneous packets. For each of these channel realizations we applied our proposed method on all of the four transmissions of each packet. The reference curves corresponding to the mother code rate or reference code rate are shown in Figure 3. These are the curves which are then shifted according to the \( R \) to compensate for the coding gain and SNR gain. The main results are shown in Figure 4 to 8 where the results for all four transmissions after the abstraction are shown. We calculated the \( \xi \) for these mcs from the rate matching algorithm of LTE and calculated the \( \gamma^\dagger_{\text{eff}} \) for the calculation of \( R \). We only show a few results here just to make this clear that a simple method can provide better results as well.

\(^1\)http://www.openairinterface.org/

VII. CONCLUSION

We have presented a semi-analytical unified link abstraction model for LTE which is not only simple but also accurate for modeling the performance of IR-HARQ from the link level simulations. In terms of low complexity it should be noted that our approach makes use of the information already available in the physical downlink control channel (PDCCH) of each transmission for a specific codeword and only a very little information is required to be stored a priori. It is clear from the results that our method is capable of modeling
the effects of IR-HARQ from the link level simulations with the help of very few reference curves and can be used for performance modeling at system level. Moreover due to the analytical formulation of effective code rate it is also possible to model the gains if a different code rate is used for the retransmissions. For future work this model is to be extended for the MIMO systems and interference aware receivers.

REFERENCES


