3GPP TSG RAN WG1 #42 R1-050718
London, U.K., Aug 29 – Sept 2, 2005

Source: **Motorola**
Title: **EESM Link Error Prediction for EUTRA System Evaluation**
Agenda Item: **10.7.2**
Document for: **Discussion and Decision**

1 Introduction

System simulations require a performance estimate for each link, but it is typically infeasible to conduct full link-level emulation within a reasonable amount of time. Therefore, it is critical to have a simple link error probability predictor that accurately models coded OFDM performance.

The exponential ESM (EESM) method [1] was shown to be a very good link error prediction method for multi-carrier systems in the case of single transmission [2]. Here, the EESM method is extended to cover Hybrid HARQ (HARQ) techniques: Chase combining, Incremental Redundancy (IR) or an arbitrary combination of both. Further simplifications are introduced to greatly reduce the amount of memory necessary to utilize EESM for HARQ transmissions.

2 EESM Notation and Model

The effective SNR, γ_{eff}, is defined as

$$\gamma_{eff} = -\beta.\ln\left(\frac{1}{N_{tot}}\sum_{i=1}^{N_{tot}} e^{-\gamma_i/\beta}\right)$$

The index i is used to denote the transmitted (or received) codeword bit *indices*[1], where i ranges from 1 to N_{tot}. N_{tot} is the total number of transmitted codeword bit indices. The γ_i correspond to received SNR values (in linear units) for transmitted codeword bit index i. Since codeword indices are unique, the γ_i received SNR values are 'combined' (in linear units) from all the transmissions of codeword bit i. For example, for an encoder packet size of EPS bits and a mother code rate of R_{MC}, N_{tot} can be at most EPS/R_{MC} bits.

Consider a total of N_{TX} transmissions, with k the transmission index ($1 \le k \le N_{TX}$), and define N_k as the number of bits transmitted on the kth transmission. For transmission k, a received SNR (in linear units) for bit index i is denoted as $\gamma_{i,k}$. Define V_k as the set of indices where a coded bit was transmitted on transmission k. Assuming that codeword bits are not repeated within a single transmission, $card(V_k) = |V_k| = N_k$. Define the indicator function for codeword bit index i for the set V_k as

[1] the processing is assumed to be bit level in all of the following, but is also applicable to symbol processing

$$\begin{cases} I_{i,k} = 0 \, for \;\; i \notin V_k \\ I_{i,k} = 1 \, for \;\; i \in V_k \end{cases}$$

Then, again assuming that codeword bits are not repeated within a transmission,

$$\gamma_i = \sum_{k=1}^{N_{TX}} I_{i,k} \cdot \gamma_{i,k} \; .$$

Define also $U_k = \bigcup_{j=1}^{k} V_j$, where U_k is the set of unique bit indices transmitted up to (and including) transmission k.

The total set of transmitted unique bit indices is $U = \bigcup_{k=1}^{N_{TX}} U_k = \bigcup_{k=1}^{N_{TX}} V_k$. The total number of transmitted codeword bit indices is $N_{tot} = card(U) = |U|$.

Then,

$$\gamma_{eff} = -\beta . \ln\left(\frac{1}{N_{tot}} \sum_{i=1}^{N_{tot}} e^{-\gamma_i / \beta} \right) = -\beta . \ln\left(\frac{1}{N_{tot}} \sum_{i \in U} \exp\left(-\frac{1}{\beta} \sum_{k=1}^{N_{TX}} I_{i,k} \cdot \gamma_{i,k} \right) \right)$$

Note that in the equations above a single β value is used, with the implication that the modulation is the same for all transmissions. A later section will discuss how to modify the equation to handle a change of modulation. The β used in the above equations is computed (and dependent) on the effective code rate EPS/N_{tot} after N_{TX} transmissions. For Chase transmission the effective code rate is EPS/N_1 , so β is unchanged regardless of the number of retransmissions. For IR, β may vary with the code rate; however, in many cases β is only loosely dependent on the code rate and a single value may be sufficient.

3 EESM with HARQ Transmissions

In this section, it is assumed that the modulation used for HARQ transmissions is always the same. A later section will discuss how changes of modulation can be handled using a demapping penalty.

3.1 IR Only

In this section, it is assumed that each codeword bit is uniquely transmitted, with no wrap-around effects (i.e., $R_{MC} = 0$). In this case the indicator function degenerates (since no combining occurs), and $N_{tot} = \sum_{k=1}^{N_{TX}} N_k$.

The EESM effective SNR is then

$$\gamma_{eff} = -\beta.\ln\left(\frac{1}{N_{tot}}\sum_{k=1}^{N_{TX}}\sum_{i\in U_k \setminus U_{k-1}} e^{-\gamma_{i,k}/\beta}\right) = -\beta.\ln\left(\frac{1}{N_{tot}}\sum_{k=1}^{N_{TX}}\sum_{i\in V_k} e^{-\gamma_{i,k}/\beta}\right),$$

since the sum over the set of all unique indices U can be broken down by summing the number of unique indices for transmission k, $U_k \setminus U_{k-1}$, which for IR equals V_k.

3.2 Chase (or Partial Chase) Only

In this section, it is assumed that on the first transmission, all the coded bits are transmitted, and that on the subsequent transmissions, some or all of these initial bits are retransmitted so that $U = U_1 = V_1$ and $N_{tot} = N_1$. The EESM effective SNR is then

$$\gamma_{eff} = -\beta.\ln\left(\frac{1}{N_1}\sum_{i\in V_1}\exp\left(-\frac{1}{\beta}\sum_{k=1}^{N_{TX}} I_{i,k}.\gamma_{i,k}\right)\right)$$

3.3 Performance of the EESM with HARQ

Simulation results are shown to illustrate that EESM works well for HARQ. Results are presented for a few representative cases and are not meant to be exhaustive. However, EESM has been tested for many other scenarios (in terms of modulation, information frame size, channel type, etc.) and the performance is consistently very good. For this particular case, the system bandwidth is 20 MHz, turbo codes are used with a max-log-MAP decoding kernel and the considered channel is GSM TU.

Figure 1 shows the performance of EESM in an IR scenario, where the FER after two transmissions is modeled for a link with k=384, QPSK, rate 1/2 for each transmission, and rate 1/4 after combining the two transmissions. The red ('static') curve is the link error performance given by EESM, and each of the green curves represents one particular instance of the channel. From Figure 1, it is clear that EESM provides an accurate FER prediction.

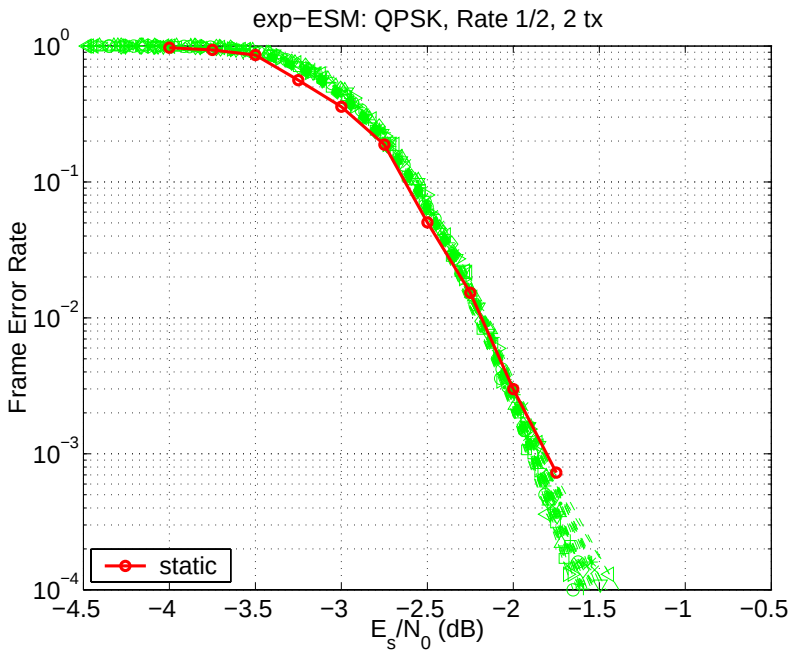

Figure 1. Performance of the EESM method with IR.

Similarly, EESM works very well with Chase combining. Figure 2 shows the performance of EESM for Chase combining for a similar scenario as in Figure 1 but for QPSK R=3/4. Clearly, EESM gives a good FER prediction for Chase combining.

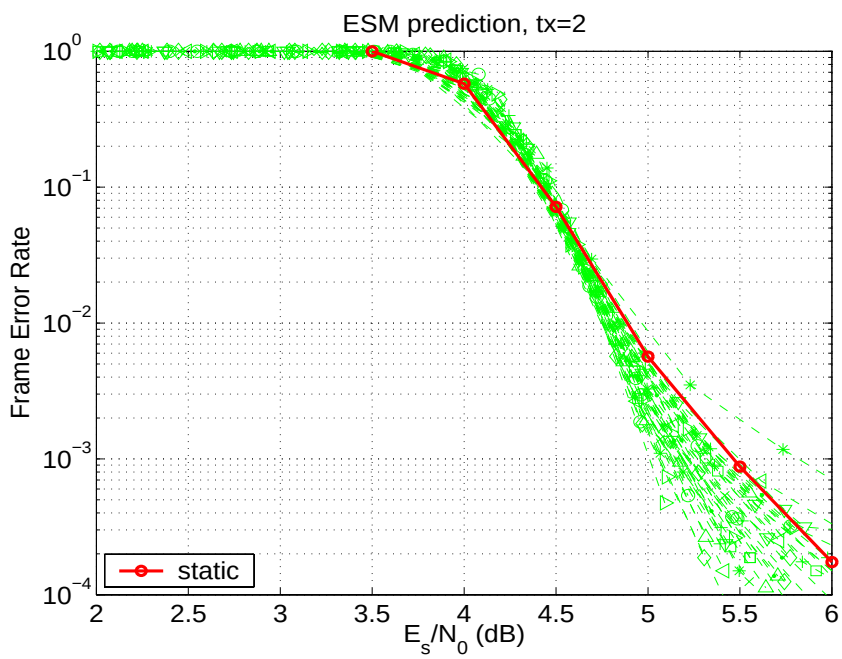

Figure 2. Performance of the EESM method with Chase combining.

As it has been shown on these two examples, EESM works very well with HARQ. As already stated earlier, the two cases presented here were just representative examples.

EESM also works well for many other cases such as more than two HARQ transmissions, other channel types, other block sizes, and other modulations.
Consequently, EESM is a good LEP that can be used to model link performance for EUTRA system simulations.

4 Recursive EESM Method

EESM provides good accuracy for LEP within a system simulation and eliminates the need for full link-level simulations within a system simulation. However, since HARQ adds significant additional storage and computational requirements, further simplifications of the EESM method are needed to make the computational and storage complexity of the system simulations more manageable. For example, consider that a coded IP packet is in the order of 12,000 uncoded bits[2]. If the mother code rate is 1/5, the buffer size for the HARQ process needs to be larger than 60,000 bits *per packet*. This requires a significant amount of memory especially for a multi-cell, multi-user system simulation. Hence, it is desirable to relax the memory requirements for the EESM method. One attractive approach to reduce complexity is to break the EESM computation into a recursive computation, where the effective SNR after the current retransmission is a simple function of the current received SNR values and the previously computed effective SNR for the transmissions up to the current transmission. However, breaking the computation into individual transmissions for the arbitrary combination of transmissions results is not straightforward, since the expression for IR processes per transmission unique bits on the transmission, and the expression for Chase looks over all transmissions. In what follows similar recursive expressions will be developed for IR (exact) and Chase (close approximation).

4.1 IR

With the recursive set defined as follows, it possible to establish that $\eta_{N_{TX}} = \gamma_{eff}$:

- $\eta_1 = -\beta.\ln\left(\frac{1}{N_1}\sum_{i\in V_1} e^{-\gamma_{i,1}/\beta}\right)$, and
- $\eta_{j+1} = -\beta.\ln\left(\frac{1}{\sum_{k=1}^{j+1} N_k}\left(\left(\sum_{k=1}^{j} N_k\right)e^{-\eta_j/\beta} + \sum_{i\in V_{j+1}} e^{-\gamma_{i,j+1}/\beta}\right)\right)$ for $1 < j < N_{TX}$

For example, looking at the second transmission

$$\eta_2 = -\beta.\ln\left(\frac{1}{N_1+N_2}\left(\sum_{i\in V_1} e^{-\gamma_{i,1}/\beta} + \sum_{i\in V_2} e^{-\gamma_{i,2}/\beta}\right)\right) = -\beta.\ln\left(\frac{1}{N_1+N_2}\left(N_1 e^{-\eta_1/\beta} + \sum_{i\in V_2} e^{-\gamma_{i,2}/\beta}\right)\right),$$

[2] The situation is even worse if multiple packets can be concatenated for transmission. For example, with 64QAM R=5/6 and 5MHz bandwidth then 34,500 uncoded bits are possible given a 2ms multiframe with 28 OFDM symbols (i.e. four 0.5ms subframes) where 23 of the OFDM symbols can support data symbols where each OFDM symbol has 300 subcarriers each.

since from the definition of η_1, $e^{-\eta_1/\beta} = \frac{1}{N_1}\sum_{i\in V_1} e^{-\gamma_{i,1}/\beta}$.

The practical consequence is that to estimate the effective SNR after q transmissions, we only need to keep track of the effective SNR after q-1 transmissions and of the number of transmitted bits, thereby requiring a minimal amount of memory.

4.2 Chase

Looking at the IR example, it can be seen that the effective SNR for the *j+1* transmission can be computed using assuming flat fading on the *j* first transmissions (with an SNR set at the effective SNR value on these *j* transmissions), and using the actual selective received SNR for the *j+1* transmission. It is *conjectured* that the same process can be applied for Chase with little degradation:

- Compute effective SNR $SNR_{eff}^{(j)}$ for the first *j* transmissions
- Assume flat fading with an SNR of $SNR_{eff}^{(j)}$ for the signals received on first *j* transmissions
- Compute effective SNR for transmission *j+1* based on $SNR_{eff}^{(j)}$ and the selective signal received on transmission *j+1*

The $SNR_{eff}^{(k)}$ values are computed using the following recursive process:

- $SNR_{eff}^{(1)} = -\beta.\ln\left(\frac{1}{N_1}\sum_{i\in V_1} e^{-\gamma_{i,1}/\beta}\right)$
- $SNR_{eff}^{(j+1)} = -\beta.\ln\left(\frac{1}{N_1}\sum_{i\in V_{j+1}} e^{-\frac{1}{\beta}\left(SNR_{eff}^{(j)} + I_{i,j+1}.\gamma_{i,j+1}\right)}\right)$ for $1 < j < N_{TX}$

Conjecture: $SNR_{eff}^{(N_{TX})} \approx \gamma_{eff}$

Link and system simulations will be shown later in this section that illustrate the validity of this conjecture.

4.3 Arbitrary Combination of Chase/IR

Combining the results using pure Chase and pure IR, the recursive set is defined as follows:

- $\mu_1 = -\beta.\ln\left(\frac{1}{N_1}\sum_{i\in V_1} e^{-\gamma_{i,1}/\beta}\right)$
- $\mu_{j+1} = -\beta.\ln\left(\frac{1}{|U_{j+1}|}\left(|U_j|.\sum_{i\in U_j} e^{-\frac{1}{\beta}\left(\mu_j + I_{i,j+1}.\gamma_{i,j+1}\right)} + \sum_{k\in V_{j+1}\setminus U_j} e^{-\gamma_{k,j+1}/\beta}\right)\right)$ for $1 < j < N_{TX}$

Conjecture: $\mu_{N_{TX}} \approx \gamma_{eff}$

4.4 Performance

The equations obtained for recursive algorithm for IR are exact. Since no approximation was made, the recursive approach does not bring any degradation. However, for Chase, the recursive approach needs to be validated. Figure 3 shows the performance of the EESM method for Chase with the recursive approach with the same parameters as of Figure 2. There is clearly a degradation in performance as when compared with the exact EESM method. However, the recursive EESM still gives a very good FER prediction.

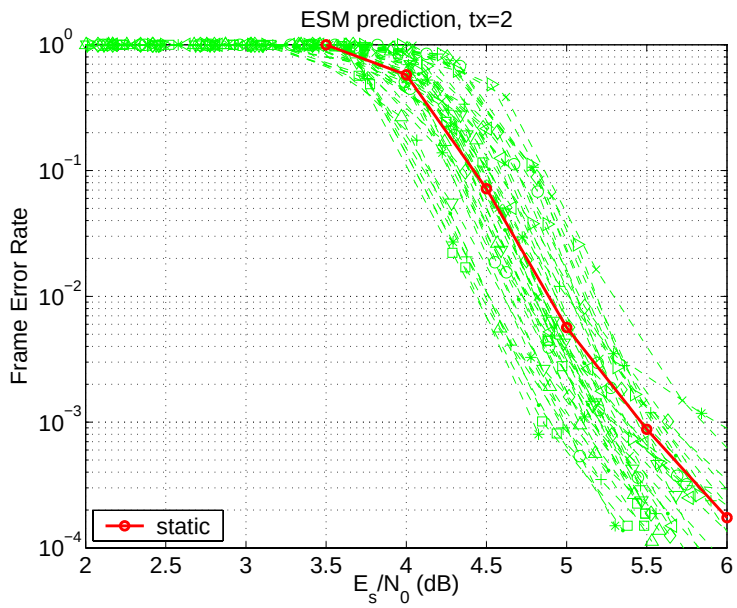

Figure 3. Performance of the recursive EESM method for Chase combining.

4.5 Summary – Recursive EESM Approach

Based on the result from the previous sections, the EESM method can be used in a recursive way. Practically speaking, this means that instead of storing *all* the *SNR* values for *all* the transmitted bits, the only variables that need to be stored in memory between transmissions are:

- The single effective SNR value obtained taking into account all the previous transmissions
- The number of bits transmitted so far

The EESM method can then be performed with minimal memory requirements.

5 Improvements

Several additional improvements can be made to the EESM and recursive EESM approaches described above. As described in the sections that follow, these include:

- **Binning to reduce memory**
- **Handling modulation changes on retransmissions**

5.1 Chase/IR with binning

In the previous recursive approximation of SNR for Chase combining, the effective SNR for the first *j* transmissions is assumed to be flat across bits in computing the SNR for the *j+1* transmission. This assumption results in a significant savings in both memory and computation. However, the accuracy of this approximation may be reduced as the frequency-selectivity of the channel increases. Therefore, binning can be used to extend the approximation to be more robust, with a little more memory consumed. The motivation for binning is that the channel quality is correlated over a group of OFDM subcarriers and/or OFDM symbol periods, so that the quality can be approximated by a single value over a bin of the appropriate dimensions. Thus, binning provides an effective tool for managing the tradeoff between channel tracking accuracy and memory/computational requirements.

The mother codeword is split into L bins, each of which consists of $\frac{N_{tot}}{L}$ bits (here it is assumed that N_{tot} can be divided by L). Each bin is denoted as a set $B_l, (1 \le l \le L)$. Instead of assuming the effective SNR of the mother codeword after the first *j* transmissions to be flat, the effective SNR of each bin is assumed to be flat. This approach requires the SNR of each bin to be stored in each transmission. With a properly chosen bin size, the memory cost does not increase significantly, but the accuracy for Chase combining can be improved for highly dispersive channels. Note that binning can be used over frequency (e.g., a bin being defined as a group of adjacent subcarriers over which the channel is somewhat correlated), and/or over time (e.g., a group of adjacent OFDM symbol periods on a particular subcarrier).

Suppose that the effective SNR for bin l after the j transmissions is $\mu_j^{(l)}$. The SNR recursion is then given by

$$\mu_1^{(l)} = -\beta \ln\left(\frac{1}{|B_l|}\sum_{i \in B_l \cap V_1} e^{-\frac{\gamma_{i,1}}{\beta}}\right)$$

$$\mu_{j+1}^{(l)} = -\beta \ln\left(\frac{1}{|B_l|}\left(\sum_{i \in U_j \cap B_l} e^{-\frac{1}{\beta}\left(\mu_j^{(l)} + I_{i,j+1}\cdot\gamma_{i,j+1}\right)} + \sum_{k \in \left(V_{j+1} \setminus U_j\right) \cap B_l} e^{-\gamma_{k,j+1}/\beta}\right)\right) \quad for\ 1 \le j < N_{tx}$$

$$\mu_j = -\beta \ln\left(\frac{1}{|U_j|}\sum_{1 \le l \le L, B_l \cap U_j \ne \phi} |B_l| e^{-\frac{1}{\beta}\mu_j^{(l)}}\right) \quad for\ 1 \le j \le N_{tx}$$

As shown in Figure 4, as the number of bins increases, the probability that a frame erasure decision is different with the approximation for Chase than in the ideal case is 14% without binning (i.e., a single bin) for 30km/h, and heads towards zero as the number of bins increases. For 3km/h the difference in frame erasure decisions is only 6% with a single bin. However, as was shown in Figure 3, while an individual decision to cause a frame to be erased or not may be off by 14%, the recursive EESM approximation is still quite accurate with a small error that is *symmetrically* distributed about the AWGN

curve. Therefore, as shown in Figure 4, the difference in user throughput for 3km/h observed at the output of the system simulator is within 1% with or without binning. Differences on sector throughput for 3km/h are ~5% without binning, which can be reduced to ~2% with binning. For higher speeds like 30km/h were there is more subcarrier SNR variation then user and sector throughput differences are brought within 5% of ideal by using at least 4 bins. For system simulations it is possible to choose the number of bins to use per user based on its assigned speed. For 3km/h a single SNR bin could be used while for 30km/h at least 4 SNR bins are needed.

Figure 5, **Figure 6**, and **Figure 7** show additional details on the frame erasure decisions. **Figure 5** shows that the difference in effective SINR decreases as the number of bins increases. Note large SNR errors do not necessarily mean a difference in the frame erasure decision. **Figure 6** shows for Chase and Nbins=256 the PDF of frame erasure decision conditioned on Effective SINR error. Integration shows that only 4.6% of the time there is a disagreement in the frame erasure decision between the approximation and ideal. **Figure 7** provides the information as in **Figure 6** but for IR. As expected from the derivation, the recursive IR expression is exact until the mother code rate is reached and wrap around occurs.

To summarize, the recursive EESM without binning is exact with IR and provides ~1% difference in system simulator user throughput results with Chase. If desired, binning may be used to improve the already accurate sector throughput results at 3km/h from within 5% to 2% of ideal. At higher speeds (e.g. 30km/h) then more than 1 bin is needed to drop the differences below 5% (e.g by using 4 SNR bins) or below 2% (with 16 bins) of ideal.

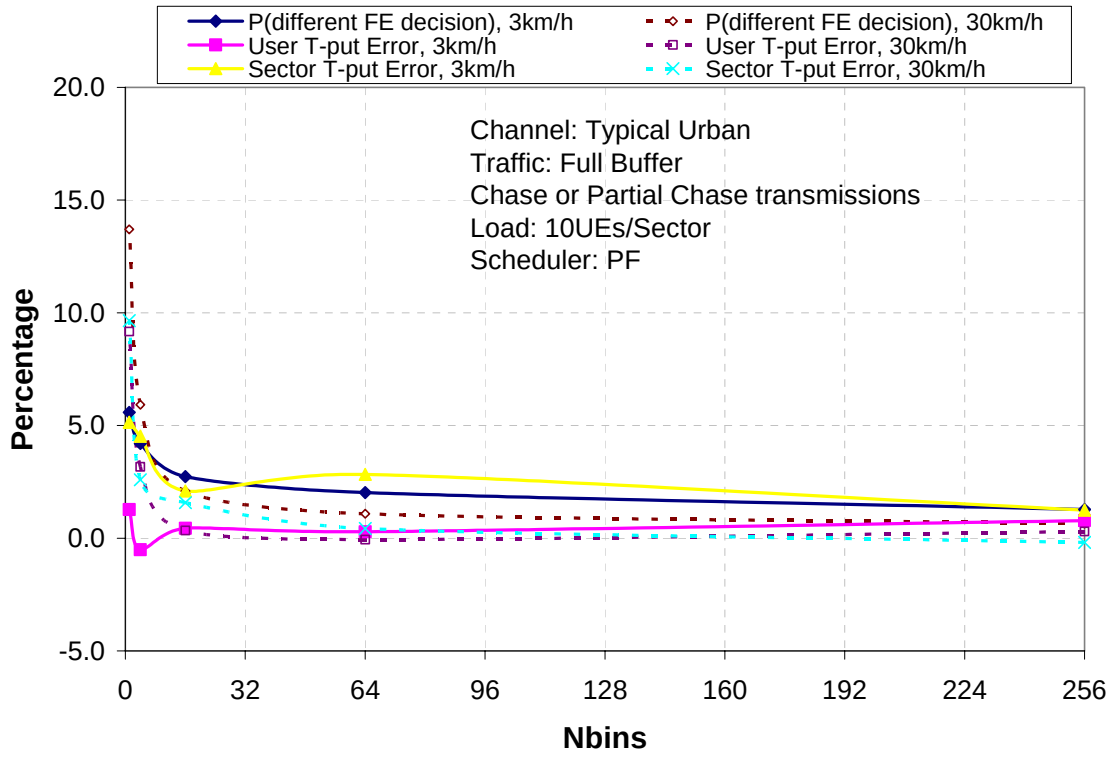

Figure 4 – Chase system simulation results comparing Ideal vs. multiple binned EESM approach

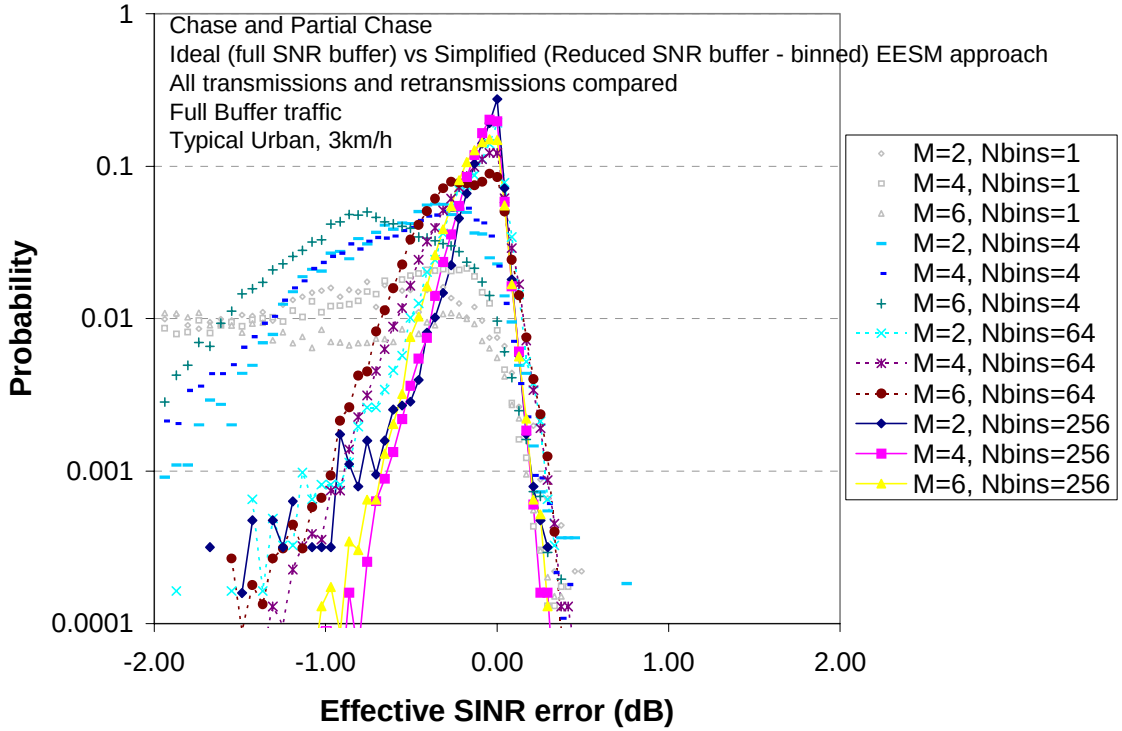

Figure 5 – PDF of effective SINR error using Ideal vs. multiple binned EESM approach for Chase

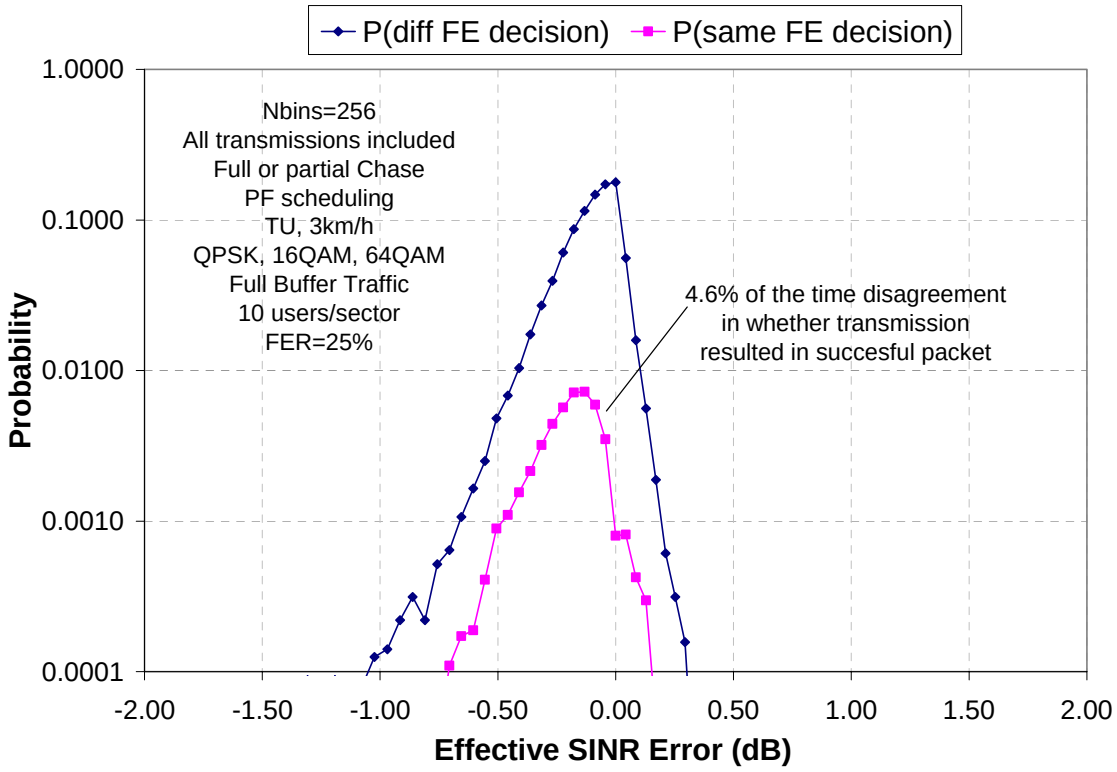

Figure 6 – PDF of FE decision conditioned on Effective SINR using Chase and for Nbins=256

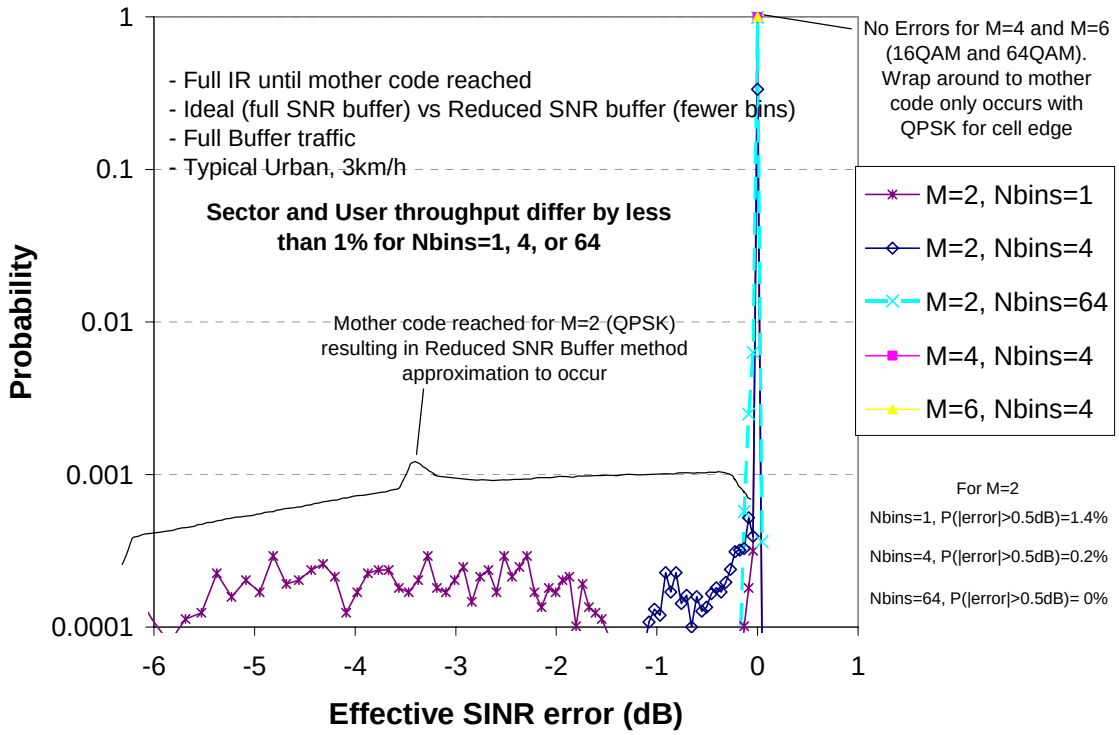

Figure 7 – PDF of Effective SINR Error for Ideal vs. multiple binned EESM Approach using IR

5.2 Chase/IR with Modulation change on retransmission

For HSDPA it is possible to change modulation on retransmissions, and this flexibility may be useful for EUTRA as well. To evaluate system performance with modulation changes on packet retransmissions using the EESM link error prediction method an extension to that technique is necessary. One possible extension is to use a demapping penalty such that QPSK AWGN curves are always referenced even if a higher order modulation is used for a given packet transmission. Since it is possible to avoid using a demapping penalty as long as the modulation does not change for a given packet's transmissions then the approach proposed is to use the appropriate MQAM AWGN curves along with the MQAM effective SNR until the modulation changes and then switch to using the demapped effective SNR with the QPSK AWGN curves as described in Figure 1 below. For the demapped effective SNR only a single beta value is used corresponding to QPSK since only the QPSK AWGN link curves are used, and the EESM beta factor is largely insensitive to coding rate especially for QPSK.

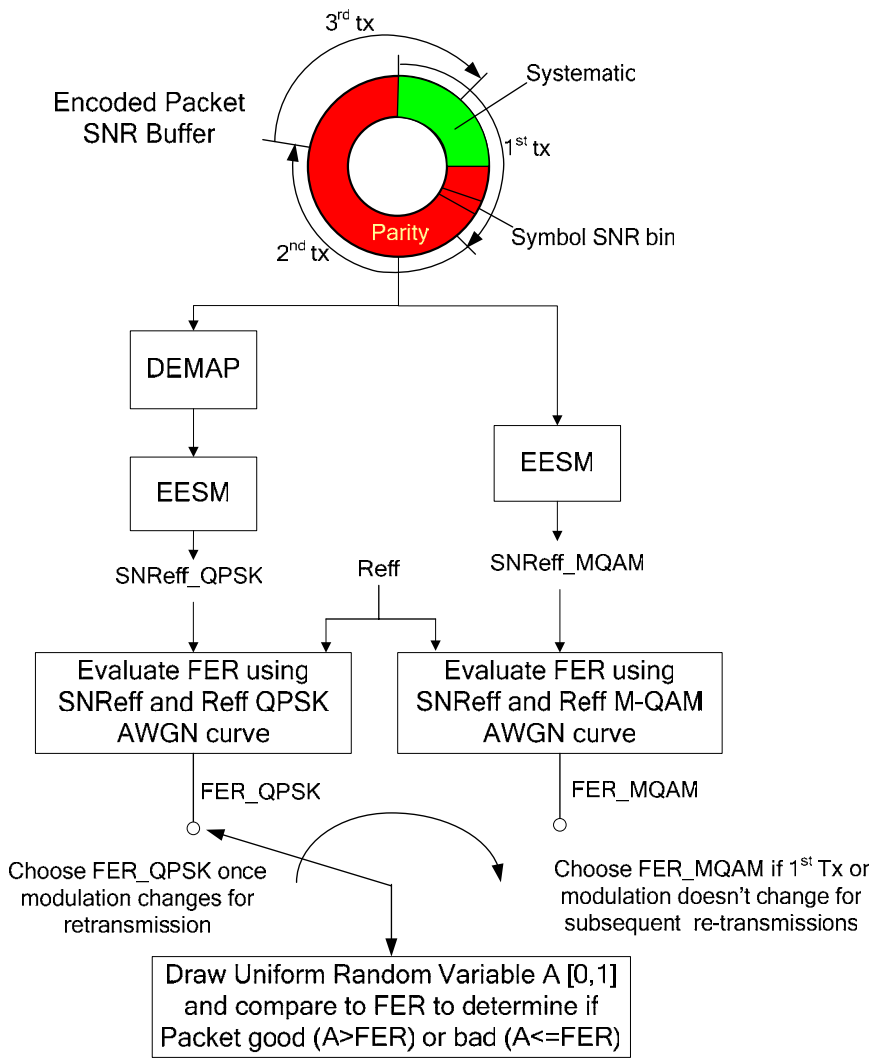

Figure 8 – EESM link mapping procedure for fixed and varying modulation on re-transmissions

6 Conclusion

Based on the results from the previous sections, the EESM method can be used to model HARQ with both IR and Chase combining and can be implemented in a recursive way to minimize memory requirements. A recursive implementation means that instead of storing *all* the *SNR* values for *all* the transmitted bits (full SNR buffer approach), the only variables that need to be stored in memory are:

- The effective SNR binned values obtained taking into account all the previous transmissions
- The number of bits transmitted so far

The recursive EESM approach is exact for IR and is shown to provide system simulation results within 1% user (5% sector) throughput of ideal for Chase at 3km/h. It is expected that this accuracy is sufficient for comparison of different techniques for EUTRA at low speeds like 3km/h. However, if additional accuracy for sector throughput values is desired or higher speeds are used, binned recursive EESM with a small amount of additional memory (e.g., 4 or 16 bins) may be used. In either case, a demapping EESM

path is provided and suggested to be used only on retransmissions where the modulation change occurs. This simplifies (and improves accuracy) of the primary path of 'normal' EESM.

The memory savings for the recursive EESM approach can be substantial. For a single "packet", the savings may be on the order of 172,500 soft values (e.g., 1 byte per code bit).[3]. In addition, the system simulation must allocate HARQ memory for multiple packets. While only a few users might be scheduled at the same time for best effort services such that the aggregate system simulation memory requirement is not too extravagant this is not true for streaming or VoIP services where many more users are active at the same time. Trying to optimize SNR buffer sizes to minimize memory with a mixture of VoIP and best effort users in a given cell would likely be problematic. Attempting to minimize memory requirements by dynamically allocating HARQ channel SNR buffers on a per packet basis is also difficult due to the need for a heap manager to keep heap memory de-fragmented. Such a manager is not necessary if all of the SNR buffers are chosen to be the same size but then the worst case SNR buffer size based on the maximum packet size possible must be chosen.

7 References

[1] Ericsson, "System-level evaluation of OFDM – initial evaluation," 3GPP TSG-RAN WG1 #35, R1-031304, Nov. 17-21, 2003.

[2] Nortel Networks, "Effective SIR computation for OFDM system-level Simulations," 3GPP TSG-RAN WG1 #35, R1-031370 Nov. 17-21, 2003.

[3] With Rmc=1/5 and a maximum buffer size of 34500/Rmc. A packet size of 34500 information bits is supported with a 2ms multi-frame composed of four 0.5ms subframes with 23 OFDM symbols of 300 subcarriers available for data symbols (and 5 OFDM symbols for pilot and control) such that the first transmission is based on using 64QAM with encoding rate 5/6.

8 Annex A

8.1 Change in β

In the description of EESM for HARQ, it was assumed that β did not change. This assumption is reasonable most of the time. For example, the value for QPSK (and even 16-QAM) varies little as a function of code rate (e.g., R=1/5 to R=3/4 is ~1.5 for QPSK, and ~3 to 6 for 16-QAM). However, with IR, the effective coding rate changes when the number of retransmission increases. Depending on the scenario, the β values may change. For increased accuracy in cases where the β value does change significantly between retransmissions, a simple procedure may be used to modify the effective SNR generated by the previous recursion iteration using an old β value. The current recursion is then computed using the modified effective SNR value from the previous iteration and the new β value.

The procedure for handling changes in β is shown below.
It has been observed that the curve $10*\log_{10}(\gamma_{eff})$ vs. β_{dB} is fairly smooth and can be approximated with an excellent accuracy by a quadratic approximation:
$10*\log_{10}(\gamma_{eff}) = a + b.\beta_{dB} + c.\beta_{dB}^2$,
Where a, b, and c are coefficients that need to be determined for the current channel realization. Figure 9shows such an example of approximation. As it can be observed, the quadratic approximation is excellent

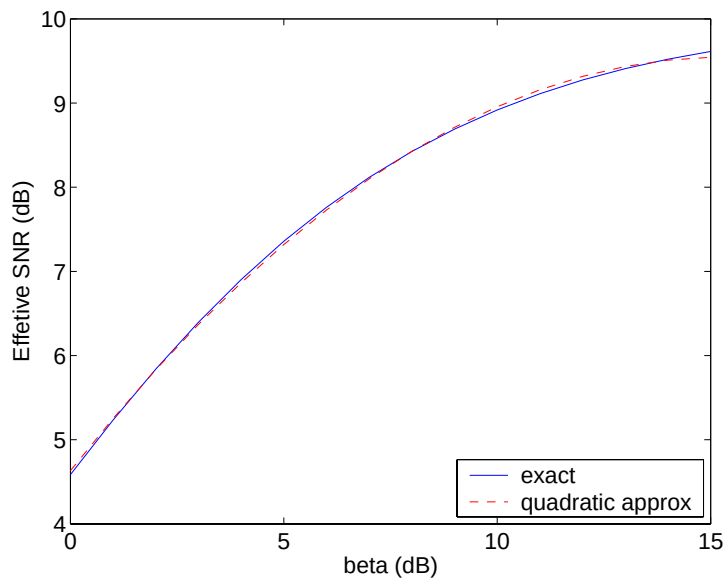

Figure 9. Quadratic approximation of $10*\log_{10}(\gamma_{eff})$ vs. β_{dB} in a TU50 channel realization.

By storing this curve for every bin or, at the price of a small loss of accuracy, by storing an "averaged" curve, it is possible to get the effective SNR for each β value (or for each corresponding MCS).