Improving QoS by Predictive Channel Quality Feedback for LTE

Xiang Xu, Mingjian Ni and Rudolf Mathar
Institute for Theoretical Information Technology, RWTH Aachen University
Aachen, Germany 52074
Email: {xu,ni,mathar}@ti.rwth-aachen.de

Abstract—In Long Term Evolution (LTE) systems, the quality of service (QoS) heavily relies on the accurate feedback of channel quality indicator (CQI). However, the time-varying channel and transmission delay make the outdated CQI unreliable. In this work, CQI prediction schemes based on Wiener filter, cubic spline extrapolation and short-term average are studied. The effects of using prediction schemes on data throughput are compared with simulation. A simple adaptation technique is found to make the prediction effective for both high speed and low speed mobile stations.

Keywords—QoS, CQI prediction, SINR, extrapolation, LTE

I. INTRODUCTION

In LTE systems, to enhance the QoS, in terms of data throughput and error protection, adaptive modulation and coding scheme (AMC) is used [1]. AMC requires knowledge of the wireless channel between eNodeB (eNB) and user equipments (UE). The channel information is normally measured by the UE and sent to eNB through a feedback channel. According to 3GPP standards, the feedback of channel information consists rank indicator (RI), pre-coding matrix indicator (PMI), and most importantly, CQI.

CQI is a 4-bit quantized value that indicates an estimation of the modulation and coding scheme (MCS) that the UE can receive reliably from eNB. It reflects the SINR on different physical resource blocks. Several different feedback schemes are introduced by 3GPP. In wideband feedback, only one CQI is provided for the whole frequency band. To facilitate frequency domain scheduling, two other CQI feedback schemes are specified. The UE can report either the CQI for all the subbands or a selected sub set of them. Only subband level feedback is considered in this work. More details of CQI signaling can be found in [2].

Due to the time spent on transmitting and processing signals, a delay between the CQI report and its application in eNB is inevitable. As the channel condition changes constantly over time, the mismatched CQI could consequently result in a reduction of the data throughput. The performance degradation caused by CQI delay is pointed out in [3].

Numerous prediction methods have been investigated to compensate CQI delay. In [4], the prediction error is assumed to follow a normal distribution and the prediction process is actually an order-one autoregressive filtering. A finite-state Markov chain (FSMC) model was proposed in [5], where the states of CQI are modeled with transition probabilities. However, the average signal to interference plus noise ratio (SINR) of all the active UE in a cell must be known as prerequisite. The normalized least mean square (NLMS) filtering was presented in [6] and further improved in [7]. The main drawback of NLMS filtering is its computational complexity and the necessity of training. The more frequent the channel fluctuates, the more training sequences are needed. In addition, NLMS filtering is ineffective when the feedback delay is large. Another scheme, which simply takes a short-term average as predicted value, was also given in [7]. Although seems rough, short-term average is an effective method for high speed UE. The reason is, that with high speed UE, the feedback delay could be much greater than the channel coherence time, thus, the optimal prediction converges to the mean CQI [8].

In this work, several schemes are introduced and compared. Moreover, a prediction scheme based on extrapolation of SINR is proposed. Cubic spline extrapolation can provide near optimal results for low speed UE, however has serious numerical problem when the UE speed is high. Therefore, a simple heuristic using sample autocorrelation is adopted to combine the extrapolation output with the historical average, which produces an accurate prediction of CQI for high speed UE. This approach does neither assume channel stationarity nor require knowledge of moving speed.

II. PRELIMINARIES

A. System model

Consider a LTE downlink, where a given cell is surrounded by $N_{int}$ interfering cells. Assuming all wireless channel responses are flat in one physical resource block (PRB). Without loss of generality, the basic time unit of PRB is assumed to be subframe (1 ms) instead of time slot (0.5 ms). The SINR can be written as

$$\gamma_k = \frac{|h_{0,k}|^2 L_{0,k}}{\sum_{i=1}^{N_{int}} |h_{i,k}|^2 P_{i,k} + \sigma_w^2},$$

(1)

where $k$ is the frequency domain index, $P$ is the transmission power, $L$ is pathloss and $h$ is the complex channel gain. The noise power is denoted by $\sigma_w^2$. The commonly used Rayleigh fading model is applied to the channel response, where the frequency correlation depends on the power-delay...
Due to frequency selectivity, different PRBs generally have different SINRs. To find a proper estimation of the CQI, SINRs on different PRBs must be mapped into an effective SNR. The block error rate (BLER) of the effective SNR in AWGN channel should match the BLER of the original PRB set as

\[ p_e^{(\text{AWGN})}(\gamma_s) \triangleq p_e(\gamma_k, k \in B(s)), \]  

where \( B(s) \) is the set of PRB in subband \( s \).

Using exponential effective signal to noise ratio mapping (EESM) [10], the effective SNR for subband \( s \) can be calculated by

\[ \gamma_s = -\beta \ln \left( \frac{1}{q} \sum_{k \in B(s)} e^{-\frac{\gamma_k}{q}} \right), \]  

where \( \beta \) is a parameter calibrated as a function of the MCS. In this work, \( \beta \) is obtained through extensive simulations, as shown in Tab. I. Each CQI corresponds to a unique combination of MCS, and the BLER of such MCS in AWGN channel can be obtained from simulation, accordingly. In LTE, a BLER smaller than 0.1 is required for AMC. Based on the BLER for different MCS shown in Fig. 1 (a), the CQI feedback can be determined by a step function \( Q_s = f(\gamma_s, dB) \) as depicted in Fig. 1 (b).

### B. Prediction noise

The noise of the prediction procedure consists of the following parts: (1) Estimation noise On UE side, due to imperfect SINR estimation, the estimated SINR is deviated from the true SINR. (2) Compression noise The SINR to effective SNR mapping is a lossy compression. (3) Calibration noise In the SINR to effective SNR mapping, the parameter \( \beta \) is empirically calibrated. Different implementations of the turbo decoder could result in a calibration mismatch. (4) Quantization noise The real valued effective SNR is mapped to a 4-bit CQI index, with a lot of information loss. (5) Noise caused by delay The channel is changing during the time spent on transmitting and processing the feedback information. Thus, the CQI value used for the scheduler might already be outdated.

In this work, the SINR before EESM is assumed to be noiseless. Prediction algorithms aim at minimize the inaccuracy caused by delay, which is a crucial problem when the channel is changing fast or the delay is large. The prediction algorithms can be either based on previous SINR values or previous CQI values. For algorithms, which are based on previous CQIs, the compression noise, calibration noise and quantization noise can be enhanced in the prediction procedure.

### C. Objective function

The most important QoS metrics related to channel quality feedback is throughput and error rate.

For maximizing throughput, the objective function for SINR-based prediction in time interval \([p_{\min}, p_{\max}]\) can be written as

\[ \gamma_k(n) = \arg \max_{\gamma_s(n)} \sum_{n=\min_{\gamma_s}}^{\max_{\gamma_s}} T_s(n), \]  

where \( T_s(n) \) is the total throughput of an UE at discrete time \( n \). It can be calculated from

\[ T_s(n) = \sum_{s=1}^{N_{\text{sub}}} N_{\text{PRB}} N_{\text{sc}} \Delta f R_s(n), \]  

where \( N_{\text{sub}} \) is the number of subbands, \( N_{\text{PRB}} \) is the number of PRB in each subband and \( N_{\text{sc}} \) is the number of subcarriers in one PRB. The subcarrier spacing is denoted by \( \Delta f \). And spectral efficiency, denoted by \( R_s(n) \), is associated with CQI as specified in Tab. I.

Since the error rate has a direct relationship with the accuracy of channel information [9], the minimization of error rate leads to the minimization of the mean square error (MSE) of SINR \( E\{[\gamma_k(n) - \hat{\gamma}_k(n)]^2\} \).

Although intuitively, maximizing throughput and minimizing MSE are similar. They are not necessarily equivalent, especially considering the case with hybrid automatic repeat request (H-ARQ). Mismatched CQIs causes occasional retransmissions. However, retransmission with higher order MCS could have better average throughput than constantly successful transmissions with lower order MCS [12]. Furthermore, in multi-user systems, the throughput depends not only on the CQI feedback but also on the scheduling algorithm. For the sake of simplicity, this work is restricted to single user system.

![Fig. 1](image.png)
The SINR value is only available on the UE side. Thus, SINR-based prediction can only be conducted at the UE. Furthermore, the limited battery life on mobile devices restrict the computational complexity. As an alternative, the prediction can also be based on CQI. The CQI-based prediction can be performed on both UE and eNB side. And since $Q_t$ belongs to a finite set with small cardinality, the computational complexity is rather low. However, the noise enhancements discussed in last section lead to very limited precision.

### III. Prediction Algorithms

In this section, prediction schemes based on filtering and extrapolation are presented. The schemes are explained with SINR prediction, but can be adapted to CQI prediction in a straight forward manner.

#### A. Wiener filter

Most of the existing prediction methods assume the SINR to be stationary and the statistics are simply available. These are very strong assumptions, since with wide-sense stationary (WSS) Rayleigh channels, the SINR is not stationary [13]. Moreover, considering the shift in UE location caused by movements, the SINR varies not only temporally but also spatially. Although the SINR is not a stationary process, it is safe to assume it is short-term stationary, and use sample autocorrelation instead of autocorrelation function (ACF). The sample autocorrelation of window $[n-n_0, n]$ is

$$r_{\gamma\gamma}(n, \tau) = \mathbb{E}\{\gamma_k(n')\gamma_k(n' + \tau)\}, n' \in [n-n_0, n-\tau],$$

where $\tau$ is feedback delay and $n_0 > \tau$ is essential.

The predicted SINR using a order $p$ Wiener filter is given by

$$\hat{\gamma}_k,\text{Wiener}(n + \tau) = \sum_{l=0}^{p-1} w(l)\gamma_k(n - l),$$

where $w$ is the Wiener filter coefficient. The mean square error (MSE) of the predicted SINR can then be written as

$$\varepsilon(n) = \mathbb{E}\{(\gamma_k(n + \tau) - \hat{\gamma}_k,\text{Wiener}(n + \tau))^2\}.$$  

Using the orthogonality principle, the following equation can be obtained [14]:

$$\mathbb{E}\{(\gamma_k(n + \tau) - \hat{\gamma}_k,\text{Wiener}(n + \tau))\gamma_k(n - l)\} = 0 \quad l = 0, \cdots, p - 1.$$  

Substituting (7) into (9), the optimum prediction filter can be obtained as

$$w = R^{-1}r,$$

where

$$w = [w(0), w(1), \cdots, w(p - 1)]^T,$$

$$R = \begin{pmatrix}
    r_{\gamma\gamma}(n, 0) & r_{\gamma\gamma}(n, 1) & \cdots & r_{\gamma\gamma}(n, p - 1) \\
    r_{\gamma\gamma}(n, 1) & r_{\gamma\gamma}(n, 0) & \cdots & r_{\gamma\gamma}(n, p - 2) \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{\gamma\gamma}(n, p - 1) & r_{\gamma\gamma}(n, p - 2) & \cdots & r_{\gamma\gamma}(n, 0)
\end{pmatrix},$$

and

$$r = [r_{\gamma\gamma}(n, \tau), r_{\gamma\gamma}(n, \tau + 1) \cdots, r_{\gamma\gamma}(n, \tau + p - 1)]^T.$$  

It is clear that the performance of Wiener filter depends on the accuracy of the sample autocorrelation. In the ideal case, when the UE moves slowly, or the delay is small, with sufficiently large number of observed samples, the Wiener filter gives minimum mean square error. The computational complexity concentrates on the calculation of weighting factors. It depends to the filter order and how often the sample ACF needs to be updated.

#### B. Extrapolation

Without assuming known statistics, extrapolation is a commonly used tool for forecasting missing values in time-series. The predicted SINR is a function of the measured SINR in a finite window $[n - \kappa, n]$ as

$$\hat{\gamma}_{k,\text{ext}}(n + \tau) = S(n + \tau),$$

where $S(\cdot)$ is the extrapolation function, which can be determined by $\gamma_k(n - \kappa), \gamma_k(n - \kappa + 1), \cdots, \gamma_k(n)$. Most commonly used extrapolation functions are linear extrapolation and cubic spline extrapolation. Linear interpolation is proved to be effective for low UE speed and short delay [15]. In this work, cubic spline extrapolation is considered.

Cubic spline functions are piecewise defined third degree polynomials of the form

$$S(x) = \begin{cases}
    s_1(x) & \text{if } x_1 \leq x < x_2, \\
    s_2(x) & \text{if } x_2 \leq x < x_3, \\
    \vdots & \\
    s_k(x) & \text{if } x_\kappa \leq x < x_{\kappa+1},
\end{cases}$$

where

$$s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i.$$  

Assuming the spline function $S(x)$ and its first and second derivatives $S'(x)$ and $S''(x)$ are continuous on the interval $[x_1, x_{\kappa+1}]$, the unknown parameters can be solved, with the natural boundary condition $S''(x_1) = S''(x_{\kappa+1}) = 0$ [16].

As shown in Fig. 2 (a), with a feedback delay of 10 ms, the predicted SINR $\hat{\gamma}_{k,\text{ext}}$ using cubic spline function matches the original SINR $\gamma_k$ perfectly, when the UE is moving at 3 km/h. However, at higher speed, the mismatch of prediction can be observed. Additionally, high peaks caused by insufficient sampling rate grow, when the UE moves faster as shown in Fig. 2 (b) and (c). The numerical errors can be reduced by introducing additional boundary conditions.

Noticing that the differences between neighboring CQI have small chance to be larger than 1, as shown in Tab. II, an artificial restriction can be added to the spline extrapolation. According to the linear relationship between $\gamma_{s,\text{dB}}$ and $Q_x$, the difference between adjacent $\gamma_{s,\text{dB}}$ should be smaller than $f^{-1}(\Delta Q)$ in dB. Since $\gamma_s$ is effectively an average of $\gamma_k$, the same constraint is applied to $\hat{\gamma}_{k,\text{ext}}$.

<table>
<thead>
<tr>
<th>Speed [km/h]</th>
<th>$p(\Delta Q &gt; 1)$</th>
<th>$p(\Delta Q &gt; 2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$9.97 \times 10^{-2}$</td>
<td>$8.37 \times 10^{-2}$</td>
</tr>
<tr>
<td>10</td>
<td>$1.37 \times 10^{-2}$</td>
<td>$2.51 \times 10^{-2}$</td>
</tr>
<tr>
<td>30</td>
<td>$7.82 \times 10^{-3}$</td>
<td>$3.88 \times 10^{-3}$</td>
</tr>
<tr>
<td>50</td>
<td>$4.52 \times 10^{-2}$</td>
<td>$4.68 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

**TABLE II**: Statistics of differences between neighboring CQI.
C. Combined prediction

As mentioned before, when UE moves fast, the short-term average offers a fairly good prediction. Thus, a proper combination of extrapolation and short-term average should deliver good performance for both low speed UE and high speed UE. The weighted combination can be written as

$$\hat{\gamma}_k(n + \tau) = w_{\text{ext}}(n, \tau) \hat{\gamma}_{k,\text{ext}}(n + \tau) + w_{\text{avg}}(n, \tau) \hat{\gamma}_k(n - \kappa, n),$$

(17)

where $\hat{\gamma}_k$ is the short-term average of $\gamma_k$ from $n - \kappa$ to $n$, $w_{\text{ext}}$ and $w_{\text{avg}}$ are weighting factors, respectively.

As a matter of fact, the ACF is a good indicator of the extrapolation quality. Generally speaking, when the autocorrelation is larger, MSE of extrapolation is smaller. Thus, the ACF can be used as weighting factors.

Consequently, the final prediction can be written as

$$\hat{\gamma}_k(n + \tau) = r_{\gamma\gamma}(n, \tau) \hat{\gamma}_{k,\text{ext}}(n + \tau) + \sqrt{1 - r_{\gamma\gamma}(n, \tau)^2} \hat{\gamma}_k(n - \kappa, n),$$

(18)

which essentially converges to $\hat{\gamma}_{k,\text{ext}}(n + \tau)$ for stationary UE and $\hat{\gamma}_k(n - \kappa, n)$ for very fast UE.

IV. SIMULATION RESULTS

A single user system is simulated. The simulated area is covered by 7 eNBs, which are located in regular hexagonal grids and each equipped with a 3-sector antenna. The most important simulation parameters are summarized in Tab. III. Since the channel of slow UE has less fluctuation and 50 km/h is a common speed limit for most of the urban area, the UE speed range of 10-50 km/h is of interest. For Wiener filter, the sample autocorrelation is evaluated once for 1000 subframes, the filter order is 10. For extrapolation, the prediction window is 10 ms. The short-term average value is also compared, where the predicted SINR is the average of the latest 10 subframes. Another important parameter is the feedback delay. To the authors’ best knowledge, considering all the measurement delay, transmission delay and processing delay, the overall feedback delay is set to 10 ms, which is quite realistic, comparing to most of the earlier works which assume the feedback delay to be 1-5 ms [4] [5] [6] [7].

The performance of different algorithms is evaluated by the throughput. The throughput with perfect CQI feedback is given as theoretical limit. The cumulative density function (CDF) of throughput are compared for UE with speed 10 km/h in Fig. 3, where the throughput gap between no prediction and perfect CQI feedback is relatively small. It can be seen that Wiener filter and extrapolation can provide a prediction gain, while the other schemes performs even worse than no prediction. The extrapolation approach gives a near-optimum performance.

When the UE speed increases, the throughput loss due to feedback delay also increases. The gap between no prediction and perfect CQI feedback is larger. The performance of Wiener
filter degrades rapidly. When the UE speed reaches 50 km/h, the Wiener filter performs rather poor, due to the weak and unreliable autocorrelation of the input data. As a result of the numerical problem discussed in Sec. III, the extrapolation also become worse than no prediction. However, the short-term average offers a better performance, whereas the combined prediction delivers the highest throughput.

The problem about Wiener filter is its sensitivity to the correlation between the filter input and output. As illustrated in Fig. 5, for small delay of 1 ms, the correlation between the filter input and output is strong, and the Wiener filter performs as good as extrapolation. However, the throughput degradation is larger, as the delay increases.

The average throughput of all the aforementioned algorithms is compared in Fig. 6. According to this result, using prediction schemes, the average throughput can be boosted by a factor of up to 15% − 25% percent. The extrapolation scheme dominates the low speed range, while the combined prediction is advantageous when UE is faster.

A simple conclusion can be deduced, that is, to achieve highest throughput, the UE should choose the extrapolation scheme or the combined prediction scheme, according to its moving speed. And the speed can be estimated by observing the correlation of past CQI.

V. CONCLUSION

In this paper, to improve the QoS, a number of CQI prediction schemes for LTE are explained and compared. None of these schemes make strong assumptions on the propagation environment. The Wiener filter is only effective for low speed and small delay. The extrapolation and combined prediction scheme provide good throughput gain for low speed UEs and high speed UEs, respectively. Therefore, an adaptive prediction approach can be made from these two schemes.

REFERENCES