Predicting CSI for Link Adaptation Employing Support Vector Regression for Channel Extrapolation

Allaeddine Djouama*§, Erich Zöchmann†‡, Stefan Pratschner ‡, Markus Rupp ‡, Fatiha Youcef Ettoumi§

*Department of Electronic Engineering, Chonbuk National University, Korea, *adjouama@jbnu.ac.kr*
†Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion
‡Institute of Telecommunications, TU Wien, Austria
§Department of Electronic Engineering, University of Science and Technology Houari Boumedien, Algeria

***Abstract*—Link adaptation in LTE-A is based on channel state information (CSI). For time-selective channels, CSI might be outdated already in the next subframe. Hence, CSI prediction must be employed. This paper investigates support vector regression (SVR) for channel extrapolation and prediction. SVR is applied for learning from the previous channel estimates in order to predict the CSI of the following ones. Simulation results show that the proposed method performs better than simple linear prediction methods and close to minimum mean square error prediction especially in a reasonable signal to noise ratio regime.**
***Keywords*—Support Vector Machines, Channel Estimation, LTE, MMSE, interpolation, extrapolation, CSI prediction**

I. Introduction

The most recent standards for wireless cellular networks employ Orthogonal Frequency Division Multiplexing (OFDM) as multiple access technique. In such systems, the data rate of each user is defined by adaptive modulation and coding (AMC) according to the channel state. In Long Term Evolution (LTE) downlink (DL) the User Equipment (UE) feedbacks the quantized CSI while in the uplink (UL) transmission, the Base Station (BS) can directly estimate the UE channel. To work with non-quantized channel estimates, we restrict ourselves to the UL. Then, the BS performs link adaptation which results in a delay due to the time taken by the physical layer to process the information. The 3GPP reports a delay of at least 5 ms [1].

Few research has been published regarding the process delay effect such as [5], in which the authors proposed a cubic spline extrapolation to obtain a prediction horizon that allows extending the reliability of the channel quality evaluation along time. However, they did not show the limit of their prediction in terms of time-selectivity. Their results are given at 3 km/h. To address this issue, we will investigate the performance of various extrapolation / prediction methods in an LTE-A compliant environment using the Vienna LTE-A Link Level Simulator [6]. The present study follows the approach of [7] which employs support vectors regression (SVR) to perform channel interpolation and extrapolate it over the time horizon.

SVR is related to statistical learning theory [8] and it became popular because of its success in handwritten digit recognition. Hence, it has been investigated in different areas and recently has been applied for interpolation in LTE channel estimation [7], [9], [10], both in UL and DL. It has been shown in [7] that SVR interpolation for channel estimation in LTE outperforms spline and linear interpolation.

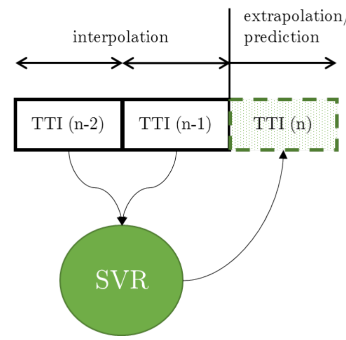

Fig. 1: Illustration of the terms interpolation and extrapolation / prediction.

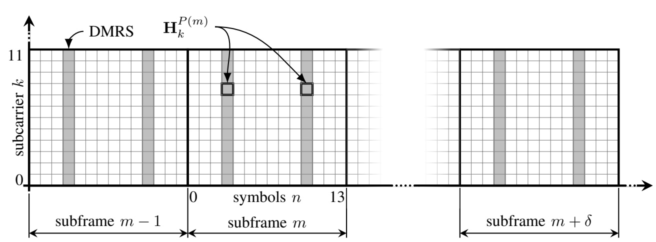

Fig. 2: The LTE-A uplink reference symbol allocation.

The rest of the paper is structured as follows. Section II briefly reviews the LTE-A uplink system model, which continues with the performance of LTE-A (without channel prediction) under various process delays in Section III. SVR is then highlighted in Section IV and some competing methods are introduced. Section V shows a performance comparison with all methods, to show realizable gains. In Section VII we draw some conclusions.

II. System Model

The physical layer of the 3GPP LTE-A uplink employs Single Carrier Frequency Division Multiplexing (SC-FDM). This modulation method is basically a DFT-spreaded version of Orthogonal Frequency Division Multiplexing (OFDM). Further, channel estimation in this system is pilot based, employing demodulation reference signals (DMRS). These are multiplexed in the time-frequency resource grid at every 4^{th} and 11^{th} OFDM symbol of each subframe m, allocating all scheduled subcarriers as illustrated in Figure 2. A subframe, consisting of 14 OFDM symbols, assuming a normal cyclic prefix length, therefore includes two DMRS. As these DMRS are multiplexed after the pre-spreading at the transmitter, and exploited before de-spreading at the receiver for the purpose of channel estimation, the received signal at the channel estimator at symbol time n and subcarrier k is given by

$$\boldsymbol{y}[n,k] = \boldsymbol{H}[n,k]\boldsymbol{x}[n,k] + \boldsymbol{z}[n,k] \ , \tag{1}$$

with transmitted symbol $\boldsymbol{x}[n,k]$, channel coefficient $\boldsymbol{H}[n,k]$ and i.i.d. Gaussian noise $\boldsymbol{z}[n,k] \sim \mathcal{CN}(0,\sigma_z^2)$. For a more detailed derivation of a LTE-A uplink system model we refer to [12] for the SISO case, and to [13], [6] for a MIMO model.

Channel estimates at the pilot positions are obtained by direct channel estimation, such as minimum mean square error (MMSE) estimation as explained in [11], which will be the underlying estimation method exploited in the remainder of this work. To improve readability we denote these channel estimates at pilot positions by vectors $\boldsymbol{H}_k^{P(m)} \in \mathbb{C}^2$ which incorporates all pilot positions of subframe m at subcarrier k, as illustrated in Figure 2.

In order to obtain channel estimates at positions where there are no DMRS allocated in the time-frequency resource grid, interpolation has to be carried out. Due to the special DMRS allocation in the LTE-A uplink, channel estimates at each subcarrier are obtained inherently by estimation. We will therefore only investigate 1D interpolation and extrapolation methods in time domain. While interpolation in the current subframe is necessary for coherent detection at the receiver, extrapolation/prediction is carried out to compensate for processing delay when calculating link adaptation parameters at the base station. For this purpose, channel estimates of the current m^{th} subframe are exploited to extrapolate the channel of a future subframe.

III. Performance without Prediction

Our channels were generated with the Vienna LTE-A simulator [6], which employs an extended Rosa Zheng model [2] introduced in [3]. The Jakes Doppler spectrum results in an autocorrelation function given by a zero'th-order Bessel function of the first kind [4]. The time when the autocorrelation function falls below 0.5 calculates approximately to

$$T_c \approx \frac{9}{16\pi f_d} \ , \tag{2}$$

where $f_d = \frac{v}{c} f_0$ is the maximum Doppler shift dependent on the transmitter speed v, the carrier frequency f_0 and the speed of light c.
If we assume a process delay of 5 ms, we would like to have the current channel realization correlated to the estimated one. Through rearranging Equation (2) and assuming $f_0 = 2\,\text{GHz}$ we obtain a maximum tolerable speed of

$$v = \frac{9}{16\pi} \frac{c}{(5\,\text{ms}) f_0} \approx 5.6 \frac{\text{m}}{\text{s}} \approx 20.2 \frac{\text{km}}{\text{h}} \ . \tag{3}$$

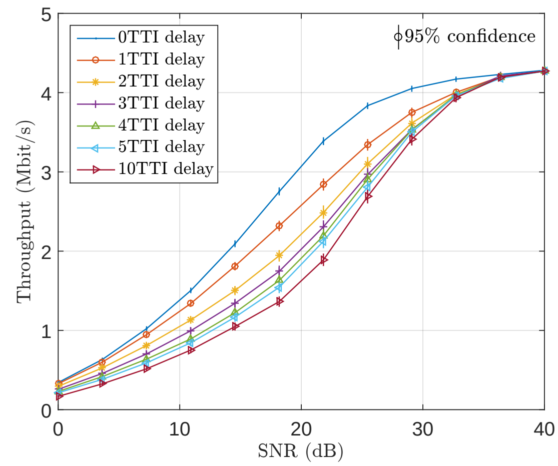

Fig. 3: Throughput with various link adaptation (process) delays and MMSE channel estimation at 20 km/h without prediction.

We notice from Figure 3 that the strongest performance drop occurs for 1-3 Transmission Time Intervals (TTIs) processing delay. If the process delay comes close to the coherence time given in Equation (2) the correlation is so weak, that link adaptation should rather be performed with statistical measures.

IV. Support Vector Regression

SVR is a modified version of the support vector classifier (SVC), which was invented by Vladimir N. Vapnik [8]. In principle, SVR has the same features as SVC, i.e., maximizes the margin and minimizes the errors. The basic idea of Mercer's theorem, as stated in [8], is that a vector in a finite dimensional space (the input space) can be mapped to a higher dimensional space $\mathcal{H}$, which can be an infinite-dimensional Hilbert space provided with a dot product through a nonlinear transformation, $\varphi(.)$. However, the transformation $\varphi(.)$, usually remains unknown (i.e., it is not necessary to know the explicit transformation, $\varphi(.)$). Hence, only the dot product of the corresponding space is required and can be stated as a function of the input vectors as follows:

$$K(u,v) = \langle \varphi(u), \varphi(v) \rangle \tag{4}$$

Such spaces are known as reproducing kernel Hilbert spaces (RKHSs) where $K(u,v)$ is the kernel that should satisfy the conditions of Mercer's theorem, meaning it is the inner product of a Hilbert space; u, v denotes the input vectors (i.e., complex reference symbols). In this work, we are comparing the Gaussian (or Radial Basis Function, RBF) and linear kernels. The Gaussian kernel has been widely used in the SVM framework mainly for its ability to map the input data into an infinite-dimensional space. However, it is very sensitive to speed and require an accurate parameter selection. Thus, we also employ the linear kernel which is a special case of RBF kernel known for its robustness and insensitivity against noisy data. In this paper, we are exploiting the index of the reference symbols in order to estimate the channel frequency response at these positions using minimum mean

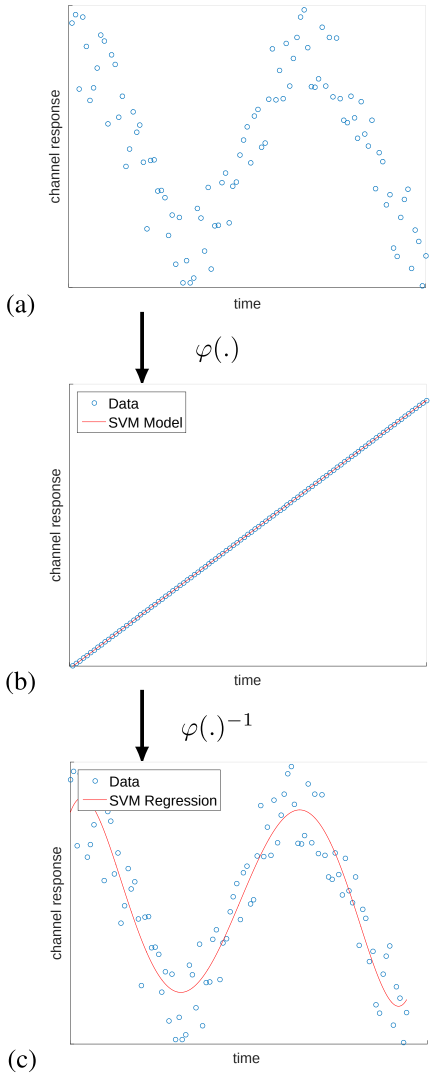

Fig. 4: Steps of SVM regression model. (a) The blue dots illustrate the channel response (input data). (b) Kernel trick for mapping data. (c) The regression model (red solid curve).

square error (MMSE) estimation. Afterwards, we perform the extrapolation of the SVR model constructed in the learning step by obtaining the channel frequency responses for the data symbols at the subframe $m+\delta$. The Gaussian kernel can be stated mathematically as follows:

$$K(u,v)=\exp\left(-\gamma\|u-v\|^2\right) \tag{5}$$

The term $\|u-v\|^2$ is recognized as the squared Euclidean distance between two input vectors; γ is a free parameter that can be chosen after acquiring knowledge about the problem or by running some cross-validation methods. Figure 4 illustrates the behavior of an SVM with nonlinear data using the kernel trick.

The regression model for extrapolation can be illustrated mathematically as follow;

$$\hat{\boldsymbol{H}}_k^{(m+\delta)}=\boldsymbol{w}^T\varphi(\hat{\boldsymbol{H}}_k^{P(m)},m+\delta)+b+e \tag{6}$$

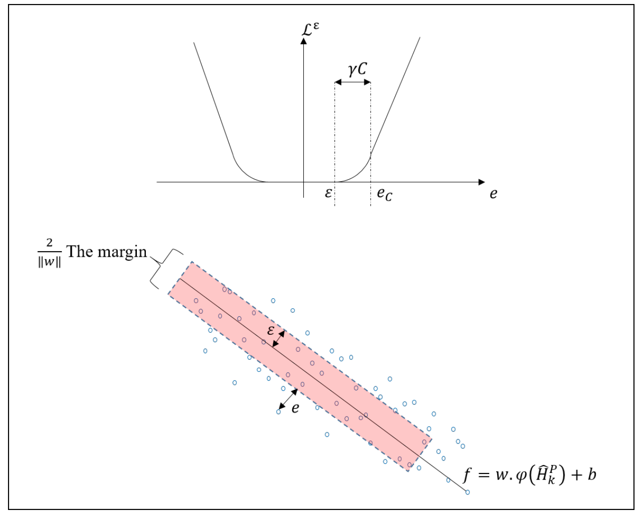

Fig. 5: The ε-insensitivity concept.

where $\hat{\boldsymbol{H}}_k^{(m+\delta)}$, $\hat{\boldsymbol{H}}_k^{P(m)}$ are the channel frequency response on data at subframe $m+\delta$ and reference symbols at subframe m for subcarrier k, respectively. In SVR, the collected information is the values of the parameters $\boldsymbol{w}$ and b, which are the weight vector and the bias term, respectively, and are well known in SVM literature [8]. The error e includes the approximation errors and noise (also known as the residuals). As previously stated, the main goal of the SVM is to maximize the margin (i.e., the weight $\boldsymbol{w}$) and to minimize error. The errors signify a regularized cost function of the residuals, often referred to as Vapniks ε-insensitivity cost function. We use the ε-insensitivity cost function, as illustrated in Figure 5.

$$\mathfrak{L}^{\varepsilon}(e)=\begin{cases}0, & |e|\le\varepsilon\\ \frac{1}{2\gamma}\left(|e|-\varepsilon\right)^2, & \varepsilon\le|e|\le e_c\\ C(|e|-\varepsilon)-\frac{1}{2}\gamma C^2, & e_c\le e\end{cases} \tag{7}$$

where $e_c=\varepsilon+\gamma C$; ε is the insensitivity zone in which the errors are ignored, and γ and C are the parameters that control the trade-off between regularization and loss. The cost function is linear for errors above e_c and is quadratic for errors between ε and e_c. Note that the ℓ_2-norm value of the errors is applied for the quadratic cost zone. A detailed mathematical derivation of SVR for interpolation was published in our previous paper in [7].

In general, the RBF kernel is a reasonable first choice. This kernel nonlinearly maps samples into a higher dimensional space. For the linear kernel, the same principles of the Gaussian Kernel remains since the linear kernel is considered as a special case of the Gaussian kernel. However, the loss function for the linear kernel ignores the quadratic zone as shown in (8).

$$\mathfrak{L}^{\varepsilon}(e)=\begin{cases}0, & |e|\le\varepsilon\\ |e|-\varepsilon, & \text{otherwise}\end{cases} \tag{8}$$

In order to choose the best kernel with the appropriate parameters, we compared the RBF and linear kernel in terms

TABLE I: SVM Parameters

	Parameter set 1	Parameter set 2
C	1	100
γ	0.0618	0.00005
ε	0.000618	25×10^{-11}

of mean square error (MSE) for interpolation with two sets of parameters as shown in Table 1. One should keep in mind that the parameters selection was performed manually after gaining knowledge of the problem. Figure 6 shows the MSE performance comparison between the linear and RBF kernel using 2 and 4 points. We clearly observe that at a relatively low speed ($\leq 50\,\mathrm{km/h}$), using more points for the linear kernel improves the performance. However, the RBF kernel is almost insensitive to an increasing number of points at low speed and it even results in worse performance at higher speed because of over-fitting. Figure 7 shows the same effect concerning the number of points and we clearly recognize that a linear kernel is almost insensitive to parameter change which is not the case for the RBF kernel. This effect is due to the γ value which is an important parameter for the RBF kernel that needs to be chosen carefully. As the 2 points RBF kernel with parameter set 2 performs close to 4 points linear kernel, we prefer the RBF kernel since there's less limitations on frequency hopping; the scheduled resource blocks are allowed to change from subframe to subframe.

A. Linear MMSE Interpolation

In this section a 1D linear MMSE (LMMSE) interpolation filter is described as in [14]. By LMMSE interpolation in time domain, exploiting time correlation within the current subframe m, channel estimates at the data positions are obtained by

$$\hat{\boldsymbol{H}}[n,k] = \left(\boldsymbol{r}^t_{Hn}\right)^H \left(\boldsymbol{R}^t_{HH} + \frac{\sigma_n^2}{\sigma_x^2}\boldsymbol{I}\right)^{-1} \hat{\boldsymbol{H}}_k^{P(m)} \qquad (9)$$

with the time domain channel autocorrelation at pilot symbol positions $\boldsymbol{R}^t_{HH} = \mathbb{E}\{\boldsymbol{H}_k^{P(m)}(\boldsymbol{H}_k^{P(m)})^H\}$ and the time domain channel cross-correlation of symbol times in between pilot positions $\boldsymbol{r}^t_{Hn} = \mathbb{E}\{\boldsymbol{H}_k^{P(m)}\boldsymbol{H}[k,n]^\star\}$. By applying Equation (9) on time positions n that do not lie in between the two DMRS of the current subframe m, LMMSE extrapolation is inherently performed. As the LMMSE extrapolation serves as a performance bound for comparison of simulation results, the channel time correlation is assumed to be perfectly known. Since Jakes' Doppler spectrum is assumed, as described in Section III, the correlation functions are given by the zero'th order Bessel function of first kind as in [14].

B. Linear Least Square Fit

Due to the structural similarity with the linear kernel, we compare SVR to a linear Least Square (LS) fit as well. The LS fit was performed with the minimum of two pilot symbols, which actually resembles linear interpolation.

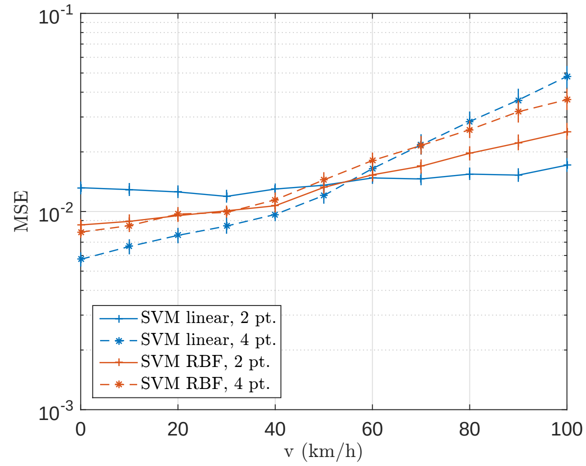

Fig. 6: MSE comparison for parameter set 1.

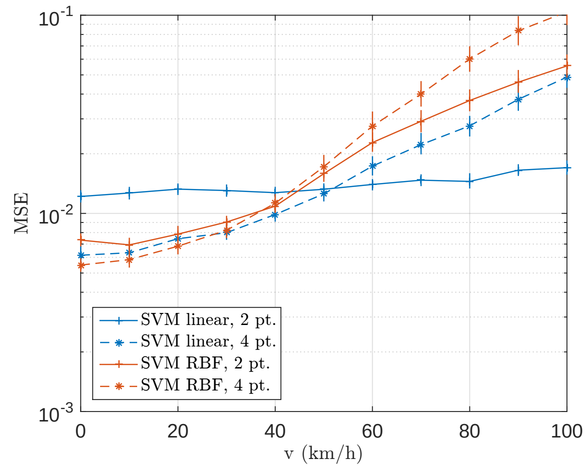

Fig. 7: MSE comparison for parameter set 2.

V. Performance with Prediction

Within this section we compare the performance of the various prediction schemes utilizing two metrics:

- Signal to Interference and Noise Ratio Mean Square Error (SINR MSE): Our link adaptation scheme is fully described in [6] and is essentially a function of the SINR only. Hence, in order to adapt the link correctly, the average SINR must be predicted precisely.

$$\text{SINR MSE} = \sigma_z^2\,|\text{SINR}_{\text{predicted}} - \text{SINR}_{\text{true}}| \quad (10)$$

- Throughput: This metric is directly related to the SINR MSE. Those prediction methods with lower SINR MSE come with higher throughput. In order to compare the gains quantitatively, we also provide the throughput results.

For the simulations we employed the parameters specified in Table II. A real system can never adapt a link faster than 1 TTI. Therefore, it reflects the absolute minimum possible

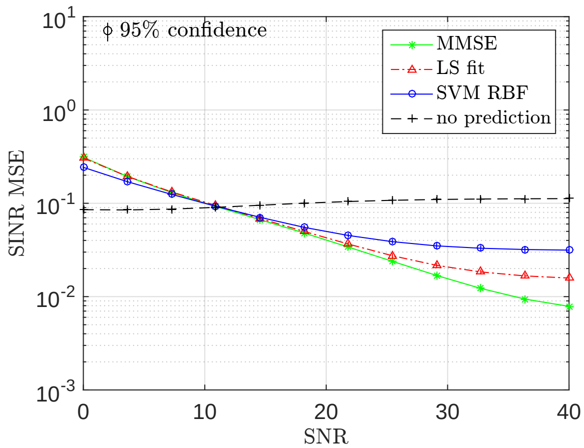

Fig. 8: SINR MSE for link adaptation delay of 1 TTI.

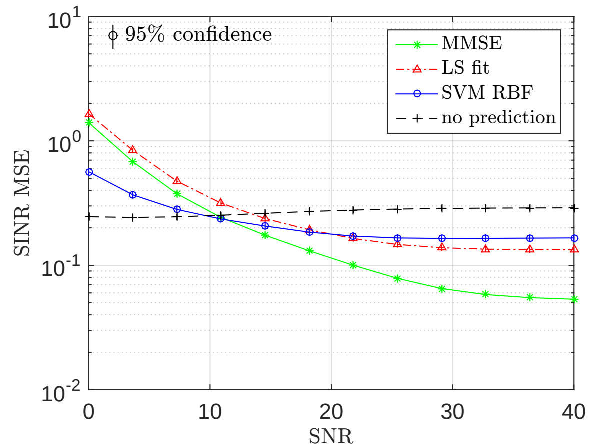

Fig. 10: SINR MSE for link adaptation delay of 3 TTIs.

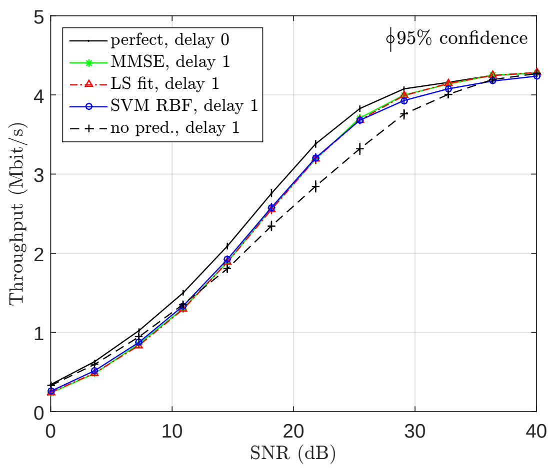

Fig. 9: Throughput MSE for link adaptation delay of 1 TTI.

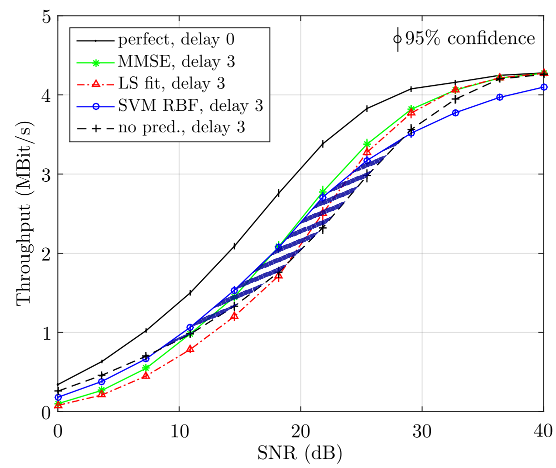

Fig. 11: Throughput for link adaptation delay of 3 TTIs.

delay; 3 TTIs process delay reflects a fast link adaptation and 5 TTIs delay is the already mentioned 3GPP recommendation.

TABLE II: Simulation Parameters

Parameter	Value
System bandwidth	1.4 MHz
Number of subcarriers N_{SC}	72
Process delay	1,3,5 TTI
Channel model	Typical Urban (TU)
Antenna configuration	1×1 uplink
Receiver	MMSE
Channel estimator	MMSE
Modulation scheme	Adaptive
UE speed	20 km/h

For the case of 1 TTI link adaptation delay, the SINR MSE is in the same order for all prediction methods, cf. Fig 8. The slight difference in SINR MSE is hardly visual in the obtained throughput, as illustrated in Fig. 9. In such very fast link adaptation scenarios one would certainly prefer the low complexity LS fit approach.

The situation changes a bit in Fig. 10 as channel correlation weakens. For a 3 TTI process delay SVR outperforms MMSE and LS prediction in the low SNR regime. However, no prediction would be beneficial there. In the region between 10 and 25 dB SNR SVR performs as good as MMSE prediction and much better than LS, cf. Fig. 11. As this region is a likely point of operation and a 3 TTI link adaptation delay is feasible, we see the major gain of our method there. The shaded region shows the gain obtained by SVM against no prediction. Between 10 and 25 dB SNR prediction amounts to an SNR gain of roughly 3 dB. A gain of this order was also reported in [15].

Once we come close to the point, where the autocorrelation drops below half, we recognize that the SINR MSE is generally very high and does not differ significantly from no prediction, cf. Fig 12. The gains in throughput depicted in Fig. 13 are almost negligible. The saturation of SVM at higher SNR as compared to other methods stems from a conservative SINR prediction of SVM. To further explain this effect, the corresponding BLock Error Ratio (BLER) is depicted in Figure 14. SVM prediction leads to a relatively low BLock Error Ratio (BLER) compared to other methods. This is contrary to the

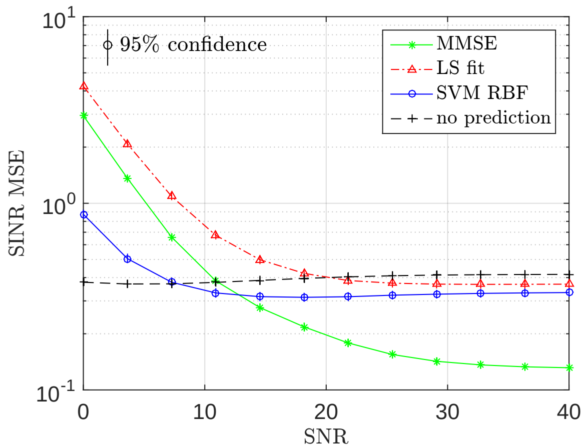

Fig. 12: SINR MSE for link adaptation delay of 5 TTIs.

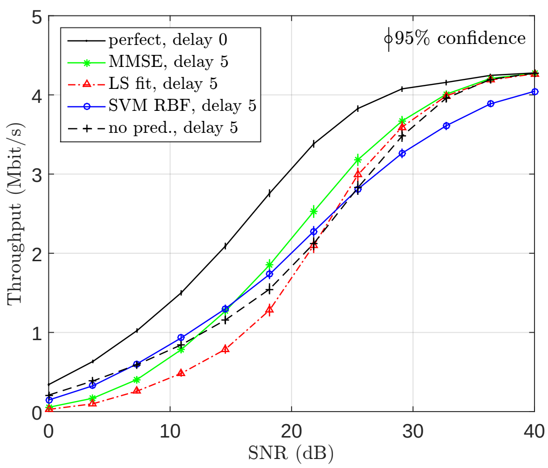

Fig. 13: Throughput for link adaptation delay of 5 TTIs.

LS fit which generally tends to predict too high SINR, visual in the higher BLER.

VI. Restrictions for MIMO transmissions

In LTE-A uplink the pilots are inserted before the spatial precoding, so that effective channels (including the precoders) are estimated [6], [11]. If the transmission rank is lower than the number of transmit antennas, the effective channel will not be recovered by knowledge of the precoder. To apply our method on MIMO transmissions, the applied precoder must stay constant over a certain time. Our prediction strategies are operating on the effective channel matrices then.

VII. Conclusions

For an LTE-A uplink transmission model we investigated the performance of SVR for channel prediction / extrapolation. It has been shown that SVR performs close to MMSE prediction. In the low SNR regime it is beneficial to not consider prediction at all, where as in the high SNR regime simpler methods such as LS fitting are superior to SVR. SVR offers some gains in the medium SNR range and as long

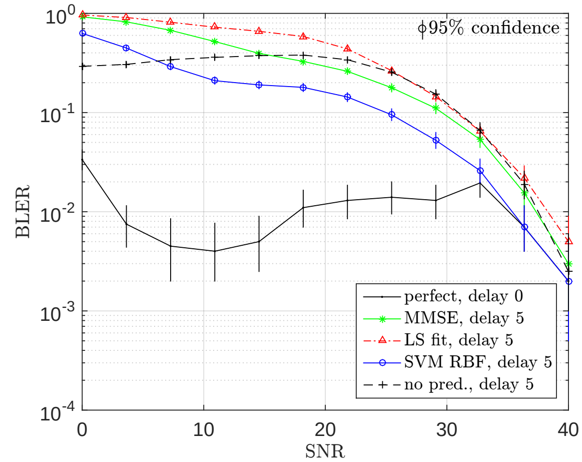

Fig. 14: Deviation from the target BLER (0.1) for various prediction schemes.

as the channels to be predicted are still strongly correlated to the channel at hand. The results obtained are promising for improving the throughput performance in case of link adaptation process delay.

VIII. Acknowledgment

This work has been funded by UNESCO Keizo Obuchi young researchers' fellowship programme. A1 Telekom Austria AG and the Institute of Telecommunications, TU Wien. The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged.

References

[1] 3GPP "LTE-Advanced Evaluation Workshop", Dec. 17-18, 2009

[2] Zheng, Y.R., Chengshan Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," in IEEE Transactions on Communications, vol.51, no.6, pp.920-928, 2003

[3] Zemen, T.; Mecklenbräuker, C.F., "Time-Variant Channel Estimation Using Discrete Prolate Spheroidal Sequences," in IEEE Transactions on Signal Processing, vol.53, no.9, pp.3597-3607, 2005

[4] Clarke, R.H., "A statistical theory of mobile-radio reception," in The Bell System Technical Journal, vol.47, no.6, pp.957-1000, 1968

[5] Lema, M. A., García – Lozano, M., and Ruiz, S. "Cubic spline extrapolation for uplink channel quality prediction in LTE–A with carrier aggregation." in Transactions on Emerging Telecommunications Technologies, early view, 2015

[6] Zöchmann, E., et al. "Exploring the Physical Layer Frontiers of Cellular Uplink – The Vienna LTE-A Simulator." arXiv preprint arXiv:1509.02351 , 2015

[7] Djouama, A., Myoung-Seob L., and Fatiha Y. E. "Channel Estimation in Long Term Evolution Uplink Using Minimum Mean Square Error–Support Vector Regression." Wireless Personal Communications vol.79.3, pp.2291-2304, 2014

[8] Vapnik, Vladimir. The nature of statistical learning theory. Springer Science & Business Media, 2000.

[9] Garcia, M. J. F. G., et al. "Support vector machines for robust channel estimation in OFDM." in IEEE Signal Processing Letters, vol. 13.7, pp.397-400, 2006

[10] Charrada, Anis. ”Support Vector Machines Regression for MIMO-OFDM Channel Estimation.” in IAES International Journal of Artificial Intelligence, vol. 1.4, pp.214-224, 2012

[11] Pratschner, S., Zöchmann E., and Rupp M. ”Low complexity estimation of frequency selective channels for the LTE-A uplink.” in IEEE Wireless Communications Letters, vol.4, no.6, pp. 673-676, 2015

[12] Zöchmann, E., et al., ”Limited feedback in OFDM systems for combating ISI/ICI caused by insufficient cyclic prefix length,” in Proceedings of 48th Asilomar Conference on Signals, Systems and Computers, 2014

[13] Zöchmann, E., et al. ”MIMO Transmission over High Delay Spread Channels with Reduced Cyclic Prefix Length.” in Proceedings of 19th International ITG Workshop on Smart Antennas; VDE, 2015

[14] Dong, X., Lu, W. S., and Soong, A. C., ”Linear Interpolation in Pilot Symbol Assisted Channel Estimation for OFDM,” in IEEE Transactions on Wireless Communications, vol.6, no.5, pp.1910-1920, 2007

[15] Sahlin, Henrik. ”Channel prediction for link adaptation in LTE uplink.” in IEEE Vehicular Technology Conference (VTC Fall), 2012