

Link Adaptation in MIMO-OFDM with Practical Impairments

Alberto Rico-Alvarino

Signal Theory and Communications Department
Universidade de Vigo 36310 Vigo, Spain
alberto@gts.uvigo.es

Robert W. Heath Jr.

Dept. of Electrical and Computer Engineering
The University of Texas at Austin, TX, USA
rheath@utexas.edu

Abstract—Performing link adaptation in multiple-input-multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems is challenging due to the problem of mapping the channel state information to a frame error rate (FER) value. This difficulty comes from the spatial and frequency selectivity of the channel, which makes the different symbols in a codeword to observe different signal to noise ratio values. Moreover, practical impairments like non-Gaussian noise, different frame lengths, or channel nonlinearities can severely affect the adaptation procedure in real scenarios. In this paper we study different FER prediction techniques, which can be classified in *parametric*, *non-parametric* and *semi-parametric*. We evaluate the performance of the FER predictors under practical impairments, and compare the achieved throughput when used in conjunction with link adaptation algorithms.

I. INTRODUCTION

The increasing complexity of modern wireless communication systems makes their performance characterization a difficult task. The use of simple metrics like average signal to noise ratio (SNR) does not suffice to characterize systems with complex channel coding operating under frequency and space selective channels. Performance characterization (in terms of frame error rate - FER) of the physical layer (PHY) of wireless communication systems, or *PHY abstraction*, is crucial to perform some tasks such as adaptive modulation and coding (AMC) or system level evaluations without resorting to time-consuming simulations.

Classic performance metrics used for single-carrier coded systems [1], e.g. SNR, are no longer valid for multiple-input-multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) systems. Approaches using *indicator functions* that map the set of SNR values (one for each carrier and spatial stream) to an *effective SNR metric* (ESM) were proposed as an alternative [2], [3].

One of the main criticisms to ESM is the impossibility to incorporate practical impairments or different transmission parameters (like codeword length or noise distribution) into the FER prediction. In IEEE 802.11ac [4], for example, the frames have variable length, which is going to affect the performance of the system. For a constant channel state, longer packets will

This work was performed while A. Rico-Alvarino was visiting the Wireless Networking and Communications Group at The University of Texas at Austin. Work by A. Rico-Alvarino supported by a Fundacion Pedro Barrie de la Maza graduate scholarship. R. W. Heath Jr. was supported in part by the National Science Foundation under Grant No.1218338 and 1319556.

experience higher FER values [5]. Also, different noise distributions will lead to different FER values, even with the same SNR value. Although the assumption of Gaussian noise is common in system design, a generalized Gaussian distribution can model more accurately some communication scenarios with Laplacian noise [6] or tropospheric impulse noise [7]. Learning-based approaches were developed to overcome these problems [8]–[10]. Within the learning framework, the FER prediction is based on observed FER samples, so the effect of impairments is already captured in these measurements. In this setting, the effect of having FER samples with different codeword length or different noise distribution has not been studied.

Previous work [2], [3], [8] focused on predicting the FER from SNR information, without taking into account possible changes in the noise distribution or frame length. In this paper, we design FER predictors that are able to capture the effect of these practical impairments. We review ESM and machine learning FER predictors, and propose modifications to include additional parameters in the prediction. We design a new *semi-parametric* FER predictor by combining ideas from ESM and machine learning. The results show that ESM loses accuracy when practical impairments are present, and that the proposed methods can be used to overcome this problem.

II. SYSTEM MODEL

Consider a point to point communication system where a transmitter, equipped with T antennas, communicates with a receiver, equipped with R antennas. Communication takes place over an N -carrier OFDM physical layer. For every carrier $n = 1, \dots, N$ we denote by $\mathbf{H}_n \in \mathbb{C}^{R \times T}$ the MIMO channel, by $\mathbf{n}_n \in \mathbb{C}^R$ the received noise vector, and by $\mathbf{x}_n \in \mathbb{C}^T$ the transmit signal. The received signal $\mathbf{y}_n \in \mathbb{C}^R$ is

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n + \mathbf{v}_n \quad n = 1, \dots, N. \quad (1)$$

We restrict our analysis to transmitters using linear precoders $\mathbf{F}_n \in \mathbb{C}^{T \times M}$ to spatially conform the transmit symbols $\mathbf{s}_n \in \mathbb{C}^M$ and receive equalizers $\mathbf{G}_n \in \mathbb{C}^{M \times R}$. The number of spatial streams (NSS), also called the *mode*, is denoted by M .

We assume perfect channel state information at both transmit and receive ends. Therefore, we can apply singular value decomposition (SVD) precoding so

$$\mathbf{r}_n = \mathbf{G}_n \mathbf{H}_n \mathbf{F}_n \mathbf{s}_n + \mathbf{G}_n \mathbf{v}_n = \mathbf{\Lambda}_n \mathbf{s}_n + \mathbf{w}_n \quad n = 1, \dots, N \quad (2)$$

where $\mathbf{\Lambda}_n$ is a diagonal matrix including the first M singular values of matrix \mathbf{H}_n , and $\mathbf{w}_n \triangleq \mathbf{G}_n \mathbf{v}_n$. Design of power allocation is out of the scope of this work, and is assumed to be included in $\mathbf{\Lambda}_n$. The effect of SVD precoding is the decomposition of the MIMO channel in a set of M scalar channels, each one with an input-output relationship described by

$$r_{n,i} = \lambda_{n,i} s_{n,i} + w_{n,i}. \quad (3)$$

Therefore, each symbol $s_{n,i}$ passes through a flat fading channel with an SNR value of $\gamma_{n,i} = \frac{|\lambda_{n,i}|^2}{\sigma^2}$. We define the SNR vector $\boldsymbol{\gamma}$ as

$$\boldsymbol{\gamma} = [\gamma_{1,1}, \dots, \gamma_{N,1}, \dots, \gamma_{N,M}]^T. \quad (4)$$

In this paper, we consider zero-mean generalized complex Gaussian noise [11]. This generalized model allows to treat Laplacian and Gaussian noise as special cases. The probability density function of the real and imaginary parts of the noise is

$$f(x) = \frac{\rho}{2\beta\Gamma(1/\rho)} \exp\left(-\left|\frac{x}{\beta}\right|^\rho\right) \quad (5)$$

with β the scale parameter, and ρ the shape parameter. Roughly speaking, the parameter ρ changes the rate of decay of the tails of the probability distribution, and β changes the variance for a fixed ρ . For example, if $\rho = 1$ the noise is Laplacian, and if $\rho = 2$ the noise is Gaussian. The variance of the complex noise is

$$\sigma^2 = \frac{\beta^2 \Gamma(3/\rho)}{2 \Gamma(1/\rho)}. \quad (6)$$

The transmitted symbols $s_{m,i}$ are the result of processing blocks of bits. Every block of bits is independently processed, and constitutes a frame. The size L of the frame is variable, and depends on the size of higher layer protocol data units. The frame is constituted after performing forward error correction (FEC) coding over blocks of bits, interleaving, and constellation mapping. The transmitter selects the modulation and coding scheme (MCS) from a discrete set $\mathcal{C} = \{c_1, \dots, c_C\}$.

The problem we address in this paper is how to estimate the FER associated to a channel state, MCS, ρ and L . Particularly, for each MCS value, we are interested in a function

$$\eta(\boldsymbol{\gamma}, \rho, L) \quad (7)$$

that maps the set of SNR values $\boldsymbol{\gamma}$, the frame length L and the shape parameter ρ to a FER value. We focus first on the case of Gaussian noise and fixed codeword length L to illustrate two different approaches to FER prediction. We extend these metrics in Section IV to deal with practical impairments.

III. FER PREDICTION TECHNIQUES

The involved structure of practical coding schemes makes the analytical study of the FER function complicated. In this paper, we classify the FER prediction approaches into *parametric* and *non-parametric*. Although both approaches need to use empirical FER results, the main difference is that the parametric approaches require adjusting some parameters

following a mean square error (MSE) fitting, for example, while non-parametric methods require adjusting the actual model, usually by cross-validation techniques. In this section, we review ESM and learning FER predictors.

A. Parametric FER prediction

Parametric techniques assume some functional relationship between the SNR values and the FER, with some parameters to be adjusted according to empirical measurements. This functional relationship is usually expressed as the composition of two different mappings, λ and γ_{eff} [3]

$$\eta(\boldsymbol{\gamma}) = \lambda(\gamma_{\text{eff}}(\boldsymbol{\gamma})). \quad (8)$$

The first one is a mapping from the SNR vector $\boldsymbol{\gamma}$ to an ESM γ_{eff} , defined as the SNR value of an AWGN channel with the same FER as the fading channel under study. This mapping is as a generalized mean that maps the SNR values to a *quality* domain, averages the quality measurements, and maps the value back to the SNR domain. If we denote by $\Theta(\cdot)$ the quality mapping, γ_{eff} is defined as

$$\gamma_{\text{eff}}(\boldsymbol{\gamma}) = \Theta^{-1}\left(\frac{1}{M} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \Theta(\gamma_{i,j})\right). \quad (9)$$

In this paper, we consider the Exponential Effective SNR metric (EESM) due to its analytical tractability and good accuracy. For example, the WiMAX forum recommended the EESM as the default method for FER prediction [12] in IEEE 802.16e. In EESM, the quality mapping is

$$\Theta(x) = e^{-\beta x} \quad (10)$$

with β the parameter to be adjusted with empirical information.

The second mapping is a function from the SNR domain to the FER domain. More precisely, $\lambda(x)$ is the FER of an AWGN channel with SNR x . This mapping is usually performed by the use of look-up tables (LUT) containing simulation results for the AWGN channel. To make a fair comparison with the non-parametric approach, which does not assume any prior FER information in AWGN, in this paper we consider a functional relationship between SNR and FER in AWGN. Particularly, we consider a generalized sigmoid function

$$\lambda(x) = \frac{1}{(1 + \exp(b(x - m)))^{1/\nu}} \quad (11)$$

with ν , b and m to be fitted to empirical measurements. We verified by simulations that the performance of the FER predictor with the sigmoid function (11) and a LUT is similar.

B. Non-parametric FER prediction

We follow some ideas from [8], and propose to use learning-inspired methods to perform FER prediction. Based on some past samples (the training data), a regressor tries to estimate the function value (the FER in our case) in a different set of samples (the test data). Note that our approach differs from [8], [9], where the objective was to discriminate whether the FER

is above or below a certain threshold, instead of predicting the FER value.

If we assume ideal interleaving, the FER is going to be invariant to permutations in γ , so the *ordered* SNR vector $\hat{\gamma} = \mathbf{P}\gamma$ suffices to obtain the FER performance. \mathbf{P} is a permutation matrix such that $[\hat{\gamma}]_i \geq [\hat{\gamma}]_{i+1} \forall i$.

The regression problem exploits the information in a set of *training data* $\{(\hat{\gamma}_i, y_i)\}_{i=1}^S$, consisting of duples of SNR of different channel realizations $\hat{\gamma}_i$ and its associated FER value y_i . The objective of the regression function is to obtain the FER y_0 associated to a different channel realization $\hat{\gamma}_0$. This regression problem involves two steps: dimensionality reduction and regression.

We select a reduced dimension *feature vector* \mathbf{f}_i from the data $\hat{\gamma}_i$ to avoid the curse of dimensionality [13]. We restrict our study to affine operations, so $\mathbf{f}_i = \mathbf{R}(\hat{\gamma}_i - \mathbf{r})$, with \mathbf{R} a dimensionality reduction matrix, and $\mathbf{f}_i \in \mathbb{R}^D$. We study the following dimensionality reduction techniques

- 1) **Subset selection** This method simply selects some entries of vector $\hat{\gamma}_i$. Thus, $\mathbf{r} = \mathbf{0}$ and \mathbf{R} is a sparse matrix with D rows taken from of the canonical base of \mathbb{R}^{NM} . Although the subset of selected entries could be optimized to gain some performance, in this paper we reduce our analysis to matrices \mathbf{R} selecting equidistant SNR positions, including the first and last ones.
- 2) **Subset selection with feature scaling** This method selects some entries of the vector $\hat{\gamma}_i$, but performs first an affine transformation to make the different entries of $\hat{\gamma}_i$ zero mean and unit variance. We define the empirical mean and variance of the entries of $\{\hat{\gamma}_i\}_{i=1}^S$ as $\mu_k \triangleq \frac{1}{S} \sum_{i=1}^S [\hat{\gamma}_i]_k$ and $\sigma_k^2 \triangleq \frac{1}{S} \sum_{i=1}^S |[\hat{\gamma}_i]_k - \mu_k|^2$. The dimensionality reduction operation is

$$\mathbf{f}_i = \mathbf{R}\Sigma(\hat{\gamma}_i - \boldsymbol{\mu}) \quad (12)$$

with Σ a diagonal matrix with entries $[\Sigma]_{k,k} = \frac{1}{\sqrt{\sigma_k^2}}$,

$\boldsymbol{\mu} = [\mu_1, \dots, \mu_{NM}]^T$, and \mathbf{R} a selection matrix.

- 3) **Principal component analysis** Principal component analysis (PCA) estimates the mean and covariance matrix of the samples as $\boldsymbol{\mu} = \frac{1}{S} \sum_{i=1}^S \hat{\gamma}_i$, $\mathbf{C} = \frac{1}{S} \sum_{i=1}^S (\hat{\gamma}_i - \boldsymbol{\mu})(\hat{\gamma}_i - \boldsymbol{\mu})^T$. The feature set is obtained by projecting the training set onto the dominant eigenmodes: let $\mathbf{C} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ be the eigendecomposition of \mathbf{C} with the eigenvalues sorted in decreasing order, and let $\tilde{\mathbf{U}}$ be the matrix containing the first D columns of \mathbf{U} . The dimensionality reduction operation is $\mathbf{f}_i = \tilde{\mathbf{U}}^T(\hat{\gamma}_i - \boldsymbol{\mu})$.

After dimensionality reduction is performed, we build a regression function based on the reduced dimension training data. The reduced dimension training data is the set $\{(\mathbf{f}_i, y_i)\}_{i=1}^S$. Although there are a wide variety of non-parametric regression methods, we choose local linear regression (LLR) for its simplicity. LLR approximates the function around a point \mathbf{f}_0 by a linear model.

$$\hat{y}_0 = \alpha + \boldsymbol{\beta}^T \mathbf{f}_0. \quad (13)$$

The parameters α and $\boldsymbol{\beta}$ are obtained from weighted least squares (WLS) fitting. The weights of the WLS problem depend on the distance between \mathbf{f}_0 and the different training samples \mathbf{f}_i . The WLS problem is

$$\min_{\alpha, \boldsymbol{\beta}} \sum_{i=1}^S K_\lambda(\mathbf{f}_0, \mathbf{f}_i) \left\| \alpha + \boldsymbol{\beta}^T \mathbf{f}_i - y_i \right\|^2 \quad (14)$$

where K_λ is a kernel function parametrized by the value λ . We use the radial basis function kernel:

$$K_\lambda(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\lambda}\right). \quad (15)$$

The value of λ determines a point in the bias vs variance tradeoff [14].

IV. INCLUSION OF PRACTICAL IMPAIRMENTS

The methods described in the previous section perform FER prediction based only on SNR information. In real systems, however, there are other factors that impact the performance of a receiver. In this section we describe how to include practical impairments in parametric and non-parametric techniques, and present a new semi-parametric approach.

A. Parametric FER prediction

Parametric approaches are difficult to adapt to include practical impairments, as a functional relationship between the FER and the practical impairment has to be obtained or approximated. In the case of variable frame length, for example, previous work assumed the availability of a different FER predictor for every length [3], [5]. In this paper, we assume only one FER predictor for every MCS, so the parametric FER predictors do not take into account the practical impairments.

B. Non-parametric FER prediction

The non-parametric FER prediction techniques offer a flexible way to deal with practical impairments, since they do not assume any functional relationship between the FER and the channel state. Thus, we can define an *extended channel state* that includes the practical impairment as part of the channel state. Assume that we have a training set where each sample is associated to a channel state vector and to a practical impairment vector. Denote $\mathbf{p}_i \triangleq [p_{i,1}, \dots, p_{i,P}]^T$ as the practical impairment vector of the i -th sample. The training set is a set of tuples $\{\hat{\gamma}_i, \mathbf{p}_i, y_i\}_{i=1}^S$. The extended channel state vector is defined as $\mathbf{e}_i = [\hat{\gamma}_i^T, \mathbf{p}_i^T]^T$. With this definition, we can redefine our training set as $\{\mathbf{e}_i, y_i\}_{i=1}^S$ and apply the dimensionality reduction methods in Section III-B.

The inclusion of practical impairments in the channel state vector increases the dimensionality of the problem and, therefore, makes the use of dimensionality reduction even more important. In the following section, we design a semi-parametric approach that combines ideas from ESM and machine learning approaches to avoid the problem of dimensionality reduction.

C. Semi-parametric FER prediction

Parametric and non-parametric methods have some advantages and drawbacks. On the one hand, parametric methods use simple functional relationships between FER and SNR, work with a relatively low number of empirical samples, but are not flexible to accommodate practical impairments. On the other hand, non-parametric methods can deal with practical impairments in a straightforward manner, but need a large number of training samples to include additional features.

One key observation is that ESMs are designed to be *good* dimensionality reduction techniques, i.e., ideally the effective SNR $\gamma_{\text{eff}}(\hat{\gamma})$ is a sufficient statistic for FER estimation. Thus, we propose to use an alternative extended channel state vector, defined as

$$\mathbf{e}_i = [\gamma_{\text{eff}}(\hat{\gamma}_i), \mathbf{p}_i^T]^T. \quad (16)$$

The estimation process involves two steps. In the first one, the optimum ESM parameter β is obtained from the training samples without taking into account the practical impairments. In the second one, the value of β is used to build the extended channel state vector (16), and an LLR estimator is trained following the procedure in III-B. Dimensionality reduction is not performed (the number of practical impairments is expected to be small, and we have already reduced the size of the SNR vector to one), but feature scaling might be necessary.

V. SIMULATION RESULTS

We evaluated the described FER prediction methods for a MIMO-OFDM system with a 4-antenna transmitter-receiver pair and 52 carriers, emulating the PHY of IEEE 802.11ac [4]. The MCS and its associated rates can be found in [15].

We performed different experiments to compare the performance of the FER prediction methods. First, we compared parametric and non-parametric methods when no practical impairments are present. We generated 6000 different realizations with SNR values between 0 and 30dB, and a 4-tap MIMO channel with Gaussian entries in the time domain. The frame length was set to $L = 1024$, and the noise distribution was Gaussian. We simulated the transmission over the channel with QPSK modulation, rate 3/4 convolutional code, and 4 spatial streams. We divided the 6000 data points into two different parts: the training data, comprising 80% of the points, and the test data, with 20% of the points. We trained our regressor with the 4800 samples, and tested it against the remaining 1200. We compared the ESM FER predictor with LLR with the 3 different types of dimensionality reduction: LLR-PCA (dimensionality reduction with PCA), LLR-SS (dimensionality reduction with subset selection), LLR-SS-SC (dimensionality reduction with subset selection and feature scaling). Also, we evaluated LLR-SS and LLR-SS-SC with the SNR values in decibels instead of natural units.

We selected the kernel parameter λ following a K -fold cross-validation approach [14], with $K=4$. This implies that every iteration in this cross-validation used 3600 samples to train the regressor and 1200 to test it. After selecting the value of λ , the complete training set was used to train the LLR.

In Figure 1 we show a plot of the FER estimation MSE as a function of the dimension of the feature vectors \mathbf{f}_i . ESM MSE is plot as a constant for comparison. We see that LLR-SS with SNR in dB outperforms ESM for number of features above 8. LLR-SS with SNR values in linear scale performs worse than ESM, and PCA offers a poor performance.

We performed similar experiments introducing practical impairments in the system. In all the cases, the division of the available samples into training, test and cross validation sets was the same as in the no-impairments case. In Figure 2 we show the results for different frame length. We generated 9000 realizations with the same channel and SNR distribution as before, Gaussian noise, but now varying the frame length between $L = 128$ and $L = 16386$. In this case, the estimation accuracy of ESM is hindered by the lack of information of the frame size. The inclusion of practical impairments reduces the accuracy of ESM in almost one order of magnitude. LLR with SNR in dB with L as a feature outperforms ESM, and the proposed semi-parametric approach (LLR-ESM(dB)) offers the best performance. The semi-parametric approach is built as a 2-feature LLR, with the ESM in dB, i.e., $10 \log_{10}(\gamma_{\text{eff}}(\gamma))$ with γ in natural units.

A similar behavior is shown for the generalized Gaussian distribution in Figure 3. We generated 12000 channel realizations with a constant frame length of $L = 1024$ but varying the ρ parameter of the Generalized gaussian distribution between 0.1 and 4. In this case, the feature vector contains the ρ value of the corresponding channel.

In Figure 4 we show the result of applying parametric and semi-parametric approaches to the problem of link adaptation with variable codeword length. A FER predictor was built for every MCS, and the MCS with a higher throughput meeting a FER constraint of $p_0 = 0.1$ was selected. The frame length was randomly selected between 128 and 13684 bits. The proposed semi-parametric approach offers up to 11% throughput gain for moderate SNR values. Also, it was observed that ESM did not meet the FER constraint in some cases due to the FER prediction inaccuracy. This results shows the advantage of taking into account practical impairments in link adaptation.

VI. CONCLUSIONS

In this paper we compared different approaches for FER prediction when practical impairments are present in the system. Machine learning FER estimators can incorporate the practical impairments as additional features, while traditional parametric approaches are less flexible. Classical ESM loses accuracy in the presence of different frame length or noise distribution. We proposed a semi-parametric approach that combines the good properties of both parametric and non-parametric methods. The results show the importance of incorporating practical impairments into the FER predictors by the use of non-parametric or semi-parametric methods.

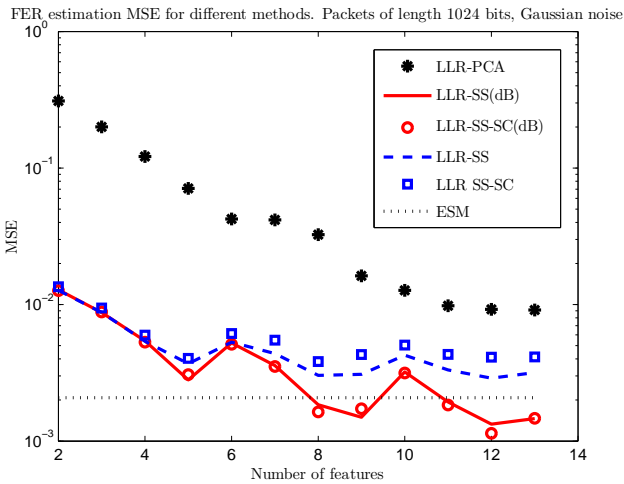


Fig. 1: FER prediction without practical impairments.

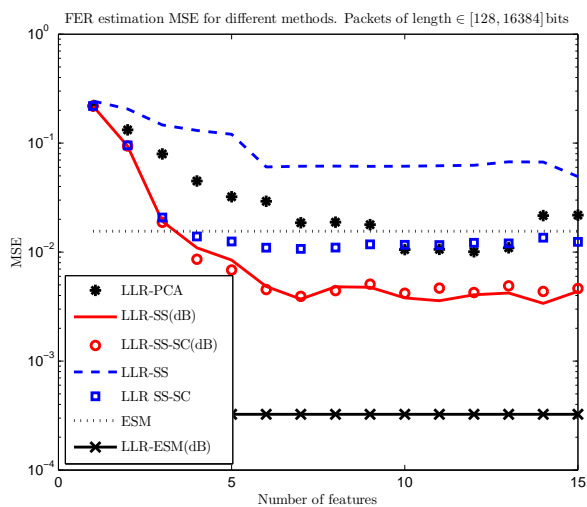


Fig. 2: FER prediction with variable frame length.

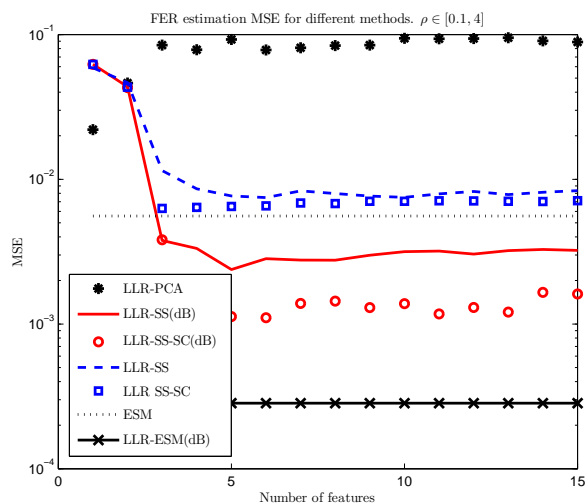


Fig. 3: FER prediction with different noise distribution.

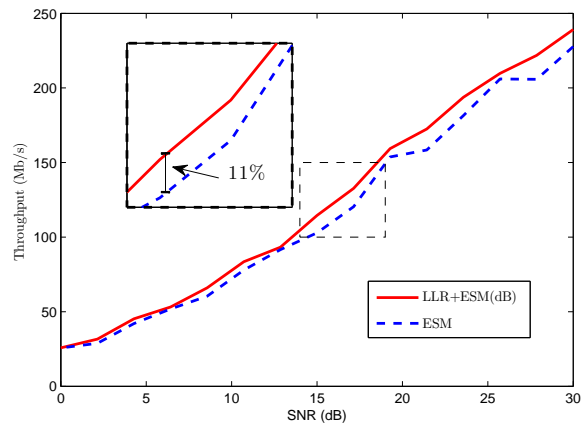


Fig. 4: Link adaptation throughput for variable frame length.

REFERENCES

- [1] M. Lampe, H. Rohling, and W. Zirwas, "Misunderstandings about link adaptation for frequency selective fading channels," in *Proc. IEEE PIMRC*, vol. 2, (Lisboa, Portugal), pp. 710 – 714 vol.2, Sept. 2002.
- [2] M. Lamarca and F. Rey, "Indicators for PER prediction in wireless systems: A comparative study," in *Proc. IEEE VTC Spring*, vol. 2, (Stockholm, Sweden), pp. 792–796 Vol. 2, May 2005.
- [3] Y. Blankenship, P. Sartori, B. Classon, V. Desai, and K. Baum, "Link error prediction methods for multicarrier systems," in *Proc. IEEE VTC Fall*, vol. 6, (Los Angeles, CA), pp. 4175–4179 Vol. 6, Sept. 2004.
- [4] "IEEE draft standard for IT - telecommunications and information exchange between systems - LAN/MAN - specific requirements - part 11: Wireless LAN medium access control and physical layer specifications - amd 4: Enhancements for very high throughput for operation in bands below 6ghz," *IEEE P802.11ac/D3.0*, June 2012, pp. 1–385, June 2012.
- [5] T. Jensen, S. Kant, J. Wehinger, and B. Fleury, "Mutual information metrics for fast link adaptation in IEEE 802.11n," in *Proc. IEEE ICC*, (Beijing, China), pp. 4910 –4915, May 2008.
- [6] H. Shao and N. Beaulieu, "Block coding for impulsive laplacian noise," in *Proc. IEEE ICC*, (Cape Town, South Africa), pp. 1–6, May 2010.
- [7] J. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-gaussian noise," *IEEE Trans. Inf. Theory*, vol. 18, no. 2, 1972.
- [8] R. Daniels, C. Caramanis, and R. W. Heath Jr., "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 114–126, Jan. 2010.
- [9] R. Daniels and R. W. Heath Jr., "Online adaptive modulation and coding with support vector machines," in *Proc. European Wireless Conference*, (Lucca, Italy), pp. 718–724, Apr. 2010.
- [10] S. Wahls and H. V. Poor, "Link adaptation for bicm-ofdm through adaptive kernel regression," in *Proc. IEEE ICASSP*, (Vancouver, BC), pp. 5136–5140, IEEE, 2013.
- [11] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.
- [12] R. Jain, C. So-In, and A.-k. Al Tamimi, "System-level modeling of IEEE 802.16E mobile wimax networks: Key issues," *IEEE Wireless Commun. Mag.*, vol. 15, pp. 73–79, Oct. 2008.
- [13] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, pp. 55 – 63, Jan. 1968.
- [14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2 ed., Feb. 2009.
- [15] A. Rico-Alvarino and R. W. Heath Jr., "Learning-based adaptive transmission for limited feedback multiuser MIMO-OFDM," *Submitted to IEEE Transactions on Wireless Communications*, June 2013. Available at <http://arxiv.org/abs/1306.0502>.