<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *21st IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*.

Citation for the original published paper:

Pellaco, L., Saxena, V., Bengtsson, M., Jaldén, J. (2020)
Wireless link adaptation with outdated CSI —a hybrid data-driven and model-based approach
In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-273703>

Wireless link adaptation with outdated CSI — a hybrid data-driven and model-based approach

Lissy Pellaco¹, Vidit Saxena^{1,2}, Mats Bengtsson¹, and Joakim Jaldén¹

¹KTH Royal Institute of Technology, Stockholm, Sweden

{pellaco, vidits, matben, jalden}@kth.se

²Ericsson Research, Stockholm, Sweden

Abstract—*Link adaptation* provides high spectral efficiency in wireless communications by selecting appropriate transmission parameters, e.g., the modulation and coding scheme (MCS), based on the instantaneous wireless channel. However, link adaptation suffers from impairments due to channel state information (CSI) feedback delay. In this paper, we extend the data-driven MCS selection scheme in our previous work to the case of outdated CSI, by assuming that CSI history is available to the system. We present two approaches that leverage the CSI history to optimally select the MCS for the current channel, i.e., i) an *end-to-end* (E2E) machine learning approach and ii) a *hybrid* data-driven and model-based approach. The E2E method uses the CSI history as input to a neural network for MCS selection. Conversely, the hybrid method uses a lower-dimensionality sufficient statistic for the instantaneous CSI, computed from the CSI history, as input to a neural network for MCS selection. We demonstrate that replacing the CSI history with the sufficient statistic comes without loss of generality. Moreover, by means of numerical experiments, we show that both approaches effectively compensate for the feedback delay. However, we advocate the hybrid approach as it comes with the additional benefits of i) a smaller neural network, which in turn requires a lower amount of data and training time, ii) improved explainability, and iii) better insights into optimization choices.

Index Terms—Link adaptation, MCS selection, channel prediction, artificial neural network, sufficient statistic.

I. INTRODUCTION

The wireless spectrum is a scarce resource. Hence, maximizing the spectral efficiency is of crucial importance. State-of-the-art cellular networks address spectral efficiency by dynamically adjusting the link transmission parameters, e.g., the modulation and coding scheme (MCS), in response to the wireless channel state. This adjustment of link transmission parameters is called *link adaptation*. The optimal link transmission parameters depend on the instantaneous wireless channel. Therefore, the transmitter relies on periodically reported *channel state information* (CSI) [1]. However, the optimal transmission parameters depend nontrivially on the CSI due to the data processing at the transmitter and receiver and due to channel impairments and stochasticity. In light of this, several link adaptation approaches have been proposed in the literature, including heuristics-based [2] and data-driven [3]–[5] techniques. Another major roadblock is the time delay between pilot transmission, CSI feedback reception at the transmitter, and data transmission. During this time, which

we will refer to as the *feedback delay*, the channel can vary substantially, nullifying the benefits of link adaptation [6], [7].

In this paper, we extend the data-driven MCS selection scheme in our previous work [4] to the case of outdated CSI. In [4], we trained a neural network to perform the MCS selection on the basis of the instantaneous CSI. By doing so, we demonstrated a performance gain of the neural network over the commonly used MCS selection based on effective SINR [8]. Conversely, in this paper, we assume that only the past CSI history is available to the system. We present two different approaches to leverage the CSI history for optimal MCS selection for the current channel, i.e., i) an *end-to-end* (E2E) machine learning approach, analogous to recently studied data-driven methods [9]–[11], and ii) a *hybrid* data-driven and model-based approach, which provides another early example of such hybrid methods in the physical layer [12]. In the E2E approach, the CSI history is directly fed to a neural network that is trained to select the optimal MCS for the current channel. In the hybrid approach, the CSI history is fed to a minimum mean square error (MMSE) estimator for predicting the instantaneous CSI. Then, the estimated CSI is fed to a neural network that is trained to select the optimal MCS for the current channel.

We compare the two approaches from a theoretical and a practical standpoint. From a theoretical standpoint, we prove that the two-part hybrid approach comes without loss of generality, as the output of the MMSE estimator is a sufficient statistic for the instantaneous CSI when the channel evolves as a Gaussian process. From a practical standpoint, by means of extensive simulation results, we show that both approaches are able to effectively compensate for the feedback delay. However, the hybrid approach i) has fewer network parameters, requiring a lower amount of data and training time to achieve a given performance, ii) is able to leverage the channel correlation better than the E2E approach for moderate delays, iii) mitigates the lack of explainability of pure data-driven techniques, and iv) provides better insights into optimization choices. Hence, we advocate the hybrid approach.

II. LINK ADAPTATION

A. System Model

We consider bit-interleaved coded modulation (BICM) or-
thogonal frequency division multiplexing (OFDM) links used

by fourth and fifth generation cellular networks [13]. The transmitter encodes and bit-interleaves T_k information bits into a frame according to an MCS index $k \in \{1, \dots, K\}$. The MCS index determines the number of bits in a frame, the code rate of an error protection code applied to the information bits, as well as the subsequent mapping to complex symbols. These complex symbols are loaded onto frequency-domain subcarriers to generate an OFDM signal that is transmitted over a noisy wireless channel and decoded at the receiver. The probability of successful/unsuccessful frame decoding depends on the CSI of the noisy channel at transmission time. The CSI consists of the vector of channel coefficients, here denoted by $\mathbf{h}$. For simplicity, in the remainder of the paper, we will use the terms CSI and channel interchangeably. We use a binary random variable E_k to represent the event of successful/unsuccessful frame decoding at the receiver with the choice of the k^{th} MCS. In particular, $E_k = 0$ if the frame is correctly decoded, while $E_k = 1$ if an error occurs. Hence, the probability of an error event with the choice of the k^{th} MCS can be expressed as a conditioned probability on the CSI at transmission time, i.e., $P_{E_k|\mathbf{H}}(E_k = 1|\mathbf{h})$.

B. Optimal MCS Selection

As previously stated, the probability of an error event, i.e., unsuccessful frame decoding, depends on the channel state at transmission time and on the selected MCS. If one could know the probability of an error event conditioned on the channel in effect at transmission time for all the available MCSs, then the selection of the optimal MCS would be straightforward. In our previous work [4], we used an artificial neural network to learn the mapping from the channel $\mathbf{h}$ to the conditional probability of an error event. In particular, the k^{th} output of the neural network corresponds to the conditional probability of error if the k^{th} MCS is adopted for transmission. In other words, we denote by $\rho_k = \rho_k(\mathbf{h}; \boldsymbol{\theta})$ the k^{th} output of a neural network with input $\mathbf{h}$ and with parameters $\boldsymbol{\theta}$ such that $\rho_k(\mathbf{h}; \boldsymbol{\theta}) = P_{E_k|\mathbf{H}}(E_k = 1|\mathbf{h})$ for $k = 1, 2, \dots, K$, where K is the number of available MCSs. It must be noted that, even though approximation theory says that neural networks can approximate any continuous function arbitrarily well [14], the k^{th} output $\rho_k(\mathbf{h}; \boldsymbol{\theta})$ of a neural network should be interpreted as an estimate of the conditional error probability. We train the neural network parameters using N input-output pairs, $\{(\mathbf{h}^n, e_k^n)\}_{n=1}^N$, where $\mathbf{h}^n$ is the instantaneous channel at the time of transmission of the n^{th} frame and $e_k^n \in \{0, 1\}$ is the observed successful/unsuccessful decoding of the n^{th} frame at the receiver when the k^{th} MCS is used for transmission. The cross-entropy

$$\sum_{k=1}^K \frac{1}{N} \left\{ \sum_{n=1}^N -e_k^n \ln \rho_k(\mathbf{h}^n; \boldsymbol{\theta}) - (1 - e_k^n) \ln(1 - \rho_k(\mathbf{h}^n; \boldsymbol{\theta})) \right\}$$

is adopted as the training cost function, where $\boldsymbol{\theta}$ denotes the trainable neural network parameters. In [4] and [15], it is shown that in the limit of infinite training samples, i.e., $N \rightarrow \infty$, the cross-entropy cost function above is minimized

by the true conditional error probability for $k = 1, \dots, K$. In other words, the k^{th} output of the neural network, ρ_k , can indeed be interpreted as the estimated conditional probability of an error event with the choice of the k^{th} MCS for transmission. Therefore, given the channel $\mathbf{h}$ and given the estimated conditional error probabilities for all the available MCSs, the MCS maximizing the expected spectral efficiency is given by

$$k^* = \underset{k}{\operatorname{argmax}} \left(\frac{T_k}{B} (1 - \rho_k) \right), \quad (1)$$

where T_k is the number of information bits sent on the channel with the k^{th} MCS and where B is the transmission bandwidth.

C. E2E Learning Approach

In [4], the transmitter is assumed to have perfect knowledge of the instantaneous channel for which the MCS needs to be selected. Such assumption is impractical because the feedback loop used to send the CSI to the transmitter introduces a delay. However, we can realistically assume that the transmitter stores some channel history, i.e., a collection of past channel realizations denoted by the vector $\boldsymbol{\psi}$. In this case, the transmitter could instead use the stored channel history to train an end-to-end (E2E) neural network that learns the mapping from $\boldsymbol{\psi}$ to the conditional error probability for all the K available MCSs, i.e., $P_{E_k|\boldsymbol{\Psi}}(E_k = 1|\boldsymbol{\psi})$. We simply use the cross-entropy

$$\sum_{k=1}^K \frac{1}{N} \left\{ \sum_{n=1}^N -e_k^n \ln \rho_k(\boldsymbol{\psi}^n; \boldsymbol{\theta}) - (1 - e_k^n) \ln(1 - \rho_k(\boldsymbol{\psi}^n; \boldsymbol{\theta})) \right\}$$

to train the network with input-output pairs $\{(\boldsymbol{\psi}^n, e_k^n)\}_{n=1}^N$, where $\boldsymbol{\psi}^n$ is the channel history prior to the transmission of the n^{th} frame. Intuitively, one can think of this neural network as internally performing the prediction of the channel that the transmitted frame will actually experience. However, considering that the input of this E2E approach consists of the channel history rather than of a single realization, the neural network has a larger parameter set to train. Such increased dimensionality results not only in the need for a larger training set and in a longer training time, but also in a more difficult-to-train neural network [16]. In addition, this approach suffers from a lack of explainability, typical of any E2E method: we do not have access to any intermediate output because the prediction of the channel and the estimate of the conditional error probabilities are strictly intertwined.

D. Hybrid Approach

To overcome the limitations of the E2E approach, we also consider an alternative hybrid data-driven and model-based approach, which does not require any increment in the dimensionality of the parameter set to train with respect to our previous work [4]. Moreover, it partially mitigates the lack of explainability of the E2E approach. The fundamental idea is to replace the channel history used as input to the neural network with a lower-dimensionality sufficient statistic [17] for the instantaneous channel computed from the channel history

Fig. 1. Overall hybrid approach. The past channel history ψ is used to optimally estimate, via Wiener filter, the channel in effect at transmission time. Then, the estimate $\hat{h}$ is mapped by the neural network to the corresponding set of conditional error probabilities, i.e., $\hat{\rho}_1, \dots, \hat{\rho}_K$, one for each MCS. The final step is the selection of the MCS which gives the maximum expected spectral efficiency. Each layer of the network is parameterized by a weight matrix $\mathbf{W}$ and a bias vector $\mathbf{b}$ such that the output of the i^{th} layer, i.e., $\alpha^{(i)}$, is given by $\alpha^{(i)} = \Phi(\mathbf{W}^{(i)} \alpha^{(i-1)} + \mathbf{b}^{(i)})$ where Φ is a non-linear activation function applied element-wise, where $\mathbf{W}^{(i)} \in \mathbb{R}^{D_i \times (D_{i-1})}$, where $\mathbf{b} \in \mathbb{R}^{D_i}$, and where D_i is the dimension of the i^{th} layer.

itself. This comes without loss of generality with respect to the E2E approach, as Section III demonstrates. In particular, given the assumption that the channel coefficients evolve as a Gaussian random process, we apply a linear MMSE estimator, i.e., the FIR Wiener filter, which is proven to be a sufficient statistic for the instantaneous channel [18]. The FIR Wiener estimate of $\mathbf{h}$ is given by

$$\hat{\mathbf{h}} = g(\psi) \triangleq R_{h\psi} R_{\psi\psi}^{-1} \psi, \quad (2)$$

where $R_{h\psi} = \mathbb{E}\{\mathbf{h}\psi^T\}$ and where $R_{\psi\psi} = \mathbb{E}\{\psi\psi^T\}$.

Bringing all the pieces together, in order to compensate for the feedback delay, we use the MMSE predictor to output an estimate of the instantaneous channel, i.e., $\hat{\mathbf{h}}$. Then, analogously to [4], we train a neural network to learn the mapping from $\hat{\mathbf{h}}$ to the conditional error probability, here denoted by $\hat{\rho}_k(\hat{\mathbf{h}}; \theta) = P_{E_k|\hat{\mathbf{H}}}(E_k = 1|\hat{\mathbf{h}})$, for all the available MCSs. For training, we use the cross-entropy

$$\sum_{k=1}^K \frac{1}{N} \left\{ \sum_{n=1}^N -e_k^n \ln \hat{\rho}_k(\hat{\mathbf{h}}^n; \theta) - (1 - e_k^n) \ln(1 - \hat{\rho}_k(\hat{\mathbf{h}}^n; \theta)) \right\}$$

as our cost function, given the training set $\{(\hat{\mathbf{h}}^n, e_k^n)\}_{n=1}^N$. Finally, the optimal MCS is selected according to (1). Fig. 1 schematizes the overall approach.

III. OPTIMALITY OF THE HYBRID APPROACH

In this section, we provide theoretical justification of the hybrid approach. In fact, it can be proven that we can optimally estimate the event error probability based only on a sufficient statistic of the instantaneous channel given the channel history.

We can describe the dependence of the probability of an error event on the channel with the following Markov chain $\Psi \rightarrow \mathbf{H} \rightarrow E_k$, where Ψ is the random vector representing the channel history, where $\mathbf{H}$ is the random vector representing the instantaneous channel, and where E_k is the random variable representing the successful/unsuccessful

frame decoding at the receiver, with the choice of the k^{th} MCS for transmission. The Markov chain above is based on the assumption that the probability of an error event depends on the channel in effect at transmission time, rather than on the past channel history. In the E2E approach, we train the neural network parameters θ by minimizing the cross-entropy cost function

$$\sum_{k=1}^K \frac{1}{N} \left\{ \sum_{n=1}^N -e_k^n \ln \rho_k(\psi^n; \theta) - (1 - e_k^n) \ln(1 - \rho_k(\psi^n; \theta)) \right\}.$$

In the limit $N \rightarrow \infty$, the cost function can be rewritten as

$$\sum_{k=1}^K \mathbb{E}\{-E_k \ln \rho_k(\Psi; \theta) - (1 - E_k) \ln(1 - \rho_k(\Psi; \theta))\},$$

where the expectation is taken over $P(E_k, \Psi)$. From [4] and [15], we know that, among all possible functions $\delta(\psi) = [\delta_1(\psi) \dots \delta_K(\psi)]^T$, one that minimizes the cross-entropy in the limit $N \rightarrow \infty$, i.e.,

$$\sum_{k=1}^K \mathbb{E}\{-E_k \ln \delta_k(\Psi) - (1 - E_k) \ln(1 - \delta_k(\Psi))\},$$

has $\delta_k(\psi) = P(E_k = 1|\Psi = \psi)$ as components for $k = 1, 2, \dots, K$. Hence, the k^{th} output of the neural network can be interpreted as the probability of an error event conditioned on the channel history, i.e., $P(E_k = 1|\Psi = \psi)$. In practice, the network output will only provide an approximation, given the inability of a finite-sized neural network to exactly describe a given function and the inability of the training procedure to yield globally optimal parameters.

It can be proven that each component of $\delta(\psi)$ can be written as a function of the sufficient statistic for $\mathbf{H}$ given Ψ . In particular, it can be noted that

$$\begin{aligned} P(E_k = 1|\Psi = \psi) &= \\ &= \int_{\mathbf{h}} P(E_k = 1|\Psi = \psi, \mathbf{h}) p(\mathbf{h}|\Psi = \psi) d\mathbf{h} \\ &= \int_{\mathbf{h}} P(E_k = 1|\mathbf{h}) p(\mathbf{h}|g(\psi)) d\mathbf{h} \\ &= \int_{\mathbf{h}} P(E_k = 1, \mathbf{h}|g(\psi)) d\mathbf{h} \\ &= P(E_k = 1|\hat{\mathbf{H}} = g(\psi)), \end{aligned}$$

where $P(E_k = 1|\Psi = \psi, \mathbf{h}) = P(E_k = 1|\mathbf{h})$ is due to the Markov chain assumption and where $p(\mathbf{h}|\Psi = \psi) = p(\mathbf{h}|g(\psi))$ is due to the assumption that $\hat{\mathbf{h}} = g(\psi)$ (see (2)) is a sufficient statistic for $\mathbf{H}$ given Ψ [18]. Therefore, as $P(E_k = 1|\Psi = \psi) = P(E_k = 1|\hat{\mathbf{H}} = g(\psi))$, we can without loss of generality estimate the event error probability based only on the optimally predicted instantaneous channel $\hat{\mathbf{h}}$, rather than on the past channel history ψ .

Moreover, it can be proven for the hybrid approach, following the same steps as in [4], that in the limit of $N \rightarrow \infty$ the k^{th} output of the network can be interpreted as the probability of error conditioned on the sufficient statistic $\hat{\mathbf{h}} = g(\psi)$, i.e., $P(E_k = 1|\hat{\mathbf{H}} = \hat{\mathbf{h}})$. Note that a similar argument has also

been used in [19] to relate hierarchical structures present in physics to neural network architectures.

IV. NUMERICAL RESULTS

We present numerical results obtained by applying the link adaptation schemes in Section II to a simulated LTE downlink operating over an industry-standard radio channel model [20, Table 5]. We work with the complex-valued channel in the frequency domain, where $M = 72$ denotes the number of OFDM subcarriers. We represent the channel as a real-valued vector in $\mathbb{R}^{2M}$ by stacking the real and the imaginary parts. We assume that the channel coefficients evolve as a Gaussian process and we assume they remain constant during each LTE subframe of 1 ms duration. We consider ψ to store the 10 most recent CSI values and we consider integer feedback delays Δ , equally spaced from 0 ms to 9 ms, coherently with LTE specifications [21]. Hence, $\psi^T(i) = [\mathbf{h}^T(i - \Delta), \dots, \mathbf{h}^T(i - \Delta - 9)]$, where i is the time index. To compute Wiener filter coefficients, we adopt the Jakes' scattering model [22], according to which the channel cross-correlation function is given by

$$R_{hh}(\tau) = 2J_0(2\pi f_d \tau), \quad (3)$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind and where f_d is the maximum doppler frequency of the channel. The latter is given by $f_d = \frac{f_c}{c}v$, where f_c is the carrier frequency, i.e., 2 GHz, where $c = 3 \cdot 10^8$ m/s is the speed of light, and where v is the relative velocity between the base station and the mobile user equipment. We consider seven relative velocities v of 30 km/h, 45 km/h, 60 km/h, 75 km/h, 90 km/h, 105 km/h, and 120 km/h. We consider 30 different signal-to-noise ratio (SNR) values, equally spaced from 0 dB to 29 dB. The base station can select $K = 29$ different MCSs from the LTE specifications, as listed in [21, Table 7.1.7.1-1]. We use open source Python libraries and Keras. To encourage reproducibility, the code used to reproduce the plots herein is available at [23].

A. Experimental scenarios

We apply the hybrid approach and the E2E learning approach under two different scenarios that serve as performance reference. Under *Scenario I*, we assume that the SNR and the relative velocity v are fixed and known, hence the operating conditions are compliant with the assumptions of our theoretical analysis (see Section III). Conversely, under *Scenario II*, we assume that the SNR and v are not known a priori. In this case, the SNR and v are random variables uniformly distributed over their (discrete) support used in the simulations, and thus are random variables not jointly Gaussian with the channel coefficients. Hence, our theoretical analysis does, strictly speaking, not hold anymore for the applied FIR Wiener filter. However, such a scenario is interesting i) for analyzing the performance of the hybrid approach under off-design operating conditions and ii) for comparing the hybrid approach with the performance of the E2E approach, which is not based on any Gaussian assumption and, in theory, is able to model non-linear impairments. Moreover, the results obtained

under the two scenarios correspond to boundary cases: in intermediate conditions, such as when estimates of SNR and/or v are available, the performance will presumably be between Scenario I and Scenario II.

B. Architecture and Training

Under Scenario I in the hybrid approach, we design a different Wiener filter for each combination of delay, SNR, and v , and we train a different neural network for each combination. We consider the full range of delays, SNRs of 5 dB, 15 dB, and 25 dB, and a relative velocity v of 60 km/h. Likewise, under Scenario I in the E2E approach, we train a different neural network for each combination of delay, SNR, and v mentioned above. Under Scenario II in the hybrid approach, for each delay, we design a single Wiener filter and train a single neural network. The Wiener filter is designed to work on the full range of SNR and v values, by averaging the cross-correlation function (3) over the considered ranges. The neural network is analogously trained on the entire range of SNR and v values. Under Scenario II in the E2E approach, we train different neural networks for each delay that we consider, each over the entire range of SNR and v values. All networks, under both scenarios, have the same architecture and dimensioning, except for the input layer. The networks consist of three fully connected *hidden layers* of 1024, 512, and 1024 neurons respectively, and rectified linear unit (ReLU) activation functions. The *output* layer of the networks has 29 neurons with sigmoid activation functions. In the hybrid approach, the *input layer* has $2M = 144$ neurons with ReLUs, while, in the E2E approach, it has 10 times more neurons, leading to a larger parameter set to train. We train the networks with an Adam optimizer with learning rate of 0.001 using cross-entropy as the cost function. Under Scenario I, the training and the test sets for each network consist of $2.1 \cdot 10^5$ and $4 \cdot 10^3$ channel realizations, respectively. Under Scenario II, the training and the test sets for each network consist of $2.1 \cdot 10^5$ and $8.4 \cdot 10^5$ channel realizations, respectively. Due to space limitation, for further details on training process and simulation parameters see [23].

C. Results and Discussion

We compare the spectral efficiency obtained with the hybrid approach, with the E2E approach and with a *delay-blind* approach. The last approach consists of the same link adaptation scheme adopted in the hybrid method, but without the channel prediction step, as in [4].

1) *Scenario I – Known SNR and v* : As can be seen in Fig. 2 (a), the hybrid and the E2E approaches outperform the delay-blind approach. Hence, disregarding the feedback delay results in degraded performance. Moreover, we observe that for moderate delays, i.e., up to around 4 ms, the hybrid approach is able to leverage the channel correlation and achieves a performance gain with respect to the E2E approach. Thus, the fewer network parameters of the hybrid approach result in easier training and a higher spectral efficiency in the mentioned range of delay. After around 4 ms, the spectral efficiencies of

Fig. 2. Spectral efficiency obtained with the hybrid approach, with the E2E approach, and with the delay-blind approach at different SNR values, with relative user velocity v of 60 km/h, under Scenario I (a) and Scenario II (b).

the three methods flatten out because the channel becomes nearly uncorrelated with the delayed CSI.

2) *Scenario II – Unknown SNR and v* : As for Scenario I, the hybrid and the E2E approaches outperform the delay-blind approach (see Fig. 2 (b)). For large delays, we observe a drop in performance with respect to Scenario I, especially for the hybrid approach, as the Wiener filter coefficients are obtained by averaging the cross-correlation function (3) on the full range of SNR and v . However, we can still observe the gain in performance of the hybrid method over the E2E method for short delays. This result is of interest if we consider that i) the hybrid approach is applied under off-design operating conditions and that ii) the E2E approach is in theory favoured to handle difficult-to-model conditions, like Scenario II.

V. CONCLUSION

We addressed the issue of outdated CSI for link adaptation by extending a data-driven method proposed in our previous work [4]. We presented and compared two different extensions, i.e., a data-driven E2E approach and a hybrid data-driven and model-based approach. We proved that using a sufficient statistic, as in the hybrid approach, rather than the channel history, as in the E2E approach, comes without loss of generality. We showed, by experimental results on a simulated LTE channel, that the hybrid and the E2E approaches reach similar spectral efficiency. However, for moderate delays, the hybrid approach

achieves a performance gain due to the lower-dimensionality sufficient statistic that reduces the neural network complexity. This, together with the higher explainability compared to pure data-driven methods, makes such hybrid methods attractive for applications in the physical layer.

ACKNOWLEDGEMENTS

The work was partially supported by the European Research Council project AGNOSTIC (742648) and by Wallenberg AI, Autonomous Systems and Software Program (WASP).

REFERENCES

- [1] A. Molisch, *Wireless Communications*. Wiley, 2010.
- [2] G. Ku and J. M. Walsh, "Resource allocation and link adaptation in LTE and LTE advanced: A tutorial," *IEEE communications surveys & tutorials*, vol. 17, no. 3, pp. 1605–1633, 2014.
- [3] R. C. Daniels, C. M. Caramanis, and R. W. Heath, "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering," *IEEE Transactions on vehicular Technology*, vol. 59, no. 1, pp. 114–126, 2009.
- [4] V. Saxena, J. Jaldén, M. Bengtsson, and H. Tullberg, "Deep learning for frame error probability prediction in BICM-OFDM systems," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 6658–6662.
- [5] V. Saxena, J. Jaldén, J. E. Gonzalez, M. Bengtsson, H. Tullberg, and I. Stoica, "Contextual multi-armed bandits for link adaptation in cellular networks," in *Proceedings of the 2019 Workshop on Network Meets AI & ML*. ACM, 2019, pp. 44–49.
- [6] J. Francis and N. B. Mehta, "Impact of feedback delays on EESM-based wideband link adaptation: Modeling and analysis," in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 3952–3957.
- [7] S. Valentin and T. Wild, "Studying the sum capacity of mobile multiuser diversity systems with feedback errors and delay," in *2010 IEEE 72nd Vehicular Technology Conference - Fall*, Sep. 2010, pp. 1–5.
- [8] Ericsson, "Effective SNR mapping for modelling frame error rates in multiple-state channels," 3GPP2-C30-20030429-010, Tech. Rep., 2003.
- [9] R. Liao, H. Wen, J. Wu, H. Song, F. Pan, and L. Dong, "The Rayleigh fading channel prediction via deep learning," *Wireless Communications and Mobile Computing*, 2018.
- [10] H. Ye, G. Y. Li, and B. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, Feb. 2018.
- [11] W. Jiang and H. D. Schotten, "Neural network-based channel prediction and its performance in multi-antenna systems," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Aug. 2018, pp. 1–6.
- [12] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 77–83, 2019.
- [13] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The next generation wireless access technology*. Academic Press, 2018.
- [14] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*, 1st ed. Cambridge, MA, USA: MIT Press, 1995.
- [15] M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate bayesian a posteriori probabilities," *Neural Computation*, 1991.
- [16] M. Verleysen, D. François, G. Simon, and V. Wertz, "On the effects of dimensionality on data analysis with neural networks," 06 2003.
- [17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.
- [18] J. Forney, G. David, "Shannon Meets Wiener II: On MMSE Estimation in Successive Decoding Schemes," in *42th Annual Allerton Conference on Communication, Control, and Computing*, 2004.
- [19] H. W. Lin, M. Tegmark, and D. Rolnick, "Why does deep and cheap learning work so well?" *Journal of Statistical Physics*, vol. 168, no. 6, pp. 1223–1247, Sep. 2017.
- [20] *ITU-R Recommendation M.1225*, ITU Std., 1997.
- [21] *3GPP TS 36.213*, Release 13, ETSI Std., 2016.
- [22] Y. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, no. 6, pp. 920–928, June 2003.
- [23] L. Pellaco and V. Saxena, "LinkAdaptationCSI," GitHub repository, <https://github.com/lpkg/LinkAdaptationCSI>, 2020.