

# Data-Driven Soft Demapping for Residual Impairments Channels

Elena-Iulia Dobre, *Student Member, IEEE*, and Lutz Lampe, *Senior Member, IEEE*

**Abstract**—Deep learning based solutions are being integrated into the physical and link layers of wireless networks. They often effect an improvement in transmission reliability and/or efficiency when there is a model or an algorithm deficit. In this letter, we propose a deep learning-aided soft demapper, consisting of a fully-connected deep neural network (DNN), to alleviate a channel model deficit. We apply it in microwave backhaul transmissions affected by impairments generated by the local oscillator and power amplifier. The proposed DNN soft demapper learns the best approximation for the log-likelihood ratios (LLRs). The learned LLRs show gains over model-based impairment-aware LLRs, as they capture the actual channel as observed through data. We implement weight pruning and periodical retraining to adapt to statistical changes and make our proposed approach fit for practical cost-aware applications.

**Index Terms**—Soft demapper, deep learning, log-likelihood ratios, residual phase noise, residual nonlinearity.

## I. INTRODUCTION

**B**IT-interleaved coded modulation (BICM) and large constellations are adopted to fulfil the requirements for higher rates in communication networks. The soft demapper of a BICM receiver processes the received symbols into log-likelihood ratios (LLRs) or soft bits, which are subsequently used by the soft-decision forward error correction (SD-FEC). Exact demapping entails the knowledge of the conditional probability density function (PDF) of the channel output given the input. In practice, LLRs are computed via an approximation of the optimal log-maximum a-posteriori (log-MAP) rule, namely the max-log-MAP rule.

Hardware components in commercial transceivers often generate nonideal waveforms which affect the transmission quality. Two sources of impairments are the local oscillator (LO) and the power amplifier (PA). Phase noise (PN) stems from the LOs, while amplitude and phase distortions result from operating the PA near the saturation regime. PA nonlinearity digital predistortion (DPD) and PN estimation followed by compensation are usually employed to counteract the effects of these transceiver impairments.

Residual impairments, often disregarded in demodulation, arise due to imperfect estimation and DPD of PN and PA nonlinearity, respectively. The demapper typically assumes a Gaussian distribution for the channel PDF, and omits nonlinear multiplicative terms or memory effects. It thus outputs LLRs that are mismatched to the actual channel, and in turn, the decoding process carried out by the SD-FEC is suboptimal.

Various methods were proposed to mitigate the channel mismatch in the LLRs. In [1], the authors scale the LLRs to improve the generalized mutual information. References [2],

[3] present closed-form expressions of the LLRs with residual PN. In [4], an amplitude and phase distortion aware demapper is designed for transmission under nonlinear PAs.

The above works correct the mismatched LLRs via model-based approaches. But the employed transition probabilities often cannot or do not describe the true channel. Learning from data can alleviate this channel model deficit. A well-trained neural network (NN) learns the input-output mapping of the demapper, and generates LLRs matched to the actual channel. In [5], an NN is employed to learn the functionality of the log-MAP rule demapper, thereby reducing its complexity. References [6] and [7] propose fully-connected (FC) deep NNs (DNNs) and recurrent NNs as demappers, respectively, for optical channels with nonlinear impairments. To combat the effects of PA nonlinearity, DNNs have been employed to learn the characteristics of PAs and DPD [8], [9].

In this letter, we propose a learned demapper architecture to replace model-based demappers. Our proposed solution builds upon the works presented in [5], [6]. Unlike the demapper in [5], or its model-based counterpart in [10], our method does not approximate the exact LLRs, since these are suboptimal in case of a channel model mismatch. As an application case, we examine a microwave backhaul transmission impaired by the combined residual PN and PA nonlinearity. The motivation for this is that microwave backhaul links typically experience an additive white Gaussian noise (AWGN) channel at a signal-to-noise ratio (SNR) that is sufficiently high to support modulations such as 1024-quadrature amplitude modulation (QAM) or larger. For such a scenario, LO and PA nonidealities are the main cause for performance limitations. We assume the use of a practical but suboptimal DPD at the transmitter, while at the receiver, we employ PN estimation using pilot symbol assisted modulation (PSAM) and phase interpolation between pilots. Following PN compensation, samples are affected by residual PN and residual distortion from the nonlinear PA. We use the residual PN aware demapper from [3], as well as a demapper designed for an ideal AWGN channel, as benchmarks. We note the learned demapper does not replace preceding signal processing blocks that mitigate the propagation channel interference, but it is being used in tandem with them.

The model-based demapper in [3] adjusts to varying PN and additive noise variances, which we refer to as channel statistics, through estimating them. We train the learned demapper offline, across a wide range of channels with different statistics, to make it effective for a range of environments, and prune it for lower complexity. During transmission, we fine-tune only a small part of the learned demapper periodically, to adapt its parameters to varying statistics with small training overhead.

We summarize below the key contributions of our work.

Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada. Email: {elenad, lampe}@ece.ubc.ca

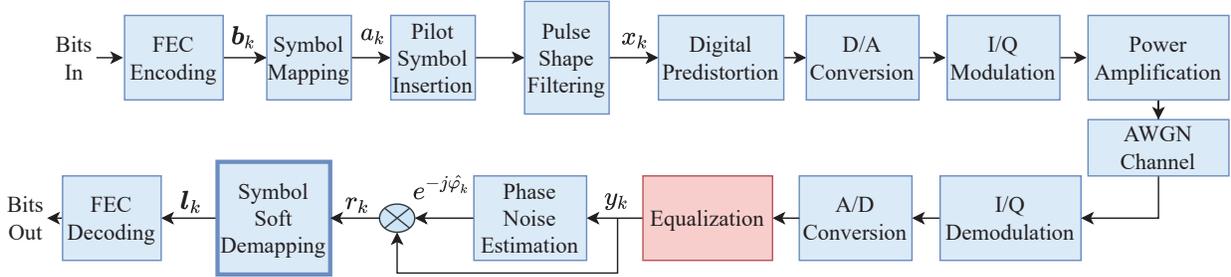


Fig. 1. Microwave backhaul transmission link. DPD and PN estimation are applied to mitigate PA nonlinearity and LO-induced PN, respectively. The red-shaded equalizer block is only used in the case of SSF-shaped transmission [11].

- 1) We design a learned demapper to overcome the performance limitation due to residual PN and PA nonlinearity in microwave backhaul links. We use relevant model-based benchmark methods for performance comparison.
- 2) We propose a novel procedure to tailor the demapper for practical use. We train it offline on a statistically diverse dataset, prune it up to a similar complexity of simple demappers, and adapt a minimal set of weights online.

This letter continues with the system model and benchmark demappers in Section II. In Section III, we introduce the proposed learned demapper, followed by simulation results in Section IV. We wrap up with conclusions in Section V.

## II. SYSTEM MODEL

The transmission link block diagram is shown in Fig. 1. A sequence of independent and uniformly distributed bits is FEC encoded. Each  $m$ -dimensional vector of coded bits  $\mathbf{b}_k = [b_{k,1}, \dots, b_{k,m}]^T$  is mapped to an  $M$ -QAM symbol  $a_k$ , where  $m = \log_2(M)$ . Equally-spaced pilots used for PN estimation are inserted, followed by pulse shape filtering. The pulse shaping filter can be either a Nyquist pulse, such as a root-raised cosine (RRC), or a non-Nyquist pulse, as for example a spectrum skirt-filling (SSF) pulse introduced in [11] for microwave systems. We purposefully consider the latter case, as it allows us to highlight the effect of model mismatch in the results section. The signal path continues with digital-to-analog (D/A) conversion and in-phase/quadrature (I/Q) modulation, followed by propagation through the AWGN channel, I/Q demodulation and analog-to-digital (A/D) conversion.

The LOs of the I/Q modulation/demodulation generate non-ideal carrier signals of the form  $f_{LO}(t) = \exp(j2\pi f_c t + \varphi(t))$ , where  $\varphi(t)$  is the PN random process. PN estimation consists of an interpolation of the PN estimates obtained from each two consecutive pilots. This simple PN estimation method is applied in practical microwave transmission systems.

The estimated PN  $\hat{\varphi}_k$  is applied to the A/D converter or, in the case of SSF transmission, the equalizer output sample  $y_k$ , respectively (see Fig. 1). The PN compensated sample  $r_k = y_k e^{-j\hat{\varphi}_k}$  is used by the demapper to produce the LLRs. The signal  $r_k$  is impaired by residual PN due to imperfect PN estimation and residual PA nonlinearity due to imperfect DPD. For the presentation of results in Section IV, we emulate the combined effect of DPD and nonlinear PA through a residual PA nonlinearity model. The SSF-shaped transmission

experiences equalization-enhanced PN [12], whose effect will be discussed in the numerical results section.

The demapper computes the LLRs  $\mathbf{l}_k$  associated to the bits of the transmitted symbol  $a_k$ , where  $\mathbf{l}_k = [l_{k,1}, \dots, l_{k,m}]^T$ . The LLR of the  $i^{\text{th}}$  bit of  $a_k$  is defined as

$$l_{k,i} \triangleq \log(p_{R|B_i}(r_k|1)) - \log(p_{R|B_i}(r_k|0)) \quad (1)$$

where  $p_{R|B_i}(r|b_i)$  denotes the PDF of the observed output  $r_k$  for the input bit  $b_{k,i} \in \{0, 1\}$ . An exact computation requires knowledge of the PDF  $p_{R|A}(r|a)$  of the effective channel from  $a_k$  to  $r_k$ .

### A. AWGN Channel Soft Demapper

A commonly used demapper assumes an ideal AWGN channel with output  $R = A + N$ , where  $N$  is complex zero-mean circularly symmetric Gaussian noise with variance  $\sigma_n^2$ . The corresponding max-log LLR approximation

$$l_{k,i} \approx \frac{1}{\sigma_n^2} \left( \min_{j \in \mathcal{I}_i^0} (\|r_k - a_j\|^2) - \min_{j \in \mathcal{I}_i^1} (\|r_k - a_j\|^2) \right), \quad (2)$$

where  $\mathcal{I}_i^b$  is the set of indices of the QAM symbols with the  $i^{\text{th}}$  bit equal to  $b$ , builds on this channel model.

### B. Channel-Aware Soft Demapper

While the demapper in (2) is mismatched to a channel with residual impairments, the likelihood  $p_{R|A}(r|a)$  used in a channel-aware demapper is often difficult to model. The likelihood can, however, be approximated for simpler scenarios. For example, for an AWGN channel and residual PN with independent Gaussian distributed samples, the likelihood  $p_{R|A}(r|a)$  is found in [2]. Reference [3] derives  $p_{R|A}(r|a)$  for the case where the LO-induced PN follows a Wiener process and PN estimation is obtained through PSAM with linear interpolation. However, the LLRs computed with this  $p_{R|A}(r|a)$  are still mismatched to the true channel in the presence of distortion that does not match the model assumptions.

## III. DEEP NEURAL NETWORK-BASED SOFT DEMAPPER

We now propose a feed-forward FC DNN as an adaptive learned demapper. Our solution mitigates a model deficit when the effective channel including residual impairments cannot be well approximated and tracked through analytical expressions.

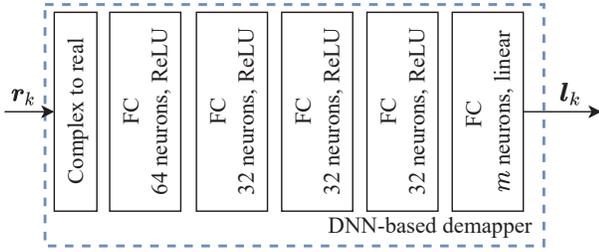


Fig. 2. Structure of the DNN-based soft demapper.

### A. Network Structure and Training Procedure

The proposed DNN structure is shown in Fig. 2. The features are the real and imaginary parts of the symbols stacked in  $\mathbf{r}_k = [r_{k-L_c}, \dots, r_{k+L_c}]^T$ , where  $L_c$  is an estimate of the channel memory length. The last layer outputs the LLRs  $l_k$  of the transmitted symbol  $a_k$ . Predicting the LLRs is a regression problem in which we train the DNN in a supervised manner. We note that

$$\Pr(b_{k,i} = 1 | r_k) = \sigma(l_{k,i}), \quad (3)$$

which follows from (1) for uniformly distributed data bits, and  $\sigma$  denotes the logistic function. The trainable parameters  $\theta$  of the DNN are optimized via stochastic gradient descent (SGD) with the binary cross-entropy loss

$$\mathcal{L}(\theta) = -\frac{1}{K} \sum_{k=1}^K \sum_{i=1}^m b_{k,i} \log(\sigma(l_{k,i})) + (1 - b_{k,i}) \log(1 - \sigma(l_{k,i})), \quad (4)$$

where  $K$  is the mini-batch size of the SGD algorithm.

### B. Pruning and Online Training

The statistical description of the effective channel between transmitted data  $a_k$  and demapper input  $r_k$  will not be known perfectly at the time of training of the DNN, which we assume is performed before deployment in a transmission link. Furthermore, the channel statistics can be time-varying, as for example the link SNR may change. One solution to this problem would be to train a bank of DNNs and switch between them at the time of operation, based on a performance indicator.

We choose a more compact approach. First, we train one base DNN across multiple channel realizations, each with different AWGN variances and PN levels set by sweeping through value ranges which are typical for microwave backhaul links. Second, we reduce the computational complexity of demapping associated with the trained base DNN by gradually pruning its low-magnitude weights until a desired sparsity level, defined as  $1 - \frac{\text{pruned DNN size}}{\text{original DNN size}}$ , is reached. Third, we deploy the pruned base DNN for transmission and set its output layer to be trainable while in use. This online update allows us to adjust the learned demapper to the specific channel statistics including impairments not present during offline training and/or varying over time. The choice of making the output layer's parameters trainable is motivated by the fact that it typically learns specialized features of the training dataset, and that complexity for training is low. The online supervised training is done periodically using pilots. Fig. 3 illustrates (a)

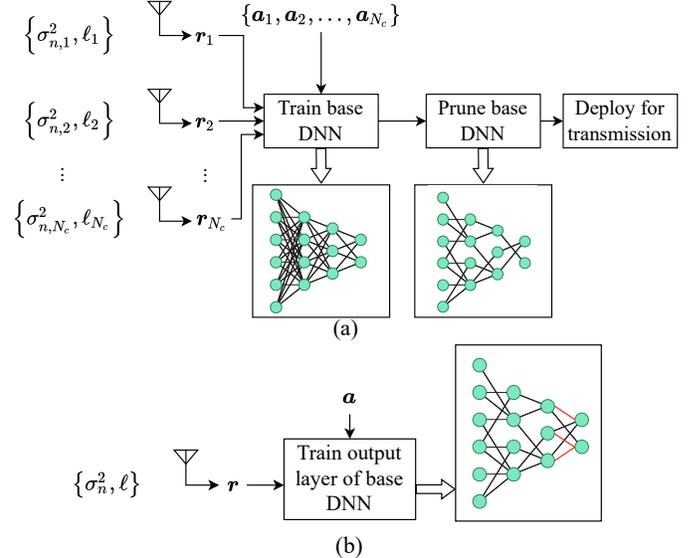


Fig. 3. (a) Offline training of the base DNN across  $N_c$  channel instances with different statistics  $\{\sigma_{n,i}^2, \ell_i\}$ , where  $\sigma_{n,i}^2$  and  $\ell_i$  denote the AWGN variance and PN level of the  $i^{\text{th}}$  channel, respectively, followed by weight pruning. (b) Periodic online training of the output layer. We allow new connections (red lines) to emerge in the output layer.

the offline training and pruning of the base DNN, and (b) the online training via periodic updates of its output layer.

## IV. SIMULATION RESULTS AND DISCUSSION

In this section, we demonstrate the performance benefits of the proposed data-driven soft demapping by simulating microwave backhaul transmission with hardware impairments.

### A. Transmission Setup and DNN Demapper

We consider transmission with a 1024-QAM constellation and RRC/SSF pulse shaping with the practical filter parameters as given in [11]. The QAM order and shaping filter parameters are fixed throughout training and inference. To mimic the LO-induced PN, we adopt the commonly used Wiener model, e.g. [3], [11]. Accordingly,  $\varphi_k = \varphi_{k-1} + \delta_\varphi$ ,  $\delta_\varphi \sim \mathcal{N}(0, \sigma_\varphi^2)$  and  $\sigma_\varphi^2 = 10^{\ell/10} 4\pi^2 f_{\text{offset}}^2 T_s$ , where  $f_{\text{offset}}$  and  $T_s$  denote the frequency relative to the carrier at which  $\ell$  is measured and the symbol period, respectively. For the effect of the PA nonlinearity when applying DPD we consider the residual nonlinearity function  $g_{\text{PA}}(|x|) = \frac{\alpha_{\text{AMAM}}|x|}{1 + \beta_{\text{AMAM}}|x|^2} \exp(j \frac{\alpha_{\text{AMPM}}|x|^2}{1 + \beta_{\text{AMPM}}|x|^2})$ , which we refer to as the residual Saleh model (RSM). The parameters  $\alpha_{(\cdot)}$  and  $\beta_{(\cdot)}$  control the memoryless nonlinearity strength, and  $g_{\text{PA}}$  is applied on the pulse shaped signal. As a second model, we consider the soft amplitude limiter (SAL)  $g_{\text{PA}}(x) = x$  for  $|x| \leq A$ , and  $g_{\text{PA}}(x) = A \exp(j\angle x)$  for  $|x| > A$ , where  $A$  denotes the clipping level [13].

The first four layers of the DNN-based soft demapper (see Fig. 2) have 64, 32, 32, 32 units, respectively, with a rectified linear unit (ReLU) activation. The output layer has  $m$  units, each for one LLR, with linear activation. We use  $L_c = 0$  for the experiments, given the memoryless residual impairments. The DNN is trained by minimizing the loss in (4) using the Adam optimizer with  $K = 64$  symbols, cyclic learning rates

TABLE I  
TRAINING PROCEDURE HYPERPARAMETERS.

	Stage 1	Stage 2	Stage 3
Learning rate	$10^{-3}$	-	-
Number of epochs	300	300	300
Cyclic learning rate base, max, step size	-	$10^{-4}, 9 \times 10^{-4}$ ,	$10^{-5}, 9 \times 10^{-5}$ ,
Early stopping pa- tience, $\Delta_{\min}$	-	50, $10^{-3}$	50, $10^{-3}$

and early stopping. Training is carried out in three stages with hyperparameters outlined in Table I. We use training datasets of 129,600 and 324,000 QAM symbols, of which 25% is reserved for validation, to initially train the base DNN and train it while pruning, respectively.

### B. BER for Independently Trained Models

In our first set of results, we compare the bit-error rate (BER) performances when using the proposed learned and the non-learned benchmark demappers. For this, a rate-2/3 low-density parity check (LDPC) code according to the DVB-S2 standard is used for FEC. The PN level is set to  $\ell = -90$  dBc/Hz at  $f_{\text{offset}} = 100$  kHz. The symbol rates of the RRC and SSF transmissions are  $R_{\text{RRC}} = 25.6$  Msymbol/s and  $R_{\text{SSF}} = 51.2$  Msymbol/s, leading to PN increments variances of  $\sigma_{\varphi, \text{RRC}}^2 = 1.5 \times 10^{-5}$  rad<sup>2</sup> and  $\sigma_{\varphi, \text{SSF}}^2 = 3.87 \times 10^{-6}$  rad<sup>2</sup>, respectively. The parameters for the RSM are  $\alpha_{\text{AMAM}} = 1.05$ ,  $\beta_{\text{AMAM}} = 0.05$ ,  $\alpha_{\text{AMPM}} = 0.2$ ,  $\beta_{\text{AMPM}} = 0.2$ . These parameters indicate a milder PA nonlinearity compared to a realistic non-predistorted one with, for example,  $\alpha_{\text{AMAM}} = \alpha_{\text{AMPM}} = 2$ ,  $\beta_{\text{AMAM}} = \beta_{\text{AMPM}} = 1$  [4]. The clipping level for the SAL is  $A = 0.8$ .

We assume the idealized scenario that the channel statistics do not change between training and testing. Accordingly, we train the DNN in Fig. 2 independently for each SNR, yielding different trained models. We also assume that the non-learned demappers know the SNR and the residual PN level perfectly. This scenario allows us to highlight the potential of the DNN-based demapper to overcome model deficits of the non-learned solutions. Fig. 4 shows the BERs as functions of SNR when using mismatched LLRs from (2), LLRs corrected for residual PN according to [3], and LLRs from the proposed DNN. The BER for the AWGN unimpaired channel is shown as reference.

First, we consider the results for RRC pulse shape. We observe that the LLRs corrected for residual PN do not improve the BERs compared to those for the mismatched LLRs from (2). For the RSM, the model mismatch is even compounded by trying to address the PN impairment but ignoring PA nonlinearity. The proposed DNN-based demapper on the other hand achieves BERs fairly close to those of the unimpaired AWGN channel in both cases. For the subsequent results, we use the RSM when considering residual PA nonlinearity.

For the case of SSF-shaped transmission, we turn-off the PA nonlinearity and only PN is active. One would expect that the PN-aware demapper from [3] is now an ideal match for the effective channel. However, due to the presence of the equalizer and the resulting equalizer-enhanced PN, a model

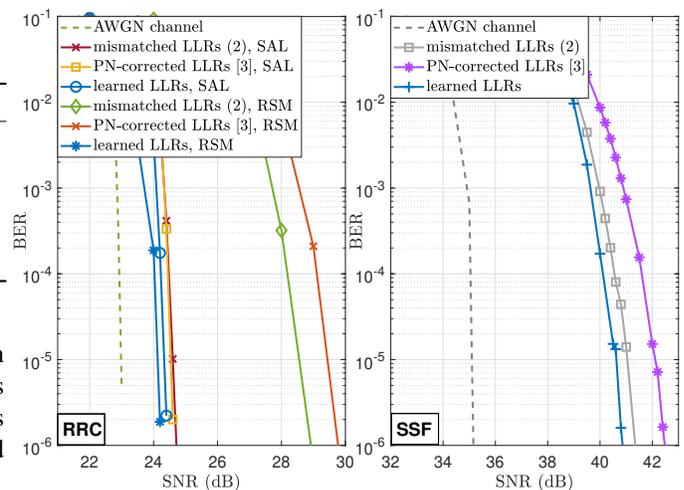


Fig. 4. Coded BER for different demappers and impairments. Left: RRC transmission, PN and PA nonlinearity. Right: SSF transmission, only PN.

mismatch occurs and the demapper with corrected LLRs again results in a BER which is worse than for max-log LLRs (2). The learned demapper again provides the best BER performance, albeit with a gap to transmission without impairments. We attribute the latter to the increased sensitivity of SSF-shaped transmission to PN as parts of the data are transmitted in spectrum skirts [11].

### C. AIR for Pruned Base Models with Online Training

In our second set of results, we demonstrate the effectiveness of the proposed DNN architecture with a pruned base DNN and online training of the output layer weights as introduced in Section III-B and illustrated in Fig. 3. The base DNN is trained on data for SNR and PN levels ranging from 21 dB to 35 dB, and from  $-94$  dBc/Hz to  $-85$  dBc/Hz, respectively, and pruned to 20% (sparsity level = 80%) of its original size. We assume that 7,000 samples are available to periodically retrain only the output layer. The latter is not a critical assumption as statistics of the microwave backhaul channel change very slowly compared to the data rate. For the non-learned demappers, we again assume the SNR is known for the mismatched LLRs (2), and the SNR and residual PN level are known for the PN-corrected LLRs [3].

We consider the RRC pulse-shaped transmission with residual PN and PA nonlinearity (RSM) impairments, where the statistics for the latter are fixed as specified for the previous results set. The achievable information rate (AIR) [1, Eq. (3)] is adopted as a compact measure for the quality of the produced LLRs. Fig. 5 (top) shows the AIRs for different demappers as functions of time, where SNR (middle) and PN level (bottom) are varying over time. The duration of time intervals between which SNR and PN level change is identified as  $\delta_t$ . In case of learned demappers, we include the result for the pruned base DNN with and without online training and for the idealized case of the full-sized DNN that is trained for the specific SNR and PN level. The latter constitutes the optimal performance benchmark. We observe that the DNN without retraining experiences a performance degradation compared to the optimal benchmark for some (SNR, PN-level) pairs. For

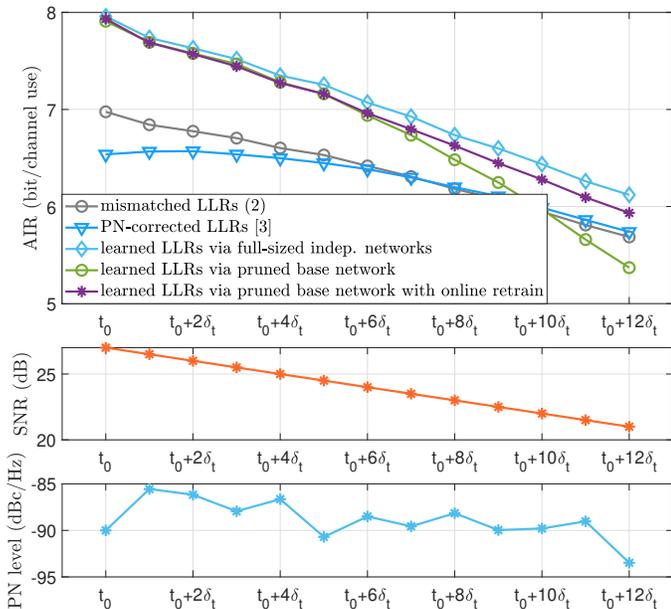


Fig. 5. Top: AIR for RRC transmission over residual impairments channels with time-varying statistics. Middle: Variation of SNR. Bottom: Variation of PN level.

low SNR, it performs even worse than when using mismatched LLRs. The proposed online learning however is able to follow closely the AIR curve for known SNRs and PN levels over the entire parameter range. In particular, the performance benefit over the non-learned solutions is retained.

A performance versus complexity plot is depicted in Fig. 6. Its purpose is two-fold: to compare the cost of the mismatched and learned demappers, and to show the benefit of pruning the base DNN and online retraining of the output layer. We present the AIR measured for two (SNR, PN-level) settings from Fig. 5, versus the number of real operations per received symbol. As it can be seen, using a single base DNN trained across a range of channel statistics and pruned to a size of just 20% of the full-size DNN may cause a performance loss. However, online retraining recovers the performance benefits of learned demapping with only a slight increase in the number of real operations. Overall, the proposed combination of pruning together with online retraining achieves the performance of the idealized benchmark case of learning with known channel statistics at a complexity that is in between the mismatched demapper and its max-log approximation. This renders it an attractive solution for practical use.

## V. CONCLUSIONS

We have studied data-driven soft demapping for channels with residual PN and PA nonlinearity impairments. The use case that motivated our work is microwave backhaul transmission, whose performance is limited by those impairments. We have proposed an FC DNN architecture that exhibits moderate complexity and adaptability to channel statistics due to the interplay of weight pruning and online retraining. This has been demonstrated through numerical results that show the performance gains over non-learned demappers including those designed for channels with residual PN.

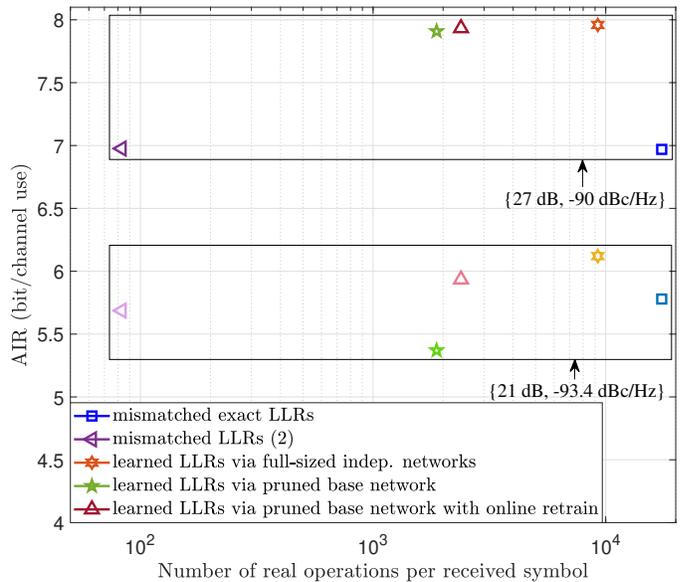


Fig. 6. AIR (from Fig. 5) versus the number of real operations per data symbol for two (SNR, PN-level) pairs.

## REFERENCES

- [1] A. Alvarado, L. Szczecinski, T. Fehenberger, M. Paskov, and P. Bayvel, "Improved soft-decision forward error correction via post-processing of mismatched log-likelihood ratios," in *Eur. Conf. Opt. Commun.*, 2016.
- [2] T. Koike-Akino, D. S. Millar, K. Kojima, and K. Parsons, "Phase noise-robust LLR calculation with linear/bilinear transform for LDPC-coded coherent communications," in *Conf. on Lasers and Electro-Optics*, 2015.
- [3] P. Neshastegaran and A. H. Banihashemi, "Log-likelihood ratio calculation for pilot symbol assisted coded modulation schemes with residual phase noise," *IEEE Trans. Commun.*, no. 5, pp. 3782–3790, 2019.
- [4] Z. Gülgün and A. Ö. Yılmaz, "Detection schemes for high order  $M$ -ary QAM under transmit nonlinearities," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4825–4834, 2019.
- [5] O. Shental and J. Hoydis, "Machine LLRning: Learning to softly demodulate," in *IEEE Globecom Workshops*, 2019.
- [6] M. Schädler, G. Böcherer, and S. Pächnicke, "Soft-demapping for short reach optical communication: A comparison of deep neural networks and Volterra series," *J. Lightwave Technol.*, vol. 39, no. 10, pp. 3095–3105, 2021.
- [7] M. Schädler *et al.*, "Recurrent neural network soft-demapping for nonlinear ISI in 800Gbit/s DWDM coherent optical transmissions," *J. Lightwave Technol.*, vol. 39, no. 16, pp. 5278–5286, 2021.
- [8] P.-Y. Chen *et al.*, "Learning to compensate: A deep neural network framework for 5G power amplifier compensation," in *IEEE Int. Conf. Commun. (ICC)*, 2021.
- [9] Y. Wu, U. Gustavsson, A. G. I. Amat, and H. Wymeersch, "Low complexity joint impairment mitigation of IQ modulator and PA using neural networks," *IEEE J. Select. Areas Commun.*, vol. 40, no. 1, pp. 54–64, 2021.
- [10] Y. Yao, Y. Su, J. Shi, and J. Lin, "A low-complexity soft QAM demapper based on first-order linear approximation," in *IEEE Intl. Symp. Pers., Indoor, and Mobile Radio Commun. (PIMRC)*, 2015, pp. 446–450.
- [11] E.-I. Dobre, A. Mostafa, L. Lampe, H. Shahmohammadian, and M. Jian, "Efficient spectrum utilization via pulse shape design for fixed transmission networks," *IEEE Trans. Commun.*, no. 3, pp. 1510–1523, 2020.
- [12] W. Shieh and K.-P. Ho, "Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing," *Opt. Exp.*, vol. 16, no. 20, pp. 15 718–15 727, 2008.
- [13] P. Händel and D. Rönnow, "Dirty MIMO transmitters: Does it matter?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5425–5436, 2018.