Optimal Neuro-Fuzzy Equalizers for Nonlinear Channels of the Perpendicular Magnetic Recording System

Rati Wongsathan[1†], Non-member and **Pornchai Supnithi**[2], Member

ABSTRACT

Nonlinear distortions caused by partial erasure and nonlinear transition shifts interacting with inter-symbol interference, are a major hindrance to data storage systems, since they degrade detector performance. This work aims to design and optimize the neuro-fuzzy equalizer (NFE) using the multi-objective genetic algorithm (MOGA) to detect nonlinear high-density magnetic recording (MR) channels. Through the GA-assisted back-propagation algorithm and least mean square optimization, the complexity in terms of decision rules is reduced by 25% and significantly provides 65% lower signal processing computation. When applied to the perpendicular (MR) system, the proposed NFE outperforms existing equalizers such as the neural network-based equalizer, fuzzy logic equalizer, and conventional NFE for the Volterra and jitter media noise channels using 1–3 dB and 1.5–3.5 dB signal-to-noise ratio gains at the bit-error-rate of 10^{-4}, respectively. Furthermore, compared to the other models, the NFE provides a more effective output mean square error performance for retrieving the original bit data.

Keywords: Equalizer, Fuzzy Logic, Magnetic Recording Channel, Neural Network, Neuro-Fuzzy

1. INTRODUCTION

The magneto-resistive (M-R) heads, among others, provide a high signal-to-noise ratio (SNR) through the conversion of electromagnetically stored data in disk media to resistance, thereby introducing nonlinear distortions (NLD) into the read-channel signal path. They interact with the inter-symbol interference (ISI) from adjacent stored bits causing nonlinear ISI in the high magnetic recording (MR) density channels. To mitigate these nonlinearities, nonlinear equalizers have been proposed under the presence of nonlinear amplitude distortion (NLAD) [1]. Based on the Volterra channel of the perpendicular MR (PMR) system, the neuro-fuzzy equalizer (NFE) slightly outperforms the neural network equalizer (NNE) and the fuzzy logic equalizer (FLE). However, due to the conventional design (i.e., human tuning of the parameters), the NFE may not achieve optimal performance and has the disadvantage of being computationally complex.

Besides the NLAD causing partial erasure due to bit percolation of the readback signal, the nonlinear transition shift (NLTS) [2] created by the demagnetizing field of the previously written transitions affects bit location during the writing process. The severity of the NLTS depends on the data patterns and head-media parameters. On the transmission side, the NLTS can be reduced by the writing of pre-compensation techniques [3], such as dibit and two-level pre-compensations, which require the optimized offsets to align with the timing of written transitions. However, it is difficult to determine the number of NLTSs prior to writing the transition bits. Moreover, there is no NLTS measurement device on the user side. On the other hand, to retrieve the distorted signals on the receiving side, the NLDs, consisting of NLAD and NLTS, are mitigated by the equalization technique.

Until now, various equalization and detection techniques have been implemented to mitigate the NLDs in magnetic read channels, ranging from the conventional maximum likelihood sequence estimation (MLSE) to the symbol-based equalization. The MLSE, based on partial response maximum likelihood (PRML) equalization [4], has been implemented using linear and nonlinear Viterbi detectors (VD) where the trellis is designed according to the channel target coefficients. To improve the performance of PRML/VD, the detector incorporates the pattern-dependent noise predictor (PDNP) and mean-adjusted PDNP (MA-PDNP) into the filter [5]. Moreover, the VD is modified by including an offset and the mean of PDN into the branch metric calculation. However, by using VD, the computational complexity grows exponentially with the length of nonlinear channel taps, which is also

Manuscript received on July 15, 2020 ; revised on January 23, 2021 ; accepted on February 23, 2021. This paper was recommended by Associate Editor Piya Kovintavewat.

[1]The author is with the North-Chiang Mai University, Chiang Mai 50230, Thailand.

[2]The author is with the School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.

[†]Corresponding author: rati@northcm.ac.th

Digital Object Identifier 10.37936/ecti-eec.2021192.241449

a major disadvantage in practice. In addition, when prior channel information is not provided, the symbol-by-symbol detection [6] outperforms sequence detection when applied to nonlinear PMR channels.

The symbol detector can be divided into linear and nonlinear processing depending on the implemented circuit. The least mean square (LMS) and recursive least square (RLS) linear adaptive filters [7] improve the BER performance of the linear channels; however, they deteriorate when applied to the nonlinear channels. Meanwhile, the nonlinear symbol detectors based on nonlinear classification decisions, such as Volterra equalizer (VE) [8], neural network equalizer (NNE) [9], hybrid VE-NNE [8], fuzzy logic equalizer (FLE) [10], and NFE [11] have gained more attention in various nonlinear channels, including MR channels, due to the significant BER improvement.

Among them, various NNEs have been successful in detecting noisy signals. However, their performances are slightly improved when compared to the increase in complexity. Since NN models are inspired by biological NNs, there is no solid answer for determining the proper structure, or the most suitable training parameters to be used. On the other hand, the FLE is based on many decision rules, which accounts for the high dynamic changes in the readback signals, providing better performance than that produced from only one rule of the NNE. However, the use of the FLE is limited due to it having no closed-form expression for determining the optimal rules and selected parameters. Many works adjust and adapt them using expert experience and trial and error, which may not achieve optimal performance. Optimizing the fuzzy rules while fine-tuning the FLE parameters automatically using the genetic algorithm (GA) can overcome the problem [10]. However, the drawbacks are slow convergence and difficulty in choosing optimal GA parameters. Besides, particle swarm optimization (PSO) has been employed to design the fuzzy rules [11]. Alternatively, an NFE [1] that derives the fuzzy rules using the NN to learn the input-output patterns outperforms the single models of both NNE and FLE, at the expense of complexity and computational load. It is notable that all the self-generated fuzzy rules maybe insufficient for BER improvement and some are redundant while increasing the computational complexity.

Based on the above-mentioned review, this paper proposes the optimal NFE for the nonlinear MR channels of the PMR system at high recording density. The contributions of this paper are as follows. (1) The hyperparameters (structural and training), fuzzy rules, genetic parameters, and relevant parameters, are simultaneously optimized and tuned by the multi-objective genetic algorithm (MOGA) through the specified fitness function. The parameters of NFE (i.e., connected weights) are learned using the back-propagation algorithm (BPA), and the LMS method. (2) The robustness, effectiveness, and generalization of the proposed NFE are examined under two different nonlinear PMR channels: (i) the Volterra channel based on NLAD [12], and (ii) the media jitter noise channel under NLTS.

Table 1: *List of symbols.*

Symbol	Description
a_k and $\hat{a}_{k-D}$	Channel input and delayed-output bit of equalizer $\in \{-1, 1\}$
$\tilde{a}_{k-D}$	Delayed-output data of equalizer $\in [-1, 1]$
L	Channel memory length
$\underline{\mathbf{S}}_{\mathbf{k}}$ and S_{k-i}	Column vector and the i^{th}-scalar delayed readback signal of the equalizer input
c_i^l and σ_i^l	The centre and width of the membership function of the l^{th}-linguistic variable for the i^{th}-fuzzy input
F_i^l and G_j^o	The l^{th}- and o^{th}-linguistic variable of the i^{th}-fuzzy input and the j^{th}-output fuzzy, respectively
θ and p	The consequent parameters of M-FIS and S-FIS
$\prod_{i=0}^{N-1} MF_i$	The total number of fuzzy rules
w	Weight parameter
μ	Degree of membership function

The performances of the proposed NFEs and the other existing equalizers are evaluated through two error metrics: BER and mean square error (MSE). The remainder of the paper is structured as follows. In Section 2, Volterra and media noise channel modeling are introduced, with the existing equalizers briefly described. The design of the proposed NFE as well as GA optimization is detailed in Section 3. The simulation results and discussions are provided in Section 4. Section 5 summarizes the paper and points to the direction of future works. For clarity, the numerous variables used in this study are listed in Table 1.

2. CHANNEL MODEL AND EXISTING NONLINEAR EQUALIZERS

2.1 Channel Models

Considering a nonlinear read channel of the PMR system in Fig. 1, let $a_k \in \{\pm 1\}$ be the k^{th} information data bit to the channel. The transition sequence $d_k \in \{0, \pm 1\}$ is generated by the differentiator $(1 - D)/2$, where '0' means no transition at position k and '$\pm$1' means a transition in the opposite direction. Let the transition response $h(t)$ be [13]

$$h(t) = V \tanh\left(\frac{(\ln 3)\, t}{PW_{50}}\right) \tag{1}$$

where $V = lim_{t\to\infty} h(t)$ is the amplitude and PW_{50} is the width of $dh(t)/dt$ at half the maximum peak,

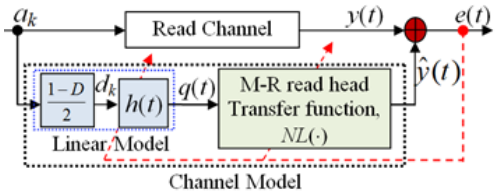


Fig. 1: *Identification of a nonlinear channel model in the MR read heads of the PMR system.*

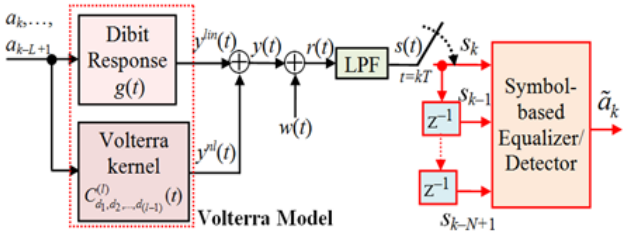


Fig. 2: *Volterra channel model and symbol-based equalizer/detector.*

reflecting the extent of the ISI.

The noiseless linear signal $q(t)$ can be written as

$$q(t) = \sum_{k=-\infty}^{\infty} d_k h(t - kT) \tag{2}$$

where T is the duration of one bit. It is passed through the nonlinear function $NL(\cdot)$, characterized by the M-R heads, causing a reduction in amplitude. For the MR channels, a normalized recording density (ND), defined as PW_{50}/T refers to the number of recorded bits within the length of PW_{50}.

To accurately identify the channel models, the error is minimized by adapting the coefficients of the linear and nonlinear models to fit the observed channel inputs and outputs. Nonlinear channel models can be modeled by the Volterra model (VM) [14]. In the VM, Volterra series expansion is used to separate the interaction order of nonlinearities. In this paper, the VM (Fig. 2) is used to physically model nonlinear PMR channels involving nonlinear distortion terms up to order 3.

In an MR channel with the finite-memory length $L = u + v + 1$, the nonlinear signal $y(t)$ involving the symbols $[a_{i-u}, \ldots, a_{i+v}]$ in T can be expressed by the M^{th}-order VM, i.e., with the power series of Volterra kernel $C^{(l)}_{d_1,\ldots,d_{l-1}}(t)$ characterizing the l^{th}-order nonlinearity, $l = 0, \ldots, M$, as

$$y(t) = \sum_{l=0}^{M} \sum_{k} \sum_{d_1=1}^{L-l} \cdots \sum_{d_{l-1}=d_{l-2}+1}^{L-1} C^{(l+1)}_{d_1,\ldots,d_{l-1}}(t-kT) a_k \prod_{j=1}^{l} a_{k-d_j} \tag{3}$$

where $M = 1, 2, \ldots, L-1$. The superscript of C represents the number of bits involved, whereas $d = 1, \ldots, L-1$ in the subscript of C is the distance of these bits from a_k.

Eq. (3) can be rewritten in a matrix form as, $\mathbf{Y} = \mathbf{H} \times \mathbf{F}$, where $\mathbf{F}$ and $\mathbf{Y}$ represent the $2^L \times \text{OSR}$ matrix of Volterra kernel chips and signal chips, respectively, OSR is the oversampling ratio and $\mathbf{H}$ is $2^L \times 2^L$ Hadamard matrix. The matrix $\mathbf{H}$ contains the elements of all the 2^L distinct L-tuples generated from a pseudorandom binary sequence (PRBS) of the L^{th}-order polynomial generator, corresponding to the signal chips listed in $\mathbf{Y}$. Thus, the inverse of $\mathbf{H}$ is its transpose, and therefore, $\mathbf{F} = \mathbf{H}^T \times \mathbf{Y}/2^L$. The 2^{L-1} unknown Volterra kernels are derived from the sum of non-overlapping 2^L Volterra kernel chips $f^{(l)}_{k_1,k_2,\ldots,k_l}$ of $\mathbf{F}$ as [14],

$$C^{(l+1)}_{d_1,\ldots,d_{l-1}}(t) = \sum_{k=-v}^{u-d_{l-1}} f^{(l)}_{k+d_{l-1},\cdots,k+d_1,k}(t - kT) \tag{4}$$

For example, in the case of $L = 3$ and $M = 2$, there are eight kernel chips $f^{(0)}$, $f^{(1)}_1$, $f^{(1)}_0$, $f^{(1)}_{-1}$, $f^{(2)}_{1,0}$, $f^{(2)}_{1,-1}$, $f^{(2)}_{0,-1}$, $f^{(3)}_{1,0,-1}$, and four Volterra kernels of $C^{(0)}$, $C^{(1)}$, $C^{(2)}_1$, and $C^{(3)}_{1,2}$, where $C^{(0)}(t) = \left[f^{(0)}(t)\right]$, $C^{(1)}(t) = \left[f^{(1)}_{-1}(t+T), f^{(1)}_0(t), f^{(1)}_1(t-T)\right]$, $C^{(2)}(t) = \left[f^{(2)}_{1,-1}(t+T), f^{(2)}_{1,0}(t), f^{(2)}_{0,-1}(t-T)\right]$, and $C^{(3)}(t) = \left[f^{(3)}_{1,0,-1}(t)\right]$ characterizing the data as independent, linear, second, and third-order nonlinear terms, respectively. It can be observed that the complexity of VM increases with L and M. Therefore, the truncated VM (i.e., $M < L-1$) with only the significant Volterra kernels can reduce the complexity.

On the receiver front, the noisy readback signal, $r(t) = y(t) + w(t)$, is filtered and sampled to the input of the equalizer, where $w(t)$ is the noise from the electronic circuits in the M-R head (i.e., transducer) usually modeled as additive white Gaussian noise (AWGN). Then, the SNR is defined as $\text{SNR}_1\,(\text{dB}) = 20\log(V/\sigma_w)$, where σ_w^2 is the power of AWGN.

To further examine the robustness and generalization of the proposed NFE, the MR channel in the presence of media noise as a major current MR technology source [3] is used for testing. The media noise arises from the zigzag transition boundary of tiny magnetic grains in thin film media. Among the components of the media noise, the jitter noise of the transition position is the main disturbance in the high ND of the PMR system and degrades the detector's performance. It is included in the readback signal (Fig. 3) using the first-order Taylor series approximation which can be written as

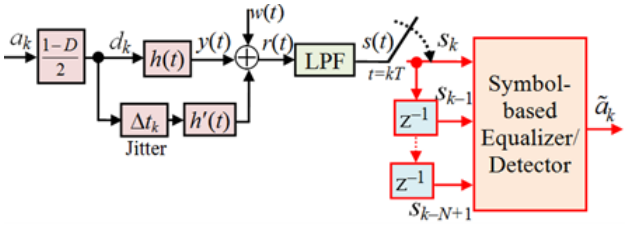
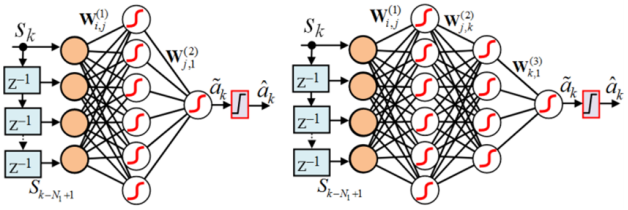


Fig. 3: *Jitter media channel model and symbol-based equalizer/detector.*



(a) Single hidden layer (b) Double hidden layers

Fig. 4: *The MLPNNE structure.*

$$r(t) \approx \sum_{k=-\infty}^{\infty} d_k h(t-kT) + \sum_{k=-\infty}^{\infty} \Delta t_k d_k h'(t-kT) + w(t) \tag{5}$$

where Δt_k is a transition jitter affecting the transition sequence d_k. If $d_k = 0$ (no transition) then $\Delta t_k = 0$. When d_k is non-zero, Δt_k is modeled as $N(0, \sigma_\Delta^2)$, and truncated to $|\Delta t_k| < T/2$. It is defined as $(\sigma_\Delta/T)\%$. Then the total noise power at the equalizer input is $\sigma^2 = \sigma_w^2 + \sigma_\Delta^2$. Therefore, $\mathrm{SNR}_2\,(\mathrm{dB}) = 20\log_{10}(V/\sigma)$, and here, σ_Δ^2/σ^2 is defined as the jitter-to-noise ratio (JNR).

Besides the ISI, the inter-track interference (ITI) arising with an increase in ND is modeled as a linear signal and added to the main track signal. In addition, thermal asperities (TAs) are another problem in the MR system, caused by defects on the medium surfaces, resulting in long burst errors. The signal processing techniques for mitigating the ITI and TA in the PMR system are detailed in [15] and [16], respectively.

2.2 Existing Nonlinear Equalizers

Although various PRML methods have been used for equalizing and detecting, they are insufficiently effective to mitigate the NLDs in the current MR technology. Therefore, nonlinear equalizers based on nonlinear signal processing have been increasingly used in MR systems, such as VE, various NNEs, hybrid VE-MLPNNE, FLE, and NFE, similar to those applied in communication channels. Among them, MLPNNE and FLE have been implemented in the Volterra and jitter channels [10], which are further used in comparison with the proposed NFE.

2.2.1 Multilayer perceptron NNE (MLPNNE)

MLPNNE, one class of NNE, is a nonlinear adaptive equalizer, which can form arbitrarily nonlinear decision boundaries for addressing the equalization problem. For example, the MLPNNE with VD [17] is implemented in the media noise channel of the PMR. However, noise correlation exists at the equalizer output, which degrades the performance of VD. To address the problem, the MLPNNE with a noise-whitening filter [18] is proposed to improve the VD performance. Recently, the MLPNNE integrated with the MLSE [19] has been applied to the Volterra channels. However, the single and double hidden layers of the MLPNNE classifier with a simple threshold detector [19] (Fig. 4), simple circuit implementation, and without noise correlation, yield a similar BER performance to the ML-MLPNNE. In addition, the double hidden layer MLPNNE and hybrid VE-NNE [8] are equivalent in equalization performance and outperform the single hidden layer MLPNNE.

Let $\underline{\mathbf{S_k}} = [S_k, \ldots, S_{k-N_1+1}]^T$ be an N_1-element column vector of the delayed sampled readback which is used as the equalizer input, where $\underline{\mathbf{S_k}} \in$ UOD (universe of discourse) $\equiv [C_0^-, C_0^+] \times \cdots \times [C_{N_1-1}^-, C_{N_1-1}^+] \subset R^{N_1}$. C_i^- and C_i^+ are the lower and upper limits of S_{k-i}. In Fig. 4(a), the estimated bit through the nonlinear transformations of the linearly weighted combination can be expressed as

$$\tilde{a}_{k-D} = f\left(\sum_{j=0}^{N_2-1} w_{j,1}^{(2)} f_j\left(\sum_{i=0}^{N_1-1} w_{i,(j+1)}^{(1)} s_{k-i}\right)\right) \tag{6}$$

where $w_{i,j}^{(1)}$ and $w_{j,1}^{(2)}$ are the connected weights of the i^{th} input neuron to the j^{th} hidden neuron and the j^{th} hidden neuron to the neuron output, respectively. D is the unit delay operator, and $f(\cdot)$ is the activation function.

Finally, $\tilde{a}_{k-D}$ is passed through the two-level threshold unit to obtain the output bit $\hat{a}_{k-D}$.

In training MLPNNE, given the maximum iteration ($Iter_{\max}$) the training input-output data of $\{\underline{\mathbf{S_k}}, a(k-D)\}, k = 1, \ldots, N_{\text{train}}$, where the scalar $a(k-D)$ is the desired output, the training of Eq. (6) optimizes the objective function J from

$$\underset{\mathbf{w}}{\operatorname{argmin}}\, J = \frac{1}{Iter_{\max}} \sum_{i=1}^{Iter_{\max}} \delta^{Iter_{\max}-i} \sum_{k=1}^{N_{\text{train}}} (a_{k-D} - \tilde{a}_{k-D})^2 \tag{7}$$

where $\delta \in (0, 1]$ is the forgetting factor that helps to avoid local solution traps. The BPA updates the weights as $\mathbf{W} \leftarrow \mathbf{W} + \eta\,(\partial \mathbf{J}/\partial \mathbf{W})$, where η is the learning rate.

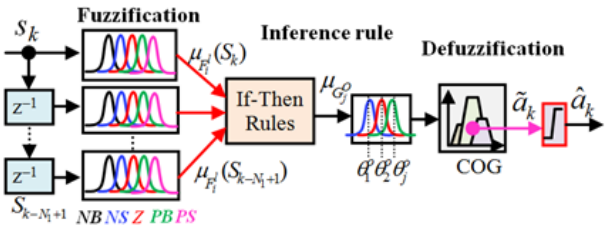


Fig. 5: *Block diagram of the FLE-based M-FIS.*

2.2.2 Fuzzy logic equalizer (FLE)

The FLE has a similar structure to the NNE, except for using many rules in making the decision output. It is implemented in various aspects, such as the fuzzy adaptive filter [20] and complex FLE [21]. Furthermore, the optimized FLE using the extended Kalman filter adaption [22] and GA [10] outperforms the BER compared to the others. The design of the FLE (Fig. 5) is composed of three main stages. 1) In the fuzzification stage, the Gaussian membership functions (GMFs) among other types are constructed as the basis for spreading over the specified UOD, each labeled by the linguistic variable F^l, e.g., negative and positive big (NB and PB), zero (Z), and negative and positive small (NS and PS), where $l = 1, \ldots, MF$ (the number of membership functions). The crisp input of S_{k-i} is mapped into the degree of GMF as

$$\mu_{F_i^l}(S_{k-i}) = \exp\left[-\left(\frac{S_{k-i} - c_i^l}{\sqrt{2}\sigma_i^l}\right)^2\right] \qquad (8)$$

where $\mu \in [0, 1]$, c and σ are the center and width of GMF.

2) In the fuzzy inference system (FIS) stage, the Mamdani or Sugeno FIS (M-FIS or S-FIS), which are different in their output forms, are used for nonlinear mapping of the FLE. However, no theory guarantees which one is better. For the existing FLE based on M-FIS (i.e., the outputs are fuzzy sets) (Fig. 5), the fuzzy input values are mapped into the fuzzy outputs through the fuzzy rules containing the If/Then rules, for example, the j^{th}-rule: **IF** S_k is F_1^l and, ..., and S_{k-N+1} is F_N^l, **THEN** $\tilde{a}_{k-D,j}$ is G_j^o, where $j = 1, \ldots, \prod_{i=0}^{N-1} MF_i$, G_j^o is the o^{th}-linguistic variable of the fuzzy output, and $o = 1, \ldots, MF_o$ (the number of membership functions of fuzzy output, $MF_o \leq \prod_{i=0}^{N-1} MF_i$).

Lastly, 3) In the defuzzification stage, a degree of GMF, $\mu_{G_j^o}(\tilde{a}_{k-D}) = \max\left\{\min\{\mu_{F_i^l}(S_k), \ldots, \mu_{F_i^l}(S_{k-N_1+1})\}\right\}$ of fuzzy output is defuzzified to the crisp output by the center of gravity (COG) method [10], i.e.,

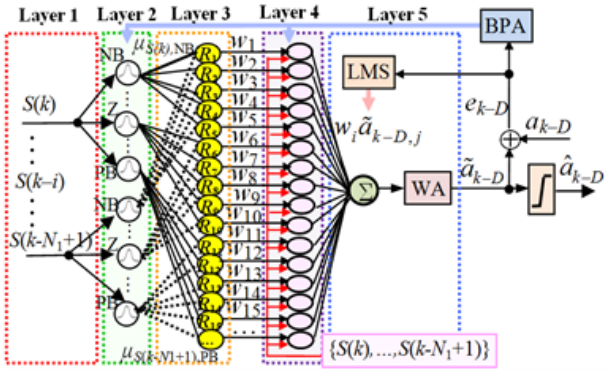


Fig. 6: *Block diagram of the proposed NFE.*

$$\tilde{a}_{k-D} = \frac{\sum_{j=1}^{\prod_{i=0}^{N-1} MF_i} \theta_j^o \left(\prod_{i=0}^{N-1} \mu_{F_i^l}(S_{k-i})\right)}{\sum_{j=1}^{\prod_{i=0}^{N-1} MF_i} \left(\prod_{i=0}^{N-1} \mu_{F_i^l}(S_{k-i})\right)} \qquad (9)$$

where θ_j^o is the parameter in the range of $[-1, 1]$ in which $\mu_{G_j^o}$ achieves the maximum value.

3. THE PROPOSED NEURO-FUZZY EQUALIZER

The neuro-fuzzy represents an FIS implemented in the framework of an adaptive NN. For channel equalization applications, NFEs have been proposed in various communication systems. For the MR system, it seems to be limited due to the difficulty in design without an exact closed form. The proposed NFEs are detailed in the following sub-sections.

3.1 Neuro-fuzzy Equalizer (NFE) Scheme

The NFE (Fig. 6) has a similar structure to the MLPNNE, typically consisting of five layers: input, output, and three hidden layers. It is based on the S-FIS (i.e., using a linear combination of inputs with an additional constant term as the fuzzy output of each rule instead of using fuzzy sets).

The steps for setting up the NFE are as follows. The crisp input vector of $\underline{\mathbf{S_k}}$ having an N_1-element of S_{k-i} is scaled into the range $[-1, 1]$ and fed into the input neurons of the first layer. The link weights at this layer are fixed to unity. On the second layer, they are converted to the linguistic variable F_i usually represented by the GMF of M_i-partition, and their degree of MF is determined using Eq. (8) through adaptive neurons based on two parameters of GMFs (c and σ) referred to as premise parameters. In the following third layer, each neuron represents a single j^{th}-fuzzy rule, where $j = 1, \ldots, NR$ (the total self-generated rules). The output firing strength or weight (w_j) is determined by the product T-norm operator for the antecedent part of the fuzzy rule. In the fourth

layer, the neurons represent the fuzzy sets used in the consequent part of fuzzy rules. The consequence rule outputs $\tilde{a}_{k-D,j}$ corresponding to their weights are evaluated by combining all crisp inputs from the first layer using the linear combination.

Therefore, the j^{th}-fuzzy rule in the third layer can be represented by, for example, **IF** S_k is NB, and S_{k-1} is NB, ..., and S_{k-N_1+1} is NB, **THEN** $\tilde{a}_{k-D,j} = p_{j,0}S_k + p_{j,1}S_{k-1} + \ldots + p_{j,N_1-1}S_{k-N_1+1} + p_j$, where $p_{j,i}$ is the consequent parameter. Finally, in the fifth layer, the equalizer output is determined using the weight average (WA) method which can be expressed as

$$\tilde{a}_{k-D} = \frac{\sum_{j=1}^{NR} w_j \tilde{a}_{k-D,j}}{\sum_{j=1}^{NR} w_j} = \frac{\sum_{j=1}^{NR}\prod_{i=0}^{N_1-1}\mu_{F_i^l}(S_{k-i})\left(\sum_{i=0}^{N_1-1} p_{j,i}S_{k-i}\right)}{\sum_{j=1}^{NR}\prod_{i=0}^{N_1-1}\mu_{F_i^l}(S_{k-i})} \quad (10)$$

where $NR = \prod_{i=1}^{N_1} M_i$, where $i = 1, \ldots, N_1$ and $l \in \{NB, \ldots, Z, \ldots, PB\}$. For the conventional NFE, given the training input vector and corresponding delayed-output data of $\{(\underline{\mathbf{S}}_\mathbf{k}, a(k-D))_i\}$, the premise parameters (c_i^l and σ_i^l) are updated through BPA, $(c_i^l, \sigma_i^l) \leftarrow (c_i^l, \sigma_i^l) + \eta\partial J/\partial(c_i^l, \sigma_i^l)$ whereas, the consequent parameters ($p_{j,i}$) are updated using the least square method (LSM), $p \leftarrow p + \eta e\underline{\mathbf{S}}^T$.

However, these parameters are increased depending on the self-generated rule (NR-rule) that grows exponentially with the number of inputs and their MFs. The accuracy of some self-generated rules may not sufficiently improve, while becoming redundant and increasingly computationally complex. To minimize the size of the NFE structure, the number of inputs (N_1) should not exceed three, i.e., S_k, S_{k-1}, and S_{k-2}, whereas the number of MFs (M) varies enough in the specific UOD by starting with at least two. For example, for $N_1 = 3$, $M = 5$, therefore $NR = 5^3 = 125$, the number of premise and consequent parameters are $3 \times 5 = 15$, and 125×4 ($p_{j,0}$, $p_{j,1}$, $p_{j,2}$, and p_j) $= 500$, respectively. A large number of adapted parameters BPA and LSE require suitable initial values in the iteration process and may not converge optimally to provide the best solutions and tend to be subject to overfitting. Here, $N_1 = 2$ is the best choice when designing the NFE.

3.2 NFE Optimized by Genetic Algorithm (GA)

Since NN is a bio-inspired model, there is no solid answer to identifying the hyperparameters, such as the learning rate (η), number of inputs (N_1), membership functions (MF), fuzzy rules (NR), and epoch ($Epoch$). Choosing the proper hyperparameters is key to maximizing the NFE performance. In this work, the GA based on evolutionary high dimensional searching yields the best fit and is alternatively employed to accelerate BPA and LSE. To maintain the degree of freedom in optimization, the multi-objective GA (MOGA) is used to optimize these hyperparameters while simultaneously reducing the redundant fuzzy rules and fine-tuning the system parameters of the NFE, including other relevant parameters. There are typically four stages to the GA: constructing the chromosome (solution), evaluating and ranking, selection, and genetic operations; the probability of crossover and mutation (P_c and P_m).

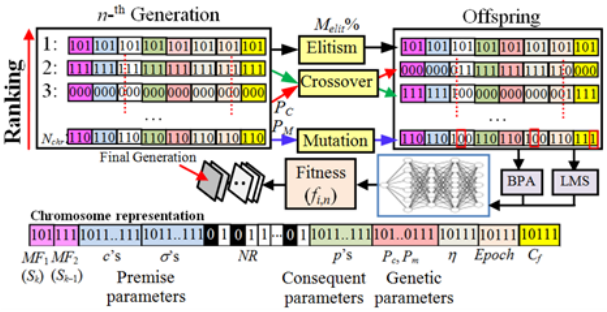


Fig. 7: *Optimization of the NFE using MOGA.*

In the optimization of NFE by GA as shown in Fig. 7, the number of individual chromosomes per generation (N_{chr}) as a vector possessing certain genes (parameters) is randomly generated and encoded in the $N_{bit,k}$-length binary string of the k^{th}-gene. The parameters c, σ, p, P_c, P_m, η, $Epoch$, and C_f are normalized within the predefined range $[Gene_j^{\min}, Gene_j^{\max}]$ as

$$Gene(i,j) = \frac{\left(Gene_j^{\max} - Gene_j^{\min}\right)}{2^{N_{bit}}} \times y(i,j) + Gene_j^{\min} \quad (11)$$

where $y(i,j)$ is the equivalent decimal number converted from the binary string of the j^{th}-gene. The hyperparameters MF_1 and MF_2 having eight levels (2–9) are each represented by {000, 001, 011, ..., 111} and the set of NR-fuzzy rules is randomly generated from the integer number of 0 and 1, where '0' means the dropout of relevant rules while '1' means taking these rules into consideration. The individual chromosome is assessed using the average MSE $= E\left[(e_{k-D})^2\right]$ of each nonlinearity case to equalize the channel under the following specified fitness function

$$f_{i,n} = \log \text{MSE}_{i,n} + C_{f_{i,n}} \frac{NR_{i,n}}{NR_{i,n-1}} \quad (12)$$

where $f_{i,n}$ is the fitness of the i^{th}-chromosome at the n^{th}-generation considering both accuracy improvement in terms of MSE (Eq. (7)) and complexity

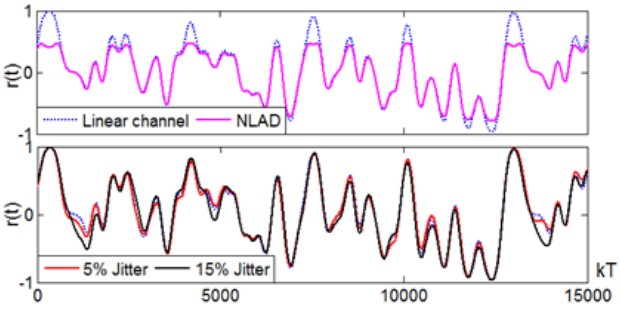


Fig. 8: *The readback signal under NLAD of VM, and NLTS with 5% and 15% jitter noise at the JNR = 100%.*

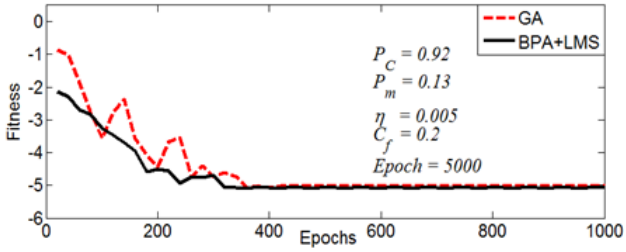


Fig. 9: *Convergence of the best chromosome in GA-assisted BPA and LMS optimization.*

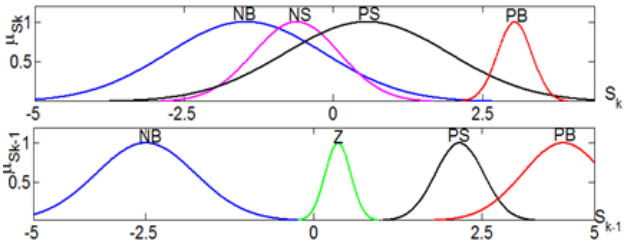


Fig. 10: *The best shape for GMF.*

reduction in terms of the number of rules at the previous iteration ($NR_{i,n-1}$) with respect to the current iteration ($NR_{i,n}$), and C_f is the weighted constant providing the percentage of the total fitness function. According to Eq. (12), the lower the fitness value, the higher the scoring chromosome. To maintain some of the higher chromosomes by unconditionally passing to the next generation, the elitism strategy is used to keep the dominant chromosome at M_{elit}%. On the other hand, the remaining chromosomes are selected for the next generation by assigning a higher probability to individuals with lower fitness values. Two individuals from the preserved and selected chromosomes are employed to change the information on each other in the crossover stage by the probability of P_c. In order to escape the local minima, the newly generated chromosomes are mutated in a random position with the probability of P_m. After the genetic operation, they are further trained using BPA and LMS with initial values according to the best solution obtained from GA, with the average of MSE and the number of significant rules taken for the fitness function. The process of GA is then repeated until the maximum generation ($Gen_{\max}$) is met. To achieve the maximum number of nine MFs, up to 347 parameters are optimized by GA (i.e., $9 \times 2 = 18$ premise parameters c and σ, $9^2 = 81$ rules, $81 \times 3 = 243$ consequent parameters of p, and 5 parameters of the rest).

4. SIMULATION RESULTS AND DISCUSSIONS

4.1 Experimental Setup

In the channel simulations of the PMR system at $ND = 3$, the observed channel output contains part of the transmitted 7-symbol $(a_k, \ldots, a_{k-6})$, i.e., $L = 7$ (PRBS $x^7 + x^3 + 1$). The Volterra channel (Eq. (3)) in the presence of NLAD by the third-order nonlinearity function, $NL(q(t)) = q(t) - 0.35q^2(t) - 0.55q^3(t)$ (reduced amplitude) and AWGN is modeled by the third-order reduced complexity VM ($M = 3$) using significant Volterra kernels of $C^{(1)}, C_1^{(2)}, \ldots, C_4^{(2)}, C_{1,2}^{(3)}, C_{1,3}^{(3)}, C_{1,4}^{(3)}$, and $C_{2,3}^{(3)}$. The readback signal generated from the VM is shown in Fig. 8 (top). Besides, the readback signal in the presence of NLTS is due to the media noise model Eq. (5) with Δt_k varying in the jitter range of 5–15% and AWGN generated as shown in Fig. 8 (bottom). To minimize the NFE structure by limiting $N_1 = 2$, the datasets of noisy sampled readback signals S_k and S_{k-1} with the corresponding delayed bit a_{k-D} are recorded and separated into training/testing sets.

On the receiver front, the datasets of $\{\underline{\mathbf{S_k}}, a_{k-D}\}$ are used to formulate the equalizers, including the proposed NFE through the MOGA, by assisting the BPA and LMS in fine-tuning the system parameters. All parameters in the individual chromosome are varied according to their searching range as follows: $c \in [-3, 3]$, $\sigma \in [0, 3]$, $p \in [-1, 1]$, $P_c \in [0.7, 0.95]$, $P_m \in [0.05, 0.25]$, $\eta \in [0.001, 0.05]$, $Epoch \in [100, 10^4]$, and $C_f \in [0.01, 0.5]$. The fixed GA parameters used in the training consist of $N_{chr} = 10$, $N_{bit} = 10$, $M_{elit} = 0\%$, and $Gen_{\max} = 1,000$. In addition, multiple-point crossover and mutation are used in the genetic process. After the optimization of GA to assist the BPA and LMS, the chromosome containing the best set of NFE parameters is presented in Table 2, whereas the other relevant parameters together with the convergence are depicted in Fig. 9. The best shape of four GMFs for each S_k and S_{k-1} based upon the premise parameters c and σ is shown in Fig. 10, although a number of consequent p parameters are also obtained, but not shown here. Furthermore, the existing equalizers: MLPNNE (Eq. (6)), FLE (Eq. (9)), and NFE (Eq. (10)) generated by the conventional training methods and detailed in [1, 10] are also

Table 2: *Simulation setup parameters for equalizers.*

Parameter	Setting value and variable			
	MLPNNE	FLE	NFE	Proposed NFE
Inputs	$[S_k, \ldots, S_{k-6}]$	$[S_k, S_{k-1}]$	$[S_k, S_{k-1}]$	$[S_k, S_{k-1}]$
Transformation function	Hyperbolic tangent	GMF	GMF	GMF
Hidden layer	1	1	3	3
GMF input	–	[5, 5]	[5, 5]	[4, 4]
FIS	–	M-FIS	S-FIS	S-FIS
Hidden nodes	25	25	25, 25, 25	16, 12, 12
Single output	$\tilde{a}_{k-D}$	$\tilde{a}_{k-D}$	$\tilde{a}_{k-D}$	$\tilde{a}_{k-D}$
GMF output	–	5	–	–
Fuzzy rules	–	25	25	12
Consequent parameters	–	–	75	36
Training method	BPA	Conventional	BPA and LMS	GA assisted BPA and LMS
Equalizer parameters	w and b	c, σ, and rules	c, σ, and p	c, σ, and p
Total parameters	200	50	95	52
Multiplications in processing*	275	135	151	89

*In determining computational complexity in terms of multiplication counts, the degree of GMF (Eq. (8)) is represented by the third-order term Taylor series as $x = u$, i.e., $e^{x-u} = 1 + (x-u) + \frac{1}{2}(x-u)^2 + O(x^3)$, where $|x-u| < 1$, requires $5 \times \sum_i M_i$. The computation of weight in the fourth layer and WA (Eq. (10)) provides $NR \times N_1$, and $(2 \times NR + 1)$, respectively.

presented in Table 2. Moreover, from the optimized fuzzy rule produced by the MOGA, the self-generated rules of $4 \times 4 = 16$ reduced to 12 means, represents a 25% reduction in complexity. Comparing the computational complexity using a trade-off between the total number of parameters for a given NFE and multiplication (Table 2), results in the proposed NFE being 35–65% less complex than the rest.

4.2 Equalizer Detection Performance

The simulation results, in terms of BER, for the equalizers under Volterra and media noise channels are illustrated in Fig. 11. It can be observed that the proposed NFE outperforms the rest in both channels. For the Volterra channel, it improves the BER performance by providing about 1, 2, and 3 dB over the conventional NFE, FLE, and MLPNNE, respectively. For the media noise channel with 15% jitter and 100% JNR, it improves the BER performance by providing about 1.5, 2.5, and 3.5 dB over the FLE, conventional NFE, and MLPNNE, respectively.

Furthermore, the proposed NFE requires lower SNR to achieve a BER of 10^{-4} than the conventional NFE in the overall cases of JNR (0–100%) (Fig. 12), demonstrating an improvement of about 6 dB SNR. However, both have been designed at 100% of JNR and cannot significantly reduce the correlated noise for small JNR. It can be observed that the SNR improvement is obtained at over 60% of JNR.

Furthermore, the burst error statistics at the detector output should be monitored to make the corrections using the appropriate error correction

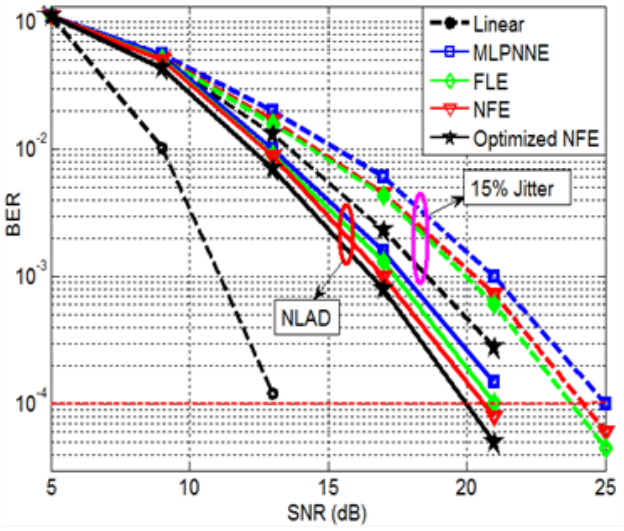


Fig. 11: *BER performance results of equalizers.*

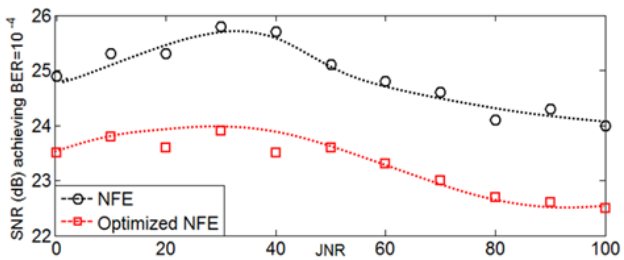


Fig. 12: *Required SNR at each JNR level.*

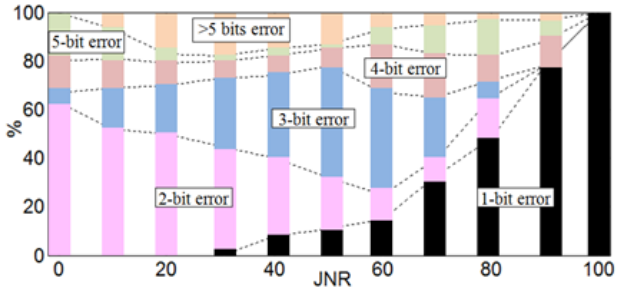


Fig. 13: *Burst error events as a function of JNR.*

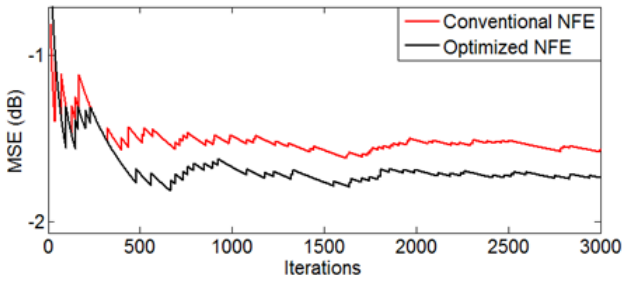


Fig. 14: *The average MSE of the equalizer outputs.*

codes. By running an additional simulation, the findings reveal that the burst error length at the output of the proposed NFE for the media noise channel (at 15% jitter) rises to 28 at BER = 10^{-3}. The error pattern against the JNR is depicted in Fig. 13. It can be observed that a 1-bit error dominates when the JNR is increased, while 2–5 bit errors and others are at the lower JNR. Besides the BER performance, the average results for MSE at the equalizer output of the proposed NFE for both channels (Fig. 14) are improved over those of the conventional NFE at the steady-state condition in which SNR = 15 dB.

5. CONCLUSION

The NLAD and NLTS in the nonlinear PMR channel at high recording density are modeled by the Volterra and the media noise model. The NFE among the other existing equalizers has been proposed for mitigating these nonlinear distortions. In this work, the performance of the NFE is improved by using MOGA optimization to assist the BPA and LMS in searching for optimal system parameters while simultaneously reducing the complexity. When implemented in the Volterra and jitter media noise channels, the proposed NFE outperforms the conventional NFE, as well as the FLE and MLPNNE in SNR gain. Moreover, it improves the MSE performance compared to the rest. To further improve the performance of the detector, the recurrent neural network (RNN) with long-short-term memory (LSTM) based on machine learning shows success in applying channel equalization for various communication channels. In future work, it will be developed for the MR channels.

ACKNOWLEDGEMENT

This work is sponsored by the Program Management Unit-B (PMU-B) under The Office of National Higher Education Science Research and Innovation Policy Council (Grant No. B05F630018).

REFERENCES

[1] R. Wongsathan and P. Supnithi, "The performance of neuro-fuzzy detection on nonlinear magnetic recording channels," in *34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Jeju, Korea, 2019.

[2] Z. Wu, P. H. Siegel, J. K. Wolf, and H. N. Bertram, "Analysis of nonlinear transition shift and write precompensation in perpendicular recording systems," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 2, pp. 158–166, 2010.

[3] Y. Wang, R. H. Victora, and M. F. Erden, "Two-dimensional magnetic recording with a novel write precompensation scheme for 2-D nonlinear transition shift," *IEEE Transaction on Magnetics*, vol. 51, no. 4, pp. 1–7, 2015.

[4] H. Yang and G. Mathew, "Joint design of optimum partial response target and equalizer for recording channels with jitter noise," *IEEE Transaction on Magnetics*, vol. 42, no. 1, pp. 70–77, 2006.

[5] Z. Wu, P. H. Siegel, J. K. Wolf, and H. N. Bertram, "Mean-adjusted pattern-dependent noise prediction for perpendicular recording channels with nonlinear transition shift," *IEEE Transaction on Magnetics*, vol. 44, no. 11, pp. 3761–3764, 2008.

[6] T. R. Oenning and J. Moon, "The effect of jitter noise on binary input intersymbol interference channel capacity," in *ICC 2001 IEEE International Conference on Communication*, Helsinki, Finland, August 2002.

[7] S. V. Vaerebergh, J. Via, and I. Santamana, "A sliding-window kernel RLS algorithm and its application to nonlinear channel identification," in *International conference on Acoustics, Speech, and Signal Processing, ICASSP-88*, 1988, pp. 789–792.

[8] R. Wongsathan, W. Phakphisut, and P. Supnithi, "Performance of the hybrid MLPNN based VE (hMLPNN-VE) for the nonlinear PMR channels," *AIP Advances*, vol. 8, pp. 1–5, 2018.

[9] M. Yamashita *et al.*, "Modeling of writing process for two-dimensional magnetic recording and performance evaluation of two-dimensional neural network equalizer," *IEEE Transaction on Magnetics*, vol. 48, no. 11, pp. 4586–4589, 2012.

[10] R. Wongsathan and P. Supnithi, "Fuzzy logic-based adaptive equalizer for non-linear perpendicular magnetic recording channels," *IET Communications*, vol. 13, no. 9, pp. 1304–1310, 2019.

[11] T. Cavdar, "PSO tuned ANFIS equalizer based on fuzzy C-means clustering algorithm," *International Journal Electronics and Communications (AEÜ)*, vol. 70, no. 6, pp. 799–807, 2016.

[12] W. R. Eppler and I. Ozgunes, "Channel characterization method using dipulse extraction," *IEEE Transaction on Magnetics*, vol. 42, no. 2, pp. 176–181, 2006.

[13] M. Madden, M. Oberg, Z. Wu, and R. He, "Read channel for perpendicular magnetic recording," *IEEE Transaction on Magnetics*, vol. 40, no. 1, pp. 241–246, 2004.

[14] R. Hermann, "Volterra modeling of digital magnetic saturation recording channels," *IEEE Transaction on Magnetics*, vol. 26, pp. 2125–2127, Sept. 1990.

[15] W. Tan and J. Cruz, "Signal processing for perpendicular recording channels with intertrack interference" *IEEE Transaction on Magnetics*, vol. 41, no. 2, pp. 730–735, 2005.

[16] G. Mathew and I. Tjhia, "Thermal asperity suppression in perpendicular recording channels," *IEEE Transactions on Magnetics*, vol. 41, no. 10, pp. 2878–2880, 2005.

[17] H. Osawa, T. Shimizu, T. Nakaoka, Y. Okamoto, H. Saito, H. Muraoka, and Y. Nakamura, "Simplification of neural network equalizer for perpendicular magnetic recording," *Electronics and Communications in Japan (Part II: Electronics)*, vol. 89, no. 2, pp. 19–27, 2006.

[18] H. Osawa, M. Hino, N. Shinohara, Y. Okamoto, Y. Nakamura, and H. Muraoka, "Simplified neural network equalizer with noise whitening function for GPRML system" *IEEE Transaction on Magnetics*, vol. 44, no. 11, pp. 3777–3780, 2008.

[19] R. Wongsathan, W. Phakphisut, and P. Supnithi, "Neural networks equalizers for nonlinear magnetic recording channels" in *14^{th} International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Phuket, Thailand, 2018.

[20] L. X. Wang and J. M. Mendel, "Fuzzy adaptive filters, with application to nonlinear channel equalization," *IEEE Transaction on Fuzzy Systems*, vol. 1, no. 3, pp. 161–170, 1993.

[21] K. Y. Lee, "Complex fuzzy adaptive filter with LMS algorithm" *IEEE Transaction on Signal Processing*, vol. 44, no. 2, pp. 424–427, 1996.

[22] W.K. Wong and H.S. Lim, "A robust and effective fuzzy adaptive equalizer for power-line communication channels," *Neurocomputing*, vol. 71, no. 1–3, pp. 311–322, 2007.

Rati Wongsathan received his Bachelor degree and Master degree in Electrical Engineering from Chiang Mai University, Thailand. He is currently pursuing his Ph.D. in Electrical Engineering at King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. His main research interests are communication channel model, channel detector design, and AI.

Pornchai Supnithi received Ph.D. in electrical Engineering from Georgia Institute of Technology, USA in 2002. He is currently a full professor at Telecommunication Engineering Department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. His research interests are in the area of Telecommunications, Ionospheric and GNSS, Data storage and Engineering Education. His laboratory is maintaining 7 observation stations in Thailand (with ionosonde, beacon receivers, magnetometer, and GNSS receivers) as well as Thai GNSS and Space Weather Information website. He has published over 50 journal articles with impact factors, 100 conference papers, and 3 books/chapters. He was a mid-career Thailand Research Fund scholar from 2013–2015. He currently served as a subcommittee member in the National Research Council of Thailand (NRCT), International Reference Ionosphere (IRI) Committee (under URSI, COSPAR), and 2nd Vice President of ECTI Association.