Online regularization of complex-valued neural networks for structure optimization in wireless-communication channel prediction

Tianben Ding\textsuperscript{a,b}, Akira Hirose\textsuperscript{a,*}

\textsuperscript{a}Department of Electrical Engineering and Information Systems, The University of Tokyo, Tokyo, Japan
\textsuperscript{b}Department of Electrical and Systems Engineering, Washington University in St. Louis, MO, United States

Abstract

This paper proposes an online-learning complex-valued neural network (CVNN) to predict future channel states in fast fading multipath mobile communications. CVNN is a framework suitable for dealing with a fading communication channel, which is complex-valued, as a complex-valued entity to realize accurate channel prediction by utilizing its high generalization ability in the complex domain. However, in actual communication environments having rapid and irregular changes, an empirically selected stationary network gives only limited prediction accuracy. In this paper, we introduce regularization in the update of the CVNN weights to develop online dynamics that can self-optimize its equivalent network size by responding to such channel-state changes. It realizes online adaptive, highly accurate and robust channel prediction with dynamical adjustment of the network size. We demonstrate its online adaptability in simulations and real wireless-propagation experiments.

Keywords: Adaptive communications, channel prediction, complex-valued neural network (CVNN), fading

*Correspondence authors.
Email addresses: tding@wustl.edu (Tianben Ding), ahirosee.ee.t.u-tokyo.ac.jp (Akira Hirose)
1. Introduction

Performance of mobile communications always suffers from signal degradation due to path loss, shadowing, interference and channel state changes caused by movement of users (Cho et al., 2010). In principle, fading, the most serious disturbance, can be mitigated by pre-equalization such as zero-forcing (Ho et al., 2017) or minimum-mean-square-error (MMSE) equalization (Eraslan et al., 2013). Transmission power control is another countermeasure against the fading phenomenon (Ren et al., 2018). These methods rely on accurate estimation of channel state information at the communication ends. However, in practical mobile communications, the channel state, or simply channel, changes rapidly and irregularly due to the movement of mobile users and their surroundings, resulting in time-varying multipath environment. The time fluctuation outdates the estimated channel and degrades the communication quality significantly. Channel prediction is an effective way to overcome this problem by forecasting channel changes in time based on preceding information. An accurate channel prediction is required to improve the communication quality and further adaptive transmission in the next-generation communications (Duel-Hallen, 2007; Bui et al., 2017).

There exist several works on the channel prediction in mobile communications based on, for example, linear (Maehara et al., 2003; Bui et al., 2013) and autoregressive (AR) model extrapolation (Eyceoz et al., 1998; Arredondo et al., 2002; Duel-Hallen et al., 2006; Sharma & Chandra, 2007). Although the low computational complexity in these methods is suitable for real-time operation in mobile communications, such simple linear or AR-based methods provide limited performance on predicting rapid changes of channel (Ding & Hirose, 2014a). Neural-network-based channel prediction methods have also been studied very actively due to the recent successful development of artificial neural networks in various engineering fields. The generalization ability of neural networks provides flexible representation of complicated channel-state changes and high prediction capability. For instance, an echo-state-network (ESN) based (Zhao et al., 2017)
and an extreme-learning-machine (ELM) based (Sui et al., 2018) as well as real-valued recurrent-neural-network (RNN) based (Liu et al., 2006; Potter et al., 2010) prediction methods have been reported, and their prediction performance has been evaluated in some simulated communication situations. To realize a high-precision prediction in practical mobile communications, the authors also proposed a method (Ding & Hirose, 2014a) based on a multiple-layer complex-valued neural network (ML-CVNN) by focusing rotary motion of the channel state in the complex plane. This method gave us superb channel prediction performance in several practical communication scenarios.

Generally, in neural-network-based applications, network size is critical to the application performance because it affects the generalization characteristics and calculation cost (Hirose, 2012; Ramachandram & Taylor, 2017). For example, a too small network is not enough to represent the complexity of targets, showing low convergence properties. On the other hand, a too large network requires expensive calculation costs, and most importantly, it causes overfitting. Despite its importance, the structure of the network is typically defined based on a rule of thumb manually. One may start with an arbitrary structure, and evaluate its learning performance using a large amount of training data by increasing or decreasing the number of neurons and network connections until the best structure is found. This is also the present state of the art for the neural-network-based methods in the channel prediction. For example, in our previous prediction method, we empirically set the structure of the CVNN (the number of input terminals and neurons in the hidden layer) based on its prediction accuracy in a series of simulations with several communication situations. Although the structure shows a high prediction performance on some simulated and experimentally observed fading channels (Ding & Hirose, 2014a), this manual pre-tuning of the network parameters is time consuming and not efficient. Moreover, mobile communications in the real world is forced to work in more diverse communication environments, and experiences more rapid and various fluctuations than those in simulations. As a result, an a priori tuned structure is no longer optimal for other practical communication environment,
but the most suitable neural-network structure is dynamically changing accordingly. This motivates us to develop a dynamics to realize online adaptive and dynamically optimized neural-network structures for the channel prediction.

In this paper, to realize a dynamically optimized network structure to suit best to the fading channel at each moment, we propose a new ML-CVNN-based channel prediction scheme by introducing regularization. We work with a large-size network platform and then let it automatically find, or self-adjust to, a suitable structure within the platform that uses only a limited portion of the network in order to achieve a good generalization. The self-adjustment is performed by imposing a sparse constraint \cite{Elad2010} to the connection weight updates. The sparse constraint suppresses the redundant connection weights to be zeros, and equivalently constructs a smaller scaled network using only the remaining non-zero connections \cite{Ding2014b}. In order to follow the time fluctuation in the channel state to make the network structure optimized, we develop an online training-and-prediction framework. We update the network by using a set of the most recent channel immediately before the prediction with a small learning iteration number. We keep the updated network structure temporarily for the next training-and-predicting time frame. In this way, we change the non-zero connection distribution from time-to-time in the structure so that it keeps the most suitable size of the network for the situations of prediction.

In each training phase, we use a backpropagation of teacher signal (BPTS) \cite{Hirose1996}, rather than the standard error-backpropagation. The BPTS-based update method is simpler with a lower computation cost, which is preferred for mobile communications. We demonstrate that the new channel prediction method with the online adaptive CVNN structure presents highly accurate predictions under fluctuating communication environment not only in a series of simulations but also using actually observed fading channel in experiments. We precisely observe and discuss the effects of its dynamically changing structure on the bit-error rate performance.

The major contributions of our study can be summarized as follows.
1. Proposal of a scheme to update the network structure online to follow
dynamically changing environment at each moment;
2. Design of a new channel prediction method based on a ML-CVNN with the
proposed online dynamic network structure and the BPTS for an adaptive
prediction;
3. Verification of the fact that the proposed fast fading prediction has a per-
formance superior to other approaches on simulated and experimentally
observed channel states.

This paper is organized as follows. Section 2 briefly introduces the channel
model theory and path separation in the frequency domain. After reviewing the
conventional CVNN-based channel prediction in Section 3, we propose a novel
prediction method based on a ML-CVNN with the dynamically changing struc-
ture in Section 4. Then, Sections 5 and 6 present its performance in simulations
and experiments, respectively. Finally, Section 7 provides the conclusion.
2. Channel Model and Multipath Separation in Frequency Domain

Channel states of communications are distorted mainly by multipath interference caused by scattering in the communication environment. In addition, movement of mobile users and/or scatterers causes rapid and irregular channel changes in time. Fig. 1 shows an example of a fading channel states in actual mobile communications. The curves demonstrate irregularity and nonlinearity of channel changes in the complex domain, and expresses difficulty of channel prediction because of its irregularly rotation-like changes. Generally, a signal received at a communication end $y(t)$ at time $t$ is modeled with time-varying channel $c(t)$ as

$$y(t) = c(t)s(t) + n(t) \quad (1)$$

where $s(t)$ and $n(t)$ are transmitted signal and additive white Gaussian noise (AWGN), respectively. According to the Jakes model [Jakes, 1994], fading channel $c(t)$ as a function of time $t$ is modeled as a summation of individual $M$ complex signal paths $c_m(t)$ at a receiver and expressed as

$$c(t) = \sum_{m=1}^{M} c_m(t) = \sum_{m=1}^{M} a_m e^{j(2\pi f_m t + \phi_m)} \quad (2)$$

where $a_m$, $f_m$, and $\phi_m$ are amplitude, Doppler frequency, and phase shift of each single path $m$, and $M$ is the total path number. The Doppler frequency due to movement of a mobile user is given by

$$f_m = \frac{f_c v}{c} \cos \psi \quad (3)$$

where $v$ and $c$ are speed of the mobile user and the speed of light, respectively, $f_c$ is the carrier frequency of the communication, and $\psi$ is the incident radio wave angle with respect to the motion of the mobile user.

Observed channel $c(t)$ in an actual communication can be decomposed into multiple path components $c_m(t)$ in the frequency domain based on this model. Different path components with different incident angles $\psi$ appear as separate peaks in a Doppler frequency spectrum. Hence, the parameters of each path
component can be estimated by finding peak amplitudes and Doppler frequencies for $a_m$ and $f_m$ in the Doppler spectrum and the corresponding phase shifts for $\phi_m$ in the phase spectrum. Chirp z-transform (CZT) with a Hann window provides low calculation cost and a smooth frequency-domain interpolation useful for an accurate estimation of the parameters in the region close to zero frequency (Tan & Hirose, 2009). By sliding the Hann window in the past and by repeating the parameter estimation process, we can obtain separated path components at different time points. We focus on the fact that the separated channel states $c_m(t)$ have rotary locus in the complex plane and, then, predict its change in time for obtaining the future channel by using CVNNs.

3. Conventional CVNN-Based Channel Prediction with a Pre-Defined Network Structure

The changes in the separated channel components $c_m(t)$ can be predicted by ML-CVNNs (Ding & Hirose, 2014a). CVNN is a framework suitable for treating signal rotation and scaling adaptively in the complex plane by use of its high generalization ability (Hirose, 2012; Hirose & Yoshida, 2012). It has been receiving more attentions in various applications that intrinsically require dealing with complex values (Hara & Hirose, 2004; Kawata & Hirose, 2005; Valle, 2014; Arima & Hirose, 2017). With a basic ML-CVNN consisting of a layer of $I_{ML}$ input terminals, a hidden-neuron layer with $K_{ML}$ neurons and an output neuron, we can predict the complex-valued $c_m(t)$ from a set of past channel components, $c_m(t-1), \ldots, c_m(t-I_{ML})$ for paths $m = 1, \ldots, M$. The input terminals distribute input signals, $c_m(t-1), \ldots, c_m(t-I_{ML})$, to the hidden-layer neurons as their inputs $z_0$. In the same way, the outputs of the hidden-layer neurons $z_1$ are passed to the output-layer neuron as its input. The neurons in the hidden layer are fully connected with the input terminals and the output-layer neuron. The output of the output-layer neuron $z_2$ is the prediction result $\tilde{c}_m(t)$. The connection weight $w_{lki}$ to $i$th input at $k$th neuron in layer $l$ is expressed by its amplitude $|w_{lki}|$ and phase $\theta_{lki}$. The internal state $u_{lk}$ of $k$th
neuron in \( \ell \)th layer is obtained as the summation of its inputs \( z_{(\ell-1)} \) weighted by \( \mathbf{w}_{\ell k} = [w_{\ell ki}] \), i.e.,

\[
u_{\ell k} \equiv \sum_i w_{\ell ki} z_{(\ell-1)i} \equiv \sum_i |w_{\ell ki}| |z_{(\ell-1)i}| e^{j(\theta_{\ell ki} + \theta_{(\ell-1)i})} \tag{4}
\]

where \( z_{(\ell-1)i} = |z_{(\ell-1)i}| e^{j\theta_{(\ell-1)i}} \). The output \( z_{lk} \) is given by adopting an amplitude-phase-type activation function \( f_{ap} \) to \( u_{\ell k} \) as

\[
z_{lk} = f_{ap}(u_{\ell k}) = \tanh(|u_{\ell k}| e^{j\text{arg}(u_{\ell k})}) \tag{5}
\]

In our previous work, the connection weights \( \mathbf{W}_{\ell} = [\mathbf{w}_{\ell k}] = [w_{\ell ki}] \) in the ML-CVNN are updated as follows. The ML-CVNN regards the past known channel component \( \hat{c}_m(t) \) as an output teacher signal, while the preceding channel components associated with the same path \( \hat{c}_m(t-1), ..., \hat{c}_m(t-I_{\text{ML}}) \) as the input teacher signals. The weights are updated based on the steepest descent method so that they minimize the difference

\[
E_{\ell} \equiv \frac{1}{2} |z_{\ell} - \hat{z}_{\ell}|^2 \tag{6}
\]

where \( z_{\ell} \) and \( \hat{z}_{\ell} \) denote temporary output signals and the teacher signals, respectively, in layer \( \ell \). The teacher signals in the hidden layer \( \hat{z}_1 \) are the signals obtained through the backpropagation of the teacher signal (BPTS) of the output layer \( \hat{z}_2 \) \cite{Hirose1994, HiroseEckmiller1996, Hirose2012}. The weight updates are performed at each estimated channel components by sliding the teacher signal and the input set in the time domain. We stop the update at a certain small number of iteration \( R_{\text{ML}} \) in the update process for \( \hat{c}_m(t) \) and keep the updated weights as the initial values in the following weight update for \( \hat{c}_m(t+1) \). With this procedure, we reduce the learning cost to follow the weak regularity of the separated channel components \( c_m(t) \) and to achieve a channel prediction with high accuracy.

4. Proposal of Online Self-Optimizing CVNN

There are a number of preceding studies to get optimized structures of neural networks in general \cite{Ishikawa1996, Tzyy-ChyangLu2013, Ramachan-
Fig. 2 shows the construction of the CVNN. We want a CVNN that changes its connections according to the prediction situations, and dynamically keeps a suitable network structure in a series of predictions without manual tuning. To realize such a network, we introduce a constraint for sparsity to the weight updates in order to restrict the connections of networks in a suitably small size. The $L_0$-norm is an exact sparsity measure, and our problem can be redefined as minimizing the error function of the weights with the $L_0$-norm constraint on the connection weights. However, this problem has been shown to be NP-hard.
in general. Fortunately, under some conditions, the $L_1$-norm can serve as sparsity measure for substituting the $L_0$-norm (Donoho & Elad, 2003; Gribonval & Nielsen, 2003). The $L_1$-norm of the weights is a practical sparsity measure since it is convex so that we can perform optimization more easily (Candes et al., 2006; Donoho & Tanner, 2008; Elad, 2010). By introducing the sparse constraint as a penalty function, the objective function we use to update the weights in layer $l$ is expressed as

$$\arg \min_{W_l} E_l' = \arg \min_{W_l} \left( \frac{1}{2} |z_l - \hat{z}_l|^2 + \alpha \|W_l\|_1 \right)$$  \hspace{1cm} (7)$$

where $\alpha$ is a coefficient to express degrees of the penalty. Minimizing this term means restricting non-zero weight number to get its minimal number in the network. This is effectively equivalent to the pruning. In other words, the penalty function introduces sparsity to the weight updates so that the remaining connection weights form an effective structure for representing the output signal.

We use the steepest descent method in the complex domain to update the weights here (Hirose, 2012). Thus, the weight amplitude $|w_{lki}|$ and the phase $\theta_{lki}$ are renewed as

$$|w_{lki}|(r + 1) = |w_{lki}|(r) - \kappa_1 \frac{\partial E_l'}{\partial |w_{lki}|}$$

$$= |w_{lki}|(r) - \kappa_1 \left\{ (1 - |z_k|^2) \times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \right\}$$

$$\times (|z_{lk}| - |\hat{z}_{lk}| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\theta_{lki}(r + 1) = \theta_{lki}(r) - \kappa_2 \frac{1}{|w_{lki}|} \frac{\partial E_l'}{\partial \theta_{lki}}$$

$$= \theta_{lki}(r) - \kappa_2 \left\{ (1 - |z_k|^2) \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\theta_{lki}(r + 1) = \theta_{lki}(r) - \kappa_2 \left\{ (1 - |z_k|^2) \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$\times (|z_k| - |\hat{z}_k| |\cos(\theta_{lki} - \hat{\theta}_{lki})| |z_{(l-1)i}|^\text{rot} \left\{ |z_{lk}|^\text{rot} \right\}$$

$$10$$
where \( \theta_{ik_i}^r = \theta_{lk} - \theta_{(l-1)k} - \theta_{lk_i} \), \( r \) is the index of learning iteration, and \( \kappa_1 \) and \( \kappa_2 \) are learning constants. This update rule has an additional term \(+\alpha\) in the amplitude \( |w_{ki}^r| \) update in comparison to the previous ML-CVNN-based method because of the penalty term. For simplicity and lower computational consumption, the BPTS is kept to use in this work for getting the teacher signal \( \hat{z}_1 \) in the hidden layer from the teacher signal in the output layer \( \hat{z}_2 \) as (Hirose, 1994)

\[
\hat{z}_1 = (f_{ap}(\hat{z}_2^*W_2))^*
\]

(10)

where \((\cdot)^*\) represents the complex conjugate or hermite conjugate.

To predict the channel, we update the connection weights by time-sliding the input and output teacher signals by using the channel components \( \hat{c}_m(t) \) estimated sequentially from the Doppler spectrum as we did in the previous work (Ding & Hirose, 2014a). That is, a set of updated weights using the complex-valued estimation \( \hat{c}_m(t) \) as the output teacher signal (prediction \( \hat{c}_m(t+1) \) in Fig. 2) and \( \hat{c}_m(t-1), \ldots, \hat{c}_m(t-I_{\text{ML}}) \) as the input teacher signals are kept in the network and used as the initial weights in the following update by regarding \( \hat{c}_m(t+1) \) as the new output teacher signal (prediction \( \hat{c}_m(t+1) \)) and \( \hat{c}_m(t), \ldots, \hat{c}_m(t-I_{\text{ML}}+1) \) as the new input teacher signals. The weight update is performed until the latest channel component is used for prediction of the
future channel. The combination of the penalty term and the prediction scheme in the time domain is expected to keep the structure with a suitable size for the channel prediction depending on the fluctuating communication environment.

5. Numerical Experiments

In the following two sections, we evaluate the performance of the channel prediction based on the ML-CVNN with the penalty in simulations and experiments. We assume orthogonal frequency-division multiplexing (OFDM) with quadrature phase shift keying (QPSK) modulation, and time division duplex (TDD) as the communication scheme in this paper. Table 1 lists the system parameters.

In this section, we characterize the influence of the degree of penalty $\alpha$ on the neural network size and prediction accuracy for simulated fading channels. The geometrical setup is shown in Fig. 3. We consider communications between a base station (BS) and a mobile user (MU) moving away from the BS at 12 m/s with a certain moving angle. There are two scatterers making 2 paths in addition to the line-of-sight path. The carrier frequency is 2 GHz here.
We predict channel changes in a TDD frame based on its preceding channel states. The past path characteristics are estimated by using CZT with the Hann window. A window with 8-TDD-frame length is applied to the past channel states for estimating the path parameters, \( a_m(t) \), \( f_m(t) \), \( \phi_m(t) \), based on peaks in Doppler spectra and corresponding phase spectra. Then, the past path characteristics \( c_m(t) \) are composed by using the parameters, and assigned as the estimated characteristics at the center of the window. We shift the window center at a TDD-frame interval for estimating multipath characteristics at every TDD frame. The details of the time frames are explained in our previous work (Ding & Hirose, 2014a).

To evaluate the performance in various channel changes, we change the scatterer distance \( \Delta x \) shown in Fig. 3 from \( \Delta x = 0.5 \) to 20 m with 0.5 m step, and performed 100 independent predictions at each scatterer distance along with the movement of MU. We start with the neural network with the parameters listed in Table 2. The penalty prunes and grows the network connections, \( 30 \times 30 \) in the hidden layer and \( 30 \times 1 \) in the output layer, online as the communication situation changes.

Fig. 4(a) shows the mean of the network size at each scatterer distance condition. A connection weight is counted as non-zero if its amplitude satisfies

\[
|w_{lki}| \geq \max(|W_l|)/100
\]  

(11)

Otherwise, the weight was considered as a zero weight. If a weight in the output layer \( (l = 2) \) is counted as a zero weight, all the weights in the hidden layer connecting themselves to the corresponding neurons are also considered
as zero weights in order to fairly compare the penalty effect on the entire network. The network sizes of the ML-CVNN with various penalty coefficients ($\alpha = 0, 10^{-5}, 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 2 \times 10^{-2}$) have been evaluated, and the mean connection numbers for 100 trials in each condition have been normalized by
the maximum possible connections to show the non-zero connection ratio. Corresponding prediction accuracy is calculated by accumulating predicted phase errors within the prediction frame. Fig. 4(b) presents the maximum estimated phase errors in each communication condition, showing stability of the prediction.

We find in Fig. 4(a) that the non-zero weight number consisting effective network decreases as the penalty coefficient $\alpha$ increases as we expect, whereas a network without the penalty ($\alpha = 0$) keeps almost all of the connections active for all communication conditions. In Fig. 4(b), the smaller networks achieved by the penalty show better prediction stability compared to the conventional CVNN-based method ($\alpha = 0$). The results also presents that the proposed prediction method reaches its best performance with a penalty coefficient around $\alpha = 5 \times 10^{-4} \sim 10^{-3}$, and that $\alpha$ larger than this value introduces instability to the channel prediction again. These results show that the proposed prediction method with an appropriate $\alpha$ can prune redundant connections in its network automatically to achieve higher prediction accuracy even in prediction conditions difficult for the conventional method.

6. Experiments in Actual Communication Environment

In this section, we further demonstrate adaptability of the proposed method in prediction with actually observed fading channels. We experimentally observed fading channels in a communication situation shown in Fig. 5. There are a MU as a transmitter and a BS as a receiver in the experimental site with some objects, such as buildings and trees, consisting typical mobile communication environment in an urban area. The MU moves in the direction of the arrows shown in Figs. 5(a) and (b) with a velocity around 12.5 m/s and transmits 1.297 GHz nonmodulated wave from a monopole antenna, whereas the BS receives the wave by using another monopole antenna. The received channel signal was mixed with 1.287 GHz local oscillator wave after an amplifier, and then extracted as a signal at an intermediate frequency of 10 MHz. After pass-
ing it to another amplifier and a band-pass filter with 2 MHz bandwidth, we sample the channel information at 500 k Sample/s. The channel change already shown in Fig. 1 is an example of the fading captured in this communication situation. It is a time-sequential data showing irregular rotation of the channel in the complex plane received at the BS. The channel state gives roughly 2 distinct paths. The channel changes in a TDD frame are predicted based on preceding
Figure 6: Actual propagation-experiment results showing (a) observed phase value, (b) change rate of the phase in every $2 \times 10^{-6}$ ms, and (c) non-zero weight ratios in the CVNN with the penalty degree set at $\alpha = 0$ (conventional) and $\alpha = 5 \times 10^{-4}$ (proposed).

First, we evaluate the time variation of the ML-CVNN size to demonstrate the online dynamics. Fig. 6 shows the sequential changes of the channel phase and the neural network size in the prediction process. The actually received phase value and its change rate are shown in Figs. 6(a) and (b), respectively. Non-zero connection weights are counted by using the same scheme described in Section 5 with (11), and plotted against time in Fig. 6(c). In order to
Figure 7: Actual propagation-experiment results (zoomed into 0.41 – 0.49 ms in Fig. 6) showing (a) observed phase values, (b) change rate of the phase, and (c) change rates of non-zero weight ratio with the penalty coefficient set at $\alpha = 0$ (conventional) and $\alpha = 5 \times 10^{-4}$ (proposed).

demonstrate the impact of the penalty function on the network size change, $\alpha = 5 \times 10^{-4}$ has been used based on the discussion in the previous section. For a comparison, the operation with $\alpha = 0$ is also characterized as the conventional method.

Fig. 6(b) shows that the channel does not always change in the same manner but there are sporadic fast changes among relatively stable states. The fast
changes cause difficulty in the channel prediction and degrade the performance. In Fig. 6(c), we can observe that the proposed method with the penalty function \((\alpha = 5 \times 10^{-4})\) increases its effective structure size at and/or after the large channel changes while the entire trend of the size is kept to be relatively compact through the process. On the other hand, the conventional method without the penalty \((\alpha = 0)\) does not change its network size significantly in any part of the update procedure, and no correlation with the channel changes is observed.

For further discussion, we focus on a prediction period containing three fast phase jumps. Figs. 7(a) and (b) present the channel phase and its change in the period included in Fig. 6 by zooming into 407 ms - 496 ms. Fig. 7(c) shows the change rates of the non-zero weight ratio in the period calculated by taking difference of the weight ratios at two consecutive TDD frames. It
is obvious that the CVNN with the regularization ($\alpha = 5 \times 10^{-4}$) increases its non-zero weight connections in its update process synchronously with the large channel changes in order to adapt its network structure to these difficult prediction parts, and decreases the connections after them. In contrast, the conventional method ($\alpha = 0$) is not sensitive to the channel changes. These results represent that the CVNN with the penalty function accommodates itself to such large and irregular channel changes by increasing its weight connections while it reduces the connections when the channel changes steadily. In other words, the proposed method has the ability to change the network structure dynamically and adaptively online according to the degree of difficulty in the channel prediction.

Finally, we compare prediction accuracy in various channel prediction methods in actual communications. In this test, respective methods predict fading in a TDD frame using channel information prior to that in the same manner as before. The predicted channel states are used for compensating the true fading in a communication situation with the OFDM system described is Section 5. A randomly generated binary sequence has been converted into 1612 QPSK symbols and modulated into transmission signals based on the OFDM parameters shown in Table 1. The signals are assumed transmitted through communication environment with the true fading channel and different levels of additive white Gaussian noise. The received signals are compensated by the predicted channel states with various methods before demodulation of OFDM and QPSK. We independently performed this process for 80 times with multiple prediction methods, namely, linear prediction directly in the time domain, AR-model-based prediction using channel characteristics estimated by CZT, the conventional CVNN-based prediction, and the proposed method. For demonstrating the performance of the proposed method, we also evaluate the conventional CVNN-based method with a smaller network consisting of input terminals $I_{ML} = 30$ and hidden-neuron number $K_{ML} = 5$, CVNN (small, $\alpha = 0$), in addition to the larger network structure listed in Table 2.

Fig. 8 shows the BER curves against bit-energy to noise-power-density ratio
$E_b/N_0$. Here, CVNN (large, $\alpha = 0$) shows the result of the conventional CVNN-based method with the large network, whereas CVNN (large, $\alpha = 5 \times 10^{-4}$) presents that of the proposed method with the same large structure. Communication without any prediction method, that is, channel compensation using channel states in the most recent TDD frame, is also performed for a comparison. Communication bit errors with respective prediction methods are obtained in each process, and accumulated over all iterations for the final bit-error-rate (BER) calculation. The black solid curve in the figure represents BER if the true future channel is perfectly known, thus showing the lower bound of the BER with the considered OFDM setup. We can see the difficulty of the channel prediction on the actual fading as the deviation of the BER curves corresponding to the linear, AR-model-based, and the conventional CVNN-based prediction methods. The larger error rates of these methods than that without any prediction (no prediction) implies that such rapid and irregular changes of the channel states cause the failure of the conventional methods. In contrast, the proposed method with the regularization, CVNN (large, $\alpha = 5 \times 10^{-4}$), achieves accurate prediction even in such difficult communication situations and gives $10^{-5}$ BER at $E_b/N_0 = 22$ dB. The results show that the proposed online adaptive CVNN with the regularization provides higher channel prediction performance due to its dynamically changing structure.

7. Conclusion

In this paper, we proposed the online adaptive channel prediction method based on ML-CVNNs with self-optimizing dynamic structures. The penalty function based on $L_1$-norm of the CVNN weights realizes the adaptive CVNN structure without large increase of calculation costs for the weight updates to achieve highly accurate and robust channel prediction of rapidly changing fading. Simulations and experiments demonstrated that the proposed CVNN automatically changes its effective connection number depending on the channel variation so that it keeps an appropriate network size to achieve accurate chan-
nel prediction. The results presented in the experiments for actually observed channels showed that the proposed method can provide accurate prediction even in the situations difficult for conventional methods including the time-domain linear, the AR-model-based, and the former CVNN-based predictions.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant No. 18H04105.

References

Arima, Y., & Hirose, A. (2017). Performance Dependence on System Parameters in Millimeter-Wave Active Imaging Based on Complex-Valued Neural Networks to Classify Complex Texture. *IEEE Access, 5*, 22927–22939. doi:10.1109/ACCESS.2017.2751618

Arredondo, A., Dandekar, K., & Guanghan Xu (2002). Vector channel modeling and prediction for the improvement of downlink received power. *IEEE Transactions on Communications, 50*, 1121–1129. doi:10.1109/TCOMM.2002.800827

Barakat, M., Druaux, F., Lefebvre, D., Khalil, M., & Mustapha, O. (2011). Self adaptive growing neural network classifier for faults detection and diagnosis. *Neurocomputing, 74*, 3865–3876. doi:10.1016/j.neucom.2011.08.001

Bui, H. P., Ogawa, Y., Nishimura, T., & Ohgane, T. (2013). Performance Evaluation of a Multi-User MIMO System With Prediction of Time-Varying Indoor Channels. *IEEE Transactions on Antennas and Propagation, 61*, 371–379. doi:10.1109/TAP.2012.2214995

Bui, N., Cesana, M., Hosseini, S. A., Liao, Q., Malanchini, I., & Widmer, J. (2017). A Survey of Anticipatory Mobile Networking: Context-Based Classification, Prediction Methodologies, and Optimization Techniques. *IEEE Communications Surveys & Tutorials, 19*, 1790–1821. doi:10.1109/COMST.2017.2694140 arXiv:1606.00191
Candès, E. J., Romberg, J. K., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59, 1207–1223. doi:10.1002/cpa.20124 arXiv:0503066.

Cho, Y. S., Kim, J., Yang, W. Y., & Kang, C. G. (2010). MIMO-OFDM Wireless Communications with MATLAB®. Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/9780470825631 arXiv:arXiv:1011.1669v3.

Ding, T., & Hirose, A. (2014a). Fading Channel Prediction Based on Combination of Complex-Valued Neural Networks and Chirp Z-Transform. *IEEE Transactions on Neural Networks and Learning Systems*, 25, 1686–1695. doi:10.1109/TNNLS.2014.2306420.

Ding, T., & Hirose, A. (2014b). Fading Channel Prediction Based on Self-optimizing Neural Networks. In *Lecture Notes in Computer Science* (pp. 175–182). volume 8834. doi:10.1007/978-3-319-12637-1_22.

Donoho, D. L., & Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. *Proceedings of the National Academy of Sciences*, 100, 2197–2202. doi:10.1073/pnas.0437847100.

Donoho, D. L., & Tanner, J. (2008). Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22, 1–53. doi:10.1090/S0894-0347-08-00600-0.

Duel-Hallen, A. (2007). Fading Channel Prediction for Mobile Radio Adaptive Transmission Systems. *Proceedings of the IEEE*, 95, 2299–2313. doi:10.1109/JPROC.2007.904443.

Duel-Hallen, A., Hallen, H., & Tung-Sheng Yang (2006). Long range prediction and reduced feedback for mobile radio adaptive OFDM systems. *IEEE Transactions on Wireless Communications*, 5, 2723–2733. doi:10.1109/TWC.2006.04219.
Elad, M. (2010). *Sparse and redundant representations: From theory to applications in signal and image processing*. Springer. doi:10.1007/978-1-4419-7011-4 [arXiv:1210.8191]

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48, 71–99. doi:10.1016/0010-0277(93)90058-4

Eraslan, E., Daneshrad, B., & Lou, C.-Y. (2013). Performance Indicator for MIMO MMSE Receivers in the Presence of Channel Estimation Error. *IEEE Wireless Communications Letters*, 2, 211–214. doi:10.1109/WCL.2013.012513.120824 [arXiv:1210.8191]

Eyceoz, T., Duel-Hallen, A., & Hallen, H. (1998). Deterministic channel modeling and long range prediction of fast fading mobile radio channels. *IEEE Communications Letters*, 2, 254–256. doi:10.1109/4234.718494

Gribonval, R., & Nielsen, M. (2003). Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49, 3320–3325. doi:10.1109/TIT.2003.820031.

Hara, T., & Hirose, A. (2004). Plastic mine detecting system using complex-valued self-organizing map that deals with multiple-frequency interferometric images. *Neural Networks*, 17, 1201–1210. doi:10.1016/j.neunet.2004.07.012

Hirose, A. (1994). Applications of complex-valued neural networks to coherent optical computing using phase-sensitive detection scheme. *Information Sciences -Applications-, 2*, 103–117. doi:10.1016/0169-0115(94)90014-0

Hirose, A. (2012). *Complex-valued neural networks, 2nd Edition* volume 400. (2nd ed.). New York: Springer-Verlag. doi:10.1007/978-3-642-27632-3

Hirose, A., & Eckmiller, R. (1996). Coherent optical neural networks that have optical-frequency-controlled behavior and generalization ability in the frequency domain. *Applied Optics*, 35, 836. doi:10.1364/AO.35.000836.
Hirose, A., & Yoshida, S. (2012). Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Transactions on Neural Networks and Learning Systems, 23*, 541–551. doi:10.1109/TNNLS.2012.2183613.

Ho, C. D., Ngo, H. Q., Matthaiou, M., & Duong, T. Q. (2017). On the Performance of Zero-Forcing Processing in Multi-Way Massive MIMO Relay Networks. *IEEE Communications Letters, 21*, 849–852. doi:10.1109/LCOMM.2017.2648795 arXiv:1701.00845.

Ishikawa, M. (1996). Structural learning with forgetting. *Neural Networks, 9*, 509–521. doi:10.1016/0893-6080(96)83696-3.

Jakes, W. (1994). *Microwave Mobile Communications*. (2nd ed.). New York: Wiley. doi:10.1109/9780470545287.ch1.

Karnin, E. (1990). A simple procedure for pruning back-propagation trained neural networks. *IEEE Transactions on Neural Networks, 1*, 239–242. doi:10.1109/72.80236.

Kawata, S., & Hirose, A. (2005). Frequency-multiplexed logic circuit based on a coherent optical neural network. *Applied Optics, 44*, 4053–4059. doi:10.1364/AO.44.004053.

Liu, W., Yang, L.-L., & Lajos, H. (2006). Recurrent Neural Network Based Narrowband Channel Prediction. In *2006 IEEE 63rd Vehicular Technology Conference* (pp. 2173–2177). IEEE volume 5. doi:10.1109/VETECS.2006.1683241.

Maehara, F., Sasamori, F., & Takahata, F. (2003). Linear predictive maximal ratio combining transmitter diversity for OFDM-TDMA/TDD systems. *IEICE Transactions on Communications, E86-B*, 221–229.

Potter, C., Venayagamoorthy, G. K., & Kosbar, K. (2010). RNN based MIMO channel prediction. *Signal Processing, 90*, 440–450. doi:10.1016/j.sigpro.2009.07.013.
Ramachandram, D., & Taylor, G. W. (2017). Deep Multimodal Learning: A Survey on Recent Advances and Trends. *IEEE Signal Processing Magazine, 34*, 96–108. doi:10.1109/MSP.2017.2738401

Reed, R. (1993). Pruning algorithms-a survey. *IEEE Transactions on Neural Networks, 4*, 740–747. doi:10.1109/72.248452

Ren, X., Wu, J., Johansson, K. H., Shi, G., & Shi, L. (2018). Infinite Horizon Optimal Transmission Power Control for Remote State Estimation Over Fading Channels. *IEEE Transactions on Automatic Control, 63*, 85–100. doi:10.1109/TAC.2017.2709914 arXiv:1604.08680

Sharma, P., & Chandra, K. (2007). Prediction of State Transitions in Rayleigh Fading Channels. *IEEE Transactions on Vehicular Technology, 56*, 416–425. doi:10.1109/TVT.2007.891421

Sui, Y., Yu, W., & Luo, Q. (2018). Jointly Optimized Extreme Learning Machine for Short-Term Prediction of Fading Channel. *IEEE Access, 6*, 49029–49039. doi:10.1109/ACCESS.2018.2868480

Tan, S., & Hirose, A. (2009). Low-calculation-cost fading channel prediction using chirp Z-transform. *Electronics Letters, 45*, 418. doi:10.1049/el.2009.3472

Tzyy-Chyang Lu, Gwo-Ruey Yu, & Jyh-Ching Juang (2013). Quantum-Based Algorithm for Optimizing Artificial Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems, 24*, 1266–1278. doi:10.1109/TNNLS.2013.2249089

Valle, M. E. (2014). Complex-Valued Recurrent Correlation Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems, 25*, 1600–1612. doi:10.1109/TNNLS.2014.2341013

Zhao, Y., Gao, H., Beaulieu, N. C., Chen, Z., & Ji, H. (2017). Echo State Network for Fast Channel Prediction in Ricean Fading Scenarios. *IEEE Communications Letters, 21*, 672–675. doi:10.1109/LCOMM.2016.2632120