Learning and Analysis of Damping Factor in Massive MIMO Detection Using BP Algorithm With Node Selection

JUNTA TACHIBANA1, (Student Member, IEEE), AND TOMOAKI OHTSUKI2, (Senior Member, IEEE)
1Graduate School of Science and Technology, Keio University, Kanagawa 223-8522, Japan
2Department of Information and Computer Science, Keio University, Kanagawa 223-8522, Japan
Corresponding author: Tomoaki Ohtsuki (ohtsuki@ics.keio.ac.jp)

ABSTRACT In a massive multiple-input multiple-output (MIMO) system, belief propagation (BP) detection is known as a method to separate and detect received signals. In BP detection, a MIMO channel is represented by a factor graph and the transmitted symbols are estimated by message passing. However, the convergence property of BP deteriorates due to multiple loops included in the MIMO channel. As a method to improve the convergence property and the detection performance, the damped BP that averages two successive messages with a weighing factor (called damping factor) is known. To train the damping factors off-line for each antenna configuration, deep neural network-based damped BP (DNN-dBP) has been reported. The problem with DNN-dBP is that the detection performance deteriorates when there is a difference of the channel correlation between training and test. This is because the optimal damping factors vary with the channel correlation. In this paper, to solve this issue, we derive the damping factors of BP with the node selection (NS) method that selects nodes to be updated to lower spatial correlation using DNN-dBP. By applying the NS method, the channel correlation among the selected nodes in BP detection is lowered. Therefore, the proposed method can improve the detection performance deterioration due to the mismatches of the channel correlations between training and test in DNN-dBP. In addition, the convergence property of BP is improved by applying the NS method. Therefore, the proposed method has the same detection performance with low computational complexity as the conventional DNN-dBP. By computer simulation, it is shown that the proposed method significantly improves the bit error rate (BER) performance deterioration due to the mismatches of the channel correlations between training and test in DNN-dBP. The results also show that the proposed method can show the same BER performance with low computational complexity as the conventional DNN-dBP. We also investigate the distribution of the trained damping factors and evaluate the tendency of that.

INDEX TERMS Massive MIMO, BP detection, neural network, 5G.

I. INTRODUCTION
Multiple-input multiple-output (MIMO) is a technology in which multiple antennas are installed for both transmitter and receiver, and signals are spatially multiplexed. By transmitting multiple signals simultaneously, high data rates can be realized without increasing the bandwidth and transmission power. On the other hand, massive MIMO, which uses several tens to hundreds of antenna elements on the transmitter side, is attracting attention as a key technology of 5G wireless communications [1]. Massive MIMO realizes high spectral efficiency and large-capacity communication and enables simultaneous connection to a large number of user terminals. In MIMO, different signals are transmitted using the same frequency at the same time. Thus, at the receiver side, multiple interfering signals are received. Therefore, it is necessary to separate and detect each signal based on channel state information (CSI). However, in a massive MIMO system, the base station (BS) communicates with a large number of user terminals at the same time, making accurate signal detection difficult and complexity of detection high.

The optimal MIMO detection method is maximum-likelihood detection (MLD) [2]. However, the computational complexity of MLD grows exponentially with the number...
of transmit antennas. Therefore, it is unsuitable for a massive MIMO detection method. Even the simple linear detection methods such as zero forcing (ZF) and minimum mean squared error (MMSE) need the inverse matrix calculation according to the number of antenna elements, which is a significant computational burden for massive MIMO systems [3].

In recent years, the signal detection method based on the belief propagation (BP) algorithm (called BP detection) has been paid attentions [4]–[6]. BP is an algorithm to calculate marginal probability of unobserved variables efficiently by message passing in a factor graph.

It has been reported that BP detection exhibits near-optimal detection performance by exchanging and updating log-likelihood ratio (LLR) of each symbol as message [4]–[6]. Moreover, BP detection is matrix-inversion free, which makes it very attractive for massive MIMO systems. However, there is a problem that the convergence property and the detection performance degrade due to multiple loops included in the MIMO channel as the nature of the factor graph.

The damped BP that averages two successive messages with a weighting factor (called damping factor) is known as a method to improve the convergence property of BP [7]. However, calculating the optimal damping factors at the time of detection results in high computational burden. As a method to solve this issue, deep neural network-based damped BP (DNN-dBP) has been reported [8]. In DNN-dBP, the damping factors can be trained offline for each antenna configuration. The problem with DNN-dBP is that the detection performance deteriorates when there is a difference of the channel correlation between training and test. This is because the optimal damping factors vary with the channel correlation [6]. However, training for each channel correlation requires enormous time and parameters, which is not practical. As a method to mitigate the effects of the channel correlation, the node selection (NS) method has been reported [3]. NS is the method that selects the nodes to be updated in BP such that the selected nodes have low spatial correlation.

In this paper, to solve the problem of the conventional DNN-dBP, we derive the damping factors of BP to which the NS method is applied using DNN-dBP. When the NS method is applied to BP, the channel correlation among the selected nodes in BP detection is lowered. Therefore, the proposed method can improve the detection performance degradation due to the mismatches of the channel correlations between training and test in DNN-dBP. This means that the proposed method does not need to train for each channel correlation. A part of this derivation method of damping factors was presented in [9]. This paper explains this method in more detail. In addition there are two differences between this paper and [9]. First, this paper evaluates the computational complexity of the proposed method and compares it with that of the conventional DNN-dBP. The convergence property of BP is improved by applying the NS method, because the number of loops in the MIMO channel is reduced. Therefore, the proposed method has the same detection performance with low computational complexity as the conventional DNN-dBP. Second, this paper investigates the distribution of the trained damping factors to examine whether the trained damping factors have regularity. By computer simulation, it is shown that the proposed method significantly improves the bit error rate (BER) performance degradation due to the mismatches of the channel correlations between training and test in DNN-dBP. Although the trained damping factors cannot guarantee the optimality, the results show that the proposed method improves the BER performance compared to BP without damping. It is also shown that the proposed method has the same BER performance with low computational complexity as the conventional DNN-dBP. Moreover, we evaluate the distribution of the trained damping factors.

**FIGURE 1. System model.**

### II. SYSTEM MODEL

In a massive MIMO system consisting of $M$ user terminals each with a single antenna and a single BS with $N$ receive antennas, we consider uplink communication where each user transmits data independently (Fig. 1). The received signal $r_j$ at the receive antenna $j$ is given by the following equation.

$$r_j = \sum_{i=1}^{M} h_{ij} x_i + n_j,$$

(1)

where $h_{ij}$ is the channel response between the user terminal $i$ and the receive antenna $j$, and $x_i$ is the transmitted symbol of the user terminal $i$. Also, $n_j$ denotes additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2$ at the receive antenna $j$. The transmitted symbol $x_i$ is selected from the possible symbols $\{s_0, \ldots, s_{Q-1}\}$ for a modulation level $Q$. The channel matrix is known at the BS, and our problem is to estimate the transmitted symbols $\{x_1, \ldots, x_M\}$ from the received signals $\{r_1, \ldots, r_N\}$.

The MIMO channel can be represented by a factor graph as shown in Fig. 2. The transmit nodes and the receive nodes are referred to as symbol nodes and observation nodes, respectively.

### III. SIGNAL DETECTION USING BP

This section briefly explains the algorithm of BP detection. In BP detection, transmitted symbols are estimated by
exchanging and updating a priori information and a posteriori information of transmitted symbols among symbol nodes and observation nodes. The a priori LLR of $x_i$ being passed from the symbol node $j$ to the observation node $i$ at iteration $l$ is represented by $\alpha^{(l)}_{i \rightarrow j} = [\alpha^{(l)}_{i \rightarrow j}(s_1), \ldots, \alpha^{(l)}_{i \rightarrow j}(s_{Q-1})]$, the a priori probability of $x_i$ by $p^{(l)}_{i \rightarrow j} = [p^{(l)}_{i \rightarrow j}(s_1), \ldots, p^{(l)}_{i \rightarrow j}(s_{Q-1})]$, the a posteriori LLR of $x_i$ being passed from the observation node $j$ to the symbol node $i$ at iteration $l$ by $\beta^{(l)}_{j \rightarrow i} = [\beta^{(l)}_{j \rightarrow i}(s_1), \ldots, \beta^{(l)}_{j \rightarrow i}(s_{Q-1})]$. $\alpha^{(l)}_{i \rightarrow j}(s_k)$ and $\beta^{(l)}_{j \rightarrow i}(s_k)$ are given by the following equations.

$$\alpha^{(l)}_{i \rightarrow j}(s_k) = \ln \frac{p^{(l)}(x_i = s_k)}{p^{(l)}(x_i = s_0)}.$$  

$$\beta^{(l)}_{i \rightarrow j}(s_k) = \ln \frac{p^{(l)}(r_i | x_i = s_k)}{p^{(l)}(r_i | x_i = s_0)}.$$  

(A. PROCESSING AT OBSERVATION NODES)

In the observation nodes, the a posteriori information of the transmitted symbols is updated based on the message (a priori information of the transmitted symbol) passed from the connected symbol nodes.

The received signal $r_j$ at the receive antenna $j$ can be divided into a desired signal and interference signals as follows.

$$r_j = h_{j,i}x_i + \sum_{k=1,k \neq i}^{N} h_{j,k}x_k + n_j = h_{j,i}x_i + z_{j,i},$$  

where $z_{j,i}$ denotes the interference term following the Gaussian distribution with mean $\mu_{z_{j,i}}$ and variance $\sigma_{z_{j,i}}^2$.

$$\mu_{z_{j,i}} = \sum_{k=1,k \neq i}^{M} h_{j,k}E\{x_k\},$$  

$$\sigma_{z_{j,i}}^2 = \sum_{k=1,k \neq i}^{M} h_{j,k}^2 \text{Var}\{x_k\} + \sigma^2.$$  

$E\{x_k\}$ and $\text{Var}\{x_k\}$ represent the mean value and variance of the transmitted symbol $x_k$ respectively, and are calculated from the message $p_{x_i \rightarrow r_j}$ from the symbol node. Here, $p(r_j | x_i)$ is given by the following equation.

$$p(r_j | x_i) = \frac{1}{\pi \sigma_{z_{j,i}}^2} \exp \left[ \frac{-|r_j - \mu_{z_{j,i}} - h_{j,i}x_i|^2}{\sigma_{z_{j,i}}^2} \right].$$  

From Eqs. (3) and (7), the message passed from the observation node $j$ to the symbol node $i$ at iteration $l$, that is, $\beta^{(l)}_{j \rightarrow i}(s_k)$ can be represented by the following equation.

$$\beta^{(l)}_{j \rightarrow i}(s_k) = \frac{(r_j - \mu_{z_{j,i}} - h_{j,i}x_i)^2}{2\sigma_{z_{j,i}}^2} - \frac{(r_j - \mu_{z_{j,i}} - h_{j,i}x_i)^2}{2\sigma_{z_{j,i}}^2} + 2h_{j,i}(r_j - \mu_{z_{j,i}} - h_{j,i}x_i)(s_k - s_0) - h_{j,i}(s_k - s_0)^2. \tag{8}$$

(B. PROCESSING AT SYMBOL NODES)

At the symbol nodes, the a priori information of the transmitted symbols is updated based on the message (a posteriori information of the transmitted symbol) passed from the connected observation nodes.

The LLR of each symbol at the symbol node $i$ at iteration $l$ is given by the following equation.

$$\gamma^{(l)}(x_i = s_k) = \sum_{i=1}^{N} \beta^{(l)}_{i \rightarrow j}(s_k). \tag{9}$$

Also, the message passed from the symbol node $i$ to the observation node $j$, that is, $\alpha^{(l)}_{i \rightarrow j}(s_k)$ can be represented by the following equation.

$$\alpha^{(l)}_{i \rightarrow j}(s_k) = \sum_{i=1, i \neq j}^{N} \beta^{(l)}_{i \rightarrow j}(s_k). \tag{10}$$

From Eq. (2), the a priori probability $p^{(l)}_{i \rightarrow j}$ can be expressed by the following equation.

$$p^{(l)}_{x_i \rightarrow r_j} = \frac{\exp(\alpha^{(l)}_{i \rightarrow j}(s_k))}{\sum_{m=0}^{Q-1} \exp(\alpha^{(l)}_{i \rightarrow j}(s_m))}.$$

(C. TRANSMITTED SYMBOLS ESTIMATION)

After $L$ iterations, the symbol $s_k$ is determined to maximize $\gamma^{(L)}(x_i = s_k)$ at each symbol node.

(IV. RELATED RESEARCH ON BP DETECTION)

One of the problems with BP detection is that in the factor graph of the MIMO channel model, all the nodes are connected and many loops are included. It is known that the convergence property and the detection performance of BP degrades when loops exist in the factor graph [10]. In massive MIMO system, the randomness of the connectivity factor (channel response) is improved by the increase of the number of nodes, and the convergence is sufficiently achieved by reducing short loops. However, since the randomness of the connectivity factor is degraded in the correlated channel, the convergence property is also significantly degraded [3].

(A. DAMPED BP)

The damped BP is known as an effective method to improve the degradation of the convergence property due to loops [7]. This method improves the convergence property by averaging...
two successive messages with a damping factor. The message \(p^{(l)}_{ij}\) from the symbol node \(i\) to the observation node \(j\) at iteration \(l\) is expressed as follows, using the message at iteration \(l - 1\).

\[
p^{(l)}_{ij} = (1 - \delta) p^{(l)}_{ij} + \delta p^{(l-1)}_{ij},
\]

where \(\delta\) indicates a damping factor between 0 and 1 inclusive. This makes it possible to improve the detection performance degradation due to the influence of loops in BP detection.

However, it is difficult to find the optimal damping factor. This is because calculating the optimal \(\delta\) at each iteration on-line at time of the detection results in increased computational complexity.

\[
\text{FIGURE 3. The structure of DNN-dBP.}
\]

B. DNN-dBP

Eq. (12) is converted as follows, considering that we determine the damping factor for each iteration and message to improve the convergence property of the damped BP [8].

\[
p^{(l)}_{ij} = (1 - \delta^{(l)}_{ij}) p^{(l)}_{ij} + \delta^{(l)}_{ij} p^{(l-1)}_{ij}.
\]

However, calculating the optimal damping factors \(\delta = \{\delta^{(l)}_{ij}\}\) online results in high computational burden. As a method to solve this issue, deep neural network-based damped BP (DNN-dBP) has been reported [8]. Here, \(\delta\) is trained off-line by replacing BP’s iterative algorithm with a neural network structure. By applying the BP algorithm using the trained damping factors at the detection, it is possible to improve the convergence property without increasing the computational complexity online. The structure of DNN-dBP is shown in Fig. 3. The hidden layers are divided into the layers corresponding to the symbol nodes and those corresponding to the observation nodes, which are referred to as symbol layers and observation layers, respectively. The numbers of symbol layers and observation layers correspond to that of BP iterations, \(L\). The numbers of neurons in the input layer and the output layer are \(M\). The number of neurons in the symbol layers is \(M\), and that in the observation layers is \(N\). In BP detection, \(\beta^{(l)}\) is updated at the observation node, using messages \(\alpha^{(l-1)}\) and \(p^{(l-1)}\) of one iteration before. Then, \(\alpha^{(l)}\) and \(p^{(l)}\) are updated from \(\beta^{(l)}\) at the symbol node in iteration \(l\). In DNN-dBP, the processing in BP iterations is replaced with the calculation in each hidden layer of the neural network.

The damping factor \(\delta\) is trained in the following procedure.

(1) Randomly generate the transmitted symbol \(x\), and the channel matrix \(H\) and calculate the received signal \(r\).

(2) Input the initial value of a priori LLR of each symbol \(\alpha^{(0)} = 0\) and the initial value of the damping factor \(\delta\). The initial value of \(\{\delta^{(l)}_{ij}\}\) is set to 0.5.

(3) In each hidden layer, the messages are updated according to the damped BP algorithm. The processing in each layer is represented by the following equations.

\[
\begin{align*}
\beta^{(l)} &= f^{(l)}_o(\alpha^{(l-1)}), \\
\beta^{(l)} &= f^{(l)}_s(\beta^{(l)}; \delta^{(l)}), \\
y_i &= \sum_{l=1}^{N} \beta^{(l)}_{il}, \\
o_i &= \varphi(y_i).
\end{align*}
\]

\(f^{(l)}_o\) and \(f^{(l)}_s\) represent the mappings of the inputs to the outputs in the \(l\)th observation layer and symbol layer, respectively. In the \(l\)th observation layer, \(\beta^{(l)}\) is output from the inputs \(\{\alpha^{(l-1)}, p^{(l-1)}\}\) based on Eq. (8). In the \(l\)th symbol layer, \(\{\alpha^{(l)}, p^{(l)}\}\) are output from the inputs \(\beta^{(l)}\) and the training parameter \(\delta^{(l)}\) based on Eqs. (10), (11), (13). \(\varphi\) is a sigmoid function, and the output \(O\) is obtained from \(y\) which is the LLR of each transmitted symbol calculated based on Eq. (9) in the \(l\)th symbol layer.

(4) The cross entropy is adopted to express the loss function. The loss function \(L(X, O)\) is defined as follows using the output \(O\) and the one-hot vector \(X\) generated from the known transmitted symbol vector.

\[
L(X, O) = -\frac{1}{M} \sum_{i=1}^{M} \sum_{k=0}^{Q-1} X_i(s_k) \log (O_i(s_k)),
\]

where \(X\) is given by the following equation.

\[
X : X_i(s_k) = \begin{cases} 
1 & x_i = s_k \\
0 & x_i \neq s_k.
\end{cases}
\]

Since the value of \(L(X, O)\) decreases as \(O\) approaches the value of the expected output, \(\delta\) is updated and trained to minimize \(L(X, O)\) based on the stochastic gradient descent.

\(\delta\) is trained for each antenna configuration \((M, N)\), and during test, the received signals are detected using the trained damping factors.

C. NODE SELECTION (NS) METHOD

Node selection (NS) has been reported as a method to reduce the channel correlation among the nodes to be updated in BP detection [3]. Normal BP updates the message \(\beta\) based on Eq. (8) at all observation nodes at each iteration. However, when the number of observation nodes is large enough, it is known that sufficient detection reliability can be expected by updating LLRs at not all but some observation nodes [3]. In the NS method, short loops are reduced and the randomness of the connectivity factors is improved by selecting...
observation nodes to be updated with low spatial correlation. Therefore, by applying the NS method, the convergence property of BP is improved and the channel correlation among the selected nodes in BP detection is lowered. The number of selected nodes is $R$, and the indexes of the selected and non-selected nodes are $g(r) \ (r = 1, \ldots, R)$, $g(s) \ (s = 1, \ldots, N - R)$, respectively. The message $\alpha^{(l)}_{i,j}$ from the symbol node $i$ to the observation node $j$ at iteration $l$ is given by the following equation.

$$\alpha^{(l)}_{i,j} = \sum_{r=1, r \neq j}^{R} p^{(l)}_{g(r),i} + \sum_{s=1, s \neq j}^{N-R} p^{(l-1)}_{g(s),i}. \quad (17)$$

As a selection rule of nodes, we select observation nodes to be updated at interval $K$ ($K = N/R$) to reduce the spatial correlation of the selected nodes. Then, we shift the selected nodes by one node at each iteration.

V. PROPOSED METHOD

In the DNN-dBP, the damping factors are trained for each antenna configuration ($M, N$), and the received signals are detected using the trained damping factors. However, in the correlated channel, even for the same antenna configuration, the optimal damping factor values differ depending on the magnitude of the channel correlation. Therefore, the detection performance deteriorates when there is a difference of the channel correlation between training and test. However, training for each channel correlation requires enormous time and parameters, which is not practical. To lower the channel correlation among the nodes to be updated in BP detection, we focus on the NS method. In this paper, to solve the issue of the conventional DNN-dBP, we derive the damping factors of BP to which the NS method is applied using DNN-dBP.

A. THE STRUCTURE OF DNN-dBP IN THE PROPOSED METHOD

When we apply message damping to BP with the NS method, it can be realized by converting the message $p^{(l)}_{g(r),i}$ in each iteration into that denoted by Eq. (13) as in the normal damped BP. In addition, we derive the damping factors in this case using DNN-dBP. The structure of DNN-dBP in the proposed method with the node selection interval $K = 2$ is shown in Fig. 4. The neural network structure is an expansion of the iterative algorithm of BP to which the NS method is applied. The training algorithm is the same as that of the conventional DNN-dBP, but the mapping function of Eq. (14) is changed. In the hidden layer, the messages are passed only to the neurons corresponding to the selected nodes. Therefore, only $\beta^{(l)}$ corresponding to the selected nodes index $g(r)$ is updated in the mapping function $f^{(l)}_o$. The neurons corresponding to the unselected nodes pass the messages from neurons in the previous observation layer. Therefore, the equation that updates $\alpha^{(l)}$ in the mapping function $f^{(l)}_s$ is converted to Eq. (17).

$$\{\alpha^{(l)}, p^{(l)}\} = f^{(l)}_s (\beta^{(l)}_{g(r)}, \beta^{(l-1)}_{g(s)}, \delta^{(l)}). \quad (18)$$

In detection, the damping factors trained for each antenna configuration are used in the damped BP to which the NS method is applied with the same node selection interval as in training.

In the conventional DNN-dBP, the detection performance degrades when there is a difference of the channel correlation between training and test. By applying the NS method, the channel correlation among the selected nodes in BP detection is lowered. Thus, the proposed method improves the detection performance degradation due to the mismatches of the channel correlations between training and test in DNN-dBP. This means that the proposed method does not need to train for each channel correlation. In addition, the convergence property of BP is improved by applying the NS method. Therefore, the proposed method has the same detection performance with low computational complexity as the conventional DNN-dBP.

B. EFFECT ON COMPUTATIONAL COMPLEXITY BY APPLYING THE NS METHOD

In the normal BP to which the NS method is not applied, all variable nodes and observation nodes are connected. Then, a posteriori LLR $\beta$ is updated at all observation nodes based on the messages passed from the variable nodes. On the other hand, in BP to which the NS method is applied, at each iteration, the messages are passed from the variable nodes to only the selected observation nodes, and a posteriori LLR is updated. From the unselected observation node, the a posteriori LLR one iteration before is passed to the variable node. Therefore, if the node selection interval is $K$, the number of the a posteriori LLRs updated at each iteration is $1/K$ of that when all variable nodes and observation nodes are connected. When the amount of computational complexity of BP is calculated based on the number of multiplication, the computational complexity of BP applying the NS method with the node selection interval $K$ per iteration is $1/K$ of that of BP not applying the NS method.

BP with the NS method requires more iterations to have the same detection performance as the normal BP.
However, by applying the NS method, the convergence property of BP is improved, because the effects of the MIMO channel loops are reduced. Therefore, the proposed method shows the same detection performance with low computational complexity as the conventional DNN-dBP.

VI. SIMULATION RESULTS
A. SIMULATION PARAMETERS
In this paper, we use the Kronecker model for the correlated channel matrices [11]. For the correlation factor $\rho$ ($0 \leq \rho \leq 1$), the correlation matrix at the transmitter side is $R_t$, the correlation matrix at the receiver side is $R_r$. The correlated channel matrix can be described as $R_t^j H R_r^j$, where $H$ is the i.i.d. Rayleigh-fading channel matrix following independent Gaussian distribution. The correlation factor $\rho$ is an index that indicates the strength of the correlation between adjacent antennas.

Table 1 lists the simulation parameters. “w/o NS” indicates the case where NS is not applied to BP, and “w/ NS” indicates the case where NS is applied. Both the number of transmit antennas and that of receive antennas are 16. QPSK is used as the modulation scheme. The number of BP iterations is determined so that the detection performance converges by computer simulation [4]. When NS is not applied, the node selection interval is $K = 1$; when NS is applied, the node selection interval is $K = 4$. The node selection interval in “w/ NS” is determined to sufficiently mitigate the effects of the channel correlation without damping by computer simulation. The computational complexity of BP is proportional to $L$. However, compared with the computational complexity in “w/o NS” per iteration, that in “w/ NS” per iteration is $1/K$. We simulate in the correlated channels with the correlation factor $\rho = 0.3$. Also, no channel coding is considered.

Next, we describe the details of training in DNN-dBP. As an optimization algorithm, Adam [12] which is one of the stochastic gradient descent is used. The SNR per receive antenna used for training is 0 dB to 20 dB (every 5 dB), and the damping factors are trained using 18000 randomly generated data samples (3600 for each SNR value). In this paper, we evaluate the influence of difference of the channel correlation between training and test. When training the damping factors for simulation in correlated channels, we consider the case where the correlation factors match (trained with $\rho = 0.3$) and do not match (trained with $\rho = 0.0$).

B. COMPARISON OF EFFECTS OF DIFFERENCE OF CHANNEL CORRELATIONS BETWEEN TRAINING AND TEST
Fig. 5 shows the BER performance of each BP detection in the correlated channel ($\rho = 0.3$). The horizontal axis shows SNR (dB) per receive antenna and the vertical axis shows BER. “w/o NS, w/ DNN-dBP” represents the conventional DNN-dBP with the node selection interval $K = 1$ and the number of iterations $L = 7$. “w/ NS, w/ DNN-dBP” represents the proposed method with the node selection interval $K = 4$ and the number of iterations $L = 28$. The computational complexity of BP is proportional to the number of iterations and inversely proportional to the node selection interval. Therefore, the computational complexities in “w/o NS” and “w/ NS” are almost the same in this case. From the figure, it can be seen that the BER performance is improved by applying NS. However, without damping, the BER performance is degraded at high SNR. By applying DNN-dBP, the BER degradation at high SNR can be improved. Comparing the cases when the correlation factors match ($\rho = 0.3$) and not match ($\rho = 0.0$) between training and test, in the conventional DNN-dBP, the BER performance when the correlation factors do not match degrades compared to that when they match. This is because the optimal damping factors vary with the channel correlation, and the detection performance deteriorates due to the mismatches of the channel correlations in the conventional DNN-dBP. However, in the proposed method, the case when the correlation factors do not match shows the BER performance almost equal to that in the case when they match. This means that the proposed method significantly improves the detection performance deterioration due to the mismatches of the channel correlation between training and test in DNN-dBP. This is because the channel correlation among the selected nodes in BP detection is lowered by applying NS.
C. THE COMPARISON OF THE COMPUTATIONAL COMPLEXITY

We evaluate the effects of the computational complexity of BP by applying the NS method. Fig. 6 shows the BER performance of each BP detection in the correlated channel ($\rho = 0.3$). On the basis of the BER performance of the conventional DNN-dBP, we compare the computational complexity of the proposed method with the number of iterations $L$ is 12, 16, and 20, respectively. The proposed method with $L = 16$ has the same BER performance as the conventional DNN-dBP with $L = 7$. Since the computational complexity of BP is proportional to the number of iterations and inversely proportional to the node selection interval, it can be seen that the proposed method shows similar BER performance with $4/7$ computational complexity compared to the conventional DNN-dBP. Although the proposed method has offline computational burden by training damping factors, it can reduce online computational complexity of BP because the convergence property is improved by applying the NS method.

D. DISTRIBUTION OF THE TRAINED DAMPING FACTORS

In this paper, we derived the damping factors for each iteration and each message in BP detection using DNN-dBP. To examine whether the distribution of the trained damping factors have regularity, we calculate the average value of the trained damping factors for all messages at each iteration. Fig. 7 shows the average value of the damping factors for all messages at each iteration in “w/o NS” ($K = 1, L = 7$), and Fig. 8 shows that in “w/ NS” ($K = 4, L = 28$). Since the message damping is represented by Eq. (13), the larger the damping factor, the more the message one iteration before is emphasized.

From Fig. 7, the average value of the damping factors for each message decreases as the number of BP iterations increases. This is because the difference between successive messages becomes smaller as the message converges, so the convergence of BP is accelerated by emphasizing the current messages in Eq. (13). Also, the average value of the damping factors is larger when training at $\rho = 0.3$ than that when training at $\rho = 0.0$. This is because the larger the channel correlation, the more likely the message diverges, and the convergence of BP is accelerated by emphasizing the messages one iteration before in Eq. (13).

From Fig. 8, even when the NS method is applied, the average value of the damping factors for each message decreases as the number of BP iterations increases, and converges about 20 iterations. In addition, the difference of the average value of the damping factors between the case of training at $\rho = 0.0$ and the case of training at $\rho = 0.3$ is smaller than when the NS method is not applied. This is because the effects of the channel correlation are reduced by applying the NS method, and as a result, the difference in the magnitude of the damping factors for improving the convergence property of BP is reduced.

VII. CONCLUSION

In this paper, we focused on belief propagation (BP) detection as a matrix-inversion free and near-optimal detection method for massive multiple-input multiple-output (MIMO) systems. The deterioration of the convergence property and the
detection performance of BP due to multiple loops included in a MIMO channel can be reduced by message damping. As a method to train the damping factors off-line, deep neural network-based damped BP (DNN-dBP) has been reported. However, the optimal damping factors vary with the channel correlation, and training for each channel correlation requires enormous time and parameters. To solve this issue, we derived the damping factors of BP to which the node selection (NS) method is applied using DNN-dBP. In DNN-dBP, the detection performance deteriorates when there is a difference of the channel correlation between training and test. By applying the NS method, the channel correlation among the selected nodes in BP detection is lowered significantly. We unfolded the iteration algorithm of BP to which the NS method is applied into the neural network structure to train the damping factors. Then, the damping factors trained for each antenna configuration are used in the damped BP to which the NS method is applied with the same node selection interval as in training. By computer simulation, it was shown that the proposed method significantly improves the BER performance deterioration due to the mismatches of the channel correlations between training and test in DNN-dBP. The results also showed that the proposed method has the same detection performance with low computational complexity as the conventional DNN-dBP. From the above, we got the results that the proposed method improves the detection performance compared to the conventional DNN-dBP in channels with various correlation values with one off-line training for each antenna configuration. We also investigated the distribution of the trained damping factors and evaluated the tendency of that.

REFERENCES

[1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 742–758, Oct. 2014.

[2] X. Zhu and R. D. Murch, “Performance analysis of maximum likelihood detection in a MIMO antenna system,” IEEE Trans. Commun., vol. 50, no. 2, pp. 187–191, Feb. 2002.

[3] W. Fukuda, T. Abiko, T. Nishimura, T. Ohgane, Y. Ogawa, Y. Ohwatarai, and Y. Kishiyama, “Low-complexity detection based on belief propagation in a massive MIMO system,” in Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring), Dresden, Germany, Jun. 2013, pp. 1–5.

[4] J. Yang, C. Zhang, X. Liang, S. Xu, and X. You, “Improved symbol-based belief propagation detection for large-scale MIMO,” in Proc. IEEE Workshop Signal Process. Syst. (SiPS), Oct. 2015, pp. 1–6.

[5] F. Long, T. Lv, R. Cao, and H. Gao, “Single edge based belief propagation algorithms for MIMO detection,” in Proc. 34th IEEE Sarnoff Symp., Princeton, NJ, USA, May 2011, pp. 1–5.

[6] Y. Gao, H. Niu, and T. Kaiser, “Massive MIMO detection based on belief propagation in spatially correlated channels,” in Proc. IEEE 11th Int. ITG Conf. Syst. Commun. Coding, Feb. 2017, pp. 1–6.

[7] P. Som, T. Datta, A. Chockalingam, and B. S. Rajan, “Improved large-MIMO detection based on damped belief propagation,” in Proc. IEEE Inf. Theory Workshop (ITW), Jan. 2010, pp. 1–15.

[8] X. Tan, W. Xu, K. Sun, Y. Xu, Y. Be’ery, X. You, and C. Zhang, “Improving massive MIMO message passing detectors with deep neural network,” IEEE Trans. Veh. Technol., vol. 69, no. 2, pp. 1267–1280, Feb. 2020.

[9] J. Tachibana and T. Ohtsuki, “Damping factor learning of bp detection with node selection in massive MIMO using neural network,” in Proc. IEEE Veh. Technol. Conf. (VTC-Spring) Antwerp, Belgium, May 2020.

[10] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[11] M. Chiani, M. Z. Win, and A. Zanella, “On the capacity of spatially correlated MIMO Rayleigh-fading channels,” IEEE Trans. Inf. Theory, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

[12] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Represent., Dec. 2015, pp. 1–15.

[13] N. Samuel, T. Diskin, and A. Wiesel, “Deep MIMO detection,” in Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Jul. 2017, pp. 690–694.