Joint Maximum Likelihood Detection in Far User of Non-Orthogonal Multiple Access*  

Kenji ANDO†a, Student Member, Yukitoshi SANADA†b, and Takahiko SABA††c, Senior Members

SUMMARY Non-orthogonal multiple access (NOMA) enables multiple mobile devices to share the same frequency band. In a conventional NOMA scheme, the receiver of a far user detects its desired signal without canceling the signal for a near user. However, the signal for the near user acts as interference and degrades the accuracy of likelihood values for the far user. In this paper, a joint maximum likelihood detection scheme for the far user of the NOMA downlink is proposed. The proposed scheme takes the interference signal into account in calculating the likelihood values. Numerical results obtained through computer simulation show that the proposed scheme improves the performance by from 0.2 dB to 3.1 dB for power allocation coefficients of 0.2 to 0.4 at a bit error rate (BER) of $10^{-2}$ relative to the conventional scheme.

key words: non-orthogonal multiple access, joint maximum likelihood detection, successive interference cancelation

1. Introduction

Due to the rise in the penetration rate of mobile devices such as smartphones and tablet PC, demands for higher data rates and larger capacity have been increasing in cellular systems. The report released by ITU-R has shown that the amount of mobile data traffic will increase 1000 times by 2020 [1]. Technical solutions to resolve the problems caused by this explosive traffic growth have been investigated, and candidate system concepts for future radio access beyond LTE-Advanced have been discussed. One solution for increasing the capacity is non-orthogonal multiple access (NOMA) [2]–[16]. NOMA enables multiple mobile users to share the same frequency band. By sending signals with different transmission powers, a receiver can extract its desired signal. It has been shown that the total channel capacity increases if a receiver can remove the overlaid interference signal [17].

1.1 Non-Orthogonal Multiple Access Model

In NOMA, mobile users that are located near and far from a base station are assigned the same frequency resource and allowed to share it through a scheduling algorithm. Since the amount of propagation loss depends on the distance from the base station, the base station transmits the signal for the far user with larger power than the signal for the near user.

The receiver of the near user cancels the signal for the far user by means of successive interference cancelation (SIC) and then detects its desired signal [6]–[15]. On the other hand, the receiver of the far user demodulates the signal without canceling the signal for the near user and employs maximum likelihood detection (MLD). This is because the signal for the near user is expected to be attenuated and it is not regarded as interference by the receiver of the far user.

However, the signal for the near user actually interferes with the signal for the far user and degrades the accuracy of likelihood values in the MLD. Thus, a joint detection scheme in the receiver of the far user has been proposed [16]. In [16], the sum of the constellation constrained capacities (CCCs) of a NOMA downlink with the joint detection scheme in both the near and the far users is derived. In deriving the CCCs, the signal for the near user is treated as probabilistic interference. However, the derivation of CCCs does not assume the superposition of the constellation points of the signals for the near and the far users. Thus, this paper investigates direct effects of the joint MLD in the far user of the NOMA system on the link-level performance with specific settings of transmission parameters including modulation schemes, coding rates, and power allocation coefficients. These numerical results show that the joint MLD in the far user of the NOMA system works effectively with various transmission parameter values.

This paper is organized as follows. Section 2 describes the system model and the proposed scheme. In Sect. 3, the bit error rate (BER) curves obtained through computer simulation are presented. Sect. 4 gives our conclusions.

2. System Model

2.1 Non-Orthogonal Multiple Access Model

In this paper, a NOMA downlink is assumed as a system model, in which two mobile users share the same frequency band. A base station and each of mobile terminals are equipped with a single antenna. A block diagram of the base station is presented in Fig. 1. In the LTE standard, a systematic parallel concatenated convolutional code is adopted [18]. In a turbo encoder, two 8-state encoders and one interleaver are used. With the input of $N_c$ bits, the turbo encoder generates three length-$N_c$ streams, $d_n^{(0)}$, $d_n^{(1)}$, and $d_n^{(2)}$. They are referred as the “Systematic”, “Parity 1”, and “Parity 2” streams, respectively. At the end of each stream, four tail bits for trellis termination are appended.

Manuscript received March 11, 2016.  
Manuscript revised June 17, 2016.  
Manuscript publicized July 29, 2016.  
†The authors are with the Dept. of Electronics and Electrical Engineering, Keio University, Yokohama-shi, 223-8522 Japan.  
††The author is with the Dept. of Computer Science, Chiba Institute of Technology, Narashino-shi, 275-0016 Japan.  
*The part of this paper is submitted to IEEE APWCS 2015.  
a) E-mail: kando@snd.elec.keio.ac.jp  
b) E-mail: sanada@elec.keio.ac.jp  
c) E-mail: saba@cs.it-chiba.ac.jp  
DOI: 10.1587/transcom.2016EBP3098
In order to support a higher data rate, bit puncturing is implemented in the turbo codes. In the LTE standard, puncturing is conducted in a rate matching (RM) block [18]. In the RM block, each of the three output streams is rearranged in a sub-block interleaver. Afterward, a single output buffer with the length of $3(N_s + 4)$ is filled by placing the rearranged “Systematic” bit stream in the beginning, followed by the two rearranged bit streams of “Parity 1” and “Parity 2” interlaced in a bit-by-bit fashion. The contents of the buffer are then passed to a circular buffer for bit selection and puncturing. The output of the circular buffer forms a single bit stream with the length of $3(N_s + 4)$. After bit puncturing, the RM block generates a single bit stream with the length of $(N_s + 4)/r$, where $r$ is the code rate and the minimum of $r$ is 1/3. The outputs of the RM blocks for the near and the far users are mapped to $2^{M_n} QAM$ and $2^{M_f} QAM$ symbols through Gray coding, where $M_n$ and $M_f$ are the numbers of bits per symbol of the near and the far users, respectively.

After symbol mapping, they are put into serial-to-parallel (S/P) converters and summed together on each subcarrier. Here, $S_n[k]$ and $S_f[k]$ are the symbols to be transmitted for the near and the far users on the $k$th subcarrier. The transmit power is divided with the coefficients of $\alpha$ and $(1 - \alpha)$. The symbol to be transmitted on the $k$th subcarrier is composed of the sum of the symbols of the near and the far users as follows:

$$S[k] = \sqrt{\alpha}S_n[k] + \sqrt{1 - \alpha}S_f[k] \quad (0 \leq k \leq N - 1),$$

where $N$ is the size of the IDFT. After the IDFT and the parallel-to-serial (P/S) conversion, the non-orthogonal OFDM signal is generated. The last part of each OFDM symbol is replicated and inserted at the beginning of the symbol as a guard interval (GI). The generated OFDM signal is put into a transmit filter and is transmitted. At the receiver side, the received signal first goes through a receive filter. In an analog-to-digital (A/D) converter, the received signal is digitized to discrete samples. The signal on the $k$th subcarrier after the removal of the GI and the $N$-sample DFT is given as

$$Y[k] = \sum_{n=0}^{N-1} y[n] \exp\left(-\frac{2\pi nk}{N}\right) = H[k]S[k] + W[k],$$

where $y[n]$ is the $n$th received signal in the time domain, $Y[k]$ is the received signal, $H[k]$ is the frequency response, and $W[k]$ is the noise in the frequency domain on the $k$th subcarrier, respectively. The output of the DFT on the $k$th subcarrier is sent to a detector to obtain soft information for a turbo decoder.

At the detector, the likelihood values for turbo decoding are calculated. After the likelihood calculation in the detector, bit depuncturing is carried out. The likelihood sequence is then passed to an inverse circular buffer. Afterward, the likelihood sequence is divided into three streams that correspond to “Systematic”, “Parity 1”, and “Parity 2” streams, respectively. Each of the three output streams is rearranged in a sub-block deinterleaver and is then put into the turbo decoder to obtain information bits.

2.2 Conventional Detection Scheme

In the NOMA system, the signals for the near and the far users are superposed on each subcarrier. Since the signal power of the far user is larger than that of the near user, the conventional detection scheme in the far user employs MLD. In the conventional detection scheme, the signal for the near user is expected to be attenuated and it is not regarded as interference by the receiver of the far user.

The block diagram of the conventional detection scheme in the far user is shown in Fig. 2. The likelihood calculations by the MLD are given as

$$\Delta_{f,1}^{MLD}[k] = \sum_{S_f[k] \in \{S_f\}'} \exp\left(-\frac{1}{N_0} \|Y[k]\right) - H[k]\sqrt{1 - \alpha S_f[k]^2}),$$

where $y[n]$ is the $n$th received signal in the time domain, $Y[k]$ is the received signal, $H[k]$ is the frequency response, and $W[k]$ is the noise in the frequency domain on the $k$th subcarrier, respectively. The output of the DFT on the $k$th subcarrier is sent to a detector to obtain soft information for a turbo decoder.

At the detector, the likelihood values for turbo decoding are calculated. After the likelihood calculation in the detector, bit depuncturing is carried out. The likelihood sequence is then passed to an inverse circular buffer. Afterward, the likelihood sequence is divided into three streams that correspond to “Systematic”, “Parity 1”, and “Parity 2” streams, respectively. Each of the three output streams is rearranged in a sub-block deinterleaver and is then put into the turbo decoder to obtain information bits.

2.2 Conventional Detection Scheme

In the NOMA system, the signals for the near and the far users are superposed on each subcarrier. Since the signal power of the far user is larger than that of the near user, the conventional detection scheme in the far user employs MLD. In the conventional detection scheme, the signal for the near user is expected to be attenuated and it is not regarded as interference by the receiver of the far user.

The block diagram of the conventional detection scheme in the far user is shown in Fig. 2. The likelihood calculations by the MLD are given as

$$\Delta_{f,1}^{MLD}[k] = \sum_{S_f[k] \in \{S_f\}'} \exp\left(-\frac{1}{N_0} \|Y[k]\right) - H[k]\sqrt{1 - \alpha S_f[k]^2}),$$

where $y[n]$ is the $n$th received signal in the time domain, $Y[k]$ is the received signal, $H[k]$ is the frequency response, and $W[k]$ is the noise in the frequency domain on the $k$th subcarrier, respectively. The output of the DFT on the $k$th subcarrier is sent to a detector to obtain soft information for a turbo decoder.

At the detector, the likelihood values for turbo decoding are calculated. After the likelihood calculation in the detector, bit depuncturing is carried out. The likelihood sequence is then passed to an inverse circular buffer. Afterward, the likelihood sequence is divided into three streams that correspond to “Systematic”, “Parity 1”, and “Parity 2” streams, respectively. Each of the three output streams is rearranged in a sub-block deinterleaver and is then put into the turbo decoder to obtain information bits.

2.2 Conventional Detection Scheme

In the NOMA system, the signals for the near and the far users are superposed on each subcarrier. Since the signal power of the far user is larger than that of the near user, the conventional detection scheme in the far user employs MLD. In the conventional detection scheme, the signal for the near user is expected to be attenuated and it is not regarded as interference by the receiver of the far user.

The block diagram of the conventional detection scheme in the far user is shown in Fig. 2. The likelihood calculations by the MLD are given as

$$\Delta_{f,1}^{MLD}[k] = \sum_{S_f[k] \in \{S_f\}'} \exp\left(-\frac{1}{N_0} \|Y[k]\right) - H[k]\sqrt{1 - \alpha S_f[k]^2}),$$

where $y[n]$ is the $n$th received signal in the time domain, $Y[k]$ is the received signal, $H[k]$ is the frequency response, and $W[k]$ is the noise in the frequency domain on the $k$th subcarrier, respectively. The output of the DFT on the $k$th subcarrier is sent to a detector to obtain soft information for a turbo decoder.

At the detector, the likelihood values for turbo decoding are calculated. After the likelihood calculation in the detector, bit depuncturing is carried out. The likelihood sequence is then passed to an inverse circular buffer. Afterward, the likelihood sequence is divided into three streams that correspond to “Systematic”, “Parity 1”, and “Parity 2” streams, respectively. Each of the three output streams is rearranged in a sub-block deinterleaver and is then put into the turbo decoder to obtain information bits.
and then the information bits are recovered. The LLR calculated as above is passed to the turbo decoder obtained from the received signal

\[
\Delta_{f,0}^{MLD}[k] = \sum_{\hat{S}_f[k] \in \{S_f[k]\}_{m=0}} \exp(-\frac{1}{N_0} \|Y[k] - \hat{S}_f[k]\|^2) - H[k]\sqrt{1 - \alpha} S_f[k]|^2), \tag{4}
\]

where \(\Delta_{f,0}^{MLD}[k]\) and \(\Delta_{f,0}^{MLD}[k]\) are the sums of the likelihood values for the \(m\)th \((1 \leq m \leq M_f)\) bit being “1” and “0” in the symbol for the far user \(S_f[k]\) on the \(k\)th subcarrier, \(N_0\) is the power spectrum density of the white Gaussian noise, \(\hat{S}_f[k]\) is the symbol candidate for the far user on the \(k\)th subcarrier, and \({\{S_f[k]\}_{m=0}}\) and \({\{S_f[k]\}_{m=0}}\) are the sets of symbols of which the \(m\)th bit is equal to “1” and “0”, respectively. From Eqs. (3) and (4), the log likelihood ratio (LLR) is calculated as

\[
L(b_{mf}|Y[k]) = \log \frac{\Delta_{f,1}^{MLD}[k]}{\Delta_{f,0}^{MLD}[k]}, \tag{5}
\]

where \(L(b_{mf}|Y[k])\) is the LLR of the \(m\)th bit of the symbol obtained from the received signal \(Y[k]\) on the \(k\)th subcarrier. The LLR calculated as above is passed to the turbo decoder and then the information bits are recovered.

2.3 Codeword Level SIC in the Far User

The conventional detection scheme with the MLD in the far user neglects the signal for the near user. However, when the signal power for the near user is larger than the noise power at the far user, interference owing to the signal for the near user degrades the accuracy of the likelihood values. Meanwhile, in [6]–[15], the codeword level SIC is implemented in the receiver of the near user. Thus, it is assumed here that the codeword level SIC is also applied to the far user.

The block diagram of the codeword level SIC in the far user is presented in Fig. 3. The receiver of the far user employs the MLD and then obtains the decoded bits for the far user. From the decoded bits, the receiver generates a replica signal for the far user. The receiver then subtracts the replica signal for the far user from the received signal. The received signal after the interference cancelation is given as

\[
Y_{n}^{SIC}[k] = Y[k] - H[k]\sqrt{1 - \alpha} \hat{S}_f[k], \tag{6}
\]

where \(Y_{n}^{SIC}[k]\) is the received signal after the codeword level SIC and \(\hat{S}_f[k]\) is the replica signal for the far user on the \(k\)th subcarrier.

The receiver of the far user carries out the MLD against the remaining signal. The likelihood calculations for the signal for the near user are given as,

\[
\Delta_{n,1}^{MLD}[k] = \sum_{\hat{S}_n[k] \in \{S_n[k]\}_{m=1}} \exp(-\frac{1}{N_0} \|Y_n^{SIC}[k] - H[k]\sqrt{1 - \alpha} \hat{S}_n[k]|^2), \tag{7}
\]

\[
\Delta_{n,0}^{MLD}[k] = \sum_{\hat{S}_n[k] \in \{S_n[k]\}_{m=0}} \exp(-\frac{1}{N_0} \|Y_n^{SIC}[k] - H[k]\sqrt{1 - \alpha} \hat{S}_n[k]|^2), \tag{8}
\]

Fig. 3 Codeword level SIC in far user.

\[
\Delta_{n,1}^{MLD}[k] = \sum_{\hat{S}_n[k] \in \{S_n[k]\}_{m=1}} \exp(-\frac{1}{N_0} \|Y_n^{SIC}[k] - H[k]\sqrt{1 - \alpha} \hat{S}_n[k]|^2), \tag{9}
\]

where \(\Delta_{n,1}^{MLD}[k]\) and \(\Delta_{n,0}^{MLD}[k]\) are the sums of the likelihood values for the \(m\)th \((1 \leq m \leq M_n)\) bit of \(S_n[k]\) for the near user on the \(k\)th subcarrier, \(\hat{S}_n[k]\) is the symbol candidate for the near user on the \(k\)th subcarrier, and \({\{S_n[k]\}_{m=1}}\) and \({\{S_n[k]\}_{m=0}}\) are the sets of symbols of which the \(m\)th bit is equal to “1” and “0”, respectively. From the sum of the likelihood values, the LLR is calculated. The LLR is passed to the decoder and then the decoded bits for the near user are obtained.

The codeword level SIC generates the replica signal for the near user. Applying the codeword level SIC again, the received signal after the interference cancelation is given as

\[
Y_f^{SIC}[k] = Y[k] - H[k]\sqrt{\alpha} \hat{S}_n[k], \tag{10}
\]

where \(Y_f^{SIC}[k]\) is the received signal after the codeword level SIC for the far user and \(\hat{S}_n[k]\) is the replica signal for the near user on the \(k\)th subcarrier. The receiver of the far user carries out the MLD against the remaining signal to obtain the decoded bits for the far user. The likelihood calculations for the far user are given as,

\[
\Delta_{f,1}^{MLD}[k] = \sum_{\hat{S}_f[k] \in \{S_f[k]\}_{m=1}} \exp(-\frac{1}{N_0} \|Y_f^{SIC}[k] - H[k]\sqrt{1 - \alpha} \hat{S}_f[k]|^2), \tag{11}
\]

\[
\Delta_{f,0}^{MLD}[k] = \sum_{\hat{S}_f[k] \in \{S_f[k]\}_{m=0}} \exp(-\frac{1}{N_0} \|Y_f^{SIC}[k] - H[k]\sqrt{1 - \alpha} \hat{S}_f[k]|^2), \tag{12}
\]
where $D_{f,1}^{MLD}[k]$ and $D_{f,0}^{MLD}[k]$ are the sums of the likelihood values for the $m$th bit being “1” and “0” in the symbol $S_f[k]$ on the $k$th subcarrier, respectively. From the sum of the likelihood values, the LLR is calculated as the same manner as Eq. (5). The decoded bits for the far user are then decoded.

2.4 Proposed Joint MLD Scheme

In the conventional scheme, the receiver of the far user employs the MLD by neglecting the signal for the near user. Alternatively, as described in Sect. 2.3, the receiver may employ the codeword level SIC by treating the signal for the near user as the interference. On the other hand, in the proposed scheme, the receiver of the far user applies joint MLD to the received signal. The joint MLD treats the received signal as the superposed signal of the near and the far users with a larger number of constellation points. In the joint MLD, the signal for the near user is taken into account in the likelihood calculation. The LLR is calculated with the coordinates of the candidate constellation points and the received signal point in each symbol. It does not depend on the code rate of the signal for the near user because the far user does not decode the received signal for the near user. Thus, the performance of the far user depends only on the code rate of the signal for the far user and the modulation schemes of the signals for the near and the far users.

The block diagram of the proposed detection scheme in the far user is shown in Fig. 4. The likelihood calculations of the joint MLD are given as,

$$D_{f,1}^{JMLD}[k] = \sum_{\hat{S}_n[k] \in (S_n \cup \{S_f[k] \})_{b_f=1}} \exp \left( -\frac{1}{\sigma_r^2} ||Y[k] - H[k](\sqrt{\alpha} \hat{S}_n[k]) + \sqrt{1 - \alpha} \hat{S}_f[k])||^2 \right),$$

$$D_{f,0}^{JMLD}[k] = \sum_{\hat{S}_n[k] \in (S_n \cup \{S_f[k] \})_{b_f=0}} \exp \left( -\frac{1}{\sigma_r^2} ||Y[k] - H[k](\sqrt{\alpha} \hat{S}_n[k]) + \sqrt{1 - \alpha} \hat{S}_f[k])||^2 \right),$$

where $D_{f,1}^{JMLD}[k]$ and $D_{f,0}^{JMLD}[k]$ are the sums of the likelihood values for the $m$th bit being “1” and “0” in the symbol $S_f[k]$ on the $k$th subcarrier, respectively. Since the signals for the near and the far users are taken into account at the same time, the likelihood values are more accurate through the joint MLD. From Eqs. (12) and (13), the LLR is calculated as

$$L^{JMLD}(b_n^m|Y[k]) = \log \frac{D_{f,1}^{JMLD}[k]}{D_{f,0}^{JMLD}[k]},$$

where $L^{JMLD}(b_n^m|Y[k])$ is the LLR of the $m$th bit in the received symbol of the far user obtained from the received signal $Y[k]$ on the $k$th subcarrier. The LLR calculated as above is passed to the decoder and then the information bits for the far user are recovered.

2.5 Detection Complexity and Processing Delay

The difference of the detection complexity between the MLD and the joint MLD is the number of the candidate constellation points to be considered. The number of the constellation points in the MLD is equal to the modulation order of the signal for the far user whereas the number of the constellation points in the joint MLD is the product of the modulation orders of the signals for the near and the far users.

Meanwhile, the MLD is also employed in the codeword level SIC. Even though it is applied three times, the detection complexity of the MLD in the codeword level SIC is smaller than that of the joint MLD. However, the codeword level SIC also requires the turbo decoding three times and the replica generation two times. These processes at least cause a delay which is larger than that caused by the proposed scheme though the amount of delay depends largely on the implementation of the decoding process.

3. Numerical Results

3.1 Simulation Conditions

Numerical results obtained through computer simulation are presented in this section. Simulation conditions are shown in Table 1. An 8-state memory turbo code is employed and the size of the interleaver is 4800 [18]. The code rates of the signals for the near and the far users are selected from 1/3, 1/2, 2/3, and 5/6 through the RM function. Encoded bits are modulated with QPSK, 16QAM or 64QAM for the near user and QPSK for the far user on each subcarrier. The power allocation coefficient has been examined in a range of up to 0.4 in [15]. Thus, the link-level performance of the receiver of the far user is evaluated under power allocation coefficients of 0.1 to 0.4. In this paper, the power allocation coefficient, $\alpha$, is set to 0.2 or 0.35 unless it is specified. The signals for the near and the far users are superposed on the subcarrier to form the non-orthogonal signal. Other specifications such as the channel bandwidth, the subcarrier spacing, the number

---

1 Different combinations of modulation orders in the near and the far users are also examined, but the observed results show the same tendency as those presented in the following sections.
### Table 1 Simulation conditions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward error coding</td>
<td>turbo code</td>
</tr>
<tr>
<td>Interleaver size</td>
<td>4800</td>
</tr>
<tr>
<td>Code rate</td>
<td>1/3, 1/2, 2/3, 5/6</td>
</tr>
<tr>
<td>Modulation scheme for near user</td>
<td>QPSK/16QAM/64QAM + OFDM</td>
</tr>
<tr>
<td>Power allocation coefficient</td>
<td>$\alpha = 0.2, 0.35$</td>
</tr>
<tr>
<td>Channel Bandwidth</td>
<td>2.5 MHz</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>15 kHz</td>
</tr>
<tr>
<td>Number of subcarriers</td>
<td>256</td>
</tr>
<tr>
<td>Number of data subcarriers</td>
<td>151</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>3.84 MHz</td>
</tr>
<tr>
<td>GT length</td>
<td>5.21 $\mu$s (1st symbol)</td>
</tr>
<tr>
<td></td>
<td>4.69 $\mu$s (2nd–7th symbols)</td>
</tr>
<tr>
<td>Decoding algorithm</td>
<td>Log-MAP algorithm</td>
</tr>
<tr>
<td>Decoding iterations</td>
<td>8</td>
</tr>
<tr>
<td>Channel model</td>
<td>6 Taps GSM-TU Model</td>
</tr>
<tr>
<td>Channel estimation</td>
<td>Ideal</td>
</tr>
<tr>
<td>Number of trials</td>
<td>$4.8 \times 10^8$ bits</td>
</tr>
</tbody>
</table>

of DFT/IDFT points, and the sampling frequency follow those of the LTE standard. The channel bandwidth is set to 2.5 MHz and the subcarrier spacing is set to 15 kHz. The number of DFT/IDFT points is set to 256 and 151 subcarriers are used for data transmission. The sampling frequency is set to 3.84 MHz. The guard interval is set to 5.21 $\mu$s for the first symbol and 4.69 $\mu$s for the following six symbols. The Log-MAP algorithm is employed for decoding and the number of decoding iterations is eight. The channel model is assumed to be the 6-tap GSM typical urban (TU) model [19]. The channel response on each subcarrier is assumed to be ideally estimated. As detection schemes, MLD, codeword level SIC, and joint MLD are employed. The number of trials is set to $4.8 \times 10^8$ bits for each plot. Furthermore, the performance of an orthogonal multiple access (OMA) scheme without the superposition of the signal for the near user is also presented for comparison [15].

### 3.2 Comparison of BER Performance

The BER performance versus $E_b/N_0$ in the far user is presented in Figs. 5–8. Encoded bits are modulated with 16QAM or 64QAM for the near user and QPSK for the far user on each subcarrier. The power allocation coefficient, $\alpha$, is set to 0.2 or 0.35 in Figs. 5–8.

The BER performance versus $E_b/N_0$ in the far user is shown in Fig. 5. The code rate of the signal for the far user is set to 1/3. The proposed scheme outperforms the MLD by 0.2 dB at a BER of $10^{-2}$. This is because the joint MLD takes the signal for the near user into account when calculating the likelihood values, whereas the MLD neglects it. On the other hand, the performance of the codeword level SIC with a code rate of 1/3 in the near user is worse by 0.2 dB at a BER of $10^{-2}$ as compared with that of the MLD. Different code rates of the signals for the near user such as 1/2 and 5/6 are also examined and the same BER tendency has been observed.

Although the MLD is also employed in the codeword level SIC, the likelihood values are not accurate, which results in incorrect replica signals. This is due to the attenuation of the signal power as well as the residual interference after cancelation. Thus, the cancelation of the signal for the near user increases the interference to the signal for the far user.

The BER performance versus $E_b/N_0$ in the far user with a power allocation coefficient of 0.35 is shown in Fig. 6.
code rate of the signal for the far user is set to 1/3. For a power allocation coefficient of 0.35, the proposed scheme outperforms the MLD by 1.5 dB at a BER of $10^{-2}$. The improvement in $E_b/N_0$ due to the proposed scheme with a power allocation coefficient of 0.35 is larger than that with a power allocation coefficient of 0.2. This is because when the power allocation coefficient increases, the interference to the signal for the far user becomes large. The receiver with the MLD suffers from the interference while the receiver with the joint MLD can calculate the likelihood values accurately. On the other hand, the performance of the codeword level SIC with a power allocation coefficient of 0.35 is worse by 2.8 dB at a BER of $10^{-2}$ as compared with that of the MLD. The amount of the degradation with a power allocation coefficient of 0.35 is larger than that with a power allocation coefficient of 0.2. This is because the MLD is also employed at the first stage in the codeword level SIC and then decoding errors occur due to the large interference. In the codeword level SIC, decoding errors increase the residual interference after cancelation.

The BER performance versus $E_b/N_0$ in the far user with a code rate of 1/2 in the signal for the far user is shown in Fig. 7. For a code rate of 1/2, the proposed scheme outperforms the MLD by 0.5 dB at a BER of $10^{-2}$. On the other hand, the performance of the codeword level SIC with a code rate of 1/2 in the signal for the near user is worse by 0.8 dB at a BER of $10^{-2}$ as compared with that of the MLD. The performance of the codeword level SIC for the case of 64QAM in the signal for the near user is worse by 0.6 dB at a BER of $10^{-2}$ as compared with that of the MLD. The BER is worse for the case of 16QAM in the signal for the near user. The residual interference after cancelation due to decoding errors is smaller for the case of 64QAM than for the case of 16QAM though the BERs of the signal for the near user after the MLD are almost equivalent as shown in Fig. 13 in Sect. 3.5. This is because the minimum distance between the signal constellation points is smaller for the case of 64QAM in terms of the same symbol power. Thus, the influence of decoding errors for the case of 64QAM is smaller than that for the case of 16QAM.

The BER performance versus $E_b/N_0$ in the far user with a code rate of 5/6 is also shown in Fig. 8. For a code rate of 5/6, the proposed scheme outperforms the MLD by 0.5 dB at a BER of $10^{-2}$. The performance of the codeword level SIC is worse and the BER curve shows the error floor with the increase in $E_b/N_0$. The reason is that the error correction capability is too low for a code rate of 5/6 and bit errors remain after decoding of the signal for the near user.

### 3.3 Effect of Code Rate

The required $E_b/N_0$ at a BER of $10^{-2}$ versus the code rate is shown in Fig. 9. As the code rate increases, the conventional detection scheme requires more $E_b/N_0$ than the proposed scheme. This is because when the code rate increases, the error correction capability becomes low. As shown in Sect. 3.2, the performance of the codeword level SIC is worse than that of the MLD.

### 3.4 Effect of Power Allocation Coefficient

The required $E_b/N_0$ at a BER of $10^{-2}$ versus the power allocation coefficient, $\alpha$, is shown in Fig. 10. When the power allocation coefficient is large, the signal for the near user interferes with the signal for the far user more significantly and the required $E_b/N_0$ increases. The improvement in the
Fig. 9  Required $E_b/N_0$ at BER of $10^{-2}$ versus code rate (near user: 16QAM, far user: QPSK, power allocation coefficient $\alpha = 0.2$).

Fig. 10  Required $E_b/N_0$ at BER of $10^{-2}$ versus power allocation coefficient $\alpha$ (near user: 16QAM, far user: QPSK, code rate: 1/3).

required $E_b/N_0$ due to the joint MLD becomes larger when the power allocation coefficient increases. Concretely, the amount of the improvement is 0.2 dB under the condition that the power allocation coefficient is set to 0.2 and it becomes 0.8 dB, 1.5 dB, and 3.1 dB when the power allocation coefficient is set to 0.3, 0.35, and 0.4, respectively. It is clear that the proposed scheme is more effective when larger interference is caused by the signal for the near user. The amount of the performance deterioration of the codeword level SIC is especially large when the power allocation coefficient is at around 0.3. With a power allocation coefficient of 0.3,

3.5 Effect of Modulation Scheme of Near User

The required $E_b/N_0$ at a BER of $10^{-2}$ versus the various modulation schemes for the near user is presented in Fig. 12. The amount of the improvement due to the joint MLD for the case of 16QAM in the signal for the near user is larger than that for the case of QPSK. The amplitude of QPSK is constant while that of 16QAM varies. The receiver of the far user occasionally suffers from larger interference owing to 16QAM symbols even though the mean power of the interference is the same. The variation of the interference am-

Fig. 11 Constellation points of superposed signals for near and far users (near user: 16QAM, far user: QPSK, power allocation coefficient $\alpha = 0.3$).

Fig. 12 Required $E_b/N_0$ at BER of $10^{-2}$ versus modulation schemes for near user (power allocation coefficient $\alpha = 0.2$, code rate: 1/2).
The BER performance after the first codeword level SIC is shown in Fig. 13. As shown in Fig. 13, the BER performance of the signal for the near user for the cases of 16QAM and 64QAM is worse than that for the case of QPSK. The amplitude of the QPSK signal is constant, whereas those of the 16QAM and the 64QAM signals vary. For the cases of 16QAM and 64QAM in the signal for the near user, the receiver of the far user suffers from larger amplitude symbols as compared with the receiver for the case of QPSK even though the mean power of the signal is the same at the input of the codeword level SIC. Therefore, for the cases of 16QAM and 64QAM, decoding errors occur more easily than for the case of QPSK, which results in incorrect interference cancelation.

Furthermore, the minimum distance between the signal constellation points is another factor as explained in Sect. 3.2. The BER performance is better for the case of 64QAM in the signal for the near user than for the case of 16QAM in Fig. 7. This is because the minimum distance between the signal constellation points is smaller for the case of 64QAM in terms of the same symbol power in average. Thus, the influence of decoding errors for the case of 64QAM is smaller than that for the case of 16QAM.

### 3.6 System-Level Aspect

The system-level simulation of the NOMA downlink with the joint detection scheme in both the near and the far users is conducted as described in [20]. The conditions of the system-level simulation are presented in Table 2. A 19-hexagonal macrocell model is assumed. The cell radius of the macrocells is set to 289 m (inter-site distance = 500 m). Users are dropped randomly with a uniform distribution. As a propagation model, distance-dependent path loss with a decay factor of 3.76 and lognormal shadowing with a standard deviation of 8 dB are assumed. The shadowing correlation between the sites is set to 0.5. A six-path Rayleigh fading channel model with an exponential decay profile is assumed. The root-mean-square (RMS) delay spread is set to 1 μs and the maximum Doppler frequency is set to 5.55 Hz. The transmit power is set to 42 dBm and the receiver noise density is set to −174 dBm/Hz. The power allocation coefficient is selected from between 0 and 1 with a step size of 0.05. In the OMA scheme, it is set to 1.0. The throughputs of the conventional and proposed schemes are evaluated with CCC through Monte Carlo simulation [16]. Proportional fairness (PF) scheduling is applied and the time interval of the PF scheduling is set to 100 [21]. The number of resource blocks (RBs) is 24 and the number of subcarriers in one RB is 12.

The system bandwidth is 4.32 MHz. A QPSK, 16QAM, or 64QAM symbol is transmitted on each subcarrier. The number of symbols per trial is 100. The number of symbols per trial is 100.

<table>
<thead>
<tr>
<th>Table 2: Conditions of system-level simulation.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell layout</strong></td>
</tr>
<tr>
<td><strong>Inter-site distance</strong></td>
</tr>
<tr>
<td><strong>Minimum distance between users and cell site</strong></td>
</tr>
<tr>
<td><strong>Distance dependent path loss</strong></td>
</tr>
<tr>
<td>$r$: Distance [km]</td>
</tr>
<tr>
<td><strong>Shadowing standard deviation</strong></td>
</tr>
<tr>
<td><strong>Shadowing correlation</strong></td>
</tr>
<tr>
<td><strong>Channel model</strong></td>
</tr>
<tr>
<td><strong>Number of users per cell</strong></td>
</tr>
<tr>
<td><strong>Scheduling algorithm</strong></td>
</tr>
<tr>
<td><strong>Time interval</strong></td>
</tr>
<tr>
<td><strong>Transmit power</strong></td>
</tr>
<tr>
<td><strong>Power allocation coefficient</strong></td>
</tr>
<tr>
<td><strong>Receiver noise density</strong></td>
</tr>
<tr>
<td><strong>System bandwidth</strong></td>
</tr>
<tr>
<td><strong>Resource block bandwidth</strong></td>
</tr>
<tr>
<td><strong>Number of RBs</strong></td>
</tr>
<tr>
<td><strong>Modulation scheme</strong></td>
</tr>
<tr>
<td><strong>User drop</strong></td>
</tr>
<tr>
<td><strong>Trial per user drop</strong></td>
</tr>
<tr>
<td><strong>Number of symbols per trial</strong></td>
</tr>
</tbody>
</table>
drop is 30. The channel response is renewed each trial. The number of transmitted symbols per trial is 100 and the last 80 symbols are used for throughput evaluation.

The relationship between the number of users in each cell versus the system throughput per subcarrier is shown in Fig. 14. Here, 19 hexagonal cell sites are assumed and proportional fairness scheduling is employed for user assignment. The numerical results obtained through system-level simulation show that the joint MLD in the far user increases the system throughput by 0.3 bit/subcarrier.

The probability distribution function (PDF) of the power allocation coefficient, $\alpha$, is presented in Fig. 15. The probability that the power allocation coefficient is within the range of 0.25 to 0.5 is higher in the far user with the joint MLD whereas it is 0 in the far user without joint MLD. The reason is that the required $E_b/N_0$ reduces with the proposed scheme for the same BER in that range of the power allocation coefficient as presented in Fig. 10. The PF scheduling then selects the power allocation coefficient with higher probabilities at the range of 0.25 to 0.5.

4. Conclusions

In this paper, a joint MLD scheme in the NOMA downlink for the far user has been proposed and its BER performance investigated. In the conventional NOMA system, the receiver of the far user employs the MLD and neglects the signal for the near user even though its interference degrades the accuracy of likelihood values. On the other hand, the proposed joint MLD scheme takes the signal for the near user into account when calculating the likelihood values. It improves the accuracy of the likelihood values for the signal for the far user and thus improves its BER. Numerical results obtained through computer simulation have shown that the proposed scheme improves the performance by from 0.2 dB to 3.1 dB for power allocation coefficients of 0.2 to 0.4 at a BER of $10^{-2}$ relative to the conventional scheme the conventional scheme. In addition, the proposed scheme is more effective if the power allocation coefficient increases or a higher order modulation scheme is employed for the near user.

Acknowledgments

This work is supported in part by a Grant-in-Aid for Scientific Research (C) under Grant No.25420382 from the Ministry of Education, Culture, Sport, Science, and Technology in Japan.

References

[8] Y. Kishiyama, A. Benjebbour, and T. Nakamura, “Initial views on non-orthogonal multiple access based radio interface for future radio


Kenji Ando was born in Shizuoka, Japan in 1992. He received his B.E. degree in electronics engineering from Keio University, Japan in 2015. Since April 2015, he has been a graduate student in School of Integrated Design Engineering, Graduate School of Science and Technology, Keio University. His research interests are NOMA system.

Yukitoshi Sanada was born in Tokyo in 1969. He received his B.E., M.E., and Ph.D. degrees all in electrical engineering from Keio University, Yokohama, Japan in 1992, 1994, and 1997, respectively. From 1994 to 1997, he was a Special Researcher of Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists. From 1997 to 1998, he was with the Department of Electrical and Computer Engineering, Nagoya Institute of Technology, Nagoya, Japan, as a Research Assistant. From 1998, he joined the Department of Computer Science, Chiba Institute of Technology, Narashino, Japan, where is now a Professor. In 2008, he was a Visiting Associate Professor of University of British Columbia, B.C., Canada. From 2015, he is a Vice President of Chiba Institute of Technology. He received Distinguished Contributions Award from Communications Society of IEICE in 2009 and 2013. From 2013 to 2015, he was an Editor-in-Chief of IEICE Transactions on Communications (Japanese Ed.). His current research interests include wireless communications and physical layer security. He is a member of IEEE.

Takahiko Saba was born in Tokyo in 1969. He received his B.E., M.E., and Ph.D. degrees all in electrical engineering from Keio University, Yokohama, Japan in 1992, 1994, and 1997, respectively. From 1994 to 1997, he was a Special Researcher of Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists. From 1997 to 1998, he was with the Department of Electrical and Computer Engineering, Nagoya Institute of Technology, Nagoya, Japan, as a Research Assistant. From 1998, he joined the Department of Computer Science, Chiba Institute of Technology, Narashino, Japan, where is now a Professor. In 2008, he was a Visiting Associate Professor of University of British Columbia, B.C., Canada. From 2015, he is a Vice President of Chiba Institute of Technology. He received Distinguished Contributions Award from Communications Society of IEICE in 2009 and 2013. From 2013 to 2015, he was an Editor-in-Chief of IEICE Transactions on Communications (Japanese Ed.). His current research interests include wireless communications and physical layer security. He is a member of IEEE.