Optimization of Quantization Levels for Quantize-and-Forward Relaying with QAM Signaling

Xinyue Ling* Shinsuke Ibi* Kenji Miyamoto†
Seiichi Sampei* Jun Terada† and Akihiro Otaka†
*Graduate School of Engineering, Osaka University, Suita, Japan
E-mail: ling@wcs.comm.eng.osaka-u.ac.jp, {ibi, sampei}@comm.eng.osaka-u.ac.jp
†NTT Access Network Service Systems Laboratories, NTT Corporation, Yokosuka, Japan

***Abstract*—This paper investigates the uplink coordinated multiple point (CoMP) reception with quantize-and-forward (QF) relaying. In this relaying system, two relay nodes directly quantize the received symbols to integer numbers based on modulo-lattice and parallelly forward these numbers to the destination via optical fiber. In order to reduce the traffic load of optical fiber without sacrificing CoMP gain, we propose an optimization strategy from the viewpoint of mutual information which adaptively controls the quantization level at the relay nodes. We have demonstrated that the proposed optimization is capable of helping the system achieve high throughput and low traffic load of optical fiber at the same time by computer simulations.**

I. INTRODUCTION

With the assistance of massive deployment of small cell base stations (BS), network densification plays a key role to expand the network capacity in 5G [1]. For feasible massive deployment, the function of traditional BS should split to central unit (CU) and distributed unit (DU). To reduce costs, CU processing should be shared and centralized in cloud radio access network (C-RAN) [2-4]. The links between CU and DU, which are connected by optical fibers, are referred to as mobile fronthaul (MFH). Open interface of CU/DU connected by MFH is now actively studied in 3GPP (Third Generation Partnership Project) [5].

In C-RAN architecture, not only the system capacity expansion, but also the negative impacts of path loss and fading as well as shadowing are remedied with the aid of coordinated multiple point (CoMP) transmission in the downlink and reception in the uplink [6]. The conventional C-RAN splits the BS functions between the physical (PHY) layer functions and the radio frequency (RF) functions using option 8 defined in [5]. In this case, the in-phase and quadrature-phase (IQ) samples of the baseband signals are forwarded in the MFH with the common public radio interface (CPRI) [7]. Unfortunately, it is subject to huge traffic load in the MFH for forwarding IQ samples. If IQ samples are forwarded even in 5G systems, the required MFH transmission rate will exceed 100 Gbps [8].

In this paper, we focus on the split option 7-2, which is defined in [5], in uplink CoMP receptions where one source (S) node communicates with one destination (D) node via two relay (R) nodes, and each R node is connected to the D node by the MFH over an optical fiber. In the uplink of the option 7-2, fast Fourier Transform (FFT), cyclic prefix (CP) removal, resource de-mapping, and possibly pre-filtering functions reside in the DU, whereas the rest of PHY functions reside in the CU. The details of the meaning of pre-filtering were not discussed in the study phase. This split point is capable of reducing traffic load in the MFH without sacrificing detection capability of CoMP receptions if the pre-filtering is appropriately designed.

Generally, relaying strategy is classified into two types: non-regenerative relaying and regenerative relaying. In amplify-and-forward (AF) strategy as a type of non-regenerative relaying, the signal waveform from the S node is directly amplified at the R node as the pre-filtering, and IQ samples are forwarded, resulting in huge traffic load in the MFH. On the other hand, decode-and-forward (DF) strategy, as a type of regenerative relaying, conducts hard decision at the R node. However, demodulation and decode are not allowed in the option 7-2.

As an alternative solution for pre-filtering, we consider quantization-and-forward (QF) strategy, which quantizes the received signal to integer numbers at the R node and forward the quantized symbol to the D node. As a form of QF, the received signal at the R node is converted into log-likelihood ratio (LLR) by the symbol demapper, and then quantized and forwarded [9]. However, when a multilevel modulation such as 16 quadrature amplitude modulation (QAM) is used, the distribution of LLR differs for each bit of the symbol which is a disadvantage for determining the appropriate quantization threshold. Therefore, in this paper, we directly map the received signals to an integer grid on the complex plane and employ quantization based on modulo-lattice [10].

A primary focus of this paper is on the optimization of quantization level at the R node without sacrificing CoMP gain. When the suitable modulation and coding scheme (MCS) is decided according to the wireless channel states, there should also be a minimum quantization level assigned to each R node to maintain the bit detection capability. Thus, we will propose an optimization strategy for QF relaying to achieve high throughput and low traffic load in the MFH at the same time.

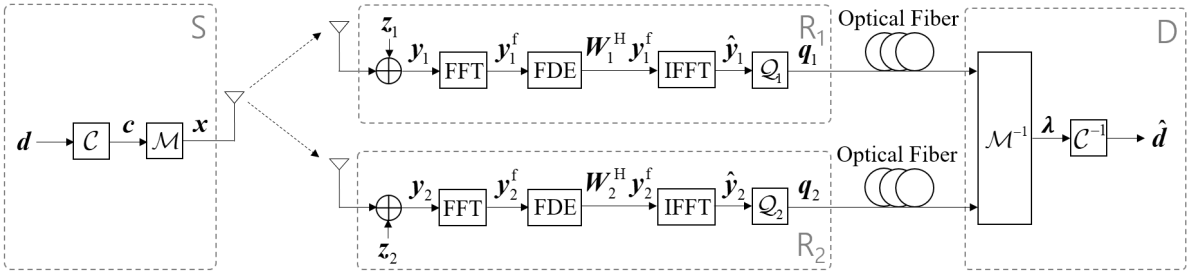


Fig. 1. A block diagram of QF relaying system.

The rest of this paper is organized as follows. In Sect. II, the problems to be solved in this paper are specified in the context of a system model of QF relaying. A quantization grid based on modulo-lattice is proposed in Sect. III. Then, the MCS and quantization level optimization strategy for QF relaying is discussed in Sect. IV. The effectiveness of proposed scheme is clarified by computer simulations in Sect. V. Finally, the paper is concluded by a summary in Sect. VI.

Throughout this paper, vectors and matrices are denoted by lower- and upper-case bold-face letters, respectively. Furthermore, $\cdot^{\mathrm{T}}$ and $\cdot^{\mathrm{H}}$ are transpose and conjugate transpose operators, respectively. tr[·] means the trace of the matrix. $\mathbb{E}\{\cdot\}$ denotes an expectation value. $\boldsymbol{I}_K$ is an identity matrix of size $K \times K$. $\mathbb{C}^{a\times b}$ denotes complex fields of size $a \times b$. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ indicate real and imaginary parts of complex values.

II. Signal Model of QF Relaying

A schematic of QF relaying system, which consists of one source (S), two relay (R), and one destination (D) nodes, is illustrated in Fig. 1. Let us consider here block transmissions. At the transmitter in the S node, information bits embedded in a block $\boldsymbol{d} = [d(1), d(2), \ldots, d(M)]^{\mathrm{T}} \in \{0, 1\}^{M\times 1}$ are encoded by the channel encoder ($\mathcal{C}$) with coding rate R_{c} for yielding a block of coded bits $\boldsymbol{c} = [c(1), c(2), \ldots, c(N)]^{\mathrm{T}} \in \{0, 1\}^{N\times 1}$, where M and N ($=M/R_{\mathrm{c}}$) denote length of information and code blocks, respectively. The resultant coded block $\boldsymbol{c}$ is mapped to the transmitted block $\boldsymbol{x} = [x(1), x(2), \ldots, x(K)]^{\mathrm{T}} \in \mathcal{X}^{K\times 1} = \{\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_{L-1}\}^{K\times 1}$ by the symbol mapper ($\mathcal{M}$), where L is the number of selectable symbols in each symbol generation process, and $K = N/\log_2 L$ is the number of symbols included in each block. The average energy of transmitted symbols $\mathbb{E}\{|x(k)|^2\}$ is E_{s}. After appending cyclic prefix (CP), the resultant block is simultaneously sent to two relay nodes R_i (i = 1, 2) via frequency selective fading channels.

After removing the CP part at the R_i node, the received symbol block $\boldsymbol{y}_i = [y_i(1), y_i(2), \ldots, y_i(K)]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ is represented by

$$\boldsymbol{y}_i = \boldsymbol{H}_i\boldsymbol{x} + \boldsymbol{z}_i, \tag{1}$$

where $\boldsymbol{z}_i = [z_i(1), z_i(2), \ldots, z_i(K)]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ denotes a complex-valued additive white Gaussian noise (AWGN) vector, whose elements obey zero mean and variance of one-sided noise power spectrum density N_0. $\boldsymbol{H}_i \in \mathbb{C}^{K\times K}$ is a circulant channel matrix between the S and R_i nodes based on a length K vector of the channel impulse response (CIR). The first column of the circulant matrix $\boldsymbol{H}_i$ is denoted by $\boldsymbol{h}_{1,i} = [h_i(1), h_i(2), \ldots, h_i(T), 0, \ldots, 0]^{\mathrm{T}}$ where $h_i(t) \in \mathbb{C}$ is a complex-valued fading coefficient of the t-th channel tap, and T is the channel memory of frequency selective fading channel.

In the receiver at each R_i node, frequency-domain equalization (FDE) is locally applied to mitigate the negative impact of frequency selective fading. The pre-filter output of adaptive equalizer $\hat{\boldsymbol{y}}_i = [\hat{y}_i(1), \hat{y}_i(2), \ldots \hat{y}_i(K)]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ based on minimum mean square error (MMSE) criterion is given by

$$\hat{\boldsymbol{y}}_i = \boldsymbol{H}_i^{\mathrm{H}}\left[\boldsymbol{H}_i\boldsymbol{H}_i^{\mathrm{H}} + N_0\boldsymbol{I}_K\right]^{-1}\boldsymbol{y}_i. \tag{2}$$

Thanks to the circularity of the channel matrix $\boldsymbol{H}_i$, the eigenvalue decomposition of $\boldsymbol{H}_i$ is represented by

$$\boldsymbol{H}_i = \boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Xi}_i\boldsymbol{F}, \tag{3}$$

where $\boldsymbol{\Xi}_i$ is a diagonal matrix. $\boldsymbol{F}$ is the fast Fourier transform (FFT) matrix of size $K \times K$ whose (k, k')-th element is defined as $(1/\sqrt{K})\exp(-\mathrm{j}2\pi(k-1)(k'-1)/K)$ and $\boldsymbol{F}^{\mathrm{H}}$ is the inverse FFT (IFFT) matrix. Substituting (3) into (2), we have

$$\hat{\boldsymbol{y}}_i = \boldsymbol{F}^{\mathrm{H}}\boldsymbol{W}_i^{\mathrm{H}}\boldsymbol{y}_i^{\mathrm{f}}, \tag{4}$$

where $\boldsymbol{y}_i^{\mathrm{f}}$ is obtained by FFT of the observed vector $\boldsymbol{y}_i$ as

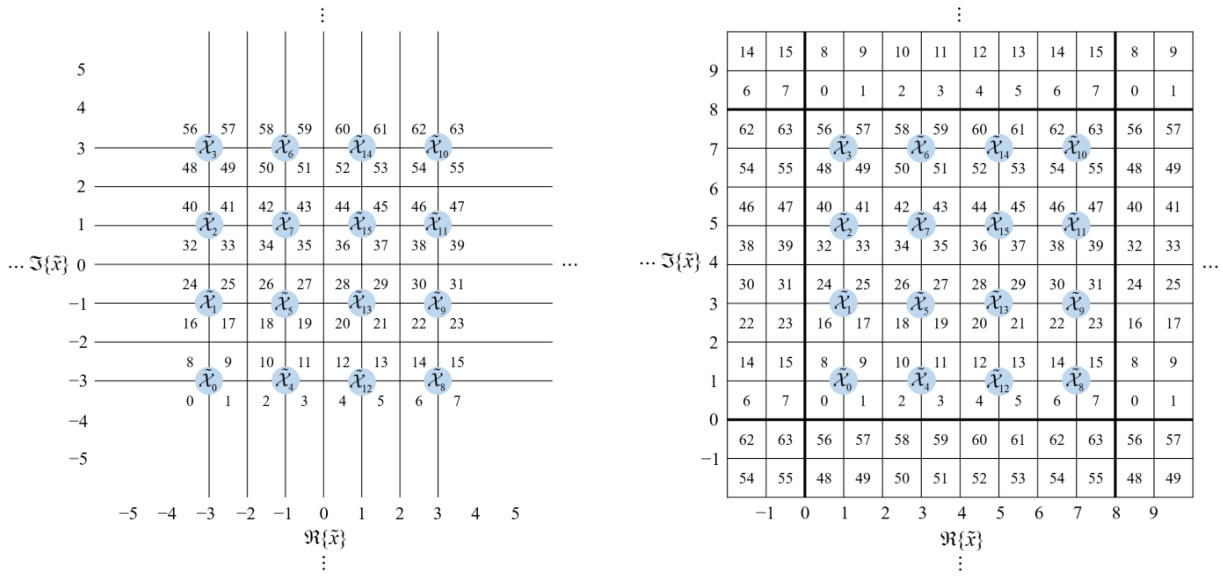
$$\boldsymbol{y}_i^{\mathrm{f}} = \boldsymbol{F}\boldsymbol{y}_i, \tag{5}$$

and FDE weight matrix $\boldsymbol{W}_i$ is given by

$$\boldsymbol{W}_i = \left[\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\mathrm{H}} + N_0\boldsymbol{I}_K\right]^{-1}\boldsymbol{\Xi}_i. \tag{6}$$

Note that the matrix inversion in (6) is calculated without requiring large computational efforts since $\boldsymbol{\Xi}_i$ is a diagonal matrix. Thus, the MMSE pre-filter output $\hat{\boldsymbol{y}}_i$ is derived from IFFT of the product of the weight matrix $\boldsymbol{W}_i^{\mathrm{H}}$ and received spectrum $\boldsymbol{y}_i^{\mathrm{f}}$.

The estimated symbol $\hat{y}_i(k)$ is quantized to an integer symbol $q_i(k) \in \{0, 1, \ldots, Q_i - 1\}$ at the quantizer ($\mathcal{Q}$) before forwarding the estimated information to the D node. We assume here that the D node is capable of capturing the quantized vector $\boldsymbol{q}_i = [q_i(1), q_i(2), \ldots, q_2(K)]^{\mathrm{T}}$ without any error via optical fibers. It is obvious that QF is equivalent to



(a) Typical quantization grid ($\alpha = \sqrt{10/E_s}$, $\beta = 0$)

(b) Quantization grid based on residue class ring ($\alpha = \sqrt{10/E_s}$, $\beta = 4$)

Fig. 2. Design of grids for 16QAM with 64-level quantization.

DF when $Q_i = L$. If an infinite resolution of quantization is allowed as $Q_i \to \infty$, the estimated symbol $\hat{y}_i(k)$ can be completely delivered to the D node without any quantization error. However, at the same time, the number of bits $\log_2 Q_i$ to be forwarded is infinite. Therefore, it is practical to set Q_i as a reasonable level.

At the D node, a block of extrinsic LLR $\boldsymbol{\lambda} = [\lambda(1), \lambda(2), \ldots, \lambda(N)]^{\mathrm{T}}$ is computed at the symbol demapper ($\mathcal{M}^{-1}$) before channel decoding ($\mathcal{C}^{-1}$). Finally, the channel decoder detects the block of information bits $\hat{\boldsymbol{d}} = [\hat{d}(1), \hat{d}(2), \ldots, \hat{d}(M)]^{\mathrm{T}}$.

Two problems arise here as follows:

A) How to compute the LLR $\boldsymbol{\lambda}$ with the aid of quantized symbols $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ at the D node,

B) How to minimize the number of forwarded bits $\log_2 Q_1 + \log_2 Q_2$ without sacrificing end-to-end throughput from the S node to D node.

In Sects. III and IV, we will solve these problems.

III. Quantization Rule and LLR Computations

A. LLR computations on Cartesian coordinates

As an example, we consider Gray-coded 16QAM ($L = 16$) as the symbol mapping rule and the quantization level $Q_i = 64$. This example can be easily extended to the other modulation level L, as well as different quantization level Q_i. Hereafter, the index (k) is omitted for ease of mathematical notations.

Let us consider scaling and shifting of symbol constellations for the intuitive quantization, which is represented as

$$\tilde{x} = \alpha x + \beta \in \tilde{\mathcal{X}} = \{\tilde{\mathcal{X}}_0, \tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_{L-1}\}, \tag{7}$$

where α and β are scaling and shifting parameters, respectively. Fig. 2(a) shows an integer grid for the quantization of $Q_i = 64$ in a case where $\alpha = \sqrt{10/E_s}$ and $\beta = 0$. In this case, the complex plane is divided into 64 cells. For adjusting the estimated symbol $\hat{y}_i$ on the quantization grid, scalar Gaussian approximation (SGA) is applied to (4) as

$$\hat{y}_i(k) = \mu x(k) + \nu_i(k), \tag{8}$$

where

$$\mu_i = \frac{1}{K}\mathrm{tr}\left[\boldsymbol{W}_i^{\mathrm{H}}\boldsymbol{\Xi}_i\right], \tag{9}$$

$$\sigma_i^2 = \mathbb{E}\left\{|\nu_i(k)|^2\right\} = \frac{N_0}{K}\mathrm{tr}\left[\boldsymbol{W}_i^{\mathrm{H}}\boldsymbol{W}_i\right]. \tag{10}$$

According to the value of μ_i, the scaling and shifting operations should be applied to $\hat{y}_i$ as

$$\tilde{y}_i = \alpha_i \hat{y}_i + \beta \triangleq \tilde{x} + \tilde{\nu}_i, \tag{11}$$

where $\alpha_i = \sqrt{10/E_s}/\mu_i$ and $\beta = 0$. In this case, the variance of $\tilde{\nu}_i$ is given by

$$\tilde{\sigma}_i^2 = \mathbb{E}\left\{|\tilde{\nu}_i(k)|^2\right\} = \alpha_i^2\sigma_i^2 = \frac{10N_0}{\mu_i^2 E_s K}\mathrm{tr}\left[\boldsymbol{W}_i^{\mathrm{H}}\boldsymbol{W}_i\right]. \tag{12}$$

Under the SGA, the probability density function (PDF) of (11) is expressed as

$$p(\tilde{y}_i \mid \tilde{x}) = \frac{1}{\pi\tilde{\sigma}_i^2}\exp\left[-\frac{|\tilde{y}_i - \tilde{x}|^2}{\tilde{\sigma}_i^2}\right]. \quad (13)$$

Let $D(q_i)$ be the region of the cell belonging to the quantized symbol q_i. By integrating the PDF over $D(q_i)$, probability mass function (PMF) of q_i is derived as

$$P\left[q_i \mid \tilde{x}\right] = \int_{D(q_i)} p\left(\xi \mid \tilde{x}\right) d\xi. \quad (14)$$

In the symbol demapper ($\mathcal{M}^{-1}$), on the basis of PMF (14), LLR is computed by

$$\lambda(n) = \ln\left[\frac{\sum_{\tilde{x}^+ \in \{\tilde{\mathcal{X}} \mid c(n)=1\}} \prod_{i=1}^{2} P\left[q_i(k) \mid \tilde{\mathcal{X}}^+\right]}{\sum_{\tilde{x}^- \in \{\tilde{\mathcal{X}} \mid c(n)=0\}} \prod_{i=1}^{2} P\left[q_i(k) \mid \tilde{\mathcal{X}}^+\right]}\right], \quad (15)$$

where $\{\tilde{\mathcal{X}} \mid c(n) = 0 \text{ or } 1\}$ means the subset of constellations belonging to $c(n) = 0$ or 1. It is obvious that the computational complexity involved in finding the LLR is unacceptable due to the integral in (14).

B. LLR computations on modulo-lattice

To simplify the calculation of the PMF (14), we consider a coset (residue class) based on modulo-lattice as shown in Fig. 2 (b) with the shifting parameter $\beta = 4$. Let $q_i^{\mathrm{I}} \in \{0, 1, \ldots, 7\}$ and $q_i^{\mathrm{Q}} \in \{0, 1, \ldots, 7\}$ be the remainder and the quotient when q_i is divided by $\sqrt{Q_i} = 8$, namely, $q_i^{\mathrm{I}} = \mathrm{mod}[q_i, 8]$ and $q_i^{\mathrm{Q}} = \mathrm{quo}[q_i, 8]$, respectively. Furthermore, the modulo operation is applied to the estimated symbol $\tilde{y}_i$, which is expressed as

$$\begin{aligned} &\underline{y}_i = \underline{y}_i^{\mathrm{I}} + \mathrm{j}\underline{y}_i^{\mathrm{Q}}, \\ &\underline{y}_i^{\mathrm{I}} = \Re\{\tilde{y}_i\} + \tau^{\mathrm{I}}\sqrt{Q_i}, \quad \underline{y}_i^{\mathrm{Q}} = \Im\{\tilde{y}_i\} + \tau^{\mathrm{Q}}\sqrt{Q_i}, \end{aligned} \quad (16)$$

where τ^{I} and τ^{Q} are integer numbers for satisfying $0 \le \underline{y}_i^{\mathrm{I}} < \sqrt{Q_i}$ and $0 \le \underline{y}_i^{\mathrm{Q}} < \sqrt{Q_i}$, respectively. Under the SGA, $\underline{y}_i^{\mathrm{I}}$ and $\underline{y}_i^{\mathrm{Q}}$ are stochastically independent, resulting in

$$p(\underline{y}_i \mid \tilde{x}) = p(\underline{y}_i^{\mathrm{I}} \mid \tilde{x}^{\mathrm{I}}) p(\underline{y}_i^{\mathrm{Q}} \mid \tilde{x}^{Q}), \quad (17)$$

where $\tilde{x}^{\mathrm{I}} = \Re\{\tilde{x}\}, \tilde{x}^{\mathrm{Q}} = \Im\{\tilde{x}\} \in \{1,3,5,7\}$ and

$$p(\underline{y}_i^{\phi} \mid \tilde{x}^{\phi}) = \frac{1}{\sqrt{\pi}\tilde{\sigma}_i}\exp\left[-\frac{(\underline{y}_i^{\phi} - \tilde{x}^{\phi})^2}{\tilde{\sigma}_i^2}\right], \quad (18)$$

$\phi \in \{\mathrm{I}, \mathrm{Q}\}$.

In this case, the PMF of (14) is readily given by

$$\begin{aligned} P\left[q_i \mid \tilde{x}\right] &= P\left[q_i^{\mathrm{I}} \mid \tilde{x}^{\mathrm{I}}\right] P\left[q_i^{\mathrm{Q}} \mid \tilde{x}^{\mathrm{Q}}\right] \\ &= \left[\sum_{z=-Z}^{Z} f(q_i^{\mathrm{I}}, \tilde{x}^{\mathrm{I}}, z)\right]\left[\sum_{z=-Z}^{Z} f(q_i^{\mathrm{Q}}, \tilde{x}^{\mathrm{Q}}, z)\right], \end{aligned} \quad (19)$$

Table I. MCS Table (L=16, K=2,048).

MCS index u	Mod. level L_u	Code rate R_u	Throughput efficiency $\eta_u = R_u \log_2 L_u$	Required MI I_u
7	16	0.369	1.4766	0.512
8	16	0.478	1.9141	0.644
9	16	0.602	2.4063	0.792

$$\begin{aligned} f(q_i^{\phi}, \tilde{x}^{\phi}, z) &= \int_{q_i^{\phi} + z\sqrt{Q_i}}^{q_i^{\phi} + z\sqrt{Q_i}+1} p\left(\xi \mid \tilde{x}^{\phi}\right) d\xi \\ &= \frac{1}{2}\mathrm{erf}\left(\frac{q_i^{\phi} - \tilde{x}^{\phi} + z\sqrt{Q_i} + 1}{\tilde{\sigma}_i^2}\right) \\ &\quad - \frac{1}{2}\mathrm{erf}\left(\frac{q_i^{\phi} - \tilde{x}^{\phi} + z\sqrt{Q_i}}{\tilde{\sigma}_i^2}\right), \end{aligned} \quad (20)$$

where Z is enough range of recursive coset for capturing Gaussian behavior, and erf(·) denotes the error function. Substituting (19) into (15), the integral is replaced by the simple error function.

IV. MCS AND QUANTIZATION LEVEL OPTIMIZATION

In typical broadband wireless communications system, adaptive modulation and coding (AMC) [11], [12] is applied for selecting suitable modulation level L and coding rate R_c according to the given wireless channel states. MCS for 16QAM specified in LTE is summarized in Tab. I. In this paper, we consider a block transmission of 16QAM with K=2,048. Furthermore, we assume that the selectable quantization levels are $Q_i \in \{0, 16, 64, 256, 1024, 4096\}$, i.e. $\log_2 Q_i \in \{\mathrm{NaN}, 4, 6, 8, 10, 12\}$ bits. Note that the algorithm proposed in this section can be extended to the other modulation levels, as well as the other quantization levels.

To determine the optimal MCS u^* at the MCS selector, mutual information (MI) is a rational metric in terms of extrinsic information transfer (EXIT) analysis [13]. For decoding with MCS u while satisfying block error rate (BLER)<10^{-3}, input MI I_λ to the channel decoder should be higher than the required MI I_u in Tab. I. In this QF relaying system, on the basis of PMF (19) with the aid of knowledge on $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$, the decoder input MI, i.e. demapper output MI, is calculated by [14]

$$\begin{aligned} I_\lambda &= \mathcal{T}\left(L, Q_1, Q_2\right) \\ &= 1 - \frac{1}{L\log_2 L}\sum_{l=0}^{L-1}\left\{\sum_{q_1=0}^{Q_1-1}\sum_{q_2=0}^{Q_2-1}\left[P\left[q_1 \mid \tilde{\mathcal{X}}_l\right]P\left[q_2 \mid \tilde{\mathcal{X}}_l\right]\right.\right. \\ &\quad \left.\left.\cdot\log_2\left(\frac{\sum_{l'=0}^{L-1} P\left[q_1 \mid \tilde{\mathcal{X}}_{l'}\right]P\left[q_2 \mid \tilde{\mathcal{X}}_{l'}\right]}{P\left[q_1 \mid \mathcal{X}_l\right]P\left[q_2 \mid \mathcal{X}_l\right]}\right)\right]\right\}. \end{aligned} \quad (21)$$

Notifications of $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ to the MCS selector are assumed to be perfect. When Q_i is high, the resolution of PMF is high, resulting in high I_λ. Therefore, the MCS u^* may be optimized by

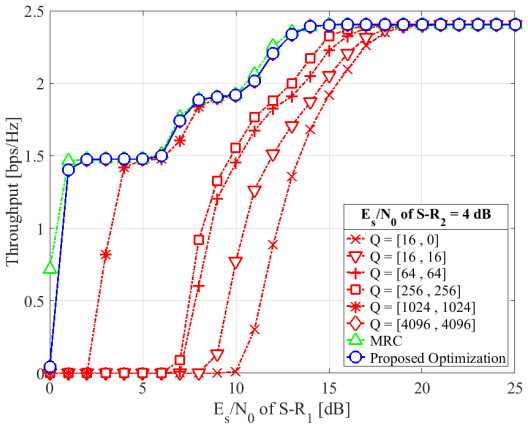


Fig. 4. Throughput capacity of S-R link.

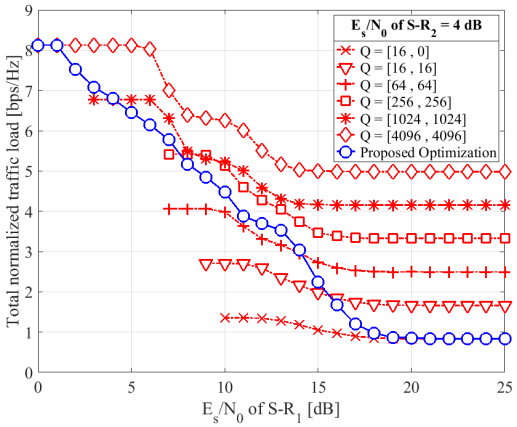


Fig. 5. Total normalized traffic load of R-D link.

$$u^* = \arg\max_u \eta_u \mathbb{I}[I_u < \mathcal{T}(L_u, 4096, 4096)], \tag{22}$$

where η_u is the throughput efficiency shown in Table I and $\mathbb{I}[v]$ is an indicator function which outputs 1 if v is true, otherwise 0.

On the other hand, it is essential to take the traffic load of R_i-D link into considerations. Higher quantization level results in lower BLER but higher traffic load. In fact, we do not always need higher quantization level because the information bits can be detected without errors only by hard decision values when the SNR of S-R_i link is sufficiently high. This feature brings us to a concept that it is not necessary for all R_i nodes to use the same quantization level. In other words, adaptively controlling the quantization level can help reduce of the traffic load in the MFH. For minimizing the traffic load without sacrificing the throughput optimized in (22), the following optimization should be solved by exhaust search.

$$\begin{aligned} &[Q_1^*, Q_2^*] = \arg\min_{Q_1,Q_2} Q_1 + Q_2, \\ &\text{s.t.} \quad I_{u^*} < \mathcal{T}(L_{u^*}, Q_1, Q_2). \end{aligned} \tag{23}$$

The optimized MCS u^* in (22) is notified to the S node, and the optimized Q_1^* and Q_2^* are notified to the R_1 and R_2 nodes, respectively. As are result of the proposed optimization, the traffic load in the MFH can be minimized without sacrificing end-to-end throughput.

V. Performance Evaluation

Computer simulations have been performed in order to confirm the effectiveness of the proposed method. First of all, Gray-coded 16QAM was assumed in $\mathcal{M}$. The channel encoder was turbo code defined by 3GPP with transmitted block of K=2,048 symbols. We used the 8-tap frequency selective Rayleigh fading channel which was normalized to have total average gain as 1, i.e. $|\boldsymbol{h}_{1,i}|^2 = 1$ with the aid of transmit power control. Max-Log-maximum a-posteriori (MAP) algorithm was applied for channel decoding with 5 iterations at the D node. The signal-to-noise power ratio (SNR) of S-R_2 link $\gamma_2 = |\boldsymbol{h}_{1,2}|^2 E_s / N_0 = E_s / N_0$ is fixed at 4 dB.

Fig. 4 shows the end-to-end throughput given by $\mathbb{E}\{\eta_{u^*} \mathbb{I}[\boldsymbol{d} = \hat{\boldsymbol{d}}]\}$ versus the SNR γ_1 of S-R_1 link. Q_i = 16 indicates that DF is conducted and Q_i = 0 is corresponding to the case where CoMP reception is not applied. When the throughput optimization in (22) is conducted, the throughput capacity is actually the same as the highest quantization level of $\boldsymbol{Q}$ = [4096, 4096] without any degradation, although it is hard to be seen clearly from the figure. As a reference curve, global MMSE equalizer in AF relaying, namely option 8 of functional splitting is shown in the figure, where SNR of the optical fiber is assumed to be ∞. More, specifically, the equalizer output is derived by

$$\hat{\boldsymbol{y}} = \boldsymbol{F} \begin{bmatrix} \boldsymbol{\Xi}_1 \\ \boldsymbol{\Xi}_2 \end{bmatrix}^{\mathrm{H}} \left[\begin{bmatrix} \boldsymbol{\Xi}_1 \\ \boldsymbol{\Xi}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Xi}_1 \\ \boldsymbol{\Xi}_2 \end{bmatrix}^{\mathrm{H}} + \frac{N_0}{E_s} \boldsymbol{I}_N \right]^{-1} \begin{bmatrix} \boldsymbol{F}^{\mathrm{H}} \boldsymbol{y}_1 \\ \boldsymbol{F}^{\mathrm{H}} \boldsymbol{y}_2 \end{bmatrix}, \tag{24}$$

and then LLR is computed. Meanwhile, the curve of global MMSE implies the upper limit of the throughput. Comparing the proposed optimization and the global MMSE, the average degradation is within an acceptable range of less than 5%.

On the other hand, Fig. 5 shows the total normalized traffic load of R-D link given by $\mathbb{E}\{(\log_2 Q_1 + \log_2 Q_2)/2\eta_{u^*}\}$ versus the SNR γ_1 of S-R_1 link. The figure explicitly demonstrates the fact that the traffic load of R-D link is significantly reduced through the quantization level optimization in (23). At first, the traffic load of the proposed optimization is equal to that of $\boldsymbol{Q}$ = [4096, 4096] since lower quantization levels cannot satisfy the required MI in Tab. I. When the quantization level $\boldsymbol{Q}$ = [1024, 1024] achieves the same throughput capacity as $\boldsymbol{Q}$ = [4096, 4096] at the SNR of around 5 dB, this lower quantization level is chosen and as a result, the traffic load is reduced. This optimization works effectively as the SNR of S-R_1 link increases. Finally, the average number of quantization bits required to transmit an information bit under the BLER of 10^{-3} is reduced to approximately 16.7% of that of $\boldsymbol{Q}$ = [4096, 4096] when the SNR of S-R_1 link is higher than 20 dB.

VI. CONCLUSIONS

In this paper, a quantization grid based on modulo-lattice which can help simplify the calculation of LLR was proposed for QF relaying. The primary problem to be solved was how to optimize the MCS and the quantization level at the relay nodes according to the wireless channel states when taking the end-to-end throughput and the traffic load of optical fiber into consideration. We proposed a unified optimization criterion based on mutual information and confirmed its effectiveness by computer simulations. The results demonstrated that the proposed optimization is capable of reducing at most 83.3% of the traffic load of optical fiber without scarifying any throughout degradation.

REFERENCES

[1] X. Ge, S. Tu, G. Mao, C. X. Wang, and T. Han, "5G Ultra-Dense Cellular Networks," in IEEE Wireless Communications, vol. 23, no. 1, pp. 72-79, Feb. 2016.

[2] C. L. I, J. Huang, R. Duan, and C. Cui, "Recent Progress on C-RAN Centralization and Cloudification," in IEEE Access, vol. 2, pp. 1030-1039, 2014.

[3] China Mobile Research Institute, "C-RAN the Road towards Green RAN," White Paper, v.3.0, Dec. 2013.

[4] NGMN, "Suggestions on Potential Solutions to C-RAN," The Next Generation Mobile Networks (NGMN) Alliance, Technical Report, Jan. 2013.

[5] 3GPP, TR38.801, v14.0.0, "Study on New Radio Access Technology: Radio Access Architecture and Interfaces (Release 14)," Mar. 2017.

[6] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson, "The Role of Small Cells, Coordinated Multipoint, and Massive MIMO in 5G," IEEE Communications Magazine, vol. 52, no. 5, pp.44-51, May 2014.

[7] CPRI, "CPRI Specification v7.0," Oct. 2015.

[8] U. Dötsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE," Bell Labs Technical Journal, vol. 18, no. 1, pp. 105-128, June 2013.

[9] Y. Nogawa, S. Ibi, and S. Sampei, "An Iterative Unified Decoding for Decode-Quantize-Forward Relaying," 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), Dresden, 2013, pp. 1-5.

[10] R. Zamir, Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory, Cambridge; Cambridge University Press, 2014.

[11] A. J. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," in IEEE Trans. Comm., vol. 46, no. 5, pp. 595-602, May 1998.

[12] C. Sunu and E. Aktaş, "Adaptive modulation and coding (AMC) in cooperative communication channels," 2014 22nd Signal Processing and Communications Applications Conference (SIU), Trabzon, 2014, pp. 1571-1574.

[13] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," in IEEE Trans. on Commun., vol. 49, no. 10, pp. 1727-1737, Oct. 2001.

[14] G. Ungerboeck, "Channel coding with multilevel/phase signals," in IEEE Trans. on Information Theory, vol. 28, no. 1, pp. 55-67, Jan. 1982.