Spectral and Energy Efficient Power Allocation for MIMO Broadcast Channels with Individual Delay and QoS Constraints

Jae-Hong Kwon, Jungil Cho, Byunggil Yu, Seongju Lee, Inha Jung, Chanho Hwang, and Young-Chai Ko

Abstract: **In this paper, we propose a novel power allocation algorithm to maximize the spectral efficiency (SE) and energy efficiency (EE) of multiple-input multiple-output (MIMO) broadcast channels under individual quality of service (QoS) constraints. In several wireless applications, one of the most important factors for QoS guarantee is delay outage. To address the impact of delay outage occurring in the link layer, effective capacity (EC) that means the maximum constant arrival rate satisfying statistical delay-QoS constraint is considered. Using a novel performance metric, effective-EE, which is defined as EC divided by total power consumption, we formulate an EC and effective-EE maximization problem with QoS constraints as an adaptive power allocation problem. By applying Lagrangian method, we solve optimization problem and propose an optimal power allocation algorithm. Simulation results demonstrate that our proposed algorithm can improve the EC and effective-EE performance.**

Index Terms: **Delay outage, effective capacity, energy efficiency, MIMO broadcast channels, power allocation, quality of service.**

I. INTRODUCTION

MULTIPLE -input multiple output (MIMO) technique has been widely deployed in wireless communication systems such as cellular system, wireless local area network (WLAN) standards etc. Traditionally, studies of MIMO predominantly focus on the improvement of spectral efficiency (SE) and achieving higher data rate for the extensive demands of mobile users. However, increasing the number of antennas, as well as deploying additional deploying base stations (BSs) densely consumes a great deal of circuit power. For this reason, energy efficiency (EE), defined as the achievable rate divided by power consumption, has attracted considerable attention for the design of future MIMO systems such as millimeter wave (mmWave) and massive MIMO systems [1]–[3].

Manuscript received October 12, 2020; approved for publication October 13, 2020. This paper is specially handled by EIC and Division Editor with the help of three anonymous reviewers in a fast manner.

This work was supported by a grant-in-aid of HANWHA SYSTEMS. This paper was presented in part at the IEEE Conference on Standards for Communications and Networking, Granada, November, 2019.

J.-H. Kwon and Y.-C. Ko are with the School of Electrical Engineering, Korea University, Seoul, South Korea, email: {hugokwon, koyc}@korea.ac.kr.

J. Cho, B. Yu, S. Lee, and I. Jung are with the C4I R&D Center, Communication-EW Center, Hanwha Systems, Sungnam, South Korea, email: {jungil01.cho, byunggil77.yu, sungju1015.lee, inha.jung}@hanwha.com.

C. Hwang is with the 2nd R&D Institute, Agency for Defense Development, Daejeon, South Korea, email: hwangchanho@add.re.kr

Y.-C. Ko is the corresponding author.

Digital Object Identifier: 10.1109/JCN.2020.000027

As described previously in [4], general energy efficient precoding method used in the MIMO system has been asymptotically investigated in several special cases, including slow fading environments or cases with a high signal-to-noise ratio (SNR). Specifically, for a MIMO singular value decomposition (MIMO-SVD) which is known as the optimal scheme in terms of SE, power allocation problem for the EE maximization has been studied in [5]. Although the EE of a MIMO system is not a convex problem, it is known that direct solving EE function does not is available since EE function is convex-concave fractional problems [6]. Furthermore, an EE optimization problem can be transformed into an equivalent convex optimization problem, which can be the efficient method in certain cases [7].

Nevertheless, works found in [4], [5] and [7] focus principally on point-to-point MIMO systems. In general downlink scenario, however, MIMO broadcast channels (MIMO-BC) have mainly been examined instead of point-to-point MIMO for spatial multiplexing gains since mobile units generally have a limited number of antennas due to both battery and size constraints. Similar to issues of point-to-point MIMO systems, studies concerning EE maximization problems in MIMO-BC also exist [8], [9]. In [8], energy efficient design of massive MIMO-BC was proposed by adopting an iteration algorithm for the number of transmit antenna and transmit power when zeroforcing (ZF) beamforming or maximal ratio transmission (MRT) beamforming were applied. For ta general situation, not specific beamforming cases, optimal and suboptimal beamforming techniques for EE in the MIMO-BC systems were investigated [9].

Meanwhile, the guarantee of quality of service (QoS) requirements, i.e., delay constraints, for ultra-reliable communication is also one main objective of future cellular networks. To address the delay requirements in a QoS of link layer, the concept of effective capacity (EC) was proposed as a QoS-aware link metric [10]. Considering the aspects of both physical layer and link layer, EC provides a maximum constant arrival rate with a delay-outage probability constraint that is specified by the delay exponent θ based on a large deviation principle theorem. In [11], an optimal power allocation method was proposed that adaptively determines transmit bits according to the variation of channel in each time slot. This work focused on the tradeoff between transmission power and delay of link layer. In [12], the authors applied a different approach that considered EE under the delay-QoS constraint using an EC metric and analyzed the tradeoff between EE and SE. The analysis of [12], however, did not consider circuit power consumption and therefore, it could only provide insights about the tradeoff between EE

1229-2370/19/$10.00 © 2020 KICS

and SE under asymptotic conditions such as a large transmission power scenario where circuit power consumption is negligible. The performance analysis of an effective-EE was undertaken in [13] by using the realistic power consumption model which includes the circuit power consumption. Furthermore, the maximum effective-EE with the minimum rate constraint was proposed in [14]. The delay outage probability effect on the EE was also analyzed in [15] and EC maximization problem with the EE constraint under the Nakagami-m fading channel was considered in [16]. The works in [11]–[16], however, only considered single user point-to-point communication, while studies on the EE of MIMO-BC with delay-QoS constraints were limited.

In this paper, we propose two power allocation algorithm to maximize the EC and effective-EE for MIMO-BC with individual QoS constraints. As mentioned above, extensive studies on the conventional SE and EE maximization problem in MIMO-BC without any consideration to delay outage in link layer already exist [9]. The objective of this work is to propose a novel spectral and energy efficient power allocation method, which takes into consideration the fact that each user has different network traffic and wireless applications. Our proposed algorithm has potential to be applied to certain scenarios that consider not only physical layer but also delay outage occurring in link layer, which is far more practical than adopting existing power allocation algorithms for SE and EE maximization in MIMO-BC. Rather than using conventional SE, we use a novel performance metric, EC. Also, effective-EE is adopted as an objective function, which is defined as the EC divided by the total power consumption [16] and formulate optimization problem to include the effect of the delay outage in link layer. We impose an EC threshold constraint that guarantee QoS for all the selected users, including a total power consumption constraint at the BS. By solving this optimization problem, we propose the power allocation algorithm to improve the EC and effective-EE performance and verify our result through simulation.

The rest of this paper is organized as follows. In section II, we describe the system model and explain the performance metric, EC and energy efficiency. Section III presents the optimal power allocation strategy to maximize the EC under individual QoS constraints. Also, section IV illustrates the optimal power allocation to maximize effective-EE. Section V shows simulation results and demonstrates our proposed algorithm. Finally, section VI summarizes the paper.

The following mathematical notations will be used throughout this paper. Uppercase and lower case boldfaces are used to denote matrices and vectors, respectively. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$,tr$(\cdot)$ and $\mathbb{E}[\cdot]$ represent the transpose, conjugate transpose, trace and expectation, respectively.

II. SYSTEM MODEL

A. Signal model

We consider a system model for MIMO-BC with single cell shown in Fig. 1. Here the BS is equipped with N_t antennas and each K user has a single antenna. We assume that the channel matrix from the BS to each K users as being $\mathbf{H} \in \mathbb{C}^{K\times N_t}$ where all the channel coefficients are independent and identically distributed (i.i.d.) complex Gaussian random variables $CN(0,1)$. When $\mathbf{x}$ is a transmitted signal vector satisfying $|x_i| = 1$ for $i = \{1, 2, \cdots, K\}$, the received signal vector, $\mathbf{y}$, with transmit beamforming is given by

$$\mathbf{y} = \mathbf{HFP}^{\frac{1}{2}}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathbb{C}^{K\times 1}$ is the additive white Gaussian noise (AWGN) vector with zero mean and variance σ^2 per entry, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_K] \in \mathbb{C}^{N_t\times K}$ is a beamforming matrix which satisfies $||\mathbf{f}_m|| = 1$ for all m, and $\mathbf{P} = \mathrm{diag}(\mathbf{p}) \in \mathbb{C}^{K\times K}$ with $\mathbf{p} = [p_1, p_2, \cdots, p_K]^{\mathrm{T}}$ is the power allocation matrix. The received signal for the ith user is expressed as

$$y_i = \sqrt{p_i}\mathbf{h}_i^{\mathrm{T}}\mathbf{f}_i x_i + \sum_{j=1, j\neq i}^{K} \sqrt{p_j}\mathbf{h}_i^{\mathrm{T}}\mathbf{f}_j x_j + n_i. \tag{2}$$

We assume that channel state information (CSI) for the ith user, $\mathbf{h}_i$, can be perfectly estimated at the ith user and all CSI is fed back to the BS with no error. After CSI is received from all K users, the BS can design the transmit beamforming matrix $\mathbf{F}$. The achievable rate for ith user is given by

$$R_i = \log_2\left(1 + \frac{p_i\left|\mathbf{h}_\mathbf{i}^{\mathrm{T}}\mathbf{f}_\mathbf{i}\right|^2}{\sum_{j=1,j\neq i}^{K} p_j\left|\mathbf{h}_i^{\mathrm{T}}\mathbf{f}_j\right|^2 + \sigma^2}\right). \tag{3}$$

We consider that BS applies zero-forcing beamforming (ZFBF) without generality [17]. Although minimum mean squared error (MMSE) beamforming is known as the optimal technique in the MIMO-BC systems, ZFBF can also achieved near optimal performance with high SNR regime. In this case, the transmit beamforming matrix is given by

$$\mathbf{F} = \frac{1}{\sqrt{\gamma}}\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H\right)^{-1}, \tag{4}$$

where $\gamma = \mathrm{tr}\left[\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\right]$ is the power normalization factor. Note that $1/\gamma$ becomes the effective channel gain for all K users. Assuming that the noise variance is equal to one for simplicity, we can write the achievable rate of the ith user as $R_i = \log_2(1 + p_i/\gamma)$.

B. Conventional definition of EE

The EE is defined as the channel capacity divided by the power consumption [18]. For MIMO-BC, we can present EE as [19]

$$\mathrm{EE} = \frac{\sum_{n=1}^{K} R_n}{\alpha\sum_{n=1}^{K} p_n + N_t P_{\mathrm{BS}} + K P_{\mathrm{user}}} \quad \text{(bits/J/Hz)}. \tag{5}$$

Note that P_{BS} and P_{user} are circuit power consumption per one radio frequency (RF) chain required for BS and user, respectively, while α is an efficiency of power amplifier (PA) at a BS. The realistic circuit power consumption model is generally given by [20]

$$P_{\mathrm{BS}} = P_{\mathrm{DAC}} + P_{\mathrm{mix}} + P_{\mathrm{filt}} \tag{6}$$

$$P_{\mathrm{user}} = P_{\mathrm{LNA}} + P_{mix} + P_{\mathrm{IFA}} + P_{\mathrm{filr}} + P_{\mathrm{ADC}}, \tag{7}$$

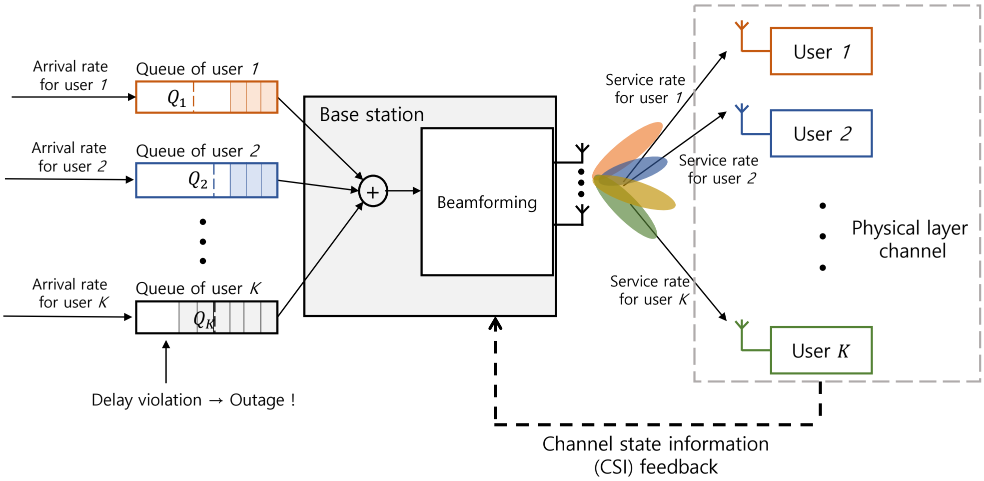

Fig. 1. A system model for queueing length where Q_n is the threshold of the queue length for the nth user.

where P_{DAC}, P_{ADC}, P_{mix}, P_{filt}, P_{filr}, P_{LNA} and P_{IFA} are the power which is consumed in the digital-to-analog converter (DAC), analog-to-digital converter (ADC), mixer, filters of transmitter, filters of receiver, low noise amplifier and intermediate frequency amplifier, respectively. According to [20], the power consumed by ADC and DAC is significantly larger than the power consumed by other parts. In the previous works on EE of the MIMO-BC [21], [22], the performance metric of (5) was used. However, when we consider delay outage in link layer, an EE metric would need to be modified by taking the effect of delay into account.

C. Effective capacity

In the physical layer, the Shannon capacity can provide insights on the maximum data rate for wireless channel. In most of the existing works that consider conventional physical layer channel metrics, QoS was generally defined as the SNR of the received signal or the Shannon capacity for each user. Studies based on this assumption have been conducted without considering that each user uses different wireless applications and has different characteristics of network traffic. In order to overcome this limitation, authors in [10] mathematically analyzed the outage probability caused by the queueing delay at the link layer and proposed a new performance metric called effective capacity (EC). EC describes the performance of a physical layer wireless channel with the parameters of link layer. According to the large deviation theory, a buffer violation probability is approximately modeled as

$$P_r\left\{Q \geqq q_{max}\right\} \approx e^{-\theta q_{max}}, \tag{8}$$

where Q is the steady state queue length at the transmitter and q_{max} is the delay threshold [10]. The EC is defined as the maximum constant arrival rate that a given service process can support to guarantee statistical delay requirements which is specified by delay exponent θ. Under the block fading channel assumption, the EC with the length of fading block, T_f, is de-

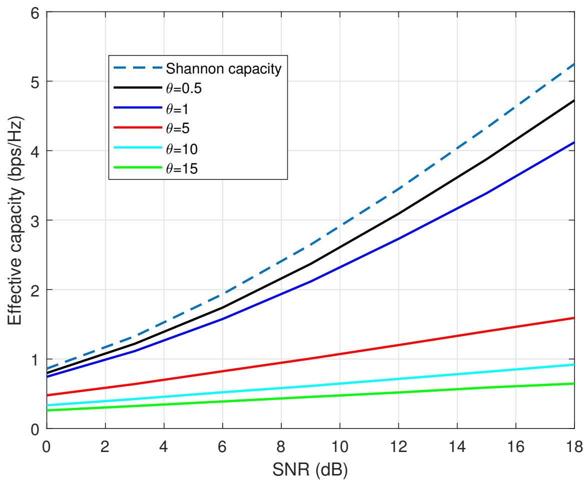

Fig. 2. Effective capacity w.r.t. SNR under different QoS constraints.

scribed as [10]

$$\mathrm{EC}\left(\theta\right) = -\frac{1}{\theta T_f} \ln \mathbb{E}_{\gamma}\left[e^{-\theta T_f R}\right],$$

where R and θ is an Shannon capacity and delay exponent, respectively. When $\theta \to 0$, there is no delay constraint and EC is equivalent to the Shannon capacity. Fig. 2 describes the EC under different QoS constraint.

III. EFFECTIVE CAPACITY MAXIMIZATION

A. Problem formulation

In this section, we consider the power allocation to maximize the sum effective capacity (EC) with individual delay constraints in the MIMO-BC systems. We assume that each user has a different QoS requirement such as a delay outage probability constraint, while the case that all the selected users have the same requirement of QoS was considered in previous work of effective-EE of the MIMO-BC [23]. In other words, a more general assumption was imposed on each user with a different value of delay exponent, θ [24]. In this case, the EC of the ith user is given by

$$\mathrm{EC}\left(\theta_i, p_i\right) = -\frac{1}{\theta_i T_f} \ln \mathbb{E}_\gamma \left[e^{-\theta_i T_f R_i}\right], \text{for } \ i = 1, 2, \cdots, K.$$

The total EC of the K users, EC_{tot}, can be expressed as the total sum of each user's EC as follows.

$$\mathrm{EC}_{\text{tot}}\left(\mathbf{p}\right) = \sum_{n=1}^{K} \mathrm{EC}\left(\theta_n, p_n\right) \tag{9}$$

Let us denote the EC threshold to avoid the delay violation as Ω_{thr}. Hence we can formulate optimization problem as follows

$$(\mathrm{OP1}) \quad \max_{\mathbf{p}} \mathrm{EC}_{\text{tot}}\left(\mathbf{p}\right), \tag{10}$$

$$\text{subject to} \quad \mathrm{EC}\left(\theta_n, p_n\right) \geqq \Omega_{\text{thr}} \quad \forall\, n, \tag{11}$$

$$P_t\left(\mathbf{p}\right) \leqq P_{\max}, \tag{12}$$

$$p_n \geqq 0, \forall\, n, \tag{13}$$

where $P_t\left(\mathbf{p}\right) = \sum_{n=1}^{K} p_n$. Note that inequality constraint of (11) indicates that each user should have the minimum EC performance for reliable communication. (12) means that the sum of transmit power to K users is less than maximum power and (13) is the nonnegative power constraint. According to [25], [26], it was proved that $\mathrm{EC}\left(\theta_n, p_n\right)$ is a convex function with regard to p_n. Since the objective function of (OP1) is the sum of convex functions and constraints are linear or convex, (OP1) is a strictly convex problem with regard to p_n. Thus, we can apply standard convex optimization method such as interior point method.

To obtain the solution, we need to compute $\mathrm{EC}\left(p_n\right)$ with given p_n and θ_n as

$$\mathrm{EC}\left(\theta_n, p_n\right) = -\frac{1}{\theta_n T_f} \ln \int_0^\infty \left(1 + \frac{p_n}{\gamma}\right)^{\frac{-\theta_n T_f}{\ln 2}} f_\gamma(\gamma) d\gamma,$$

where $f_\gamma(\cdot)$ is the probability density function (PDF) of γ. Unfortunately, it is complicated to obtain the closed form expression of $\mathrm{EC}\left(\theta_n, p_n\right)$ due to the unknown $f_\gamma(\cdot)$. Hence we apply a numerical method such as bisection searching.

B. Low complexity power allocation to maximize EC

To reduce the computational complexity, we first remove the individual QoS constraint through statistical analysis. If we define the minimum transmission power to satisfy the QoS constraint as $p_{\text{thr}}(\theta_n)$, as follows:

$$p_{\text{thr}}\left(\theta_n\right) = \left(\frac{-\beta_n \ln 2\Omega_{\text{thr}}}{\mathbb{E}\left[\gamma^{\beta_n}\right]}\right)^{-\frac{1}{\beta_n}}, \tag{14}$$

where $\beta_n = \theta_n T_f / \ln 2$.

Proposition 1: The threshold power $p_{thr}(\theta_n)$ can be approximately expressed as $\left(\frac{-\beta_n \ln 2\Omega_{thr}}{\mathbb{E}[\gamma^{\beta_n}]}\right)^{-\frac{1}{\beta_n}}$.

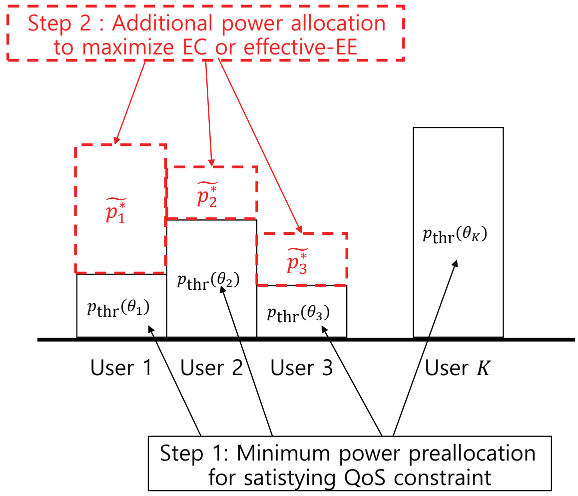

Fig. 3. The concept of our proposed algorithm.

proof: To obtain the minimum power to satisfy the QoS constraint, we apply simple lower bound of EC as

$$\mathrm{EC}\left(\theta_n, p_n\right) = -\frac{1}{\beta_n \ln 2} \ln \mathbb{E}_\gamma \left[\left(1 + \frac{p_n}{\gamma}\right)^{-\beta_n}\right] \tag{15}$$

$$\geqq \log_2 p_n + \frac{1}{\beta_n \ln 2} \ln\left(\mathbb{E}\left[\gamma^{\beta_n}\right]\right). \tag{16}$$

From (11), we can obtain

$$p_{\text{thr}}\left(\theta_n\right) \approx \left(\frac{-\beta_n \ln 2\Omega_{\text{thr}}}{\mathbb{E}\left[\gamma^{\beta_n}\right]}\right)^{-\frac{1}{\beta_n}}. \tag{17}$$

If we can predetermine the $p_{\text{thr}}(\theta_n)$ and allocate to total K users previously, we can remove the individual delay-QoS constraint of (11). Under the assumption of $P_{\max} \gg P_t\left(\mathbf{p}_{\text{thr}}\right)$, we can additionally allocate the remaining power after allocating the minimum transmit power, $P_{\max} - P_t\left(\mathbf{p}_{\text{thr}}\right)$, to each of the K users to maximize the EC. Thus, we can reformulate optimization problem as

$$(\mathrm{OP2}) \quad \max_{\widetilde{\mathbf{p}}} \quad \mathrm{EC}_{\text{tot}}\left(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}\right), \tag{18}$$

$$\text{subject to} \quad \widetilde{p}_n \geqq 0 \quad \forall\, n, \tag{19}$$

$$P_t\left(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}\right) \leqq P_{\max}. \tag{20}$$

Note that $\mathbf{p}_{\text{thr}} = \left[p_{\text{thr}}\left(\theta_1\right), p_{\text{thr}}\left(\theta_2\right), \cdots, p_{\text{thr}}\left(\theta_K\right)\right]^{\mathrm{T}}$ and $\widetilde{\mathbf{p}} = \left[\widetilde{p}_1, \widetilde{p}_2, \cdots, \widetilde{p}_K\right]^{\mathrm{T}}$ where $\widetilde{p}_n$ is the additional power allocated to the nth user to improve the EC performance. Fig. 3 describes the concept of our algorithm. Since objective function of (OP2) is still convex, we can adopt Karush-Kuhn-Tucker (KKT) condition to achieve the global optimum. To solve this problem, we consider the Lagrangian function, $L_1\left(\widetilde{\mathbf{p}}, \nu\right)$, with multipliers ν as

$$L_1\left(\widetilde{\mathbf{p}}, \nu\right) = \mathrm{EC}_{\text{tot}}\left(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}\right) + \nu\left(P_{\max} - P_t\left(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}\right)\right). \tag{21}$$

To calculate the optimal additional power allocated for the nth user from dual problem, we have

$$\widetilde{p}_n^* = \left[\underset{p}{\arg\min}\left(\mathcal{M}_\zeta\left(1-\beta_n,\gamma\right) - \frac{\mathcal{M}_{\phi_n}\left(-\beta_n,\gamma\right)}{\nu \ln 2}\right)\right]^+, \tag{22}$$

where $[x]^+ = \max\{x,0\}$. Note that $\mathcal{M}_X(\tau,t) = \mathbb{E}\left[X(t)^{\tau-1}\right]$ represents the Mellin transform where $\phi_n(\gamma) = \gamma^{\beta_n+1} + \gamma^{\beta_n}p$ and $\zeta(\gamma) = 1 + p\gamma^{-1}$. To obtain the Lagrangian dual variable ν, we adopt well known subgradient method as follows

$$\nu^{(j+1)} = \nu^{(j)} - \Delta_\nu^{(j)}\left[P_{\max} - P_t\left(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}^{(j)}\right)\right], \tag{23}$$

where j is an iteration number and $\Delta_\nu^{(j)}$ is corresponding non-negative step size. This process is similar to the well known water-filling method. In our approach, we need to compute additional power for $N \leqq K$ users who have less value of θ at each iteration and N is monotonically decreased until convergence. Also, since the number of constraints is less than (OP1), we can reduce the computational complexity.

IV. EFFECTIVE EE MAXIMIZATION

A. Problem formulation

In [9], work on the EE-optimal power allocation method for MIMO-BC focused on the maximization of conventional EE without a delay violation probability in the link layer. It means that this work was only valid in the physical layer of the wireless communication system. Considering the effect of queueing delay, we need to use the performance metric as effective EE that is defined as the total EC divided by total power consumption. Using the power consumption model of (5), we can define the effective-EE, η, as

$$\eta = \frac{\text{EC}_{\text{tot}}(\mathbf{p})}{\alpha P_t(\mathbf{p}) + P_c(N_t,K)}. \tag{24}$$

where $P_c(N_t,K) = N_t P_{BS} + K P_{UE}$ is the sum of transmit power and circuit power consumption. Similar to the relationship between EC and Shannon capacity, when θ is equal to 0, effective-EE is close to the conventional EE. Here, we use (24) instead of (5) to include the effect of the delay as well as the QoS of users.

In this section, we formulate the optimization problem in terms of the effective-EE under the QoS constraint for each user. Similar to EC maximization case, we can formulate the effective-EE maximization problem as follows

$$\max_{\mathbf{p}} \eta, \tag{25}$$

$$\text{subject to} \quad \text{EC}(\theta_n, p_n) \geqq \Omega_{\text{thr}} \quad \forall\, n, \tag{26}$$

$$P_t(\mathbf{p}) \leqq P_{\max}. \tag{27}$$

$$p_n \geqq 0,\ \forall\, n. \tag{28}$$

Three constraints are the same as the (11)–(13).

B. Optimal power allocation with Dinkelbach's transformation

The objective function of (25) is called a nonlinear fractional problem which is generally non convex function. To handle the fractional problem, we adopt well known *Dinkelbach's transform* technique [27]. We reformulate with auxiliary variable z as follows.

$$\text{(OP3)} \quad \max_{\mathbf{p}} \text{EC}_{\text{tot}}(\mathbf{p}) - z\left(\alpha P_t(\mathbf{p}) + P_c(N_t,K)\right), \tag{29}$$

$$\text{subject to} \quad \text{EC}(\theta_n, p_n) \geqq \Omega_{\text{thr}} \quad \forall\, n, \tag{30}$$

$$P_t(\mathbf{p}) \leqq P_{\max}, \tag{31}$$

$$p_n \geqq 0, \ \forall\, n. \tag{32}$$

Note that objective function of (OP3) is a strictly convex problem since the $\text{EC}_{\text{tot}}(\mathbf{p})$ is convex and $\alpha P_t(\mathbf{p}) + P_c(N_t,K)$ is a linear function. Hence we can optimal P^* with fixed z. The proof of equivalence between (OP3) and (25)–(28) is in [27]. To obtain the value of auxiliary variable z, we use iterative manner as

$$z[t+1] = \frac{\text{EC}_{\text{tot}}(\mathbf{p}[t])}{\alpha P_t(\mathbf{p}[t]) + P_c(N_t,K)}, \tag{33}$$

where t is the iteration index. First, we have initial value as $z[1] = 0$ and solve the (OP2) to obtain the $\mathbf{p}[1]$. In the next step, we compute $z[2]$ using (33) and solve (OP2) again. We continue these processes until $z[t]$ is converged. At each step, we adopt standard convex optimization algorithm such as interior-point method. As iterations continue, the value of z is monotonically increased and fast converge to global optimum of original problem.

C. Low complexity power allocation for effective-EE

In this section, we take the alternative low complexity power allocation method into account. To reduce the computational complexity, we first remove the individual QoS constraint through statistical analysis similar to EC maximization case. We then reformulate the problem of (25)–(28) as

$$\text{(OP4)} \quad \max_{\widetilde{\mathbf{p}}} \frac{\text{EC}_{\text{tot}}(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}})}{\alpha P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}) + P_c(N_t,K)}, \tag{34}$$

$$\text{subject to} \quad \widetilde{p}_n \geqq 0 \quad \forall\, n, \tag{35}$$

$$P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}) \leqq P_{\max}. \tag{36}$$

The objective function of (OP4) is a concave-convex fractional problem since the numerator is concave and the denominator is convex with respect to allocated power. When both the numerator and denominator of the concave-convex fractional problem are differentiable, the stationary point of this function is then known to be at a global maximum and Karush-Kuhn-Tucker (KKT) conditions are sufficient [6]. Therefore, we can directly solve our optimization problem by applying KKT condition even though the effective-EE function is not a convex problem. The Lagrangian with multiplier λ, $L_2(\widetilde{\mathbf{p}},\lambda)$, is described as

$$L_2(\widetilde{\mathbf{p}},\lambda) = \eta + \lambda\left(P_{\max} - P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}})\right).$$

The partial derivatives of L_2 with respect to $\widetilde{p}_n$ is then given by $\frac{\partial L_2}{\partial \widetilde{p}_n} = \frac{\partial \eta}{\partial \widetilde{p}_n} - \lambda$. First, we need to calculate partial derivative,

$g_\eta(\widetilde{p}_n) \triangleq \frac{\partial \eta}{\partial \widetilde{p}_n}$, as follows.

$$g_\eta(\widetilde{p}_n) = \frac{\mathcal{M}_{\phi_n}(-\beta_n, \gamma)\left(\alpha P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}) + P_c(N_t, K)\right)}{\ln 2 \mathcal{M}_\zeta(1-\beta_n, \gamma)} + \frac{\alpha \ln \mathcal{M}_{\phi_n}(-\beta_n, \gamma)}{\beta_n \ln 2} \quad (37)$$

For the KKT condition, allocating $p_m = p_m^{\text{thr}}$ is an optimal solution for the mth user in the case of $g_\eta(\widetilde{p}_m = 0) < 0$. Considering that $\mathcal{K}$ is a set of all selected users, we define $\mathcal{S}$ as a set of users who receive additional transmission power, and $\mathcal{Y}$ as a set of users who do not receive additional transmission power, that is, $\mathcal{S} = \mathcal{K} - \mathcal{Y}$. Then, the optimal additionally power allocation for the nth user, $\widetilde{p}_n^*$, is given by

$$\widetilde{p}_n^* = \begin{cases} \arg\min\limits_{p_n} \; g_\eta(\widetilde{p}_n) & \text{for } n \in \mathcal{S} \\ 0, & \text{for } n \in \mathcal{Y}. \end{cases} \quad (38)$$

If we denote the cardinality of $\mathcal{Y}$ as M, we would then need to calculate $K - M$ simultaneous equations since the objective function of (OP4) becomes the function of all $\widetilde{p}_n$ for $n \in \mathcal{S}$.

After computing $\widetilde{p}_n^*$ for all n, when we define $p_n^* = \widetilde{p}_n^* + p_n^{\text{thr}}$ for n, we take two cases into the consideration for the total power constraint.

C.1 $P_{\max} \geqq P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}_n^*)$ case

In the case above, p_n^* obtained from (38) is a valid solution for the nth user. Hence, the inequality constraint of (36) is no longer needed. Therefore, we can easily obtain the solution by solving all $K - M$ equations in (38).

C.2 $P_{\max} < P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}_n^*)$ case

In this case, we need to modify the inequality constraint (36) to $P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}_n^*) = P_{\max}$. When the total average transmit power is fixed, the effective-EE maximization problem is equivalent to the EC maximization problem since the denominator of the effective-EE is unchanged. Thus, when the maximum transmission power available at the transmitter is not sufficiently large, the method maximizing the EC performance forms the solution.

To get the insight of system design, we need to find the point of $P_t(\mathbf{p}_{\text{thr}} + \widetilde{\mathbf{p}}_n^*) = P_{\max}$ statistically. Considering $\text{EC}_{\text{tot}}(\mathbf{p}) \leqq \sum_{n=1}^K \log_2(1 + \gamma p_n)$ when $\theta \to 0$, we have $\mathbb{E}\left[\sum_{n=1}^K \log_2(1 + \gamma p_n)\right] \leqq \sum_{n=1}^K \log_2(1 + \mathbb{E}[\gamma] p_n)$ due to the Jensen's inequality. Using the Then the expectation of η is computed as

$$\mathbb{E}[\eta] \leqq \frac{\sum_{n=1}^K \log_2(1 + \mathbb{E}[\gamma] p_n)}{\alpha \sum_{n=1}^K p_n + P_c} \approx \frac{K \log_2\left(1 + \frac{P_t(N_t - K)}{K^2}\right)}{\alpha P_t + P_c}.$$

Note that the expectation of trace of inverse wishart matrix is given by $\mathbb{E}[\gamma] = K/N_t - K$. Differentiating $\mathbb{E}[\eta]$ with regard to P_t, we can obtain the closed-form expression of P_t^* as follows.

$$P_t^* \approx \frac{K^2}{N_t - K}\left(e^{\mathcal{W}\left(\frac{(N_t - K)P_c}{K^2 \alpha e} - \frac{1}{e}\right) + 1} - 1\right), \quad (39)$$

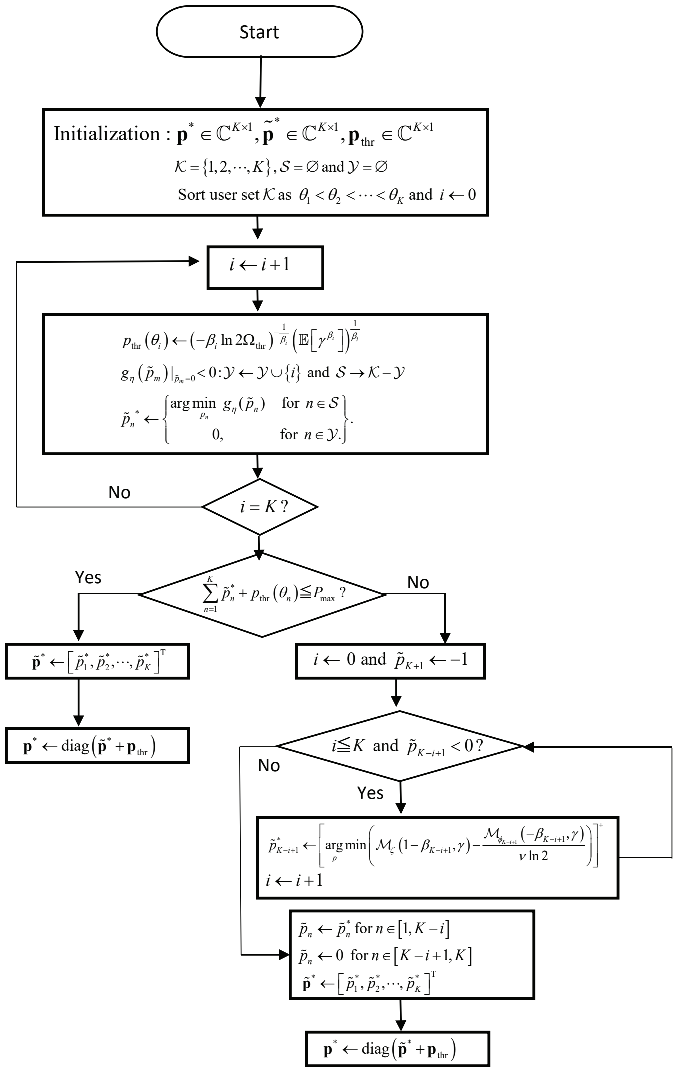

Fig. 4. Effective-EE optimization algorithm.

where $\mathcal{W}(\cdot)$ is *Lambert W* function. For the optimal system design in terms of effective EE, P_t^* becomes a sufficient maximum power of BS.

We can summarize optimal power allocation algorithm as shown in fig 4. Our proposed algorithm is more efficient than the ZFBF scheme for maximizing conventional EE as proposed in [9] for practical purpose since effective EE metric includes the impact of queueing delay outage in the link layer.

D. Computational complexity comparison

The computational complexity of (OP1)–(OP4) is evaluated and compared. We have total K decision variables for objective function and $(2K + 1)$ convex and linear constraints in (OP1) and (OP3). Thus, the computational complexity is asymptotically expressed as order of $\mathcal{O}\left(K^3(2K + 1)\right)$ for (OP1) and $\mathcal{O}\left(IK^3(2K + 1)\right)$ for (OP3) where $I > 1$ is the number of iterations for variable z. On the other hand, proposed low complexity algorithm has limited number of decision variables and constraints due to the predetermined minimum power to satisfy the statistical QoS constraint. Since the number of optimization variables is $N << K$ and the number of constraint is $(1+N)$, the computational complexity is denoted as $\mathcal{O}\left(N^3(1+N)\right)$ for both (OP2) and (OP4).

E. *The effect of* θ

In the concept of EC introduced by [10], the value of θ determines the delay constraint of the user's wireless application. In other words, users have low value of θ in loose delay-constrained system and high value in tight delay-constrained system. When delay constraint becomes tighter, BS needs to allocate more power to users to transmit accumulated information in the queue to escape the queueing delay outage. We can see that $p_{\text{thr}}(\theta_n)$ is proportional to the value of θ. If the selected users have low θ value, the optimal solution in terms of effective-EE is that BS does not use all available transmit power. However, for the tight-constrained case, BS needs to utilize all available transmit power to satisfy delay-QoS constraint for each user. Hence we can expect that performance gap between our proposed algorithm and other schemes such as equal power allocation is decreased when delay constraint becomes tighter.

V. SIMULATION RESULTS

In this section, we demonstrate that our proposed algorithm improves EC and effective-EE performance through Monte-Carlo simulations under the Rayleigh channel assumption. We assume $T_f = 1$ and $\Omega_{\text{thr}} = 1$. For the power consumption model, we set the parameters by referring to [20] and references therein. Although this parameter setting has an impact on the effective-EE performance, the impact of these parameters on the effective-EE using our proposed algorithm is little.

In Fig. 5, we show the EC performance with regard to the number of transmit antennas. For the low θ group, θ is uniformly distributed in $(0, 10]$ and for the high θ group, θ has uniformly distributed in $(0100]$. Compared to the equal allocation case, our proposed algorithm can improve the EC performance. Similar to Fig. 5, Fig. 6 shows the EC versus SNR. It can be also demonstrated improvement of our proposed EC maximization algorithm. From both Fig. 5 and Fig. 6, it can be shown that our proposed algorithm achieves the performance of original optimization method

In Fig. 7, we show the effective-EE performance versus the number of transmit antennas, N_t, for the different values of delay exponent θ. It confirms that our proposed algorithm has the ability to improve the effective-EE performance. Furthermore, we observe that the optimal number of transmit antennas for the effective-EE exists, since deployment of multiple antennas can increase the amount of the circuit power consumption at the BS.

Fig. 8 illustrates the effective-EE performance with regard to the maximum transmission power available on the BS for different N_t and K values. In the low maximum transmit power region, it is shown that the effective-EE performance using our proposed algorithm is the same as that of the EC-optimal scheme since the solution to both schemes are equal. However, over the high maximum transmit power region, the effective-EE performance remained unchanged under our algorithm while those of other schemes decrease. Even if the BS has the ability to use a higher transmit power, the BS is needed to use $\sum_{n=1}^{K} \widetilde{p}_n^* + p_n^{\text{thr}}$ which is less than $P_{\max}$ to maximize effective-EE performance. Hence, in the high maximum transmit power region, the effective-EE performance of our proposed scheme is

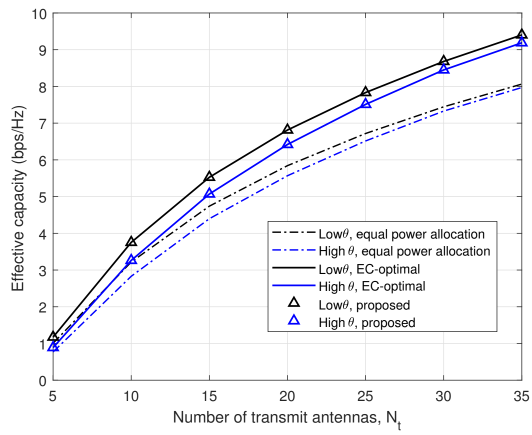

Fig. 5. Effective capacity with regard to N_t for different value of θ.

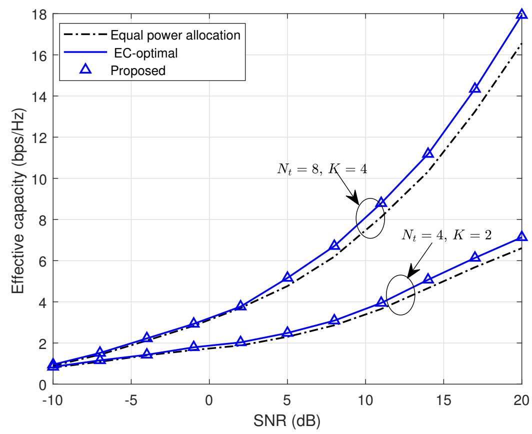

Fig. 6. Effective capacity with regard to SNR for different value of N_t and K.

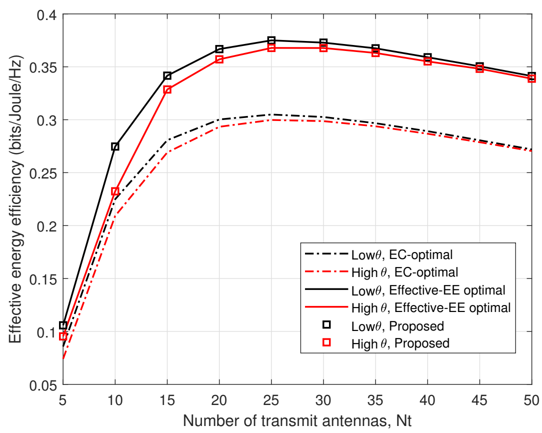

Fig. 7. Effective-EE w.r.t. the number of transmit antennas, N_t, for $K = 4$, $P_{\max} = 10$ dB and $\theta_{\text{low}} = [1, 2, 3, 4]$ and $\theta_{\text{high}} = [1, 11, 21, 31]$.

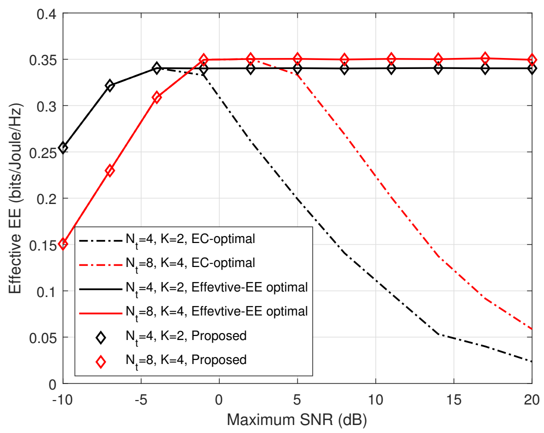

Fig. 8. Effective-EE w.r.t. the maximum transmission power available at the BS, P_{max}, for different vales of N_t when $\theta = [10, 20, 30, 40]$ and $K = 4$.

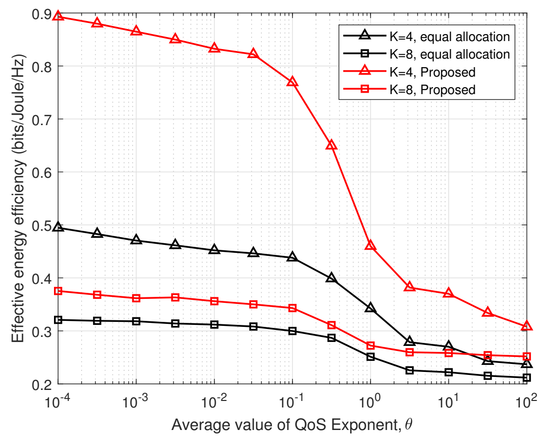

Fig. 10. Effective-EE w.r.t. the θ_{max}, for different vales of K when $N_t = 16$ and $P_{\text{max}} = 10$ dB.

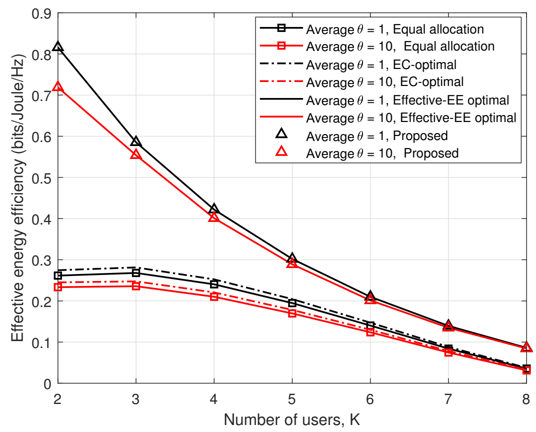

Fig. 9. Effective-EE w.r.t. the number of users, K, for different vales of θ_{max} when $N_t = 8$ and $P_{\text{max}} = 10$ dB.

saturated regardless of the maximum transmit power at BS.

In Fig. 9, we describe the effective-EE with regard to the number of users, K for three different schemes. We assume that the delay exponent value, θ, is uniformly distributed in $(0, \theta_{\text{max}}]$ and thus plot two cases, the first when $\theta_{\text{max}} = 1$ and the second when $\theta_{\text{max}} = 10$. In this case, our proposed algorithm achieves a higher effective-EE performance compared to the other two schemes. On the other hand, the performance gap between the three schemes decreases when K is increased. For example, the effective-EE performance of our proposed scheme is about 0.55 bits/Joule/Hz higher than that of the equal power allocation scheme when $K = 2$. However, the performance gap between the two scheme is decreased to 0.04 bits/Joule/Hz when $K = 8$ $(K \leqq N_t)$. The reason for this phenomenon is due to the power consumption of the user increasing as the number of users who simultaneously receive signals increases. In this case, it is expected that the effect of different power allocation for each user in the BS is relatively reduced.

We further plot the effective-EE versus θ_{max} for $K = 4$ and $K = 8$ where $N_t = 16$ and $P_{\text{max}} = 10$ dB in Fig. 10. The value of θ is also assumed to be uniformly distributed in $(0, \theta_{\text{max}}]$. This figure shows that the improvement of effective-EE performance of our proposed algorithm becomes larger when the number of selected users becomes smaller. Incidently, for tight delay-constrained systems (e.g. large θ), the performance gap between the two schemes is small since the power required to satisfy QoS constraint of each user is increased, while the remaining power to improve effective-EE performance is decreased. In short, the performance improvement caused by applying our proposed algorithm is reduced for a large θ scenario.

VI. CONCLUSION

In this paper, we proposed an energy efficient power allocation algorithm in MIMO broadcast channels under individual QoS constraints. We define individual QoS constraints as the minimum effective capacity which is a novel performance metric that considers the impact of a delay outage in the link layer. We formulated an optimization problem in terms of the effective energy efficiency which is defined as the effective capacity divided by the power consumption. In this problem, we set two constraints, one of which is the maximum total transmission power, and the other is a threshold of the effective capacity for each user. By considering the two constraints, we adopted the Lagrangian method and found the power allocation algorithm. Our proposed algorithm can be applied to the energy efficient MIMO broadcast channels with the consideration of the QoS guarantees for mobile users.

REFERENCES

[1] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, “On the energy-efficiency of hybrid analog-digital transceivers for single- and multi-carrier large antenna array systems,” *IEEE J. Sel. Areas Commun.,* vol. 35, no. 9, pp.1980–1995, Sept. 2017.

[2] R. Zi, X. Ge, J. Thompson, C. X. Wang, H. Wang, and T. Han, “Energy efficiency optimization of 5G radio frequency chain systems,” *IEEE J. Sel. Areas Commun.,* vol. 34, no.4 , pp.758–771, Apr. 2016.

[3] P. Cao and J. Thompson, “Low complexity energy efficiency analysis in millimeter wave communication systems,” in *Proc. WiOpt,* May 2017.

[4] E. V. Belmega and S. Lasaulce, “Energy-efficient precoding for multiple-antenna terminals,” *IEEE Trans. Signal Process.,* vol. 59, no. 1, pp.329–340, Jan. 2011.

[5] R. S. Prabhu and B. Daneshrad, "Energy-efficient power loading for a MIMO-SVD system and its performance in flat fading," in *Proc. IEEE GLOBECOM,* Dec. 2010.

[6] Schaible, S., "Fractional programming," *Zeitschrift für Operations-Research,* vol. 27, no. 1, pp.39–54, Dec. 1983.

[7] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with onformed transmitter," *IEEE Trans. Wireless Commun.,* vol. 11, no. 8, pp.2946–2957, Aug. 2012.

[8] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?," in *Proc. IEEE WCNC,* Apr. 2014.

[9] O. Tervo, L. N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.,* vol. 63, no. 20, pp.5574–5588, Oct. 2015.

[10] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.,* vol. 2, no. 4, pp.630–643, July 2003.

[11] J. Lee and N. Jindal, "Energy-efficient scheduling of delay constrained traffic over fading channels," *IEEE Trans. Wireless Commun.,* vol. 8, no. 4, pp.1866–1875, Apr. 2009.

[12] M. C. Gursoy, D. Qiao, and S. Velipasalar, "Analysis of energy efficiency in fading channels under QoS constraints," *IEEE Trans. Wireless Commun.,* vol. 8, no. 8, pp.4252–4263, Aug. 2009.

[13] L. Musavian and Tho Le-Ngoc, "Energy-efficient power allocation for delay-constrained systems," in *Proc. IEEE GLOBECOM,* Dec. 2012.

[14] B. Liu, L. Qiu and J. Xu, "QoS aware energy efficiency analysis in the cellular networks," *IEICE Trans. Commun.,* vol. E94-B, no. 10, pp.2925–2928, Oct. 2011.

[15] D. Qiao, "The impact of statistical delay constraints on the energy efficiency in fading channels," *IEEE Trans. Wireless Commun.,* vol. 15, no. 2, pp.994–1007, Feb. 2016.

[16] L. Musavian and Q. Ni, "Effective capacity maximization with statistical delay and effective energy efficiency requirements," *IEEE Trans. Wireless Commun.,* vol. 14, no. 7, pp.3824–3835, July 2015.

[17] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.,* vol. 24, no. 3, pp.528–541, Mar. 2006.

[18] H. Kwon and T. Birdsall, "Channel capacity in bits per joule," *IEEE J. Ocean. Eng.,* vol. 11, no. 1, pp.97–99, Jan. 1986.

[19] K. M. S. Huq, S. Mumtaz, J. Rodriguez, and R. L. Aguiar, "Energy efficiency optimization in MU-MIMO system with spectral efficiency constraint," in *Proc. IEEE ISCC,* June 2014.

[20] H. Li, L. Song, D. Zhu, and M. Lei, "Energy efficiency of large scale MIMO systems with transmit antenna selection," in *Proc. IEEE ICC,* June 2013.

[21] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO Systems: Is massive MIMO the answer?," *IEEE Trans. Wireless Commun.,* vol. 14, no. 6, pp.3059–3075, June 2015.

[22] L. Liu, G. Miao and J. Zhang, "Energy-efficient scheduling for downlink multi-user MIMO," in *Proc. IEEE ICC,* June 2012.

[23] J. Chen and K. K. Wong, "Improving energy efficiency for multiuser MIMO systems with effective capacity constraints," in *Proc. IEEE VTC spring,* Apr. 2009.

[24] Z. Feng, G. Wen, and C. W. Chen, "Multiuser effective capacity analysis for queue length based rate maximum wireless scheduling," in *Proc. IEEE ICCC,* Aug. 2012.

[25] D. Qiao, M. C. Gursoy, and S. Velipasalar, "Transmission strategies in multiple-access fading channels with statistical QoS constraints," *IEEE Trans. Inf. Theory,* vol. 58, no. 3, pp.1578–1593, Mar. 2012.

[26] L. Liu, Y. Yi, J. F. Chamberland, and J. Zhang, "Energy-efficient power allocation for delay-sensitive multimedia traffic over wireless systems," *IEEE Trans. Veh. Technol.,* vol. 63, no. 5, pp.2038–2047, June 2014.

[27] K. Shen and W. Yu, "Fractional programming for communication systems—part I: Power control and beamforming," *IEEE Trans. Signal Process.,* vol. 66, no. 10, pp. 2616-2630, May, 2018.

Jae-Hong Kwon received the B.S. and M.s degrees in Electrical Engineering from Korea University, Seoul, Korea, in 2014 and 2016. He is currently working toward the Ph.D. degree in the School of Electrical Engineering at Korea University. His research interests include energy efficiency of communication systems, millimeter wave communications and beamforming techniques.

Jungil Cho received M.S degrees in Wireless Communications from the INHA University, Incheon, Korea, in 2011. He is currently a C4I R&D Center, Communication-EW Center Senior Engineer at HANWHA SYSTEMS. His research interests are in phased array signal processing for tactical communication, RADAR and mobile communication.

Byunggil Yu received his Ph.D degrees in Radio Eng. from Kwangwoon University, Korea, in 2010. He is currently a C4I R&D Center, Communication-EW Center/Chief Engineer at HANWHA SYSTEMS. His research interests include active phased array antenna systems and metamaterials.

Seongju Lee received his M.S. degrees in Information and Communication Engineering from Kyungnam University in 2003.From 2003 to 2005, he was with plasma tech. center at Institute for Advanced Engineering working on high power transmitter for satellite communication system. He is currently working as a senior Engineer at Comm. & EW center at HANWHA SYSTEMS. His research interests include mmW, RF, mixed signal circuits design for wireless communication systems.

Inha Jung received a M. S degree in Wireless Communications from the University of Southampton in UK. He is currently a Senior Engineer at HANWHA SYSTEMS. His research interests are in architectures of phased array antenna systems for tactical communication, RADAR and mobile communication.

Chanho Hwang received his M.S degrees in Radio Communication Eng. from Korea Maritime & Ocean University, Busan, Korea, in 2015. He is currently a Senior Researcher at Agency for Defense Development, Daejeon, Korea. His research interests are in digital communication and signal processing.

Young-Chai Ko received the B.Sc. degree in Electrical and Telecommunication Engineering from the Hanyang University, Seoul, Korea and the M.S.E.E. and Ph.D. degrees in Electrical Engineering from the University of Minnesota, Minneapolis, MN in 1999 and 2001, respectively. He was with Novatel Wireless as a research scientist from January 2001 to March 2001. In March 2001, he joined the Texas Instruments, Inc., wireless center, San Diego, CA, as a senior engineer. He is now with the school of Electrical Engineering at Korea University as a Faculty Member. His current research interests include the performance analysis and the design of wireless communication systems.