more complicated and expensive hardware, including analog-to-digital converters with higher sampling rate, filters with wider bandwidth, larger RAM to store the digital samples, and faster CPU to process the digital samples.

REFERENCES

[1] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proc. ACM MobiCom*, Los Angeles, CA, USA, 2006, pp. 358–365.

[2] S. C. Liew, L. Lu, and S. Zhang,*A Primer on Physical-Layer Network Coding*. San Rafael, CA, USA: Morgan & Claypool, 2015.

[3] L. Lu and S. C. Liew, "Asynchronous physical-layer network coding," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 819–831, Feb. 2012.

[4] Q. Yang and S. C. Liew, "Asynchronous convolutional-coded physical-layer network coding," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1380–1395, Mar. 2014.

[5] X. Wu, C. Zhao, and X. You, "Joint LDPC and physical-layer network coding for asynchronous bi-directional relaying," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1446–1454, Aug. 2013.

[6] P.-C. Wang, Y.-C. Huang, and K. R. Narayanan, "Asynchronous physical-layer network coding with quasi-cyclic codes," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 2, pp. 309–322, Feb. 2015.

[7] D. Wang, S. Fu, and K. Lu, "Channel coding design to support asynchronous physical layer network coding," in *Proc. IEEE GLOBECOM*, Honolulu, HI, USA, 2009, pp. 1–6.

[8] F. Rossetto and M. Zorzi, "On the design of practical asynchronous physical layer network coding," in *Proc. IEEE SPAWC*, Perugia, Italy, 2009, pp. 469–473.

[9] L. Lu, T. Wang, S. C. Liew, and S. Zhang, "Implementation of physical-layer network coding," *Phys. Commun.*, vol. 6, pp. 74–87, 2013.

[10] L. Lu, L. You, and S. C. Liew, "Network-coded multiple access," *IEEE Trans. Mobile Comput.*, vol. 13, no. 12, pp. 2853–2869, Dec. 2014.

[11] Q. Yang, "Wireless communication by exploiting multi-user interference," Ph.D. dissertation, Dept. Inform. Eng., Faculty of Engineering, Chinese Univ. Hong Kong, Shatin, Hong Kong, Apr. 2015. [Online]. Available: http://wireless.ie.cuhk.edu.hk/yq/paper/QingYangDissertation.pdf

[12] X. Dang, Q. Li, and X. Yu, "Symbol timing estimation for physical-layer network coding," *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 755–758, May 2015.

[13] R. W. Heath, Jr., "Digital wireless communication: Physical layer exploration lab using the NI USRP," Nat. Instrum., Austin, TX, USA, 2015. [Online]. Available: http://www.ni.com/white-paper/13309/en/

[14] J. G. Proakis and M. Salehi,*Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2007.

[15] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531–532, Jul. 1972.

[16] L. Lu *et al.*, "Real-time implementation of physical-layer network coding," in *Proc. ACM SRIF*, 2013, pp. 71–76.

[17] R. P. Brent,*Algorithms for Minimization Without Derivatives*. Mineola, NY, USA: Dover, 2013.

[18] S. L. Miller, "Physical layer network coding with spectrally efficient pulse shapes," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 665–668, Jun. 2015.

Energy-Efficient Transmission Design in Non-orthogonal Multiple Access

Yi Zhang, *Student Member, IEEE*,
Hui-Ming Wang, *Senior Member, IEEE*,
Tong-Xing Zheng, and Qian Yang

***Abstract*—Non-orthogonal multiple access (NOMA) is considered as a promising technology for improving the spectral efficiency in fifth-generation systems. In this correspondence, we study the benefit of NOMA in enhancing energy efficiency (EE) for a multiuser downlink transmission, wherein the EE is defined as the ratio of the achievable sum rate of the users to the total power consumption. Our goal is to maximize EE subject to a minimum required data rate for each user, which leads to a nonconvex fractional programming problem. To solve it, we first establish the feasible range of the transmitting power that is able to support each user's data rate requirement. Then, we propose an EE-optimal power allocation strategy that maximizes EE. Our numerical results show that NOMA has superior EE performance in comparison with conventional orthogonal multiple access.**

***Index Terms*—Energy efficiency (EE), fractional programming optimization, non-orthogonal multiple access (NOMA), power allocation.**

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) has been recognized as a promising candidate for fifth-generation communication systems [1]. In contrast with conventional orthogonal multiple access (OMA), e.g., time-division multiple access (TDMA), NOMA serves multiple users simultaneously via power-domain division. Early literature on NOMA has focused mainly on the improvement of spectral efficiency (SE). For example, in [2], Ding *et al.* analyzed the ergodic sum rate and the outage performance of a single-input single-output (SISO) NOMA system with randomly deployed users. In [3], the impact of user pairing on two-user SISO NOMA systems was considered. Moreover, the power allocation among users in a SISO NOMA system was investigated in [4] from the perspective of user fairness.

In addition to SE, energy efficiency (EE) has recently drawn significant attention since the information and communication technology accounts for around 5% of the entire world energy consumption [5], which is becoming one of the major social and economical concerns worldwide. Currently, only a few works have studied NOMA from the perspective of EE. In [6], EE optimization was performed in a fading multiple-input–multiple-output NOMA system. However, the number of users is limited and fixed as two in [6], which greatly restrains the application of NOMA.

Motivated by the aforementioned observations, in this correspondence, we study EE optimization in a downlink SISO NOMA system

Manuscript received March 27, 2016; revised June 2, 2016; accepted June 4, 2016. Date of publication June 9, 2016; date of current version March 10, 2017. This work was supported in part by the Fund for the Author of National Excellent Doctoral Dissertation of China under Grant 201340; by the New Century Excellent Talents Support Fund of China under Grant NCET-13-0458; and by the Young Talent Support Fund of Science and Technology of Shaanxi Province under Grant 2015KJXX-01. The review of this paper was coordinated by Prof. G. Mao. *(Corresponding author: Hui-Ming Wang.)*

The authors are with the Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: yi.zhang.cn@outlook.com; xjbswhm@gmail.com; txzheng@stu.xjtu.edu.cn; qian-yang@outlook.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2016.2578949

with multiple users, whereby each user has its own quality of service (QoS) requirement guaranteed by a minimum required data rate. We first determine the minimum transmitting power that is able to support the required data rate for each user. Then, an energy-efficient power allocation strategy is proposed to maximize the EE by solving a nonconvex fractional programming problem. This optimization is further decoupled into two concatenate subproblems and solved one by one: 1) a nonconvex multivariate optimization problem that is solved in closed form; and 2) a strict pseudoconcave univariate optimization problem that is solved by the bisection method. Our numerical results show that NOMA has superior EE performance compared with conventional OMA.

II. System Model

Consider a downlink transmission scenario wherein one single-antenna base station (BS) simultaneously serves K single-antenna users. The channel from the BS to the kth user, i.e., $1 \le k \le K$, is modeled as $h_k = g_k d_k^{-\alpha/2}$, where g_k is the Rayleigh fading coefficient, d_k is the distance between the BS and the kth user, and α is the path-loss exponent. The instantaneous channel state information (CSI) of all users is known at the BS. Without loss of generality, we assume that the channel gains are sorted in ascending order, i.e., $0 < |h_1|^2 \le |h_2|^2 \cdots \le |h_K|^2$.

According to the principle of NOMA [1], [2], the BS broadcasts the superposition of K signals to its K users via power-domain division. We denote P as the total power available at the BS and a_k as the kth user's power allocation coefficient, which is defined as the ratio of the transmitting power for the kth user's message to the total power P. At receivers, successive interference cancelation is used to eliminate multiuser interference. Specifically, the kth user first decodes the ith user's message, i.e., $i < k$, and then removes this message from its received signal, in the order $i = 1, 2, \ldots, k-1$; the messages for the ith user, i.e., $i > k$, are treated as noise [2]. The achievable rate of the kth user R_k and the achievable sum rate of the system R are given by

$$R_k = \log_2\left(1 + \frac{P|h_k|^2 a_k}{P|h_k|^2 \sum_{i=k+1}^{K} a_i + \sigma^2}\right) \quad (1)$$

$$R = \sum_{k=1}^{K} R_k \quad (2)$$

respectively [2], where σ^2 is the power of the additive noise.

III. Problem Formulation

As done in [5] and [6], EE is defined as the ratio of the achievable sum rate of the system to the total power consumption, which is given by $\mathrm{EE} \triangleq R/(P_t + P_c)$, where $P_t \triangleq \sum_{k=1}^{K} a_k P$ is the actually consumed transmitting power, and P_c is the constant power consumption of circuits.

Our design is based on providing QoS guarantees for all users. Each user has a minimum required data rate, which is denoted as R_k^{Min} for $1 \le k \le K$, i.e.,

$$R_k \ge R_k^{\mathrm{Min}}, \quad 1 \le k \le K \quad (3)$$

which can be further transformed into

$$a_k \ge A_k\left(\sum_{i=k+1}^{K} a_i + \frac{\sigma^2}{P|h_k|^2}\right), \quad 1 \le k \le K \quad (4)$$

where $A_k \triangleq 2^{R_k^{\mathrm{Min}}} - 1$. Thereby, the EE maximization problem is formulated as

$$\max_{P_t, a_k, 1 \le k \le K} \ \mathrm{EE} \quad (5a)$$

$$\text{s.t.} \quad P_t \le P \quad \text{and} \quad P_t = \sum_{k=1}^{K} a_k P \quad (5b)$$

$$(4). \quad (5c)$$

Due to the minimum data rate constraints in (5c), problem (5) might be infeasible when the total power P is not sufficiently large. Accordingly, there must exist a minimum transmitting power P_{Min} that satisfies all users' data rate requirements; then, problem (5) is feasible only under the condition $P \ge P_{\mathrm{Min}}$. Therefore, it is important to first establish the feasible range of P, the derivation of which is discussed as follows.

A. Minimum Required Transmitting Power P_{Min}

Denote P_k as the power allocated to the kth user's message, then the problem of figuring out P_{Min} is formulated as

$$P_{\mathrm{Min}} \triangleq \min_{P_k, 1 \le k \le K} \ \sum_{k=1}^{K} P_k \quad (6a)$$

$$\text{s.t.}\ P_k \ge A_k\left(\sum_{i=k+1}^{K} P_i + \frac{\sigma^2}{|h_k|^2}\right), \quad 1 \le k \le K \quad (6b)$$

where (6b) comes from the minimum data rate constraints in (4). Problem (6) is solved by the following theorem.

Theorem 1: The optimal solution to problem (6), which is denoted by $\{P_k^{\mathrm{Min}}\}_{k=1}^{K}$, is given as

$$P_k^{\mathrm{Min}} = A_k\left(\sum_{i=k+1}^{K} P_i^{\mathrm{Min}} + \frac{\sigma^2}{|h_k|^2}\right), \quad 1 \le k \le K. \quad (7)$$

Proof: It can be seen that problem (6) is convex; thus, the following Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for its optimal solution:

$$1 + \sum_{i=1}^{k-1} \mu_i A_i = \mu_k, \qquad 1 \le k \le K \quad (8)$$

$$\mu_k\left[A_k\left(\sum_{i=k+1}^{K} P_i + \frac{\sigma^2}{|h_k|^2}\right) - P_k\right] = 0, \ 1 \le k \le K \quad (9)$$

$$\mu_k \ge 0, \qquad 1 \le k \le K \quad (10)$$

where $\{\mu_k\}_{k=1}^{K}$ are the Lagrange multipliers for the constraints in (6b). According to (8), we have $\mu_k > 0$ for $1 \le k \le K$, because $\{A_k\}_{k=1}^{K}$ and $\{\mu_k\}_{k=1}^{K}$ are all nonnegative numbers. This indicates that the constraints in (6b) are all satisfied at equality. Furthermore, by setting the constraints in (6b) to be active for $1 \le k \le K$, the closed-form expressions of $\{P_k^{\mathrm{Min}}\}_{k=1}^{K}$ are given by (7). Specifically, $\{P_k^{\mathrm{Min}}\}_{k=1}^{K}$ are sequentially calculated in the order $k = K, K-1, \ldots, 1$. Then, the proof is complete. ■

According to Theorem 1, with the instantaneous CSI, $\{P_k^{\mathrm{Min}}\}_{k=1}^{K}$ are calculated in the order $k = K, K-1, \ldots, 1$ by using (7). Afterward, $P_{\mathrm{Min}} = \sum_{k=1}^{K} P_k^{\mathrm{Min}}$ can be used as a threshold to verify whether P is large enough to meet the constraint on data rate for each user.

IV. Energy Efficiency Maximization

Here, we solve problem (5) under the condition $P \ge P_{\mathrm{Min}}$, which guarantees the feasibility of problem (5).

Substituting (1) into (2), we first reformulate the achievable sum rate R as follows:

$$R = \log_2\left(P|h_1|^2\sum_{i=1}^{K} a_i + \sigma^2\right) + \sum_{k=1}^{K-1}\left[\log_2\left(P|h_{k+1}|^2\sum_{i=k+1}^{K} a_i + \sigma^2\right) - \log_2\left(P|h_k|^2\sum_{i=k+1}^{K} a_i + \sigma^2\right)\right] - \log_2(\sigma^2). \quad (11)$$

For notational simplicity, we further define

$$C_k \triangleq P|h_k|^2, \quad 1 \le k \le K \quad (12a)$$

$$\theta \triangleq \sum_{i=1}^{K} a_i = \frac{P_t}{P} \quad (12b)$$

$$x_k \triangleq \sum_{i=k+1}^{K} a_i, \;\; 1 \le k \le K-1 \quad (12c)$$

$$F_k(x_k) \triangleq \log_2(C_{k+1}x_k + \sigma^2) - \log_2(C_k x_k + \sigma^2). \quad (12d)$$

By using these notations, R in (11) is recast as

$$R = \log_2(C_1\theta/\sigma^2 + 1) + \sum_{k=1}^{K-1} F_k(x_k) \quad (13)$$

and the original problem (5) is rewritten as

$$\max_{\theta, a_k, 1\le k\le K} \frac{\log_2(C_1\theta/\sigma^2+1) + \sum_{k=1}^{K-1} F_k(x_k)}{\theta P + P_c} \quad (14a)$$

$$\text{s.t.} \quad \theta \le 1 \quad \text{and} \quad \theta = \sum_{k=1}^{K} a_k \quad (14b)$$

$$(4). \quad (14c)$$

Here, we emphasize that θ is the ratio of the actually consumed transmitting power P_t to the total power available at the BS P. In particular, θ might be less than 1 for maximizing the EE. Problem (14) can be further decoupled into two concatenate subproblems as follows:

$$\max_{\theta} \frac{\log_2(C_1\theta/\sigma^2+1) + \max_{a_k, 1\le k\le K} \sum_{k=1}^{K-1} F_k(x_k)}{\theta P + P_c} \quad (15a)$$

$$\text{s.t.} \quad \theta \le 1 \;\text{ and }\; \theta = \sum_{k=1}^{K} a_k \quad (15b)$$

$$(4). \quad (15c)$$

The inner optimization problem is performed over arguments $\{a_k\}_{k=1}^K$ by taking θ as a constant, the solution of which is a function of θ. Afterward, the outer optimization problem is taken over θ. These two subproblems are sequentially solved in Sections IV-A and B, respectively. To be specific, in Section IV-A, taking θ as a constant, we propose a power allocation strategy to solve the inner optimization problem and at the same time obtain closed-form expressions for the optimal power allocation coefficients $\{a_k^*(\theta)\}_{k=1}^K$. In Section IV-B, we prove that the outer optimization problem is a strict pseudoconcave optimization problem with respect to (w.r.t.) the unique argument θ. Then, the bisection method is applied to find the optimal θ^* that maximizes the EE.

A. Optimal Power Allocation Strategy

By fixing θ in the feasible range $P_{\text{Min}}/P \le \theta \le 1$, the constraint $\theta \le 1$ in (15b) can be eliminated and then the inner optimization problem in (15) is rewritten as

$$\max_{a_k, 1\le k\le K} \sum_{k=1}^{K-1} F_k(x_k) \quad (16a)$$

$$\text{s.t.} \quad \sum_{k=1}^{K} a_k = \theta \;\text{ and }\; (4). \quad (16b)$$

Remark 1: In fact, by regarding θ as a constant in $P_{\text{Min}}/P \le \theta \le 1$, the nature of the inner optimization problem (16) is to maximize the EE subject to the constraint that the transmitting power should exactly be θP.

From (16), we can see that the objective function in (16a) is the summation of $K-1$ nonconvex subfunctions sharing similar forms. Based on this observation, we propose an optimization algorithm to solve (16), which can be elaborated in two steps as follows. *Step 1*: We individually maximize each subfunction $F_k(x_k)$ subject to the constraints in (16b). *Step 2*: We demonstrate that the optimal solution set of each maximization problem possesses a *unique common solution*. That is, we can find a unique solution that simultaneously maximizes $F_k(x_k)$ for $1 \le k \le K-1$ with all the constraints in (16b) satisfied. Thereby, this unique solution is the optimal solution to problem (16). Mathematically, denoting Φ_k as the optimal solution set for maximizing $F_k(x_k)$ subject to the constraints in (16b), we will show that

$$\Phi_1 \cap \Phi_2 \cap \cdots \cap \Phi_{K-1} = \left\{\{a_i^*(\theta)\}_{i=1}^K\right\} \quad (17)$$

where $\{a_i^*(\theta)\}_{i=1}^K$ is the unique common solution of the $K-1$ optimization problems.

Step 1: We now solve these $K-1$ optimization problems. First, the first-order derivative of $F_k(x_k)$ w.r.t. x_k is given as

$$\frac{dF_k(x_k)}{dx_k} = \frac{(C_{k+1} - C_k)\sigma^2}{\ln 2(C_{k+1}x_k + \sigma^2)(C_k x_k + \sigma^2)} \ge 0 \quad (18)$$

which demonstrates that $F_k(x_k)$ is a monotonically increasing function of x_k. Therefore, maximizing $F_k(x_k)$ is equivalent to maximizing x_k. As a result, we can uniformly formulate the aforementioned $K-1$ optimization problems as

$$\max_{a_k, 1\le k\le K} x_{K_0} \quad (19a)$$

$$\text{s.t.} \quad \sum_{k=1}^{K} a_k = \theta \quad (19b)$$

$$(4) \quad (19c)$$

where $1 \le K_0 \le K-1$ is the index for the $K-1$ optimization problems. Problem (19) is solved by the following proposition.

Proposition 1: Problem (19) is solved when the constraints in (19c) are active for $1 \le k \le K_0$, and the closed-form expressions of $\{a_k\}_{k=1}^{K_0}$ and x_{K_0} are given by

$$a_k = D_k\left(\theta - \sum_{i=1}^{k-1} a_i\right) + \frac{D_k\sigma^2}{P|h_k|^2}, \quad 1 \le k \le K_0 \quad (20a)$$

$$x_{K_0} = \theta - \sum_{k=1}^{K_0} a_k \quad (20b)$$

respectively, where $D_k = A_k/2^{R_k^{\text{Min}}}$.

Proof: See Appendix A.

Step 2: Based on Proposition 1, the following theorem further gives a closed-form expression for the unique solution to problem (16).

Theorem 2: The optimal power allocation coefficients $\{a_k^*(\theta)\}_{k=1}^K$ that maximize the objective function in (16a) are given by

$$a_k^*(\theta) = \begin{cases} D_k\left(\theta - \sum_{i=1}^{k-1} a_i^*(\theta)\right) + \frac{D_k\sigma^2}{P|h_k|^2}, & k \neq K \\ \theta - \sum_{i=1}^{K-1} a_i^*(\theta), & k = K. \end{cases} \quad (21)$$

Proof: According to (20a) in Proposition 1, arguments $\{a_k\}_{k=1}^{K_0}$ are uniquely and sequentially determined in the order $k=1,2,\ldots,K_0$ for maximizing x_{K_0}. This implies that more power allocation coefficients will be determined when K_0 increases. That is, the size of the optimal solution set of problem (19), i.e., Φ_{K_0}, becomes smaller as K_0 increases, which can be characterized by

$$\Phi_1 \supset \Phi_2 \supset \cdots \supset \Phi_{K-1} \quad (22a)$$

$$\Phi_1 \cap \Phi_2 \cap \cdots \cap \Phi_{K-1} = \Phi_{K-1}. \quad (22b)$$

Accordingly, Φ_{K-1} is the optimal solution set that simultaneously maximizes $F_{K_0}(x_{K_0})$ for $1 \le K_0 \le K-1$, consequently solving problem (16). By setting K_0 to $K-1$ in (20), the first $K-1$ optimal arguments $\{a_k^*(\theta)\}_{k=1}^{K-1}$ are uniquely and sequentially determined in the order $k=1,2,\ldots,K-1$ by using (20a). Furthermore, we have $a_K^*(\theta) = \theta - \sum_{k=1}^{K-1} a_k^*(\theta)$ from (19b). As a result, the closed-form expressions of $\{a_k^*(\theta)\}_{k=1}^K$ that maximize the objective function in (16a) are given by (21). Then, the proof is complete. ■

From Theorem 2, we find that the inner optimization problem (16) is solved when the minimum data rate constraints in (4) are active for $1 \le k \le K-1$, which implies that the optimal power allocation strategy is to use the extra power $(\theta P - P_{\mathrm{Min}})$ only for increasing the Kth user's data rate because the Kth user has the largest channel gain, and it achieves the highest data rate among all users with the same amount of power. That is, the Kth user can use power more efficiently than the other users do. As a result, when the transmitting power is fixed as θP, the nature of maximizing the EE is to enlarge the data rate of the user with the largest channel gain as much as possible. However, this does not signify that the extra power $(\theta P - P_{\mathrm{Min}})$ should be wholly allocated to the Kth user, since its signal also interferes with the other $K-1$ users. More explicitly, the following corollary further reveals the essence of the proposed optimal power allocation strategy.

Corollary 1: $\{da_k^*(\theta)/d\theta\}_{k=1}^K$ are positive constants.

Proof: See Appendix B. ■

Corollary 1 means that the power of the kth user's signal $a_k^*(\theta)P$ linearly increases as θ increases. This implies that the extra power $(\theta P - P_{\mathrm{Min}})$ is allocated to the kth user with the constant proportion $da_k^*(\theta)/d\theta$.

B. Optimal Transmitting Power $\theta^* P$ for Maximizing the EE

In the previous section, the inner optimization problem (16) is solved with the closed-form solution in (21), of which θ is the unique argument. Consequently, the outer optimization problem in (15) is transformed into a univariate optimization problem w.r.t. θ, which is given by

$$\max_{\theta} \quad \frac{\log_2(C_1\theta/\sigma^2+1) + \sum_{k=1}^{K-1} F_k\left(x_k^*(\theta)\right)}{\theta P + P_c} \quad (23a)$$

$$\text{s.t.} \quad P_{\mathrm{Min}} \le \theta P \le P \quad (23b)$$

where $x_k^*(\theta) = \sum_{i=k+1}^K a_i^*(\theta) = \theta - \sum_{i=1}^k a_i^*(\theta)$, and the constraint in (23b) indicates the feasible range of θ.

Theorem 3: Denote the objective function in (23a) as $\mathrm{EE}(\theta)$, then $\mathrm{EE}(\theta)$ is a strict pseudoconcave function w.r.t. θ.

Proof: It can be easily verified that the second-order derivative of $F_k(x_k)$ w.r.t. x_k is nonpositive, which indicates that $F_k(x_k)$ is a concave function of x_k for $1 \le k \le K-1$. Based on this property, we further conclude that $F_k(x_k^*(\theta))$ is a concave function of θ. This is because $\{a_i^*(\theta)\}_{i=1}^K$ and $\{x_k^*(\theta)\}_{k=1}^{K-1}$ are all affine mappings according to their linear expressions, which preserves the convexity of $F_k(x_k^*(\theta))$ w.r.t. θ. Moreover, it can be easily verified that $\log_2(C_1\theta/\sigma^2+1)$ is a strict concave function of θ. As a result, the numerator of $\mathrm{EE}(\theta)$, which is the summation of $\sum_{k=1}^{K-1} F_k(x_k^*(\theta))$ and $\log_2(C_1\theta/\sigma^2+1)$, must be a strict concave function w.r.t. θ, since the convexity is preserved by the addition operation. By now, we have proved that $\mathrm{EE}(\theta)$ has a strict concave numerator and an affine denominator, which ensures that $\mathrm{EE}(\theta)$ is a strict pseudoconcave function w.r.t. θ [5, Proposition 6]. ■

According to Theorem 3, $\mathrm{EE}(\theta)$ is a strict pseudoconcave function of θ and, thus, admits a unique maximizer that is the unique root of the equation $d\mathrm{EE}(\theta)/d\theta = 0$ [5, Proposition 5]. The expression of $d\mathrm{EE}(\theta)/d\theta$ is given by (24), shown at the bottom of the page. Then, the bisection method[1] can be applied to find out θ^* that maximizes $\mathrm{EE}(\theta)$ with polynomial complexity.

V. Simulation Results

Here, we numerically evaluate the proposed energy-efficient power allocation strategy, which is labeled as "EEPA." Moreover, another strategy that uses full power P for maximizing the SE of the system is also presented, which is labeled as "MaxSE." This "MaxSE" strategy is actually the solution of the inner optimization problem (16) with $\theta = 1$. For the comparison between NOMA and conventional OMA, we use a TDMA system as a baseline, where the time slots with equal duration are individually allocated to users and the transmit power is fixed, of which the maximum EE is obtained via exhaustive search on the transmit power.

We solve problem (5) for 10 000 times with random channel realizations. The parameter setting is: $\{g_k\}_{k=1}^K \sim \mathcal{CN}(0,1)$, $\alpha = 3$, $\sigma^2 = -70$ dBm, and $P_c = 30$ dBm. In particular, when the total power P is not large enough for guaranteeing all users' minimum required date rates, the BS will not send messages, and the EE is set to zero for this case.

Fig. 1 shows the average EE versus P. We can see that there exists a "Green Point" at which the maximum EE is achieved by both "EEPA" and "MaxSE" strategies. When P is smaller than the Green Point's

[1]Problem (23) can be also solved by Dinkelbach's algorithm or Charnes–Cooper transform (see, e.g., [7] and references therein).

$$\frac{d\mathrm{EE}(\theta)}{d\theta} = \frac{\frac{1}{\ln 2}\left(\frac{C_K \frac{da_K^*(\theta)}{d\theta}}{\sigma^2 + C_K a_K^*(\theta)}\right)(\theta P + P_c) - \left[\log_2(C_1\theta/\sigma^2+1) + \sum_{k=1}^{K-1} F_k\left(x_k^*(\theta)\right)\right] P}{(\theta P + P_c)^2} \quad (24)$$

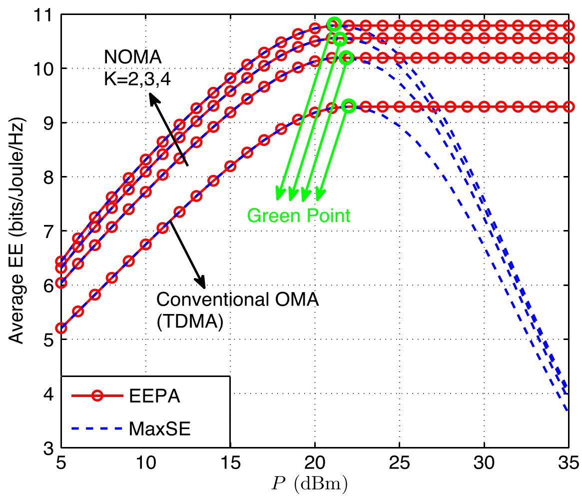

Fig. 1. Average EE (bits/J/Hz) versus total power available at the BS P (dBm). $R_k^{\mathrm{Min}} = 1$ bits/s/Hz, and $d_k = 80$ m, where $1 \le k \le K$.

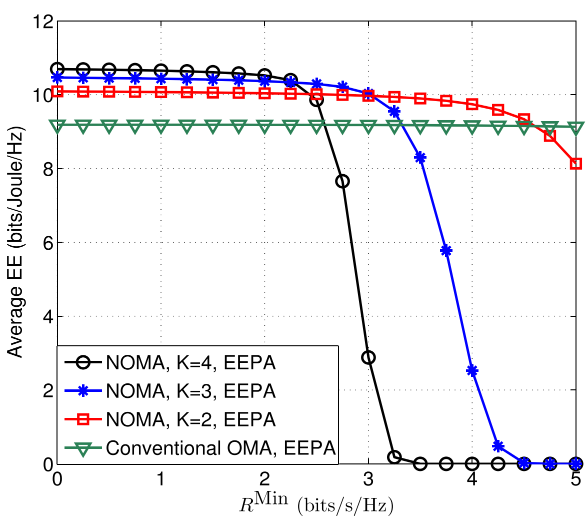

Fig. 2. Average EE versus minimum required data rate R^{Min} for different numbers of users, where $\{R_k^{\mathrm{Min}}\}_{k=1}^K = R^{\mathrm{Min}}$, and $P = 20$ dBm.

corresponding power on the horizontal axis, the increase in SE will simultaneously bring an increase in EE. However, when P is larger, using full power P is not optimal from the perspective of EE. Moreover, NOMA is superior to OMA in terms of EE, and the performance gains of NOMA become more significant as K increases because when more users are simultaneously served, higher diversity gains and higher SE can be achieved.

By setting $\{R_k^{\mathrm{Min}}\}_{k=1}^K$ to the same value, which is denoted by R^{Min}, Fig. 2 shows the average EE versus R^{Min}. We can see that as R^{Min} increases, it is more difficult to achieve a high EE value because the increase of R^{Min} requires the BS to allocate more power to the users with worse channel conditions, which consequentially degrades the EE performance. It can be further seen that as R^{Min} becomes very large, the EE approaches zero faster for NOMA. This is because P is not large enough for satisfying the highly demanding data rate requirements. Moreover, the BS does not send messages, which implies that NOMA is more suitable for low-rate communications and less robust for the increase in data rate requirements in comparison with conventional OMA.

Fig. 3 shows the influence of user locations on EE. First, there is no doubt that the system must have a low EE when all users

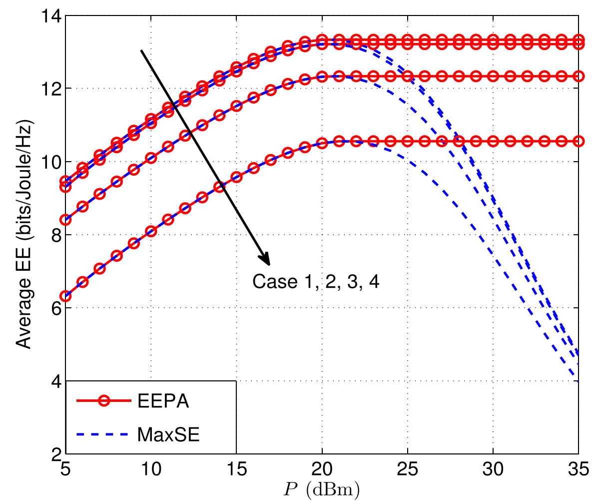

Fig. 3. Average EE (bits/J/Hz) versus total power available at the BS P (dBm), for different cases of user locations. Case 1: $d_1 = 60$ m, $d_2 = 50$ m, $d_3 = 40$ m, $(d_1 + d_2 + d_3)/3 = 50$ m. Case 2: $d_1 = 70$ m, $d_2 = 55$ m, $d_3 = 40$ m, $(d_1 + d_2 + d_3)/3 = 55$ m. Case 3: $d_1 = 60$ m, $d_2 = 55$ m, $d_3 = 50$ m, $(d_1 + d_2 + d_3)/3 = 55$ m. Case 4: $d_1 = 80$ m, $d_2 = 80$ m, $d_3 = 80$ m, $(d_1 + d_2 + d_3)/3 = 80$ m.

locate far from the BS (shown as case 4). More importantly, we can see that 1) case 1 and case 2 have very close EE and that 2) case 2 outperforms case 3, although they have an equal average user distance. These observations imply that the EE performance is mainly determined by the user with the closest distance to the BS since this user is most likely to have the largest channel gain to use energy most efficiently, which validates our analysis in Section IV-A.

VI. Conclusion

In this correspondence, we have studied the EE optimization in a SISO NOMA system wherein multiple users have their own data rate requirements. An energy-efficient power allocation strategy has been proposed to maximize the EE. Our numerical results have shown that NOMA has superior EE performance compared with conventional OMA because, in NOMA, multiple users are simultaneously served via power-domain division, which allows energy to be more efficiently used.

Appendix A
Proof of Proposition 1

Since problem (19) is convex, the following KKT conditions are necessary and sufficient for the optimality of problem (19):

$$\lambda = \begin{cases} \mu_k - \sum_{i=1}^{k-1} \mu_i A_i, & 1 \le k \le K_0 \\ \mu_k - \sum_{i=1}^{k-1} \mu_i A_i + 1, & K_0 + 1 \le k \le K \end{cases} \tag{25}$$

$$\mu_k \left[A_k \left(\sum_{i=k+1}^{K} a_i + \frac{\sigma^2}{P|h_k|^2} \right) - a_k \right] = 0,\ 1 \le k \le K \tag{26}$$

$$\mu_k \ge 0, \quad 1 \le k \le K \tag{27}$$

where λ and $\{\mu_k\}_{k=1}^K$ are the Lagrange multipliers for constraints (19b) and (19c), respectively. In the following, we prove that $\{\mu_k\}_{k=1}^{K_0}$ are positive numbers, which is equivalent to that the constraints in (19c) are active for $1 \le k \le K_0$.

First, we demonstrate $\mu_1 > 0$ by contradiction: Suppose $\mu_1 = 0$, then we have $\lambda = \mu_1 = 0$ by setting $k = 1$ in (25). Accordingly, for

$1 \leq k \leq K_0$ in (25), we can further obtain $\mu_k = \sum_{i=1}^{k-1} \mu_i A_i$, which indicates that $\{\mu_k\}_{k=1}^{K_0}$ are all zeros, since $\mu_k = 0$ can be calculated in the order $k = 2, 3, \ldots, K_0$. However, by setting $k = K_0 + 1$ in (25), we have

$$\mu_1 = \lambda = \mu_{K_0+1} - \sum_{i=1}^{K_0} \mu_i A_i + 1 = \mu_{K_0+1} + 1 > 0 \tag{28}$$

which contradicts the assumption that $\mu_1 = 0$. As a result, we have proved that $\lambda = \mu_1 > 0$.

Afterward, for $2 \leq k \leq K_0$ in (25), we have $\mu_k = \sum_{i=1}^{k-1} \mu_i A_i + \lambda$, which indicates that $\mu_k > 0$ for $2 \leq k \leq K_0$. Thereby, constraints (19c) must be active for $1 \leq k \leq K_0$. We set constraints (19c) to be active for $1 \leq k \leq K_0$ and replace $\sum_{i=k+1}^{K} a_i$ by $(\theta - \sum_{i=1}^{k} a_i)$ in (19c), then the closed-form expressions of $\{a_k\}_{k=1}^{K_0}$ and x_{K_0} are derived and given by (20a) and (20b), respectively. Specifically, $\{a_k\}_{k=1}^{K_0}$ are sequentially calculated in the order $k = 1, 2, \ldots, K_0$.

Appendix B
Proof of Corollary 1

First, $da_k^*(\theta)/d\theta$ can be derived from (21), i.e.,

$$\frac{da_k^*(\theta)}{d\theta} = \begin{cases} D_k \left(1 - \sum_{i=1}^{k-1} \frac{da_i^*(\theta)}{d\theta}\right), & k \neq K \\ 1 - \sum_{i=1}^{K-1} \frac{da_i^*(\theta)}{d\theta}, & k = K. \end{cases} \tag{29}$$

It can be seen that $0 \leq \sum_{i=1}^{k-1} (da_i^*(\theta)/d\theta) < 1$ is a sufficient condition for $da_k^*(\theta)/d\theta > 0$ due to $0 \leq D_k < 1$ for $1 \leq k \leq K$. In the following, we use mathematical induction to prove

$$0 \leq \sum_{i=1}^{k-1} \frac{da_k^*(\theta)}{d\theta} < 1, \quad 1 \leq k \leq K. \tag{30}$$

It is obvious that (30) holds when $k = 1$. When $k = N + 1$, we have

$$\begin{aligned} \sum_{i=1}^{N} \frac{da_i^*(\theta)}{d\theta} &= \sum_{i=1}^{N-1} \frac{da_i^*(\theta)}{d\theta} + D_N \left(1 - \sum_{i=1}^{N-1} \frac{da_i^*(\theta)}{d\theta}\right) \\ &= (1 - D_N) \sum_{i=1}^{N-1} \frac{da_i^*(\theta)}{d\theta} + D_N. \end{aligned} \tag{31}$$

By using the induction hypothesis that (30) holds when $k = N$, i.e., $0 \leq \sum_{i=1}^{N-1} (da_i^*(\theta)/d\theta) < 1$, we have $0 \leq \sum_{i=1}^{N} (da_i^*(\theta)/d\theta) < 1$. Thereby, we have proved $da_k^*(\theta)/d\theta > 0$. Moreover, $\{da_k^*(\theta)/d\theta\}_{k=1}^{K}$ are sequentially calculated in the order $k = 1, 2, \ldots, K$.

References

[1] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System level performance evaluation of downlink Non-Orthogonal Multiple Access (NOMA)," in *Proc. IEEE Annu. Symp. PIMRC*, London, U.K., Sep. 2013, pp. 611–615.

[2] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[3] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[4] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[5] A. Zappone, P. Lin, and E. A. Jorswieck, "Energy efficiency in secure multi-antenna systems," *IEEE Trans. Signal Process.*, submitted for publication. [Online]. Available: http://arxiv.org/abs/1505.02385

[6] Q. Sun, S. Han, C.-L. I, and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *Proc. IEEE ICC*, London, U.K., Jun. 2015, pp. 2668–2673.

[7] A. Zappone and E. A. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, pp. 185–396, Jan. 2015.

Spectral-Efficient Quadrature Spatial Modulation Cooperative Amplify and Forward Spectrum-Sharing Systems

Ali Afana, Salama Ikki, Raed Mesleh, and Ibrahem Atawi

***Abstract*—Quadrature spatial modulation (QSM) is a recent multiple-input–multiple-output (MIMO) digital transmission paradigm. Combining QSM with cooperative relaying in spectrum-sharing systems improves the overall spectral efficiency and enhances the communication reliability. In this paper, we study the performance of QSM-MIMO amplify-and-forward (AF) cooperative relaying spectrum-sharing systems, in which a multiantenna secondary source communicates with a secondary receiver with the help of a secondary AF relay in the presence of multiple primary receivers. In particular, a closed-form expression for the average pairwise error probability (PEP) of the secondary system is derived and used to obtain a tight upper bound of the average bit error probability (ABEP) over Rayleigh fading channels. In addition, a simple asymptotic, yet accurate, expression is derived and analyzed to show the effect of key parameters. Simulation results are presented to validate numerical analysis. Results reveal that QSM with cooperative relaying improves the spectrum-sharing systems' performance.**

***Index Terms*—Amplify and forward (AF), cognitive radio (CR), multiple-input–multiple-output (MIMO) systems, quadrature spatial modulation (QSM), spectrum sharing.**

I. Introduction

Quadrature spatial modulation (QSM) and spectrum sharing in cognitive radio (CR) systems are emerging technologies for next-generation (5G) wireless networks [1]. Both techniques promise significant enhancement in the overall system performance without sacrificing power or bandwidth. Hence, studying the performance of the QSM-CR systems is timely and of significant importance for future development. The transmission schemes of QSM [2], spatial modulation (SM) [3], and space-shift keying (SSK) [4] are proposed as low-complexity and spectral-efficient implementations that avoid conventional multiple-input–multiple-output (MIMO) systems' drawbacks, including inter-channel interference and high receiver complexity [5], [6].

Manuscript received February 10, 2016; revised April 30, 2016; accepted June 4, 2016. Date of publication June 14, 2016; date of current version March 10, 2017. The review of this paper was coordinated by Dr. J.-C. Chen. This work was supported in part by NSERC-DG grant and Lakehead University Research Development Fund.

A. Afana and S. Ikki are with the Department of Electrical Engineering, Faculty of Engineering, Lakehead University, Thunder Bay, ON P7B 5E1, Canada (e-mail: aafana@lakeheadu.ca; sikki@lakeheadu.ca).

R. Mesleh is with the Electrical and Communication Engineering Department, School of Electrical Engineering and Information Technology, German Jordanian University, Amman 11180, Jordan (e-mail: raed.mesleh@gju.edu.jo).

I. Atawi is with the Department of Electrical Engineering, Faculty of Engineering, University of Tabuk, Tabuk 71491, Saudi Arabia (e-mail: ieatawi@ut.edu.sa).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2016.2580512