# JOINT DEVICE-TO-DEVICE TRANSMISSION ACTIVATION AND TRANSCEIVER DESIGN FOR SUM-RATE MAXIMIZATION IN MIMO INTERFERING CHANNELS

*Jingran Lin**, *Qingjiang Shi*†, *and Qiang Li**

*School of Comm. & Info. Eng., University of Elect. Sci. & Tech. of China, Chengdu China, 611731
†School of Info. & Sci. Tech., Zhejiang Sci-Tech University, Hangzhou China, 310018

**ABSTRACT**

Consider a network that consists of one multi-antenna base station (BS) and multiple pairs of multi-antenna user equipments (UEs). In each UE pair, the communication between transmitter and receiver is established either through BS or via device-to-device (D2D) link. All the D2D transmission and the uplink transmission of BS relaying share the same resources, while causing interference to each other. To improve the network throughput, we consider a sum-rate maximization problem by jointly optimizing the transmission mode of each UE pair and the corresponding transceivers. Due to the NP-hardness of the problem, we seek for some efficient approximate solutions to it. To this end, we first reformulate the problem by the weighted MMSE (WMMSE) approach, and then fit it into the alternating direction method of multipliers (ADMM) framework. Finally, an efficient distributed algorithm, which converges to a stationary solution, is developed. In particular, each step of the algorithm can be computed in closed form, thus giving it very low complexity.

***Index Terms***— Device-to-device activation, transceiver design, sum-rate, MIMO interfering channel, alternating direction method

## 1. INTRODUCTION

As a promising technology in future wireless networks, device-to-device (D2D) communication can considerably improve the network throughput, power efficiency and reliability [1]. However, many critical challenges such as D2D mode switch and interference management arise at the same time to achieve these potential benefits [2]. That is, we should properly activate the D2D transmission for some users and carefully design the transceivers.

So far, a variety of schemes have been proposed to manage D2D communications. Most early works consider the D2D mode switch and power control problems in SISO networks consisting of one cellular user and one pair of D2D users [3, 4]. They usually choose the best transmission mode and power allocation by exhaustive search. In multi-user networks, more advanced techniques such as game theory [5], graph theory [6], and nonlinear optimization [7] are adopted to help select the D2D communication mode and mitigate the interference. These problems are challenging mixed-integer programs in general, and thus many heuristic algorithms [7, 8] are proposed to reduce the complexity. Although their performances have been validated by simulations, there is still a lack of theoretic support to these heuristic methods. As far as the D2D MIMO transmission is concerned, beamforming is integrated for further performance improvement. Many works employ zero-forcing (ZF) beamforming to avoid interference between cellular users and D2D users [9, 10]. In [11], a more complicated scheme is proposed where the transceivers for one cellular user and one D2D pair are jointly designed. It outperforms the ZF strategy, but does not consider the D2D mode switch.

In this paper, we consider a network consisting of one base station (BS) and multiple pairs of user equipments (UEs), where the BS and UEs are all equipped with multiple antennas. Each UE pair includes one transmitter (UE_T) and one receiver (UE_R). Information is forwarded from UE_Ts to their UE_Rs either through BS or over D2D link. We assume the interfering channel (IC) model and the D2D transmission reuses the uplink resources of BS relaying. In this scenario, we aim at maximizing the network throughput by joint D2D transmission activation and linear transceivers design.

Unfortunately, this problem is NP-hard and thus we turn to finding some efficient approximate solutions. Specifically, we first apply the weighted MMSE (WMMSE) reformulation [12, 13] to our problem such that it can be solved via alternating optimization. Considering that a distributed algorithm is preferred for multi-user networks, we further fit the problem into the framework of alternating direction method of multipliers (ADMM) [14, 15] and develop a distributed algorithm finally. With each step being computed in closed form, the algorithm is highly efficient and converges to a stationary solution.

## 2. SYSTEM MODEL AND PROBLEM STATEMENT

Consider a network consisting of one BS and $M$ UE pairs (one UE_T and one UE_R). The BS is equipped with $N_b$ antennas, and each UE with $N_u$ antennas. The UE_Ts transmit data to their UE_Rs in either BS relaying mode or D2D mode. Specifically, in BS relaying mode, the BS employs a decode-and-forward strategy [15] in data relaying. We assume the D2D communication shares the uplink resources of BS relaying. When the D2D transmission of some UE pair is activated, the UE_R switches to the uplink timeslot and/or frequency band to receive data from UE_T directly.

Define $\mathbf{G}_{b,m} \in \mathbb{C}^{N_b \times N_u}$ and $\mathbf{H}_{m,b} \in \mathbb{C}^{N_u \times N_b}$ as the orthogonal uplink and downlink channels between UE pair $m$ and BS, respectively; $\mathbf{F}_{m,n} \in \mathbb{C}^{N_u \times N_u}$ as the D2D channel between UE_T$n$ and UE_R$m$ for $m, n = 1, 2, \cdots, M$. Let $\mathbf{v}_m \in \mathbb{C}^{N_u \times 1}$ denote the transmit beamformer of UE_T$m$; $\mathbf{u}_{b,m}$ and $\mathbf{u}_{d,m} \in \mathbb{C}^{N_u \times 1}$ denote the receive beamformers of UE_R$m$ in BS relaying mode and D2D mode, respectively. Let $\mathbf{x}_m$ and $\mathbf{z}_m \in \mathbb{C}^{N_b \times 1}$ denote the receive and transmit beamformers of BS for UE_T$m$ and UE_R$m$, respectively. Let $p_m$ denote the transmit power budget of UE_T$m$ for $m = 1, 2, \cdots, M$, and $P_b$ is the total power budget of BS. Define $\sigma^2$ as the power of noise at BS and UE_Rs.

In D2D mode, the rate of UE pair $m$ can be written as

$$R_{d,m} = \log \left[ 1 + \frac{|\mathbf{u}_{d,m}^H \mathbf{F}_{m,m} \mathbf{v}_m|^2}{\sum_{n \neq m} |\mathbf{u}_{d,m}^H \mathbf{F}_{m,n} \mathbf{v}_n|^2 + \sigma^2 \|\mathbf{u}_{d,m}\|_2^2} \right] \quad (1)$$

This work was supported in part by the national NSFC, Grant 61302076, 61401073 and 61471103, and by the Applied Basic Research Programs of Sichuan Province, China, Grant 2015JY0102.

where $(\cdot)^H$ denotes the conjugate transpose. Similarly, in BS relaying mode, the uplink and downlink rates of UE pair $m$ are

$$R_{b,m}^{ul} = \log\left[1 + \frac{|\mathbf{x}_m^H \mathbf{G}_{b,m}\mathbf{v}_m|^2}{\sum_{n\neq m}|\mathbf{x}_m^H \mathbf{G}_{b,n}\mathbf{v}_n|^2 + \sigma^2\|\mathbf{x}_m\|_2^2}\right] \quad (2)$$

$$R_{b,m}^{dl} = \log\left[1 + \frac{|\mathbf{u}_{b,m}^H \mathbf{H}_{m,b}\mathbf{z}_m|^2}{\sum_{n\neq m}|\mathbf{u}_{b,m}^H \mathbf{H}_{n,b}\mathbf{z}_n|^2 + \sigma^2\|\mathbf{u}_{b,m}\|_2^2}\right] \quad (3)$$

We further define an $M\times 1$ binary vector $\mathbf{d} = [d_1, d_2, \cdots, d_M]^T$ to indicate the communication modes of the UE pairs. UE pair $m$ works in D2D mode if $d_m = 1$, and in BS relaying mode if $d_m = 0$. The achievable rate of UE pair $m$ can thus be expressed as

$$R_m = d_m R_{d,m} + (1-d_m)\min\{R_{b,m}^{ul}, R_{b,m}^{dl}\},\ \ \forall m \quad (4)$$

**Remark 1** *To simplify the problem formulation, we assume that the D2D pairs keep idle during the BS downlink transmission. The full time D2D problem will be discussed in the journal version.*

Since the links between UEs are not connected to BS, obtaining $\mathbf{F}_{m,n}$'s is relatively difficult and requires heavy backhaul overhead. To alleviate the network overhead, we further consider limiting the number of active D2D pairs from the sparse optimization perspective. Now our sum-rate maximization problem is described as

$$\max_{\{\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Z},\mathbf{d}\}} \sum_{m=1}^{M} R_m - \lambda\|\mathbf{d}\|_0 \quad \text{(P1)}$$

$$\text{s.t.} \quad \sum_{m=1}^{M}\|\mathbf{z}_m\|_2^2 \le P_b,\ \|\mathbf{v}_m\|_2^2 \le p_m,\ \forall m \quad (5)$$

$$d_m \in \{0,1\},\quad \forall m \quad (6)$$

where $\mathbf{U}$, $\mathbf{V}$, $\mathbf{X}$ and $\mathbf{Z}$ are the collections of $\{\mathbf{u}_{d,m}, \mathbf{u}_{b,m}\}_{m=1}^M$, $\{\mathbf{v}_m\}_{m=1}^M$, $\{\mathbf{x}_m\}_{m=1}^M$ and $\{\mathbf{z}_m\}_{m=1}^M$, respectively; $\lambda > 0$ is the tunable factor to control the number of active D2D pairs.

**Remark 2** *Note in $R_{b,m}^{dl}$, we assume that the BS transmits data to all the UE_Rs. If UE_Rm receives data via D2D link, i.e., $d_m = 1$, then $\mathbf{z}_m$ will automatically be optimized to zero to reduce its interference to other UE_Rs and thus improve the throughput.*

## 3. PROBLEM REFORMULATION AND ALGORITHM

It is well known that the sum-rate problem in MU-MIMO networks is NP-hard [16]. In our formulation, the additional binary variables $d_m$'s further complicate the problem, rendering (P1) more challenging to solve. As a compromise, we try to find some efficient approximate solutions to it. To this end, we first relax $d_m$ as a continuous variable $0 \le d_m \le 1$, and approximate the $l_0$-norm by $l_1$-norm.

We next deal with the complicated $R_m$ terms. Noticing the function $h(\omega,\nu) = \omega^2\nu^{-1}$ is convex for $\nu > 0$ [17], we introduce auxiliary variables $\{t_{d,m}, t_{b,m}\}_{m=1}^M$, and then (P1) is reformulated as

$$\max_{\{\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Z},\mathbf{d},\mathbf{t}\}} \sum_{m=1}^{M}\left(t_{d,m}^2 + t_{b,m}^2 - \lambda|d_m|\right) \quad \text{(P2)}$$

s.t. (5)

$$0 \le d_m \le 1,\ \forall m \quad (7)$$

$$R_{d,m} \ge \frac{t_{d,m}^2}{d_m + \epsilon},\ \forall m \quad (8)$$

$$\min\{R_{b,m}^{ul}, R_{b,m}^{dl}\} \ge \frac{t_{b,m}^2}{1 - d_m + \epsilon},\ \forall m \quad (9)$$

where $\mathbf{t}$ is the collection of $\{t_{d,m}, t_{b,m}\}_{m=1}^M$; $\epsilon$ is a small positive number used to handle the numerical problems of zero denominator.

We further utilize the WMMSE technique [12,13] to reformulate the $\log(\cdot)$ terms in $R_m$, and (P2) can be approximated by

$$\max_{\{\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Z},\mathbf{W},\mathbf{d},\mathbf{t}\}} \sum_{m=1}^{M}\left(t_{d,m}^2 + t_{b,m}^2 - \lambda|d_m|\right) \quad \text{(P3)}$$

s.t. (5) and (7)

$$\log(w_{d,m}) - w_{d,m}e_{d,m} + 1 \ge \frac{t_{d,m}^2}{d_m + \epsilon},\ \forall m \quad (10)$$

$$\log(w_{b,m}^{ul}) - w_{b,m}^{ul}e_{b,m}^{ul} + 1 \ge \frac{t_{b,m}^2}{1 - d_m + \epsilon},\ \forall m \quad (11)$$

$$\log(w_{b,m}^{dl}) - w_{b,m}^{dl}e_{b,m}^{dl} + 1 \ge \frac{t_{b,m}^2}{1 - d_m + \epsilon},\ \forall m \quad (12)$$

where $\{e_{d,m}, e_{b,m}^{ul}, e_{b,m}^{dl}\}$ are the MSE values [12] of UE pair $m$ in D2D link, BS relaying uplink and BS relaying downlink, respectively; $\mathbf{W}$ is the collection of weight factors $\{w_{d,m}, w_{b,m}^{ul}, w_{b,m}^{dl}\}_{m=1}^M$.

Next, we develop an iterative algorithm for (P3). Instead of directly maximizing the non-concave objective of $(t_{d,m}^2 + t_{b,m}^2)$, we solve sequential concave approximations by exploiting the idea of difference-of-convex (DC) programming [18]. Specifically, in each DC programming iteration, we solve the problem

$$\min_{\{\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Z},\mathbf{W},\mathbf{d},\mathbf{t}\}} obj_{\text{DC}} \quad \text{(P4)}$$

$$\text{s.t.} \quad obj_{\text{DC}} = \sum_{m=1}^{M}(\lambda|d_m| - 2\hat{t}_{d,m}t_{d,m} - 2\hat{t}_{b,m}t_{b,m})$$

(5), (7), (10), (11) and (12)

where $\hat{t}_{d,m}$ and $\hat{t}_{b,m}$ are the iterates of $t_{d,m}$ and $t_{b,m}$ in the previous iteration, respectively.

Applying the alternating optimization framework of WMMSE, (P3) can be solved by Algorithm-1 given in Table 1.

**Table 1:** Algorithm-1 Summary

1) Initialize $\{t_{d,m}, t_{b,m}\}$;
2) **Repeat** (*outer loop*)
3) $\{\hat{t}_{d,m}, \hat{t}_{b,m}\} \leftarrow \{t_{d,m}, t_{b,m}\},\ \forall m$;
4) Initialize $\mathbf{v}_m$'s;
5) **Repeat** (*inner loop*)
6) $\mathbf{u}_{d,m} \leftarrow (\sum_n \mathbf{F}_{m,n}\mathbf{v}_n\mathbf{v}_n^H\mathbf{F}_{m,n}^H + \sigma^2\mathbf{I})^{-1}\mathbf{F}_{m,m}\mathbf{v}_m$;
7) $\mathbf{x}_m \leftarrow (\sum_n \mathbf{G}_{b,n}\mathbf{v}_n\mathbf{v}_n^H\mathbf{G}_{b,n}^H + \sigma^2\mathbf{I})^{-1}\mathbf{G}_{b,m}\mathbf{v}_m$;
8) $\mathbf{u}_{b,m} \leftarrow (\sum_n \mathbf{H}_{n,b}\mathbf{z}_n\mathbf{z}_n^H\mathbf{H}_{n,b}^H + \sigma^2\mathbf{I})^{-1}\mathbf{H}_{m,b}\mathbf{z}_m$;
9) $w_{d,m} \leftarrow (1 - \mathbf{u}_{d,m}^H\mathbf{F}_{m,m}\mathbf{v}_m)^{-1}$;
10) $w_{b,m}^{ul} \leftarrow (1 - \mathbf{x}_m^H\mathbf{G}_{b,m}\mathbf{v}_m)^{-1}$;
11) $w_{b,m}^{dl} \leftarrow (1 - \mathbf{u}_{b,m}^H\mathbf{H}_{m,b}\mathbf{z}_m)^{-1}$;
12) $\{\mathbf{V},\mathbf{Z},\mathbf{d},\mathbf{t}\} \leftarrow \text{argmin}_{\{\mathbf{V},\mathbf{Z},\mathbf{d},\mathbf{t}\}}\ obj_{\text{DC}}$
s.t. (5), (7), (10), (11), (12);
13) **Until** $\{\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Z},\mathbf{W},\mathbf{d},\mathbf{t}\}$ converges;
14) **Until** $\|\mathbf{t} - \hat{\mathbf{t}}\|^2 <$ some given threshold;
15) Quantize $\{d_m\}$ to identify the UEs' transmission modes;

**Proposition 1** *Every accumulation point of the iterates generated by Algorithm-1 is a stationary solution of the problem (P2).*

*Proof:* Here we just outline the proof due to the space limitation.

First, we show the sequence of $\{\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Z},\mathbf{d},\mathbf{t}\}$, generated by the WMMSE iterations in the inner loop, converges to a stationary solution of (P4). This can be established by checking the KKT conditions of (P4) at any limit point of $\{\mathbf{U},\mathbf{V},\mathbf{X},\mathbf{Z},\mathbf{d},\mathbf{t}\}$. Second, we show the outer loop produces non-decreasing objective values of (P3). Since the feasible set is compact and continuous, we conclude the convergence of the DC programming iterations. By checking the KKT conditions at the limit points of the outer loop, we claim that Algorithm-1 converges to a stationary point of (P3). Lastly, based on the properties of WMMSE [12,13], we know the stationary point of (P3) is also a stationary point of (P2). □

Except for step 12), Algorithm-1 can be implemented distributively. In the next section, we further develop a distributed algorithm for step 12) by fitting it into the ADMM framework.

## 4. DISTRIBUTED ALGORITHM BASED ON ADMM

To decouple $\mathbf{v}_m$'s and $\mathbf{z}_m$'s in $R_{d,m}$'s, $R^{ul}_{b,m}$'s and $R^{dl}_{b,m}$'s, we introduce $2M$ copies of $\{\mathbf{v}_m\}_{m=1}^M$ and $M$ copies of $\{\mathbf{z}_m\}_{m=1}^M$. Specifically, we introduce $\mathbf{q}^d_{m,n} = \mathbf{v}_m$, $\mathbf{q}^b_{m,n} = \mathbf{v}_m$ and $\mathbf{y}_{m,n} = \mathbf{z}_m$ for $m, n = 1, 2, \cdots, M$.

We next introduce two series of auxiliary indicators, i.e., $s_{d,m} = d_m + \epsilon$ and $s_{b,m} = 1 - d_m + \epsilon$, $m = 1, 2, \cdots, M$.

We further introduce three series of intermediate variables, i.e., $r_{d,m} = a_{d,m}$, $r^{ul}_{b,m} = a_{b,m}$ and $r^{dl}_{b,m} = a_{b,m}$, $m = 1, 2, \cdots, M$.

Updating $\{\mathbf{V}, \mathbf{Z}, \mathbf{d}, \mathbf{t}\}$ in step 12) is then equivalent to solving the following problem

$$\min_{\{\mathbf{V},\mathbf{Z},\mathbf{d},\mathbf{t},\mathbf{Q},\mathbf{Y},\mathbf{s},\mathbf{r},\mathbf{a}\}} \sum_{m=1}^{M} (\lambda|d_m| - 2\hat{t}_{d,m}t_{d,m} - 2\hat{t}_{b,m}t_{b,m}) \quad \text{(P5)}$$

$$\begin{aligned}
\text{s.t.}\quad & (5) \text{ and } (7),\\
& \log(w_{d,m}) - w_{d,m}e_{d,m}(\mathbf{Q}^d_m) + 1 \geq r_{d,m},\ \forall m \quad (13)\\
& \log(w^{ul}_{b,m}) - w^{ul}_{b,m}e^{ul}_{b,m}(\mathbf{Q}^b_m) + 1 \geq r^{ul}_{b,m},\ \forall m \quad (14)\\
& \log(w^{dl}_{b,m}) - w^{dl}_{b,m}e^{dl}_{b,m}(\mathbf{Y}_m) + 1 \geq r^{dl}_{b,m},\ \forall m \quad (15)\\
& a_{d,m} \geq t^2_{d,m}s^{-1}_{d,m},\ s_{d,m} \geq \epsilon,\ \forall m \quad (16)\\
& a_{b,m} \geq t^2_{b,m}s^{-1}_{b,m},\ s_{b,m} \geq \epsilon,\ \forall m \quad (17)\\
& \mathbf{q}^d_{m,n} = \mathbf{v}_m,\ \mathbf{q}^b_{m,n} = \mathbf{v}_m,\ \mathbf{y}_{m,n} = \mathbf{z}_m,\ \forall m, n\\
& r_{d,m} = a_{d,m},\ r^{ul}_{b,m} = a_{b,m},\ r^{dl}_{b,m} = a_{b,m},\ \forall m\\
& s_{d,m} = d_m + \epsilon,\ s_{b,m} = 1 - d_m + \epsilon,\ \forall m
\end{aligned}$$

where $\mathbf{Q}^d_m$ is the collection of $\{\mathbf{q}^d_{l,m}\}_{l=1}^M$; similarly, $\mathbf{Q}^b_m$, $\mathbf{Y}_m$, $\mathbf{s}$, $\mathbf{r}$ and $\mathbf{a}$ are the collections of respective variables.

The partial augmented Lagrangian function [14] of (P5) can be defined as (18), where $c > 0$ is the penalty factor, and $\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\theta}, \boldsymbol{\tau}$ and $\boldsymbol{\kappa}$ are the associated Lagrangian multipliers. Dividing the variables $\{\mathbf{V}, \mathbf{Z}, \mathbf{d}, \mathbf{t}, \mathbf{Q}, \mathbf{Y}, \mathbf{s}, \mathbf{r}, \mathbf{a}\}$ into two blocks of $\{\mathbf{V}, \mathbf{Z}, \mathbf{t}, \mathbf{s}, \mathbf{a}\}$ and $\{\mathbf{Q}, \mathbf{Y}, \mathbf{r}, \mathbf{d}\}$, we can apply the ADMM framework [14] in Table 2 to solve (P5) iteratively.

In ADMM framework, (P5) is divided into two simple problems, i.e. updating $\{\mathbf{V}, \mathbf{Z}, \mathbf{t}, \mathbf{s}, \mathbf{a}\}$ and updating $\{\mathbf{Q}, \mathbf{Y}, \mathbf{r}, \mathbf{d}\}$. Moreover,

**Table 2:** ADMM framework for solving (P5)

| |
|---|
| ***Repeat*** |
| Update $\{\mathbf{V}, \mathbf{Z}, \mathbf{t}, \mathbf{s}, \mathbf{a}\}$ *with other variables fixed* |
| $\{\mathbf{V}, \mathbf{Z}, \mathbf{t}, \mathbf{s}, \mathbf{a}\} \leftarrow \text{argmin}_{\{\mathbf{V},\mathbf{Z},\mathbf{t},\mathbf{s},\mathbf{a}\}} L_c(\cdot)$ s.t. (5), (16) and (17) |
| Update $\{\mathbf{Q}, \mathbf{Y}, \mathbf{r}, \mathbf{d}\}$ *with other variables fixed* |
| $\{\mathbf{Q}, \mathbf{Y}, \mathbf{r}, \mathbf{d}\} \leftarrow \text{argmin}_{\{\mathbf{Q},\mathbf{Y},\mathbf{r},\mathbf{d}\}} L_c(\cdot)$ s.t. (7), (13), (14) and (15) |
| Update Lagrangian variables $\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\theta}, \boldsymbol{\tau}$ and $\boldsymbol{\kappa}$ |
| ***Until*** the iterations converge |

these two problems can be further separated into smaller and simpler problems. For instance, updating $\{\mathbf{V}, \mathbf{Z}, \mathbf{t}, \mathbf{s}, \mathbf{a}\}$ can be divided into $(3M + 1)$ smaller problems, which update $\{\mathbf{v}_m\}_{m=1}^M$, $\{\mathbf{Z}\}$, $\{a_{d,m}, t_{d,m}, s_{d,m}\}_{m=1}^M$ and $\{a_{b,m}, t_{b,m}, s_{b,m}\}_{m=1}^M$, respectively; updating $\{\mathbf{Q}, \mathbf{Y}, \mathbf{r}, \mathbf{d}\}$ can be divided into $4M$ smaller problems, which update $\{\mathbf{Q}^d_m, r_{d,m}\}_{m=1}^M$, $\{\mathbf{Q}^b_m, r^{ul}_{b,m}\}_{m=1}^M$, $\{\mathbf{Y}_m, r^{dl}_{b,m}\}_{m=1}^M$ and $\{d_m\}_{m=1}^M$, respectively. More interestingly, these smaller problems can all be calculated with closed-form solutions.

### 4.1. Updating $\{\mathbf{V}, \mathbf{Z}, \mathbf{t}, \mathbf{s}, \mathbf{a}\}$

Due to the separable problem structure, we update $\{\mathbf{v}_m\}_{m=1}^M$, $\{\mathbf{Z}\}$, $\{a_{d,m}, t_{d,m}, s_{d,m}\}_{m=1}^M$ and $\{a_{b,m}, t_{b,m}, s_{b,m}\}_{m=1}^M$ independently.

The problem of $\mathbf{v}_m$ is a simple convex quadratically constrained quadratic program (QCQP), and can be solved directly as

$$\mathbf{v}_m = \frac{\sum_{n=1}^M \left[\boldsymbol{\phi}^d_{m,n} + \boldsymbol{\phi}^b_{m,n} + c\left(\mathbf{q}^b_{m,n} + \mathbf{q}^b_{m,n}\right)\right]}{2(cM + \alpha_m)},\ \forall m \quad (19)$$

$$\alpha_m = \left[\frac{\|\boldsymbol{\xi}_m\|_2}{2\sqrt{p_m}} - cM\right]^+,\ \forall m \quad (20)$$

$$\boldsymbol{\xi}_m = \sum\nolimits_{n=1}^M [\boldsymbol{\phi}^d_{m,n} + \boldsymbol{\phi}^b_{m,n} + c(\mathbf{q}^b_{m,n} + \mathbf{q}^b_{m,n})],\ \forall m \quad (21)$$

where $\alpha_m$ is the Lagrangian multiplier of $\|\mathbf{v}_m\|_2^2 \leq p_m$, and $[\omega]^+$ returns the larger value between 0 and $\omega$.

Similarly, $\mathbf{Z}$ is solved as

$$\mathbf{z}_m = \frac{\sum_{n=1}^M \left(\boldsymbol{\psi}_{m,n} + c\mathbf{y}_{m,n}\right)}{cM + 2\mu},\ \forall m \quad (22)$$

$$\mu = \left[\frac{\sqrt{\sum_{m=1}^M \left\|\sum_{n=1}^M \left(\boldsymbol{\psi}_{m,n} + c\mathbf{y}_{m,n}\right)\right\|_2^2}}{2\sqrt{P_b}} - \frac{cM}{2}\right]^+ \quad (23)$$

where $\mu$ is the Lagrangian multiplier of $\sum_{m=1}^M \|\mathbf{z}_m\|_2^2 \leq P_b$.

$$\begin{aligned}
&L_c(\mathbf{V}, \mathbf{Z}, \mathbf{d}, \mathbf{t}, \mathbf{Q}, \mathbf{Y}, \mathbf{s}, \mathbf{r}, \mathbf{a}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\kappa}) = \sum\nolimits_{m=1}^M \left(\lambda|d_m| - 2\hat{t}_{d,m}t_{d,m} - 2\hat{t}_{b,m}t_{b,m}\right) \quad (18)\\
&+ \sum\nolimits_{(m,n)=(1,1)}^{(M,M)} \left[\text{Re}\left([\boldsymbol{\phi}^d_{m,n}]^H(\mathbf{q}^d_{m,n} - \mathbf{v}_m) + [\boldsymbol{\phi}^b_{m,n}]^H(\mathbf{q}^b_{m,n} - \mathbf{v}_m)\right) + \frac{c}{2}\left(\|\mathbf{q}^d_{m,n} - \mathbf{v}_m\|_2^2 + \|\mathbf{q}^b_{m,n} - \mathbf{v}_m\|_2^2\right)\right]\\
&+ \sum\nolimits_{(m,n)=(1,1)}^{(M,M)} \left[\text{Re}\left(\boldsymbol{\psi}^H_{m,n}(\mathbf{y}_{m,n} - \mathbf{z}_m)\right) + \frac{c}{2}\|\mathbf{y}_{m,n} - \mathbf{z}_m\|_2^2\right]\\
&+ \sum\nolimits_{m=1}^M \left[\theta_m(r_{d,m} - a_{d,m}) + \tau^{ul}_m(r^{ul}_{b,m} - a_{b,m}) + \tau^{dl}_m(r^{dl}_{b,m} - a_{b,m}) + \frac{c}{2}[(r_{d,m} - a_{d,m})^2 + (r^{ul}_{b,m} - a_{b,m})^2 + (r^{dl}_{b,m} - a_{b,m})^2]\right]\\
&+ \sum\nolimits_{m=1}^M \left[\kappa_{d,m}(s_{d,m} - d_m - \epsilon) + \kappa_{b,m}(s_{b,m} - 1 + d_m - \epsilon) + \frac{c}{2}[(s_{d,m} - d_m - \epsilon)^2 + (s_{b,m} - 1 + d_m - \epsilon)^2]\right]
\end{aligned}$$

For the problem of updating $\{a_{d,m}, t_{d,m}, s_{d,m}\}$, we explore the first-order optimality conditions, and then obtain

$$\begin{cases} a_{d,m} = r_{d,m} + c^{-1}(\theta_{d,m} + \beta_{d,m}) \\ s_{d,m} = \max\{d_m + \epsilon + c^{-1}(\hat{t}^2_{d,m}\beta^{-1}_{d,m} - \kappa_{d,m}),\ \epsilon\} \\ t_{d,m} = \hat{t}_{d,m}\beta^{-1}_{d,m}s_{d,m} \end{cases} \tag{24}$$

where $\beta_{d,m}$ is the Lagrangian multipliers of $a_{d,m} \geq t^2_{d,m}s^{-1}_{d,m}$, and should be chosen properly to satisfy the KKT complementarity condition. For our specific problem, $\beta_{d,m}$ can be calculated in closed form by solving a cubic equation with respect to $\beta_{d,m}$. The details are omitted due to the space limitation.

Updating $\{a_{b,m}, t_{b,m}, s_{b,m}\}$ is quite similar as (24).

### 4.2. Updating $\{\mathbf{Q}, \mathbf{Y}, \mathbf{r}, \mathbf{d}\}$

Due to the separable problem structure, we update $\{\mathbf{Q}^d_m, r_{d,m}\}^M_{m=1}$, $\{\mathbf{Q}^b_m, r^{ul}_{b,m}\}^M_{m=1}$, $\{\mathbf{Y}_m, r^{dl}_{b,m}\}^M_{m=1}$ and $\{d_m\}^M_{m=1}$ independently.

Updating $\{\mathbf{Q}^d_m, r_{d,m}\}$ is also a simple QCQP and is solved as

$$\begin{cases} \mathbf{q}^d_{m,m} = \boldsymbol{\chi}^d_{m,m}(c\mathbf{v}_m - \boldsymbol{\phi}^d_{m,m} + 2\delta_{d,m}w_{d,m}\mathbf{F}^H_{m,m}\mathbf{u}_{d,m}) \\ \mathbf{q}^d_{n,m} = \boldsymbol{\chi}^d_{n,m}(c\mathbf{v}_n - \boldsymbol{\phi}^d_{n,m}),\ \forall n \neq m, \\ r_{d,m} = a_{d,m} - c^{-1}(\theta_m + \delta_{d,m}) \\ \boldsymbol{\chi}^d_{m,m} = (2\delta_{d,m}w_{d,m}\mathbf{F}^H_{m,m}\mathbf{u}_{d,m}\mathbf{u}^H_{d,m}\mathbf{F}_{m,m} + c\mathbf{I})^{-1} \\ \boldsymbol{\chi}^d_{n,m} = (2\delta_{d,m}w_{d,m}\mathbf{F}^H_{m,n}\mathbf{u}_{d,m}\mathbf{u}^H_{d,m}\mathbf{F}_{m,n} + c\mathbf{I})^{-1} \end{cases} \tag{25}$$

where $\delta_{d,m}$ is the Lagrangian multiplier of (13), and should be chosen to satisfy the KKT conditions (e.g., by bisection search).

Updating $\{\mathbf{Q}^b_m, r^{ul}_{b,m}\}$ and $\{\mathbf{Y}_m, r^{dl}_{b,m}\}$ is quite similar as (25). Again, we omit the details due to the space limitation.

The problem of $d_m$ is a simple quadratic problem (QP) and can be easily solved as

$$d_m = \left[\frac{c + cs_{d,m} - cs_{b,m} + \kappa_{d,m} - \kappa_{b,m} - \lambda}{2c}\right]^1_0,\ \forall m \tag{26}$$

where $[\,\cdot\,]^1_0$ denotes the projection on the range of $[0, 1]$.

**Remark 3** *Due to the separable structure of (P5) in ADMM framework, it can be solved distributively.*

**Remark 4** *The complexity of the ADMM algorithm is mainly determined by updating $\{\mathbf{Q}^d_m, r_{d,m}\}$, $\{\mathbf{Q}^b_m, r^{ul}_{b,m}\}$ and $\{\mathbf{Y}_m, r^{dl}_{b,m}\}$. Its complexity is about $\mathcal{O}(M^2N^2_u + 2M^2N_uN_b)$, while the complexity of updating $\{\mathbf{V}, \mathbf{Z}, \mathbf{d}, \mathbf{t}\}$ directly, e.g., by the interior-point method, is $\mathcal{O}(M^3(N_u + N_b)^3)$. Obviously, the ADMM algorithm is more efficient and its efficiency will be further improved if the distributed implementation and parallel computation can be exploited.*

## 5. NUMERICAL RESULTS AND CONCLUSIONS

Consider a hexagonal cell with the distance between adjacent corners being $1000m$. There are $M = 10$ UE pairs randomly deployed in the cell and one BS located in the center. Each UE is equipped with $N_u = 2$ antennas, and the BS with $N_b = 4$ antennas. The channel model we use is Rayleigh channel with zero mean and variance $L_s(\frac{200}{dis})^3$, where $dis$ is the distance between two terminals, and $10\log_{10}(L_s) \sim \mathcal{N}(0, 64)$. The noise power is $\sigma^2 = 1$, and the transmit power budgets of BS and UE_T$m$ are $P_b = 3000$ and $p_m = 100, m = 1, 2, \cdots, M$, respectively. The parameter $\epsilon$ is set to 1e-6 in the following simulations.

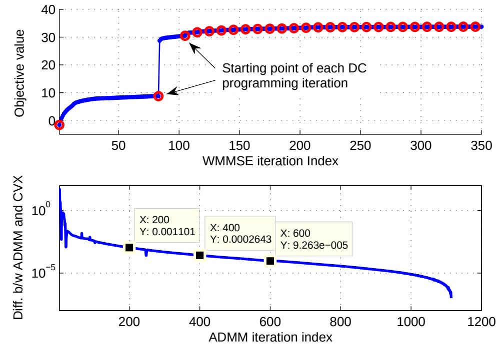


**Fig. 1**: Typical converging traces of the propose algorithm for $\lambda = 0$, $M = 10$, $N_u = 2$, $N_b = 4$, $\sigma^2 = 1$, $P_b = 3000$ and $p_m = 100$.

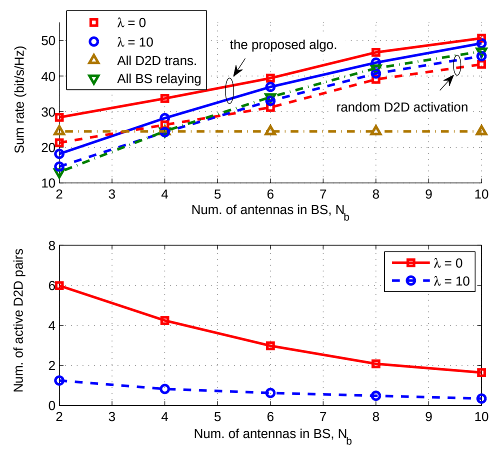


**Fig. 2**: Performance comparison with different BS antenna numbers for $M = 10$, $N_u = 2$, $\sigma^2 = 1$, $P_b = 3000$ and $p_m = 100$.

First, the converging curves of the proposed algorithm are shown in Fig. 1. Typically, the DC programming loop converges in 10 iterations; the WMMSE loop converges in 100 iterations; the ADMM loop converges in 600 iterations, with a stopping criterion (the difference between ADMM and CVX results) of 1e-4.

Next, we compare the performances of 4 algorithms, i.e., (1) the proposed distributed algorithm based on WMMSE and ADMM, (2) the algorithm based on all BS relaying strategy, (3) the algorithm based on all D2D strategy, and (4) the algorithm based on random D2D activation which randomly activates the same number of D2D pairs as that of the proposed algorithm. The results are shown in Fig. 2. The proposed algorithm outperforms the other three algorithms in sum-rate substantially by performing a network-wise optimization. As expected, less D2D UE pairs will be activated for larger value of $\lambda$. Moreover, as $N_b$ increases, more UE pairs tend to adopt the BS relaying mode to benefit from the extra spatial diversity.

In summary, we aim to maximize the network throughput in this paper via joint D2D transmission activation and transceiver design. Our scenario is featured by multi-user, MIMO IC and equally prioritized D2D transmission and BS relaying. After some approximation works, we develop an algorithm, based on WMMSE and ADMM, to find a stationary solution efficiently and distributively. In particular, each step of the algorithm can be computed in closed form.

## 6. REFERENCES


[1] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Survey & Tutorials*, vol. 16, no. 4, pp. 1801–1819, Nov. 2014.

[2] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Enable device-to-device communications underlaying cellular networks: challenges and research aspects," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 90–96, Jun. 2014.

[3] H. Min, W. Seo, J. Lee, S. Park, and D. Hong, "Reliability improvement using receive mode selection in device-to-device uplink period underlaying cellular networks," *IEEE Trans. Wireless Communications*, vol. 10, no. 2, pp. 413–418, Feb. 2011.

[4] C.-H, Yu, K. Doppler, C.B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Communications*, vol. 10, no. 8, pp. 2752–2763, Aug. 2011.

[5] P. Cheng, L. Deng, H. Yu, Y. Xu, and H. Wang, "Resource allocation for cognitive networks with D2D communication: an evolutionary approach," in *Proc. IEEE WCNC*, Shanghai, China, April 2012, pp. 2671–2676.

[6] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference-aware graph based resource sharing for device-to-device communications underlaying cellular networks,", in *Proc. IEEE WCNC*, Shanghai, China, April 2013, pp. 140–145.

[7] M.-H. Han, B.-G. Kim, and J.-W. Lee, "Subchannel and transmission mode scheduling for D2D communication in OFDMA networks," in *Proc. IEEE VTC-Fall*, Quebec City, Canada, Sept. 2012, pp. 1–5.

[8] M. Belleschi, G. Fodor, and A. Abrardo, "Performance analysis of a distributed resource allocation scheme for D2D communications,", in *Proc. IEEE GLOBECOM*, Houston, USA, Dec. 2011, pp. 358–362.

[9] P. Janis, V. Koivunen, C.B. Ribeiro, K. Doppler, and K. Hugl, "Interference-avoiding MIMO schemes for device-to-device radio underlaying cellular networks," in *Proc. IEEE PIMRC*, Tokyo, Japan, Sept. 2009, pp. 2385–2389.

[10] J.C.F. Li, M. Lei, and F. Gao, "Device-to-device (D2D) communication in MU-MIMO cellular networks,", in *Proc. IEEE GLOBECOM*, Anaheim, USA, Dec. 2012, pp. 3583–2587.

[11] D. Zhu, W. Xu, H. Zhang, C. Zhao, J.C.F. Li, and M. Lei, "Rate maximized tranceiver optimization for multi-antenna device-to-device communications," in *Proc. IEEE WCNC*, Shanghai, China, April 2013, pp. 4152–4157.

[12] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel", *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[13] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "Linear transceiver design for a MIMO interfering broadcast channel achieving max-min fairness," *Signal Processing*, vol. 93, no. 12, pp. 3327–3340, 2013.

[14] M. Fukushima, "Application of alternating direction method of multipliers to separable convex programming problems," *Computational Optimization and Applications*, vol. 1, no. 1, pp. 93–111, Oct. 1992.

[15] J. Lin, Q. Li, C. Jiang, and H. Shao, "Joint multi-relay selection, power allocation and beamformer design for multi-user decode-and-forward relay networks," to appear in *IEEE Trans. Vehicular Technology*, DOI: 10.1109/TVT.2015.2462117.

[16] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: complexity and duality", *IEEE Journal of Selected Topics in Signal Processing*, vol. 20, no. 1, pp. 57–73, Feb. 2008.

[17] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge, UK: Cambridge University Press, 2004.

[18] R. Horst and N.V. Thoai, "DC programming: overview," *Journal of Optimization Theory and Applications*, vol. 103, no. 1, pp. 1–43, 1999.