Hierarchical Mean Field/Stackelberg Game Power Control for D2D Cooperative Relaying Networks

Zheng Wang* and Brian L. Mark
Dept. of Electrical and Computer Engineering
George Mason University
4400 University Drive, MS 1G5 Fairfax, VA
email: zwang23@gmu.edu; bmark@gmu.edu

***Abstract*—Device-to-device (D2D) communications is widely adopted as the underlay to cellular primary networks. By reusing the licensed spectrum and directly transmitting to other nearby devices, D2D users can significantly enhance system spectrum efficiency. However, D2D communications usually faces the restrictions of short transmission distance and limited energy supply. Mutual interference among different devices increases energy consumption and exacerbates communication quality-of-service degradation. In this paper, we consider a dense D2D network where the D2D devices communicate using cooperative relaying. We propose a hierarchical game framework consisting of a Stackelberg game to model intra-tier interactions within each D2D link and a mean field game to model inter-tier interactions among different D2D links. A finite difference method is used to derive an optimal power control scheme, and numerical results are presented to demonstrate the performance of the proposed hierarchical power control scheme. We also make comparisons to a proportional power control scheme, where the experimental result shows clear advantage of the proposed hierarchical scheme.**

***Index Terms*—D2D communications, cooperative relaying, power control, mean field game, Stackelberg game.**

I. Introduction

In recent years, device-to-device (D2D) communications has been widely used as the underlay to cellular primary systems to enhance power and spectrum efficiency [1], [2]. Due to the proximity between D2D devices, direct communication is achievable by reusing spectrum licensed to the primary devices [3]. Yet several challenges remain: 1) Network performance is often limited by the short transmission range of D2D devices [4]. 2) The mutual interference caused by a large number of densely deployed D2D devices has a significant impact [5], [6]. 3) D2D devices are usually powered by batteries with finite energy supply, which makes power control a critical issue [7].

Cooperative relaying has been proposed to extend transmission range and enhance system capacity in D2D networks [8]. Few works have considered power control schemes that can satisfy the requirements for dense D2D cooperative relaying networks. In this paper, we aim to coordinate the interactions among a large number of D2D devices and control the power usage. Both inter-tier interactions among different D2D pairs and intra-tier interactions among different hops for a D2D pair exist in this scenario [9]. Power control is a dynamic process that determines the energy consumption of each D2D device, which also needs to be effectively characterized.

Game theory has been widely adopted to coordinate the interactions among D2D devices and decide their transmission strategies [10]. For intra-tier coordination, Stackelberg game (SG) can be a powerful framework due to its capability of modeling two interacting rational players [11]. Based on the Stackelberg game outcomes, we propose to use a mean field game (MFG) framework to model the inter-tier interactions between a single rational player and the statistical behavior of all other players, since the number of D2D devices is large in a dense network [12], [13]. The statistical behavior of all D2D devices can be considered as the mean field which affects the state of each player.

The contribution of our work is two-fold: 1) We develop a hierarchical game framework for modeling dense D2D cooperative relaying networks incorporating Stackelberg game outcomes into a mean field game. 2) We develop an iterative finite difference algorithm to derive an optimal power control policy for D2D links, which takes into account both transmitter and relay node energy. Our numerical results show that the proposed hierarchical game scheme significantly outperforms a proportional power control scheme that does not take into account intra-tier interactions.

The remainder of the paper is organized as follows. In Section II, we describe the network model and formulate the power control problem. In Section III, we propose our Stackelberg game model for intra-tier coordination. In Section IV, we formulate the MFG framework for inter-tier coordination. In Section V, we develop the finite difference algorithm for distributed power control. Simulation results are presented in Section VI and conclusions are given in Section VII.

II. System Model

We consider a dense D2D network macrocell in which a large number of D2D pairs are deployed and share uplink spectrum resources of macro user equipments (MUEs). Each

This work was supported in part by the U.S. National Science Foundation under Grant No. 2034616.

*Z. Wang is currently with the Bradley Dept. of Electrical and Computer Eng., Virginia Tech, Blacksburg, VA.

D2D pair employs cooperative relaying to increase its transmission range. A macrocell base station can communicate with all the MUEs in the network. Each D2D transmitter aims to transmit to its corresponding receiver using a relay node, and the numbers of D2D pairs and relay nodes are both N. At most one relay is used, as a larger number of relays introduces significant overhead without much performance benefit.

We assume full frequency reuse for the D2D transmitters and relay nodes when they do not interfere with each other. To satisfy such demand, two MUEs provide their channels to allow D2D transmitters and relay nodes to transmit. Therefore, D2D transmitters will only cause interference to relay nodes, and the relay nodes will only cause interference to D2D receivers. Decode-and-forward (DF) relaying is adopted in this work. We use D_i^T, D_i^R and R_i to denote the i-th D2D transmitter, D2D receiver and relay node, respectively. We assume a Rayleigh channel model in this work, and the additive noise follows a complex Gaussian distribution. At time t, the inter-tier interference powers $I_{R,i}(t)$ and $I_{D,i}(t)$ at R_i and D_i^R are given by

$$I_{R,i}(t)=\sum_{\substack{j=1\\j\neq i}}^{N}p_{d,j}(t)g_{dr,j,i}(t),\ \ I_{D,i}(t)=\sum_{\substack{j=1\\j\neq i}}^{N}p_{r,j}(t)g_{rd,j,i}(t),\quad (1)$$

where $p_{d,j}(t), p_{r,j}(t)$ are the transmit powers of D_j^T and R_j, respectively, and $g_{dr,j,i}(t), g_{rd,j,i}(t)$ denote the channel gains from D_j^T to R_i and from R_j to D_i^R, respectively. The signal to interference-plus-noise ratio (SINR) at R_i and at D_i^R can be written as

$$\gamma_{R,i}(t)=\frac{p_{d,i}(t)g_{dr,i,i}(t)}{I_{R,i}(t)+\sigma^2},\ \gamma_{D,i}(t)=\frac{p_{r,i}(t)g_{rd,i,i}(t)}{I_{D,i}(t)+\sigma^2},\quad (2)$$

respectively, where $g_{dr,i,i}(t), g_{rd,i,i}(t)$ are the channel gains between D_i^T and R_i, and between R_i and D_i^R, respectively, and σ^2 is the additive channel noise power.

We assume that the D2D transmitter and relay node have the same maximum available energy $E_{\max}$, and the power control policy is conducted over a time period $[0, T]$. The system uses time division multiplexing (TDM). The power control policy is defined as a sequence of transmit powers for both the D2D transmitter and relay node in each time slot, i.e., at time t, the power control policy is a pair $(P_{i,d}^*(t), P_{i,r}^*(t))$. The power control problem impacts the dynamics of device energy and can be formulated under a differential game framework [14].

We assume that the D2D transmitter and the relay node have the same maximum available energy $E_{\max}$ at $t = 0$. Therefore, the energy of D_i^T, $E_{d,i}(t)$ and R_i, $E_{d,i}(t)$ satisfy $0 \le E_{d,i}(t), E_{r,i}(t) \le E_{\max}$ during transmission. The energy dynamics can be written as

$$dE_{d,i}(t) = -p_{d,i}(t)dt, \quad dE_{r,i}(t) = -p_{r,i}(t)dt. \quad (3)$$

The state process of the i-th D2D pair is then given by

$$s_i(t) = [E_{d,i}(t), E_{r,i}(t)]. \quad (4)$$

The cost function is the optimization objective over $[0, T]$ with respect to both SINR performance and transmit power. Each D2D pair interacts with other D2D pairs while deciding its own power control policy to minimize the cost. According to the assumptions, we aim to find an optimal control policy for D2D transmitter and relay node power. We assume an SINR threshold γ_{th} for both D2D receiver and relay node. The cost function of the i-th D2D pair is defined as

$$c_i(t) = [\gamma_{D,i}(t) - \gamma_{th}]^2 + [\gamma_{R,i}(t) - \gamma_{th}]^2 + \beta p_{d,i}(t), \quad (5)$$

where the coefficient β is introduced to balance the squared SINR difference and the transmit power. The differential game aims to coordinate inter-tier interactions among different D2D pairs, while the Stackelberg game coordinates intra-tier interactions within a D2D pair.

III. Stackelberg Game Analysis

The Stackelberg game models two rational players as a leader-follower pair and the interactions between them as a trade relationship [11], [15]. In DF transmission, a relay node will select an appropriate transmit power to forward the signal to the next hop. The D2D transmitter is willing to pay some price to the relay node in DF and the relay node can gain some revenue by helping to establish the link. For each D2D pair, we model the relay node as the leader and D2D transmitter as the follower. The leader announces the unit price for its transmit power, and the follower reacts to the leader by deciding the amount of transmit power it would like to buy and makes a payment. The leader and follower attempt to maximize their own utilities simultaneously. The Stackelberg game Nash equilibrium (SGNE) is achieved if neither the leader nor the follower can achieve higher utility by making unilateral changes to its own transmission strategy.

We assume that the transmit power $p_{d,i}(t)$ of D_i^T is known a priori. The follower aims to gain as much benefit as possible at the least possible expense. Consider the bottleneck in cooperative relaying, the i-th follower utility is given by

$$\begin{aligned} U_{F,i}(t) =& \frac{1}{2}\min\left\{\log_2(1+\gamma_{R,i}(t)), \log_2(1+\gamma_{D,i}(t))\right\} \\ &- \alpha_i p_{r,i}(t), \end{aligned} \quad (6)$$

where α_i is the unit price for relay node transmit power, $U_{F,i}(t)$ is related to the link capacity of the D2D pair, and $-\alpha_i p_{r,i}(t)$ denotes the expense paid for relay transmit power. The larger α_i is, the less willing the follower is to buy relay transmit power, which leads to a link capacity decrease. As for the leader, the relay node can benefit by selling its transmit power. The i-th leader utility function can be defined as [16]

$$U_{L,i}(t) = (\alpha_i - \alpha_0)p_{r,i}(t), \quad (7)$$

where α_0 is a constant which denotes the unit cost of relay transmit power. The leader can make profits when α_i is higher than the cost and relay transmit power is greater than zero. The leader and follower aim to maximize their utilities as follows:

$$\max_{\{p_{r,i}(t)\}} \ U_{F,i}(t) \qquad \text{subject to: } p_{r,i}(t) \ge 0, \quad (8)$$

$$\max_{\{\alpha_i\}} \ U_{L,i}(t) \qquad \text{subject to: } \alpha_i \ge \alpha_0. \quad (9)$$

The price α_i charged by the leader and the relay transmit power $p_{r,i}(t)$ the follower would like to buy constitute a strategy pair $(\alpha_i, p_{r,i}(t))$. Based on the definition of Stackelberg game, the proposed strategy pair is also the SGNE if neither the leader nor the follower can make more profits by making unilateral changes to its own strategy, i.e.,

$$U_{F,i}(\alpha_i, p_{r,i}(t), t) \geq U_{F,i}(\alpha_i, p'_{r,i}(t), t), \tag{10}$$

$$U_{L,i}(\alpha_i, p_{r,i}(t), t) \geq U_{F,i}(\alpha'_i, p_{r,i}(t), t). \tag{11}$$

The SGNE can be solved by backward induction, i.e., the analysis of follower strategy is carried out first, then the analysis of leader strategy is carried out based on follower outcomes to derive the optimal price. To simplify notations, we define $C := \frac{(\sum_{j=1,j\neq i}^{N} p_{r,j} g_{rd,j,i} + \sigma^2) g_{dr,i,i}}{(\sum_{j=1,j\neq i}^{N} p_{d,j} g_{dr,j,i} + \sigma^2) g_{rd,i,i}}$, $D := \frac{g_{rd,i,i}}{\sum_{j=1,j\neq i}^{N} p_{r,j} g_{rd,j,i} + \sigma^2}$ and rewrite $\alpha_{i,\max} = \frac{D}{2\ln 2}$. If the leader sets the price too high, the follower will not buy its power and this leads to an insufficient outcome. Therefore, by setting the optimal power estimate as non-negative, the price is upper bounded by $\alpha_{i,\max} = \frac{D}{2\ln 2}$. On the other hand, as long as its revenue can cover the cost, the leader can always profit from forwarding the signal. Based on the arguments, the SGNE can be characterized as follows[1]:

1) $\alpha_{i,\max} > \alpha_0$: The optimal price $\hat{\alpha}_i$ is searched within $[\alpha_0, \alpha_{i,\max}]$, and the optimal transmit power is always positive. Let $\hat{p}_{r,i} = \frac{1}{\sqrt{2\alpha_0 D \ln 2}} - \frac{1}{D}$. Based on the relaying bottleneck in (6), we have two cases:
   - $\hat{p}_{r,i} < C p_{d,i}$: The second hop is the bottleneck and the follower chooses $\hat{p}_{r,i}$ as the optimal transmit power. Therefore, the SGNE is $\left(\sqrt{\frac{\alpha_0 D}{2\ln 2}}, \frac{1}{\sqrt{2\alpha_0 D \ln 2}} - \frac{1}{D}\right)$.
   - $\hat{p}_{r,i} \geq C p_{d,i}$: The first hop is the bottleneck and the follower utility becomes $\frac{1}{2}\log_2(1+\gamma_{R,i}(t)) - \alpha_i p_{r,i}(t)$. Thus, the optimal transmit power is $C p_{d,i}$. The leader utility becomes $U_{L,i} = (\alpha_i - \alpha_0) C p_{d,i}$ and the optimal price is as $\alpha_{i,\max}$. The SGNE is $\left(\frac{D}{2\ln 2}, C p_{d,i}\right)$.
2) $\alpha_{i,\max} \leq \alpha_0$: The relay sets its power to 0 as it does not profit from transmission and the optimal price is α_0. Thus, the SGNE is given by $(\alpha_0, 0)$.

IV. Mean Field Game for Power Control

In D2D cooperative relaying, each player y determines an optimal power control profile $Q_i^*(t)$ to minimize its cost function in (5) over $[0, T]$. According to Stackelberg game analysis, the relay transmit power can be uniquely determined if the D2D transmit power and channel parameters are known. The optimal power control profile can be written as

$$Q_i^*(t) = \arg\min_{p_{d,i}(t)} \mathrm{E}\left[\int_0^T c_i(p_{d,i}(t), \boldsymbol{p}^*_{d,-i}) dt + c_i(T)\right] \tag{12}$$

where $\mathrm{E}[\cdot]$ denotes expectation and $\boldsymbol{p}^*_{d,-i}$ denotes the transmit power of all D2D links except the i-th link. The objective is to determine the optimal power control path over $[0, T]$ that minimizes the cost function. Meanwhile, we define a value function as follows:

$$u_i(t, s_i(t)) = \min_{p_{d,i}(t)} \mathrm{E}\left[\int_t^T c_i(t) dt + u_i(T, s_i(T))\right], \tag{13}$$

where $u_i(t, s_i(t))$ is the value at time t and achieves $u_i(T, s_i(T))$ when $t = T$. In the MFG, the Nash equilibrium exists and yields the power control profile (12) if and only if the constraints (3) and (4) hold (see [13]).

The mean field is defined as the statistical distribution of the proposed state process in the limiting regime where the number of players goes to infinity. Given the state process $s_i(t) = [E_{d,i}(t), E_{r,i}(t)]$ in (4), the corresponding mean field $m(s, t)$ is the probability that a certain state $s_i(t)$ is likely to appear at time t, i.e., $m(s,t) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f_i(s_i(t), s)$, where $f_i(s_i(t), s)$ equals 1 if $s_i(t) = s$ and 0 otherwise. To derive the mean field given above, mean field approximation is applied to the interference received by the relay node and D2D receiver. When N is sufficiently large, (1) can be rewritten as

$$I_{R,i}(t) \approx (N-1)\bar{p}_d(t)\bar{g}_{dr,j,i}(t), I_{D,i}(t) \approx (N-1)\bar{p}_r(t)\bar{g}_{rd,j,i}(t)$$

where $\bar{p}_d(t), \bar{p}_r(t)$ are known test powers and $\bar{g}_{dr,j,i}(t), \bar{g}_{rd,j,i}(t)$ are the mean channel gains. We set the test power equal to $E_{\max}/T$ for all D2D transmitters and relay nodes to obtain the mean channel gain estimates. The powers received by R_i and D_i^R, denoted by $p_{r,i}^R(t), p_{d,i}^R(t)$, are approximated by

$$p_{r,i}^R(t) \approx \bar{p}_d(t) g_{dr,i,i}(t) + (N-1)\bar{p}_d(t)\bar{g}_{dr,j,i}(t) \tag{14}$$

$$p_{d,i}^R(t) \approx \bar{p}_r(t) g_{rd,i,i}(t) + (N-1)\bar{p}_r(t)\bar{g}_{rd,j,i}(t) \tag{15}$$

The mean field approximation of the channels gains can then be obtained by solving (14) and (15) for $\bar{g}_{dr,j,i}(t)$ and $\bar{g}_{rd,j,i}(t)$, respectively.

Since each D2D pair aims to decide the appropriate transmission strategy by itself, the Stackelberg game outcomes need to be derived using mean field approximation as well. By substituting (14) and (15) into the expressions for C and D, we obtain the approximations $\bar{C} \approx \frac{(p_{d,i}^R(t) - \bar{p}_r(t) g_{rd,i,i}(t) + \sigma^2) g_{dr,i,i}(t)}{(p_{r,i}^R(t) - \bar{p}_d(t) g_{dr,i,i}(t) + \sigma^2) g_{rd,i,i}(t)}$, $\bar{D} \approx \frac{g_{rd,i,i}(t)}{p_{d,i}^R(t) - p_{r,i}(t) g_{rd,i,i}(t) + \sigma^2}$. The SGNE $(\hat{\alpha}_i(t), \hat{p}_{r,i}(t))$ is chosen from the following estimates (see Section III):

$$\left\{\left(\sqrt{\frac{\alpha_0 \bar{D}}{2\ln 2}}, \frac{1}{\sqrt{2\alpha_0 \bar{D} \ln 2}} - \frac{1}{\bar{D}}\right), \left(\frac{\bar{D}}{2\ln 2}, \bar{C} p_{d,i}\right), (\alpha_0, 0)\right\}, \tag{16}$$

where $\hat{\alpha}_i(t), \hat{p}_{r,i}(t)$ are the optimal price and relay transmit power under mean field approximation, respectively. Hence, the estimated SINR received by R_i and D_i^R are given by

$$\bar{\gamma}_{R,i}(t) \approx \frac{\hat{p}_{d,i}(t) g_{dr,i,i}(t)}{p_{r,i}^R(t) - \bar{p}_d(t) g_{dr,i,i}(t) + \sigma^2}, \tag{17}$$

$$\bar{\gamma}_{D,i}(t) \approx \frac{\hat{p}_{r,i}(t) g_{rd,i,i}(t)}{p_{d,i}^R(t) - \bar{p}_r(t) g_{rd,i,i}(t) + \sigma^2}, \tag{18}$$

[1] Due to space limitations, we omit the detailed proofs.

respectively. Then the cost function (5) is approximated by

$$\hat{c}_i(t) = [\bar{\gamma}_{D,i}(t) - \gamma_{th}]^2 + [\bar{\gamma}_{R,i}(t) - \gamma_{th}]^2 + \beta p_{d,i}(t). \quad (19)$$

The MFG can be formulated as a dynamic coupled system which consists of two coupled differential equations called Hamilton-Jacobi-Bellman (HJB) and Fokker-Plank-Kolmogorov (FPK) given as follows [13]:

$$\text{HJB}: \quad -\partial_t u(t,s) = \min_{p_d(t)} [c(t) + \partial_t s(t) \cdot \nabla u(t,s)], \quad (20)$$

$$\text{FPK}: \quad \partial_t m(t,s) + \nabla(m(t,s) \cdot \partial_t s(t)) = 0, \quad (21)$$

where s denotes the system state. The Nash equilibrium (i.e., the power control profile) of the proposed MFG can be derived by jointly solving the HJB backward from T to 0 and the FPK forward from 0 to T. The value $u(t,s)$ and the mean field $m(t,s)$ are the analytical solutions for HJB and FPK, respectively. The mean field equilibrium (MFE) is denoted by a stable pair $(u^*(t,s), m^*(t,s))$.

V. Distributed Power Control Scheme

The search space is discretized into a grid, where the three axes are time T, D2D transmitter energy E_d, and relay node energy E_r. We assume that $T \in [0, T_{\max}]$, $E_d \in [0, E_{\max}]$, $E_r \in [0, E_{\max}]$, and the step sizes in the three dimensions are $\delta_t = T_{\max}/X$, $\delta_{E_d} = E_{\max}/Y$, and $\delta_{E_r} = E_{\max}/Z$, respectively, where X, Y, Z denote the numbers of grids in each dimension. The solution to the FPK equation evolves in the three-dimensional space, while the optimal control path can be derived by solving the coupled HJB and FPK equations.

By substituting $s(t)$ from (4) into (21), we obtain

$$\partial_t m(t,s) + \nabla_{E_d} m(t,s) E_d(t) + \nabla_{E_r} m(t,s) E_r(t) = 0, \quad (22)$$

where $E_d(t), E_r(t)$ denote the remaining energy in the D2D transmitter and the relay node, respectively. This equation can be iteratively solved by applying the Lax-Friedrichs scheme [17]. We define $M(i,j,k)$, $P_d(i,j,k)$, $P_r(i,j,k)$ as the mean field, D2D transmitter power level and relay node power level on the (i,j,k)th grid in the discretized search space, respectively. Here, $i \in \{0, \dots, X\}$, $j \in \{0, \dots, Y\}$, and $k \in \{0, \dots, Z\}$ are indices corresponding to T, E_d, and E_r, respectively. The mean field update is given by

$$\begin{aligned} M(i{+}1,j,k) = {}& \frac{1}{2}[M(i,j{+}1,k) + M(i,j{-}1,k) + M(i,j,k{+}1) \\ & + M(i,j,k{-}1)] + \frac{\delta_t}{2\delta_{E_d}}[M(i,j{+}1,k)P_d(i,j{+}1,k) \\ & - M(i,j{-}1,k)P_d(i,j{-}1,k)] + \frac{\delta_t}{2\delta_{E_r}}[M(i,j,k{+}1)P_r(i,j,k{+}1) \\ & - M(i,j,k{-}1)P_r(i,j,k{-}1)]. \end{aligned} \quad (23)$$

If we consider the dynamics of the value function $u(t,s)$ in HJB and use the optimal path of the mean field $m(t,s)$ in FPK as an equality constraint, the problem can be framed as

$$\begin{aligned} \min_{p_{d,i}(t)} \quad & \mathrm{E}\left[\int_0^T c_i(t)dt + c_i(T)\right] \\ \text{s.t. } & \partial_t m(t,s) + \nabla_{E_d} m(t,s) E_d(t) + \nabla_{E_r} m(t,s) E_r(t) = 0. \end{aligned} \quad (24)$$

Since the FPK equation is regarded as an equality constraint, the proposed optimization problem can be solved by introducing a Lagrange multiplier to the objective function. Assuming $\lambda(i,j,k)$ is the Lagrange multiplier at the (i,j,k)th grid, $C(i,j,k)$ is the cost function at the (i,j,k)th grid. The discretized Lagrangian L yields

$$\begin{aligned} L(i,j,k) = {}& \delta_t \delta_{E_d} \delta_{E_r} \sum_{i=1}^{X+1} \sum_{j=1}^{Y+1} \sum_{k=1}^{Z+1} [M(i,j,k)C(i,j,k) \\ & + \lambda(i,j,k)(I + J + K)], \text{ where} \end{aligned} \quad (25)$$

$$\begin{aligned} I = {}& \frac{1}{\delta_t}\Big[M(i{+}1,j,k) - \frac{1}{2}(M(i,j{+}1,k) + M(i,j{-}1,k) \\ & + M(i,j,k+1) + M(i,j,k-1)\Big), \end{aligned} \quad (26)$$

$$\begin{aligned} J = {}& \frac{1}{2\delta_{E_d}}[M(i,j+1,k)P_d(i,j+1,k) \\ & - M(i,j-1,k)P_d(i,j-1,k)], \end{aligned} \quad (27)$$

$$\begin{aligned} K = {}& \frac{1}{2\delta_{E_r}}[M(i,j,k+1)P_r(i,j,k+1) \\ & - M(i,j,k-1)P_r(i,j,k-1)]. \end{aligned} \quad (28)$$

Thus, the optimal D2D transmit power can be derived from $\frac{\partial L}{\partial P_d(i,j,k)} = 0$ at each grid point (i,j,k). The discretized equation for $\frac{\partial L}{\partial P_d(i,j,k)} = 0$ is

$$\begin{aligned} \frac{\partial L(i,j,k)}{\partial P_d(i,j,k)} = {}& [\lambda(i,j+1,k) - \lambda(i,j-1,k)]\Bigg[\frac{M(i,j,k)}{2\delta_{E_d}} \\ & + \frac{M(i,j,k)}{2\delta_{E_r}} \frac{\partial P_r(i,j,k)}{\partial P_d(i,j,k)}\Bigg] + M(i,j,k)\frac{\partial C(i,j,k)}{\partial P_d(i,j,k)} = 0, \end{aligned} \quad (29)$$

where the power $P_d(i,j,k)$ can be solved and $P_r(i,j,k)$ can be derived subsequently. Similarly, the Lagrangian multiplier $\lambda(i,j,k)$ can be updated by solving $\frac{\partial L(i,j,k)}{\partial M(i,j,k)} = 0$. As the HJB function is solved backward in time, we can derive $\lambda(i{-}1,j,k)$ in each iteration as follows:

$$\begin{aligned} \lambda(i-1,j,k) = {}& \frac{1}{2}\big[\lambda(i,j+1,k) + \lambda(i,j-1,k) \\ & + \lambda(i,j,k+1) + \lambda(i,j,k-1)\big] - \frac{\delta_t}{2}\left[\frac{P_d(i,j,k)}{\delta_{E_d}} + \frac{P_r(i,j,k)}{\delta_{E_r}}\right] \\ & \big[\lambda(i,j+1,k) - \lambda(i,j-1,k)\big] - \delta_t C(i,j,k)]. \end{aligned} \quad (30)$$

Algorithm 1 performs updates according to Eqs. (23), (29), and (30) with an overall computational complexity of $O(XYZ)$ per iteration. The mean field distribution is initialized at $t = 0$, while the Lagrangian and the power distribution are initialized at $t = T_{\max}$. The mean field $M(i,j,k)$ evolves from 0 to $T_{\max}$, while $\lambda(i,j,k), P_d(i,j,k)$ evolves backward from $T_{\max}$ to 0. The algorithm stops when the mean field distributions converge under some criterion.

VI. Numerical Results

We consider a D2D network in which the D2D links are randomly and uniformly deployed within a cell of radius $R = 50$ m. The distance of each D2D hop is randomly chosen in the interval $[10 \text{ m}, 20 \text{ m}]$. The bandwidth of each hop is set to 20 MHz and the background noise to 2×10^{-9} W, as the

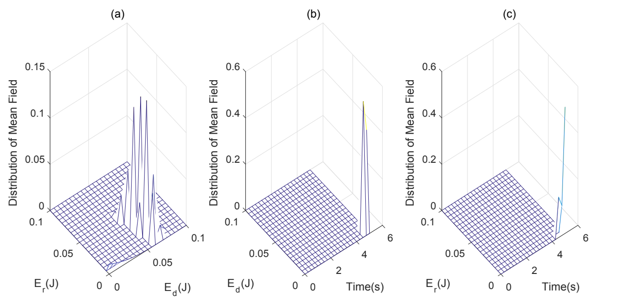

Fig. 1. Mean field distributions for each parameter space slice: (a) $T = T_{\max} = 5$ s, (b) $E_r = E_{\max} = 0.1$ J, (c) $E_d = E_{\max} = 0.1$ J.

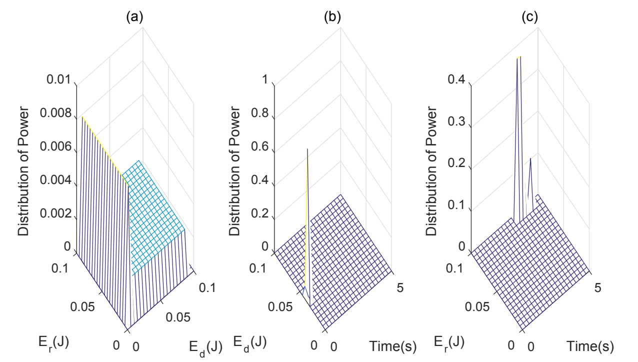

Fig. 2. Power distributions for each parameter space slice: (a) $T = T_{\max} = 5$ s, (b) $E_r = E_{\max} = 0.1$ J, (c) $E_d = E_{\max} = 0.1$ J.

noise spectral density is -174 dBm/Hz. The total number of D2D links is 50 and $T_{\max} = 5$ s. The maximum energy of the D2D transmitter and relay node is 0.1 J and the path loss exponent is 4. We set $X = Y = Z = 20$ in the discretized parameter space and $\alpha_0 = 0.2$ as the unit power cost.

A. Power Control Performance Study

We consider the following scenarios: (a) fixed time $T = T_{\max} = 5$ s; (b) fixed D2D transmitter energy $E_d = E_{\max} = 0.1$ J; (c) fixed relay node energy $E_r = E_{\max} = 0.1$ J. The mean field distributions are shown in Fig. 1 and the D2D transmit power distributions are shown in Fig. 2. In Fig. 1, the mean field in case (a) converges at $T = 5$ to a random distribution, as the relay node transmit power is affected by the random distributed channel parameters. For cases (b) and (c), the mean field distributions converge to 0 in the energy state space at $T = 5$, showing that the D2D transmitter and the relay node consumed their energy during the power control process. Similarly, for the power distributions in Fig. 2, it can be seen that the energy states of both the D2D transmitter and the relay node also affect the distributions of transmit power.

Algorithm 1 Distributed Power Control Scheme

Initialize: $M(0,:,:)$, $\lambda(X+1,:,:)$, $P_d(X+1,:,:)$
repeat
  for $i = 1:1:X, \forall j \in \{1,...,Y\}, \forall k \in \{1,...,Z\}$ **do**
    Update $M(i+1,j,k)$ using Eq. (23) // Forward
  end for
  $M(i+1,j+1,k+1) = 0$
  if $P_d(i,j+1,k) = 0$ **then**
    $M(i+1,j+1,k+1) = M(i,j,k)$
  end if
  $\forall i$, normalize $M(i,:,:)$
  for $i = X+1:-1:1, \forall j \in \{1,...,Y\}, \forall k \in \{1,...,Z\}$ **do**
    Update $\lambda(i-1,j,k)$ using Eq. (30) // Backward
  end for
  for $\forall i \in \{1,...,X\}, \forall j \in \{1,...,Y\}, \forall k \in \{1,...,Z\}$ **do**
    Update $P_d(i,j,k)$ using Eq. (29) // Backward
  end for
  for $\forall i \in \{1,...,X\}, \forall j \in \{1,...,Y\}, \forall k \in \{1,...,Z\}$, **do**
    Update $P_r(i,j,k), C(i,j,k)$
  end for
until Convergence criterion is met

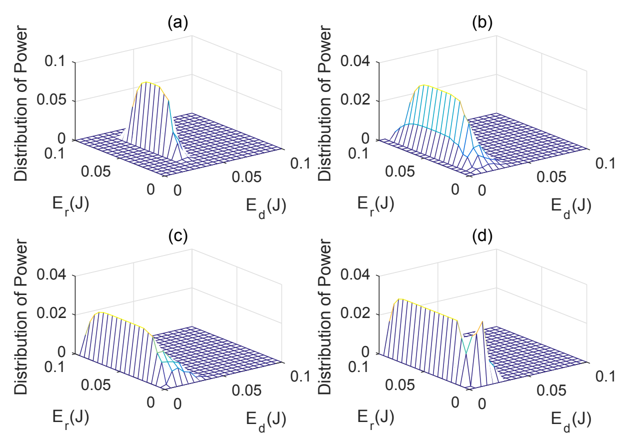

Fig. 3. Power distributions at $T = 1,2,3,4$: (a) $T = 1$; (b) $T = 2$; (c) $T = 3$; (d) $T = 4$.

To demonstrate the convergence process of the power control policy, we depict the transmit power distributions in time slots $T = 1,2,3,4$, as shown in Fig. 3. The power distribution converges to the zero state in the transmitter energy dimension as time grows, which shows that the transmitter consumed its remaining energy during the power control process.

Finally, we plot the cross-sections of the distributions at different energy states to show the randomness of the distributions. We study the following cases: (a) the distributions at time slot $T = 3$; (b) the distributions at time slot $T = 4$; (c) the distributions at time slot $T = 5$. The corresponding cross-sections at each time slot are depicted in Fig. 4, while the cross-sections of power distributions are shown in Fig. 5. The randomness of the channel parameters mainly affects the mean field distributions. Fig. 5 shows that the power distributions over different energy states appear to be more uniform. Thus, the power control policy yields a stable outcome.

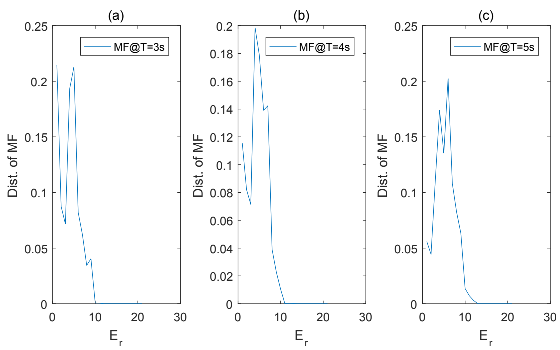

Fig. 4. Cross-sections of mean field distributions at different time: (a) $T = 3$ s; (b) $T = 4$ s; (c) $T = T_{\max} = 5$ s.

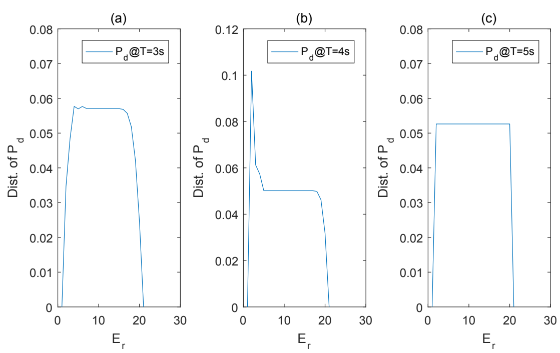

Fig. 5. Cross-sections of D2D transmitter power distributions at different times: (a) $T = 3$ s; (b) $T = 4$ s; (c) $T = T_{\max} = 5$ s.

B. Hierarchical Game vs. Proportional Power Control

We compare the proposed hierarchical scheme and a scheme in which the relay node power is proportional to the D2D transmitter power (i.e., $P_r = \alpha P_d, \alpha \in (0, 1]$) at each time instant such that the devices will not run out of battery during the power control process. As shown in Table I, the follower (D2D transmitter) utility in the hierarchical game slightly outperforms that in proportional power control. Meanwhile, the relay node benefits significantly from adopting intra-tier coordination in the hierarchical game compared to proportional power control. As α decreases, both the leader and the follower see a decrease in their utilities. Clearly, the hierarchical game achieves much better overall performance.

TABLE I
HIERARCHICAL GAME VS. PROPORTIONAL POWER CONTROL UTILITIES

	D2D Transmitter Utility (Follower)	Relay Node Utility (Leader)
Hierarchical	-3.8324	3.7621
Proportional, $a = 1$	-4.1418	0.4470
Proportional, $a = 0.8$	-4.1390	0.3875
Proportional, $a = 0.6$	-4.1755	0.3204
Proportional, $a = 0.4$	-4.5593	0.2279

VII. CONCLUSION

We formulated a hierarchical mean field/Stackelberg game-theoretic framework to model the interactions and behaviors of a dense D2D cooperative relaying network. Based on this framework, we proposed a distributed power control policy which incorporates the Lax-Friedrichs finite method to jointly solve the coupled HJB and FPK equations. Our numerical results characterized the mean field and power distributions over the parameter space, which showed the effectiveness of the power control policy. The proposed hierarchical scheme also outperforms a proportional power control policy in both leader and follower utilities.

REFERENCES

[1] R. Zhang, Y. Li, C.-X. Wang, Y. Ruan, Y. Fu, and H. Zhang, "Energy-spectral efficiency trade-off in underlaying mobile D2D communications: An economic efficiency perspective," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4288–4301, 2018.

[2] Z. Zhou, M. Dong, K. Ota, J. Wu, and T. Sato, "Energy efficiency and spectral efficiency tradeoff in device-to-device (D2D) communications," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 485–488, 2014.

[3] L. Song, D. Niyato, Z. Han, and E. Hossain, *Wireless device-to-device communications and networks*. Cambridge University Press, 2015.

[4] B. Ying and A. Nayak, "A power-efficient and social-aware relay selection method for multi-hop D2D communications," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1450–1453, 2018.

[5] Z. Zhou, M. Dong, K. Ota, and Z. Chang, "Energy-efficient context-aware matching for resource allocation in ultra-dense small cells," *IEEE Access*, vol. 3, pp. 1849–1860, 2015.

[6] B. Soret, K. I. Pedersen, N. T. Jørgensen, and V. Fernández-López, "Interference coordination for dense wireless networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 102–109, 2015.

[7] W.-K. Lai, Y.-C. Wang, H.-C. Lin, and J.-W. Li, "Efficient resource allocation and power control for LTE-A D2D communication with pure D2D model," *IEEE Trans. Veh. Technol.*, vol. 69, pp. 3202–3216, 2020.

[8] C. Xu, J. Feng, Z. Zhou, Z. Chang, Z. Han, and S. Mumtaz, "Two-stage matching for energy-efficient resource management in D2D cooperative relay communications," in *IEEE Globecom*, 2017, pp. 1–6.

[9] R. I. Ansari *et. al*, "5G D2D networks: Techniques, challenges, and future prospects," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3970–3984, 2017.

[10] C. Gao, Y. Li, Y. Zhao, and S. Chen, "A two-level game theory approach for joint relay selection and resource allocation in network coding assisted D2D communications," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2697–2711, 2016.

[11] Y. Yuan, T. Yang, H. Feng, and B. Hu, "An iterative matching-Stackelberg game model for channel-power allocation in D2D underlaid cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7456–7471, 2018.

[12] C. Yang, J. Li, P. Semasinghe, E. Hossain, S. Perlaza, and Z. Han, "Distributed Interference and Energy-Aware Power Control for Ultra-Dense D2D Networks: A Mean Field Game," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 1205 – 1217, Feb. 2017.

[13] P. Semasinghe and E. Hossain, "Downlink power control in self-organizing dense small cells underlaying macrocells: A mean field game," *IEEE Trans. Mobile Comput.*, vol. 15, no. 2, pp. 350 – 363, Feb. 2016.

[14] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese Journal of Mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[15] Z. Chu, T. A. Le, H. X. Nguyen, A. Nallanathan, and M. Karamanoglu, "A Stackelberg-game approach for disaster-recovery communications utilizing cooperative D2D," *IEEE Access*, vol. 6, pp. 10 733–10 742, 2017.

[16] H. Al-Tous and I. Barhumi, "Joint power and bandwidth allocation for amplify-and-forward cooperative communications using Stackelberg game," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1678–1691, 2012.

[17] M. Burger and J. M. Schulte, "Adjoint methods for Hamilton-Jacobi-Bellman equations," in *Westfälische Wilhelms-Universität Münster*, 2010.