Energy Saving Dynamic Relaying Scheme in Wireless Cooperative Networks Using Markov Decision Process

Yifei Wei, Chaowei Wang, Mei Song, Yue Ma, Xiaojun Wang
Beijing Key Laboratory of Work Safety Intelligent Monitoring
School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, P.R. China
Email: weiyifei@bupt.edu.cn, wangchaowei@bupt.edu.cn, songm@bupt.edu.cn

***Abstract*—Energy saving becomes one of the most important design considerations in wireless cooperative networks which are composed of nodes typically powered by batteries that can supply only a finite amount of energy. In this paper, we propose a dynamic relaying scheme based on relay selection and physical-layer power control with the objective of minimizing the energy consumption for data transmission. We first develop a mathematical model for the cooperative relaying and analyze the total battery energy consumption to forward a symbol. Based on the analytic results and Markov channel model, we formulate the optimization problem that minimizes total average energy consumption as a Markov decision process, with which we can decide an optimal relay and the transmission power distributely. In the proposed scheme, the relay selection process and cooperation mode starts only when the direct transmission between source and destination node failed, which is energy efficient from a network sense. Numerical simulation show that the proposed scheme achieves significant energy savings.**

I. Introduction

It is well known that cooperative relaying technology can be used to increase cell coverage, improve average user throughput and cell-edge user throughput. In addition, relaying technology splits the direct transmission from source to destination into two or more hops, and the total energy consumption is expected to be greatly reduced, since the transceiver distances are smaller than that of the direct link and the path loss is thus significantly reduced [1]. However, the analyses in [2] and [3] show that it is not always true due to the extra transceiver circuit energy consumed by the relay nodes, and the cooperation schemes determine the energy consumption.

Compared with multi-node cooperative communication schemes, single relay cooperation schemes need neither cooperative beamforming nor distributed space time coding, where only the selected relay from a set of candidates participates in the data transmission. Single relay cooperation schemes are easy to implement and incur less cooperation overhead, and can potentially achieve high energy efficiency. Hence, relay selection strategies are practically appealing and have been discussed extensively. However, most of previous papers on selective single relay cooperation schemes [4]–[6] focus on system performance optimization, such as diversity gain or outage probability, and ignore the energy issue. Energy consumption is an important problem in wireless networks, and the energy efficiency of selective relay cooperation is investigated in [7], which showed that cooperative communication with single relay selection is a simple but effective communication scheme for energy constrained networks.

The authors in [8] developed an energy efficient relay selection scheme, which considers the MAC layer protocol and the power control strategy at the physical layer jointly, to extend the life time of wireless sensor networks. Energy allocation or power control can further improve the energy efficiency. Optimal energy distribution among cooperative nodes is studied in [9] from an information theoretic point of view based on the outage probability analysis. To minimize the energy consumption and maximize the network lifetime in energy constrained network, C. Y. Lee and G. U. Hwang [10] proposed a single relay selection in a distributed fashion. The authors in [11] considered a simple relay network with two nodes transmitting to a common destination and forwarding other nodes information, and propose power control policies to minimize total average energy consumption under an outage probability constraint. A fully distributed power allocation scheme was described in [12], where the power at each relay node is decided only on the basis of its instantaneous channel gain to the destination.

Although some works have been done for relay selection from a energy saving perspective, the foregoing schemes assume that the channel fading is slow enough such that the channel conditions remain in the same state from the current frame to the next. The estimated channel condition of the current frame is simply taken as the predicted channel condition for the next frame. However, this assumption is often not realistic in wireless mobile environment, since wireless channels are random variables due to fading, interference and users' movement. In this paper, the finite-state Markov channel (FSMC) models are used to predict the upcoming channel state, and the procedure of relay selection and transmission power decision is formulated as a Markov decision process. Moreover, the relay selection process starts before each packet transmission in existing schemes, while the relay selection process and cooperation mode starts only when the feedback from the destination node indicates failure of the direct transmission in the proposed scheme, which can save energy and make more efficient use of wireless channel.

The rest of this paper is organized as follows. In Section II, the system models are described. Section III formulates

This work is jointly supported by the National Natural Science Foundation of China under Grant 61101107, the Scientific Research and Innovation Plan for the Youth of BUPT under Grant 2011RC0305, and the National International Science and Technology Cooperation Project under Grant 2010DFA11320.

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

the problem as a Markov decision process and solves the problem. Section IV discusses the distributed dynamic relaying scheme. Some numerical and simulation results are presented in Section V. Finally, we conclude this study in Section VI.

II. System Models

We consider a two-hop relay network consisting of a source, a destination and N relay nodes. Each node operates in half-duplex mode and are equipped with single transmit and receive antennas. In our proposed scheme, cooperative relaying communication procedure can be divided into three phases: The first phase is transmission phase, the source node transmits (broadcasts) a data packet to its destination (and overhead by relay nodes). The second phase is channel estimation and relay selection phase, the destination node replies (broadcasts) a feedback signal indicating the success or failure of the direct transmission, and relay selection policy selects the optimal relay and determines its corresponding transmission energy per symbol based on local measurements of the feedback signal. The third phase is retransmission phase, the selected relay retransmits in an attempt to exploit spatial diversity.

Since the decode-and-forward (DF) relay scheme has advantages in digital processing and avoids noise amplification, the DF protocol is used in the relay network where the relay node first decodes the information from the source and then encodes and forwards the information to the destination. The set of relay candidates is defined as the relay nodes which can decode the data from the source node correctly, and denoted as $\mathcal{M} = \{1, 2, \ldots, M\}$. In this paper, the duration of transmission is divided into T equal-length time slots which correspond to the time interval between two continuous decision, and time $t \in \mathcal{T} = \{0, 1, \ldots, T-1\}$ stands for time instant at which decision need to be made. The relay selection policy must automatically decide which relay to be active and its corresponding transmission energy per symbol at each time slot, and such decision depends on the state of wireless channel.

A. Channel Model

We assume that the wireless channels are frequency non-selective channels that undergo independent Rayleigh fading. Thus, the channel gains from source to relays, denoted by h_{SR_m}, and from relays to destination, denoted by $h_{\mathrm{R}_m\mathrm{D}}$, are independent exponentially distributed random variables. The system operates in slotted time and the channel gain between relay node and destination node is denoted by $h_{\mathrm{R}_m\mathrm{D}}(t)$. We assume a block fading model [13] for the channel gains so that their value remains constant during one time slot and changes at the beginning of the next time slot.

We use finite-state Markov channel (FSMC) model to represent the time-varying behavior of the Rayleigh fading channel, since the use of a first-order Markov process to model the correlation structure of a Rayleigh channel has been shown to be a good approximation in [14]. FSMC is based on the partitioning of the received signal-to-noise ratio (SNR) in a finite number of states [15]. Since the received SNR is proportional to the channel gain, we partition the range of the channel gain into K levels. That is, the channel gain between relay node m and destination node $h_{\mathrm{R}_m\mathrm{D}}$ varies over a set of states, denoted as $\mathcal{H} = \{\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_{K-1}\}$, at each time slot according to a set of Markov transition probabilities. Let $\phi_{x_m y_m}$ denote the probability that $h_{\mathrm{R}_m\mathrm{D}}$ moves from state x_m to state y_m. The $K \times K$ channel state transition probability matrix is defined as:

$$\Phi = [\phi_{x_m y_m}]_{K \times K}\,, \tag{1}$$

where $\phi_{x_m y_m} = \Pr\left(h_{\mathrm{R}_m\mathrm{D}}(t+1) = y_m \mid h_{\mathrm{R}_m\mathrm{D}}(t) = x_m\right)$, and $x_m, y_m \in \mathcal{H}$. Relay m can estimate its current channel state $h_{\mathrm{R}_m\mathrm{D}}(t)$ after receiving feedback signal from the destination at time t.

B. Energy model

Given the channel gain between transmit-node i and receive-node j, denoted by h_{ij}, and the transmission energy per symbol, denoted by E_t, the received signal energy will be $E_r = E_t|h_{ij}|^2$. We assume that a certain level of quality of service (QoS) requirement such as bit error rate (BER) in the relay network is given. To meet the target average BER, the average received SNR or the received signal energy must exceeds a threshold given the modulation and coding scheme. We assume that the threshold of received signal energy is denoted as E_r^*, thus the selected relay node has to forward the signal to the destination with transmission energy per symbol:

$$E_t = \frac{E_r^*}{|h_{\mathrm{RD}}|^2} \tag{2}$$

Therefore, there is an adaptive power control scheme in the selected relay node which adaptively changes the transmission energy per symbol E_t based on its channel gain.

In [16], a realistic nonlinear model is adopted to describe the battery discharge process, and the total battery energy consumption is formulated as a quadratic function of the transmission energy for transmitting a symbol:

$$\mathcal{E} = \frac{\omega\gamma_p(1+\alpha)^2}{V\eta^2}{E_t}^2 + \frac{1+\alpha}{\eta}E_t + \frac{P_{ct}T_p + P_{cr}T_d}{\eta} \tag{3}$$

where ω is the battery efficiency factor, γ_p is a parameter determined by the normalized transmit pulse shape, V is the battery voltage, η is the transfer efficiency of the DC/DC converter, α is the extra power loss factor of the power amplifier, P_{ct} and P_{cr} are the transmitter and receiver circuit power consumption respectively, T_p is the pulse duration at transmitter and T_d is the demodulation duration in which the receiving circuit needs to be turned on. We assume that these parameters are constant, thus the energy consumption $\mathcal{E}$ for forwarding the signal to the destination can be expressed as an explicit function of channel gain between relay node m and destination node $h_{\mathrm{R}_m\mathrm{D}}$:

$$\mathcal{E}(h_{\mathrm{R}_m\mathrm{D}}) = \frac{C_4}{|h_{\mathrm{R}_m\mathrm{D}}|^4} + \frac{C_2}{|h_{\mathrm{R}_m\mathrm{D}}|^2} + C_0 \tag{4}$$

Apparently, total battery energy consumption $\mathcal{E}$ consists of three parts corresponding to the three respective terms in (4):

the excess energy loss due to the nonlinear battery discharge process which is proportional to the square of the energy of the transmitted signal, the energy carried by the transmitted signal, and the transceiver circuit energy consumption which is a constant depending on the power of the circuit and the duration. The optimization problem is to select the best relay and decide the transmission energy per symbol in each time slot based on its channel gain with the objective of minimizing the energy consumption.

III. Problem Formulation

A Markov decision process can be described as follows: the system m is in one of its states $h_m(t) \in \mathcal{H}_m$ in each time slot t, an action or selection decision $a_m(t)$ is taken according to its state, then the system reward $R_{h_m(t)}^{a_m(t)}$ is earned, and their states change to another state according to its transition probability matrix.

The fully distributed relay selection process is performed based on the channel gains from relays to destination. That is, each relay node estimates the channel gain from relay to destination through feedback signal broadcast by the destination node in time slot t, the system will select the optimal relay and determine its corresponding transmission energy per symbol, denoted as $a_m(t) \in \{0(\text{not selected}), 1(\text{selected})\}$. The system reward must represent the optimization objective. Therefore, we formulate the energy consumption $\mathcal{E}$ for forwarding the signal to the destination as the system reward:

$$R_{h_m(t)}^{a_m(t)} = a_m(t)\mathcal{E}(h_{\mathrm{R}_m\mathrm{D}}), \tag{5}$$

where $h_{\mathrm{R}_m\mathrm{D}}(t)$ is the system state defined in (1). The instantaneous system reward $R_{h_m(t)}^{a_m(t)}$ is earned for relay m in state $h_m(t)$ when it takes action $a_m(t)$ in time slot t.

For a stochastic process, the maximum immediate reward does not mean the maximum expected long-term accumulated reward. Thus the objective of the dynamic relaying scheme is to find a policy that minimize the energy consumption for data transmission, and the optimum value is:

$$\mathcal{Z}^* = \max_{u \in \mathcal{U}} E_u \left[\sum_{t=0}^{T-1} (R_{h_1(t)}^{a_1(t)} + R_{h_2(t)}^{a_2(t)} + \cdots + R_{h_M(t)}^{a_M(t)})\beta^t \right]. \tag{6}$$

where $u \in \mathcal{U}$ is the Markovian policy, and β is a discount factor that means rewards receives later in time will have less value than an equivalent reward received closer to the present.

The discounted Markov decision process is formulated as the linear program [17]:

$$(\mathrm{LP}) \quad \mathcal{Z}^* = \max_{\boldsymbol{x} \in X} \sum_{m \in \mathcal{M}} \sum_{h_m \in \mathcal{H}_m} \sum_{a_m \in \{0,1\}} R_{h_m}^{a_m} x_{h_m}^{a_m}, \tag{7}$$

where $X = \{\boldsymbol{x} = (x_{h_m}^{a_m}(u))_{h_n \in \mathcal{H}_n, a_m \in \{0,1\}, m \in \mathcal{M}} \mid u \in \mathcal{U}\}$ is the corresponding performance region spanned by performance vector $\boldsymbol{x}$ under all admissible policies $u \in \mathcal{U}$, and the performance measure $x_{h_m}^{a_m}(u)$ represents the total expected discounted time that relay m take action a_m in state h_m under admissible policy u. Let α_{h_m} denote the probability that the initial state is h_m, for $h_m \in \mathcal{H}_m$, and the initial state probability vector $\boldsymbol{\alpha} = (\alpha_{h_m})_{h_m \in \mathcal{H}_m}$ is given. The first-order relaxation is formulated as the linear program in [17]:

$$(\mathrm{LP}^1) \quad \mathcal{Z}^1 = \max \sum_{m \in \mathcal{M}} \sum_{h_m \in \mathcal{H}_m} \sum_{a_m \in \{0,1\}} R_{h_m}^{a_m} x_{h_m}^{a_m}$$

subject to

$$\begin{aligned} &\boldsymbol{x}_m \in \mathscr{P}_m^1, \quad m \in \mathcal{M}, \\ &\sum_{m \in \mathcal{M}} \sum_{h_m \in \mathcal{H}_m} x_{h_m}^1 = \frac{1}{1-\beta}. \end{aligned} \tag{8}$$

The authors of [17] interpreted the primal-dual heuristic as a priority-index heuristic under some mixing assumptions on active and passive transition probabilities. Please refer to [17] for details. The obtained dynamic relaying policy has an indexable property that reduces the computational complexity dramatically. We use this priority-index rule, which set active the relay that has the smallest index and determine its corresponding transmission energy per symbol. For relay m in state $h_{\mathrm{R}_m\mathrm{D}}(t)$, we denote by the index $\delta_m(h_m)$. At each epoch, the relay with the smallest index $\delta_m(h_m)$ is set to be active and the transmission energy per symbol is calculated with (2), while other networks are passive.

IV. Distributed Dynamic Relaying Scheme

The proposed scheme selects the optimal relay node and determines the average energy per symbol to aid the source-to-destination communication. We assume that wireless channels are reciprocal which means that channel from relay node to destination node is the same from destination node to relay node. This assumption is fulfilled in time division duplexing systems where the round-trip duplex time is much shorter than the coherence time of the channel, or in frequency division duplexing systems where the frequency duplex separation is smaller than the coherence bandwidth. We also assume that all nodes can adjust their instantaneous transmission power P_t, and denote the symbol duration as T_s which is fixed, and use the transmission energy per symbol $E_t = P_t * T_s$ as the control parameter. The overall relaying communication process for one packet transmission consists of three phases.

The first phase is transmission phase. When a source node has packet to transmit, it broadcasts the packet to the destination and relay nodes $\mathcal{N} = \{1, 2, \ldots, N\}$ simultaneously using a fixed transmission energy. Therefore, the source node doesn't involved in the relay selection process, and the broadcast nature of the wireless channel can be exploited to save energy. Depending on the channel states, some or all of the relay nodes successfully decode the packet from the source.

The second phase is channel estimation and relay selection phase. The destination node indicates success or failure by broadcasting a feedback signal to the source and relay nodes, the set of relays that both have decoded the packet from the source successfully and received the feedback signal from the destination constitutes the relay candidates, and denoted as $\mathcal{M} \subseteq \mathcal{N}$. If the source to destination channel is sufficiently good, the feedback indicates success of the direct transmission, and the relay does nothing. If the source to destination channel is not sufficiently good for successful direct transmission, the feedback requests that the relay forward what it has received from the source. Based on local measurements of the feedback signal by relay nodes, the instantaneous channel state between the relay node and the destination node $h_{\mathrm{R}_m\mathrm{D}}$ can be abtained. With the current state $h_m(t)$, each relay candidate calculates its index $\delta_m(h_m)$ and broadcasts a candidate index packet (CIP) containing its index. When other relay candidates receive this CIP packet, they will compare the received index with their own index and broadcast their own CIP packet only if their own index is smaller than the received index, otherwise they will keep silent. We can simply set a timer on each node to stop receiving the delayed CIP packets. After the timer expires, the relay with the smallest index is selected to cooperate with.

The third phase is retransmission phase. The selected relay retransmits the decoded packet to the destination with the required transmission energy per symbol computed with (2). The destination tries to combine the two transmissions from the source and selected relay for joint decoding. The transmitted signals in the first phase one and the third phase will have the same length and format, but with different energy.

The proposed scheme saves energy and makes more efficient use of the degrees of freedom of the channel, because they repeat only rarely, and the best relay is selected in a distributed fashion with minimum signaling overhead.

V. Numerical Results and Discussions

In this section, we compare the proposed dynamic relaying scheme with the traditional opportunistic relaying scheme which selects relay before the packet transmission and the selection decision is made based on the current estimated channel condition. We set the battery efficiency factor $\omega = 0.05$, the battery voltage $V = 3.5$, the transfer efficiency of the DC/DC converter $\eta = 0.8$, the extra power loss factor of the power amplifier $\alpha = 0.33$, the transmitter and receiver circuit power consumption $P_{ct} = 100mW$ and $P_{cr} = 50mW$, the pulse duration at transmitter $T_p = 1.5 \times 10^{-4}s$ and the demodulation duration $T_d = 2.5 \times 10^{-4}s$. All nodes can adjust their instantaneous transmission power within the range of [0, 50mW], the symbol duration is set $T_s = 10^{-4}$s and the time slot is $T = 1$s, thus the transmission energy per symbol will be controlled within the range of [0, $50 \times 10^{-7}J$]. We assume that the state of wireless channel can be partitioned into four states and set the state transition probability matrix as:

$$\Phi = \begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.15 & 0.60 & 0.15 & 0.10 \\ 0.10 & 0.15 & 0.60 & 0.15 \\ 0.05 & 0.15 & 0.20 & 0.60 \end{pmatrix}.$$

We set the discount factor $\beta = 0.8$ in the simulations. The initial states of the relays are random, and 10 runs with different seed numbers are conducted for each simulation and output data are averaged over these runs. We run the simulations for 400 seconds, with $M = 8$ relay candidates. Fig. 1 is the total average energy consumption per slot time. It can be seen that the proposed dynamic relaying scheme consumes less energy than traditional opportunistic relaying scheme, since there are some time slots that the relay does nothing when the feedback indicates success of the direct transmission, and the proposed dynamic relaying scheme can select the optimal relay and set the transmission energy per symbol for the subsequent frame at almost every decision time.

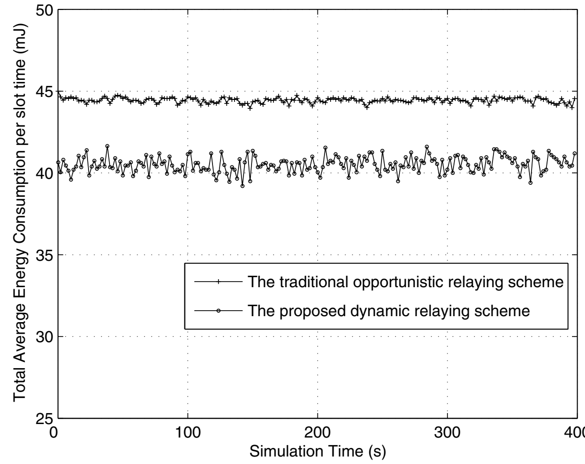

Fig. 1. Total average energy consumption per slot time

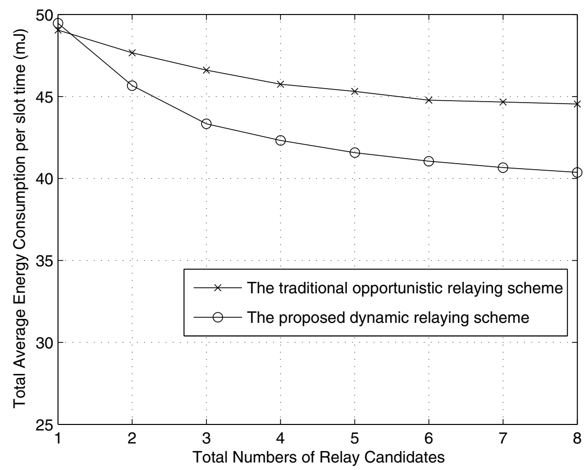

Fig. 2. Total average energy consumption per slot time vs. M

Fig. 2 is the total average energy consumption per slot time with different numbers of relay candidates M. As the number of relay candidates increases, the probability that there exists at least one relay with the best channel become higher, and thus the total average energy consumption per slot time decreases. It can be seen that the proposed dynamic relaying scheme outperforms the traditional opportunistic relaying scheme, since it saves the energy when the direct transmission is successful.

Since the traditional opportunistic relaying scheme selects the optimal relay for the subsequent packet transmission based on the current observed value, which may change in the

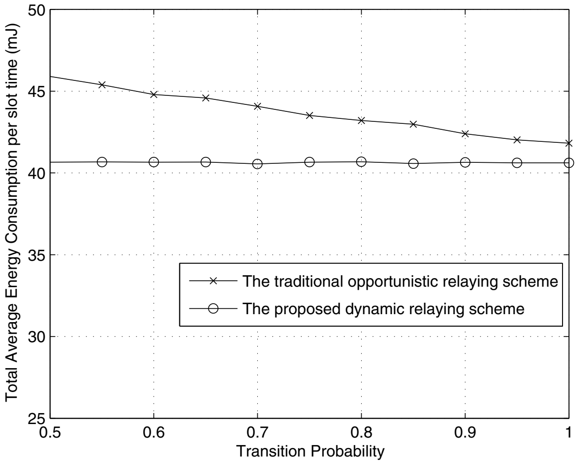

Fig. 3. Total average energy consumption per slot time vs. ϕ_{ii}

subsequent time slot with the probability of 0.4 under the given parameter Φ. We conduct simulations under different state transition probabilities, which means the probability of staying in the same state. Fig. 3 shows the total average energy consumption per slot time when increasing ϕ_{ii} from 0.5 to 1. We can see that the traditional opportunistic relaying scheme is getting closer to the proposed dynamic relaying scheme with the increase of the probability of staying in the same state.

VI. Conclusions

In this paper, we propose a dynamic relaying scheme based on relay selection and physical-layer power control with the objective of minimizing the energy consumption for data transmission. Finite-state Markov channel model is used to predict the upcoming channel state, and the procedure of relay selection and transmission power decision is formulated as a Markov decision process. The problem can be solved using linear programming and primal dual-index heuristic algorithm, and the obtained solution has an indexability property that dramatically simplifies the computation and implementation of the scheme. Moreover, the relay selection process starts before each packet transmission in existing schemes, while the relay selection process and cooperation mode starts only when the feedback from the destination node indicates failure of the direct transmission in the proposed scheme, which can save energy and make more efficient use of wireless channel. Numerical and simulation results are presented to show that the proposed scheme is energy saving.

References

[1] S. Lakkavalli, A. Negi, and S. Singh, "Stretchable architectures for next generation cellular networks," in *Proc. Intl. Symp. Advanced Radio Technol.*, pp. 59–65, 2003.

[2] J. Song, H. Lee, and D. Cho, "Power consumption reduction by multi-hop transmission in cellular networks," in *Proc. IEEE Veh. Technol. Conf.*, pp. 3120–3124, 2004.

[3] K. Schwieger and G. Fettweis, "Power and energy consumption for multihop protocols: a sensor network point of view," in *Proc. International Workshop Wireless Ad-Hoc Networks*, pp. 1–5, 2005.

[4] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. R. Liu, "Relay selection in multi-node cooperative communications: When to cooperate and whom to cooperate with?," in *Proc. IEEE Globecom*, pp. 1–5, Nov. 2006.

[5] M. M. Fareed and M. Uysal, "A novel relay selection method for decode-and-forward relaying," in *Proc. Canadian Conference on Electrical and Computer Engineering (CCECE 2008)*, (Niagara Falls, ON, Canada), pp. 135–140, May 2008.

[6] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 659–672, Mar. 2006.

[7] R. Madan, N. Mehta, A. Molisch, and J. Zhang, "Energy-efficient cooperative relaying over fading channels with simple relay selection," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3013–3025, August 2008.

[8] Z. Zhou, S. Zhou, J. Cui, and S. Cui, "Energy-efficient cooperative communications based on power control and selective relay in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3066–3078, August 2008.

[9] M. O. Hasna and M.-S. Alouini, "Optimal power allocation for relayed transmissions over rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1999–2004, Nov. 2004.

[10] C. Y. Lee and G. U. Hwang, "Minimum energy consumption design of a two-hop relay network for qos guarantee," *Wireless Telecommunications Symposium (WTS)*, pp. 1–6, April 2010.

[11] T. Zhang, S. Zhao, L. Cuthbert, and Y. Chen, "Energy-efficient cooperative relay selection scheme in mimo relay cellular networks," in *Proc. IEEE International Conference on Communication Systems 2010*, (Singapore), pp. 269–273, Nov. 2010.

[12] M. Chen, S. Serbetli, and A. Yener, "Distributed power allocation strategies for parallel relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 552–561, Feb. 2008.

[13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[14] C. C. Tan and N. C. Beaulieu, "On first-order markov modeling for the rayleigh fading channel," *IEEE Trans. Computers*, vol. 48, pp. 2032–2040, Dec. 2000.

[15] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, 1999.

[16] D. Duan, F. Qu, L. Yang, A. Swami, and J. C. Principe, "Modulation selection from a battery power efficiency perspective," *IEEE Trans. Commun.*, vol. 58, pp. 1907–1911, July 2010.

[17] D. Berstimas and J. Niño-Mora, "Restless bandits, linear programming relaxations, and a primal dual index heuristic," *Operations Research*, vol. 48, no. 1, pp. 80–90, 2000.