On Energy Efficiency and Optimum Resource Allocation of Relay Transmissions in the Low-Power Regime

Yingwei Yao, *Member, IEEE*, Xiaodong Cai, *Senior Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—Relay links are expected to play a critical role in the design of wireless networks. This paper investigates the energy efficiency of relay communications in the low-power regime under two different scenarios: when the relay has unlimited power supply and when it has limited power supply. A system with a source node, a destination node, and a single relay operating in the time division duplex (TDD) mode was considered. Analysis and simulations are used to compare the energy required for transmitting one information bit in three different relay schemes: amplify and forward (AnF), decode and forward (DnF), and block Markov coding (BMC). Relative merits of these relay schemes in comparison with direct transmissions (direct Tx) are discussed. The optimal allocation of power and transmission time between source and relay is also studied.**

***Index Terms*—Amplify and forward, block Markov coding, decode and forward, energy efficiency, relay transmission.**

I. Introduction

IN RECENT years, relay transmissions have attracted a lot of attention due to the emergence of wireless ad hoc links as a flexible architecture capable of supporting many different applications in the context of data networks, home networks, and sensor networks [1]. Unlike cellular networks and most wireless local area networks (WLANs), where nodes communicate through a central node (base station/access point) linked to a wired backbone, in wireless ad hoc networks, communication between different nodes relies on intermediate nodes acting as relays. It has been shown in [2] that when using simple point-to-point relays (with no source-relay cooperation), the capacity per node in a wireless network goes to zero as the number of nodes grows large. So, it is imperative to explore more efficient relay strategies.

Manuscript received December 2, 2003; revised August 1, 2004; accepted October 18, 2004. The editor coordinating the review of this paper and approving it for publication is H. Yanikomeroglu. This work was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011 and was also supported by the USDoD ARO Grant No. W911NF-05-1-0283. This paper was presented in part at the IEEE International Symposium on Information Theory, Chicago, IL, June 27–July 2, 2004. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.

Y. Yao is with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: yyao@uic.edu).

X. Cai is with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33124 USA (e-mail: x.cai@miami.edu).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TWC.2005.858294

Due to the broadcasting nature of wireless transmissions, the information transmitted from source to relay can also be received by the destination node. Relay schemes exploiting this characteristic can achieve better performance and/or transmission rate over point-to-point relaying, and will be the focus of this paper. Capacity for the relay nonfading channel was investigated in [3], where upper and lower capacity bounds were obtained using the max-flow-min-cut theorem with superimposed block Markov coding (BMC). More recently, the same methods have been used to study relay capacity in a more realistic scenario, where the relay uses frequency division duplexing (FDD) or time division duplexing (TDD) instead of transmitting and receiving simultaneously over the same frequency band [4]. Outage performance and throughput of various simpler relay strategies have been studied in several papers [5]–[8]. Design issues pertaining to power allocation and coding have also been considered in [7] and [9]–[12].

In many wireless ad hoc networks, individual nodes are powered by battery and transmit at a low data rate. For example, the microsensor system specification for a machine monitoring application given in [13] requires each sensor to process about 20 2-B messages/s and to operate for more than 5 years on a single AA battery. For such a system, energy efficiency should dictate its design and operation. Systems limited by energy constraints do not necessarily operate at a low data rate. Ultra-wideband (UWB) systems, which attract a lot of research activities lately, have the potential of supporting high data rate multimedia transmissions while operating in the power-limited region because of the huge bandwidth available.

As pointed out in [14], reliance on relay transmissions and energy constraints are two of the most important characteristics of wireless networks. However, the energy efficiency of different relay schemes has not been studied until very recently. The minimum achievable energy per bit for the relay channel coding proposed in [3] was investigated in [15] for nonfading channels. In addition to minimum energy per bit, the wideband slope has been shown to be an important performance criterion that leads to better understanding of the bandwidth–energy tradeoff in the low-power regime [16]. Achievable rates of block Markov transmission in the low power regime were obtained in [17] using the approach in [16], which leads to the optimal power allocation between source and relay. In this paper, we will use similar tools to compare the energy efficiency of several different relay strategies, including amplify and forward (AnF), decode and forward (DnF), and BMC. Optimal resource allocation will also be addressed.

The rest of this paper is organized as follows. In Section II, we introduce the system model for relay transmissions. In Section III, we investigate the energy efficiency of various relay strategies and derive optimal resource allocation schemes. Based on our analysis, numerical results and simulations are then used in Section IV to study the relative merits of different relay methods and their implications to the design of energy constrained systems. Finally, we draw our conclusions in Section V.

II. System Model

Consider a system in which one node (source) tries to send information to another node (destination) with the help of an intermediate node (relay). Since a relay cannot transmit and receive data over the same time–frequency band, practical relay systems typically rely on FDD or TDD operation. Without loss of generality, we assume that the relay operates in TDD mode. In particular, during the first time slot containing $N_1 := \alpha N$ symbol periods, the source transmits an $N_1 \times 1$ block of symbols $\boldsymbol{x}_{s1}$ that is received at the relay and the destination as

$$\boldsymbol{y}_r = h_{sr}\sqrt{\frac{P_{s1}}{\alpha}}\boldsymbol{x}_{s1} + \boldsymbol{n}_r \tag{1}$$

$$\boldsymbol{y}_d = h_{sd}\sqrt{\frac{P_{s1}}{\alpha}}\boldsymbol{x}_{s1} + \boldsymbol{n}_d \tag{2}$$

where the $N_1 \times 1$ vectors $\boldsymbol{y}_r$ and $\boldsymbol{y}_d$ are the received blocks of symbols at the relay and the destination, respectively, and h_{sr} and h_{sd} denote the channels between source and relay, and source and destination, respectively. In the second time slot comprising $N_2 := (1-\alpha)N$ symbol periods, the source transmits an $N_2 \times 1$ block of symbols $\boldsymbol{x}_{s2}$ while the relay transmits $\boldsymbol{x}_r$ that is received by the destination as

$$\boldsymbol{y}_d^r = h_{sd}\sqrt{\frac{P_{s2}}{1-\alpha}}\boldsymbol{x}_{s2} + h_{rd}\sqrt{\frac{P_r}{1-\alpha}}\boldsymbol{x}_r + \boldsymbol{n}_d^r \tag{3}$$

where h_{rd} is the channel between relay and destination. We have normalized the transmitted vectors such that $E[\|\boldsymbol{x}_{s1}\|^2] = \alpha N$ and $E[\|\boldsymbol{x}_{s2}\|^2] = E[\|\boldsymbol{x}_r\|^2] = (1-\alpha)N$. So, the average source transmit power per symbol is

$$\frac{1}{N}E\left[\frac{P_{s1}}{\alpha}\|\boldsymbol{x}_{s1}\|^2 + \frac{P_{s2}}{1-\alpha}\|\boldsymbol{x}_{s2}\|^2\right] = P_{s1} + P_{s2} := P_s \tag{4}$$

and its counterpart at the relay is given by

$$\frac{1}{N}E\left[\frac{P_r}{1-\alpha}\|\boldsymbol{x}_r\|^2\right] := P_r. \tag{5}$$

We assume that the additive noise terms $\boldsymbol{n}_r$, $\boldsymbol{n}_d$, and $\boldsymbol{n}_d^r$ are independent white Gaussian distributed with zero mean and power spectral density N_0. The channel coefficients h_{sr}, h_{sd}, and h_{rd} are independent complex random variables. Each time slot is assumed to be large enough so that the channel fading processes are ergodic. We further assume that receivers have perfect channel information while transmitters only have statistical information of the channels; namely, the mean and the variance of $|h_{sr}|^2$, $|h_{rd}|^2$, and $|h_{sd}|^2$.

III. Energy Efficiency

Using linear quadratic approximate expansion, it has been shown that in the low-power low-spectral-efficiency regime the energy efficiency of a system depends on its spectral efficiency C as [16]

$$10\log_{10}\frac{E_b}{N_0}(\mathsf{C}) = 10\log_{10}\frac{E_b}{N_0}_{\min} + \frac{\mathsf{C}}{\mathcal{S}_0}10\log_{10}2 + o(\mathsf{C}) \tag{6}$$

where $f(x) \sim o(x) \Rightarrow \lim_{x\to 0} f(x)/x = 0$ and

$$\frac{E_b}{N_0}_{\min} := \frac{\log_e 2}{\dot{C}(0)} \tag{7}$$

$$\mathcal{S}_0 := \frac{2\left[\dot{C}(0)\right]^2}{-\ddot{C}(0)} \tag{8}$$

with $\dot{C}$ and $\ddot{C}$ denoting the first-order and second-order derivatives of $C(\mathsf{SNR})$, a function of the transmit signal-to-noise ratio $\mathsf{SNR} := P_s/(N_0W)$ over bandwidth W. Note that we have adopted the notations in [16], where $C(\mathsf{SNR})$ is a function of the SNR and is in nats/s/Hz, while C is a function of E_b/N_0 and is in bits/s/Hz.

In the ensuing analysis, we will use (6) to approximate the energy needed for transmitting one information bit in different relay transmission systems. We will consider two different scenarios: when the relay has unlimited power and when the relay has limited power supply. But first, let us examine the direct transmission (direct Tx) scheme, which we will use as a benchmark.

A. *Direct Tx*

The spectral efficiency of the direct Tx is

$$C = E_{h_{sd}}\log\left(1 + \frac{P_s|h_{sd}|^2}{N_0W}\right) := E\log\left(1 + |h_{sd}|^2\mathsf{SNR}\right). \tag{9}$$

It is easy to see that at zero SNR, $\dot{C}(0) = E[|h_{sd}|^2]$ and $\ddot{C}(0) = -E[|h_{sd}|^4]$. Defining $\lambda_{sd} := E[|h_{sd}|^2]$ and $\kappa_{sd} := E[|h_{sd}|^4]/[E[|h_{sd}|^2]]^2$, we have [cf. (7), (8)]

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\lambda_{sd}} \tag{10}$$

and

$$\mathcal{S}_0 = \frac{2}{\kappa_{sd}}. \tag{11}$$

At low spectral efficiencies (small C), we can skip the quadratic term in (6) to arrive at

$$10\log_{10}\frac{E_b}{N_0}(\mathsf{C}) \cong 10\log_{10}\frac{\log_e 2}{\lambda_{sd}} + (10\log_{10}2)\frac{\kappa_{sd}}{2}\mathsf{C}. \tag{12}$$

Equation (12) shows that at low spectral efficiency values, the energy efficiency of direct Tx is simply characterized by the second-order and fourth-order moments of the fading channel gain.

B. AnF

In the AnF strategy, the relay simply amplifies the signal received from the source in the first time slot and forwards it to the destination in the second time slot. Both slots have identical duration, i.e., $\alpha = 0.5$. Because the source in AnF does not transmit in the second time slot, $P_s = P_{s1}$. The transmit signal of the relay is then

$$\boldsymbol{x}_r = \frac{1}{\sqrt{2|h_{sr}|^2 P_s + N_0 W}} \boldsymbol{y}_r. \tag{13}$$

The spectral efficiency of such a scheme is

$$C = \frac{1}{N} \sup_{p(\boldsymbol{x}_{s1})} E\left[I\left(\boldsymbol{x}_{s1}; \boldsymbol{y}_d, \boldsymbol{y}_d^r | h_{sd}, h_{sr}, h_{rd}\right)\right] \tag{14}$$

which is achieved when $\boldsymbol{x}_{s1}$ is Gaussian distributed and can be shown to be [7]

$$C = \frac{1}{2} E\left[\log\left(1 + \frac{2P_s|h_{sd}|^2}{N_0 W} + f\left(\frac{2P_s|h_{sr}|^2}{N_0 W}, \frac{2P_r|h_{rd}|^2}{N_0 W}\right)\right)\right] \tag{15}$$

where $f(x, y) := xy/(1 + x + y)$.

Given (15), we wish to determine the constants in (7) and (8) for the AnF setup under different power considerations at the relay.

1) Relay With Unlimited Power: In certain wireless networks, some nodes may have access to fixed power sources while others have to rely on battery power. In a home network, for instance, portable devices like laptops, personal digital assistants (PDAs), and cordless phones often need to use battery, while desktop PCs may not have to worry about the power supply. Another interesting example is the multihop cellular architecture recently proposed in [18]. It has been shown that mobile stations serving as relays can significantly improve throughput of cellular networks. These mobile stations typically have wired links to the base stations and thereby power is not at a premium. Apparently, for such applications, it is meaningful to focus on the impact of relay methods on the source power consumption, assuming that the source is a portable device relying on battery.

Defining $\mathsf{SNR} := P_s/(N_0 W)$, we have [cf. (15)]

$$C(\mathsf{SNR}) = \frac{1}{2} E\left[\log\left(1 + 2|h_{sd}|^2 \mathsf{SNR} + f\left(2|h_{sr}|^2 \mathsf{SNR}, \frac{2P_r|h_{rd}|^2}{N_0 W}\right)\right)\right]. \tag{16}$$

It is straightforward to show that

$$\dot{C}(0) = \lambda_{sd} + E\left[2\frac{|h_{sr}|^2 |h_{rd}|^2 P_r}{N_0 W + 2|h_{rd}|^2 P_r}\right] \tag{17}$$

and $\ddot{C}(0)$ is given by (18) at the bottom of the page. If we let $P_r \to \infty$, using Lebesgue's dominated convergence theorem, we have

$$\lim_{P_r \to \infty} \dot{C}(0) = \lambda_{sd} + \lambda_{sr} \tag{19}$$

and

$$\lim_{P_r \to \infty} \ddot{C}(0) = -2E\left[\left(|h_{sd}|^2 + |h_{sr}|^2\right)^2\right] \tag{20}$$

where $\lambda_{sr} := E[|h_{sr}|^2]$. So, for large relay power, we have

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\lambda_{sd} + \lambda_{sr}} \tag{21}$$

$$\mathcal{S}_0 = \frac{(1 + \beta)^2}{\kappa_{sd} + \kappa_{sr}\beta^2 + 2\beta} \tag{22}$$

where $\beta := \lambda_{sr}/\lambda_{sd}$, $\lambda_{sr} := E[|h_{sr}|^2]$, and $\kappa_{sr} := E[|h_{sr}|^4]/\lambda_{sr}^2$.

As expected, the energy efficiency constants in (21) and (22) again depend on the second- and fourth-order moments of the source–relay and source–destination channels, but not on the relay–destination link since the relay can afford arbitrarily large power.

2) Relay With Limited Power: In applications such as sensor networks, almost all nodes depend on battery power. In this case, we are interested in relay strategies that conserve total energy consumption at the source and the relay.

Modifying now the SNR definition as $\mathsf{SNR} := (P_s + P_r)/(N_0 W)$, we have from (15)

$$C(\mathsf{SNR}) = \frac{1}{2} E\left[\log\left(1 + 2\gamma|h_{sd}|^2 \mathsf{SNR} + \frac{4\gamma(1-\gamma)|h_{sr}|^2 |h_{rd}|^2 \mathsf{SNR}^2}{1 + 2(1-\gamma)|h_{rd}|^2 \mathsf{SNR} + 2\gamma|h_{sr}|^2 \mathsf{SNR}}\right)\right] \tag{23}$$

$$\ddot{C}(0) = -E\left[\frac{2|h_{sd}|^4 + 4|h_{rd}|^2\left(|h_{sd}|^4 + |h_{sd}|^2|h_{sr}|^2 + |h_{sr}|^4\right)\frac{2P_r}{N_0 W} + 2|h_{rd}|^4\left(|h_{sd}|^2 + |h_{sr}|^2\right)^2\left(\frac{2P_r}{N_0 W}\right)^2}{\left(1 + |h_{rd}|^2 \frac{2P_r}{N_0 W}\right)^2}\right] \tag{18}$$

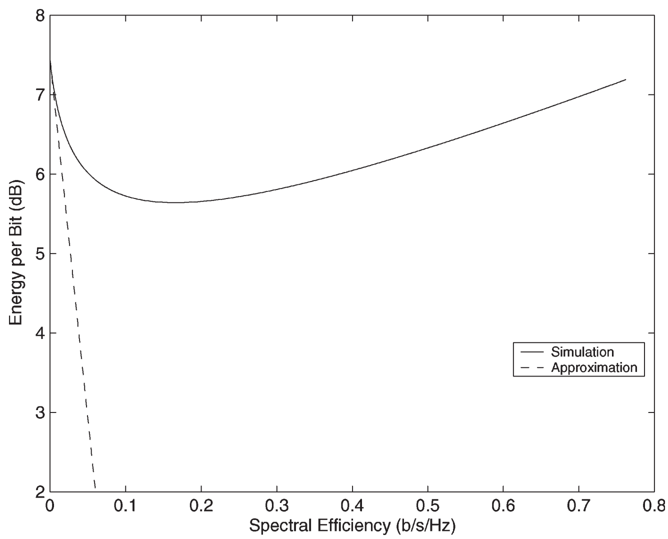

Fig. 1. Energy per bit versus spectral efficiency of AnF scheme.

where $\gamma := P_s/(P_s + P_r)$. After some algebra, we obtain

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\gamma\lambda_{sd}} \tag{24}$$

$$\mathcal{S}_0 = \frac{1}{\kappa_{sd} - \frac{2(1-\gamma)\lambda_{sr}\lambda_{rd}}{\gamma\lambda_{sd}^2}}. \tag{25}$$

Notice that different from (21) and (22), the power of the relay and the second- and fourth-order moments of the relay–destination link appear also in (24) and (25). Hence, for small values of C, the energy efficiency of AnF can be expressed as

$$10\log_{10}\frac{E_b}{N_0}(\gamma, \mathsf{C}) \cong 10\log_{10}\frac{\log_e 2}{\gamma\lambda_{sd}} + 10(\log_{10} 2)\left(\kappa_{sd} - \frac{2(1-\gamma)\lambda_{sr}\lambda_{rd}}{\gamma\lambda_{sd}^2}\right)\mathsf{C}. \tag{26}$$

From (26), we can see that E_b/N_0 might decrease as C increases if $\kappa_{sd} < c2(1-\gamma)\lambda_{sr}\lambda_{rd}/(\gamma\lambda_{sd}^2)$. So the E_b/N_0 at zero SNR given by (24) may not be the minimum achievable energy per bit, although we still use the notation $(E_b/N_0)_{\min}$ for notational consistency.

In Fig. 1, we show how E_b/N_0 changes with C. In our simulation, we assume Rayleigh fading channels with path loss coefficient 2. Relay is placed at the midpoint between source and destination, and $\gamma = 1/2$, implying equal power at the source and relay. We observe that E_b/N_0 does decrease initially as C increases, but reaches a minimum at about $\mathsf{C} = 0.15$ bit/s/Hz, and then starts increasing with C. So, the approximation of (26) is valid only when the SNR and spectral efficiency values are very low.

Close examination of (23) reveals that $\ddot{C}(0)$ is typically much larger than $\dot{C}(0)$ and the third-order derivative of C at zero SNR is much larger than $\ddot{C}(0)$, meaning that the quadratic and possibly higher order terms in (6) are not negligible. Also, this condition becomes even more severe as γ goes to 0. So, the first-order approximation given in (26) cannot be used in the optimization of γ, and we will have to rely on Monte Carlo simulations to obtain the optimum power allocation γ^*.

C. DnF

In DnF strategies, $\boldsymbol{x}_{s1}$ is first decoded by the relay, and then regenerated and forwarded to the destination during the second time slot. Again, the source does not transmit in the second time slot, and the first and second time slots have the same duration. The spectral efficiency for such a system is [7]

$$C = \min\{C_1, C_2\} \tag{27}$$

where $C_1 := 0.5E[\log(1 + (2P_s|h_{sd}|^2 + 2P_r|h_{rd}|^2)/(N_0W)]$ is determined by the maximum rate at which the receiver at the destination can reliably decode $\boldsymbol{x}_{s1}$ from $\boldsymbol{y}_d$ and $\boldsymbol{y}_d^r$, and $C_2 := 0.5E[\log(1 + 2P_s|h_{sr}|^2/(N_0W))]$ is determined by the capacity of the source–relay link.

1) Relay With Unlimited Power: Defining $\mathsf{SNR} := P_s/(N_0W)$, we have $C_1(\mathsf{SNR}) = 0.5E[\log(1 + 2|h_{sd}|^2\mathsf{SNR} + |h_{rd}|^2(P_r/N_0W))]$. When $P_r \neq 0$, $C_1(0) = 0.5E[\log(1 + |h_{rd}|^2(P_r/N_0W))] > 0$. So, for $\mathsf{C} \leq 0.5E[\log_2(1 + |h_{rd}|^2(P_r/N_0W))]$, we have $(E_b/N_0)_1 = 0$, where $(E_b/N_0)_1$ is the required source energy per bit such that $C_1 \geq \mathsf{C}$. Hence, for large P_r and small C values, we have

$$10\log_{10}\left(\frac{E_b}{N_0}\right) = 10\log_{10}\left(\frac{E_b}{N_0}\right)_2 \cong 10\log_{10}\frac{\log_e 2}{\lambda_{sr}} + 10\log_{10}(2)\kappa_{sr}\mathsf{C} \tag{28}$$

where $(E_b/N_0)_2$ is the source energy consumption per bit needed for C_2 to be greater than or equal to C.

2) Relay With Limited Power: By defining $\mathsf{SNR} := (P_s + P_r)/(N_0W)$, we have

$$C_1(\mathsf{SNR}) = \frac{1}{2}E\left[\log\left(1 + 2\gamma|h_{sd}|^2\mathsf{SNR} + 2(1-\gamma)|h_{rd}|^2\mathsf{SNR}\right)\right] \tag{29}$$

and

$$C_2(\mathsf{SNR}) = \frac{1}{2}E\left[\log\left(1 + 2\gamma|h_{sr}|^2\mathsf{SNR}\right)\right]. \tag{30}$$

It is then straightforward to show that

$$\frac{E_b}{N_0} = \max\left\{\left(\frac{E_b}{N_0}\right)_1, \left(\frac{E_b}{N_0}\right)_2\right\} \tag{31}$$

where

$$10\log_{10}\left(\frac{E_b}{N_0}\right)_1 \cong 10\log_{10}\frac{\log_e 2}{\gamma\lambda_{sd} + (1-\gamma)\lambda_{rd}} + (10\log_{10} 2)\frac{\kappa_{sd} + 2\frac{(1-\gamma)\lambda_{rd}}{\gamma\lambda_{sd}} + \left(\frac{(1-\gamma)\lambda_{rd}}{\gamma\lambda_{sd}}\right)^2\kappa_{rd}}{\left(1 + \frac{(1-\gamma)\lambda_{rd}}{\gamma\lambda_{sd}}\right)^2}\mathsf{C} \tag{32}$$

and

$$10\log_{10}\left(\frac{E_b}{N_0}\right)_2 \cong 10\log_{10}\frac{\log_e 2}{\gamma\lambda_{sr}} + (10\log_{10}2)\kappa_{sr}\mathsf{C}. \quad (33)$$

Remark: Note that (31) is not an affine function of C. To obtain an affine approximation, we need to compute the first- and second-order derivatives of $C(\mathsf{SNR}) = \min\{C_1(\mathsf{SNR}), C_2(\mathsf{SNR})\}$, instead of computing the derivatives of $C_1(\mathsf{SNR})$ and $C_2(\mathsf{SNR})$ separately. Since the derivatives of $C(\mathsf{SNR})$ at the zero SNR point are the derivatives of whichever of $C_1(\mathsf{SNR})$ and $C_2(\mathsf{SNR})$ is smaller in a small neighborhood of the zero SNR, the affine approximation does not contain as much information as the approximation given by (31). We would like to point out, however, that in most cases [when (32) and (33) do not intersect in the range of C we are interested in] the approximations given by the affine approximation and by (31) coincide with each other.

Since $(E_b/N_0)_2$ decreases monotonically as γ increases from 0 to 1, and $(E_b/N_0)_2 < (E_b/N_0)_1$ when $\gamma = 0$, the optimal power allocation γ^* between source and relay should be in the interval $[\gamma_0, 1]$, where γ_0 is the largest $\gamma \in [0, 1]$ that satisfies

$$\left(\frac{E_b}{N_0}\right)_1 = \left(\frac{E_b}{N_0}\right)_2 \quad (34)$$

if such a solution exists; otherwise, $\gamma^* = \gamma_0 = 1$. In the case of Rayleigh fading channels, (34) is equivalent to

$$\log(1+x) + \frac{2x\log_e 2}{(1+x)^2}\mathsf{C} = \log_e\frac{\lambda_{sr}}{\lambda_{sd}} \quad (35)$$

where $x := (1-\gamma)\lambda_{rd}/(\gamma\lambda_{sd}) \in [0, +\infty]$. A nonnegative solution of (35) exists if and only if (iff) $\lambda_{sr} \geq \lambda_{sd}$. So, as expected, we should allocate power to the relay iff the quality of the source–relay channel is better than that of the direct link between source and destination.

To show how to find the exact value of γ^*, we will focus on the most common scenario of $\lambda_{sr} < \lambda_{sd}$ and $\lambda_{rd} < \lambda_{sd}$ (other cases can be handled similarly). First, γ_0 can be found using root finding algorithms. Noting that $(E_b/N_0)_1 < (E_b/N_0)_2$ when $\gamma = 1$, the same must be true for all $\gamma \in (\gamma_0, 1]$, and we need only to examine the behavior of $(E_b/N_0)_1$ on this interval. Taking the first-order derivative of $\log_e(E_b/N_0)_1$, we have

$$\frac{d\log_e\left(\frac{E_b}{N_0}\right)_1}{d\gamma} = \frac{-(\lambda_{rd}-\lambda_{sd})^3\gamma^2 + A\gamma + B}{[(\gamma-1)\lambda_{rd} - \gamma_{sd}]^3} \quad (36)$$

where $A := 2\lambda_{rd}^3 - (4+\mathsf{C}\log_e 4)\lambda_{rd}^2\lambda_{sd} + (2-\mathsf{C}\log_e 4)\lambda_{rd}\lambda_{sd}^2$ and $B := -\lambda_{rd}^3 + (1+C\log_e 4)\lambda_{rd}^2\lambda_{sd}$. If the denominator on the right-hand side of (36) does not have real roots, then $\gamma^* = \gamma_0$; otherwise, ordering the roots so that $\gamma_1 \leq \gamma_2$, we have

$$\gamma^* = \begin{cases} \gamma_0, & \gamma_0 \geq \gamma_2 \\ \min\{1, \gamma_2\}, & \gamma_1 \leq \gamma_0 < \gamma_2 \\ \arg\min_{\{\gamma_0, \min\{1,\gamma_2\}\}}\left(\frac{E_b}{N_0}\right)_1, & \gamma_0 < \gamma_1. \end{cases} \quad (37)$$

Remark: It is well known that the performance of DnF can be improved if the relay can determine if it has decoded the source information correctly, either through measuring the received SNR or by checking the cyclic redundancy check (CRC) codes. For instance, while DnF achieves only a diversity of order 1 under high SNR, a simple adaptive protocol that makes the relay refrain from forwarding in the event of a decoding error enables full diversity [7]. In our case, the adaptive protocol amounts to switching back to direct Tx when communication via relay becomes more expensive than direct Tx. So, the E_b/N_0 for such a scheme will be the smaller one of the E_b/N_0 of DnF and that of direct Tx.

D. BMC

Instead of simply forwarding the decoded source information to the destination node, cooperative coding between source and relay can be used to improve system throughput. The basic idea is for the relay to send resolution information that helps the destination decode the source information being transmitted at a rate higher than the direct link capacity [19]. As in previous relay schemes, the relay operates under TDD mode. However, unlike AnF and DnF, the time slot in which the relay transmits and the time slot in which it receives do not have to have the same duration. In particular, suppose that during time instants $n = 1, \ldots, \alpha N$, the source transmits $\boldsymbol{x}_{d1}$, while during $n = \alpha N + 1, \ldots, N$, the source transmits $\boldsymbol{x}_{d2}$ and the relay transmits $\boldsymbol{x}_r$. Since the transmitters know only the channel statistics, no phase synchronization is performed at the transmitters and $\boldsymbol{x}_{d2}$ and $\boldsymbol{x}_r$ have independent phases. The achievable rate (lower bound) has been obtained in [10], i.e.,

$$C_{lb} = \min\{C_1, C_2\} \quad (38)$$

where

$$C_1 = \alpha E\left[\log\left(1 + \frac{P_{s1}|h_{sr}|^2}{\alpha N_0 W}\right)\right] + (1-\alpha)E\left[\log\left(1 + \frac{P_{s2}|h_{sd}|^2}{(1-\alpha)N_0 W}\right)\right] \quad (39)$$

and

$$C_2 = \alpha E\left[\log\left(1 + \frac{P_{s1}|h_{sd}|^2}{\alpha N_0 W}\right)\right] + (1-\alpha)E\left[\log\left(1 + \frac{P_{s2}|h_{sd}|^2}{(1-\alpha)N_0 W} + \frac{P_r|h_{rd}|^2}{(1-\alpha)N_0 W}\right)\right]. \quad (40)$$

1) Relay With Unlimited Power: Denoting $\mathsf{SNR} := (P_{s1} + P_{s2})/(N_0 W)$ and $\theta := P_{s1}/(P_{s1} + P_{s2})$, we have

$$C_1(\mathsf{SNR}) = \alpha E\left[\log\left(1 + \frac{\theta}{\alpha}|h_{sr}|^2\mathsf{SNR}\right)\right] + (1-\alpha)E\left[\log\left(1 + \frac{1-\theta}{1-\alpha}|h_{sd}|^2\mathsf{SNR}\right)\right] \quad (41)$$

and

$$C_2(\mathsf{SNR}) = \alpha E\left[\log\left(1+\frac{\theta}{\alpha}|h_{sd}|^2\mathsf{SNR}\right)\right] + (1-\alpha)E\left[\log\left(1+\frac{1-\theta}{1-\alpha}|h_{sd}|^2\mathsf{SNR}+\frac{P_r|h_{rd}|^2}{(1-\alpha)N_0W}\right)\right]. \quad (42)$$

It is easy to see that for large P_r and small C, $(E_b/N_0)_2 = 0$. Since $\dot{C}_1(0) = \theta\lambda_{sr} + (1-\theta)\lambda_{sd}$ and $-\ddot{C}_1(0) = \theta^2 E[|h_{sr}|^4]/\alpha + (1-\theta)^2 E[|h_{sd}|^4]/(1-\alpha)$, we have for small C

$$10\log_{10}\frac{E_b}{N_0}(\mathsf{C}) \cong 10\log_{10}\frac{\log_e 2}{\theta\lambda_{sr}+(1-\theta)\lambda_{sd}} + (10\log_{10}2)\frac{\frac{\kappa_{sd}}{1-\alpha}+\frac{\theta^2}{\alpha(1-\theta)^2}\left(\frac{\lambda_{sr}}{\lambda_{sd}}\right)^2\kappa_{sr}}{2\left(1+\frac{\theta\lambda_{sr}}{(1-\theta)\lambda_{sd}}\right)^2}\mathsf{C}. \quad (43)$$

Given θ, (43) attains the minimum when $1/\alpha = 1 + (1-\theta)\lambda_{sd}/(\theta\lambda_{sr})\sqrt{\kappa_{sd}/\kappa_{sr}}$, i.e.,

$$10\log_{10}\frac{E_b}{N_0}(\mathsf{C}) \cong 10\log_{10}\frac{\log_e 2}{\theta\lambda_{sr}+(1-\theta)\lambda_{sd}} + (10\log_{10}2)\frac{\kappa_{sd}\left(1+\frac{\theta}{1-\theta}\frac{\lambda_{sr}}{\lambda_{sd}}\sqrt{\frac{\kappa_{sr}}{\kappa_{sd}}}\right)^2}{2\left(1+\frac{\theta\lambda_{sr}}{(1-\theta)\lambda_{sd}}\right)^2}\mathsf{C}. \quad (44)$$

Equation (44) summarizes nicely how energy efficiency in the BMC relay strategy depends on the spectral efficiency through the fraction of power per time slot (captured through θ) and the source–relay and source–destination channel statistics.

While a closed-form solution for the optimal value of the power allocation ratio θ exists, it is rather cumbersome. If the source–relay channel quality is at least as good as the source–destination channel quality ($\lambda_{sr} \geq \lambda_{sd}$ and $\kappa_{sr} \leq \kappa_{sd}$), however, a much simpler solution can be obtained. In this case, (44) achieves the minimum when $\theta^* = 1$, i.e.,

$$10\log_{10}\frac{E_b}{N_0}(\mathsf{C}) \cong 10\log_{10}\frac{\log_e 2}{\lambda_{sr}} + (10\log_{10}2)\frac{\kappa_{sr}}{2}\mathsf{C}. \quad (45)$$

2) Relay With Limited Power: Defining $\mathsf{SNR} := (P_{s1} + P_{s2} + P_r)/(N_0W)$, we have

$$C_1(\mathsf{SNR}) = \alpha E\left[\log\left(1+\frac{\theta\gamma}{\alpha}|h_{sr}|^2\mathsf{SNR}\right)\right] + (1-\alpha)E\left[\log\left(1+\frac{(1-\theta)\gamma}{1-\alpha}|h_{sd}|^2\mathsf{SNR}\right)\right] \quad (46)$$

and

$$C_2(\mathsf{SNR}) = \alpha E\left[\log\left(1+\frac{\theta\gamma}{\alpha}|h_{sd}|^2\mathsf{SNR}\right)\right] + (1-\alpha) \times E\left[\log\left(1+\frac{(1-\theta)\gamma}{1-\alpha}|h_{sd}|^2\mathsf{SNR}+\frac{1-\gamma}{1-\alpha}|h_{rd}|^2\mathsf{SNR}\right)\right]. \quad (47)$$

After some algebra, we can show that

$$\left(\frac{E_b}{N_{0\,\min}}\right)_1 = \frac{\log_e 2}{\theta\gamma\lambda_{sr}+(1-\theta)\gamma\lambda_{sd}} \quad (48)$$

$$\mathcal{S}_{01} = \frac{2}{\kappa_{sd}}\frac{\left(1+\frac{\theta\lambda_{sr}}{(1-\theta)\lambda_{sd}}\right)^2}{\frac{1}{1-\alpha}+\frac{\kappa_{sr}}{\alpha\kappa_{sd}}\left(\frac{\theta\lambda_{sr}}{(1-\theta)\lambda_{sd}}\right)^2} \quad (49)$$

$$\left(\frac{E_b}{N_{0\,\min}}\right)_2 = \frac{\log_e 2}{\gamma\lambda_{sd}+(1-\gamma)\lambda_{rd}} \quad (50)$$

and $\mathcal{S}_{02}$ is given by (51) at the bottom of the page. So, for BMC relay strategy, we have

$$\frac{E_b}{N_0} = \max\left\{\left(\frac{E_b}{N_0}\right)_1, \left(\frac{E_b}{N_0}\right)_2\right\} \quad (52)$$

where

$$10\log_{10}\left(\frac{E_b}{N_0}\right)_i \cong 10\log_{10}\left(\frac{E_b}{N_{0\,\min}}\right)_i + (10\log_{10}2)\frac{\mathsf{C}}{\mathcal{S}_{0i}}, \quad i=1,2. \quad (53)$$

$(E_b/N_0)_{\min}$ for BMC transmissions in the FDD mode was also studied in (15). For the case of $\lambda_{sr}, \lambda_{rd} \geq \lambda_{sd}$, we can see from (48) and (50) that $(E_b/N_0)_{\min}$ is minimized by $\theta = 1$. It can be verified that if we set $\theta = 1$ and minimize the maximum of (48) and (50) over γ, the resulting expression for $(E_b/N_0)_{\min}$ coincides with the corresponding one given in [15], namely

$$\left(\frac{E_b}{N_0}\right)_{\min} = \frac{\lambda_{sr}+\lambda_{rd}-\lambda_{sd}}{\lambda_{sr}\lambda_{rd}}\log_e 2. \quad (54)$$

The optimal values of θ, γ, and α that minimize (52) can be obtained using a three-dimensional (3-D) search with computational complexity $O(M^3)$, assuming M quantized values of each parameter. Alternatively, we can iteratively update these

$$\mathcal{S}_{02} = \frac{2}{\kappa_{sd}}\frac{\left(1+\frac{(1-\lambda)\lambda_{rd}}{\gamma\lambda_{sd}}\right)^2\alpha(1-\alpha)}{\alpha-2\theta\alpha+\theta^2+\frac{\alpha}{\kappa_{sd}}\left[\kappa_{rd}\left(\frac{(1-\gamma)\lambda_{rd}}{\gamma\lambda_{sd}}\right)^2+2(1-\theta)\frac{(1-\gamma)\lambda_{rd}}{\gamma\lambda_{sd}}\right]} \quad (51)$$

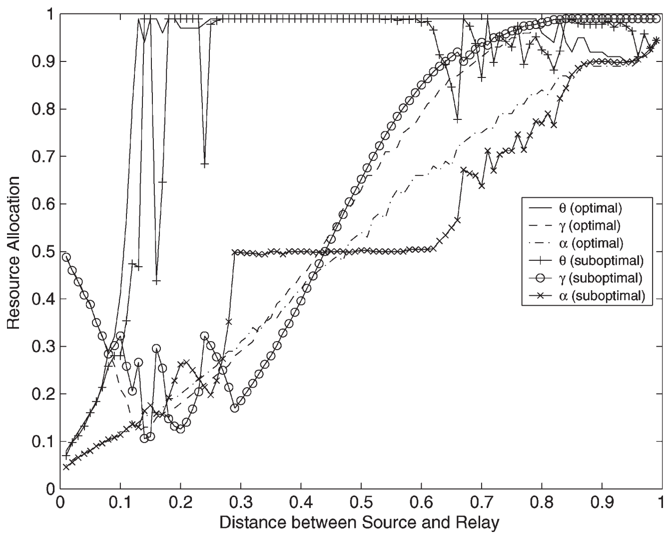

Fig. 2. Optimal and suboptimal resource allocation ($d = 2, C = 0.4$ bit/s/Hz).

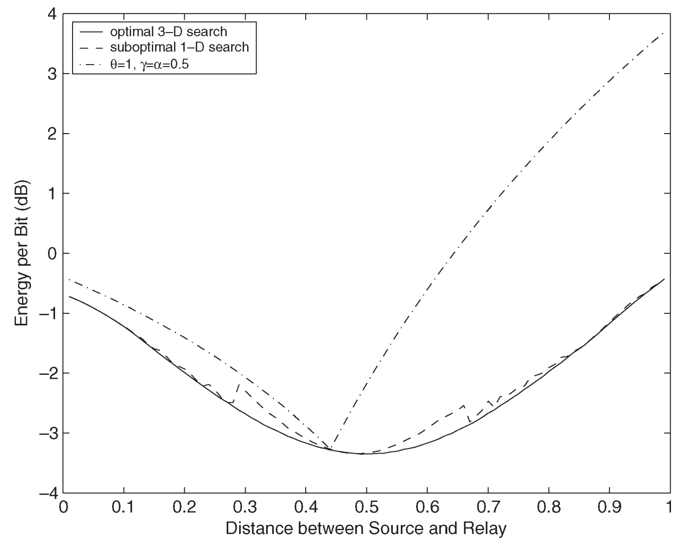

Fig. 3. Energy efficiency of various resource allocation schemes ($d = 2, C = 0.4$ bit/s/Hz).

three parameters sequentially using one-dimensional (1-D) searches. This iterative scheme is suboptimal but has lower computational complexity in the order of $O(M)$.

In Fig. 2, the values of θ, γ, and α obtained by optimal 3-D and suboptimal 1-D searches are given. We have assumed Rayleigh fading channels with path loss coefficient 2. The source–destination distance is 1, while the relay is placed on the line between them. The spectral efficiency has been set to $\mathsf{C} = 0.4$ bit/s/Hz. Fig. 2 suggests that $\theta = 1$ except when the source–relay distance is very small, meaning that in most cases, the source should only transmit in the first time slot. The duration allocated to the first time slot increases as the relay moves away from the source toward the destination node. This is intuitively reasonable, because as the source–relay distance increases, the relay–destination link improves relative to the source–relay link, and as a result we can accommodate higher data rates. Similar reasons can be used to explain why most power should be allocated to the source node when the relay is close to the destination.

An interesting case occurs when the source–relay distance becomes very small. In this case, the source–relay link is so reliable that the source can convey its information to the relay in a very short time using very low power, so both α and P_{s1} can be very small. In the second time slot, both relay and source transmit data to the destination. Since the link qualities of source–relay and relay–destination are very similar in this scenario, P_{s2} and P_r should be approximately equal, according to the theory of parallel Gaussian channel [20]. So, for small source–relay distances, the optimal θ and α should be small, while γ should approach 0.5.

From Fig. 2, we also verify that the parameters obtained by suboptimal 1-D searches are mostly quite close to their optimal values. The energy per bit of the optimal and suboptimal resource allocation is plotted in Fig. 3. For comparison, we also plot the energy per bit achieved by setting $\theta = 1$ and $\alpha = \gamma = 0.5$. We observe that the gap between optimal and suboptimal schemes is small, and both of these schemes achieve much better energy efficiency than the direct Tx benchmark.

E. Capacity Upper Bound

A capacity upper bound for the relay channel can be obtained by the max-flow-min-cut method, and is shown to be

$$C_{ub} = \min\{C_1, C_2\} \tag{55}$$

where

$$C_1 = \alpha E\left[\log\left(1 + \frac{P_{s1}(|h_{sr}|^2 + |h_{sd}|^2)}{\alpha N_0 W}\right)\right] + (1-\alpha)E\left[\log\left(1 + \frac{P_{s2}|h_{sd}|^2}{(1-\alpha)N_0 W}\right)\right] \tag{56}$$

and

$$C_2 = \alpha E\left[\log\left(1 + \frac{P_{s1}|h_{sd}|^2}{\alpha N_0 W}\right)\right] + (1-\alpha)E\left[\log\left(1 + \frac{P_{s2}|h_{sd}|^2}{(1-\alpha)N_0 W} + \frac{P_r|h_{rd}|^2}{(1-\alpha)N_0 W}\right)\right]. \tag{57}$$

Due to space limitations and because we use this capacity upper bound solely as a benchmark, we will not derive the approximation for the corresponding E_b/N_0, which is a lower bound for the energy required to transmit 1 bit. Instead, we will rely on simulation to generate the benchmark for comparison.

IV. Comparisons and Simulations

A. Relay With Unlimited Power

Consider the case where the source–relay channel is as good as or better than the source–destination channel. From (21), (28), and (45), we observe that among AnF, DnF, and BMC, AnF achieves the smallest $(E_b/N_0)_{\min}$, while BMC always has the same or better energy efficiency than DnF [this can be verified by comparing (45) with (28)]. To gain more insights on the relative energy consumption of various relay schemes,

TABLE I
ENERGY EFFICIENCY OF VARIOUS SCHEMES (UNLIMITED RELAY POWER)

Scheme	$\frac{E_b}{N_0}_{\min}$ (dB)	$\mathcal{S}_0$ (b/s/Hz/3dB)
Direct Tx	-1.59	1
AnF	$-1.59 - 3d - 10\log_{10}(1 + 2^{-d})$	decreases from $\frac{2}{3}$ to $\frac{1}{2}$ as d goes from 0 to ∞
DnF	$-1.59 - 3d$	$\frac{1}{2}$
BMC	$-1.59 - 3d$	1

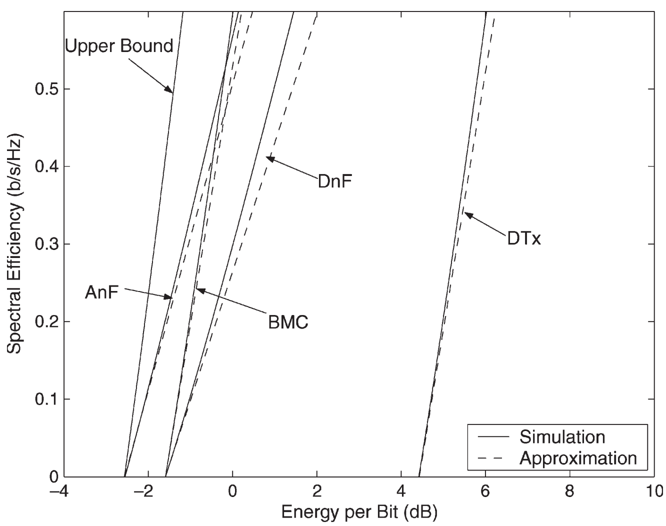

Fig. 4. Spectral efficiency versus energy per bit for various relay strategies ($d = 2$, relay has unlimited power).

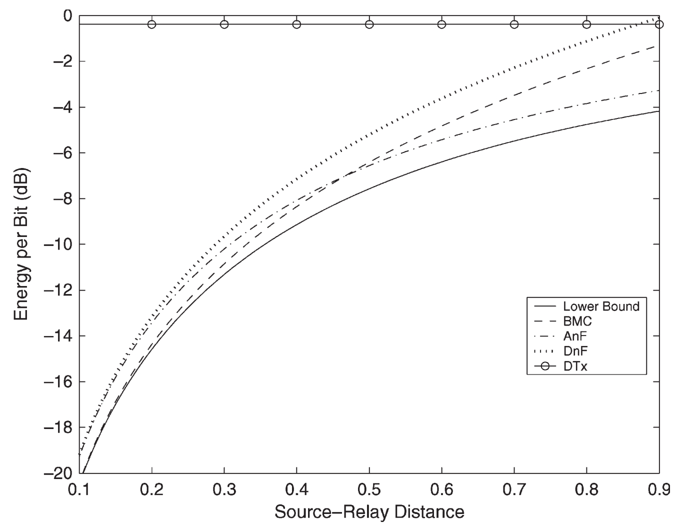

Fig. 5. Dependence of energy per bit on relay location ($d = 2, C = 0.4$ bit/s/Hz, relay has unlimited power).

let us examine the following numerical example. Assume that the channels between nodes experience Rayleigh fading, the source–destination distance is 1, and the relay has been placed at the middle point on the line between them. The path loss coefficient is denoted as d. Applying the analytical results of Section III to this paradigm, we have obtained the minimum energy per bit and the slope of spectral efficiency versus E_b/N_0 of different relay schemes, and we summarize our results in Table I.

Comparing these schemes, we can see that both DnF and BMC improve the minimum energy per bit by 3 dB over the direct Tx. The AnF scheme achieves the lowest $(E_b/N_0)_{\min}$, but the gap between AnF and the other two relay methods disappears as the path loss coefficient grows large. As to the slope $\mathcal{S}_0$, both direct Tx and BMC achieve 1 bit/s/Hz/3 dB, while for the DnF scheme, it is 1/2 bit/s/Hz/3 dB. The slope for AnF depends on d: as d goes from 0 to ∞, $\mathcal{S}_0$ decreases from 2/3 to 1/2. Summarizing, AnF and DnF achieve better energy efficiency than direct Tx when C is small, while BMC always outperforms direct Tx. At extremely low spectral efficiency, AnF is the most efficient scheme, while BMC is more favorable if $\mathsf{C} < (1 + 2^d)^2 \log_2(1 + 2^{-d})/(1 + 2^{2d})$.

To verify the conclusions we draw from the analysis, we perform Monte Carlo simulations. As stated before, the source, relay, and destination are placed on a straight line, with the relay being the middle point. The simulation results for $d = 2$ are plotted in Fig. 4. We observe that the simulation results fit the approximation given by our analysis very well. At very low spectral efficiency, AnF achieves an energy efficiency close to that obtained from the capacity upper bound; when the spectral efficiency exceeds 0.5 bit/Hz/s, however, BMC becomes most energy efficient among these three schemes.

Our analysis quantifies how the energy efficiency of relay transmissions depends on the relative link quality between source and relay, and between source and destination. In Fig. 5, we depict the dependence of various schemes' energy per bit on the distance between source and relay. Again, all three nodes are placed on a straight line. The path loss coefficient is $d = 2$ and the spectral efficiency is $\mathsf{C} = 0.4$ bit/s/Hz. We observe that BMC has the lowest E_b/N_0 when the source–relay link quality is good; otherwise, AnF outperforms other schemes.

B. *Relay With Limited Power*

Since for a relay with limited power the optimal resource allocation parameters usually depend on the target spectral efficiency, comparing the E_b/N_0 versus C slope of different relay schemes may not be very meaningful. For this reason, we focus on the zero SNR E_b/N_0 of various schemes. Again, we consider a Rayleigh fading channel with path loss coefficient d. The source–destination distance is 1, while the relay is placed on the line between source and destination with source–relay distance x. From the analytical results in Section III, the zero SNR E_b/N_0 for different schemes is as shown at the bottom of the next page.

We can see that DnF and BMC achieve the best $(E_b/N_0)_{\min}$. The gain of their energy efficiency over direct Tx increases with the path loss coefficient d. The AnF scheme does not improve the zero SNR E_b/N_0 over the direct Tx, which is a contrast to the case of unlimited power relay, where AnF

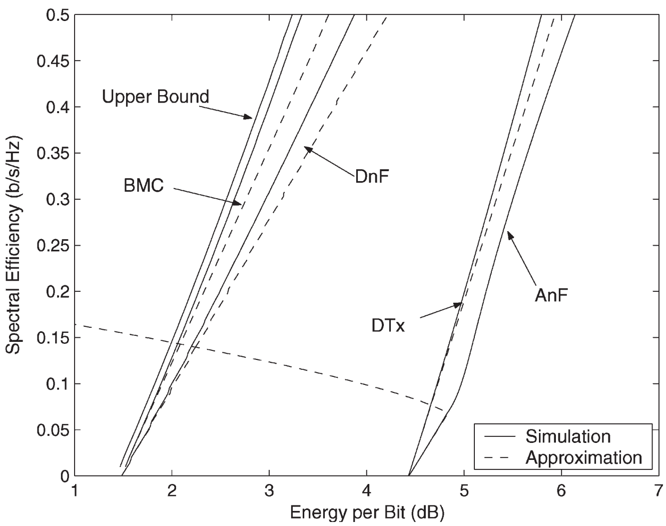

Fig. 6. Spectral efficiency versus energy per bit for various relay strategies ($d = 2$, relay has limited power).

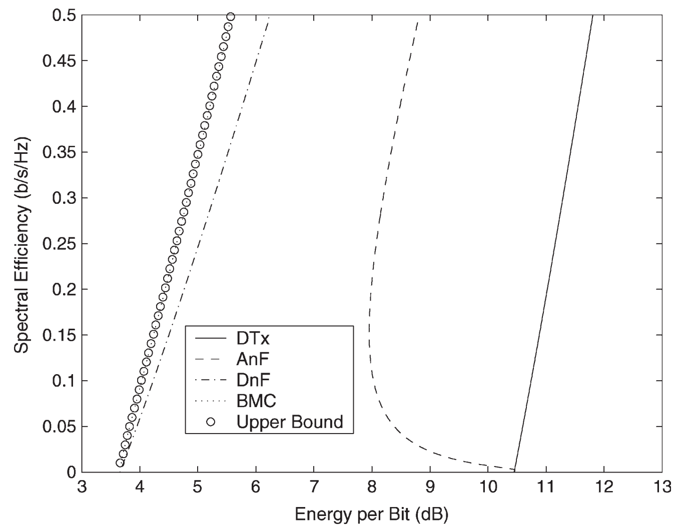

Fig. 7. Spectral efficiency versus energy per bit for various relay strategies ($d = 4$, relay has limited power).

achieves the best $(E_b/N_0)_{\min}$ among all schemes. To compare the energy efficiency of different schemes at nonzero spectral efficiency, numerical computation is needed in general. We are able to establish, however, that the energy efficiency of BMC is always better than or as good as that of DnF. This can be verified after setting $\theta = 1$ and $\alpha = 0.5$ in (48)–(51) and comparing them with (32) and (33).

In Figs. 6 and 7, we compare the energy efficiency of various transmission schemes for $d = 2$ and $d = 4$, respectively. Optimal resource allocation has been adopted for all relay schemes. We observe that BMC achieves the best energy efficiency in both scenarios, while DnF comes in a fairly close second. They both achieve energy efficiency close to the one corresponding to the capacity upper bound, and both markedly outperform the direct Tx and AnF transmission. Somewhat unexpectedly, in the $d = 2$ case, AnF's spectral efficiency for a given E_b/N_0 is even lower than that of the direct Tx. Our explanation is that the benefit of AnF relaying is not large enough to overcome the spectral efficiency loss due to the loss of source transmission time. When the path loss is higher ($d = 4$), AnF does bring significant gain over direct Tx. We also observe that E_b/N_0 initially decreases as C increases. This suggests that when using AnF, if the target data rate is very low and there is no peak power constraint, we can achieve improved energy efficiency by using a bursty transmission (BT), i.e., by transmitting during only a fraction of the time at a higher data rate.

The approximations involved in our analysis have also been quantified in Fig. 6. We verify that except for the AnF scheme,

---

$$
\begin{aligned}
\text{Direct Tx} \quad & 10\log_{10}\frac{\log_e 2}{\lambda_{sd}} = -1.59 \text{ dB} \\
\text{AnF} \quad & 10\log_{10}\frac{\log_e 2}{\gamma\lambda_{sd}} \text{ that achieves its minimum value} \\
& 10\log_{10}\frac{\log_e 2}{\lambda_{sd}} = -1.59 \text{ dB when } \gamma = 1 \\
\text{DnF} \quad & 10\log_{10}\left(\min_{\gamma}\max\left\{\frac{\log_e 2}{\gamma\lambda_{sd}+(1-\gamma)\lambda_{rd}}, \frac{\log_e 2}{\gamma\lambda_{sr}}\right\}\right) \\
& = 10\log_{10}\frac{\lambda_{sr}+\lambda_{rd}-\lambda_{sd}}{\lambda_{rd}\lambda_{sr}} \\
& = -1.59 \text{ dB} + 10\log_{10}\left[x^d + (1-x)^d - x^d(1-x)^d\right] \\
\text{BMC} \quad & 10\log_{10}\left(\min_{\gamma,\theta}\max\left\{\frac{\log_e 2}{\gamma\lambda_{sd}+(1-\gamma)\lambda_{rd}}, \frac{\log_e 2}{\theta\gamma\lambda_{sr}+(1-\theta)\gamma\lambda_{sd}}\right\}\right) \\
& = 10\log_{10}\frac{\lambda_{sr}+\lambda_{rd}-\lambda_{sd}}{\lambda_{rd}\lambda_{sr}} \text{ that is identical to that of DnF}
\end{aligned}
$$

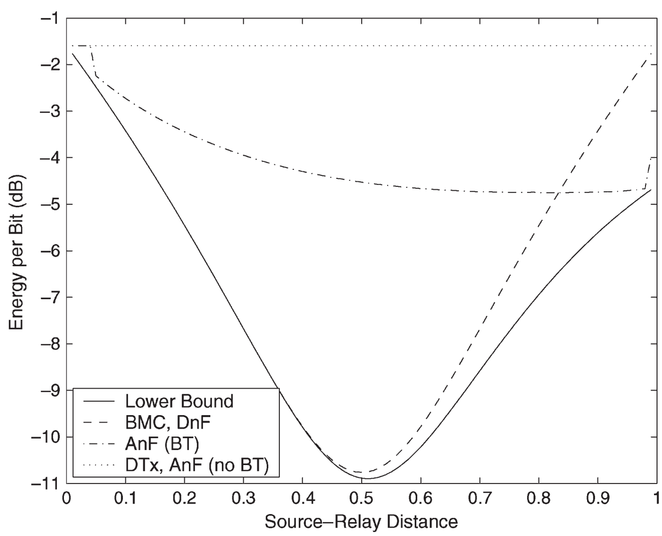

Fig. 8. Minimum energy per bit versus relay location ($d = 2$, relay has limited power, optimal resource allocation).

the simulations match with our analytical results very well. The approximation for AnF predicts the simulation results fairly well when C is very small, but is inadequate when C grows higher.

In Fig. 8, we examine how the energy per bit for various relay transmission methods depends on the source–relay distance at the wideband limit. We have fixed the source–destination distance as 1 and adopted optimal resource allocation values. We observe that BMC and DnF achieve the same $(E_b/N_0)_{\min}$, which almost coincides with the lower bound when the source–relay distance is less than 0.5; as the source–relay distance becomes larger, however, the gap between their energy efficiency and the lower bound increases. We also observe the benefit of using BT in AnF: without BT, AnF has the same energy efficiency as direct Tx at the wideband limit and is always outperformed by DnF and BMC; with BT, more than 4 dB of energy is saved and AnF achieves better energy efficiency than DnF and BMC when the source–relay distance is large.

V. Conclusion

In this paper, we studied the energy required for transmitting one information bit under three different relay schemes: AnF, DnF, and BMC. Two different scenarios one may encounter in practice have been considered. In the first scenario, the relay has access to a fixed power source, and hence theoretically unlimited power supply. For these applications, analysis and simulation indicate that: 1) AnF achieves the best energy per bit at the wideband limit; 2) BMC is always at least as energy efficient as DnF; and 3) AnF is most efficient when the spectral efficiency is low or when the relay–destination distance is small; otherwise, BMC offers the best energy efficiency. In the second scenario, the relay relies on battery with limited power, and hence its power consumption must also be taken into account. For this class of applications, we demonstrated that: 1) BMC always has the same or better energy efficiency than DnF; 2) the energy efficiency of AnF can be substantially improved using BT in the power-limited regime; and 3) when the relay–destination distance is small, AnF with BT offers the best energy efficiency; otherwise, BMC and DnF are more energy efficient.

In studying various relay methods, we have adopted the perspective of minimizing the power consumption for a prescribed data rate and bandwidth. We have shown that relay transmissions save the energy consumption by several decibels, which leads to a much longer battery life. In applications where bandwidth is at a premium, one may want to examine the impact of relay transmissions on bandwidth requirements given fixed rate and power [16]. From the analytical and numerical results in this paper, we can see that the gain of relay transmission over direct Tx is even more significant from this perspective: properly chosen relay strategies can often reduce the bandwidth requirement by several orders of magnitude.

While the potential gain of relay links is exciting, in practice many factors other than the transmit power must be considered in deciding when to relay and what scheme to choose. For example, we will need to investigate how different relay methods will impact the efficiency of linear amplification at the device layer. Energy consumption overhead caused by having to perform additional channel estimation, decoding, and encoding, and having to wake up a relay node possibly in a sleep mode must also be taken into account. These and many other challenging issues must be resolved before the potential of wireless networks can be realized.

References

[1] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 8–27, Aug. 2002.

[2] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 388–404, Mar. 2000.

[3] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[4] A. Høst-Madsen, "On the capacity of wireless relaying," in *Proc. IEEE Vehicular Technology Conf. (VTC) Fall*, Vancouver, BC, Canada, Sep. 24–28, 2002, vol. 3, pp. 1333–1337.

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity—Part 1: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[6] ——, "User cooperation diversity—Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[7] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[8] M. O. Hasna and M.-S. Alouini, , "Application of the harmonic mean statistics to the end-to-end performance of transmission systems with relays," in *Proc. IEEE Global Telecommunications (GLOBECOM) Conf.*, Taipei, Taiwan, R.O.C., Nov. 17–21, 2002, vol. 2, pp. 1310–1314.

[9] ——, "Optimal power allocation for relayed transmissions over Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Orlando, FL, Apr. 22–25, 2003, vol. 4, pp. 2461–2465.

[10] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for the wireless relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020–2040, Jun. 2005.

[11] T. Hunter and A. Nosratinia, "Cooperative diversity through coding," in *Proc. Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, Jun. 30–Jul. 5, 2002, p. 220.

[12] B. Zhao and M. C. Valenti, "Distributed turbo coded diversity for relay channel," *Electron. Lett.*, vol. 39, no. 10, pp. 786–787, May 2003.

[13] A. Y. Wang, S. Cho, C. G. Sodini, and A. P. Chandrakasan, "Energy efficient modulation and MAC for asymmetric microsensor systems," in *Proc. Int. Symp. Low Power Electronics and Design*, Huntington Beach, CA, Aug. 6–7, 2001, pp. 106–111.

[14] A. Ephremides, "Energy concerns in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 48–59, Aug. 2002.

[15] A. El Gamal and S. Zahedi, "Minimum energy communication over a relay channel," in *Proc. Int. Symp. Information Theory (ISIT)*, Yokohama, Japan, Jun. 29–Jul. 4, 2003, p. 344.

[16] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[17] X. Cai, Y. Yao, and G. B. Giannakis, "Achievable rates in low-power relay-links over fading channels," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 184–194, Jan. 2005. [Online]. Available: http://spincom.ece.umn.edu/publications.htm

[18] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: A new architecture for wireless communications," in *Proc. Information Communications (INFOCOM)*, Tel-Aviv, Israel, Mar. 26–30, 2000, vol. 3, pp. 1273–1282.

[19] C. M. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1160–1165, Nov. 1989.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

Yingwei Yao (M'03) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2002.

From 2002 to 2004, he was a Postdoctoral Researcher at the University of Minnesota, Minneapolis. Since August 2004, he has been an Assistant Professor at the University of Illinois at Chicago. His research interests are in the areas of communication theory, wireless networks, and statistical signal processing.

Xiaodong Cai (S'00–M'00–SM'05) received the Ph.D. degree in electrical engineering from the New Jersey Institute of Technology, Newark, in 2001.

From February 2001 to June 2001, he was a Member of the Technical Staff at Lucent Technologies, Murray Hill, NJ. From July 2001 to October 2001, he was a Senior System Engineer at Sony Technology Center, San Diego, CA. From November 2001 to July 2004, he was a Postdoctoral Research Associate at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. Since August 2004, he has been an Assistant Professor at the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL. His research interests lie in the areas of communication theory, signal processing, and networking. Current research focuses on cooperative communications and collaborative signal processing for wireless ad hoc and sensor networks.

Georgios B. Giannakis (S'84–M'86–SM'91–F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1981, and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been a Professor at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 220 journal papers, 380 conference papers, and two edited books. Current research focuses on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space-time coding, multi-carrier ultrawide band wireless communication systems, cross-layer designs, and sensor networks.

Dr. Giannakis is the (co)recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies (1992, 1998, 2000, 2001, 2003, and 2004). He also received the IEEE Signal Processing Society's Technical Achievement Award in 2000 and European Association for Signal, Speech and Image Processing (EURASIP) Technical Achievement Award in 2005. He has served as Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the Signal Processing Conference Board, as Member of the Signal Processing Publications Board, as Member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the Signal Processing for Communications Technical Committee, and as a Member of the IEEE Fellows Election Committee. He has also served as a Member of the IEEE-Signal Processing Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and the Steering Committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.