On the Optimum Number of Hops in Linear Wireless Networks

Marcin Sikora, J. Nicholas Laneman, Martin Haenggi, Daniel J. Costello, Jr., and Thomas Fuja
Department of Electrical Engineering, University of Notre Dame
275 Fitzpatrick Hall, Notre Dame, IN 46556
msikora@nd.edu

Abstract — **We consider a wireless communication system with a single source node, a single destination node, and multiple relay nodes placed equidistantly between them. We limit our analysis to the case of coded TDMA multihop transmission, i.e., the nodes do not cooperate and do not try to access the channel simultaneously. Given a global constraint on bandwidth, we determine the number of hops that achieves a desired end-to-end rate with the least total transmission power. Furthermore, we examine how the optimum number of hops changes when an end-to-end delay constraint is introduced using the sphere-packing bound and computer simulations. The analysis demonstrates that the optimum number of hops depends on the end-to-end rate and the path-loss exponent. Specifically, we show the existence of an asymptotic per-link spectral efficiency, which is the preferred spectral efficiency in TDMA multihop transmission.**

I. Introduction

Layered architectures for communication networks have led to significant progress and technological advances. For emerging wireless networks such as ad hoc, multihop cellular, and sensor networks, momentum is growing behind multihop routing at the network layer, distributed channel access at the link layer, and powerful channel coding at the physical layer. These advances have been studied largely in isolation, whereas this paper focuses on their interaction, especially in delay-constrained scenarios.

To illustrate the main concepts, we consider a linear wireless network of $N+1$ nodes. Transmission between the end nodes can occur in a single hop, or up to N hops. Proponents of multihop routing argue that more short hops are preferable to fewer long hops, because the minimum signal-to-noise ratio (SNR) along the route is larger for multihop. As indicated in [1] and [2] this observation does not take into account the important practical issues of resource allocation, end-to-end delay, error propagation, and interference induced by extra transmissions. Consider the following motivating observations and questions:

- Because wireless transceivers cannot both receive and transmit at the same time on the same frequencies, multi-hop requires excess bandwidth when compared to single-hop. Do the costs of this excess bandwidth outweigh the benefits of SNR gain due to multihop?
- In a delay-constrained environment, the accumulated delay incurred when coded packets are decoded and re-encoded at every hop can become unacceptable. For a given end-to-end delay, is it better to decode/re-encode weaker codes (each with small delay) over many short hops or to use stronger codes (each with large delay) over fewer long hops?

[1] This work was supported in part by NASA Grant NAG5-12792, NSF Grant CCR02-05310, and NSF Grant ECS03-29766.

To address these issues, we compare single hop and multi-hop transmission under bandwidth and delay constraints. First, we examine the impact of bandwidth constraints alone on channel capacity. Then we incorporate delay constraints by employing the sphere-packing bound. Finally, we illustrate the performance of multihop coding schemes based upon convolutional and turbo codes and compare them to theoretical performance predictions.

Our results indicate that the benefits of multihop are eroded by bandwidth constraint, especially for high spectral efficiency (a similar conclusion was reported independently in [3]), but the impact of a delay constraint is much less severe than anticipated. We further show the existence of an asymptotic per-link spectral efficiency, which is the preferred spectral efficiency in a time division multiple access (TDMA) multi-hop transmission. Choosing the number of hops for which the required per-link spectral efficiency is closest to this asymptotic value is an optimal strategy for most end-to-end rates. In the simple case of deciding between one-hop and two-hop transmission, one hop is generally preferred when the rate (in b/s/Hz) is larger than the path-loss exponent.

II. System Description

A System model

The communication system under consideration is illustrated in Figure 1. It consists of a source node **S** and a destination node **D** separated by a distance L, and $N-1$ intermediate relay nodes $\mathbf{F}_i$, $i = 1, ..., N-1$, placed equidistantly on a line from **S** to **D**. The objective of the system is the reliable delivery of bits generated at the source node at a rate of $1/T_b$ bits per second to the destination node using coded transmission. The resources available for this task comprise a band of radio frequencies allowing for a signaling rate of $1/T_s$ complex-valued symbols per second and a total transmit power P_T.

The channel is modeled to attenuate the transmitted signal and corrupt it with additive white Gaussian noise (AWGN) with a one-sided spectral density N_0. The attenuation depends on the distance l between the transmitter and the receiver (for neighboring nodes $l = L/N$) according to

$$P_R = P_T c l^{-\alpha}, \qquad (1)$$

where P_T is the transmitted signal power, P_R is the received signal power, α is the path loss exponent (typically taking values between 2 and 4), and c is a constant. This model usually holds only for distances l for which $cl^{-\alpha} \ll 1$. However, since the constant terms in (1) do not affect the analysis that follows in terms of relative performance, we will assume that $c = 1$ and $L = 1$ in order to simplify the notation.

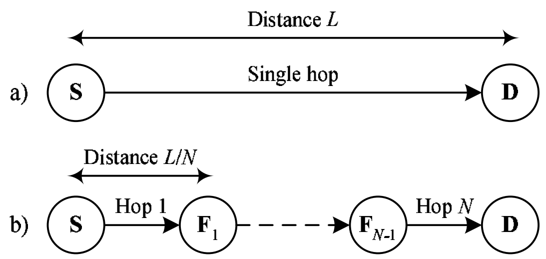

Figure 1: Single-hop and multi-hop communication systems.

B TDMA Access Mode

If all nodes in the system were perfectly synchronized and could coherently receive all transmitted signals, the system presented in Figure 1 could be interpreted as a multiple Gaussian relay channel. The data rates achievable in such a channel can always be increased by increasing the number of relays—at least as long as the distances L/N between neighboring nodes stay in a range in which the power law (1) holds [5]. In most practical ad hoc networks, however, neither the synchronization nor the complexity required to achieve such rates can be met. As a result attention must be directed to simpler, but more robust, operating modes.

In a TDMA multihop system, the end-to-end transmission is split into N partial transmissions, called hops, between neighboring nodes. At any point in time only one node is transmitting, and there is no interference at any receiver. Since, by assumption, the distances between neighboring nodes are equal, all hops are identical and should be assigned equal portions of the available resources (channel time and power). Hence, each of the N hops can utilize $(NT_s)^{-1}$ channel uses per second, each involving symbols with transmitted energy $E_s = P_T T_s$.

For single-hop transmission the received symbol energy is the same as the transmitted energy due to (1) and the assumptions we made about c and L. For multiple hops, however, the distance between neighboring nodes is L/N, and so the received energy is $E_s N^\alpha$. The SNR per hop can thus be expressed as $\mathrm{SNR} = E_s N^\alpha / N_0$.

C Performance Measures

In order to fairly compare the performance of systems involving different numbers of hops, the performance measures must be chosen carefully. To measure the bandwidth efficiency, we will use the bandwidth-normalized rate R defined as $R = T_s/T_b$ bits per channel use. Note that this refers to end-to-end, not node-to-node, transmission; the spectral efficiency at each hop equals NR due to the N-fold reduction in channel uses in TDMA operation. Power efficiency will be evaluated in terms of E_b^*/N_0, where the total energy per bit E_b^* is defined as the sum of the energies per bit spent over all N hops, i.e., the energy spent to deliver one bit from node **S** to node **D**. Hence, $E_b^* = NE_s/(NR) = E_s/R$.

The performance of a TDMA linear ad hoc network with N hops will be characterized by the highest achievable bandwidth-normalized rate R for a given E_b^*/N_0. Without delay constraint, the achievable rate is understood as the highest rate for which an arbitrarily small bit error rate (BER) can be obtained using forward error correction coding. When the end-to-end delay is limited, a rate is achievable if there exists a coding scheme with appropriate latency operating with a BER or block error rate (BLER) not exceeding some prescribed value.

III. Number of Hops Without Delay Constraints

When no delay constraint is imposed on the system, the highest achievable end-to-end transmission rate is the channel capacity. For a single hop, this can be simply expressed using Shannon's well-known formula [4]

$$R = \log_2\left(1 + \frac{E_s}{N_0}\right). \tag{2}$$

Switching from a single hop to N time-shared hops has the following consequences. The transmitted energy per symbol remains unchanged, but the received energy at each hop is N^α times higher due to reduced attenuation over a smaller distance. At the same time, each of the hops must accommodate the transmission of the same number of information bits in $1/N$-th of the channel uses available in the single hop mode—thus increasing the required per-hop spectral efficiency N-fold. The first of the two effects gives multihop an improvement over single hop transmission at low SNR, while the latter penalizes multihop at high SNR. Hence we can write

$$R = \frac{1}{N}\log_2\left(1 + \frac{E_s}{N_0}N^\alpha\right). \tag{3}$$

The curves corresponding to (2) and (3), with $E_b^* = E_s/R$, for several values of N are presented in Figure 2 for the cases $\alpha = 2$ and $\alpha = 4$. It can be observed that no single curve provides the best achievable rate for all E_b^*/N_0, but every curve dominates in some range of E_b^*/N_0 values.

It might be tempting to ask for the value of E_b^*/N_0 for which an N-hop system performs best. Note, however, that if we consider an analogous plot for a system in which the distance L is different from 1, the curves in Figure 2 will be shifted on the E_b^*/N_0 scale by $-10\alpha\log_{10} L$ dB. Hence, it makes more sense to determine the ranges of the end-to-end bandwidth-normalized rate R for which each curve dominates.

The value of R for which E_b^*/N_0 is the same for both N hops and $N+1$ hops will be referred to as the crossover rate and denoted R_N. By equating E_s/N_0 in (3) for N and $N+1$, R_N can be obtained as a solution to the polynomial equation

$$(N+1)^{-\alpha}(2^{R_N})^{N+1} - N^{-\alpha}(2^{R_N})^N + N^{-\alpha} - (N+1)^{-\alpha} = 0. \tag{4}$$

For $N = 1$ the above equation yields $R_1 = \log_2(2^\alpha - 1)$. Hence, as a simple rule, if the required end-to-end rate exceeds α, then the transmission from **S** to **D** should be performed in a single hop. For the calculation of crossover rates for higher N, we resort to numerical methods.

Figure 3 presents the optimal number of hops $N_{\mathrm{opt}}(R)$ for a given bandwidth-normalized end-to-end rate R, as well as the corresponding per-hop spectral efficiencies $RN_{\mathrm{opt}}(R)$ at each hop. The latter plots are particularly interesting, since they suggest the existence, for any value of α, of an asymptotic per-hop spectral efficiency

$$S(\alpha) = \lim_{R\to 0} RN_{\mathrm{opt}}(R) \tag{5}$$

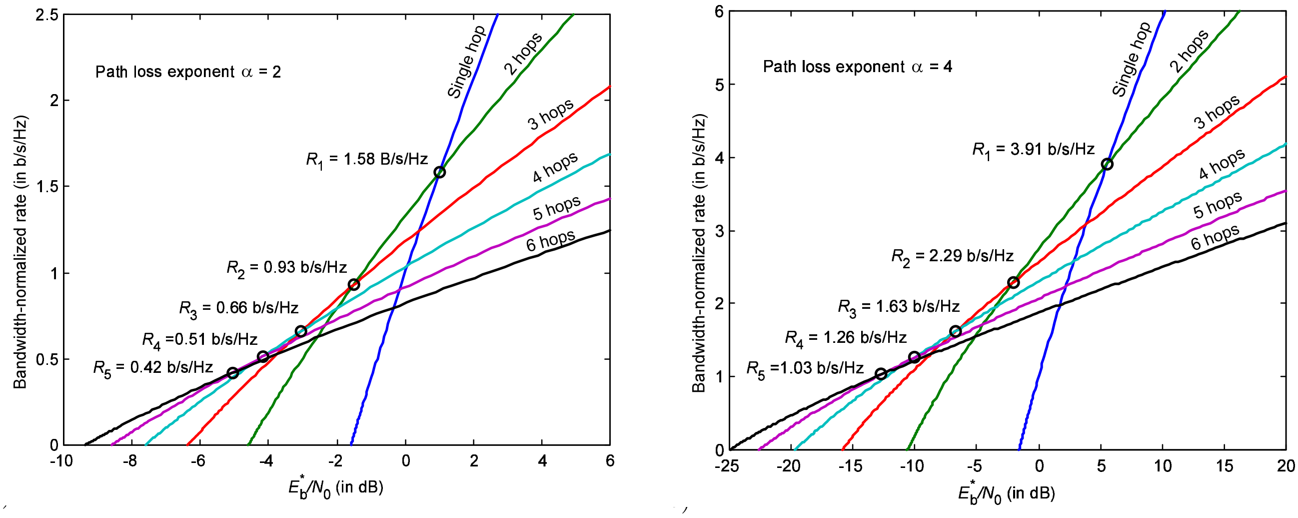

Figure 2: End-to-end bandwidth-normalized rates achievable with zero to five intermediate nodes for a) $\alpha = 2$ and b) $\alpha = 4$.

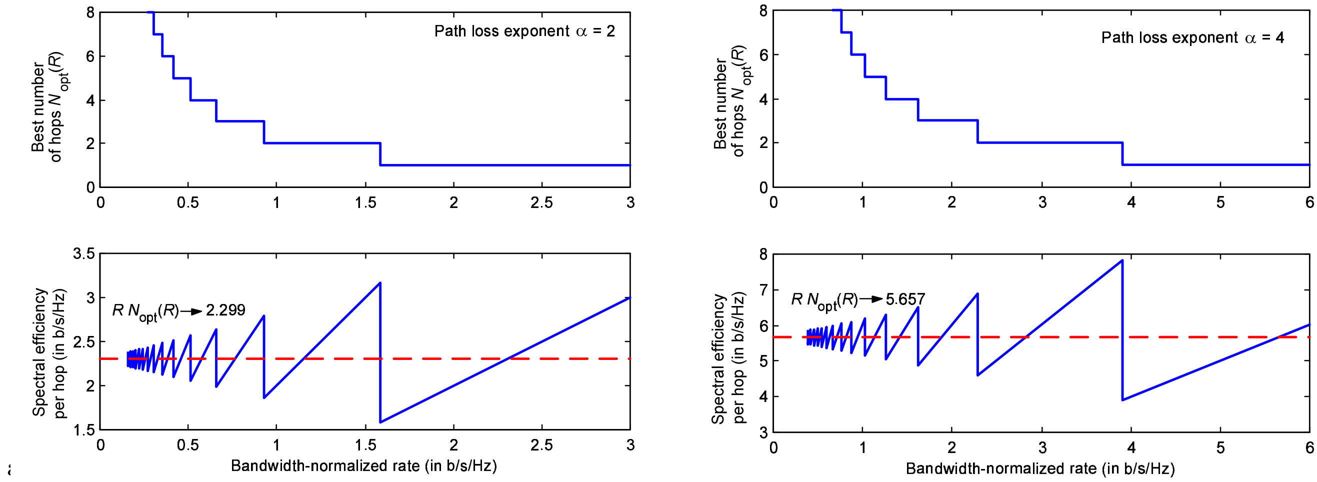

Figure 3: Optimal numbers of hops $N_{opt}(R)$ and the corresponding per-hop spectral efficiencies for a) $\alpha = 2$ and b) $\alpha = 4$.

preferred by the system. Moreover, an accurate prediction of the optimal number of hops can be obtained from

$$N_{\rm opt}(R) \approx [S/R]\,, \tag{6}$$

where by $[\cdot]$ we denote the nearest positive integer. However, we were unable to obtain from (4) a closed form expression for $S(\alpha)$. The values obtained numerically are plotted in Figure 4 for a range of path loss exponents.

IV. Number of Hops with a Delay Constraint

The predictions about the preferred number of hops made in the previous section were based on the assumption that the block lengths used by channel codes can be arbitrarily large. In many applications, however, there is a strict limit on the tolerable transmission delay, or end-to-end latency. There are several sources of latency in a communication system:

- Waiting for the data source to emit enough bits to form a block of a desired length (for channel coding);
- Processing delay caused by encoding the information bits for transmission;
- Transmission and reception of the whole encoded message;
- Processing delay caused by decoding.

If the communication system involves multiple hops, the latter three elements are repeated several times, increasing overall latency. To compensate for this, shorter block lengths must be used—at a cost of reduced error-correcting capabilities at each link.

In this paper we will consider only the delay caused by the need to receive the whole block of symbols before decoding can proceed. We also assume that the symbol duration T_s is much larger than the actual propagation delay. We denote the largest tolerable end-to-end delay as D, which corresponds to $n_s = D/T_s$ channel uses. In what follows we assume that by choosing $n = n_s/N$ as the block length (in term of coded symbols) in an N-hop system, the delay constraint is satisfied. Since this reduction in block length by the factor $1/N$

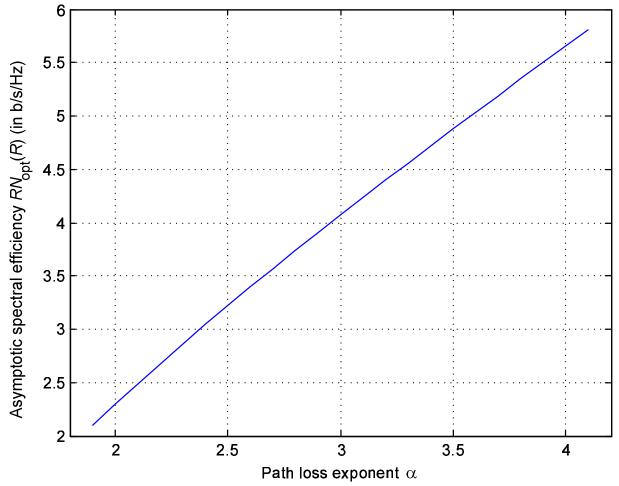

Figure 4: Asymptotic per-hop spectral efficiencies S for a range of path loss exponents.

is accompanied by an N-fold increase in code rate due to the TDMA mode of operation, we can code for an N-hop system by squeezing a fixed number Rn_s of information bits into n_s/N channel symbols.

Error-correcting codes with a limited block length cannot achieve arbitrarily low error rates. Hence, for a rate R to be achievable in a multi-hop system, a coding scheme must exist for which the end-to-end block error rate (BLER) falls below some tolerable value P_e. Since a block error at any of the N hops will result in a block error at the destination, except for the unlikely event of several block errors canceling each other out, we can instead require that the BLER at each hop does not exceed P_e/N. Hence, in a system with a delay constraint, increasing the number of hops is penalized by requiring coding schemes that operate at a higher rate, with a smaller block length, and achieving a lower BLER, while at the same time enjoying an SNR increase by a factor N^α.

The sphere-packing bound introduced by Shannon [6] is a useful tool that relates the BLER to the code rate, the SNR, and the block length for the real-valued AWGN channel. Following the notation in [7], the probability of a block error on the real-valued AWGN channel has a lower bound

$$P_e \geq Q_n(\theta, A) \tag{7}$$

depending on the signal amplitude parameter A, an angle θ, and the block length n in real-valued channel symbols. The angle θ is the half-angle of the n-dimensional cone encompassing a fractional solid angle 2^{-rn} and relates to the block length n and code rate r in bits per real-valued symbol according to

$$\Omega_n(\theta) = 2^{-rn}, \tag{8}$$

so that Ω_n must be inverted in order to obtain θ. Since the functions $Q_n(\theta, A)$ and $\Omega_n(\theta)$ appearing in (7) and (8) do not have closed-form expressions, we will use their asymptotic forms, which are accurate for $n > 20$ [7]:

$$\Omega_n(\theta) \approx \frac{\sin^{n-1}\theta}{\sqrt{2\pi n}\cos\theta} \tag{9}$$

Table 1: Simulated coding schemes.

Spectral efficiency (b/s/Hz)	Convolutional codes (decoding delay 100 bits)	Turbo codes (decoding delay 1000 bits)
0.5	QPSK + CC R=1/4, memory 10, [8]	QPSK + Turbo code R=1/4, [9]
1	QPSK + CC R=1/2, memory 10, [8]	QPSK + Turbo code R=1/2, [9]
2	8PSK + TCM R=2/3, memory 7, [8]	16QAM + Turbo code R=1/2, [9]
4	32CROSS + TCM R=4/5, memory 6, [8]	32CROSS + Turbo TCM R=4/5, [10]

and

$$Q_n(\theta, A) \approx \frac{\left[G\sin\theta e^{-\left(A^2 - AG\cos\theta\right)/2}\right]^n}{\sqrt{n\pi}\sqrt{1+G}\sin\theta\left[AG\sin^2\theta - \cos\theta\right]}, \tag{10}$$

where

$$G(\theta, A) = \frac{1}{2}\left(A\cos\theta + \sqrt{A^2\cos^2\theta + 4}\right). \tag{11}$$

The variables used in the above formulas relate to the variables introduced earlier as follows: $A^2 = 2E_sN^\alpha/N_0$, $n = 2n_s/N$, and $2r = R/N$, with the factor of 2 in the last two formulas accounting for the fact that we are using a complex-valued AWGN channel. We will write $Q(n_s/N, R/N, E_sN^\alpha/N_0)$ to denote $Q_n(\theta, A)$ combined with an inverse of (9) and with the above substitutions.

We computed the crossover points R_N by numerically searching for the solutions to the following set of equations:

$$\left\{\begin{array}{ccc} \frac{P_e}{N} & = & Q\left(\frac{n_s}{N}, RN, \frac{E_s}{N_0}N^\alpha\right) \\ \frac{P_e}{N+1} & = & Q\left(\frac{n_s}{N+1}, R(N+1), \frac{E_s}{N_0}(N+1)^\alpha\right) \end{array}\right. \tag{12}$$

with $P_e = 10^{-4}$. The results of this search are shown in Figure 5 for $\alpha = 2$ and $\alpha = 4$. The plots indicate that the crossover rates remain quite stable for most practical values of the delay constraint D. This means that the simpler predictions about R_N based on the capacity formulas in the previous section remain fairly accurate.

V. Simulation Results

The sphere-packing bound allows us to estimate the achievable performance of channel codes under limited transmission power and block length constraints. In a practical system one more constraint is present: code complexity. Therefore we repeated the process of finding the first crossover rate R_1 using the simulated BLER vs. E_b^*/N_0 performance obtained for two common families of codes: convolutional codes and turbo codes. Each family includes codes with increasing spectral efficiencies and decreasing block lengths. The parameters of the simulated codes are listed in Table 1.

The E_b^*/N_0 values needed to achieve a BLER of 10^{-4} have been established for each code and plotted vs. the spectral efficiency in Figure 6—once assuming they were used in a single hop system and once in a two-hop system (with $\alpha = 2$). The crossover rates obtained from these curves are also included in Figure 5a at their corresponding block lengths. These rates fall somewhat below the values predicted by the sphere-packing bound. Although the method we used in determining

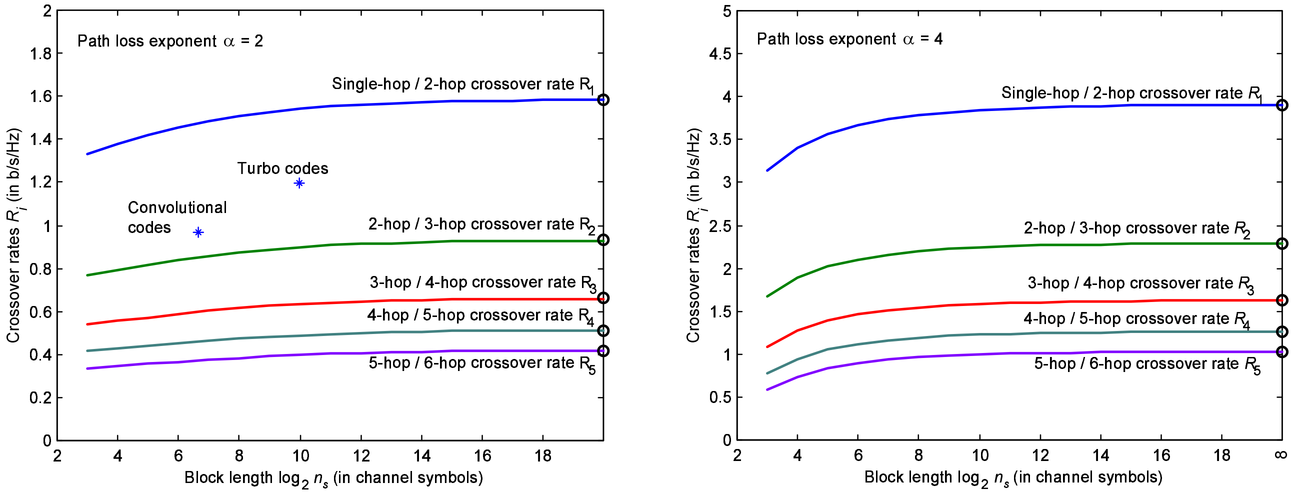

Figure 5: Crossover rates for delay-constrained transmission for a) $\alpha = 2$ and b) $\alpha = 4$.

the crossover rates is very sensitive to the choice of representative codes, this result suggests that in practical scenarios there are additional incentives to using a smaller number of hops that cannot be assessed by just comparing the capacities.

VI. Conclusion

For wireless multihop networks, such as ad hoc, multihop cellular, and sensor networks, a fundamental question is whether it is advanta-geous to route over many short hops or over a smaller number of longer hops. The benefits of short-hop routing include SNR gain and the reduction in interference. This paper addresses the first point and shows that, depending on the path loss exponent, the SNR gain may be offset by the required increase in spectral efficiency. In particular, for the one-dimensional scenario studied here, single-hop routing outperforms two-hop routing for bandwidth-normalized rates larger than the value of the path loss exponent. For delay-constrained transmission, this rate threshold for single hop to be preferable is somewhat lower.

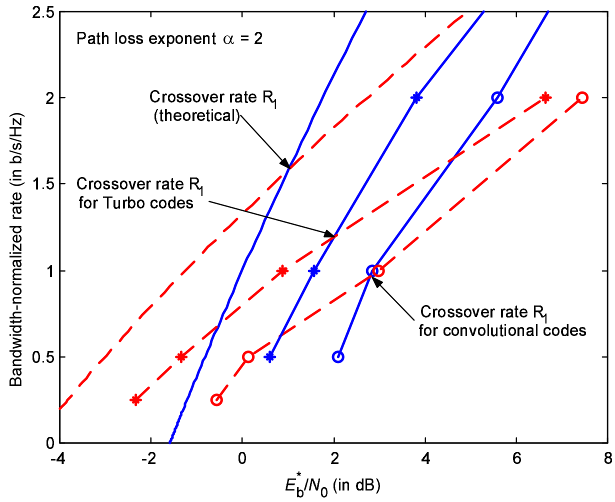

Figure 6: Crossover point between single-hop (solid lines) and two-hop (dashed lines) systems for practical coding schemes ($\alpha = 2$).

References

[1] M. Haenggi, "Twelve reasons not to route over many short hops," in *Proc. IEEE Vehicular Technology Conference (VTC'04)*, Los Angeles, CA, Sep. 2004.

[2] M. Haenggi, "On routing in random Rayleigh fading networks," accepted to *IEEE Transactions on Wireless Communications*, 2004. Available at `http://www.nd.edu/~mhaenggi/routing.pdf`.

[3] Z. Dawy and P. Leelapornchai, "Optimal Number of Relay Nodes in Wireless Ad Hoc Networks with Non-Cooperative Accessing Schemes," in *Proc. Int. Symp. on Inform. Theory ISITA'02*, Oct. 2002.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[5] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 748–767, May 2004.

[6] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.

[7] S. Dolinar, D. Divsalar, and F. Pollara, "Code Performance as a Function of Block Size," TMO Progress Report, May 1998.

[8] S. Lin and D.J. Costello, Jr., *Error Control Coding, 2nd Edition*, Prentice Hall, New Jersey, 2004.

[9] 3GPP2 C.S0002-B, Physical layer standard for cdma2000 spread spectrum systems, Release B, ver. 1.0, April 2002.

[10] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 206–218, Feb. 1998.