On the Throughput Enhancement of the Downstream Channel in Cellular Radio Networks Through Multihop Relaying

Jaeweon Cho, *Member, IEEE,* and Zygmunt J. Haas, *Senior Member, IEEE*

Abstract—In this paper, we study the effect of multihop relaying on the throughput of the downstream channel in cellular networks. In particular, we compare the throughput of the multihop system with that of the conventional cellular system, demonstrating the achievable throughput improvement by the multihop relaying. We also propose a hybrid control strategy for the multihop relaying, in which we advocate the use of both, the direct transmission and the multihop relaying. Our study shows that most of the throughput gain can be obtained with the use of a two- and three-hop relaying scheme. Substantial throughput improvement could be additionally obtained by operating the concurrent relaying transmission in conjunction with the nonconcurrent transmission. We also argue here that the multihop relaying technology can be utilized for mitigating unfairness in quality-of-service (QoS), which comes about due to the location-dependent signal quality. Our results show that the multihop system can provide more even QoS over the cell area. The multihop cellular network architecture can also be utilized as a self-configuring network mechanism that efficiently accommodates variability of traffic distribution. We have studied the throughput improvement for the uniform, as well as for the nonuniform traffic distribution, and we conclude that the use of multihop relaying in cellular networks would be relatively robust to changes in the actual traffic distribution.

Index Terms—Ad hoc network, cellular network, code-division multiple access (CDMA), concurrent transmission, downstream channel, fairness, multihop relaying, self-configuring network, throughput.

I. INTRODUCTION

MULTIHOP cellular networks have been proposed as an extension to the conventional single-hop cellular network by combining the fixed cellular infrastructure with the multihop relaying technology that is usually used in ad hoc

J. Cho was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with the Telecommunication R&D Center, Global Standards Strategy Team, Samsung Electronics Co., Ltd., Suwon 442-600, Korea (e-mail: jaeweon.cho@ieee.org).

Z. J. Haas is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: haas@ece.cornell.edu).

networks. Due to the potential of the multihop relaying to enhance coverage, capacity and flexibility, the multihop cellular networks have been attracting considerable attention. This approach of augmenting cellular communication with multihop relaying was also used in the standardization effort to include the multihop relaying into the third-generation (3G) mobile communication systems [1].

The primary advantage of the multihop relaying comes from the reduction in the overall path loss between a base station (BS) and a mobile station (MS) [2]. However, the penalty for employing multihop relaying is in the need for additional radio channels. Another benefit of the multihop relaying is the path diversity gain that can be achieved by selecting the most favorable multihop path in the shadowed environment. This diversity gain can increase with the number of MSs, as then the number of potentially relaying candidates increases and the possibility of finding a relay with lower path loss increases as well. In addition, system capacity can additionally increase by allowing concurrency among the multihop transmissions. However, such concurrency also increases the interference. So, the overall effect is not immediately clear.

As we saw above, the performance of the multihop cellular networks is governed by various tradeoffs. Thus, to benefit from the multihop relaying, the various tradeoffs should be comprehensively studied. However, the exploration of such tradeoffs in the literature is very limited. In particular, the analysis of the tradeoff caused by the concurrent transmissions between the interference and the channel reuse efficiency is of vital importance. Toumpis and Goldsmith [3] showed that the concurrent transmission can enhance the system capacity of the multihop cellular networks. However, their results were obtained for a single cell system and for just two cases of network topology, i.e., a linear topology and a single realization of a random topology. Hence, those results are insufficient to demonstrate, in general, the concurrency tradeoff. Moreover, several studies reported that it is not easy to enhance the capacity of code-division multiple-access (CDMA) systems by the use of the multihop relaying [5]–[9]. This is mainly due to the interference increase resulting from the concurrent transmissions. Such interference might be the most significant factor limiting the network capacity. Hence, the impact of the concurrent transmission should be carefully investigated.

The multihop relaying technology can provide a significant flexibility in the design and the operation of the cellular network. In the multihop cellular networks, MS can choose to utilize the multihop relaying instead of the single-hop direct transmission.

Such a hybrid operation can be exploited for various purposes; one of which is to mitigate the unfairness in the quality of service (QoS) among the users. In cellular networks, there is a tradeoff problem between system throughput and QoS fairness [4]. Since the received signal quality depends on the user location, it is not easy to provide an even QoS over the whole cell service area and to maximize the system throughput at the same time. On the other hand, the use of multihop relaying, instead of a direct link, can improve the QoS of the users with poor direct link who are located near the cell boundary or in a deep shadowed region. Therefore, the fairness, as well as system throughput in the cellular network can be improved through the use of the multihop relaying.

Another application exploiting the flexibility of the multihop relaying technology is to mitigate the inefficiency due to the temporal changes in traffic demand in cellular networks. To optimize the performance of a cellular system, finding the optimum positions of the cell sites is a crucial problem in interference-limited systems, such as the CDMA type systems. However, due to the ever-changing traffic demand patterns, optimal placement of cell site is a difficult problem. Even if the traffic distribution could be estimated, it would still be difficult to optimally plan the radio network, as the fixed cell sites cannot be relocated whenever the traffic distribution changes. Hence, there is a need for a self-configuring network, which would be capable of automatically coping with the changes in traffic distribution. In the multihop cellular network, the alternative selection of the multihop path can allow flexible design of the cell site, which is particularly important in the case of nonuniform traffic distribution. Thus, the multihop cellular network architecture can be utilized as a self-configuring network mechanism that can efficiently accommodate the spatial and temporal variability of traffic patterns.

Although the multihop relaying technology has been proposed as one of the key technologies for the self-configuring cellular networks [10], [11], there have been only few numerical results clarifying how the self-configuring feature achieved through the multihop relaying can improve the system capacity for nonuniformly distributed traffic case. Although Wu *et al.*[12] evaluated the capacity of the multihop system with nonuniform traffic, such that when the traffic between adjacent cells is unbalanced, that paper focused only on the channel borrowing between adjacent cells through multihop relaying. Moreover, some features of the multihop relaying, such as path-loss reduction and path diversity, were not considered in their analysis. Hence, their results may not be able to fully explain the behavior of the self-configuring capability of multihop relaying itself.

In this paper, we examine the effect of multihop relaying, in particular, focusing on the impact of the concurrent transmissions on the system capacity. Our study, in addition to the investigation of the fairness performance in the multihop systems, also compares the throughput of the multihop system with that of the conventional system, for both, the nonuniform, as well as the uniform traffic distribution. This paper focuses on the best-effort (delay-tolerant) type of service in the downstream direction, i.e., from the BS to the target MS. The nature of data traffic for most services on the Internet is asymmetric; for instance, recently standardized systems, which evolved

from the 3G systems, such as the 1x evolution data optimized (1xEV-DO) system [14] and the high-speed downlink packet access (HSDPA) system [15], aim at supporting high data rate packet services on the downlink. For this reason, efficient utilization of resource on the downlink is becoming more and more important.

The remainder of this paper is organized as follows. In Section II, we describe our system model. The *throughput gain*s for various types of the multihop relaying are derived in Section III. The numerical results of the performance of the multihop relaying are presented in Section IV. The impact of the concurrency on the throughput is demonstrated in Section V. In Section VI, we present some examples of applications that could benefit from the flexibility of the multihop relaying. Finally, we summarize the work and draw conclusions in Section VII.

II. System Model

Let us consider a cellular system where a number of BS-s are placed according to a hexagonal grid pattern. Every MS in the system is assumed to be capable of multihop relaying. We also assume a fully loaded system and an infinite buffer at the network nodes. The upstream and the downstream channels are assumed to work independently. The downstream transmission may involve multiple relaying MSs, in which case packets are relayed between neighboring MSs on the same downstream path through the use of time-division duplex (TDD), as in [13]. The channels assigned to each cell share the same frequency spectrum by CDMA. Time-division multiplexing (TDM) is selected for multiplexing downstream channels at the BS, i.e., each BS provides one downstream channel to one target MS at a time.[1]

In our system, as in the 1xEV-DO and the HSDPA systems, rate control is utilized based on the channel condition. Note that for the best effort type data services, rate control is preferable to power control, because in the rate control scheme a high signal-to-interference ratio (SIR) available over a large portion of the cell area can be exploited to provide higher data rate. In our model, the transmission power P is fixed, and the data rate assigned to each link R is determined by the SIR. *Shannon* capacity formula is then used in the rate control model to evaluate the network capacity [3][2]

$$R = W \log_2(1 + \text{SIR}) \tag{1}$$

where W is the channel bandwidth and the SIR at the receiver of the transmission from node i is given by

$$\text{SIR} = \frac{L_i P_i}{\sum_{j \in \mathcal{J}, j \neq i} L_j P_j} \tag{2}$$

where the subscript j denotes the transmitting node (BS or MS) and $\mathcal{J}$ is the set of transmitting nodes at a given time. In the

---

[1]In multirate CDMA systems, packet traffic channels can be assigned in a code-division multiplexing (CDM) or TDM fashion; e,g, TDM is used in the 1xEV-DO system, and TDM and/or CDM in HSDPA system.

[2]In this paper, the interference limited scenario is considered, and so the impact of the thermal noise on the received signal quality is neglected.
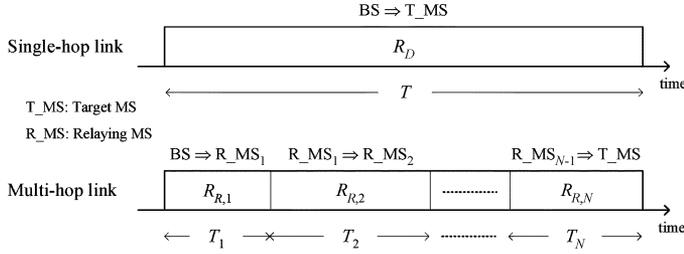
Fig. 1. Transmission data rates in the cases of a single-hop transmission and a multihop transmission.

equation, L_i denotes the link gain, so that at distance d_i from the node i, it can be represented as

$$L_i = A \cdot d_i^{-\gamma} \cdot 10^{\frac{\zeta_i}{10}} \tag{3}$$

where A is a constant, γ is the path loss exponent, and ζ_i is the shadowing random variable.

III. THROUGHPUT GAIN

The downstream throughput can be calculated as the *effective data rate* on the direct link or on the multihop path between a BS and a target MS. We define the *decision parameter* and the *throughput gain* based on this effective data rate, and we present the performance metrics for three different relaying schemes. The first one is the *simple relaying* scheme, in which only one downstream channel for a target MS is provided in each cell at any time. The others are two types of relaying with concurrent transmission: one allows concurrency among the different hops along the path to one target MS, a scheme which we term *type A concurrency*; the second type allows concurrency among the hops on the different downstream paths to multiple target MSs, a scheme which is referred to here as *type B concurrency*. Each concurrent transmission is assumed to employ different spreading code.

A. Simple Relaying Without Concurrency

Let us assume that a time period for data transmission T is assigned to the target MS. In the conventional system, the effective data rate is identical to the data rate of the direct single-hop link, R_D. Let U_R be the effective data rate of the relaying path comprised of multihop links. Each hop is assigned the downstream channel for a short time duration, as shown in Fig. 1. We assume that the relaying MSs act as forwarder only during the time T. Hence, the transmitted amount of data for each time slot should be identical for all the nodes on the relaying path,[3] i.e.,

$$T_1 R_{R,1} = T_2 R_{R,2} = \cdots = T_N R_{R,N} \tag{4}$$

where $R_{R,n}$ denotes the data rate on the nth hop of the relaying path $(n = 1, \cdots, N)$, and N is the number of hops on the downstream path. Since $T = \sum_{n=1}^{N} T_n$ and

---

[3]In this paper, the time slot means a *logical time slot* which has a variable length.

$T_n = T_l R_{R,l} / R_{R,n} (, 1 \leq l \leq N)$, each time slot length can be determined by the given period T and the data rates as

$$T_l = \frac{T}{\sum_{n=1}^{N} \frac{R_{R,l}}{R_{R,n}}}. \tag{5}$$

Then, U_R can be given by the number of effective transmitted bits per T as

$$U_R = \frac{T_l R_{R,l}}{T} = \frac{1}{\sum_{n=1}^{N} \frac{1}{R_{R,n}}}. \tag{6}$$

Now, we define the *decision parameter D* as the ratio of the two effective data rates, the effective data rate in the multihop system and the effective data rate in the conventional system

$$D \equiv \frac{U_R}{R_D} = \frac{1}{\sum_{n=1}^{N} \frac{R_D}{R_{R,n}}}. \tag{7}$$

Note that our network model is a hybrid of the conventional system and the multihop system; either the direct link or the multihop path can be selected, whichever case provides higher throughput. That is, the multihop path is selected if $D > 1$, otherwise, the direct link is selected. Then, the *throughput gain G* in a network that uses such a hybrid scheme can be defined as the ratio of throughput of the most favorable case (the direct link or the multihop path) to that of the direct link. Therefore

$$G \equiv \begin{cases} D, & \text{if } D > 1 \\ 1, & \text{otherwise} \end{cases}. \tag{8}$$

B. Type A Concurrency

In the case of the *type A concurrency*, while only one target MS can be provided with a downstream path at any time, the concurrent transmissions on this path are allowed on the different hops. Note that even though the concurrency occurs, all the transmitted data during the time T is for one target MS only. Hence, the transmitted amount of data in each time slot is limited by the hop whose data rate is the minimum among the multiple concurrent hops, and it should be equal to the transmitted amount of data in the other time slots. So as in (4)
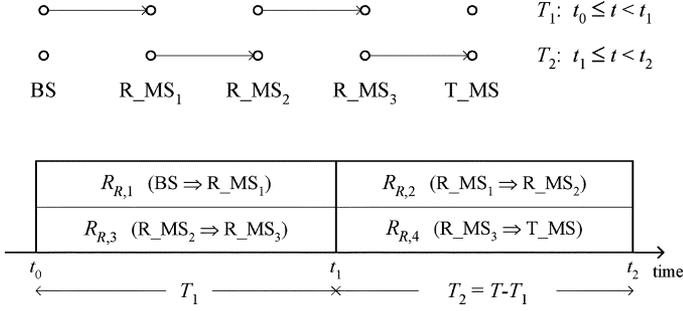
$$T_1 \cdot \min_{n \in \mathcal{T}_1} \{R_{R,n}\} = \cdots = T_M \cdot \min_{n \in \mathcal{T}_M} \{R_{R,n}\} \tag{9}$$

where $\mathcal{T}_m$ is the set of the hops sharing the mth time slot $(m = 1, \cdots, M)$, and $M (2 \leq M \leq N)$ is the number of time slots during T. Then, U_R can be represented by

$$U_R = \frac{T_m \cdot \min_{n \in \mathcal{T}_m} \{R_{R,n}\}}{T}$$
$$= \frac{1}{\frac{1}{\min_{n \in \mathcal{T}_1} \{R_{R,n}\}} + \cdots + \frac{1}{\min_{n \in \mathcal{T}_M} \{R_{R,n}\}}}. \tag{10}$$

An example of the *type A concurrency* for a four-hop relaying path is presented in Fig. 2. In this example, the time period T is divided into two slots, i.e., $M = 2$, and in each time slot

Fig. 2. Example of the *type A concurrency* for a four-hop path.

the transmission concurrency occurs on the two different hops which are separated by one-hop distance. Hence, (9) becomes

$$T_1 \cdot \min\{R_{R,1}, R_{R,3}\} = T_2 \cdot \min\{R_{R,2}, R_{R,4}\}. \tag{11}$$

Then, U_R is given by

$$U_R = \frac{1}{\frac{1}{\min\{R_{R,1}, R_{R,3}\}} + \frac{1}{\min\{R_{R,2}, R_{R,4}\}}}. \tag{12}$$

Note that U_R for the *simple relaying* is [see (6)]

$$U_R = \frac{1}{\frac{1}{R_{R,1}} + \frac{1}{R_{R,2}} + \frac{1}{R_{R,3}} + \frac{1}{R_{R,4}}}. \tag{13}$$

In these two last expressions for U_R, one can easily identify the increase in time slot reuse by the *type A concurrency*: the denominator of (13) is the sum of four terms but that of (12) is the sum of two terms only, resulting in an increase in U_R. However, one should also consider the interference caused by the *concurrency*. The SIR of the BS-to-R_MS$_1$ link may be dominated by the inference from R_MS$_2$, so that the data rate $R_{R,1}$ could be lower than in the *simple relaying*. The received signal at R_MS$_3$ is also interfered by the transmitting power of BS. The numerical results of the tradeoff between the improvement in the time slot reuse and the increase in the interference will be presented in Section IV.
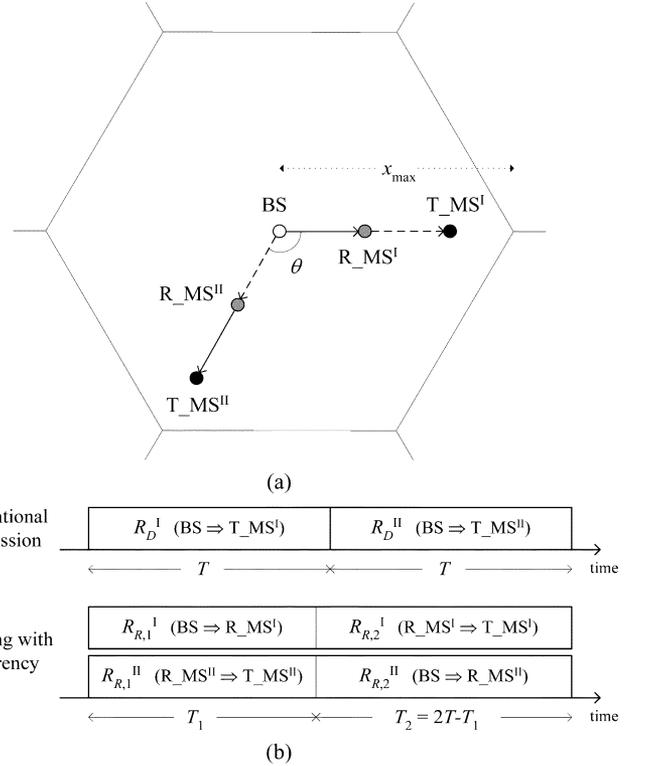
We use the same criterion for the multihop path selection as applied to the case of the *simple relaying*, and the same definition for the metrics, i.e., $D \equiv U_R/R_D$ and $G \equiv \max\{D, 1\}$.

C. Type B Concurrency

With the *type B concurrency*, the hops of different downstream paths to multiple target MSs can share the same time slots. For simplicity, we consider two downstream paths (to two target MSs) with two-hop each.

The multihop relaying with the *type B concurrency* is illustrated in Fig. 3, where the superscripts I and II denote target MSs. In this example, the BS and R_MS^{II} can transmit concurrently to R_MS^{I} and T_MS^{II}, respectively. Likewise, concurrent transmissions in the same time slot can occur on the BS-to-R_MS^{II} and the R_MS^{I}-to-T_MS^{I} links. Such transmission concurrency can improve the reuse of the time slot, but can also lead to an increase in the interference. Note that the interference increases as the angular separation between the two paths θ decreases. Hence, the SIR and the data rate of each link increase with θ.

In the example of Fig. 3, since the relaying MSs act as forwarder only, the transmitted amounts of data during T_1

Fig. 3. Illustration of the concurrent transmission in the case of the *type B concurrency*. (a) Topology. (b) Transmission timing.

and T_2 should be the same; i.e., $T_1 R_{R,1}^{I} = T_2 R_{R,2}^{I}$ and $T_1 R_{R,1}^{II} = T_2 R_{R,2}^{II}$. However, $R_{R,1}^{I}/R_{R,2}^{I}$ is generally not equal to $R_{R,1}^{II}/R_{R,2}^{II}$. Thus, the two equations cannot be satisfied simultaneously with the same time division of the two paths. Rather, we should assign different time division to each of the two paths, so that for both paths the transmitted amounts on the first hop is equal to that on the second hop.

We illustrate the transmission timing with the different time division for each path in Fig. 4. In the figure, T^{I} and T^{II} denote the transmission periods of the BS-to-R_MS^{I} link on the path from BS to T_MS^{I} and of the R_MS^{II}-to-T_MS^{II} link on the path from BS to T_MS^{II}, respectively. According to the length of T^{I} and T^{II}, the time division is divided into two cases. We first consider the case of $T^{I} \leq T^{II}$. From Fig. 4(a), we can formulate the following set of equations:

$$R_{R,1}^{I} T^{I} = (T^{II} - T^{I}) R_{R,2}^{I} + (2T - T^{II}) R_{R,3}^{I} \tag{14}$$

$$R_{R,1}^{II} T^{I} + (T^{II} - T^{I}) R_{R,2}^{II} = (2T - T^{II}) R_{R,3}^{II}. \tag{15}$$

Solving the two equations for T^{I} and T^{II}, and we obtain

$$T^{I} = \frac{\left(R_{R,3}^{I} R_{R,2}^{II} + R_{R,2}^{I} R_{R,3}^{II}\right) 2T}{C_a} \tag{16}$$

$$T^{II} = \frac{\left[\left(R_{R,2}^{II} - R_{R,1}^{II}\right) R_{R,3}^{I} + \left(R_{R,2}^{I} + R_{R,1}^{I}\right) R_{R,3}^{II}\right] 2T}{C_a} \tag{17}$$

where

$$C_a \equiv \left(R_{R,1}^{I} + R_{R,2}^{I}\right) \left(R_{R,2}^{II} + R_{R,3}^{II}\right)$$
$$- \left(R_{R,3}^{I} - R_{R,2}^{I}\right) \left(R_{R,1}^{II} - R_{R,2}^{II}\right). \tag{18}$$

Fig. 4. Transmission timing in the case of the *type B concurrency*. (a) Case: $T^{\mathrm{I}} \leq T^{\mathrm{II}}$. (b) Case: $T^{\mathrm{I}} > T^{\mathrm{II}}$.

Hence, the effective data rate during the $2T$ time period for the multihop paths is given by

$$
\begin{aligned}
U_R^{\mathrm{I}} &= \frac{T^{\mathrm{I}} R_{R,1}^{\mathrm{I}}}{2T} \\
&= \frac{\left(R_{R,3}^{\mathrm{I}} R_{R,2}^{\mathrm{II}} + R_{R,2}^{\mathrm{I}} R_{R,3}^{\mathrm{II}}\right) R_{R,1}^{\mathrm{I}}}{C_a}
\end{aligned} \tag{19}
$$

$$
\begin{aligned}
U_R^{\mathrm{II}} &= \frac{(2T - T^{\mathrm{II}}) R_{R,3}^{\mathrm{II}}}{2T} \\
&= \frac{\left(R_{R,1}^{\mathrm{I}} R_{R,2}^{\mathrm{II}} + R_{R,2}^{\mathrm{I}} R_{R,1}^{\mathrm{II}}\right) R_{R,3}^{\mathrm{II}}}{C_a}.
\end{aligned} \tag{20}
$$

Now, let us proceed with the case of $T^{\mathrm{I}} > T^{\mathrm{II}}$, which is illustrated in Fig. 4(b). Note that during the period from t_1 to t_3, the BS transmits to two relaying MS in the TDM fashion. (Note that in this case, we can apply CDM with the BS's power evenly assigned to the two channels and the BS transmitting to both MSs simultaneously. However, we continue to employ TDM here as well, to be consistent with our model.) From Fig. 4(b), we can formulate the following set of equations

$$
R_{R,1}^{\mathrm{I}} T^{\mathrm{II}} + (T^{\mathrm{I}} - T^{\mathrm{II}}) R_{R,2}^{\mathrm{I}} = (2T - 2T^{\mathrm{I}} + T^{\mathrm{II}}) R_{R,3}^{\mathrm{I}} \tag{21}
$$

$$
R_{R,1}^{\mathrm{II}} T^{\mathrm{II}} = (T^{\mathrm{I}} - T^{\mathrm{II}}) R_{R,2}^{\mathrm{II}} + (2T - 2T^{\mathrm{I}} + T^{\mathrm{II}}) R_{R,3}^{II}. \tag{22}
$$

Solving the two equations for T^{I} and T^{II}, and we obtain

$$
T^{\mathrm{I}} = \frac{\left[\left(R_{R,1}^{\mathrm{II}} + R_{R,2}^{\mathrm{II}}\right) R_{R,3}^{\mathrm{I}} - \left(R_{R,1}^{\mathrm{I}} - R_{R,2}^{\mathrm{I}}\right) R_{R,3}^{\mathrm{II}}\right] 2T}{C_b} \tag{23}
$$

$$
T^{\mathrm{II}} = \frac{\left(R_{R,3}^{\mathrm{I}} R_{R,2}^{\mathrm{II}} + R_{R,2}^{\mathrm{I}} R_{R,3}^{\mathrm{II}}\right) 2T}{C_b} \tag{24}
$$

where

$$
\begin{aligned}
C_b \equiv{} & \left(R_{R,1}^{\mathrm{II}} + R_{R,3}^{\mathrm{II}}\right) R_{R,2}^{\mathrm{I}} + \left(R_{R,1}^{\mathrm{I}} + R_{R,3}^{\mathrm{I}}\right) R_{R,2}^{\mathrm{II}} \\
& + 2\left(R_{R,3}^{\mathrm{I}} R_{R,1}^{\mathrm{II}} - R_{R,1}^{\mathrm{I}} R_{R,3}^{\mathrm{II}}\right).
\end{aligned} \tag{25}
$$

Hence, the effective data rates are

$$
\begin{aligned}
U_R^{\mathrm{I}} &= \frac{(2T - 2T^{\mathrm{I}} + T^{\mathrm{II}}) R_{R,3}^{\mathrm{I}}}{2T} \\
&= \frac{\left(R_{R,1}^{\mathrm{I}} R_{R,2}^{\mathrm{II}} + R_{R,2}^{\mathrm{I}} R_{R,1}^{\mathrm{II}}\right) R_{R,3}^{\mathrm{I}}}{C_b}
\end{aligned} \tag{26}
$$

$$
\begin{aligned}
U_R^{\mathrm{II}} &= \frac{T^{\mathrm{II}} R_{R,1}^{\mathrm{II}}}{2T} \\
&= \frac{\left(R_{R,3}^{\mathrm{I}} R_{R,2}^{\mathrm{II}} + R_{R,2}^{\mathrm{I}} R_{R,3}^{\mathrm{II}}\right) R_{R,1}^{\mathrm{II}}}{C_b}.
\end{aligned} \tag{27}
$$

The choice between the above two cases should be made according to the locations of MSs. From (16), (17), (23), and (24), we can see that the ratios of $T^{\mathrm{I}}/T^{\mathrm{II}}$ for the two cases are

$$
\frac{T^{\mathrm{I}}}{T^{\mathrm{II}}} = \frac{1}{1 + \frac{R_{R,1}^{\mathrm{I}} R_{R,3}^{\mathrm{II}} - R_{R,3}^{\mathrm{I}} R_{R,1}^{\mathrm{II}}}{R_{R,3}^{\mathrm{I}} R_{R,2}^{\mathrm{II}} + R_{R,2}^{\mathrm{I}} R_{R,3}^{\mathrm{II}}}}. \tag{28}
$$

Hence, the inequality $T^{\mathrm{I}} \lessgtr T^{\mathrm{II}}$ becomes identical to $R_{R,1}^{\mathrm{I}} R_{R,3}^{\mathrm{II}} \gtrless R_{R,3}^{\mathrm{I}} R_{R,1}^{\mathrm{II}}$. Note that these four data rates are affected not by the applied time division but by the locations of MSs. Therefore, by evaluating the inequality $R_{R,1}^{\mathrm{I}} R_{R,3}^{\mathrm{II}} \gtrless R_{R,3}^{\mathrm{I}} R_{R,1}^{\mathrm{II}}$, we can determine the ratio of T^{I} to T^{II} and so define which case of the time division should be applied.

The effective data rates for the direct link are given by $U_D^{\mathrm{I}} \equiv R_D^{\mathrm{I}}/2$ and $U_D^{\mathrm{II}} \equiv R_D^{\mathrm{II}}/2$. Unlike in the case of the *simple relaying* and in the case of the *type A concurrency*, the effective data rate should be calculated as half of the data rate of the link. This is for a fair comparison, because for direct transmission in the case of the *type B concurrency* each downstream path exploits relatively only half duration of $2T$, i.e., T, while for multihop relaying with the *type B concurrency* both downstream paths share the whole period of $2T$.

For the case of the *type B concurrency*, we can apply two different criteria for the selection of the multihop path. For each criterion, we state a condition; if the criterion is used and the corresponding condition is met, the multihop paths are selected for both target MSs. Otherwise, the direct transmission is used for both target MSs. The conditions that correspond to each criterion are given below.

- *Criterion 1*: If $U_R^{\mathrm{I}} + U_R^{\mathrm{II}} > U_D^{\mathrm{I}} + U_D^{\mathrm{II}}$, select the multihop paths.
- *Criterion 2*: If $U_R^{\mathrm{I}} > U_D^{\mathrm{I}}$ and $U_R^{\mathrm{II}} > U_D^{\mathrm{II}}$, select the multihop paths.

For both criterions, *throughput gain* can be represented by

$$
G \equiv \begin{cases} \frac{(D^{\mathrm{I}} + D^{\mathrm{II}})}{2}, & \text{if the multihop paths are selected} \\ 1, & \text{otherwise} \end{cases} \tag{29}
$$

where $D^{\mathrm{I}} \equiv U_R^{\mathrm{I}}/U_D^{\mathrm{I}}$ and $D^{\mathrm{II}} \equiv U_R^{\mathrm{II}}/U_D^{\mathrm{II}}$.

IV. PERFORMANCE ANALYSIS OF SIMPLE RELAYING

In this section, we investigate the effect of the *simple relaying* on the downstream throughput for various cases. The system model for the performance evaluation is comprised of 19 cells. The hexagonal cell radius $x_{\max}$ is 500 m. Since the thermal noise is assumed to be negligible in (2), we need only to define

the ratio of the transmission power of a BS to the power of an MS, which is set to 4 in this work. The path loss exponent γ is assumed to be 4.

The shadowing in decibels, ζ follows the *Gaussian* distribution with zero-mean and standard deviation $\sigma = 10$ dB. Both autocorrelation and cross correlation are applied to the shadowing variables of the link from a BS to a MS [16]. The autocorrelation coefficient is defined by an exponential function [17]

$$\rho_A(\Delta \underline{p}) \equiv \frac{E\left[\zeta_k(\underline{p})\zeta_k(\underline{p} + \Delta \underline{p})\right]}{\sigma^2} = \exp\left(-\frac{|\Delta \underline{p}|}{d_{\text{cor}}} \ln 2\right) \tag{30}$$

where the subscript k denotes the kth BS, $\underline{p}$ is a position vector of the MS, and d_{cor} is the decorrelation length, which is set to 20 m. For simplicity, the cross-correlation coefficient is assumed to be a constant, as in [18]

$$\rho_C \equiv \frac{E\left[\zeta_k(\underline{p})\zeta_q(\underline{p})\right]}{\sigma^2} = 0.5 \qquad (k \neq q). \tag{31}$$

We assume that the shadowing variables between the MS-to-MS links or between the MS-to-MS link and the BS-to-MS link are independent one from another.

A *Monte Carlo* computer simulation has been developed to evaluate the performance of the multihop relaying scheme. The shadowing variables are generated with the autocorrelation and the cross correlation, and then we calculate SIR, the data rate, and the metrics. For simplicity, the multihop relaying is employed only in the center cell. In the other cells, the direct transmission mode alone is used. The simulation collects statistical data from the center cell.

A. Tradeoff Between the Path-Loss Reduction and the Resource Demand

We first investigate the tradeoff between the reduction in the path loss and the increase in the number of required channels that are used for relaying. As the number of hops increases, the overall path loss is reduced, but also the available transmission period per hop is now shorter. Thus, unless concurrent transmissions are allowed, the effective transmission period per downstream path decreases with the number of hops. We study this tradeoff, while varying the location of the target MS. The target MS is located at the given position $(x, y = 0)$ in *Cartesian* coordinates with respect to the center cell site. The number of relaying MSs is set to $N - 1$, and they are linearly placed between the BS and the target MS, so that the ith relay MS is located at $(i \cdot x/N, 0)$. We first consider the case without shadowing and without concurrency.

Fig. 5 shows the decision parameter D for various locations of the target MS. We observe first that D increases with the distance from the cell site. This is because of the nonlinearity of the applied rate-control model, i.e., the *Shannon* curve. The received SIR on each hop increases with the number of hops due to the path-loss reduction. However, for the *Shannon* curve, the data rate increment with SIR is larger in the low-SIR region than in the high-SIR region. Therefore, in the vicinity of the cell site, the reduction in the effective transmission period due to the multihop relaying is more dominant than the SIR improvement caused by the path-loss reduction. Hence, D decreases as the target MS is closer to the cell site, and the *throughput gain* $G = \max\{D, 1\}$ is dependent on the location of the target MS, i.e.,
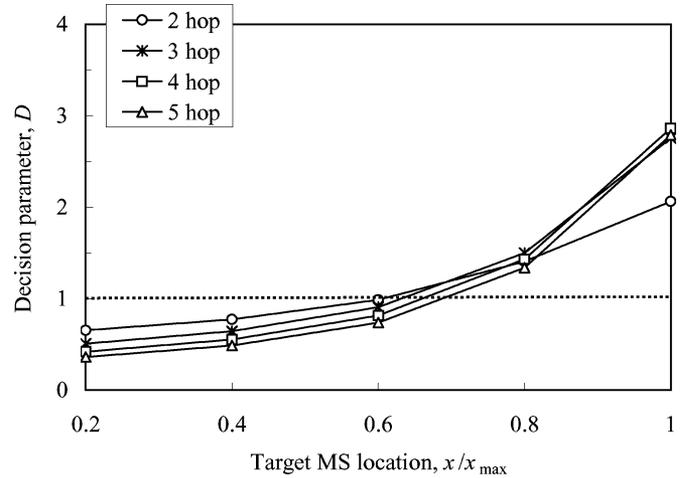

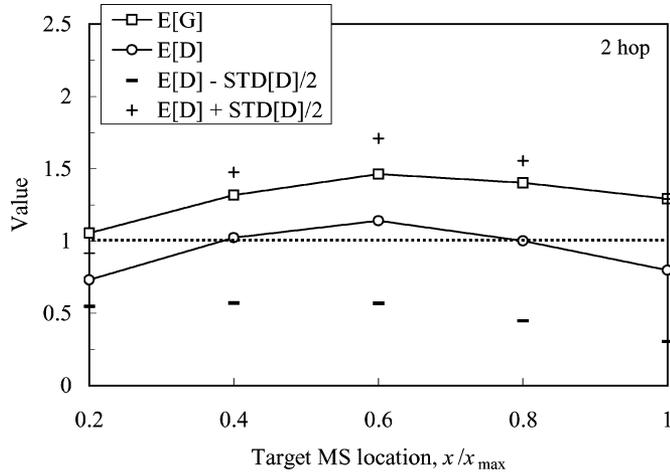Fig. 5. *Decision parameter* for linear topology with no shadowing.

the closer the mobile is placed to the cell boundary, the larger the throughput improvement is. Another observation from the figure is that the *throughput gain* is not strongly affected by the number of hops on the downstream paths. This is due to the tradeoff between the path-loss reduction and the decrease in the effective transmission period.
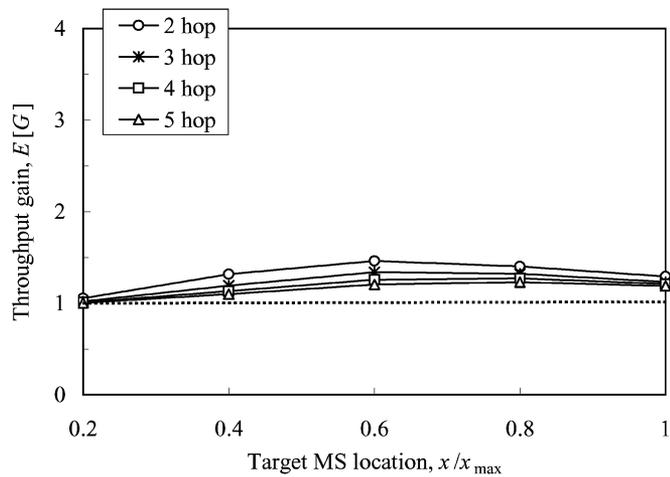
B. Effect of Shadowing

The results in Fig. 5 change when shadowing is included in the propagation model. The cellular network inherently implements a macrodiversity scheme, i.e., the BS with the stronger signal is selected by the MS. When a MS powers on, it first searches for the most favorable link to a BS (the one with the smallest path loss) by monitoring the received power level of the broadcasting control channel. The effect of such *best BS selection* in the multihop system can be found in Fig. 6(a), which shows the results for D and G in the shadowed environment. Note that D and G are now random variables due to the stochastic nature of the shadowing process. Thus, we consider their expected values and standard deviations, i.e., $E[\cdot]$ and $STD[\cdot]$. We can see that, unlike the results in Fig. 5, $E[D]$ now decreases at the cell boundary. *The best BS selection* yields a macrodiversity gain, so that the path loss of the direct link at the cell boundary decreases, compared with the nonshadowing case. Consequently, the throughput ratio of the multihop path to the direct link is reduced.[4] Fig. 6(a) also shows that the value of $(E[D] - STD[D]/2)$ is below 1 for all the MS locations. This demonstrates the fact that the direct link can, under some circumstances, be better than the multihop path. This fact justifies the need for a hybrid system that is able to select either the direct link or the multihop path.

Fig. 6(b) shows the effect of the number of hops in the shadowed environment. We can see that $E[G]$ slightly decreases as the number of hops increases for all the choices of the MS locations, including the cell boundary. There are two reasons for this behavior. First, there is the effect of *the best BS selection*, as explained above. The second reason is the effect of *the worst*

---

[4]In our system scenario, the macrodiversity gain achieved by *the best BS selection* can be exploited only for the direct link. For simple routing, we assumed that the multihop path is provided via only the relaying MSs located in the same cell that the target MS has selected.
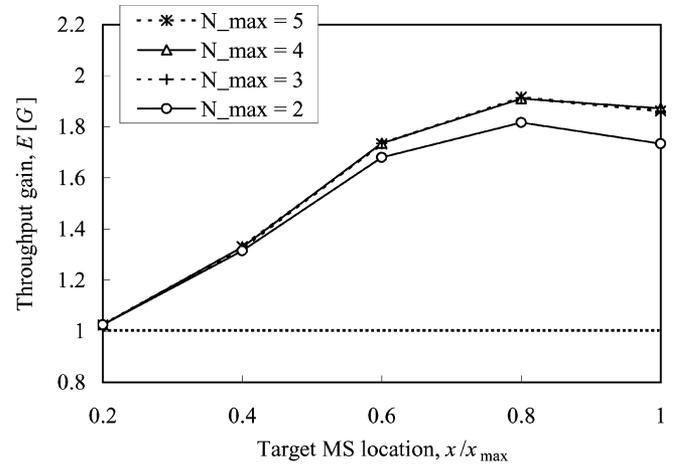
Fig. 6.   Effect of shadowing in linear topology. (a) Number of hops $= 2$. (b) With various numbers of hops.

Fig. 7.   Path diversity gain with the multihop relaying. (a) *Throughput gain.* (b) Probability of selecting each N-hop path.
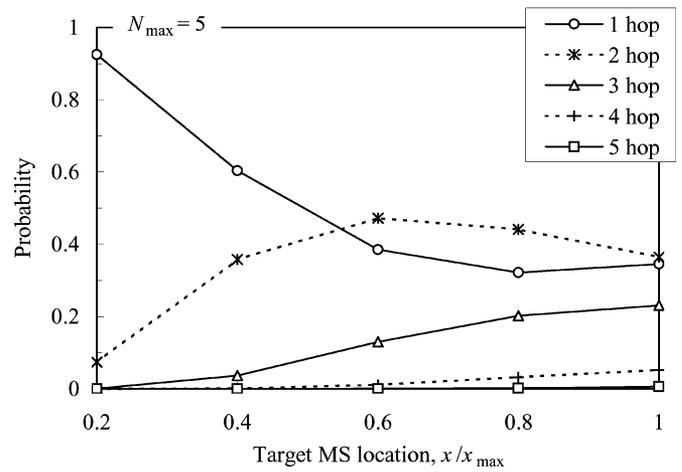
hop of the multihop path; we refer to the hop having the minimum data rate as *the worst hop*. Even if every hop in a multihop path has the same length, the data rate of each hop may be different because of the differences in the value of shadowing. The overall throughput of the multihop path is limited by the data rate of *the worst hop*. Since statistically the chances of the data rate of *the worst hop* of a path to be lower increases with the number of hops, the throughput of the multihop path degrades as well. Consequently, $E[G]$ decreases with the number of hops.

C. Path Diversity

Although we have shown that the *throughput gain* of multihop relaying is reduced in the shadowed environment, we need to remember that this result was derived for the linear topology with preselected relaying MSs. In a general topology, multihop relaying nodes benefit from the possibility of selecting the best relaying MS out of all the candidates and, thus, securing the most favorable multihop path, i.e., *path diversity gain*. We now assume that a number of MSs which are capable of relaying are randomly distributed over the cell area. We then find the best path, one with the maximum throughput, subject to the limitation on the maximum number of hops in the path, $N_{\max}$.

Through the use of a *Monte Carlo* simulation, we collected the statistical data of the performance metrics.

Fig. 7(a) shows $E[G]$ while varying the number of the maximum allowable hops on the path. These results were calculated for the case of 20 MSs in the center cell, i.e., 19 candidate relaying MSs are uniformly distributed throughout the cell. We can see that $E[G]$ significantly increases as the target MS is closer to the cell boundary. Since the further away the MS is from the cell site, the larger is on the average the number of favorable relaying MSs that it can see (recall that a mobile close to the cell site has higher chances of using the direct link). Thus, more diversity gain is obtained when the MS is located further away from the cell site.

We also observe that no additional *throughput gain* can be achieved when the number of hops is over three; in most cases, the throughput of a four- or five-hop path is not larger than that of a two- or three-hop path. This fact is confirmed by Fig. 7(b), where the probability of selecting each N-hop path for the case of $N_{\max} = 5$ is shown. The results show that the multihop relaying occurs more frequently as the target MS is closer to the cell boundary, but the four- or five-hop path is rarely selected. Therefore, we have concluded that the throughput enhancement of multihop relaying comes mostly from two-hop relaying. This

implies that the optimal multihop path can be achieved with relatively small number of hops. Consequently, we can expect that the selected multihop path will not undergo too frequent breakage due to the mobility and that the reconfiguration of the path is not expected to be a significant factor in the network stability.

V. PERFORMANCE IMPROVEMENT BY CONCURRENCY

We investigate and compare the effect of the *type A* and the *type B* concurrent transmissions on the throughput. The concurrency improves the resource reuse, but also introduces interference among the concurrent transmissions. This tradeoff will be presented in this section. We also discuss the impact of the concurrency on the capacity of the conventional power-controlled CDMA systems. The same system model and the same propagation model that were used in the analysis of the *simple relaying* continue to be employed in this section as well.
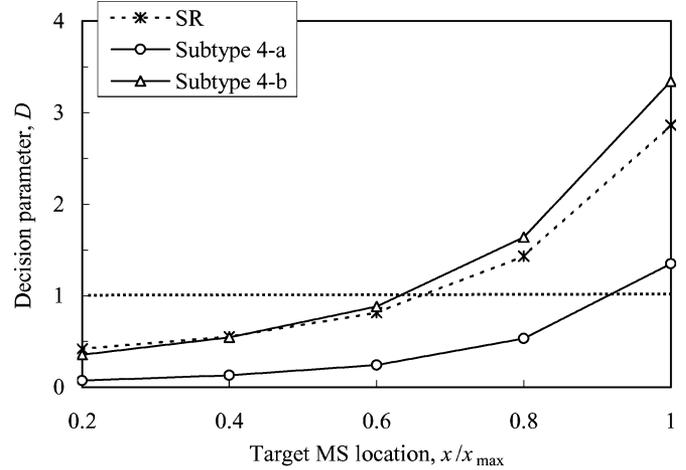
A. Type A Concurrency

We present the impact of the *type A concurrency* on the throughput in Fig. 8. These results were calculated for the case of $N = 4$ and for the same linear topology that was used in Sections III-A and B. We compare the *simple relaying* with two subtypes of the *type A concurrency*, the *subtype 4-a* and the *subtype 4-b*. For detailed description of these subtypes the reader is referred to Appendix.
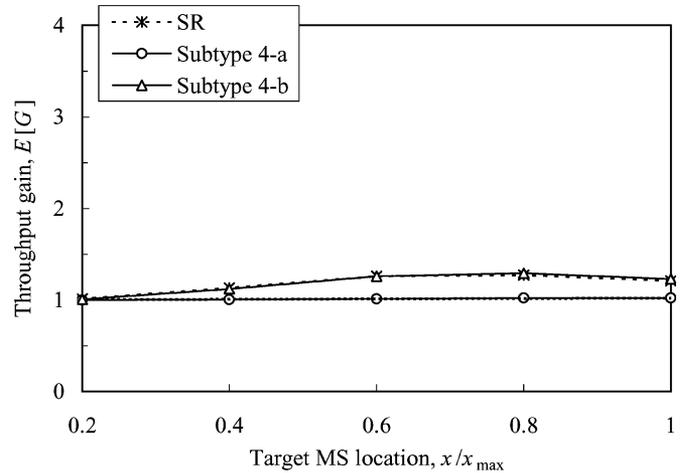
The results for D without shadowing are shown in Fig. 8(a). We observe that the performance of the *subtype 4-a* is worse than that of *simple relaying*. It is so, because the concurrent links are so close that the throughput gain resulting from reusing time slot can be wiped out by the increased interference. On the other hand, we can see that the *subtype 4-b* improves the throughput at the cell boundary. Note that the concurrent links with the *subtype 4-a* are separated by one-hop distance, but those with the *subtype 4-b* by two-hop distance. Even though the *subtype 4-b* does not utilize the time slots as much as the *subtype 4-a*, it provides more gain due to its lower interference.

The *throughput gain* $E[G]$ in the shadowed environment is shown in Fig. 8(b). Comparing to Fig. 8(a), we can see that the *throughput gain* of the *subtype 4-b* disappears at the cell boundary, and so the performance of the *subtype 4-b* is almost identical to *simple relaying*. It is so, because of the effects of *the best BS selection* and the *worst hop*, similar to the results in Fig. 6. Due to *the best BS selection*, the throughput gain for both, the *subtype 4-b* and *simple relaying*, is significantly reduced at the cell boundary. In addition, the effect of the *worst hop* in the case of the *type A concurrency* is more severe than in the case of *simple relaying*; note the minimum function in the denominator of U_R in (10) and the fact that fewer number of terms is present. Therefore, U_R for the *type A concurrency* is more susceptible to the data rate of the *worst hop* than in the case for *simple relaying*.

From the results in Fig. 8, we have found that the throughput increase of the *type A concurrency* is not significant, but we need to remember that those results are for the linear topology. The analysis of the *type A concurrency* for the random topology will be presented in the sequel.

Fig. 8. Improvement due to the *concurrency type A* in linear topology. (a) Without shadowing. (b) With shadowing.

B. Type B Concurrency

In this section, we study the impact of spatial separation of concurrent links with the *type B concurrency*. We place the target MS I at the position $(x, y = 0)$ given in *Cartesian* coordinates with respect to the center cell site, as shown in Fig. 3(a). The target MS II is placed at $(x \cdot \cos\theta, -x \cdot \sin\theta)$, so that the two MSs are separated by the angle θ. Each relaying MS is placed at the midway point between the BS and the associated target MS.

We first consider this case without shadowing. Fig. 9(a) shows D for the *type B concurrency* for various locations of the target MSs. The result for the *simple relaying* with the two-hop path is also shown in the figure and is denoted by *SR*. Note that by symmetry, $U_R^{\text{I}} = U_R^{\text{II}}$ and $U_D^{\text{I}} = U_D^{\text{II}}$, and so the *criterion 1* and the *criterion 2* result in the same decision. The results in Fig. 9 show that D increases with the separation θ, as expected. However, we can see that, in some cases, the concurrency may not provide gain over *simple relaying*. In particular, when $\theta = 60°$, *simple relaying* outperforms the concurrency except at the cell boundary. Hence, the results in this figure confirm that if the concurrent links are close to each other, the concurrency should not be used.
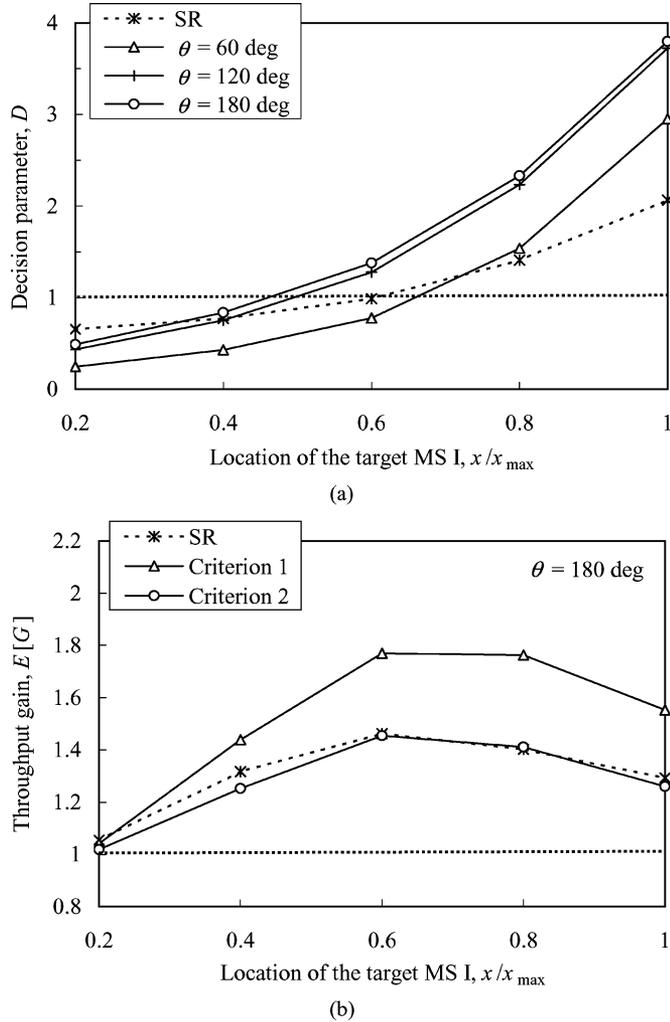
Fig. 9. Improvement due to *type B concurrency* in linear topology. (a) Without shadowing. (b) With shadowing: $\theta = 180°$.

We now continue our evaluation by incorporating shadowing into our system. Fig. 9(b) shows the results of G for the shadowed environment. These results were calculated for $\theta = 180°$. We can see that $E[G]$ can be increased by concurrency when the *criterion 1* is used. However, since this criterion is based on the calculation of the total throughput, the throughput of one of the two target MSs may be lower than that with the direct single-hop transmission. This fact can be confirmed by examining the following probability:

$$P_A \equiv \Pr\left(U_R^{\mathrm{I}} < U_D^{\mathrm{I}} \text{ or } U_R^{\mathrm{II}} < U_D^{\mathrm{II}} | U_R^{\mathrm{I}} + U_R^{\mathrm{II}} > U_D^{\mathrm{I}} + U_D^{\mathrm{II}}\right).$$
(32)

The results for P_A that we have obtained are listed in Table I and demonstrate that P_A is over 50% for all locations. This means that even if the total throughput for the concurrency case with the *criterion 1* is higher, the throughput of one of the two target MS's is mostly lower than that for the conventional system. In order to avoid the throughput decrease for any of the two target MSs, the *criterion 1* should be replaced by the *criterion 2*. However, we note that, since the *criterion 2* is more stringent in selecting the multihop path case, the performance of the *criterion 2* is worse than *criterion 1*, as shown in

TABLE I
P_A FOR $\theta = 180°$ WITH LINEAR TOPOLOGY

x/x_{max}	0.2	0.4	0.6	0.8	1.0
P_A	0.649	0.551	0.544	0.590	0.654

Fig. 9(b). Nevertheless, the *criterion 2* should be employed in order to ensure the superiority of the relaying system over the conventional system at all times.

The fact that $E[G]$ with the *criterion 2* is not higher than that with *simple relaying*, as shown in Fig. 9(b), is due to the flexibility in the path selection in the case of *simple relaying*. The *type B concurrency* provides only two options, such that, whether the multihop path or direct link is chosen, this choice is assigned to both target MSs. On the other hand, with *simple relaying*, we can select and utilize the multihop path for *each* target MS individually. Such freedom of the path selection can lead to larger throughput gain than the concurrency.

From the results above, one can expect that a hybrid control of the *criterion 2* and *simple relaying* will exploit both the concurrency and the flexible path selection. Besides, since the throughput improvement by the concurrency is highly dependent on the separation of the concurrent links, the concurrency cannot always ensure higher throughput gain than in the case of nonconcurrency. For an environment with severe interference, *simple relaying* would be a better choice. Therefore, we conclude that a hybrid scheme would provide better overall performance.

C. Hybrid of Concurrency and Nonconcurrency

We present and compare the performances of various relaying schemes in random topology with shadowing. All of the candidate relaying MSs and the target MSs are assumed to be uniformly distributed over the cell area. We find the best path for each target MS, one with the maximum throughput.

In the hybrid control of the *type B concurrency* and *simple relaying*, we compare the three effective data rates for each target MS: of the direct single-hop link, of the multihop path with *simple relaying*, and of the multihop path with the *type B concurrency*. Then, we select the best scheme, the one with the maximum sum of the effective data rates for two target MSs. Considering the *criterion 2*, if any of the two effective data rates of the multihop paths with the *type B concurrency* is smaller than that of the direct single-hop link, the *type B concurrency* is eliminated from the selection. Note that in this hybrid control, the effective data rate of the multihop path with *simple relaying* should be considered for $2T$, and it is given by $1/(2 \cdot (1/R_{R,1} + 1/R_{R,2}))$.

We also obtain and compare the results for the hybrid control of the *type A concurrency* and *simple relaying*. They are calculated for the case that the number of maximum allowable hops is five. We consider six subtypes of the *type A concurrency* shown in Table II in the Appendix and find the best one with the maximum throughput. The nonconcurrent transmission may still be chosen if it would more improve the throughput. Note that except for this hybrid control case, for the other cases—the *simple relaying*, the stand-alone *type B concurrency*, and the hybrid of
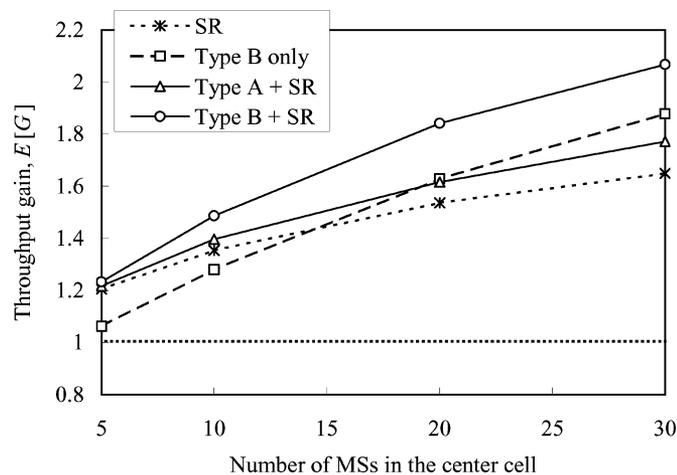
Fig. 10. Performance comparison of the concurrency in random topology.

the *simple relaying* and the *type B concurrency*—the number of hops is limited to 2.

Fig. 10 shows $E[G]$ while varying the number of the total MSs in the center cell. As expected, the throughputs for all the schemes increase with the number of MSs. Compared with the *simple relaying*, the stand-alone *type B concurrency* does not give significant gain, similar to the results in the previous subsection. In particular, for the small number of MSs, the performance of the *type B concurrency* is even worse than that of the *simple relaying* case. However, we observe that the throughput is considerably increased by the use of the proposed hybrid control scheme of the *type B concurrency* and *simple relaying*. The best scheme for a certain topology may be either *type B concurrency* or *simple relaying*, according to the MS location and the shadowing. This comparison result support our claim that a hybrid scheme should be used in the general case.

The results in Fig. 10 also show that the amount of increase offered by the *type A concurrency* is much smaller than that offered by the *type B concurrency* case. Since the concurrent transmissions in the *type B concurrency* can be farther away from each other than in the *type A concurrency*, the increase in interference power can be smaller in the *type B concurrency*. From the results, we conclude that in order to achieve the best improvement, the concurrency should be applied between different downstream paths, rather than within the same downstream path.

D. Impact on the Capacity of the Conventional Power-Controlled System

We now discuss the impact of the multihop relaying on the capacity in the conventional CDMA system, where the power control is employed to provide a constant bit rate while maintaining the required quality, and the channel multiplexing within a cell is based on CDM (for downlink) or CDMA (for uplink).

Previous works showed that the coverage of those systems can be increased by the multihop relaying, but that it is not easy to enhance the capacity under the interference-limited conditions, such as the case in the high-loaded systems with small

cells [5]–[9].[5] [6] This is mainly due to the interference increase from the concurrently relayed transmissions. Recall that in the conventional systems, the near–far problem can be avoided by power control. However, in the multihop system, it is possible that the interfering transmitter, rather than the intended transmitter, is closer to the receiver. In such a case, any power control scheme would be ineffective and could, in fact, result in very high interference level, possibly leading to total outage.[7] Although such interfering transmission may be avoided by the use of a sophisticated algorithm for routing and time slot assignment, optimization of both, the selections of the relaying MS and the time slot seems to be difficult to implement, because of a large number of transmissions that would normally occur concurrently within the same cell.

In this work, we have realized the concurrency gain by allowing only two downstream paths at any time and also by utilizing both, concurrent and nonconcurrent transmissions. Since the downstream paths are implemented in the TDM fashion, in such a way that any period $2T$ is dedicated to distinct pair of MSs, the path selection for the two MSs does not affect the signal quality of the other MSs in the same cell. So we have only to consider the interference between the two paths for these two MSs. If this interference is too high, it can be avoided by choosing the direct paths or the nonconcurrently relaying paths. Clearly, such control will be easier in the TDM system than in the CDM system where the path selection for each MS affects one another. More comprehensive comparison of TDM and CDM in the multihop cellular networks is left as a further study. Though, we can claim at least that the capacity gain by multihop relaying can be more easily implemented in the rate-controlled CDMA system with TDM for multiplexing the same cell channels, than in the conventional power-controlled CDMA system.

VI. APPLICATION OF MULTIHOP RELAYING

The multihop relaying technology provides a significant flexibility in design and operation of cellular networks. In multihop cellular network, MS may choose to utilize multihop relaying instead of single-hop direct transmission. As an example of an application that could benefit from the multihop extension of cellular network, we present the improvement in fairness and robustness in the presence of variability of traffic distributions.

A. Fairness Improvement

One of the concerns in the design of cellular networks is the tradeoff between throughput and fairness [4]. System

---

[5]Although Rouse *et al.* [9] evaluated the capacity improvement achieved through multihop relaying over the conventional system, such an improvement does not occur for all nonlocal traffic cases in which the source and the target are in different cells, or only one of the two resides in the wireless network. Note that in our paper, the nonlocal traffic is assumed to be much more dominant than the local traffic whose source and target are in the same cells.

[6]Zadeh *et al.* [13] demonstrated the capacity improvement achievable by multihop relaying, unlike in their previous work [6]. However, interference cancellation scheme such as multiuser detection (MUD) was employed in [13]. Note that in our work, we do not apply any scheme for the interference cancellation.

[7]Such near–far problems may also arise in mobile ad hoc networks (MANET's) [19].
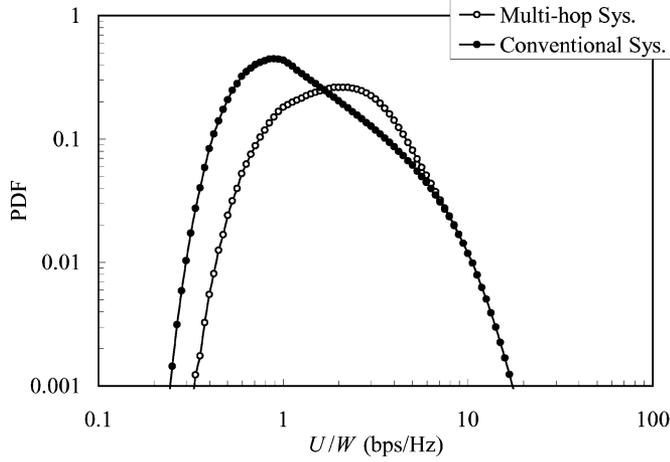
Fig. 11.   PDF of the effective data rate normalized to the bandwidth.

Fig. 12.   Improvement of the tradeoff between fairness and throughput.

throughput can be maximized by allocating more radio resources to a user with higher SIR; e.g., a user that is close to BS. However, a user with lower SIR will experience higher latency. If more resources are assigned to the lower SIR user to achieve better fairness, the total system throughput will decrease. Therefore, it is not easy to provide QoS fairly over the whole service area and, at the same time, to maximize the system throughput. This tradeoff caused by the location-dependent SIR is an inherent feature of the cellular network.

In Section IV, we showed that the *throughput gain* achievable with the multihop relaying increases as the MS is closer to the cell boundary. This implies that the degraded QoS due to low SIR can be compensated for through the use of the multihop relaying. This can be confirmed through the comparison of two distributions of the effective data rates for the multihop system and for the conventional system.

Fig. 11 shows the probability density functions (PDFs) of the effective data rate for the multihop system and for the conventional system. The effective data rate U is here defined as $\max\{U_R, R_D\}$ for the multihop system, whereas it is identical to R_D for the conventional system. The result of the multihop system was calculated for the case in which the number of hops is two and the number of MSs in the center cell is 20. In Fig. 11, we can see that, compared with the conventional system, the low data rate provided in the multihop system is increased, while the high data rate remains the same. Such improvement in the low rate region could result in an improvement in fairness: e.g., the standard deviation normalized to the mean is about 0.74 for the multihop system and 0.94 for the conventional system.

We now establish the improvement in fairness by demonstrating the tradeoff curve of throughput and fairness in a multihop system. The simple scheduling model and the performance metrics provided in [20] are used for this analysis. We assume that, after the effective data rate is calculated from the data rate of each hop, the transmission period T is assigned under the following rule:

$$T = \tau \cdot U^{-\delta} \tag{33}$$

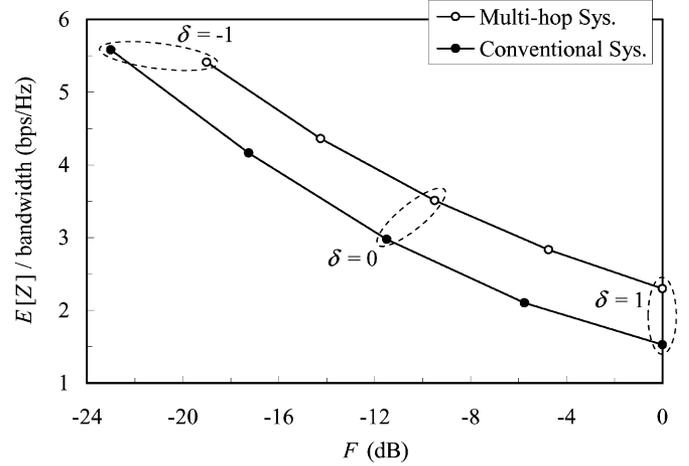where δ is the fairness control parameter $(\delta \leq 1)$ and τ is a constant.

Fairness can be controlled with δ, and various scheduling algorithms can be implemented accordingly. For instance, with $\delta = 0$, equal amount of service time is assigned to all users, i.e., even scheduling. A fair scheduling can also be implemented with $\delta = 1$, because the assigned amount of service time is inversely proportional to the provided data rate. On the other hand, if δ is set to a very large negative value, almost the entire service time is assigned to one user with the highest data rate, i.e., SIR-based scheduling.

Let F be the *fairness performance factor* reflecting the user QoS discrimination. We have chosen the transmitted amount of data during one round-robin period, S, as the QoS metric. Then, S can be represented by

$$S = T \cdot U = \tau \cdot U^{1-\delta}. \tag{34}$$

Hence, we can define F as the ratio of the minimum and the maximum values of S

$$F \equiv 10 \log \frac{S_{\min}}{S_{\max}} = 10(1 - \delta) \log \frac{U_{\min}}{U_{\max}} \tag{35}$$

where $S_{\min}$ and $S_{\max}$ are defined as the χ percentile and the $(1 - \chi)$ percentile of S, respectively, and $U_{\min}$ and $U_{\max}$ follow the same definition.

Let us determine the system throughput from the viewpoint of BS. The system throughput Z can be defined as the effective data rate provided by a BS. Let $g(U)$ be the probability that the effective data rate U occupies the downstream path which the BS currently provides. Then, the expected value of Z is given by

$$E[Z] \equiv \int U \cdot g(U) dU = \frac{\int U^{1-\delta} \cdot f(U) dU}{\int U^{-\delta} \cdot f(U) dU} \tag{36}$$

where $f(U)$ is the PDF of U.

To demonstrate the tradeoff curve, we have calculated the system throughput and the *fairness performance factor* by using the PDF shown in Fig. 11 and the formulae derived above. The $E[Z]$ is plotted in Fig. 12 as a function of F with $\{\delta : -1, -0.5, 0, 0.5, 1\}$. The confidence level χ is set to 0.05. As expected, the two metrics show a tradeoff relation for both systems: F increases with δ and $E[Z]$ decreases accordingly. We can also see that the tradeoff curve for the multihop system
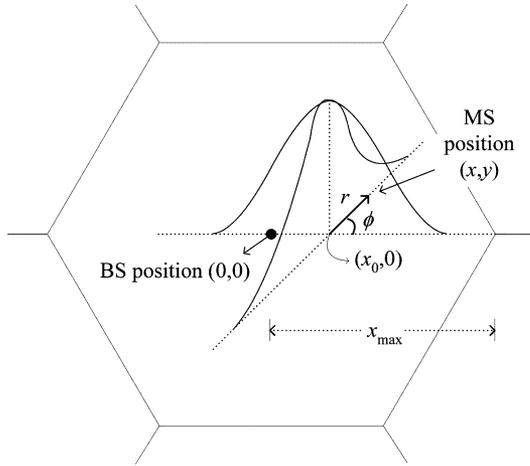
Fig. 13.   Bell-shaped traffic distribution.

Fig. 14.   *Throughput gain* for the nonuniform traffic distribution.

is positioned in the right and upper direction relative to the conventional system. This implies that fairness and throughput are improved at the same time. We have, therefore, concluded that the multihop relaying can improve the fairness performance, as well as increase the system throughput.

B. Robustness to the Variability of Traffic Distributions

Another main application of the multihop relaying technology is to mitigate the inefficiency due to the changes in traffic distribution in cellular networks. Traffic distribution in a practical system is not uniform. Moreover, it may be frequently changing as well. Suppose that a traffic hot spot moves away from the center of a cell toward its boundary. Then, the system performance degrades, because the QoS in the cellular network is dependent on the location. It seems very difficult to cope with such variability of traffic demand in the cellular architecture with fixed BS.

In Fig. 7, we have already confirmed that the *throughput gain* increases as the MS gets closer to the cell boundary. In addition, as the relaying MSs are closer to the target MS, the shorter relaying path can be secured. Thus, higher *throughput gain* can be obtained when more MSs are concentrated at the cell boundary. One can expect that multihop systems be able to more efficiently accommodate the nonuniform traffic, as compared with the conventional cellular system. We next discuss the performance when all the MSs are nonuniformly distributed throughout the cell area.

We used the bell-shaped distribution shown in Fig. 13 as our nonuniform traffic model. The distance from the center of the distribution to the location of the MS r follows the *Gaussian* distribution with zero-mean and standard deviation σ_B. The azimuth ϕ is modeled as a uniformly distributed random variable on $[0, \pi)$.

Fig. 14 shows the $E[G]$ for different settings of the center of the bell-shaped distribution x_0 and for different values of the standard deviation σ_B. These results were calculated for the case of 20 MSs in the center cell. The maximum number of hops is limited to two. We can see that $E[G]$ is larger than 1 for all the presented cases. This means that the multihop network always outperforms the conventional network with respect
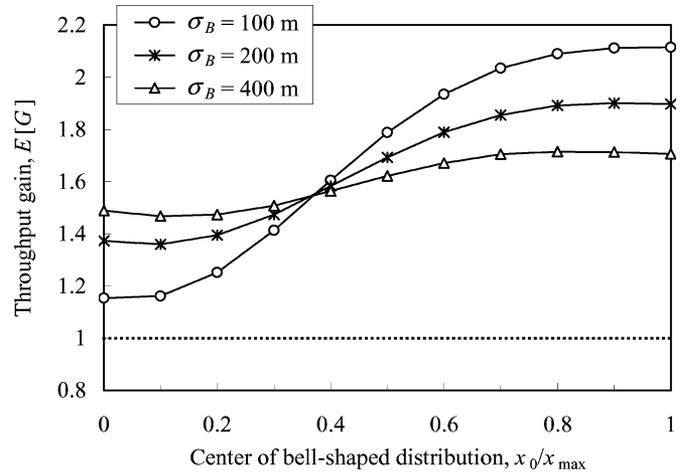
to throughput. The figure also shows that $E[G]$ increases as the distribution moves toward the cell boundary. This implies that the throughput improvement with multihop relaying increases as the BS is placed less optimally; in other words, the multihop relaying can reduce the performance degradation caused by nonoptimal positioning of the BS. Hence, multihop relaying can allow for more flexible design of the cell site and make up for suboptimal placement of the cell site. Multihop relaying also makes the network relatively robust to changes in the traffic distribution.

So far, for simplicity of the analysis, we have used the system model where the multihop relaying is employed only in the center cell. We refer to this model as the *simple system model*. We now extend this *simple system model* to the model where the multihop relaying is employed in all the 19 cells. We term this model as the *extended system model*. In the *extended system model*, two relay transmissions in the adjacent cells may be positioned near the cell boundary and, thus, may interfere with each other. Next, we study the impact of such interference.

We set the *extended system model* as follows: One traffic hot spot, modeled as the bell-shaped distribution introduced earlier, is placed in every cell. Hence, the traffic distribution in the *extended system model* is comprised of 19 bell-shaped distributions. The center of each bell-shaped distribution is uniformly selected within its cell area, except for the center cell, where the location of the bell-shaped distribution is given as x_0, as in the *simple system model*. The standard deviations of all bell-shaped distributions are all equal to σ_B. We also assume that every cell has the same number of MSs.[8] The simulation collects statistical data from the only center cell, as in the *simple system model*.

In Fig. 15, we present the *throughput gains* in the *extended system model* and in the *simple system model*, for different settings of the location of the bell-shaped distribution in the center cell x_0. These results are calculated for the case of 20 MSs in each cell and for $\sigma_B = 100$ m. The figure shows that $E[G]$ in the *extended system model* increases with x_0, but it slightly decreases when x_0/x_{max} is over 0.8. The $E[G]$ near the cell

[8]This assumption is made for the purpose of focusing on the impact of the nonuniform traffic distribution, rather than on the impact of an unbalanced loading among cells. The study of the unbalanced load condition is left for future work.

boundary is also smaller compared with the result of the *simple system model*. This is due to the impact of the interference between the close relay transmissions in the adjacent cells.[9] However, we observe that, except for areas close to the cell boundary, $E[G]$ in the *extended system model* still increases as the distribution moves from the cell site toward the cell boundary, similar to the result of the simple system model. Therefore, this result shows that our claim still holds for large fraction of the cell area: the multihop relaying can reduce the performance degradation caused by nonoptimal positioning of the BS. Finally, we note that one can devise a packet scheduling and multiplexing scheme which detects a close relay transmission in the adjacent cell and defers the relay transmission, so as to avoid the high interference.

VII. CONCLUSION

The results that we have obtained in this paper show that the system throughput can be significantly enhanced through the use of multihop relaying. By examining the results, we have derived the rules for the multihop relaying with the rate-controlled downstream path. First, one should use a hybrid mode of operation of the direct transmission and the multihop relaying, rather than the relaying mode alone. Next, since the throughput enhancement with paths of length longer than three hops is not significant, the multihop relaying scheme with only two- and three-hop paths is sufficient to obtain most of the *throughput gain*. Finally, to obtain an additional improvement, the concurrency should be applied among different downstream paths, rather than within the same downstream path, and also both scheme, concurrent and nonconcurrent transmission should be considered to obtain optimal performance.

This paper has demonstrated that multihop relaying provides significant performance gain in the cellular networks, and that it can be converted into an improvement of QoS. Our results have shown that unfairness in the QoS due to the location-dependent SIR can be mitigated in a multihop system. That is, through the use of multihop relaying, the QoS can be provided more evenly over the cell area. Therefore, we have found that multihop relaying provides not merely an increase in the system throughput, but also an improvement in fairness as well.

Our results also suggest that the multihop cellular network, as a self-configuring network mechanism, can efficiently accommodate the spatial and the temporal variability in the traffic distribution. The results have shown that the throughput can be significantly enhanced in the case of nonuniform traffic distribution, as well as for the uniform traffic distributions. Thus, a multihop cellular network is relatively robust to changes in the traffic distribution. The importance of this conclusion is in the fact that it allows less stringent design and planning of such multihop cellular systems, as well as to improve the performance of the existing systems that are subject to temporal traffic changes.

Fig. 15. *Throughput gain* in the extended system model.

TABLE II
SUBTYPES FOR *TYPE A CONCURRENCY*

N	M	Formation	Q	Name
3	2	$\{\{\mathcal{T}_1 : 1, 3\}, \{\mathcal{T}_2 : 2\}\}$	1	**3-a**
4	2	$\{\{\mathcal{T}_1 : 1, 3\}, \{\mathcal{T}_2 : 2, 4\}\}$	1	**4-a**
	3	$\{\{\mathcal{T}_1 : 1, 4\}, \{\mathcal{T}_2 : 2\}, \{\mathcal{T}_3 : 3\}\}$	2	**4-b**
5	2	$\{\{\mathcal{T}_1 : 1, 3, 5\}, \{\mathcal{T}_2 : 2, 4\}\}$	1	**5-a**
	3	$\{\{\mathcal{T}_1 : 1, 4\}, \{\mathcal{T}_2 : 2, 4\}, \{\mathcal{T}_3 : 3\}\}$	2	**5-b**
	4	$\{\{\mathcal{T}_1 : 1, 5\}, \{\mathcal{T}_2 : 2\}, \{\mathcal{T}_3 : 3\}, \{\mathcal{T}_4 : 4\}\}$	3	**5-c**

The results presented in this paper provide a significantly better understanding of the tradeoffs in the design of multihop cellular networks than what is currently available in the technical literature, and that our results will contribute to the design of protocols for such networks.

APPENDIX
SUBTYPES FOR *TYPE A CONCURRENCY*

We describe here the subtypes of the *type A concurrency* that were considered in this work. Each subtype can be represented by gathering the set $\mathcal{T}_m$. For instance, the subtype shown in Fig. 2 can be represented by $\{\{\mathcal{T}_1 : 1, 3\}, \{\mathcal{T}_2 : 2, 4\}\}$. Recall that the element of the set $\mathcal{T}_m$, n, denotes the nth hop from the BS. In a similar way, we can represent other subtypes for each N-hop path, as shown in Table II. In the table, Q denotes the separation of the concurrent links, and its unit is the hop-distance. Note that the *type A concurrency* is not applicable for two-hop path, because TDD is applied as a duplex scheme for relaying.

For simplicity, we have considered only the six subtypes listed in Table II, though there are total 18 subtypes: one for three-hop path, four for four-hop path, and 13 for five-hop path. Consideration of the whole subtypes requires too much computation for finding the best relaying path. For example, if the number of candidate relaying MSs is X, we should check

---

[9]A relay transmission of the adjacent cell near the cell boundary may interfere with the receiver located anywhere in the center cell. However, if the relaying mobile in the adjacent cell is far away from the center cell boundary, such interference would be reduced. For this reason, the *throughput gain* in the case of the uniform traffic are the almost same; our results for the uniform traffic distribution show that $E[G]$s in the *extended system model* and in the *simple system model* are about 1.59 and 1.54, respectively.
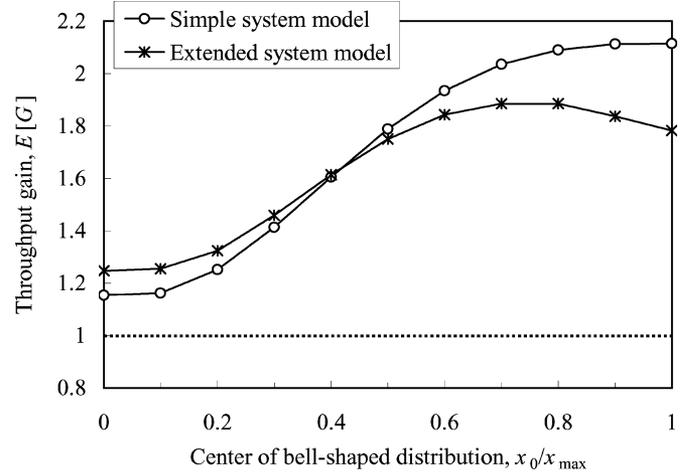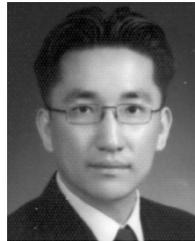
$\Pi_{i=0}^{N-2}(X-i)$ combinations even for one subtype of N-hop path. On the contrary, in the case of *type B concurrency* with two-hop path, the total number of combinations to be searched is just $X(X-1)$. Enlarging the considered subtypes might improve the performance of the *type A concurrency* over the results shown in the paper. However, considering the large increase in the computation, clearly, the *type A concurrency* is still less favorable than the *type B concurrency*.

REFERENCES

[1] 3GPP TR25.924, "Opportunity driven multiple access," 3GPP, ver. 1.0.0, Dec. 1999.

[2] TR 101 146, "UMTS; UTRA; Concept evaluation (UMTS 30.06)," ETSI, ver. 3.0.0, Dec. 1997.

[3] S. Toumpis and A. Goldsmith, "Capacity regions for wireless ad hoc networks," in *Proc. IEEE ICC 2002*, New York, Apr. 2002, pp. 3168–3173.

[4] M. Airy and K. Rohani, "QoS and fairness for CDMA packer data," in *Proc. IEEE Vehicular Technology Conf. 2000 Spring*, Tokyo, Japan, May 2000, pp. 450–454.

[5] A. Fujiwara *et al.*, "Area coverage and capacity enhancement by multihop connection of CDMA cellular network," in *Proc. IEEE Vehicular Technology Conf. 2002 Fall*, Vancouver, BC, Canada, Sept. 2002, pp. 2371–2374.

[6] A. N. Zadeh and B. Jabbari, "Performance analysis of multihop packet CDMA cellular networks," in *Proc. IEEE GLOBECOM 2001*, San Antonio, TX, Nov. 2001, pp. 2875–2879.

[7] T. J. Harrold and A. R. Nix, "Performance analysis of intelligent relaying in UTRA TDD," in *Proc. IEEE Vehicular Technology Conf. 2002 Fall*, Vancouver, BC, Canada, Sept. 2002, pp. 1374–1378.

[8] T. Rouse, S. McLaughlin, and H. Haas, "Coverage-capacity analysis of opportunity driven multiple access (ODMA) in UTRA TDD," in *Proc. IEE 3G Mobile Communication Technologies*, Mar. 2001, pp. 252–256.

[9] T. Rouse, I. Band, and S. McLaughlin, "Capacity and power investigation of opportunity driven multiple access (ODMA) networks in TDD-CDMA based systems," in *Proc. IEEE ICC 2002*, New York, Apr. 2002, pp. 3202–3206.

[10] A. G. Spilling, A. R. Nix, M. A. Beach, and T. J. Harrold, "Self-organization in future mobile communications," *Electron. Commun. Eng. J.*, pp. 133–147, June 2000.

[11] R. Becher, M. Dillinger, M. Haardt, and W. Mohr, "Broad-band wireless access and future communication networks," *Proc. IEEE*, vol. 89, pp. 58–75, Jan. 2001.

[12] H. Wu, C. Qiao, W. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2105–2115, Oct. 2001.

[13] A. N. Zadeh, B. Jabbari, R. Pickholtz, and B. Vojcic, "Self-organizing packet radio ad hoc networks with overlay (SOPRANO)," *IEEE Commun. Mag.*, vol. 40, pp. 149–157, June 2002.

[14] Q. Bi *et al.*, "Performance of 1xEV-DO third-generation wireless high-speed data systems," *Bell Labs Tech. J.*, vol. 7, pp. 97–107, Mar. 2003.

[15] A. Das *et al.*, "Evolution of UMTS toward high-speed downlink packet access," *Bell Labs Tech. J.*, vol. 7, pp. 47–68, Mar. 2003.

[16] F. Graziosi and F. Santucci, "A general correlation model for shadow fading in mobile radio systems," *IEEE Commun. Lett.*, vol. 6, pp. 102–104, Mar. 2002.

[17] TR 101 112, "UMTS; Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03)," ETSI, ver. 3.2.0, Apr. 1998.

[18] A. J. Viterbi, A. M. Viterbi, K. S. Gilhousen, and E. Zehavi, "Soft handoff extends CDMA cell coverage and interferences reverse link capacity," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1281–1288, Oct. 1994.

[19] A. Muqattash and M. Krunz, "CDMA-based MAC protocol for wireless ad hoc networks," in *Proc. ACM MobiHoc 2003*, Annapolis, MD, June 2003, pp. 153–164.

[20] J. Cho and D. Hong, "Tradeoff analysis of throughput and fairness on CDMA packet downlinks with location-dependent QoS," *IEEE Trans. Veh. Technol.*, submitted for publication.

Jaeweon Cho (S'95–M'03) received the B.Sc. (*magna cum laude*), M.Sc., and Ph.D. degrees in electronic engineering from Sogang University, Seoul, Korea, in 1995, 1997, and 2002, respectively.

From 1997 to 1998, he was with the R&D Center, Dacom Corporation, Daejeon, Korea. From 2002 to 2003, he was a Postdoctoral Associate in the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. In 2003, he joined the Telecommunication R&D Center, Samsung Electronics Co., Ltd., Gyeonggi-do, Korea, as a Member of Technical Staff. His research interests are in wireless communications and the next-generation system issues, including the system architectures, control algorithms, and performance analysis.

Dr. Cho is a member of the Korean Institute of Communication Sciences (KICS).

Zygmunt J. Haas (S'84–M'88–SM'90) received the B.Sc. and M.Sc. degrees in electrical engineering, in 1979 and 1985, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, in 1988.

He then joined AT&T Bell Laboratories, Network Research Department, where he pursued research on wireless communications, mobility management, fast protocols, optical networks, and optical switching. From September 1994 to July 1995, he was with AT&T Wireless Center of Excellence, where he investigated various aspects of wireless and mobile networking, concentrating on TCP/IP networks. Since August 1995, he has been with the faculty of the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. He is an author of numerous technical papers and holds 15 patents in the fields of high-speed networking, wireless networks, and optical switching. His interests include mobile and wireless communication and networks, personal communication service, and high-speed communication and protocols.

Dr. Haas has organized several workshops, delivered numerous tutorials at major IEEE and ACM conferences, and serves as Editor of several journals and magazines, including the IEEE TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the *IEEE Communications Magazine*, and the *ACM/Kluwer Wireless Networks Journal*. He has been a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS issues on Gigabit Networks, Mobile Computing Networks, and Ad Hoc Networks. He is a voting member of the Association for Computing Machinery (ACM), and the Chair of the IEEE Technical Committee on Personal Communications.