Abstract

This article presents an overview of soft handoff, an idea which is becoming quite important because of its use in the IS-95 code-division multiple access (CDMA) cellular phone standard. The benefits and disadvantages of using soft handoff over hard handoff are discussed, with most results drawn from the available literature. The two most well-known benefits are fade margin improvement and higher uplink capacity, while disadvantages include increased downlink interference and more complex implementation. Handoff parameter optimization is extremely important, so various studies on the trade-offs to be considered when selecting these parameters are surveyed, from both the link quality and resource allocation perspectives. Finally, research directions and future trends are discussed.

Soft Handoffs in CDMA Mobile Systems

Daniel Wong, Stanford University
Teng Joon Lim, National University of Singapore

Handoff is an essential component of mobile cellular communication systems. Mobility causes dynamic variations in link quality and interference levels in cellular systems, sometimes requiring that a particular user change its serving base station. This change is known as a *handoff*. In first-generation cellular systems like the Advanced Mobile Phone System (AMPS) [1, 2], handoffs were relatively simple. Second-generation cellular systems like the Global System for Mobile Communications (GSM) and the Personal Access Communication System (PACS) [2, 3] are superior to first-generation ones in many ways, including the handoff algorithms used. More sophisticated signal processing and handoff decision procedures have been incorporated in these systems. The control/decision structures have been improved so that in progressing from network-controlled toward mobile-assisted handoffs (MAHO) or mobile-controlled handoffs (MCHO), the handoff decision delay has been substantially reduced. Another idea that has been proposed for improving the handoff process is soft handoff, the subject of this article.

Our purpose is to provide an overview of soft handoff, from the point of view of the performance benefits available and the trade-offs involved in selecting system parameters. Recent research on various aspects of soft handoff are reviewed and discussed. It is not the objective of this article to provide comprehensive coverage of every paper referenced, and the interested reader is encouraged to refer to the tables themselves for details. We do present a broad survey of the technical issues involved, with some details of the highlighted issues to assist the reader in developing a good understanding of those issues, and understanding the place of soft handoff in modern cellular systems.

This article is organized as follows. In the rest of this introductory section, we will discuss what soft handoffs are and why one might want to implement them in a cellular system. In the second section, we will look at benefits of soft handoff. However, in order to reap the benefits of soft handoff, it is necessary that the handoff parameters be well set. So in the third section we will examine the tradeoffs involved and discuss the setting of soft handoff parameters based on an understanding of the tradeoffs. A general discussion and conclusions will be given in the fourth section.

What Is Soft Handoff?

Soft handoff is so called to distinguish it from the more traditional hard handoff process. With hard handoff, a definite decision is made on whether to handoff or not. On a positive decision, the handoff is initiated and executed without the user attempting to have simultaneous traffic[1] channel communication with the two base stations.[2] With soft handoff, a *conditional decision* is made on whether to hand off. Depending on the changes in pilot signal strength from the two or more base stations involved, a hard decision will eventually be made to communicate with only one. This normally happens after it is clear that the signal from one base station is considerably stronger than those from the others. In the interim period, the user has simultaneous traffic channel communication with all candidate base stations.

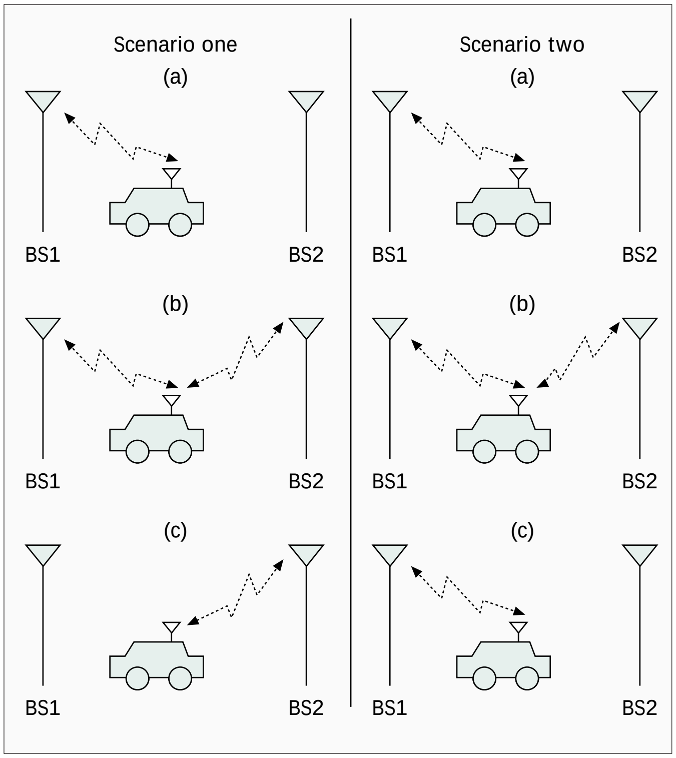
The difference between hard and soft handoffs is like the difference between swimming relay events and track-and-field relay events. In swimming relays, the next swimmer starts just as the preceding one touches the wall, analogous to the switch from one base station to another in a hard handoff. In track-and-field relays, the baton is passed from one runner to the next after the second runner starts running, and so for a short time they are both running together, analogous to a soft handoff.

---

[1] *In both soft and hard handoffs, there will normally be some simultaneous* control *channel communication between the two base stations and the user according to the signaling protocol in use, so we must look at traffic channels to distinguish between hard and soft handoffs.*

[2] *Handoff can also be between two channels at one base station. If the two channels are in two sectors of a sectorized cell, and the kind of handoff used is soft, this is sometimes known as "softer handoff" [4]. The IS-95 cellular standard does not discuss intersector handoff in much detail, and mostly assumes that all handoffs are intercell. In this article, we take handoff to involve two or more base stations unless otherwise stated.*

1070-9916/97/$10.00 © 1997 IEEE



■ **Figure 1.** *A simple soft handoff situation with two different outcomes. a) The column on the left shows a user at first having only BS1 in its active set; b) followed by a period time in which the pilot signal strengths of both BS1 and BS2 are strong, so the user has both BS1 and BS2 in its active set. c) Eventually, the signal strength of BS1 declines and the signal strength of BS2 increases to the point where BS1 is removed from the active set. The column on the right shows the same situation, except that at the end BS2 is removed from the active set.*

It also helps to think of soft handoff in terms of membership in an *active set*, which is the set of base stations with which a user is communicating at any given time. It normally consists of one base station, but other base stations are added when the signal strength between them and the user exceeds the predefined *add threshold*. Base stations are removed from the active set when the power of the signals received from them drop below the *drop threshold* and remain there for at least T_{drop} s, where T_{drop} is a preset constant. A system using soft handoff allows for more than one base station in the active set during the handoff period, while in hard-handoff systems, only one base station at a time is ever in the active set.

A simple case of soft handoff is shown in Fig. 1. There are only two base stations involved in the example. The same basic setup has two outcomes, one of which is shown in the left column as "scenario one," and the other in the right column as "scenario two." Scenario one shows a soft handoff from base station BS1 to base station BS2. Scenario two shows a case in which the user goes back to BS1 after a period of time in soft handoff. This might be the case when the mobile is temporarily obstructed from the line of sight of BS1. The equivalent scenario with hard handoff would be a hard handoff to BS2 and then another hard handoff back to BS1, wasting valuable network resources in the process of carrying out these unnecessary handoffs.

Why Implement Soft Handoffs?

It is desirable to implement soft handoffs in power-controlled code-division multiple access (CDMA) systems because implementing hard handoffs is potentially difficult in such systems, as will be explained shortly. There are also some other advantages and disadvantages arising from the use of soft handoff, which will be mentioned later in this section.

Power Control and Soft Handoffs — A system with power control attempts to dynamically adjust transmitter powers while in operation. Power control is closely related to soft handoff. One reason for this is that the IS-95 CDMA standard [5], the only major cellular standard that implements soft handoff, uses both power control and soft handoff as interference-reduction mechanisms (more details on IS-95 handoffs are in Appendix A). Power control is the main tool used in IS-95 to combat the near-far problem.[3] It is theoretically unnecessary to have power control if one can successfully implement a more intelligent receiver than that used in IS-95, which is the subject of the field of multi-user detection. However, the work of Verdu and others in this field [6–8] have not yet demonstrated that such receiver structures are practical. The problems faced include code sequence mismatch (examined for one of the simpler multiuser detectors in [9, 10]), various other synchronization issues, an over-simplified channel model (which ignores or does not adequately model small-scale Rayleigh fading effects,[4] for example), and the sheer computational complexity of the proposed receivers.

Therefore, at present power control is necessary in order for a CDMA system to achieve a reasonable level of performance in practice. However, the use of power control in CDMA systems necessitates the use of soft handoff when the original and new channels occupy the same frequency band. Because of the nature of CDMA, the channels are wideband channels, and many channels must necessarily occupy the same frequency band in order for the system to use bandwidth efficiently. Therefore, the overwhelming majority of CDMA handoffs are of this type, and must use soft handoff.

(Note that it *is* possible for the old and new channels to occupy different frequency bands, and in such a case hard handoff is feasible, if the frequency bands are far enough apart and the transmit/receive filters have sharp cutoffs. The IS-95 standard provides for hard handoffs for these cases. Furthermore, because of hardware limitations, hard handoff is the *only* kind of handoff that can be performed in such cases.)

The reason is that for power control to work properly, the user must attempt to be linked at all times to the base station from which it receives the strongest signal. If this does not happen, a positive power control feedback loop could inadvertently occur, causing system problems. Soft handoff can guarantee that the user is indeed linked at all times to the base station from which it receives the strongest signal (it may possibly be linked to other base stations simultaneously as well), whereas hard handoff cannot guarantee this.

The problem will be illustrated by an example using the setup shown in Fig. 2. There are only two base stations, B1 and B2, and two users, M1 and M2. Power control and hard

[3] *The performance of CDMA systems is very sensitive to differences in received signal powers from various users on the uplink. Due to the nonorthogonality of the spreading codes used by different users, a strong interfering signal may mask out a weak desired signal, causing unreliable detection of the latter. This is known as the near-far problem.*

[4] *Small-scale fading is also known as fast fading, but we follow [11] in calling it small-scale fading because the term "fast fading" is also used to refer to a rapidly time-varying channel with a large Doppler spread. More details on small-scale and large-scale fading, and fast and slow fading, can be found in [11, 12].*

handoff are used. The serving base station of user M1 is B1, and the serving base station of user M2 is B2. A case will be examined in which no handoff is performed, even though the link between B1 and M2 is better than that between B2 and M2, and the link between B2 and M1 is better than that between B1 and M1; this is because the difference is not great enough to overcome the hysteresis (see Appendix B) built into the hard handoff algorithm used. The uplink is examined, and all signal powers listed are the received signal powers (in dB) at the base stations. The received signal powers are denoted by $S_{i,j}$, where i = 1,2 denotes the user and j = 1,2 denotes the base station. There will also be power changes caused by user movement, but for clarity in this example it will be assumed that the users are stationary. Thus, changes in power levels are due only to power control. The hysteresis margin is Δ = 4 dB. Power control is implemented as follows:

- Assume that received "signal'' power must be greater than received "interference'' power for the communication quality on an uplink to be acceptable.
- Therefore, the following updating periodically occurs:
  1. (Processed at B1) If $S_{1,1} \leq S_{2,1}$, $\varepsilon_1 = 1$; otherwise, $\varepsilon_1 = -1$. Instruct M1 to change its transmitted power such that S1,1 changes by ε_1 at the next iteration. If $S_{2,1} - S_{2,2} > \Delta$ then have M2 hand off to B1.
  2. (Processed at B2) If $S_{2,2} \leq S_{1,2}$, $\varepsilon_2 = 1$; otherwise, $\varepsilon_2 = -1$. Instruct M2 to change its transmitted power such that $S_{2,2}$ changes by ε_2 at the next iteration. If $S_{1,2} - S_{1,1} > \Delta$, then have M1 hand off to B2.

Table 1 shows the transmitted signal powers at each power control iteration. Both base stations keep instructing their respective users to increase their transmit powers, but the relative power levels do not change, thereby creating a positive feedback effect. At the same time, the total interference caused by these two users to others in the two cells increases with each iteration. The natural correcting mechanism of handoff fails to work here because $S_{2,1} - S_{2,2}$ and $S_{1,2} - S_{1,1}$ are both smaller than Δ at every iteration.

Advantages and Disadvantages of Soft Handoff — One reason for implementing soft handoffs has already been given above (i.e., because power control is used). Other advantages of using soft handoff are listed below. In the following list, there is an implicit comparison with traditional hard handoff systems using a simple threshold-with-hysteresis decision criterion, described in Appendix B.

Advantages —

- Soft handoff reduces/eliminates the "ping-pong'' effect common in hard handoff. This results in:
  - Less load on the network from handoff signaling and overhead.
  - Smoother user communications without the "clicks'' typical of hard handoff when speech transmissions are stopped momentarily during handoffs.
- With soft handoff, there is no hysteresis margin, resulting in less delay and equivalent to "instantaneous'' macroscopic selection diversity. This is accomplished by "instantaneous'' switching to the best base station signal during a soft handoff (uplink), and avoids the additional interference associated with handoffs with hysteresis. Hence:
  - Keeping base station separations and (base station and user) transmitter powers fixed, the overall uplink interference is reduced, leading to:
    a) better communication quality for a given number of users.
    b) more users (i.e., greater capacity) for the same required E_c/I_0 (ratio of received energy per chip to total received spectral density).
    c) smaller required uplink transmitter powers, further reducing uplink interference.
  - Keeping required outage probability and base station separation fixed, the system fade margins are reduced. This leads to smaller required downlink transmitter powers and downlink interference.
  - Keeping the same required outage probability and fade margins, base station separations increase.

  More details on these points are given in the second section.
- Soft handoff imposes fewer time constraints on the network. There is a longer mean queuing time to get a new channel from the target base station, so this helps reduce blocking probability or probability of dropped calls.

Against these advantages, soft handoff faces the following drawbacks.

Disadvantages —

- Additional network resources are used during a soft handoff. These resources thus become unavailable for use elsewhere.
- Soft handoff is more complex.
- Downlink interference (to other users) increases when soft handoff is in progress, since several base stations are transmitting what would otherwise be transmitted by one base station. This can add to the uplink interference too, if the same frequency is used for uplink as for downlink. The interference-increasing effect should normally be slight, if it is assumed that only a small fraction of the duration of a typical call is spent in soft handoff.

Clearly, it is difficult to conclude that one type of handoff is better in absolute terms. System

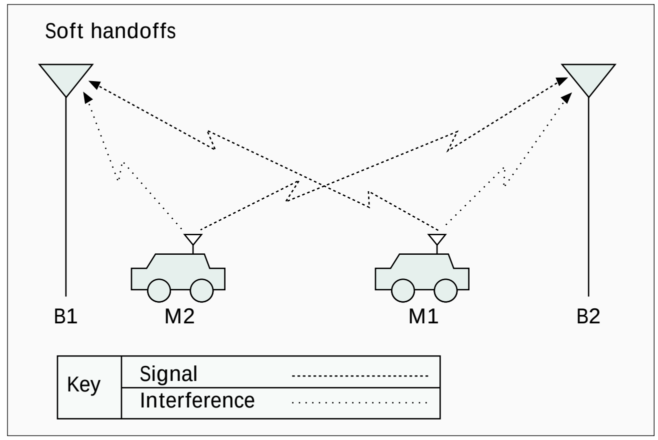


Figure 2. *The power control problem.*

Iteration number	At B1		At B2		Resulting changes	
	Sig. $S_{1,1}$	Int. $S_{2,1}$	Sig. $S_{2,2}$	Int. $S_{1,2}$	ε_1	ε_2
1	5	7	4	6	1	1
2	6	8	5	7	1	1
3	7	9	6	8	1	1
4	8	10	7	9	1	1

Table 1. *Numerical example of the positive feedback loop problem in power control with hard handoffs.*

designers have to determine whether the advantages outweigh the disadvantages for their particular system.

Quantitative Results on Soft Handoff

Only in recent years have attempts been made to quantify the benefits obtainable with soft handoff, as opposed to merely stating them on qualitative reasoning. Most published research results may be divided into two broad categorie, one focusing on fade margin improvement and the other on uplink capacity increase. Both types of analysis are described in this section.

Fade Margin Improvement

The outage probability in a system is defined as the probability of dropping a call before it is terminated by either party, and is denoted by P_{out}. This quantity is a function of distance from the base station r, the minimum acceptable received signal power $S_{r,\min}$, the path loss exponent μ (the power of the signal decays as $r^{-\mu}$), the shadow fading component ζ, the availability of macroscopic diversity, and the transmitted signal power S_t. A system's outage probability must satisfy $P_{\text{out}}(r_0) \leq P_{\text{out,max}}$, where r_0 is the radius of the service area within which a minimum quality of service (QoS) — of which P_{out} is a parameter — is guaranteed. The smallest S_t required to meet the P_{out} requirement is the parameter of interest in analyses of soft handoff algorithms focusing on fade margins, a term to be defined shortly.

For the rest of this section, all signal powers are in dB and r is normalized to the nominal cell radius, so the distance from base station to the cell edge is $r = 1$, and the distance between adjacent base stations is $r = 2$. For convenience, it will sometimes also be assumed that if a user is a distance r_1 from base station 1, it will be a distance $r_2 = 2 - r_1$ from adjacent base station 2 (i.e., approximately collinear with the base stations). Handoff execution delay is represented by d, and hysteresis by Δ (dB). Numerical subscripts on each variable generally refer to base stations, with "1'' representing the original serving base station.

Let $S_r(r)$ be the power of the strongest signal received by the mobile from the base stations in the active set when it is distance r away from the strongest base station. Then it is required that

$$P_{\text{out}}(r_0) = P[S_r(r_0) \leq S_{r,\min}] \leq P_{\text{out,max}}.$$

However, $S_r(r_0) = S_t - 10\mu \log_{10} r_0 - \zeta(r_0)$, and its only random component is $\zeta(r_0)$, which is assumed to be normally distributed with a variance of σ^2 and a mean of zero.[5]

Defining the fade margin as $\gamma(r) = E[S_r(r)] - S_{r,\min}$, where $E[S_r(r)] = S_t - 10\mu\log_{10}r$, the outage probability requirement is equivalent to

$$P[\zeta(r_0) \geq \gamma(r_0)] \leq P_{\text{out,max}}. \quad (1)$$

For given $P_{\text{out,max}}$, r_0 and σ^2, the minimum value of $\gamma(r_0)$ which satisfies this inequality is known as the system fade margin, γ_0.

The minimum transmitted signal power S_t is directly related to the fade margin since

$$S_t = \gamma(r_0) + S_{r,\min} + 10\mu\log_{10}r_0; \quad (2)$$

thus, smaller fade margins mean smaller required transmitted power at the mobile handset.

[5] *Because ζ has a p.d.f. symmetric about 0, it is also correct to write $S_r(r) = S_t - 10\mu\log r + \zeta(r)$. However, $-\zeta$ was chosen for notational conformity with papers in the literature such as [13, 14].*

It has been argued that the system fade margin for hard handoff is larger than the corresponding value for soft handoff in an identical environment with identical service requirements, and therefore using soft handoff leads to performance improvement. It will be discussed later how this fade margin gain can be used for system improvement.

It is assumed that signal propagation reciprocity holds. For consistency, just the downlink is examined, bearing in mind that the same holds for the uplink too. As for power control, it is assumed that in the "worst-case'' scenarios typically examined (i.e., at the edges of acceptable coverage areas), transmitted power is as high as it can be, and link quality cannot be further enhanced by adjustments in transmitted power. Depending on the choices of r_0, type of diversity combining, P_{out} expression, and method of calculation of P_{out}, various differences in system fade margin between hard and soft handoff systems have been reported in the literature, ranging from as low as 1 or 2 dB to as high as 8 dB.

Handoff delay can be divided into handoff decision delay and handoff execution delay, where handoff decision delay is the time between when the user should hand off to when the decision is made to hand off, and handoff execution delay is the time between when the decision is made to hand off and when the corresponding handoff is completed. Most of the published results consider the difference in handoff decision delay between hard and soft handoff to be one of the reasons that less fade margin is required for soft handoffs (the greater the hysteresis margin, the greater the handoff decision delay), as will be seen in the following section. Handoff execution delay, d, on the other hand, is often assumed to be small and negligible, or at least of comparable order of magnitude for both hard and soft handoffs. An exception is [15], which considers hard handoff to suffer from both a greater decision delay and a greater execution delay than soft handoff, and finds the fade margins accordingly. This will also be explained.

Examples of Fade Margin Analysis — Some examples of approaches that have been taken to calculate fade margins for hard and soft handoffs are given below. These demonstrate how different assumptions might lead to different results.

Macroscopic Diversity in Soft Handoff, but Not in Hard Handoff — There is a choice of base stations at all times with soft handoffs, so the best one can be picked (macroscopic selection diversity), whereas this choice does not exist for hard handoffs. Hysteresis and its associated delays preclude counting on other base stations to provide macroscopic diversity coverage for users, because it could take too long to hand off to those base stations, increasing the probability that the call will be dropped. One design philosophy might therefore be to rely on coverage from a single serving base station, at least up to the nominal cell boundary, and thus $r_0 = 1$. Then γ_0 becomes $\gamma(1)$. To further differentiate between the fade margins required in hard and soft handoffs, they will be denoted by $\gamma_{h,0}$ and $\gamma_{s,0}$, respectively.

With these definitions, $P_{\text{out,max}}$ for hard handoffs is

$$P_{\text{out,max}} = P[S_{r,1}(1) - S_{r,\min} \leq 0] = P[\gamma_{h,0} \leq \zeta_1],$$

whereas for soft handoffs, again using $r_0 = 1$ but with selection diversity,

$$\begin{aligned} P_{\text{out,max}} &= P[S_{r,1}(1) - S_{r,\min} \leq 0, S_{r,2}(1) - S_{r,\min} \leq 0] \\ &= P[\gamma_{s,0} < \zeta_1, \gamma_{s,0} < \zeta_2]. \end{aligned} \quad (3)$$

In order to compare the hard and soft handoff margins, it is necessary to compute Eq. 3. With $\sigma = 8$ and $P_{\text{out,max}} = 0.1$ as in [14], $\gamma_{h,0} = 10.3$ dB. If independence of the shadow fading terms is assumed, it can be shown that $\gamma_{s,0} = 3.8$ dB for a fade margin difference of 6.5 dB. If, however, the shadow fading terms are not independent, but have a cross-correlation given by

$$\frac{E(\zeta_i \zeta_j)}{\sigma^2} = a^2, \tag{4}$$

then letting $a^2 = 0.5$, we have $\gamma_{s,0} = 6.2$ dB [14], which yields a fade margin difference of 4.1 dB.

No Macroscopic Diversity in Hard Handoffs, and Hysteresis Requires Extra Margin — It may also be argued that because of hysteresis and the associated delays in hard handoffs, coverage from the original serving base station must extend even beyond the boundaries of a cell, so r_0 is larger than unity. For example, if the ideal handoff point is at the nominal cell boundary but the handoff is delayed because of hysteresis, the affected user goes outside the cell boundary and faces increased outage probability while waiting to hand off. In this case, $\gamma_{s,0}$ is unchanged, and $\gamma_{h,0}$ changes as follows:

$$\begin{aligned} P_{\text{out,max}} &= P[S_{r,1}(r_0) - S_{r,\min} \le 0] \\ &= P[\gamma_{h,0} - 10\mu \log r_0 \le \zeta_1] \end{aligned} \tag{5}$$

For example, let $P_{\text{out,max}} = 0.1$, $\mu = 4$ and $\sigma = 8$ dB, with $r_0 = 1.2$ in Eq. 5; then $\gamma_{h,0} = 13.5$ dB. With $a = 1/\sqrt{2}$ in Eq. 4, the corresponding margin for soft handoff is $\gamma_{s,0} = 6.2$ dB. There is a 6–8 dB difference in margins, corresponding to a cell area increase of 3–4 dB.

In this type of analysis, first introduced in [14], the two major differences in the calculation of fade margin for hard and soft handoffs are:

- Hard handoff systems are based on single base station coverage, whereas soft handoff systems are based on coverage when some form of macroscopic diversity combining is used.
- Hard handoff systems must also account for delay caused by hysteresis, so the single base station coverage must extend beyond the nominal cell boundaries, whereas soft handoff systems use no hysteresis and need not be designed for extended base station coverage.

This worst-case look at hard handoff margins puts soft handoff in its most favorable light. Actually, hard handoff systems can and do enjoy some of the benefits of macroscopic diversity, albeit not as much as soft handoff systems due to the delay in handoff decision. A way of modeling this in fade margin calculations is considered next.

Restricted Macroscopic Diversity for Hard Handoff — Consider a system in which the mobile is within range of two base stations at most, and are at distances r_1 and r_2 from them, with base station 1 being the current serving base station. Allowing for a hysteresis margin of Δ dB, this system calls for a handoff whenever the signal from base station 2 is stronger than that from base station 1 by Δ dB. The outage probability is then given by

$$P_{\text{out}}(r_1, r_2) = \begin{cases} P[S_{r,1} - S_{r,\min} \le 0] & \text{if } S_{r,2} - S_{r,1} \le \Delta \\ P[S_{r,2} - S_{r,\min} \le 0] & \text{if } S_{r,2} - S_{r,1} > \Delta \end{cases},$$

where $S_{r,1}$ and $S_{r,2}$ are understood to be evaluated at distances r_1 and r_2 from base stations 1 and 2, respectively.

The system fade margin, which is evaluated at distances $r_{0,1}$ and $r_{0,2}$ from the two base stations, is therefore

$$P_{\text{out,max}} = \begin{cases} P[\gamma_0 \le \varsigma_1] & \text{if } S_{r,2}(r_{0,2}) - S_{r,1}(r_{0,1}) \le \Delta \\ P[\gamma_0 \le \varsigma_2] & \text{if } S_{r,2}(r_{0,2}) - S_{r,1}(r_{0,1}) > \Delta \end{cases}. \tag{6}$$

Hysteresis is taken into account by not allowing a better signal to be used unless it is better than the one to the serving base station by an amount at least equal to the hysteresis margin. The fade margins for both hard and soft handoffs can be found using Eq. 6: for soft handoff Δ is set to 0, while for hard handoff Δ is simply the handoff hysteresis margin. Therefore, $\gamma_{h,0}$ and $\gamma_{s,0}$ can be found and compared.

This is the method employed in [13], but just for $r_{0,1} = r_{0,2} = 1$. In obtaining numerical results, [13] uses $a^2 = 1/2$, $\Delta = 8$ for hard handoff, $\sigma = 8$ and $\mu = 4$. Noteworthy is their find of only a 2 dB difference in required margins for 90 percent reliability (a criterion also used in [14]), as well as 96 and 99 percent reliability.

The analysis methods that have been presented so far have approached the problem in a particular way. Coverage is examined in a very statistical sense, and assumptions are made about user distributions on the average. What is considered is a system with a set of users who need a certain quality of coverage up to some distance from the base station. Another valid style of analysis is to look at the probabilities related to each single user moving from one base station to another. This type of approach has the advantage of being able to consider correlations in the signal power levels received by each user as it moves, but has the disadvantage of generally being more complex than the statistical coverage approach.

A Different Approach: Tracking a User — Various statistical quantities, such as probability of being in communication with a given base station, are evaluated based on a user traveling in a straight line between two base stations. Intersample correlation of the shadow fading component, based, for example, on Gudmundson's exponential spatial correlation model of [16], is incorporated. The possibility of multiple handoffs back and forth between the two base stations is allowed to influence the calculations leading to fade margins. Thus, the fade margins for hard handoffs are not as pessimistic as those obtained from the first or second method above.

Doing this type of analysis, the outage probability is found by [15] to be[6]

$$P_{\text{out}} \approx \int_{-\infty}^{-\gamma(r_0)} \pi_1(y)dy + \int_{-\infty}^{-\gamma(r_0)+10\mu\log_{10}((2-r_0)/r_0)} \pi_2(y)dy \tag{7}$$

where $\pi_1(y)$ and $\pi_2(y)$ are expressions involving $Q(.)$ functions, the propagation parameters, and the handoff parameters. The first integral is the probability of outage while connected to base station 1 when it is r_0 away from the first base station, and likewise the second integral is the probability of outage while connected to base station 2 at that time.

The same analysis is used to model both soft and hard handoff, and the difference is only to be found in a difference in the settings of Δ and d. $\Delta = 0$ and $d = 0$ (all delays are normalized to the sampling interval) is equivalent to soft handoff. In [15], soft handoff is compared to hard handoffs with $\Delta = 6$ and $d = 0,2,5$, and also to the hard handoff model of [14], in which an outage happens whenever the signal strength from the first base station drops below the required level, regardless of the signal strength from the second base station. The comparisons are performed with $r_0 = 1$ (at the nominal cell boundaries), for uncorrelated and 0.5-correlated shadow fading. Difference in margin for soft and hard handoffs is less (by a few dB) than that reported in [14].

Why might d vary for hard handoffs? It has to do with system design. For example, part of the handoff execution delay comes from the time it takes delivery of the handoff message to occur, so it is important to know how often handoff mes-

[6] *We have omitted the details of the derivation, and refer the reader to [15].*

sages are sent or active sets are updated.[7] In a typical CDMA system, the active set updates occur every 100 ms [13], which is about how long it takes for MCHO to occur. In hard handoff systems that use MAHO instead of MCHO, however, messaging occurs only on the order of every 0.5 s (e.g., in GSM). Hence $d > 0$ would more accurately model such a system. While this type of analysis, such as in [15], is arguably more accurate than that in [13, 14], the computation of fade margins is harder because it involves nontrivial iterations or the evaluation by numerical integration of equations like Eq. 7.

Simulations — Chopra *et al.* have performed simulations of soft handoff to determine the cell coverage extension due to soft handoff [13]. Their results show the difference between CDMA handoffs and GSM handoffs. They model the sampling timing of pilot strength, the timing of the active, candidate, neighbor, and remaining set updates, as well as some of the thresholds. They assume a lightly loaded system in computing I_0, the total received power spectral density, assuming zero loading or no interference. The GSM simulation also closely follows the GSM specifications. However, the assumption is made that handoff is possible as early as half a second after the last handoff, which might not always be possible if there is too much network delay in the handoff execution. It is also assumed that the user measurements can be transferred to the serving base station without corruption (since GSM uses MAHO, this is a potential problem). Simplified simulations suggest that an additional margin of about 1 dB might be needed to account for this [13], although the threshold at which signaling breaks down is normally lower than that at which voice breaks down. After running simulations in different conditions with varying propagation parameters, it is concluded that the difference in required fade margin for IS-95 CDMA and GSM is about 3 dB, slightly higher than what their rough analysis indicates (2 dB), and slightly lower than the results from the analysis of [15].

Under what circumstances will these fade margin advantages of soft handoff be useful? They can be translated to a downlink benefit, that is, smaller base station transmitter power on the downlink. It is unclear whether the downlink capacity or uplink capacity is more critical in CDMA systems. Several papers have been written on downlink power control/capacity-related issues (e.g., [17]). However, it is generally believed that the uplink is more critical. Thus it might be preferable to view fade margin gains in terms of cell coverage extensions instead of downlink gains. The cell coverage gains are generally more applicable to a noise-limited environment or lightly loaded system. They might be helpful in a rural/suburban area. In a more heavily loaded system, interference limits the system, and bigger cells may be undesirable. Instead, it may be desired to keep the cell sizes the same, or even to have smaller cells (microcells). Hence, one wishes to examine how soft handoff affects the relative interference levels.

Interference and Uplink Capacity

The key to the claimed soft handoff interference advantage is that hysteresis margins are used in hard and not soft handoffs. Hysteresis results in a delay in switching to the best base station. However, interference caused, and experienced, by the user is generally larger during the time of the delay when the user is in communication with an inferior base station. So there should be less overall interference in soft handoff systems than in hard handoff systems. However, this interference reduction is counter-balanced by an increase in downlink interference caused by the simultaneous redundant downlink transmission from several base stations during soft handoff, whereas only one would be transmitting if soft handoff were not in use. The resultant effect on the overall downlink interference situation is therefore unclear and is probably quite sensitive to various parameter settings.

It may be argued that uplink interference is more important in affecting overall capacity. One way of analyzing it will be presented. There are two categories of interference in a CDMA system, same-cell and other-cell interference. Other-cell interference will be considered first. Since power control is used, S_t is proportional to the propagation loss between the user and the base station. Assuming reciprocity of radio links and normalizing so the received uplink power is 1, a user's S_t simply equals its propagation loss. Hence, using the notation introduced earlier,

$$S_t = \text{path loss} = 10\mu\log r_1 + \zeta_1,$$

where r_1 is the distance to the serving base station. Letting r_2 represent the user distance from another base station at which its signal is an interference signal, and letting I be the interference power, then

$$\begin{aligned} I &= S_t - \text{path loss to the other base station} \\ &= 10\mu\log r_1 + \zeta_1 - 10\mu\log r_2 - \zeta_2. \end{aligned} \tag{8}$$

Assuming interference adds in power, total other-cell interference is

$$\sum_{\text{all interfering users}} 10^{I_j/10} \tag{9}$$

where I_j is the interference from the jth interferer, and the interference from each interferer is given by Eq. 8. Equation 9 is applicable for both hard and soft handoffs in the presence of perfect power control. Variations of Eq. 9 appear in papers looking at interference in power-controlled systems, such as [4, 14].

A modified, continuous integral version of Eq. 9 is used in [14] to analyze the increased uplink capacity of a power-controlled soft handoff system over a hypothetical power-controlled CDMA system with hard handoff.[8] A uniform density of users, κ per unit area, is assumed, and the term in Eq. 9 is multiplied by κ and integrated over the regions where the interferers are located. The difference between hard and soft handoff is then to be found in the differences in the assumed locations of the interferers.

As for same-cell interference, by the same normalization defined above, each user's interference is 1 and the total same-cell interference is just the number of other users in the same cell.

Assumptions need to be made in order to evaluate Eq. 9 or a variation of it. For example, in [14] it is assumed that:

- In a system with hard handoff, all users remain within the strict cell boundaries at all times. Correspondingly, all users outside of a given cell are served by another base station and are interferers to the given cell.
- In a system with soft handoff, certain regions are chosen within which the user is assumed to possibly be in soft handoff (the reader is referred to [14] for illustrations of these regions for the different possible cases).

Because of shadow fading effects, these assumptions are not always valid. However, if σ is small, they are reasonably

[7] *It could be said that these delays contribute to the decision delay, but we conceptualize that the decision is already made once the measurements are made and processed, and the decision criteria are satisfied, whether or not the system has formally "decided" to hand off at that time.*

[8] *We have already pointed out the inherent instability in such a system which uses power control and hard handoff. This instability is not noted in [14].*

accurate for approximating what happens in real systems. Using these assumptions, results are found showing that the interference does indeed go down when soft handoff is used. It is concluded that soft handoff increases uplink capacity by a factor of 2 to 2.5.

It is difficult to find a good and simple analysis of other-cell interference in the uplink. Despite the nature of the assumptions made in [14], it is a relatively good analysis, with a good balance between a reasonable model and complex analysis. For example, the analysis of uplink interference in [18], while simple, does not even consider shadow fading, whereas works like [19] tend to get very complicated. As mentioned earlier, the downlink interference situation is complicated. If uplink and downlink channels are on the same frequency bands, this also adds to the uplink interference. The interference levels are affected by the choice of handoff parameters, so the trade-offs and parameter optimization must be examined.

Parameter Optimization

Soft handoff is more complex to implement than hard handoff, one reason being that finding optimal settings for various soft handoff parameters is difficult. We will introduce some of the parameters which affect the performance of soft handoff, and then discuss how these might relate to some performance indicators.

Parameters

The parameters include:

- Add threshold — The threshold for membership in the active set.
- Drop threshold — The threshold for dropping of membership in the active set (see T_{drop}).
- T_{drop} — In order for an active set member to be dropped from the active set, the signal level of that member must be below the drop threshold for a period of time at least equal to T_{drop}.
- Soft handoff window (SHW) — The difference between the add and drop thresholds (after [21]). It can also be seen as an indication of how long a soft handoff will take on average. The larger the window, the longer the average soft handoff.
- The ratio a — Defined (after [22]) as

$$a = \frac{\text{area of handoff region}}{\text{area of a cell}}$$

which can be varied, for example, by adjusting the distance between base stations.

Performance Indicators

The performance indicators are of two types.

Link Quality Indicators —

- Average downlink E_c/I_0 for a given system load
- Average uplink E_c/I_0 for a given system load

Good Resource Allocation Indicators —

- T_C, carried traffic: a dimensionless quantity measured in Erlangs, the expected number of channels occupied in each cell.
- P_B, the (new call) blocking probability: probability that a new call (i.e., not a handoff call) is blocked.
- P_{CB}, probability that all channels are occupied in the new cell in a handoff.
- NO_{BS}, expected number of base stations in the active set (after [24]): note that $NO_{BS} = 1$ for hard handoff and $NO_{BS} \rightarrow 2$ as handoffs get softer and softer (i.e., SHW increases and/or the add/drop thresholds decrease). This is a measure of system resource utilization.
- TRE, trunking resource efficiency (after [22]): expected system "efficiency'', where efficiency is 1/(size of active set). A complement of NO_{BS}, TRE = 1 for hard handoff and TRE < 1 if soft handoffs are used.
- NO_{update}, expected number of changes in active set (after [24]): a measure of network loading.

Research Results

Research on the trade-offs and parameter settings has been mostly in the form of simulation studies. Some of the published results are compared in Table 2. Wang and Wang note in [4] that little or no overhead exists for "softer handoff," so there is little or no problem with trade-offs for "softer handoff," and the focus should be on soft handoffs. Equation 9 is modified to include factors proportional or inversely proportional to the antenna gains along the directions from the particular interferer to its base station and to the base station at which it is interfering. The modified equation is the basis of the simulations.

Seïté's simulation study [21] uses the large-scale propagation model for microcellular environments of Berg, Bownds, and Lotse [23]. A relatively high lognormal standard deviation of 8 dB is used (it is normally around 3–4 dB for microcellular environments [23]). Unlike the other simulation studies discussed in this article, the bit error rate (BER) is also simulated. Perfect synchronization is assumed between spreading sequences in the ideal Rake receiver simulated. The channels have five paths, but only three Rake branches.

Su and Chen [22] analyze resource allocation where each cell has C channels available for usage, of which C_h are reserved for handoff calls. In a soft handoff, one channel from each cell involved is used. These channels are never lent to other cells.[9] Each cell is divided into two regions, the normal region and the handoff region, where the handoff region is a fixed portion of any cell such that whenever a user enters a handoff region, it tries to find a channel at another cell to which to hand off. Soft handoff is restricted to happen only in the handoff regions. System performance is analyzed as a birth-death process with Markov chains and exponential arrivals/departures. Handoff requests are placed in a handoff request retry queue while waiting for an available channel, and rejected if the queue is full, or if they are on the queue for too long without handing off successfully.

Zhang and Holtzman [24] provide analytical tools for analyzing the performance trade-offs of soft handoff, following the basic framework of [26, 27]. Their model considers just two base stations and a user moving between them in a straight line, without any interference. Doing comparable analysis for multiple base stations, more complicated traffic patterns, and with interference is prohibitively complicated. Differences between their model and IS-95 include the use of pilot signal strength rather than E_c/I_0, and the use of just the Active set and the complement of the active set.

Asawa and Stark [20] believe that add and drop thresholds are not powerful enough decision criteria for active set maintenance. Hence, the problem is formulated as a reward/cost

[9] Unlike the case for other cellular multiple access schemes, channel borrowing does not make sense for CDMA systems if the frequency reuse interval is 1, which it often is (many channel borrowing schemes and variants have been proposed and discussed in the literature, e.g., [25]).

	Asawa, Stark [20]	Chopra *et al.* [13]	Seïté [21]	Su, Chen [22]	Wang, Wang [4]
Power control	N/A	N/A	Imperfect	N/A	Perfect
Voice activity detection	N/A	N/A	Yes	N/A	Yes
Antenna sectorization	No	No	Perfect	No	Imperfect
Softer handoff	No	No	No	No	Yes
Active, candidate, neighbor, remainder sets	No	Yes	No	No	No
Add/drop thresholds	No[1]	Yes	Yes	N/A	Yes
More than two base stations in system	Yes	Yes	Yes	Yes	No
More than two base stations possible for soft handoffs	Yes	Yes	Yes	No	No
Network resources considered	No	No	No	Yes	No
Propagation environment	Cellular	Cellular	Microcellular	Cellular	Cellular
Path loss exponent	3	3.6 and 4	2–4[2]	N/A	4
Shadow fading variance	6	6.5 and 8	8	N/A	8
Small-scale fading	No	Yes	Yes	N/A	No

Notes: [1] The decision criteria in [20] are different from the add/drop thresholds used in IS-95, and are claimed to be better than the IS-95 criteria.

[2] It is approximately free-space loss close to the base station and fourth-power loss after a "breakpoint." This is typical of microcellular environments [23].

Table 2. *Comparison of soft handoff simulation studies.*

stochastic optimization problem where the reward is associated with good signal and the cost with soft handoff overhead and so on. This is an extension of Asawa and Stark's hard handoff optimization model [28]. The system model consists of N base stations and one user which is allowed to have up to all N base stations in its active set.

The nature of today's cellular systems is such that optimization over many parameters is very difficult. Therefore, optimization is performed over a small number of parameters. If the assumption is made that link quality is what limits capacity (in a CDMA system, this generally means that the system is interference-limited, not noise-limited) and that users are quite evenly distributed throughout the system, the focus might be on optimizing link quality. If, however, we assume a system with plenty of radio coverage margin but maybe an uneven user distribution and/or limited network resources, the focus might be on optimizing resource allocation.

Link Quality — From the perspective of link quality, there is an uplink-downlink interference trade-off, alluded to previously. Increasing SHW (by reducing the add threshold or drop threshold or increasing T_{drop}) increases the downlink interference because the time when several base stations are transmitting the same downlink signals increases. However, decreasing SHW increases the uplink interference because the uplink interference reduction properties of soft handoff are reduced. This is verified in [21], which reports that downlink E_c/I_0 is better for smaller SHWs, while uplink E_c/I_0 is better for larger SHWs. For the downlink, system performance is worse than without soft handoff when SHW exceeds 7 dB. The values of the add and drop thresholds used are not reported, nor is how performance varies if both thresholds are increased or decreased by the same amount, keeping SHW fixed.

On the other hand, [4] only considers SHW = 0, where the add threshold is equal to the drop threshold. The thresholds are specified with respect to the signal level from the minimum loss base station, so they are in negative dB. It is concluded that a soft handoff threshold (for both adding and dropping) of between –3 dB and –5 dB makes the best trade-off. In general, for a given SHW, jointly increasing the add/drop thresholds together will reduce the percentage of time spent in soft handoff, while reducing the add/drop thresholds together will increase the percentage of time spent in soft handoff, with the same effects on link quality as mentioned earlier. For fixed cell sizes, increasing SHW (or reducing add/drop thresholds together) also increases a, and decreasing SHW (or increasing add/drop thresholds together) also decreases a. But changing a has an effect on resource allocation too, and this will be examined in the next subsection.

The reward/cost stochastic optimization model [20] associates a higher reward with a diversity-combined signal than with a signal to/from only one base station. There are also costs associated with being in a soft handoff, such as a penalty for using more than one channel and penalties for signaling. Based on the chosen rewards and costs, it is found that this optimization method works better than one using add/drop thresholds. However, it involves more computation.

Resource Allocation — From the perspective of resource allocation, the main trade-off is between efficient utilization of limited network resources and improved performance in terms of indicators like P_B. As a increases (e.g., by changing SHW or spacing between base stations), it is expected that P_{CB}, P_B, and NO_{update} improve by decreasing. However, NO_{BS} and TRE deteriorate by increasing and decreasing, respectively. These issues are investigated by varying a, for two cases [22]:

- N_c is allowed to increase as a increases, going along with the increase in uplink capacity according to [14]. N_c is the number of channels available per cell.
- N_c remains fixed no matter what a is.

For both cases, it is verified that increasing a gives a lower P_B and P_{CB}, and these improvements are greater in the case that N_c increases with a. However, it is observed that TRE is lower for larger a. As for T_c, the simulations show that when N_c is fixed, T_c drops as a increases, reflecting the taking away of needed channels from the fixed N_c channels by the soft handoff process. However, since N_c should increase with a because of the uplink capacity increase effect, the simulations have found that this counterbalances the T_c-decreasing effect, especially for higher new-call arrival rates. Changes in a on the order of 0.2 can lead to changes by factors of 10 or more for these performance indicators. The ideal a is unknown. The effects of varying T_{drop} may also be investigated [24]. A large T_{drop} is expected to have an effect similar to a large SHW or large a. The analytical model is checked by simulations, which show a close fit. There is a parametric trade-off curve between $\text{NO}_{\text{update}}$ and NO_{BS}, where T_{drop} is the parameter. By increasing T_{drop}, $\text{NO}_{\text{update}}$ decreases significantly, while NO_{BS} increases slowly.

It appears that the reward/cost stochastic optimization method of [20] has the flexibility to model both the link quality and resource allocation factors in a unified framework and could perhaps be used to further investigate the various trade-offs.

The trade-offs involved in the setting of soft handoff parameters are still not very well understood. Parameter optimization for soft handoff is difficult, but a better understanding of the issues is crucial because some performance indicators are very sensitive to some of the parameters. Moreover, the settings of the parameters are dependent on many factors, including the overall system design, the propagation environment, and the traffic and calling patterns. (The parameters are typically fine-tuned by cellular system operators based on their particular situation.) There is much potential for future research in this area.

Discussion and Conclusion

Soft handoff is an intriguing technology. It promises better performance than hard handoff, through the exploitation of macroscopic diversity and not having to use hysteresis margins. Attempts have been made to substantiate these claims by augmenting the qualitative arguments with quantitative results. Several papers have shown that soft handoff has some fade margin gain over hard handoff, and that there is a possible uplink capacity increase in power-controlled systems with soft handoff over power-controlled systems with hard handoff. However, these are limited comparisons which do not adequately balance the main advantages and disadvantages of soft handoff.

Because of the complexity of soft handoff, another main area of research on soft handoff is on the different trade-offs involved in the handoff parameter settings. Most of the research in this area, with the notable exception of [24], is in the form of simulations. There are many variables and performance indicators involved, and it is often unclear what is optimal for a particular system. However, studies have indicated that system performance may be very sensitive to the settings of some parameters, so a deeper understanding of the trade-offs and optimal parameter settings is essential to the successful implementation of soft handoff.

Future Directions

Many things about soft handoff are still not well understood. The quantitative trade-offs between the various advantages and disadvantages of soft handoff need to be further investigated, as do the parameter settings. Also, future studies could look at forms of macroscopic diversity combining other than selection combining. [15] uses the same analysis for both soft and hard handoffs, taking $\Delta = 0$, $k = 0$ as being equivalent to soft handoff. The resulting analysis, as well as that of others like [14], assumes the use of selection diversity combining. The uplink in IS-95 CDMA does use selection diversity combining, but if more sophisticated combining schemes such as that in [29] are used, the analysis should also be modified.

Future studies could make more careful assumptions, to enhance the understanding of the issues in practical systems. For example, several papers (e.g., [15, 24]) assume that samples (of signal strength, etc.) are spatial samples, taken at fixed spatial intervals rather than at fixed time intervals. However, such samples are difficult to obtain in today's cellular systems because of varying user speeds. Possible solutions to this problem include Kennemann's hidden Markov model [30] and Austin and Stüber's adaptive sampling scheme [31].

The notion of a gentler, softer type of handoff than hard handoff has aroused much interest in the wireless communications research community. The issues are beginning to be studied quantitatively, and we enthusiastically await further developments.

Acknowledgments

The authors thank Hiroki Sugimoto of Oki Techno Centre (Singapore) Pte. Ltd. for providing information on soft handoffs in the IS-95 standard. The authors also thank the reviewers for their helpful comments. The first author thanks Prof. Donald Cox of Stanford University for informative discussions.

References

[1] *Bell Sys. Tech. J.*, Special Issue on Advanced Mobile Phone Service, Jan. 1979.
[2] J. E. Padgett, C. G. Günther, and T. Hattori, "Overview of Wireless Personal Communications," *IEEE Commun. Mag.*, vol. 33, no. 1, Jan. 1995, pp. 28–41.
[3] D. J. Goodman, "Trends in Cellular and Cordless Communications, *IEEE Commun. Mag.*, vol. 29, no. 6, June 1991, pp. 31–40.
[4] S.-W. Wang and I. Wang, "Effects of Soft Handoff, Frequency Reuse and Non-Ideal Antenna Sectorization on CDMA System Capacity," *Proc. IEEE VTC*, Secaucus, NJ, May 1993, pp. 850–54.
[5] TIA, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System," July 1993.
[6] M. Honig, U. Madhow, and S. Verdu, "Blind Adaptive Multiuser Detection," *IEEE Trans. Info. Theory*, vol. 41,no. 4, July 1995, pp. 944–60.
[7] R. Lupas and S. Verdu, "Near-Far Resistance of Multiuser Detectors in Asynchronous Channels," *IEEE Trans. Commun.*, vol. 38,no. 4, Apr. 1990, pp. 496–508.
[8] S. Verdu, "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels," *IEEE Trans. Info. Theory*, vol. IT-32, no. 1, Jan. 1986, pp. 85–96.
[9] S. Parkvall, E. Strom, and B. Ottersten, "The Impact of Timing Errors on the Performance of Linear DS-CDMA Receivers," *IEEE JSAC*, vol. 14, no. 8, Oct. 1996, pp. 1660–68.
[10] F.-C. Zheng and S. Barton, "On the Performance of Near-Far Resistant CDMA Detectors in the Presence of Synchronization Errors," *IEEE Trans. Commun.*, vol. 43,no. 12, Dec. 1995, pp. 3037–45.
[11] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Upper Saddle River, NJ: Prentice Hall, 1995.
[12] W. C. Jakes, Ed., *Microwave Mobile Communications*, New York: Wiley, 1974; republished by IEEE Press, 1994.
[13] M. Chopra, K. Rohani, and J. Reed, "Analysis of CDMA Range Extension Due to Soft Handoff," *Proc. IEEE VTC*, Chicago, IL, July 1995, pp. 917–21.
[14] A. J. Viterbi *et al.*, "Soft Handoff Extends CDMA Cell Coverage and Increases Reverse Link Capacity," *IEEE JSAC*, vol. 12, no. 8, Oct. 1994, pp. 1281–87.
[15] K. M. Rege *et al.*, "Analysis of Fade Margins for Soft and Hard Handoffs," *Proc. IEEE Int'l. Symp. Pers., Indoor, and Mobile Radio Commun.*, Toronto, Canada, Sept. 1995, pp. 829–35.
[16] M. Gudmundson, "Correlation Model for Shadow Fading in Mobile Radio Systems," *Elect. Lett.*, vol. 27, no. 23, Nov. 1991, pp. 2145–46.
[17] M. Ismail and T. Rahman, "Forward-Link Frequency Reuse Efficiency of Power Controlled CDMA Cellular System," *IEEE Int'l. Symp. Pers., Indoor, and Mobile Radio Commun.*, Toronto, Canada, Sept. 1995, pp. 441–45.
[18] M.-S. Kwok and H.-S. Wang, "Adjacent Cell Interference Analysis of Reverse-Link in CDMA Cellular Radio Systems," *IEEE Int'l Symp. Pers., Indoor, and Mobile Radio Commun.*, Toronto, Canada, Sept. 1995, pp. 446–50.
[19] J. Zou and V. K. Bhargava, "On Soft Handoff, Erlang Capacity and Service Quality of a CDMA Cellular System: Reverse Link Analysis," *IEEE Int'l. Symp. Pers., Indoor, and Mobile Radio Commun.,* Toronto, Canada, Sept. 1995, pp. 603–7.

[20] M. Asawa and W. E. Stark, "Optimal Scheduling of Soft Handoffs in DS/CDMA Communication Systems," *Proc. Annual Joint Conf. IEEE Comp. and Commun. Socs.*, Boston, MA, Apr. 1995, pp. 105–12.
[21] P. Seïté, "Soft Handoff in a DS-CDMA Microcellular Network," *IEEE VTC*, Stockholm, Sweden, June 1994, pp. 530–34.
[22] S. L. Su and J. Y. Chen, "Performance Analysis of Soft Handoff in Cellular Networks," *IEEE Int'l. Symp. Pers., Indoor, and Mobile Radio Commun.*, Toronto, Canada, Sept. 1995, pp. 587–91.
[23] J.-E. Berg, R. Bownds, and F. Lotse, "Path Loss and Fading Models for Microcells at 900 Mhz," *Proc. IEEE VTC*, Denver, May 1992, pp. 666–71.
[24] N. Zhang and J.M. Holtzman, "Analysis of a CDMA Soft Handoff Algorithm," *Proc. IEEE Int'l. Symp. Pers., Indoor, and Mobile Radio Commun.*, Toronto, Canada, Sept. 1995, pp. 819–23.
[25] T.-T. Su, P.-C. Huang, and C.-J. Chang, "A Channel Borrowing Protection Scheme for Handoffs in a Cellular-Based PCS System," *Proc. IEEE Int'l. Conf. Universal Pers. Commun.*, Tokyo, Japan, Nov. 1995, pp. 657–61.
[26] R. Vijayan and J. M. Holtzman, "A Model for Analyzing Handoff Algorithms," *IEEE Trans. Vehic. Tech.*, vol. 43, no. 3, Aug. 1993, pp. 351–56.
[27] N. Zhang and J. M. Holtzman, "Analysis of Handoff Algorithms Using Both Absolute and Relative Measurements," *Proc. IEEE VTC*, Stockholm, Sweden, June 1994, pp. 82–86.
[28] M. Asawa and W. E. Stark, "A Framework for Optimal Scheduling of Handoffs in Wireless Networks," *Proc. IEEE Global Telecommun. Conf. and Exhibition*, San Francisco, CA, Nov. 1994, pp. 1669–73.
[29] W. Papen, "Improved Soft Handoff and Macro-Diversity for Mobile Radio," *Proc. ICC '95*, Seattle, WA, June 1995, pp. 1828–33.
[30] O. Kennemann, "Pattern Recognition by Hidden Markov Models for Supporting Handover Decisions in the GSM System," *Proc. 6th Nordic Sem. Digital Mobile Radio Commun.*, June 1994, pp. 195–202.
[31] M. D. Austin and G. L. Stüber, "Velocity Adaptive Handoff Algorithms for Microcellular Systems, *IEEE Trans. Vehic. Tech.*, vol.43, no.3, pt. 1, Aug. 1994, pp. 549–61.
[32] R. C. Bernhardt, "Macroscopic Diversity in Frequency Reuse Radio Systems," *IEEE JSAC*, vol. SAC-5, no. 5, June 1987, pp. 862–70.
[33] G. P. Pollini, "Trends in Handover Design," *IEEE Commun. Mag.*, vol. 34, no. 3, Mar. 1996, pp. 82–90.

Biographies

K. DANIEL WONG (daniel@isl.stanford.edu) received a B.S.E. degree in electrical engineering (with highest honors) from Princeton University, New Jersey, in 1992, and an M.S. degree in electrical engineering from Stanford University, California, in 1994. He is currently a Ph.D. candidate in electrical engineering working on handoff algorithms for cellular communications systems at Stanford University. His research interests include handoff algorithms in cellular systems and statistical signal processing in communications applications, especially wireless applications. He is also a member of Tau Beta Pi, Sigma Xi, and Phi Beta Kappa.

TENG JOON LIM received a B.Eng. degree with honors in electrical engineering from the National University of Singapore in 1992, and a Ph.D. degree from Cambridge University, England, in 1995. Since September 1995 he has been a member of technical staff at the Centre for Wireless Communications in Singapore. His current research interests include detection and parameter estimation in multiuser CDMA systems, statistical signal processing applied to mobile communications, handoff algorithms in CDMA cellular systems, and multicarrier modulation.

Appendix A
A Case Study: IS-95 Soft Handoff

This is an example of how soft handoff may be implemented in a system.

Channel Set Maintenance

Users maintain lists of channels, specified by pseudo-noise (PN) offset and frequencies, whose members depend on various threshold criteria. These lists include the active set, the candidate set, the neighbor set, and the remaining set. The active set contains the currently used channel(s) and has more than one member during a soft handoff. The candidate set contains channels which are almost as good as the one(s) in the active set, and it is from the candidate set that new channels are chosen for soft handoff. The neighbor set is the set of channels which do not meet the criteria to be included in the active and candidate sets, but are reasonably strong. The remaining set contains everything else.

Idle Handoff

The user is always associated with a base station, even when idle. After the user unit is turned on, it acquires the pilot signal and timing from the best base station around and goes into a user idle state. It continuously monitors pilot channel signals. So if, while idle, the user detects a sufficiently strong pilot channel signal other than the current base station's, an idle handoff occurs.

In idle state, the paging channel protocol allows a user to communicate with a base station over a paging channel. The paging channel is divided into 80 ms slots. There are two modes possible. One is nonslotted mode, which means the user must monitor all the slots and communication can be in any of them. The other is slotted mode, which is like TDMA of the paging channel frequency, and has advantages such as the fact that the user can be turned off between scheduled slots. A user is required to operate in the nonslotted mode, while performing an idle handoff, until the user has received at least one valid message on the new paging channel; it can then revert to the slotted mode, and the idle handoff can be considered complete.

Regular Handoff

There are three handoff procedures possible in the traffic channel state (regular communication link established). They are:
- CDMA-to-CDMA soft handoff — Between CDMA channels on identical frequencies
- CDMA-to-CDMA hard handoff — Between CDMA channels on different frequencies
- CDMA-to-analog hard handoff — When the handoff is to one of the coexisting analog channels

Both the CDMA-to-CDMA soft and hard handoffs using the same frequency are normally initiated by the user. The user makes the measurements on the pilot channel (each base station continuously transmits a pilot channel for every CDMA frequency it supports; there could be pilot channels at other frequencies), without changing frequencies. The user searches for usable multipath components in ranges of PN offsets (known as search windows) specified by the base station. When a user detects a pilot of sufficient strength that not one of its current downlink channels (and there could be several, in which case the user provides diversity combining), the measurement is sent to the serving base station, which can then make the appropriate decision, perhaps to handoff. The information sent to the base station by the user includes the following.

"Strength of Pilot" — This is computed by adding ratios of received pilot energy per chip to total received spectral density, from at most k usable multipath components, where the number k is the number of correlators implemented by the user for demodulating; that is,

$$\text{strength of pilot} = \left(\frac{E_c}{I_0}\right)_1 + \ldots + \left(\frac{E_c}{I_0}\right)_k$$

Handoff Drop Timer — The user maintains a handoff drop timer for each pilot of the channels it is using and good alternatives. The timer is started when the pilot strength drops below a threshold. The timer is reset when signal strength rises above the threshold.

PN Phase Measurements — These might be used by the base station to make an estimate of propagation delay to the user,

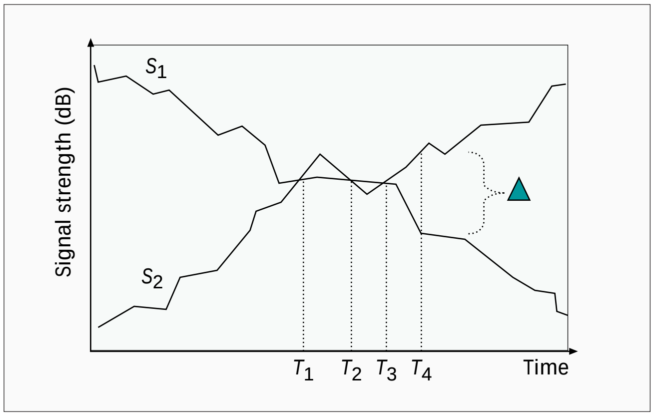


Figure 3. *Hysteresis and the "ping-pong" effect in hard handoffs.*

for faster uplink traffic channel acquisition time (by more intelligent setting of its correlator delays, for example).

Power Control in Soft Handoff

What happens to the various control functions in the system during a soft handoff? What happens to power control is of particular interest. Normally, when not in soft handoff, a user communicates with one base station, and that is the base station which gives it power control instructions, to increase or decrease transmitted power. In soft handoff, one possibility is that the different base stations independently transmit power control instructions, requiring that the user arbitrate between the instructions if they are different. Another possibility is that they all agree on and transmit the same power control instructions.

In IS-95, something between the above-mentioned possibilities is used. All downlink traffic channels associated with pilots in the user's active set carry the same modulation symbols, except for the power control subchannel. Sets of downlink traffic channels carry identical power control information, but each set could be different from the others. The user performs diversity combining on each set (because the information is identical), and looks at all the resulting power control bits obtained from all the sets. If even one of them instructs the user to decrease transmitted power, it obeys and decreases it; otherwise, it increases it. This is because if there is at least one instruction to decrease power, there is at least one base station able to provide coverage.

Appendix B
Hysteresis and the "Ping-Pong" Effect in Hard Handoffs

We consider a hard handoff situation. Let S_1 be the signal strength from the original serving base station B1 and S_2 from B2, to which the user hands off. Figure 3 shows an example of signal strength measurements against time. Suppose that at regular intervals of time, the following handoff decisions occur:

- If $S_2 - S_1 > 0$ and serving base station is B1, then handoff to B2.
- If $S_1 - S_2 > 0$ and serving base station is B2, then handoff to B1.
- Otherwise do not handoff.

Using this handoff algorithm, the user will hand off three times back and forth between B1 and B2. At T_1, the user hands off from B1 to B2, at T_2 from B2 back to B1, and at T_3 from B1 to B2 again, finally remaining with B2. This handing off back and forth several times between two base stations in a relatively short period of time is sometimes known as the "ping-pong'' problem, analogous to the movement of a ping-pong ball between the two ends of the table in a ping-pong game. The problem is that each time a handoff is executed, there is some overhead in the network. Signaling must be done, and varying amounts of authentication, database updates, circuit switching and bridging are performed with each handoff. It is an undesirable use of network resources to go through the whole handoff process more than actually necessary, but that is what happens with the handoff algorithm described above.

In order to reduce the ping-pong' effect, a standard feature of hard handoff algorithms is the incorporation of hysteresis. By this it is meant that the basic algorithm is modified to become:

- If $S_2 - S_1 > \Delta_1$ and serving base station is B1, then handoff to B2.
- If $S_1 - S_2 > \Delta_2$ and serving base station is B2, then handoff to B1.
- Otherwise do not handoff.

Δ_1 and Δ_2 are hysteresis margins, and generally $\Delta_1 = \Delta_2 = \Delta$. Hysteresis allows the system to wait till it is more certain that a handoff should be performed before it does so, and thus reduces the ping-pong' effect. A variation of this algorithm is used in the GSM standard.

Advantages and disadvantages of using hysteresis are both illustrated in Fig. 3. With hysteresis, only one handoff is performed, at time T_4, rather than three handoffs. This shows that handoff is advantageous in reducing the ping-pong effect. However, T_4 is later than T_1, T_2, and T_3. A disadvantage of using hysteresis is that the handoff decision is delayed, and this delay increases with hysteresis margin. Two primary functions performed by a handoff are to exchange a weak, deteriorating link for a stronger one, and to reduce the interference to other base stations caused by a user far from its serving base station (the interference problem is particularly severe in a system with power control). Delay in handoff reduces its effectiveness in performing both functions, increasing the probability that the call will be dropped because of too poor a link, and also reducing the interference reduction efficiency. Example parametric trade-off curves between delay and number of unnecessary handoffs are given in [26], where the hysteresis margin is the parameter.

Soft handoff is an attempt to have both a small hysteresis margin (equal to or close to 0, depending on the threshold settings) without the overhead of the "ping-pong'' effect, or at least less of it. It can do this because in the critical period when the signal strength of two or more base stations is quite similar, it does not have to decide to complete a handoff to one of them, unlike in a hard handoff situation.

Appendix C
Macroscopic Diversity and Soft Handoff

When macroscopic diversity is discussed, it refers to there being more than one base station which can be used (i.e., a diversity of base stations), and is distinguished from microscopic diversity, which is generally on a smaller scale [12, 11]. There are two situations to which we apply the term macroscopic diversity. The first is when two or more base stations are reasonable potential candidates to be the serving base station of a user. Since the user has a choice of base stations, the outage probability at a given distance is reduced, or cell coverage is increased over the single base station cell coverage range for the same outage probability. However, only one base station provides actual service at a given time. The second situation is more specific and only possible in some system designs like CDMA systems: when more than one base station can provide actual service simultaneously, even if they might individually be unable to sustain service to the user. For example, in IS-95 this type of macroscopic diversity is provided on the downlink in a soft handoff situation, when two Rake receiver fingers in the user might be receiving from one base station, and another finger from another base station. On the uplink, however, only selection combining is used.

It is sometimes said that the major advantage of soft handoff is macroscopic diversity. As currently designed, CDMA systems are able to effectively use a 0 dB hysteresis threshold (equivalent to “instant’’ switching to the best base station at all times by the base station controller during a soft handoff) without suffering from the ping-pong effect. This enables the benefits of macroscopic diversity to be realized. In TDMA or frequency-division multiple access (FDMA) systems, the prohibitive cost of the ping-pong effect does not allow a 0 dB hysteresis threshold.

However, soft handoff macroscopic diversity benefits are not only for CDMA systems. Similar benefits are available to TDMA/FDMA systems willing to accept the accompanying cost in resource usage. Macroscopic diversity for general “frequency reuse’’ systems is discussed by Bernhardt in [32]. In fact, the process of handoff itself can be seen as a use of macroscopic diversity. TDMA/FDMA systems typically employ hard handoff algorithms which restrict their utilization of macroscopic diversity by implementing hysteresis. But could TDMA/FDMA systems implement soft handoff? Not as easily, because it is more costly, perhaps involving troublesome synchronization of time slots and multiple receptions and transmissions at different frequencies. Because of the nature of CDMA, it does not face these problems.

Papen [29] notes that in IS-95 uplinks, the signals from the different base stations after deinterleaving and channel decoding are combined with selection combining. Reference [29] proposes combining prior to deinterleaving and channel decoding. That is, each base station transmits the hard bits and quality information necessary for deinterleaving and channel decoding to the combiner in the base station with the highest signal strength. This base station then uses the information together with soft decisions from its demodulator/equalizer. It is found in simulation that this scheme is up to 5 dB better than traditional combining in terms of E_c/I_0 for a given BER. Papen notes that his scheme is applicable not only to CDMA systems, but to any system which implements some form of macroscopic diversity combining.