Duplexing, resource allocation and inter-cell coordination: design recommendations for next generation wireless systems


Bell Laboratories, Lucent Technologies

Summary

Coexistence of different access technologies, hierarchical cellular deployment, a wide variety of data services, requirements for transparent operation across different technologies, adaptivity to varying network conditions and mobility and quality of service (QoS) constraints introduce a number of challenges in the design of future generation systems and the specification of new air interfaces, such as efficiency and flexibility in the utilization of spectrum, dynamic resource allocation and exploitation of the multiuser diversity and reconfigurable interference management and inter-cell coordination. In this paper, three critical issues for the design of next generation systems are addressed: (i) duplexing, (ii) scheduling and resource allocation and (iii) interference and inter-cell coordination. A number of research directions are presented, which constitute promising potential candidates for next generation systems specification. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: channel state information; duplexing; radio resource management; scheduling; interference co-ordination

1. Introduction

Next generation wireless systems are expected to deliver a wide variety of data services in a heterogeneous communication network environment, which supports transparent operation across a number of different technologies, hierarchical and ad hoc structures, adaptivity to varying traffic and propagation conditions and satisfies certain quality of service (QoS) constraints. To this end, the design of next generation systems and the specification of a new air interface will have to rely on the exploitation of new resources, such as the channel state information (CSI), cross layer and contextual information, and the implementation of optimization strategies for the efficient and flexible utilization of the spectrum available, the dynamic resource allocation exploiting all types of diversity (time, frequency, code, space, multiuser) and the reconfigurable interference management and inter-cell coordination.

In this paper, the challenges associated with these critical issues for the design of next generation systems are addressed and a number of research directions are presented, which constitute promising
potential candidates for next generation systems specification.

First, in Section 2, the duplexing scheme selection is analyzed and the major benefits and drawbacks of the two traditional candidates, namely time division duplex and frequency division duplex, are presented in terms of link reciprocity, link budget and synchronicity and guard requirements. A new duplexing approach is proposed, which flexibly combines the features of the two.

Then, in Section 3, the resource allocation issue is first analyzed in a cellular network setup in the context of an orthogonal frequency division multiplexing (OFDM) air-interface and a number of possible options on how resources can be assigned across adjacent cells are discussed. Considering resource allocation within a cell, three promising scheduling techniques are then presented: a near capacity multiantenna multiuser transmission scheme, the so-called sphere-encoded multiple messaging, a distributed scheduling scheme supporting service differentiation and a joint opportunistic beamforming and scheduling scheme exploiting multiuser diversity.

In Section 4, interference and inter-cell coordination are addressed. The delay sensitivity of handoff algorithms and the benefits of fast cell switching, as an alternative to soft handoff, are discussed along with the impact on system performance of the structure of out-of-cell interference, in the absence of inter-cell coordination. The benefits of superposition coding as a throughput-optimal encoding technique for compound channels, in which the random state of the channel, such as the out-of-cell interference, is unknown to the transmitter, are considered in a realistic scenario, in which the transmission rates of the superposition code have to be chosen from a fixed and predetermined rate set.

The importance of self-organization in future wireless networks is emphasized, as a means of achieving adaptive and reconfigurable operation, with base stations probing the environment around them and adjusting accordingly a number of parameters, such as their antenna configuration and transmit power.

Finally, the paper is concluded in Section 5.

2. Duplexing

Traditionally, the decision on how to partition the resources available for communication between uplink and downlink has boiled down to two clear possibilities: either time or frequency separation. Each contending option has accepted advantages as well as clear drawbacks. Time division duplex (TDD) scheme uses the same frequency band but alternates the transmission direction in time. Frequency division duplex (FDD) scheme requires separate frequency bands for uplink and downlink transmission. An implicit understanding developed over the years is that TDD was attractive in microcell systems while FDD was preferred in wide-area systems.

In this section, the duplexing issue is considered in the context of a new air interface design, in an effort to go beyond the traditional paradigm where only pure FDD and TDD are evaluated, bringing other options into consideration.

2.1. Paired Versus Unpaired Spectrum

The allocation of spectrum is a highly political issue, performed by certain national and international agencies and mostly beyond the control of equipment manufacturers. However, the allocation of spectrum strongly conditions the duplexing choice because, if unpaired spectrum is allocated, FDD cannot be used. Although unpaired spectrum is indeed easier to find, the historical tendency is nonetheless to assign paired spectrum for wide-area systems, in which case both options are possible.

2.2. Link Asymmetry

An immediate advantage of TDD is that it enables an asymmetric allocation of degrees of freedom between uplink and downlink although in general not dynamically on a cell-by-cell basis, but rather on a system-wide basis. A similar asymmetry in FDD would require uneven spectrum blocks, highly unlikely and very rigid. The relevance of link asymmetry, nonetheless, is still unclear. The UMTS forum forecasts a 2.3 ratio of downlink over uplink for 2010 but this will depend largely on emerging and yet-to-be-envisioned applications. In that sense, TDD may be considered as a favored option but not decisively so because of the difficulty in predicting whether the asymmetry ratio will be significantly different from unity.

2.3. Link Reciprocity

Link reciprocity is usually regarded as the most attractive feature of TDD, which naturally enables it, at least for low-to-moderate normalized Doppler spreads. As a result of reciprocity, sophisticated transmit processing schemes that necessitate instantaneous
channel information become feasible [1]. The lack of reciprocity in FDD, in turn, makes these schemes dependant on the relay of CSI through feedback, which tends to incur unacceptable delays if conventional transmission techniques are employed.

Two questions arise concerning the link reciprocity issue:

- Whether the majority of users fall within the Doppler range where reciprocity holds with sufficient accuracy;
- What is, quantitatively, the value of channel information availability at the transmitter;

In a cell operating individually, the advantage in terms of downlink throughput is sizeable if the number of base station antennas is sufficient [22].

2.4. Link Budget

Let us consider an FDD system that is radiating a steady power level \( P \). Consider now TDD. If the power level during the active part of the duplex is kept at \( P \) (same amplifiers used), there is a 3 dB increase in the thermal noise floor because of the doubling in bandwidth. With a path loss exponent of 3.8, for instance, this results in a 17% reduction in range. In exchange, the ‘average’ transmit power over time is halved, which has no impact on the base station but would extend the battery life in a mobile terminal. If, in contrast, it is preferred to keep the range unchanged with respect to the FDD case, the power radiated during the active part must be doubled. This would require bigger amplifiers and it would also result in the same ‘average’ transmit power and thus the same battery life.

2.5. Synchronicity and Guards

In FDD, uplink and downlink are orthogonal in frequency, provided there is sufficient separation between the corresponding blocks. Each side of the link requires guard bands to accommodate filter roll-offs (Figure 1). In TDD, temporal orthogonality is only possible if cells have synchronized uplink and downlink switch patterns plus guard times to account for propagation delays. This is in addition to guard bands roughly equal to those in FDD (Figure 2).

Orthogonality is essential in wide-area systems, otherwise catastrophic interference may take place.¹

¹If the system has a microcell component with street-level or indoor bases, these may not need to be synchronous.

The overhead represented by the guard times can be made as small as desired by extending the duplex time (time that either link is active). Guard times are therefore an issue only because the duplex time should be kept short in order to minimize physical-layer contribution to latency and also to ensure channel reciprocity over the widest possible range of Doppler spreads.

Synchronicity is essential in TDD. This is an inconvenience with respect to FDD but not insurmountable given the availability of low-cost GPS technology. In the following, we present a guard time evaluation in order to provide a quantitative insight on the issue.

Consider the possibility of base-to-base interference due to lack of synchronicity or else due to propagation delays. To first-order, propagation between elevated (over the clutter) base stations is dominated by a direct path and a reflected path. The resulting path-loss versus distance is plotted in Figure 3. Remarkably,
the path-loss exponent for base-to-base propagation (beyond 1–2 km) is roughly the same as for mobile-to-base, with such mobile in the clutter. The intercept, however, is about 47 dB higher. Coupled with higher transmit power and antenna gain at bases with respect to mobiles, this confirms that base-to-base interference would be catastrophic and must be avoided. Specifically, it takes about 32 rings of cells for the base-to-base path-loss to have grown to the level of mobile-to-base path-loss at the second ring. Therefore, guard times should protect against interference originating up to several tens of rings, which might be tens of km depending on cell size and base height. Fortunately, though, earth curvature poses a limit to the detrimental consequences of base-to-base propagation. For typical base heights, this effect sets in at about 30–35 km and hence it suffices to guard up to roughly that distance. At 3.3 m/s (speed of light), this distance corresponds to about 100 μs guard time. (For added margin, this could be increased to 150–200 μs.)

The amount of overhead can now be estimated depending on the time between uplink–downlink transitions, which we refer to as duplex time. For 1 ms duplex time, for instance, the overhead would amount to 10–20%. For 2 ms duplex time, it would constitute 5–10%. The above analysis provides reasonable assumptions when design choices are made.

2.6. Link Continuity

A drawback of TDD comes from the periodic interruptions in the links, which are active only for a part of the time (usually but not necessarily 50%). Interestingly, this issue did not exist in circuit-switched voice systems and thus has not been part of the traditional discussions on FDD versus TDD duplexing. This is a new issue that is caused by, and central to, packet-switched data traffic.§

Besides higher bandwidth efficiencies, one of the central goals in the design of future generation systems is to achieve an order-of-magnitude reduction in latency. This is being recognized as a necessary condition for the support of certain envisioned applications (such as gaming). With discontinuous links, no message—not even a 1-bit acknowledgement—can be relayed back with a delay inferior to the duplex time. This implies that the time taken by a basic roundtrip at the physical layer level cannot go below a few ms and thus the aggregate delay experienced by a packet running through a scheduler and subject to ARQ (Automatic Repeat Request) can easily be on the order of 10 ms. This latency propagates through the protocol stack posing serious problems to the upper layers and causing bottlenecks. As a result, some of the throughput improvements enabled at the physical layer may not be realized, an issue that becomes increasingly important as data rates grow.

In summary, the primary issues on which the choice of a duplexing scheme rests appear to be link reciprocity and link continuity, each of which favors a different choice. Similarly, the remaining issues (synchronicity, link budget, symmetry etc.) do not point to a clear preferred choice either. In light of these facts, the question that naturally arises is whether it is possible to combine FDD and TDD in such a way that the best of each is preserved.

2.7. Band Switching Duplexing

Band switching duplexing has been proposed in [2] and can be described as follows. Given paired spectrum blocks, instead of reserving a block for uplink and the other for downlink, alternate their use every Ts, as depicted in Figure 4. With this scheme reciprocity is achieved and the channel can be estimated in each band when it is used for uplink and then exploited when it is used for downlink. Synchronicity and guard times are still needed, as in TDD. Both links are always active (except on guard times). The switching time T must be selected, so that the channel can be measured in one time period and then the measurement can be used in the following time period, under the assumption that the channel has not changed considerably during that time. In terms of channel measurement accuracy the switching time needs to be selected as small as possible. The minimum switching time is determined by the down time necessary between reception and transmission and inter-base-station synchronization time requirements.

Note that band switching is both TDD and FDD. It is TDD because every unit of bandwidth is used, alternatively, half of the time for uplink and half of the time for downlink; it is FDD because, at every point in time, half the spectrum is used for uplink and half for downlink.

Fig. 4. Band switching duplexing.
Implementation of this duplexing scheme may introduce some challenges that need to be assessed. Nevertheless, it provides the best set of trade-offs and at the same time the TDD alternative remains an option in case unpaired spectrum is allocated as many parameters, such as guard times, synchronicity etc. are reusable. If link reciprocity fails to provide the expected gains band switching can be easily reduced to conventional FDD.

3. Scheduling and Resource Allocation

The efficient utilization of spectrum requires dynamic resource allocation strategies with the flexibility to adapt to varying wireless network conditions, user requirements and QoS constraints, will be one of the major criteria for the design of a new air interface. In this section, resource allocation strategies are first analyzed in a cellular network setup in the context of an OFDM air-interface. Not restricted to a specific air-interface, three promising resource allocation within the cell and multiuser scheduling techniques are then considered.

3.1. Resource Allocation Across Adjacent Cells within the Context of an OFDM Air-Interface

OFDM is a promising modulation scheme for future wireless communications, as it can provide large data rates with sufficient robustness to radio channel impairments. In OFDM, a large number of orthogonal, overlapping, narrow band sub-channels or sub-carriers, transmitted in parallel, divide the available transmission bandwidth. The separation of the sub-carriers is theoretically minimal, such that there is a very compact spectral utilization. The attraction of OFDM is mainly due to how the system handles the multipath interference at the receiver. Multipath generates two effects: frequency selective fading and intersymbol interference (ISI). The ‘flatness’ perceived by a narrow-band channel overcomes the former and modulating at a very low symbol rate, which makes the symbols much longer than the channel impulse response, diminishes the latter. Using powerful error correcting codes together with time and frequency interleaving yields even more robustness against frequency selective fading, and the insertion of an extra guard interval between consecutive OFDM symbols can reduce the effects of ISI even more.

When Orthogonal Frequency Division Multiple Access (OFDMA) schemes are employed, user-specific subcarriers (tones) are assigned. In a multicell environment the challenge is to manage resources—in this case different tones—in an efficient manner, so that inter-cell interference effects are minimized. In this section, different resource allocation strategies and their relative merits are discussed.

A taxonomy of the various options on how resources can be assigned across adjacent cells in an OFDM system is depicted in Figure 5. Specifically, the issue of utilizing the sub-carriers or tones of a given carrier across adjacent cells is addressed. Each option has a distinct out-of-cell interference characteristic that determines performance. In the following paragraphs, these options are outlined and their relative merits discussed. The broad classification at the

*Fig. 5. Taxonomy of resource allocation options in multicell orthogonal frequency division multiplexing (OFDM).*
highest level is based on how the sub-carriers or tones are assigned and used in different base stations.

According to Option A, all base stations transmit on all the tones with equal power. For variable rate transmission using adaptive coding and modulation, this can be achieved by variable spreading through a combination of coding and symbol repetition for cases when very low coding rate is required to use all the tones. Since all tones are used to signal to mobiles significant frequency diversity gains can be obtained if necessary. This approach does not benefit from frequency hopping since power is transmitted from all tones at all times. Frequency reuse will not be required with proper design of link layer rates. Thus no frequency planning at the time of deployment is required with proper design of link layer rates. Thus there will be significant interference fluctuation across code blocks making it difficult to do adaptive modulation. Thus fast hopping is preferred in those cases where the bandwidth and hence the number of sub-carriers is not large.

In frequency hopping, each base station transmits on a subset of tones at any given time. The tone assignment is done independently in each cell and is changed over time using random hopping sequences. Thus no frequency planning is required in this case. Frequency hopping provides frequency diversity and interference averaging. In slow frequency hopping (Option C), the tone assignment is fixed over the duration of an entire channel code block that spans multiple OFDM symbols while in fast hopping (Option D) frequency hopping is done within a code block from symbol to symbol. The main difference between slow hopping and fast hopping occurs when the number of tones occupied by a channel code block is small. In this case, with slow hopping, there is insufficient interference averaging within the code block since there is no hopping within the code block. Thus there will be significant interference fluctuation across code blocks making it difficult to do adaptive modulation. Thus fast hopping is preferred in those cases where the bandwidth and hence the number of sub-carriers is not large.

Alternate to random hopping, tones can be dynamically assigned across cells in a coordinated manner by taking into account the specific terminals and their path loss from the different base stations to which transmission is scheduled. This is Option E in Figure 5. This results in optimum management of out-of-cell interference and maximizes throughput. However, the implementation requires exchange of messages between base stations and the radio network controller at a very fine time scale and thus may not be practical. An alternative is to achieve partial coordination in a distributed manner. For example, this could be achieved through feedback from the mobile, providing information on its measured strength of signals from other base stations. Another approach is a hybrid of fixed frequency planning and hopping, in which base stations start using tones that are disjoint and as traffic grows they start using additional tones that are reserved for neighboring base stations. Such a scheme avoids interference at low traffic conditions and will require frequency planning.

When evaluating the options presented above in the design of next generation systems, one should take into consideration that one of the main objectives is deployment flexibility. In fact, the base stations should be able to autonomously configure themselves. Thus,
it is logical not to consider options that involve any frequency planning. Hence, we reject Options B and F from further consideration. Next, consider Option E that involves inter-cell coordination. It is likely that inter-cell coordination does not become practical because of the complexity involved in the coordination process especially in a mobile environment. Furthermore, we lose scalability or have to restrict the amount of coordination. Thus, we are inclined to reject Option E from further consideration.

We are thus left with the options of random tone hopping or spreading the power across all the tones. Both these approaches do not require any frequency planning. In either case frequency diversity is obtained as well. On the uplink, the main reason that drives us towards the frequency hopping approach is that mobiles have typically limited power. By restricting each mobile to transmit only on a few tones at any given time with multiple mobiles transmitting simultaneously on different tones, we maximize the throughput. The alternate approach of spreading each mobile’s signal across all the sub-carriers, which also allows multiple mobiles to transmit simultaneously, will lead to loss of orthogonality between the signals of the different mobiles in the presence of multipath. It is therefore better to have each mobile transmit on a subset of the tones and then use frequency hopping to average interference. On the downlink, the option where each base spreads its power across all the tones is preferred. This can potentially exploit frequency diversity better than frequency hopping because of coherent combining across carriers (as opposed to getting frequency diversity through coding in the frequency hopping case) for low rate transmissions or through optimum assignment of users to tones.

The use of universal frequency reuse is justified in Figure 6, where we compare different frequency reuse options in a cellular environment on the downlink with 40 dBm transmit power and 0.5 km cells. Standard path loss model [3] and shadow fading standard deviation of 8 dB were used to generate the results. Figure 6(a) shows the cumulative distribution function of the SINR that is achieved with various reuse factors and Figure 6(b) shows the normalized rate distribution that can be achieved in a typical 20 MHz system. We see that while reuse improves the SINR, the throughput that can be achieved in a given amount of bandwidth is lower with frequency reuse than with universal reuse. Thus, universal reuse is superior from system capacity point of view besides not requiring frequency planning at the time of deployment.

3.2. Resource Allocation within the Cell and Multiuser Scheduling

3.2.1. Near capacity multiantenna multiuser communication

The use of multiple antennas to communicate with many users simultaneously (multiuser or ‘broadcast’ link) has recently received a great amount of attention, especially in wireless local area network (WLAN) environments such as IEEE 802.11, where channel conditions change slowly and there is sufficient time for all parties to learn their channel conditions.

The capacity of an $M$-transmit, $N$-receive antenna link grows linearly in a Rayleigh fading environment
with the minimum of $M$ and $N$ when the receiver knows the channel [4]. It is also shown in [4] that $K$ users, each with one antenna, can transmit to a single receiver with $M$ antennas and the sum-capacity (total of transmission rates to all $K$ users) grows linearly with the minimum of $M$ and $K$. In a broadcast link case where the $M$ antennas are used to transmit to the $K$ users, it was shown in [5] that the sum-capacity grows linearly with $\min(M,K)$, provided both the transmitter and receivers know the channel.

Transmission schemes achieving the sum-capacity in multiantenna multiuser links have been recently proposed, based on the ‘dirty-paper coding’ concept. Dirty-paper coding [6] is introduced for Gaussian interference channels, where it is shown that the capacity of a channel with interference, where the interfering signal is known at the transmitter, is the same as in the case of the channel with no interference. In this approach, the interference is seen as ‘dirt’ and the desirable signal as ink. This information-theoretic solution is not to combat the dirt, but to use a code that aligns itself as much as possible with the dirt. Dirty-paper techniques are natural candidates for achieving sum-capacity in multiantenna multiuser links because the transmitted signal for one user can be viewed as interference for another user, and this interference is known but out of the control of the transmitter.

Let us now consider the scenario where the transmitter creates the signals, and thereby can also control the interference seen by all the users. It is shown in [7] that a suitably modified form of channel inversion can achieve near-sum-capacity performance. This technique multiplies the vector-signal to be transmitted by the inverse of the channel matrix; the result is an ‘equalized’ channel to each user. Regularization can be employed in the inversion in order to overcome the noise amplification problem. Another way to make sure that the transmitted data do not lie along the singular values of the inverse channel is to introduce a vector perturbation of the signal to be transmitted to all the receivers. A combination of regularization and vector perturbation can be employed to achieve better results. ‘Sphere encoding’ techniques such as the Fincke–Pohst algorithm [8] can be used in selecting the desired vector perturbation. The perturbation algorithm has the simple interpretation of placing the largest signal components along the smallest singular values of the inverse channel, and the smallest signal components along the largest singular values of the inverse channel.

This multiantenna multiuser transmission scheme, denoted Sphere-Encoded Multiple Messaging, allows for a base station that is equipped with an $M$-element transmit antenna array to send simultaneously and selectively $K$ messages to $K$ autonomous single-antenna terminals. The technique can also be used when $M$ base stations, each equipped with a single transmit antenna, are ‘wired-together’ to function as an $M$-element transmit array. The total throughput, for a constant total power and spectrum, is proportional to $\min(M,K)$. When $K$ becomes comparable to $M$, this technique significantly outperforms simple channel inversion with diagonal loading. Figure 7 illustrates the effectiveness of Sphere Encoded Multiple Messaging for the scenario of $M=10$ transmit antennas sending ten different messages to $K=10$ single-antenna terminals. The scheme employs a rate 3/4 turbo code and 16-QAM to produce a throughput per user of $4 \times 3/4 = 3$ bits/s/Hz, or a total throughput of $3 \times K = 30$ bits/s/Hz. The solid curve displays the bit error rate as a function of the expected SNR at each terminal, and the vertical dashed line shows the SNR (11.2 dB) such that the sum-capacity is equal to 30 bits/s/Hz. Note that Sphere Encoded Multiple Messaging performs within 4 dB of sum-capacity.

Fast CSI, which gives the transmitter knowledge of the complex downlink propagation matrix in near real-time, is the essential enabler for Sphere-Encoded Multiple Messaging. Fast CSI transfer is relatively straightforward for TDD [1] where reciprocity holds. In the case of FDD, there is no reciprocity, the terminals have to estimate the downlink channel from received pilots, and each terminal then has to send its own portion of the CSI to the base station on
the reverse link. Recently, it was shown that the FDD CSI transfer can be accomplished simply and effectively, with minimal delay, by means of uncoded analog linear modulation [9]. The scheme exploits the fact that the uplink channel is a multiaccess MIMO channel. Figure 8 illustrates the number of uplink symbols per user for TDD and for FDD, for $M = 2, 4, 8, 16$ transmit antennas, that are needed to achieve a stipulated relative root mean square (rms) estimation error. The product of the uplink SNR and the number of users is equivalent to 10 dB. Higher-quality CSI (i.e. small estimation error) requires a longer transmission interval. It is remarkable that the FDD CSI transfer becomes easier as the number of transmit antennas increases: doubling the number of transmit antennas doubles the amount of CSI that each user must transmit, but the throughput of the uplink channel more than doubles. While FDD CSI transfer does require more effort than TDD CSI transfer, the efforts become comparable for large numbers of antennas.

Three extensions of the above-described transmission scheme need to be considered, which (a) handle users with different SNRs and permit unequal rate transmissions, (b) handle delay-spread channels through an OFDM version, and (c) introduce scheduling considerations.

### 3.2.2. Distributed scheduling in wireless data networks with service differentiation

In recent years, WLANs have been rapidly deployed all over the world. An important design issue in such data networks is that of distributed scheduling. The lack of centralized control leads to multiple users competing for channel access. This leads to significant throughput degradation. Existing approaches, such as Slotted Aloha [10] and IEEE 802.11 DCF (distributed coordination function) [11], also fail to provide differentiated service to users. The upcoming IEEE 802.11e enhanced DCF aims to address these issues, but, by supporting only strict priority classes, it is unable to provide dynamic service differentiation.

A class of distributed scheduling algorithms for regulated contention medium access control (RCMAC) has been proposed in [12], which provides dynamic prioritized access to users for service differentiation. Unlike Aloha, in RCMAC the user’s access probability is not fixed but dynamically varies with a differentiation function. The latter depends on the user’s dynamic weight associated with metrics, such as the user’s priority, queue length or delay of packets and a certain access threshold. For the differentiation function, two approaches are considered in [12]. In the first, it is proportional to the user’s dynamic weight and is called weight proportional (WP), whereas in the second, it reflects whether the dynamic weight exceeds or not a certain threshold and is called threshold based regulation (TBR). The performance of RCMAC, for both the cases of WP and TBR, is compared with that of Slotted Aloha and IEEE 802.11 DCF in Figure 9, in terms of service rates obtained in the two-user case, and in Figure 10, in terms of total throughput—expressed as the fraction of slots with successful transmission for that user —versus the number of users for symmetric arrival rates. RCMAC-TBR is shown to achieve

---

Fig. 8. Number of training symbols per user for TDD and for FDD, for $M = 2, 4, 8, 16$ transmit antennas, that are needed to achieve a certain rms estimation error.

Fig. 9. Achieved service rates in the two-user case.
significant improvements in system throughput over a sizable range of the number of nodes in the system. By regulating multiuser contention, RCMAC achieves higher throughput when traffic is bursty, as is typically the case. In addition to WLANs, the basic concepts of RCMAC also have applications in multi-hop cellular networks, mesh networks and sensor networks.

3.2.3. Jointly opportunistic beamforming and scheduling (JOBS) for downlink packet access

In a multiuser scenario, the so-called opportunistic approaches have recently attracted considerable attention. The basic idea is to multiplex users by granting the channel to those with higher chances of completing a successful transmission and thereby achieve an overall, rather than individual link, throughput maximization. For specular spatial channels, opportunistic beamforming approaches [13] point at the user with the highest SINR out of those present in the system. On the other hand, in rich-scattering scenarios opportunistic approaches will implicitly exploit fading by granting access to those with highest instantaneous capacity [14]. Instead of traditional strategies aiming at stabilizing individual links against channel (or interference) fluctuations by using multiple antennas, artificial fading may need to be introduced for slowly time-varying channels on the transmit side by randomly changing transmit weights. More sophisticated multiple-access can be derived by combining space and code diversity. In that case, additional issues like efficient user grouping arise.

Opportunistic beamforming is particularly attractive, because no additional over-the-air signaling is required and the additional processing required at the mobile is minimal. However, the issue of delay and delay jitter was not considered, which is also a crucial parameter of a packet data access system. In fact, in packet data systems—like HSDPA (high speed downlink packet access) in UMTS—there is a tradeoff between throughput and delay. When the throughput gains, the delay performance deteriorates.

Jointly opportunistic beamforming and scheduling (JOBS) [15] is a simple and robust downlink packet access technique for mobile scenarios where fading is expected. It is particularly suited for delay tolerant data services to mobiles moving at moderate and high velocities. JOBS combines channel aware scheduling with ‘dumb’ beamforming [13]. The delay performance can be dramatically improved and at the same time the throughput is also improved by employing JOBS, which exploits all past mobile reports to ‘learn’ the current preferred beam of each mobile, and puts this information to use by providing priority to mobiles waiting longer for the next packet.

An example of the achieved enhancements in terms of system throughput is given in Figure 11, where system throughput versus delay performance is compared for proportional fair (PF) scheduling, standard opportunistic beamforming (sOBF), that is an opportunistic beamforming system with simple scanning of beams, and JOBS. JOBS is shown to achieve 10% higher throughput as compared to sOBF.

Several open issues remain that require further investigation, such as the effects of system latency. As in the case of packet data systems, the performance of fast moving mobiles degrades rather fast with
system latency. In order to limit the built-in latency, a variety of schemes need to be investigated following a passive or active approach. An example of an active scheme is predicting the future state of the channel of the selected mobile, based on its past reports. An example of a passive scheme is using a certain back-off, i.e. forcing the base station to reduce its transmit data rate to allow the mobile to receive the packet, even when the channel has deteriorated somewhat by the time the transmission actually takes place. Moreover, signaling overhead requirements, for example certain pilot requirements must be evaluated.

4. Interference and Inter-Cell Coordination

The benefits of dynamic resource allocation and reconfigurable scheduling need to be validated in challenging wireless network setups, where coexistence of different access technologies, hierarchical cellular deployment, a wide variety of envisioned services, with increasing demand for data services and mobility requirements result in challenging interference scenarios and introduce the requirement for adaptive and reconfigurable inter-cell coordination.

Some insight on the impact of resource allocation across cells on the out-of-cell interference characteristics has been already given in Section 3.1, where a number of resource allocation options have been analyzed in the context in an OFDM system. In what follows in this section, the delay sensitivity of handoff algorithms and the benefits of fast cell switching as an alternative to soft handoff, the impact on system performance of the structure of out-of-cell interference in the absence of inter-cell coordination and the benefits of superposition coding, as a throughput-optimal encoding technique, are discussed. It is finally emphasized that self-organization of base stations will play a critical role in the design of adaptive and reconfigurable wireless networks.

4.1. Handoff-Delay Sensitivity Considerations

The act of transferring support of a mobile from one base station to another is termed handoff. In traditional hard handoff, the connection to the current cell is broken, and a new connection to another cell is made. However in soft handoff the mobile can communicate with multiple base-stations simultaneously.

Current CDMA based cellular systems employ soft handoff for voice and circuit data on the downlink and cell switching for packet data. Soft handoff is implemented on the uplink for all services. While soft handoff is beneficial for the current system from the point of view of air-interface system capacity, it imposes severe constraints on the design of the backhaul network through timing requirements and QoS requirements. If handoff requests could be processed instantaneously, transmission from the base station with the best link to the user would achieve significant fraction of the macro-diversity gain without utilizing additional resources. However, in practical systems there is a non-zero execution delay and soft handoff provides the required robustness to delays, although it comes at the expense of additional network resources.

The delay sensitivity of the soft and hard handoff schemes has been studied in [16] in various channel conditions. Figure 12 shows the relative performance of the two algorithms, soft and hard handoff, under the urban and sub-urban channel conditions. Also shown are the performance of the power-optimal algorithm and the one where both the transmitters are always on. The optimal algorithm selects the transmitter with the stronger signal instantaneously and when the signal is so weak that the system is still in outage, the second transmitter is also selected as in soft handoff. The performance of the power-optimal handoff scheme provides the lower bound on the power requirement. Both the urban and the sub-urban channels are assumed to be log-normally distributed with identical variances equal to 4.5 dB. The urban channel however differs from the sub-urban channel in its auto-correlation coefficient. The sub-urban channel conditions change less frequently than those of the urban channel. Both handoff algorithms are shown to be more sensitive to actual handoff delays in case of urban channels.

Fast cell switching, which exploits the selection diversity gain by appropriately choosing the base station for a particular user based on instantaneous channel conditions, is a promising alternative to soft handoff for delay sensitive packet service [17]. Simulation results indicate that, with reduced switching delays, it is possible to achieve comparable performance on the downlink. Flat all-IP network architectures envisioned for next generation systems become much simpler to implement without soft handoff.

4.2. Out-of-Cell Interference Considerations on the Downlink and Uplink

The structure of out-of-cell interference, in the absence of inter-cell coordination, impacts the system
performance significantly. One can consider a system design where out-of-cell interference is bursty, less predictable or smoother by averaging over many users.

A detailed simulation study [18] comparing these options was carried out for the uplink, assuming the use of incremental redundancy schemes to deal with unpredictable interference, which showed that interference averaging is not always essential. Three types of transmission strategies studied, were classified according to the resulting interference characteristics observed at the receiving base station: (a) no in-cell interference, averaged out-of-cell interference, (b) no in-cell interference, bursty out-of-cell interference and (c) averaged in-cell interference, averaged out-of-cell interference. The interference patterns created by these transmission strategies translate to different interference statistical behavior, ranging from small mean and wide variance to quasi-static but large interference. Examples of transmission strategies resulting in interference pattern (a) include synchronous CDMA transmission or OFDM transmission with fast hopping. In these cases, users within the cell are orthogonal to each other and thus there is no in-cell interference. Each user is subjected to interference from all the users transmitting in the surrounding cells and the out-of-cell interference does not fluctuate significantly over short time durations. Examples of transmission strategies resulting in interference pattern (b) are a CDMA or OFDM system, in which transmission among users within a cell is time multiplexed. Since in-cell users transmit at different times there is no in-cell interference. The out-of-cell interference comes from a single user in each cell and its power is unpredictable or bursty. An example of a transmission strategy resulting in interference pattern (c) is the asynchronous CDMA system with multiple simultaneous users transmitting within the cell. Each user suffers interference from many in-cell users as well as a large number of out-of-cell users, resulting in averaged in-cell and out-of-cell interference.

Performance in the presence of these interference patterns was evaluated in terms of system throughput under identical maximum allowable in-cell rise-over-thermal (IROT) ratio criteria. The results show [18] that option (b) clearly outperforms the transmission option (c) with any number of users. Furthermore, in spite of the large variance in the interference power statistics, performance of option (b) is comparable to that of option (a) that requires in-cell orthogonality. Finally, the results also indicate that by allowing multiple user transmission for users near the cell

![Fig. 12. Sensitivity of various handoff techniques for urban (left) and suburban channels (right).](image-url)
edge and single user transmission for users closer to the base station, so that the maximum transmit power does not impose a limit on the IROT achieved, can further increase throughput. Similar studies need to be performed in the case when interference cancellation techniques are applied at the receiver in order to evaluate critical interference patterns and resulting performance merits in this case.

4.3. Superposition Coding for Unknown Interference

Superposition coding has been proposed [19] and shown to be a throughput-optimal encoding technique for compound channels, in which the random state of the channel is unknown to the transmitter. Such a scenario arises for example in a wireless network when the out-of-cell interference from neighboring base stations, and therefore the signal-to-noise ratio at the receiver, is unknown and time-varying due to power control and scheduling decisions made by neighboring base stations.

The main idea behind superposition coding is to make sure that a minimum data rate can be supported even when the interference is large; a higher data rate can be received when the interference is small. In the case of transmission to \( K \) receivers (corresponding in our scenario to virtual receivers that can decode the received signal at different noise levels) under total power constraint \( P \), when deploying superposition coding the transmitter generates \( K \) codebooks with certain powers and rates, and transmits the sum of \( K \) codewords encoding the symbols intended for the \( K \) receivers. The receiver with the worst channel decodes the codeword with the corresponding rate and all the other codewords appear as noise to this receiver. The next-worst receiver is able to decode the rate corresponding to the worst receiver and subtract it from the received signal. Then, it decodes the codeword of its corresponding rate, while it sees all other codewords as additional noise and so on, until the best receiver decodes all codewords.

The benefit of applying superposition coding was investigated under the scenario in which the actual experienced interference is unknown in every time slot, but the statistical behavior of the interference is known to the transmitter. In addition, compared to previously known theoretical results, we impose the practical constraint that the transmission rates of the superposition codes have to be chosen from a fixed and pre-determined rate set. The maximum achievable throughput using superposition coding was explicitly calculated and compared to that of single-rate coding [20]. Determining the throughput-maximizing superposition code involves choosing the optimal subset of transmission rates, the corresponding decoding order, associated discrete noise and interference levels (at which the different components of the superposition code can be reliably decoded) and their corresponding power levels. Superposition coding was found to achieve larger throughputs than single-rate coding, as it effectively improves the granularity of the code and thereby makes more efficient use of the available power.

Preliminary results comparing the performance of superposition coding with infinite rate granularity with retransmission techniques such as incremental redundancy with a limited number of retransmissions indicate that the benefit of superposition coding over incremental redundancy may not be significant for delay-constrained applications. Further conclusions on the comparison between incremental redundancy and superposition coding can only be drawn based on case studies with a fixed and finite rate set, in particular conducted in the context of a power-controlled versus a scheduled mode of transmission in the uplink.

4.4. Self-Organization of Base Stations

Significant expense is incurred in the configuration and management of base stations. In the future it is envisioned that base stations will probe the environment around them and adjust accordingly a number of parameters, such as their antenna configuration and transmit power. Algorithms and protocols for such self-organization scheme are currently under investigation [21]. Centralized control (bunching) of base stations, intelligent relaying, dynamic cell sizing and intelligent handover are only a few examples of self-organization procedures. The information used to enable these procedures can be classified in four categories:

- Geographical information (e.g. location-related parameters and propagation characteristics);
- Spatial/temporal information (e.g. available coverage, capacity and interference patterns);
- Network information (e.g. services offered, traffic);
- Contextual information (e.g. user profile).

Logical sensors located in the corresponding parts of the network would probe and buffer a number of parameter values associated with the above categories and compute the required adjustments in a dynamic fashion.
5. Conclusions

In this paper, three critical issues for the design of next generation wireless systems have been addressed, namely duplexing, resource allocation and inter-cell interference coordination. The objective was to understand the major requirements and the challenges involved in the design of future generation networks and present a number of research directions that appear to be promising potential candidates.

The major benefits and drawbacks of TDD and FDD were discussed in terms of link reciprocity, link budget, synchronicity and guard requirements. A new scheme, band switching duplexing, was proposed, which flexibly alternates between the two and combines their features.

Some insight on resource allocation across cells and the inter-cell interference implications associated with a number of possible options was given in the context of an OFDM air-interface. Considering resource allocation within the cell, Sphere-Encoded Multiple Messaging—a near capacity multi-antenna multi-user transmission scheme, Regulated Contention Medium Access Control—a distributed scheduling scheme supporting service differentiation, and Joint Opportunistic Beamforming and Scheduling—a scheme combining the benefits of space and multiuser diversity, were presented as potential candidates for future systems design.

In an inter-cell coordination context, the delay sensitivity of handoff algorithms was addressed and the benefits of fast cell switching, as an alternative to soft handoff, were discussed. Aiming to assess the impact of the structure of out-of-cell interference, in the absence of inter-cell coordination, on performance, three different scenarios depending on the smooth or bursty interference behavior have been considered. Superposition coding, as a throughput-optimal encoding technique in the presence of unknown interference, was presented as a possible candidate.

Finally, self-organization of base stations is believed to play an important role in achieving adaptive and reconfigurable operation in future wireless networks.

References

11. IEEE 802.11 WG. Wireless LAN medium access control and physical layer specifications.
Authors’ Biographies

Angeliki Alexiou received the diploma in Electrical and Computer Engineering from the National Technical University of Athens in 1994 and her Ph.D. in Electrical Engineering from Imperial College of Science, Technology and Medicine, University of London in 2000. Since January 1999, she has been a member of technical staff in the Wireless Research Department of Bell Laboratories, Lucent Technologies, Swindon, U.K., where she is involved in the development of advanced signal processing techniques for wireless communications. Her current research interests include multiple antenna communications, signal processing for future generation wireless systems, spatio-temporal channel modeling and re-configurable wireless systems supporting adaptivity to varying network conditions. Dr Alexiou is a co-recipient of Bell Labs President’s Gold Award in 2002 for contributions to Bell Labs Layered Space-Time (BLAST) project and the Central Bell Labs Teamwork Award in 2004 for role model teamwork and technical achievements in the IST FITNESS project. Dr Alexiou is the vice chair of the Working Group on New Air Interfaces, Relay-based systems and Smart antennas of the Wireless World Research Forum. She is a member of the IEEE, the IEE and the Technical Chamber of Greece.

Dan Avidor was born in Tel Aviv, Israel, on 18 October 1936. He received the B.Sc. degree in Electrical Engineering from the Technion-Israel Institute of Technology, Haifa, Israel in 1958 and the M.Sc. degree and his Ph.D. in Engineering from the University of California, Los Angeles (UCLA), in 1970 and 1981 respectively. After completing his studies for his Ph.D., he returned to Israel, where he served as a Research and Development Department Head in the Israeli defense forces. He is currently a Distinguished Member of Technical Staff in the Wireless Research Laboratory, Lucent Technologies, Bell Labs, Holmdel, NJ. His current research interests are in adaptive arrays, signal processing and simulation techniques for wireless systems.

Peter Bosch’s background is in Operating System, File System and Real-time System’s research. He has worked on the distributed operating system Amoeba in the early 1990s, he has built a real-time file system for the operating system Nemesis and Linux in the late 1990s and now works on and with the operating system Plan9 by Bell Laboratories. More recently, Peter et al. in the computing sciences research group at Bell Laboratories have teamed up with wireless research groups, also with Bell Laboratories, to address cellular wireless system problems.

Suman Das was born in India in 1972. He received his undergraduate degree in Computer Science and Engineering from Indian Institute of Technology, Kharagpur in 1994 and M.S. degree and his Ph.D. from Rice University, Houston, TX in 1997 and 2000 respectively. Since then he has been a member of technical staff in the wireless technology research department, Bell Laboratories, Lucent Technologies at Murray Hill, NJ. His primary research interest are in algorithms and architectures for signal processing and communication and design and analysis of future wireless networks.

Piyush Gupta received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Bombay, India, in 1989 and the M.S. degree in Computer Science and Automation from the Indian Institute of Science, Bangalore, India, in 1996. He received his Ph.D. in Electrical and Computer Engineering from the University of Illinois, Urbana-Champaign, in 2000. From 1993 to 1994, he worked as a design engineer at the Center for Development of Telematics, Bangalore. Since September 2000, he has been a Member of Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Lucent Technologies, Murray Hill, NJ. His research interests include wireless networks, network information theory and learning and intelligent systems.

Bertrand Hochwald was born in New York, NY. He received his undergraduate education from Swarthmore College, Swarthmore, PA and the M.S. in Electrical Engineering from Duke University, Durham, NC. From 1986 to 1989, he worked for the Department of Defense at Fort Meade, MD. In 1989, he enrolled at Yale University, New Haven, CT, where he received the M.A. in Statistics and the Ph.D. in Electrical Engineering. In 1995–1996, he was a research associate and visiting assistant professor at the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign. He joined the Mathematics of Communications Research Department at Lucent Technologies Bell Laboratories in September 1996, where he is now a Distinguished Member of the Technical Staff. He is the recipient of several achievement awards while employed at the Department of Defense and the Prize Teaching Fellowship at Yale. He has several patents in the field of wireless communication.

Thierry E. Klein was born in Ettelbruck, Luxembourg, on 7 October 1971. From 1990 to 1992, he attended the Classes Préparatoires at Lycée Louis-Le-Grand in Paris, France, in preparation for the National Admission Contest for Ecole Centrale. He received the B.S. and the M.S. degrees in Mechanical Engineering from the Université de Nantes, Nantes, France, in 1993 and 1994 respectively, and the Electrical Engineer’s degree in Automatics from Ecole Centrale de Nantes, Nantes, France in 1995 (ranked first in class). From 1995 to 2000, he was a Research Assistant in the Laboratory for Information and Decision
Thomas L. Marzetta was born in Washington, D.C. He received his Ph.D. in Electrical Engineering from the Massachusetts Institute of Technology in 1978. His dissertation extended the three-way equivalence of autocorrelation sequences, minimum-phase prediction error filters and reflection coefficient sequences to the two-dimensional case. He worked for Schlumberger-Doll Research from 1978 to 1987 to modernize geophysical signal processing for petroleum exploration. From 1987 to 1995, he performed research and development at Nichols Research Corporation under contracts from the U.S. Department of Defense, NASA and Schlumberger; he headed a group that improved automatic target recognition, radar signal processing and video motion detection. Since 1995, he has been with Bell Laboratories (formerly AT&T, now Lucent Technologies), currently in the Mathematical Sciences Research Center where he heads the Mathematics of Communications Research Department. He specializes in multiple antenna wireless with particular emphasis on techniques for realizing extremely high throughputs with large numbers of antennas. Dr Marzetta was a member of the Sensor Array and Multi-channel Technical Committee of the IEEE Signal Processing Society (2000–2003), and he has served as associate editor for two IEEE journals, and as guest associate editor for the IEEE Transactions on Signal Processing Special Issue on Signal Processing Techniques for Space-Time Coded Transmissions (October 2002) and for the IEEE Transactions on Information Theory Special Issue on Space-Time Transmission, Reception, Coding and Signal Design (October 2003). He was the recipient of the 1981 ASSP Paper Award from the IEEE Signal Processing Society. He was elected a Fellow of the IEEE in January, 2003.

Jonathan Ling received the B.Sc. degree in Electrical Engineering from Rutgers, NJ in 1992 and M.Sc. in Computer Science from Stevens Institute in 1999. He is currently pursing his Ph.D. in Electrical Engineering at Stevens. Jonathan joined the Wireless Communications Research Department of Bell Laboratories, Holmdel, NJ in 1994. His research has focused on radio propagation measurement and modeling, and physical layer performance of wireless systems. He reported the first outdoor large-scale measurement of MIMO capacity for Manhattan in April 2003. Recently, he has developed a technique to improve both throughput and delay in opportunistic scheduling packet data system.

Angel Lozano was born in Manresa, Spain, in 1968. He received the engineer degree in Telecommunications (with honors) from the Polytechnical University of Catalonia, Barcelona, Spain, in 1992 and the Master of Science and his Ph.D. in Electrical Engineering from Stanford University, Stanford, CA, in 1994 and 1998 respectively. Between 1996 and 1998, he worked for Pacific Communication Sciences, Inc. and for Conexant Systems in San Diego, CA. Since January 1999, he has been with Bell Laboratories (Lucent Technologies) in Holmdel, NJ. He has authored over 40 papers and two book chapters and holds six patents. Since October 1999, Dr. Lozano has served as associate editor for IEEE Transactions on Communications.

Thomas L. Marzetta

Sayandev Mukherjee was born in Bangalore, India in 1970. He received the Bachelor of Technology degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India, in 1991, and the M.S. degree and his Ph.D. in Electrical Engineering from Cornell University, Ithaca, NY, in 1994 and 1997 respectively. Since 1996, he has been a Member of Technical Staff in the Wireless Research Laboratory, Bell Labs, Lucent Technologies, NJ, USA. His research interests include stochastic models, wireless system simulations and intelligent resource allocation in wireless systems.

Sape Mullender’s background is in Operating Systems and Distributed Systems. He was a principal designer of the Amoeba distributed system, and a contributor to the Nemesis multimedia system and the Plan 9 operating system. He is a full-time member of Technical Staff at Bell Laboratories and holds a part-time chair in the Computer Science Department of the University of Twente in the Netherlands. He is the editor of Distributed Systems, published by Addison Wesley.

Constantinos B. Papadias was born in Athens, Greece, in 1969. He received the diploma of Electrical Engineering from NTUA in Athens in 1991 and his doctorate from ENST in Paris in 1995. From 1992 to 1995, he was teaching and research assistant at Institut Eurécom, France. From 1995 to 1997, he was postdoctoral researcher at Stanford University’s Smart Antennas Research Group. Since 1997, he has been with Bell Labs’ (Lucent Technologies) Wireless Research Lab, first as Member of Technical Staff and since 2002 as Technical Manager. His current research interests lie in the areas of space-time and next generation wireless systems. He is a Senior Member of IEEE, a member of the Technical Chamber of Greece and he represents Lucent Technologies at the Steering Board of the Wireless Worlds Research Forum (WWRF). He is an associate editor of the IEEE Transactions of Signal Processing and he recently received the IEEE Signal Processing Society’s 2003 Young Author Best Paper Award. He is also currently an adjunct associate professor at Columbia University, teaching a class on space-time wireless systems.
Reinaldo A. Valenzuela obtained his B.Sc. degree from University of Chile and his Ph.D. from the Imperial College of Science and Technology of the University of London, England. At Bell Laboratories, he studied indoor microwave propagation and modeling, packet reservation multiple access for wireless systems and optical WDM networks. He became Manager, Voice Research Department, at Motorola Codex, involved in the implementation integrated voice and data packet systems. On returning to Bell Laboratories, he led a multidisciplinary team to create a software tool for Wireless System Engineering (WiSE), now in widespread use in Lucent Technologies. He received the Distinguished Member of Technical Staff award and is Director of the Wireless Communications Research Department. He is interested in microwave propagation measurements and models, intelligent antennas, third generation wireless system and space-time systems achieving high capacities using transmit and receive antenna arrays. He has published over 80 papers and has 12 patents. He is a Fellow of the IEEE. He is editor for the IEEE Transactions on Communications and the IEEE Transactions on Wireless.

Harish Viswanathan received the B.Tech. degree from the Department of Electrical Engineering, Indian Institute of Technology, Chennai, India in 1992 and the M.S. degree and his Ph.D. from the School of Electrical Engineering, Cornell University, Ithaca, NY in 1995 and 1997 respectively. He was recepient of the Cornell Sage Fellowship. He is presently with Lucent Technologies Bell Labs, Murray Hill, NJ. His research interests include information theory, communication theory, wireless networks and signal processing.