1 Adaptive Resource Allocation in Cooperative Cellular Networks

Wei Yu, [†]Taesoo Kwon and [†]Changyong Shin
University of Toronto, [†]Samsung Electronics

1.1 Introduction

The cellular structure is a central concept in wireless network deployment. A wireless cellular network is comprised of base-stations geographically located at the centre of each cell serving users within its cell boundary. The assignment of users to base-stations depends on the relative channel propagation characteristics. As a mobile device can usually observe signals from multiple base-stations, the mobile is typically assigned to the base-station with the strongest signal; signals from all other base-stations are then regarded as intercell interference. However, at the cell edge, it is often the case that the propagation path losses from two or more base-stations are similar. In this case, the signal-to-noise-and-interference ratio (SINR) would have been close to 0dB, even if the mobile is assigned to the strongest base-station. To avoid excess intercell interference in these cases, traditional cellular networks employ a fixed frequency reuse pattern so that neighbouring base-stations do not share the same frequency. In this manner, neighbouring cells are separated in frequency so that cell-edge users do not interfere with each other.

The traditional fixed frequency reuse patterns are effective in minimizing intercell interference, but it is also resource intensive in the sense that each cell requires substantial amount of nonoverlapping bandwidth, so that only a fraction of the total bandwidth can be made available for each cell. Consequently, many standards for future wireless systems have targeted on maximal frequency reuse, where all cells use the same frequency everywhere. In these systems, it is crucial to manage intercell interference using dynamic power control, frequency allocation, and rate allocation methods.

Wireless channels are fundamentally impaired by fading, by propagation loss, and by interference. In the past decade, intense research has focused on the mitigation of short-term fading, where spatial, temporal and frequency diversity techniques have been devised to combat the short-term variation of the channel over time. Large-scale fading, propagation loss and intercell interference, however, call for different approaches. As large-scale channel and noise characteristics can often be estimated at the receivers and made available at the transmitter, rather than combating large-scale fading the right approach is to adapt to it.

Toward this end, cooperative communication has emerged as a promising future technology for dealing with the large-scale channel impairments. This chapter considers two types of cooperative networks that specifically address the issues of intercell interference and path loss.

- *Base-station Cooperation:* This type of cooperative network explores the possibility of coordinating multiple base-stations. In a traditional cellular network, each base-station operate independently. In particular, each base-station adapts to the channel propagation condition within each cell without considering its intercell interference onto the neighbouring cells. The intercell interference is always treated as a part of the background noise. A network with base-station cooperation is a network in which the transmission strategies among the multiple base-stations is designed jointly. In particular, the base-stations may cooperate in their power, frequency, and rate allocations in order to jointly mitigate the effect of intercell interference for users at the cell edge. Such a cooperative network can also be thought of as an adaptive frequency reuse scheme where the frequency usage and transmission power spectrum are designed specifically according to the mobile locations and user traffic patterns.
- *Relay Cooperation:* This type of cooperative network explores the use of relays to aid the direct communication between the base-station and the remote subscribers. The path loss is a fundamental characteristic of the wireless medium. The path loss exponent, which is determined by the physics of electromagnetic wave propagation environment, typically ranges from 2 to 6. Consequently, propagation distance is the most crucial factor that affects the capacity of the wireless channel. The use of cooperative relays in a cellular network can be thought of as a method for reducing the propagation distance. Instead of adding more base-stations to the network (which is costly), the idea of a cooperative relay network is to deploy relay stations within each cell so that the mobile users may connect to the nearest relay, rather than the base-station which may be far away. Relay deployment substantially improves the area-spectral efficiency of the network by improving the network topology.

In both types of cooperative networks, resource allocation is expected to be a crucial issue. In a network with base-station cooperation, base-stations must jointly determine their respective power and bandwidth allocation for the purpose of minimizing intercell interference. In a cooperative relay network, power and bandwidth assignments need to be made for each of the base-station-to-relay and relay-to-mobile links. The optimal allocation of these network resources has a significant impact in the overall network performance.

This chapter provides an optimization framework for power, bandwidth and rate allocation in cooperative cellular networks. The network is assumed to employ orthogonal frequency-division multiple-access (OFDMA) which provides flexibility in power, subchannel, and rate assignment for each link. This chapter covers both the theory and the practice of cooperative network design, and makes a case that cooperative communication is a key future technology that could improve the overall capacity of wireless networks.

Throughout this chapter, it is assume that the network employs an initial channel estimation phase so that the frequency selective channel gain between any arbitrary pair of transmitter and receiver can be estimated and made known throughout the network. The assumption of channel knowledge is necessary in order to optimize the allocation of power, bandwidth and rate in the network. This chapter further assumes that channel estimation is perfect. In practical situations where channel estimation error exists, robust optimization design would be needed. The impact of imperfect channel knowledge on the resource allocation of cooperative cellular networks has been dealt with in [1, 2], but is not directly addressed in this chapter.

It should be noted that this chapter considers cooperative networks in which transmitting nodes cooperate in their transmission strategies (i.e. power, bandwidth) only, but not in actual signals. It is possible to envision a network-wide cooperative system where all the antennas from all the base-stations are pooled together as a single antenna array. Such a network multi-input multi-output (MIMO) system is capable of achieving the ultimate area-spectral efficiency limit of the network, but is outside of the scope of this chapter.

1.2 System Model

This chapter considers wireless cellular networks employing an OFDMA scheme, where the total bandwidth is divided into a large number of subchannels, and where arbitrary scheduling, as well as power, frequency and rate allocation may be made for any transmitter-receiver pairs throughout the network. The flexibility of OFDMA system in assigning resources throughout the network is one of its a key advantages, but it also presents challenge in resource optimization, as the total number of dimensions (and therefore optimization variables) is typically quite large in a realistic network. This section presents a system model for an OFDMA network and formulate an optimization problem at the network level.

1.2.1 Orthogonal Frequency-Division Multiple Access

The OFDM system is originally conceived as a solution to combat the multipath or frequency-selective nature of the wireless channel. By utilizing an N-point Inverse Fast Fourier Transform (IFFT) at the transmitter and an N-point Fast Fourier Transform (FFT) at the receiver, the available frequency band is divided into N orthogonal subchannels onto which independent data transmissions take place. The orthogonalization of frequency dimensions relies on the use of cyclic prefix, and on the assumption that the channel is stationary within each OFDM symbol, which is assumed throughout this chapter.

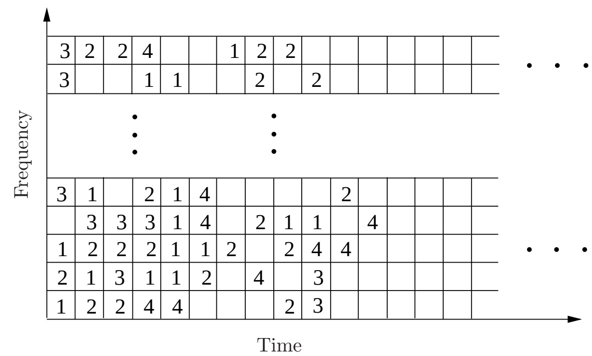

Figure 1.1 In an OFDMA system, the time and frequency dimensions are partitioned can be assigned arbitrarily to multiple users in the cell.

The OFDM system can also be thought of as a multiple-access scheme, in which multiple users may occupy orthogonal frequency subchannels without interfering with each other. For example, in a cellular network, different mobile users may communicate with the base-station on nonoverlapping sets of frequency tones, so that different users' signals are separated in frequency. This is known as OFDMA. The idea of OFDMA can also be extended to a cooperative relay network scenario in which different transmitter-receiver pairs in the network use nonoverlapping sets of frequency tones to communicate.

Orthogonalization within each cell is in general a good idea, as whenever a receiver is close in range to a non-intended transmitter, orthogonalization is needed to avoid mutual interference. The use of OFDM enables the orthogonalization in the frequency domain, which along with scheduling (which is essentially orthogonalization in the time domain) allows an arbitrary division of orthogonal dimensions among users within each cell. The assignment of dimensions can be visualized in a time-frequency map as shown in Figure 1.1.

It is implicitly assumed in the preceding discussion that when multiple transmitter-receivers pairs use OFDMA, the FFT at each receiver not only orthogonalizes the intended transmit signal, but also all the interfering signals. For this to happen, the received OFDM symbols from all transmitters must be symbol synchronized, as otherwise, a leakage would occur from one tone to its neighbouring tones [3]. For the downlink cellular setting, symbol synchronization is automatic. For the uplink, transmit timing offset can be introduced to ensure synchronization at the receiver. A more challenging case is the relay cooperative network, where it is possible to have one relay communicating with a mobile on one set of frequency tones, while another relay communicates with a different mobile on adjacent tones. In this case, simultaneous symbol synchronization at two different receivers becomes difficult. While it is possible to use advanced techniques (such as a cyclic suffix in addition to a cyclic prefix [4]) to correct for these effects, for simplicity, the rest of this chapter assumes that leakage of this type is sufficiently small, so that its effect can be ignored.

1.2.2 Adaptive Power, Spectrum and Rate Allocation

An OFDM system allows arbitrary assignment of power, modulation format and rates across the frequency domain for each transmitter-receiver pair. Assuming a fixed type of modulation, e.g. quadrature amplitude modulation (QAM), and a fixed target probability of error, the maximum bit rate in each OFDM tone is a function of the SINR on that tone. The overall achievable rate of the link is the sum of the bit rates across the tones, which can be expressed as follows:

$$R = \sum_{n=1}^{N} \log\left(1 + \frac{\mathrm{SINR}(n)}{\Gamma}\right) \tag{1.1}$$

where $\mathrm{SINR}(n)$ is the ratio of the received signal power to the noise and interference at the receiver in tone n, and Γ is the SNR gap, which is a measure of the efficiency of the particular modulation and coding scheme employed. With strong coding, Γ can be made to be close to the information-theoretical limit of 0dB, but in practical wireless system, Γ can range around 6dB-12dB. The exact value of Γ depends on the modulation scheme, coding gain, and the probability of error target.

The use of SNR gap to relate SINR with transmission rate is an approximation which is accurate for moderate and high SNRs. The exact relation between SINR and rate depends on a detailed probability of error analysis, and would give rise to complex functional forms. The value of the SNR gap approximation is that the resulting functional relation is amenable to analytic optimization, and that it closely resembles the Shannon capacity formula for the additive white Gaussian noise channel.

Note that because of the presence of intercell and intracell interference within the network, the SINRs of different links in a cellular network are interdependent. For this reason, the optimization of achievable rates over all users across the network is in general a nontrivial problem.

1.2.3 Cooperative Networks

This chapter considers two types of cooperative cellular networks: networks where multiple base-stations from different cells may cooperatively set their power allocation across the frequency tones, and networks where relay stations may be deployed to transmit and receive information from the mobile users. It is assumed that an OFDMA scheme is used within each cell, so that no two links within each cell can use the same frequency tone and at the same time. This eliminates intracell interference. Intercell interference is still present especially for cell-edge users. Given a fixed frequency and power allocation for all transmitters, the SINR for each link in every frequency tone can be easily computed as a function of the transmit powers and the direct and interfering channel path looses. The network optimization problem is then that of coordinating the allocation of frequency tones and time slots to different links within each cell and the allocation of power in each time/frequency dimension subject to total and peak power constraints, so that an overall network objective function is maximized.

1.3 Network Optimization

1.3.1 Single-User Water-filling

The OFDM transmit power optimization problem for a single link case has a classic solution known as water-filling. In this section, we briefly review the optimization principle behind water-filling, and set the stage for subsequent development for multiuser network optimization. For the single-link problem where the noise and interference are assumed to be fixed, the optimization of the achievable rate subject to a total power constraint can be formulated as

$$\begin{aligned} \text{maximize} \quad & \sum_{n=1}^{N} \log\left(1 + \frac{|h(n)|^2 P(n)}{\Gamma\sigma^2(n)}\right) \\ \text{subject to} \quad & \sum_{n=1}^{N} P(n) \le P_{total} \\ & 0 \le P(n) \le P_{max} \end{aligned} \tag{1.2}$$

where the optimization is over $P(n)$, the transmit power on the frequency tone n. The channel path loss $|h(n)|^2$ and the combined noise and interference $\sigma^2(n)$ are assumed to be fixed. The optimization is subject to a total power constraint P_{total} and a per-tone maximum power-spectral-density constraint P_{max}.

The water-filling solution arises from solving the above optimization problem via its dual. Let λ be the dual variable associated with the total power constraint, the Lagrangian of the above optimization problem is:

$$L(P(n), \lambda) = \sum_{n=1}^{N} \log\left(1 + \frac{|h(n)|^2 P(n)}{\Gamma\sigma^2(n)}\right) - \lambda\left(\sum_{n=1}^{N} P(n) - P_{total}\right) \tag{1.3}$$

The constrained optimization problem is now reduced to a unconstrained one where λ can be interpreted as the power price. Optimizing the above Lagrangian subject to peak power constraints by setting its derivative to be zero gives the following:

$$P(n) = \left[\frac{1}{\lambda} - \frac{\Gamma\sigma^2(n)}{|h(n)|^2}\right]_0^{P_{max}} \tag{1.4}$$

where $[\cdot]_a^b$ denotes a limiting operation with lower bound a and upper bound b. The optimal λ can then be found based on the total power constraint, either by a

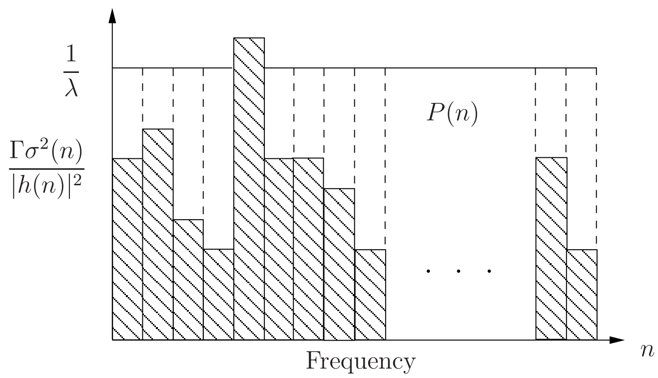

Figure 1.2 Single-user water-filling solution.

bisection or by using algorithms based on the sorting of the subchannels by their effective noises. Eq. (1.4) is the celebrated water-filling solution for transmit power optimization over a single link. The name waterfilling comes from the interpretation that the effective noise and interference $\Gamma\sigma^2(n)/|h(n)|^2$ can be thought of as the bottom of a bowl, $1/\lambda$ can be thought of as the water level, and the power allocation process can be thought of as that of pouring water into the bowl, as illustrated in Figure 1.2. The optimal power is the difference between the water level and the bottom of the bowl.

The fundamental reason that an (almost) analytic and exact solution exits for this problem is that the objective function of the optimization problem (1.2) is a concave function of the optimization variables and the constraints are linear. Therefore, convex optimization techniques such as Lagrangian dual optimization can be applied.

Modern wireless communication systems often implement adaptive power control and adaptive modulation schemes that emulate the optimal water-filling solution. It should be noted that the exact shape of the optimal power allocation is not important. If one approximates the optimal solution by a constant power allocation where all subchannels that would receive positive power in the optimal solution receive equal power in this approximate solution, the value of the objective function would be close to the optimum [5, 6]. This is because the water-filling relation, i.e. Eq. (1.4), operates on a linear scale on $P(n)$, while the rate expression, i.e. Eq. (1.2), is a logarithmic function of $P(n)$, which is not sensitive to the exact value of $P(n)$ for large value of $P(n)$. Thus, in the implementation of water-filling in practice, while it is important to identify the minimum channel-gain-to-noise ratio beyond which transmission should take place, the exact value of $P(n)$ is not as important.

In a cellular setting, whenever a particular cell implements water-filling, it changes its interference pattern on its neighbours. Thus, when every cell implements water-filling at the same time, the entire network effectively reaches a simultaneous water-filling solution, where the optimal power allocation in each cell is the water-filling solution against the combined noise and interference from all other cells. Such simultaneous water-filling solution can typically be reached via an iterative water-filling algorithm in a system-level simulation where the water-filling operation is performed on a per-cell basis iteratively [7]. Mathematically, the most general condition for convergence of such iterative algorithm is not yet known, but iterative water-filling has been observed to converge in most practical situations.

1.3.2 Network Utility Maximization

In a network with multiple users, the transmit power, bandwidth and rate allocation problem becomes considerably more complicated. This is because the achievable rates of various users become in general interdependent.

There are two consequences of this interdependency. First, the improvement in the rate of one user would in general come at the expense of other users in the network. For example, in a multicell OFDMA setup, to improve the rate of one user, one has to either increase its frequency allocation or increase its transmit power. The former case comes at the expense of the bandwidth allocation for other users within the cell. The latter comes at the expense of more interference for users in adjacent cells. In both cases, there exists a tradeoff between the rates of various users. The concept of rate region is often used to characterize such a tradeoff between the rates of various users as a function of their power and bandwidth allocations. This is often considered as a tradeoff in the physical layer.

Secondly, a fixed amount of rate improvement brings different amount of benefit to different users, depending on the application layer rate requirement. For example, a rate improvement could result in either higher video quality for one user engaged in a video-on-demand service, or a faster file transfer by a different user. The network must decide which of the two alternatives above is more preferable. Such a choice not only depends on the nature of the application, but can also depends on revenue considerations. This is often considered as a tradeoff in the application layer.

Network utility maximization (NUM) is an optimization framework that captures both the physical-layer and the application-layer tradeoffs [8]. In this framework, each user has an associated utility, which is a function of its (windowed) average rate, denoted as $U_i(\bar{R}_i)$. The utility function is increasing and assumed to be concave; it captures the desirability of having the rate $\bar{R}_i$. The network utility maximization problem is that of maximizing the sum utility subject to the achievability of these rates in the physical layer, i.e.

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{K} U_i(\bar{R}_i) \\ \text{subject to} \quad & (R_1, R_2, \cdots, R_K) \in \mathcal{R} \end{aligned} \tag{1.5}$$

where the constraint is that each instantaneous rate tuple must be inside $\mathcal{R}$, the achievable rate region at each time instance, which is defined as the convex hull of the union of all achieveable rate-tuples.

It is implicitly assumed in the above problem formulation that the utility function for all the users in the network is separable. This is a realistic assumption for the case where each user runs a separate application. In a specialized network, such a sensor network, where users collaborate in a specific task, it is conceivable that the utility of the network could depend jointly on all the rates, i.e. the objective is to maximize $U(\bar{R}_1, \bar{R}_2, \cdots, \bar{R}_K)$. A generalization of NUM in this setting has been treated in [9].

1.3.3 Proportional Fairness

A common choice of the utility function is the logarithm function. The choice of log-utility leads to a proportional fair rate allocation, which is described in detail below.

The network's objective is to maximize the sum utility of the average rates of different users in the network. The averaging is typically done in a windowed fashion or more commonly, exponentially weighted as below:

$$\bar{R}_i = (1-\alpha)\bar{R}_i + \alpha R_i \tag{1.6}$$

where $0 < \alpha < 1$ is the forgetting factor. Assuming that αR_i is small, the new contribution of the instantaneous rate R_i to the overall utility can be approximated as:

$$U_i\left((1-\alpha)\bar{R}_i + \alpha R_i\right) \approx U_i\left((1-\alpha)\bar{R}_i\right) + \left.\frac{\partial U_i}{\partial R_i}\right|_{R_i=\bar{R}_i} (\alpha R_i) \tag{1.7}$$

Under this approximation, the maximization of the sum utility, which is equivalent to the maximization of the incremental utility, becomes the maximization of weighted sum rate, where the weights are determined by the derivative of the utility function at the present rate. When the utility function is the logarithm function, the equivalent maximization problem reduces to

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{K} w_i R_i \\ \text{subject to} \quad & (R_1, R_2, \cdots, R_K) \in \mathcal{R} \end{aligned} \tag{1.8}$$

where $w_i = \frac{1}{\bar{R}_i}$. This is a weighted rate sum maximization problem for instantenous rates with weights equal to the inverse of the rate average. As these weights change over time, and as the rate region $\mathcal{R}$ changes over time (due to user mobility and the fading characteristics of the underlying fading channel), the above optimization problem needs to be solved repeatedly.

The proportional fair rate allocation is originally devised in the context of user scheduling [10]. The above discussion shows that this is also applicable to the power allocation problem. The use of proportional fairness utility is not the only way to reduce the network utility maximization to a weighted rate maximization problem. Alternatively, one may consider a system in which each user has an associated input queue, and where weights of the rate maximization problem are set as a function of the input queue length of each respective user [11, 12]. The important point is that both approaches decouples the application layer demand for rates (expressed either in the utility function or in queue length) from the physical layer provision of rates. In both cases, the physical-layer problem is reduced to a weighted rate sum maximization problem.

1.3.4 Rate Region Maximization

The reduction of the network utility maximization problem to a weighted sum rate maximization problem is a crucial step for OFDM based networks. The network utility is in general a nonlinear function of the rate of each link, which is the sum of bit rates over the frequency tones. As shown later in this section, the reduction of the utility maximization problem to a weighted sum rate maximization linearizes the objective function, which decouples the objective function on a per-tone basis and simplifies the problem significantly.

The rate region maximization problem also often has constraints that couple across the frequency tones. For example, each user may have a power constraint across the frequency. In addition, there is typically a constraint that no two users should occupy the same frequency tone within each cell. To solve the rate region maximization problem efficiently, it is important to decouple these constraints across the frequency tones as well.

A key technique to achieve such decoupling is to utilize the Lagrangian duality theory in optimization. For example, consider the case where

$$R_i = \sum_{n=1}^{N} \log \left(1 + \frac{|h_{ii}(n)|^2 P_i(n)}{\Gamma(\sum_{j \neq i} |h_{ji}(n)|^2 P_j(n) + \sigma^2(n))} \right) \tag{1.9}$$

subject to a power constraint

$$\sum_{n=1}^{N} P_i(n) \leq P_{i,total} \tag{1.10}$$

where $P_i(n)$ denote the transmit power of user i in tone n, and $h_{ij}(n)$ is the complex channel gain from the transmitter of user i to the receiver of the user j. The weighted rate sum maximization problem subject to the power constraint can be alternatively solve by dualizing with respect to the power constraint. This results in a dual function $g(\lambda_i)$ defined as follows:

$$g(\lambda_i) = \max_{P_i(n)} \left\{ w_i R_i - \lambda_i \left(\sum_{n=1}^{N} P_i(n) - P_{i,total} \right) \right\}. \tag{1.11}$$

The point is that when the objective function is a weighted rate sum and the constraint is linearized via the use of Lagrangian dual variable λ, the above optimization problem reduces to N per-tone problems:

$$\max_{P_i(n)} \left\{ w_i \log \left(1 + \frac{|h_{ii}(n)|^2 P_i(n)}{\Gamma(\sum_{j \neq i} |h_{ji}(n)|^2 P_j(n) + \sigma^2(n))} \right) - \lambda_i P_i(n) \right\} \qquad (1.12)$$

Just as in single-user water-filling, where the solution to a convex optimization problem reduces to solving the problem for each λ, then finding the optimal λ, a similar algorithm based on λ-search is applicable here. The reduction to N per-tone optimization problem ensures that the computational complexity for each step of this optimization problem with fixed λ_i is linear in the number of tones.

The theoretical justification for the above duality approach is convexity. For convex optimization problems where the feasible set has a non-empty interior (which is almost always true in engineering applications), the maximum value of the original objective is equal to the minimum of the dual optimization problem

$$\min_{\lambda_i \geq 0} g(\lambda_i). \qquad (1.13)$$

The optimum λ_i can be found using search based on the ellipsoid method (which is a generalization of bisection search to higher dimensions.)

An interesting fact is that this duality technique remains applicable even when the functional form of the rate expression is nonconvex as is the case in Eq. (1.9), as long as the OFDM system has a large number of dimensions in the frequency domain, which allows an effective convexification of the achievable rate region as a function of the power allocation. A rigorous proof of this statement can be found in [13, 14]. This fact allows the duality technique to be used in a wide variety of applications.

To summarize, under the network utility maximization framework, the network optimization problem for an OFDMA network under per-user power constraints reduces to a per-tone weighted sum rate maximization problem with a linear power penalty term. The weights in weighted rate sum maximization are determined by the utility function. The power penalty weighting can be found using a generalization of bisection. The rest of this chapter is devoted to two examples of cooperative networks where the adaptive scheduling, and power, frequency and rate allocation problem can be solved efficiently in OFDMA networks.

1.4 Network with Base-Station Cooperation

1.4.1 Problem Formulation

Consider an OFDMA-based cellular network as shown in Figure 1.3 in which base-stations cooperate in setting their downlink transmit power and the mobiles likewise jointly set their uplink transmit power in order to avoid excessive interference between the neighbouring cells. The optimal design of this multicell coop-

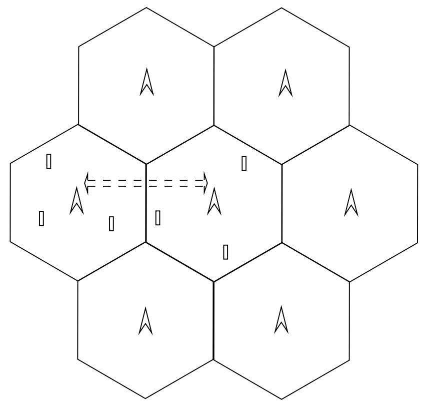

Figure 1.3 A cellular network in which the base-stations cooperate in user scheduling and power allocation across the frequency spectrum.

erative network becomes that of designing a joint scheduling and power allocation scheme that decides in each frequency tone:

- Which user should be served in each cell?
- What is the appropriate uplink and downlink transmit power levels?

Scheduling can be thought of either as the optimal partitioning of the frequency among the users within each time, or as the optimal assignment of users in each time slot for each frequency tone. Scheduling and power allocation need to be considered jointly to reach an optimal solution for the entire network.

Assuming a proportional fairness objective function, the network optimization problem for the downlink is a weighted rate sum maximization problem

$$\max \sum_{l=1}^{L} \sum_{k=1}^{K} w_{D,lk} R_{D,lk} \tag{1.14}$$

where $R_{D,lk}$ is the instantaneous downlink rate of kth user in lth cell in a network consisted of L cells with K users per cell. The weights $w_{D,lk} = \frac{1}{\bar{R}_{D,lk}}$ are the proportional fairness variable determined by the exponentially weighted average rate. Let $k = f_D(l,n)$ be the downlink scheduling function, which assigns a user k to the nth frequency tone in the lth base-station. The downlink rate expression $R_{D,lk}$ is then

$$R_{D,lk} = \sum_{n \in \mathcal{D}_{lk}} \log\left(1 + \frac{P_{D,l}^n |h_{llk}^n|^2}{\Gamma(\sigma^2 + \sum_{j \neq l} P_{D,j}^n |h_{jlk}^n|^2)}\right) \tag{1.15}$$

where the summation is over frequency tones assigned to the kth user in the lth cell, i.e. $\mathcal{D}_{lk} = \{n | k = f_D(l,n)\}$. Here h_{jlk}^n is the channel transfer function from the jth base-station to the kth user in the lth cell and the nth tone, and $P_{D,l}^n$ is the downlink power allocation for the lth base-station in nth tone. The optimization is over $P_{D,l}^n$. The weighted rate maximization problem is to be solved under the per base-station power constraint for the downlink

$$\sum_{n=1}^{N} P_l(n) \leq P_{l,total} \tag{1.16}$$

as well as possibly peak power constraints:

$$0 \leq P_{D,l}^n \leq S_D^{max}. \tag{1.17}$$

In addition, there is the OFDMA constraint that no two users should occupy the same frequency tone within each cell. Finally, note that a similar optimization problem with corresponding rate and power expressions can be written down for the uplink.

The joint scheduling and power allocation problem formulated above has been studied in several recent works [15, 16, 17], where key ideas such as iterative optimization of scheduling and power allocation and numerical methods for power adaptation have been proposed. The following section outlines the approach based on these recent works and provides a performance projection for networks with base-station power cooperation.

1.4.2 Joint Scheduling and Power Allocation

The dual decomposition technique outlined in the previous section can be used to tackle the joint scheduling and power allocation above. The key fact is that after dualizing with respect to the total power constraint, the optimization problem decouples on a tone-by-tone basis:

$$\begin{aligned} \text{maximize} \quad & \sum_l w_{D,lk} \log\left(1 + \frac{P_{D,l}^n |h_{llk}^n|^2}{\Gamma(\sigma^2 + \sum_{j \neq l} P_{D,j}^n |h_{jlk}^n|^2)}\right) - \lambda_{D,l} P_{D,l}^n \\ \text{subject to} \quad & 0 \leq P_{D,l}^n \leq S_D^{max} \qquad \forall l. \end{aligned} \tag{1.18}$$

where the optimization is over both the power variables $P_{D,l}^n$ as well as the scheduling function $k = f_D(l,n)$ across the L base-stations for a fixed tone n.

The duality theory for OFDMA networks states that if the above per-tone optimization problem can be solved exactly for each set of fixed $\lambda_{D,l}$'s, an ellipsoid or subgradient search over $\lambda_{D,l}$ in an outer loop can be carried out to find the optimal $\lambda_{D,l}$, which then leads to the global optimum of the overall network optimization problem.

The optimization problem (1.18) is a mixed integer programming problem with nonconvex objective function. Finding the global optimum for such an optimization problem is known to be a difficult task. However, many approximation algorithms exist that can reach at least a local optimum. Although strictly speaking, the duality theory for OFDMA networks requires the global optimum solution of the per-tone optimization problem, the local optimum solution already works quite well in practice. The rest of this section focuses on solving (1.18) using local optimal approaches.

Observe first that for the downlink, the scheduling step and the power allocation step can be carried out separately. This is because the scheduling choices at each cell does not affect the amount of intercell interference on its neighbours. The intercell interference is a function of the power allocation only. Thus, an iterative algorithm can be devised so that one can find the best schedule for a fixed power allocation, then find the best power allocation for the fixed schedule [15]. The iteration always increases the objective function monotonically, so it is guaranteed to converge to at least a local optimum of the joint scheduling and power allocation problem.

For the downlink, because the intercell interference is independent of the scheduling decisions at each cell, finding the best schedule for a fixed power allocation is a per-cell optimization problem. In other words, each base-station only needs to find the user in each tone that maximizes the weighted rate. This amounts to a simple search among the K users.

For a fixed user schedule, the optimal power allocation problem becomes that of solving (1.18) for fixed k's. This is a nonconvex problem with potentially multiple local optima. The first-order condition for this optimization problem can be found by taking the derivative of the objective function and setting it to be zero:

$$\frac{w_{D,lk}|h^n_{llk}|^2}{P^n_{D,l}|h^n_{llk}|^2 + \Gamma(\sigma^2 + \sum_{j\neq l} P^n_{D,j}|h^n_{jlk}|^2)} = \sum_{j\neq l} t^n_{D,jl} + \lambda_{D,l}, \tag{1.19}$$

where $k = f_D(l,n)$ for $l = 1, \cdots, L$, and

$$t^n_{D,jl} = w_{D,jk'} \frac{\Gamma|h^n_{ljk'}|^2}{P^n_{D,j}|h^n_{jjk'}|^2} \left(\frac{(\mathrm{SINR}^n_{D,j})^2}{1+\mathrm{SINR}^n_{D,j}} \right), \tag{1.20}$$

and

$$\mathrm{SINR}^n_{D,j} = \frac{P^n_{D,j}|h^n_{jjk'}|^2}{\Gamma(\sigma^2 + \sum_{i\neq j} P^n_{D,i}|h^n_{ijk'}|^2)}, \tag{1.21}$$

with $k' = f_D(j,n)$.

The first-order condition gives a water-filling like condition if the terms $t^n_{D,jl}$ are considered to be fixed. In this case, (1.19) suggests that the following power allocation is a local optimum of the per-tone optimization problem:

$$P^n_{D,l} = \left[\frac{w_{D,lk}}{\sum_{j\neq l} t^n_{D,jl} + \lambda_{D,l}} - \frac{\Gamma(\sigma^2 + \sum_{j\neq l} P^n_{D,j}|h^n_{jlk}|^2)}{|h^n_{llk}|^2} \right]_0^{S_D^{max}} \tag{1.22}$$

where $k = f_D(l,n)$. Note that this is similar to the single-user water-filling power allocation (1.4), except that the power is allocated with respect to the combined noise and interference, and that the water-filling level $\lambda_{D,l}$ is modified by the

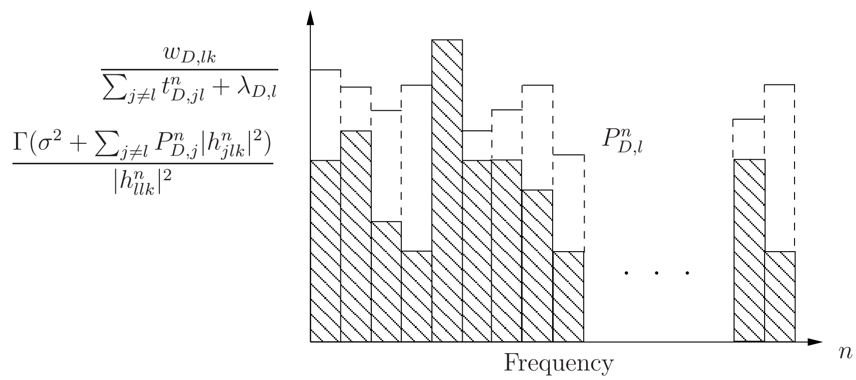

Figure 1.4 Water-filling where the water-filling level is modified by the $t^n_{D,jl}$ pricing terms.

additional $t^n_{D,jl}$ terms. This process is called modified water-filling [18] and is illustrated in Figure 1.4.

The $t^n_{D,jl}$ term can be interpreted as the summary of the effect of allocating additional power at the lth base-station on the downlink rate at the neighbouring jth cell. A larger value of $t^n_{D,jl}$ signals a larger effect of interference from the lth cell to the jth cell. The multiuser water-filling condition in (1.22) implies that when interference is present, the water-filling level needs to be modified. The water-filling level should decrease if the effect of interference is strong, which suggests that the power allocation should be reduced. Note that the water-filling level is also affected by the proportional fairness weights $w_{D,lk}$. A larger weight suggests a higher water-filling level.

The terms $t^n_{D,jl}$ also have a pricing interpretation [19, 20, 21, 22], which comes from the fact that $t^n_{D,jl}$ is the derivative of the jth base-station's data rate with respect to the lth base-station's power, weighted by the proportional fairness variable. A higher value of $t^n_{D,jl}$ suggests that the lth base-station must pay a high price for allocation its power in tone n, which is reflected in the modification of the water-filling level.

The water-filling condition (1.22) suggests that one way to coordinate multiple base-stations in a cooperative cellular network is to allow base-stations to exchange values of $t^n_{D,jl}$ with their neighbours. Note that the value of $t^n_{D,jl}$ depends on the ratio of the direct and the interfering channel gains, which can be easily estimation using pilot signals.

Knowing $t^n_{D,jl}$, each base-station may use (1.22) to update its power allocation. This results in an iterative process. When it converges, it will reach a local optimum of the weighted rate sum maximization problem (1.18). This procedure is known as the modified iterative water-filling algorithm [18]. In practice, it may be necessary to damp the iteration to ensure convergence [17].

Alternatively, one may resort to a direct numerical optimization of (1.18) [17, 16]. Starting from an initial power allocation, one may compute a gradient or Newton's increment direction for the optimization objective, then successively improve the objective function until a local optimum is reached. The gradient can again be computed located at each base-station based on the pricing terms $t^n_{D,jl}$.

To summarize for the downlink, the coordination of base-stations can be efficiently implemented using an approach that iterates between coordinated scheduling and coordinated power allocation. The scheduling step for the downlink can be efficiently implemented on a per-cell bases; the power allocation step can be implemented if certain exchange of pricing information is allowed among the base-stations. This iterative process, together with an outer loop that finds the optimal power prices $\lambda_{D,l}$, finds a local optimum of the weighted rate sum maximization problem.

Much of the discussion in this section is also applicable to the uplink, except that optimal scheduling is no longer a per-cell problem. In the uplink, the assignment of users in each cell directly affects the interference in neighbouring cells, so an optimal uplink scheduler needs to consider the effect of the interference as well. However, there is evidence suggesting that if one uses identical schedulers for both the uplink and the downlink, the network often already performs very well [17]. This can be justified in part by the fact that there is a duality between uplink and downlink channels. The capacity regions of the uplink and downlink channels are identical under the same power constraint.

1.4.3 Performance Evaluation

To illustrate the performance of the proportionally fair joint scheduling and power allocation method described in the previous section, this section presents simulation results on a multicell network with base-station cooperation. The simulated network consists of 19 cells hexagonally tiled with 40 users per cell, occupying a total bandwidth of 10 MHz partitioned into 256 subchannels using OFDMA. For simplicity, a maximum transmit power spectral density (PSD) of -27 dBm/Hz is imposed at both the base-stations and the remote users, but no total power constraint is imposed. A multipath fading channel model is used with 8dB of log-normal shadowing. The channel path loss is modelled as a function of distance d as $128.1 + 37.6\log_{10}(d)$ (in dB). The background noise level is assumed to be -169 dBm/Hz.

The joint proportionally fair scheduling and adaptive power allocation is expected to provide the largest performance improvement for users at the cell edge where intercell interference is dominant. To illustrate the performance gain for cell-edge users, in the simulation below users are placed at the cell edge on purpose. Table 1.1 illustrates a comparison of the achievable sum rates over all users in 19 cells for the adaptive power allocation algorithm vs. the constant transmit PSD scheme with proportionally fair scheduling. These results have

Table 1.1. Uplink (UL) and downlink (DL) sum rates over 19 cells with 40 cell-edge users per cell with proportional fairness joint scheduling and cooperative power allocation among the base-stations [17].

Base-to-base Distance	2.8km		1.4km	
	UL	DL	UL	DL
Fixed Power Spectrum	125 Mbps	129 Mbps	137 Mbps	142 Mbps
Adaptive Power Spectrum	185 Mbps	181 Mbps	228 Mbps	227 Mbps
Improvement	48%	40%	66%	60%

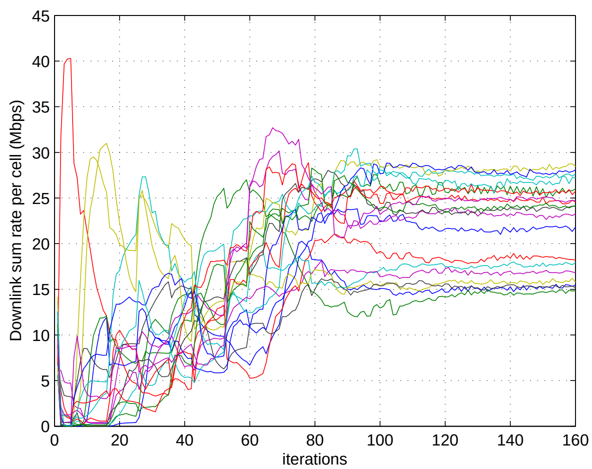

Figure 1.5 Convergence of downlink sum rates in each of the 19 cells using modified iterative water-filling with proportional fairness joint scheduling and power allocation.

been reported in [17] and are consistent with other studies in this area [16]. It can be seen that depending on the base-station to base-station distance, a sum rate improvement between 40%-60% is possible. The improvement is larger when base-stations are closer, because in this case the intercell interference is also larger. It is worth emphasizing that the sum rate improvement reported in Table 1.1 is for cell-edge users. If averaged over all users uniformly placed over the cell, the sum rate improvement would have been about 15%-20%.

Figure 1.5 illustrates the convergence behaviour of the joint proportional fair scheduling and power allocation algorithm. Each iteration here consists of either an adaptive power allocation step or a scheduling step. Up to 10 sub-iterations are performance within each power allocation step. The sum rates of each of the 19 cells are plotted. Note that the proportional fairness weights are also updated in each iteration. These weights ensure that rates are allocated to all users with fairness.

The simulation results clearly illustrate the value of coordinating base-station power spectral densities in an interference limited multicell environment. The projected performance improvement is obtained by allowing base-stations to exchange pricing information with each other, and by iteratively converging to a joint network-wide optimum.

1.5 Cooperative Relay Network

Base-station cooperation addresses the intercell interference for cell-edge users in a cellular network, but the cell-edge users' performances are also fundamentally affected by the path loss, which is distance dependent. A viable approach to deal with the path loss is to deploy relay stations throughout the cell, so that a mobile user may connect to the base-station via the relay, thereby reducing the effective path distance. This section addresses the network optimization problem for the cooperative relay network.

The resource allocation problem for the cooperative relay network has attracted much attention in the wireless cellular communication literature (e.g. [23, 24, 25, 26]). There are many different ways in which a relay may help the communication between a pair of transmitter and receiver (also known as the source and the destination in the relay literature). In a decode-and-forward protocol, the relay decodes the message from the source then reencode and transmit to the destination. Alternatively, a relay may amplify and forward, or quantize its observation and forward to the destination.

In general, decode-and-forward is a sensible strategy when the relay is located closer to the source than to the destination, while amplify-and-forward and quantize-and-forward are more suitable when the relay is closer to the destination. However, the question of which strategy is the most suitable is a complicated one, as it also depends on the power allocation at the source and at the relay, as well as the end-user's rate requirement or its utility function. An optimization framework for choosing the best cooperation strategy has been dealt with for a single-relay link in [27], but the general optimization of relay strategies for a cellular network is likely to be computationally complex.

The chapter focuses instead on a simplified model where only the decode-and-forward protocol is used. This is done for the following reasons. First, the primary focus of this chapter is the use of relay for enhancing cellular coverage at the cell edge, in which case a sensible relay location within the cell is somewhere close to the half-way point between the base-station and the cell edge. In this case, for both uplink and downlink transmissions, the distances of both the source-relay and the relay-destination paths are about the same, making decode-and-forward a suitable strategy. Second, decode-and-forward offers a digital approach to relaying. It eliminates the noise enhancement problem inherent in amplify, or quantize-and-forward. Third, this chapter considers the deployment of fixed infrastructure-based relay stations. These relay stations typically have the computational resources to perform decoding and reencoding.

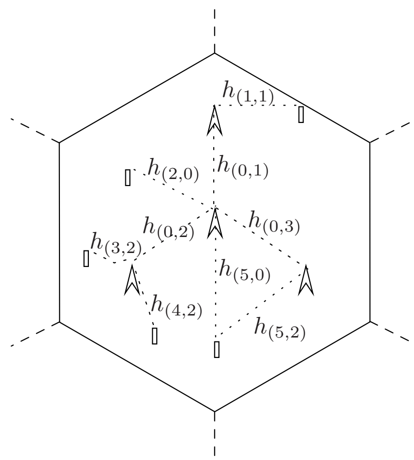

Figure 1.6 A cooperative relay network in which the mobiles may connect directly to the base-station or through the relays.

Further, as the primary focus here is the use of relay to combat distance-dependent path loss, this chapter restricts attention to a two-hop relay strategy, where the direct path from the source to the destination is ignored (as it is typically very weak), and the relay is a simple repeater. In the first hop, the source transmits information to the relay. In the second hop, the relay decodes and retransmits the same information to the destination.

Under these assumptions, the capacity of a single source-relay-destination link is simply the minimum of the source-relay and the relay-destination link capacities. The characterization of capacity becomes more involved if one considers the possibility that a single relay deployed in a cellular network may help multiple mobiles at the same time. Further, each mobile has a choice of either connecting to the base-station direction, or through relays. The mobile may even choose to use different mobiles for different frequency tones. These possibilities are coupled with the problem of bandwidth and power allocation across the frequency tones.

The rest of this section uses a network optimization framework introduced earlier and provides a solution based on the duality theory to solve the bandwidth, rate and power allocation problem for OFDMA relay networks. This methodology used here is the one as first proposed in [23] and also in [26].

1.5.1 Problem Formulation

Consider a wireless cellular network in which each cell is equipped with M relay stations located at the midpoint between the base-station and the cell edge and at angles $360/M$ degrees apart from each other. There are K mobiles in each cell. Each mobile may either connect directly to the base-station or through one of the relays in each frequency tone, (but the mobile can possibly use different relays in different tones.) There are a total of $M+1$ links emanating from each mobile. These mobile-originated links, plus the M links connecting the relays to the base-station, gives a total of $K(M+1)+M$ links in the entire cellular network.

Label the base-station as node 0, and the mobiles as nodes $k=1,\cdots,K$. Label the links by a pair of indices as follows: the base-station to relay links are labelled as $(0,m)$, with $m=1,\cdots,M$; the mobile to base-station links are labelled as $(k,0)$ and the mobile to relay links are labelled as (k,m) with $k=1,\cdots,K$ and $m=1,\cdots,M$. This labelling convention is illustrated in Figure 1.6.

This chapter considers a setup in which each cell employs an OFDMA scheme. Further, it is assumed that in each frequency tone, at most one link may be active at any given time. This assumption allows the intracell interference to be avoided completely, and simplifies the numerical solution considerably. For simplicity, the chapter also assumes that scheduling and power allocation are done on a per-cell basis (i.e. without base-station cooperation). The problem formulation presented here can be extended to a more general setting where spatial reuse is eabled within each cell, or where intercell cooperation is enabled across the cells, but the resulting optimization problem would become considerably more complex.

Consider now the uplink scenario. Define the scheduling function as a mapping from the frequency tone to the link index, i.e., $f_U(n)=(i,j)$. With the assumptions stated above, the achievable rate for each link (i,j), denoted as $r_{U,(i,j)}$ can be expressed as

$$r_{U,(i,j)} = \sum_{n\in\mathcal{U}_{(i,j)}} \log\left(1+\frac{P^n_{U,(i,j)}|h^n_{(i,j)}|^2}{\Gamma\sigma_n^2}\right) \tag{1.23}$$

where the summation is over all frequency tones assigned to that link, i.e. $\mathcal{U}_{(i,j)} = \{n|(i,j)=f_U(n)\}$, $P^n_{(i,j)}$ denotes the transmit power and $|h^n_{(i,j)}|^2$ denotes the channel gain for the link (i,j) at tone n, σ_n^2 denotes the combined intercell interference and noise.

The achievable rate for each user, denoted as R_k, is the sum of achievable rates of all links emanating from the mobile, i.e.

$$R_{U,k} = \sum_{j=0}^{M} r_{U,(k,j)}. \tag{1.24}$$

At each relay, a flow conservation constraint must be satisfied so that all the incoming traffic can be forwarded to the base-station. This results in M constraints as follows:

$$\sum_{k=1}^{K} r_{U,(k,m)} \le r_{U,(0,m)}. \tag{1.25}$$

The above equation is an example of the general flow conservation formulation [28, 23].

It is now straightforward to write down the uplink per-cell optimization problem for the cooperative relay network, which consists of both the allocation of power and bandwidth for each link, and the routing of the information within each cell. Under the network utility maximization framework, the optimization problem can be reduced to a weighted rate sum maximization problem across the K users with weights $w_{U,k} = \frac{1}{R_{U,k}}$:

$$\begin{aligned}
\text{maximize} \quad & \sum_{k=1}^{K} w_{U,k} \left(\sum_{m=0}^{M} r_{U,(k,m)} \right) \\
\text{subject to} \quad & \sum_{k=1}^{K} r_{U,(k,m)} \leq r_{U,(0,m)}, && \forall m = 1 \cdots M \\
& \sum_{n=1}^{N} p^n_{U,(0,m)} \leq P^{\max}_{U,R,m}, && \forall m \\
& \sum_{m=0}^{M} \sum_{n=1}^{N} p^n_{U,(k,m)} \leq P^{\max}_{U,M,k}, && \forall m \\
& 0 \leq p^n_{U,(k,m)} \leq S^{\max}_{U,(k,m)}, && \forall k, m, \quad \forall n \\
& p^n_{U,(k,m)} p^n_{U,(k',m')} = 0 && \forall (k,m) \neq (k',m'), \quad \forall n
\end{aligned} \tag{1.26}$$

where $r_{U,(k,m)}$ is as expressed in (1.23), and the optimization is over power allocations $p^n_{U,(k,m)}$, which is subject to the per-mobile total power constraint $P^{\max}_{U,M,k}$, the per-relay total power constraint $P^{\max}_{U,R,m}$, as well as the peak power-spectral-density constraints at both the mobiles and the relays $S^{\max}_{U,(k,m)}$. The last constraint ensures that no two links share the same frequency tone within each cell. Note that the downlink problem can be formulated in a similar fashion.

1.5.2 Joint Routing and Power Allocation

The network utility maximization problem for the cooperative relay network is essentially a joint routing and power allocation problem, as each mobile has the option of either transmitting its information bits directly to the base-station or routing through one or more of the relays.

One way to solve this problem is to dualize with respect to the flow conservation constraint so that the objective function becomes a new weighted rate sum maximization problem over all link rates (rather than the end-user rates in (1.26)). The new objective function is now

$$\begin{aligned}
& \sum_{k=1}^{K} w_{U,k} \left(\sum_{m=0}^{M} r_{U,(k,m)} \right) - \sum_{m=1}^{M} \mu_m \left(\sum_{k=1}^{K} r_{U,(k,m)} - r_{U,(0,m)} \right) \\
& \qquad = \sum_{m=1}^{M} \mu_m r_{U,(0,m)} + \sum_{k=1}^{K} \sum_{m=0}^{M} (w_{U,k} - \mu_m) r_{U,(k,m)}
\end{aligned} \tag{1.27}$$

subject to the peak and total power constraints in (1.26). Let $g(\mu_1, \cdots, \mu_M)$ denote the maximum value of (1.27) subject to power constraints for any fixed set of μ_m's. Because of the zero-duality-gap property of the OFDMA system, the solution to the original problem then reduces to the maximization of $g(\mu_1, \cdots, \mu_M)$ over all μ_m's.

The dual variables μ_m's enter the new objective function as weights to the weighted rate sum maximization problem over link rates. Roughly speaking, a higher value for μ_m indicates congestion in the link between the base-station and the mth relay, and that more rate should be allocated to that link to release congestion. Similarly, a lower value of μ_m indicates the opposite.

The weighted link rate sum maximization problem subject to power constraints can itself be solved by yet another dual decomposition step with respect to the total power constraints, as treated earlier in this chapter. In this case, the optimization problem is completely decoupled on a tone-by-tone basis. Because of the assumption that only one link may be active in any given time slot and frequency tone, the weighted rate sum maximization then reduces to the selection of the best link for each frequency tone, which involves a simple search.

Finally, the optimization of $g(\mu_1, \cdots, \mu_M)$ over all μ_m's can be handled by either an outer loop using the ellipsoid or the subgradient method. The search over the optimal set of μ_m's balances the incoming and outgoing flows at each relay.

1.5.3 Performance Evaluation

The optimization framework described above is used to evaluate the effectiveness of deploying relay stations in a cellular network. To take into account the cost of relay deployment, the performance of a baseline system with cell diameter of 1.4km is compared with a relay network in which the cell area is doubled (with diameter 1.98km), but with 3, 4 or 5 relays deployed within each cell. The rationale is that if a relay station costs roughly 1/3 to 1/5 of a base-station, then the deployment cost of both systems would be approximately the same. In both cases, the achievable uplink transmission throughput is computed for a 7-cell system with users uniformly placed within the area. The user densities in both cases are the same. When the relays are deployed, they are located between $\frac{1}{3}r$ to $\frac{2}{3}r$ from the base-station, symmetrically in the angular direction. Again, standard cellular channel models are used with both distance-dependent path loss and log-normal shadowing components. These simulation results have been presented in [26].

The simulation results presented in this section are based on the weighted rate-sum maximization formulation. However, instead of maximizing a proportionally fair utility function, the results here pertains to a maximization of the minimum rate over all users in the system (similar to [23]), which requires a slight modification of the optimization problem (1.26). First, additional constraints that each user must have a rate larger than some minimal rate $R_{\min}$ are added, where $R_{\min}$

Table 1.2. The achievable minimum and sum rates for a 7-cell network: the baseline vs. relay scenarios (RS) with varying number of relays per cell and relay locations [26].

Scenario	Baseline	RS 1	RS 2	RS 3	RS 4
Relays Per Cell	0	3	4	5	3
Relay Distance	n/a	$\frac{2}{3}r$	$\frac{2}{3}r$	$\frac{2}{3}r$	$\frac{1}{3}r$
Mobiles Per Cell	9	18	20	20	18
Cell Diameter (km)	1.4	1.98	1.98	1.98	1.98
Cell Area (km^2)	1.54	3.08	3.08	3.08	3.08
Mininum Rate (Mbps)	0.193	0.583	0.972	0.705	0.578
Sum Rate (Mpbs)	96.4	75.4	80.2	72.7	87.2

is a constant. Then, the resulting optimization problem is solved with successively larger $R_{\min}$'s until the problem becomes infeasible. The largest such $R_{\min}$ is the maximum minimal rate. In practice, the maximum $R_{\min}$ can be found efficiently using a bisection.

For simplicity, the adaptive scheduling, and power and rate allocation here is implemented on a per-cell basis. For simulation purposes, the network throughput is computed using an iterative approach, in which the intercell interference is updated in each iteration, and the cellular network eventually reaches an equilibrium.

Table 1.2 shows the maximum minimal rate for the baseline network without relays and a number of relay scenarios. These are the results as first presented in [26]. The most interesting feature here is that the addition of relays to the infrastructure improves the minimal rate dramatically, however it does not make much a difference in the sum rate at all. This illustrates that the benefit of relays concentrates on users at the cell-edge. As far as the sum rate is concerned, when the users are uniformly distributed throughout the cell, the sum rate would be dominated by the rate of those users closest to the base-station, which are not helped by the relays.

1.6 Conclusions

This chapter presents a network utility maximization framework for cooperative networks employing OFDMA. It is shown that the objective of maximizing the utility function of multiple users in a multicell network can be efficiently carried out using various techniques, including proportional fairness scheduling, dual optimization, descent method for local optimization, and the network flow conservation principle. A central observation here that because the OFDM scheme partitions the frequency domain into many parallel subchannels, the network utility maximization problem often decomposes into a tone-by-tone optimization problem, which is considerably easier to solve.

This chapter focuses on two types of cooperative networks, and formulates the corresponding joint scheduling, power adaptive and rate allocation problems in each case. For networks with base-station cooperation, it is shown that adaptively adjusting power allocation across the base-stations has the effect of reducing intercell interference, hence improving the overall throughput of the network. Intercell interference can also be reduced by deploying relays throughout the cells. The relays have the effect of enhancing the coverage at the cell edge, which improves the minimal service rate within each cell.

Base-station cooperation and relay deployment are technologies with the potential to significantly enhance the performance of traditional wireless cellular network structure, especially at the cell edge. The benefits brought by these cooperative techniques are particularly valuable to network service providers, because cell-edge users are the bottleneck in the current generation of wireless networks.

References

[1] I. C. Wong and B. L. Evans, "Optimal resource allocation in the ofdma downlink with imperfect channel knowledge," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 232–241, Jan. 2009.

[2] Y. Cui, V. K. N. Lau, and R. Wang, "Distributive subband allocation, power and rate control for relay-assisted OFDMA cellular system with imperfect system state knowledge," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5096–5102, Oct. 2009.

[3] V. M. K. Chan and W. Yu, "Multiuser spectrum optimization for discrete multitone systems with asynchronous crosstalk," *IEEE Trans. Signal Processing*, vol. 55, no. 11, pp. 5425–5435, Nov. 2007.

[4] F. Sjöberg, M. Isaksson, R. Nilsson, P. Ödling, S.K. Wilson, and P.O. Börjesson, "Zipper: a duplex method for vdsl based on dmt," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1245 –1252, aug 1999.

[5] P. S. Chow, *Bandwidth optimized digital transmission techniques for spectrally shaped channels with impulse noise*, Ph.D. thesis, Stanford Univ., Stanford, CA, 1993.

[6] W. Yu and J. M. Cioffi, "Constant power water-filling: Performance bound and low-complexity implementation," *IEEE Trans. Commun.*, vol. 54, no. 1, pp. 23–28, Jan. 2006.

[7] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Select. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, June 2002.

[8] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[9] J. Yuan and W. Yu, "Joint source coding, routing and power allocation in wireless sensor networks," *IEEE Trans. Commun.*, vol. 56, no. 6, pp. 886–898, June 2008.

[10] E. F. Chaponniere, P. J. Black, J. M. Holtzman, and D. N. C. Tse, "Transmitter directed, multiple receiver system using path diversity to equitably maximize throughput," U.S. Patent 6,449,490, filed July 1999.

[11] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Automat. Contr.*, vol. 37, no. 12, pp. 1936–1949, Dec. 1992.

[12] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.

[13] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 1310–1322, June 2006.

[14] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Select. Areas Signal Processing*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[15] L. Venturino, N. Prasad, and X. Wang, "Coordinated scheduling and power allocation in downlink multicell OFDMA networks," *IEEE Trans. Veh. Technol.*, vol. 6, no. 58, pp. 2835–2848, July 2009.

[16] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *INFOCOM*, Apr. 2009.

[17] W. Yu, T. Kwon, and C. Shin, "Joint scheduling and dynamic power spectrum optimization for wireless multicell networks," in *Proc. Conference Info. Science Sys. (CISS)*, Princeton, NJ, U.S.A., Mar. 2010.

[18] W. Yu, "Multiuser water-filling in the presence of crosstalk," in *Inform. Theory and Appl. Workshop*, San Diego, U.S.A., 2007.

[19] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 5, May 2006.

[20] C. Shi, R. A. Berry, and M. L. Honig, "Distributed interference pricing for OFDM wireless networks with non-separable utilities," in *Proc. Conference Info. Science Sys. (CISS)*, Mar. 2008, pp. 755–760.

[21] F. Wang, M. Krunz, and S. Cui, "Price-based spectrum management in cognitive radio networks," *IEEE J. Sel. Top. Signal Processing*, vol. 1, no. 2, pp. 74–87, Feb. 2008.

[22] J. Yuan, *Optimization Techniques for Wireless Networks*, Ph.D. thesis, University of Toronto, Toronto, Canada, 2007.

[23] S.-J. Kim, X. Wang, and M. Madihian, "Optimal resource allocation in multi-hop ofdma wireless networks with cooperative relay," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1833–1838, May 2008.

[24] R. Kwak and J. M. Cioffi, "The subchannel-allocation for OFDMA relaying downlink systems with total power constraint," in *Proc. IEEE Globecom*, Dec 2008.

[25] T. Ji, D. Lin, A. Stamoulis, A. Khandekar, and N. Bhushan, "Relays in heterogeneous networks," in *Information Theory and Applications Workshop*, Feb 2009.

[26] J. Ji and W. Yu, "Bandwidth and routing optimization in wireless cellular networks with relays," in *Proc. The 5th workshop on Resource Allocation, Cooperation and Competition in Wireless Networks (RAWNET/WNC3)*, Seoul, Korea, June 2009.

[27] T. C.-Y. Ng and W.Yu, "Joint optimization of relay strategies and resource allocations in a cooperative cellular network," *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 328–339, Feb. 2007.

[28] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, pp. 1136–1144, July 2004.