Abstract—The next generation of wireless networks (4G) will use OFDMA (Orthogonal Frequency-Division Multiple Access) technology on both forward and reverse links. In addition to high speed data services, these networks will also support voice using the Voice Over IP (VoIP) protocol. In this paper we focus on the capacity of VoIP service for the OFDMA uplink. Although resources allocated to users within a sector are orthogonal, transmissions are affected by interference from transmissions in neighboring sectors. Therefore the user capacity is determined by the uplink bandwidth, the maximum transmission power of the mobile station, the uplink channel conditions, the background noise and the intersector interference. As the transmission power of each mobile increases, the interference caused also increases. Eventually the cell edge users can no longer support the rate required for the VoIP connection and goes into outage. For a given resource utilization per active user we determine the maximum number of users that can be supported. We first determine the centralized solution and then show that a simple distributed algorithm can also achieve this optimal solution. We then show how user capacity can be improved by varying the resource utilization of users based on their channel conditions. Numerical results are provided to illustrate algorithm performance.

I. INTRODUCTION

In the uplink of an OFDMA network [1], [2], the transmission power of each Service Subscriber (SS) (or sometimes simply called user) must be chosen so as to provide the desired rate of the user while at the same time maintaining low interference levels in adjacent sectors. For best effort traffic this can be accomplished through rate adjustments but for real time services where a delay constraint must be met (i.e., the rate must be maintained) then admission controls are needed.

Although papers have been written on this topic for CDMA networks [3], fewer have been written for OFDMA networks. Furthermore, the work that has been done for OFDMA networks consider the case of data traffic where transmission powers can be reduced in response to high interference levels in neighboring sectors. However, in the case of VoIP, transmission powers cannot be reduced since this will affect the required performance level of the service (in terms of packet loss rate and/or delay). Therefore, interference management is not feasible for VoIP traffic and so admission controls must be used to maintain acceptable service levels for those users already being served. In this paper we focus on the factors that affect VoIP capacity.

In the OFDMA uplink, frequency-time resource blocks (called tiles) are allocated to users. Each such tile consists of a subset of subcarriers and symbols. Since each tile is allocated to at most one user (i.e., we do not consider the case of Space Division Multiple Access in this paper) then transmissions within the sector are orthogonal. However, since users in adjacent sectors may also use this tile then those transmissions interfere with the transmission within the concerned sector. In the case of VoIP, if the interference is sufficiently high then even if the user were to transmit with maximum power it may not be able to achieve the required received SINR. This results in excessive frame losses and/or unacceptable delays causing the user to go into the outage state (i.e., cannot achieve acceptable performance levels). VoIP capacity is defined as follows. A user is said to be in outage if the probability of being in the outage state exceeds some specified threshold. User capacity is defined as the maximum number of users that can be supported such that no more than a specified percentage of users are in outage. In addition, users must also satisfy certain delay and jitter requirements. We assume that users are provided persistent resource assignments and that sufficient power is used to ensure that the probability that a VoIP packet is delayed is sufficiently small so that the delay and jitter requirements are satisfied.

Downlink power allocation has been investigated by Foschini [4] where it is shown that a simple distributed algorithm converges to the optimal power allocation. Uplink interference management was addressed in Hosein [5] and Rao [6] but for the case of data traffic. Because of the delay tolerant nature of data traffic it is possible to adjust each user’s rate in order to manage uplink interference. Our work is different in that we consider the uplink of an OFDMA network and focus on VoIP traffic for which interference management is inapplicable. In practice both time sensitive and delay tolerant traffic must be supported and hence both interference management as well as admission controls are needed.

Note that the shared resources that can be optimized by the base station are the bandwidth of each tile and the tile allocation rate. Given these values the SS uses sufficient power to achieve the desired rate. We will first assume that the bandwidth and tile allocation rate are fixed for all users and determine optimal values for each. We first do this given global knowledge of the system (a centralized approach) and then show that a simple distributed algorithm can achieve the same answer. Next we propose schemes for re-balancing resources to improve user capacity. First we do this by varying the size of the bandwidth-time resources allocated to a user based on the user’s channel conditions. Next we discuss Fractional Frequency Reuse approaches as well as Multiple Input Multiple Output (MIMO) antenna schemes.
which will be the decision variable. Given the sector supported which in turn determines the maximum number of simultaneous active sessions that can be allocated only to active sessions. Our focus is on the period (when the user is talking). We assume that resources are allocated only for the active period and hence resources need to be allocated only to active sessions. Our focus is on the maximum number of simultaneous active sessions that can be supported which in turn determines the maximum number of VoIP connections. Assume that \( \tau \) tiles per second are required to support the rate \( r \) required for an active VoIP session (e.g., the rate for full rate VoIP frames). In Fig. 1 we illustrate a simplified example of the resource allocation structure. In this case there are four channels per frame so that \( B/b = 4 \). Two tiles are transmitted per user in each VoIP period to that the tile rate is \( \tau = 100 \). Since there are four frames per VoIP period then the frame length is \( d = 5 \text{ ms} \).

A. The Centralized Algorithm

We focus on the power allocation for a specific tile. We assume \( N \) sectors and that a single user is chosen for a VoIP packet transmission in each sector. Therefore the same index is used for the sector as well as for the user allocated within that sector. We assume that the concerned tile is shared by the same subset of users so that the interference measured in one tile is approximately the same experienced by the transmission made within that tile in the following VoIP period. Let \( g_{ij} \) denote the average channel gain from user \( i \) to sector \( j \). Therefore if user \( i \) transmits with power \( p_i \), then the received signal strength at sector \( i \) is \( p_i g_{ii} \) while the interference it incurs on sector \( j \neq i \) is \( p_j g_{ij} \). For convenience, we define the variable \( x_i = p_i g_{ii} \) which will be the decision variable. Given \( x_i \) we can then determine the transmission power of user \( i \). We also define the total interference experienced by sector \( j \) by

\[
x_j \equiv \sum_{i \neq j} p_i g_{ij}.
\]

The background noise experienced at sector \( j \) will be denoted by \( N_j \) so that the total interference plus noise at sector \( j \) is \( X_j + N_j \). The rate achieved by user \( i \) can be determined from the received SINR of its transmitted packet. We assume that each user must achieve at least rate \( r \) to maintain acceptable packet loss rates and delay. Therefore we can write

\[
r = b \tau \log_2 \left( 1 + \frac{x_i}{X_i + N_i} \right).
\]

Define \( \kappa = 2^{r/(br)} - 1 \) then the constraint can be stated as

\[
\kappa = \frac{x_i}{X_i + N_i}.
\]

so that some fixed SINR must be achieved to maintain acceptable performance for the user. Note that as the bandwidth-time resource, \( br \), utilized by a user is decreased then more users can be accommodated and hence the capacity increases. However, as the resource utilization \( br \) is decreased, the required SINR \( \kappa \) (SS transmission power) must be increased. Therefore there is a limit on the maximum transmission power, \( p_{max} \), of a SS and this will limit how large we can make \( \kappa \). Hence we must find the largest value of \( \kappa \) such that the probability that a SS has insufficient power to achieve rate \( r \) is no more than the outage probability. We first determine the power allocations for a given \( \kappa \). We can write:

\[
x_i = \kappa(X_i + N_i) \quad (4)
\]

\[
= \kappa \left( N_i - x_i + \sum_{j=1}^{N} p_j g_{ji} \right) \quad (5)
\]

\[
= \kappa \left( N_i - x_i + \sum_{j=1}^{N} x_j g_{ij} / g_{jj} \right) \quad (6)
\]

Define the \( N \times N \) matrix \( G \) with elements given by

\[
G(i,j) = \frac{g_{ij}}{g_{ii}} \quad 1 \leq i \leq N, \quad 1 \leq j \leq N.
\]

Let \( \bar{x} \) and \( \bar{N} \) denote the decision variable and noise vectors and \( I \) the identity matrix. We have

\[
\bar{x} = \kappa(\bar{N} - \bar{x} + \bar{x}G) \quad (8)
\]

\[
\bar{x}^T[(1 + \kappa)I - \kappa G] \geq \kappa\bar{N} \quad (9)
\]

\[
\bar{x} = \kappa\bar{N}[I - \kappa(G - I)]^{-1}. \quad (10)
\]

Therefore we can explicitly solve for \( \bar{x} \) as a function of the rate per user \( r \), the bandwidth per tile \( b \), the number of allocated tiles per second \( \tau \) and the user geometries \( G \).

Note that since the elements of the matrix \( G \) are random variables then the elements of the vector \( \bar{x} \) are stochastic. As \( \kappa \) is increased, the SS transmission powers increase. Therefore there is some probability that the power required for a particular SS exceeds its maximum power, \( p_{max} \). When
this occurs the SS is in outage. For a given \( \kappa \) we can therefore compute this outage probability.

Next we compute the capacity for a given \( \kappa \). Denote the frame duration by \( d \) ms. In a 20ms VoIP period a total of \( 20/d \) tiles are available per channel of bandwidth \( b \). Since the rate at which tiles are allocated to a user is \( \tau \) then the number of tiles allocated during this period is \( 0.02\tau \) and hence each channel can support \((20/d)/(0.02\tau) = 1000/d\tau \) users. Since there are \( B/b \) such channels the number of simultaneous active sessions is given by \( 1000B/(b\tau d) \). If we denote this by \( C \), we can substitute for \( bt \) using Eq. 2 and we can replace the SINR by \( \kappa \) using Eq. 3 to obtain

\[
C = \frac{1000B}{rd} \log_2(1 + \kappa). \tag{11}
\]

Since \( B, r \) and \( d \) are fixed then \( C \) is maximized by maximizing \( \kappa \). Since the outage probability grows with \( \kappa \) we must determine the value of \( \kappa \) for which the outage probability is equal to the specified value. Unfortunately a closed form expression is not possible and hence we will determine this optimal value, which we denote by \( \hat{\kappa} \) through Monte Carlo simulations.

We performed some simple Monte Carlo simulations since our focus is on relative performance numbers as opposed to absolute performance numbers. We consider the case of 49 cells with one sector per cell and a frequency reuse of one. A single user is randomly dropped within each of the cells (the user to be allocated to the concerned tile). We compute the path loss between each user and each base station and focus on the outage probability of the center cell. For a given \( \kappa \) we use Eq. 10 to determine if the power required for the center cell exceeds the maximum value. If it does then the transmission is declared to be in outage and we compute the probability of this occurrence over 20000 different user drops. For this given \( \kappa \) we can use Eq. 11 to determine the corresponding user capacity. We repeated this for different values of \( \kappa \) and in Fig. 2 we plot the resulting outage probability (in percentage) as a function of \( C \). As expected, the outage probability increases with \( C \). If we define the capacity as the maximum number of users that can be supported with an outage probability of at most 2% then in this particular case we obtain a capacity of about 103 users. Note that this is the number of users with active VoIP connections and hence the true VoIP capacity (i.e., including the users in a silent period) is much larger.

### B. The Distributed Algorithm

In the previous section we assumed that global knowledge was available and so the optimal power allocations can be determined. In practice this is not the case and distributed algorithms are preferable in which each sector determines the power allocation based solely on local information. This local information includes the intersector interference experienced by the sector over the associated tile as well as the channel conditions of the allocated user.

We show that, if a feasible solution exists (i.e., that power allocations can be made so that each user achieves their desired rates and all user powers are less than the maximum possible) then this solution can be obtained with the following distributed algorithm. For each frame transmission, the SS uses the exact power needed to achieve the required rate for the VoIP service assuming the prevailing intersector interference.

**Theorem 1:** Assume that each user autonomously adjusts its transmission power to achieve its specified rate. For any feasible initial power allocations, the system converges to the optimal power allocations.

**Proof:** Denote the initial decision variable vector by \( \vec{x}(0) \) for which each power allocation is at most \( p_{max} \). We assume that after each VoIP period each sector computes the intersector interference experienced over the concerned tile and uses this value to compute the power allocation needed for the transmission in the following VoIP period. Let \( \vec{x}(n) \) denote the power allocation for the \( n \)th VoIP period. This power allocation depends on the intersector interference of the previous period which depends on the transmission powers \( \vec{x}(n-1) \) used in that period. Therefore using Eq. 8 we can write

\[
\vec{x}(n) = \kappa(n)\vec{x}(n-1)(G - I). \tag{12}
\]

We can recursively write \( \vec{x}(n-1) \) in terms of \( \vec{x}(n-2) \) and repeat all the way back to \( \vec{x}(0) \). If we do this we obtain

\[
\vec{x}(n) = \kappa^n\vec{x}(0)(G - I)^n + \kappa N \sum_{k=0}^{n-1} (\kappa(G - I))^k. \tag{13}
\]

Note that if the inverse in Eq. 10 exists (i.e., \( \kappa \) is sufficiently small) then the eigenvalues of \( \kappa(G - I) \) are less than one and hence in the limit as \( n \) goes to infinity the first term on the right hand side goes to zero. This means that the initial allocation does not affect the asymptote. The second term on the right hand side is the geometric series and again if the eigenvalues of \( \kappa(G - I) \) are less than one then this sum exists and can be computed. In the limit as \( n \) goes to infinity the second term on the right converges to \( \kappa N(I - \kappa(G - I)) \) and hence we have

\[
\lim_{n \to \infty} \vec{x}(n) = \kappa N(I - \kappa(G - I)) = \vec{x}^*. \tag{14}
\]

Therefore the power allocation vector for the distributed algorithm converges to that of the centralized algorithm. ■

---

**Fig. 2. Outage probability versus user loading**
In order to illustrate the convergence of the distributed algorithm to the optimal power allocation we ran the following simulation. We simulate 49 cells but focus only on the power allocation in the central cell. All sectors become active at time 0 with initial transmission powers computed assuming no intersector interference. At the end of each VoIP period we compute the intersector interference values and power allocations and use those for the subsequent VoIP period. We also computed the centralized solution. We then looked at the user that is furthest away from its optimal power allocation by taking the smallest ratio, over all users, of each user’s distributed power allocation to their centralized power allocation. This metric indicates how close the distributed power allocation vector is to the optimal one. In Figure 3 we plot this error metric as a function of time. We find that after 5 VoIP frame periods the distributed power allocations are all within 1% of the optimal values. Note that, in this example, major changes in the system are made over a short period of time. This will not be the case in practice and hence the time to convergence is expected to be much smaller for more gradual changes.

III. Capacity for Variable Resource Allocations

From the previous section we find that the VoIP capacity is limited by the cell edge users since those are the ones that first run out of power. However, even at capacity the users within the cell do not fully utilize their power resources. Therefore capacity can be increased if we were to shift resources from the center users to those at the edge. There are two ways to do this, allocate more bandwidth and/or higher tile rates to users at the edge than users at the center. In this case \( b \) and \( \tau \) become user dependent values.

Consider a cell edge user that is limiting capacity. Consider the following two options (a) allocate two tiles instead of one in each frame for which an allocation is made, or (b) double the tile rate for the user. Note that in both cases the resources allocated to the user is doubled. However there is a difference in the two cases. In the bandwidth expansion case, the power must be shared between the two allocated tiles while in the time expansion case full power can be used for each tile transmission. Hence the total energy received at the receiver is much larger in the time expansion case than in the bandwidth expansion case. One must however note that in the bandwidth expansion case the interference caused by the user within its allocated tiles is lower and this in turn will result in a lower interference level for the transmission but this gain is quite small. Therefore we find that if additional resources are to be allocated to the edge users then it should be done in the time domain so that full power can be used for each tile transmission.

Again we focus on the center cell and run the same scenario that was used to produce Fig. 2. We make the following resource allocation changes. If the power required for the center cell is larger than \( 2p_{\text{max}} \) an outage is recorded. If however it is larger than \( p_{\text{max}} \) but less than \( 2p_{\text{max}} \) then we reduce the required SINR by half and record the fact that the tile rate was increased. After 20000 drops we compute the capacity as the probability that the tile rate was not increased times the corresponding capacity plus the probability that the tile rate was doubled times half the user capacity since twice as many tiles are needed for this user. In Figure 4 we plot the outage probability versus the user loading. Note that there is a significant increase in capacity. At 2% outage a user loading of 123 users can be supported. Therefore this corresponds to a 20% increase in capacity.

IV. Management of Other Resources

In this section we investigate other potential means of improving the VoIP capacity by improving the allocation of resources. Besides bandwidth and time that was investigated in the previous section, the rate depends on the transmission power (which is given constraint), path gain (which can be improved by using multiple transmit and/or receive antennas), and intersector interference (which can be reduced by adjusting the frequency reuse factor).

A. Dynamic Fractional Frequency Reuse

If we can reduce the interference for the edge users then we can increase their performance. Suppose we do the following.
A subset of resources are reserved for frequency reuse. These resources are used during the odd frame index transmissions by the odd numbered base stations and the remaining frames are used by the even numbered base stations. This subset is to be used by those users that would have been placed in the outage state if allocated within the normal set of tiles. Therefore, whenever such a user transmits, the intersector interference it experiences is half of what it would normally experience since only half of the base stations are receiving transmissions. Let us compare this approach with the time expansion approach. In the time expansion case the received SINR is given by $\frac{2p_{\text{max}} g}{(I + N)}$, since two transmissions with full power are supported. In the frequency reuse case the SINR is given by $\frac{p_{\text{max}} g}{(I/2 + N)}$. Therefore, if the noise is negligible then both approaches achieve the same SINR but in general the time expansion approach is expected to provide better results. Furthermore, the frequency reuse approach has additional drawbacks. Each time a SS can potentially be in the outage state it must be relocated to the special frequency reuse subset. This requires additional signaling. Furthermore coordination is required between base stations for the frequency reuse subset in that if this subset needs to be adjusted in size then all base stations must be informed.

### C. Spatial Division Multiple Access

The point to point MIMO case required increasing the number of antennas at the SS. It is still possible to achieve spatial multiplexing gains by allowing more than one SS to transmit within the same tile. If these transmissions are uncorrelated then one can achieve significant gain in capacity. Let us focus on the case of at most two users per tile. In the case of VoIP where the number of users is large, the probability of always finding a pair of users that are almost uncorrelated is high. However, we should also point out that the intersector interference also increases. Since each SS must transmit with higher power to overcome any intrasector interference of its paired user then the total interference is more than doubled that of the single user case. The mode of operation would typically be single user transmissions for edge users and paired transmissions for users at the center. Since the performance gains of this approach depends on several factors (the scheduling algorithm for choosing pairs, the degree of correlation between users, etc.) then a detailed analysis is outside the scope of this work.

### V. SUMMARY AND FUTURE WORK

We considered the problem of maximizing the user capacity for VoIP service on the uplink of an OFDMA network. We computed the optimal power allocation for the centralized case and showed that a distributed algorithm can be used to obtain this optimal power allocation. We pointed out that the capacity was limited by the poor performance of the users at the cell edge even though users within the cell had more than enough resources. We therefore focused on re-balancing resources so that more were provided to the edge users than the center users. One such approach was increasing the tile allocation rate for users at the edge. We showed that this can improve capacity by about 20%. Finally we evaluated some other methods that could potentially be used for re-balancing resources to improve capacity. We found that fractional frequency reuse does not provide sufficient gains to warrant the increased complexity. However we believe that SDMA shows potential with little additional overhead or cost. We plan to evaluate the performance gains for SDMA in our future work.

### REFERENCES