Energy-Aware Resource Allocation Strategies for LTE Uplink with Synchronous HARQ Constraints

Dan J. Dechene, *Member*, *IEEE*, and Abdallah Shami, *Senior Member*, *IEEE*

Abstract—In this paper, we propose a framework for energy efficient resource allocation in multiuser localized SC-FDMA with synchronous HARQ constraints. Resource allocation is formulated as a two-stage problem where resources are allocated in both time and frequency. The impact of retransmissions on the time-frequency problem segmentation is handled through the use of a novel block scheduling interval specifically designed for synchronous HARQ to ensure uplink users do not experience ARQ blocking. Using this framework, we formulate the optimal margin adaptive allocation problem, and based on its structure, we propose two suboptimal approaches to minimize average power allocation required for resource allocation while attempting to reduce complexity. Results are presented for computational complexity and average power allocation relative to system complexity and data rate, and comparisons are made between the proposed optimal and suboptimal approaches.

Index Terms—SC-FDMA, energy efficiency, margin adaptive, resource allocation

1 Introduction

Energy efficient communication is a long-standing issue in modern wireless communication systems. The mobile device radio accounts for a large portion of a mobile device's battery life, and as such, efficient use of radio resources can dramatically improve mobile device energy consumption. With the increased proliferation of smaller and faster devices, it has become essential to efficiently utilize mobile battery resources.

General radio resource allocation problems, particularly for systems such as orthogonal frequency division multiplexing (OFDM) fall in to two major classifications, namely the rate and margin adaption (RA and MA) problems [1]. RA problems try to allocate resources to maximize system throughput for a given power constraint, while MA problems try to minimize transmission power and maintain a minimum throughput guarantee. The latter of which is essential for energy efficient scheduling.

In recent years, MA problems have been well studied for a general OFDMA transmission system [1], [2]. However, more modern systems, such as 3GPP-LTE, utilize localized single carrier frequency division multiple access (SC-FDMA) at the physical layer for uplink transmissions. With localized SC-FDMA, subcarriers can only be allocated contiguously in frequency [3], [4]. The use of localized SC-FDMA has been shown to offer improved peak to average power ratio (PAPR) compared to OFDM. This however imposes a limitation of contiguous frequency block assignment [5], and thus eliminates direct application of traditional MA framework as described above. Furthermore, the finite set of modulation and coding schemes (MCS) used in modern communication systems dramatically increases the optimal allocation complexity.

Recent works have tried to address various parts of the overall scheduling/resource allocation problem that deal with the complexity associated with SC-FDMA [5], [6], [7], [8]. For example, [5] focused on SC-FDMA uplink scheduling for the case of delay-limited traffic. While in contrast, the authors of [8] approached the sum-rate maximization problem using a similar approach to [9] and further proposed several heuristic methods to deal with the complexity issue. However, the objective function in this case is to maximize the cell transmission rate rather than minimize transmission power.

Nevertheless, the more promising approaches for such allocation schemes with manageable computational complexity have been to distil the resource allocation problem [10], [11], [12] into two phases in an effort to reduce resource allocation complexity. Under this regime, first a subset of users is chosen in time to transmit during a frame, i.e., time-domain packet scheduling (TDPS). Second, these users are allocated frequency resources within a given time frame to maximize throughput (or minimize energy resources), i.e., frequency-domain packet scheduling (FDPS).

The difficulty with this type of approach is that the practicality of retransmissions needs to be considered and that 3GPP's LTE also employs synchronous hybrid automatic repeat request (HARQ) for retransmission in the uplink. In the nonadaptive configuration, transport blocks (TBs) are retransmitted at the same rate as the original transmission exactly ARQ_w frames following the original transmission. Consequently, retransmission resources must

- *D.J. Dechene is with the IBM Semiconductor Research and Development Center, Hopewell Junction, NY. E-mail: ddechene@ieee.org.*
- *A. Shami is with the Department of Electrical and Computer Engineering, University of Western Ontario, London, ON, Canada. E-mail: ashami2@uwo.ca.*

Manuscript received 21 Apr. 2012; revised 22 Aug. 2012; accepted 6 Dec. 2012; published online 12 Dec. 2012.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2012-04-0204.
Digital Object Identifier no. 10.1109/TMC.2012.256.

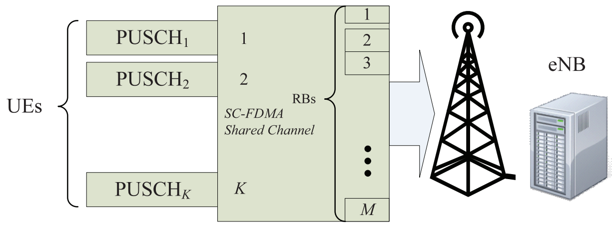


Fig. 1. SC-FDMA shared channel.

be properly considered by resource allocation algorithms to enable HARQ.

In this paper, we propose a method of performing resource allocation that exploits the periodicity of the HARQ process in scheduler design. We propose the use of a block time-frequency domain packet scheduler (BTFDPS). This approach reduces the amount of scheduling decisions required for uplink traffic in addition to simplifying incorporation of synchronous HARQ into the framework. At each block time-frequency frame, resources are allocated dynamically relative to the required throughput, priority, and number of on-going transmissions. Power allocation is minimized per unit time to meet constraints resulting in minimizing energy expended for transmission.

The remainder of this paper is divided as follows: In Section 2, we overview the details of the employed uplink system model including the channel and scheduling models. In Section 3, we describe the scheduling methodology, while in Section 4, we formulate the optimal energy efficient allocation problem with the given HARQ constraints as a power allocation minimization problem. In Section 5, we propose two suboptimal methods to improve computational tractability. In Section 6, simulation results are provided while in Section 7, conclusions are drawn on this work.

2 System Model

The simplified MAC model of a multiuser SC-FDMA system used for uplink in 3GPP's LTE, [13] is shown in Fig. 1. Here, each user has access to a single physical uplink shared channel (PUSCH) for transmission of their uplink data. Within a single cell, there are K users (UEs) and a single base station (eNB). For the purpose of our work, it is assumed that intercell interference is negligible. The cell spectrum consists of N_{sub} narrow band subcarriers grouped into M resource blocks (RBs) of 12 subcarriers each. Without loss of generality, there is an integer number (M) of RBs. The system is assumed to be operating in frequency division duplexed (FDD) mode.

There are N_{sym}[1] symbols per subcarrier in a given subframe where the exact number of subcarriers depends on the uplink configuration. The PUSCH is used for transmission of uplink data and shared between UEs (as shown in Fig. 1).

In 3GPP LTE, CQI values describe a range of targeted MCSs and are given in Table 3 reproduced from [14]. The overall TB size is given as the effective spectral efficiency combined with the number of allocated RBs. Consequently, the overall number of data bits, the TB size, that can be transmitted per subframe over a set $\mathcal{N}$ of RBs in frequency given the assumptions above is

$$\eta(b, \mathcal{N}) = \lfloor 12(N_{sym} - N_{overhead})b \cdot |\mathcal{N}| \rfloor, \quad (1)$$

where b denotes the spectral efficiency in bits per symbol, $\mathcal{N}$ denotes the set of RBs in frequency and $|\cdot|$ is the size of a set. The quantity $N_{overhead} >= 0$ allows for consideration of any additional in-channel overhead per TB within the described framework. For the remainder of the paper and without loss of generality, we employ the notation $\varphi = 12(N_{sym} - N_{overhead})$. Further, the minimum and maximum spectral efficiencies for b are shown in Table 3 as $b_{min} = 0.1523$ and $b_{max} = 5.5547$.

Localized SC-FDMA is employed in LTE uplink [3], [4] and in our system model. In this technology, unlike with OFDMA only adjacent subcarrier blocks can be allocated to a given station (UE) at one time. As a result, the application of localized SC-FDMA constrains allocated RBs to any UE such that they are adjacent in frequency. Individual UEs are further limited to a single TB using a common MCS mode per subframe[2] and retransmissions employing HARQ must be transmitted over the same number of resource blocks using the same MCS after exactly eight (ARQ_w) subframes have passed from original transmission (nonadaptive HARQ).

2.1 Scheduling Model

The minimum duration of time that can be allocated to a single UE is a scheduling block (SB). The nth SB is in the time duration bounded by $[nT_f, (n+1)T_f)$, where T_f is the subframe duration (equal to 1 ms) and is one RB in frequency.

Resource allocation decisions in LTE are made at each scheduling epoch. A scheduling epoch is the interval of time over which resource allocation decisions can be made. While in general the scheduling epoch can be equal to the scheduling block (i.e., $T_e = T_f$), we define the mth scheduling epoch as the time duration in $[mT_e, (m+1)T_e)$ and each epoch consists of T_e/T_f subframes in time, where m is the index of the scheduling epoch. The time T_e is chosen to be equal to

$$\frac{ARQ_w T_f}{TXOP}, \quad (2)$$

where $TXOP$ denotes the maximum number of new TBs that can be allocated per ARQ_w subframes where

$$TXOP = \frac{ARQ_w}{Max_{tx}}, \quad (3)$$

and Max_{tx} is the maximum number of times a packet can be (re)transmitted. The justification behind this selection T_e is discussed in later sections. Fig. 2 shows an example of the uplink frame layout with relevant time durations. Successful transmissions are shown in gray, while failures are red. As shown, retransmissions occur exactly ARQ_w subframes following the initial transmission using the same number of resource blocks.

1. $N_{sym} = 14$ with regular cyclic prefix in LTE [14].

2. For single layer (non-MIMO) transmissions [14], which is assumed in this work.

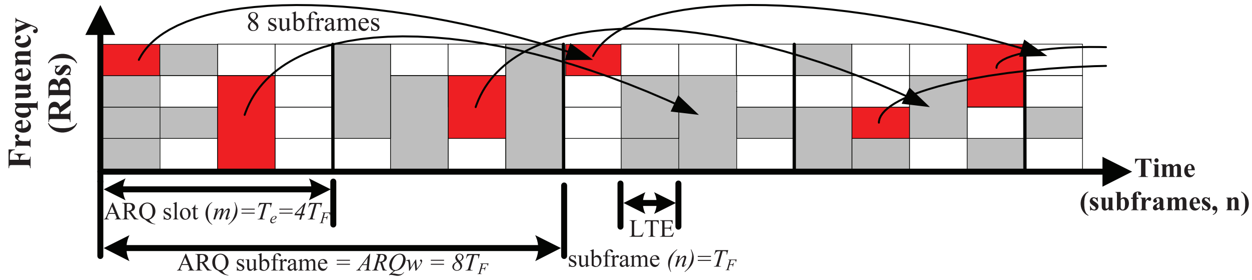


Fig. 2. Frame layout.

We assume that each UE has a single radio data bearer established between itself and the core network. Each bearer has an access network bit rate of χ_i bit per unit time or an average newly allocated TB size of $\overline{T}_i$ bits per ARQ slot.

2.2 Channel Model

The channels between each UE and the eNB, and from RB to RB are independent. A block fading model is assumed, where the channel is static for the duration of the decision epoch T_e and independent from epoch to epoch. Channel estimation is assumed to be error free and available at the eNB and each channel follows the Rayleigh SNR distribution given as

$$p(\gamma) = \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right), \quad (4)$$

where $\gamma_{i,k}(m)$ is the instantaneous SNR in the uplink channel of user i over RB k in epoch m and where γ_0 denotes average SNR for a reference applied power level.

The block error rate (BLER) is the average failure rate of TBs. This depends on $\gamma_{i,eff}$ (the effective signal to noise ratio, SNR, of a transmission), T_i (the TB size), and b_i (the effective MCS). To the best of our knowledge, there are no analytical solutions to compute the BLER of a general coded transmission. In practical implementations, estimates of BLER are obtained through receiver measurements. The measure of information outage probability (IOP) derived in [15] can be used as an approximate measure for BLER. The IOP was shown to closely model BLER for moderate block lengths. Any extensions to our proposed framework are trivial when estimates of the BLER can be more accurately obtained for a particular implementation using appropriate training and calibration techniques.

Following from [15], we can obtain the expression for BLER as

$$BLER(\gamma, \mathcal{N}_i, T_i) \approx Q\left(\frac{\log(1+\gamma) - \frac{\log(2)T_i}{\varphi|\mathcal{N}_i|}}{\sqrt{\frac{2}{\varphi|\mathcal{N}_i|}\frac{\gamma}{1+\gamma}}}\right), \quad (5)$$

where $\mathcal{N}_i$ is set of RBs allocated to UE i, $Q(\cdot)$ is the well-known Q-function and $\overline{\gamma}_{i,eff}$ is the effective SNR. The quantity $\varphi|\mathcal{N}_i|$ and $\frac{T_i}{\varphi|\mathcal{N}_i|}$ corresponds to the number of modulated symbols and effective spectral efficiency per allocation, respectively, described earlier. The factor $\log(2)$ is required and follows from [15] in that we measure spectral efficiency in bits per symbol rather than nats per symbol. Due to the monotonicity of the Q-function arguments, the above can be solved efficiently using bisection techniques for the required SNR. Alternatively, a more computationally efficient method is to obtain a least-squares approximation to the above as a function of data rate, target BLER and the number of RBs allocated (similar to the approach in [16]). We found that the following fitting function closely approximates the SNR as a function of data rate

$$\gamma_{i,eff}^{(r)} \approx a_x \exp(b_x T_i) - \gamma_{0,x}, \quad (6)$$

where $x = |\mathcal{N}_i|$. The values of a_x, b_x, and $\gamma_{0,x}$ are given in Table 1 for $BLER_{tgt} = 10\%$ for up to 24RBs. Using the above, along with channel estimates, it is easy to obtain the required applied power to achieve a target block error rate. The accuracy of this approximation is verified in Section 6.

The measured effective SNR for a reference power level of a transmission is related to the SNR of the RBs comprising it. As in [17], the effective SNR of an SC-FDMA symbol cannot be approximated using EESM or MIESM (as

TABLE 1
Least-Squares Approximate Model Parameters

RBs (x)	a_x	$b_x \cdot 10^3$	$\gamma_{0,x}$
1	1.1748	5.2471	1.1019
2	1.1208	2.624	1.0723
3	1.0977	1.7495	1.0591
4	1.0841	1.3122	1.0512
5	1.0749	1.0498	1.0458
6	1.0682	0.8749	1.0418
7	1.063	0.7499	1.0387
8	1.0588	0.6562	1.0362
9	1.0553	0.5833	1.0342
10	1.0524	0.525	1.0324
11	1.0499	0.4772	1.0309
12	1.0478	0.4375	1.0296
13	1.0459	0.4038	1.0284
14	1.0441	0.375	1.0274
15	1.0426	0.35	1.0265
16	1.0412	0.3281	1.0256
17	1.04	0.3088	1.0249
18	1.0388	0.2917	1.0242
19	1.0378	0.2763	1.0235
20	1.0368	0.2625	1.0229
21	1.0359	0.25	1.0224
22	1.0351	0.2386	1.0219
23	1.0343	0.2283	1.0214
24	1.0336	0.2188	1.0209

TABLE 2
SNR Gain for Retransmissions [18]

Modulation	rv_{idx}	$\mu \cdot 10^3$	ϵ
QPSK	1	0.8024	2.89
	2	1.628	4.57
	3	2.006	5.62
16-QAM	1	0.420	1.1
	2	8.435	0.74
	3	9.464	1.15
64-QAM	1	8.996	-1.23
	2	12.288	-0.71
	3	12.728	0.15

TABLE 3
CQI Indices (Modulation/Coding Schemes) [14]

Index	Modulation	Coding Rate	Spectral Efficiency (b)
0	—	—	0 bits
1	QPSK	78/1024	0.1523
2	QPSK	120/1024	0.2344
3	QPSK	193/1024	0.3770
4	QPSK	308/1024	0.6016
5	QPSK	449/1024	0.8770
6	QPSK	602/1024	1.1758
7	16QAM	378/1024	1.4766
8	16QAM	490/1024	1.9141
9	16QAM	616/1024	2.4063
10	64QAM	466/1024	2.7305
11	64QAM	567/1024	3.3223
12	64QAM	666/1024	3.9023
13	64QAM	772/1024	4.5234
14	64QAM	873/1024	5.1152
15	64QAM	948/1024	5.5547

in OFDM), but rather it can be approximated as the average SNR over the set of RBs allocated or as

$$\gamma_{i,eff}^{(0)}(m, \mathcal{N}_i) = \frac{1}{|\mathcal{N}_i|} \sum_{k \in \mathcal{N}_i} \frac{\gamma_{i,k}(m)}{|\mathcal{N}_i|}, \tag{7}$$

Here, $\mathcal{N}_i$ is the set of RBs in the computation and the superscript (x) (i.e., $x = 0$ above) is used to denote the xth retransmission number, where in the above case 0 denotes the initial transmission. The additional $|\mathcal{N}_i|$ in the above equation describes that power is equally allocated across the all assigned resource blocks. We emphasize here that extension of the above is trivial where measurements/estimates of the effective SNR are obtained and known for a given implementation.

2.3 Retransmission Model

Synchronous HARQ is used for uplink retransmission to improve the probability that a TB can be decoded at the base station while limiting signalling overhead required for a synchronous HARQ. In the LTE implementation, a TB is first encoded with a 1/3 rate turbo code. From this, four possible code blocks are obtained denoted as redundancy versions (RVs). Each redundancy version is a different set of coded bits from the TB. Each retransmission of a TB utilizes a different RV, so that retransmissions can be combined with the initial transmission to improve overall reliability.

Consequently, this improved reliability can be considered as an increase in the measured effective SNR as done in [18], which is also used here. We employ this model with coefficients for SNR gains given in Table 2 but note that for a given implementation, these gain factors can be obtained through receiver calibration measurements. The effective SNR of retransmission z can be computed as (in dB)

$$\begin{aligned}\gamma_{i,eff,(dB)}^{(z)}(m, \mathcal{N}_i, k_i) = \gamma_{i,eff,(dB)}^{(0)}(m, \mathcal{N}_i) \\ + \mu(z, k_i) \cdot R_{code} + \epsilon(z, k_i),\end{aligned} \tag{8}$$

where R_{code} is the code rate $\times$1,024 which is given from [14] (i.e., 78, 120, ...), $\mu(z, k_i)$, and $\epsilon(z, k_i)$ are given from [18] for each retransmission and modulation scheme, k_i is the modulation scheme, where $\gamma_{i,eff}^{(0)}(m, \mathcal{N}_i)$ is the effective SNR for UE i given in (7). These parameters are summarized in Table 2. The modulation scheme is derived from the CQI mode in Table 3 yielding the next highest spectral efficiency mode than $\frac{T_i}{\varphi|\mathcal{N}_i|}$.

2.4 Transmission Power Selection

For a given target BLER of $BLER_{tgt}$, the required power is simply

$$P_i(\mathcal{N}_i, T_i) = \frac{\overline{\gamma}_{i,eff}(\mathcal{N}_i, T_i)}{\gamma_{i,eff}^{(z)}(m, \mathcal{N}_i)}, \tag{9}$$

where $\overline{\gamma}_{i,eff}(\mathcal{N}_i, T_i)$ is the γ argument in (5) when (5) is set equal to $BLER_{tgt}$ for given arguments $\mathcal{N}_i$ and T_i (or using the approximation given in (6)). The above corresponds to the power needed for allocation.

3 Scheduling Methodology

Recent research, particularly for LTE systems [10], [11], [12], proposes segmenting the packet scheduling problem into a TDPS and FDPS problem. Such methods choose a set of users at each scheduling interval to schedule (TDPS) and allocate them in frequency (FDPS) to maximize cell throughput (rate adaptive, RA).

In the energy limited regime, the goal is to minimize transmission energy while meeting throughput requirements. In this regime, the optimal allocation for MA differs from that of an RA problem. As a result, RA methods are not directly applicable in solving the MA problem. Further, due to limitations imposed by synchronous HARQ, the above described approaches may experience ARQ blocking since the TDPS is limited in knowledge to a single subframe in time. ARQ blocking here is defined as the inability to allocate a new TB to transmit during a given subframe due to the requirement of retransmitting a previously transmitted TB as part of the synchronous HARQ process.

A simple to implement method to eliminate ARQ blocking is to limit the number of new transmissions a station can initiate in each ARQ window. In synchronous HARQ, retransmissions are limited to exactly ARQ_w

subframes following the initial transmission (where $ARQ_w = 8$ in LTE uplink), and the maximum number of times a packet can be transmitted is 4. Based on these observations, the maximum number of new transmissions that any station should initiate during any ARQ_w subframes is $ARQ_w/4 = 2$ to eliminate any ARQ blocking. Based on this, we can define an ARQ slot as the duration of time such that each UE can be allotted at most one new TB for transmission. Consequently, each ARQ slot m has a duration of four subframes or $1/2$ of an ARQ_w (i.e., $T_e = 4T_f$).

Using this segmentation approach, we can formulate the resource allocation problem by segmenting it into a two-stage formulation, namely a TDPS and BTFDPS. At the beginning of scheduling epoch (ARQ slot) m, the TDPS chooses which UEs can access the channel (and with how much data, $T_i(m)$) based on the buffers and priority of each UE over ARQ slot m. The BTFDPS then allocates the UEs within the time-frequency grid of an ARQ slot based on channel conditions and knowledge about on-going retransmissions. The benefit of such an approach as described above is that it reduces the number of scheduling decisions needed in time, as well as, eliminates ARQ blocking. The drawback of this, however, is that it requires estimates of the channel for the overall ARQ slot.

3.1 Time Domain Packet Scheduling

During ARQ slot m, the TDPS determines the amount of new data to be transmitted over the channel. The TDPS decisions can be done in a number of ways. Existing works have proposed TDPS designs which demonstrate methods of exploiting information including but not limited to buffer occupancy, and quality of service requirements. The focus of this work is on the design of the BTFDPS and for the remainder of this paper, we assume a static TDPS where all UEs are allocated a new TB of size $T_i(m)$ during ARQ slot m.

Further, the TDPS may define a static (or dynamic) weight parameter $\alpha_i(m)$ denoting the relative priority of each station. Without loss of any generality we can assume $\sum_{i=1}^{K} \alpha_i(m) = 1, \forall m$. Weights can be designed to satisfy fairness or priority criteria. In our work, we do not focus on the design of weights and simply assume they are known.

3.2 Block Time-Frequency Domain Packet Scheduling

The goal of the BTFDPS is to allocate both retransmissions and new TBs within the time-frequency grid of an ARQ slot. For a given set of TBs from each UE $\{T_1(m), T_2(m), \ldots, T_K(m)\}$ for all m, channel state information γ, and knowledge about TBs for retransmission, the BTFDPS applied power minimization problem can be formulated to solve for resource assignment. The resource assignment policy is comprised of the resource element assignment ($S_{i,j,n}(m)$ and $\widetilde{S}_{i,j,n}(m), \forall i, j, n, m$), power assignment ($P_{i,n}(m), \forall i, n, m$), and rate assignment ($k_{i,n}(m), \forall i, n, m$), where $n = 0, 1, 2, 3$ are the individual subframes within an ARQ slot. The quantities $S_{i,j,n}(m)$ and $\widetilde{S}_{i,j,n}(m)$ are binary indicators used to denote whether a given RB j during subframe n is allocated to a given UE i for both new transmissions and retransmissions, respectively. The indicators $S_{i,j,n}(m)$ and $\widetilde{S}_{i,j,n}(m)$ are equal to 0 if a resource is not allocated to a given UE and 1 if it is for initial transmission and retransmisions, respectively. The quantity $P_{i,n}(m)$ denotes the power allocated to UE i and $b_{i,n}(m)$ denotes the number of bits allocated to a given channel for an MCS chosen for UE i all during subframe n of ARQ slot m.

The design of the BTFDPS policy requires the following constraints:

- Minimum rate constraints (to ensure the TB is allocated).
- HARQ constraints (to ensure the retransmitted TB is allocated 8 LTE subframes following its failed transmission).
- New transmission limitation (only a single TB can be allocated per ARQ slot m).
- Contiguity constraints (to ensure RBs for a single TB are allocated contiguously in frequency).
- Allocation constraints (to ensure at most one UE can occupy a given RB during any subframe).
- BLER constraints (to ensure power is adjusted to meet $BLER_{tgt}$).

Retransmissions are handled as follows: Due to the synchronous HARQ mechanism, any transmission that was erroneous, and has not exceeded the maximum number of transmissions, is rescheduled exactly 8 (ARQ_w) subframes following its original transmission. These transmissions are scheduled using the same set of RBs[3] and using the same MCS.

4 Optimal Allocation Framework

Due to the dependence on previous retransmission, m, it is not possible to solve the optimal allocation for all time m. We therefore formulate the global allocation problem to solve the locally optimal allocation problem in each ARQ slot m. This approach is described below. For clarity to the reader, a summary of the frequently used notation that follows is found in Table 4.

4.1 Optimization Problem

The per-subframe power-optimal BTFDPS is formulated as follows: First, the objective function can be given as

$$\min\left(\sum_{i=1}^{K}\sum_{j=1}^{M}\sum_{n=0}^{3} \alpha_i(m)(S_{i,j,n}(m) + \widetilde{S}_{i,j,n}(m))P_{i,n}(m)\right), \quad (10)$$

where $\alpha_i(m)$ is the relative importance parameter of power minimization for UE i and subframe $n \in \{0, 1, 2, 3\}$ within ARQ slot m defined by the TDPS. The above denotes the overall weighted transmission power allocated in ARQ slot m for a given set of allocation parameters (i.e., $S_{i,j,n}(m)$, $\widetilde{S}_{i,j,n}(m)$, and $P_{i,n}(m)$).

3. While nonadaptive HARQ limits retransmission to the same MCS, number of resource blocks and subframe, there is no limitation on which resource blocks are assigned. In this paper, we further limit retransmissions to the same RBs in frequency. This reduces signalling requirements and problem complexity. Retransmissions can easily be incorporated in the proposed model by considering retransmission blocks as additional users. The impact of such an assumption is discussed in the results section.

TABLE 4
Frequently Used Notation

Quantity	Symbol	Quantity	Symbol
Number of users (UEs)	K	Number of resource blocks (RBs)	M
Duration of subframe	T_F	Duration of scheduling epoch	T_e
Scheduling block index	n	Resource block index	m
ARQ window size	ARQ_w	UE priority weighting	α_i
Best-N depth	N	BTransport block size for UE i	T_i
Reference SNR	γ_0	SNR for UE i in RB k	$\gamma_{i,k}$
Physical rate assignment for UE i in subframe n	$k_{i,n}$	Power allocation for UE i in subframe n	$P_{i,n}$
Power level gain for UE i	PLG_i	Block error rate	$BLER$
Bit allocation for UE i in subframe n	$b_{i,n}$	Resource block allocation for UE i	$\mathcal{N}_i$

4.1.1 Rate Constraints

The amount of new information that must be transmitted during a frame is constrained as

$$\sum_{j=1}^{M}\sum_{n=0}^{3} S_{i,j,n}(m) b_{i,n}(m) \geq T_i(m), \quad \forall i, \tag{11}$$

where $b_{i,n}(m)$ denotes the number of bits allocated by UE i in subframe n and $T_i(m)$ is the total amount of data requiring transmission during ARQ slot m.

4.1.2 HARQ Constraints

The ARQ constraints ensure that resource allocation cannot occur in a subframe where a retransmission exists. This means $S_{i,j,n}(m) = 0, n \in \mathcal{S}_i(m-2), \forall i$, where $\mathcal{S}_i(m-2)$ is the set of subframes during ARQ slot $m-2$ whose transmissions were unsuccessful and have not reached the maximum number of retransmissions. Further, the number of RBs allocated to a user during a retransmission must equal the original amount of RBs assigned, i.e.,

$$\begin{aligned}\sum_{j=1}^{M} \widetilde{S}_{i,j,n}(m) = \sum_{j=1}^{M}(&\widetilde{S}_{i,j,n}(m-2) \\ &+ S_{i,j,n}(m-2)), n \in \mathcal{S}_i(m-2), \forall i.\end{aligned} \tag{12}$$

4.1.3 New Transmission Limitation

To ensure ARQ blocking does not occur, each station is limited to a new transmission in a ARQ slot. This constraint is given as

$$\sum_{n \in \mathcal{S}_i^C(m-2)} I\left(\sum_{j=1}^{M} S_{i,j,n}(m)\right) \leq 1, \quad \forall i, \tag{13}$$

where $I(x) = 0$ when $x = 0$ and 1 otherwise and $\mathcal{S}_i^C(m-2)$ is the complementary set of $\mathcal{S}_i(m-2)$.

4.1.4 Allocation and Contiguity Constraints

Additionally, localized SC-FDMA is limited to RB allocations in contiguity, this constraint can be given jointly as

$$\begin{aligned}\sum_{i=1}^{K}(S_{i,j,n}(m) + \widetilde{S}_{i,j,n}(m)) \leq 1, S_{i,j,n}(m), \widetilde{S}_{i,j,n}(m) \\ \in \{0,1\}, \forall j, n\end{aligned} \tag{14}$$

ensuring only a single user can occupy an RB during an instant of time and from [19] as

$$\begin{aligned}&S_{i,j,n}(m) - S_{i,j+1,n}(m) + S_{i,x,n}(m) \leq 1, \\ &\quad x = j+2, \ldots, M, \forall i, j, n \in \mathcal{S}_i^C(m-2),\end{aligned} \tag{15}$$

$$\begin{aligned}&\widetilde{S}_{i,j,n}(m) - \widetilde{S}_{i,j+1,n}(m) + \widetilde{S}_{i,x,n}(m) \leq 1, \\ &\quad x = j+2, \ldots, M, \forall i, j, n \in \mathcal{S}_i(m-2).\end{aligned} \tag{16}$$

4.1.5 Power Level Constraints

The applied power level constraints are given from (9) and expressed as

$$P_{i,n}(m) = \frac{\overline{\gamma}_{i,eff}(\mathcal{N}_{i,n}, T_i(m))}{\gamma_{i,eff}^{(0)}(m, \mathcal{N}_{i,n})}, \forall i, n \in \mathcal{S}_i^C(m-2), \tag{17}$$

$$P_{i,n}(m) = \frac{\overline{\gamma}_{i,eff}(\bar{\mathcal{N}}_{i,n}, T_i(m))}{\gamma_{i,eff}^{(z_{i,n})}(m, \bar{\mathcal{N}}_{i,n})}, \forall i, n \in \mathcal{S}_i(m-2), \tag{18}$$

where $z_{i,n}(m)$ is the retransmission number in subframe n for UE i during ARQ slot m and $\mathcal{N}_{i,n} = \{j \mid S_{i,j,n}(m) = 1\}$ and $\bar{\mathcal{N}}_{i,n} = \{j \mid \widetilde{S}_{i,j,n}(m) = 1\}$.

Combining (10)-(18) forms the optimal BTFDPS allocation for inputs $\alpha_i(m), T_i(m), \forall m$. Due to the time dependence on m and the instantaneous channel conditions, the above must be solved for each ARQ slot m online.

4.2 Optimization Formulation Using Binary Programming

The above problem can be formulated using a similar approach as that used in [9]. In this fashion, the contiguity constraints are exploited in a manner that reduces the binary search space. The basic concept behind this approach is in formulating the problem above as a binary programming problem where each possible solution vector describing contiguous allocations is enumerated and selection is made to minimize the average weighted power allocation. Due to the limited feasible contiguous allocations, the search space is low relative to the noncontiguous case.

Using the above approach, the optimization problem is solved at each ARQ slot m. For the remainder of this section, the index m is dropped, however, all quantities are assumed to be specific to ARQ slot m. The problem can be expressed as a general set-packing problem and formulated using binary programming as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{1}_{4M},\ \mathbf{A}_{eq}\mathbf{x} = \mathbf{1}_K, \quad x_j \in \{0,1\}, \forall j \in \mathbf{x}, \end{aligned} \tag{19}$$

where $\mathbf{c}$ is the real-valued vector containing the weighted power of choosing a given allocation, $\mathbf{x}$ is the vector of allocation selections, $\mathbf{A}_{eq}$ is a binary equality constraint matrix of K rows and $\mathbf{A}$ is a binary inequality constraint matrix of $4M$ rows. Each nonzero entry of the solution vector $\mathbf{x}$ corresponds to selecting the corresponding column allocation in $\mathbf{A}$.

4.2.1 Inequality Constraints

The matrix $\mathbf{A}$ describes the set of potential RB allocations for all users. It is comprised of individual allocations given as

$$\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_K], \tag{20}$$

where $\mathbf{A}_i$ is a matrix containing the set of feasible allocations for UE i. Each column of $\mathbf{A}_i$ corresponds to a feasible allocation while each row corresponds to a specific resource. Within a given column, the kth row corresponds to the $(k \bmod M)$th RB and subframe $n = \lfloor \frac{k-1}{M} \rfloor$ of ARQ slot m.

Each entry in $\mathbf{A}_i$ can take a value of $\{0,1\}$. An entry is 1 if the particular resource is required by a UE for that allocation and 0 otherwise.

The set of possible allocation is determined as follows for each UE. During any ARQ slot, a UE is allocated a TB of T_i bits. For a given T_i, the maximum and minimum number of RBs can be found from Table 3 using b_{min} and b_{max} as

$$N_{min} = \left\lceil \frac{T_i}{\varphi b_{max}} \right\rceil, \tag{21}$$

$$N_{max} = \min\left(\left\lceil \frac{T_i}{\varphi b_{min}} \right\rceil, M \right). \tag{22}$$

The effective MCS scheme is a function of the number of RBs and T_i. For each possible contiguous block of RBs of size $[N_{min}, N_{max}]$, the power level needed to maintain $BLER_{tgt}$ for all possible allocations of contiguous resource blocks is found using (5). For each such possible contiguous blocks above and each subframe of ARQ slot m, a column allocation in given $\mathbf{A}_i$ with corresponding transmission power in the corresponding entry of $\mathbf{c}$.

Example. To demonstrate how to generate $\mathbf{A}_i$, consider the following example. Suppose during ARQ slot m and for UE i, $T_i(m) = 1{,}000$ bits and there are 10 uplink resource blocks ($M = 10$). Using the set of CQIs in Table 3, we observe that 2 and 10 are the minimum and maximum unique quantity of RBs that satisfy the above. Therefore, we let all eligible RB allocations fall within 2 and 10 RBs inclusive $[2, 10]$ (we denote this set $\mathcal{E}$) with spectral efficiencies given as solving (1) set equal to $T_i(m)$.

Suppose we consider $M'(b_i) = 6$. It can easily be shown that for $M = 10$ (to simplify the example), there are five $(M - M'(b_i) + 1 = 10 - 6 + 1)$ ways to allocate six RBs contiguous in frequency. As a result, we define a submatrix $\bar{\mathbf{A}}_i^{(6)}$ computed as the five possible allocations or as

$$\bar{\mathbf{A}}_i^{(6)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{23}$$

Next, we define ${}^{(n)}\bar{\mathbf{A}}_i^{(6)}$ as the eligible allocations of six RBs for subframe n. It is given as

$${}^{(n)}\bar{\mathbf{A}}_i^{(6)} = \begin{cases} \bar{\mathbf{A}}_i^{(6)}, & \text{if User } i \text{ has no ReTx in } n, \\ \emptyset, & \text{otherwise,} \end{cases} \tag{24}$$

where here $\emptyset$ is used to denote a 0×0 (empty) matrix. From this we obtain $\mathbf{A}^{(6)}$ as

$$\mathbf{A}_i^{(6)} = \begin{bmatrix} {}^{(0)}\bar{\mathbf{A}}_i^{(6)} & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\bar{\mathbf{A}}_i^{(6)} & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\bar{\mathbf{A}}_i^{(6)} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\bar{\mathbf{A}}_i^{(6)} \end{bmatrix}. \tag{25}$$

where ${}^{(n)}\mathbf{0}$ is a matrix of zeros identical to size of ${}^{(n)}\mathbf{A}_i^{(6)}$. The above corresponds to the feasibility of allocating any user in any subframe of the ARQ slot in which it does not have an ongoing retransmission. Finally we obtain $\mathbf{A}_i$ as

$$\mathbf{A}_i = \left[\mathbf{A}_i^{(2)}, \mathbf{A}_i^{(3)}, \mathbf{A}_i^{(4)}, \mathbf{A}_i^{(5)}, \mathbf{A}_i^{(6)}, \mathbf{A}_i^{(7)}, \mathbf{A}_i^{(8)}, \mathbf{A}_i^{(9)}, \mathbf{A}_i^{(10)}\right], \tag{26}$$

where the above is found by concatenating each submatrix obtained from each entry of $\mathcal{E}$.

4.2.2 Equality Constraints

The equality matrix $\mathbf{A}_{eq}$ is simply a matrix of K rows constraining the number of selected allocations such that each UE is only allotted one allocation selection from their matrix $\mathbf{A}_i$. This is given as

$$\mathbf{A}_{eq} = \begin{bmatrix} \mathbf{1}_{C_1}^T & \cdots & \mathbf{0}_{C_K}^T \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{C_1}^T & \cdots & \mathbf{1}_{C_K}^T \end{bmatrix}, \tag{27}$$

where C_i is the number of columns in $\mathbf{A}_i$ and $\mathbf{1}_x$ and $\mathbf{0}_x$ are column vectors of length x.

4.2.3 Cost Function

The objective function vector $\mathbf{c}^T = [\mathbf{c}_1^T, \ldots, \mathbf{c}_K^T]$ is simply the cost of choosing the corresponding allocation for each $\mathbf{c}_i$. From (19) we see the cost is simply the weighted power of choosing an allocation. Since the power is directly a function of the set of resources, the channel, and the TB size, individual entries of $\mathbf{c}_i$ can be then be given as

$$c_{i:j_i} = \alpha_i P_i(\mathcal{N}_i(j_i), T_i),\ j_i = 1, 2, \ldots, C_i, \tag{28}$$

where the function $P_i(\cdot)$ is given in (9), α_i is the priority weight of UE i and where $\mathcal{N}_i(j_i) = \{x \mid a_{i:x,j_i} = 1, x = 1, 2, \ldots, M\}$. The quantity $a_{i:x,j_i}$ denotes the $\{x, j_i\}$ entry in $\mathbf{A}_i$ and j_i is the j_ith column of $\mathbf{A}_i$.

The above problem is solved at each ARQ slot m online for given $T_i(m), \forall i$.[4] The above will yield the optimal solution at each decision epoch, and is less computationally intensive than the OFDM MA method with additional contiguity constraints as the search space is reduced precomputation. The resulting framework is still computationally prohibitive for online operation. In the following section, we propose two methods to reduce the problem complexity.

5 Suboptimal Resource Allocation Schemes

Due to the rather large search space above, we propose two suboptimal methods for the above approach. The first approach is a subset reduction technique, which is herein referred to as the Best-N scheme. In this method, we exploit the potential of choosing a subset of the best allocations (in terms of minimizing transmission power) for each user and allocating resources to satisfy this set. The second method, herein denoted as the iterative allocation method, allocates time-frequency resources to all users to maximize the *power level gain* (PLG) at each iteration. Both methods are described in the following sections.

5.1 Method 1 - Best-N Subset Reduction

The Best-N allocations represent the N allocations (set of RBs, MCS mode, and power allocated) that are the most energy efficient for the UE to use for transmission. This is essentially a precomputed subset matrix of the previously described $\mathbf{A}_i$. We denote this new constraint matrix as $\mathbf{B}_i$.

Once efficient allocations have been computed, the revised set-packing problem can be formulated using binary programming as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{d}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{B}\mathbf{x} \leq \mathbf{1}_{4M}, \mathbf{B}_{eq}\mathbf{x} = \mathbf{1}_K, \quad x_j \in \{0,1\}, \forall j \in \mathbf{x}, \end{aligned} \tag{29}$$

where $\mathbf{d}$ is a real-valued vector, $\mathbf{x}$ is the vector of allocation selections, $\mathbf{B}_{eq}$ is the equality constraint matrix of K rows and $\mathbf{B}$ is the inequality constraint matrix of $4M$ rows. Each nonzero entry of the solution vector $\mathbf{x}$ corresponds to selecting the corresponding column allocation in $\mathbf{B}$.

The matrix $\mathbf{B}$ is found as

$$\mathbf{B} = [\mathbf{B}_1, \cdots, \mathbf{B}_K], \tag{30}$$

where entries $\mathbf{B}_i$ are given as

$$\mathbf{B}_i = \begin{bmatrix} {}^{(0)}\bar{\mathbf{B}}_i & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\bar{\mathbf{B}}_i & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\bar{\mathbf{B}}_i & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\bar{\mathbf{B}}_i \end{bmatrix}, \tag{31}$$

where $\mathbf{0}$ is a matrix of zeros identical to size of ${}^{(n)}\bar{\mathbf{B}}_i$ and ${}^{(n)}\bar{\mathbf{B}}_i$ is

$${}^{(n)}\bar{\mathbf{B}}_i = \begin{cases} \bar{\mathbf{B}}_i, & \text{if User } i \text{ has no ReTx in } n, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{32}$$

The matrix $\bar{\mathbf{B}}_i$ is an $M \times N$ matrix containing the set of Best-N allocations. The matrix $\bar{\mathbf{B}}_i$ is obtained as follows: Let $\mathcal{N}(i,n)$ be the set of RB allocations for the nth best[5] allocation of i. Entries of $\bar{\mathbf{B}}_i$ are therefore given as

$$b_{i:n,j} = \begin{cases} 1, & j \in \mathcal{N}(i,n) \\ 0, & \text{otherwise} \end{cases}, j = 1, 2, \ldots, M. \tag{33}$$

Consequently, the matrix $\mathbf{B}_{eq}$ is given as

$$\mathbf{B}_{eq} = \begin{bmatrix} \mathbf{1}_{\bar{C}_1}^T & \cdots & \mathbf{0}_{\bar{C}_K}^T \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{\bar{C}_1}^T & \cdots & \mathbf{1}_{\bar{C}_K}^T \end{bmatrix}, \tag{34}$$

where $\bar{C}_i$ is used to denote the number of columns in $\mathbf{B}_i$. The vector $\mathbf{d}$ is obtained identical to $\mathbf{c}$ for entries corresponding to allocations in $\mathbf{B}$.

Algorithm 1. Iterative Resource Allocation: Initialization.

1: $\mathcal{K} = \{1, 2, \ldots, K\}$
2: *Allocate ReTx with required power*
3: $\mathcal{N}_n^{(r)} \leftarrow$ *Set of RBs in n with ReTx,* $n = 0, 1, 2, 3$
4: $\mathcal{Z}_i \leftarrow$ *Set of subframes with ReTx for UE i,* $n = 0, 1, 2, 3 \;\; \forall i \in \mathcal{K}$
5: $\mathbf{n}(i) = -1, \forall i \in \mathcal{K}$
6: $\mathcal{N}_n^{(a)} = \{1, 2, \ldots, M\} \setminus \mathcal{N}_n^{(r)}, \; n = 0, 1, 2, 3$
7: $\mathcal{N}_i = \emptyset, \forall i \in \mathcal{K}$
8: **for** $i \in \mathcal{K}$ **do**
9: $\quad \mathcal{N}_i^{(f)} = \bigcup_{n=0, n \neq \mathcal{Z}_i}^{3} \mathcal{N}_n^{(a)}$
10: $\quad N_{min,i} = \lceil \frac{T_i(m)}{\varphi b_{max}} \rceil$
11: $\quad N_{max,i} = \min(\lceil \frac{T_i(m)}{\varphi b_{min}} \rceil, M)$
12: **end for**

Algorithm 1. Iterative Resource Allocation: Iterative Component.

13: **while** $\max(PLG_i) > 0$ && $\bigcup_{n=0}^{3} \mathcal{N}_n^{(a)} \neq \emptyset$ **do**
14: $\quad$ **for** $i \in \mathcal{K}$ **do**
15: $\quad\quad$ **if** $\mathcal{N}_i \neq \emptyset$ && $|\mathcal{N}_i| < N_{max,i}$ **then**
16: $\quad\quad\quad$ **for** $j \in \mathcal{N}_i^{(f)} \cap \mathcal{N}_{\mathbf{n}(i)}^{(a)}$ **do**
17: $\quad\quad\quad\quad \Delta p_{i,j} = \alpha_i (P(\mathcal{N}_i, T_i, \gamma) - P(\mathcal{N}_i \cup j, T_i, \gamma))$
18: $\quad\quad\quad$ **end for**
19: $\quad\quad\quad PLG_i = \max(\{\Delta p_{i,j} | j \in \mathcal{N}_i^{(f)} \cap \mathcal{N}_{\mathbf{n}(i)}^{(a)}\})$
20: $\quad\quad$ **else**
21: $\quad\quad\quad \mathcal{N}_i^{(c)} = \mathcal{G}(N_{min,i}, \{\mathcal{N}_n^{(a)} | \forall n\})$
22: $\quad\quad\quad$ **if** $|\mathcal{N}_i^{(c)}|_x == 1$ **then**
23: $\quad\quad\quad\quad i^* = i$
24: $\quad\quad\quad\quad x^* = 0$
25: $\quad\quad\quad\quad$ *Go To Line 38*
26: $\quad\quad\quad$ **end if**
27: $\quad\quad\quad$ **for** $x = 0; x < |\mathcal{N}_i^{(c)}|_x$ **do**
28: $\quad\quad\quad\quad p_{i,x} = \alpha_i P(\mathcal{N}_{i,x}^{(c)}, T_i, \gamma)$
29: $\quad\quad\quad$ **end for**
30: $\quad\quad\quad \mathcal{N}^{(B)} = \{0, 1, \ldots, |\mathcal{N}_i^{(c)}|_x - 1\}$

4. $T_i(m)$ can be a constant or vary in time.

5. The nth best allocation refers to the allocation that achieves the nth smallest required power allocation subject to any additional constraints.

31: $PLG_i = \min(\{p_{i,x}, x \in \mathcal{N}^{(B)} \setminus \arg\min_{x^* \in \mathcal{N}^{(B)}}(p_{i,x^*})\}) - \min(\{p_{i,x}, \forall x \in \mathcal{N}^{(B)})$
32: **end if**
33: **end for**
34: **if** $\max(PLG_i) < 0$ **then**
35: **break**
36: **end if**
37: $i^* = \arg\max(PLG_i)$
38: **if** $\mathbf{n}(i^*) \stackrel{i}{=}= -1$ **then**
39: $x^* = \arg\max_x(\{p_{i,x} | i = i^*\})$
40: $\mathcal{N}_{i^*} = \mathcal{N}^{(c)}_{i^*,x^*}$
41: $\mathbf{n}(i^*) = \mathcal{F}(x^*, i^*)$
42: $n^* = \mathbf{n}(i^*)$
43: $\mathcal{N}^{(a)}_{n^*} = \mathcal{N}^{(a)}_{n^*} \setminus \mathcal{N}^{(c)}_{i^*,x^*}$
44: **else**
45: $n^* = \mathbf{n}(i^*)$
46: $j^* = \arg\max_j(\{\Delta p_{i,j} | i = i^*\})$
47: $\mathcal{N}_{i^*} = \mathcal{N}_{i^*} \cup j^*$
48: $\mathcal{N}^{(a)}_{n^*} = \mathcal{N}^{(a)}_{n^*} \setminus j^*$
49: **end if**
50: $\mathcal{N}^{(f)}_{i^*} = \{\min(\mathcal{N}_{i^*}) - 1, \max(\mathcal{N}_{i^*}) + 1\} \cap \mathcal{N}^{(a)}_{n^*}$
51: **end while**

In rare cases the solution may be infeasible during a given scheduling epoch as we do not exhaustively allow selection of every possible allocation. To minimize the probability of this occurrence, the following steps can be taken:

- Employ appropriate admission control design in the TDPS.
- Limit the maximum number of resource blocks that can be allocated to a single user as a function of available resources and required data rate.
- Enforce overlap restrictions on the Best-N selection scheme such that no two possible allocations for a single UE may share any resource blocks (employed in this implementation).
- Increment N and resolve the problem if the solution is found infeasible (employed in this implementation).

There are no guarantees that the Best-N method will result in a solution without large increases in N. For example, if the system is highly loaded (size of TBs and number of users is large), the computational complexity may tend toward that of the optimal allocation by incrementing N until a solution is found. As a result, it is not a robust implementation for heavily loaded systems from the perspective of low computation time. In the following section, we propose an additional method that iteratively allocates RBs to each user which can be used for such systems where computational complexity is not impacted by the size of TBs.

5.2 Method 2—Iterative Allocation

The second suboptimal method tries to iteratively allocate resources to all UEs. At each iteration, resources are allocated to the user to maximize the *power level gain*. This method is described mathematically in Algorithm 1. The function $\mathcal{F}(x, i)$ returns the subframe index corresponding to index x for UE i and $\mathcal{G}(N, \mathcal{N}_i, \mathcal{Z}_i)$ returns a set $\mathcal{N}_i$ of subsets $\mathcal{N}_{i,x}$ containing all unique contiguous RBs of size N in all subframes $n \notin \mathcal{Z}_i$ (subframes which are not allocated for retransmission for UE i). The number of subsets in $\mathcal{N}_i$ is $|\mathcal{N}_i|_x$ and the size of each subset $\mathcal{N}_{i,x}$ is $|\mathcal{N}_{i,x}| = N$.

The proposed algorithm operates as follows: The initialization of the algorithm is described by lines 1-12. The iterative allocation portion of the algorithm (described in lines 13-51 inclusive) is divided into two major components: the power level gain computation (lines 13-33) and the allocation stage (lines 37-50).

5.2.1 Initialization Procedure

The initialization stage is described by lines 1-12. As with the optimal and Best-N allocation schemes, retransmissions are allocated first to determine the residual resources available for new transmissions. Next, the minimum and maximum amount of number of resource blocks for each UE i is determined.

5.2.2 Power Level Gain

The *power level gain* stage operates as follows: For users who have been already allocated resources, the PLG of a resource is calculated as the difference in power allocated if the new resource is added to the current resource allocation block compared to the power allocation required with the existing allocated resources (line 17). The newly added resource is constrained to those resources within a users feasible allocation set ($\mathcal{N}^{(f)}_i$). Alternatively, for any user who has yet to be allocated a resource, the PLG of that user is measured as the difference between the power allocation required if that best available resource is allocated and that of the second best available resource (line 31).

If at any instant of time, a user has only one eligible allocation, the user is allocated that resource (described in lines 22-25).

5.2.3 Resource Allocation Stage

The user with the maximum PLG at any iteration is allocated the corresponding resource (line 38-50). For users being allocated an initial resource, this constrains the feasible resources for allocation within that given subframe n.

The overall allocation continues until the maximum PLG is negative or there are no additional resources for transmission.

6 Numerical Evaluation

Simulation results are provided measuring the power and computation timing of the optimal, Best-N and iterative schemes. In addition, we measure the impact of static scheduling of retransmissions. Simulation parameters are given in Table 5.

6.1 Least-Squares Fit Function

The justification behind use of the predescribed fit function is shown in Fig. 3. Here we show the results for 2, 4, and 8 RBs with $BLER_{tgt} = 10\%$. The least-squares approximation

TABLE 5
Simulation Parameters

Parameter	Value	Parameter	Value
γ_0	10 dB	T_f	1ms
$\alpha_i(m)$	1 ($\forall i$)	T_e	4ms
$N_{overhead}$ (symbols)	3	N (Best-Ndepth)	3
N_{sym} (symbols)	14	$\bar{T}_i$	250 bits
Number of Subframes	20000	$BLER_{tgt}$	10%

is shown to hold tightly to the actual BLER function; providing for a more tractable computation of the required SNR level and justifying its use as a suitable alternative in computation of the required SNR.

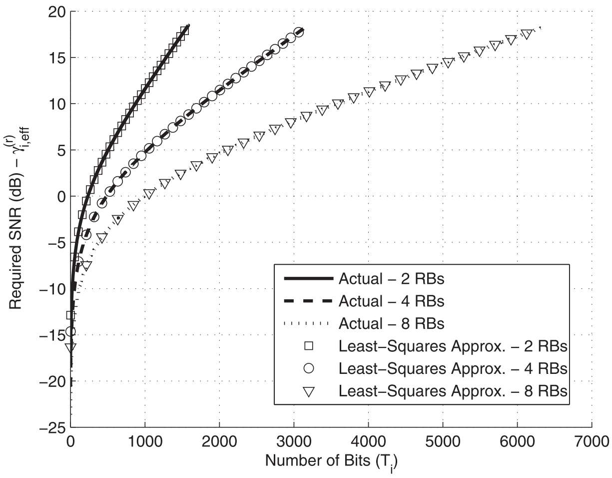


Fig. 3. Least-squares approximation accuracy.

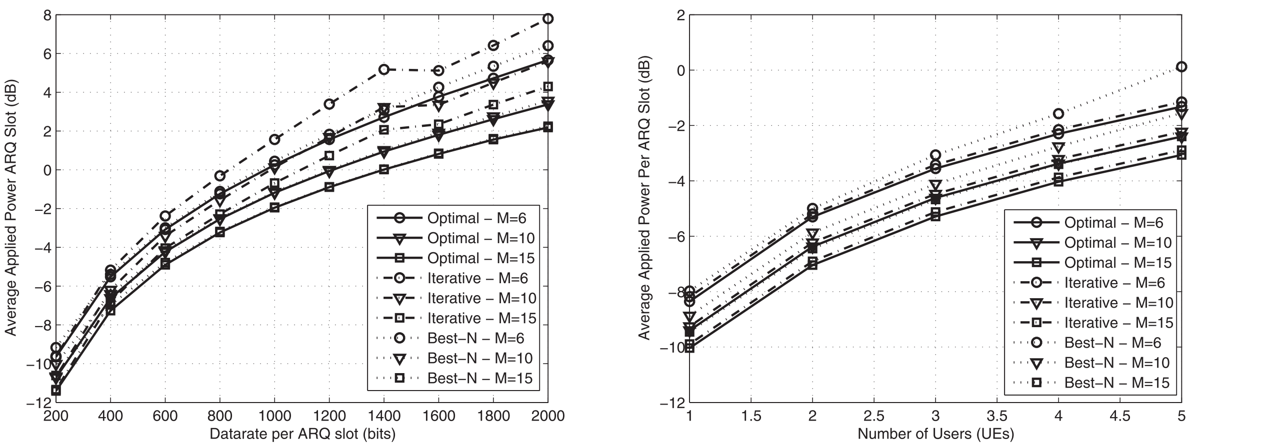
6.2 Retransmission Power—Static Scheduling

Worth noting is the important impact of static scheduling on retransmissions. While using the above framework, packet error rates are 10 percent, the overall power allocated for retransmission packets accounts for over 50 percent of allocated power. As a result, in some simulations, the Best-N method requires less overall power to be allocated than the optimal method as a result of power expended for retransmissions. Overall, we observe the performance of the iterative method tends to perform worse in this regime. This is a result of the average number of resource blocks being assigned to any user being lower than with the other two methods. This is particularly detrimental if a user experiences a low quality channel in those RB(s) as the channel is not averaged over a larger set of resources.

This overall result highlights the importance of reallocating TBs for retransmission within a given ARQ slot. To accomplish this, the previously described mechanisms could be modified as follows: For both the optimal and Best-N scheme, each user with a retransmission can be considered as an additional user to allocate. Unlike new transmissions, these transmission have a fixed number of RBs, in addition are limited to a single subframe. As a result, there is a relatively small complexity increase for these additional users. For the iterative mechanism, retransmissions can be incorporated by considering them as additional users. In this case, $N_{min} = N_{max} = |\mathcal{N}_i|$ where here $\mathcal{N}_i$ denotes the set of resource blocks for the initial transmission of a given TB (replacing lines 10 and 11 in).

6.3 Optimal Power Expenditure Gap

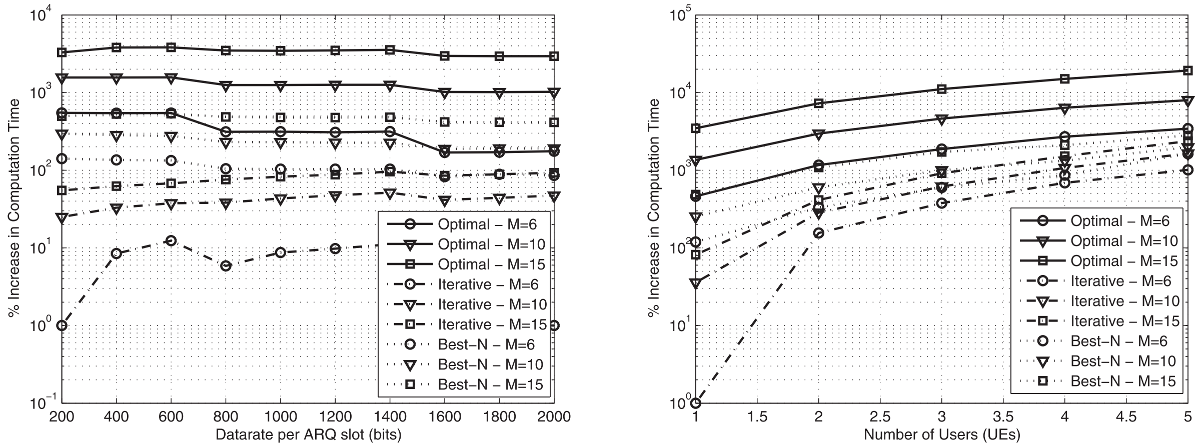
The average power expenditure per ARQ slot for the initial transmission of a TB is shown in Fig. 4. Here we observe the following: In general the amount of power expended increases with both data rate and the number of UEs. For lower data rates, the iterative mechanism outperforms the Best-N scheme, however, as the per ARQ slot data rate increases, Best-N performance surpasses that of the iterative method. At approximately $T_i = 1{,}400$ bits, the iterative method experiences an uncharacteristic trend. This is a result of an increase to N_{min} at that data rate, forcing the iterative scheme to maximize *power level gain* of more than one RB for the initial allocation. In this way, the scheduler is better able to allocate a block of RBs. The effect of the number of UEs on the power between the two suboptimal methods is negligible as it largely depends on the data rate. Both suboptimal methods obtain near optimal per ARQ slot performance.



(a) versus Datarate (K=1)

(b) versus Number of UEs

Fig. 4. Applied power for initial transmission.



(a) versus Datarate (K=1)

(b) versus Number of UEs

Fig. 5. Computation time.

6.4 Complexity Comparison

The system complexity (measured in relative computation time) is shown in Fig. 5. This is measured as the percentage increase in computation time taken compared to the base case (iterative method and $M = 6$ for both figures and $T_i = 200$ and $K = 1$ for Figs. 5a and 5b, respectively). As expected, the iterative method has the lowest complexity of all schemes. Further, the number of UEs and/or RBs results in a large increase in complexity, while the data rate is shown to have negligible impact.

7 Conclusion

In this work, we have proposed framework for energy efficient resource allocation in the SC-FDMA multiuser uplink. First, we proposed an alternative method of selecting an appropriate scheduling epoch based on the impact of synchronous HARQ. By utilizing the proposed method, we can reduce the number of allocation procedures in time and ensure users can always initiate a new transmission during any frame (i.e., do not experience ARQ blocking).

Second, we proposed two suboptimal power efficient resource allocation methods. Both methods were compared to the optimal method in terms of complexity and power efficiency. We found that the suboptimal methods closely obtain the power efficiency of the optimal allocation with reduced complexity. Further, we found that the efficiency of the power allocation scheme is dramatically reduced when static scheduling is employed for retransmissions. In our future work, we will study the impact of intercell interference on the proposed framework.

Acknowledgments

D.J. Dechene was with the Department of Electrical and Computer Engineering, The University of Western Ontario, London, ON, Canada, during this work. Part of this work was presented at the IEEE International Conference on Communications, 2011.

References

[1] C.Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Selected Areas in Comm.*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999.

[2] M. Bohge, J. Gross, A. Wolisz, and M. Meyer, "Dynamic Resource Allocation in OFDM Systems: An Overview of Cross-Layer Optimization Principles and Techniques," *IEEE Network*, vol. 21, no. 1, pp. 53-59, Jan. 2007.

[3] H.G. Myung, J. Lim, and D.J. Goodman, "Single Carrier FDMA for Uplink Wireless Transmission," *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 30-38, Sept. 2006.

[4] L.R. de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen, "Channel-Aware Scheduling Algorithms for SC-FDMA in LTE Uplink," *Proc. IEEE 19th Int'l Symp. Personal, Indoor and Mobile Radio Comm. (PIMRC '08)*, pp. 1-6, Sept. 2008.

[5] O. Delgado and B. Jaumard, "Scheduling and Resource Allocation for Multiclass Services in LTE Uplink Systems," *Proc. IEEE Sixth Int'l Conf. Wireless and Mobile Computing, Networking and Comm. (WiMob '10)*, pp. 355-360, 2010.

[6] Z. Li, C. Yin, and G. Yue, "Delay-Bounded Power-Efficient Packet Scheduling for Uplink Systems of LTE," *Proc. Fifth Int'l Conf. Wireless Comm., Networking and Mobile Computing (WiCom '09)*, pp. 1-4, 2009.

[7] A. Ahmad and M. Assaad, "Polynomial-Complexity Optimal Resource Allocation Framework for Uplink SC-FDMA Systems," *Proc. IEEE Global Telecomm. Conf. (GlobeCom '11)*, pp. 1-5, 2011.

[8] F. Sokmen and T. Girici, "Uplink Resource Allocation Algorithms for Single-Carrier FDMA Systems," *Proc. IEEE Wireless Conf. (EW), European*, pp. 339-345, 2010.

[9] I. Wong, O. Oteri, and W. Mccoy, "Optimal Resource Allocation in Uplink SC-FDMA Systems," *IEEE Trans. Wireless Comm.*, vol. 8, no. 5, pp. 2161-2165, May 2009.

[10] F. Calabrese, P. Michaelsen, C. Rosa, M. Anas, C. Castellanos, D. Villa, K. Pedersen, and P. Mogensen, "Search-Tree Based Uplink Channel Aware Packet Scheduling for UTRAN LTE," *Proc. IEEE Vehicular Technology Conf.*, pp. 1949-1953, May 2008.

[11] F. Calabrese, C. Rosa, M. Anas, P. Michaelsen, K. Pedersen, and P. Mogensen, "Adaptive Transmission Bandwidth Based Packet Scheduling for LTE Uplink," *Proc. IEEE Vehicular Technology Conf.*, pp. 1-5, Sept. 2008.

[12] S. Jungsup, G. Gye-Tae, and K. Dong-Hoi, "Packet-Scheduling Algorithm by the Ratio of Transmit Power to the Transmission Bits in 3GPP LTE Downlink," *EURASIP J. Wireless Comm. Networking*, vol. 2010, article 6, Jan. 2010.

[13] "Evolved Universal Terrestrial Radio Access (E-UTRA); Long Term Evolution (LTE) Physical Layer; General Description," 3GPP TS 36.201, 2009.

[14] "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures," 3GPP TS 36.213, 2008.

[15] D. Buckingham, "Information-Outage Analysis of Finite-Length Codes," PhD dissertation, West Virginia Univ., http://wvuscholar.wvu.edu:8881/exlibris/dtl/d3_1/apache_media/13988.pdf, 2008.
[16] Q. Liu, S. Zhou, and G. Giannakis, "Queuing with Adaptive Modulation and Coding over Wireless Links: Cross-Layer Analysis and Design," *IEEE Trans. Wireless Comm.*, vol. 4, no. 3, pp. 1142-1153, May 2005.
[17] F. Calabrese, "Scheduling and Link Adaptation for Uplink SC-FDMA Systems," PhD dissertation, Aalborg Univ., http://vbn.aau.dk/files/19156393/PhD_Thesis_FrancescoDavideCalabrese_final_print.pdf, 2009.
[18] J. Ikuno, M. Wrulich, and M. Rupp, "Performance and Modeling of LTE H-ARQ," *Proc. IEEE Workshop Smart Antennas*, pp. 1-6, Feb. 2009.
[19] M. Al-Rawi, R. Jantti, J. Torsner, and M. Sagfors, "On the Performance of Heuristic Opportunistic Scheduling in the Uplink of 3G LTE Networks," *Proc. IEEE Personal, Indoor and Mobile Radio Comm.*, pp. 1-6, Sept. 2008.

Dan J. Dechene received the BEng degree in electrical engineering from Lakehead University, Thunder Bay, Ontario, Canada, in 2006. He received the master's and PhD degrees in electrical and computer engineering from the University of Western Ontario, London, Canada, in 2008 and 2012, respectively, with a research focus on energy efficient resource allocation techniques and multiple antenna systems. He is an engineer at IBM's Semiconductor Research and Development Center in Hopewell Junction, New York. He is a member of the IEEE.

Abdallah Shami received the BE degree in electrical and computer engineering from Lebanese University, Beirut, Lebanon, in 1997, and the PhD degree in electrical engineering from the Graduate School and University Center, City University of New York, in September 2002. In September 2002, he joined the Department of Electrical Engineering at Lakehead University, Thunder Bay, Ontario, Canada, as an assistant professor. Since July 2004, he has been with Western University, Canada, where he is an associate professor in the Department of Electrical and Computer Engineering. His current research interests are in the area of wireless/optical networking. He is an associate editor for *IEEE Communications Letters* and *Wiley Journal of Communications Systems*. He has chaired key symposia for IEEE GlobeCom, IEEE ICC, IEEE ICNC, and ICCIT. He is a senior member of the IEEE.