We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter 6

Simulated Annealing and Multiuser Scheduling in Mobile Communication Networks

Raymond Kwan, M. E. Aydin and Cyril Leung

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/45869

1. Introduction

Adaptive modulation and coding (AMC) is an effective way for improving the spectral efficiency in wireless communication systems. By increasing the size of the modulation scheme constellation, the spectral efficiency can be improved, generally at the cost of a degraded error rate. A similar trade-off is possible by using a higher rate channel code. By an appropriate combination of the modulation order and channel code rate, we can design a set of modulation and coding schemes (MCSs), from which an MCS is selected in an adaptive fashion in each transmission-time interval (TTI) in order to maximize system throughput under different channel conditions. The use of AMC yields a rich variety of scheduling strategies [25]; [12]. In practice, a commonly encountered constraint is that the probability of erroneous decoding of a Transmission Block should not exceed some threshold value [11].

Multiple orthogonal channelization codes (multicodes) can be used to transmit data to a single user, thereby increasing the per-user bit rate and the granularity of adaptation [11, 16]. In Wideband Code-Division Multiple Access (WCDMA), the channelization codes are often referred to as Orthogonal Variable Spreading Factor (OVSF) codes. The number of OVSF codes per base station (BS) is quite limited due to the orthogonality constraint [11] and thus OVSF codes and transmit power are scarce resources. Fig. 1 shows the number of OVSF codes as a function of the spreading factor for WCDMA. Note that a lower value of spreading factor corresponds to a higher bit rate and vice versa. According to Fig. 1, if a spreading factor of 2 is needed, the system can allocate at most two such OVSF codes. On the other hand, if a spreading factor of 4 is required, a total of 4 such codes can be allocated. In High Speed Downlink Packet Access (HSDPA), a fixed spreading factor of 16 has been specified, thereby limiting the number of OVSF codes to 16[1].

The allocation of the number of OVSF codes (or multicodes) and the MCS level for each user depends on the strength of the received signal at the user, which, in turn, depends on 1) the quality of the wireless channel, and 2) the level of the transmit power to the respective user.

[1] In principle, 16 OVSF codes can be used. However, one code is allocated for other purposes such as signalling. Thus, a maximum of 15 codes can be allocated for data traffic [11].

As shown in Fig. 2, if the spreading factor and the number of OVSF codes are fixed, one way to increase the user bit rate is to increase the MCS level. As each MCS level is associated with a specific signal quality requirement, the highest MCS level that can be allocated depends on the channel quality at the user receiver, which is stochastic by nature. Thus, at a given channel quality at the receiver, the MCS level can be increased by increasing the transmit power to the user. Another way to increase the user bit rate at a fixed spreading factor is to fix the MCS level while increasing the number of OVSF codes allocated to the respective user, as shown in Fig. 3. In order to achieve a given signal quality requirement for each OVSF code, a higher power level is required to be allocated to this user. Thus, the general problem of HSDPA resource allocation boils down to the joint allocation of user-specific MCS, number of OVSF codes, and power level over all users connected to a given base station, subject to the constraints of code and power resources as shown in Fig.4. It is important to note that this allocation is done very rapidly (on the order of two or more milliseconds) in order to exploit the channel diversity of the users. HSDPA is based on a shared channel concept, in which multiple users share the channel in a time-multiplexed fashion, and the process of resource allocation is performed at regular time intervals. Note that in HSDPA, as shown in Fig. 5, the shared channel is a dual to the dedicated channel in which the bit rate for each user is kept constant over a relatively long time period by appropriate closed-loop power control.

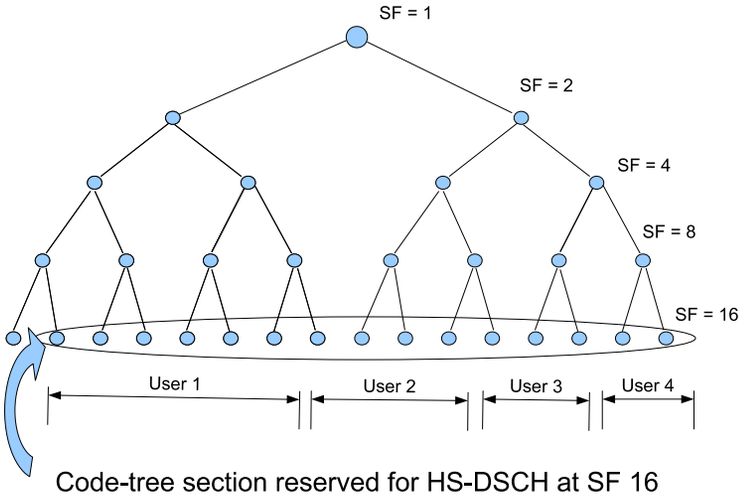

Figure 1. Orthogonal Variable Spreading Factor (OVSF) code tree.

For simplicity, the downlink transmit power is normally held constant (or slowly changing)[2] in HSDPA [11]. A number of scheduling algorithms have been proposed for HSDPA [7]. The most commonly encountered ones are (1) round-robin in which users are allocated resources in turn, regardless of channel conditions (2) Max C/I in which resources are allocated to the user with the best channel condition (3) proportional fair in which resources are assigned to the user with the *relatively* best channel condition. Other schedulers include minimum bit rate (MBR), MBR with proportional fairness and minimum delay (MD).

In exploiting multiuser diversity, a common way to achieve the best network throughput is to assign resources to a user with the largest signal-to-noise ratio (SNR) among all backlogged

[2] In some cases, the specification stipulates a slight power reduction for a mobile with an exceptionally good channel quality [22].

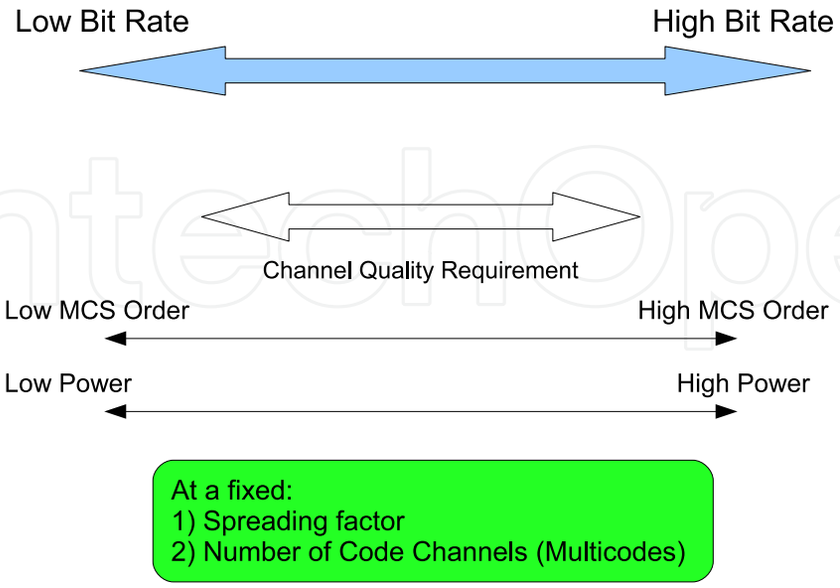

Figure 2. Bit rate and channel quality requirement trade-off for HSDPA.

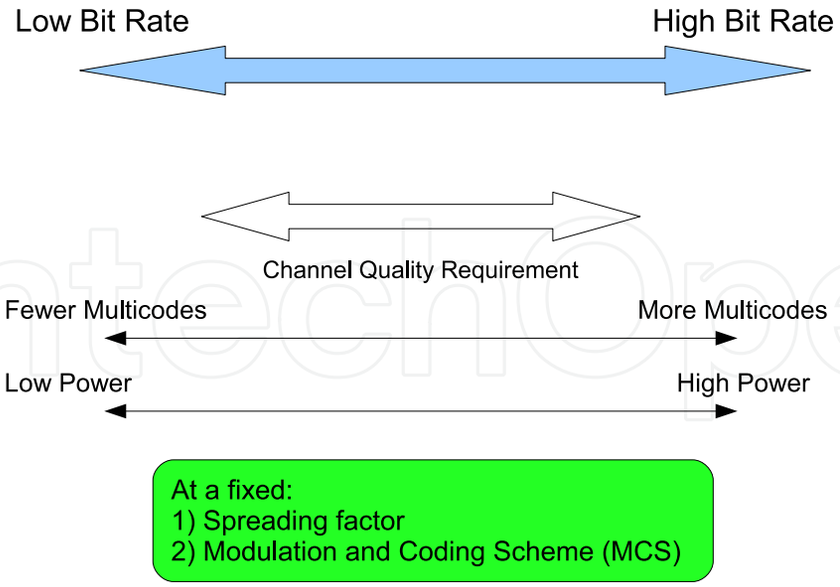

Figure 3. Bit rate and channel quality requirement trade-off for HSDPA.

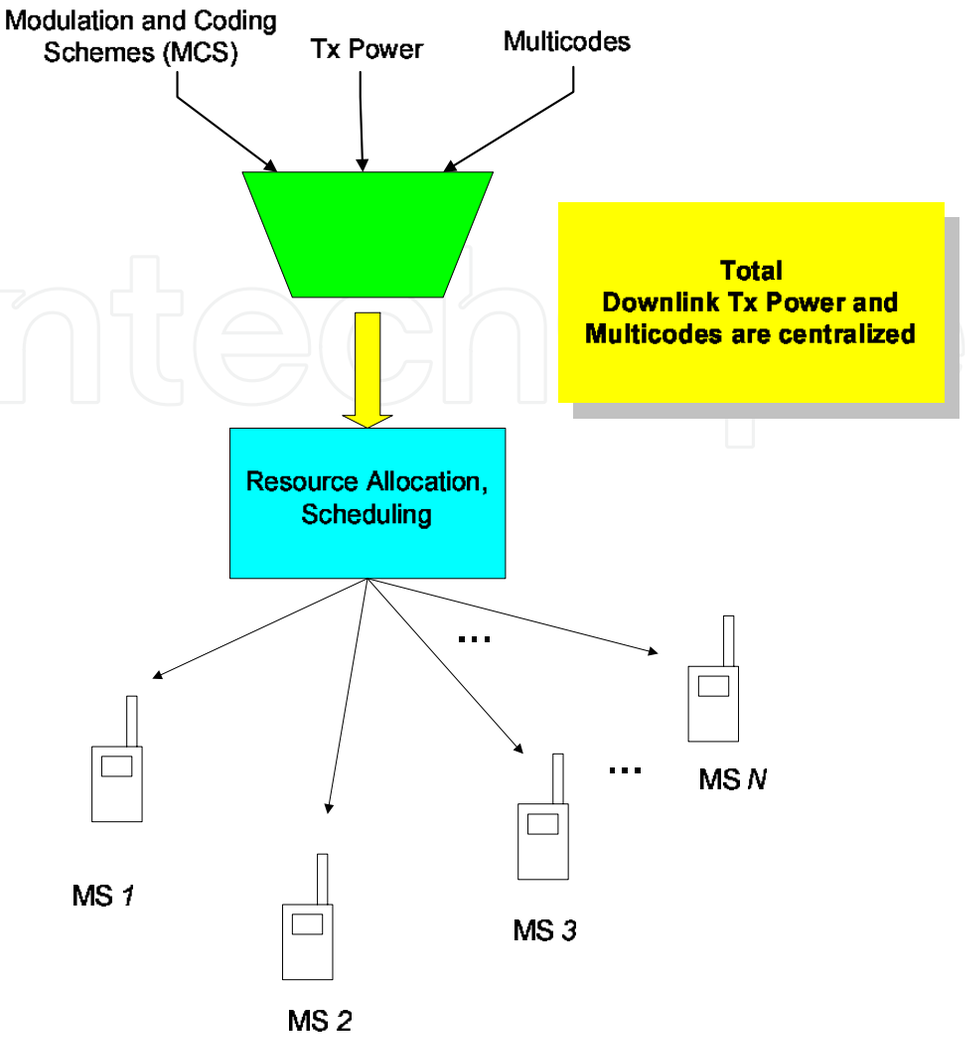

Figure 4. Resource allocation in HSDPA.

users (i.e. users with data to send) at the beginning of each scheduling period [5]. However, due to limited mobile capability, a user might not be able to utilize all the radio resources available at the BS. Thus, transmission to multiple users during a scheduling period may be more resource efficient. In [2, 15], the problem of downlink multiuser scheduling subject to limited code and power constraints is addressed. It is assumed in [15] that the exact path-loss and received interference power at every TTI for each user are fed back to the BS. This would require a large bandwidth overhead.

In this chapter, the problem of optimal (maximum aggregate throughput) multiuser scheduling in HSDPA is addressed[3]. The MCSs, numbers of multicodes and power levels for all users are jointly optimized at each scheduling period, given that only limited CSI information, as specified in the HSDPA standard [22], is fed back to the BS. This problem corresponds to a general resource allocation formulation for HSDPA, as compared to those in the existing literature, where resources are typically not allocated jointly among users for simplicity. Due to the inherent complexity, an integer programming formulation is proposed for the above problem. Due to the complexity of obtaining an optimal solution, an

[3] The materials presented here are mostly based on the contents in [14].

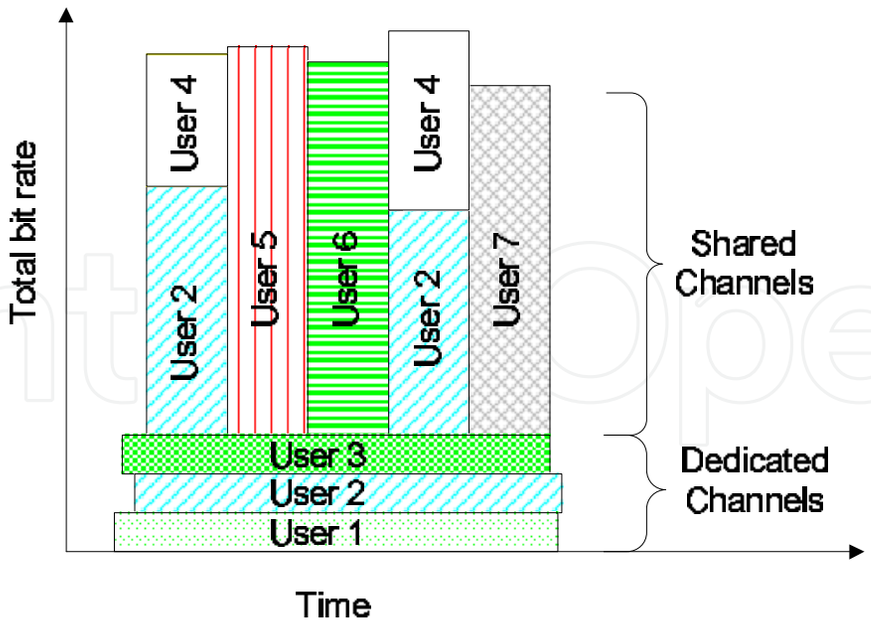

Figure 5. Resource allocation for shared and dedicated channels.

evolutionary simulated annealing (ESA) approach is explored. It is shown that ESA can provide a near-optimum performance with significantly reduced complexity.

2. System model

In a communication system, the quality of the channel is often quantified by the Signal to Interference and Noise Ratio (SINR), which is defined as the ratio of the received signal power relative to the power contribution from interference and noise. A better channel quality is represented by a higher received signal power and a smaller interference and noise power.

We consider downlink transmissions from a BS to a number of mobile users. Let P_i denote the downlink transmit power to user i, h_i denote the link gain from the BS to user i, and I_i be the total received interference and noise power at user i. The received SINR for user i is then given by

$$\gamma_i = \frac{h_i P_i}{I_i}, \; i = 1, \ldots, N, \tag{1}$$

where N is the number of users and

$$\sum_{i=1}^{N} P_i \leq P_T \tag{2}$$

where P_T is the total HSDPA power constraint. Ideally, the user would measure the received SINR, and report its value back to the BS. Upon receiving the SINR value for this user, the BS would decide what MCS and the number of multicodes that the user can be allocated, taking into account all the resource constraints that the BS has. However, if each user i were to send back its exact SINR value γ_i to the BS, the required feedback channel bandwidth would be impractically large. As specified in [22], the channel quality information fed back by a

mobile, also known as the *channel quality indicator* (CQI), can only take on a finite number of non-negative integer values $\{0, 1, \ldots, K\}$. According to [22], the CQI is provided by the mobile via the High Speed Dedicated Physical Control Channel (HS-DPCCH). Each CQI value maps directly to a maximum bit rate[4] that a mobile can support, based on the channel quality and mobile capability [23], while ensuring that the block error rate (BLER) does not exceed 10%. Finally, upon receiving the CQIs from all the users, the BS decides on the most appropriate combination of MCS and number of multicodes for each user.

Although the mapping between the CQI and the SINR is not specified in [22], it has been discussed in various proposals [18]; [13]. In [8], a mapping is proposed in which the system throughput is maximized while the BLER constraint is relaxed. Let $\tilde{\gamma}_i = 10\log_{10}(\gamma_i)$ be the received SINR value, in dB, for user i and let q_i be the CQI value that user i reports back to the BS via HS-DPCCH. The mapping between q_i and $\tilde{\gamma}_i$ can generally be expressed as a piece-wise linear function [6, 8, 18]

$$q_i = \begin{cases} 0 & \tilde{\gamma}_i \leq t_{i,0} \\ \lfloor c_{i,1}\tilde{\gamma}_i + c_{i,2} \rfloor & t_{i,0} < \tilde{\gamma}_i \leq t_{i,1} \\ q_{i,max} & \tilde{\gamma}_i > t_{i,1} \end{cases} \tag{3}$$

where the terms $\{c_{i,1}, c_{i,2}, t_{i,0}, t_{i,1}\}$ are model and mobile capability dependent constants, and $\lfloor . \rfloor$ denotes the floor function. Due to the quantization operation implied in (3), $\tilde{\gamma}_i$ cannot generally be recovered exactly from the value of q_i alone. It should be noted that the region $t_{i,0} < \tilde{\gamma}_i \leq t_{i,1}$ is the operating region for the purpose of link adaptation. This region should be chosen large enough to accommodate the SINR variations encountered in most practical scenarios [11], i.e. the probability that $\tilde{\gamma}_i$ falls outside this region should be quite small. As part of our proposed procedure, $\tilde{\gamma}_i$ is approximated as

$$\tilde{\gamma}_i^{\dagger} = \tilde{\gamma}_i^{(l)} + \left(\tilde{\gamma}_i^{(u)} - \tilde{\gamma}_i^{(l)}\right)\xi, \tag{4}$$

where

$$\tilde{\gamma}_i^{(l)} = \frac{q_i - c_{i,2}}{c_{i,1}}, \tag{5}$$

$$\tilde{\gamma}_i^{(u)} = \frac{q_i + 1 - c_{i,2}}{c_{i,1}}, \tag{6}$$

and ξ is a uniformly distributed random variable, i.e. $\xi \sim U(0,1)$. In a more conservative design, the value of ξ could be set to 0. Note that this approximation assumes that $\tilde{\gamma}_i$ is uniformly distributed between $\tilde{\gamma}_i^{(l)}$ and $\tilde{\gamma}_i^{(u)}$ for a given value of q_i.

For $q_i = 0$ and $q_i = q_{i,max}$, $\tilde{\gamma}_i$ could be approximated as $t_{i,0}$ and $t_{i,1}$ respectively, or more generally as $t_{i,0} - \xi_{i,0}$ and $t_{i,1} + \xi_{i,1}$ respectively, with $\xi_{i,0}$ and $\xi_{i,1}$ following some pre-defined probability distributions. Finally, the estimated value of γ_i is given by $\hat{\gamma}_i = 10^{\tilde{\gamma}_i^{\dagger}/10}$. We refer to the mapping from SINR to q_i in (3) as the *forward mapping*, and the approximation of SINR based on the received value of q_i in (4) as the *reverse mapping*.

[4] In this chapter, the bit rate refers to the transport block size, i.e. the maximum number of radio link control (RLC) protocol data bit (PDU) bits that a transport block can carry, divided by the duration of a TTI, i.e. 2 ms [11].

3. Joint optimal scheduling

Note that the value of the channel quality, q_i, reported by user i indicates the rate index that is associated with the maximum bit rate that the user can be supported by the BS, and is related jointly to a required number of OVSF codes (multicodes) and MCS. The number of multicodes and MCS assigned to each user as well as the estimated SINR values of the users determine the transmit power required by the BS. Since the number of multicodes and transmit power are limited, the BS might not be able to simultaneously satisfy the bit rate requests for all users as indicated by $\{q_i, i = 1, \ldots, N\}$. Therefore, for a set $\{q_i, i = 1, \ldots, N\}$, the BS must calculate a set of *modified* CQIs, $\{J_i, i = 1, \ldots, N\}$, for all users by taking into account the transmit power and number of multicodes constraints.

From the *forward* and *reverse* mappings in (3) and (4), the modified CQIs are chosen as

$$J_i = \min\left(\max\left(\eta_i(\tilde{\gamma}_i^{\dagger}, \phi_i), 0\right), q_{i,max}\right), i = 1, \ldots, N \tag{7}$$

where ϕ_i is the power adjustment factor for user i, i.e. $\hat{\gamma}_i \mapsto \phi_i \hat{\gamma}_i$, and

$$\eta_i(\tilde{\gamma}_i^{\dagger}, \phi_i) = \lfloor c_{i,1}\left(\tilde{\gamma}_i^{\dagger} + 10\log_{10}\phi_i\right) + c_{i,2}\rfloor, \tag{8}$$

$$0 \leq \phi_i \leq 10^{\left(\frac{q_{i,max} - (c_{i,1}\tilde{\gamma}_i^{\dagger} + c_{i,2})}{10c_{i,1}}\right)}. \tag{9}$$

Fig. 6 summarizes the conversion process from the received CQI, q_i, to the final assigned rate index, J_i.

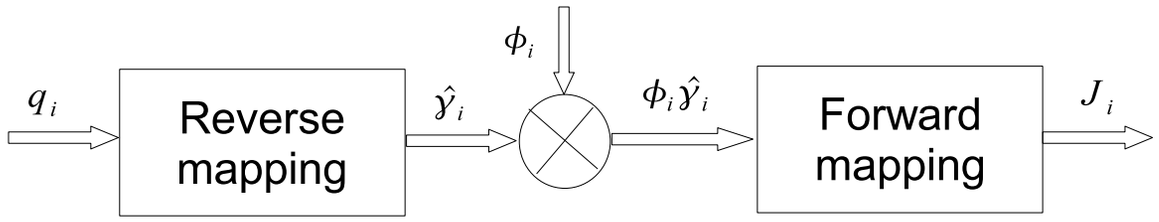

Figure 6. The conversion process from the received CQI, q_i, from the mobile to the assigned rate index J_i at the base station. ([14]©IET)

The multiuser joint optimal scheduling problem **P1** can be expressed as

$$\mathbf{P1}: \quad \max_{\mathbf{A},\underline{\phi}} \sum_{i=1}^{N}\sum_{j=0}^{J_i} a_{i,j} r_{i,j} \tag{10}$$

subject to (7)-(9) and

$$\sum_{j=0}^{J_i} a_{i,j} = 1, \ \forall i, \tag{11}$$

$$\sum_{i=1}^{N}\sum_{j=0}^{J_i} a_{i,j} n_{i,j} \leq N_{max}, \tag{12}$$

$$a_{i,j} \in \{0, 1\}, \tag{13}$$

$$\sum_{i=1}^{N} \phi_i \leq N. \tag{14}$$

In (10), J_i is the maximum allowable CQI value for user i and $r_{i,j}$ denotes the achievable bit rate for user i and CQI value j [22]; the decision variable $a_{i,j}$ is equal to 1 if rate index j is assigned to user i; otherwise, $a_{i,j} = 0$. In (12), N_{max} is the maximum number of multicodes available for HSDPA at the BS and $n_{i,j}$ is the required number of multicodes for user i and CQI value j [22]. Depending on multicode availability, the assigned combination of MCS and the number of multicodes may correspond to a bit rate that is smaller than that permitted by J_i. The constraint in (14) can be obtained by substituting $P_i = \phi_i P_T / N$ into (2). The objective in the optimization problem **P1** is to choose $\mathbf{A} = \{a_{i,j}\}$ and $\underline{\phi} = \{\phi_i\}$ at each TTI so as to maximize the sum bit rate for all users, subject to (7)-(9) and (11)-(14).

4. Linearization

It is important to note that the quantity J_i, which appears in the upper summation index in (10), is itself a function of the decision variable ϕ_i. On the other hand, ϕ_i is related to J_i via a non-linear relationship (9). Thus, the problem **P1** is not a standard linear integer programming problem. As such a problem is highly non-linear, a global optimal solution is very difficult to obtain. To solve problem **P1** with linear integer programming methods, the problem needs to be appropriately transformed into a linear problem by introducing additional auxiliary variables. The first step is to re-formulate the model as follows:

$$\mathbf{P1}' : \qquad \max_{\mathbf{A},\underline{\phi}} \sum_{i=1}^{N} \sum_{j=0}^{q_{i,max}} b_{i,j} a_{i,j} r_{i,j} \tag{15}$$

subject to

$$\sum_{j=0}^{q_{i,max}} b_{i,j} a_{i,j} = 1, \ \forall i, \tag{16}$$

$$\sum_{i=1}^{N} \sum_{j=0}^{q_{i,max}} b_{i,j} a_{i,j} n_{i,j} \leq N_{max}, \tag{17}$$

together with (7)-(9),(13)-(14), where the new variable $b_{i,j}$

$$b_{i,j} = \begin{cases} 0 \, , j > J_i \\ 1 \, , j \leq J_i \end{cases} \tag{18}$$

is introduced to limit the rate index j to no higher than J_i.

After the above re-formulation, Problem $\mathbf{P1}'$ is still non-linear due to terms involving the product $a_{i,j} b_{i,j}$, and the presence of the floor function $\lfloor . \rfloor$ and the logarithm in (7). Subsequent linearlization of Problem $\mathbf{P1}'$, involves introducing a new decision variable $m_{i,j} = a_{i,j} b_{i,j}$ and re-writing the problem as follows:

$$\mathbf{P1}'' : \qquad \max_{\mathbf{A},\mathbf{B},\underline{\phi}} \sum_{i=1}^{N} \sum_{j=0}^{q_{i,max}} m_{i,j} r_{i,j} \tag{19}$$

subject to

$$\sum_{j=0}^{q_{i,max}} m_{i,j} = 1, \ \forall i, \tag{20}$$

$$\sum_{i=1}^{N} \sum_{j=0}^{q_{i,max}} m_{i,j} n_{i,j} = \leq N_{max}, \tag{21}$$

$$m_{i,j} \leq b_{i,j}, \ \forall \, i, j \tag{22}$$

$$m_{i,j} \leq a_{i,j} M, \ \forall \, i, j \tag{23}$$

$$m_{i,j} \geq b_{i,j} - (1 - a_{i,j}) M, \ \forall \, i, j \tag{24}$$

$$e_{i,j} - \phi_i \leq (1 - \omega_{i,j}) M, \ \forall \, i, j, \tag{25}$$

$$\phi_i - e_{i,j+1} \leq (1 - \omega_{i,j}) M, \ \forall \, i, j, \tag{26}$$

$$J_i = \sum_{j=0}^{q_{i,max}} j \omega_{i,j}, \forall \, i, \tag{27}$$

$$j - J_i \leq (1 - b_{i,j}) M, \ \forall \, i, j \tag{28}$$

$$J_i + 1 - j \leq b_{i,j} M, \ \forall \, i, j \tag{29}$$

$$\sum_{j=0}^{q_{i,max}} \omega_{i,j} = 1 \tag{30}$$

$$b_{i,j}, a_{i,j}, m_{i,j}, \omega_{i,j} \in \{0, 1\} \tag{31}$$

together with (14), where

$$e_{i,j} = \begin{cases} -M & \text{if } j = 0 \\ 10^{\left(\frac{j - (c_{i,1} \tilde{\gamma}_i^{\dagger} + c_{i,2})}{10 c_{i,1}} \right)} & \text{if } 1 \leq j \leq q_{i,max} \\ M & \text{if } j = q_{i,max} + 1, \end{cases} \tag{32}$$

and M is a large number. In this new formulation, constraints (22)-(24) model the product $a_{i,j} b_{i,j}$, while (25)-(27) models the floor and the max functions in (7) and the constraint in (9). The constraints (28)-(30) are used to linearize the expression defined in (18). Note that the introduction of auxiliary variables increases the size of the model, and, thereby, increases the complexity of the problem. The linearized Problem **P1**$''$ was solved using a commercial optimization software package implementing the branch-and-bound method [19]. An alternative method to solve Problem **P1** is to use meta-heuristics and is discussed in Section 5.

5. Simulated annealing

Meta-heuristic approaches attract much attention with their success in solving hard combinatorial and optimization problems. One of these successful approaches is Simulated annealing (SA) [24], which is proven to be a powerful meta-heuristic approach used for optimization in many combinatorial problems. In SA, a probabilistic decision making process is typically involved, in which a control parameter, often known as temperature τ, is used to control the probability of accepting a poorer solution in the neighborhood of the current

solution. The idea is to provide the possibility of reaching a better solution by diversifying the search within search space, redirecting the search in a new neighborhood when the chance of discovering a better solution within the old neighborhood is not high. The algorithm explores the solution space through a simulated cooling process from a given initial (hot) temperature to a final (frozen) temperature. At a higher temperature, the probability of selecting the poorer solution is higher, and thereby allowing the search to be more extensive within the solution space. However, as the temperature decreases, the search becomes more confined in the region near the desirable solution within the solution space, and thereby providing a refinement to the existing solution. Essentially, the search is conducted through two nested loops; the outer one decreases the temperature using a particular cooling schedule, while the inner one repeats the search at the same temperature. Within the inner loop, a sequence of solutions are obtained by manipulating the current solution. Each solution is the result of one iteration.

Let $\mathbf{x}_n$ be the solution at iteration n, and $\mathbf{x}_n{}' = N(\mathbf{x}_n)$, where $N(\mathbf{x}_n)$ is some neighbor function of $\mathbf{x}_n$. The next solution in the search process is a probabilistic function of $\mathbf{x}_n$, and is given by [24]

$$\mathbf{x}_{n+1} = \begin{cases} \mathbf{x}'_n, \text{ if } s(\mathbf{x}'_n) > s(\mathbf{x}_n) \\ \mathbf{x}'_n, \text{ if } r < e^{-\Delta s/\tau_k} \\ \mathbf{x}_n, \text{ otherwise} \end{cases}, \tag{33}$$

where τ_k corresponds to the k^{th} temperature level, $s(\cdot)$ is the objective function to be maximized, $\Delta s = s(\mathbf{x}_n) - s(\mathbf{x}'_n)$, and r is the outcome of a random variable which is uniformly distributed in [0,1]. This method for choosing a new solution is commonly referred to as the Metropolis rule [24]. The motivation is to diversify the search process, thereby reducing the possibility of locally optimal solutions. As the temperature decreases, so does the probability of accepting a worse solution.

SA has become a basis which inspires different algorithmic variations for solving a large variety of optimization problems. *Evolutionary* Simulated Annealing (ESA) is one of these recently developed population-based SA algorithm enhanced with evolutionary operators [3]. Instead of manipulating a single solution, ESA makes use of a population of solutions in order to combine the advantages of both SA and population-based approaches. A single instance of SA is devised to act as an evolutionary operator and is invoked successively starting at a fixed initial temperature each time. Each invocation of SA is commonly referred to as a *generation*. ESA evolves the population of solutions with the SA operator alongside the selection and replacement operators *generation-by-generation*. The idea is to decrease the temperature during each SA operation and raise it back to the fixed initial value whenever the SA operator is invoked. These artificially induced fluctuations in temperature allows the solution space to be explored more thoroughly, ad thereby reducing the possibility of being trapped in local optima. Recently, a comprehensive study on implementing ESA for facility location problems has been reported in [26].

In this chapter, we propose to use the ESA algorithm to solve the multiuser scheduling problem as outlined in (7)-(14) due to its ability to cope with the highly non-linear nature of the problem. Our ESA implementation consists of two components - an initial solution, and an SA operator. The SA operator is invoked once per generation for G generations. For each SA operation, a search is conducted over N_τ different temperatures. Starting with an initial temperature τ_{hot}, the k-th temperature level is given by

$$\tau_k = \tau_{hot}\theta^{k-1}, k = 1, 2, \ldots, N_\tau, \tag{34}$$

where $\theta \in (0,1)$ is the cooling coefficient. The N_τ-th temperature level is also referred to as the frozen temperature τ_{frozen}. Note that the combined use of (33) and (34) corresponds to a variant of simulated annealing known as the *Simulated Quenching* (SQ) [9].

In the proposed scheme, each temperature level is associated with U iterations. For each iteration, a user i is randomly selected, resulting in a solution to the problem of the form

$$\mathbf{x}_n = (\phi_1, \ldots, \phi_i, \ldots, \phi_N, \mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_N), \tag{35}$$

at iteration $n \in \{1, \ldots, U\}$, where $\mathbf{a}_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,J_i})$.

By applying function $f(.)$ to the current value of ϕ_i, new solutions can be generated. Each new value, ϕ_i', is obtained by applying the function $f(.)$ successively until ϕ_i' satisfies (9) and (14). Once a suitable ϕ_i' is available, the corresponding J_i is then obtained using (7). Subsequently, the value of a_{i,J_i} is set to 1 while the remaining elements are set to 0, i.e. $\mathbf{a}_i = \{\overbrace{0, 0, ..., 1}^{J_i+1}\}$. If constraint (12) is violated, $\mathbf{a}_i$ is cyclically shifted to the left by one position, i.e. $\mathbf{a}_i = \{\overbrace{0, 0, ..., 1, 0}^{J_i+1}\}$. This process is repeated until (12) is satisfied, resulting in an updated vector $\mathbf{a}_i'$. Subsequently, the new solution is given by

$$\mathbf{x}_n' = (\phi_1, \ldots, \phi_i', \ldots, \phi_N, \mathbf{a}_1, \ldots, \mathbf{a}_i', \ldots \mathbf{a}_N). \tag{36}$$

The computational complexity of ESA in terms of the number, N, of users at given values of G and N_τ is discussed as follows. The number of iterations, U, of solutions explored at each temperature is chosen to be N. Thus, the time required to check whether a solution $\mathbf{x}_n'$ in (36) satisfies constraints (12) and (14) has a complexity of $\mathcal{O}(N)$. Subsequently, due to the fact that the time to verify these constraints grows linearly with N, the complexity of ESA is $\mathcal{O}(N^2)$.

As both SA and ESA are heuristic algorithms, solutions obtained by these methods may not necessarily be optimal, but are generally close to optimal. The "temperature" parameter determines trade-off between the speed of convergence towards the optimal value, and how far the solution is relative to that of the optimal. Generally, a higher temperature allows the optimal value to be approached faster, while a lower temperature provides improved fine-tuning, and thereby improving the solution quality. ESA converges faster than conventional SA since the search is diversified by periodically re-increasing the temperature. As with any heuristic algorithm, a detailed convergence study requires the determination of appropriate parameter values through experimentation.

6. Simulation results

A number of different simulation cases were used to illustrate the effectiveness of ESA for the multiuser HSDPA resource allocation problem. The first case involves $N = 2$ users, and the following values $t_{i,0} = -4.5$, $t_{i,1} = 25.5$, $c_{i,1} = 1$, $d_{i,1} = 4.5$, and $q_{i,max} = 30$ for $i = 1, 2$ are used for the parameters in (3). These values are obtained from [18], assuming that the mobiles are of category 10 (i.e. have a wide CQI range) as defined in [22]. Values for $n_{i,j}$ and $r_{i,j}$ are obtained from [22]. The fading channel following the general Nakagami model [21] is assumed so that $\{\gamma_i, i = 1, 2\}$ in (1) are outcomes of Gamma distributed random variables

$\{\Gamma_i, i = 1, 2\}$ with pdfs given by

$$f_{\Gamma_i}(\gamma) = \begin{cases} \left(\frac{\alpha_i}{\overline{\Gamma}_i}\right)^{\alpha_i} \frac{\gamma^{\alpha_i - 1}}{\Gamma(\alpha_i)} \exp\left(\frac{-\alpha_i \gamma}{\overline{\Gamma}_i}\right) & \gamma \geq 0 \\ 0 & \gamma < 0 \end{cases}, \tag{37}$$

where $\Gamma(.)$ is the Gamma function, α_i is the fading figure, and $\overline{\Gamma}_i$ is the mean of Γ_i. The parameter values are listed in Table 1. Let $\mathbf{\Gamma} = \{\Gamma_1 \quad \Gamma_2\}$, and let the aggregate transport block size (TBS), T, per TTI be

Scenario	User i	$\overline{\Gamma}_i$ (in dB)	α_i
I	1	5	6.5
	2	6	3
II	1	12	6.5
	2	10	3
III	1	17	6.5
	2	16	3

Table 1. List of parameter values used ([14]©IET)

$$T = \sum_{i=1}^{N} \sum_{j=0}^{J_i} a_{i,j} r_{i,j}. \tag{38}$$

The performance improvements of the proposed Joint Global Optimum (JGO) and ESA approaches, as discussed in sections 4 and 5 respectively, are compared to that of a simple greedy (SG) algorithm. The idea behind the SG algorithm is to allocate resources to users in decreasing order of their estimated SINR values, $\hat{\gamma}_i$. In other words, the user with the highest $\hat{\gamma}$ is first allocated as much resources as it can possibly use. Subsequently, remaining resources that can be productively used are then assigned to the user with the next highest $\hat{\gamma}$. This allocation of resources continues until resources are exhausted.

Fig. 7 shows the cumulative distribution function (CDF) of T for the three different scenarios in Table 1. The results in this figure are obtained based on two thousand channel realizations. It can be seen that ESA can achieve a performance that is close to the optimal. A summary of the average performance, i.e. $E_{\mathbf{\Gamma}}[T]$, for all three schemes is presented in Table 2. It can be seen that both JGO and ESA can provide a good throughput improvement over SG.

The second study involves the comparison of performance among the three algorithms by increasing the number of users from 2 to 5. The average computation times per TTI required by JGO, ESA, and SG on a personal computer with an Intel Core™ 2 Duo T5500 processor are plotted as a function of the number of users in Fig. 8. As expected, SG is the fastest and JGO is the slowest. The running for ESA is about 0.1 s per TTI and increases slowly with the number of users. With the use of parallel dedicated processors at the BS, ESA becomes a viable alternative to SG. Fig. 9 shows the average computation time, normalized to the 2-user case, as a function of the number of users; the average is taken over only twenty channel realizations due to the long simulations times needed for JGO. The following parameter values are used: $\alpha_1 = \alpha_2 = \ldots = \alpha_N = 5$ and $\overline{\Gamma}_1 = \overline{\Gamma}_2 = \ldots = \overline{\Gamma}_N = 8.45$ dB. It can be seen clearly that even though JGO provides a globally optimal solution, the complexity increases very rapidly with

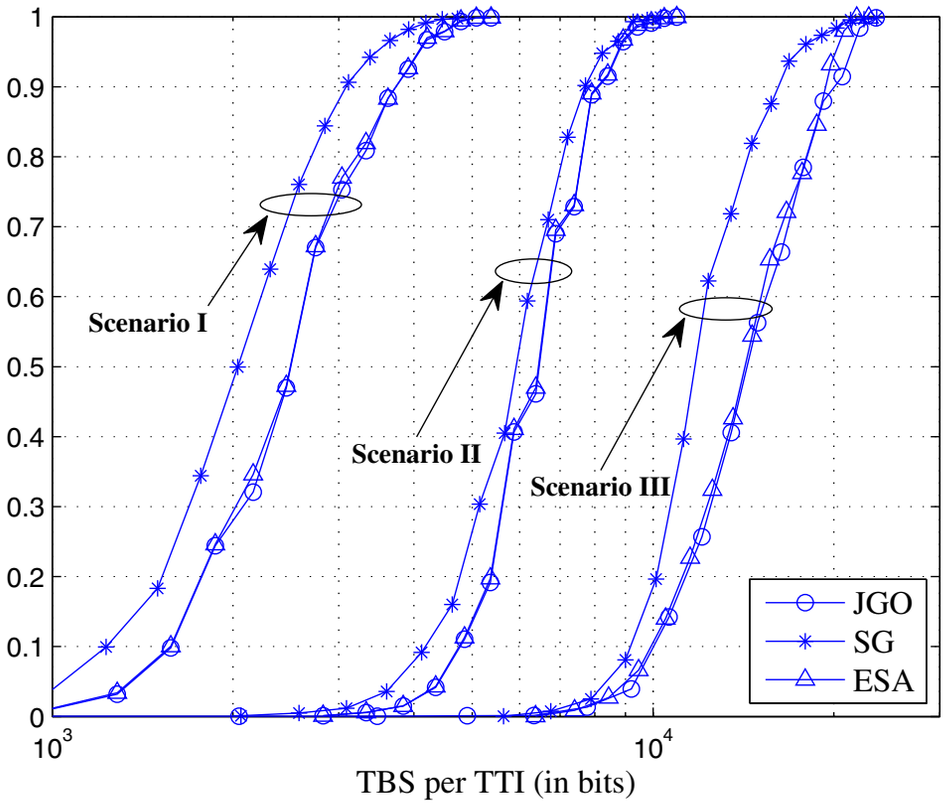

Figure 7. CDF of the aggregate bit rate in TBS per TTI. ([14]©IET)

Scenario	Scheme	ATBS/TTI (kbits)	Gain (%)
I	SG	2.151	0
	ESA	2.588	20.03
	JGO	2.591	20.42
II	SG	6.031	0
	ESA	6.492	7.64
	JGO	6.507	7.88
III	SG	12.472	0
	ESA	14.587	16.95
	JGO	14.944	19.82

Table 2. Average rate (in TBS per TTI) for the three different algorithms under the three scenarios. ([14]©IET)

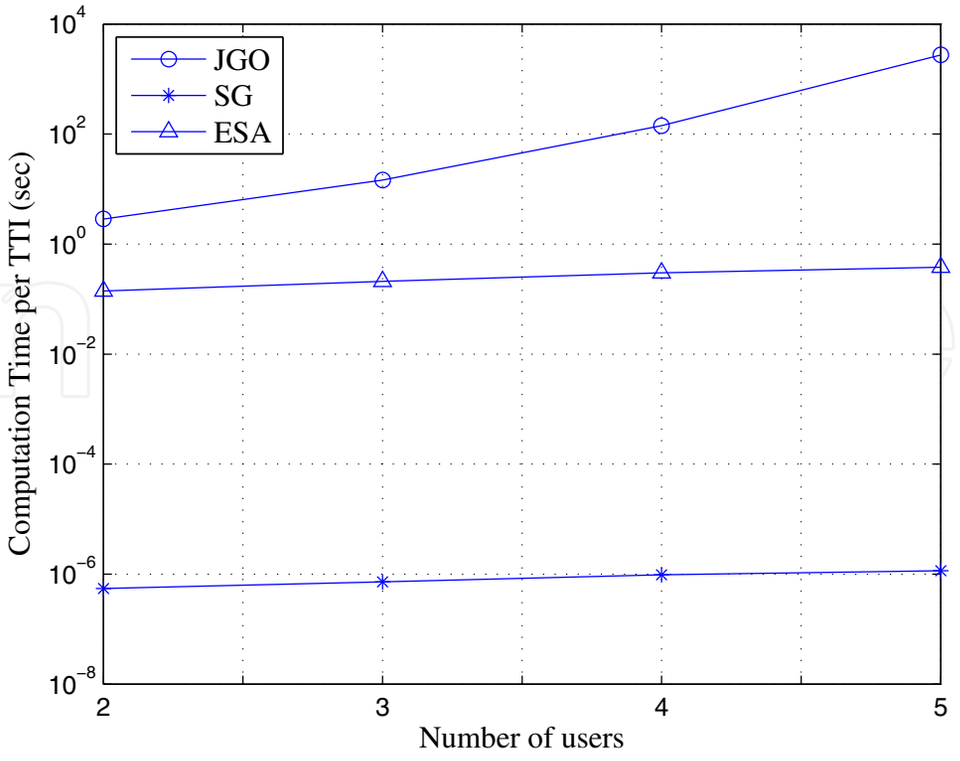

Figure 8. Average computation time per TTI as a function of the number of users. ([14]©IET)

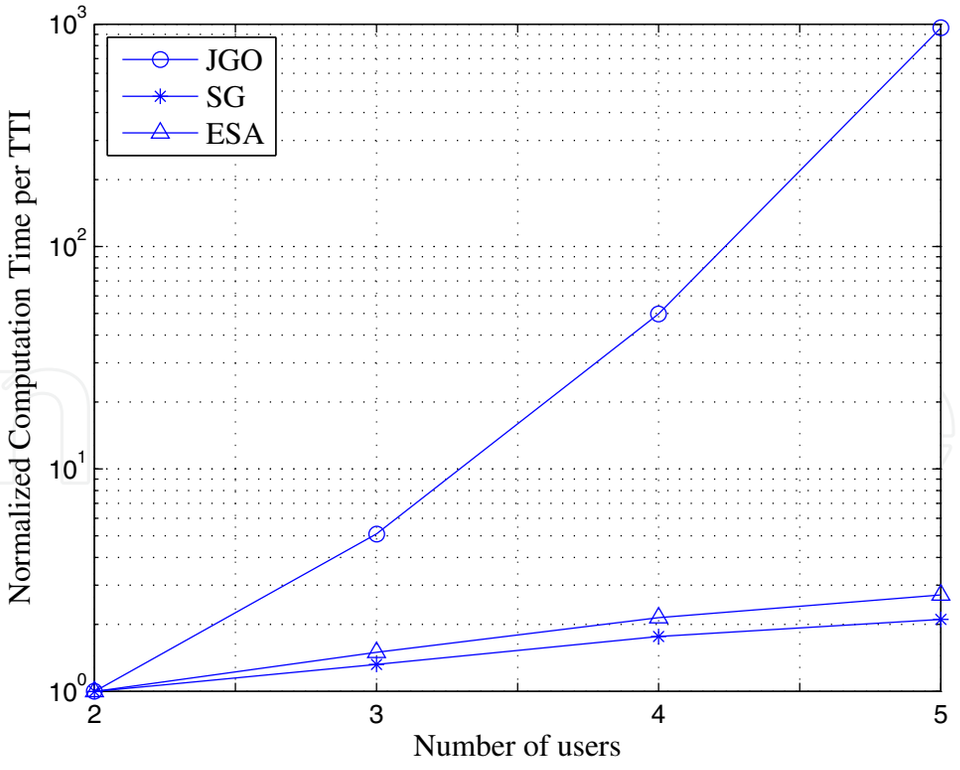

Figure 9. Average computation time per TTI, normalized to the 2-user case, as a function of the number of users. ([14]©IET)

the number of users. On the other hand, ESA offers a very similar performance at a much reduced complexity.

Fig. 10 shows the CDFs of T for SG and ESA for different numbers of users. As in Fig. 9, $\alpha_1 = \alpha_2 = \ldots = \alpha_N = 5$ and $\overline{\Gamma}_1 = \overline{\Gamma}_2 = \ldots = \overline{\Gamma}_N = 8.45$ dB. The CDF curves for JGO are not shown due to the excessively long running times but are expected to be close to those for ESA. Table 3 shows the average aggregate TBS per TTI values for SG and ESA. It can be seen that the performance improvement of ESA over SG increases with the number of users.

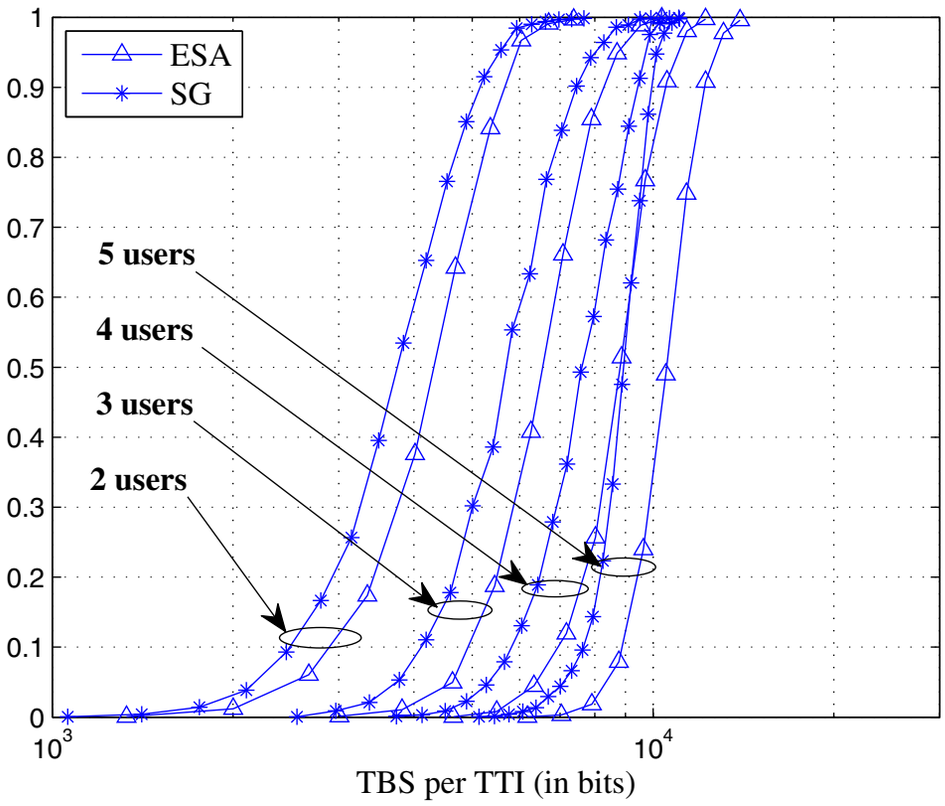

Figure 10. CDF of the aggregate bit rate in TBS per TTI for SG and ESA with different number of users. ([14]©IET)

No. of Users	Scheme	ATBS/TTI (kbits)	Gain (%)
2	ESA	4.324	12.21
	SG	3.853	0
3	ESA	6.676	14.75
	SG	5.818	0
4	ESA	8.865	15.41
	SG	7.681	0
5	ESA	10.651	20.00
	SG	8.877	0

Table 3. Average rates (in TBS per TTI) for the ESA and SG algorithms with different number of users. For each user, $\alpha = 5$ and $\overline{\Gamma} = 8.45$ *dB*. ([14]©IET)

The CDFs of T for ESA, SG, max C/I, and Round Robin (RR) with five users are plotted in Fig. 11. Note that max C/I is a special case of SG in which, for each TTI, resources are allocated only to the user with the best channel condition. On the other hand, RR refers to a base-line scheduling scheme, whereby each user is scheduled one at a time in a round robin fashion. Under such a scheme, the respective channel quality of the users are not used during the scheduling. The corresponding values of $E_{\mathbf{\Gamma}}[T]$ for ESA, SG, max C/I, and RR are 10.6, 8.8, 2.8, and 1.9 bits per TTI respectively. The prominent steps in the CDFs for RR and max C/I are due to the coarse quantization resulting from allocation of resources to only one user in each TTI. It can be seen that RR has the lowest bit rate as it does not take into account user channel conditions.

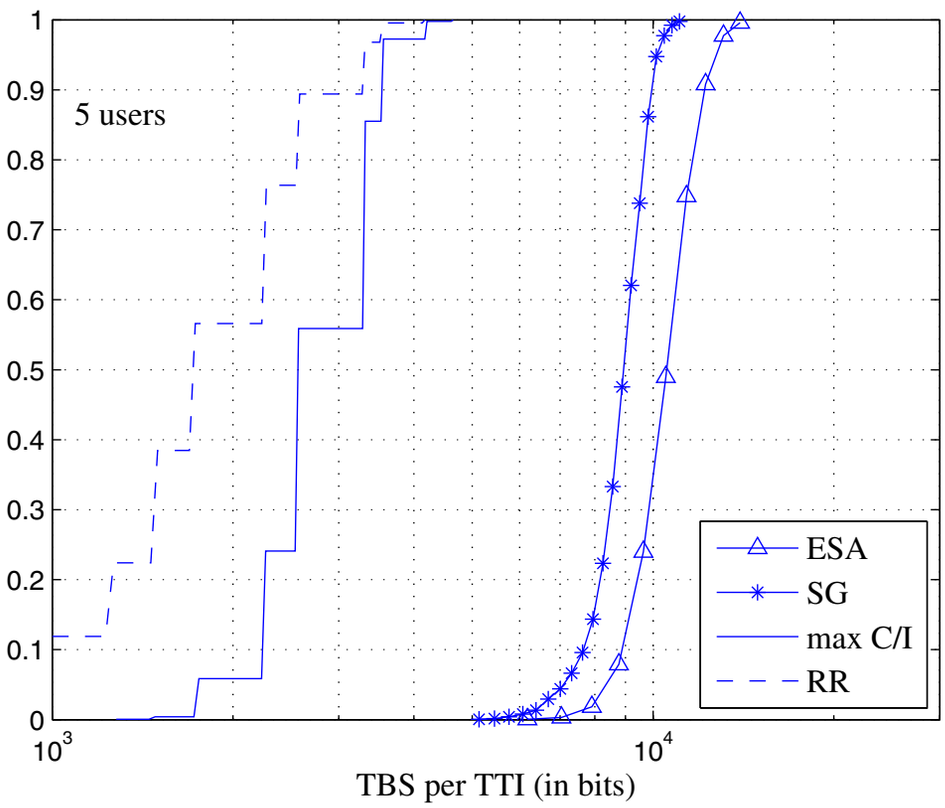

Figure 11. CDF of the aggregate bit rate in TBS per TTI for ESA, max C/I, and Round Robin (RR) with five users. ([14]©IET)

7. Conclusion

In this chapter, the issue of allocating resources to multiple users simultaneously in HSDPA has been examined via a number of optimization methods. The problem formulation based on the channel feedback scheme specified in the WCDMA standard has been presented. Simulation results have shown that both the global optimal and the simulated annealing-based methods can provide a substantial throughput improvement over a more simple greedy algorithm. It has been observed that the method based on simulated annealing can achieve a bit rate that is very close to that of the global optimal, with a much lower

computational complexity. The advantage of simulated annealing-based method relative to the simple greedy method increases with the number of users. In this chapter, it was assumed that the user SINR values, on which the channel quality indicators are based, can be accurately estimated. One research direction would be to study the performance degradation due to noisy SIR estimates and methods in reducing this degradation. Another potential study item is to investigate the benefits and trade-offs of other heuristic optimization methods for the same problem.

Acknowledgement

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant OGP0001731, by the UBC PMC-Sierra Professorship in Networking and Communications and by a Marie Curie International Incoming Fellowship PIIF-GA-2008-221380.

Author details

Raymond Kwan and M. E. Aydin
University of Bedfordshire, United Kingdom

Cyril Leung
University of British Columbia, Canada

8. References

[1] Abedi, S. [2005]. Efficient Radio Resource Management for Wireless Multimedia Communications: A Multidimensional QoS-Based Packet Scheduler, *IEEE Transactions on Wireless Communications* 4(6): 2811 – 2822.

[2] Aniba, G. & Aissa, S. [2005]. Resource Allocation in HSDPA using Best-Users Selection Under Code Constraints, *Proc. of IEEE Vehicular Technology Conference, Spring*, Vol. 1, pp. 319 – 323.

[3] Aydin, M. E. & Fogarty, T. C. [2004]. A Distributed Evolutionary Simulated Annealing Algorithm for Combinatorial Optimisation Problems , *Journal of Heuristics* 10(3): 269 – 292.

[4] Baum, K. L., Kostas, T. A., Sartori, P. J. & Classon, B. K. [2003]. Performance Characteristics of Cellular Systems with Different Link Adaptation Strategies, *IEEE Transactions on Vehicular Technology* 52(6): 1497 – 1507.

[5] Bedekar, A., Borst, S. C., Ramanan, K., Whiting, P. A. & Yeh, E. M. [1999]. Downlink Scheduling in CDMA Data Network, *Proc. of IEEE Global Telecommunications Conference, GLOBECOM '99*, Vol. 5, pp. 2653–2657.

[6] Brouwer, F., de Bruin, I., Silva, J. C., Souto, N., Cercas, F. & Correia, A. [2004]. Usage of Link-Level Performance Indicators for HSDPA Network-Level Simulation in E-UMTS, *Proc. of International Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*, Sydney, Australia.

[7] Dahlman, E., Parkvall, S., Sköld, J. & Beming, P. [2007]. *3G HSPA and LTE for Mobile Broadband*, Academic Press.

[8] Freudenthaler, K., Springer, A. & Wehinger, J. [2007]. Novel SINR-to-CQI Mapping Maximizing the Throughput in HSDPA, *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, Hong Kong, China.

[9] H.A. Oliveira, J., Ingber, L., Petraglia, A., Petraglia, M., Machado, M. & Petraglia, M. [2012]. *Stochastic global optimization and its applications with fuzzy adaptive simulated annealing*, Springer.
[10] Haleem, M. A. & Chandramouli, R. [2005]. Adaptive Downlink Scheduling and Rate Selection: A Cross-Layer Design, *IEEE Journal on Selected Areas in Communications* 23(6): 1287 – 1297.
[11] Holma, H. & Toskala, A. (eds) [2006]. *HSDPA/HSUPA For UMTS High Speed Radio Access for Mobile Communications*, John Wiley & Sons.
[12] Jeon, W. S., Jeong, D. G. & Kim, B. [2004]. Packet Scheduler for Mobile Internet Services using High Speed Downlink Packet Access, *IEEE Transactions on Wireless Communications* 3(5): 1789 – 1801.
[13] Ko, K., Lee, D., Lee, M. & Lee, H. [2006]. Novel SIR to Channel-Quality Indicator (CQI) mapping method for HSDPA System, *Proc. of IEEE Vehicular Technology Conference (VTC)*, Montreal, Canada.
[14] Kwan, R., Aydin, M. E., Leung, C. & Zhang, J. [2009]. Multiuser Scheduling in High Speed Downlink Packet Access, *IET Communications* 3(8): 1363 – 1370.
[15] Kwan, R. & Leung, C. [2007]. Downlink Scheduling Schemes for CDMA Networks with Adaptive Modulation and Coding and Multicodes, *IEEE Transactions on Wireless Communications* 6(10): 3668 – 3677.
[16] Lee, S. J., Lee, H. W. & Sung, D. K. [1999]. Capacities of Single-Code and Multicode DS-CDMA Systems Accommodating Multiclass Service, *IEEE Trans. on Vehicular Technology* 48(2): 376 –384.
[17] Liu, Q., Zhou, S. & Giannakis, G. B. [2005]. Cross-Layer Scheduling with Prescribed QoS Guarantees in Adaptive Wireless Networks, *IEEE Journal on Selected Areas in Communications* 23(5): 1056 – 1066.
[18] Motorola & Nokia [2002]. Revised CQI Proposal, *Technical Report R1-02-0675*, 3GPP RAN WG1.
[19] Rardin, R. [1998]. *Optimization in Operations Research*, Prentice Hall, Upper Saddle River, NJ.
[20] Rhee, J.-H., Holtzman, J. M. & Kim, D. K. [2004]. Performance Analysis of the Adaptive EXP/PF Channel Scheduler in an AMC/TDM System, *IEEE Communications Letters* 8(8): 497 – 499.
[21] Simon, M. K. & Alouini, M.-S. [2000]. *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*, John Wiley & Sons.
[22] *Universal Mobile Telecommunications Systems (UMTS); Physical Layer Procedures (FDD)* [2007]. *Technical Specification 3GPP TS25.214*, 3rd Generation Partnership Project.
[23] *Universal Mobile Telecommunications Systems (UMTS); UE Radio Access Capabilities* [2007]. *Technical Specification 3GPP TS25.306*, 3rd Generation Partnership Project.
[24] Van Laarhoven, P. J. M. & Aarts, E. H. [1987]. *Simulated Annealing: Theory and Applications*, Kluwer Academic Publishers, New York.
[25] Yang, J., Khandani, A. K. & Tin, N. [2005]. Statistical Decision Making in Adaptive Modulation and Coding for 3G Wireless System, *IEEE Transactions on Vehicular Technology* 54(6): 2006 – 2073.
[26] Yigit, V., Aydin, M. E. & Turkbey, O. [2006]. Solving large-scale uncapacitated facility location problems with evolutionary simulated annealing , *International Journal of Production Research* 44(22): 4773 – 4791.
[27] Zerlin, B., Ivrlac, M., Utschick, W., Nossek, J., Viering, I. & Klein, A. [2005]. Joint Optimization of Radio Parameters in HSDPA, *Proc. of IEEE Vehicular Technology Conference, Spring*, Vol. 1, pp. 295 – 299.