Downlink Scheduling and Resource Allocation for 5G MIMO Multicarrier Systems

Ankur Vora and Kyoung-Don Kang
State University of New York at Binghamton, NY, USA.
{avora4, kang}@binghamton.edu

Abstract—Emerging 5G wireless communication technology is envisioned to significantly enhance the performance. In this paper, we propose a new algorithm for effective cross-layer downlink scheduling and resource allocation (SRA) considering the channel and queue state, while supporting fairness. We also integrate our cross-layer SRA scheme with filter-bank multicarrier/offset quadrature amplitude modulation (FBMC/OQAM) to leverage the higher spectral efficiency. Our performance evaluation results show that our SRA method outperforms a novel SRA algorithm [1] by up to approximately 60%, 2.6%, and 1.6% in terms of goodput, goodput fairness, and delay fairness, respectively.

I. INTRODUCTION

The CISCO predicts that the global mobile data traffic is expected to grow by eight fold from 2015 to 2020. For example, it increased from 4.4 exabytes per month at the end of 2015 to 7.2 exabytes per month at the end of 2016 [2]. The upcoming 5G technology is envisioned to revolutionize mobile services and many industries, e.g., telecommunication, automobile, health care, and entertainment, by significantly enhancing the performance, e.g., throughput and latency. This motivates new research to investigate 5G new radio technology for high performance and spectrum efficiency at the data link control (DLC) and physical (PHY) layers. In 5G, a scheduling and resource allocation (SRA) unit should effectively allocate resources of the base station (BS), e.g., time slots and frequency sub-bands (subcarriers). The importance of SRA is increasing rapidly as the demand and expectation for performance rise sharply. However, effective SRA is challenging, since many SRA problems are NP-hard [3], [4].

Feminias et al. [1] propose a novel cross-layer SRA framework that incorporates the downlink SRA across the DLC and PHY layers by extending their previous work [5]. They aim to optimize the utility, i.e., the weighted goodput, via effective cross-layer SRA. They provide a cross-layer optimization over the PHY and DLC layers to support effective SRA for massive multiple-input and multiple-output (MIMO) systems by taking advantage of higher bandwidth [6] and adaptive modulation and coding schemes (MCSSs) [7] in 5G. Their SRA algorithm, however, is greedy. Thus, it may produce sub-optimal results. To address the problem, in this paper, we propose an optimal SRA algorithm based on dynamic programming and show that its time complexity is polynomial in a practical sense.

For 5G new radio, two dominant waveform technologies are: 1) orthogonal frequency division modulation (OFDM) and 2) filter-bank multicarrier/offset quadrature amplitude modulation (FBMC/OQAM). OFDM is widely deployed in 4G and 802.11. In OFDM, the subcarrier frequencies are chosen to make the subcarriers orthogonal to each other and eliminate crosstalk between the subchannels and intercarrier guard bands [8]. This greatly simplifies the design of the transmitter and the receiver too, since a separate filter is not required for each sub-channel. However, OFDM requires time and frequency synchronization at the receiver, which is fulfilled by adding a cyclic prefix (CP) at the beginning of the OFDM waveform. Although the CP removes the inter-symbol interference and inter-channel interference (ISI/ICI), it considerably increases the spectral inefficiency [8]. To overcome the spectral inefficiency caused by CPs, FBMC/OQAM uses a series of filters with a higher spectral selectivity to minimize crosstalk instead of using CPs [9].

In this paper, we integrate our cross-layer SRA scheme based on dynamic programming with FBMC/OQAM to exploit its higher spectral efficiency than OFDM’s. Although significant work has been done to compare the performance of OFDM and FBMC/OQAM at the PHY layer [10], [11], [12], much less work has been done to analyze their performance at both the PHY and DLC layers, considering cross-layer optimization via SRA [1], [5]. Zhang et al. [13] compare the link level performance of various advanced multicarrier waveforms including OFDM and FBMC using MIMO wireless communication channel models. They found that FBMC with its linear filtering offered significantly lower out-of-band emissions than OFDM did. Xenakis et al. [14] propose a dynamic SRA that allocates a contiguous collection of resources to the pending users in each step to improve the throughput and spectral efficiency. The work of Feminias et al. [1] is closest to our work presented in this paper in that they propose a unifying cross-layer SRA framework that incorporates downlink SRA with OFDM and FBMC/OQAM, respectively.

For performance evaluation, we have undertaken extensive experiments using three well-established multipath fading, 1) the Extended Pedestrian A; 2) the Extended Vehicular A, and 3) the Extended Typical Urban models [15] in Matlab. We have also considered different Doppler frequencies and channel correlations for thorough performance analysis. The results indicate that our approach outperforms the greedy
SRA algorithm [1] by up to 60% in terms of goodput (the total number of bits successfully transmitted per unit time). In terms of goodput, our approach consistently outperforms the greedy algorithm [1] in all experiments even under high channel correlations and Doppler frequencies. Further, in terms of Jain’s goodput and delay fairness index (JFI) [16], our proposed algorithm supports 0.96–1 where 1 indicates perfect fairness. Our goodput and delay JFIs are higher than those of [1] by up to approximately 2.6% and 1.6%, respectively.

The remainder of this paper is organized as follows. Section II formulates the SRA problem. In Section III, our SRA algorithm is described and the time complexity is analyzed. In Section IV, the performance of the proposed SRA algorithm is thoroughly evaluated. Finally, Section V concludes the paper.

II. PROBLEM FORMULATION

A. Downlink Time-Slotted MIMO System

The overall system architecture and problem formulation are inspired by [1]. We assume that the BS has $N_{Tx}$ transmit antennas and $N_{Rx}$ receiving antennas with total transmission power $P_T$. The BS offers service to $N_{MS}$ active mobile stations (MSs) of users. Each MS is equipped with the same number of receive antennas. Transmissions between the BS and an active MS are organized in fixed time-frequency resource allocation units, called resource blocks (RBs), where one RB consists of a certain number of resource elements, each of which takes one time slot and frequency sub-band (subcarrier). In the time axis, each RB holds a fixed time slot, $T_{PHY}^u$, where $PHY$ represents either OFDM or FBMC/OQAM symbols. Each of these slots consists of a fixed number of symbol $N_{sym}$ of OFDM or FBMC symbols with duration $T_p$, and sub-carrier bandwidth $\Delta f = \frac{1}{T_p}$.

Since OFDM uses the CP to cancel out ISI, there are $N_{sym}^long$ OFDM symbols prefixed with a long CP of duration $T_{CP}^long$, and $N_{sym}^short$ symbols prefixed with a short CP of duration $T_{CP}^short$. Thus, for OFDM, the fixed time slot size is:

$$T_{OFDM}^u = N_{sym}^long \times T_p + N_{sym}^long \times T_{CP}^long + N_{sym}^short \times T_{CP}^short$$

However, for FBMC/OQAM, it is:

$$T_{FBMC}^u = N_{sym}^long \times T_p$$

since FBMC/OQAM uses no CP. Hence, FBMC/OQAM can enhance the spectral efficiency and improve the network performance. We assume that the SRA process happens at the beginning of a transmission time interval (TTI) between two consecutive time slots, similar to [1].

B. SRA Problem Formulation

Let us assume that the BS has $Q_u(t)$ bits in the queue for MS $u$ at the beginning of TTI $t$. Also, we assume that the SRA scheme of the BS allocates $L_u$ spatial streams to user $u$ and a transmission capacity, $r_{u,l}(t, N_{Bu})$ bits, to spatial stream $l \in L_u = \{1, ..., L_u\}$ where $N_{Bu}$ is the set of RBs to user $u$. At the end of the TTI, the total queue length $Q_u(t)$ of user $u$ at the BS is:

$$Q_u(t + 1) = Q_u(t) + A_u(t) - S_u(t)$$

where $A_u(t)$ and $S_u(t)$ represent the number of the arriving data bits to transmit to user $u$ during TTI $t$ and that successfully transmitted to the user, respectively.

When there are $N_{MS}$ mobile stations (users), at the beginning of TTI $t$, the SRA unit of the BS is required to derive the RB allocation set $N_B = \{N_{B1}, ..., N_{BN_{MS}}\}$, where $N_{Bu}$ is the number of RBs allocated to MS $u$, and the MCS allocation set $\mu = \{\mu_1, ..., \mu_{N_{MS}}\}$, where $\mu_u = \{\mu_{u,1}, ..., \mu_{u,L_u}\}$ represents a set of MCSs assigned to each spatial stream $l$ of MS $u$, to effectively allocate RBs and MCSs to MSs, respectively.\footnote{For simplicity, $t$ is dropped in our problem formulation presented hereafter.}

We formulate the SRA optimization problem to maximize the total utility $V$, i.e., the total weighted goodput, as follows:

$$V = \max_{N_{B}, \mu} \sum_{u=1}^{N_{MS}} \sum_{l=1}^{L_u} w_u r_{u,l}(N_{Bu}) \left[1 - BLER_{u,l}(\mu_u) \right] (N_{Bu})$$

subject to $N_{Bk} \cap N_{Bj} = \emptyset \ \forall k \neq j$

$$\sum_{l=1}^{L_u} r_{u,l}(N_{Bu}) \leq Q_u \ \forall (u, l)$$

$$BLER_{u,l}(\mu_u) (N_{Bu}) \leq BLER_0 \ \forall (u, l)$$

where $w_u$ is the weight of user $u$ and $BLER_{u,l}(\mu_u)$ is the block error rate of user $u$’s spatial stream $l$ to which the MCS $\mu_{u,l}$ is assigned.

The SRA unit is required to maximize $V$ subject to three constraints:

- An RB should be exclusively assigned to one user.
- SRA should allocate user $u$ no more transmission capacity than the number of bits in its queue (frugality constraint).
- The average BLER of $u$ does not exceed the upper bound, $BLER_0$, for a minimum quality guarantee.

The proposed methodology in [1] uses an adaptive MCS [5] for each user $u$ considering its channel state information (CSI) and queue state information (QSI). Moreover, the authors propose a suboptimal greedy algorithm to efficiently allocate RBs to users. Essentially, it allocates the first RB to the user with the largest utility increase. It repeats this greedy approach until the set of non-allocated RBs becomes empty or there is no more active user to allocate RBs to. In this paper, we propose a cost-effective algorithm based on dynamic programming to optimally allocate RBs to active users, while applying the same adaptive MCS scheme used in [1].

III. SCHEDULING AND RESOURCE ALLOCATION

In general, a greedy algorithm makes a choice deemed best according to a certain criterion regardless of the choices it
made before or will make in the future. Although it may find an effective solution in a reasonable amount of time, it also results in a suboptimal solution when a series of local decisions fails to lead to a global optimum. The basic idea for dynamic programming is to solve subproblems optimally only once and store the results and look up the stored optimal solutions to the subproblems instead of recomputing them to compute the optimal solution for a given problem efficiently [17].

In this paper, we design a new SRA algorithm by adapting the dynamic programming method for the 0/1 knapsack problem to optimize the utility of SRA for allocating RBs to active users with non-empty queues. It is challenging to design a cost-effective algorithm for RB allocation, since the 0/1 knapsack problem is NP-complete in general. In this section, we design an efficient dynamic programming algorithm to maximize the utility defined in Section II-B and analyze the time complexity.

To this end, we first design the recursive structure of utility function $V$ to optimally allocate free RBs, $N_{free}$, to an arbitrary user $u$ where $Q_u > 0$ as follows:

$$V[u,k] = \begin{cases} \max(V[u-1,k], V[u-1,k-m[u]] + w[u] \times r_{u,l}(m[u])) & \text{if } m[u] \leq k; \\
V[u-1,k] & \text{otherwise.} \end{cases}$$

where $k$ is the number of the available RBs, $m[u]$ is the number of RBs required by MS $u$, $w[u]$ is the weight of MS $u$, and $r_{u,l}(m[u])$ is the transmission capacity provided to MS $u$ by $m[u]$ RBs.

If $m[u] \leq k$, MS $u$ can be assigned the required number of RBs. In this case, our dynamic programming method for SRA optimizes the total utility by assigning $m[u]$ RBs to MS $u$, if $V[u-1,k-m[u]] + w[u] \times r_{u,l}(m[u]) > V[u-1,k]$ and updates the total utility as $V[u-1,k-m[u]] + w[u] \times r_{u,l}(m[u])$. Otherwise, it does not assign the RBs to MS $u$ and maintains the utility as $V[u-1,k]$.

If $m[u] > k$, however, the RBs required by MS $u$ is unavailable; therefore, our approach cannot meet the requirement of MS $u$. As a result, the utility remains as $V[u-1,k]$.

We design the dynamic programming algorithm for SRA based on these recursive properties as shown in Algorithm 1.

The time complexity of Algorithm 1 is $O(N_{MS} \times N_{all})$ where $N_{all}$ is the total number of the RBs in a wireless communication frame at the BS. In general, when the number of the items to consider is $n$ and the total capacity of the knapsack is $W$, the time complexity of the dynamic programming algorithm for the 0/1 knapsack problem is $O(nW)$. $O(nW)$ has pseudo-polynomial complexity, since $W$ may not be a polynomial function of $n$ but could be arbitrarily large (e.g., exponential with respect to $n$). In practice, however, $N_{all}$ during a wireless communication frame is a fixed constant known a priori. For example, in LTE, one frame is 10ms and $N_{all}$ is 6 and 100 when the channel bandwidth is 1.4MHz and 200MHz, respectively.

**Algorithm 1: SRA via Dynamic Programming**

**Data:** Set of active users $U := \{u|Q_u \neq \emptyset\}$

$N_{MS}$: Number of MSs ($=|U|$)

$N_{free}$: Number of free RBs

$m[1..N_{MS}]$: Array of RBs required by MSs

$w[1..N_{MS}]$: Array of the MS weights

**Result:** SRA via Dynamic Programming

```plaintext
knapsack(U, N_{free}, m, w) {
  for j = 0; j \leq N_{free}; j++ do
    V[0, j] = 0; /* no MS */
  end
  for (u = 1; u \leq N_{MS}; u++) do
    V[u,0] = 0; /* no RB */
    for (k = 1; k < N_{free}; k++) do
      if (m[u] \leq k) then
        V[u, k] = max(V[u-1, k], V[u-1, k-m[u]] + m[u] \times w[u] \times r_{u,l}(m[u]));
      else
        V[u, k] = V[u-1, k];
      end
    end
  end
}
```

5G standardization is still on-going. As long as $N_{all}$ remains to be a constant or is a polynomial function of $N_{MS}$ in practical implementations of the upcoming 5G standard, the time complexity of our algorithm is polynomial.

**IV. PERFORMANCE EVALUATION**

As the 5G standardization is still underway, no standard 5G evaluation procedure is available yet. Hence, we follow the latest evaluation procedure recommended for LTE [15] and used in recent 5G research, including [1], [5]. We compare the performance of our approach to that of the novel greedy method presented in [1] in terms of goodput, delay and fairness in a single cell using the LTE toolbox in Matlab. In the rest of this section, the greedy method [1] and our dynamic programming approach are called Greedy and Dynamic for short. Specifically, for performance comparisons, we consider four approaches:

- **Greedy-OFDM:** In this baseline, Greedy [1] and OFDM are used for cross-layer SRA and PHY waveform.
- **Greedy-FBMC:** Greedy and FBMC are employed in this baseline. Greedy-OFDM and Greedy-FBMC are proposed in [1]. In this paper, they are used as the state-of-the-art baselines.
- **Dynamic-OFDM:** In this baseline, Dynamic and OFDM are employed together.
- **Dynamic-FBMC:** This is our proposed approach that integrates Dynamic and FBMC to create synergy via optimal SRA and spectral efficiency supported by Dynamic and FBMC, respectively.

[Each RB consists of 84 resource elements when each RB consists of 7 symbols (time slots) in the time axis and 12 subcarriers (15kHz each) in the frequency axis [18].]
For performance evaluation, we use three multipath fading channel models: 1) the Extended Pedestrian A (EPA), 2) the Extended Vehicular A (EVA), and 3) the Extended Typical Urban (ETU) models. The channel models, evaluation strategy, and chosen parameters summarized in Table I are as per [15]. As specified in the table, we consider low, medium, and high correlation at the BS and MS represented by $\alpha$ and $\beta$, respectively. Also, for the EVA and ETU channels, we consider two different Doppler frequencies according to [15]. In this paper, we present the average of 1000 Monte Carlo simulations. In the performance result graphs, the dashed and solid lines represent Greedy and Dynamic, respectively.

**Goodput for Different Channel Models.** Figure 1 shows the goodput for the increasing number of users (MSs) in a cell. In Figure 1a, Dynamic-FBMC outperforms Dynamic-OFDM, Greedy-FBMC and Greedy-OFDM by up to approximately 6Mbps ($\approx 11.7\%$), 13Mbps ($\approx 29.5\%$) and 17Mbps ($\approx 42.5\%$) for three users, respectively. For six or more users, the difference between Dynamic and Greedy reduces to approximately 2Mbps ($\approx 2.7\%$), but the difference between Dynamic-FBMC and Dynamic-OFDM ranges between approximately 7Mbps–8Mbps ($\approx 10.6\%–11.2\%$). In all the performance evaluation results, Dynamic-FBMC achieves the highest goodput. In the remainder of this paper, we mainly compare the performance of Dynamic-FBMC and Greedy-FBMC that support higher goodput than Dynamic-OFDM and Greedy-OFDM do, respectively.

In Figure 1b, Dynamic-FBMC outperforms Greedy-FBMC by up to 16Mbps ($\approx 44.4\%$) and 15Mbps ($\approx 44.1\%$) for the 5Hz and 50Hz Doppler frequency, when four users are considered respectively. For six or more users, Dynamic-FBMC enhances the goodput by approximately 1–2Mbps compared to Greedy-FBMC.

In Figure 1c, Dynamic-FBMC outperforms Greedy-FBMC by up to 13Mbps ($\approx 31.7\%$) and 12.5Mbps ($\approx 32.1\%$) for the 70Hz and 50Hz Doppler frequency, when four users are considered respectively. Dynamic-FBMC enhances goodput by approximately 1–2Mbps compared to Greedy-FBMC for six or more users.

Further, in Figure 1, we observe that as the number of users and contention for RBs increase, the spectrum efficiency provided by FBMC generally becomes more important in terms of enhancing goodput. Thus, it is required for the BS to efficiently utilize both the RBs and PHY waveform to improve goodput across different situations as done in Dynamic-FBMC.

<table>
<thead>
<tr>
<th>Channel Model</th>
<th>Doppler Frequency (Hz)</th>
<th>Correlation Profiles</th>
<th>Greedy-FBMC</th>
<th>Greedy-OFDM</th>
<th>Dynamic-FBMC</th>
<th>Dynamic-OFDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPA</td>
<td>5</td>
<td>Low</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td>EVA</td>
<td>5, 50</td>
<td>Low</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td>ETU</td>
<td>70, 300</td>
<td>Low</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>Alpha</td>
<td>Beta</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
</tbody>
</table>

**TABLE I: Channel Model Parameters [15]**

![Figure 1](image_url)
Impact of Correlations on Goodput. Incoming signals at the antenna array can be correlated, potentially reducing the overall network performance. Hence, the correlation coefficient for any system is important for performance evaluation. Thus, we use the standard correlation parameters in Table I accordingly [15]. Figure 2 shows the impact of correlation on goodput for different numbers of users in a cell. In Figure 2a, all the tested approaches show lower goodput for higher correlation. Also, Dynamic-FBMC consistently outperforms Greedy-FBMC for the increasing number of users especially for the medium and high correlation cases. For example, Dynamic-FBMC enhances goodput compared to Greedy-FBMC by approximately 1-10Mbps under high correlation for 1 - 16 users in Figure 2a. The biggest goodput enhancement is more than 60% for the two user case (the black solid and dotted curves in Figure 2a). It improves goodput even more significantly in a consistent manner as shown in Figures 2b and 2c. (The biggest improvement is observed for four users in Figures 2b and 2c.) These results are desirable in that Dynamic-FBMC achieves consistently higher goodput than Greedy-FBMC and all the other approaches do in noisy environments that impose significant challenges for wireless communication. Overall, we observe that, among the tested approaches, Dynamic-FBMC provides the highest performance in terms of goodput in all the experiments.

Jain’s Fairness Index (JFI). SRA should also be fair in terms of resource allocation. In this paper, we use JFI matrices [16]. The JFI ranges between 0 and 1 where 1 indicates perfect fairness. Figures 3 and 4 show that the JFIs provided by Greedy and Dynamic range between 0.96–1. Moreover, Dynamic consistently achieves higher goodput and delay fairness than Greedy does when they use the same PHY waveform as shown in Figures 3 and 4. More specifically, it enhances the goodput and delay JFIs by up to 0.026 and 0.016 in Figures 3c and 4a, respectively. Notably, FBMC considerably enhances goodput as shown in Figures 1 and 2 by trading orthogonality off for more spectral efficiency. However, the JFI loss of Dynamic-FBMC (resp. Greedy-FBMC) compared to Dynamic-OFDM (resp. Greedy-OFDM) is at most 0.027 (Figure 3c) and 0.02 (Figure 4a).

V. CONCLUSIONS

Emerging 5G wireless communication technology is envisioned to significantly enhance the performance. In this paper, we propose a new algorithm for effective cross-layer downlink SRA considering the channel and queue state. Further, we integrate our SRA algorithm with filter-bank multicarrier/offset quadrature amplitude modulation (FBMC/OQAM), which supports more spectral efficiency than OFDM does. In the performance evaluation, our SRA method outperforms the state-of-the-art greedy algorithm by up to approximately 60%, 2.6%, and 1.6% in terms of goodput, goodput fairness, and delay fairness, respectively. In the future, we will also explore other important 5G research
issues, such as index modulation and re-sampling, while further improving our scheduling and resource allocation scheme.

ACKNOWLEDGMENT

This work was supported, in part, by NSF grant CNS-1526932.

REFERENCES


