A Graph Coloring based Inter-Slice Resource Management for 5G Dynamic TDD RANs

Emmanouil Pateromichelakis, Konstantinos Samdanis
Huawei Technologies, German Research Center, Munich Germany

Abstract—**This paper studies the notion of network slicing in the emerging 5G Time Division Duplex (TDD) networks, considering different applications with diverse service requirements. The proposed solution relies on flexible slicing considering the entire spectrum band, wherein service oriented slices can independently adopt and adjust an UL/DL ratio reflecting traffic conditions. To this end, this paper initially describes and analyses the inter-slice resource allocation problem taking into account different TDD patterns, the slice traffic load and the inter-node interference, which originates by multiple and diverse sources. In this context, a graph-based framework is proposed to reduce the complexity of the problem by decoupling it in two sub-problems: the selection of TDD patterns to find which links should be activated at particular time instances and accordingly perform graph-based resource allocation for all slice traffic collectively. Our performance evaluations show significant enhancements of the UL and DL throughput respectively.**

Keywords—5G, TDD, RRM, Network Slicing

I. INTRODUCTION

5G is expected to offer a communication ecosystem that enables a plethora of diverse services in a cost effective manner. By 2020, 5G is anticipated to support 5 billion subscribers globally, with a data traffic of 30.6 exabytes per month, and 1 billion Internet of Things (IoT) devices [1][2]. To analyze 5G services and corresponding requirements, 3GPP has conducted a study in [3] considering enhanced mobile broadband services, critical communications and massive IoT for connecting a high volume of stationary devices. Each category enables services with diverse and often conflicting requirements, in terms of the Uplink (UL) and Downlink (DL) demand, latency, reliability, etc., which are difficult to be fulfilled simultaneously by the current 4G "one fit all" network architecture. Hence, 5G networks require enhanced flexibility to assure efficient resource usage with minimal infrastructure upgrades. Network slicing allows the deployment of multiple logical, self-contained networks, offering 3rd parties and verticals customized services on the top of a shared infrastructure [4][5].

This paper adopts the notion of network slicing in TDD networks, considering multiple services with asymmetric traffic conditions. Network slicing can eliminate the negative performance impact of one type of service to another, due to the isolation property. However, this comes at the cost of loss in multiplexing gain, since slice resources are kept for explicit use [6]. Therefore, the key in balancing isolation and efficient resource management is flexibility, which allows a separate UL/DL ratio, i.e. a different TDD frame, per slice and a dynamic inter-slice resource adjustment, including also the capability of selecting resource blocks for each slice considering the entire spectrum band.

A service-oriented TDD network slicing framework is introduced in [7] allocating a fixed amount of resources per slice, which occupy specific isolated sub-carriers, i.e. realizing a "hard" spectrum slicing. This paper enhances such TDD slicing framework introducing flexibility in slice allocations bringing up the following contributions:
- Study the problem of resource flexibility assuming the allocation of resource blocks per slice from the entire spectrum band, while considering inter-slice interference.
- Formulate the generic problem for multi-slice multi-user and multi-cell UL/DL resource optimization.
- Translate the original problem into two sub-problems (P1 and P2) to reduce complexity. P1 involves link activation selection per time instance, given the slice/traffic requirements and the TDD patterns. P2 considers the link selection and aims to optimize per time-slot the performance by allocating resource blocks to the active links based-on the desired performance.
- Propose solutions for P1 and P2. For P1 it involves a constraint-based greedy algorithm, whereas for P2 the problem is solved by a novel bi-partite graph-coloring scheme, which adapts the frequency partitioning per time slot to avoid interference due to resource conflicts, while maintaining resource utilization efficiency high.
- Provide a case study of a dense Radio Access Network (RAN) with overlapping small cells, for evaluating our solutions elaborating the gains over the state of the art in terms of UL and DL user throughput.

The remainder of the paper is structured as follows. Section II provides the related work and Section III discusses the inter-slice resource management model. Section IV analyzes and translates the problem in two sub-problems and Section V provides the corresponding solutions. Section VI, presents the evaluation study and finally Section VII concludes the paper.

II. RELATED WORK

As bursty data services such as IoT, social media, etc., with asymmetric UL/DL demands are widely adopted, TDD systems gain momentum for 5G networks considering dense or hot spot deployments [8]. 3GPP has specified 7 different frame configurations offering diverse UL/DL ratios [9]. Initial deployments focused on synchronized base stations with the

same UL/DL ratio based-on average traffic demands reducing cross slot interference at the cost of restricting utilization efficiency and performance. To overcome these limitations, a cell specific dynamic frame re-configuration (CSDR) [10] allows each base station to adjust the UL/DL ratio dynamically, reflecting evolving traffic demands, avoiding DL interference by regulating the power of sub-frames. 3GPP studied the adoption of dynamic UL/DL ratio among neighbor cells considering power control for interference management, issuing a specification on enhanced Interference Management and Traffic Adaptation (eIMTA) [11]. Interference mitigation for dynamic TDD can also be realized via base station cooperation that cancels cross-link interference by introducing an aligned reverse frame structure [12].

The notion of dynamic TDD in the context of ultra-dense networks is explored in [13] focusing on an uncoordinated greedy operation, referred to as ``blind'', that eliminates the need for fast inter-cell communication for reducing co-channel interference leveraging the benefits of beam-steering. A game theoretic approach, where each eNB re-configures its TDD frame via self-organizing means in order to minimize the UL/DL delay by estimating the cross-slot interference and transmitting load is studied in [14]. To enhance the system performance at the cell edge, a solution termed as "virtual cells" [15], allows users with overlapping coverage to utilize sub-frames from neighbor eNBs facilitating customized TDD frames that match best application specific UL/DL demands. The notion of efficient user multiplexing with highly diverse service requirements via a flexible TDD frame structure is elaborated in [16] focusing on different transmission time intervals considering also the radio conditions.

Although dynamic TDD offers flexibility, handling services with conflicting performance requirements on a common infrastructure that follows the same TDD frame configuration may not yield the desired service performance, especially when network resources are scarce. To overcome such a limitation there is a need for further customization, which can be achieved via the means of network slicing. A TDD service-oriented network slicing solution is introduced in [7], allowing each slice to employ and adjust independently an UL/DL ratio for the duration of a service request. This paper aims to enhance such TDD slicing solution allowing a higher flexibility in the resource allocation considering the entire spectrum band instead of the a-priori static spectrum allocation. Our solution performs an efficient service-oriented resource management relying on: (i) a constraint-based greedy algorithm that considers the slice requirements and the desired UL/DL ratio, and (ii) a novel bi-partite graph-coloring scheme to avoid inter-slice interference.

III. INTER-SLICE INTERFERENCE MANAGEMENT MODEL

For analyzing the proposed inter-slice interference management approach, a multi-cell OFDMA network is considered that consists of a dense deployment of small cells. Each small cell is served by a single, randomly located, antenna denoted as

small-cell Access Point (s-AP). The entire network comprises of $a \in A$ s-APs, $b \in B$ users and the spectrum, which can be shared among all small cells and divided in $n=1,2,...N$ sub-bands or sub-channels. $S_b, \forall b$, is the set of serving s-APs for each user $b=1,..,B$, based on the received signal power, whereas $I_b \neq S_b, \forall b$, is the set of interferer s-APs for each user b in the DL direction. Let also define as $I_a = \{I_a \subseteq B | b \in I_a \Leftrightarrow a \notin S_b\}, \forall a \in A$ the set of interferer users in the UL direction for each s-AP.

The signal strength in the DL direction from each s-AP a to a user b can be formulated as $\gamma_{a,b,n}^{DL} = P_{a,n} G_{a,b,n} L(d_{a,b}{}^\rho)$, where $P_{a,n}$ is the transmit power of the s-AP at sub-channel n. $G_{a,b,n}$ accounts for the antenna gain and $L(d_{a,b}{}^\rho)$ is the distance dependent channel loss, taking into account the distance $d_{a,b}$ and the path loss exponent ρ of each link among s-AP and UE. For the UL direction the signal strength from each user b to s-AP a can be formulated as $\gamma_{b,a,n}^{UL} = P_{b,n} G_{b,a,n} L(d_{a,b}{}^\rho)$, where $P_{b,n}$ is the transmit power of a user at sub-channel n and $G_{b,a,n}$ accounts for the antenna gain of the corresponding link. In a similar manner, the signal power for cross-links (between s-APs and users respectively) can be defined as $\gamma_{a,a',n}^{c1}, \forall a, a' \in A$ and $\gamma_{b,b',n}^{c2}, \forall b, b' \in B$. Based on these formulations, we derive the SINR for UL and DL as follows:

$$SINR_{b,n}^{DL} = \frac{\sum_{a \in S_b} \gamma_{a,b,n}^{DL}}{\sum_{i \neq a \in I_b} \gamma_{i,b,n}^{DL} + \sum_{j \neq b \in I_a} \gamma_{b,j,n}^{c2} + \sigma^2} \quad (1)$$

$$SINR_{a,n}^{UL} = \frac{\gamma_{b,a,n}^{UL}}{\sum_{i \neq b \in I_a} \gamma_{b,i,n}^{UL} + \sum_{j \neq a \in I_b} \gamma_{a,j,n}^{c1} + \sigma^2}, \forall a \in S_b \quad (2)$$

From Eq. (1) and Eq. (2), the achievable data rate for DL and UL can be re-presented as $R^{DL}{}_{b,n} = BW log_2(1 + SINR_{b,n}^{DL})$ and $R^{UL}{}_{a,n} = BW log_2(1 + SINR_{a,n}^{UL})$ respectively, where BW accounts for the allocated bandwidth. Furthermore, let $\{w^{DL}{}_{b,n}, b \in B, n \in N\}$ be arbitrary user weights taking into account the scheduling policies for DL and let $\{w^{UL}{}_{a,n}, a \in A, n \in N\}$ be arbitrary cell weights taking into account the corresponding policies for the UL to ensure fairness.
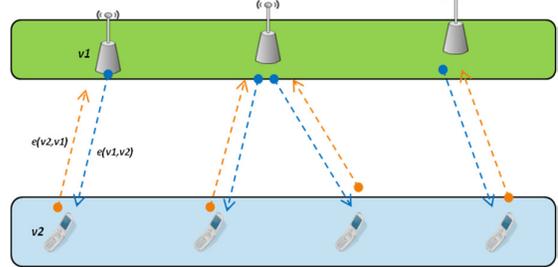

Fig.1: Bi-directional bi-partite graph between s-APs and UEs

Let $G(V, E)$ be a graph consisting of a set of V nodes (s-APs or UEs) and a set of E bi-directional edges. An edge $e \in E$ is a connection between any two nodes $v1, v2 \in V$ each belonging to

a different set (i.e. the set of s-APs and UEs) $V1, V2$: $V1 \cup V2 \subseteq V$, provided that a UE is within the coverage of a s-AP. The presence of an edge $e \in E$ indicates that data can be exchanged between $v1$ and $v2$ considering also the transmission direction (e.g. UL, DL). So, this graph can be seen as a bi-partite graph as shown in Fig. 1. Let the set of all bi-partite sub-graphs belonging to $G(V,E)$, denoted as $\boldsymbol{\Omega}$. Each Ω_i sub-graph represents a combination of link activations (one set of the bi-partite graph is the transmitter nodes and the other set is the receiver nodes). Each of these sub-graphs is associated with a weight factor $w_{\Omega i}$ which represents the fraction of time that this combination of active links lasts. We also define a binary indicator variable $I_{e,\Omega}$ which is 1 if the edge e is included in sub-graph Ω_i and 0 otherwise.

IV. PROBLEM DESCRIPTION: TDD SLICE ALLOCATION

A. Combinatorial Problem Formulation

Provided that the DL/UL configuration patterns are given ($conf_{DL}$, $conf_{UL}=1 - conf_{DL}$) and assuming multiple slices sharing the RAN, the generic combinatorial optimization problem can be formulated as follows:

$$maximize \sum_{t \in T} \{ \sum_{s \in S} conf_{DL}(s) \sum_{n=1}^{N} \sum_{b=1}^{B} w^{DL}{}_{b,n}(s) \cdot R^{DL}{}_{b,n}(s) \tag{3}$$
$$\cdot \left(\sum_{a=1}^{A} v1_{a,b,n,s}(t) \right)$$
$$+ (1 - conf_{DL}(s)) \sum_{n=1}^{N} \sum_{a=1}^{A} w^{UL}{}_{a,n}(s) \cdot R^{UL}{}_{a,n}(s)$$
$$\cdot \left(\sum_{b=1}^{B} v2_{b,a,n,s}(t) \right) \},$$

Subject to:

$$\sum_{n} w^{DL}{}_{b,n}(s) = 1, \forall b \in B, s \in S, \tag{4}$$

$$\sum_{n} w^{UL}{}_{a,n}(s) = 1, \forall a \in A, s \in S \tag{5}$$

$$\sum_{n \in N} R^{DL}{}_{b,n}(s) \geq R_{min,b,s}, \forall b \in B, s \in S, \tag{6}$$

$$\sum_{n \in N} R^{UL}{}_{a,n}(s) \geq R_{min,a,s}, \forall a \in A, s \in S, \tag{7}$$

$$\sum_{s} v1_{a,b,n,s}(t) \neq \sum_{s} v2_{b,a,n,s}(t), \forall n \in N, \forall a, b, s \tag{8}$$

$$v1_{a,b,n,s1}(t) \neq v1_{a,b,n,s2}(t), v2_{b,a,n,s1}(t) \neq v2_{b,a,n,s2}(t) \tag{9}$$
$$\forall s1, s2 \neq s1 \in S, \forall n \in N, \forall a, b, \forall s \in S$$

where $v1$ and $v2$ are binary allocation variables for DL and UL, i.e. if $v1_{a,b,n,s}(t) = 1$, it means that the link (a, b) in DL is allocated to sub-channel n at time-slot t, otherwise this variable is 0 (the same applies for $v2$). Constraints (4) and (5) are the weight constraints for UL and DL ($w^{DL}{}_{b,n}, w^{UL}{}_{a,n}$), which are defined per slice and can be different taking into account slice specific requirements. Further, (6) and (7) are the minimum rate constraints for DL and UL respectively considering the Service Level Agreement (SLA), while (8) provides the half duplex

constraint (i.e. each link can transmit in the UL or DL direction at a time). Finally, (9) shows that in case multiple slices are supported by a link, one sub-channel can be allocated at most to one slice at a time instance.

B. Problem Translation

Since the original problem is an NP-hard non-convex combinatorial, we translate it to a graph based problem in order to lower the complexity and enhance modularity by de-coupling it in two sub-problems (i.e. **P1** and **P2**). **P1** aims to solve the link allocation for UL and DL traffic flows in a slice aware manner with the objective to meet the service requirements for all slices, whereas **P2** controls the allocation of resources (i.e. frequency chunks or sub-bands) for active links to ensure that the traffic demands are met. In **P1**, all slices identify the link combinations to be active in the time window T and the associated weight (i.e. fraction of time), subject to: a) traffic demands per slice and bi-directional link allocations, b) per slice UL/DL configurations. Initially, we define the potential traffic demand of a slice s over the link e as $l(e, s)$, to determine the rate requirement that will contribute to the load of that link over a time window, if this link is used to carry the traffic related to that particular slice. Each link e has a upper-bound capacity that corresponds to the maximum rate for each time instance, defined as $c_e, \forall e \in E$. Using the notion of link capacity and link load, a new parameter, which captures the number of time-slots required for a link to satisfy its target rate is defined as follows:

$$f_{e,s} = \left\lceil \frac{l(e,s)}{c_e} \right\rceil \tag{10}$$

where $\lceil \cdot \rceil$ accounts for the ceiling function, i.e. the least integer that is greater than or equal to x. In other words f_e can be interpreted as a cost function of each link, which captures how many timeslots are required to meet the desired data rate requirements per slice. Given the cost definition, **P1** can be formulated as an integer programming problem which aims to find the optimal link activation in a way that the total cost (corresponding to the timeslots required to meet a certain edge demand) is minimized. The problem is formulated as follows:

$$w^*{}_{\Omega_i} := min \sum_{t \in T} \sum_{s \in S} \sum_{e \in E} f_{e,s} x_{e,s}(t), \forall \Omega_i \tag{11}$$

Subject to:

$$\sum_{s \in S} x_{e=\{i,j\},s}(t) \neq \sum_{s \in S} x_{e=\{j,i\},s}(t), \forall i, j \in V \tag{12}$$

$$\sum_{t \in T} f_{e,s} x_{e,s}(t) \leq D_{max}(s), \forall s \in S, \forall e \in E \tag{13}$$

$$f_{e=\{i,j\},s} \leq T conf_{DL}(s) \sum_{i=1}^{|\Omega|} w_{\Omega_i} 1_{e,s,\Omega_i}, \forall e \in E, \forall s \in S \tag{14}$$

$$f_{e=\{j,i\},s} \leq T(1 - conf_{DL}(s)) \sum_{i=1}^{|\Omega|} w_{\Omega_i} 1_{e,s,\Omega_i}, \forall e \in E, \forall s \in S \tag{15}$$

$$\sum_{i=1}^{|\Omega|} w_{\Omega_i} = 1, w_{\Omega_i} \geq 0 \tag{16}$$

where (12) accounts for the half duplex constraint and (13) sets an upper bound per slice delay ($D_{max}(s)$) threshold. Moreover, (14) and (15) set the maximum cost constraint per link for DL and UL respectively considering a function of the time window T, the configuration ratio ($conf_{DL}$ or $conf_{UL}$) and the summation of weights w_{Ω_i} per link, corresponding to the activation ratio of sub-graphs Ω_i enclosing that link. In this context, $1_{e,s,\Omega_i}$ is defined as a binary variable, which is 1 only if the link is enclosed in Ω_i sub-graph. Finally, (16) shows that the summation of all w_{Ω_i} should be equal to 1.

P2 in turn performs a multi-channel interference-aware resource allocation in activated bi-directional links to optimize the perceived user performance. In particular, this problem aims to find the optimal Resource Block (RB) allocation for each link, based on inter-cell and cross-link interference constraints, i.e. via estimated SINR based on active interferers. Knowing that each sub-graph is active for a fraction of time (based on the optimal weight $w^*_{\Omega_i}$) and that slice constraints are incorporated in **P1** for each sub-graph, this problem can be solved independently without considering the time dimension and slice constraints. Then, the optimal achievable capacity, which can be extracted from a sub-graph, is normalized using the weight factor. The problem can be formulated as follows:

$$\underset{k}{maximize} \sum_i w^*_{\Omega_i} \sum_{n=1}^{N} \sum_{rx \in \Omega_i(Rx)} \sum_{tx \in \Omega_i(Tx)} k_{tx,rx,n} log_2 \left(1 \right. \tag{17}$$
$$\left. + \frac{P_{tx,rx,n} G_{tx,rx,n} L(d_{tx,rx}{}^\rho)}{\sum_{tx' \neq tx \in \Omega_i(Tx)} P_{tx',rx,n} G_{tx',rx,n} L(d_{tx',rx}{}^\rho) + \sigma^2} \right)$$

s.t. (16) and

$$k_{tx,rx,n} \in \{0,1\}, \forall tx \in \Omega_i(Tx), \forall rx \in \Omega_i(Rx), \forall n \in N \tag{18}$$

$$\sum_{n=1}^{N} P_{tx,rx,n} \leq P_{tx,max}, \tag{19}$$

$$\sum_{rx \in \Omega_i(Tx)} k_{tx,rx,n} \leq 1, \forall tx \in \Omega_i(Tx), \forall n \in N, \forall \Omega_i \tag{20}$$

where $K = \{k_{tx,rx,n} | k_{tx,rx,n} \in \{0,1\}\}$ is a binary variable corresponding to the allocation decision for the sub-channel n to the link between the transmitting and receiving ends tx, rx of Ω_i sub-graph. Furthermore, (19) is the total transmit power constraint for each tx, which is at most equal to $P_{tx,max}$ and (20) is the constraint that restricts the same sub-channel to be allocated to a link that originates from a single transmitter towards different receivers.

V. A SERVICE-ORIENTED TDD SLICE FRAMEWORK

A. Solution to P1: A Slice-aware TDD pattern Activation

For **P1** a heuristic solution is provided as illustrated in Algorithm 1 for activating the links in a time window based on the slice demand and aforementioned constraints. Initially in Step 0, a list of permitted timeslots for UL and DL is introduced per slice considering the TDD configuration pattern where a link can be activated only for a given Transmission Time Interval (TTI). A weight $f_{e,s}$ is also defined based on the slice traffic demand and a list of conflicting links, taking into account the half duplex constraint. In Step1, a random link is chosen to be included in a Candidate List (CL) for the first TTI and then the next link is identified with the minimum demand, provided that it does not violate the above rules. Once selecting a link, in Step 3 it is added to the CL and reduce its weight by 1. This is repeated in Step 4 and Step 5, till no more links exist for this TTI and then this process is repeated till all TTIs are considered (Step 6).

Algorithm 1: Slice-aware TDD pattern Activation

Step 0: ∀ slice (s):
- Set list of allowable timeslots per slice for DL: *AM_DL (s, TTI)* based on *confDL(s) and* for UL: *AM_UL (s, TTI)* based on *confUL(s)*.
- Set vectors of Links (*E*) and Traffic Demand per link: *fe,s*
- Set List of conflicting links for each link *e: Conf (e, s) and CL={}*

Step 1: Start from random link e₀, add to *CL={e₀}*
Step 2: Add the link e* with the lowest *f(e*,s)* to *CL* list $\forall e^*: e^* \notin AM\_DL, AM\_UL$ or $\forall e^*: e^* \notin Conf(\{CL\}) \rightarrow CL=CL+\{e^*\}$
Step 3: Reduce *f(e*,s)* by 1. If *f(e*,s)* is 0, remove *e** from *E* list
Step 4: Go to Step 2 till E={ } or no link can be added
Step 5: Store *CL* as *FL(i)* and reset *CL={}. i=i+1* and repeat Steps 1-4
Step 6: Stop when *i=T*

B. Solution to P2: Graph-based Resource Allocation

For **P2** a graph theoretic approach is considered. The outcome of the solution to **P1** gives an allocation of links to TTIs. However it is still unknown how many and which resources can be assigned to these links in order to avoid inter-cell and cross-link interference assuring the desired slice performance. The proposed **P2** solution is illustrated in Algorithm 2.

Algorithm 2: Graph-based Resource Allocation

- Set **FL** as [#Links x #TTI] matrix from **Algorithm 1**
- Set a color set *Color* and maximum number of colors **Cmax** and *Clist={}*

for TTI=1:T
  *if **FL**(1:links, TTI)==y≤ Cmax*
    Set randomly $y \in Colors$ different colors for the links connecting to TTI
  end if
    Store color indices for all links for TTI in a matrix as:
    Coloring(Link, Color Index, TTI)
end for
for color_index=1:Cmax
  CList(color_index) =Coloring (1:links,color_index,1:T)
end for
for bands=1:RB and *color_index=1:Cmax*
  Map **bands** to **CList(color_index)** that maximizes weighted sum-rate
end for
  end for

Initially, a bi-partite graph is created including the set of links and the set of TTIs. Based on this bi-partite graph, we translate the resource allocation problem in a time-tabling problem [11], where a number of activate links are required to occupy a number of different TTIs. A s-AP has to create a time-table according to its availability in a way that no collision occurs in each slot. A graph-coloring is adopted to assign different colors, so as to restrict the allocation of links to conflicting TTIs in distinct sub-channels. As shown in Fig.2, a bi-partite graph is translated to a line coloring graph, where each node is a combination of link and TTI (edge of the bi-partite graph).
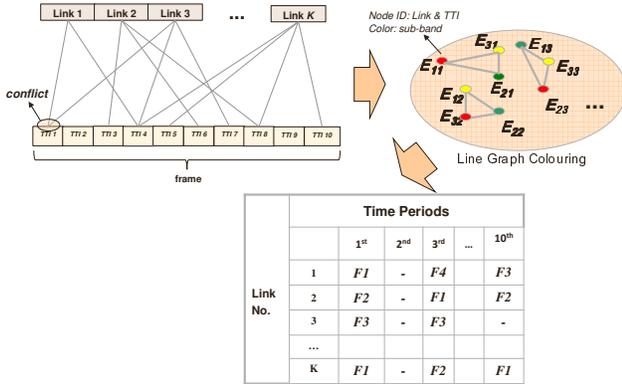

Fig.2: Graph-based Algorithm

The edge between two nodes in the line coloring graph appears only if a conflict exists at the receiving end of the bi-partite graph, which is equivalent of having two or more links being assigned to the same TTI. The graph-coloring algorithm assigns a different color to a node only in case of a conflict, which means that different sub-bands will be scheduled to avoid interference. Based on this algorithm, the output is a time-table where each link is assigned to different bands (e.g. *F1,F2,...Fn*), within distinct TTIs to ensure interference-free transmission/reception. In fact, this algorithm provides a flexible dynamic adaptation, where different parameters like the number of users, slice performance and resource availability can be altered accordingly.

VI. EVALUATION ANALYSIS

In the evaluation study we focus on an outdoor small cell deployment of 4 s-APs covering a hotspot area, using the 3GPP LTE as baseline for our simulations (24 users uniformly distributed, 3GPP UMi channel, ideal backhaul). In each s-AP the corresponding users (6 users per cell) are randomly distributed. We run Matlab Monte Carlo simulations and random user drops for 500 snapshots. We assume 4 slices, whereas each slice has different TDD pattern (Slice 1: 80/20, Slice 2: 70/30, Slice 3: 60 /40, Slice 4: 50/50). At each snapshot, we randomly select 6 users out of 4 cells to be connected to each slice and we also have a random traffic demand (1-10Mbps per user for both UL and DL). For our simulation comparison we consider:

- *Benchmark 1* is the CSDR [10] without slicing where each s-AP can adopt a different TDD pattern, while using the same spectrum resources, with inter-cell and cross-link interference potentially deteriorating performance.
- *Benchmark 2* is the service-oriented TDD slicing [7], where slices are assigned a constant amount of resources (¼ of resource blocks in our simulations) and different TDD patterns are used independently for each slice. This solution provides a high spectral efficiency due to the interference isolation, but at the cost of lower resource utilization, which can limit the peak throughput.
- In our proposal, **Algorithms 1** and **2** are used to select links and allocate resources over the entire range of resource blocks, while keeping interference at low levels.
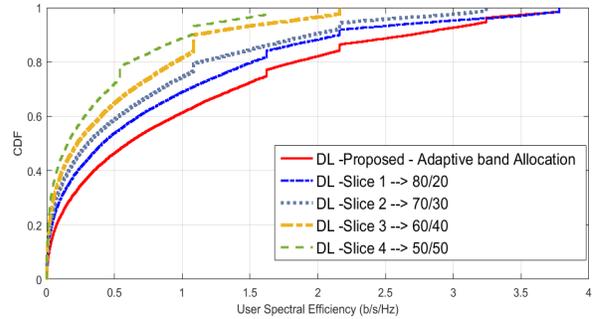

Fig.3: CDF of user Spectral Efficiency

Initially, we show the comparison of spectral efficiency, which is an ideal metric for illustrating the effect of the interference mitigation at the user side (especially at the tail of the CDF). We compare this with fixed partitioning of resources per slices as in [7] and we show the trend for different TDD patterns, which affects the spectral efficiency and the gain of the proposed scheme enabling a flexible allocation of sub-bands to flows of different slices. As can be seen in Fig.3, the TDD pattern plays an important role at the CDF spectral efficiency. For example, focusing on DL, the higher the percentage of time allocation, the higher is the average spectral efficiency per user, since more resources are allocated for the DL links. In addition, if we assume adaptive partitioning of resources to different slices, we can achieve a higher spectral efficiency by allowing orthogonal frequency resources for potential conflicting DL links for each time slot, where the DL links are active.

Fig.4 and Fig.5 show the comparison of CDF curves of DL throughput per user as well as for UL throughput respectively (averaging it over the allocated TTIs) for all snapshots. In Fig.4, we observe that the proposed solution outperforms Benchmarks 1 and 2 since it better addresses the trade-off between interference isolation vs resource utilization. Benchmark 1 shows the worst case interference scenario, whereas Benchmark 2, orthogonalizes resources for different slices. For Benchmark 2, we aggregated the DL rate for all slices collectively and it is shown that for the median and the 90th percentile of the CDF, the average throughput can be increased by more than 150%. The proposed solution shows a significant

gain even over the 2nd Benchmark, due to the fact that it achieves higher spectral efficiency with more resources being allocated to DL links based on the corresponding demand (in Benchmark 2, some resources may be wasted).
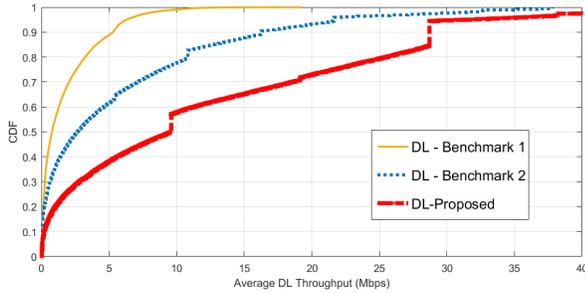

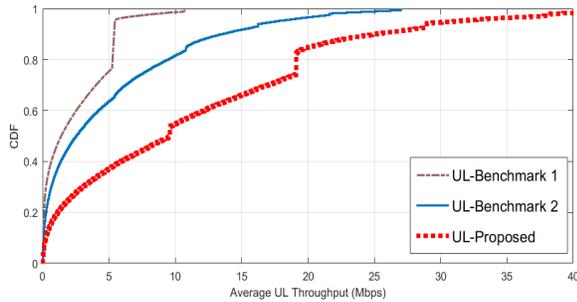**Fig.4: CDF of Av. User Throughput (DL)**

Fig.5: CDF of Av. User Throughput (UL)

In Fig.5, a similar trend is observed for the CDF of the UL throughput. Our proposed solution shows similar performance at the 10th percentile of the CDF (cell-edge performance), whereas at median and 90th percentile (cell-center) it outperforms both Benchmarks 1 and 2 respectively. This gain is mainly due to the fact that we can achieve a better UL spectral efficiency, and at the same time allocate more resources to links based on the actual demand, so as to maximize the total performance.

VII. CONCLUSIONS

This paper introduces a resource optimization framework for allocating resources enhancing the benefits of service-oriented network slicing for emerging 5G TDD networks. Our proposal provides customization allowing each slice to adopt a separate UL/DL ratio, while at the same time it introduces flexibility in allocating resource blocks towards different slices considering the entire spectrum band. To cope with interference, our solution adopts a slice-aware TDD pattern activation given a time window for particular slice demands and a graph-based resource allocation, which avoids inter-cell and cross-link interference when assigning resources. The simulation study considering a test case of a dense small cell deployment shows that our proposal significantly outperforms the benchmarks (i.e. [7][10]) in UL and DL spectral efficiency and throughput. This gain comes from the efficient utilization of resources, which maximizes the performance, while mitigating inter-cell interference.

REFERENCES

[1] GSMA, The Mobile Economy, 2017.

[2] Cisco Visual Networking, Global Mobile Data Traffic Forecast Update, 2015-2020, White Paper, Feb. 2016.

[3] 3GPP TR 22.891, Feasibility Study on New Services and Markets Technology Enablers, Stage 1, Rel.14, Jun 2016.

[4] NGMN Alliance, NGMN 5G White Paper, Tech. Rep. Ver. 1, Feb. 2015

[5] V. Sciancalepore, et.al, "Mobile Traffic Forecasting for Maximizing 5G Network Slicing Resource Utilization", IEEE Infocom, Atlanta, May 2017.

[6] T. Shimojo, et.al., "Future Mobile Core Network for Efficient Service Operation'", IEEE NetSoft, London, Apr. 2015.

[7] S. Costanzo, et.al., "Service-Oriented Resource Virtualization for Evolving TDD Networks Towards 5G", IEEE WCNC, Doha, Apr. 2016.

[8] S. Chen, J. Zhao and Y. Peng, "The development of TD-SCDMA 3G to TD-LTE-advanced 4G from 1998 to 2013", IEEE Wireless Communications, Vol.21, No.6, Dec. 2014.

[9] 3GPP TS 36.300, E-UTRA and E-UTRAN; Overall description; Stage2, Rel.14, Sep. 2017.

[10] Z. Shen, et.al, "Dynamic Uplink-Downlink Configuration and Interference Management in TD-LTE", IEEE Communication Magazine, Vol.50, No.11, Nov. 2012.

[11] S. Pirzada and A. Dharwadker, "Applications of Graph Theory", Journal of The Korean Society for Industrial and Applied Mathematics (KSIAM), vol. II, no. 4, pp. 19-38, 2007

[12] K. Lee, et.al., "Aligned Reverse Frame Structure for Interference Mitigation in Dynamic TDD Systems'", IEEE Transactions on Wireless Communications, doi: 10.1109/TWC.2017.2735398.

[13] H. Celik and K. W. Sung, "On the Feasibility of Blind Dynamic TDD in Ultra-Dense Wireless Networks'", IEEE VTC-Spring, Glasgow, May 2015.

[14] M. S. ElBamby, M. Bennis, W. Saad and M. Latva-aho, "Dynamic Uplink-Downlink optimization in TDD-based Small Cell Networks'", IEEE ISWCS, Barcelona, Aug. 2014.

[15] K.Samdanis, et.al, "Virtual Cells: Enhancing the Resource Allocation Efficiency for TD-LTE'", IEEE VTC Fall, Vancouver, Sep. 2014.

[16] K.I. Pedersen, G. Berardinelli, F. Frederiksen and P. Mogensen, "A Flexible 5G Wide Area Solution for TDD with Asymmetric Link Operation", IEEE Wireless Communications, Vol.24, No.2, Apr. 2017.