Intelligent User-Centric Networks: Learning-Based Downlink CoMP Region Breathing

Li Wang, *Senior Member, IEEE*, Gunnar Peters, Ying-Chang Liang, *Fellow, IEEE* and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—In the presence of irregular transmission/reception point (TRP) topologies and non-uniform user distribution, the user-to-node association optimization is a rather challenging process in real user-centric networks, especially for the joint transmission aided coordinated multipoint (CoMP) technique. The grade of challenge further escalates, when taking the dynamic user scheduling process into account in order to enhance the system capacity attained. To tackle the above-mentioned problem, we holistically optimize the system by conceiving joint user scheduling and user-to-node association. Then, for the sake of striking a significantly better balance between the network capacity and coverage quality, we propose a generalized reinforcement learning assisted framework intrinsically amalgamated both with neural-fitted Q-iteration as well as with ensemble learning and transfer learning techniques. Consequently, a powerful policy can be found for dynamically adjusting the set of TRPs participating in the joint transmission, thus allowing the CoMP-region to breathe, depending on both the temporal and geographical distribution of the tele-traffic load across the network. To facilitate the prompt learning of the global policy supporting flexible scalability, the overall network optimization process is decoupled into multiple local optimization phases associated with a number of TRP clusters relying on iterative information exchange among them. Our simulation results show that the proposed scheme is capable of producing a policy achieving a network-edge throughput gain of up to 140% and a network capacity gain of up to 190% under the challenging scenario of having a non-uniform geographical UE distribution and bursty traffic.**

***Index Terms*—single frequency network, machine learning, user-centric, CoMP.**

I. Introduction

A. Background & Motivations

INTERFERENCE is one of the main problems in the design of mobile communication systems. Interference limits the reusability of radio resources and has traditionally led to a host of different techniques to combat its detrimental impact. Spatial frequency reuse and its diverse variants [1–3] limits the reuse of the same frequency in close-by cells, for reducing the destructive impact of interference, hence guaranteeing a certain minimum service level across the network, albeit at the expense of reducing the area spectral efficiency (ASE) in bit/sec/Hz/km^2 [4].

The adoption of a radical frequency reuse factor of 1 and base station (BS) densification results in strongly interference-limited spots in near-future networks, when aiming for offering increased capacity to meet the escalating data-rate demands right across the network. As an answer to this problem, more advanced interference avoidance techniques relying on multi-cell coordination [5–8] have been devised in different dimensions, such as for example radio resource allocation and transmit power control. These have resulted in different tradeoffs between the cell's coverage area and its ASE. A compelling alternative approach to interference avoidance relying on coordination is interference exploitation, which effectively turns the interference into useful signal either during its transmission or after its reception with the aid of advanced coordinated multi-point techniques (CoMP) [9–11].

An efficient solution to the cell edge problem is constituted by the advanced concept of user-centric networks [12–14], but a high user-density combined with coherent multi-point cooperation comes with its own set of new problems, which have only been partially addressed. A fundamental problem is the user-to-transmission-reception-point (TRP) association. Cellular networks have traditionally relied on the association of the users to the specific BS, which provides the strongest downlink (DL) signal. This simple rule may however create load-balancing problems, which become particularly severe in the face of realistic non-uniformly distributed bursty traffic load across the network.

Thus, if we aim for improving the overall network capacity, other factors, like the geographic load distribution and bandwidth utilization should be taken into account, in addition to the received signal strength. When load balancing is taken into account, the problem of user association to a single TRP has already been proven to be NP-complete [10, 15]. Hence, the problem of user association to multiple TRPs, or equivalently the clustering of cooperative TRPs in CoMP scenarios considered becomes even more complex [16–18]. In [17] the authors consider the user association to a joint (multi-cell) HetNet scenario with the specific aim of 1) minimizing the sum-load and/or 2) minimizing the maximum-load while taking into account the inter-dependence of resource consumption due to mutual interference in deriving their model. The algorithms proposed in [17] are capable of achieving on the order of 20% sum-load reduction and 40% maximum load reduction compared to the case of no joint transmission (JT), when using the objective function of 1) and 2), respectively. By contrast, in this treatise we firstly tackle the challenging problem of user-centric multi-TRP-association in the context of the so-called single frequency network (SFN) by carrying out a joint TRP association and user scheduling design, leading to a novel dynamic SFN (DSFN) scheme.

On the other hand, radio access networks (RANs) are data-rich environments, where data is continuously gathered in the form of radio measurements or other system observations by thousands of user equipments (UE) and network entities. Nonetheless, the above-mentioned user-centric TRP-association and scheduling invoked in CoMP scenarios - similarly to many other resource management (RRM) algorithms - derive little insight from such data, since data is utilized as input to run rule-based algorithms, which is then swiftly discarded either upon aging (e.g., after a few milliseconds) or

The work of YC Liang was supported by the National Natural Science Foundation of China under Grants 61631005 and U1801261, and the 111 Project under Grant B20064.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/N004558/1, EP/P034284/1, EP/P034284/1, EP/P003990/1 (COALESCE), of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom.

when a user moves across radio cells. Therefore, the decisions taken using such data, which are currently treated as a short-lived and localized commodity by RANs, are circumscribed in time and space.

Fortunately, the rapid advances in the field of machine learning (ML) combined with the abundance of data available in RANs and the recent advances in hardware specialized to handle large data sets, present not only an opportunity to deploy cutting-edge communication technology, but also a chance for creating a paradigm shift both in the RRM architecture and algorithm design [19–21], in order to meet the stringent requirements of future cellular networks. Therefore, in this treatise we devise a novel ML-aided CoMP region breathing mechanism for the proposed DSFN scheme in order to further enhance its performance in the context of dynamically changing system conditions. The proposed framework exploits the data available in the network in the form of radio measurements and other system observations, for conceiving a sophisticated learning-aided user-TRP association strategy, which is shown to significantly outperform the conventional human-engineered ones.

More specifically, we pursue an entirely different approach. Rather than aiming for a simplified and hence potentially inaccurate model of the system constructed for optimization from a single-user's perspective, end-to-end learning is carried out by adopting a joint multi-user optimization perspective invoking an efficient reinforcement learning (RL) framework known as Neural-Fitted Q-Iteration (NFQ) [22], which is combined here with ensemble learning and transfer learning. Our simulation results quantify the dramatically improved performance compared to the conventional SFN and DSFN schemes.

B. *Contributions & Organization*

Against the above background, the novel contributions of this treatise are summarized as follows:

- For the sake of enhancing the capacity of SFN networks, the so-called downlink signal-to-interference ratio (SIR) protection level is defined, based on which the DSFN concept is proposed relying on a single-user perspective.
- Based upon the theoretical sum-rate analysis of the DSFN system, our scheduling process is jointly designed with the TRP selection, which can further improve the network's capacity at a slightly increased computational complexity, but without compromising the network's coverage quality.
- In order to further boost both the network's coverage quality and capacity, we address the rather challenging problem of user-centric TRP-grouping and user-association from the multi-user perspective by proposing ML based techniques to embed artificial intelligence (AI) into the DSFN system, which also facilitates the employment of our ML based solutions in other RRM optimization problems.
- We intrinsically amalgamated the transfer learning and ensemble learning based scenario decomposition techniques with iterative decision exchange and load density function (LDF) based information compression. Furthermore, adaptive-step-size based actions are devised under the neural-fitted Q-iteration (NFQ) framework, which successfully circumvent the 'curse of dimensionality' [23] in our large-scale context.

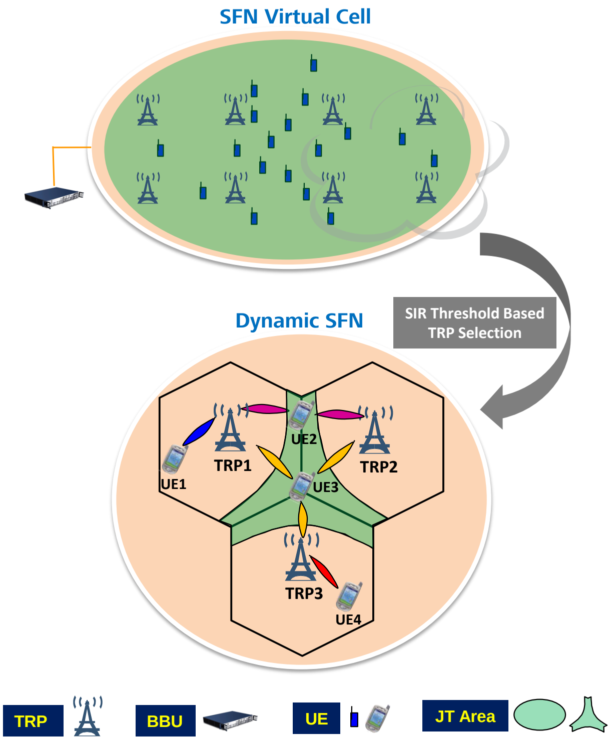


Fig. 1. Illustration of SFN and DSFN networks.

- We quantify the performance limits of the above learning-based CoMP region breathing scheme in the context of tele-traffic statistics obtained from an operational network in Singapore. Hence, we capture the characteristics of a dynamically changing user distribution, large bursty packets of FTP and of small packets (web traffic). We also considered mixed tele-traffic as well as irregular TRP deployment in conjunction with a wide range of inter-site distances. The simulation results demonstrate significant gains, which are as high as 140% and 190% for the network's coverage quality and average user throughput, respectively, under high-system-load scenarios.

The rest of this paper is organized as follows. The principles of both SFN and DSFN are introduced in Section II, where our generalized system model is developed. Based upon the multiple-user sum-rate analysis, the TRP selection and user scheduling are jointly designed in Section III. An attractive learning-based CoMP region breathing framework is proposed in Section IV along with a series of key algorithmic component designs. Finally, our simulation results and performance analysis are discussed in Section V, followed by our concluding remarks provided in Section VI.

II. Dynamic Single Frequency Network

A. *SFN Principles*

The single-frequency network (SFN) concept as a transmitter-side macro-diversity technique has been conceived for eliminating the traditional cell edges, thus improving the coverage area and quality. This SFN concept has been widely deployed both in indoor and outdoor scenarios, such as for example, office buildings, airports, train stations, stadiums etc. In the traditional cellular network where each base station (BS) has its own unique cell ID, no macro-diversity is attained, unless the 'make-before-break' soft-handover concept is exploited. By contrast, the compelling concept of hyper-cell based edgeless connectivity relies on the SFN philosophy, where a group of TRPs shown in Fig. 1 shares the same cell

ID and thus transmits the same cell-specific reference symbol. Moreover, all TRPs of an SFN-based hyper-cell having its own independent baseband processing unit co-located in the centralized baseband unit (BBU) pool will transmit the same data to a specific UE. No other UEs roaming in the same hyper-cell can be scheduled to transmit on the same radio resource blocks (RBs) in a particular transmit time interval (TTI). In addition to the CoMP-based signal strength enhancement, the UEs also benefit from a significantly reduced handover-imposed signaling overhead, which reduces the call-dropping ratio, since no handover is needed when roaming within the hyper-cell. However, this is achieved at the expense of lost scheduling opportunities in the absence of frequency reuse within an SFN hyper-cell.

B. Dynamic SFN

In order to enhance the network's capacity and to exploit the fact that each TRP has its own baseband processing capability, the so-called DSFN is conceived in this paper for improving the frequency-reuse strategy with the aid of sophisticated TRP selection, as also illustrated in Fig. 1. Explicitly, from a single UE perspective certain TRP(s) may be removed from the joint-transmission-aided (JT) TRP set at the cost of a moderate signal-strength degradation. Our proposed TRP selection may be carried out centrally in the BBU for each UE relying on the uplink (UL) reference signal's received power (RSRP) measured at each TRP, which will be elaborated on in Section III-A. Then, by exploiting the dissimilar path loss and shadowing effects, the UEs located in different areas may have distinct JT TRP sets and UEs having non-overlapped JT TRP sets are allowed to be scheduled for transmission on the same RB(s) within a specific TTI, potentially yielding a higher cell throughput in hot-spot scenarios. As exemplified in Fig. 1, the same temporal and spectral resources can be assigned to UE1 and UE4 as a result of the above-mentioned TRP selection.

C. System Model

In a network constituted by K SFN-hyper-cells, the kth ($k \in \mathcal{K} = \{\text{SFN}_k, k \in 1, 2, \cdots, K\}$) hyper-cell is composed of a set of TRPs $\mathcal{L}_k$, which are connected to a centralized BBU pool. The uth UE in the set of all UEs $\mathcal{Q}_k$ roaming within the kth SFN hyper-cell is jointly served by all TRPs belonging to its selected JT TRP cluster $\mathcal{L}_u^k$. Since frequency-division duplexing (FDD)[1] is considered in this treatise, no channel state information (CSI) is assumed to be available at the BSs for joint downlink beamforming or for multiuser MIMO transmission.

In the case of a SFN, the JT TRP cluster constituted for the uth UE in the kth hyper-cell is constantly configured for including all TRPs of the current hyper-cell. Explicitly, we have $\mathcal{L}_u^k = \{\text{TRP}_l, l \in 1, 2, \cdots, |\mathcal{L}_k|\}$. By contrast, the JT TRP set chosen for UEs might not include all TRPs belonging to the hyper-cell in the case of our DSFN system, thus for the sake of attaining a capacity enhancement using DSFN, multiple UEs can be scheduled for simultaneous transmission on the same RB as long as there is no overlap among their own selected JT RRU clusters. In general, the cardinality of the scheduled UE set $\mathcal{U}_k$ of the kth hyper-cell on a specific resource block (RB) in a certain TTI can be expressed as:

$$|\mathcal{U}_k| \begin{cases} > 1, & \text{if } \exists\ u \in \mathcal{Q}_k : \mathcal{L}_u^k \cap \mathcal{L}_v^k = 0\ \forall v \neq u \in \mathcal{Q}_k \\ = 1, & \text{otherwise.} \end{cases} \tag{1}$$

Assuming OFDM transmission with both the subcarrier- and TTI-indices omitted for notational simplicity, the generalized per-subcarrier per-TTI system model for the uth UE of the kth hyper-cell may be formulated as:

$$\mathbf{y}_u^k = \sum_{l \in \mathcal{L}_u^k} \widetilde{\mathbf{H}}_{u,l}\mathbf{s}_u + \underbrace{\sum_{v \in \mathcal{U}_k/u} \sum_{i \in \mathcal{L}_v^k} \widetilde{\mathbf{H}}_{u,i}\mathbf{s}_v}_{\text{Intra-hyper-cell interf.}} + \underbrace{\sum_{q \in \mathcal{K}/k} \sum_{v \in \mathcal{U}_q} \sum_{j \in \mathcal{L}_v^q} \widetilde{\mathbf{H}}_{u,j}\mathbf{s}_v}_{\text{Inter-hyper-cell interf.}} + \mathbf{n}_u, \tag{2}$$

where $\widetilde{\mathbf{H}}_{u,l} \in \mathbb{C}^{N_r \times M_t}$ represents the equivalent channel matrix, which can be expressed as:

$$\widetilde{\mathbf{H}}_{u,l} = g_{u,l}\mathbf{H}_{u,l}\mathbf{P}_l, \tag{3}$$

with $g_{u,l}$ being the distance-based path-loss component plus shadow fading. Furthermore, $\mathbf{H}_{u,l} \in \mathbb{C}^{N_r \times M_t}$ is the fast fading matrix associated with the channels connecting the uth N_r-antenna-element-assisted UE and the lth RRU equipped with M_t transmit antennas, with its elements drawn from an ensemble of independent and identically distributed (i.i.d.) complex random variables having zero mean and unit variance. To elabroate, $\mathbf{P}_l \in \mathbb{R}^{M_t \times M_t}$ is a matrix having diagonal elements representing the power radiated from the M_t antennas. Moreover, $\mathbf{n}_u \in \mathbb{C}^{N_r \times 1}$ of Eq.(2) denotes the additive white Gaussian noise (AWGN).

As observed from Eq.(2), depending on the TRP selection strategy, the signal transmitted in the DSFN system for the uth UE may be contaminated not only by the interference generated by all the other hyper-cells, but also by the one emanating from certain TRPs within the serving hyper-cell. Essentially, the DSFN system may achieve a higher frequency-reuse factor in pursuit of increasing the system capacity at the cost of reducing SIR owing to the presence of intra-hyper-cell interference. The instantaneous SIR may be calculated as:

$$\gamma_u^k = \frac{\sum\limits_{l \in \mathcal{L}_u^k} \left\| \widetilde{\mathbf{H}}_{u,l}\mathbf{s}_u \right\|^2}{\sum\limits_{v \in \mathcal{U}_k/u} \sum\limits_{i \in \mathcal{L}_v^k} \left\| \widetilde{\mathbf{H}}_{u,i}\mathbf{s}_v \right\|^2 + \sum\limits_{q \in \mathcal{K}/k} \sum\limits_{v \in \mathcal{U}_q} \sum\limits_{j \in \mathcal{L}_v^q} \left\| \widetilde{\mathbf{H}}_{u,j}\mathbf{s}_v \right\|^2}. \tag{4}$$

All the key variables defined above and in the ensuing sections are summarized in Table I.

III. TRP-Selection and Scheduling Co-Design

A. Single-User Perspective TRP Selection

A spatial domain isolation based JT TRP clustering strategy will be proposed for the DSFN system relying on its associated UL RSRP measured at every TRP of the serving hyper-cell. Now let us first define the so-called downlink SIR-protection

[1] In fact, a SFN can be readily conceived with time-division duplex (TDD) systems as well and then the SFN downlink transmission technique may invoke diverse schemes for exploiting the channel knowledge at the transmitter. However, this may only be realized with more advanced UEs supporting transmission mode (TM) 9 and/or TM10. Considering the currently rather low TM9 and TM10 UE penetration, our discourse is confined to the classic SFN. Therefore, the discussion on its derivatives is beyond the scope of this paper.

TABLE I
Glossary of Key Variable Definitions

Variable	Definition	
$\mathcal{K}$	Set of all SFN hyper-cells in the network	
$\mathcal{L}_k$	The full set of TRPs of kth hyper-cell	
$\mathcal{Q}_k$	Set of all UEs served by kth hyper-cell	
$\mathcal{L}_u^k$	Selected JT TRP cluster for uth UE of kth hyper-cell	
$\mathcal{U}_k$	Set of UEs scheduled on the same temporal and spectral resources inside kth hyper-cell	
M_t	Transmit antenna number of TRP	
N_r	Receive antenna number of UE	
$\mathbf{y}_u^k$	Received signal of uth UE inside kth hyper-cell	Sec. II
$\widetilde{\mathbf{H}}_{u,l}$	Equivalent channel matrix absorbing path loss, shadowing, fast fading and transmit power associated with uth UE and lth TRP	
$\mathbf{s}_u$	Downlink transmit signal to uth UE	
$\mathbf{n}_u$	AWGN noise at uth UE	
$g_{u,l}$	Path loss and shadowing factor	
$\mathbf{H}_{u,l}$	Fast fading matrix associated with uth UE and lth TRP	
$\mathbf{P}_l$	Transmit power diagonal matrix	
$\rho_{u,l}$	Spatial domain isolation level associated with uth UE and lth TRP	
$\mathcal{L}_{u,l+}^k$	Set of TRPs inside kth hyper-cell having higher values of $E\left\{g_{u,i}^2 \left\Vert\mathbf{P}_i\right\Vert^2\right\}$ than lth TRP	
$\mathcal{L}_{u,l-}^k$	Set of TRPs inside kth hyper-cell having lower values of $E\left\{g_{u,i}^2 \left\Vert\mathbf{P}_i\right\Vert^2\right\}$ than lth TRP	
γ_u	Instantaneous SINR of uth UE	
B	Total RB number of system bandwidth	
N_u	Average number of RBs assigned to uth UE per TTI	Sec. III
$N_{\text{JT},l}$	Average number of RBs assigned to JT-mode UE per TTI, who is primarily served by lth TRP	
$N_{\text{SD},l}$	Average number of RBs assigned to SD-mode UE per TTI, who is primarily served by lth TRP	
$\mathcal{Q}_{\text{JT},l}$	Set of JT-mode UEs primarily served by lth TRP	
$\mathcal{Q}_{\text{SD},l}$	Set of SD-mode UEs primarily served by lth TRP	
C	UE sum rate of a DSFN cell	
$\mathcal{P}_k$	Set of decoupled twin-TRP sub-scenarios of kth hyper-cell	
$\widetilde{\mathcal{L}}_p$	Set of two TRPs of pth sub-scenario	
$\mathcal{G}_{p,l}$	Set of $\triangle$RSRP grids into which the area around lth TRP of pth sub-scenario is divided	
$\mathcal{Q}_{p,l}$	Set of UEs considered to be primarily served by lth TRP in pth sub-scenario	Sec. IV
$\mathcal{Q}_{p,l,n}$	Set of UEs falling inside nth $\triangle$RSRP grid around lth TRP of pth sub-scenario	
T	RL action period	
$B_{u,t}$	RB number allocated to uth UE at time t	

TABLE II
Pseudo code of single-user-perspective TRP selection

```
Algorithm 1: Single-User-Perspective TRP Selection
Repeat with a certain period
  Initiate: $\mathcal{L}_u^k = \emptyset$, where $u \in \mathcal{Q}_k$ and set fixed value to $T_{\text{SIR}}$.
  Loop $u \in \mathcal{Q}_k$
    Loop $l \in \mathcal{L}_k$
      1) Generate the sorted lists, $\mathcal{L}_{u,l+}^k$ and $\mathcal{L}_{u,l-}^k$ for the UE-TRP
         pair $(u,l)$, based on the uplink RSRP measurements;
      2) Compute the downlink spatial domain isolation factor between
         $u$th UE and $l$th TRP using Eq. (6).
    End loop
    Select TRPs for $u$th UE based on the following criterion:
      $\mathcal{L}_u^k = \left\{l \in \{1,2,\cdots,|\mathcal{L}_k|\} | \rho_{u,l}^k \le T_{\text{SIR}}\right\}.$   (7)
  End loop
End repeat
```

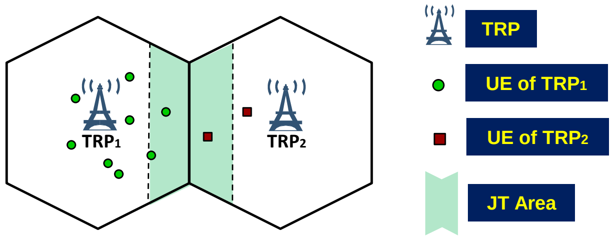


Fig. 2. Illustration of CoMP region in DSFN determined by T_{SIR}.

level, $\rho_{u,l}^k$, associated with the uth UE and lth TRP pair within the kth hyper-cell as follows:

$$\rho_{u,l}^k = \lg \frac{\mathbb{E}\left\{\left\Vert \sum\limits_{i\in\mathcal{L}_{u,l+}^k} \widetilde{\mathbf{H}}_{u,i} \right\Vert^2\right\}}{\mathbb{E}\left\{\left\Vert \widetilde{\mathbf{H}}_{u,l} + \sum\limits_{j\in\mathcal{L}_{u,l-}^k} \widetilde{\mathbf{H}}_{u,j} \right\Vert^2\right\}} = \lg \frac{\sum\limits_{i\in\mathcal{L}_{u,l+}^k} \mathbb{E}\left\{g_{u,i}^2 \left\Vert\mathbf{P}_i\right\Vert^2\right\}}{\mathbb{E}\left\{g_{u,l}^2 \left\Vert\mathbf{P}_l\right\Vert^2\right\} + \sum\limits_{j\in\mathcal{L}_{u,l-}^k} \mathbb{E}\left\{g_{u,j}^2 \left\Vert\mathbf{P}_j\right\Vert^2\right\}}, \tag{5}$$

where $\mathcal{L}_{u,l+}^k$ associated with $|\mathcal{L}_{u,l+}^k| = a$ and $\mathcal{L}_{u,l-}^k$ with $|\mathcal{L}_{u,l-}^k| = b$ denote all the TRPs of the kth hyper-cell having higher and lower values of $\mathbb{E}\left\{g_{u,i} \left\Vert\mathbf{P}_i\right\Vert^2\right\}$ $(i \neq l)$ than the lth TRP, respectively. Assuming the allocation of equal power, the downlink SIR-pretection level of Eq. (5) may be simplified to

$$\rho_{u,l}^k = \lg \frac{\sum\limits_{i\in\mathcal{L}_{u,l+}^k} g_{u,i}^2}{g_{u,l}^2 + \sum\limits_{j\in\mathcal{L}_{u,l-}^k} g_{u,j}^2}, \tag{6}$$

where the path loss and shadowing factors can be estimated based on the uplink RSRP measured at all TRPs of the hyper-cell of interest. It is important to note that the SIR-protection level defined in Eq. (6) for a specific UE-and-TRP pair is dependent not only on the relative position between themselves but also on their relative positions with respect to all the other TRPs of the hyper-cell. The physical meaning of the SIR protection level of Eq. (6) between the uth UE and the lth TRP is the maximum downlink intra-hyper-cell average SIR that can potentially be achieved by removing the lth TRP plus a fixed number of additional TRPs (i.e. b) from the original full-JT based TRP cluster $|\mathcal{L}_k|$.

Upon evaluating each TRP's SIR-protection level with respect to the uth UE using Eq. (6), a single-user perspective based TRP selection mechanism can be devised for attaining a guaranteed intra-hyper-cell SIR target T_{SIR}, where the centralized BBU simply selects all TRPs having a $\rho_{u,l}^k, (l \in \{1,2,\cdots,|\mathcal{L}_k|\})$ value lower than the SIR target T_{SIR} in order to form the JT TRP cluster $\mathcal{L}_u^k$ for the uth UE. Consequently, the resultant intra-hyper-cell SIR for the uth UE is guaranteed to be no lower than T_{SIR}.

B. Multi-UE Perspective Based Sum-Rate Analysis

The above JT TRP clustering process is continued in the same manner independently for all UEs inside the hyper-cell of interest with the configured T_{SIR} essentially imposing practically imperceivable CoMP JT region boundaries. Without loss of generality, a twin-TRP hyper-cell case is considered here with the hyper-cell ID k omitted for notation simplicity in order to carry out our sum-rate analysis based on a multi-UE perspective, thus shedding light on the ensuing TRP-selection-and-scheduling co-design as well as on the learning-based CoMP region breathing algorithm design. As illustrated in Fig. 2, the UE served by both its primary and neighboring TRPs is regarded as being in *JT mode*, while the one only served by its primary TRP is considered to be in *spatial division (SD) mode*. Furthermore, all UE downlink transmissions within the same hyper-cell are jointly scheduled by the centralized BBU based on the classic proportional fairness (PF) criterion.

Lemma 1: The PF scheduler employed in the conventional non-SFN cellular network, which prioritizes the spectral resource allocation for UEs in each TTI according to the descending order of their PF metric, tends to equally divide the total amount of temporal and spectral resources among UEs in the long run.

Proof of Lemma 1: The well-known PF metric of the uth UE, which is defined as the ratio between its instantaneous spectral efficiency and its average transmission rate per TTI, can be calculated using the function $f_{\text{PF}}(\cdot)$ defined as:

$$f_{\text{PF}}(\gamma_u) = \frac{r_u}{\bar{R}_u} = \frac{f_{\text{SE}}(\gamma_u)}{\mathbb{E}\left\{f_{\text{SE}}(\gamma_u)\right\} \cdot N_u}, \tag{8}$$

where $f_{\text{SE}}(\gamma_u)$ is the function quantifying the uth UE's spectral efficiency based on its instantaneous signal-to-interference-plus-noise ratio (SINR) γ_u, whereas N_u denotes the average number of RBs allocated to the uth UE per TTI. The PF scheduler prioritizes transmissions to the UEs according to the descending order of their PF metrics. Essentially, the PF scheduler aims for making the expected values of the UEs' PF metrics identical, where we have:

$$\begin{aligned}
&\mathbb{E}\left\{\frac{f_{\text{SE}}(\gamma_u)}{\mathbb{E}\left\{f_{\text{SE}}(\gamma_u)\right\} \cdot N_u}\right\} = \mathbb{E}\left\{\frac{f_{\text{SE}}(\gamma_v)}{\mathbb{E}\left\{f_{\text{SE}}(\gamma_v)\right\} \cdot N_v}\right\} \\
&\Rightarrow \frac{\mathbb{E}\left\{f_{\text{SE}}(\gamma_u)\right\}}{\mathbb{E}\left\{f_{\text{SE}}(\gamma_u)\right\} \cdot N_u} = \frac{\mathbb{E}\left\{f_{\text{SE}}(\gamma_v)\right\}}{\mathbb{E}\left\{f_{\text{SE}}(\gamma_v)\right\} \cdot N_v} \\
&\Rightarrow N_u = N_v.
\end{aligned} \tag{9}$$

This completes the proof of Lemma 1, which is valid for the traditional non-DSFN cellular network deployment.

With the aid of Lemma 1, our sum-rate analysis can be carried out for the DSFN system as below under the constraint of Eq. (1) imposed by the TRP selection. As shown in Fig. 2, the set $\mathcal{Q}$ of all UEs in the twin-TRP hyper-cell can be classified into four sets $\mathcal{Q}_{s,l}$, with $s \in \{\text{JT, SD}\}$ and $l \in \{1, 2\}$ corresponding to the UE's mode and to its primary TRP, respectively. According to Eq. (1), the SD-mode UEs belonging to the TRP$_2$ (i.e. UEs in the set $\mathcal{Q}_{\text{SD},2}$) can be scheduled simultaneously with the SD-mode UE associated with the TRP$_1$ (i.e. UEs in the set $\mathcal{Q}_{\text{SD},1}$) on the same RB. Additionally, if the number of SD-mode UEs served by TRP$_\iota$ (e.g. $\iota = 1$ in the case of Fig. 2) is higher than that served by TRP$_\kappa$ (e.g. $\kappa = 2$), the UEs in the set $\mathcal{Q}_{\text{SD},\kappa}$ will be scheduled more frequently than those from the set $\mathcal{Q}_{\text{SD},\iota}$, hence leading to a higher average number of RBs being allocated per TTI, namely:

$$N_u > N_v, \;\; (u \in \mathcal{Q}_{\text{SD},\kappa}, v \in \mathcal{Q}_{\text{SD},\iota} : |\mathcal{Q}_{\text{SD},\iota}| > |\mathcal{Q}_{\text{SD},\kappa}|). \tag{10}$$

Based on Eq. (9) and (10), it can be readily seen that the long-term average PF priority of UEs in $\mathcal{Q}_{\text{SD},\kappa}$ is expected to be constantly lower than that for the UEs in $\mathcal{Q}_{\text{SD},\iota}$:

$$\mathbb{E}\left\{f_{\text{PF}}(\gamma_u)\right\} < \mathbb{E}\left\{f_{\text{PF}}(\gamma_v)\right\}, \tag{11}$$

where $u \in \mathcal{Q}_{\text{SD},\kappa}$ and $v \in \mathcal{Q}_{\text{SD},\iota}$.

Consequently, the UEs from the set $\mathcal{Q}_\kappa = \mathcal{Q}_{\text{SD},\kappa}$ will essentially avoid participating in the competition with all the other UEs within the same hyper-cell (i.e. UEs from the union $\mathcal{Q}_\iota = \mathcal{Q}_{\text{JT},1} \cup \mathcal{Q}_{\text{JT},2} \cup \mathcal{Q}_{\text{SD},\iota}$) for scheduling opportunities. According to Lemma 1, the PF scheduler will on average equally share the temporal and spectral resources among the competing UEs in the set $\mathcal{Q}_\iota$ and $\mathcal{Q}_\kappa$, respectively. Therefore, the average number of RBs allocated per TTI to the UE from $\mathcal{Q}_\iota$ and to that from $\mathcal{Q}_\kappa$ can be expressed as:

$$N_{\text{JT},1} = N_{\text{JT},2} = N_{\text{SD},\iota} = \frac{B}{|\mathcal{Q}_{\text{JT},1}| + |\mathcal{Q}_{\text{JT},2}| + |\mathcal{Q}_{\text{SD},\iota}|} \tag{12}$$

and

$$N_{\text{SD},\kappa} = \frac{B}{|\mathcal{Q}_{\text{JT},1}| + |\mathcal{Q}_{\text{JT},2}| + |\mathcal{Q}_{\text{SD},\iota}|} \times \frac{|\mathcal{Q}_{\text{SD},\iota}|}{|\mathcal{Q}_{\text{SD},\kappa}|}, \tag{13}$$

respectively. Then, the expected sum-rate (i.e. the hyper-cell capacity) C may be expressed as:

$$\begin{aligned}
C &= \sum_{v \in \mathcal{Q}_\iota} f_{\text{SE}}(\gamma_v) \times N_{\text{SD},\iota} + \sum_{u \in \mathcal{Q}_\kappa} f_{\text{SE}}(\gamma_u) \times N_{\text{SD},\kappa} \\
&= \underbrace{\frac{1}{|\mathcal{Q}_\iota|} \sum_{v \in \mathcal{Q}_\iota} f_{\text{SE}}(\gamma_v)}_{\text{Average spectral efficiency}} \times \underbrace{B}_{\text{System bandwidth}} \\
&+ \underbrace{\frac{1}{|\mathcal{Q}_\kappa|} \sum_{u \in \mathcal{Q}_\kappa} f_{\text{SE}}(\gamma_u)}_{\text{Average spectral efficiency}} \times \underbrace{\frac{B \times |\mathcal{Q}_{\text{SD},\iota}|}{|\mathcal{Q}_{\text{JT},1}| + |\mathcal{Q}_{\text{JT},2}| + |\mathcal{Q}_{\text{SD},\iota}|}}_{\text{Frequency-Reuse Gain}},
\end{aligned} \tag{14}$$

given the fact that $|\mathcal{Q}_\iota| = |\mathcal{Q}_{\text{JT},1}| + |\mathcal{Q}_{\text{JT},2}| + |\mathcal{Q}_{\text{SD},\iota}|$ and $|\mathcal{Q}_\kappa| = |\mathcal{Q}_{\text{SD},\kappa}|$. The fourth term of Eq. (14) is actually the frequency-reuse gain achieved by the DSFN system in comparison to its SFN counterpart. It can also be observed from Eq. (14) that since the average UE spectral efficiency is expected to be reduced owing to the appearance of SD-mode UEs in either the set Q_ι or Q_κ, the DSFN problem of maximizing the system capacity is equivalent to striking a desirable balance between the average UE spectral efficiency and the available spectral resources by optimizing the CoMP region via the T_{SIR} configuration.

C. TRP Selection and Scheduling Co-Design Assisted DSFN

Since the DSFN system aims for enhancing the networks capacity, while maintaining its coverage achieved by its SFN counterpart, it is crucial to guarantee that the frequency-reuse-induced extra scheduling opportunities are sufficient to compensate for the SINR reduction suffered by the SD-mode UEs, especially for those located near the boarder of the CoMP region. On the other hand, based on Eq. (1) and Lemma 1, the average number of RBs assigned to each JT-mode UE per TTI within a specific hyper-cell of the SFN system is:

$$N_{\text{JT},l} = \frac{B}{|\mathcal{Q}|}, \tag{15}$$

where $\mathcal{Q}$ is the set of all UEs within the hyper-cell of interest. Thus, by comparing Eq. (12) and (15) in the context of a twin-TRP hyper-cell of Fig. 2, it can be inferred that the average number of RBs allocated per TTI is expected to be increased even for the JT-mode UE with the aid of the DSFN scheme in comparison to that in the SFN system by the amount of:

$$\triangle N_{\text{JT},l} = \frac{(|\mathcal{Q}| - 1)\, B}{|\mathcal{Q}|^2 - |\mathcal{Q}_{\text{SD},\kappa}||\mathcal{Q}|} \geq 0, \tag{16}$$

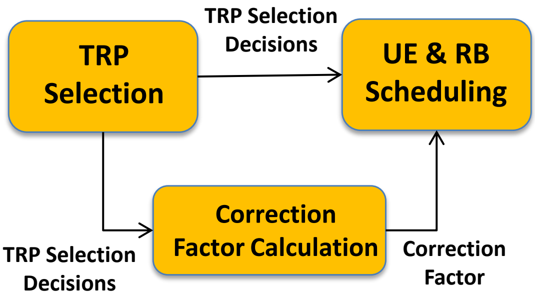


Fig. 3. TRP selection and scheduling co-design.

since we have $|\mathcal{Q}| \geq |\mathcal{Q}_{\text{SD},\kappa}| \geq 0$ and $\mathcal{Q} \neq \emptyset$.

It is observed based upon above analysis that the total amount of increased spectral resources in a DSFN hyper cell due to frequency reuse, which is achieved at the expense of the SD-mode UE's SINR, may also be partially consumed by JT-mode UEs. Hence, a co-design between TRP selection and scheduling is needed based on the policy of "no pain no gain" by limiting all frequency-reuse-induced extra spectral resources to SD-mode UEs. As such, the margin for further shrinking the CoMP region in pursuit of a higher cell capacity can be maximized without compromising the throughput of the SD-mode UE situated in the vicinity of the CoMP region, which to a large extent reflects the coverage performance of the hyper-cell of interest.

Hence, as depicted in Fig. 3, a new module is added for coupling the TRP selection and UE scheduler modules, where the correction factor of Eq. (17) seen in Table III will be calculated for each UE based on all UEs' specifically selected serving TRP sets, which in turn will then be used by the scheduler to adjust the UE's PF metric shown in Eq. (18) of Table III, thus optimizing the resource allocation. Again, the pseudo code for this co-design is described in Table III. More specifically, upon exploiting the global knowledge of the TRP selection outcomes for all the UEs, the increased amount of scheduling resources available for a particular UE as well as the number of UEs competing with it are computed, which correspond to steps 1) and 2) of the correction factor calculation loop of Table III, respectively. Then, the correction factor can be calculated at step 3) using Eq. (17), which physically indicates how many folds the scheduling resources should be increased for a particular UE on average, given the current multi-user TRP selection decisions. It is worth noting that after coupling the TRP selection with the ensuing scheduling process via the correction factor of Eq. (17) computed, the equality observed for the conventional PF resource allocation among UEs as seen in Eq. (12) and (13) will be broken and the extra scheduling resources obtained by increasing the frequency-reuse level in the DSFN hyper-cell will be only used by the SD-mode UEs who have created them at the expense of their SINRs. Indeed, it can be readily verified based on Eq. (17) that the calculated correction obeys $c_u = 1$ if the uth UE is in JT-mode and $c_u > 1$ if it is in SD-mode.

The performance enhancement achieved by the proposed TRP-selection-and-scheduling co-design aided DSFN is verified by our simulations in the context of a 6-hyper-cell network assuming the data-buffer is always full, where each hyper-cell has 20 UEs uniformly distributed and has a square-shaped 2×2-TRP deployment with an inter-TRP distance of 130m. It is observed in Fig. 4 that the proposed co-design is capable of boosting the network's capacity by 28% for the DSFN system without compromising its coverage area

TABLE III
PSEUDO CODE OF TRP SELECTION AND SCHEDULING CO-DESIGN.

```
Algorithm 2: Co-Design of TRP Selection & Scheduling
Repeat as soon as the $T_{\text{SIR}}$ is newly configuration.
  Initiate: $\mathcal{L}_u = \emptyset$, where $u \in \mathcal{Q}$, which is the set of all UEs within the hyper-cell of interest.
  Conduct TRP selection Algorithm 1 of Section III-A to generate the selected TRP set for each UE, i.e. $\mathcal{L}_u, u \in \mathcal{Q}$.
  Loop $u \in \mathcal{Q}$
    1) Generate the list, $\mathcal{J}_u$, of UEs not competing with the $u$th UE for temporal and spectral resources according to the rule: $\mathcal{J}_u = \{v \in \mathcal{Q} : \mathcal{L}_v \cap \mathcal{L}_u = \emptyset\}$;
    2) Loop $v \in \mathcal{J}_u$
         Count the UE number $J_v$ within the hyper-cell that can be scheduled on the same RB simultaneously with the $v$th UE, namely: $J_v = |\mathcal{J}_v|$, where $\mathcal{J}_v = \{i \in \mathcal{Q} : \mathcal{L}_i \cap \mathcal{L}_v = \emptyset\}$;
       End Loop
    3) Calculate correction factor for the $u$th UE:
         $c_u = 1 + \sum_{v \in \mathcal{J}_u} \frac{1}{J_v}.$   (17)
  End loop
End repeat
Repeat at every TTI
The UE and RB scheduling is conducted based on each UE's PF metric modified using the correction factor as:
    $f_{\text{PF}}(\gamma_u) = \frac{c_u \cdot f_{\text{SE}}(\gamma_u)}{\bar{R}_u}$,   where $u \in \mathcal{Q}$.   (18)
End repeat
```

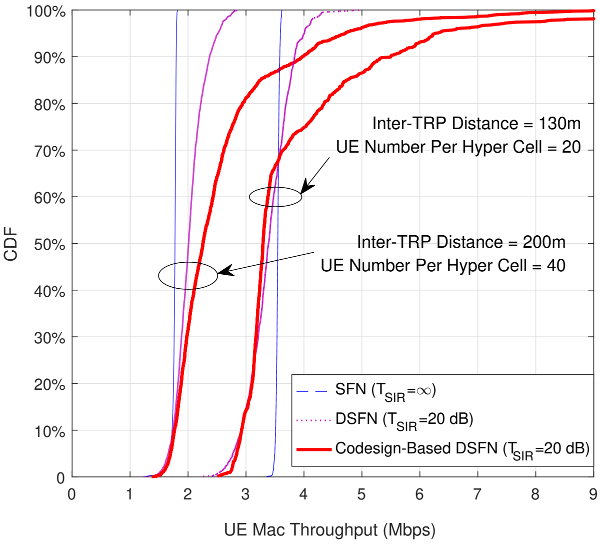


Fig. 4. Performance improvement achieved by co-design between TRP-selection and scheduling.

in terms of 5%-tile UE throughput, yielding a total capacity gain of 46.8% compared to the baseline SFN benchmark system. However, as also seen in Fig. 4, a fixed target SIR configuration, such as for example $T_{\text{SIR}} = 20$dB, cannot always guarantee a superior performance in terms of both the network's capacity and coverage for DSFN-assisted systems compared to the SFN-based benchmarks in the presence of diverse inter-TRP distances and UE/system load distributions. Therefore, in Section IV we will resort to machine learning techniques for conceiving the proposed learning-based CoMP region breathing mechanism.

IV. Learning-Based CoMP Region Breathing

As reflected by both Eq. (14) from the sum-rate point of view as well as by Eq. (17) from the TRP-selection-and-scheduling co-design perspectives, it is important but rather challenging to optimize the TRP selection from a global multi-user point of view, not to mention the extra challenge of handling the dynamically varying UE distribution as well as the irregular TRP deployment of realistic networks. Hence, we invoke both the supervised and RL [24] mechanisms for devising the proposed learning-based CoMP region breathing scheme on top of the co-design based DSFN of Section III-C, which is expected to learn from the network measurements gathered how to carry out TRP selection adaptively according to the dynamic network conditions.

A. Neural-Fitted Q Iteration Principle

RL is an area of ML dealing with how a software agent learns to behave in a given environment in order to achieve a given objective, such as maximizing a form of reward. RL has been found particularly suitable for large-scale control problems. In contrast to the model-based setting, which attempts to construct a model of the environment, we consider a model-free setup, where the Q-learning mechanism [25] consists of a set of states $s \in \mathcal{S}$ and actions $a \in \mathcal{A}$ aiming for finding an optimal policy $\pi^* : s \to a$ that maximizes the cumulative discounted rewards over an infinite horizon, given the current state $s_0 = s$, which is formulated as:

$$V(s) = \mathbb{E}\left\{\sum_{t=0}^{\infty} \gamma^t \times r(s_t, \pi^*(s_t)) | s_0 = s\right\}. \qquad (19)$$

In Eq. (19) $0 \leq \gamma < 1$ is the discount factor used for adjusting the preference for immediate rewards, while $r(\cdot)$ is the immediate reward function associated with both the state s_t and action a_t at time t. Also known as the value of the state s, the expected cumulative reward of Eq. (19) obtained under the assumption of the current state being $s_0 = s$ and taking sequential actions according to the optimal policy π^*, may be rewritten with the time index omitted (i.e. s and s' denote the current and next states, respectively) as:

$$V(s) = \max_{a \in \mathcal{A}} Q(s, a), \qquad (20)$$

where the so-called Q value associated with a 'state-action' pair can be recursively computed according to [24]:

$$Q(s, a) = \mathbb{E}\{r(s, a)\} + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}(a) \max_{b \in \mathcal{A}} Q(s', b). \qquad (21)$$

Both the expected immediate reward $\mathbb{E}\{r(s, a)\}$ and the state transition probability $P_{s,s'}(a)$ of Eq. (21) can be implicitly inferred with the aid of the classic Monte Carlo method via constant interactions between the agent and the environment. It has been proved that the Q-learning finally converges to the optimal values $V(s)$ and $Q(s, a)$ in a Markov decision process (MDP), when the optimal policy can be expressed as:

$$a^* = \pi^*(s) = \arg\max_{a \in \mathcal{A}} Q(s, a). \qquad (22)$$

In contrast to the traditional approach of relying on a large lookup table of Q-functions, we propose to combine Q-learning with a function-based approximation of the Q-function, where the parameters of a functional approximation are appropriately adjusted for minimizing the error between the predicted Q-value $Q(s, a)$ and the target value $\mathbb{E}\{r(s, a)\} + \gamma \sum_{s' \in \mathcal{S}} P_{s,s'}(a) \max_{b \in \mathcal{A}} Q(s', b)$. The proposed learning framework is capable of exploiting different functional approximation methods, such as for example artificial neural networks (ANNs) [26] and decision forests [27]. We opt for the training algorithm based on ANNs since their proven efficiency in compressing the large-dimensional inputs into low-dimensional outputs makes them one of the preferred choices of non-linear function approximations. To improve generalization, we adopt ensemble learning [28, 29] where M ANNs with distinct structures and configurations (e.g. number of layers, neurons per layer, weights initialization, etc.) are independently trained to learn alternative Q-function approximations from the same set of data. The agent follows an exploration strategy to take either an explorative action or an exploitative action selected by running majority voting on the ensemble of ANNs. While there exist a variety of ways in which ANNs can be used in an RL framework, arguably, one of the most attractive frameworks is the neural-fitted Q iteration (NFQ) algorithm [22]. More explicitly, the policy generation procedure of NFQ consists of two loops, as seen in Fig. 5. Specifically, in the ith iteration of the outer loop the inputs (s, a) and outputs $r(s, a) + \gamma \max_{b \in \mathcal{A}} Q_{i-1}(s', b)$ are generated based on the current ANN updated at the $(i-1)$st iteration. By contrast, in the inner loop the weights of the ANN are adjusted to fit the Q-function to the input-output training pairs, which is formulated as i.e. $Q_{i-1} \to Q_i$.

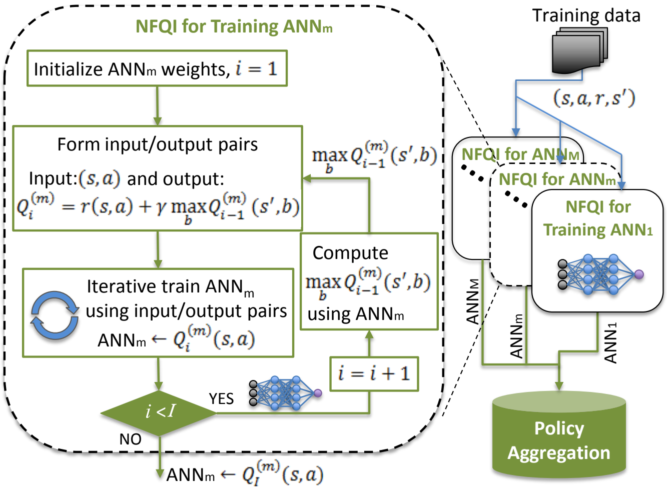


Fig. 5. Policy training based on NFQ with an ensemble of M ANNs.

B. Learning-Based CoMP Region Breathing Framework

1) Scenario Decoupling & Local Optimization: Ideally, the agent in charge of the CoMP region optimization within a hyper-cell should be trained by exposing it to diverse scenarios associated with all possible TRP deployements. However, training an agent in this way may become practically infeasible, since the state-action learning space expands exponentially as the number of TRPs increases. As a remedy, the strategy of breaking down the multi-TRP hyper-cell into multiple twin-TRP sub-scenarios is proposed, as illustrated in Fig. 6 for efficient learning space reduction. In principle, the kth hyper-cell constituted by $|\mathcal{L}_k|$ number of TRPs can be decomposed into a set of twin-TRP sub-scenarios, $\mathcal{P}_k$, with the remaining TRPs being initially treated as external interferers. To elaborate further, we have $|\mathcal{P}_k| = \binom{|\mathcal{L}_k|}{2} = |\mathcal{L}_k|!/(2!(|\mathcal{L}_k| - 2)!)$ and we denote the twin-TRP set associated with the pth sub-scenario by $\widetilde{\mathcal{L}}_p, (p \in \mathcal{P}_k, |\widetilde{\mathcal{L}}_p| = 2)$. With the aid of this scenario decomposition, only the knowledge of how to optimize the TRP selection for UEs within a twin-TRP scenario has to be learned by the so-called instance transfer learning in a centralized

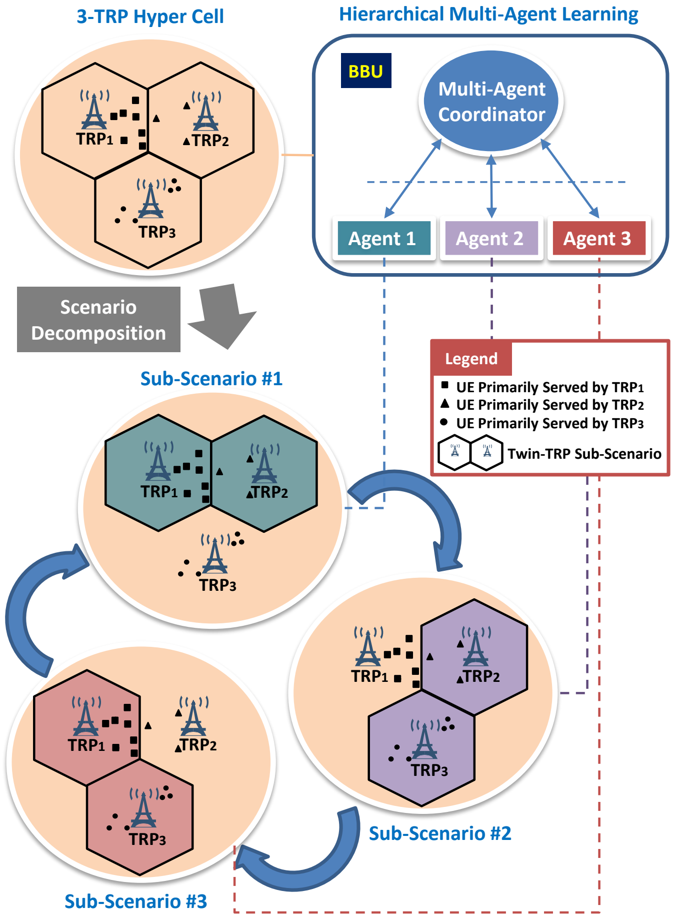


Fig. 6. Scenario decoupling based knowledge generalization.

training manner, where the training data collected from many twin-TRP scenarios in the network with distinct inter-TRP distance and user distribution are aggregated. Consequently, the above learned knowledge can then be generally applied to all the decomposed sub-scenarios both within and across hyper-cells in order to carry out the ‘local-area’ optimization. In fact, the above total number of twin-TRP sub-scenarios requiring local optimization can be significantly reduced by ignoring sub-scenarios where the two TRPs are too far apart, since the UEs therein can safely select only one out of the two TRPs (i.e. its primary TRP) without worrying about the interference emanated from the other distant TRP belonging to the same twin-TRP sub-scenario. Furthermore, in comparison to the situation without scenario decoupling, the joint action set employed by the learning agent for every twin-TRP sub-scenario’s ‘local-area’ optimization is significantly reduced, namely from $\mathcal{A}^{|\mathcal{L}_k|}$ to $\mathcal{A}^2$.

Apart from learning space reduction, it is important to note that the above scenario decomposition is also capable of converting the original TRP-specific target SIR setting $T_{\mathrm{SIR},l}, (l \in \mathcal{L}_k)$ to a per-sub-scenario, per-TRP configuration $T_{\mathrm{SIR},p,l}, (p \in \mathcal{P}_k, l \in \widetilde{\mathcal{L}}_p)$ with the benefits of both a finer optimization granularity and an increased degree of freedom. In other words, a particular TRP may have a set of different target SIR configurations associated with the decomposed twin-TRP sub-scenarios, where it belongs. As illustrated in Fig. 6, each TRP in this three-TRP hyper-cell belongs to two out of the entirely set of three decomposed twin-TRP sub-scenarios (highlighted in green field) with possibly rather distinct network conditions in terms of, for example, the inter-TRP distance and the geographical distribution of UEs. Thereby, upon optimizing the per-sub-scenario target SIR settings $T_{\mathrm{SIR},p,l}, (p \in \mathcal{P}_k, l \in \widetilde{\mathcal{L}}_p)$, the resultant shape of the CoMP region surrounding a specific TRP may become nonsymmetric and irregular.

2) Multi-Agent Coordination & Iterative Information Exchange: Due to the multi-agent nature of the learning task and the partial observability of the network state, independent and uncoordinated local optimizations for decomposed sub-scenarios are insufficient to realize the full potential of the learning based CoMP region breathing. On the one hand, if multiple logical agents, each of which is associated with a specific decomposed sub-scenario, randomly interact with the system, a mutual perturbation may be imposed on each other. As a result, the response that a specific agent observed from the environment upon taking an action at a particular time is also likely to involve the impact of other agents’ unknown actions taken at the same time, hence violating the causal relationship between state and action from a single agent’s point of view. On the other hand, the logical agent in charge of a specific twin-TRP sub-scenario can only partially observe its associated hyper-cell’s condition, which might consist of, for example, the knowledge of which TRP(s) outside the current twin-TRP sub-scenario is already serving a certain UE of interest. To address the above challenges in pursuit of achieving near-global-optimality, an efficient inter-agent coordination and communication mechanism is designed here for harmonizing multi-agent operations and for exchanging their local information.

First of all, for the sake of mitigating the mutual perturbations among logical agents and for converting the multi-agent problem back into a single-agent MDP one, the logical agents invoked will be coordinated for periodically carrying out local optimization for their own sub-scenarios in turns, obeying a global action time interval of T ms. For example, in the kth hyper cell, which is decoupled into $|\mathcal{P}_k|$ number of twin-TRP sub-scenarios, each of the associated $|\mathcal{P}_k|$ number of logical agents will take an action in turns and with an identical period of $T \cdot |\mathcal{P}_k|$. Therefore, the observation made by a specific agent after completing its action, will not be influenced by any other agents until the next agent’s action time, hence resulting in a valid observation duration of T ms for each agents.

Then, as a remedy to the partial observability problem, iterative information exchange will be invoked among the logical agents, where the periodic local TRP selection decisions made for the UEs belonging to a particular sub-scenario concerning its latest CoMP region optimization will be immediately propagated to all the other sub-scenarios having one of the two relevant TRPs. The information received may then be used for constructing and updating the state of the local environment observed by the recipient logical agents through their calculation of the UEs’ △RSRP of Eq. (23) discussed in Section IV-C1, where the design of the state representation will be detailed.

The algorithmic framework of the proposed learning-based CoMP region breathing is summarized in Fig. 7, whereas the design of its learning-related components will be discussed in the ensuing section.

C. Reinforcement Learning Agent Design

1) Information-Compression Feature Design: As revealed by Eq. (14) derived during the sum-rate analysis conducted in Section III-B, the CoMP region optimization based on a multiuser perspective relies on the key parameters, such as the

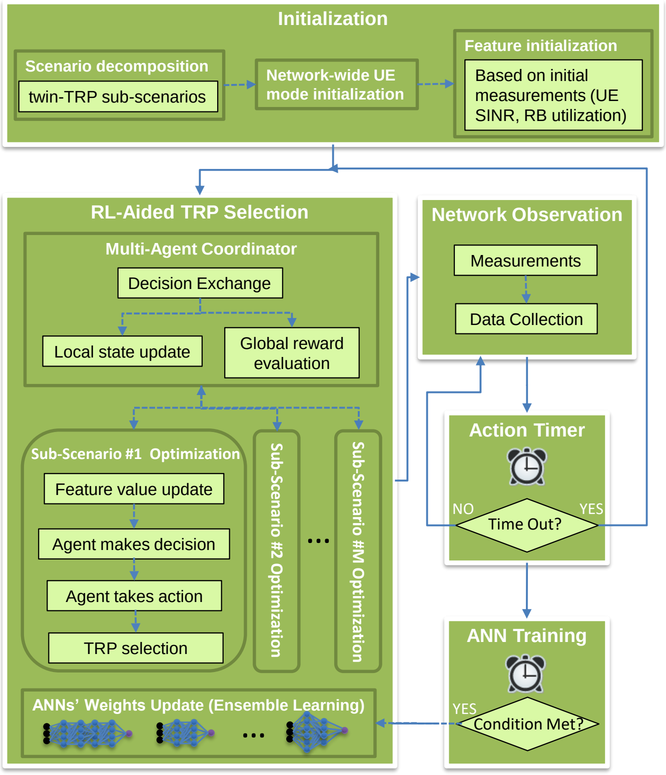
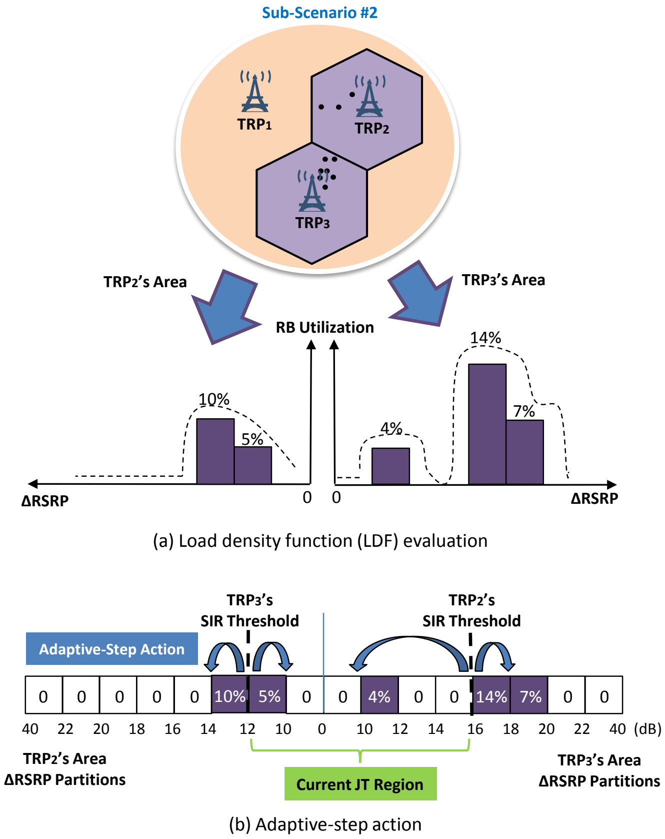


Fig. 7. Learning-based CoMP region breathing flow chart.



(a) Load density function (LDF) evaluation

(b) Adaptive-step action

Fig. 8. Information compression for feature design and adaptive-step-size action design.

UE SIRs, the UE distribution, the RB allocation and the overall system load. In the interest of capturing all the above problem-dependent information while avoiding an excessive learning space explosion, a so-called load density function (LDF) is proposed for efficiently compressing the required information, as illustrated in Fig. 8.

Basically, the LDF associated with the pth sub-scenario reflects the system's load distribution in terms of its RB utilization across the area around the two related TRPs, which are split into a set of predetermined $\triangle$RSRP partitions denoted by $\mathcal{G}_{p,l}, (l \in \widetilde{\mathcal{L}}_p)$ as dimensioned along the x-axes of Fig 8(a). In order to evaluate the RB utilization for each $\triangle$RSRP partition, the UE to $\triangle$RSRP partition association has to be first carried out based on the calculation of the UE's $\triangle$RSRP as seen below:

$$\triangle\mathrm{RSRP}_u = \sum_{i\in\mathcal{L}_u^k} \mathbb{E}\left\{g_{u,i}^2 \|\mathbf{P}_i\|^2\right\} - \sum_{j\notin\mathcal{L}_u^k} \mathbb{E}\left\{g_{u,j}^2 \|\mathbf{P}_j\|^2\right\}, \tag{23}$$

where the uth UE's selected TRP set $\mathcal{L}_u^k$ absorbs the TRP selection decisions gleaned from all other sub-scenarios with the aid of the multi-agent iterative information exchange mechanism of Section IV-B2. Then, the average RB utilization of the nth $\triangle$RSRP partition near the lth TRP of the pth twin-TRP sub-scenario is contributed by all UEs associated with that particular $\triangle$RSRP partition, namely the set of $\mathcal{Q}_{p,l,n}$, and thus may be computed based on

$$\mathrm{RBU}_{p,l,n} = \frac{\sum_{u\in\mathcal{Q}_{p,l,n}} \sum_{t\in T} B_{u,t}}{B\cdot T}, \tag{24}$$

where $B_{u,t}$ denotes the number of RBs allocated to the uth UE at the tth TTI of the current agent's observation time duration T ms. As exemplified in Fig. 8(a), the system load distribution of the sub-scenario of interest, which is evaluated separately for areas close to each of the two related TRPs, is illustrated as the RB utilization against the $\triangle$RSRP partitions. Finally, observe in Fig. 8(b) that the LDF (in terms of a vector where each component is the RB utilization encountered within a specific $\triangle$RSRP partition) can be created by concatenating the two sets of RB utilization components pertaining to the two TRPs, respectively, according to the ascending order of their primarily served UE numbers as:

$$\mathrm{LDF}_p = \mathrm{cat}\left(\mathrm{RBU}_{p,i,\{1\cdots n\cdots|\mathcal{G}_{p,i}|\}}, \mathrm{RBU}_{p,j,\{1\cdots n\cdots|\mathcal{G}_{p,j}|\}}\right), \tag{25}$$

where $|\mathcal{Q}_{p,i}| < |\mathcal{Q}_{p,j}|$. The above load-based TRP sorting between the pair of related TRPs is invoked for reducing the effective state space by half upon exploiting the symmetry of any twin-TRP scenario. The process of constructing the LDF is detailed in Table IV. Consequently, the LDF vector LDF_p of Eq. (26) together with the two TRPs' current SIR target values, i.e. $T_{\mathrm{SIR},p,l}, (l \in \widetilde{\mathcal{L}}_p)$, can be used for characterizing the state of the pth sub-scenario. Hence, the dimension of the state space designed is $\dim(\mathcal{S}) = \sum_{l\in\widetilde{\mathcal{L}}_p} |\mathcal{G}_{p,l}| + 2$.

2) Adaptive-Step-Size Action Design: Since the SIR target $T_{\mathrm{SIR},p,l}, (l \in \widetilde{\mathcal{L}}_p)$ to be optimized for the pth sub-scenario is essentially the difference between the signal-strength received at a UE from two sets of TRPs, namely, $\mathcal{L}_u^k$ and $(\mathcal{L}^k - \mathcal{L}_u^k)$, the SIR target adjustment step size should naturally be in line with the granularity of the $\triangle$RSRP partitioning conducted in the feature design phase of Section IV-C1. However, since in real networks the UEs may be sparsely distributed, they tend to yield non-uniform and discontinuous load distributions between two TRPs, as exemplified in Fig. 8(b). Furthermore,

TABLE IV
PSEUDO CODE OF LOAD DENSITY FUNCTION CONSTRUCTION.

```
Algorithm 3: Information-Compression LDF Construction.
Repeat with a period of $|\mathcal{P}_k|T$ for $p$th sub-scenario of $k$th hyper-cell
  Loop $l$th TRP within $p$th two-TRP sub-scenario
    Initiate: Divide the area around $l$th TRP into a set of predefined $\triangle$RSRP grids, i.e. $\mathcal{G}_{p,l}$.
    Loop $u \in \mathcal{Q}_{p,l}$
      1) Based on TRP selection decisions made from other sub-scenarios, compute $\triangle$RSRP for $u$th UE primarily served by $l$th TRP of $p$th sub-scenario:
         $\triangle\mathrm{RSRP}_u = \sum_{i \in \mathcal{L}_u^k} \mathbb{E}\left\{ g_{u,i}^2 \|\mathbf{P}_i\|^2 \right\} - \sum_{j \notin \mathcal{L}_u^k} \mathbb{E}\left\{ g_{u,j}^2 \|\mathbf{P}_j\|^2 \right\};$
      2) UE and $\triangle$RSRP-grid association.
    End loop
    Loop $n \in \mathcal{G}_{p,l}$
      1) Generate UE list, $\mathcal{Q}_{p,l,n}$, for $n$th $\triangle$RSRP grid;
      2) Calculate average RB utilization for $n$th $\triangle$RSRP grid:
         $\mathrm{RBU}_{p,l,n} = \frac{\sum_{u \in \mathcal{Q}_{p,l,n}} \sum_{t \in T} B_{u,t}}{B \cdot T}$
    End loop
  End loop
  Finalize: Construct LDF by concatenating $\mathrm{RBU}_{p,i,\{1 \cdots n \cdots |\mathcal{G}_{p,i}|\}}$ and $\mathrm{RBU}_{p,j,\{1 \cdots n \cdots |\mathcal{G}_{p,j}|\}}$ upon load-based TRP sorting:
    $\mathrm{LDF}_p = \mathrm{cat}\left( \mathrm{RBU}_{p,i,\{1 \cdots n \cdots |\mathcal{G}_{p,i}|\}}, \mathrm{RBU}_{p,j,\{1 \cdots n \cdots |\mathcal{G}_{p,j}|\}} \right),$
  where $|\mathcal{Q}_{p,i}| < |\mathcal{Q}_{p,j}|$.
End repeat
```

TABLE V
SUMMARY OF SIMULATION PARAMETERS

	SFN	**DSFN**	**RL-ASFN**
# of hyper-cells	6		
# of TRPs per hyper-cell	random from 3 to 5		
Inter-TRP distance	random from 150m to 300m		
Traffic type	bursty mixed traffic		
UE arrival rate	from 0.03 up to 0.2 user/second		
Mean small packet size	0.0015 Mbyte		
Mean big packet size	0.3 Mbyte		
Small/big packet user ratio	4 : 1		
Target SIR threshold T_{SIR}	∞	20 dB	Adaptive
Action time interval	N/A	N/A	2s
ANN training period	N/A	N/A	every 4000 samples

the logical agent of the pth sub-scenario employing a constant-step-size adjustment for $T_{\mathrm{SIR},p,l}$, $(l \in \widetilde{\mathcal{L}}_p)$ is likely to generate many futile actions making no influence on the TRP selection for any UE therein. This might in turn significantly slow down the speed of reaction to the system's state changes as well as potentially impose unnecessary computational and communication overheads. Hence, the load-aware adaptive step-size adjustment mechanism of Fig. 8(b) is proposed, where the legitimate action is one of three possible options for the lth TRP in the pth sub-scenario. Explicitly: 1) No $T_{\mathrm{SIR},p,l}$ adjustment; 2) Increase $T_{\mathrm{SIR},p,l}$ by a minimum amount that is 'just' sufficient to influence the TRP selection for UEs in one and only one $\triangle$RSRP partition; 3) Decrease $T_{\mathrm{SIR},p,l}$ by a minimum amount that is 'just' sufficient to influence the TRP selection for UEs in one and only one $\triangle$RSRP partition;

3) Reward Design: Given that the sub-scenarios should cooperate rather than compete with each other, the reward should be defined in a global hyper-cell-wide sense as a function of the key performance indicators (KPI)s. Therefore, we have used the following general definition of the α-fair resource allocation utility function [30]:

$$r(x) = \frac{1}{1-\alpha} \sum_{x_u \in \mathcal{X}} w_u (h_u(x_u)^{1-\alpha} - 1), \tag{26}$$

where $x_u \in \mathcal{X}$ are scalars, $\alpha \in [0, \infty)$ is a scalar coefficient, and $h_u : \mathcal{X} \rightarrow \mathbb{R}$ is a concave and increasing function. Here, x_u represents a radio measurement or a performance indicator associated with a user in the hyper cell, $\mathcal{X}$ is the set of all radio measurements or performance indicators associated with all users in the hyper cell, w_u is a weight associated with x_u and $\mathbf{x} = [x_1, \cdots, x_{|\mathcal{X}|}]$ is a vector comprising all $x_u \in \mathcal{X}$. In principle, the reward expression of Eq. (26) in conjunction with different values of α and weights w_u allows the proposed learning-based CoMP region breathing mechanism to optimize different performance metrics that can either be associated with individual UEs, with local hyper-cells, with the entire cellular radio network. For the sake of striking a beneficial system capacity vs user fairness trade-off, the so-called harmonic-mean throughput of the hyper-cell is chosen as the reward in our design, where we have $\alpha = 2, w_u = 1 \; \forall u$ and the function $h_u(x_u) = x_u$ represents the average perceived data rate of the uth user in the hyper-cell. Consequently, the reward gleaned by the agent is formulated as:

$$r(x) = 1 / \sum_{x_u \in \mathcal{X}} h_u(x_u)^{-1}. \tag{27}$$

V. PERFORMANCE EVALUATION

The performance of both the proposed DSFN and of the RL-ASFN is evaluated and compared to the baseline SFN scheme using a LTE-A/5G compliant event-driven system-level simulator, where the tele-traffic is modeled to comprise a mix of big packets (FTP) and small packets (web traffic). The data used for training the ANNs employed by the RL-ASFN scheme are generated by simulations conducted for a network consisting of 6 hyper-cells, which are heterogeneous in terms of both their system load and TPR deployment. More specifically, these hyper-cells may have rather different system load levels for different average UE arrival time intervals, which obeys Poisson distribution as well as for a variety of irregular TRP deployments associated with different inter-TRP distances. The main simulation parameters are summarized in Table V.

The CDF of the user's perceived data rate attained by the RL-ASFN scheme is benchmarked against that of the SFN in Fig. 9, which is obtained by aggregating the statistics from a number of simulations for random hyper-cell-specific average UE arrival rates ranging from 0.03 to 0.20 user/second in order to cover a broad range of system loads in terms of RB utilization. It is observed from Fig. 9 that the RL-ASFN scheme is capable of significantly outperforming its SFN counterpart, especially in the area exhibiting a relatively high traffic load. Moreover, since a periodic ANN training process takes place once every 4000 newly collected data samples, the policy learned improves as the system warms up and gathers more data. It is shown in Fig. 9 that the policy finally converges after the fourth round of training based on a total of $16,000$ data samples. This is a distinguishing aspect of a learning based system: the algorithm is not defined by static rules, but rather it evolves to a better solution as it collects more data from the network, thus having the capability of adapting to unpredictable changes in the network in terms of network deployment, channel characteristics, tele-traffic type, etc.

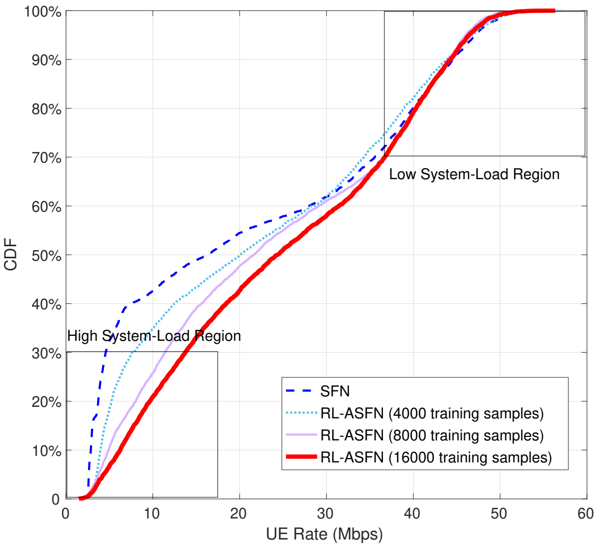


Fig. 9. Performance improvement by periodic ANN training with an increasing data amount collected through the RL-ASFN agent's interaction with the network.

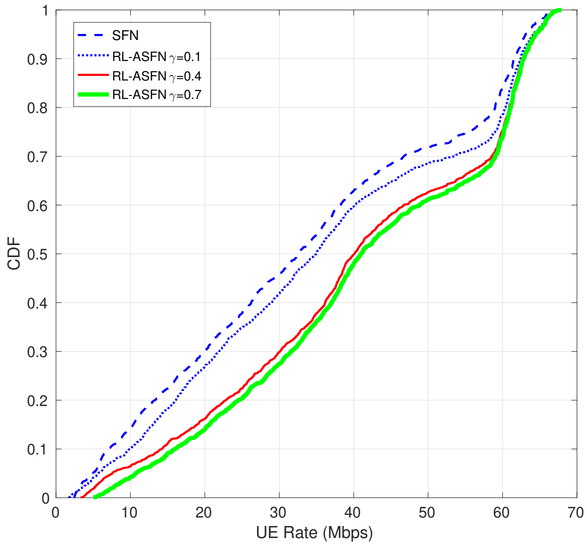


Fig. 10. Performance improvement achieved by increasing number of training iterations with increasing γ.

On the other hand, for each round of the periodic ANN training process, the NFQ of Fig. 5 is configured for $I = 3$ outer-loop iterations in conjunction with a gradually increasing value of the discount factor γ, using $\gamma = 0.1, 0.4$ and 0.7, respectively. Consequently, as shown in Fig. 10, during a certain round of the ANN training, the performance of the RL-ASFN improves and converges rapidly as the NFQ's outer-loop iteration proceeds with the horizon of the Q-function being lengthened by increasing the discount factor γ of Eq.(21). This is essentially because as γ grows, the importance of future rewards increases and so does the policy's search space, which may include an improved policy, leading to an enhanced network performance.

In addition, the impact of the reward design on the performance of the proposed RL-ASFN is characterized in Fig. 11 in terms of the network's average UE-perceived data rate and the network's 5-percentile UE-perceived rate. As expected, the reward defined as the hyper-cell's harmonic mean UE-perceived rate of Eq. (27) strikes the most favorable balance between the network's capacity and coverage in comparison to other reward designs, such as for example, employing the hyper-cell's average or minimum UE data rate as the reward. More importantly, it is also demonstrated in Fig. 11 that the implicit cooperation among multiple agents in charge of the decomposed twin-TRP sub-scenarios, which is enabled by sharing and collaboratively maximizing a global reward, such as the hyper-cell's harmonic mean UE data rate, achieves prominent performance gains over the design where a local reward is employed by the multiple agents, as exemplified by the twin-TRP sub-scenario's local harmonic mean UE data rate.

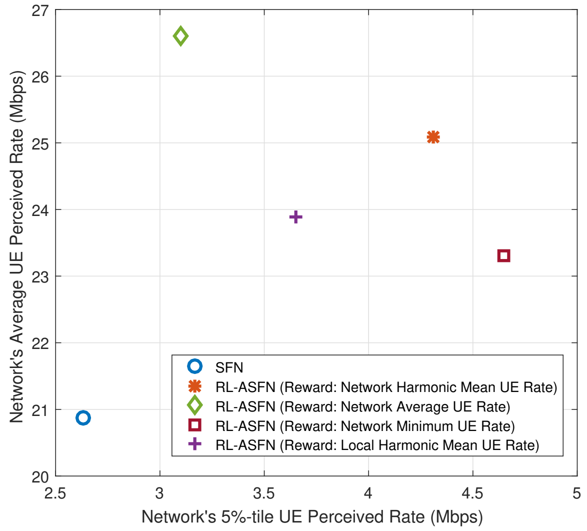


Fig. 11. Performance of the RL-ASFN in conjunction with different definitions of reward.

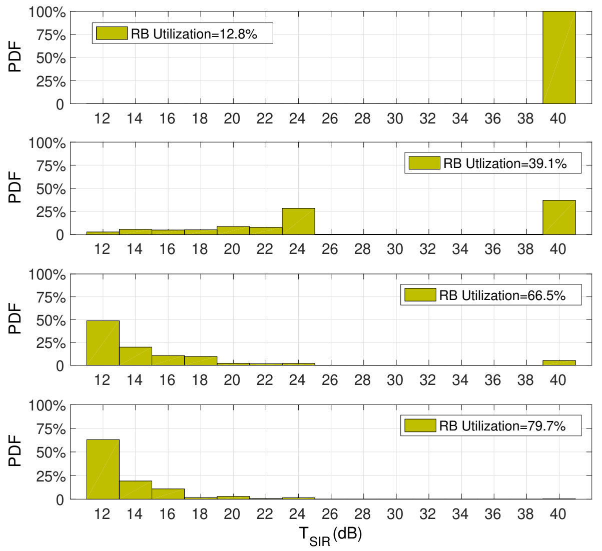
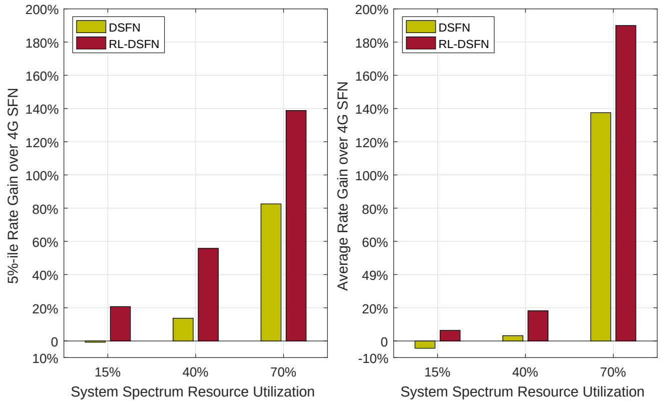


Fig. 12. Distribution of the learning-based T_{SIR} configurations. Note that the legitimate values of T_{SIR} are defined to be mainly concentrated in the range of interest from 12 to 24 dB, on top of which another legitimate possible value for T_{SIR} is 40 dB, which is defined in order to ensure that the proposed RL-ASFN encapsulate the baseline SFN as its special case.

Furthermore, the distribution of the adaptively tuned threshold T_{SIR} is plotted in Fig. 12 for different system load levels. It can be seen that the policy learned resorts to using larger values of T_{SIR} for grouping more TRPs belonging to the same hyper-cell in order to promote joint transmissions for increasing the UEs' SINRs in the context of power-limited scenarios, when the RB utilization is low. By contrast, the distribution moves towards lower values of T_{SIR} for promoting a higher frequency-reuse level, as the system becomes more bandwidth-limited due to its heavy system load.



(a) Hyper-cell-edge throughput gain.

(b) Hyper-cell capacity gain.

Fig. 13. Performance gain achieved by RL-ASFN in conjunction with codesign of TRP selection and scheduling.

Finally, Fig. 13 summarizes the performance gains achieved by both the proposed DSFN and RL-ASFN algorithms with respect to the baseline SFN scheme. Observe that at high system loads the devised DSFN scheme using $T_{\mathrm{SIR}} = 20\mathrm{dB}$ is capable of achieving up to 80% and 140% gains in terms of the 5-percentile and average user perceived rates, respectively, which is an explicit benefit of the joint TRP-selection and scheduling design of Section III. However, due to its lack of AI-assisted adaptability, it struggles to attain useful performance gains at medium system loads, whilst exhibiting some performance erosion at lower system loads in comparison to the SFN. By contrast, the RL-ASFN conceived, which combines both the co-design mechanism employed by the DSFN as well as the intelligent ML-aided adaptability become capable of outperforming both the SFN and DSFN systems right across the full range of system loads. Remarkably, in the high system load scenario, RL-ASFN achieves as much as 190% and 35% gain in system capacity against its SFN and DSFN counterparts, respectively. Viewed from a different perspective, our solution obtains 140% and 75% increase in network coverage compared to the same two benchmarks. More importantly, our RL-ASFN still attains 9% and 15% capacity gain as well as 20% and 22% coverage gains in comparison to the SFN and DSFN systems, respectively, even in the scenario of a rather low system load.

VI. Conclusions and Future Work

CoMP has emerged in recent years as an efficient solution of combating the interference affecting the performance of dense user scenarios. Moreover, owing to extending the search space of the user-to-node association, CoMP comes with its own set of problems. Designing efficient heuristic solutions by taking into account a large variety of measurements and traffic scenarios is quite a challenge. In this contribution we have taken an end-to-end learning approach, which allowed us to rapidly learn an efficient strategy, which achieves significant gains compared to both the SFN as well as to the rule-based approach of DSFN. However, as a potential problem to be solved in the future, a hyper-cell constituted of a very large number of TRPs will not only slow down the convergence of the learning process, but it will also hamper its adaptation to the dynamically fluctuating environment. To cope with this issue, rather than having fixed predetermined coordination clusters of a large number of TRPs, we may have to design a dynamic user-cluster-centric hyper-cell formation strategy. Thanks to the natural sparsity in the cellular network in terms of the user's geographical distribution, we can image that multiple hyper-cells of different size can be dynamically formed to better serve the different clusters of UEs that are located far apart from each other. Furthermore, a hierachical coalition formation mechanism can be constructed, where we can have hyper-cell-level cooperations. Although all the above-mentioned measures dealing with the coordination scalability are beyond the scope of this paper, this is likely to be one of our future research directions. On the other hand, learning-based methods could be used more extensively in the design of algorithms for future networks as a compelling design alternative to rule-based algorithms. As a natural and important extension of this work, power control and allocation over the frequency band and/or among the UEs will be considered when we design intelligent user-centric networks in our future work, which is a more challenging but critical task for further performance enhancement and energy efficiency improvement.

References

[1] 3GPP-R1-050507, "Soft frequency reuse scheme for utran LTE," *Huawei, 2005. Available (July 2012): http://www.3gpp.org/ftp/tsg ran/WG1 RL1/TSGR1 41/Docs/R1-050507.zip*.

[2] T. D. Novlan, R. K. Ganti, A. Ghosh, and J. G. Andrews, "Analytical evaluation of fractional frequency reuse for ofdma cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4294–4305, 2011.

[3] P. Yen, Q. Zhan, and H. Minn, "New fractional frequency reuse patterns for multi-cell systems in time-varying channels," *IEEE Transactions on Wireless Communications Letters*, vol. 4, no. 3, pp. 253–256, 2015.

[4] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[5] G. Li and H. Liu, "Downlink radio resource allocation for multi-cell ofdma system," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3451–3459, 2006.

[6] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1414–1425, 2010.

[7] H. Zhang, L. Venturino, N. Prasad, P. Li, S. Rangarajan, and X. Wang, "Weighted sum-rate maximization in multi-cell networks via coordinated scheduling and discrete power control," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 6, pp. 1214–1224, 2011.

[8] A. Hamza, S. Khalifa, H. Hamza, and K. Elsayed, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1642–1670, 2013.

[9] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.

[10] D. H. N. Nguyen, L. B. Le, and T. Le-Ngoc, "Optimal dynamic point selection for power minimization in multiuser downlink CoMP," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 619–633, 2017.

[11] C.-W. Chen, H.-W. Tsao, and P.-Y. Tsai, "MIMO precoder design with a compensated QR-decomposition combination for CoMP downlink scenarios," *IEEE Communications Letters*, vol. 65, no. 10, pp. 7982–7992, 2016.

[12] C. Xu, F. Zhao, J. Guan, H. Zhang, and G.-M. Muntean, "Interference management in ultradense networks: A user-centric coalition formation game approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5188–5202, 2018.

[13] Z. Chen, X. Hou, and C. Yang, “Training resource allocation for user-centric base station cooperation networks,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2729–2735, 2016.

[14] Y. Zhang, S. Bi, and Y.-J. A. Zhang, “User-centric joint transmission in virtual-cell-based ultra-dense networks,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4640–4644, 2018.

[15] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, “User association for load balancing in heterogeneous cellular networks,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, June 2013.

[16] M. M. Abdelhakam, M. M. Elmesalawy, K. R. Mahmoud, and I. I. Ibrahim, “A cooperation strategy based on bargaining game for fair user-centric clustering in cloud-ran,” *IEEE Communications Letters*, vol. 22, no. 7, pp. 1454–1457, 2018.

[17] L. You and D. Yuan, “Load optimization with user association in cooperative and load-coupled LTE networks,” *IEEE Transactions on Wireless Communications,*, Early Access, 2017.

[18] S. Bassoy, M. Jaber, M. A. Imran, and P. Xiao, “Load aware self-organising user-centric dynamic comp clustering for 5G networks,” *IEEE Access*, vol. 4, pp. 2895–2906, 2016.

[19] F. D. Calabrese, L. Wang, E. Ghadimi, G. Peters, L. Hanzo, and P. Soldati, “Learning radio resource management in RANs: Framework, opportunities, and challenges,” *IEEE Communications Magazine*, vol. 56, no. 9, pp. 138–145, 2018.

[20] Y.-J. Liu, S.-M. Cheng, and Y.-L. Hsueh, “eNB selection for machine type communications using reinforcement learning based Markov decision process,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11 330–11 338, 2017.

[21] L. T. Tan and R. Q. Hu, “Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 190–10 203, 2018.

[22] T. Gabel, C. Lutz, and M. Riedmiller, “Improved neural fitted Q iteration applied to a novel computer gaming and learning benchmark,” *IEEE Symposium on Adaptive Dynamic Programming And Reinforcement Learning*, ADPRL, 2011.

[23] R. E. Bellman, “Dynamic programming,” *Princeton University Press*, 1957.

[24] S. Sutton and A. G. Barto, “Reinforcement learning: an introduction,” *IEEE Journal on Selected Areas in Communications*, Cambridge, MA: MIT Press, 1998.

[25] C. Watkins, “Learning from delayed rewards,” *Ph.D. dissertation, University of Cambridge, England*, 1989.

[26] C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

[27] A. Criminisi, J. Shotton, and E. Konukoglu, “Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning,” *Foundations and Trends in Computer Graphics and Vision*, vol. 7, no. 2-3, pp. 81–227, 2012.

[28] Y. M. C. Zhang, *Ensemble Machine Learning: Methods and Applications*. New York, NY, USA:Springer, 2012.

[29] S. K. Goudos and G. Athanasiadou, “Application of an ensemble method to UAV power modeling for cellular communications,” *IEEE Antennas and Wireless Propagation Letters*, pp. 2340–2344, 2019.

[30] E. Altman, K. Avrachenkov, and A. Garnaev, “Generalized α-fair resource allocation in wireless networks,” *47th IEEE Conference on Decision and Control*, p. 24142419, 2008.

Li Wang (S’09-M’10-SM’18) received the Ph.D. degree from the University of Southampton, Southampton, U.K., in 2010. From 2010 to 2012, he conducted research as a Senior Research Fellow with the School of Electronics and Computer Science, University of Southampton. During his academic period, he was involved in a number of projects, such as those from UK’s EPSRC, Mobile VCE and Indian-UK Advanced Technology Centre (IU-ATC). In March 2012, he joined the R&D Center of Huawei Technologies in Stockholm, Sweden, and is currently working as a Principle Engineer in Wireless Network Algorithm Lab, leading a team focusing on 5G downlink PHY/MAC algorithm design. He has authored about 50 research papers in IEEE/IET journals and conferences, and also coauthored one JohnWiley/IEEE Press book on LTE. He has wide research interests in both radio transmission technology (RTT) and radio resource management (RRM) areas for future wireless communication technologies and networks, including cross-layer cross-module system design, millimeter Wave, communication system intelligentization and sparse signal processing. He was the recipient of Huawei Individual Contribution Award in 2015, Huawei Furture Star Award in 2017, Huawei Team Gold Medal Award in 2018 and Huawei Outstanding Individual Award in 2019.

Gunnar Peters Gunnar Peters is a Senior Expert in Network Algorithms at Huawei Sweden. He has a PhD in Mathematics and Master of Engineering in Applied Physics from the Royal Institute of Technology. He has worked in telecom for 19 years and started at Ericsson 2000 were he was an expert in Radio Performance. In 2010 he joined the Wireless organization at Huawei Sweden, were he leads a group developing radio network algorithms. Gunnar has also worked at the University of South Carolina, Uppsala University and the Royal Institute of Technology. His research covers stochastic analysis, quantum field theory, wavelet analysis, computational solid state mechanics, signal processing, and radio resource management. Gunnar has received the Individual Gold Medal Award and the Team Gold Medal Award for his work in massive MIMO, advanced receivers and the application of machine learning to RAN.

Ying-Chang Liang (F11) is currently a Professor with the University of Electronic Science and Technology of China, China, where he leads the Center for Intelligent Networking and Communications and serves as the Deputy Director of the Artificial Intelligence Research Institute. He was a Professor with The University of Sydney, Australia, a Principal Scientist and Technical Advisor with the Institute for Infocomm Research, Singapore, and a Visiting Scholar with Stanford University, USA. His research interests include wireless networking and communications, cognitive radio, symbiotic networks, dynamic spectrum access, the Internet-of-Things, artificial intelligence, and machine learning techniques. Dr. Liang has been recognized by Thomson Reuters (now Clarivate Analytics) as a Highly Cited Researcher since 2014. He received the Prestigious Engineering Achievement Award from The Institution of Engineers, Singapore, in 2007, the Outstanding Contribution Appreciation Award from the IEEE Standards Association, in 2011, and the Recognition Award from the IEEE Communications Society Technical Committee on Cognitive Networks, in 2018. He is the recipient of numerous paper awards, including the IEEE Jack Neubauer Memorial Award, in 2014, and the IEEE Communications Society APB Outstanding Paper Award, in 2012. He is a Fellow of IEEE, and a foreign member of Academia Europaea. He is the Founding Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: COGNITIVE RADIO SERIES, and the Key Founder and now the Editor-in-Chief of the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He is also serving as an Associate Editor-in-Chief for China Communications. He served as a Guest/Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, the IEEE Signal Processing Magazine, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORK. He was also an Associate Editor-in-Chief of the World Scientific Journal on Random Matrices: Theory and Applications. He was a Distinguished Lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society. He was the Chair of the IEEE Communications Society Technical Committee on Cognitive Networks, and served as the TPC Chair and Executive Co-Chair of the IEEE Globecom17.

Lajos Hanzo FREng, FIEEE, FIET, Fellow of EURASIP, DSc holds an honorary doctorate by the Technical University of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served as Governor of both IEEE ComSoc and of VTS. He has published 1900+ contributions at IEEE Xplore, 19 Wiley-IEEE Press books and has helped the fast-track career of 119 PhD students. Over 40 of them are Professors at various stages of their careers in academia and many of them are leading scientists in the wireless industry. (http://www-mobile.ecs.soton.ac.uk, https://en.wikipedia.org/wiki/Lajos_Hanzo)