User-Centric Strategy for Base-Station Virtualization in 5G Networks

Slim Zaidi*‡, Oussama Ben Smida*, Sofiène Affes*, Usa Vilaipornsawai†, Liqing Zhang†, and Peiying Zhu†
*INRS-EMT, Université du Québec, Montreal, QC, H5A 1K6, Canada, Emails: {zaidi,oussama.ben.smida,affes}@emt.inrs.ca
‡ECE Department, University of Toronto, Toronto, Ontario, M5S 3G4, Canada, Emails: slim.zaidi@utoronto.ca
†Huawei Technologies Canada Co. Ltd., Canada, Emails: {usa.vilaipornsawai,liqing.zhang,peiying.zhu}@huawei.com

***Abstract*—This paper proposes a new user-centric base-station (BS) virtualization strategy aiming to adapt users communication links to their quality of service (QoS) requirements and environments. The developed user-centric virtual base stations (uVBS)s offer substantial improvements in terms of power and spectral efficiencies while requiring minimum signaling changes at both user and network sides. They also better leverage new 5G features such as massive connectivity and extreme densification as well as new concepts such as massive MIMO and mmWave spectrum. Furthermore, our uVBSs are able to adapt to multiple network dimensions such as time, space, etc.**

***Index Terms*—Wireless/radio access virtualization, cloud-radio access network (C-RAN), user-centric architecture, dynamic adaptive clustering, massive MIMO, mmWave.**

I. Introduction

Cell-centric architectures are adopted in current 4G radio access networks (RAN)s where the cell is the network's focal point serving several users (i.e., devices: sensors, smartphones, etc., machines, vehicles) located in its coverage area [1]-[8]. Such conventional networks have limited spectrum resources and, hence, approach their limits when the services' data rate and/or the number of users increase/s. A straightforward way to circumvent this impediment is to increase the system capacity by deploying more and more transmission points (TP)s. This reduces the number of devices competing for each TP's resources and, hence, the spectrum reuse across large geographic areas. However, extreme densification results in inevitable high inter-cell interference and a poor cell-edge user experience. Some remedial solutions such as coordinated beamforming [9]-[11], inter-cell interference coordination, and fractional frequency reuse have been introduced in 4G RAN to overcome this liming factor. The latter were unfortunately unable to completely remove the cell-boundary effects, although they offer some performance gains at the cost of increased overhead and complexity.

In contrast to 4G RANs, future 5G networks will exploit wireless access virtualization (WAV) to provide boundaryless communications [1]-[8]. Indeed, using WAV, the coverage is dimensioned around the user making it the network's focal point rather than the cell. The network will then adapt the data transmission to the user's quality of service (QoS) requirement and environment, thereby creating the illusion of a moving virtual cell following it. As a result, we break away from the traditional cell-centric RAN by providing boundaryless communications where all users do not experience any cell-edge effects. This would potentially lead to substantial improvements in terms of power efficiencies and network's spectral and, therefore, to the fulfillment of 5G's pledge of ubiquitous user experience [12]. WAV will practically be enabled by capitalizing on both the massive connectivity and extreme densification to allow each user to be served by a set of optimally and carefully selected transmission points (TP)s forming a user-centric virtual base-station (uVBS).

Work supported by the NSERC/Huawei Canada/TELUS CRD Grant on 5G-WAVES (WAV Enabling Schemes), the DG and CREATE PERSWADE <www.create-perswade.ca> Programs of NSERC, and a Discovery Accelerator Supplement Award from NSERC.

Various TPs clustering approaches already exist in the literature and can be classified into two main categories: static and dynamic [13]-[17]. When static clustering is performed, uVBSs are formed using solely system information (i.e., TPs' density and positions, their available resources, etc.) and, therefore, are predetermined and rarely updated. This considerably reduces not only the complexity of static clustering, but also the extra latency, overhead, and power consumption it requires. Nevertheless, this approach usually achieves poor performance in terms of both spectral efficiency and throughput [13]. This occurs mainly because uVBSs are not adapted to the highly changing users' environments owing to the lack of user-side information such as channel quality indicator (CQI), the user's channel state information (CSI), signal-to-interference-plus-noise-ratio (SINR), etc. In turns, dynamic clustering, which exploits the latter information, provides much better performance, but incurs extra latency, overhead, and power costs that are condemned to increase even more with the massive connectivity and network densification foreseen in future 5G networks [14]-[17]. In addition, the uVBSs are usually formed using iterative greedy highly-complex algorithms that investigate all potential set constructions to ultimately settle on network partitions which are usually far from optimal. As both static clustering's low cost and dynamic clustering's high efficiency features are keys to enable efficient uVBSs, this work aims to establish a *best-of-the-two-worlds* clustering technique that combines these approaches' benefits while avoiding their drawbacks.

In this paper, we propose a new user-centric base-station (BS) virtualization strategy aiming to adapt users communication links to their quality of service (QoS) requirements and environments. The developed uVBSs offer considerable improvements in terms of power and spectral efficiencies and, further, requires minimum signaling changes at both user and network sides. They also better leverage new 5G features such as massive connectivity and extreme densification as well as new concepts such as massive MIMO and mmWave spectrum. Furthermore, our uVBSs are able to adapt to multiple network dimensions such as time, space, etc.

II. NETWORK MODEL

We consider in this paper a cloud-RAN (C-RAN) that consists of N users and M TPs connected through fiber to a central unit (CU). Each TP is equipped with K antennas while users are assumed to have a single antenna. All users are assumed to be actively communicating with the network during TP clustering.

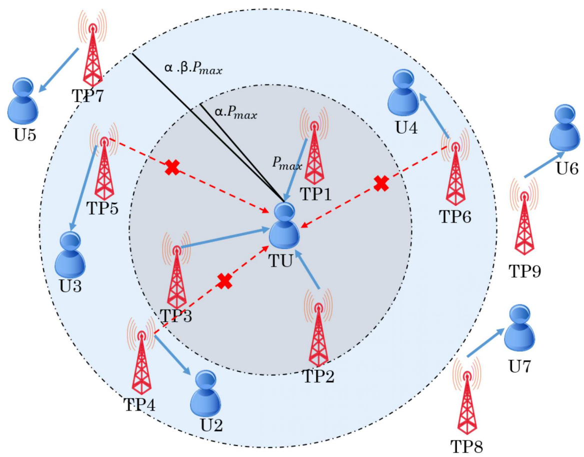


Fig. 1: Proposed TPs clustering approach.

III. PROPOSED USER-CENTRIC WAV APPROACH

In order to select the proper TP sets, we propose in this work to exploit the maximum reference signal received power (RSRP) available locally at every user (i.e., user-side information). Let $P_{\max}^{k}$ be the maximum RSRP at the k-th user given by

$$P_{\max}^{k} = \max\left\{P_{i-k}, i = 1, \ldots, M\right\}, \tag{1}$$

where P_{i-k} denotes the RSRP of the i-th TP at the k-th user.

A. Concept

Let us consider two system parameters $\alpha, \beta \in [0, 1]$ that encompass system information such as TPs' and users' densities, available resources, and positions, etc. Exploiting α and β along with (1), we can build the following two clusters from the M TPs deployed in in the C-RAN:

$$\mathrm{SC}_k = \left\{\mathrm{TP}_{i=1,\ldots,M}/\text{s.t. } \alpha P_{\max}^{k} \leq P_{i-k} \leq P_{\max}^{k}\right\}, \tag{2}$$

and

$$\mathrm{NC}_k = \left\{\mathrm{TP}_{i=1,\ldots,M}/\text{s.t. } \beta\alpha P_{\max}^{k} \leq P_{i-k} < \alpha P_{\max}^{k}\right\}. \tag{3}$$

where SC_k and NC_k are the k-th user's serving cluster (SC) and nulling cluster (NC), respectively. Accordingly, using the proposed clustering approach, TPs with RSRPs at the k-th user that are large enough to be in $\left[\alpha P_{\max}^{k}, P_{\max}^{k}\right]$ will serve it while those with moderate RSRPs in $\left[\beta\alpha P_{\max}^{k}, \alpha P_{\max}^{k}\right[$ will perform interference nulling toward it. From the k-th user perspective, all the selected TPs form then a uVBS that serve the latter and avoid interfering on it when serving other users. In turns, using the conventional single-serving TP selection, target user (TU) is only served by the TP with the highest RSRP and, hence, is subject to strong interference from other neighboring TPs. This is in contrast with the proposed approach that, as illustrated in Fig. 1, turns all high RSRP signals into useful ones and cancel the moderate RSRP signals yet strong enough to affect the TU's performance, thereby resulting in substantial throughput improvement. As α and/or β decrease/s, more TPs may join the TU's SC and/or NC which better improves its throughput. However, it is not practically feasible to indefinitely decrease these parameters without degrading the performance of other users. Indeed, if α decreases, more TPs are solicited to serve the TU and, hence, more resources are allocated to a smaller number of users. Consequently, an increasing number of users and TPs might be in outage of service or shortage of resources, respectively. As far as β is concerned, each TP has a limited nulling capability of $(K-1)$ and, therefore, it can perform simultaneous interference nulling toward at most $(K-1)$ users. The number of nulling requests received by a TP increases as β decreases and may exceed this limit, thereby hindering the performance of other users deprived of these resources.

Joint optimization of both α and β is, thus, required to guarantee both optimal resource utilization and system performance.

B. Computation of α and β

Conventional mathematical methods such as in [16]-[18] may be adopted to derive the system parameters α and β. Although having their own merits, these methods rely often on assumptions/approximations (i.e, simplified throughput expression, non-overlapping and/or single-user serving clusters, fixed or single-antenna communication [16]-[18]) that hinder the accuracy of the objective function and/or its constraints and, hence, reduce the applicability range of the obtained solution. Furthermore, this simplified version of the extremely complex clustering problem is often solved using iterative algorithms. This is actually a critical drawback since optimality and convergence time dramatically decrease with the numbers of TPs and users which are expected to be incredibly large in 5G systems. Finally, these algorithms as stated in [18] must usually run on a central unit that requires not only a high computational capability, but also the global knowledge of all network's CSI. Such information is unfortunately acquired through a frequent feedback made by all active users, thereby dramatically depleting their power and increasing the network overhead.

For all these reasons, this work opts for a purely heuristic method much more practically appealing and, hence, suitable for industrial applications. Such method consists in optimizing α and β offline using a system-level simulator. A set of α and β values are first picked from the interval $[0, 1]$ with ideally a small step before running a simulation campaign for each them. The optimal parameters are those providing the best overall network performance. This process should be repeated for different network setups (i.e., different user and TP densities). Please note that these parameters could be computed for the whole network (i.e., global parameters) or for every group of users and TPs (i.e, local parameters). Since the parameters optimization is made offline, heuristic method incurs then much lower cost and complexity than its counterparts. It is noteworthy that α and β could be also calibrated online by testing the performance gain resulting from small variations of their values.

TABLE I: uVBS implementation mechanisms.

Implementation Mechanisms	Mechanism 1	Mechanism 2	Mechanism 3
Network Side	• Broadcasts α and β • Refines or overwrites SC and NC based on global information such as traffic load, priority, QoS/QoE requirements	• Broadcasts α and β • Is not involved in the TP selection process	• Does not broadcast α and β • Decides on TP clusters based on the RSRPs feedback by all users
User Side	• Selects TP cluster(s) using the values of α and β provided by the network • Only feedbacks the RSRPs of the TPs in SC and NC to the network	• Decides local on TP cluster(s) • Informs the network of its SC and NC	• Feedbacks all its RSRPs to the network • TPs selection is completely transparent to it
Overhead	• Broadcast of α and β • Feedback of RSRPs of TPs in SCs and NCs	• Broadcast of α and β • Feedback of the IDs of TPs in SCs and NCs	• Feedback of all users RSRPs

C. uVBS implementation mechanisms

Tab. I summarizes three possible implementation mechanisms of uVBSs. In Mechanism 1, each user recommends its own SC and NC selected using α and β values provided by the network. The latter may refine or overwrite the selected TPs based on global information. Indeed, it may deny the access to some TPs for instance when their traffic load is extremely high or to serve users with higher priority or QoS requirement. In such a case, some selected TPs could be substituted or completely removed from SC and/or NC. Only the RSRPs in SC and NC are feedback using this mechanism, thereby substantially reducing both the system overhead and power costs. In turns, with Mechanism 2, the network broadcasts α and β and let the user make the final decision on its SC and NCs. The overhead is then further reduced as user needs to feedback only the selected TPs IDs. The main drawback of Mechanism 2 is that TPs selection is completely transparent to the network and, hence, it is unable to overwrite users' TP clusters to adapt to particular conditions, QoS/QoE requirements, or users' priority. This responsibility could however be handled by the user itself at the cost of additional complexity cost at its side. As far as Mechanism 3 is concerned, it consists of making the TPs selection completely transparent to the user. The latter must only to feedback al RSRPs of TPs in its vicinity. In this way, the broadcast of α and β is avoided but the incurred overhead remains prohibitive especially in ultra-dense networks context, where a huge number of users need to feedback the RSRPs of a huge number of TPs. Consequently, Mechanism 1 and 2 may be preferred at high network density as they allow substantial overhead savings while Mechanism 3 may be favored at high traffic loads in order to allow the network make some adjustments on TP clusters. Furthermore, different mechanisms could be used with different subnetworks or even different users or devices. Indeed, the denser is the subnetwork, more suitable to it will be Mechanism 1 and 2. The higher is its traffic load, more suitable to it is Mechanism 3. Moreover, privileged users may use Mechanism 2 to allow them meet their QoS/QoE requirements at any time while the rest of subscribers are only entitled to Mechanism 3.

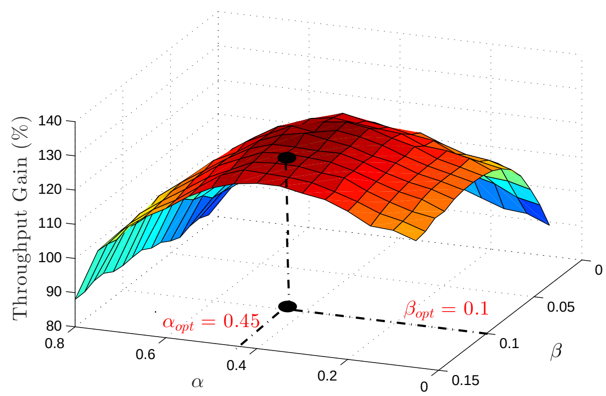


Fig. 2: Network throughput gain of the proposed clustering approach over single-serving TP selection versus α and β for $\rho = 0.31$.

IV. Simulations Results

In this section, system-level simulations are performed to analyze the performance of the proposed approach and compare it with the conventional single-serving TP selection and a static clustering solutions. The static clustering technique partitions the network into three adjacent TPs set wherein the user is served by one TP while the others perform interference nulling towards it. The heuristic method discussed in Section III-B is adopted here to optimize the parameters α and β.

In order to highlight the gains provided by uVBSs, we get rid of any form of multi-user MIMO (MU-MIMO) from our LTE standard-compliant simulator. In other words, only one user is associated with each single resource in the spatial and spectral domains. We consider, in all simulations, a channel bandwidth of 10 MHz, 7 macro-TPs whose transmit powers are 46 dBm, and 10 femto-TPs in each macro whose transmit powers are 20 dBm. In addition, users are initially (i.e., at

$t = 0$) uniformly distributed in the target area. We also assume that they are equipped with a single antenna while all TPs are equipped with two antennas (i.e., $K = 2$). We adopt a proportional fair (PF) scheduling locally at each TP. TP clustering is updated at the same rate of the dynamic point selection (DPS) (i.e., each subframe) introduced in LTE release 11 [19]. We consider in this paper that TPs in SC employ maximum ratio transmission (MRT) to jointly transmit the users data while TPs in NCs perform zero-forcing beamforming to avoid interfering on it. Please note that the choice of these particular signal combining techniques was only made for the sole sake of simplicity. Our new approach can, however, support any other advanced signal combining and/or nulling techniques [10][11].

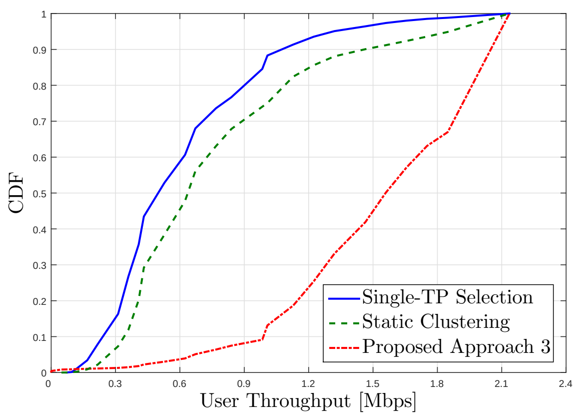
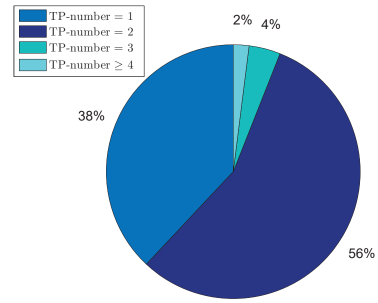


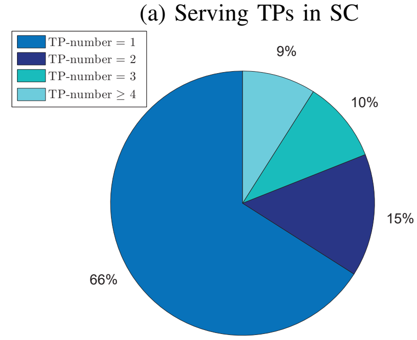
Fig. 3: CDFs of the user throughput achieved by the proposed WAV appraoch, single-serving TP selection, and static clustering when $(\alpha_{\text{opt}}, \beta_{\text{opt}}) = (0.45, 0.1)$ and $\rho = 0.31$.

Fig. 2 shows the network throughput gain achieved by the proposed approach over single-serving TP selection versus the parameters α and β for TP density $\rho = 0.31$. From this figure, we confirm the existence of optimum values $(\alpha_{\text{opt}}, \beta_{\text{opt}})$ of the parameters (α, β). We find that $(\alpha_{\text{opt}}, \beta_{\text{opt}}) = (0.45, 0.1)$ when $\rho = 0.31$. In such a case, the proposed approach achieves a throughput gain as high as 120%. On the other hand, from Fig. 2, deviations of until 10% from the optimal values of α and β results in at most 4% loss in throughput gains. This very important feature makes α_{opt} and β_{opt} robust against quantization errors. Therefore, with a low quantization level turning out to be acceptable, the overhead incurred when broadcasting α can be further reduced significantly. Therefore, with a low quantization level turning out to be acceptable, the overhead incurred when broadcasting α_{opt} and β_{opt} can be further reduced significantly.

Fig. 3 shows the CDFs of the achieved user throughput using the proposed approach, single-serving TP selection, and static clustering. We observe that the throughput achieved by 55% of the users exceeds 1.5 Mbits/s with our approach whereas only 3% and 10% of users achieve the same throughput with single-serving TP selection and static clustering, respectively. Besides these throughput gains, the proposed WAV approahc achieves significant coverage gains against its counterpart, thereby reducing (if not suppressing) the cell-edge effect.



(a) Serving TPs in SC



(b) Nulling TPs in NC

Fig. 4: Pie charts of probabilities for the number of serving and nulling TPs in each user's SC and NC for $(\alpha_{\text{opt}}, \beta_{\text{opt}}) = (0.45, 0.1)$ and $\rho = 0.31$.

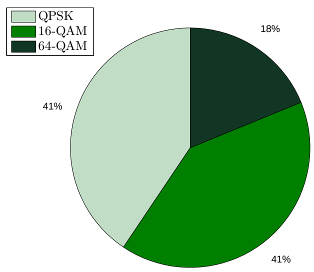
Figs. 4a and 4b illustrate the pie charts of the number of TPs in users' SCs and NCs, respectively. From Fig. 4a, 38% of users are served by a single TP whereas 56% of them are served by two TPs, 4% by three, and the rest (about 2%) by four or more. From Fig. 4b, only one TP cancels its interference towards 66% of the users whereas two TPs simultaneously cancel their interference towards 15% of them, three TPs are required for 10% of users and four or more TPs for the rest (about 9%). Hence, in most cases, each user's SC and NC cardinalities do not exceed two and therefore do not burden the network virtualization cost. Such a very suitable feature makes the proposed approach an interesting candidate for the upcoming 5G networks.

Fig. 5 shows the occurrence probabilities of QPSK, 16-QAM, and 64-QAM obtained with the proposed WAV approach, single-serving TP selection, and static clustering. We observe that 64-QAM occurs 95% of the time against 18% and 31% with single-serving TP and static clustering, respectively. This is hardly surprising since our approach offers a dramatic SINR improvement by turning the strongest interference links into useful ones and by canceling the moderate yet still problematic ones, thereby substantially increasing its link capacity. Accordingly, our proposed WAV strategy allows higher-order modulations in 5G networks in order to cope with the higher rates that better address the unprecedented demand for mobile data expected in the near future.

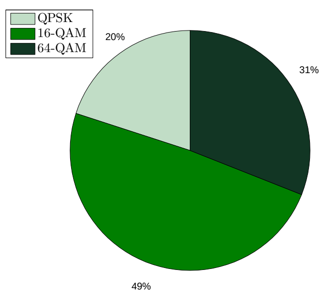
Tab. II summarizes the performance of the proposed approach and compare them with single-serving TP selection and static clustering. It lists the average throughput and 5-th percentile coverage performance of all clustering approaches. We show that the proposed approach dramatically outperforms

TABLE II: Average sum throughput and coverage achieved using the proposed approach, the single-serving TP selection, and the static approach.

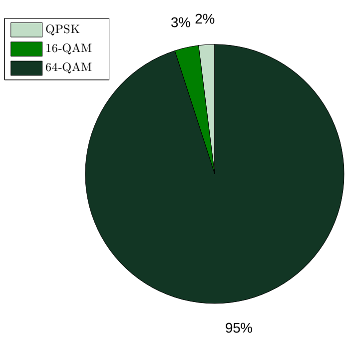
	Average Sum Throughput [Mbps]	5-th Percentile Coverage [Mbps]	QPSK modulation	16-Qam modulation	64-QAM modulation
Single-serving TP	0.703	0.175	41%	41%	18%
Static approach	0.924	0.218	20%	49%	31%
Proposed approach	1.785	0.669	2%	3%	95%



(a) Single-serving TP selection



(b) Static selection



(c) Proposed clustering approach for $(\alpha_{\text{opt}}, \beta_{\text{opt}}) = (0.45, 0.1)$ and $\rho = 0.31$.

Fig. 5: Occurrence probabilities of QPSK, 16-QAM, and 64-QAM modulations.

the single-serving TP selection in terms of both throughput and coverage. Indeed, our approach achieves an average throughput of 1.785 Mbps while single-serving TP and static clustering does not exceed 0.703 and 0.924 Mbps, respectively. This represents a throughput gain of up to 153.9% over the single-serving TP and 93.1%, respectively, over static clustering. Furthermore, according to Tab. II, our proposed approach achieves a coverage gain over single-serving TP of 282.2% and of 206.9% over static clustering. These huge performance gains highlight the efficiency of the proposed WAV approach and its significant superiority over its conventional benchmarks.

V. Discussion and Conclusions

The findings of this paper may be articulated around five desired features of any prospective WAV approach:

- **Dynamic, Adaptive**: The TP clusters formed using the proposed approach are overlapping sets whose cardinalities (i.e., the number of TPs in each set) are adjusted to each user's environment and situation. This is in contrast with former approaches wherein the clusters' cardinalities are fixed and/or the overlapping constraint is relaxed to simplify the optimization problem [16]-[18]. Due to its dynamic and adaptive SCs and NCs, the proposed WAV approach provides dramatic performance gains in terms of throughput and coverage with respect to its benchmarks.
- **Low complexity**: Our approach rely on the optimization of α and/or β (i.e., at most two parameters) for multiple users utilization in the same network or subnetwork. As discussed previously, the optimal values of these primates could be easily obtained using online calibration and/or offline simulations. We then avoid the implementation of iterative sub-optimal greedy highly-complex algorithms often required by the so far existing techniques. Table. III shows the complexity of the proposed approach and the clustering algorithms developed in [16] and [18] at both infrastructure and user sides. In all clustering solutions, all user equipments are expected to forward the information they collect each on the TPs in their vicinity. Therefore, the user-side complexity is proportional to M. On the other hand, at the infrastructure side, whereas the proposed approach require no extra processing since the parameters alpha and beta are computed offline, once for all, the conventional clustering techniques suffer from relatively huge complexity loads significantly increasing with the numbers of TPs, per TP antennas, and users.
- **Low overhead, power and latency costs**: Combining our approach with Mechanism 1 or 2, the decision on clusters is made locally at each user. This is in contrast with the existing approaches which often require that the CU has a global knowledge of all users' CSIs/SINRs to be able to form the TP clusters [16]-[18]. Therefore, the proposed approach offer significant overhead, power, and latency savings. Indeed, the overhead incurred by the conventional clustering approaches could be expressed as $B_{\text{oh}} = R_r K Q_l \sum_{i=1}^{N} M_i$ where $M_i \in \{1, \ldots, M\}$ is the number of TPs in the i-th user vicinity, Q_l is the quantization level of CSI/SINR, and R_r^{CSI} is the clusters formation refreshment rate. On the other hand, the overhead incurred by the proposed WAV approach which requires the broadcast of both α and β is

TABLE III: Complexity comparison between the proposed approaches and some benchmarks available in the literature.

	Offline processing	Infrastructure-side online processing	User-side online processing
Proposed approaches	Yes	0	$O(M)$
Approach in [16]	No	$O(KNM^2 + KM^2 + 4KNM + 4KM)$	$O(M)$
Approach in [18]	No	$O(141MK^2N^2 + 141MK^2N^2 - 39MKN^2 + 235KN^2 - 38MKN + 470KN - 65N)$	$O(M)$

$B_{\text{oh}}^{Prop} = 2R_r^{\alpha,\beta}\acute{Q}_l$ where $R_r^{\alpha,\beta}$ is the refreshment rate of α and β and $\acute{Q}_l$ is their quantization level. Assuming for extreme simplification in favor of the conventional clustering techniques that $Q_l = \acute{Q}_l{}^2$, we have then $\Omega = B_{\text{oh}}/B_{\text{oh}}^{Prop} = (R_r^{CSI}\sum_{i=1}^{N} M_i)/2R_r^{\alpha,\beta})$. Therefore, Ω substantially increases not only with the users, TPs, and antennas' numbers, but also with $R_r^{CSI}/R_r^{\alpha,\beta}$. Note here that the CSI's refreshment rate is usually in the range of milliseconds, i.e., in the TTI (transmission time interval) duration scale in LTE, while that of α and β is in the range of minutes or even hours since they depend on the numbers of TPs and users. This is actually a fundamental difference that drastically reduces the overhead and power costs. Assuming for simplicity, again in favor of conventional clustering techniques, that $R_r^{CSI}/R_r^{\alpha,\beta} = 10^3$, we measure $\Omega = 27.3\,10^3$ and $\Omega = 35.7\,10^3$ when $\rho = 0.31$ and $\rho = 0.44$, respectively, with the simulation setup described in Section IV. This means that the proposed approach, under the most unfavorable assumptions to them (i.e., equal quantization level and much smaller than expected refreshment rate ratio), still incur as much as 10^3 times less overhead, and consequently much less power and latency as well (following the same rationale) than their conventional counterparts, making them unambiguously more suitable for future 5G networks.

- **Scalability**: The performance gain achieved by the proposed approach obviously increases with the available network resources. It may then benefit from new 5G technologies such as massive MIMO and mmWave spectrum which provide high degrees of freedoms and huge spectrum, respectively. It may also benefit from advanced multi-user strategies that allow using the same resources to serve more than one user. This again in contrast with existing approaches whose complexities increase exponentially with such technologies.
- **Flexibility**: By associating different parameters to the different network dimensions, our approach pave the way towards dramatic improvements in both spectral and power efficiencies. Indeed, the definition of **user-class**-, **service**-, and **application**-based parameters allows adequate adaptation of the allocated resources to different classes of subscribers and network services and applications. Furthermore, **period**- and **location**-based parameters that properly adjust to the network conditions at different places and periods would further enhance the throughput of each user.

All these key observations unambiguously prove that the proposed approach is efficient and offers substantial performance gains while requiring negligible extra overhead, power, complexity, and latency costs, making it an interesting candidate for the upcoming 5G networks.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?," *IEEE J. Select. Areas Commun.*, vol. 32, pp. 1065-1082, June 2014.

[2] S. Zaidi, B.Hmidet, S. Affes, U. Vilaipornsawai, and L. Zhang, "User-centric wireless access virtualization strategies for future 5G networks," *IEEE ICUWB'2016,* Nanjing, China, October 16-19 2016.

[3] Z. Chang, Z. Zhou, S. Zhou, T. Chen, and T. Ristaniemi, "Towards service-oriented 5G: virtualizing the networks for everything-as-a-service,“ *IEEE Access,* vol. 6, pp. 1480-1489, December 2017.

[4] S. Zaidi, S. Affes, M. Azzakhmam, C. Despins, K. Zarifi, and P. Zhu, "Progressive hybrid greyfield wireless access virtualization with leveraged combining of cloud, fog, and legacy RANs,“ *IEEE PIMRC'2017,* Montreal, QC, Canada , October 8-13 2017.

[5] S. Zaidi, S.Affes, U. Vilaipornsawai, L. Zhang, and P. Zhu, "Wireless access virtualization strategies for future user-centric 5G networks," *IEEE GC Wkshps'2016,* Washington, DC, USA, December 4-8 2016.

[6] S. Zaidi, O. B. Smida, S. Affes, U. Vilaipornsawai, L. Zhang, and P. Zhu "QoS-based virtualization of user equipment in 5G networks (Invited paper),“ *IEEE IWCMC'2018,* Limassol, Cyprus, June 25-29 2018.

[7] I. Ahmad, T. Kumar, M. Liyanage, J. Okwuibe, M. Ylianttila, and A. Gurtov, "Overview of 5G security challenges and solutions“, *IEEE Com. Mag.,* vol. 2, pp. 36-43, March 2018.

[8] E. J. Kitindi, S. Fu, Y. Jia, A. Kabir, and Y. Wang, "Wireless network virtualization with SDN and C-RAN for 5G networks: requirements, opportunities, and challenges,“ *IEEE Access,* vol. 5, pp. 19099-19115, September. 2017.

[9] S. Zaidi and S. Affes "Distributed collaborative beamforming in the presence of angular scattering,“ *IEEE Trans. Commun.,* vol. 62, pp. 1668-1680, May 2014.

[10] S. Zaidi and S. Affes, "Distributed collaborative beamforming design for maximized throughput in interfered and scattered environments,“ *IEEE Trans. Commun.,* vol. 63, pp. 4905-4919, Dec. 2015.

[11] S. Zaidi, B. Hmidet, and S. Affes, "Power-constrained distributed implementation of SNR-optimal collaborative beamforming in highly-scattered environments," *IEEE Wireless Commun., Lett.,* vol. 4, pp. 457-460, Oct. 2015.

[12] "5G: A technology vision," White Paper, Huawei Technologies, Co. Ltd., Nov. 2013. [Online]. Available: www.huawei.com/5Gwhitepaper/

[13] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.,* vol. 8, pp. 1910-1921, Apr. 2009.

[14] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," *Proc. IEEE ICC'2008,* Beijing, China, May 19-23, 2008.

[15] J. Gong, S. Zhou, Z. Niu, L. Geng, and M. Zheng, "Joint scheduling and dynamic clustering in downlink cellular networks," *Proc. IEEE GLOBECOM'2011,* Houston, TX, USA, Dec. 5-9, 2011.

[16] A. Maaref, J. Ma, M. Salem, H. Baligh, and K. Zarifi, "Device-centric radio access virtualization for 5G Networks," *Proc. IEEE GLOBE-COM'2014,* Austin, TX, USA, Dec. 8-12, 2014.

[17] K. Zarifi, H. Baligh, J. Ma, M. Salem, and A. Maaref, "Radio access virtualization: cell follows user," *Proc. IEEE PIMRC'2014,* Washington DC, USA, Sep. 2-5, 2014.

[18] S. Fu, H. Wen, and B. Wu, "Power-fractionizing mechanism: achieving joint user scheduling and power allocation via geometric programming,“ *IEEE Trans. on Vehicular Technology,* vol. 67, no. 3, pp. 2025-2034, Mars. 2018

[19] 3GPP, " LTE Carrier Aggregation Enhancements,“ 3GPP TR 36-101, October 2014.

[20] 3GPP, "3GPP TS 36.213 V9.2.0: Evolved universal terrestrial radio access network (E-UTRA); physical layer procedures," June 2010.