Integrating Aerial Base Stations for sustainable urban mobile networks

Integrating Aerial Base Stations for sustainable urban mobile networks

Michela Meo*, Daniela Renga*, Felice Scarpa°
Department of Electronics and Telecommunications, Politecnico di Torino
*{firstname.lastname}@polito.it °felice.scarpa@studenti.polito.it

***Abstract*—The extensive densification of mobile networks is increasing the network energy consumption and leading to remarkable economical and sustainability concerns. At the same time, regulatory and physical constraints, especially in urban environments, may limit the network expansion and the free installation of Base Stations (BSs). In this context, High Altitude Platform Stations (HAPSs) are emerging as a promising solution to host aerial BSs that can provide additional capacity over a wide geographical area, to offload the on-ground mobile network and support a sustainable transition towards the 6G era. This paper investigates the potential of HAPS offloading to reduce the energy demand from the grid and the operational cost of mobile networks. Our results highlight the effectiveness of HAPS offloading in reducing the size of the RE supply that is required to achieve grid energy reduction on the terrestrial network, thus enhancing the feasibility of a sustainable evolution towards 6G networks. Different allocation strategies are designed and analyzed under several configuration settings, to dynamically adapt the HAPS capacity to the traffic variability in space and over time. A fine tuning of the strategy settings is proved effective in trading off physical constraints, operational cost, sustainability goals, and Quality of Service.**

***Index Terms*—Non-Terrestrial Networks, HAPS, 5G/6G Networks, Energy Efficiency, Renewable Energy**

I. Introduction

The current shift towards the 5G paradigm in mobile networks is significantly affecting the deployment of new generation network architectures. Indeed, especially in urban ecosystems, the Radio Access Network (RAN) infrastructures need to rapidly evolve to keep the pace with the astounding growth of traffic demand and the consequent network densification. This transition raises remarkable challenges that are bound to become even more critical in the evolution towards the future 6G networks and need to be effectively addressed. In particular, on the one side, physical constraints and regulatory restrictions may limit the extensive expansion of network infrastructures in urban scenarios. On the other side, mobile networks are reported to be responsible for 0.7% of annual electricity consumption [1] and an intense network densification process clearly poses remarkable issues in terms of operational cost and sustainability. To this extent, Resource on Demand (RoD) approaches are often adopted to reduce consumption by deactivating unneeded Base Stations (BSs) in those periods in which the traffic demand is low and local integration of Renewable Energy (RE) sources is frequently adopted to power BSs and reduce the energy demand from the grid. However, both these approaches are not sufficient to cope with the challenges mentioned above.

In this context, the introduction of aerial BSs may play a key role to provide supplementary capacity in 5G urban networks, so as to offload part of the mobile traffic and unburden the on-ground network infrastructures. In particular, High Altitude Platform Stations (HAPSs) represent a promising candidate to host aerial BSs, acting as self-sustainable network nodes that operate in the stratosphere at an altitude around 20 km [2]–[4]. HAPS bandwidth can be allocated in a flexible way by means of multiple cells, that may vary in number, coverage size, and capacity, both in space and time, to follow traffic load dynamics [5].

Various studies available in the literature demonstrate a raising interest in investigating the potential of integrating HAPSs as supplementary components in terrestrial mobile networks, steering research efforts towards the deployment of Aerial Access Networks [6]–[9]. HAPSs are proposed to expand the terrestrial mobile coverage in 5G and beyond scenarios, providing high speed and ubiquitous mobile service [2], [10]–[12]. Indeed, the integration of these aerial platforms in Smart Urban Ecosystems looks promising to enhance Ultra Reliable Low Latency Communication (URLLC) and enhanced Mobile Broadband (eMBB) services, and HAPSs are expected to become a crucial knob to support Intelligent Transport Systems, defeat coverage holes, and cover unplanned user events [2]. Furthermore, few studies investigate their key role in enabling massive Internet of Things (IoT) communication and in supporting massive access and management services for both ground and air devices, including Unmanned Aerial Vehicles (UAVs) that are exploited for large-scale monitoring purposes and cargo delivery tasks [2], [12]–[14]. Other works focus on analysing the contribution offered by the HAPSs serving as distributed aerial Data Centers, thanks to their caching and computational capabilities, to support content delivery functionalities and Multiaccess Edge Computing (MEC) in terrestrial networks [15], [16]. In addition, HAPS potential is emerging as alternative backhauling for isolated BSs as well as to handle LEO satellites handoffs, yielding seamless connectivity [2], [17], [18]. These studies confirm the crucial role that aerial platforms may play in a fully integrated three-layer Vertical Heterogeneous Network (VHetNet), that is currently under discussion as a cornerstone in the 6G network architecture, where aerial nodes represent an essential intermediate network level between the terrestrial network and the satellite network [18], [19]. Nevertheless, little attention has been devoted to investigate the potential of HAPSs to promote a sustainable deployment of future communication networks. Our previous work [20] presents a preliminary case

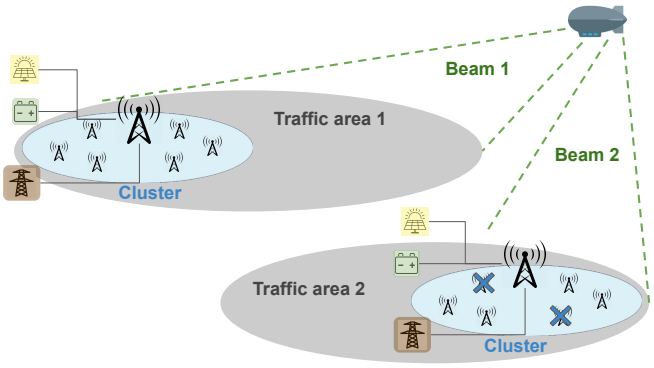


Fig. 1: Hybrid network scenario

study to show the potential benefits yielded by integrating HAPSs in a urban RAN to enhance energy consumption and efficiency of network operation. However, to the best of our knowledge, the current work is the first study that specifically and extensively investigates how the integration of HAPSs in dense 5G RANs can not solely enhance the bandwidth availability, but also, (i) contribute to lowering the energy demand from the grid and the related operational cost for Mobile Network Operators (MNOs), (ii) provide additional flexibility to resource allocation, (iii) reduce the required on-ground RE supply, (iv) pave the way for a feasible and sustainable evolution towards 6G networks.

In further detail, in this paper we investigate the integration of an aerial BS in a portion of an urban mobile access network with the following objectives. First, we analyse how the integration of HAPSs can limit the need for installing additional permanent on-ground RAN infrastructures that are required by the densification processes. Second, we evaluate the contribution of aerial BSs to limit the required on-ground RE supply sizing in those dense renewable powered RANs deployed in urban areas, where physical and regulatory constraints do not allow to install huge RE generators. Third, we examine the potential benefits of exploiting HAPS offloading in RANs to restrain the astounding growth of mobile network energy demand from the power grid and the related operational costs. Finally, we evaluate the flexibility provided by the HAPS bandwidth distribution in space and over time to trade off Quality of Service (QoS) and grid energy reduction, depending on the traffic scenario.

II. Combining aerial and terrestrial BSs

This paper studies a portion of a urban RAN where different areas can be identified depending on the data traffic patterns, as depicted in Fig. 1. The mobile access is yielded by a number of clusters that are composed by one macro cell BS and six micro cell BSs. A HAPS mounted Super Macro Base Station (SMBS) offers additional capacity to the RAN, by means of multiple spot beams, each forming a cell that covers a different traffic zone [21]. The SMBS is assumed to be self-sustainable, since it can be powered by a set of photovoltaic (PV) panels that can be hosted on the huge HAPS surface, together with some storage units. Since no reliable power model is currently available to represent the energy consumption dynamics of future SMBSs, our work does not focus on modeling the HAPS energy demand and its RE generation system, assuming, as usually done in the literature, that sufficient energy supply is constantly available to satisfy the SMBS demand [2]. Besides the electric grid, each cluster is powered by PV panels, and it is equipped with some units of lead-acid batteries to store RE that is not immediately used. Real RE generation profiles are obtained from PVWatts for the city of Milan during the typical meteorological year [22]. In our study, the linear model proposed in [23] is applied to compute the power required to operate each BS.

We consider 16 different zones in the city of Milan, whose traffic patterns are derived from real data from 2015, provided by a large Italian MNO, with granularity of 15 minutes. The traffic traces are scaled up to account for the upsurge of mobile traffic demand, leading to a 4-fold data traffic increase over the past 5 years in Western Europe [24]. We hence assume that the peak of each traffic pattern corresponds to the maximum bandwidth capacity of each BS, i.e., 150 Mbps. This is a pessimistic assumption with respect to energy saving potential, since the network capacity is typically overprovisioned. A representative subset of traffic areas is selected to be shown in the results. Our study focuses on the operation of a sample cluster for each traffic area.

A RoD strategy is applied to deactivate unneeded on-ground BSs when the traffic load is low, as depicted in Fig. 1 in the Traffic area 2. The on-ground BS clusters can be partially unloaded by transferring some traffic to the HAPS SMBS (*HAPS Offloading, HO*), thanks to direct communication with end user devices, that is enabled by the relatively low HAPS altitude along with the limited line-of-sight related channel attenuation, with no need for additional infrastructure relay [21], [25]. In case of saturation of the HAPS cell capacity, some more traffic can be moved from an on-ground micro BS to the macro BS, provided that the macro BS still yields sufficient residual capacity and the micro BS load is lower than a threshold ρ_{min} set according to the value suggested in [26], i.e., 0.37 of the maximum BS load. This specific threshold value has been proved to discriminate when it results more convenient to carry a traffic unit in the macro cell rather than in the micro cell, under the considered BS consumption model [26]. After these offloading operations, all the micro BSs that are not currently carrying any traffic can be switched off.

III. HAPS bandwidth allocation strategies

Different HAPS bandwidth allocation strategies are implemented to determine the portion of HAPS capacity that is made available for each cell covering a given area. The strategies define at each time slot t (in the numerical results we consider time slots of 1 h) the capacity $C^H_{i,t}$ to allocate to a cell i. The notation adopted in this paper is summarized in Table I.

A. Single HAPS cell operation

Constant HAPS cell capacity (CHC) According to this offloading strategy, the capacity available per HAPS cell is set a priori and maintained constant during the strategy operation,

$$C^H_{i,t} = C^H_i = C^{H^*} \quad \forall t$$

TABLE I: Notation

$C^H_{i,t}$	HAPS capacity allocated to cell i during time step t
C^{H^*}	Reference value of the HAPS capacity allocated to cell i
F_i	Scaling factor by which C^{H^*} is multiplied to obtain the actual value of $C^H_{i,t}$
D_i	Traffic load of the BS cluster served by HAPS cell i in time step t
D^{th}_i	Threshold on the traffic load of the BS cluster served by HAPS cell i, based on which $C^H_{i,t}$ is set to different values
E^C_i	Energy demand of the BS cluster served by HAPS cell i per time step t
E^R_i	Residual energy demand of a BS cluster served by HAPS cell i, in time step t, that cannot be satisfied by RE
E^{PV}_i	Amount of RE (either currently produced by the PV panel or previously stored in the battery) available at time step t to power the BS cluster served by HAPS cell i
G_R	Grid Energy Reduction
C_S	Cost Saving
R	Capacity/Demand Ratio
N^{BS}	Average number of BSs that can be deactivated in a cluster thanks to the traffic offloaded from the micro BSs to the HAPS
f_H	Fraction of the traffic volume of the BS cluster that is offloaded to the HAPS
C^{BS}	Capacity offered by a BS, i.e. 150 Mbps

where C^{H^*} is a reference value.

Variable HAPS cell capacity with Fixed Threshold (FTS) According to this approach, when the traffic demand in the served cluster is low, the available bandwidth provided by the HAPS cell can be decreased, so that additional HAPS bandwidth can be made available to other traffic areas characterized by higher traffic demand or featuring low RE availability.

The capacity available per HAPS cell is hence dynamically modified based on the load in the served cluster, denoted by D_i. Given a threshold D^{th} on the load, the value of C^H_i is reduced by a scaling factor F with respect to the baseline value C^{H^*}, when the load is below the threshold:

$$C^H_i = F \cdot C^{H^*} \quad \text{if} \quad D_i < D^{th}$$

Variable HAPS cell capacity with Dynamic Threshold (DTS) The value of C^H_i is modulated over time based on dynamic values of the threshold and the scaling factor which, in their turn, depend on the current cluster energy demand, RE availability and electricity prices.

Let D^{th} and F take N different discrete values between 0 and 1, denoted by $D^{th}(n)$ and $F(n)$, respectively, with $n = 1, 2, ..., N$. Let also the energy demand per cluster be discretized in N energy consumption levels, $L^E(n)$ (with $n = 1, 2, ..., N$), within a range defined between the minimum and maximum values of energy demand that are identified among all the clusters. Note that the cluster energy demand per time step, denoted by E^C, is defined as the energy demand assuming that all the BSs within the cluster are always active, no local RE is produced and no HO is applied. Similarly, the electricity price values are quantized in N price levels, $L^P(n)$ (with $n = 1, 2, ..., N$), having the same step size, within a range defined between the minimum and maximum values of all the price values observed during the considered simulation period. At time step t, in traffic zone i, the value of the threshold D^{th}_i is set as:

$$D^{th}_i = \alpha D^{th}(n) + (1-\alpha) D^{th}(m) \tag{1}$$

where n is the index of the current cluster energy demand level at time step t and m is the index of the price level in which the current electricity price falls. α corresponds to a weight parameter that can be arbitrarily set, depending on whether energy availability or electricity prices, respectively, should be given the largest influence in determining the threshold value. Let us denote by E^R_i the residual energy demand of a cluster, in time step t, that cannot be satisfied by RE:

$$E^R_i = max(0, E^C_i - E^{PV}_i) \tag{2}$$

where E^{PV}_i is the amount of currently available RE, including both the RE currently produced by the PV panel and the RE previously stored in the battery. And, also in this case, consider the quantization of E^R in N step and denote it by $E^R(n)$. The setting of the scaling factor is then:

$$F_i = 0 \quad \text{if} \quad E^R_i = 0$$
$$F_i = F(n) \quad \text{if} \quad E^R_i > 0$$

where n corresponds to the index of the current level of residual energy demand $E^R(n)$. The idea is that if the currently available RE is larger than the cluster energy demand, it is not necessary to offload any traffic to the HAPS, given the full self-sustainability of the cluster without the need for drawing energy from the grid.

Benchmark Allocation Strategy (BAS) This strategy is introduced as a benchmark against which the performance of CHC, FTS and DTS approaches is compared. This algorithm aims at dynamically allocating to a given area an amount of HAPS bandwidth that can satisfy the largest possible fraction of current cluster demand, in order to maximize the exploitation of the HAPS and minimize the on ground energy demand. The HAPS cell capacity at time step t is hence set as follows:

$$C^H_i = min(C^{H*}, D_i) \tag{3}$$

In addition, a portion of the remaining on-ground traffic can be further moved to the macro BS from some micro BSs, that can be unloaded and possibly deactivated. The residual cluster energy demand can be satisfied by RE, if available, and by the energy drawn from the grid. This approach allows to minimize the energy amount drawn from the grid at each time step.

B. Joint operation of multiple HAPS cells

The overall SMBS capacity can be shared by the various HAPS cells, entailing a flexible and possibly uneven bandwidth distribution among them. The *Jointed Bandwidth Allocation Algorithm (JBA)* dynamically manages the HAPS bandwidth allocation jointly considering two different traffic zones. Let us introduce the following notation: D_i, which is

the current traffic load in zone i; $D_i^{th^*}$, which is the value set for the threshold D^{th} in the traffic zone i at a given time step t, defined according to (1); $F_{i,t}^*$, which is the value of F at a given time step in zone i, derived according to the procedure described in Sec. III-A for the DTS approach. At each time step t, the scaling factor in traffic zone i, F_i, is set according to the criteria summarised in Table II. Note that, unlike under FTS and DTS approaches, the portion of bandwidth actually assigned to either of the two HAPS cells may exceed C^{H^*}. Basically, when the traffic is low in both areas, the scaling factor is set separately for each area according to the DTS strategy, since it is assumed that the need for the HAPS contribution is limited given the low mobile demand. When the traffic is unbalanced, lower than the threshold in one area and larger in the second area, some additional capacity can be transferred from the low traffic area to the high traffic area, in order not to waste HAPS capacity that could be profitably exploited in a more loaded traffic area, instead of uselessly overprovisioning a less demanding zone. When both areas are undergoing high traffic, capacity is activated in proportion to the traffic.

TABLE II: JBA configuration settings

Condition	$\mathbf{F}_1$	$\mathbf{F}_2$
$D_1 < D_1^{th^*}$, $D_2 < D_2^{th^*}$	F_1^*	F_2^*
$D_1 < D_1^{th^*}$, $D_2 \geq D_2^{th^*}$	F_1^*	$2 - F_1^*$
$D_1 \geq D_1^{th^*}$, $D_2 < D_2^{th^*}$	2-F_2^*	F_2^*
$D_1 \geq D_{1,t}^{th^*}$, $D_2 \geq D_2^{th^*}$	$\frac{2D_1}{D_1+D_2}$	$\frac{2D_2}{D_1+D_2}$

IV. Key Performance Indicators

The following Key Performance Indicators (KPIs) are introduced to measure the system performance:

a. Grid Energy Reduction (G_R) is the fraction of energy drawn from the grid that can be saved in the considered scenario with respect to a baseline scenario in which there is no HAPS and no RE is locally produced at the BSs of the cluster. Two different types of baseline scenarios are considered: (i) no RoD strategy is applied, hence all the cluster BSs are always active; (ii) RoD is applied without the contribution of HAPS offloading, hence the traffic is shifted to the macro BS only.

b. Cost Saving (C_S) is the fraction of cost reduction that is obtained under the considered system configuration with respect to the cost in the baseline scenario (ii).

c. Capacity/Demand Ratio (R) is the average ratio between the overall offered capacity in a given traffic area (including both the on-ground BS cluster and the HAPS cell, when available) and the total traffic demand in the considered traffic area. This indicator assesses the effectiveness of allocating resources on demand.

d. Number of deactivated BSs (N^{BS}) is the average number of micro BSs that can be deactivated in a cluster thanks to the traffic offloaded from the micro BSs to the HAPS;

e. HAPS Traffic Fraction (f_H) is the fraction of the traffic volume in a given traffic area that is offloaded to the HAPS and handled by the SMBS. This parameter represents a measure of the operational contribution provided by the HAPS in managing the traffic load.

To properly tune the parameters of the HAPS bandwidth allocation strategies, we formulate various performance indicators as a function of D^{th}, C^H and F, so that the strategy settings can be defined based on the performance obtained under different parameter configurations. We model the cluster demand with a random variable, D, characterized by a uniform distribution $\mathcal{U}_{[0,C^{max}]}$, where the value of C^{max} corresponds to the maximum capacity provided by the on-ground BS cluster, i.e. $(m+1)C^{BS}$, with m representing the total number of micro BSs in the cluster (in our study m=6) and C^{BS} corresponding to the BS capacity, i.e. 150 Mbps. We derive the formulation of the following performance metrics.

a) Number of deactivated micro BSs under HO, N^{BS}:

$$N^{BS} = min(m, \lfloor \frac{f_d C^H (m+1)}{d} \rfloor)$$

where d is the considered traffic demand, and the scaling factor f_d is equal to F if $d < D^{th}$, 1 otherwise, i.e. $f_d = [F \cdot \Theta(D^{th} - d) + \Theta(d - D^{th})]$.

b) Average grid energy reduction, G_R:

$$G_R = \int_D \frac{N_d^{BS}(P_0^m + \frac{d}{m+1}P_\Delta^m)}{P_0^M + \frac{d}{m+1}P_\Delta^M + mP_0^m + \frac{md}{m+1}P_\Delta^m}$$

where P_0^m and P_0^M represent the fixed component of the power consumption of each micro or macro BS, respectively, whereas P_Δ^m and P_Δ^M represent the traffic proportional component of the power consumption of each micro or macro BS, respectively.

c) Capacity/demand ratio, R:

$$R = \int_D \frac{f_d C^H + (m + 1 - N^{BS})C^{BS}}{d}$$

V. Integrating HAPSs to limit RE supply sizing

Assuming a *uniform* distribution of the SMBS bandwidth among the various traffic zones the available capacity per cluster, C^{H^*}, is equal to 150 Mbps. Our simulations are run over a period of two months. We first investigate the potential of integrating a HAPS mounted SMBS in a dense RAN to increase the available bandwidth and limit the dimensioning of the required on-ground RE supply. Fig. 2 reports the values of capacity/demand ratio R versus the reduction of energy from the grid G_R under CHC strategy for a subset of the traffic areas, represented in different colors. G_R is here computed with respect to the grid energy demand observed in the scenario in which all the BSs are always active. Both the baseline case in which HAPS offloading is coupled with RoD without the presence of any on-ground RE supply (subplot on the left) and the cases with variably sized RE generators (last three subplots) are represented: (i) Small size: S_P = 2 kWp, S_B = 2 kWh; (ii) Medium size: S_P = 4 kWp, S_B = 4 kWh; (iii) Large size: S_P = 8 kWp, S_B = 8 kWh. Each dot in the plots corresponds to a different C^H setting, ranging from 0.25 C^{H^*} to 1.5 C^{H^*}, with increasing step width of 0.25 C^{H^*}. In the baseline case, as C^H becomes higher, both R and G_R grow larger. In those areas featuring bursty traffic, with

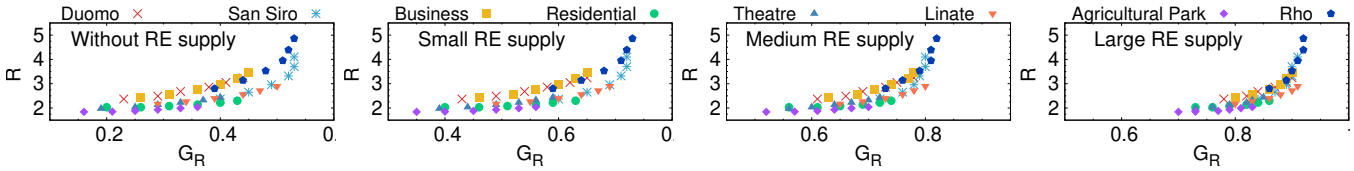


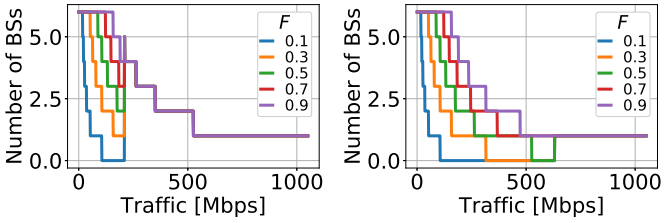
Fig. 2: Capacity/demand ratio, R, versus reduction of energy drawn from the grid, G_R, under different values of C^H for different traffic areas, without RE as well as under differently dimensioned on-ground RE generators.

peaks observed in relation to specific occasional events, like in the areas called San Siro and Rho, further raising C^H over the value of C^{H^*} does not provide significant G_R increase. Assuming the presence of a local RE generator to power the cluster BSs, the values of grid energy reduction tend to progressively shift towards larger values as the RE system size becomes larger, whereas R is not affected, since the decisions to switch BSs on/off according to the RoD strategy are taken solely based on the traffic demand, without considering the RE availability. A small on-ground RE supply is sufficient to obtain G_R over 35% in any traffic zone. By increasing the size of the RE generator, increasingly higher G_R can be achieved, resulting as high as more than 90% in the most favored areas. However, even C^H values as low as 0.25 C^{H^*} are effective in achieving more than 70% in any traffic zone. When HAPS offloading is combined with the large on-ground RE generator, a very limited value of C^H is sufficient to achieve large grid energy saving in any traffic zone.

This analysis highlights how, under the same C^H settings, the grid energy reduction may significantly vary depending on the traffic patterns. Hence, to achieve similar target performance C^H should be properly tuned based on the local traffic demand. These aspects should be taken into account when configuring the HAPS bandwidth allocation. In particular, a slightly larger RE supply could be preferably placed in a zone with lower traffic demand to reduce the required value of C^H to obtain the same grid energy reduction. Some HAPS bandwidth can hence be spared and made available over other traffic zones that feature higher peak demand and in which tighter physical constraints may prevent the installation of a RE supply.

VI. Impact of Configuration Settings

Fig. 3 reports the number of deactivated BSs, N_d^{BS}, for increasing traffic demand, with each curve representing different settings of F. Two different values of D^{th} are considered, i.e. 0.2 (Fig. 3a) and 0.6 (Fig. 3b). For traffic demand higher than D^{th}, N_d^{BS} tends to decrease similarly for any value of F as the traffic demand increases. The traffic step width increases with the demand growth, since the traffic is distributed uniformly among the BSs, hence the traffic volume that must be offloaded from each BS to enable the switch off of a single BS becomes larger as the cluster traffic demand increases. If the traffic demand is below D^{th}, lower values of the scaling factor F determine smaller values of N_d^{BS} for a given traffic demand, with a steeper descent of the curve, since the lower available HAPS bandwith allows to deactivate less BSs. Under higher setting of D^{th}, the transition point from a regime in which N_d^{BS} is affected by F to a regime where N_d^{BS} is not influenced by F is shifted towards higher values of traffic demand.



(a) D^{th}=0.2 (b) D^{th}=0.6

Fig. 3: Number of deactivated micro BS, N_d^{BS}, versus traffic demand under different settings of D^{th} and F (C^H=150Mbps).

Fig. 4 depicts the average grid energy reduction, G_R, and the average capacity/demand ratio, R, under different settings of D^{th} and F, with C^H=150Mbps. As shown in Fig. 4a, G_R increases as F grows larger, and higher values of D^{th} lead to larger grid energy reduction, up to 40%. The impact of D^{th} is more evident under low F. Conversely, the capacity/demand ratio, R, depicted in Fig. 4b, achieves the highest values for low F, rapidly decreasing as F grows larger, to increase again as F approaches to 1. This trend, that is especially evident under D^{th} as low as 0.1, depends on the combination of the cluster capacity and HAPS bandwidth, both varying not linearly with F. In general, as D^{th} becomes higher, R is increased, although the growth becomes negligible under high values of D^{th}.

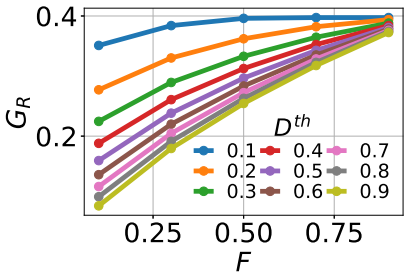
Based on these findings, the configuration parameters can be set so as to trade off grid energy reduction, capacity/demand ratio, and the need for sparing HAPS cell bandwidth to be redirected to other cells that cover more loaded traffic zones.

VII. Joint Operation of Multiple HAPS Cells

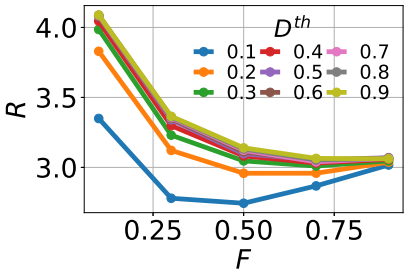
Considering now a scenario in which two traffic areas are served by two HAPS cells, the system performance is investigated under JBA strategy and compared against FTS and DTS approaches. In case of RoC, 3 kWp PV panel capacity and 4 kWh battery size are assumed for the RE generator.

A. Performance evaluation

Based on the results presented in Sec. VI, under FTS we set D^{th}=0.15 and F=0.5, to trade off the grid energy reduction, G_R, versus QoS, i.e. R. Under DTS we assume $\boldsymbol{D}^{th}$=[0.05, 0.15, 0.25, 0.35] and $\boldsymbol{F}$=[0.25, 0.50, 0.75, 1.00], to better trade off, depending on the traffic demand, between grid energy reduction and the need for keeping the HAPS cell bandwidth availability to a lower level whenever possible, so as to redirect the spared HAPS capacity to other cells. Finally, regarding the setting of α adopted for the DTS operation, in our simulations α=0.65, in order to assign energy availability



(a) G_R



(b) R

Fig. 4: Grid energy reduction, G_R, and capacity/demand ratio, R, under different D^{th} and F settings (C^H=150Mbps).

an almost two-fold larger influence with respect to electricity prices in determining the value of the threshold D^{th}.

Fig. 5 shows various KPIs to evaluate the system performance under disjointed and jointed operation, considering different couples of traffic areas, either assuming that RoC is operating in both areas (5a - light colored bars), in the Rho area only (5d, 5e), or in none of the areas (5a - dark bars, 5b, 5c). First, we investigate how JBA impacts the system performance with respect to operating separately the two HAPS cells, considering the Business and Residential areas, under different configuration settings. Fig. 5a shows the grid energy reduction, G_R, for the Business and Residential areas, assuming that the traffic areas are operated independently (scenario S, either under FTS or DTA) or jointly (scenario M, under JBA), with HO running alone (dark colored bars) or combined with RoC (light colored bars). Without any RE supply DTS provides higher G_R than FTS, similarly to JBA, that is only slightly less performing than DTS. With a local RE generator, JBA performs only slightly better than DTS. Similar savings are obtained in terms of cost (data non reported), with DTS and JBA achieving up to more than 30% higher cost saving compared to FTS. Although JBA performance results slightly worse than DTS in terms of G_R, similar gains are obtained in terms of cost saving under these strategies. Note that the cost reduction does not merely depend on the decrease of the overall amount of energy bought from the grid, but also on the timely reduction of grid energy demand, that is more convenient when the electricity prices are lower. Higher values of F under DTS and JBA highlight how the HAPS capacity is more exploited under these strategies rather than under FTS.

Fig. 5b-5c compare the system performance yielded by the proposed strategies CHC (yellow bars), FTS (orange bars), and DTS (red bars) against the performance obtained under BAS operation (green bars) in two sample areas, i.e. *Rho* and *Residential*. Considering the capacity demand ratio, R, reported in Fig. 5b, BAS yields the lowest value of R, that results decreased by up to more than 30% with respect to the other approaches, with a more significant gap in the Rho area. Conversely, from Fig. 5c we observe that the highest values of f_H are registered under CHC and in the Rho area. Note that (data not reported) up to almost 80% of operational cost can be saved under any strategy, with virtually undetectable differences between the various approaches. The combination of higher R and lower f_H under FTS and DTS suggests that, unlike BAS and CHC, more micro BSs on average are kept active, providing additional capacity without significantly affecting the energy cost during periods in which RE is available to power the on-ground nodes. At the same time, the HAPS capacity contribution can be reduced, and the spared bandwidth can be redirected to different areas.

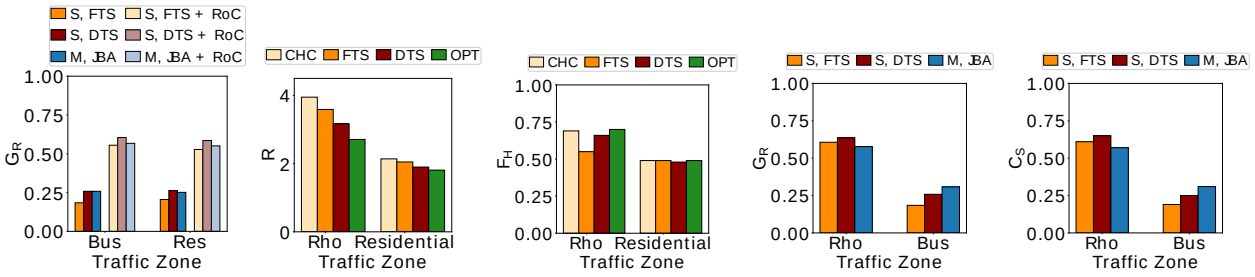
Our results show that a more efficient exploitation of the HAPS bandwidth and a better utilization of the available RE both contribute to enhance the QoS with respect to the BAS, providing significantly higher R without impairing the cost saving. This appears evident especially in those traffic area that result more underloaded, like the Rho area. In similar traffic areas, bandwidth resources tend to be overprovisioned on average, hence yielding a wider margin for an enhanced exploitation of the spared resources. However, properly designed strategies are required to achieve this objective, providing a better QoS without raising the energy bill for MNOs.

B. Uneven allocation of renewable energy supply

We now evaluate the JBA performance under uneven distribution of RE supply, considering the case in which only a traffic area is equipped with local RE supply. We consider a peripheral traffic zone, i.e. the Rho area, having bursty traffic and limitless space to freely install PV panels, together with a traffic zone in the city center, i.e. the Business area, characterized by huge traffic volumes, where physical constraints prevent the installation of a local RE generator. In the Rho area, the same RE system sizing adopted for the case analysed in Sec. VII-A is assumed. Real electricity prices from the year 2021 are considered. Fig. 5d-5e depict the system performance in the Rho and Business areas under FTS (orange bars), DTS (red bars) and JBA (blue bars). As reported in Fig. 5d, DTS slightly increases G_R in the Rho area, whereas JBA leads to lower grid energy saving with respect to FTS; in the Business area, the gain in terms of grid energy reduction under JBA is almost 70% higher than under FTS. As it can be observed from Fig. 5e, the cost saving, C_S, follows a trend that is similar to G_R under the different strategies. Although the joint operation of multiple areas under JBA allows to more effectively exploit the HAPS capacity where no RE is present, the aggregated cost over the two areas result very similar under all strategies. To enhance cost saving, further studies are required to find proper configuration settings for HAPS bandwidth allocation that optimize grid energy reduction and cost savings.

VIII. Conclusion

The integration of HAPS in terrestrial RANs introduces additional degrees of freedom to reduce grid energy and save



(a) Under/without RoC (b) R (c) f_H (d) G_R (e) C_S

Fig. 5: Performance under disjointed and jointed operation, considering different couples of traffic areas: with RoC operating in both areas (5a - light colored bars), in the Rho area only (5d, 5e), or in none of the areas (5a - dark bars, 5b, 5c).

cost, thus make 5G networks more sustainable. Even under minimal additional SMBS capacity, HO allows to limit the size of RE supply required to power terrestrial BSs, still saving grid energy and cost, and preserving QoS without the need for installing permanent underutilized on-ground network infrastructures: a pivotal benefit in view of the staggering growth of mobile traffic and the raise of electricity prices.

Properly designed strategies are crucial to flexibly adapt the available capacity to the actual traffic demand variations across different zones, and a proper setting configuration is essential, especially in highly loaded areas to efficiently exploit both HAPS capacity and RE. Our results highlight the remarkable benefits of HAPS offloading to enable a feasible deployment of 5G networks and pave the way to further investigating the potentiality of hybrid aerial-terrestrial RANs for a sustainable transition towards 6G networks. Future work is required to design enhanced allocation strategies to optimize the RE utilization over a wider time window, to better trade off energy cost, physical and CAPEX constraints, and QoS.

REFERENCES

[1] H. Pihkola, M. Hongisto, O. Apilo, and M. Lasanen, "Evaluating the energy consumption of mobile data transfer—from technology development to consumer behaviour and life cycle thinking," *Sustainability*, vol. 10, no. 7, 2018.

[2] G. K. Kurt, M. G. Khoshkholgh, S. Alfattani, A. Ibrahim, T. S. J. Darwish, M. S. Alam, H. Yanikomeroglu, and A. Yongacoglu, "A vision and framework for the high altitude platform station (haps) networks of the future," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2021.

[3] "What's up with Stratobus?," 2017. *Thales*. Accessed: 2021-04-29, https://www.thalesgroup.com/en/worldwide/space/news/whats-stratobus.

[4] F. Baurreau, R. Staraj, F. Ferrero, L. Lizzi, J.-M. Ribero, and J.-P. Chessel, "Stratospheric platform for telecommunication missions," in *2015 IEEE International Symposium on Antennas and Propagation USNC/URSI National Radio Science Meeting*, pp. 914–915, 2015.

[5] J. Wigard and I. Z. Kovacs, "Cell splitting for Non-Terrestrial Networks." International Patent WO 2020/094212 A1, May 14, 2020.

[6] M. M. Azari, S. Solanki, S. Chatzinotas, O. Kodheli, H. Sallouha, A. Colpaert, J. F. M. Montoya, S. Pollin, A. Haqiqatnejad, A. Mostaani, E. Lagunas, and B. Ottersten, "Evolution of non-terrestrial networks from 5g to 6g: A survey," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2022.

[7] H. Zhang, Z. Han, G. C. Alexandropoulos, and N. H. Tran, "Special issue on aerial access networks for 6g," *Journal of Communications and Networks*, vol. 24, no. 2, pp. 121–124, 2022.

[8] Y. He, D. Wang, F. Huang, R. Zhang, X. Gu, and J. Pan, "Downlink and uplink sum rate maximization for hap-lap cooperated networks," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2022.

[9] F. Rinaldi, H.-L. Maattanen, J. Torsner, S. Pizzi, S. Andreev, A. Iera, Y. Koucheryavy, and G. Araniti, "Non-terrestrial networks in 5g beyond: A survey," *IEEE Access*, vol. 8, pp. 165178–165200, 2020.

[10] Y. He, D. Wang, F. Huang, R. Zhang, X. Gu, and J. Pan, "Downlink and uplink sum rate maximization for hap-lap cooperated networks," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2022.

[11] Y. Hokazono, H. Kohara, Y. Kishiyama, and T. Asai, "Extreme coverage extension in 6g: Cooperative non-terrestrial network architecture integrating terrestrial networks," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 138–143, 2022.

[12] W. Wu, Q. Zhang, K. Wang, and W. Wang, "Towards Ubiquitous Coverage of High Altitude Platforms Aided 5G+ for Massive Internet of Things: A Cell Perspective," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 118–123, 2020.

[13] K. Xiao, C. Li, and J. Zhao, "Lstm based multiple beamforming for 5g haps iot networks," in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, pp. 1895–1900, 2019.

[14] W. Miao, C. Luo, G. Min, Y. Mi, and H. Wang, "Unlocking the potential of 5g and beyond networks to support massive access of ground and air devices," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 4, pp. 2825–2836, 2021.

[15] G. Vallero, D. Renga, and M. Meo, "Caching in the air: High altitude platform stations for urban environments," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2244–2249, 2022.

[16] D. S. Lakew, A.-T. Tran, N.-N. Dao, and S. Cho, "Intelligent offloading and resource allocation in hap-assisted mec networks," in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1582–1587, 2021.

[17] H. Kitanozono, J. Suzuki, Y. Kishiyama, Y. Hokazono, T. Sotoyama, M. Ouchi, R. Miura, and H. Tsuji, "Development of high altitude platform station backhaul system using 38ghz band frequency," in *2021 IEEE VTS 17th Asia Pacific Wireless Communications Symposium (APWCS)*, pp. 1–5, 2021.

[18] O. Kodheli, E. Lagunas, N. Maturo, S. K. Sharma, B. Shankar, J. F. M. Montoya, J. C. M. Duncan, D. Spano, S. Chatzinotas, S. Kisseleff, J. Querol, L. Lei, T. X. Vu, and G. Goussetis, "Satellite Communications in the New Space Era: A Survey and Future Challenges," *IEEE Communications Surveys Tutorials*, vol. 23, no. 1, pp. 70–109, 2021.

[19] X. Cao, P. Yang, M. Alzenad, X. Xi, D. Wu, and H. Yanikomeroglu, "Airborne communication networks: A survey," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 9, pp. 1907–1926, 2018.

[20] D. Renga and M. Meo, "Can high altitude platforms make 6g sustainable?," *IEEE Communications Magazine*, pp. 1–7, 2022.

[21] M. S. Alam, G. K. Kurt, H. Yanikomeroglu, P. Zhu, and N. D. ào, "High altitude platform station based super macro base station constellations," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 103–109, 2021.

[22] A. P. Dobos, *PVWatts Version 5 Manual*. Sep 2014.

[23] G. Auer, O. Blume, V. Giannini, I. Godor, M. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, *et al.*, "D2. 3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," *Earth*, vol. 20, no. 10, 2010.

[24] P. Cerwall *et al.*, "Ericsson mobility report june 2020," 2020.

[25] K. Hoshino, S. Sudo, and Y. Ohta, "A study on antenna beamforming method considering movement of solar plane in haps system," in *2019 IEEE Vehicular Technology Conference (VTC2019-Fall)*, pp. 1–5, 2019.

[26] M. Dalmasso, M. Meo, and D. Renga, "Radio resource management for improving energy self-sufficiency of green mobile networks," in *Performance Evaluation Review*, vol. 44, pp. 82–87, Sept 2016.