On Optimizing Time-, Space- and Power-Domain Energy-Saving Techniques for Sub-6 GHz Massive MIMO Base Stations

E. Peschiera*, Y. Agram†, F. Quitin†, L. Van der Perre*, and F. Rottenberg*
*ESAT-DRAMCO, Campus Gent, KU Leuven, Ghent, Belgium
†Brussels School of Engineering, ULB, Brussels, Belgium

***Abstract*—This paper addresses the optimized base station (BS) resource allocation strategy in a massive multiple-input multiple-output (MIMO)-orthogonal frequency-division multiplexing (OFDM) system, aiming to minimize energy usage while meeting fixed downlink user data rates. We explore whether it is better to save energy by minimizing the active time slots ("rush-to-sleep"), the active antennas ("rush-to-mute"), the transmit power ("awake-but-whisper"), or combining these approaches. We utilize a measurement-based parametric power consumption model of sub-6 GHz BSs. We show that the formulated problem can be optimally solved by exploiting its convexity. The performance analysis across different network loads suggests that a rush-to-mute is close-to-optimal at most network loads when the BS hardware does not include time-domain power-saving modes like micro-discontinuous transmission (μDTX). Median energy savings of 24% are achieved over the rush-to-sleep and awake-but-whisper at medium network loads. With enabled time-domain hardware power-saving modes, operating in the three energy-saving domains is the optimal strategy. The average power consumption decreases and median energy savings against the three specific schemes reach 18% at medium network loads.**

***Index Terms*—Resource allocation, power consumption, energy saving, green communications, massive MIMO.**

I. Introduction

The absolute energy consumption of fifth-generation (5G) networks is raising concerns in terms of environmental footprint and operational costs for operators [1]. Recent studies show that the average power consumption of current 5G base stations (BSs) is two to three times higher compared to fourth-generation (4G) BSs [2], [3]. Moreover, the large traffic-independent power consumption observed in current BSs prevents a substantial decrease of the consumption when the network is lightly loaded, which occurs frequently (for example during night time) [4]. This has led to research and standardization efforts towards network energy-saving techniques, with the aim of drastically reducing the BS energy consumption at low and medium traffic loads [5]–[7].

Time, frequency, space and power have been identified as the four major domains where energy-saving techniques can be applied [5]–[7]. Time-domain techniques include hardware deactivation at the microsecond scale such as micro-discontinuous transmission (μDTX) [8], as well as at longer time scales (up to seconds) with the implementations of BS advanced sleep modes [9]. The lean design of 5G New Radio (NR) also contributes to saving energy thanks to shorter signaling durations [6]. The complementary deactivation of 4G and 5G BSs based on measurements is proposed in [10]. The information-theoretic work in [11] has considered a multi-time slot system and derived the optimal number of active time slots under an average rate constraint over the frame. In the frequency domain, energy-saving techniques can leverage the long-term traffic load variations in a cell to shutdown entire component carriers [7]. These are left out of the scope of this paper. Spatial-domain techniques generally refer to adapting the number of active antennas in massive multiple-input multiple-output (MIMO) or the number of active BSs in multi-cell systems. Authors in [12], [13] minimize the absolute BS consumption by optimizing the number of utilized antennas. Our previous work [14] provides the optimal number of active antennas that minimize BS power consumption in a single-time slot system by including a model for the power amplifier (PA) operational efficiency in addition to circuit consumption. Power-domain techniques usually adapt the transmission power to achieve objectives such as improving the PA efficiency [15]. Recent works started to consider multi-domain resource allocation, *e.g.*, [16] that jointly optimizes the use of spectral and spatial resources.

A fundamental question is not addressed by previous works: is it more effective to reduce the use of time slots, antennas, or lower the transmit power in order to minimize the absolute BS consumption while satisfying a fixed rate constraint? In this paper we provide an answer to this question by considering a BS power consumption model (Section II) validated on operator measurements and manufacturers documentation [2]. The formulated optimization problem (Section III) considers fixed target user rates and per-antenna power constraints, and we show that it is reduced to a two-dimensional convex differentiable problem that has to be solved only when the channel second-order statistics or the users' target rates change. We then discuss (Section IV) the best energy-saving strategy at different network loads for two massive MIMO configurations.

II. Transmission and Power Consumption Models

A. Transmission Model

Let us consider a BS equipped with M antennas serving K single-antenna users in downlink using space-division multiple access (SDMA). The BS uses orthogonal frequency-division

This work is supported by the JU-SNS SUSTAIN-6G project under the European Union's Horizon Europe program (Grant Agreement No. 101191936).

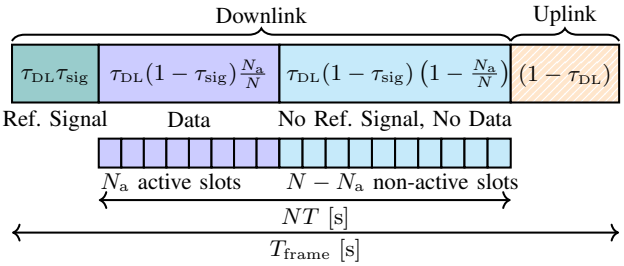

Fig. 1. Frame structure in time with time ratios of each phase.

multiplexing (OFDM) with Q subcarriers carrying users' symbols. The OFDM symbol duration including the cyclic prefix is denoted by T. Without loss of generality, we consider the transmission of one frame lasting T_{frame} seconds and including a downlink and an uplink phase. The downlink phase is further subdivided into three phases: a) reference signal transmission, b) data transmission, and c) no transmission. The corresponding time ratios of each phase relative to T_{frame} are illustrated in Fig. 1, where τ_{DL} is the downlink to frame ratio and τ_{sig} is the reference signaling to downlink ratio. In the following, we set $\tau_{\mathrm{DL}} = 0.75$ and $\tau_{\mathrm{sig}} = 1/14$ [2]. There are N available OFDM symbols (referred to as time slots) that carry users' data and we consider that N_{a} are active.

Similarly, we consider that M_{a} out of M antennas are active, where $K \leq M_{\mathrm{a}} \leq M$. The transmitted symbols, assumed uncorrelated among different users, form the vector $\mathbf{s} \in \mathbb{C}^{K\times 1}$ at one subcarrier and active time slot. The transmitted symbol of user k at every subcarrier and active time slot has zero mean and variance p_k. The channel matrix between the active antennas and the users at one subcarrier and active time slot is given by $\mathbf{H} \in \mathbb{C}^{K\times M_{\mathrm{a}}}$. Every element in the k-th row of $\mathbf{H}$ is drawn from a zero-mean circularly-symmetric complex-Gaussian distribution with variance β_k, *i.e.*, the large-scale fading coefficient of user k. We assume that per-subcarrier zero-forcing (ZF) precoding is used, giving the precoded signal vector $\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{s} \in \mathbb{C}^{M_{\mathrm{a}}\times K}$, where $(\cdot)^{\mathrm{H}}$ indicates Hermitian transpose. Considering a unitary inverse fast Fourier transform (IFFT), the average total transmitted powers in the time and frequency domains are equivalent. The latter is given by $P_{\mathrm{T}} = \mathbb{E}\left\{\mathrm{tr}\left[\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{s}\mathbf{s}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{H}\right]\right\} = \mathrm{tr}\left[\mathbb{E}\left\{(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\right\}\mathbb{E}\left\{\mathbf{s}\mathbf{s}^{\mathrm{H}}\right\}\right]$. Using the uncorrelation of users' symbols and the expression of the mean of a complex inverse-Wishart distribution [17], we obtain

$$P_{\mathrm{T}} = \frac{1}{M_{\mathrm{a}}-K}\sum_{k=1}^{K}\frac{p_k}{\beta_k}, \quad P_{\mathrm{a}} = \frac{P_{\mathrm{T}}}{M_{\mathrm{a}}} \tag{1}$$

where P_{a} is the average transmit power at each active antenna. We assume that the received signal is corrupted, after OFDM demodulation, by additive zero-mean circularly-symmetric complex-Gaussian noise where σ_k^2 is the noise power at user k. The use of ZF precoding implies that the bit rate delivered in downlink to user k averaged over the frame equals [18]

$$R_{k,\mathrm{deliv}} = \frac{Q}{T}\tau_{\mathrm{DL}}\left(1-\tau_{\mathrm{sig}}\right)\frac{N_{\mathrm{a}}}{N}\log_2\left(1+\frac{p_k}{\sigma_k^2}\right) \quad \text{[bits/s]}. \tag{2}$$

In the following, we consider that each user has a normalized per-subcarrier average target rate (in bits per OFDM symbol and per subcarrier) $R_k = R_{k,\mathrm{deliv}}/\left(\frac{Q}{T}\tau_{\mathrm{DL}}\left(1-\tau_{\mathrm{sig}}\right)\right)$.

B. Power Consumption Model

Let us consider the power consumption model [2] that expresses the average BS consumption over the frame as

$$P_{\mathrm{cons}} = \frac{\bar{P}_{\mathrm{PA}}}{\eta_{\mathrm{s/c}}^{\mathrm{PA}}} + \frac{\bar{P}_{\mathrm{AFE}}}{\eta_{\mathrm{s/c}}^{\mathrm{AFE}}} + \frac{\bar{P}_{\mathrm{DBB}}}{\eta_{\mathrm{s/c}}^{\mathrm{DBB}}} \tag{3}$$

where $\bar{P}_{\mathrm{PA}}$, $\bar{P}_{\mathrm{AFE}}$, and $\bar{P}_{\mathrm{DBB}}$ are the average powers consumed by the PAs, analog front-end (AFE), and digital baseband (DBB) across the frame, while $\eta_{\mathrm{s/c}}^{\mathrm{PA}}$, $\eta_{\mathrm{s/c}}^{\mathrm{AFE}}$, $\eta_{\mathrm{s/c}}^{\mathrm{DBB}} \in (0,1]$ are the supplying and cooling efficiencies of the PAs, AFE, and DBB.[1] The energy consumption in the frame is given by $T_{\mathrm{frame}}P_{\mathrm{cons}}$. Every term in (3) considers active and non-active components (*e.g.*, active and non-active PAs) that can be in different modes (from working to sleep). The model can be adapted to the scenario of this paper and the dependency on the considered optimization variables (*i.e.*, N_{a}, M_{a}, P_{a}) made explicit. By observing that $\bar{P}_{\mathrm{PA}}$ depends on the active time slots N_{a}, active antennas M_{a} and transmit power P_{a}, $\bar{P}_{\mathrm{AFE}}$ depends on M_{a} and $\bar{P}_{\mathrm{DBB}}$ does not depend on N_{a}, M_{a} or P_{a}, we can express

$$P_{\mathrm{cons}} = \frac{N_{\mathrm{a}}}{N}M_{\mathrm{a}}\left(\frac{P_0}{M}+\gamma P_{\mathrm{a}}^{\alpha}\right) + \frac{M_{\mathrm{a}}}{M}P_1 + P_{\mathrm{sleep}} \tag{4}$$

where P_0, P_1, P_{sleep}, γ are non-negative parameters, and $\alpha \in [0.5, 1]$. Given the space limits of this manuscript, we refer to the extended journal version [19] for the full expressions of γ, P_0, P_1 and P_{sleep}, but we summarize hereinafter their dependencies. The working-mode PA consumption is captured by $\gamma P_{\mathrm{a}}^{\alpha}$, while P_0 depends on PA discontinuous transmission (DTX) mode. Then, P_1 includes terms related to both PA and AFE non-working modes, while P_{sleep} contains the PA and AFE sleep-mode consumptions and the DBB consumption.

The model [2] utilizes reduction factors in $[0,1]$ to quantify the energy savings in each non-working mode. In this work we specifically address the reduction factors related to PA DTX mode ($\delta_{\mathrm{PA}}^{\mathrm{dtx}}$) and AFE idle mode ($\delta_{\mathrm{TRX}}^{\mathrm{idle}}$). Also, [2] is parametric such that its parameters can be changed to reflect different BS implementations. We compute the parameters of (4) by focusing on two 5G NR massive MIMO configurations (32T32R and 64T64R, with 32 or 64 TX and RX chains, PAs and antennas), utilizing the fitted parameters in [2, Table II] and varying the values of $\delta_{\mathrm{PA}}^{\mathrm{dtx}}$ and $\delta_{\mathrm{TRX}}^{\mathrm{idle}}$. This leads to two cases:

1) Time-domain hardware power-saving modes disabled, with $\delta_{\mathrm{PA}}^{\mathrm{dtx}} = 1$ and $\delta_{\mathrm{TRX}}^{\mathrm{idle}} = 1$ as in [2]. This means that the PA does not implement μDTX and the AFE does not save power in idle mode.
2) Time-domain hardware power-saving modes enabled, with $\delta_{\mathrm{PA}}^{\mathrm{dtx}} = 0.25$ and $\delta_{\mathrm{TRX}}^{\mathrm{idle}} = 0.5$. This means that PA μDTX and AFE idle-mode power savings are implemented.

The obtained values of the parameters are given in Table I.

[1]The AFE includes the transmitting (TX) and receiving (RX) radio-frequency chains, as well as analog hardware that performs per-cell signal processing. The DBB performs operations such as channel coding/decoding, modulation/demodulation, MIMO precoding/decoding.

TABLE I
PARAMETERS OF POWER CONSUMPTION MODEL (4) FOR THE TWO 5G NR MASSIVE MIMO CONFIGURATIONS.

Configuration	M	K	f_c [GHz][1]	B [MHz][2]	$P_{\max}$ [W][3]	α	γ	P_0 [W][4]	P_1 [W][4]	P_{sleep} [W]
32T32R	32	8	3.5	100	6.250	0.75	3.68	{0, 47.55}	{257.30, 120.86}	476.59
64T64R	64	8	3.5	100	3.125	0.75	3.50	{0, 53.92}	{341.57, 161.90}	550.23

[1]Carrier frequency [2]Bandwidth [3]Per-antenna maximum transmit power (considering 8 dB output power back-off)
[4]Left value is obtained with $\delta_{\text{PA}}^{\text{dtx}}=1$ and $\delta_{\text{TRX}}^{\text{idle}}=1$ as in [3], right value is obtained with $\delta_{\text{PA}}^{\text{dtx}}=0.25$ and $\delta_{\text{TRX}}^{\text{idle}}=0.5$

$$\min_{\substack{0\le N_{\text{a}}\le N\\ K\le M_{\text{a}}\le M}} P_{\text{cons}}=\frac{N_{\text{a}}}{N}\frac{M_{\text{a}}}{M}P_0+\frac{N_{\text{a}}}{N}M_{\text{a}}\gamma\left(\frac{1}{M_{\text{a}}(M_{\text{a}}-K)}\sum_{k=1}^{K}\frac{\sigma_k^2}{\beta_k}\left(2^{R_k\frac{N}{N_{\text{a}}}}-1\right)\right)^{\alpha}+\frac{M_{\text{a}}}{M}P_1+P_{\text{sleep}}\quad\text{s.t. }(\mathcal{C}_{P_{\max}}) \tag{7}$$

III. Optimal Resource Allocation in Time-, Space-, and Power-Domain

The problem that we consider in this paper is finding the optimal N_{a}, M_{a} and P_{a} that minimize the average BS consumption under per-user rate constraints

$$\min_{N_{\text{a}},M_{\text{a}},P_{\text{a}}} P_{\text{cons}}\text{ s.t. }R_k=\frac{N_{\text{a}}}{N}\log_2\left(1+\frac{p_k}{\sigma_k^2}\right),\ k=1,\dots,K. \tag{5}$$

where $0\le N_{\text{a}}\le N$, $K\le M_{\text{a}}\le M$ and $0\le P_{\text{a}}\le P_{\max}$. In the next subsections, we reformulate problem (5) as a two-dimensional one, solve it by exploiting its convexity, and identify the three energy-saving schemes that minimize the use of each of the three BS resources individually.

A. Optimal Transmit Power at Active Antennas

The user rate constraint in (5) can be used to express P_{a} as a function of N_{a} and M_{a}. We first compute the power allocation of user k as $p_k=\sigma_k^2(2^{R_kN/N_{\text{a}}}-1)$ and then substitute p_k into the expression of P_{a} in (1), yielding

$$P_{\text{a}}=\frac{1}{M_{\text{a}}(M_{\text{a}}-K)}\sum_{k=1}^{K}\frac{\sigma_k^2}{\beta_k}\left(2^{R_k\frac{N}{N_{\text{a}}}}-1\right) \tag{6}$$

where R_k gives the target bits per subcarrier and per OFDM symbol of user k. Utilizing (6) in problem (5), we obtain the two-dimensional problem (7) in M_{a} and N_{a}. Note that the user rate constraints are implicitly considered and hence satisfied. Expression (6) allows us to express also the maximal transmit power constraint $P_{\text{a}}\le P_{\max}$ as a function of M_{a} and N_{a}, *i.e.*, $(\mathcal{C}_{P_{\max}})$: $(6)\le P_{\max}$. We observe that P_{a} is minimized when $M_{\text{a}}=M$ and $N_{\text{a}}=N$, and this leads to the definition

$$P_{\text{a,min}}=\frac{1}{M(M-K)}\sum_{k=1}^{K}\frac{\sigma_k^2}{\beta_k}\left(2^{R_k}-1\right).$$

If $P_{\text{a,min}}\le P_{\max}$ the problem is feasible, *i.e.*, satisfying the user rate constraints while using all time and spatial resources leads to a transmit power per antenna below the maximum one. We can further show, by finding the two solutions of (6) in M_{a} and selecting the one with positive sign, that the constraint $(\mathcal{C}_{P_{\max}})$ can be expressed as

$$M_{\text{a}}\ge\frac{K}{2}+\frac{1}{2}\sqrt{K^2+4\sum_{k=1}^{K}\rho_k^{-1}\left(2^{R_k\frac{N}{N_{\text{a}}}}-1\right)} \tag{8}$$

where $\rho_k=P_{\max}\beta_k/\sigma_k^2$. From the above inequality we can find $M_{\text{a,min}}$, the minimum number of active antennas, as the value of M_{a} that satisfies (8) when $N_{\text{a}}=N$

$$M_{\text{a,min}}=\left\lceil\frac{K}{2}+\frac{1}{2}\sqrt{K^2+4\sum_{k=1}^{K}\rho_k^{-1}\left(2^{R_k}-1\right)}\right\rceil$$

and $N_{\text{a,min}}$, the minimum number of active time slots, as the minimum value of N_{a} that satisfies (8) when $M_{\text{a}}=M$

$$M\ge\frac{K}{2}+\frac{1}{2}\sqrt{K^2+4\sum_{k=1}^{K}\rho_k^{-1}\left(2^{R_k\frac{N}{N_{\text{a,min}}}}-1\right)}. \tag{9}$$

B. Optimal Spatial and Time Resources

We now consider a continuous relaxation of problem (7), defining the continuous variables $x=\frac{N}{N_{\text{a}}}$ and $y=M_{\text{a}}$. By removing the constant term P_{sleep}, the objective function is

$$f(x,y)=\frac{y}{x}\frac{P_0}{M}+\frac{y}{x}\gamma\left(\frac{1}{y(y-K)}\sum_{k=1}^{K}\frac{\sigma_k^2}{\beta_k}\left(2^{R_kx}-1\right)\right)^{\alpha}+\frac{y}{M}P_1.$$

We can show that $f(x,y)$ is jointly convex in the domain of interest.[2] The constraints $0\le N_{\text{a}}\le N$, $K\le M_{\text{a}}\le M$ and (8) define the feasibility domain of the problem, $\mathcal{D}$, given by

$$\mathcal{D}=\left\{1\le x,y\le M, y\ge\frac{K}{2}+\frac{1}{2}\sqrt{K^2+4\sum_{k=1}^{K}\rho_k^{-1}\left(2^{R_kx}-1\right)}\right\}.$$

We can show that a sufficient condition for the convexity of the domain $\mathcal{D}$, related to constraint (8), is[2]

$$\rho_k=\frac{P_{\max}\beta_k}{\sigma_k^2}\ge\frac{2}{K}\text{ for }k=1,\dots,K. \tag{10}$$

The above is a sufficient condition: in certain scenarios, for example when there are low target user rates, constraint (8) is not binding and the possible non-convexity of $\mathcal{D}$ does not prevent us to find the global optimum. Given the problem convexity, we can find the optimal continuous solution $(\bar{x},\bar{y})=\arg\min_{(x,y)\in\mathcal{D}} f(x,y)$ and the optimal discrete solution as

$$(N_{\text{a}},M_{\text{a}})=\left\lfloor N/\bar{x},\bar{y}\right\rceil \tag{11}$$

[2]Given the space limits of this manuscript, we refer to the extended journal version for the proof [19].

where $\lfloor\cdot\rceil$ selects among $(\lfloor\bar{x}/N_{\mathrm{a}}\rfloor, \lfloor\bar{y}\rfloor)$, $(\lfloor\bar{x}/N_{\mathrm{a}}\rfloor, \lceil\bar{y}\rceil)$, $(\lceil\bar{x}/N_{\mathrm{a}}\rceil, \lfloor\bar{y}\rfloor)$ and $(\lceil\bar{x}/N_{\mathrm{a}}\rceil, \lceil\bar{y}\rceil)$, the one that minimizes the objective function. To retrieve $(\bar{x}, \bar{y})$, we employ an unconstrained Newton's method with iterative inspections of the constraints, where its details and complexity are given in the extended journal version [19]. Important to notice is that the problem has to be solved only when $\{\beta_k\}$ or $\{R_k\}$ change. We can finally identify the three specific energy-saving regimes: (i) **rush-to-sleep**, with all spatial resources active at full transmit power giving $M_{\mathrm{a}} = M$ and $P_{\mathrm{a}} = P_{\max}$ with minimum active time slots $N_{\mathrm{a}} = N_{\mathrm{a,min}}$, (ii) **rush-to-mute**, with all time slots active at full transmit power giving $N_{\mathrm{a}} = N$ ($x = 1$) and $P_{\mathrm{a}} = P_{\max}$ with minimum number of active antennas $M_{\mathrm{a}} = M_{\mathrm{a,min}}$, and (iii) **awake-but-whisper**, with all spatial and temporal resources activated giving $N_{\mathrm{a}} = N$ ($x = 1$) and $M_{\mathrm{a}} = M$ with minimum transmit power $P_{\mathrm{a,min}}$.

IV. Numerical Evaluation

In this section, we evaluate the performance of the optimized solution (11) and compare it with the three benchmarks. We generate the large-scale fading coefficients $\{\beta_k\}$ based on a channel quality indicator (CQI) distribution derived from on-site measurements of sub-6 GHz 5G BSs [10, Figure 3], where we select the data of a deployed 5G BS operating at 3.5 GHz with 100 MHz of bandwidth. The CQI distribution is converted to a signal-to-noise ratio (SNR) distribution by using the appropriate relations [20]. The SNR of user k, SNR_k, can be generated and our rate model (2) shows that $p_k = \sigma_k^2 \mathrm{SNR}_k$. The noise power is set to $\sigma_k^2 = -84$ dBm [21]. We eventually use the expression of the total transmit power (1) and approximate $K = 1$ to compute $\beta_k = \sigma_k^2 \mathrm{SNR}_k / (P_{\mathrm{T}}(M-1))$. The on-site measurements in [10] and the BS certificate of conformity in [22] are used to set $P_{\mathrm{T}} = 20$ W.

The user target rates $\{R_k\}$ are generated as $R_k = \kappa R_{k,0}$, where κ is a rate scaling and $R_{k,0}$ is the baseline target rate for user k. The K baseline rates are drawn from a standard uniform distribution and then normalized such that $\sum_{k=1}^{K} R_{k,0} = 1$. We can compute the maximum rate scaling $\kappa_{\max}$ by setting $M_{\mathrm{a}} = M$, $N_{\mathrm{a}} = N$ and solving for $\kappa_{\max}$

$$M = \frac{K}{2} + \frac{1}{2}\sqrt{K^2 + 4\sum_{k=1}^{K} \rho_k^{-1}\left(2^{\kappa_{\max} R_{k,0}} - 1\right)}$$

for fixed $\{\beta_k\}$ and $\{R_{k,0}\}$. Given the definition of $N_{\mathrm{a,min}}$ (9), we obtain that $N_{\mathrm{a,min}} = \lceil (\kappa/\kappa_{\max}) N \rceil$. We define the network load as the ratio $\kappa/\kappa_{\max}$ and vary it, where a larger κ implies a more loaded system in terms of larger user target rates. We consider $N = 100$ time slots and the parameters in Table I. We observe that when the domain $\mathcal{D}$ is not convex, the distance from the optimized solution (that exploits convexity) to the global optimum is negligible.

A. Power Consumption versus Network Load

We consider one realization of $\{\beta_k\}$ and $\{R_{k,0}\}$ and compute P_{cons} while varying κ. We focus here on the case where time-domain hardware power-saving modes are enabled ($\delta_{\mathrm{PA}}^{\mathrm{dtx}} = 0.25$ and $\delta_{\mathrm{TRX}}^{\mathrm{idle}} = 0.5$). Fig. 2 shows that the four

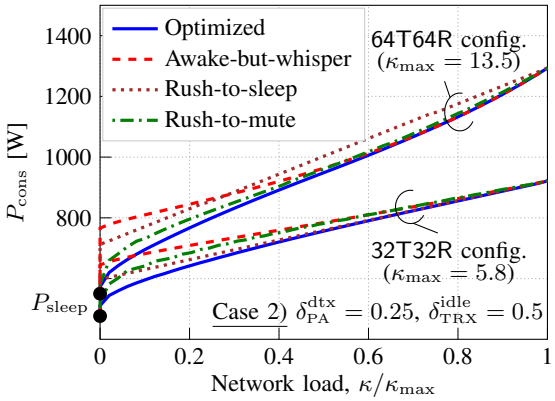

Fig. 2. BS power consumption of the four energy-saving schemes vs. network load for the two BS configurations with enabled PA μDTX and AFE idle-mode power savings. The large-scale fading coefficients and the baseline rates of the 8 users are $\{\beta_k\} = \{0.69, 7.90, 0.36, 15.30, 0.66, 0.44, 1.71, 1.92\} \cdot 10^{-14}$ and $\{R_{k,0}\} = \{0.02, 0.05, 0.14, 0.17, 0.06, 0.26, 0.10, 0.21\}$.

energy-saving schemes consume P_{sleep} at zero load and the maximum amount of power (with $N_{\mathrm{a}} = N$, $M_{\mathrm{a}} = M$ and $P_{\mathrm{a}} = P_{\max}$) at maximum load, and these two values change with the configuration. The 32T32R configuration consumes less power than the 64T64R because its $\kappa_{\max}$ (and then the delivered user rates) is lower. The advantage of utilizing a combination of time-, space-, and power-domain energy-saving techniques is clear for both BSs up to network loads of approximately 0.4, which are the most common in real networks [4]. At loads higher than 0.4, the gap between the optimized (solid blue) curve and the benchmarks diminishes. For the 32T32R BS, the rush-to-sleep is the closest to the optimized scheme at most network loads. Rush-to-mute has instead the smallest gap to the optimal scheme for the 64T64R BS. This is related to the values of P_1 (Table I), where a larger P_1 in the 64T64R case promotes the use of less BS antennas.

B. Empirical Power Consumption Distribution

We select three network loads, low at $\frac{\kappa}{\kappa_{\max}} = 0.01$, medium at $\frac{\kappa}{\kappa_{\max}} = 0.06$, and high at $\frac{\kappa}{\kappa_{\max}} = 0.18$, and obtain for each load the P_{cons} values over 10^3 realizations of $\{\beta_k\}$ and $\{R_{k,0}\}$. For a fixed network load and BS configuration, the rush-to-sleep consumes always the same amount of power given that $N_{\mathrm{a,min}}$ is fixed as $\lceil (\kappa/\kappa_{\max}) N \rceil$. Fig. 3a illustrates that at all three network loads, a rush-to-mute technique achieves the best performance among the benchmarks for the 32T32R BS with no time-domain hardware power savings, and this is more pronounced for the 64T64R BS (Fig. 3b) where the rush-to-mute provides quasi-optimal performance. When the BS hardware can enter time-domain power-saving modes, Fig. 3c shows the larger gains achieved by the optimized solution. Tendency to rush-to-mute remains at low network load, while rush-to-sleep achieves superior performance at medium and high network loads. The combination of rush-to-mute, rush-to-sleep and awake-but-whisper remains optimal for the 64T64R configuration (Fig. 3d), but the performance of the rush-to-sleep deteriorates while the rush-to-mute improves its performance. The gap between the awake-but-whisper and optimized scheme is significant in all the analyzed scenarios.

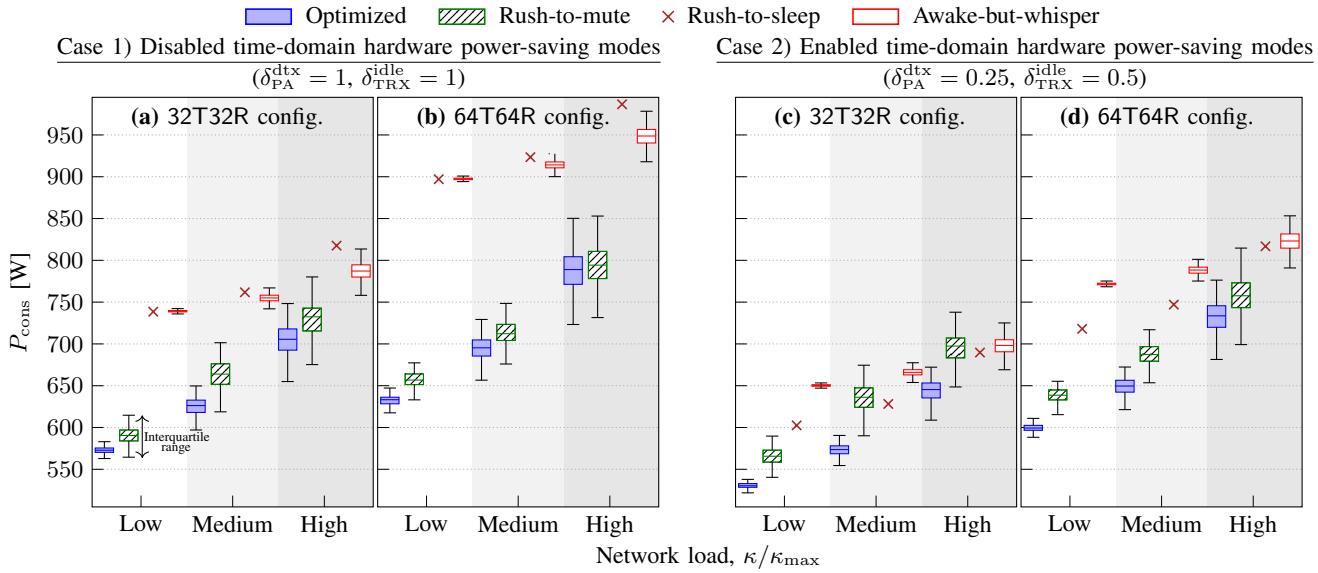

Fig. 3. Boxplots of BS power consumption of the four energy-saving schemes at low network load ($\kappa/\kappa_{\max} = 0.01$), medium network load ($\kappa/\kappa_{\max} = 0.06$), and high network load ($\kappa/\kappa_{\max} = 0.18$). **(a)** and **(b)** refer to 32T32R and 64T64R configurations with disabled PA µDTX and AFE idle-mode power savings, **(c)** and **(d)** refer to 32T32R and 64T64R configurations with enabled PA µDTX and AFE idle-mode power savings. Outliers data are not displayed.

V. Conclusion

We studied the optimal allocation of active time slots, active antennas, and transmit power for sub-6 GHz massive MIMO BSs under user-rate and power constraints using a state-of-the-art power consumption model. The optimization problem is simplified and efficiently solved. We show that minimizing the number of active antennas has the largest contribution to energy savings at most network loads when time-domain hardware power-saving modes are disabled, leading to energy savings up to 30% at low network loads. The joint optimization of active antennas, time slots and transmit power is required when the BS hardware uses time-domain power-saving modes such as PA µDTX, with median energy savings up to 22% at low network loads. Multi-domain resource allocation and practical evaluation of BS energy savings should be further explored and kept at the forefront of future network operation.

References

[1] T. Li *et al.*, "Carbon emissions of 5G mobile networks in China", *Nat. Sustain.*, vol. 6, pp, 1620–1631, Aug. 2023.

[2] L. Golard, Y. Agram, F. Rottenberg, F. Quitin, D. Bol, and J. Louveaux, "A parametric power model of multi-band sub-6 GHz cellular base stations using on-site measurements," in *Proc. IEEE 35th Int. Symp. Pers. Indoor Mob. Radio Commun.*, Valencia, Spain, 2024, pp. 1–7.

[3] A. Ahmed and M. Coupechoux, "The long road to sobriety: Estimating the operational power consumption of cellular base stations in France," in *Proc. 2023 Int. Conf. ICT Sustain.*, Rennes, France, 2023, pp. 188–196.

[4] L. Golard, J. Louveaux, and D. Bol, "Evaluation and projection of 4G and 5G RAN energy footprints: The case of Belgium for 2020–2025," *Ann. Telecommun.*, Nov. 2022.

[5] D. Laselva, S. Hakimi, M. Lauridsen, B. Khan, D. Kumar, and P. Mogensen, "On the potential of radio adaptations for 6G network energy saving," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit*, Antwerp, Belgium, 2024, pp. 1157–1162.

[6] D. López-Pérez *et al.*, "A survey on 5G radio access network energy efficiency: Massive MIMO, lean carrier design, sleep modes, and machine learning," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 653–697, Firstquarter 2022.

[7] 3GPP, "Study on network energy savings for NR (Release 18)," Tech. Rep. 38.864, Mar. 2023, version 18.1.0.

[8] J.-F. Cheng, H. Koorapaty, P. Frenger, D. Larsson, and S. Falahati, "Energy efficiency performance of LTE dynamic base station downlink DTX operation," in *Proc. IEEE 79th Veh. Technol. Conf. Spring*, Seoul, Republic of Korea, 2014, pp. 1–5.

[9] F. E. Salem, A. Gati, Z. Altman, and T. Chahed, "Advanced sleep modes and their impact on flow-level performance of 5G networks," in *Proc. IEEE 86th Veh. Technol. Conf. Fall*, Toronto, Canada, 2017, pp. 1–7.

[10] Y. Agram, F. Rottenberg, and F. Quitin, "Measurement based time-domain power saving through radio equipment deactivation on sub-6GHz base station site," in *Proc. 9th Int. Conf. Green Commun., Comput. Technol.*, Nice, France, 2024.

[11] F. Rottenberg, "Information-theoretic study of time-domain energy-saving techniques in radio access," *IEEE Trans. Green Commun. Netw.*, vol. 9, no. 2, pp. 605–620, June 2025.

[12] K. Senel, E. Björnson, and E. G. Larsson, "Joint transmit and circuit power minimization in massive MIMO with downlink SINR constraints: When to turn on massive MIMO?", *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1834–1846, Mar. 2019.

[13] N. Rajapaksha, J. Mohammadi, S. Wesemann, T. Wild, and N. Rajatheva, "Minimizing energy consumption in MU-MIMO via antenna muting by neural networks with asymmetric loss," *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 6600–6613, May 2024.

[14] E. Peschiera and F. Rottenberg, "Energy-saving precoder design for narrowband and wideband massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 4, pp. 1793–1806, Dec. 2023.

[15] H. V. Cheng, D. Persson, and E. G. Larsson, "Optimal MIMO precoding under a constraint on the amplifier power consumption," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 218–229, Jan. 2019.

[16] S. Marwaha *et al.*, "Energy efficient operation of adaptive massive MIMO 5G HetNets," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 6889–6904, Jul. 2024.

[17] D. Maiwald and D. Kraus, "On moments of complex Wishart and complex inverse Wishart distributed matrices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 5, Munich, Germany, 1997, pp. 3817–3820.

[18] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[19] E. Peschiera, Y. Agram, F. Quitin, L. Van der Perre, and F. Rottenberg, "On optimizing time-, space- and power-domain energy-saving techniques for sub-6 GHz base stations," *arXiv preprint arXiv:2505.15445*.

[20] Y. Wang, W. Liu, and L. Fang, "Adaptive modulation and coding technology in 5G system," in *Proc. Int. Wireless Commun. Mobile Comput.*, Limassol, Cyprus, 2020, pp. 159–164.

[21] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[22] Environmental Dept., Flanders Government, "Certificate of conformity of transmitting antennas," Sep. 2021, certificate no. 00108929.