

On the Resource Allocation of Hierarchical NOMA for Fog-RAN with Energy Harvesting

Vasilis K. Papanikolaou*, Nikos A. Mitsiou*, Panagiotis D. Diamantoulakis*,
Sotirios K. Goudos†, and George K. Karagiannidis*

*Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

†Department of Physics, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

e-mails: vpapanikk@auth.gr, nmitsiou@ece.auth.gr, padiaman@ieee.org, sgoudo@physics.auth.gr, geokarag@auth.gr

Abstract—In this paper, we introduce a multiple access protocol, termed *hierarchical non-orthogonal multiple access (Hi-NOMA)*, optimized for fog-radio access networks (F-RANs). Resource allocation optimization is deemed critical in order to guarantee the users' fairness in the network, while energy efficiency can be increased through energy harvesting (EH) at the user equipment (UE) nodes. Therefore, the HiNOMA protocol with energy harvesting capabilities is examined for F-RANs, leading to the optimization of the proportional fairness metric. Finally, numerical results reveal the effectiveness of the joint design and the interesting trade-off between harvested power and achievable rate in the case of F-RAN.

Index Terms—non-orthogonal multiple access (noma), hierarchical noma, convex optimization, resource allocation, energy harvesting.

I. INTRODUCTION

Recently, non-orthogonal multiple access (NOMA) has been proposed as a capacity-achieving scheme for the Gaussian broadcast channel, which can overcome the limitations of the orthogonal multiple access (OMA) schemes [1], by superimposing multiple users' messages in a single resource block. With the use of advanced signal processing techniques, such as successive interference cancellation, the interference is mitigated and the spectral efficiency is increased.

On the other side, fog-radio access network (F-RAN) has been proposed as a way to fully utilize the network edge and increase the area capacity of the wireless network. More specifically, the base station is divided into two remote parts connected via capacity limited fronthaul links, the centralized pool of baseband units (BBU) and the fog access point (FAP). User equipment (UEs) are connected to their serving FAPs and the FAPs can perform computational tasks locally or forward them to the BBU. Compared to cloud-radio access networks (C-RANs), F-RANs offer better service to delay-sensitive applications, since only centralized processing is available in C-RANs. The architectural traits that are shared between the aforementioned networks incites the use of a common term to describe them, *hierarchical networks*.

Hierarchical networks like C-RAN and F-RAN have been extensively studied in recent years, mostly in terms of rate,

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning 2014-2020" in the context of the project "Energy Autonomous Mobile Communication Systems with Wireless Power Transfer" (5047877).

resource allocation, and user scheduling. Especially, NOMA's applicability has been examined in these networks, yielding better spectral efficiency compared to OMA schemes [1]–[3]. More specifically, in [4], an F-RAN with NOMA was optimized for the weighted sum rate showcasing NOMA's superiority over OMA in this scenario in terms of user fairness. In [5], resource allocation for a NOMA based C-RAN system was optimized, while it was shown that NOMA can achieve higher spectral efficiency than conventional OMA schemes.

Inspired by the above contributions and the increasing interest in hierarchical network architectures, in this paper, we propose the use of hierarchical NOMA (HiNOMA) to increase the spectral efficiency of the network. Additionally, the energy efficiency of wireless networks have attracted huge interest towards the way to *greener* wireless networks. As such, energy harvesting (EH) offers a way to capture and utilize the RF wireless signals to charge the batteries of UEs and prolong their lifetime. In this regard, we utilize the EH technology within the proposed HiNOMA protocol so that each UE served by the network can harvest a specific amount of energy. Moreover, the fronthaul links are optimized as well, compared to the fixed links considered in [4], while, in the present work, an arbitrary number of FAPs and UEs can be examined in the hierarchical network.

II. SYSTEM MODEL

We consider the downlink transmission of an F-RAN, comprised by a BBU pool, $|\mathcal{N}| = N$ non-interfering fog access points (FAPs), each n of which serve $|\mathcal{M}_n| = M_n$ user equipment nodes (UEs) with $M = \sum_{n=1}^N M_n$, where the operator $|\mathcal{A}|$ denotes the cardinality of set \mathcal{A} . Each UE is served by only one FAP via half-duplex (HD) relaying and we further assume that both the FAPs and UEs are equipped with single antennas. The BBU utilizes an orthogonal frequency resource block B_0 to serve the FAPs and each FAP uses a corresponding B_n block to serve its users. The first hop has a duration of τ_1 for the communication between BBU and FAPs, while the second hop has a duration τ_2 for each FAP to transmit to its users for a total of $\tau_1 + \tau_2 = 1$ for a normalized timeslot duration, while power domain NOMA is utilized for all communication links.

During the transmission phase of the first hop, N signals are transmitted to each FAP from the BBU. The baseband equivalent of the received signal y_n at FAP n is given by

$$y_n = h_n \sum_{i=1}^N \sqrt{p_i} s_i + w_n, \quad (1)$$

where h_n denotes the channel gain coefficient between the n -th FAP and the BBU, p_i represents the allocated power for the i -th FAP, s_i denotes the message sent from the BBU to the n -th FAP and w_n is the additive Gaussian white noise (AWGN) at the receiver of the n -th FAP.

Since in NOMA messages for the users in a group are transmitted simultaneously, there are total energy and power consumption constraints for the timeslot that need to apply for the sum of the allocated powers. As such, for the BBU, concerning the energy and power constraints, respectively,

$$\tau_1 \sum_{n=1}^N P_n \leq E_{\text{BBU}}, \quad \sum_{n=1}^N P_n \leq P_{\text{BBU}}. \quad (2)$$

The FAP decodes the received message and transmits with power domain NOMA to its users (decode and forward relaying). In the second hop, each FAP transmits in non-interfering resource blocks the message to its respective users. The received signal is passed through a power-splitting SWIPT receiver with power-splitting factor $0 \leq \theta \leq 1$, such that the θ part of the received power is going to the energy harvesting receiver and that $(1 - \theta)$ is going to the information receiver.

Therefore, the received information signal y_{nm} of UE m that is served by FAP n is given by

$$y_{nm} = \sqrt{(1 - \theta_{nm})} h_{nm} \sum_{j=1}^{M_n} \sqrt{q_{nj}} s_j + w_{nm}, \quad (3)$$

where q_{nj} is the power coefficient of the j -th UE that is served by FAP n . Moreover, the harvester energy of UE m that is served by FAP n is given by

$$\text{EH}_{nm} = \zeta \theta_{nm} |h_{nm}|^2 \tau_2 \sum_{m=1}^{M_n} q_{nm}, \quad (4)$$

where ζ is the energy harvesting efficiency. Identically to the first hop, the total power consumption constraint for the timeslot needs to apply for the sum of the allocated powers in the second hop as well, therefore

$$\sum_{m=1}^{M_n} q_{nm} \leq P_{\text{FAP}}, \quad \tau_2 \sum_{m=1}^{M_n} q_{nm} \leq E_{\text{FAP}}, \quad \forall n \in \mathcal{N}. \quad (5)$$

The achievable data rate in downlink NOMA is determined opportunistically by the FAP's channel condition, therefore SIC is successful if the FAP's are ordered based on their channel conditions, i.e., $|h_m|^2 \geq |h_j|^2$, for $m > j$. Therefore, the achievable rate of FAP n can be expressed as

$$R_n = \tau_1 B_0 \log_2(1 + \gamma_n), \quad (6)$$

while the achievable data rate of UE m that is served by FAP n is given by

$$R_{nm} = \tau_2 B_n \log_2(1 + \gamma_{nm}), \quad (7)$$

where γ_n denotes the signal-to-interference plus noise ratio (SINR) after successive interference cancellation (SIC) at the FAP n , written as

$$\gamma_n = \frac{|h_n|^2 P_n}{|h_n|^2 \sum_{i=n+1}^N P_i + \sigma^2}, \quad (8)$$

where σ^2 denotes the variance of the AWGN at the receiver. Similarly, γ_{nm} denotes the SINR after EH and SIC at the UE m and can be expressed as

$$\gamma_{nm} = \frac{(1 - \theta_{nm}) |h_{nm}|^2 q_{nm}}{(1 - \theta_{nm}) |h_{nm}|^2 \sum_{i=m+1}^{M_n} q_{ni} + \sigma^2}. \quad (9)$$

Furthermore, it should be noted that the rate expressed in (6) and (7) is the maximum achievable rate without considering jointly the two hops.

Finally, in order for the FAPs to be able to support their respective UEs, the achievable data rate of FAP n needs to be greater or equal than the sum of the required rates of its UEs, i.e.,

$$R_n \geq \sum_{m=1}^{M_n} R_{nm}, \quad \forall n \in \mathcal{N}. \quad (10)$$

III. OPTIMAL RESOURCE ALLOCATION

In this section we optimize one of the most common fairness metric in communications, the proportional fairness (PF) [6], which is defined by the sum-log-rate of the UEs. Solutions that yield very low data rates to some UEs offer lower PF, due to the logarithm, while PF is also an increasing function of the rate. Therefore, its maximization leads to a balance between user fairness and sum throughput in the system.

Our aim is to optimize resource allocation, given by the power allocation coefficients \mathbf{p} and \mathbf{q} for every FAP and every UE respectively, the timeframe duration of the two hops, τ , and the required splitting factor θ_{nm} of every UE, so that the proportional fairness is maximized in the HiNOMA system, according to the system model description from the previous section. Thus, the optimization problem is expressed

$$\begin{aligned} & \max_{\tau, \mathbf{p}, \mathbf{q}} \sum_{n=1}^N \sum_{m=1}^{M_n} \log R_{nm}(\mathbf{q}, \tau) \\ & \text{s.t. } C_1 : \sum_{m=1}^{M_n} R_{nm}(\mathbf{q}, \tau_2) \leq R_n(\mathbf{P}, \tau_1), \quad \forall n \in \mathcal{N}, \\ & C_2 : \tau_1 + \tau_2 \leq 1, \quad C_3 : \sum_{n=1}^N P_n \leq P_{\text{BBU}}, \\ & C_4 : \sum_{m=1}^{M_n} q_{nm} \leq P_{\text{FAP}}, \quad \forall n \in \mathcal{N}, \quad C_5 : \tau_1 \sum_{n=1}^N P_n \leq E_{\text{BBU}}, \end{aligned}$$

$$\begin{aligned}
 C_6 : \quad & \tau_2 \sum_{m=1}^{M_n} q_{nm} \leq E_{\text{FAP}}, \quad \forall n \in \mathcal{N}, \\
 C_7 : \quad & 0 \leq p_n \leq P_{\text{BBU}}, \quad \text{and} \quad 0 \leq q_{nm} \leq P_{\text{FAP}}, \\
 C_8 : \quad & \tau_2 \zeta \theta_{nm} |h_{nm}|^2 \sum_{m=1}^{M_n} q_{nm} \geq \bar{E}H_{nm}.
 \end{aligned} \tag{11}$$

The optimization problem in (11) can be infeasible due to constraint C_8 that dictates a minimum amount of power has to be harvested for every user despite the stochastic nature of the channel gain. A condition for the feasibility of the problem is obtained when the maximum allowed values for the optimization variables, i.e., $\theta_{nm} = 1$ and $\tau_2 \sum_{m=1}^{M_n} q_{nm} = E_{\text{FAP}}$ are utilized in C_8 , which is expressed as $|h_{nm}|^2 \geq \frac{\bar{E}H_{nm}}{\zeta E_{\text{FAP}}}$, $\forall m \in \mathcal{M}_n, \forall n \in \mathcal{N}$. In this case, Rayleigh fading with $\sigma_R = 1$ is assumed, therefore the channel gain $|h_{nm}|^2$ of each user m served by FAP n follows an exponential distribution. As such, the probability that problem (11) is infeasible is given as

$$\begin{aligned}
 P_I &= 1 - \prod_{n=1}^N \prod_{m=1}^{M_n} \Pr \left(|h_{nm}|^2 \geq \frac{\bar{E}H_{nm}}{\zeta E_{\text{FAP}}} \right) \\
 &= 1 - \prod_{n=1}^N \prod_{m=1}^{M_n} \exp \left(-\frac{\bar{E}H_{nm}}{\zeta E_{\text{FAP}}} \right).
 \end{aligned} \tag{12}$$

Moreover, it is noted that the optimization problem in (11) is non-convex, since in the expression of the achievable rates by both the FAPs and the UEs, the term of the interference in the SINR leads to the inclusion of the power variable in the denominator. Additionally, the objective function is the logarithm of the rate and therefore it is non-concave. More importantly, in constraint C_1 a difference of logarithms appears causing the function to be non-convex. Furthermore, the inclusion of timeslot duration as a variable that is multiplied with the logarithm function causes the function to be non-concave, as well. Finally, C_5 and C_6 are not convex because of the multiplication of τ_1 with P_n and τ_2 with q_{nm} . Therefore the complexity to solve this problem is high, mainly due to the relation of the rates with the power allocation variables. Thus, it is important to prove, that the problem in (11) can be transformed to a convex one; so, the process to find a global maximum can be solved in polynomial time.

Proposition 1: The optimization problem in (11) can be formulated as a convex one and is given in (13)

Following Proposition 1, the equivalent convex problem of (11) can be expressed as follows:

$$\begin{aligned}
 \max_{\tilde{P}, \tilde{q}, \tilde{\tau}, \tilde{\theta}, \tilde{r}_{nm}, \tilde{r}_n} \quad & \sum_{n=1}^N \sum_{m=1}^{M_n} \tilde{r}_{nm} \\
 \text{s.t.} \quad C_1 : \quad & \sum_{m=1}^{M_n} e^{\tilde{r}_{nm} - \tilde{r}_n} - 1 \leq 0, \quad \forall n \in \mathcal{N}, \\
 C_2 : \quad & e^{\tilde{r}_1} + e^{\tilde{r}_2} \leq 1, \quad C_3 : \quad \sum_{n=1}^N e^{\tilde{P}_n} \leq P_{\text{BBU}},
 \end{aligned}$$

$$\begin{aligned}
 C_4 : \quad & \sum_{m=1}^{M_n} e^{\tilde{q}_{nm}} \leq P_{\text{FAP}}, \quad \forall n \in \mathcal{N}, \\
 C_5 : \quad & \sum_{n=1}^N e^{\tilde{P}_n + \tilde{r}_1} \leq E_{\text{BBU}}, \\
 C_6 : \quad & \sum_{m=1}^{M_n} e^{\tilde{q}_{nm} + \tilde{r}_2} \leq E_{\text{FAP}}, \quad \forall n \in \mathcal{N}, \\
 C_7 : \quad & \zeta (1 - e^{\tilde{\theta}_{nm}}) |h_{nm}|^2 E_{\text{FAP}} \geq \bar{E}H_{nm}, \\
 & \quad \forall m \in \mathcal{M}_n, \forall n \in \mathcal{N}. \\
 C_8 : \quad & -\tilde{P}_n - \log(|h_n|^2) + \log \left(2^{\frac{1}{B_0} \exp(\tilde{r}_n - \tilde{r}_1)} - 1 \right) \\
 & + \log \left(\sigma^2 + |h_n|^2 \sum_{k>n}^{M_n} \exp(\tilde{P}_k) \right) \leq 0, \\
 C_9 : \quad & -\tilde{q}_{nm} - \tilde{\theta}_{nm} - \log(|h_{nm}|^2) + \log \left(2^{\frac{\exp(\tilde{r}_{nm} - \tilde{r}_2)}{B_n}} - 1 \right) \\
 & + \log \left(\sigma^2 + |h_{nm}|^2 \sum_{k>m}^{M_n} \exp(\tilde{q}_{nk} + \tilde{\theta}_{nm}) \right) \leq 0,
 \end{aligned} \tag{13}$$

Proof: The proof is provided in Appendix A. \blacksquare

Due to its convexity, the optimization problem in (13) can be solved by a decomposition method. However, because of the constraints C_1 and C_2 the two hops cannot be decoupled completely, leading any solution that independently optimizes each hop to be suboptimal. As such, primal decomposition methods are not appropriate for this problem. To this end, a dual decomposition method is required to obtain a global optimum solution, which can be obtained due to the satisfaction of Slater's constraint [7]. Based on the above, problem (13) is solved with the Lagrange dual decomposition method.

IV. NUMERICAL RESULTS AND DISCUSSION

In this section, Monte Carlo simulation results with 10^5 iterations are presented for the hierarchical network with the proposed protocols. Rayleigh fading is assumed for the links between the BBU and the FAPs, as well as between the FAPs and their assigned UEs, i.e., $h_n, h_{nm} \sim CN(0, 1)$. Moreover, the FAPs are assumed to transmit with SNR lower by a factor of 20 compared with the transmit SNR of the BBU, approximately 13dB lower. This is a practical assumption, since most BBU pools have greater power capabilities than their respective FAPs in the network. Additionally, $E_{\text{BBU}}/P_{\text{BBU}} = 1$ and $E_{\text{FAP}}/P_{\text{FAP}} = 1$ are taken into account for the presented simulation results. In Fig. 1, a total of $N = 2$ FAPs are deployed, with a total of $M_1 = 2$ and $M_2 = 3$ UEs, respectively.

Benchmark Scheme: A benchmark TDMA scheme is also considered to compare the proposed protocols with conventional OMA solutions. In this TDMA scheme, the total timeslot is divided in two, one slot for each hop. During the first timeslot, τ_1 , the BBU transmits information to the FAPs with TDMA as the protocol. In the second timeslot,

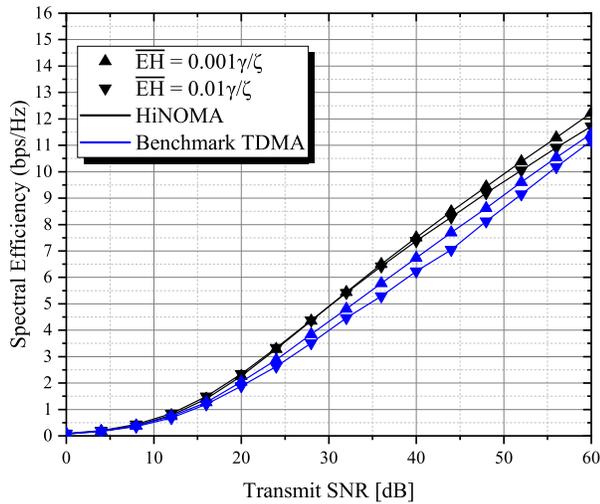


Fig. 1. Spectral Efficiency with regards to transmit SNR γ for various values of required harvested energy for the HiNOMA protocol and the benchmark TDMA scheme.

τ_2 , each FAP transmits to their assigned UEs also via a TDMA protocol. Resource allocation is also optimized in the benchmark TDMA scheme in the same way as the proposed schemes for fair comparison.

In Fig. 1, the sum rate is presented for the HiNOMA protocol versus the transmit SNR. The proposed protocol outperforms the benchmark TDMA with a gain of over 3dB for a large number of practical transmit SNR scenarios. Additionally, as expected, when the harvested energy requirements are higher, the performance of both systems is scaled down in both communication protocols.

APPENDIX A PROOF OF PROPOSITION 1

In order to transform the objective function into a concave function, two auxiliary variables are introduced, r_n and r_{nm} , which add the following constraints to the problem, $r_n \leq R_n$ and $r_{nm} \leq R_{nm}$. The problem remains non-convex, so we introduce the following transformations in order to avoid products of the optimization variables from appearing in the final expressions:

$$\begin{aligned} P_n &= \exp(\tilde{P}_n), & \forall n \in \mathcal{N}, \\ q_{nm} &= \exp(\tilde{q}_{nm}), & \forall m \in \mathcal{M}_n, \forall n \in \mathcal{N}, \\ \tau_i &= \exp(\tilde{\tau}_i), & \forall i \in \{1, 2\}, \end{aligned}$$

$$\begin{aligned} r_n &= \exp(\tilde{r}_n), \quad r_{nm} = \exp(\tilde{r}_{nm}), \quad \forall m \in \mathcal{M}_n, \forall n \in \mathcal{N}, \\ 1 - \theta_{nm} &= \exp(\tilde{\theta}_{nm}). \end{aligned} \quad (14)$$

After the transformation and some algebraic manipulations, most of the constraints are easily shown to be convex, and the objective function concave. The newly introduced constraints

due to the transformation can be similarly shown to be convex, as such, for the case of $e^{\tilde{r}_{nm}} \leq R_{nm}$.

$$\begin{aligned} & -\tilde{q}_{nm} - \tilde{\theta}_{nm} - \log(|h_{nm}|^2) + \log\left(2^{\frac{\exp(\tilde{r}_{nm} - \tilde{\tau}_2)}{B_n}} - 1\right) \\ & + \log\left(\sigma^2 + |h_{nm}|^2 \sum_{k>m} \exp(\tilde{q}_{nk} + \tilde{\theta}_{nm})\right) \leq 0, \end{aligned} \quad (15)$$

The first three terms of (15) are linear. The fifth term is convex as a log-sum-exp function. Finally, the fourth term of the left part of (15) is a function $g = \log\left(2^{\frac{1}{B_n}} \exp(\tilde{r}_{nm} - \tilde{\tau}_2) - 1\right)$ we need to examine for convexity. By considering its Hessian matrix, which has a single non-zero eigenvalue that is expressed as

$$u_1 = 2q = \frac{2^z z \log(2)(2^z - z \log(2) - 1)}{(2^z - 1)^2}, \quad (16)$$

where z is defined by $z = \frac{1}{B_n} \exp(\tilde{r}_{nm} - \tilde{\tau}_2)$. Considering also that $y = 2^z - z \log(2) - 1$ is an increasing function with respect to z and when $z \rightarrow 0$, $y \rightarrow 0$, it is shown that $u_1 \geq 0$. Then, it becomes evident that the Hessian matrix of g is positive semi-definite, due to the fact that the eigenvalues of the matrix are non-negative. Similarly for the case of $e^{\tilde{r}_n} \leq R_n$. As a result, the newly introduced constraints are convex after the transformation, as well.

Finally, for the case of C_7 , it is easy to show by contradiction that the maximum allowed energy would be consumed, as it increases the objective function and thus the optimal solution is not excluded from the possible solution set. Therefore, C_7 is transformed to reflect that and as such it is a convex constraint, as well.

Therefore, the non-convex problem of (11) can be transformed to an equivalent convex problem and the proof is completed.

REFERENCES

- [1] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.
- [2] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *2013 IEEE 24th Annu. Int. Symp. Pers. Indoor, and Mobile Radio Commun. (PIMRC)*, London, UK, Sep. 2013, pp. 611–615.
- [3] S. Timotheou and I. Krikidis, "Fairness for Non-Orthogonal Multiple Access in 5G Systems," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1647–1651, 2015.
- [4] I. Randrianantenaina, M. Kaneko, H. Dahrouj, H. ElSawy, and M. Alouini, "Interference Management in NOMA-Based Fog-Radio Access Networks via Scheduling and Power Allocation," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 5056–5071, 2020.
- [5] R. Rai, H. Zhu, and J. Wang, "Resource scheduling in non-orthogonal multiple access (NOMA) based cloud-RAN systems," in *2017 IEEE 8th Annual Ubiquitous Computing Electron. Mobile Commun. Conf. (UEMCON)*, New York, NY, USA, Oct. 2017, pp. 418–422.
- [6] V. K. Papanikolaou, P. D. Diamantoulakis, P. C. Sofotasios, S. Muhaidat, and G. K. Karagiannidis, "On Optimal Resource Allocation for Hybrid VLC/RF Networks With Common Backhaul," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 6, no. 1, pp. 352–365, 2020.
- [7] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.