Energy-Efficient Edge Computing empowered by Reconfigurable Intelligent Surfaces

Paolo Di Lorenzo

Thanks to: Mattia Merluzzi, Emilio Calvanese Strinati, Sergio Barbarossa

SAPIENZA
UNIVERSITÀ DI ROMA

CLEEN 2022 Workshop, IEEE PIMRC 2022, 12/09/2022

(Beyond) 5G services

- Beyond 5G networks aim to be the enabler of a plethora of new services and applications, e.g., artificial intelligence, industry 4.0, autonomous driving, internet of things, etc.

(Beyond) 5G services

Requirements challenges

- Poor device battery capacity
- High network energy consumption

Strict E2E latency constraints

Extremely high reliability

Mobile edge computing

- **Benefits:** low latency, distributed storage and computing resources, reduced backbone load

- **Drawbacks:** reduced available computation resources (no more virtually infinite as in the cloud!), mobility issues,….

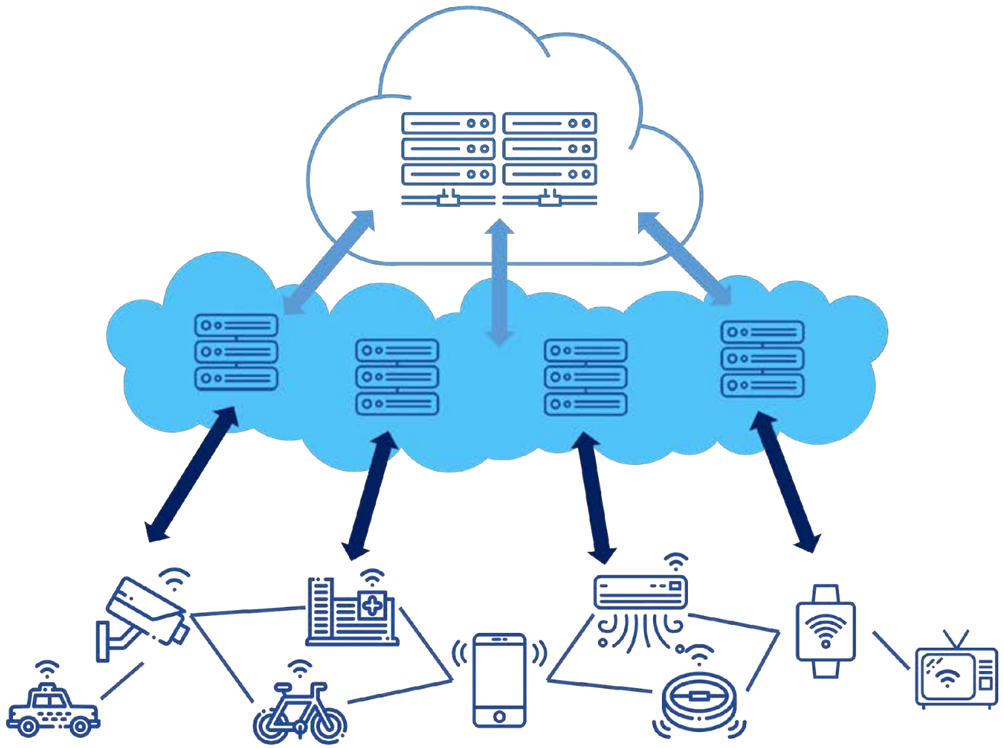


Figure: An Edge Computing architecture

- **Challenge:** effective joint management of radio and computation resources, dealing with complex and unknown dynamics of the system

Reconfigurable wireless environments

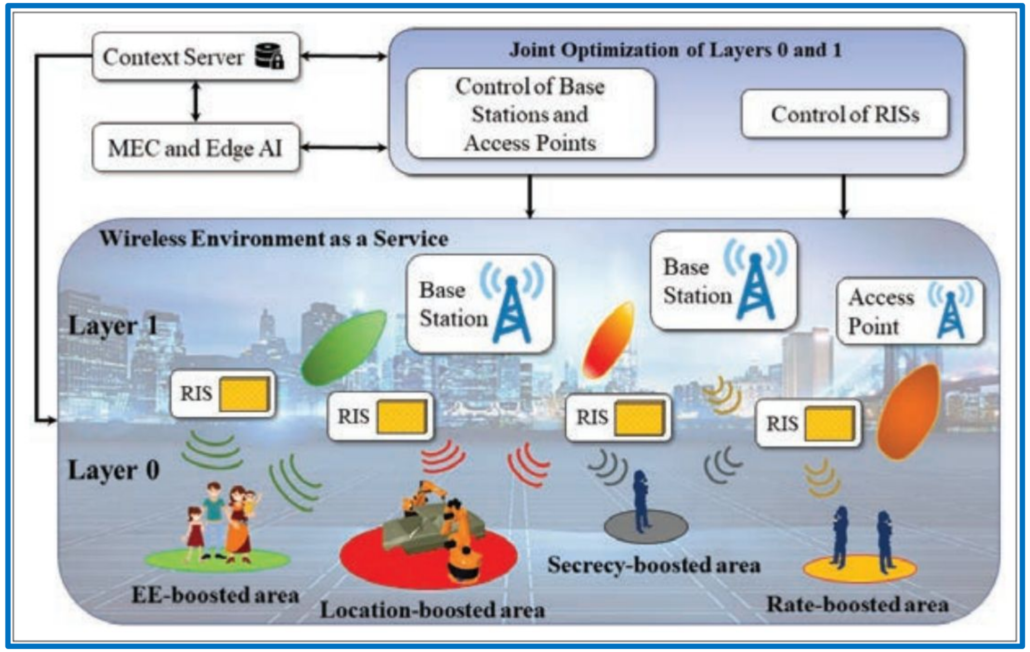
- Moving toward mmWave communications (and beyond), poor channel conditions due to mobility, dinami of the environment, and blocking events, might severely hinder the performance of MEC

- **Idea:** Shape the wireless propagation to enhance communication (and beyond) performance

https://rise-6g.eu/

- **Reconfigurable intelligent surfaces (RISs):** Structures of nearly passive elements, whose phases can be reconfigured to shape signal reflections

- **Boosted area:** Zone in space, time and frequency, where KPIs/KVIs are enhanced thanks to RISs
  - Area with enhanced capacity
  - Area with reduced Electromagnetic Field exposure
  - Area with enhanced secrecy
  - Area with enhanced localization capabilities
  - **Area with enhanced edge computing capabilities**
  - ……

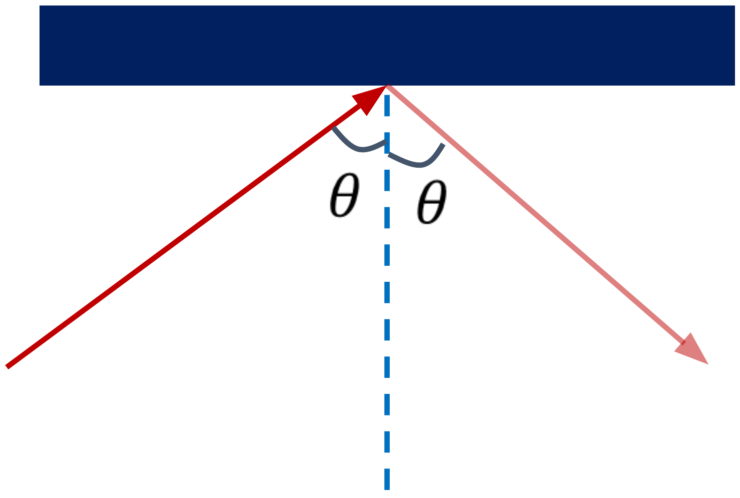


RIS-empowered network [SAW+21]

[SAW+21] E. C. Strinati *et al.*, "Reconfigurable, Intelligent, and Sustainable Wireless Environments for 6G Smart Connectivity," in *IEEE Communications Magazine*, vol. 59, no. 10, pp. 99-105, October 2021, doi: 10.1109/MCOM.001.2100070.

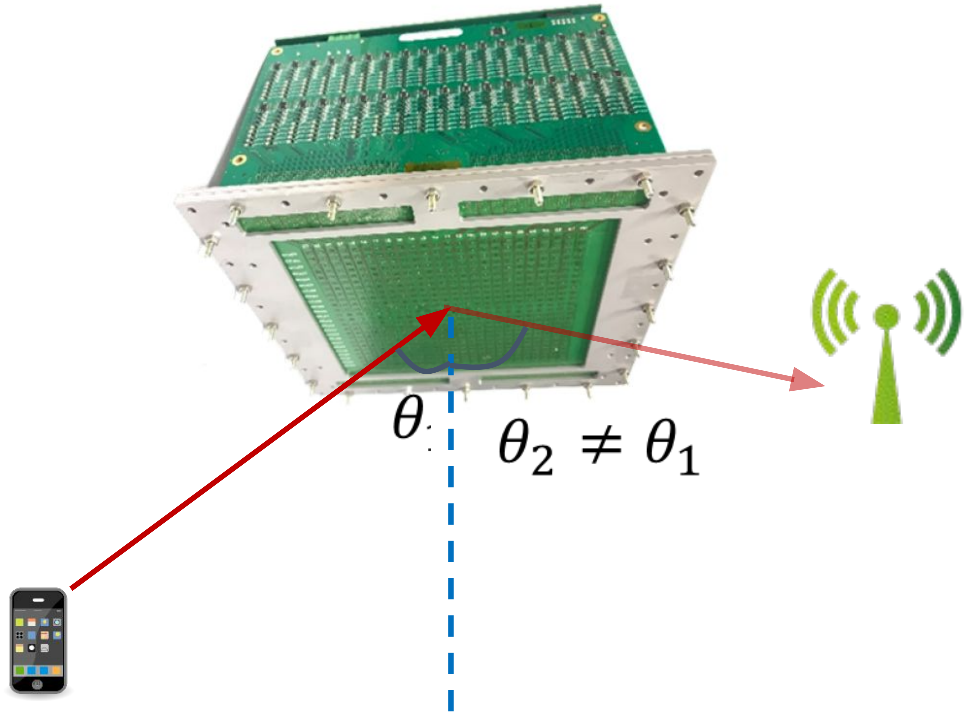
Reconfigurable Intelligent Surfaces principle

- Control the reflection of an impinging electromagnetic wave



Snell's law

RIS prototype developed at CEA-Leti [MSR+21]



Controlled reflection

[MSR+21] F. F. Manzillo, M. Smierzchalski, J. Reverdy and A. Clemente, "A Ka-band Beam-Steering Transmitarray Achieving Dual-Circular Polarization," ***2021 15th European Conference on Antennas and Propagation (EuCAP)*****, 2021, pp. 1-4, doi: 10.23919/EuCAP51087.2021.9411338.**

RIS channel model

- RIS *i* reflectivity matrix at time *t:*

$$\mathbf{\Phi} = \mathrm{diag}\{m_1 e^{j\phi_1}, \ldots, m_N e^{j\phi_N}\}$$

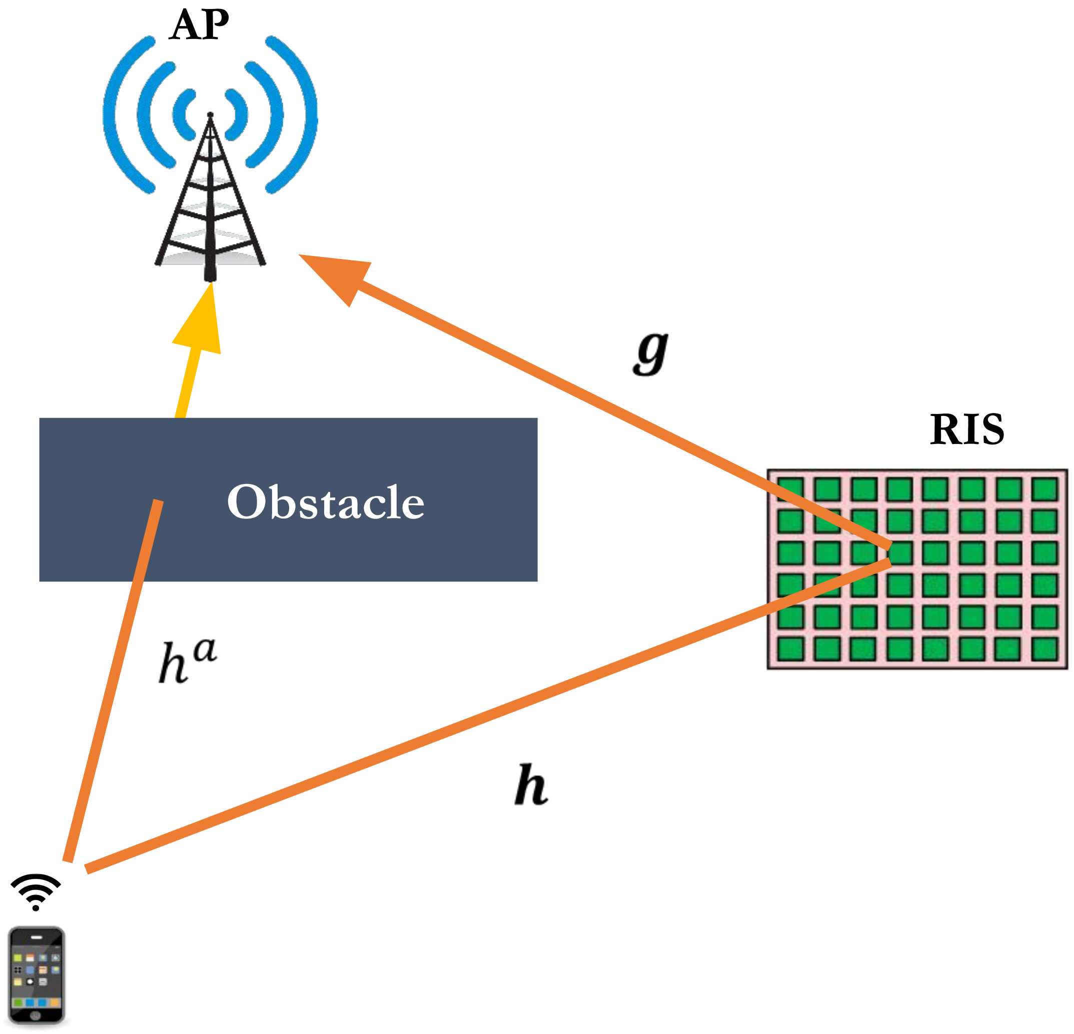
- RIS *reflection coefficients:*

$$v_l = m_l e^{j\phi_l} \in \mathcal{S} = \left[0, \left\{e^{j\frac{2k\pi}{2^b}}\right\}_{k=0}^{2^b-1}\right]$$

$$l = 1, \ldots, N$$

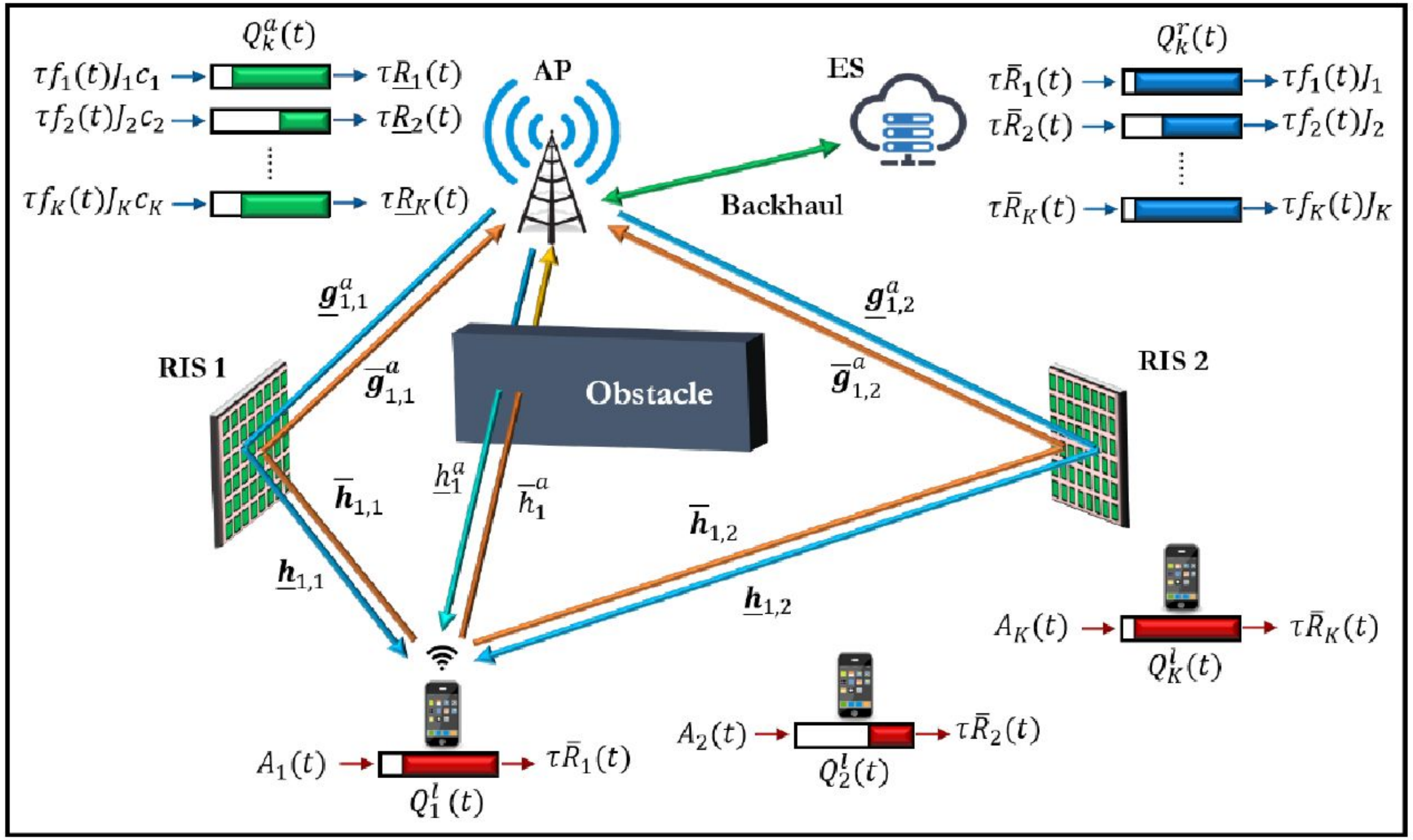
- RIS-aided SISO channel*:*

$$\boxed{\mathbf{h}^{tot} = h^a + \mathbf{h}^T \mathbf{\Phi} \mathbf{g}}$$



- Easy extension to MIMO scenarios

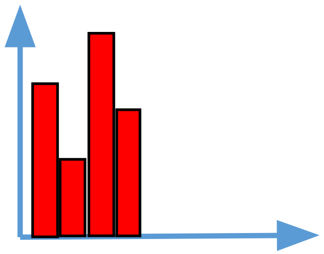
RIS-empowered dynamic mobile edge computing



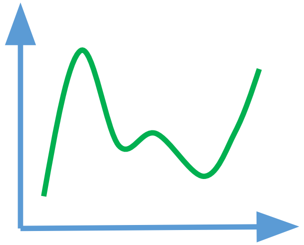
Let us consider a dynamic mobile edge computing scenario empowered by RIS

Unknown statistics of context parameters:

Data arrivals



Radio Channels



Goal: Learn a joint optimization strategy for communication, computation, and RIS resources to empower dynamic MEC

RIS-empowered communications

- RIS i reflectivity matrix at time t:

$$\boldsymbol{\Phi}_i(t) = \text{diag}\left\{m_{i,1}(t)e^{j\phi_{i,1}(t)}, \ldots, m_{i,N_i}(t)e^{j\phi_{i,N_i}(t)}\right\}$$

- RIS *reflection coefficients:*

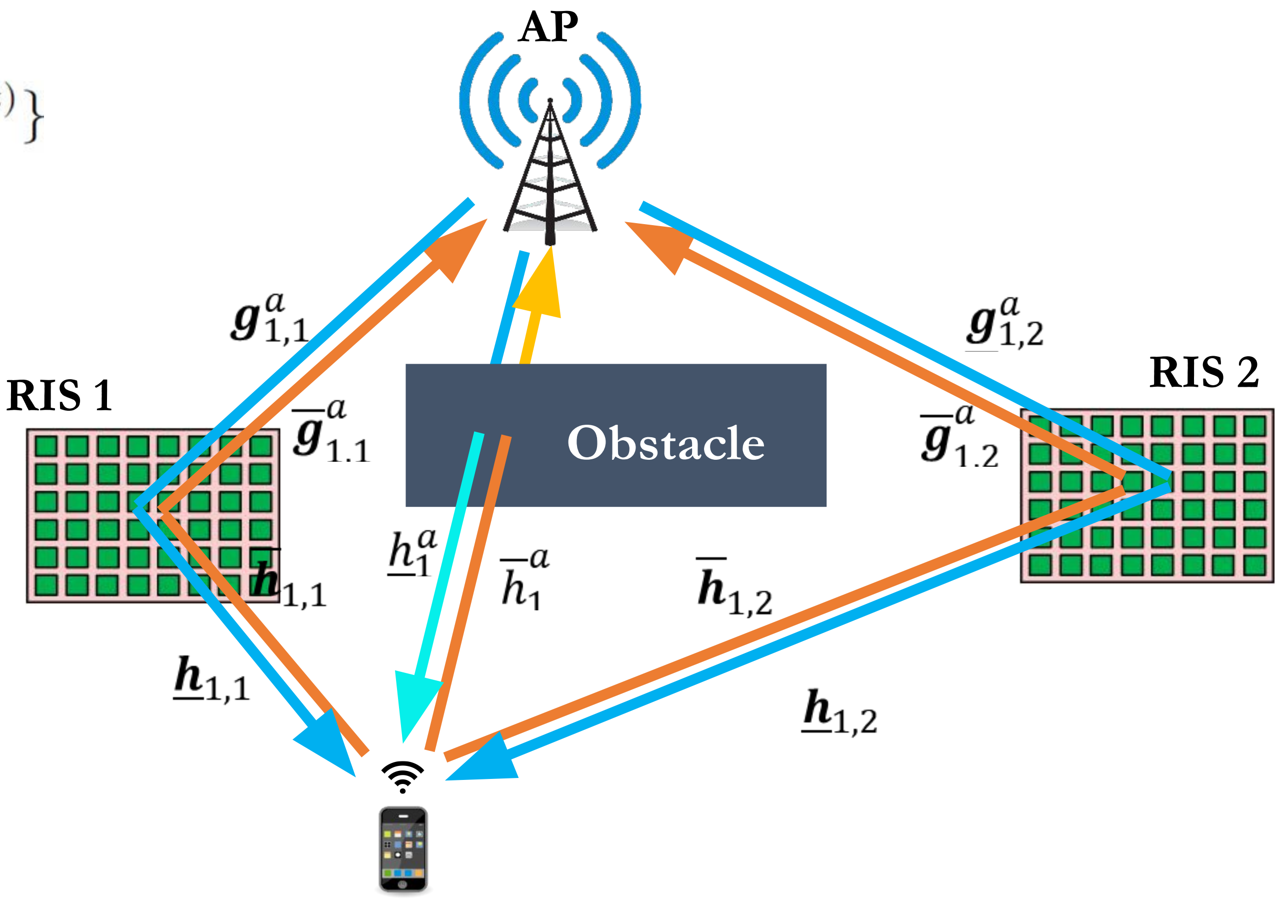
$$v_{i,l}(t) = m_{i,l}(t)e^{j\phi_{i,l}(t)} \in \mathcal{S}_i = \left[0, \left\{e^{j\frac{2m\pi}{2^{b_i}}}\right\}_{m=0}^{2^{b_i}-1}\right]$$

- Uplink rate from user k to AP:

$$\overline{R}_k(t) = \overline{B}_k \log_2\left(1 + \overline{\alpha}_k(\boldsymbol{v}(t))\overline{p}_k(t)\right)$$

$$\overline{\alpha}_k(\boldsymbol{v}(t)) = \frac{\left|\overline{h}_k^a(t) + \sum_{i=1}^{I} \overline{\boldsymbol{h}}_{k,i}(t)^T \,\text{diag}(\boldsymbol{v}_i(t))\, \overline{\boldsymbol{g}}_{k,i}^a(t)\right|^2}{N_0 \overline{B}_k}$$

RIS-dependent normalized channel coefficient



- Underline notation for downlink radio parameters

Queues dynamics

- Local **communication queue** at user *k:*

$$Q_k^l(t+1) = \max\left(0, Q_k^l(t) - \tau \overline{R}_k(t)\right) + A_k(t)$$

Uplink rate

Data arrivals

- Remote **computation queue** at the edge server*:*

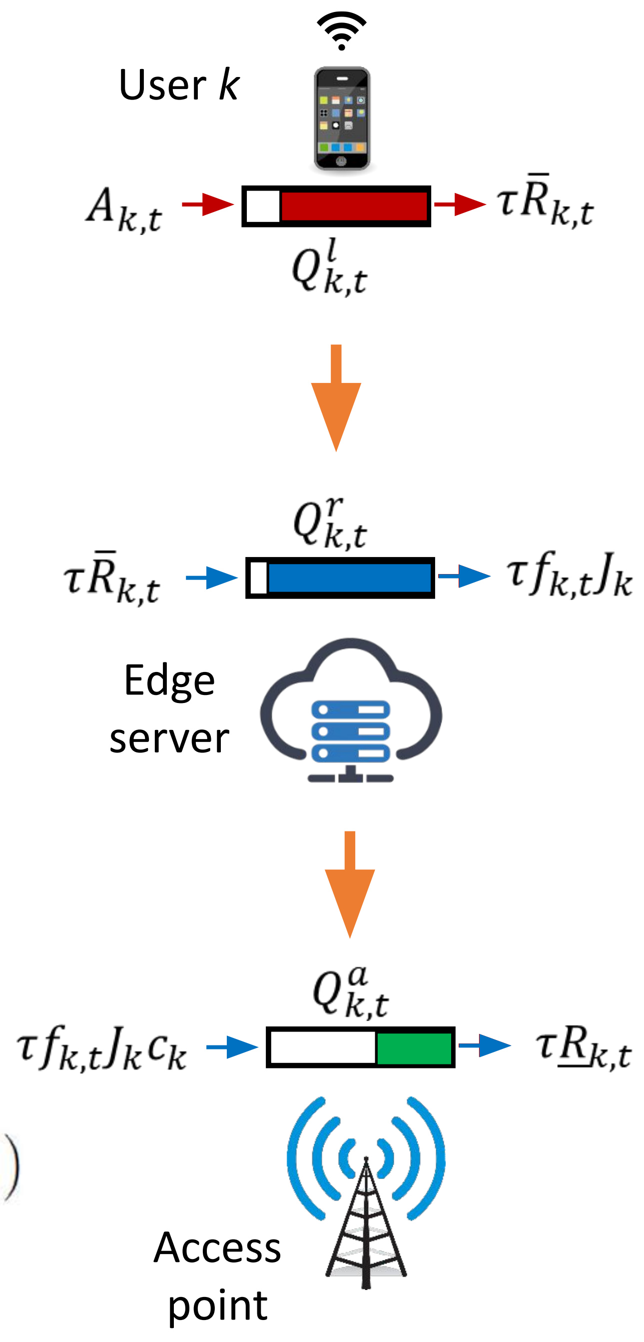
$$Q_k^r(t+1) = \max\left(0, Q_k^r(t) - \tau f_k(t) J_k\right) + \min\left(Q_k^l(t), \tau \overline{R}_k(t)\right)$$

CPU frequency assigned by the ES to user *k*

- Remote **communication queue** at the access point*:*

$$Q_k^a(t+1) = \max\left(0, Q_k^a(t) - \tau \underline{R}_k(t)\right) + c_k \cdot \min\left(Q_k^r(t), \tau f_k(t) J_k\right)$$

Downlink rate



Energy consumption

- Energy spent for **uplink transmission** by user *k* at time *t:* $e_k(t) = \tau\,\overline{p}_k(t) I_a(t)$
- Energy spent for **computation** by the edge server at time *t:* $e_c(t) = \tau \gamma_c \left(f_c(t)\right)^3$
- Energy spent for **downlink transmission** by the access point at time *t:*

$$e_a(t) = \tau\Big(I_a(t) p_a^{\text{on}} + I_a(t) \sum\nolimits_{k=1}^{K} \underline{p}_k(t) + (1 - I_a(t)) p_a^{\text{s}}\Big)$$

- **Energy spent by RIS** *i* at time *t:* $e_i^r(t) = I_a(t) \tau p^r(b_i) \sum\nolimits_{l=1}^{N_i} |v_{i,l}(t)|^2$
- **System energy consumption** at time *t:*

$$e_\sigma^{\text{tot}}(t) = \sigma \sum\nolimits_{k=1}^{K} e_k(t) + (1-\sigma)\Big(e_c(t) + e_a(t) + \sum\nolimits_{i=1}^{I} e_i^r(t)\Big)$$

Problem formulation

$$\min_{\Psi(t)} \quad \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left\{e_\sigma^{\text{tot}}(t)\right\}$$

Average system energy

$$\text{subject to} \quad (a) \quad \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left\{Q_k^{\text{tot}}(t)\right\} \le Q_k^{\text{avg}}, \quad \forall k;$$

Average latency constrains

$$\left.\begin{array}{llllll}
(b) & I_a(t) \in \{0,1\} \quad \forall t; & (c) & v_{i,l}(t) \in \mathcal{S}_i \quad \forall i,l,t; \\
(d) & |v_{i,l}(t)|^2 \le I_a(t) \quad \forall i,l,t; & (e) & 0 \le \overline{p}_k(t) \le P_k\, I_a(t), \quad \forall k,t; \\
(f) & \underline{p}_k(t) \ge 0, \quad \forall k,t; & (g) & \sum_{k=1}^{K} \underline{p}_k(t) \le P_a\, I_a(t), \quad \forall t; \\
(h) & f_k(t) \ge 0, \quad \forall k,t; & (i) & \sum_{k=1}^{K} f_k(t) \le f_c(t), \quad \forall t; \\
(j) & f_c(t) \in \mathcal{F} \quad \forall t; & &
\end{array}\right\} \mathcal{X}(t)$$

Instantaneous constraints on the variables

Difficult problem with unknown statistics → Lyapunov stochastic optimization

Dynamic solution based on stochastic optimization

Virtual queues:

$$Z_k(t+1) = \max\left\{0, Z_k(t) + \epsilon_k\left(Q_k^{\text{tot}}(t+1) - Q_k^{\text{avg}}\right)\right\}$$

Lyapunov function:

$$\mathcal{L}(t) = \mathcal{L}(\mathbf{\Theta}(t)) = \frac{1}{2}\sum_{k=1}^{K} Z_k^2(t)$$

$$\mathbf{\Theta}(t) = \{Z_k(t)\}_{k=1}^{K}$$

Drift-plus-penalty function:

$$\Delta^p(t) = \mathbb{E}\left\{\mathcal{L}(t+1) - \mathcal{L}(t) + V \cdot e_\sigma^{\text{tot}}(t) \middle| \mathbf{\Theta}(t)\right\}$$

Per-slot optimization problem:

$$\begin{aligned}\min_{\Psi(t)\in\widetilde{\mathcal{X}}(t)} \quad & \sum_{k=1}^{K}\Big[\big(4Q_k^r(t) - 2Q_k^l(t) - Z_k(t)\big)\tau\overline{R}_k(t) + \big(4c_kQ_k^a(t) - 4Q_k^r(t) - Z_k(t)\big)\tau f_k(t)J_k \\ & - \big(4Q_k^a(t) + Z_k(t)\big)\tau\underline{R}_k(t)\Big] + V \cdot e_\sigma^{\text{tot}}(t)\end{aligned}$$

The problem can be **decomposed into sub-problems** that admit **low-complexity** solution procedures

RIS and radio resource optimization

If the AP is active, the sub-problem associated with **RIS and radio resource allocation** reads as:

$$\min_{\boldsymbol{\Psi}^r(t)} \; -\sum_{k=1}^{K} \overline{U}_k(t) \log_2\left(1+\overline{\alpha}_k(\boldsymbol{v}(t))\overline{p}_k(t)\right) - \sum_{k=1}^{K} \underline{U}_k(t) \log_2\left(1+\underline{\alpha}_k(\boldsymbol{v}(t))\underline{p}_k(t)\right)$$

$$+V\left[\sum_{k=1}^{K}\left(\sigma\tau\overline{p}_k(t)+(1-\sigma)\,\tau\underline{p}_k(t)\right)+(1-\sigma)\tau p_a^{\text{on}}+(1-\sigma)\tau\sum_{i=1}^{I}p^r(b_i)\sum_{l=1}^{N_i}|v_{i,l}(t)|^2\right]$$

$$\text{subject to} \quad 0\leq\overline{p}_k(t)\leq\widetilde{P}_k(t) \quad \forall k; \qquad v_{i,l}(t)\in\mathcal{S}_i \quad \forall i,l;$$

$$0\leq\underline{p}_k(t)\leq\underline{P}_k(t), \quad \forall k; \qquad \sum_{k=1}^{K}\underline{p}_k(t)\leq P_a.$$

$$\overline{U}_k(t)=(2Q_k^l(t)-4Q_k^r(t)+Z_k(t))\overline{B}_k\tau \qquad \underline{U}_k(t)=(4Q_k^a(t)+Z_k(t))\underline{B}_k\tau$$

Complex problem with exponential complexity ➡ approximated solution

RIS and radio resource optimization

Complex problem with exponential complexity ➡ approximated solution based on the following steps:

1) Find the RISs parameters that greedily maximize

$$\mathcal{U}(t) = \{k \,|\, \overline{U}_k(t) > 0\}$$

$$\Delta^R(\boldsymbol{v}(t)) = -\sum_{k\in\mathcal{U}(t)} \overline{U}_k(t)\overline{\alpha}_k(\boldsymbol{v}(t)) - \sum_{k=1}^{K} \underline{U}_k(t)\underline{\alpha}_k(\boldsymbol{v}(t)) + V(1-\sigma)\tau\sum_{i=1}^{I} p^R(b_i)\sum_{l=1}^{N_i}|v_{i,l}(t)|^2$$

2) Closed-form solution for uplink and downlink powers, given the RISs configuration

Uplink powers

$$\overline{p}_k(t) = \begin{cases} \left[\dfrac{\overline{U}_k(t)}{V\tau\log 2} - \dfrac{1}{\overline{\alpha}_k(\boldsymbol{v}(t))}\right]_0^{\widetilde{P}_k(t)}, & \text{if } k \in \mathcal{U}_t; \\ 0, & \text{if } k \notin \mathcal{U}_t. \end{cases}$$

Downlink powers

$$\underline{p}_k(t) = \left[\frac{\underline{U}_k(t)}{[V(1-\sigma)+\nu]\log 2} - \frac{1}{\underline{\alpha}_k(\boldsymbol{v}(t))}\right]_0^{\underline{P}_k(t)}$$

Computing resource optimization

The sub-problem associated with **computing resource allocation** reads as:

$$\min_{\{f_k(t)\}_{k=1}^K,\, f_c(t)} \quad -\sum_{k=1}^{K} Y_k(t) f_k(t) + V(1-\sigma)\tau\gamma_s(f_c(t))^3$$

$$\text{subject to} \qquad 0 \le f_k(t) \le \frac{Q_k^r(t)}{\tau J_k}, \quad \forall k; \qquad \sum_{k=1}^{K} f_k(t) \le f_c(t); \qquad f_c(t) \in \mathcal{F}$$

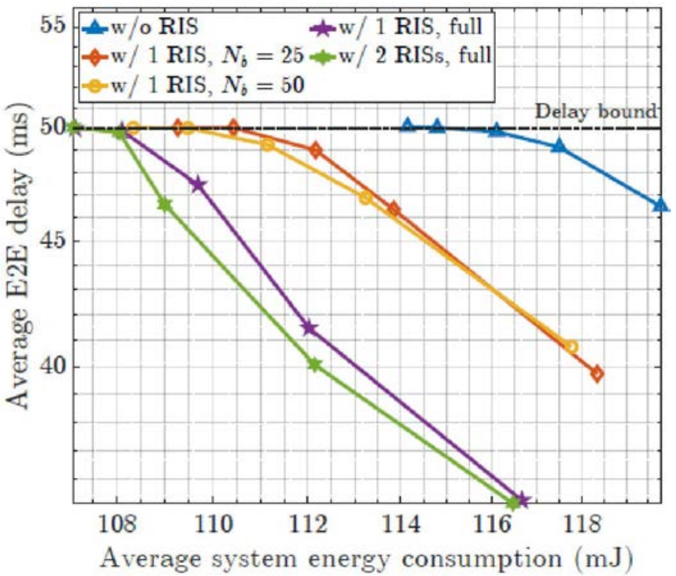
$$Y_k(t) = (-4c_k Q_k^a(t) + 4Q_k^r(t) + Z_k(t))J_k\tau$$

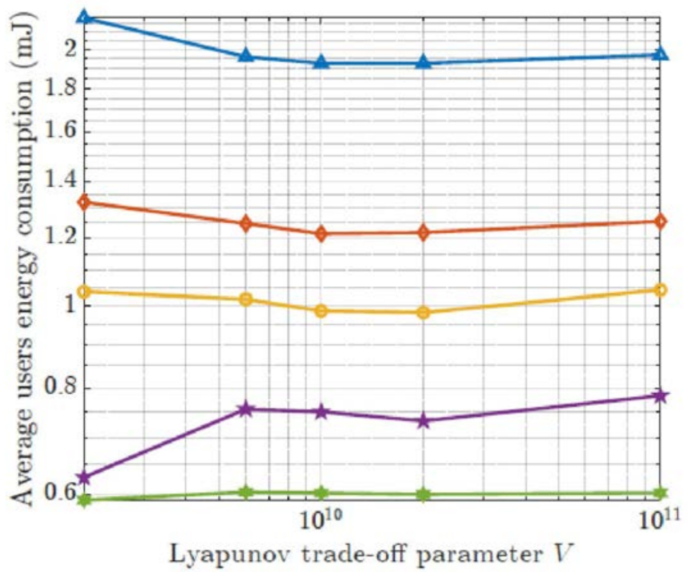
For any $f_c(t) \in \mathcal{F}$, the problem is linear in $\{f_k(t)\}_{k=1}^K$ and **the optimal frequencies can be obtained using a simple iterative procedure** (requiring at most *K* steps). Finally, the optimal ES frequency is obtained as:

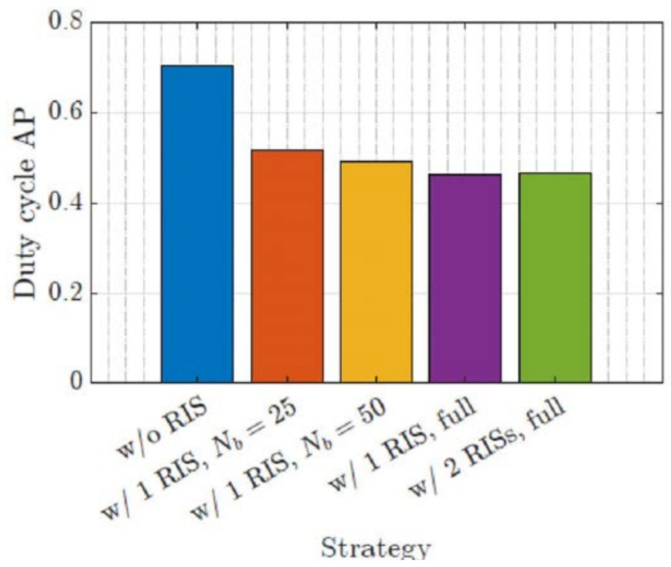
$$f_c(t) = \arg\min_{f_c \in \mathcal{F}} \quad -\sum_{k \in \mathcal{C}(t)} Y_k(t) f_k(f_c) + V(1-\sigma)\tau\gamma_s(f_c)^3$$

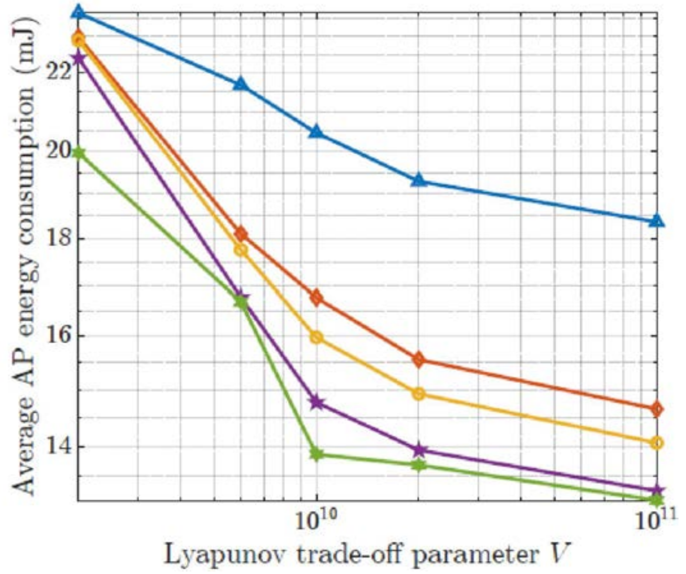
Numerical results

- K = 5users
- AP operating at 28 GHz
- Bandwidth is B = 100 MHz
- The channels between the users and RISs, and between the RISs and the AP are generated through the available tool SimRIS
- Two 8x8 RISs aid the communication
- The slot duration is equal to 10 ms.
- The average data arrival rate is 100 kbps
- The average delay constraint is 50 ms
- The ES frequency can be selected in the finite set [0; 0.01; 0.02; … ; 1] fmax, with fmax = 3.3 GHz.





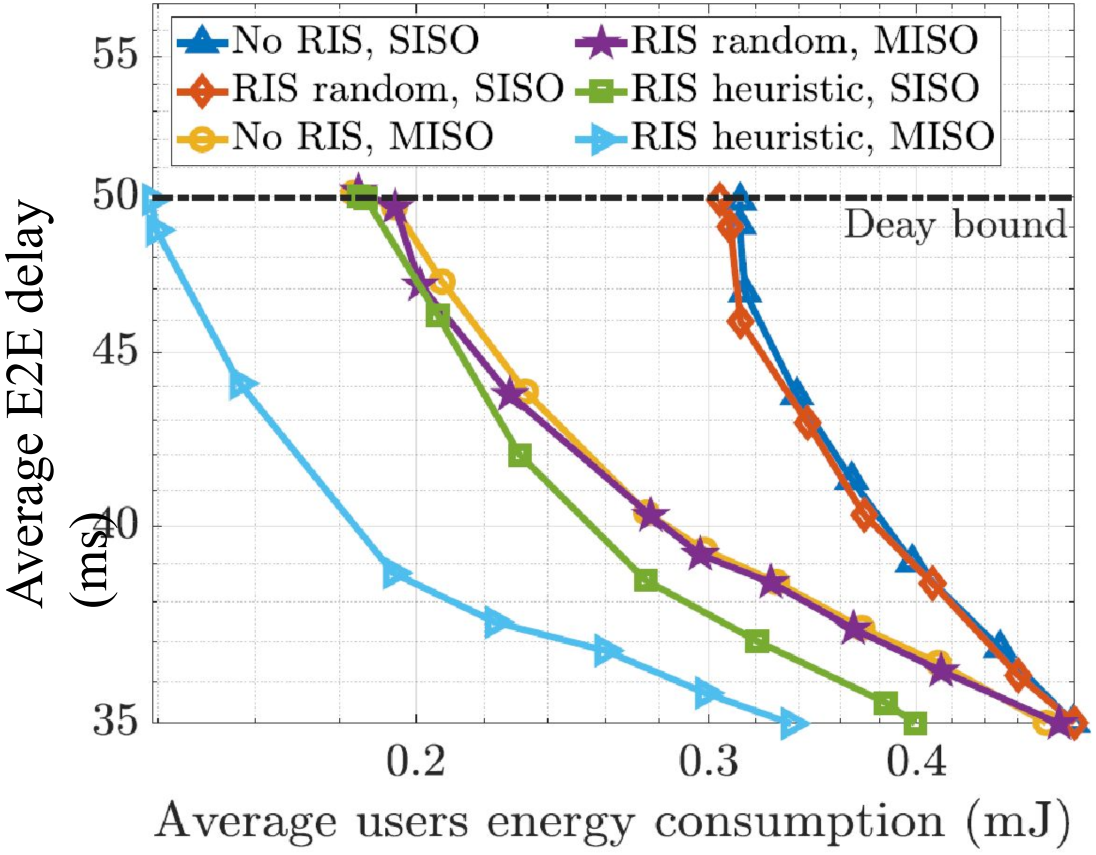


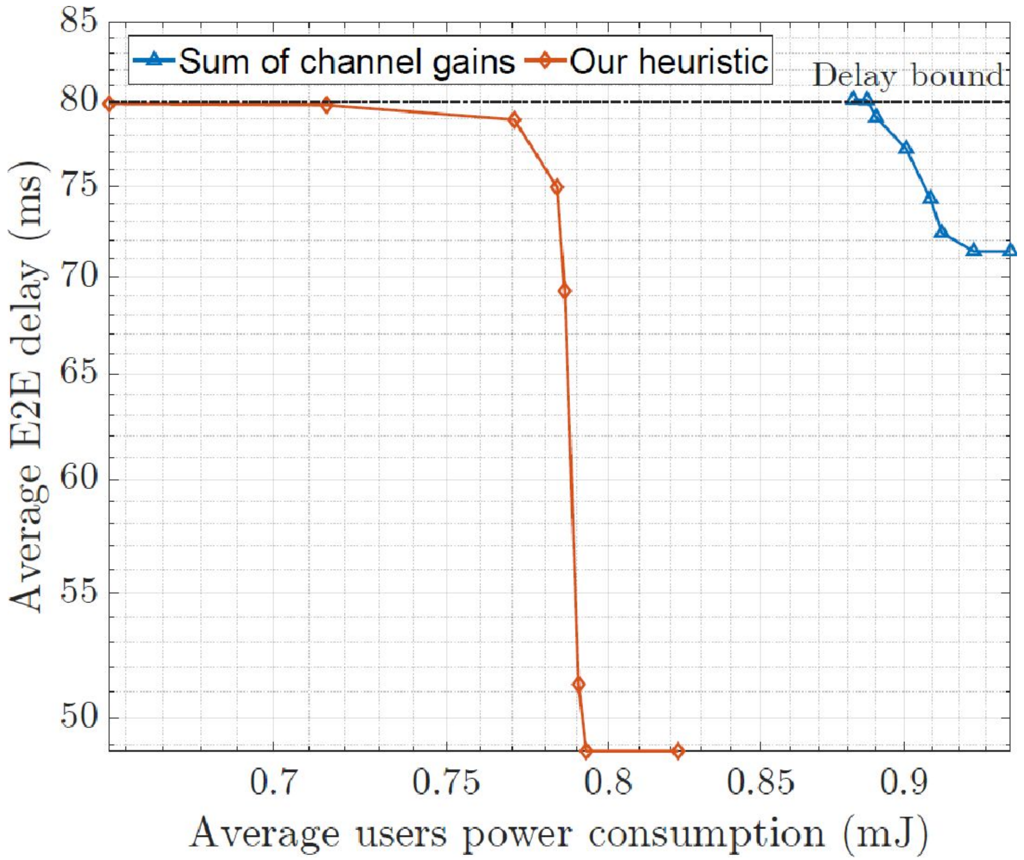


E. Basar and I. Yildirim, “Simris channel simulator for reconfigurable intelligent surface-empowered communication systems,” in 2020 IEEE Latin-American Conference on Communications, 2020, pp. 1–6.

G. C. Alexandropoulos, S. Samarakoon, M. Bennis, and M. Debbah, “Phase configuration learning in wireless networks with multiple reconfigurable intelligent surfaces,” in IEEE Globecom, 2020, pp. 1–6.

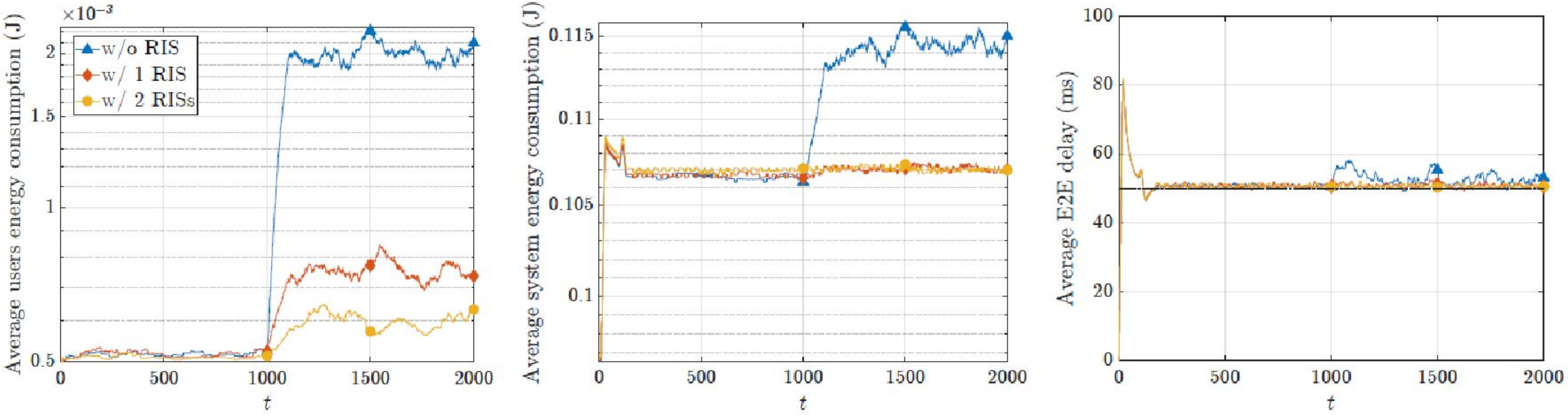
Numerical results





- RIS-aided MISO communications improve MEC performance
- The proposed RIS optimization strategy (MEC oriented) largely improves performance with respect to pure communication oriented designs

Numerical results



(a) Average users energy vs. t (b) Average system energy vs. t (c) Average delay vs. t

- **Non-stationary scenario:** A blocking occurs at iteration 1000
- The proposed method **quickly adapt** the resource allocation, illustrating the great advantage achieved by endowing MEC systems with RISs

Conclusions and research directions

- In this talk, we have illustrated the possible integration of **reconfigurable intelligent surfaces** in MEC systems
- The method hinges on **stochastic optimization** tools, allocating dynamically and jointly **the phases of RISs elements**, the **transmitting power** of users and access point, the **duty cycle** of users/RISs/AP, and the **CPU frequencies** of the edge server to enable **energy-efficient, low-latency mobile edge computing**
- Numerical results assess the performance of our dynamic offloading strategy, illustrating the potential **gain achievable by endowing MEC systems with multiple reconfigurable intelligent surfaces**
- **Current research directions:**
  - MIMO communications, control channel, frequency selective RIS models
  - Other constraints (e.g., out-of-service probability, EMF exposure, etc.)
  - RIS-empowered edge inference and learning

References

- P. Di Lorenzo, M. Merluzzi, E. Calvanese Strinati, **Dynamic Mobile Edge Computing empowered with Reconfigurable Intelligent Surfaces,** Proc. of IEEE SPAWC, Lucca, Italy, Sept. 2021.

- P. Di Lorenzo, M. Merluzzi, E. Calvanese Strinati, and S. Barbarossa, **Energy-Efficient Dynamic Edge Computing empowered by Reconfigurable Intelligent Surfaces,** submitted to EURASIP Journal on advances in Wireless Communications, 2022.