Finite-Blocklength RIS-Aided Transmit Beamforming

M. Abughalwa, H. D. Tuan, D. N. Nguyen, H. V. Poor, and L. Hanzo

***Abstract*—This paper considers the downlink of an ultra-reliable low-latency communication (URLLC) system in which a base station (BS) serves multiple single-antenna users in the short (finite) blocklength (FBL) regime with the assistance of a reconfigurable intelligent surface (RIS). In the FBL regime, the users' achievable rates are complex functions of the beamforming vectors and of the RIS's programmable reflecting elements (PREs). We propose the joint design of the transmit beamformers and PREs, the problem of maximizing the geometric mean (GM) of these rates (GM-rate) and show that this aforementioned results are providing fair rate distribution and thus reliable links to all users. A novel computational algorithm is developed, which is based on closed forms to generate improved feasible points, using its execution. The simulations show the merit of our solution.**

***Index Terms*—Reconfigurable intelligent surface, short (finite) blocklength communication, transmit beamforming, trigonometric function optimization, geometric mean maximization, non-convex optimization algorithms**

I. Introduction

Reconfigurable intelligent surfaces (RIS) constituted a planar array of passive programmable reflecting elements (PREs), intentionally augment the coverage of future wireless networks (6G) [1]–[3]. Explicitly their spectral efficiency can be maximized by the joint design of the transmit beamformer (TBF) at the BS and the RIS PREs [4]–[7].

Ultra-reliable and low-latency communication (URLLC) has also attracted recent research attention thanks to its potential applications in the internet of things (IoT), with special attention to, holographic communications, the tactile Internet, autonomous driving etc. [8], [9]. Under the URLLC framework, low-latency requires short (finite) blocklength (FBL) while ultra-reliability imposes extra low error probability constraints [10]. As a consequence, the rate function of URLLC is dependent not only on the signal-to-noise ratio (SNR) but also on the blocklength and the decoding error probability. Hence, its definition is much more computationally challenging than that of the Shannon's rate function in the long block regime. Resource allocation and transmit beamforming used for optimizing the users' rate under the FBL regime have been recently considered e.g. in [11], [12].

The work of M. Abughalwa was supported by an Australian Government Research Training Program Scholarship. The work of H. D. Tuan was supported by the Australian Research Council's Discovery Projects under Grant DP190102501. The work of H. V. Poor was supported by the U.S. National Science Foundation under Grant CNS-2128448. The work of L. Hanzo was supported by the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE), and the European Research Council's Advanced Fellow Grant QuantCom under Grant 789028.

M. Abughalwa, H. D. Tuan, and D. N. Nguyen are with the School of Electrical and Data Engineering, University of Technology Sydney, NSW 2007, Australia (email: monir.abughalwa@student.uts.edu.au, tuan.hoang@uts.ed.au, diep.nguyen@uts.ed.au).

H. V. Poor is with the Department of Electrical and Computer Engineering, Princeton University, NJ 08544, USA (email:poor@princeton.edu).

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK (email: lh@ecs.soton.ac.uk).

The authors of [13], [14] analysed RIS-aided URLLC systems of a single antenna BS and a RIS serving a single user. The more advanced joint design of the transmit beamformer at multiple BSs and RIS PREs maximizing the sum-rate subject to specific quality of service (QoS) constraints in terms of the users' rates was considered in [15]. However, the computational complexity of the algorithm proposed in [15] is extremely high, as it iterates by observing convex problems of escalating dimension. Hence Ghanem et al. [15] considered only up to 20 PREs for the RIS, even though RIS should employ very large numbers of PREs [16]. Regarding this problem, one can combine the techniques proposed in [6] and [11] to develop an algorithm, which iterates by evaluating convex problems of the same size as the original nonconvex problem. However, this size is already large for practical RIS-aided networks due to the large numbers of PREs and beamforming decision variables, which makes the computation of these convex problems not really tractable.

Against the above background, this paper provides a computationally tractable solution for the joint design of TBF and RIS PREs to optimize all users's rate in the FBL regime. Following our earlier results in [7] for optimizing all users' rates in the long blocklength (LBL) regime (Shannon rate), we now aim for maximizing the geometric means of the users' rates (GM-rate) as we explicitly demonstrate it is capable of providing a fair users' rate distribution without enforcing computationally intractable rate constraint. As a further novelty, we avoid the computationally intractable unit modulus constraints on the PREs by directly optimizing their argument. As such, the design of PREs is based on trigonometric function optimization.

The rest of the paper is organized as follows Section II is devoted to the problem statement and solution, which is supported by our simulation results provided in Section III. Finally, Section IV concludes the paper.

Notation. Only the vector variables are printed in boldface; I_N is the identity matrix of size $N \times N$; For $x = (x_1, \ldots, x_n)^T$, $\mathsf{diag}(x)$ is a diagonal matrix of the size $n \times n$ with $x_1, x_2, \ldots, x_n$ on its diagonal; $\langle x, y \rangle = x^H y$ is the dot product of the vectors x and y; The notation $X \succeq 0$ ($X \succ 0$, resp.) used for the Hermitian symmetric matrix X indicates that it is positive semi-definite (positive definite, resp.); The maximal eigenvalue of the Hermitian symmetric matrix X is denoted by $\lambda_{\max}(X)$; For a real-valued vector $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, e^{jx} is entry-wise understood, i.e. $e^{jx} = (e^{jx_1}, \ldots, e^{jx_n})^T \in \mathbb{C}^n$. $\angle x \in [0, 2\pi)$ is the argument of the complex x.

II. Problem statement

As illustrated by Fig. 1, we consider the downlink of a system, in which an M-antenna BS serves K single-antenna users (UEs) $k \in \mathcal{K} \triangleq \{1, \ldots, K\}$ with the aid of a RIS having N PREs as the BS cannot see the UEs. As the RIS is seen by the BS and the UEs are seen by the RIS line-of-sight (LoS),

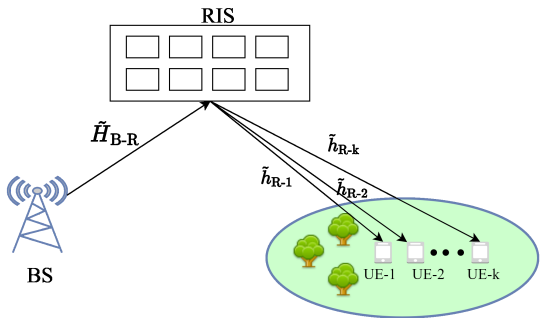

Fig. 1: System model

the channels spanning from the BS to the RIS (from the RIS to UE k, resp.) are modelled by $\tilde{H}_{\text{B-R}} = \sqrt{\beta_{\text{B-R}}} H_{\text{B-R}} \in \mathbb{C}^{N\times M}$ ($\tilde{h}_{\text{R-k}} = \sqrt{\beta_{\text{R-k}}} h_{\text{R-k}} \in \mathbb{C}^{1\times N}$, resp.), where $\sqrt{\beta_{\text{R-k}}}$ and $\sqrt{\beta_{\text{B-R}}}$ respectively represents the path-loss and large-scale fading of the RIS-to-UE k link and the BS-to-RIS link, while $h_{\text{R-k}}$ and $H_{\text{B-R}}$ are modelled by Rician fading [17].

Let $s_k \in \mathcal{C}(0,1)$ be the information symbol intended for UE k, which is beamformed by the array of weights $\mathbf{w}_k \triangleq (\mathbf{w}_k(1),\ldots,\mathbf{w}_k(M))^T \in \mathbb{C}^M$ to create the transmit signal $x = \sum_{k\in\mathcal{K}} \mathbf{w}_k s_k$. The signal received at UE k is given by

$$y_k = \mathcal{H}_k(\boldsymbol{\theta}) \sum_{k\in\mathcal{K}} \mathbf{w}_k s_k + n_k, \tag{1}$$

for $\mathcal{H}_k(\boldsymbol{\theta}) \triangleq \tilde{h}_{\text{BR-k}} \text{diag}(e^{\jmath\boldsymbol{\theta}}) H_{\text{B-R}} = \sum_{n=1}^{N} \mathcal{H}_{k,n} e^{\jmath\boldsymbol{\theta}_n} \in \mathbb{C}^{1\times M}$, with $\tilde{h}_{\text{BR-k}} \triangleq \sqrt{\beta_{\text{B-R}}}\sqrt{\beta_{\text{R-k}}} h_{\text{R-k}} \mathcal{R}_{\text{R-k}}^{1/2} \in \mathbb{C}^{1\times N}$, where $\mathcal{R}_{\text{R-k}} \in \mathbb{C}^{N\times N}$ encompasses the spatial correlation of the PREs with respect to user k [17], $n_k \in \mathcal{C}(0,\sigma)$ is the background noise, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_N)^T \in [0,2\pi)^N$ is the vector of the PREs' angles, and $\mathcal{H}_{k,n} \triangleq \tilde{h}_{\text{BR-k}} \Upsilon_n H_{\text{B-R}}$, where Υ_n is the matrix of size $N \times N$ with all-zero entries apart from $\Upsilon_n(n,n) = 1$.

For $\mathbf{w} \triangleq \{\mathbf{w}_k, k \in \mathcal{K}\}$, the effective signal-to-interference-plus-noise (SINR) at UE k is defined by

$$g_k(\mathbf{w},\boldsymbol{\theta}) = \frac{|\mathcal{H}_k(\boldsymbol{\theta})\mathbf{w}_k|^2}{\alpha_k(\mathbf{w},\boldsymbol{\theta})}, \tag{2}$$

for $\alpha_k(\mathbf{w},\boldsymbol{\theta}) \triangleq \sum_{j\in\mathcal{K}\setminus\{k\}} |\mathcal{H}_k(\boldsymbol{\theta})\mathbf{w}_j|^2 + \sigma$. In the LBL regime, the rate in nats/sec/Hz at UE k is $r_k(\mathbf{w},\boldsymbol{\theta}) = \ln[1 + g_k(\mathbf{w},\boldsymbol{\theta})]$.

Let $\mathcal{B}$ be the communication bandwidth. According to [18], by treating other terms there as Gaussian noise the achievable URLLC rate in nats/sec/Hz for the signal s_k in (1) is approximated by

$$\hat{r}_k(\mathbf{w},\boldsymbol{\theta}) \triangleq r_k(\mathbf{w},\boldsymbol{\theta}) - a v_k^{1/2}(\mathbf{w},\boldsymbol{\theta}), \tag{3}$$

where the channel included dispersion $v_k(\mathbf{w},\boldsymbol{\theta})$ under the SINR $g_k(\mathbf{w},\boldsymbol{\theta})$ is defined by [18, eq. (27)]

$$v_k(\mathbf{w},\boldsymbol{\theta}) \triangleq 2\frac{g_k(\mathbf{w},\boldsymbol{\theta})}{1+g_k(\mathbf{w},\boldsymbol{\theta})} \tag{4}$$

$$= 2\left(1 - \frac{\alpha_k(\mathbf{w},\boldsymbol{\theta})}{\beta_k(\mathbf{w},\boldsymbol{\theta})}\right), \tag{5}$$

in conjunction with

$$\beta_k(\mathbf{w},\boldsymbol{\theta}) \triangleq \alpha_k(\mathbf{w},\boldsymbol{\theta}) + |\mathcal{H}_k(\boldsymbol{\theta})\mathbf{w}_k|^2 = \sum_{j\in\mathcal{K}} |\mathcal{H}_k(\boldsymbol{\theta})\mathbf{w}_j|^2 + \sigma. \tag{6}$$

Also, $a \triangleq \frac{1}{\sqrt{\mathcal{B}t_t}} Q_G^{-1}(\epsilon^c)$, where t_t is the URLLC transmission duration, $Q_G^{-1}(.)$ is the inverse of the Gaussian Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt$, and ϵ^c is defined as an acceptable decoding error probability, which implies that under the block fading channel model considered, one out of $1/\epsilon_c$ short packets (URLLC transmissions) may experience outage.

We consider the following problem of jointly designing the beamformer $\mathbf{w}$ and the PREs $\boldsymbol{\theta}$ for maximizing the GM-rate:

$$\max_{\mathbf{w},\boldsymbol{\theta}} \hat{f}(\mathbf{w},\boldsymbol{\theta}) \triangleq \left(\prod_{k=1}^{K} \hat{r}_k(\mathbf{w},\boldsymbol{\theta})\right)^{1/K} \tag{7a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} ||\mathbf{w}_k||^2 \le P, \tag{7b}$$

where (7b) sets the transmit sum power constraint within a given power budget P. Our previous paper [7], which considered the problem

$$\max_{\mathbf{w},\boldsymbol{\theta}} f(\mathbf{w},\boldsymbol{\theta}) \triangleq \left(\prod_{k=1}^{K} r_k(\mathbf{w},\boldsymbol{\theta})\right)^{1/K} \quad \text{s.t.} \quad (7b) \tag{8}$$

in the LBL regime showed that GM-rate maximization naturally leads to fair user rate distributions without imposing the rate constraints of $r_k(\mathbf{w},\boldsymbol{\theta}) \ge \bar{r}$, which are nonconvex and thus computationally intractable. As a compelling benefit, by directly optimizing the angles $\boldsymbol{\theta} \in [0,2\pi)^N$ of PREs, both (7) and (8) circumvent the unit modulus constraints on the latters.

Compared to the rate function r_k of the LBL regime, we have the rate-reduction term $v_k^{1/2}(\mathbf{w},\boldsymbol{\theta})$ arisen in the rate function $\hat{r}_k$ in the FBL regime. Therefore, the main challenge in considering (7) is to handle this term.

Initialized by $(w^{(0)},\theta^{(0)})$ as the optimal solution of (8) computed by [7], let $(w^{(\kappa)},\theta^{(\kappa)})$ be a feasible point for (7) that is found from the $(\kappa-1)$-st round, and

$$\gamma_k^{(\kappa)} \triangleq \frac{\max_{k'\in\mathcal{K}} \hat{r}_{k'}(w^{(\kappa)},\theta^{(\kappa)})}{\hat{r}_k(w^{(\kappa)},\theta^{(\kappa)})}, k \in \mathcal{K}. \tag{9}$$

As discussed in [7], the descent iterations are based on the following problem

$$\max_{\mathbf{w},\boldsymbol{\theta}} \hat{f}^{(\kappa)}(\mathbf{w},\boldsymbol{\theta}) \triangleq \sum_{k=1}^{K} \gamma_k^{(\kappa)} \hat{r}_k(\mathbf{w},\boldsymbol{\theta}) \quad \text{s.t.} \quad (7b). \tag{10}$$

A. Beamforming descent iteration

We seek $w^{(\kappa+1)}$ satisfying:

$$\hat{f}^{(\kappa)}(w^{(\kappa+1)},\theta^{(\kappa)}) > \hat{f}^{(\kappa)}(w^{(\kappa)},\theta^{(\kappa)}). \tag{11}$$

Define $r_{b,k}^{(\kappa)}(\mathbf{w}) \triangleq r_k(\mathbf{w},\theta^{(\kappa)}), v_{b,k}^{(\kappa)}(\mathbf{w}) \triangleq v_k(\mathbf{w},\theta^{(\kappa)})$, $\hat{r}_{b,k}^{(\kappa)}(\mathbf{w}) \triangleq \hat{r}_k(\mathbf{w},\theta^{(\kappa)})$, $\alpha_{b,k}^{(\kappa)}(\mathbf{w}) \triangleq \alpha_k(\mathbf{w},\theta^{(\kappa)})$, $\beta_{b,k}^{(\kappa)}(\mathbf{w}) \triangleq \beta_k(\mathbf{w},\theta^{(\kappa)})$, $v_{b,k}^{(\kappa)}(\mathbf{w}) \triangleq v_k(\mathbf{w},\theta^{(\kappa)})$, and $\mathcal{H}_k^{(\kappa)} \triangleq \mathcal{H}_k(\theta^{(\kappa)})$. As such,

$$\hat{f}^{(\kappa)}(\mathbf{w},\theta^{(\kappa)}) = \sum_{k=1}^{K} \gamma_k^{(\kappa)} \hat{r}_{b,k}^{(\kappa)}(\mathbf{w}), \tag{12}$$

and

$$\hat{r}_{b,k}^{(\kappa)}(\mathbf{w}) = r_{b,k}^{(\kappa)}(\mathbf{w}) - a\sqrt{v_{b,k}^{(\kappa)}(\mathbf{w})}. \tag{13}$$

The following lower bounding concave approximation of $r_{b,k}^{(\kappa)}(\mathbf{w})$ was obtained in [6]:

$$r_{b,k}^{(\kappa)}(\mathbf{w}) \geq a_{k,1}^{(\kappa)} + 2\Re\{\langle b_{k,k}^{(\kappa)}, \mathbf{w}_k\rangle\} - c_{k,1}^{(\kappa)} \sum_{j=1}^{K} |\mathcal{H}_k^{(\kappa)}\mathbf{w}_j|^2, \quad (14)$$

with $a_{k,1}^{(\kappa)} \triangleq r_{b,k}^{(\kappa)}(w^{(\kappa)}) - g_k(\theta^{(\kappa)}, w^{(\kappa)}) - \sigma c_{k,1}^{(\kappa)}$, $b_{k,k}^{(\kappa)} \triangleq (\mathcal{H}_k^{(\kappa)})^H \mathcal{H}_k^{(\kappa)} w_k^{(\kappa)} / \alpha_{b,k}^{(\kappa)}(w^{(\kappa)})$, $0 < c_{k,1}^{(\kappa)} \triangleq 1/\alpha_{b,k}^{(\kappa)}(w^{(\kappa)}) - 1/\beta_{b,k}^{(\kappa)}(w^{(\kappa)})$.

Our next step is now to develop an upper bounding convex approximation of $\sqrt{v_{b,k}^{(\kappa)}(\mathbf{w})}$, which together with (14) gives a lower bounding concave approximation of $\hat{r}_{b,k}^{(\kappa)}$ in (13).

Using the inequality

$$\sqrt{x} \leq \frac{\sqrt{\bar{x}}}{2}\left(1 + \frac{x}{\bar{x}}\right) \quad \forall\, x > 0, \bar{x} > 0, \quad (15)$$

gives

$$\begin{aligned}\sqrt{v_{b,k}^{(\kappa)}(\mathbf{w})} \;\leq\; & \frac{\sqrt{v_{b,k}^{(\kappa)}(w^{(\kappa)})}}{2}\left(1 + \frac{2}{v_{b,k}^{(\kappa)}(w^{(\kappa)})}\right) \\ & - \frac{1}{\sqrt{v_{b,k}^{(\kappa)}(w^{(\kappa)})}} \frac{\alpha_{b,k}^{(\kappa)}(\mathbf{w})}{\beta_{b,k}^{(\kappa)}(\mathbf{w})}. \end{aligned} \quad (16)$$

Applying the following inequality for $x \in \mathbb{C}^n$, $\bar{x} \in \mathbb{C}^n$, $y > 0$, $\bar{y} > 0$, and $\sigma > 0$,

$$\frac{||x||^2}{y+\sigma} \geq \frac{||\bar{x}||^2}{\bar{y}+\sigma}\left(2\frac{\Re\{\bar{x}^H x\}}{||\bar{x}||^2} - \frac{y+\sigma}{\bar{y}+\sigma}\right) \quad (17)$$

yields:

$$\begin{aligned}&\frac{\alpha_{b,k}^{(\kappa)}(\mathbf{w})}{\beta_{b,k}^{(\kappa)}(\mathbf{w})} \geq \frac{\alpha_{b,k}^{(\kappa)}(w^{(\kappa)})}{\beta_{b,k}^{(\kappa)}(w^{(\kappa)})} \\ &\times \left(2\frac{\sum_{j\in\mathcal{K}\setminus\{k\}} \Re\{(w_j^{(\kappa)})^H[(\mathcal{H}_k^{(\kappa)})^H]^2\mathbf{w}_j\} + \sigma}{\alpha_{b,k}^{(\kappa)}(w^{(\kappa)})} - \right. \\ &\left. \frac{\beta_{b,k}^{(\kappa)}(\mathbf{w})}{\beta_{b,k}^{(\kappa)}(w^{(\kappa)})}\right),\end{aligned}$$

which together with (16) gives the following upper bounding convex approximation of $a\sqrt{v_{b,k}^{(\kappa)}(\mathbf{w})}$:

$$\begin{aligned}a\sqrt{v_{b,k}^{(\kappa)}(\mathbf{w})} \;\leq\; & a_{k,2}^{(\kappa)} - 2\sum_{j\in\mathcal{K}\setminus\{k\}} \Re\{\langle b_{k,j}^{(\kappa)}, \mathbf{w}_j\rangle\} + \\ & c_{k,2}^{(\kappa)} \sum_{j=1}^{K} |\mathcal{H}_k^{(\kappa)}\mathbf{w}_j|^2, \end{aligned} \quad (18)$$

for $a_{k,2}^{(\kappa)} \triangleq a\frac{\sqrt{v_{b,k}^{(\kappa)}(w^{(\kappa)})}}{2}\left(1 + \frac{2}{v_{b,k}^{(\kappa)}(w^{(\kappa)})}\right) + \sigma\frac{\alpha_{b,k}^{(\kappa)}(w^{(\kappa)})}{\beta_{b,k}^{(\kappa)}(w^{(\kappa)})\sqrt{v_{b,k}^{(\kappa)}(w^{(\kappa)})}}\left(\frac{-2}{\alpha_{b,k}^{(\kappa)}(w^{(\kappa)})} + \frac{1}{\beta_{b,k}^{(\kappa)}(w^{(\kappa)})}\right)$, $b_{k,j}^{(\kappa)} \triangleq \frac{a}{\beta_{b,k}^{(\kappa)}(w^{(\kappa)})\sqrt{v_{b,k}^{(\kappa)}(w^{(\kappa)})}}(\mathcal{H}_k^{(\kappa)})^H\mathcal{H}_k^{(\kappa)}w_j^{(\kappa)}$, $j \in \mathcal{K}\setminus\{k\}$, and $c_{k,2}^{(\kappa)} \triangleq a\frac{\alpha_{b,k}^{(\kappa)}(w^{(\kappa)})}{(\beta_{b,k}^{(\kappa)}(w^{(\kappa)}))^2\sqrt{v_{b,k}^{(\kappa)}(w^{(\kappa)})}}$.

The bounds (14) and (18) yield the following lower bounding concave approximation for $\hat{r}_{b,k}^{(\kappa)}(\mathbf{w})$ in (13):

$$\hat{r}_{b,k}^{(\kappa)}(\mathbf{w}) \geq a_k^{(\kappa)} + 2\sum_{j=1}^{K} \Re\{\langle b_{k,j}^{(\kappa)}, \mathbf{w}_j\rangle - c_k^{(\kappa)} \sum_{j\in\mathcal{K}} |\mathcal{H}_k^{(\kappa)}\mathbf{w}_j|^2, \quad (19)$$

for $a_k^{(\kappa)} \triangleq a_{k,1}^{(\kappa)} - a_{k,2}^{(\kappa)}$, and $c_k^{(\kappa)} \triangleq c_{k,1}^{(\kappa)} + c_{k,2}^{(\kappa)}$.

We now generate $w^{(\kappa+1)}$ as the optimal solution of the following problem

$$\max_{\mathbf{w}} \; \hat{f}_b^{(\kappa)}(\mathbf{w}) \quad \text{s.t.} \quad (7b), \quad (20)$$

where $\hat{f}_b^{(\kappa)}(\mathbf{w}) \triangleq \sum_{k=1}^{K}\gamma_k^{(\kappa)}a_k^{(\kappa)} + 2\sum_{k=1}^{K}\Re\{\langle b_k^{(\kappa)}, \mathbf{w}_k\rangle\} - \sum_{k=1}^{K}(\mathbf{w}_k)^H\Psi_b^{(\kappa)}\mathbf{w}_k$ with $b_k^{(\kappa)} \triangleq \sum_{j=1}^{K}\gamma_j^{(\kappa)}b_{j,k}^{(\kappa)}$, and $0 \preceq \Psi_b^{(\kappa)} \triangleq \sum_{j=1}^{K}\gamma_j^{(\kappa)}c_j^{(\kappa)}(\mathcal{H}_j^{(\kappa)})^H\mathcal{H}_j^{(\kappa)}$. It can be readily checked that

$$\hat{f}^{(\kappa)}(w^{(\kappa)}, \theta^{(\kappa)}) = \hat{f}_b^{(\kappa)}(w^{(\kappa)}). \quad (21)$$

The problem (20) admits the following closed-form solution

$$w_k^{(\kappa+1)} = \begin{cases} (\Psi_b^{(\kappa)})^{-1}b_k^{(\kappa)} & \text{if} \quad \sum_{k=1}^{K}||(\Psi_b^{(\kappa)})^{-1}b_k^{(\kappa)}||^2 \leq P \\ \left(\Psi_b^{(\kappa)} + \mu I_M\right)^{-1} b_k^{(\kappa)} & \text{otherwise,} \end{cases} \quad (22)$$

where $\mu > 0$ is chosen by bisection such that $\sum_{k=1}^{K}||\left(\Psi_b^{(\kappa)} + \mu I_M\right)^{-1}b_k^{(\kappa)}||^2 = P$.

It follows from (12) and (19) that $\hat{f}^{(\kappa)}(w^{(\kappa+1)}, \theta^{(\kappa)}) \geq \hat{f}_b^{(\kappa)}(w^{(\kappa+1)})$, while $\hat{f}_b^{(\kappa)}(w^{(\kappa+1)}) > \hat{f}_b^{(\kappa)}(w^{(\kappa)}) = \hat{f}^{(\kappa)}(w^{(\kappa)}, \theta^{(\kappa)})$, because $w^{(\kappa+1)}$ and $w^{(\kappa)}$ represent the optimal solution and a feasible point for (20). We thus have (11) as sought.

B. Programmable reflecting elements' descent iteration

We seek the next iterative point $\theta^{(\kappa+1)}$ such that

$$\hat{f}^{(\kappa)}(w^{(\kappa+1)}, \theta^{(\kappa+1)}) > \hat{f}^{(\kappa)}(w^{(\kappa+1)}, \theta^{(\kappa)}). \quad (23)$$

Define $r_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \triangleq r_k(w^{(\kappa+1)}, \boldsymbol{\theta})$, $v_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \triangleq v_k(w^{(\kappa+1)}, \boldsymbol{\theta})$, $\hat{r}_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \triangleq \hat{r}_k(w^{(\kappa+1)}, \boldsymbol{\theta})$, $\alpha_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \triangleq \alpha_k(w^{(\kappa+1)}, \boldsymbol{\theta})$, $\beta_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \triangleq \beta_k(w^{(\kappa+1)}, \boldsymbol{\theta})$, $v_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \triangleq v_k(w^{(\kappa+1)}, \boldsymbol{\theta})$, and $\ell_{k,j}^{(\kappa)} \triangleq \mathcal{H}_k(\theta^{(\kappa)})w_j^{(\kappa+1)}$, $(k,j) \in \mathcal{K}\times\mathcal{K}$. As such,

$$\hat{f}^{(\kappa)}(w^{(\kappa+1)}, \boldsymbol{\theta}) = \sum_{k=1}^{K}\gamma_k^{(\kappa)}\hat{r}_{p,k}^{(\kappa)}(\boldsymbol{\theta}), \quad (24)$$

and

$$\hat{r}_{p,k}^{(\kappa)}(\boldsymbol{\theta}) = r_{p,k}^{(\kappa)}(\boldsymbol{\theta}) - a\sqrt{v_{p,k}^{(\kappa)}(\boldsymbol{\theta})}. \quad (25)$$

Recall that $\mathcal{H}_{k,n}$ are defined in (1). The following lower bounding approximation of $r_{p,k}^{(\kappa)}(\boldsymbol{\theta})$ was obtained in [6]:

$$\begin{aligned}r_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \;\geq\; & \tilde{a}_{k,1}^{(\kappa)} + 2\Re\{\sum_{n=1}^{N}\tilde{b}_{k,1}^{(\kappa)}(n)e^{\jmath\boldsymbol{\theta}_n}\} - \\ & \tilde{c}_{k,1}^{(\kappa)}(e^{\jmath\boldsymbol{\theta}})^H\Psi_p^{(\kappa)}e^{\jmath\boldsymbol{\theta}}, \end{aligned} \quad (26)$$

with $\tilde{a}_{k,1}^{(\kappa)} \triangleq r_{p,k}^{(\kappa)}(\theta^{(\kappa)}) - g_k(\theta^{(\kappa)}, w^{(\kappa+1)}) - \sigma\tilde{c}_{k,1}^{(\kappa)}$, $0 < \tilde{c}_{k,1}^{(\kappa)} \triangleq 1/\alpha_{p,k}^{(\kappa)}(\theta^{(\kappa)}) - 1/\beta_{p,k}^{(\kappa)}(\theta^{(\kappa)})$, and $\tilde{b}_{k,1}^{(\kappa)}(n) \triangleq$

$\frac{1}{\alpha_{p,k}^{(\kappa)}(\theta^{(\kappa)})}(\ell_{k,k}^{(\kappa)})^H\mathcal{H}_{k,n}w_k^{(\kappa+1)}$, $n=1,\ldots N$, $\mathcal{W}^{(\kappa+1)} \triangleq \sum_{j=1}^K[w_j^{(\kappa+1)}]^2$, $\Psi_p^{(\kappa)}(n',n) \triangleq \langle \mathcal{H}_{k,n}\mathcal{W}^{(\kappa+1)}\mathcal{H}_{k,n'}^H\rangle$, $n'=1,\ldots,N$; $n=1,\ldots,N$.

Similarly to (18), we have

$$a\sqrt{v_{p,k}^{(\kappa)}(\boldsymbol{\theta})} \leq \tilde{a}_{k,2}^{(\kappa)} - 2\Re\{\sum_{n=1}^N \tilde{b}_{k,2}^{(\kappa)}(n)e^{\jmath\boldsymbol{\theta}_n}\} + \tilde{c}_{k,2}^{(\kappa)}\sum_{j=1}^K(e^{\jmath\boldsymbol{\theta}})^H\Psi_p^{(\kappa)}e^{\jmath\boldsymbol{\theta}}, \quad (27)$$

for $\tilde{a}_{k,2}^{(\kappa)} \triangleq a\frac{\sqrt{v_{p,k}^{(\kappa)}(\theta^{(\kappa)})}}{2}\left(1+\frac{2}{v_{p,k}^{(\kappa)}(\theta^{(\kappa)})}\right) + \sigma\frac{\alpha_{p,k}^{(\kappa)}(\theta^{(\kappa)})}{\beta_{p,k}^{(\kappa)}(\theta^{(\kappa)})\sqrt{v_{p,k}^{(\kappa)}(\theta^{(\kappa)})}}\left(\frac{-2}{\alpha_{p,k}^{(\kappa)}(\theta^{(\kappa)})}+\frac{1}{\beta_{p,k}^{(\kappa)}(\theta^{(\kappa)})}\right)$, and $\tilde{b}_{k,2}^{(\kappa)}(n) \triangleq \frac{a}{\beta_{p,k}^{(\kappa)}(\theta^{(\kappa)})\sqrt{v_{p,k}^{(\kappa)}(\theta^{(\kappa)})}}\sum_{j\in\mathcal{K}\setminus\{k\}}(\ell_{k,j}^{(\kappa)})^H\mathcal{H}_{k,n}w_j^{(\kappa+1)}$, $n=1,\ldots,N$, and $\tilde{c}_{k,2}^{(\kappa)} \triangleq a\frac{\alpha_{p,k}^{(\kappa)}(\theta^{(\kappa)})}{(\beta_{p,k}^{(\kappa)}(\theta^{(\kappa)}))^2\sqrt{v_{p,k}^{(\kappa)}(\theta^{(\kappa)})}}$. Based on (26) and (27) we obtain the following lower bound:

$$\hat{r}_{p,k}^{(\kappa)}(\boldsymbol{\theta}) \geq \tilde{a}_k^{(\kappa)} + 2\Re\{\sum_{n=1}^N \tilde{b}_k^{(\kappa)}(n)e^{\jmath\boldsymbol{\theta}_n}\} - \tilde{c}_k^{(\kappa)}(e^{\jmath\boldsymbol{\theta}})^H\Psi_p^{(\kappa)}e^{\jmath\boldsymbol{\theta}}, \quad (28)$$

where $\tilde{a}_k^{(\kappa)} \triangleq \tilde{a}_{k,1}^{(\kappa)} - \tilde{a}_{k,2}^{(\kappa)}$, $\tilde{c}_k^{(\kappa)} \triangleq \tilde{c}_{k,1}^{(\kappa)} + \tilde{c}_{k,2}^{(\kappa)}$), and $\tilde{b}_k^{(\kappa)} \triangleq \tilde{b}_{k,1}^{(\kappa)} + \tilde{b}_{k,2}^{(\kappa)}$. Then,

$$\hat{f}^{(\kappa)}(w^{(\kappa+1)},\boldsymbol{\theta}) \geq \tilde{a}^{(\kappa)} + 2\Re\{\sum_{n=1}^N \tilde{b}^{(\kappa)}(n)e^{\jmath\boldsymbol{\theta}_n}\} - (e^{\jmath\boldsymbol{\theta}})^H\hat{\Psi}_p^{(\kappa)}e^{\jmath\boldsymbol{\theta}}, \quad (29)$$

for $\tilde{a}^{(\kappa)} \triangleq \sum_{k=1}^K\gamma_k^{(\kappa)}\tilde{a}_k^{(\kappa)}$, $\tilde{b}^{(\kappa)}(n) \triangleq \sum_{k=1}^K\gamma_k^{(\kappa)}\tilde{b}_k^{(\kappa)}(n), n=1,\ldots,N$, and $0 \preceq \hat{\Psi}_p^{(\kappa)} \triangleq \left(\sum_{k=1}^K\gamma_k^{(\kappa)}\tilde{c}_k^{(\kappa)}\right)\Psi_p^{(\kappa)}$. Furthermore, we have

$$\text{RHS of (29)} \geq \hat{f}_p^{(\kappa)}(\boldsymbol{\theta}) \quad (30)$$

for $\hat{f}_p^{(\kappa)}(\boldsymbol{\theta}) \triangleq \tilde{a}^{(\kappa)} + 2\Re\{\sum_{n=1}^N(\tilde{b}^{(\kappa)}(n) - \sum_{m=1}^N e^{-\jmath\theta_m^{(\kappa)}}\hat{\Psi}_p^{(\kappa)}(m,n) + \lambda_{\max}(\hat{\Psi}_p^{(\kappa)})e^{-\jmath\theta_n^{(\kappa)}})e^{\jmath\boldsymbol{\theta}_n}\} - (e^{\jmath\theta^{(\kappa)}})^H\hat{\Psi}_p^{(\kappa)}e^{\jmath\theta^{(\kappa)}} - 2\lambda_{\max}(\hat{\Psi}_p^{(\kappa)})N$. We thus generate $\theta^{(\kappa+1)}$ as the optimal solution of the problem

$$\max_{\boldsymbol{\theta}} \ \tilde{\hat{f}}_p^{(\kappa)}(\boldsymbol{\theta}), \quad (31)$$

which admits the closed-form solution[1] of

$$\theta_n^{(\kappa+1)} = 2\pi - \angle\left(\tilde{b}^{(\kappa)}(n) - \sum_{m=1}^N e^{-\jmath\theta_m^{(\kappa)}}\hat{\Psi}_p^{(\kappa)}(m,n) + \lambda_{\max}(\hat{\Psi}_p^{(\kappa)})e^{-\jmath\theta_n^{(\kappa)}}\right), n=1,\ldots,N. \quad (32)$$

It follows from (30) that $\hat{f}^{(\kappa)}(w^{(\kappa+1)},\theta^{(\kappa+1)}) \geq \hat{f}_p^{(\kappa)}(\theta^{(\kappa+1)}) \geq \tilde{\hat{f}}_p^{(\kappa)}(\theta^{(\kappa+1)}) > \tilde{\hat{f}}_p^{(\kappa)}(\theta^{(\kappa)}) = \hat{f}_p^{(\kappa)}(\theta^{(\kappa)}) = \hat{f}^{(\kappa)}(w^{(\kappa+1)},\theta^{(\kappa)})$, confirming (23). Hence $\theta^{(\kappa+1)}$ is a better feasible point than $\theta^{(\kappa)}$.

[1] $[(\hat{\Psi}_p^{(\kappa)} - \mu I_N)e^{\jmath\theta^{(\kappa)}}](n)$ is the n-th entry of $(\hat{\Psi}_p^{(\kappa)} - \mu I_N)e^{\jmath\theta^{(\kappa)}}$

C. Algorithm

Algorithm 1 provides the pseudo-code for the proposed steep descent computational procedure of (7) as the iterations (22) and (32) seek a descent direction by seeking a better feasible point for the nonconvex problem (7).

Algorithm 1 URLLC GM-rate descent algorithm

1: **Initialization:** Use the Algorithm [7] to initialize a feasible $(w^{(0)},\theta^{(0)})$. Set $\kappa=0$.
2: **κ-th iteration:** Generate $w^{(\kappa+1)}$ by (22) and $\theta^{(\kappa+1)}$ by (32). Given the convergence tolerance ν_t, stop if $|\hat{f}(w^{(\kappa+1)},\theta^{(\kappa+1)}) - \hat{f}(w^{(\kappa)},\theta^{(\kappa)})|/\hat{f}(w^{(\kappa)},\theta^{(\kappa)}) \leq \nu_t$. Reset $\kappa \leftarrow \kappa+1$.
3: **Output** $(w^{(\kappa)},\theta^{(\kappa)})$ and URRLC rates $\hat{r}_k(w^{(\kappa)},\theta^{(\kappa)})$, $k \in \mathcal{K}$ with their GM $\left(\prod_{k=1}^K\hat{r}_k(w^{(\kappa)},\theta^{(\kappa)})\right)^{1/K}$.

Remark. One can see that the above Algorithm invokes the nonconvex problem (10) at each iteration, which is a problem of weighted sum rate maximization associated with the weights $\gamma_k^{(\kappa)}$ iteratively updated according to (9), to generate a better feasible point.

III. Numerical examples

This section evaluates the performance of the proposed algorithm using numerical examples. The set up is the same as that in [7]: the large-scale fading and the RIS-to-UE k path-loss is $\beta_{\text{R-k}} = G_{\text{RIS}} - 33.05 - 30\log_{10}(d_{\text{R-k}})$ dB, where $d_{\text{R-k}}$ is the distance between the RIS and UE k in meters, while G_{IRS} is the antenna gain of the RIS elements [16], [17]. The large-scale fading and the path-loss between BS and RIS is $\beta_{\text{B-R}} = G_{\text{BS}} + G_{\text{IRS}} - 35.9 - 22\log_{10}(d_{\text{B-R}})$ dB, where $d_{\text{B-R}}$ is the distance between the BS and RIS in meters, while G_{BS} is the BS antenna gain [16], [17]. The coordinates of the BS and the RIS are $(20,0,25)$ and $(0,30,40)$, while the users are randomly located in a $(60m\times 60m)$ area to the right of the BS and RIS. The entries of the BS-to-RIS LoS channel matrix are $[H_{B-R}]_{n,m} = e^{j\pi\left((n-1)\sin\bar{\theta}_n\sin\bar{\phi}_n + (m-1)\sin\theta_n\sin\phi_n\right)}$, where θ_n and ϕ_n are uniformly distributed over $(0,\pi)$ and $(0,2\pi)$, and $\bar{\theta}_n = \pi - \theta_n$, $\bar{\phi}_n = \pi + \phi_n$. The small-scale fading channel gain $h_{\text{R-k}}$ follows the Rician distribution having K-factor of 3. The spatial correlation matrix is $[R_{R-k}]_{n,n'} = e^{j\pi(n-n')\sin\tilde{\phi}\sin\tilde{\theta}}$, where $\tilde{\phi}$ and $\tilde{\theta}$ are the azimuth and elevation angle for UE k, respectively. Unless stated otherwise, the following parameters have been used in our simulation, $G_{\text{BS}} = G_{\text{IRS}} = 5$ dBi, $\mathcal{B} = 1$ Mhz, $\sigma^2 = -174$ dBm/Hz, $M=10$, $K=10$, $P=20$ dBm, $N=100$, $\epsilon^c = 10^{-5}$, and $t_t = 0.1$ ms which is suitable for URLLC transmission [9], and the choice of 1 ms end-to-end delay ensures having a quasi-static channel during URLLC communication [19]. The results are multiplied by $\log_2(e)$ to convert the unit nats/sec into the unit bps/Hz. Lastly we set the convergence tolerance ν_t to 10^{-3}.

Furthermore, we use the following terms for interpreting the results:

- LBR refers to the performance in LBL regime [7].
- URLLC refers to the performance by Algorithm 1.

The problem of maximizing the sum rate (SR) can also be solved by Algorithm 1 upon setting $\gamma_k^{(\kappa)} \equiv 1$ in (9). Note

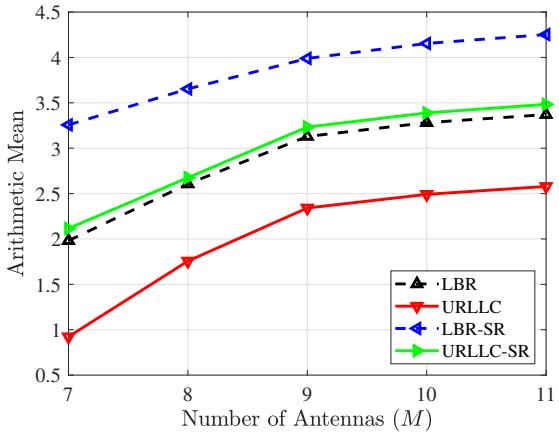

Fig. 2: AM-rate versus M

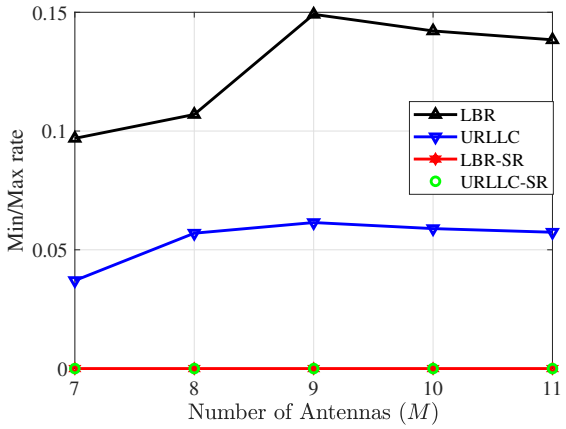

Fig. 3: The ratio of the highest and lowest users rates versus M

that the maximization of the SR represents the maximization of the arithmetic mean (AM) of the users' rates (AM-rate) because the latter is defined as the former divided by K. Fig. 2 plots the AM-rate achieved by maximizing the GM-rate and SR. As expected, the SR maximization achieves better AM-rate than GM-rate maximization. However, SR maximization is unable to provide fairness for all users, as it assigns some users having low channel quality zero rate, which can be seen in Fig. 3 that plots the ratio between the minimum user rate and the maximum user rate (RR). As it can be seen, the RR under SR maximization is always zero because SR maximization cannot avoid having zero rate in either LBR or in URLLC. This remains the case even when the number M of transmit antennas is higher than the number K of users. By contrast, GM maximization manages to assign nonzero rates to all users, even when M is lower than K. This demonstrates that using GM rate maximization is capable of improving all users' rates. Furthermore, Fig. 4 portrays the users' rate variance (URV) versus M achieved by GM-rate maximization and SR maximization. As expected the URV attained by SR-maximization is very high, as it tends to assign high fraction of the total SR to a few users. By contrast, the URV of GM-rate maximization is low and in fact it is not sensitive to the number M of transmit antennas.

Fig. 5 plots the GM rate versus M. As expected, the GM rate increases with M, since the system's ability to mitigate the multi user interference improving with M, especially when M is higher than or equal to K. The URLLC GM rate increases similarly to the LBR GM rate. However, the gap between the LBR-GM rate and the URLLC-GM rate does not decrease with M increasing.

Fig. 6 depicts the GM rate of LBR and URLLC versus P, observe that as anticipated the GM rate increases with P. Furthermore, the URLLC behaviour is similar to LBR with the gap between the rates being almost the same, which indicates that increasing P does not affect the overall URLLC rate.

Fig. 7 portrays the URLLC-GM rate against t_t, which increases with t_t but the rate increase gradually slows down. Nevertheless, even for low t_t the system is still able to achieve a good rate, which represents the advantages of our Algorithm. Note that the LBR assumes the transmission duration to be ∞, therefore it serve as an upper bound for the URLLC. Moreover

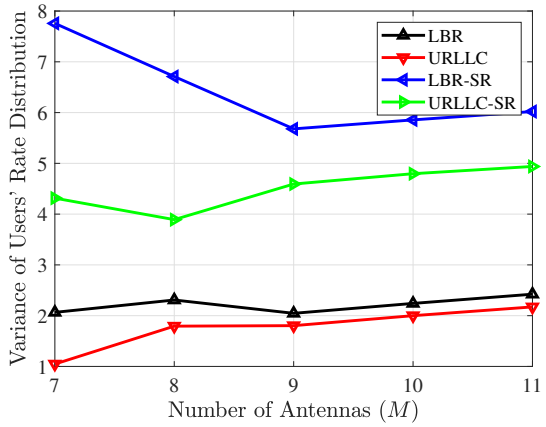

Fig. 4: Users' Rate Variance versus M

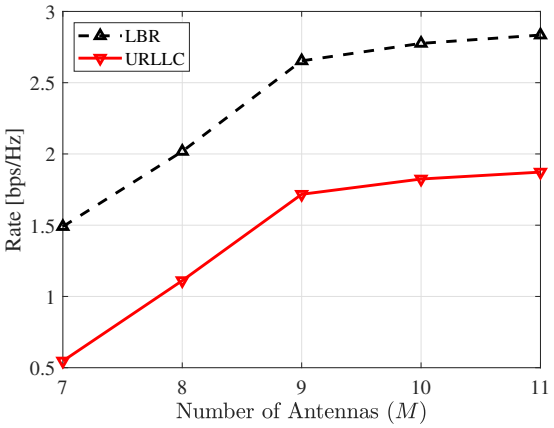

Fig. 5: GM Rate versus M

Fig. 7 can be used for choosing t_t depending on the quality of service required.

Lastly, Fig. 8 plots the RR versus t_t. The SR maximization in FBL cannot avoid having zero rate even for long transmission duration of $t_t = 0.1$ ms, while GM maximization always assigns fair rates to all the users, regardless of the transmission

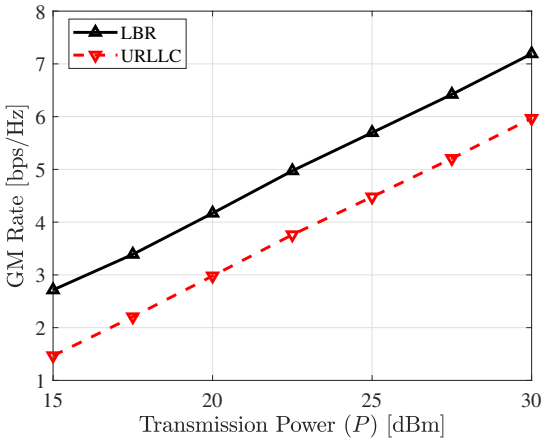

Fig. 6: GM rate versus P

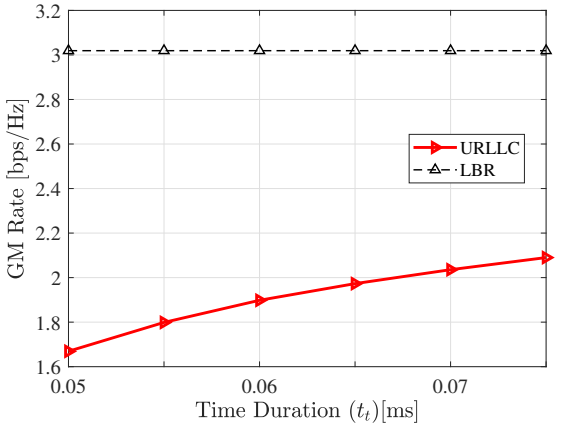

Fig. 7: GM rate versus t_t

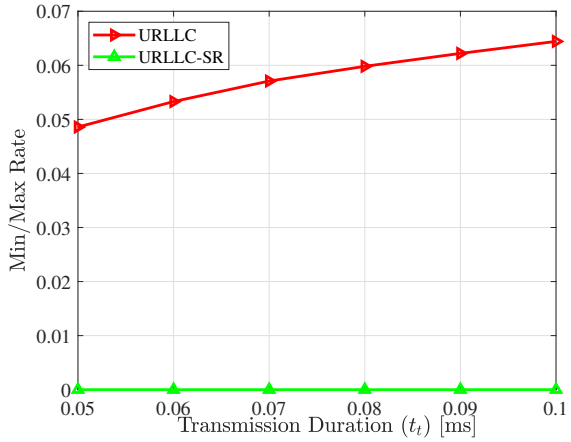

Fig. 8: The ratio of the highest and lowest users rates versus t_t

duration t_t. Hence the advantage of using GM maximization over SR maximization becomes quite convincing.

IV. Conclusions

This paper has considered the joint design of transmit beamforming at the base station and RIS PREs for RIS-aid multi-user URLLC. To guarantee the required quality-of-service in terms of downlink throughput in FBL regime while maintaining computational tractability, we developed an algorithm, which invokes closed-form expressions at each iteration for generating a better point for the maximizing the geometric means of the users' rates (GM-rate). The algorithm has been supported by simulations.

References

[1] M. Di Renzo *et al*., "Smart radio environments empowered by reconfigurable intelligent surfaces: how it works, state of research, and road ahead," *IEEE J. Sel. Areas Commun*, vol. 38, pp. 2450–2525, Nov. 2020.

[2] C. Huang *et al*., "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wirel. Commun.*, vol. 27, no. 5, pp. 118–125, 2200.

[3] Z. Cui, K. Guan, J. Zhang, and Z. Zhong, "SNR coverage probability analysis of RIS-aided communication systems," *IEEE Trans. Veh. Technol.*, vol. 70, pp. 3914–3919, Apr. 2021.

[4] Y. Han *et al*., "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, p. 8238–8242, Aug. 2019.

[5] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Select. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, 2020.

[6] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper Gaussian signaling," *IEEE J. Sel. Areas Commun.*, vol. 38, pp. 2589–2603, Nov. 2020.

[7] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Maximizing the geometric mean of user-rates to improve rate-fairness: Proper vs. improper Gaussian signaling," *IEEE Trans. Wirel. Commun.*, vol. 21, pp. 295–306, Jan. 2022.

[8] G. Brown, "Ultra-reliable low-latency 5G for industrial automation," Technol. Rep. 2, Qualcomm, 2018.

[9] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2307–2359, May 2010.

[11] A. A. Nasir, H. D. Tuan, H. H. Nguyen, M. Debbah, and H. V. Poor, "Resource allocation and beamforming design in the short blocklength regime for URLLC," *IEEE Trans. Wirel. Commun.*, vol. 20, pp. 1321–1335, Feb. 2021.

[12] A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Trans. Wirel. Commun.*, vol. 20, pp. 5861–5871, Sept. 2021.

[13] R. Hashemi, S. Ali, N. H. Mahmood, and M. Latva-aho, "Average rate and error probability analysis in short packet communications over RIS-aided URLLC systems," *IEEE Trans. Veh. Techn.*, vol. 70, pp. 10320–10334, Oct. 2021.

[14] H. Ren, K. Wang, and C. Pan, "Intelligent reflecting surface-aided URLLC in a factory automation scenario," Tech. Rep. arXiv:2103.09323, arXiv, 2021.

[15] W. R. Ghanem, V. Jamali, and R. Schober, "Joint beamforming and phase shift optimization for multicell IRS-aided OFDMA-URLLC systems," in *Proc. IEEE Wirel. Commun. Network. Conf.*, pp. 1–7, 2021.

[16] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent reflecting surface vs. decode-and-forward: How large surfaces are needed to beat relaying?," *IEEE Wirel. Commun. Lett.*, vol. 9, pp. 244–248, Feb. 2020.

[17] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wirel. Commun.*, vol. 19, Dec. 2020.

[18] J. Scarlett, V. Y. F. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-Gaussian channels," *IEEE Trans. Infor. Theory*, vol. 63, pp. 81–92, Jan. 2017.

[19] C. She, C. Yang, and T. Q. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 1, pp. 127–141, 2017.