Assignment Optimization of Reconfigurable Intelligent Surfaces

Master's Thesis

for obtaining the academic degree

Diplom-Ingenieurin

as part of the study

Information and Communication Engineering

carried out by

Vivian Onyinye Uba
matriculation number: 11922557

Institute of Telecommunications
at TU Wien

Supervision:
Associate Prof. Dipl.-Ing. Dr.techn. Stefan Schwarz
Projektass.(FWF) MSc. Martin Mark Zan

Vienna, December 2025

Abstract

In a multi-cell architecture, each Reconfigurable Intelligent Surface (RIS) is assigned to at most one Base Station (BS) at a time. We aim to maximize the worst-user SINR by optimizing the BS beamforming vectors and/or the binary RIS–BS assignment variables. This leads to a highly non-convex, non-linear mixed-integer optimization problem. To obtain tractable solutions, we develop four schemes: a large-scale MISOCP formulation, an alternating-optimization method, a quota-based allocation scheme, and an RIS optimization with fixed beamforming. All proposed schemes achieve substantial performance gains compared to a conventional no-RIS baseline.

Contents

Chapter 1

Introduction and System Overview

1.1 Motivation

Reconfigurable intelligent surfaces (RISs) are a low-power means of improving coverage and spectral efficiency in wireless networks [1, 2, 3, 4]. A RIS is a passive surface that comprises many elements that can cause phase shifts on incident electromagnetic waves. It does this by reflecting the signal from a base station (BS) towards one or more user equipments (UEs). A RIS can make line-of-sight (LOS) paths where the direct link is weak or blocked [1, 3].

In this thesis, we consider a downlink system in which several BSs share a finite number of RIS. Each RIS is assigned to at most one BS at any given time, and each BS serves multiple single-antenna users. Our research problem is how to allocate RISs to BSs and design BS beamforming such that the worst-user performance is improved in a fair way. Similar RIS-assignment problems have been studied, e.g., in [5, 6], typically under instantaneous CSI and sum-rate objectives, whereas here we focus on max-min SINR and fairness.

There are three properties that make this problem challenging:

1. The RIS–BS assignment is *discrete*. That is, each RIS either serves a given BS or it does not.

2. The performance metric is the users' signal-to-interference-plus-noise ratio (SINR), which is a ratio of quadratic forms in the beamforming vectors and depends nonlinearly on the RIS assignment.

3. Fairness is a design goal. We want a max–min SINR criterion to protect the weakest user, in line with classical notions of max–min fairness in communication networks [7, 8].

These properties result in a mixed-integer nonlinear optimization problem (MINLP), which is, in general, NP-hard and non-convex [9]. A key concept of this thesis is that the structure of the SINR expressions allow a reformulation as a *mixed-integer second-order cone program* (MISOCP), both for a large-scale (estimated) SINR model and, for an instantaneous SINR model with power split. The continuous relaxation is a second-order cone program (SOCP) [10, 11].

During the course of this work, we investigated and compared:

- **Large-scale MISOCP**: an optimization of RIS assignment based on an estimated SINR model that uses only slow-varying channel statistics, like the large-scale RIS-oriented designs surveyed in [2, 3].

- **Instantaneous joint optimization**: an optimization based on the instantaneous SINR model with power split where we considered *two algorithmic variants*:

- an *alternating optimization (AO)* variant, in which beamforming and RIS assignment are optimized in turn (fix $\{z_{r,b}\}$, optimize $\{\boldsymbol{w}_{j,r}\}$; then fix $\{\boldsymbol{w}_{j,r}\}$, optimize $\{z_{r,b}\}$), similar to other alternating RIS beamforming schemes [5, 6];
  - a *quota-based optimization* that searches the vector of per-BS RIS counts $\boldsymbol{m} = (m_1, \ldots, m_B)$, and for each quota configuration solves a MISOCP that jointly optimizes the beamforming vectors and the binary RIS–BS assignment under the constraints $\sum_r z_{r,b} = m_b$; this yields, for each $\boldsymbol{m}$, the globally optimal worst-user SINR, and the outer search over $\boldsymbol{m}$ provides certificates of global optimality within the considered quota space.

- **Fixed beamforming designs**: here, the beamforming vectors are fixed using simple heuristics (per-RIS unit beams) and only the RIS assignment is optimized via a MISOCP. We observed this design to perform well compared to the full instantaneous joint optimization, at much lower computational complexity, echoing similar findings in other low-complexity RIS resource allocation works (e.g., [12]).

Our numerical results (discussed in later chapters) show that:

- the max–min MISOCP formulation based on large-scale SINR provides a meaningful and robust RIS assignment,
- the alternating optimization can converge to local optima and fail to match the global min-rate achieved by the quota-based search,
- fixed-beamforming RIS optimization often attains a large fraction of the performance of the fully joint design, while being orders-of-magnitude simpler.

This introductory chapter establishes the system model, the binary formulation for RIS sharing, the instantaneous and large-scale SINR expressions, and the max–min fairness problems that motivate the MISOCP formulations used throughout the thesis.

1.2 Network Topology and System Assumptions

1.2.1 Topology and notation

We consider a downlink system with:

- B base stations (BSs), indexed by $b \in \{1, \ldots, B\}$,
- R RIS panels, indexed by $r \in \{1, \ldots, R\}$,
- U single-antenna users (UEs), indexed by $u \in \{1, \ldots, U\}$.

Each BS is equipped with M_{BS} antennas, and each RIS has M_{RIS} reflecting elements. We collect the sets of indices as

$$\mathcal{B} = \{1, \ldots, B\}, \quad \mathcal{R} = \{1, \ldots, R\}, \quad \mathcal{U} = \{1, \ldots, U\}.$$

The positions of BSs, RISs, and UEs are drawn from a geometric model and then held fixed, consistent with recent RIS deployment studies [2, 4]. In the simulation environment, the geometry is generated once for a given seed, and all subsequent channel realizations reuse this geometry; the small-scale fading is then drawn for NN i.i.d. snapshots. This leads to a separation between:

- *large-scale* parameters: which are constant over many coherence times,
- *small-scale* fading, which varies from snapshot to snapshot.

1.2.2 Channel model

For a given user u and BS b, we represent the direct channel by a complex vector

$$\boldsymbol{h}_{u,b}^{(\mathrm{dir})} \in \mathbb{C}^{M_{\mathrm{BS}}}.$$

For user u and RIS r we denote the RIS–UE channel by

$$\boldsymbol{h}_{u,r} \in \mathbb{C}^{M_{\mathrm{RIS}}},$$

while the BS–RIS channel from BS b to RIS r is

$$\boldsymbol{G}_{r,b} \in \mathbb{C}^{M_{\mathrm{RIS}} \times M_{\mathrm{BS}}}.$$

The RIS phase matrix is diagonal,

$$\boldsymbol{\Phi}_r = \mathrm{diag}\big(e^{\mathrm{j}\theta_{r,1}}, \dots, e^{\mathrm{j}\theta_{r,M_{\mathrm{RIS}}}}\big),$$

where $\theta_{r,m}$ is the phase applied by element m of RIS r.

The geometry determines large-scale path gains:

- $\gamma_{u,b}$: gain from BS b to user u,
- $\gamma_{u,r}$: gain from RIS r to user u,
- $\gamma_{r,b}$: gain from BS b to RIS r,

and LOS/NLOS indicators. These parameters are combined with Rician or Rayleigh scattering models to generate the small-scale channels for each snapshot, in line with common RIS channel models that have been studied [1, 3]. In the large-scale model used for the MISOCP, the effect of fading is captured through means and covariances.

Each user u is served by exactly one BS, denoted $b(u) \in \mathcal{B}$. This BS–UE association is determined from long-term channel conditions (path loss, LOS probability) and is treated as fixed throughout the optimization.

1.2.3 RIS operation and sharing across BSs

In this work, we adopt the following simplifying but realistic assumptions, consistent with many RIS-assisted cellular models [2, 4]:

- In each scheduling interval, each RIS is *assigned* to at most one BS.
- If RIS r is assigned to BS b, it reflects the signal from BS b towards one or more users associated with BS b.

These assumptions are represented using binary variables in Section 1.3.

1.3 Binary RIS–BS Assignment and Search Space

1.3.1 Binary assignment and mutual exclusion

We introduce binary variables

$$z_{r,b} \in \{0,1\}, \qquad r \in \mathcal{R},\ b \in \mathcal{B}, \tag{1.1}$$

where $z_{r,b} = 1$ means that RIS r is assigned to BS b. The mutual-exclusion constraint

$$\sum_{b\in\mathcal{B}} z_{r,b} \le 1, \qquad \forall\, r \in \mathcal{R}, \tag{1.2}$$

ensures that each RIS is assigned to at most one BS. If $\sum_b z_{r,b} = 0$, RIS r is inactive in that scheduling interval.

For each BS b, the number of assigned RISs is

$$m_b \triangleq \sum_{r\in\mathcal{R}} z_{r,b}, \qquad b \in \mathcal{B}. \tag{1.3}$$

To avoid degenerate solutions in which a BS is never assisted by any RIS, we impose lower and upper bounds

$$1 \le m_b \le m_b^{\max}, \qquad \forall\, b, \tag{1.4}$$

and possibly a global limit $\sum_b m_b \le m_{\max}^{\text{tot}}$ if the total number of used RISs shall be capped.

1.3.2 Size of the discrete search space

Even for modest values of R and B, the number of possible RIS–BS assignments is huge. Since each RIS can be assigned to any of the B BSs, the total number of assignment patterns is

$$N_{\text{assign}} = (B)^R. \tag{1.5}$$

For the typical configuration used in our simulations, with $B = 2$ BSs and $R = 10$ RISs, we already obtain

$$N_{\text{assign}} = 2^{10} = 1024$$

possible assignments.

It is sometimes convenient to think in terms of the *quota* vector $\boldsymbol{m} = (m_1, \ldots, m_B)$ of per-BS RIS counts, under the constraint (i.e. we require that every BS has at least one RIS)

$$\sum_{b\in\mathcal{B}} m_b = R, \qquad m_b \in \mathbb{Z}_{>0}, \tag{1.6}$$

then the number of feasible quota vectors is

$$N_{\text{quota}} = \binom{R-1}{B-1}. \tag{1.7}$$

For $R = 10$ and $B = 2$ we have $N_{\text{quota}} = 9$ possible (m_1, m_2) pairs. For each fixed $\boldsymbol{m}$, the number of assignments that realize this quota is

$$N_{\text{assign}|\boldsymbol{m}} = \frac{R!}{m_1! \cdots m_B!}. \tag{1.8}$$

In the example $R = 10$, $B = 2$, and say $m_1 = 4$, $m_2 = 6$, we obtain $N_{\text{assign}|\boldsymbol{m}} = \binom{10}{4} = 210$ distinct assignments.

A total search over all $z_{r,b}$ is therefore infeasible. Even a search over all quota vectors $\boldsymbol{m}$ is feasible only for small R and B. This leads us to investigate the quota-based global search used in later chapters: for each candidate vector $\boldsymbol{m}$ we solve a continuous MISOCP to find the best assignment, and then compare the resulting worst-user rates. This yields a global optimum with respect to $\boldsymbol{m}$, but the problem remains combinatorial. [9].

1.3.3 Combinatorial and computational complexity

In the worst case, MINLP solvers require time exponential in the number of binary variables [9]. In our setting, there are $R \times B$ binary variables $z_{r,b}$, so even moderate values of R and B will lead to large branch-and-bound search trees.

For each node in the branch-and-bound tree, the continuous relaxation of the problem is solved as a second-order cone program (SOCP). The interior-point methods used by solvers such as MOSEK have polynomial complexity in the size of the continuous problem:

$$\mathcal{O}\big(n_{\text{var}}^3\big),$$

where n_{var} is the number of continuous variables (beamforming vectors, auxiliary SOC variables, etc.). In our implementation, n_{var} scales with U, B, M_{BS}, and the number of RIS-related covariance terms. This scaling is typical of conic optimization methods [11, 10].

As a result, the overall computational complexity is driven by:

- the number of binary variables $R \times B$ (combinatorial explosion),
- the size of the continuous SOCP at each node (cubic complexity),
- solver parameters such as the allowed optimality gap and maximum search time.

In practice, we observed that the large-scale MISOCP is solvable for the problem sizes considered in this thesis, whereas the instantaneous joint MISOCP is substantially more demanding. This observation motivated the use of quota based optimization and fixed-beamforming schemes as lower-complexity alternatives.

1.4 Instantaneous SINR with Power Split

1.4.1 Transmit signal and power split

Each BS b transmits one data stream s_b to its associated users, with total power budget P_b. If BS b is connected to m_b RISs, then we split its power equally across the m_b beams. Similar power-splitting models are used in joint RIS beamforming designs [5, 6].

Let $\boldsymbol{w}_{b,r} \in \mathbb{C}^{M_{\text{BS}}}$ is the beamforming vector used by BS b for the path involving RIS r. For convenience, we normalize $\|\boldsymbol{w}_{b,r}\|_2 = 1$ and account for power in a scalar factor.

The transmit signal of BS b is

$$x_b = \sum_{r \in \mathcal{R}} z_{r,b} \sqrt{\frac{P_b}{m_b}}\, \boldsymbol{w}_{b,r} s_b, \tag{1.9}$$

where m_b is defined in (1.3) and clipped to at least one,

$$m_b = \max\left\{1,\ \sum_{r \in \mathcal{R}} z_{r,b}\right\}. \tag{1.10}$$

If $m_b = 1$ and no RIS is assigned to BS b, the term with $z_{r,b}$ simply reduces to a direct beam.

1.4.2 Per-path contributions and Power Split SINR

The complex channel contribution from BS b to user u via RIS r is

$$\alpha_{u,b,r} = (\boldsymbol{h}_{u,b}^{(\text{dir})})^{\text{H}} \boldsymbol{w}_{b,r} + \boldsymbol{h}_{u,r}^{\text{H}} \boldsymbol{\Phi}_r \boldsymbol{G}_{r,b} \boldsymbol{w}_{b,r}, \tag{1.11}$$

where the first term captures the direct path (if used) and the second term the RIS-reflected path.

For user u and BS b, the coherent sum of all paths that involve RISs assigned to BS b is

$$A_{u,b} \;=\; \sum_{r \in \mathcal{R}} z_{r,b}\, \alpha_{u,b,r}. \tag{1.12}$$

The received signal at user u is

$$y_u = \sqrt{\frac{P_{b(u)}}{m_{b(u)}}} A_{u,b(u)} s_{b(u)} + \sum_{b \in \mathcal{B},\ b \neq b(u)} \sqrt{\frac{P_b}{m_b}} A_{u,b} s_b + n_u, \tag{1.13}$$

where n_u is circular complex Gaussian noise with variance σ_n^2.

The instantaneous SINR of user u in a given small-scale fading snapshot is

$$\mathrm{SINR}_u = \frac{\frac{P_{b(u)}}{m_{b(u)}} \left| A_{u,b(u)} \right|^2}{\sum_{b \in \mathcal{B},\ b \neq b(u)} \frac{P_b}{m_b} \left| A_{u,b} \right|^2 + \sigma_n^2}. \tag{1.14}$$

In the simulation code, this expression is implemented directly using the precomputed $\alpha_{u,b,r}$ values and the binary matrix $z_{r,b}$.

Over NN i.i.d. fading realizations indexed by s, the ergodic rate of user u is approximated by

$$\overline{R}_u = \frac{1}{NN} \sum_{s=1}^{N_\mathrm{N}} \log_2 \left(1 + \mathrm{SINR}_{u,s}\right), \tag{1.15}$$

where $\mathrm{SINR}_{u,s}$ i the instantaneous SINR in snapshot s. In later chapters, we compare the distribution of these rates, for example via empirical CDFs, under different optimization schemes.

1.5 Large-Scale (Estimated) SINR Model

1.5.1 Direct and RIS-aided components

The instantaneous SINR model in (1.14) is accurate but it depends on the small-scale fading, which changes rapidly and is difficult to use for system-level optimization. Instead, we also use a large-scale SINR model that depends only on path-loss-like parameters and statistical scattering quantities, as is common in RIS system-level studies [2, 3].

Consider a user u served by BS $d = b(u)$. The large-scale model used in this thesis decomposes the SINR into six contributions:

- *Desired signal:*
  - direct BS–UE component,
  - coherent RIS-aided signal,
  - incoherent self-term due to residual scattering.
- *Denominator:*
  - thermal noise,
  - direct inter-cell interference from other BSs,
  - RIS-aided inter-cell interference from RISs assigned to other BSs.

Let

- $\gamma_{u,d}$: large-scale gain BS $d \to$ UE u,
- $\gamma_{u,r}$: large-scale gain RIS $r \to$ UE u,
- $\gamma_{r,d}$: large-scale gain BS $d \to$ RIS r,

- $\kappa_{u,r}$: complex coefficient capturing phase-aligned LOS components for the RIS-assisted path,
- $\bar{k}_{u,r}$: parameter capturing the non-coherent scattering strength of the RIS-assisted path,
- P_d: transmit power of BS d (per resource block),
- m_d: number of RISs assigned to BS d (as in (1.3)),
- M_{BS} and M_{RIS}: number of BS antennas and RIS elements, respectively.

Using the binary assignment variables $z_{r,d} \in \{0, 1\}$, the large-scale SINR for user u served by BS d is modeled as

$$\hat{\beta}_u(\boldsymbol{z}) = \frac{\underbrace{\gamma_{u,d}^2 P_d}_{\text{direct}} + \underbrace{\left(\sum_{r=1}^{R} z_{r,d}\, \gamma_{u,r}\, \gamma_{r,d}\, \kappa_{u,r} \sqrt{\frac{P_d M_{\mathrm{BS}}}{m_d}}\, M_{\mathrm{RIS}} \right)^2}_{\text{coherent RIS-aided signal}} + \underbrace{\sum_{r=1}^{R} z_{r,d}\, \gamma_{u,r}^2\, \gamma_{r,d}^2\, \bar{k}_{u,r}^2 \frac{P_d M_{\mathrm{BS}}}{m_d}\, M_{\mathrm{RIS}}}_{\text{incoherent self-term}}}{\underbrace{\sigma_n^2}_{\text{noise}} + \underbrace{\sum_{b \in \mathcal{B},\, b \neq d} \gamma_{u,b}^2 P_b}_{\text{direct inter-cell}} + \underbrace{\sum_{b \in \mathcal{B},\, b \neq d} \sum_{r=1}^{R} z_{r,b}\, \gamma_{u,r}^2\, \gamma_{r,b}^2\, P_b\, M_{\mathrm{RIS}}}_{\text{RIS-aided inter-cell}}}. \tag{1.16}$$

The coherent RIS term shows that RIS assignment creates a sum over r for quantities that scale with the large-scale gains and the per-BS power split. When multiple RISs are assigned to BS d, their contributions add up coherently, leading to a squared sum; this is the main source of RIS gain. The incoherent term captures the residual RIS-induced power that is not perfectly phase-aligned and therefore does not add coherently at the receiver.

1.6 Fairness and Max–Min SINR Formulation

1.6.1 Why fairness, and why max–min?

RIS panels are scarce resources. In typical deployments there are far more users than RISs, and each RIS can serve at most one BS at a time [2]. Users who never see a RIS-assisted path are at a strong disadvantage, particularly if their geometry leads to weak direct links. A purely throughput-oriented criterion such as sum rate

$$\max \sum_{u} \overline{R}_u$$

would tend to allocate RIS resources to already-strong users, because increasing their SINR yields a large gain in sum rate. This can leave users at the cell edge with poor coverage.

To explicitly protect the weakest users, we adopt a *max–min fairness* criterion on the SINR:

$$\max_{\boldsymbol{z}, \{\boldsymbol{w}_{b,r}\}} \; \min_{u \in \mathcal{U}} \; \beta_u(\boldsymbol{z}, \{\boldsymbol{w}_{b,r}\}), \tag{1.17}$$

where β_u stands either for the instantaneous SINR SINR_u in (1.14) or the large-scale SINR $\hat{\beta}_u$ in (1.16), depending on the context. Max–min fairness has a long history in networking and resource allocation [7, 8] and has recently been applied to RIS-aided systems as well [12].

This objective enforces a form of max–min fairness: the optimizer is driven to raise the SINR of the worst user as much as possible. In the context of RIS-assisted networks, this tends to allocate RISs to disadvantaged users and to balance the RIS usage between BSs, as observed in the experiments.

1.6.2 Epigraph form of the max–min problem

A standard way to handle (1.17) is to introduce a scalar variable $\tau \geq 0$ that represents the minimum SINR level across users and rewrite the problem in epigraph form:

$$\max_{\tau,\, \boldsymbol{z},\, \{\boldsymbol{w}_{b,r}\}} \quad \tau \tag{1.18}$$

$$\text{s.t.} \quad \beta_u(\boldsymbol{z}, \{\boldsymbol{w}_{b,r}\}) \;\geq\; \tau, \qquad \forall\, u \in \mathcal{U}, \tag{1.19}$$

$$z_{r,b} \in \{0,1\}, \quad \sum_b z_{r,b} \leq 1, \quad \forall\, r, \tag{1.20}$$

$$\text{additional power and topology constraints on } \boldsymbol{w}_{b,r}. \tag{1.21}$$

For fixed τ, the constraints (1.19)–(1.21) define a feasibility problem. In both the large-scale and instantaneous cases, these constraints can be expressed using second-order cones, which leads to a MISOCP [11, 9].

In practice, we use bisection on τ and solve a sequence of feasibility MISOCPs until the maximum achievable $\tau^\star$ is found to a desired accuracy.

1.6.3 Fairness interpretation

The max–min SINR solution has several desirable fairness properties:

- **Bottleneck removal:** If a particular user $u^\star$ initially has an SINR much lower than others, the optimizer will allocate RIS and beamforming resources to improve $\beta_{u^\star}$ until it becomes comparable to the rest.
- **Equalization:** In many of the simulation scenarios, the max–min solution nearly equalizes the SINR (or rate) across users, especially when the number of degrees of freedom (BS antennas and RISs) is large enough [8, 13].
- **RIS balancing across BSs:** When one BS is in a disadvantaged position (e.g., farther from the RIS cluster), the fairness criterion leads to RIS allocations that compensate this asymmetry; for example, by assigning more RISs to the weaker BS.

This behavior was visible in the results: the quota-based global search with the MISOCP favored an asymmetric, but fairness-optimal, distribution of RISs (e.g., $(m_1, m_2) = (4, 6)$ instead of $(5, 5)$), yielding a significantly higher worst-user rate than the symmetric assignment.

1.7 MISOCP Formulation and Geometric Interpretation

1.7.1 General MISOCP template

A generic MISOCP can be written as

$$\max_{\tau, \boldsymbol{w}, \boldsymbol{z}} \quad \tau \tag{1.22a}$$

$$\text{s.t.} \quad \left\| \boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i \right\|_2 \;\leq\; \boldsymbol{c}_i^\top \boldsymbol{x} + d_i, \quad i = 1, \ldots, m \tag{1.22b}$$

$$\boldsymbol{E}\boldsymbol{x} = \boldsymbol{f}, \qquad \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h}, \tag{1.22c}$$

$$\boldsymbol{z} \in \{0,1\}^p. \tag{1.22d}$$

Notation:

$$
\begin{array}{rl}
\boldsymbol{x} \in \mathbb{R}^{n} & \text{decision vector} \\
\boldsymbol{w} \in \mathbb{R}^{n_w} & \text{beamforming vector (if separate)} \\
\boldsymbol{A}_i \in \mathbb{R}^{k_i \times n},\ \boldsymbol{b}_i \in \mathbb{R}^{k_i} & \text{SOC left-hand side} \\
\boldsymbol{c}_i \in \mathbb{R}^{n},\ d_i \in \mathbb{R} & \text{SOC right-hand side} \\
\boldsymbol{E} \in \mathbb{R}^{q \times n},\ \boldsymbol{f} \in \mathbb{R}^{q} & \text{linear equalities} \\
\boldsymbol{G} \in \mathbb{R}^{r \times n},\ \boldsymbol{h} \in \mathbb{R}^{r} & \text{linear inequalities} \\
\boldsymbol{z} \in \{0,1\}^{p} & \text{binary assignment vector} \\
\tau \in \mathbb{R}_{+} & \text{objective scalar (min-SINR)}
\end{array}
$$

Objective. The objective (1.22a) maximizes τ, which in our setting represents the *minimum SINR* achievable across all users.

Second-order cone (SOC) constraints. The constraints (1.22b) are of SOC type. A standard SOC is

$$\|\boldsymbol{u}\|_2 \leq v, \qquad v \geq 0,$$

which defines a convex "ice-cream cone" set. In the SINR setting,

- $\boldsymbol{u}$ corresponds to the stacked *interference-plus-noise* terms (a vector),
- v corresponds to the *desired signal* term (a nonnegative scalar),

so the inequality reads: *"the interference+noise norm must fit under the desired signal."* Index $i = 1, \ldots, m$ typically ranges over *all users and all channel scenarios*, so these SOCs collectively enforce SINR feasibility for everyone.

Linear constraints. The constraints (1.22c) capture ordinary linear relations, e.g.,

- **Power budgets:** $\|\boldsymbol{w}_{b,r}\|_2^2 \leq P_b$ for each BS b,
- **Assignment rules:** $\sum_b z_{r,b} \leq 1$ (each RIS r to at most one BS b),
- **Candidate pruning:** force $\boldsymbol{z}_{r,b} = 0$ for disallowed RIS–BS pairs.

These do not involve cones—only linear equalities/inequalities.

Integer constraints. The condition (1.22d) imposes integrality on the assignment variables $\boldsymbol{z}$: $\boldsymbol{z} \in \{0,1\}^p$, where p is the total number of binary decisions (e.g., $p = R \times B$). This is what makes the problem *mixed-integer*; without (1.22d) the model would be a pure SOCP.

SINR cone in compact form. Writing the per-user, per-scenario SINR constraint as an SOC gives

$$\left\|[\text{interference terms}, \sqrt{\sigma^2}]\right\|_2 \leq \frac{1}{\sqrt{\tau}}\,(\text{desired signal})\,,$$

i.e., *radius* (interference+noise norm) $\leq$ *axis* (scaled desired signal).

The template (1.22a)–(1.22d) means: maximize the SINR τ; keep all users' SINRs feasible via SOC constraints; ensure power and assignment rules are linear; and require RIS assignments to be binary.

The problem can be solved globally (up to a chosen optimality gap) using branch-and-bound [9, 11]. We have represented the SINR constraint using MISOCP in Chapter 3.

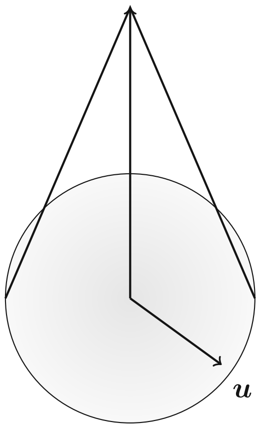

Figure 1.1: Standard second-order cone $\{(\boldsymbol{u}, v) : \|\boldsymbol{u}\|_2 \leq v\}$. The SINR constraints keep the interference-plus-noise vector inside such a cone scaled by the desired signal.

1.7.2 Why a MISOCP formulation?

From the perspective of modeling, several options are possible:

- A MINLP could be used, which would directly accept the nonlinear SINR expressions with binary decisions. However, such problems are non-convex and solvers may return only locally optimal solutions without certification.
- A mixed-integer linear program (MILP) could be used by linearizing the quadratic and ratio terms, but this would introduce approximation errors and substantially complicate the model.
- A MISOCP formulation leverages the fact that:
  - the numerator and denominator of the SINR can be written as norms and squared norms of affine functions of the beamformers,
  - the SINR constraints can be converted to second-order cones,
  - modern solvers such as MOSEK offer robust support for MISOCPs with global optimality guarantees and explicit control on the optimality gap [11, 9].

In this thesis, we therefore choose MISOCP as the main modeling framework. It provides a good balance between modeling fidelity and tractability [10, 11].

1.7.3 Geometric view: second-order cones and epigraphs

The standard second-order cone (SOC) in $\mathbb{R}^{n+1}$ is defined as

$$\mathcal{K}_2^{n+1} = \left\{(\boldsymbol{u}, v) \in \mathbb{R}^n \times \mathbb{R} : \|\boldsymbol{u}\|_2 \leq v,\ v \geq 0\right\}. \tag{1.23}$$

Geometrically, it looks like a solid ice-cream cone, with v as the vertical axis and $\boldsymbol{u}$ spanning the base. The SINR constraint is of this type: the interference-plus-noise vector $\boldsymbol{v}_u$ must lie inside a cone whose axis is scaled by the desired signal s_u [11].

For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, the *epigraph* is

$$\operatorname{epi}(f) = \left\{(\boldsymbol{x}, t) : f(\boldsymbol{x}) \leq t\right\}. \tag{1.24}$$

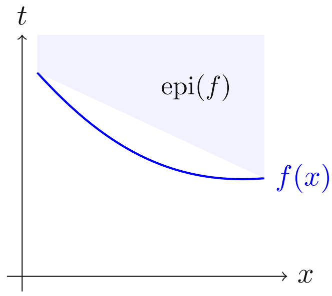

Figure 1.2: Epigraph of a convex function $f(x)$. The SOC is the epigraph of the Euclidean norm. SINR constraints are implemented as epigraph constraints on interference norms.

The set of all such $(\boldsymbol{x}, t)$ is convex. The SOC can be seen as the epigraph of the Euclidean norm:

$$\mathcal{K}_2^{n+1} = \operatorname{epi}\big(\|\cdot\|_2\big).$$

Thus, the SINR constraints are epigraph constraints of a norm: they ensure that the norm of the interference vector does not exceed a certain affine function of the beamformers and RIS assignment [11].

This geometric perspective is useful for understanding how the optimization balances desired signal and interference. Maximizing the common SINR level τ corresponds to narrowing the cones (via $1/\sqrt{\tau}$) until they touch the feasible set of beamformers and RIS assignments [11, 9].

1.8 Overview of Algorithmic Approaches

Although detailed algorithms and numerical results are deferred to later chapters, we summarize the main approaches that build on the formulation in this chapter.

1.8.1 Large-scale MISOCP with quota search

In the large-scale setting, we construct a feasibility MISOCP for a given SINR threshold γ using the large-scale model in (1.16). By bisection on γ we find the maximum achievable common SINR $\gamma^\star$.

To reduce the combinatorial complexity, we first select a quota vector $\boldsymbol{m}$ and restrict the MISOCP so that each BS b must be assigned exactly m_b RISs. We then:

1. Loop over all feasible quota vectors $\boldsymbol{m}$ (for small R and B).
2. For each $\boldsymbol{m}$, solve the corresponding MISOCP to obtain the best assignment and beamformers, and record the worst-user rate.
3. Select the $\boldsymbol{m}$ that yields the highest worst-user rate. This provides a certificate of global optimality *within the quota search space.*

This *quota-based global search* was used to demonstrate that the alternating optimization can be trapped in local optima.

1.8.2 Alternating optimization (AO)

The alternating optimization scheme iterates between:

Beamforming step: We fix the RIS assignment $\boldsymbol{z}$ and solve a beamforming SOCP to maximize the minimum SINR across users, respecting per-BS power constraints, similar to other alternating RIS beamforming methods [5, 6].

RIS allocation step: We fix the beamformers and solve a MISOCP for the RIS assignment alone, again maximizing the minimum SINR under the binary and mutual-exclusion constraints.

This procedure is relatively fast, but because the overall problem is non-convex, AO is not guaranteed to reach the global optimum. The quota-based search was used as a reference to detect cases where AO converged to a suboptimal local solution.

1.8.3 Fixed beamforming with RIS optimization

Finally, we consider schemes where the beamformers are fixed according to simple rules (e.g., per-RIS unit beams aligned with the direct channel or large-scale statistics), and only the binary RIS assignment $\boldsymbol{z}$ is optimized via a MISOCP. This reduces the continuous variable dimension and the overall complexity.

Despite this simplification, numerical experiments showed that some fixed-beamforming designs perform remarkably well, often achieving a large fraction of the worst-user rate of the fully joint AO scheme, especially when averaged over NN fading snapshots, in line with similar observations in recent RIS resource allocation works [12].

1.9 Chapter Summary

In this chapter we:

- introduced the multi-BS, multi-RIS, multi-user downlink scenario and the underlying channel and power models,
- formulated RIS–BS sharing using binary variables with mutual-exclusion constraints,
- analyzed the size of the discrete RIS assignment search space and discussed its combinatorial complexity,
- derived the instantaneous SINR with equal power split across beams and the corresponding ergodic-rate approximation,
- presented a large-scale (estimated) SINR model that decomposes the signal and interference into direct, coherent RIS-aided, incoherent self-term, noise, direct inter-cell, and RIS-aided inter-cell components,
- framed the RIS allocation problem as a max–min SINR optimization to enforce fairness across users,
- showed how the SINR constraints can be expressed as second-order cone constraints, leading to a MISOCP formulation,
- provided a geometric interpretation of second-order cones and epigraphs in the context of SINR constraints,
- briefly outlined the main algorithmic approaches studied in the thesis: large-scale MISOCP with quota search, alternating optimization, and fixed-beamforming RIS optimization.

The next chapters build on this foundation to develop, implement, and evaluate these algorithms in detail, using both large-scale and instantaneous channel models, and to compare their performance and computational complexity in realistic RIS-assisted networks.

Chapter 2

State-of-the-Art in RIS Optimization and Mixed-Integer Formulations

Early overview and tutorial works describe RISs as a key enabler for "smart radio environments". They discuss prototypical hardware architectures, and survey applications across communication, sensing, and localization [14, 15, 16, 17]. A common theme in this literature is that RIS design is fundamentally a resource-allocation and optimization problem: the system must jointly choose precoding at the base stations (BSs) and passive reflection coefficients at the RIS, under hardware constraints and quality-of-service (QoS) or fairness requirements.

In this chapter we summarize the state-of-the-art in RIS optimization and explain where binary formulations and mixed-integer optimization techniques fit into the broader scope of knowledge. We focus on downlink multi-user systems, which are closest to the setting of this thesis, and emphasize works that optimize max–min signal-to-interference ratio (SINR) or user rates, or that employ mixed-integer formulations.

2.1 Continuous Optimization of Beamforming

The first group of RIS optimization literature considered fully continuous phase shifts at each reflecting element, typically modeled as $\phi_m = e^{\mathrm{j}\theta_m}$ with $\theta_m \in [0, 2\pi)$. In this regime, the RIS design problem can be cast as a continuous (but highly nonconvex) optimization problem.

A reference is the work by Wu and Zhang on joint beamforming for RIS-aided networks [18]. They consider a multi-user downlink where a multi-antenna access point (AP) communicates with multiple users via a single RIS, and formulate transmit power minimization under per-user SINR constraints. The problem is nonconvex due to the unit-modulus constraints on the RIS coefficients and the coupling between precoding and passive phases. They propose an alternating optimization (AO) procedure that iteratively updates:

1. the beamforming vectors by solving a second-order cone program (SOCP) with fixed RIS phases; and
2. the RIS phase vector by solving a nonconvex quadratically constrained quadratic program (QCQP) using semidefinite relaxation (SDR) plus Gaussian randomization.

Despite the lack of global optimality guarantees, this AO scheme became a canonical baseline for continuous RIS beamforming.

Subsequent works generalized this AO approach to different performance criteria and system models, including:

- weighted sum-rate or sum-SINR maximization;
- power minimization with QoS constraints;

- energy-efficiency maximization, where transmit power is traded against data rate or against RIS circuit power [19];
- scenarios with statistical channel state information (CSI) only, where the RIS phases are designed based on large-scale fading statistics rather than instantaneous channels [19].

A parallel line of work investigates continuous-phase RIS optimization in multi-cell or multi-operator settings, with explicit fairness objectives. For example, Xie et al. consider a multi-cell IRS-aided MISO system and maximize the minimum (weighted) SINR across users by jointly optimizing coordinated transmit beamforming at all BSs and the RIS reflection coefficients [20]. They also employ an AO structure: the transmit beamforming subproblem is solved via SOCP, while the RIS subproblem is handled via SDR or successive convex approximation (SCA). This work is an important reference for max–min fairness in IRS-assisted systems and provides complexity benchmarks for AO-based continuous optimization. Furthermore, Schwarz invetigates a multi-operator setting where operators compete for RISs in an auction format which is based on Reinforment Learning to effectively assign RISs to operators[21].

Comprehensive surveys and tutorials consolidate these continuous-phase approaches, emphasizing AO, SDR/SCA, and manifold-based methods as the main algorithmic tools [16, 17].

2.2 Discrete RIS Constraints and Mixed-Integer Nonlinear Programs

While continuous phase shifts simplify the optimization problem, practical RIS implementations typically support only a finite number of phase levels . This leads to discrete constraints on each reflection coefficient and turns the RIS design problem into a mixed-integer nonlinear program (MINLP).

Wu and Zhang explicitly address this in their beamforming optimization with discrete phase shifts [22]. They revisit the multi-user downlink scenario and assume that each RIS element can only take values from a finite set $\{e^{\mathrm{j}2\pi b/2^B} : b = 0, \dots, 2^B - 1\}$. The joint design of continuous precoders and discrete phases under SINR constraints becomes a nonconvex MINLP. Their solution remains iterative but incorporates discrete optimization subproblems:

- for the single-user case, they develop optimal and low-complexity quantization schemes that maximize the effective channel gain;
- for the multi-user case, they combine gradient-based methods with element-wise discrete updates and show that the performance loss relative to the continuous-phase case can be bounded and remains small for moderate B.

This paper is a key reference that explicitly labels the RIS design problem as a mixed-integer non-linear program and quantifies the impact of discretization.

Other works formulate discrete RIS constraints within more general multi-user or multi-function systems. Di et al. consider hybrid beamforming for RIS-based multi-user communications with limited discrete phase shifts, where analog and digital precoders are jointly designed [23]. Chen et al. study RIS-assisted device-to-device (D2D) networks and jointly optimize transmit power and discrete RIS phases to satisfy QoS constraints, again leading to a mixed-integer, nonconvex problem [24]. In all these cases, the discrete nature of the RIS coefficients is handled by iterative heuristics (greedy search, element-wise updates, or relax-and-quantize schemes) rather than exact mixed-integer solvers.

2.3 RIS Selection, Element Activation, and Mixed-Integer Programming

Beyond discrete phase quantization, RIS-related optimization often involves *selection* decisions: which RISs (or which subsets of elements) should be used, to balance coverage, interference management, and hardware or circuit power consumption. These selection variables are binary and make the resulting optimization problems mixed-integer.

Sun et al. consider a downlink network with multiple candidate RISs and explicitly model the circuit power consumed by each active RIS [25]. They jointly design the BS beamforming vectors, the set of the used RISs, and the reflection matrices to minimize the total network power consumption under per-user rate constraints. The formulation is a mixed-integer quadratic program (MIQP), which is NP-hard. Their solution is based on alternating optimization between:

1. an SOCP to update the continuous beamforming variables; and
2. an MIQP (further relaxed to an SDP) to update the binary RIS activation pattern.

This is one of the first works to use mixed-integer optimization to decide which RISs should be used, rather than only tuning their phases.

Related works in green or energy-efficient RIS networks also introduce binary variables for element activation, antenna selection, or user admission control, leading to mixed-integer formulations that are solved via branch-and-bound, cutting-plane, or relaxation-and-rounding methods. For example, hybrid analog–digital designs with discrete RIS phases and antenna selection lead to mixed-integer programs combining continuous beamforming variables with binary configuration variables [23].

In all these references, exact mixed-integer optimization is typically restricted to small-scale instances due to the exponential growth of the search space. For larger instances, the authors resort to:

- convex relaxations (SDP or SOCP) followed by randomized rounding;
- greedy heuristics (element or RIS-add/drop rules);
- or problem-specific decomposition schemes that exploit structure in the network topology or channel statistics.

2.4 Max–Min Fairness and Binary Formulations

Max–min fairness is a performance criterion in multi-user systems where we seek to guarantee a minimum quality of service across users (or operators). In RIS-aided multi-cell downlink systems, Xie et al. formulate a max–min SINR problem and jointly optimize coordinated transmit beamforming and RIS phases under per-BS power constraints [20]. Their approach is fully continuous (no binary variables), but it establishes the computational complexity of max–min problems even without integer constraints and serves as a baseline for fairness-aware RIS design.

When RIS selection or discrete phases are introduced on top of a max–min objective, the resulting problems combine:

- nonconvex continuous variables (beamforming and complex reflection coefficients) with unit-modulus or constant-modulus constraints; and
- discrete or binary variables, such as RIS activation flags, element selection indicators, or per-BS RIS assignment variables.

This leads to mixed-integer nonlinear programs that are even more difficult than the power-minimization problems considered in [22, 25].

The binary max–min SINR formulations studied in this thesis fit into this category:

- binary variables represent RIS-to-BS assignments, possibly with per-BS quota constraints limiting the number of RISs that can be served by each BS;
- continuous variables represent per-BS beamforming vectors and/or RIS reflection coefficients;
- the objective is to maximize the minimum user (or operator) SINR over all users.

The resulting problems can be seen as mixed-integer second-order cone programs (MISOCPs) or mixed-integer conic programs, where the SINR constraints are represented as second-order cone constraints and the RIS assignment variables are binary. Compared to purely continuous AO-based designs, these formulations offer:

- *global optimality* for the binary assignment within the chosen large-scale or instantaneous model, when solved to optimality by a MISOCP solver;
- *explicit fairness guarantees*, since the max–min objective is optimized directly rather than approximated by heuristic AO schemes;
- a clear *trade-off between optimality and complexity*, controlled by the solver's optimality gap or time limit, which is well-understood in the mixed-integer optimization community.

2.5 Positioning of This Thesis within the Literature

From the above state-of-the-art, we can identify three main trends:

1. **Continuous-phase AO-based designs**: joint beamforming with continuous RIS phases, typically solved by AO with SDR/SCA-based updates [18, 19, 20].
2. **Discrete-phase and selection-based designs**: RIS design with discrete phases and/or RIS activation selection, leading to MINLP or MIQP formulations that are handled by iterative heuristics or relax-and-quantize schemes [22, 23, 24, 25].
3. **Fairness-aware formulations**: max–min SINR objectives in continuous-phase settings, showing the substantial complexity of fairness optimizationwithout integer variables [20].

The binary max–min formulations and MISOCP-based methods developed in this thesis can be seen as a bridge between these strands:

- Like [22, 25], they embrace binary decision variables (RIS assignments) and thus fall into the realm of mixed-integer optimization.
- Like [20], they target max–min SINR as the primary fairness metric, but extend it to multi-BS and multi-RIS settings with explicit RIS allocation decisions.
- By adopting a conic formulation (MISOCP) and exploiting large-scale channel approximations, they provide a clean baseline that certifies the best achievable fairness for the given system model and the chosen level of abstraction.

Subsequent chapters will build on this state-of-the-art overview and systematically compare the performance and complexity of MISOCP-based binary RIS assignment, alternating optimization, quota-based schemes, and fixed beamforming approaches.

Chapter 3

Convex Optimization Background

Convex optimization provides the mathematical and algorithmic backbone for all optimization-based methods in this thesis. The large–scale MISOCP baseline, the instantaneous SOCP beamforming designs, and the mixed–integer joint optimization of RIS allocation and beamforming are all special cases of convex (or mixed–integer convex) programs written in second–order cone form and solved numerically by generic conic solvers such as MOSEK via CVXPY [11, 9, 26, 27].

This chapter summarizes the concepts needed to understand the formulations used later, with a particular focus on epigraph modeling and second–order cone programming.

3.1 Convex Sets and Convex Functions

A set $\mathcal{C} \subseteq \mathbb{R}^n$ is *convex* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ and any $\theta \in [0, 1]$ the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$ lies in $\mathcal{C}$:

$$\theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y} \in \mathcal{C}. \tag{3.1}$$

Examples include affine subspaces, halfspaces, hyperplanes, Euclidean balls, and polyhedra given by linear inequalities [11, 9].

A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is *convex* if its domain is a convex set and

$$f(\theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y}) \;\leq\; \theta f(\boldsymbol{x}) + (1 - \theta) f(\boldsymbol{y}), \quad \forall \boldsymbol{x}, \boldsymbol{y},\ \theta \in [0, 1]. \tag{3.2}$$

Convexity is crucial because for convex problems *any* local optimum is global and can be found efficiently by interior–point methods or other algorithms [11, 9].

3.2 Epigraphs and Epigraph Formulations

The *epigraph* of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined as

$$\operatorname{epi} f \;=\; \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R} \mid f(\boldsymbol{x}) \leq t\}. \tag{3.3}$$

A fundamental characterization is

$$f \text{ is convex} \quad \Longleftrightarrow \quad \operatorname{epi} f \text{ is a convex set} \tag{3.4}$$

[11, 9].

Epigraphs are used to transform nonlinear objectives or constraints into equivalent forms that a conic solver can handle. Three basic patterns appear repeatedly in this thesis.

(a) Epigraph of a Norm

The constraint

$$\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2 \leq t \tag{3.5}$$

is the epigraph of the convex function $\boldsymbol{x} \mapsto \|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}\|_2$. It is also a second–order cone constraint (Section 3.4), so it fits directly into the MISOCP beamforming and SINR models.

(b) Epigraph of a Maximum

The pointwise maximum of affine functions

$$f(\boldsymbol{x}) = \max_{i=1,\dots,m} \left(\boldsymbol{a}_i^{\mathrm{H}}\boldsymbol{x} + b_i\right) \tag{3.6}$$

has an epigraph representation

$$f(\boldsymbol{x}) \leq \tau \quad \Longleftrightarrow \quad \boldsymbol{a}_i^{\mathrm{H}}\boldsymbol{x} + b_i \leq \tau,\ i = 1, \dots, m. \tag{3.7}$$

This pattern is used when we collect multiple per–user/per–scenario SINR constraints into a single scalar fairness variable τ in the max–min formulation (Chapters 4–5).

(c) Epigraph of the Negative Log

In rate expressions $\log_2(1+\mathrm{SINR}_u)$ we often lower–bound the SINR, e.g., by introducing τ with $\mathrm{SINR}_u \geq \tau$. Then

$$R_u = \log_2(1 + \mathrm{SINR}_u) \geq \log_2(1 + \tau), \tag{3.8}$$

so optimizing the minimum SINR τ controls the minimum user rate. In this way the nonlinearity of the log does not appear explicitly in the solver; only linear and conic constraints on τ are required.

Epigraph formulations are the main modeling tool that allow us to write max–min SINR problems in the conic form [11, 9, 27].

3.3 Standard Convex Problem Classes

A general convex optimization problem can be written as

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, m, \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{3.9}$$

where $f_0, \dots, f_m$ are convex functions and $\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ is an affine subspace [11, 9]. Several important subclasses appear in this thesis.

3.3.1 Linear Programming (LP)

A linear program has the form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^{\mathrm{H}}\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{A}_{\mathrm{eq}}\boldsymbol{x} = \boldsymbol{b}_{\mathrm{eq}}. \end{aligned} \tag{3.10}$$

LPs provide the geometric intuition (polyhedra, extreme points, simplex–type methods) and serve as building blocks for more general conic problems [28].

3.3.2 Quadratic Programming (QP)

In a quadratic program the objective is quadratic and the constraints are linear:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \tfrac{1}{2}\boldsymbol{x}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{H}}\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{A}_{\mathrm{eq}}\boldsymbol{x} = \boldsymbol{b}_{\mathrm{eq}}, \end{aligned} \tag{3.11}$$

with $\boldsymbol{P} \succeq \boldsymbol{0}$ for convexity. Classical power control problems, LMMSE designs, and regularized least–squares can all be expressed as QPs.

3.3.3 Conic Programs

A *conic program* minimizes a linear function over the intersection of an affine set with a convex cone. The RIS problems in this thesis are modeled as (mixed–integer) *second–order cone programs*, optionally with binary variables that encode RIS allocation decisions.

3.4 Second–Order Cone Programming

The (standard) second–order cone in $\mathbb{R}^k$ is

$$\mathcal{Q}^k = \left\{ (\boldsymbol{u}, v) \in \mathbb{R}^{k-1} \times \mathbb{R} \,\middle|\, \|\boldsymbol{u}\|_2 \leq v \right\}. \tag{3.12}$$

An SOCP in *inequality form* is

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^{\mathrm{H}}\boldsymbol{x} \\ \text{s.t.} \quad & \big\|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\big\|_2 \leq \boldsymbol{c}_i^{\mathrm{H}}\boldsymbol{x} + d_i, \quad i = 1, \dots, m, \\ & \boldsymbol{F}\boldsymbol{x} = \boldsymbol{g}, \end{aligned} \tag{3.13}$$

with data $(\boldsymbol{A}_i, \boldsymbol{b}_i, \boldsymbol{c}_i, d_i, \boldsymbol{F}, \boldsymbol{g})$ [29].

Many quadratic constraints arising in SINR formulations can be expressed in this form.

3.4.1 Rotated Second–Order Cones

The *rotated* second–order cone in $\mathbb{R}^k$ is

$$\mathcal{Q}_r^k = \left\{ (\boldsymbol{u}, v, w) \in \mathbb{R}^{k-2} \times \mathbb{R} \times \mathbb{R} \,\middle|\, 2vw \geq \|\boldsymbol{u}\|_2^2,\ v \geq 0,\ w \geq 0 \right\}. \tag{3.14}$$

It is especially useful to represent inequalities of the form

$$\big\|\boldsymbol{u}(\boldsymbol{x})\big\|_2^2 \leq \alpha(\boldsymbol{x})\,\beta(\boldsymbol{x}) \tag{3.15}$$

with $\alpha(\boldsymbol{x}), \beta(\boldsymbol{x}) \geq 0$. Rotated cones are one way to treat squared signal terms in SINR constraints while preserving convexity in the decision variables.

3.4.2 SINR Constraints in Cone Form

For a given user u with serving BS $b = b(u)$, the instantaneous SINR in this thesis has the form

$$\mathrm{SINR}_u = \frac{P_b/m_b\,|A_{u,b}|^2}{\sum\limits_{j \neq b} P_j/m_j\,|A_{u,j}|^2 + \sigma^2}, \tag{3.16}$$

where $A_{u,j}$ is the coherent sum of direct and RIS–aided channels from BS j to user u, and m_j is the number of active beams/RISs of BS j.

Imposing a minimum SINR requirement $\mathrm{SINR}_u \geq \tau^{-2}$ can be written as the SOCP constraint

$$\left\| \begin{bmatrix} \left\{ \sqrt{P_j/m_j}\, \Re\{A_{u,j}\} \right\}_{j \neq b} \\ \left\{ \sqrt{P_j/m_j}\, \Im\{A_{u,j}\} \right\}_{j \neq b} \\ \sigma \end{bmatrix} \right\|_2 \leq \tau \sqrt{\frac{P_b}{m_b}}\, \Re\{A_{u,b}\}, \quad \Im\{A_{u,b}\} = 0,\ \Re\{A_{u,b}\} \geq \rho, \tag{3.17}$$

for some small $\rho > 0$. Feasibility of this cone constraint guarantees $\mathrm{SINR}_u \geq 1/\tau^2$. This cone–based SINR constraint is the core ingredient of the SOCP beamforming designs used both in the large–scale MISOCP baseline and in the instantaneous joint optimization chapters.

3.5 Mixed–Integer Second–Order Cone Programming (MISOCP)

To model RIS allocation we introduce binary variables $z_{r,b} \in \{0, 1\}$ that indicate whether RIS r is assigned to BS b, with mutual–exclusion and quota constraints such as

$$\sum_{b=1}^{N_{\mathrm{BS}}} z_{r,b} \leq 1, \quad m_b = \sum_r z_{r,b}, \quad m_b \geq 1. \tag{3.18}$$

When these binaries appear only linearly, while all other constraints remain conic, the problem becomes a *mixed–integer second–order cone program* (MISOCP).

MISOCPs are solved globally by combining branch–and–bound search over the integer variables with conic relaxations on each node. The key properties are:

- **Convex continuous relaxations:** fixing the binaries $\boldsymbol{z}$ turns the problem into a pure SOCP, for which strong duality and efficient interior–point methods are available [9].
- **Global optimality:** under reasonable assumptions, modern MISOCP solvers provide certificates (e.g., primal–dual bounds and relative gaps) that guarantee global optimality of the best incumbent solution.
- **Flexible modeling:** integer variables can give on/off RIS assignments, per–BS quotas, cardinality constraints, or restricted candidate sets, all of which are used in the RIS allocation problems in this thesis.

In the large–scale baseline (Chapter 4), we exploit this structure to obtain a globally optimal RIS assignment with respect to a pathloss–based max–min SINR objective at modest problem dimensions.

3.6 Algorithmic Aspects and Implementation

Interior–point methods provide polynomial–time algorithms for LP, SOCP, and SDP, and are implemented at scale in solvers such as MOSEK [9, 27]. At the modeling level, domain–specific languages such as CVXPY allow us to express MISOCPs in a high–level, mathematically transparent form, which is then automatically converted into the cone standard form required by the solver [26].

In this thesis, all conic subproblems (SOCP and MISOCP) are implemented in Python using CVXPY with the MOSEK backend; the corresponding scripts directly realize the generic MISOCP formulation described above.

3.7 Relevance to the RIS Optimization Problems

The convex optimization concepts summarized in this chapter are used as follows in the rest of the thesis:

- Epigraph modeling and max–min reformulation are used to express user fairness via a single scalar SINR variable t.
- Second–order cone programming captures the geometry of SINR constraints and power budgets in both large–scale and instantaneous beamforming problems.
- Mixed–integer SOCP integrates binary RIS allocation decisions with continuous beamformer design, allowing us to derive global baselines and structured heuristics (alternating optimization and quota–based optimization).

These tools enable a convex framework for analyzing and optimizing RIS-assisted networks under stringent fairness and power constraints.

Chapter 4

Large–Scale MISOCP Benchmark for RIS Allocation

In this chapter we introduce a large–scale approximation of the RIS–assisted downlink and formulate the RIS–to–BS assignment as a mixed–integer second–order cone program (MISOCP). The formulation operates directly on large–scale channel gains (pathloss and Rician K–factors) and optimizes a binary assignment of RIS panels to BSs with a max–min SINR objective. Because the model is purely large–scale and the optimization problem is convex apart from the binary variables, standard MISOCP solvers can compute globally optimal assignments for the system sizes considered here.

This provides a conceptually clean *benchmark*: it reveals how much performance could be gained from RIS deployment if only the geometry and large–scale fading mattered, and if RISs could be configured with coherent combining. Later chapters compare this benchmark against more realistic microscopic optimization schemes that operate on instantaneous channel realizations.

4.1 System model and large–scale SINR

We already introduced the system topology in Chapter 1. It stays consistent in this chapter.

Large–scale propagation is described by non–negative channel power gains

$$g_{\mathrm{ud}} \quad (\text{BS } b \text{ to UE } u), \qquad g_{\mathrm{ur}} \quad (\text{RIS } r \text{ to UE } u), \qquad g_{\mathrm{rb}} \quad (\text{BS } b \text{ to RIS } r).$$

In addition, the RIS–UE links are characterized by Rician K–factors $\kappa[u,r] \geq 0$ that depend on whether the RIS–UE path is line–of–sight or non–line–of–sight.

Each user u is associated with a serving BS $b(u) \in \{0,\dots,B-1\}$, determined by large–scale geometry (for instance, strongest path gain). We assume that each BS uses a single downlink data stream and applies maximum–ratio transmit (MRT) beamforming towards its served users on the large–scale level. The RIS panels are configured to provide coherent combining towards the serving BS of each user, consistent with the large–scale Rician factors.

Binary RIS–BS assignment

RIS assignment is represented by a binary matrix $Z = [z_{r,b}] \in \{0,1\}^{R\times B}$, where

$$z_{r,b} = \begin{cases} 1, & \text{if RIS } r \text{ is assigned to BS } b, \\ 0, & \text{otherwise.} \end{cases}$$

Each RIS can be controlled by at most one BS,

$$\sum_{b=0}^{B-1} z_{r,b} \leq 1, \qquad r = 0, \dots, R-1, \tag{4.1}$$

and each BS should receive at least one RIS,

$$n_b \triangleq \sum_{r=0}^{R-1} z_{r,b} \geq 1, \qquad b = 0, \dots, B-1. \tag{4.2}$$

Here n_b denotes the number of RISs assigned to BS b.

We adopt an equal power–split model across the RIS directions of each BS. If BS b has power P_b and controls m_b RISs, the effective power per RIS beam scales as P/m_b, and the MRT gain per BS contributes a factor $\sqrt{PM_{\mathrm{BS}}/m_b}$ to the coherent signal.

Large–scale SINR expression

For a given assignment Z, the large–scale SINR of user u served by BS b can be written as

$$\widehat{\mathrm{SINR}}_u(Z) = \frac{\mathrm{signal}_u(Z)}{\mathrm{interf}_u(Z) + \sigma_n^2}, \tag{4.3}$$

with

$$\begin{aligned} \mathrm{signal}_u(Z) = \underbrace{\gamma_{u,b}^2 P_b}_{\text{direct}} &+ \underbrace{\left(\sum_{r=1}^{N_{\mathrm{RIS}}} z_{r,b}\, \gamma_{u,r} \gamma_{r,b} \kappa_{u,r} \sqrt{\frac{P_b M_{\mathrm{BS}}}{m_b}}\, M_{\mathrm{RIS}} \right)^2}_{\text{coherent RIS-aided signal}} \\ &+ \underbrace{\sum_{r=1}^{N_{\mathrm{RIS}}} z_{r,b}\, \gamma_{u,r}^2 \gamma_{r,d}^2 \bar{\kappa}_{u,r}^2 \frac{P_b M_{\mathrm{BS}}}{m_b} M_{\mathrm{RIS}},}_{\text{incoherent self-term}} \end{aligned} \tag{4.4}$$

and

$$\mathrm{interf}_u(Z) = \underbrace{\sum_{j \neq b} \gamma_{u,j}^2 P_j}_{\text{direct inter-cell}} + \underbrace{\sum_{j \neq b} \sum_{r=1}^{N_{\mathrm{RIS}}} z_{r,j}\, \gamma_{u,r}^2 \gamma_{r,j}^2 P_j M_{\mathrm{RIS}}}_{\text{RIS-aided inter-cell}}. \tag{4.5}$$

Here $\gamma_{u,b}$, $\gamma_{u,r}$, and $\gamma_{r,b}$ denote the effective large–scale channel amplitudes (square roots of g_{ud}, g_{ur}, and g_{rb}), $\kappa_{u,r} = \sqrt{(K/(1+K))}$ where K is the Rician factor of the RIS–UE link, and $\bar{\kappa}_{u,r} = \sqrt{(1/(1+K))}$ summarizes the non–coherent component. These equations are equivalent to those already introduced in 1.16. They are re-expressed for a convenience of notation in the following sections.

The equal power split between RISs shows through $1/m_d$ and $1/\sqrt{m_d}$, so the SINR is a nonlinear function of the binary vector $(z_{r,b})$.

4.2 Max–min SINR MISOCP formulation

The design goal is to choose the binary assignment matrix Z so as to maximize the worst user SINR:

$$\max_{Z \in \{0,1\}^{R \times B}} \; \min_{u=1,\dots,U} \widehat{\mathrm{SINR}}_u(Z) \quad \text{s.t. } (4.1),\ (4.2). \tag{4.6}$$

Introducing an auxiliary scalar $\tau \geq 0$ that represents the minimum SINR, we can equivalently write

$$\begin{aligned} \max_{\tau, Z} \quad & \tau \\ \text{s.t.} \quad & \widehat{\mathrm{SINR}}_u(Z) \;\geq\; \tau, \qquad u = 1, \ldots, N_{\mathrm{UE}}, \\ & Z \in \{0,1\}^{R \times B}, \\ & (4.1),\ (4.2). \end{aligned} \tag{4.7}$$

Each constraint $\widehat{\mathrm{SINR}}_u(Z) \geq \tau$ can be rearranged as

$$\mathrm{signal}_u(Z) \;\geq\; \tau\big(\mathrm{interf}_u(Z) + \sigma_n^2\big).$$

The only non–convex part is the squared coherent term, which has the form $\left(\sum_r z_{r,b} a_{u,r}\right)^2$ for suitable coefficients $a_{u,r}$. This is handled by introducing a real amplitude variable a_u for each user and an auxiliary non–negative variable y_u approximating a_u^2, together with a standard rotated second–order cone constraint of the form

$$\|[2a_u,\ y_u - 1]^\top\|_2 \;\leq\; y_u + 1,$$

which enforces $y_u \geq a_u^2$ in a convex way.

The dependence on m_b is represented using continuous variables that approximate $1/m_b$ and $1/\sqrt{m_b}$ and are linked to the binary variables $(z_{r,b})$ through linear inequalities. With these auxiliary variables, all user SINR constraints can be written as affine and second–order cone constraints in the continuous variables, while Z remains binary. The overall problem is therefore a mixed–integer second–order cone program (MISOCP) that can be solved to global optimality by modern solvers.

4.3 Algorithm for the benchmark

The benchmark evaluation follows a simple repeated procedure over N_{ep} independent episodes (here $N_{\mathrm{ep}} = 20$):

- **Initialization**
  - Fix the system parameters $(B, R, U, M_{\mathrm{BS}}, M_{\mathrm{RIS}}, P_b, \sigma_n^2)$.
  - Define a geometric deployment model for BSs, RISs, and UEs, and the corresponding large–scale channel gains and Rician K–factors.
  - Prepare storage for the per–episode, per–BS sum–rates in two configurations: without RISs and with the MISOCP assignment.
- **Per–episode steps**
  - Draw a new realization of the geometry and large–scale channels.
  - *No RIS configuration:*
    - ∗ Disable all RISs and let each BS transmit with a single MRT beam based on the direct BS–UE channels.
    - ∗ Compute the resulting large–scale SINR for each user and convert to rates $\log_2(1 + \mathrm{SINR})$.
    - ∗ For each BS, we compute the sum-rate of its associated users and average the sum-rate over NN to obtain a per-BS sum-rate sample.
  - *Large–scale MISOCP configuration:*

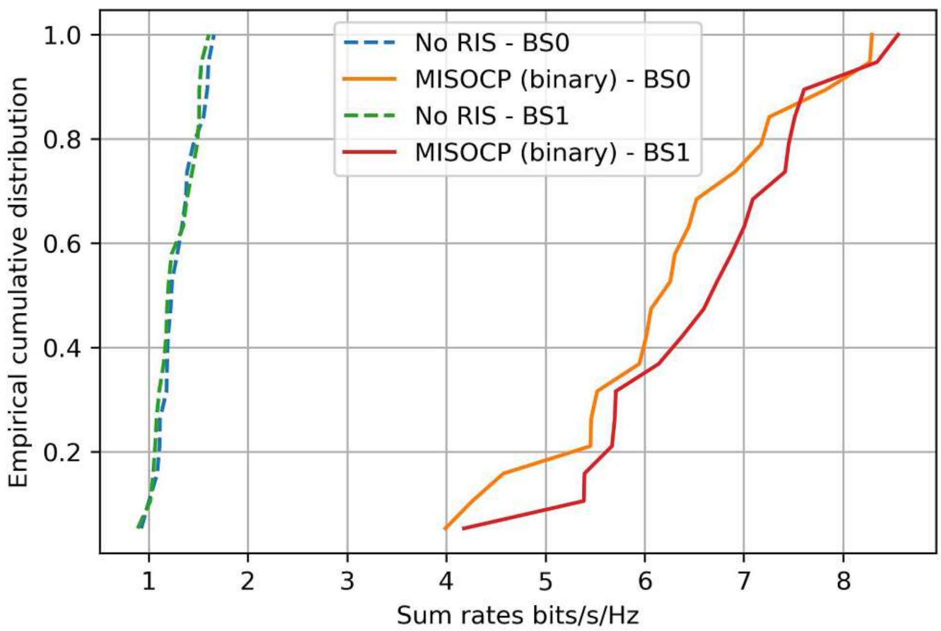

Figure 4.1: Per–BS ECDF of sum–rates for the large–scale MISOCP benchmark ("MISOCP (binary)") and for the system without RISs ("No RIS").

 * Solve the max–min MISOCP in (4.7) for the current geometry to obtain the optimal assignment $Z^\star$ and minimum SINR $t^\star$.
 * Evaluate the large–scale SINR $\widehat{\mathrm{SINR}}_u(Z^\star)$ for all users and convert to rates.
 * For each BS, average the rates of its associated users to obtain a per–BS sum–rate sample under the MISOCP benchmark.

- **Post–processing**
  - From all episodes, collect the per–BS sum–rate samples for both configurations.
  - Form empirical cumulative distribution functions (ECDFs) of these samples and plot them, either per BS or aggregated over all BSs.

This description focuses on the conceptual steps.

4.4 Per–BS ECDF of sum–rates

Figure 4.1 shows the empirical cumulative distribution functions of the per–BS sum–rates (in bit-s/s/Hz) across $N_{\mathrm{ep}} = 20$ episodes, evaluated separately for the two BSs. For each episode and each BS we record

- the mean per–user rate when all RISs are disabled (*No RIS – BS0/BS1*), and
- the mean per–user rate when RISs are assigned by the large–scale MISOCP benchmark (*MISOCP (binary) – BS0/BS1*).

The dashed blue and green curves correspond to the "No RIS" configuration for BS0 and BS1, respectively. They lie roughly in the interval 1–1.7 bits/s/Hz and are almost indistinguishable. This indicates that, without RISs, both BSs operate at similar (and relatively low) large–scale sum–rates across episodes.

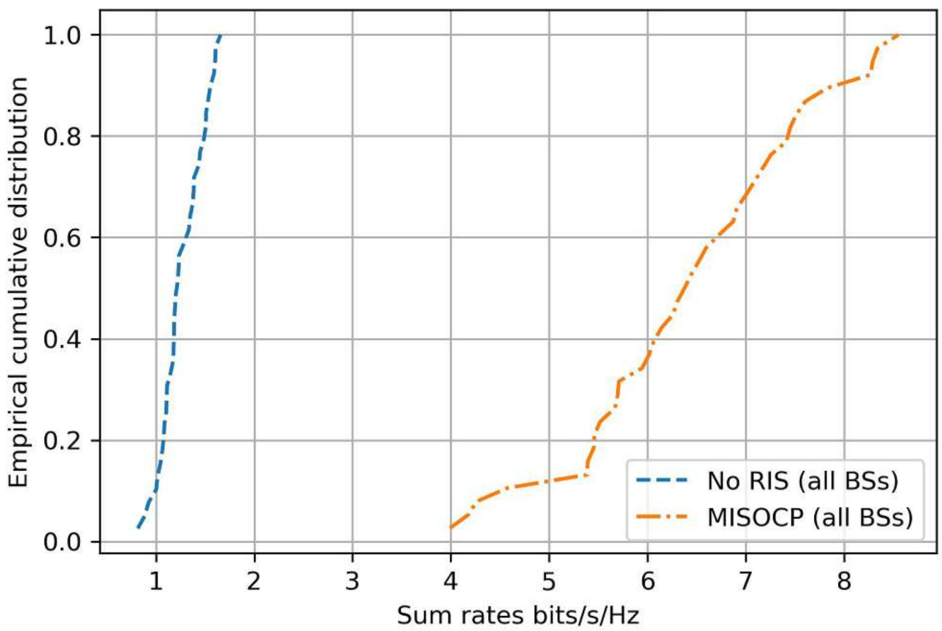

Figure 4.2: ECDF of sum–rates aggregated over all BSs and episodes, for the system without RISs and with the large–scale MISOCP benchmark.

The solid orange and red curves show the MISOCP benchmark for BS0 and BS1. Both curves are shifted far to the right, approximately into the range 4–8.5 bits/s/Hz. This right–shift of the entire ECDF means that, for every episode, the optimized RIS assignment greatly increases the per–BS sum–rate compared to the "No RIS" case.

The two MISOCP curves are not exactly on top of each other: one BS typically achieves slightly higher sum–rates at most quantiles. This asymmetry reflects the underlying geometry. Depending on the positions of the BSs, RISs, and users, the globally optimal assignment may attach more RISs—or better located RISs—to one BS than the other to maximize the worst user SINR. Nevertheless, the curves remain relatively close, indicating that the max–min objective maintains a reasonably balanced performance between the two cells even under this simplified model.

4.5 ECDF aggregated across all BSs

The second plot, Fig. 4.2, aggregates all per–BS sum–rate samples across both BSs and all episodes and shows a single ECDF for each configuration: "No RIS (all BSs)" and "MISOCP (all BSs)".

The dashed curve (No RIS) is essentially the same as in the per–BS view: almost all mass lies between 1 and 1.7 bits/s/Hz. The dash–dotted curve (MISOCP, all BSs) again lies between approximately 4 and 8.5 bits/s/Hz, with almost no overlap with the dashed curve. From a network–wide perspective, this shows that *every* cell in *every* episode benefits from the optimized RIS assignment in the large–scale model: the worst cells are much better than the worst cells without RISs, and the best cells reach significantly higher rates.

Viewing the results in this aggregated way emphasizes that the MISOCP benchmark does not only help a few favorable geometries but consistently improves performance for the entire ensemble of deployments considered in the simulations.

4.6 Discussion: benchmark nature and limitations

The large–scale MISOCP assignment provides a useful reference point:

- It is *globally optimal* for the chosen large–scale SINR model and system size, so any other scheme based on the same model and constraints cannot do better.
- It isolates the impact of geometry and pathloss, with RISs used in an idealized coherent fashion and equal power split across assigned panels.
- The ECDFs show a large potential gain from deploying and optimally assigning RISs, compared to a network without RISs.

At the same time, several limitations must be kept in mind:

- The model ignores microscopic fading and instantaneous beamforming decisions; it operates on long–term channel statistics only.
- The coherent RIS term assumes that all RIS elements can be perfectly phase–aligned for each user, which is overly optimistic in a multi–user multi–cell setting.
- Power is split uniformly across RIS beams of a BS; in practice, joint optimization of beamforming, power allocation, and RIS phases may achieve higher rates.
- The MISOCP involves binary variables and second–order cone constraints; its complexity grows quickly with the number of RISs and BSs, limiting scalability.

Because of these simplifications, the large–scale MISOCP should be interpreted as a *conceptual benchmark* for the specific large–scale model rather than a strict upper bound on what is achievable with more detailed instantaneous optimization. In fact, the joint beamforming and RIS–allocation schemes based on instantaneous SINR developed in the next chapter exploit more channel information and produce higher user rates in our simulations, underscoring that the benchmark captures only part of the available performance gain.

Chapter 5

Joint Optimization of Beamforming and RIS Allocation

In Chapter 4 we studied a large–scale baseline where the RIS–to–BS assignment was optimized using statistical channel information only, and the RIS configuration was kept fixed during evaluation. In this chapter we turn to a more challenging problem: the *joint* optimization of the BS beamforming vectors and the RIS–to–BS assignment based on instantaneous (or sample–average) channel realizations. The goal is to exploit the full microscopic channel knowledge in order to maximize a max–min fairness metric across all users, while keeping the optimization problem computationally tractable.

We describe two algorithmic approaches: an alternating optimization (AO) scheme, which mirrors the original mixed–integer structure of the problem, and a quota–based optimization (QO) scheme, which reduces the combinatorial burden by searching over per–BS RIS quotas. Finally, we present numerical results for the quota–based design and discuss the main insights and limitations of the overall approach.

5.1 Joint Max–Min Rate Formulation

The joint design problem consists of choosing the binary RIS assignment variables $\{z_{rj}\}$ and the continuous beamforming vectors $\{\mathbf{w}_{j,r}\}$ so as to maximize a fairness–oriented performance metric. In this thesis we focus on the classic max–min rate objective,

$$\begin{aligned}
\max_{\{z_{rj}\},\{\mathbf{w}_{j,r}\}} \quad & \min_{u=1,\ldots,U} \; SI\hat{N}R_u && (5.1)\\
\text{s.t.} \quad & z_{rj} \in \{0,1\}, \qquad r = 1,\ldots,R,\; j = 1,\ldots,B, && (5.2)\\
& \sum_{j=1}^{B} z_{rj} = 1, \qquad r = 1,\ldots,R, && (5.3)\\
& \sum_{r=1}^{R} \left\|\mathbf{w}_{j,r}\right\|_2^2 \leq 1, \qquad j = 1,\ldots,B. && (5.4)
\end{aligned}$$

The optimization problem (5.1)–(5.4) is a *mixed–integer nonlinear program* (MINLP) with a highly nonconvex objective. An exhaustive search over all RIS assignments would require $\mathcal{O}(B^R)$ configurations, which is prohibitive even for moderate values of B and R. Therefore, we use algorithmic decompositions that exploit the structure of the problem while providing good performance in practice.

5.2 Alternating Optimization Scheme

A conceptually straightforward way to tackle (5.1)–(5.4) is to fix one block of variables and optimize over the other. This leads to an alternating optimization (AO) algorithm that updates the beamformers and RIS assignments in turn. The AO procedure used in this work consists of two main steps, denoted Step (A) and Step (B), which are iterated until the max–min SINR stabilizes.

Step (A): Beamforming SOCP (given $\{z_{rj}\}$)

In Step (A) we keep a feasible RIS assignment $\mathbf{Z} = [z_{rj}]$ fixed and optimize the beamformers. For each BS j the number of used RISs is $m_j = \max\{1, \sum_r z_{rj}\}$. The variables of this subproblem are the beamforming vectors $\{\mathbf{w}_{j,r}\}$ and a slack variable $\tau > 0$ representing the lower bound of the worst–user SINR level.

For notational convenience we introduce

$$A_{u,j}(\mathbf{w}) \triangleq \sum_{r=1}^{R} z_{rj}\, \alpha_{u,j,r} = \sum_{r=1}^{R} z_{rj}\, \mathbf{c}_{u,j,r}^H \mathbf{w}_{j,r}, \tag{5.5}$$

where $\mathbf{c}_{u,j,r}$ denotes the effective steering vector for the composite direct and RIS–assisted channel of the triple (u, j, r). We define the effective per–user, per–BS, per–RIS channel vector

$$\boldsymbol{c}_{u,j,r}^{(s)} \triangleq \boldsymbol{h}_{u,j,\text{dir}}^{(s)} + \boldsymbol{h}_{u,j,r,\text{RIS}}^{(s)} \in \mathbb{C}^{M_{\text{BS}}}. \tag{5.6}$$

Given a fixed beamforming vector $\boldsymbol{w}_{j,r}$, the corresponding complex scalar contribution of RIS r to user u from BS j in snapshot s is

$$\alpha_{u,j,r}^{(s)} \triangleq \left(\boldsymbol{c}_{u,j,r}^{(s)}\right)^H \boldsymbol{w}_{j,r}, \qquad \alpha_{u,j,r}^{(s)} \in \mathbb{C}. \tag{5.7}$$

We also introduce user–specific phases θ_u so that the useful signal can be rotated onto the real axis, and a small constant $\rho > 0$ that avoids degenerate solutions with vanishing useful signal.

Using a standard second–order cone reformulation of the SINR constraints, the beamforming subproblem can be written as

$$\max_{\{\mathbf{w}_{j,r}\},\, \tau} \quad \tau \tag{5.8}$$

$$\text{s.t.} \quad \left\| \begin{bmatrix} \left\{\sqrt{P_j/m_j}\, \Re\{A_{u,j}(\mathbf{w})\}\right\}_{j\neq b} \\ \left\{\sqrt{P_j/m_j}\, \Im\{A_{u,j}(\mathbf{w})\}\right\}_{j\neq b} \\ \sigma \end{bmatrix} \right\|_2 \le \frac{1}{\sqrt{\tau}} \sqrt{\frac{P_b}{m_b}}\, \Re\!\left\{e^{-\jmath\theta_u} A_{u,b}(\mathbf{w})\right\}, \quad \forall u, \tag{5.9}$$

$$\Im\!\left\{e^{-\jmath\theta_u} A_{u,b}(\mathbf{w})\right\} = 0, \quad \Re\!\left\{e^{-\jmath\theta_u} A_{u,b}(\mathbf{w})\right\} \ge \rho, \quad \forall u, \tag{5.10}$$

$$\sum_{r=1}^{R} \left\|\mathbf{w}_{j,r}\right\|_2^2 \le 1, \quad j = 1, \ldots, B. \tag{5.11}$$

The SOC constraint (5.9) makes sure $SINR_u \ge 1/\sqrt{\tau}$ for all users under the equal–power model, while the phase constraints (5.10) select the best common phase rotation for the desired signal. For fixed $\mathbf{Z}$ the problem (5.8)–(5.11) is an SOCP and is solved globally using MOSEK. In the implementation, a bisection on τ is used instead of a direct maximization, which yields the same optimum max–min SINR. This works because for fixed τ, the problem in (5.1) is convex and the max function preserves convexity. Therefore, a bisection will run into a global optimum. Without these properties, it may get caught in a local optimum.

Step (B): RIS–Allocation MISOCP (given $\{\mathbf{w}_{j,r}\}$)

In Step (B) we keep the beamformers $\{\mathbf{w}_{j,r}\}$ fixed and update the RIS assignment. For each user, BS, and RIS we precompute the complex coefficients

$$\alpha_{u,j,r} = \mathbf{c}_{u,j,r}^{H}\mathbf{w}_{j,r}, \tag{5.12}$$

which are now constants. The effect of different assignments appears only through the linear combination

$$A_{u,j}(\mathbf{z}) \triangleq \sum_{r=1}^{R} z_{rj}\,\alpha_{u,j,r}. \tag{5.13}$$

The equal–power factors m_j depend on the assignment and therefore limit convexity. To keep the SINR constraints in second–order cone form we *freeze* these factors inside the cones:

$$m_j^{(\text{hold})} \triangleq \max\left\{1, \sum_{r=1}^{R} z_{rj}^{(\text{current})}\right\}, \qquad j = 1, \ldots, B, \tag{5.14}$$

where $\mathbf{Z}^{(\text{current})}$ is the assignment from the previous AO iterate. Using $m_j^{(\text{hold})}$ in the scaling of the interference terms and of the useful signal preserves convexity while capturing the power split reasonably well.

The RIS allocation subproblem then becomes

$$\max_{\{z_{rj}\},t} \quad t \tag{5.15}$$

$$\text{s.t.} \quad \left\| \begin{bmatrix} \left\{\sqrt{P_j/m_j^{(\text{hold})}}\,\Re\{A_{u,j}(\mathbf{z})\}\right\}_{j\neq b} \\ \left\{\sqrt{P_j/m_j^{(\text{hold})}}\,\Im\{A_{u,j}(\mathbf{z})\}\right\}_{j\neq b} \\ \sigma \end{bmatrix} \right\|_2 \le \frac{1}{\sqrt{t}}\sqrt{\frac{P_b}{m_b^{(\text{hold})}}}\,\Re\{e^{-\jmath\theta_u}A_{u,b}(\mathbf{z})\}, \quad \forall u, \tag{5.16}$$

$$\Im\{e^{-\jmath\theta_u}A_{u,b}(\mathbf{z})\} = 0, \quad \Re\{e^{-\jmath\theta_u}A_{u,b}(\mathbf{z})\} \ge \rho, \quad \forall u, \tag{5.17}$$

$$\sum_{j=1}^{B} z_{rj} \le 1, \quad r = 1, \ldots, R, \tag{5.18}$$

$$z_{rj} \in \{0,1\}, \quad r = 1, \ldots, R, \; j = 1, \ldots, B. \tag{5.19}$$

The SOC constraints (5.16)–(5.17) are identical to those of Step (A), but now the variables are the assignment binaries $\{z_{rj}\}$. The problem (5.15)–(5.19) is a mixed–integer second–order cone program (MISOCP) and is again solved by MOSEK.

Overall AO Algorithm and Discussion

The complete AO procedure can be summarized as follows. We start from a feasible assignment in which every BS controls at least one RIS and compute the corresponding m_j. Then we iterate:

1. **Step (A):** solve the SOCP (5.8)–(5.11) to obtain new beamformers $\{\mathbf{w}_{j,r}\}$ and a slack value t.

2. **Step (B):** freeze the power–split factors inside the cones, $m_j^{(\text{hold})} = \max\{1, \sum_r z_{rj}\}$, and solve the MISOCP (5.15)–(5.19) to obtain an updated assignment $\{z_{rj}\}$ and a new value of t.

3. update $m_j \leftarrow \max\{1, \sum_r z_{rj}\}$ and repeat.

The outer loop is terminated when the relative improvement of the objective becomes small, for example when

$$\frac{\left|t^{(\ell)} - t^{(\ell-1)}\right|}{\max\{1, t^{(\ell-1)}\}} \leq \varepsilon$$

for some tolerance $\varepsilon > 0$, or when a maximum number of AO iterations is reached. Warm–starting the SOCP in Step (A) with the beamformers from the previous iterate and the MISOCP in Step (B) with the last assignment further reduces the runtime.

The alternating scheme has several desirable properties:

- It closely mirrors the original MINLP formulation and operates directly on quantities that are meaningful from a physical perspective (instantaneous SINR, effective channels, per–BS powers).
- For a fixed assignment, the beamforming subproblem in Step (A) admits an efficient SOCP formulation with strong numerical stability.
- The resulting beamformers are explicitly max–min fair for the given RIS allocation.

However, the limitations are equally important:

- The overall problem remains nonconvex and the AO algorithm is not guaranteed to find the global optimum. The quality of the solution depends on the initialization of both $\mathbf{Z}$ and $\{\mathbf{w}_{j,r}\}$.
- Each AO iteration requires solving at least one large SOCP and one MISOCP, which becomes computationally expensive when evaluating many channel realizations.
- The AO scheme does not provide an explicit handle on the per–BS RIS quotas m_j, and thus it may lead to imbalanced operating points where one BS controls a disproportionately large number of RISs.

These limitations motivate the quota–based optimization approach introduced next.

5.3 Numerical Results for the Alternating Optimization

We evaluate the alternating optimization (AO) scheme of Section 5.2, where beamformers and RIS assignments are updated in turn on the basis of instantaneous channel realizations. The system parameters and channel model are the same as in the large–scale benchmark and in the quota–based design: two BSs, ten RISs, twenty users, and $NN = 20$ microscopic fading realizations per evaluation episode.

5.3.1 All–User ECDFs

Figure 5.1 shows the empirical cumulative distribution function (ECDF) of the per–user ergodic rate for the no–RIS baseline and for the AO–based joint RIS design. All users from all episodes are pooled together, thus providing a global, BS–agnostic view.

The orange curve (Joint RIS — all users) lies clearly to the right of the blue curve (No RIS — all users) over almost the entire support. For any desired rate threshold, a larger fraction of users achieve that rate under the AO joint design than in the no–RIS system. In the central part of the distribution, the AO scheme yields much higher typical user rates, and the upper tail extends beyond 10–14 bps/Hz, whereas the no–RIS system saturates at smaller values.

A distinctive feature of the instantaneous design is the behavior of the *left tail*. Both curves start close to zero rate, indicating that a small fraction of users experience near–outage conditions in some realizations. This is in contrast to the large–scale MISOCP benchmark of

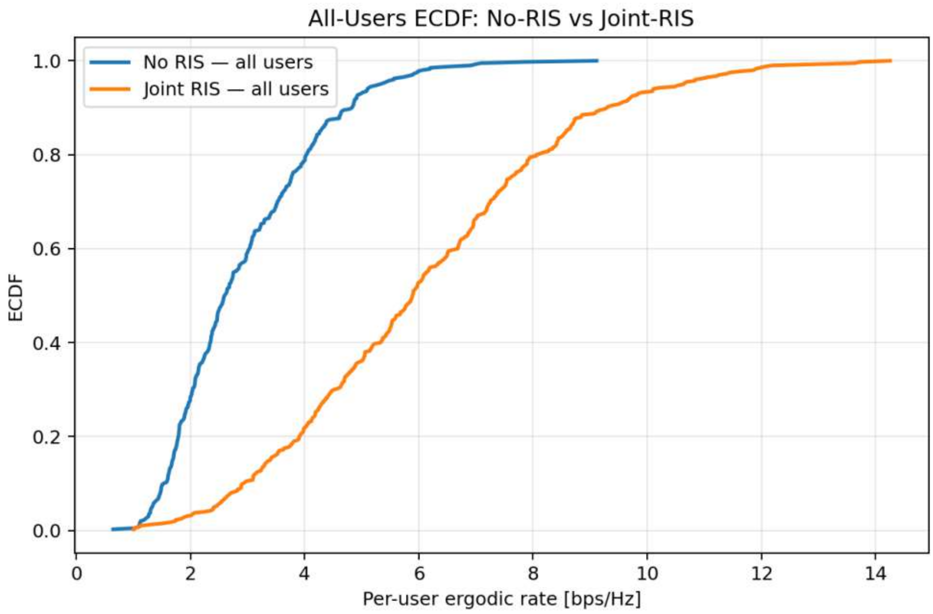

Figure 5.1: All–users ECDF of the per–user ergodic rate for the no–RIS baseline and the alternating–optimization (AO) joint RIS design.

Chapter 4, whose ECDFs started at rates well above zero because the model contained only large–scale gains together with an explicit max–min SINR constraint. Here, geometry and small–scale fading combine: users in unfavorable geometric positions who also draw poor fading realizations can end up with very low instantaneous SINRs, and the heuristic AO updates cannot completely remove these deep left tails. The role of the joint RIS design is therefore to shift and compress the whole distribution towards higher rates, not to guarantee a strict positive minimum rate in every realization.

5.3.2 Per–BS ECDFs

To inspect the cell–wise behavior we split users according to their serving BS and compute per–BS ECDFs of the ergodic rate. The resulting curves are shown in Fig. 5.2. Solid lines correspond to BS 0 and dashed lines to BS 1; blue curves show the no–RIS baseline, and orange curves the AO joint RIS design.

For each BS, the orange curve lies to the right of the corresponding blue curve, confirming that the AO scheme improves the user rates in both cells individually. Typical users of both BSs profit from the RISs, and the maximum observed ergodic rates are markedly higher than in the no–RIS system.

The per–BS plots also highlight the impact of geometry and resource allocation. In the no–RIS case the curves of BS 0 and BS 1 are not perfectly aligned; one BS tends to serve consistently weaker users, which is visible as a left shift of its ECDF. Under the AO joint design, the separation between the two orange curves is reduced but not eliminated: depending on the realization, the AO updates may assign more RISs or more favorable RISs to one BS than to the other, reflecting the underlying geometric asymmetry.

Again, both BSs exhibit deep left tails with a small proportion of users whose ergodic rates fall below 1 bps/Hz. These tails are shorter for the AO design than for the no–RIS baseline, but they are not fully removed. This is a key qualitative difference from the large–scale MISOCP benchmark, where the combination of large–scale modeling and a max–min SINR constraint essentially clipped the left tail away from zero. In the instantaneous setting, near–outage users

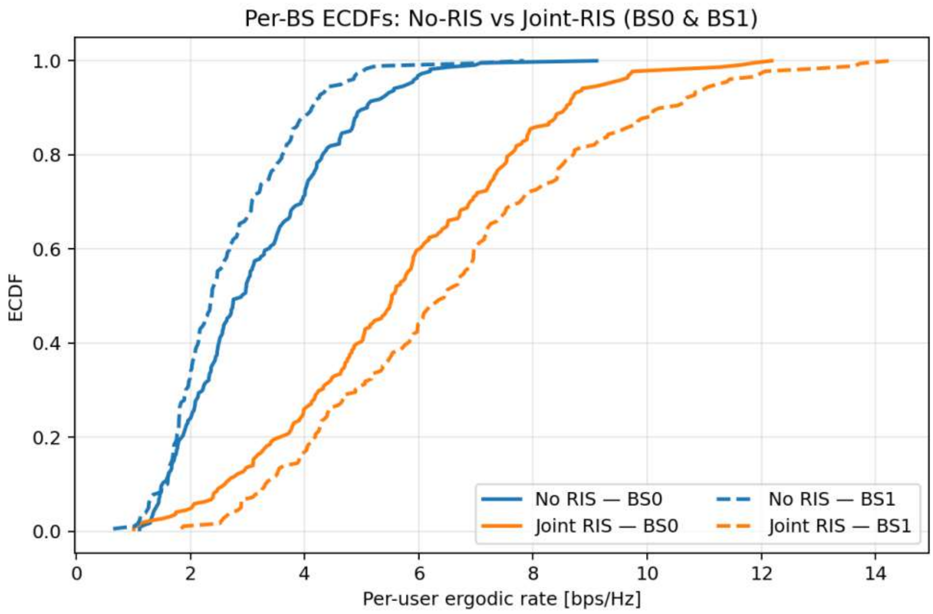

Figure 5.2: Per–BS ECDFs of the per–user ergodic rate for the no–RIS baseline and the AO joint RIS design. Solid: users of BS 0. Dashed: users of BS 1.

remain, even with joint optimization; what changes is the fraction of such users and how far into the distribution they extend.

5.4 Quota–Based Joint Optimization

The key idea of the quota–based method is to separate the global fairness across BSs from the detailed combinatorial assignment of RISs to BSs. Instead of searching directly over all binary matrices $\mathbf{Z}$, we first fix the number of RISs per BS, described by a quota vector $\mathbf{m} = (m_1, \ldots, m_B)$, and then design an assignment that respects these quotas. This reduces the search space substantially and makes it easier to reason about fairness between BSs.

5.4.1 Per–BS RIS Quotas

Given the set of user–to–BS associations $\{\mathcal{U}_j\}_{j=1}^{B}$, an initial guess for the quotas is to allocate RISs proportionally to the number of users per BS,

$$m_j^{(0)} \approx R \cdot \frac{|\mathcal{U}_j|}{\sum_{k=1}^{B} |\mathcal{U}_k|}, \qquad j = 1, \ldots, B, \tag{5.20}$$

followed by rounding and small adjustments to satisfy $\sum_j m_j^{(0)} = R$. For a given system with B BSs and R RISs, the number of possible quota vectors is equal to the number of integer compositions of R into B parts.

5.4.2 Geometry–Based RIS Assignment for Fixed Quotas

For a fixed quota vector $\mathbf{m}$ we next construct a RIS assignment matrix $\mathbf{Z}$ that respects the quotas and favors geometrically strong BS–RIS links. To this end, we draw random isotropic beamforming vectors $\mathbf{w}_{j,r}$ and compute, for each snapshot s, the effective gains $\alpha_{u,j,r}(s)$. We then define a score

$$\text{score}_{j,r} \triangleq \frac{1}{U\,NN} \sum_{u=1}^{U} \sum_{s=1}^{NN} \left|\alpha_{u,j,r}(s)\right|^2. \tag{5.21}$$

Because the random beams are isotropic, the scores $\text{score}_{j,r}$ essentially measure the average squared norm of the composite BS–RIS–UE link and thus depend mainly on the network geometry and path loss.

The RIS assignment problem for fixed quotas can now be formulated as

$$\max_{\mathbf{Z}} \quad \sum_{j=1}^{B} \sum_{r=1}^{R} z_{rj}\, \text{score}_{j,r} \tag{5.22}$$

$$\text{s.t.} \quad \sum_{j=1}^{B} z_{rj} = 1, \qquad r = 1, \dots, R, \tag{5.23}$$

$$\sum_{r=1}^{R} z_{rj} = m_j, \qquad j = 1, \dots, B, \tag{5.24}$$

$$z_{rj} \in \{0, 1\}. \tag{5.25}$$

This is a small integer linear program with a structure that also admits efficient greedy heuristics. In the implementation used for the numerical results, RISs are processed in descending order of a confidence margin and assigned to the BS that offers the highest score among those BSs that still have remaining quota.

5.4.3 Beamforming Design for Fixed Quotas and Assignment

Once the quotas $\mathbf{m}$ and the assignment matrix $\mathbf{Z}$ are fixed, the remaining task is to design beamforming vectors $\{\mathbf{w}_{j,r}\}$ that maximize the minimum achievable rate. This problem is structurally identical to the beamforming subproblem in Step (A) of the AO scheme and is solved by the same SOCP–based method. In order to keep the optimization dimension manageable, the per–snapshot channels are replaced by their sample averages across the NN realizations. The resulting beamforming vectors are then evaluated on all NN snapshots to obtain ergodic rates.

5.4.4 Searching Over Quotas and Fairness Considerations

The quota vector $\mathbf{m}$ plays a role in balancing the performance of different BSs. For each candidate vector $\mathbf{m}$ we perform the two steps above, obtain the ergodic user rates, and compute

$$R_{\min,j}(\mathbf{m}) \triangleq \min_{u \in \mathcal{U}_j} R_u(\mathbf{m}), \tag{5.26}$$

$$R_{\min}(\mathbf{m}) \triangleq \min_{j} R_{\min,j}(\mathbf{m}). \tag{5.27}$$

Here, $R_{\min,j}(\mathbf{m})$ captures per–BS fairness, while $R_{\min}(\mathbf{m})$ provides the system–wide worst–user rate. By exploring the (small) set of candidate quota vectors we can select a configuration that yields a good compromise between high overall minimum rate and balanced service across BSs.

5.5 Numerical Results for the Quota–Based Design

In this section we present empirical results for the quota–based joint optimization scheme. Unless stated otherwise, the system parameters and channel model are identical to those used in Chapter 4. In particular, we consider $B = 2$ BSs, $R = 10$ RISs, and $U = 20$ users, with

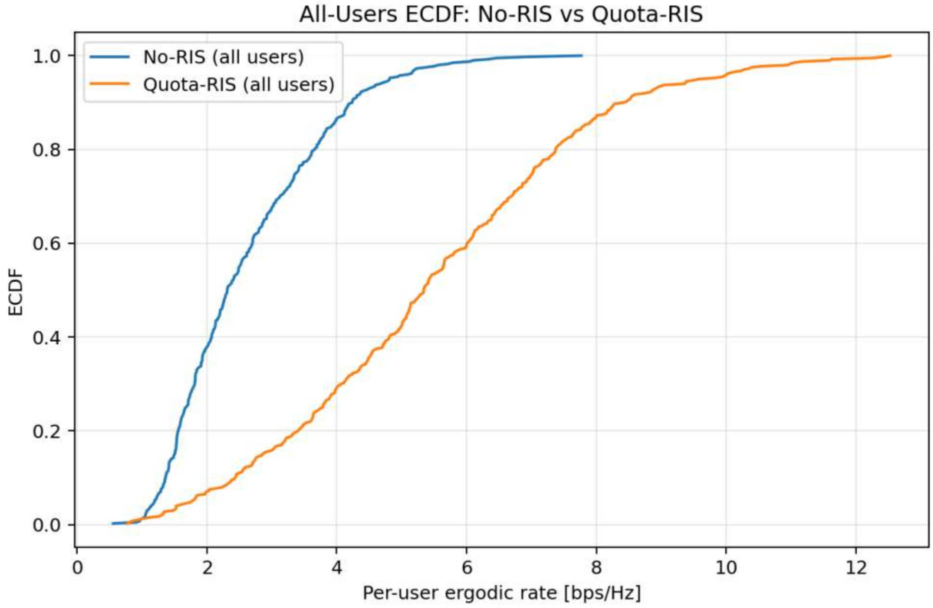

Figure 5.3: All–users ECDF of the per–user ergodic rate for the no–RIS baseline and the quota–based joint optimization scheme (BS–agnostic view).

$NN = 20$ microscopic fading realizations per evaluation episode. The baseline "no–RIS" curves correspond to a system where the RISs are removed and only the direct BS–to–UE channels are present.

5.5.1 All–User ECDFs

Figure 5.3 shows the empirical cumulative distribution function (ECDF) of the per–user ergodic rates for the no–RIS baseline and for the quota–based joint optimization scheme. The curves are obtained by aggregating the rates of all users and all evaluation episodes and then computing the ECDF.

The figure demonstrates a clear and consistent gain from the RIS–assisted joint optimization: the orange curve (Joint RIS) lies almost entirely to the right of the blue curve (No RIS). In the low–rate regime, corresponding to the worst–served 10–20% of users, the quota–based scheme approximately doubles the rates compared to the no–RIS case. Around the median, the gain remains substantial, and the upper tail of the distribution extends to more than 10 bps/Hz, whereas the no–RIS system saturates at significantly smaller values. These results indicate that the joint optimization of RIS allocation and beamforming provides benefits not only to a few lucky users, but to the entire user population.

5.5.2 Per–BS ECDFs and Inter–Cell Fairness

To analyze inter–cell fairness we separate the users according to their serving BS and compute per–BS ECDFs of the ergodic rate. The results are shown in Fig. 5.4. Solid lines correspond to BS 0, and dashed lines to BS 1, for both the no–RIS baseline (blue) and the quota–based joint optimization (orange).

For both BSs the orange curves lie systematically to the right of the corresponding blue curves, confirming that the quota–based RIS design improves the service quality for users of each BS individually. Moreover, the separation between the BS–specific curves shrinks under the RIS–assisted design: while in the no–RIS case one BS exhibits noticeably worse user rates across most quantiles, the quota–based scheme yields ECDFs that are much closer together.

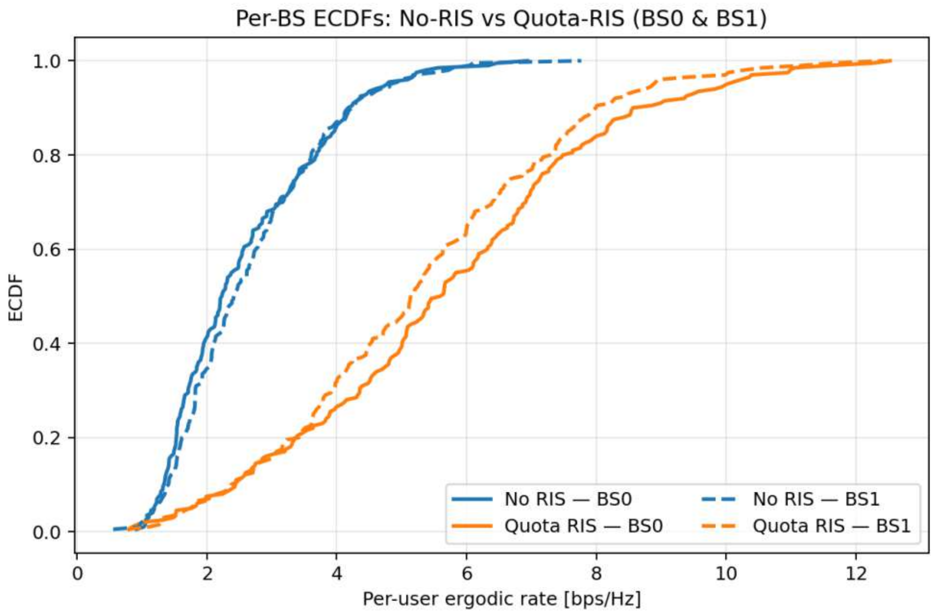

Figure 5.4: Per–BS ECDFs of the per–user ergodic rate for the no–RIS baseline and the quota–based joint optimization scheme. Solid: users of BS 0. Dashed: users of BS 1.

This behavior reflects the role of the quotas $\mathbf{m}$ in balancing the number and quality of RISs allocated to each BS.

5.5.3 Comparison of AO and Quota–Based Results

Comparing the ECDFs obtained with the alternating optimization (AO) scheme in Section 5.3 and with the quota–based (QO) scheme in Section 5.5, we observe that both methods deliver very similar gains over the no–RIS baseline. In both cases the all–user ECDFs are shifted substantially to the right, and the median and upper quantiles increase by several bps/Hz compared to the system without RISs. The per–BS ECDFs confirm that users of both BSs benefit from the presence of RISs under either design.

There are, however, subtle differences. The AO scheme can occasionally push the best users to slightly higher rates, but it may also lead to a more uneven distribution of RISs between BSs, which shows up as a somewhat larger gap between the per–BS ECDFs in some realizations. The QO scheme, by explicitly controlling the per–BS quotas, tends to produce more balanced per–BS ECDFs, with both cells operating at similar performance levels even if their geometries are asymmetric. Taken together, the results suggest that AO and QO achieve comparable overall performance, while QO offers a clearer handle on inter–cell fairness and a lower computational complexity.

5.6 Discussion and Limitations

The numerical results highlight the potential of RISs when combined with a joint design of beamforming and RIS assignment. Compared to a system without RISs, the quota–based method yields a significant right–shift of the ergodic rate ECDFs, both in a BS–agnostic view and when users are grouped by serving BS. The gains are particularly pronounced for the weakest users, which is consistent with the max–min design objective.

At the same time, several limitations must be kept in mind:

- The optimization relies on reasonably accurate CSI, including the BS–to–RIS and RIS–to–UE channels. In practice, estimation and feedback overhead may limit the achievable accuracy, especially for large RISs.
- The joint design is carried out on a slow time scale, with RIS configurations and beamformers assumed to be fixed over many coherence intervals. This is compatible with static or slowly moving users, but may be less effective in highly dynamic scenarios.
- Both the AO and QO algorithms are heuristics for the original MINLP. While the SOCP subproblems are solved globally, the overall solution is not guaranteed to be globally optimal with respect to $\{z_{rj}\}$ and $\{\mathbf{w}_{j,r}\}$.

Despite these caveats, the combination of per–BS quotas, geometry aware RIS assignment, and convex beamforming design strikes a useful balance between performance, fairness, and computational complexity.

A key difference to the large–scale MISOCP benchmark in Chapter 4 is the behavior of the ECDFs in the low–rate regime. In Chapter 4, the purely large–scale SINR model together with an explicit max–min SINR constraint kept all users above a relatively high rate threshold, so the ECDFs had almost no mass near zero and did not exhibit deep left tails. In contrast, the instantaneous joint designs of the present chapter operate under microscopic fading and are implemented via heuristic algorithms. As a result, both the no–RIS and RIS–assisted ECDFs show a small but visible fraction of users with very low ergodic rates. The role of the joint RIS optimization is therefore to substantially shift and compress these distributions towards higher rates, rather than to completely eliminate all near–outage users.

5.7 Summary

This chapter has addressed the joint optimization of BS beamforming and RIS–to–BS assignment for a downlink with RIS–assisted links. We formulated a max–min ergodic rate problem that captures the coupling between discrete RIS allocation variables and continuous beamformers through the instantaneous SINR. Due to the inherent MINLP structure, we proposed two algorithmic approaches: an alternating optimization scheme built around Step (A) (beamforming SOCP) and Step (B) (RIS–allocation MISOCP), and a quota–based optimization scheme that searches over per–BS RIS quotas and employs geometry–based scores for building RIS assignments.

The quota–based method was evaluated numerically and shown to significantly outperform a no–RIS baseline, both in terms of system–wide user rates and in terms of per–BS fairness. These results confirm that RISs, when combined with suitable joint optimization algorithms, can substantially enhance both spectral efficiency and fairness in multi–cell networks. In subsequent chapters we will build on these insights to further refine the optimization methods and to compare the instantaneous designs with the large–scale MISOCP baselines developed earlier.

Chapter 6

RIS Optimization with Fixed Beamforming

In the previous chapter we considered the setting where the beamforming vectors $\{\boldsymbol{w}_{j,r}\}$ and the binary RIS–to–BS assignment variables $\{z_{rj}\}$ are optimized jointly. This joint design maximizes the instantaneous max–min SINR, but the resulting mixed–integer non–linear program (MINLP) is computationally demanding and requires repeated SOCP/MISOCP solves.

In this chapter we investigate a more practical regime in which the beamforming vectors are kept *fixed*, and only the RIS assignment is optimized. This decouples long–term beam design from discrete RIS selection and leads to a smaller MISOCP with fewer continuous variables, while still exploiting the full geometric channel model and the equal power split across used beams.

6.1 Motivation for Fixed Beamforming

The joint alternating scheme of Chapter 5 provides high performance but has several drawbacks:

- **Algorithmic complexity:** Each outer iteration requires one SOCP (beamforming update) and one MISOCP (RIS allocation update), both built on NN fading snapshots. For realistic values (e.g., $B = 2$, $R = 10$, $U = 20$, $M_{\mathrm{BS}} = 50$ and $NN = 20$), this becomes very expensive if repeated for many geometries.
- **Coupling of layers:** The BS beamforming policy becomes tightly coupled to the RIS assignment. Any change of the RIS configuration requires a new SOCP, which is not attractive if beamformers are meant to be updated only on a slow time scale (e.g., per cell layout or long–term statistics).
- **Implementation realism:** In practice, operators often prefer fixed or slowly varying per–BS beam codebooks (e.g., MRT or ZF towards the set of associated UEs), and would like to use RIS assignment as a higher–level resource management tool on top of these beams.

This motivates a *fixed–beamforming* formulation: we precompute one or several beams per BS and RIS using only the geometry and long–term CSI, and then optimize the binary RIS assignments for these fixed beams. The goal remains to maximize a fairness–oriented max–min SINR across users, but the continuous variables $\{\boldsymbol{w}_{j,r}\}$ disappear from the optimization; they enter only as parameters.

6.2 Max–Min SINR Optimization with Fixed Beams

With fixed beams $\{\boldsymbol{w}_{j,r}\}$, the optimization variables are the binary RIS assignment indicators $\{z_{rj}\}$ and a scalar slack $\tau > 0$ representing the minimum user SINR.

We consider a max–min SINR objective as in the previous chapters,

$$\max_{\tau,\, z} \ \tau \quad \text{s.t.} \quad \mathrm{SINR}_u^{(s)}(z) \ \geq \ \tau, \ \ \forall u, \ \forall s, \tag{6.1}$$

under the constraints

$$z_{rj} \in \{0,1\}, \qquad r = 1,\ldots,R, \ j = 1,\ldots,B, \tag{6.2}$$

$$\sum_{j=1}^{B} z_{rj} \leq 1, \qquad r = 1,\ldots,R, \tag{6.3}$$

$$m_j = \max\Big\{1, \sum_{r=1}^{R} z_{rj}\Big\}, \qquad j = 1,\ldots,B. \tag{6.4}$$

Directly enforcing (6.1) for all snapshots s and all users u leads to a large mixed–integer non–linear program. In the implementation we adopt a standard conic relaxation, similar to the large–scale MISOCP in Chapter 4:

- For each user u and snapshot s, we rewrite the constraint $\mathrm{SINR}_u^{(s)}(z) \geq \tau$ in the second–order cone (SOC) form

$$\left\| \left(\{\sqrt{P_j/m_j}\, \Re\{A_{u,j}^{(s)}\}\}_{j\neq b}, \{\sqrt{P_j/m_j}\, \Im\{A_{u,j}^{(s)}\}\}_{j\neq b}, \sigma_n \right) \right\|_2 \ \leq \ \frac{1}{\sqrt{\tau}} \sqrt{\frac{P_b}{m_b}}\, \Re\{A_{u,b}^{(s)}\}, \tag{6.5}$$

  with additional linear constraints $\Im\{A_{u,b}^{(s)}\} = 0$ and $\Re\{A_{u,b}^{(s)}\} \geq \rho$, where $\rho > 0$ is a tiny guard constant to avoid degenerate solutions.

- The expressions $A_{u,j}^{(s)}$ are *affine* in the binaries $\{z_{rj}\}$ because the contributions $\alpha_{u,j,r}^{(s)}$ are precomputed from the fixed beams. This allows us to embed $\Re\{A_{u,j}^{(s)}\}$ and $\Im\{A_{u,j}^{(s)}\}$ as linear functions of z in the SOC constraints.

- The max operation in (6.4) is handled via standard linearisation tricks (e.g., auxiliary integer variables for m_j and constraints $m_j \geq 1$, $m_j \geq \sum_r z_{rj}$).

The resulting problem is a mixed–integer second–order cone program (MISOCP) with

- a small set of continuous variables (mainly τ and some slack variables hidden in the SOC representation),
- $B \times R$ binary variables $\{z_{rj}\}$,
- one SOC constraint per user and per snapshot.

6.3 Algorithmic Realization in the Simulations

In the simulations the fixed–beamforming formulation is implemented in a few conceptually simple steps which are repeated for each evaluation episode.

Step 1: Environment and Channel Generation

For each episode we generate a geometric environment with fixed positions of BSs, RISs, and UEs, and draw NN microscopic fading realizations. This provides

- large–scale quantities such as pathloss and LOS/NLOS flags,
- instantaneous BS–UE, BS–RIS, and RIS–UE channel matrices for all NN snapshots,
- a fixed association $b(u)$ that assigns each user to a serving BS.

Step 2: Construction of Fixed Beamforming Vectors

In the fixed–beamforming regime each BS designs its transmit beams once, based only on the direct BS–UE channels, and then keeps them unchanged throughout the RIS optimization:

- For each BS we form an MRT–style beam that points towards the effective channel of the users served by that BS.
- The same set of beams is reused for all fading snapshots and for all RIS configurations.

Thus the beamforming vectors are treated as *given parameters* in the subsequent RIS assignment problem.

Step 3: Effective Channels Under Fixed Beams

Given the channels and the fixed beams, we compute for each user u, BS j, RIS r, and snapshot s an effective scalar channel coefficient

$$\alpha_{u,j,r}^{(s)} = \text{(direct contribution of BS } j \text{ to } u) + \text{(contribution via RIS } r \text{ under the chosen beam)}.$$

Two levels of description are used:

- A *sample–average* coefficient $\bar{\alpha}_{u,j,r} = \frac{1}{NN} \sum_{s=1}^{NN} \alpha_{u,j,r}^{(s)}$, which enters the optimization problem as a deterministic approximation.
- The full set $\{\alpha_{u,j,r}^{(s)}\}_{s=1}^{NN}$, which is kept for evaluating the true ergodic rates after an RIS assignment has been found.

Step 4: MISOCP for RIS Assignment

With fixed beams and averaged coefficients $\{\bar{\alpha}_{u,j,r}\}$, the RIS assignment is obtained by solving a max–min MISOCP:

- The variables are the binary RIS–to–BS indicators $z_{rj} \in \{0, 1\}$ and a scalar t representing the inverse of the worst–user SINR.
- For each user, the useful signal and interference terms are expressed in terms of $A_{u,j} = \sum_r z_{rj}\, \bar{\alpha}_{u,j,r}$, and the SINR constraints are written as second–order cone (SOC) inequalities.
- Additional linear constraints enforce that each RIS is assigned to at most one BS and that every BS controls at least one RIS.

A conic mixed–integer solver then maximizes t and returns a binary assignment matrix $Z^\star$ together with the corresponding max–min SINR level. Because the beams are fixed and the conic constraints are convex, global optimality with respect to Z is certified for each episode.

Step 5: Performance Evaluation with NN Snapshots

Finally, we evaluate the true performance of the optimized assignment $Z^\star$ using the full instantaneous coefficients $\alpha_{u,j,r}^{(s)}$:

- For each snapshot s we form the instantaneous SINRs according to the fixed–beam expression (with the RIS terms switched on and off according to $Z^\star$).
- The per–user ergodic rates $\bar{R}_u(Z^\star)$ are obtained by averaging $\log_2(1 + SINR_u^{(s)})$ over the NN snapshots.
- From these rates we compute per–BS min–rates, the global min–rate, and empirical CDFs of user rates, both per BS and aggregated across all users.

For comparison, the same evaluation is carried out for a *no–RIS* baseline, where all RIS paths are disabled but the beamforming vectors and channel realizations remain unchanged. This isolates the gain due purely to RIS deployment and assignment under fixed beamforming.

6.4 Implications and Limitations

6.4.1 Implications

The fixed–beamforming RIS optimization has several noteworthy properties:

- **Practical separation of time scales:** Beamforming and RIS assignment are decoupled: beams can be designed on a slow time scale from large–scale CSI or codebook constraints, while RIS assignments are adapted on a faster geometric time scale without re–solving beamforming SOCPs.
- **Global optimality for the reduced problem:** For fixed beams, the MISOCP is convex in the continuous variables and linear in the binaries. Modern solvers can certify a globally optimal $Z^\star$ for the given beams, in contrast to the alternating scheme of Chapter 5, which only provides local guarantees.
- **Fairness across BSs:** The equal power–split model combined with max–min SINR and a global MISOCP tends to produce RIS assignments that balance the per–BS min–rates. In the numerical results, the per–BS ECDFs of user rates become much closer than for naive or geometry–only assignments.
- **Robustness to microscopic fading:** Since the optimization uses sample–averaged coefficients and is evaluated over many snapshots, the resulting RIS assignment is robust to small–scale fading fluctuations and primarily depends on the large–scale geometry.

6.4.2 Limitations

On the other hand, fixing the beamforming vectors introduces limitations:

- **Loss of joint degrees of freedom:** The BS beams cannot adapt to the RIS assignment. In the joint optimization, the SOCP can steer beams to exploit constructive interference from the selected RISs. With fixed beams, some of this potential gain is lost, especially when the optimal RIS assignment would steer the BS beams away from their precomputed directions.

- **Dependence on beam design:** Performance is sensitive to how the fixed beams are chosen. MRT based on direct channels is simple and effective in our scenario, but different geometries or interference patterns might require ZF, regularized ZF, or other robust beam designs. The RIS assignment can only compensate to some extent for a poor beam choice.

- **MISOCP complexity remains nontrivial:** Although the number of continuous variables is reduced, we still face a mixed–integer conic problem with $B \times R$ binary variables and one SOC per user and snapshot. For larger values of B and R, solver runtimes remain significant.

- **Approximate treatment of m_j:** The dependence of the per–beam power P_j/m_j on the integer counts m_j forces additional linearisation. While this is handled exactly in the MISOCP, it complicates the model and may limit scalability.

6.5 Numerical Results

We now illustrate the behaviour of the fixed–beamforming RIS optimization using Monte–Carlo simulations.

Scenario

We consider a two–cell system with $B = 2$ BSs, $R = 10$ RIS panels and $U = 20$ users. Each BS has $M_{\mathrm{BS}} = 50$ antennas and each RIS has $M_{\mathrm{RIS}} = 250$ reflecting elements. The BS transmit power is $P_j = 20\,\mathrm{dBm}$, and the thermal noise level and noise figure are the same as in the previous chapters. For each configuration we generate $NN = 20$ small–scale fading snapshots and 20 independent geometry realisations.

All–User ECDFs

Figure 6.1 shows the ECDF of the per–user ergodic rate when all users from all episodes are pooled together. The solid curve corresponds to the fixed–beamforming RIS design, while the dashed curve is the baseline without RISs.

The RIS–assisted curve is moderately shifted to the right over a large part of the distribution. Around the median, the fixed–beamforming design increases typical user rates by roughly 0.5 bits/s/Hz compared to the no–RIS system, and the upper tail reaches slightly higher maximum rates. At the same time the improvement is less pronounced than in the joint instantaneous designs of Chapter 5: the two curves still overlap substantially, and the gains are clearly smaller than those obtained with joint beamforming and RIS allocation.

Both curves exhibit a non–negligible left tail, with a small fraction of users experiencing ergodic rates close to 1 bits/s/Hz or less. In this respect the fixed–beamforming scheme behaves similarly to the joint instantaneous designs and very differently from the purely large–scale MISOCP benchmark of Chapter 4, whose ECDFs were effectively clipped away from zero by the large–scale max–min formulation.

Per–BS ECDFs

To reveal how the benefits are distributed between the two cells, Fig. 6.2 shows ECDFs of the per–user ergodic rates when users are grouped according to their serving BS. Solid lines correspond to the fixed–beamforming RIS design, dashed lines to the no–RIS baseline. Blue and orange curves refer to BS 0, and green and red curves to BS 1.

The per–BS curves show that the fixed–beamforming RIS design does not benefit both cells equally:

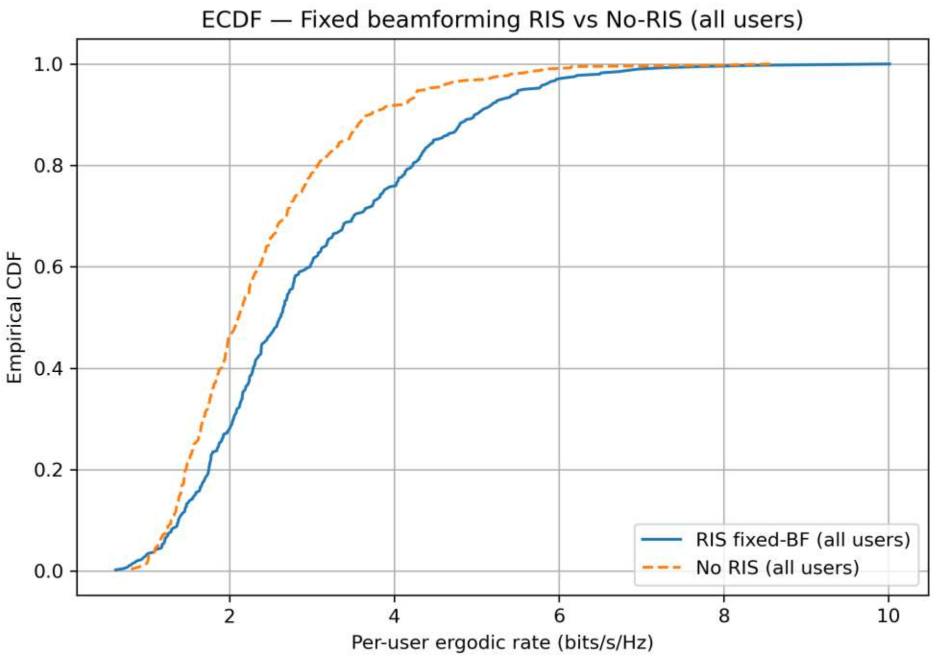

Figure 6.1: All–users ECDF of the per–user ergodic rate for the fixed–beamforming RIS optimization and the no–RIS baseline.

- For BS 1 (green vs. red), the RIS–assisted curve is clearly to the right of the baseline. Users of BS 1 see higher rates across most quantiles, with noticeable gains for both typical and high–rate users.
- For BS 0 (blue vs. orange), the situation is reversed: the RIS–assisted curve lies to the *left* of the no–RIS curve over a large part of the distribution. In other words, the fixed–beamforming RIS assignment improves BS 1 at the cost of BS 0, and some users of BS 0 are better off in the no–RIS system.

This asymmetry is consistent with the underlying geometry: in the optimised assignment more RIS panels tend to support the BS that benefits most from them (here BS 1), and with fixed beams there is no mechanism to jointly re–shape the transmit directions to compensate for the disadvantage of the other BS. In contrast, the quota–based joint optimization of Chapter 5 enforced explicit per–BS RIS quotas and optimized the beams accordingly, which led to more balanced per–BS ECDFs and simultaneous gains for both cells.

Summary of Fixed–Beamforming Performance

Taken together, the fixed–beamforming results can be summarised as follows:

- The RIS–assisted fixed–beamforming design yields a moderate improvement of the overall all–user ECDF compared to the no–RIS baseline, but the gains are smaller and less uniform than those observed with the joint instantaneous schemes.
- The improvement is not symmetric across BSs: in the considered geometry, BS 1 benefits significantly from RIS optimisation, while BS 0 experiences a degradation relative to its no–RIS performance. Thus, fixed–beamforming RIS optimisation tends to favour the geometrically advantaged cell rather than enforcing strict inter–cell fairness.
- Deep left tails remain present, as in the joint instantaneous designs, indicating that a small fraction of users still suffers from very low ergodic rates even after RIS assignment.

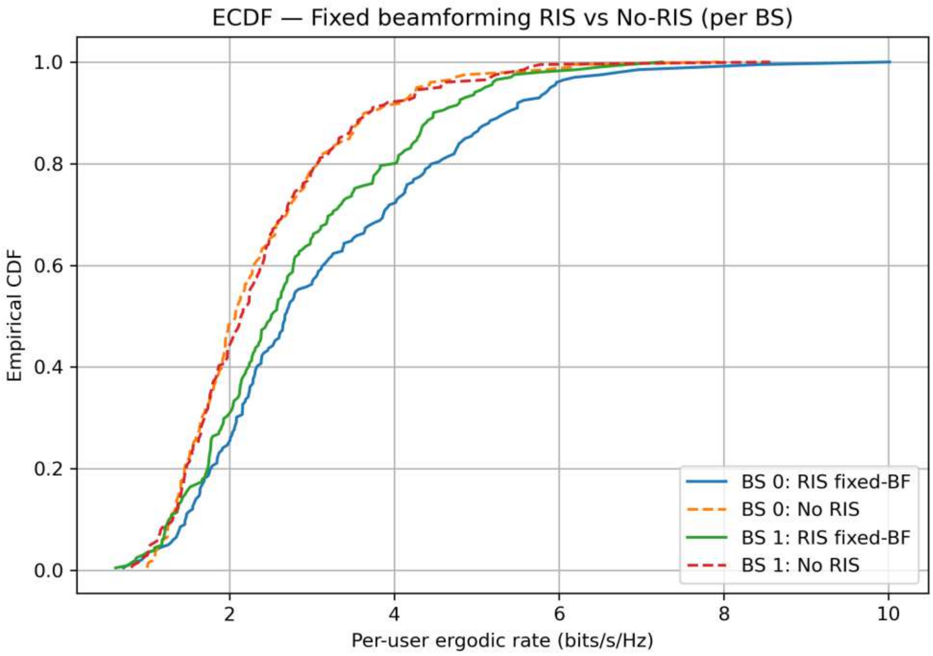

Figure 6.2: Per–BS ECDFs of the per–user ergodic rate for the fixed–beamforming RIS optimization and the no–RIS baseline. Blue / orange: users of BS 0 with / without RIS. Green / red: users of BS 1 with / without RIS.

Overall, the fixed–beamforming MISOCP lies between the purely large–scale optimisation of Chapter 4 and the fully joint designs of Chapter 5. It leverages the detailed geometric channel model and provides globally optimal RIS assignments for a given set of beams, but it cannot match the magnitude and the cell–wise balance of the gains achieved when beamforming and RIS allocation are optimised jointly.

Chapter 7

Conclusion and Outlook

7.1 Conclusion

This thesis has examined how RISs can be used to improve fairness and spectral efficiency in a small multi–cell downlink. Across the different formulations, a consistent picture appears: appropriately placed and optimised RISs can lift a large fraction of users to higher rates, but geometry and microscopic fading still create deep left tails in the rate distributions that cannot be fully removed by optimisation alone.

The comparison of the three design levels is instructive. The large–scale MISOCP benchmark shows the optimistic potential of RIS deployment when only slow fading is modelled and global optimality is available. The instantaneous joint designs demonstrate that, under realistic channels, RISs still provide substantial gains but with residual near–outage users and a clear tradeoff between performance, fairness and algorithmic complexity. The fixed–beamforming MISOCP sits in between: it delivers optimal RIS assignments for given beams and moderate gains, but these gains can be unevenly distributed between cells when beams are not jointly optimised.

7.2 Outlook

Several extensions follow from this work:

- **Larger and denser networks:** applying the proposed methods to more BSs, more RISs and denser user deployments will require scalable approximations or distributed algorithms in place of monolithic MISOCPs.
- **Imperfect CSI and robustness:** all designs assumed accurate channel state information; incorporating estimation errors or delayed feedback via robust or chance–constrained formulations is an important next step.
- **Dynamic operation and mobility:** the current optimisation is quasi–static. Multi–time–scale strategies in which RIS assignments follow long–term statistics while beams track fast fading would better reflect mobile scenarios.
- **Learning–aided control:** data–driven methods, for example predicting good quota vectors or RIS patterns from large–scale features, could approximate the optimisation results at much lower online complexity.

Overall, the thesis confirms that RISs are a powerful additional degree of freedom for future wireless systems, but also shows that their benefits depend critically on realistic modelling and on carefully designed, fairness–aware optimisation algorithms.

Bibliography

[1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2020.

[2] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, 2020.

[3] E. Björnson *et al.*, "Reconfigurable intelligent surfaces: A signal processing perspective," *IEEE Signal Processing Magazine*, 2021.

[4] T. Sharma *et al.*, "Reconfigurable intelligent surfaces for 5g and beyond wireless communications: A comprehensive survey," *Energies*, 2021.

[5] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE GLOBECOM*, 2018.

[6] X. Yu, D. Xu, and R. Schober, "Miso wireless communication systems via intelligent reflecting surfaces," in *Proc. IEEE/CIC ICCC*, 2019.

[7] R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Digital Equipment Corporation, Tech. Rep. DEC-TR-301, 1984.

[8] A. Coluccia *et al.*, "On the optimality of max–min fairness in resource allocation," *Annals of Telecommunications*, 2012.

[9] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* SIAM, 2001.

[10] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and Its Applications*, vol. 284, pp. 193–228, 1998.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge University Press, 2004.

[12] Y. Xu, "Resource allocation and beamforming design for max–min fair irs-aided swipt networks," Ph.D. dissertation, Univ. of British Columbia, 2022.

[13] F. Mehmeti *et al.*, "Max-min fair resource allocation in sd-ran," in *Proc. ACM CoNEXT*, 2022.

[14] E. Basar, M. D. Renzo, J. de Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[15] M. D. Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[16] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Towards smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020.

[17] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[18] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[19] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[20] H. Xie, J. Xu, and Y.-F. Liu, "Max–min fairness in IRS-aided multi-cell MISO systems with joint transmit and reflective beamforming," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1379–1393, 2021.

[21] S. Schwarz, "Gambling on reconfigurable intelligent surfaces," *IEEE Communications Letters*, vol. 28, no. 4, pp. 957–961, Apr. 2024.

[22] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, 2020.

[23] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1809–1822, 2020.

[24] Y. Chen, N. Huang, D. Li, W. Wu, and L. Yang, "Reconfigurable intelligent surface assisted D2D networks: Power and discrete phase shift design," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2020, pp. 1–6.

[25] S. Sun, M. Fu, Y. Shi, and Y. Zhou, "Towards reconfigurable intelligent surfaces powered green wireless networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[26] S. Diamond and S. Boyd, "CVXPY: A python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[27] E. D. Andersen and K. D. Andersen, "The MOSEK interior point optimizer for linear programming: An implementation of the homogeneous algorithm," in *High Performance Optimization*, H. Frenk *et al.*, Eds. Boston, MA, USA: Kluwer, 2000, pp. 197–232.

[28] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization.* Belmont, MA, USA: Athena Scientific, 1997.

[29] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, no. 1–3, pp. 193–228, 1998.