A SINGLE-RF ARCHITECTURE FOR MULTIUSER MASSIVE MIMO VIA REFLECTING SURFACES

Ali Bereyhi⋆, Vahid Jamali⋆, Ralf R. Müller⋆, Antonia M. Tulino†, Georg Fischer⋆, and Robert Schober⋆

⋆Friedrich-Alexander Universität Erlangen-Nürnberg
†Nokia Bell Labs, and Universita degli Studi di Napoli Federico II

ABSTRACT

In this work, we propose a new single-RF MIMO architecture which enjoys high scalability and energy-efficiency. The transmitter in this proposal consists of a single RF illuminator radiating towards a reflecting surface. Each element on the reflecting surface re-transmits its received signal after applying a phase-shift, such that a desired beamforming pattern is obtained. For this architecture, the problem of beamforming is interpreted as linear regression and a solution is derived via the method of least-squares. Using this formulation, a fast iterative algorithm for tuning of the reflecting surface is developed. Numerical results demonstrate that the proposed architecture is fully compatible with current designs of reflecting surfaces.

Index Terms— Single-RF transmitter, reflecting surfaces, least-squares, massive MIMO.

1. INTRODUCTION

A single-radio frequency (RF) transmitter consists of a monotone signal generator oscillating at the carrier frequency and a network of tunable elements [1–4]. These elements shape various output signals, depending on the value of their tunable factors. Using this architecture, downlink precoding is addressed completely in the RF stage. Due to their high cost-efficiency, these architectures have received a great deal of attention in the context of massive multiple-input multiple-output (MIMO) systems.

The conventional single-RF architectures often employ elements such as parasitic antenna arrays or load modulators [2,4]. Such settings require a physical module which connects the RF signal generator to the analogue network. This introduces undesired power-loss to the system resulting in performance degradation. An efficient way to address this issue is to implement the network of tunable elements via a reflecting surface. The connection between the RF chain (RFC) and the analog front-end in this case is carried out over the air which can lead to high power-efficiency and scalability [5].

Inspired by promising performance gains obtained via reflecting surfaces, several recent studies have proposed reflecting-surface-assisted architectures for MIMO transmission. In [6], reflecting surfaces have been employed to improve signal transmission in a point-to-point communication scenario. This idea was further extended to multiuser scenarios in [7], where a symbol-level precoding scheme is designed via intelligent reflecting surfaces for downlink transmission in multiuser MIMO settings. Some recent studies in this respect can be followed in [5–10] and the references therein.

1.1. Contributions

This paper proposes a novel reflecting-surface-assisted single-RF MIMO architecture. In this architecture, the modulation and beamforming are carried out at the reflecting surface via the phase-shifts which are applied by reflecting elements. Unlike available proposals, e.g., [7], this architecture does not impose any restriction on the system and considers a generic downlink transmission scenario. Invoking the least-squares (LS) method, a tractable algorithm for tuning the RFC and reflecting surface is developed. The investigations demonstrate that the proposed transmitter performs promisingly and is compatible with the available technology.

1.2. Notations

Throughout this paper, scalars, vectors, and matrices are represented by non-bold, bold lower case, and bold upper case letters, respectively. \( H^t \) and \( H^* \) denote the conjugate transpose and pseudo inverse of \( H \), respectively. \( I_K \) represents a \( K \times K \) identity matrix. \( \rho_{\text{max}}(H) \) returns the maximum squared singular value of \( H \). \( \|\cdot\|_F \) denote the Euclidean and Frobenius norm, respectively. \( \mathbb{R} \) is the real axis, and \( \mathbb{C} \) represents the complex plane. \( \omega \) denotes the phase of complex scalar \( s \). \( \mathbb{E}\{\cdot\} \) is mathematical expectation. For simplicity, \( \{1, \ldots, N\} \) is abbreviated by \( [N] \).

2. PROBLEM FORMULATION

Consider a multiuser setting in which a base station (BS) intends to transmit messages to \( K \) single-antenna users. The BS is equipped with a single transmit RFC and a reflecting surface with \( M \) antenna elements. It is assumed that the BS knows the channel state information (CSI) of the downlink channels from the surface to the users.

The BS pursues the following steps for transmission: First, it encodes the message of user \( k \) with a codeword of length \( N \), i.e., \( s_k(1), \ldots, s_k(N) \) for \( k \in [K] \). It then constructs transmit signal \( x(n) \in \mathbb{C}^M \) for \( n \in [N] \) from the CSI and vector of information symbols \( s(n) := [s_1(n), \ldots, s_K(n)]^T \) using a single-RF transmitter. Following the above steps, \( x(1), \ldots, x(N) \) are transmitted via the reflecting surface to the users in \( N \) distinct transmission time intervals. Let channel matrix \( H \in \mathbb{C}^{K \times M} \) contain the downlink channel coefficients. We assume that the channel experiences semi-static fading meaning that \( H \) remains unchanged during the transmission. Consequently, the vector of received signals in interval \( n \) reads

\[
y(n) = H x(n) + z(n),
\]

(1)
where \( z(n) \) is additive white Gaussian noise (AWGN) with zero mean and variance \( \sigma^2 \), i.e., sequence \( \{ z(n) : n \in [N] \} \) is independent and identically distributed (i.i.d.) and \( z(n) \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_K) \) for each \( n \in [N] \). \( y(n) \) further reads \( y(n) = [y_1(n), \ldots, y_K(n)]^\top \) with \( y_k(n) \) representing the signal received by user \( k \) in interval \( n \).

### 2.1. Transmitter Architecture

The detailed architecture of the transmitter is shown in Fig. 1. In this architecture, \( T \) is fed with a monotone signal which oscillates at the carrier frequency. The signal is first amplified and then radiated towards a reflecting surface whose center is located in distance \( R_d \) from the RFC. The elements on this surface receive attenuated and phase-shifted versions of the radiated signal. Each element reflects its received signal after applying a tunable phase-shift.

Let \( P \) be the power of the monotone signal and \( A(n) \) denote the amplification gain of the RFC in the \( n \)-th transmission time interval. For \( n \in [M] \), the signal transmitted via the \( m \)-th antenna element on the reflecting surface is given by

\[
x_m(n) = A(n) T_m \sqrt{P} \exp \{ j \omega_m + \beta_m(n) \}
\]

where \( T_m \) and \( \omega_m \) denote the attenuation and the phase-shift imposed on the radiated signal at the place of antenna element \( m \), due to propagation, respectively. \( \beta_m(n) \) is the phase-shift applied by the \( m \)-th antenna element before reflection. Depending on the properties of the surface, \( \beta_m(n) \) is chosen from \( \mathbb{B} \subseteq [-\pi, \pi] \).

Fig. 1 represents a special form of the architecture studied in [5][1] in which the digital base-band unit is omitted and all the signal processing tasks are shifted to the analog unit. Following derivations in [5], it is concluded that

\[
T_m = \frac{\lambda \sqrt{\zeta G(\theta_m, \phi_m)}}{4\pi r_m},
\]

\[
\omega_m = \frac{-2\pi r_m}{\lambda},
\]

where \( \lambda \) is the wavelength, \( \zeta \) denotes the power efficiency of the antenna elements, and \((r_m, \theta_m, \phi_m)\) is the spherical coordinate of the \( m \)-th element on the surface when the origin is set at the RFC. \( G(\theta, \phi) \) further represents the radiation pattern of the RFC with \( \theta \) and \( \phi \) being the elevation and azimuth angle, respectively.

### 3. LS-BASED SINGLE-RF TRANSMITTER

In the proposed reflecting-surface-assisted architecture, the signal pre-processing tasks are carried out via the tunable phase-shifts and amplitude variations at the RFC. Hence, the phase-shifters and amplification gain are updated in each transmission time interval based on the given information symbols. As a result, the design reduces to the problem of finding \( \beta_m(n) \) for \( m \in [M] \) and \( A(n) \) in terms of \( s(n) \) and \( H \), such that a desired performance metric is optimized. We address this problem via the method of LS.

#### 3.1. Transmission with Minimum Received MSE

The ultimate goal of transmitter design is to construct \( x(n) \), such that user \( k \) can recover its information symbol \( s_k(n) \) with minimal post-processing from the received symbol \( y_k(n) \). To quantify this requirement, let us define the received mean squared error (MSE) in this setting as

\[
\text{MSE}_R(n) := \sum_{k=1}^{K} \mathbb{E} \left\{ |s_k(n) - G_k y_k(n)|^2 \right\}
\]

for some \( G_k \in \mathbb{C} \). The received MSE determines the sum of MSEs at the user terminals when user \( k \) takes \( y_k(n) \) as the soft estimate of its information symbols after amplification and performing a phase-shift. In \( G \), \( G_k \) models linear post-processing operations on \( y_k(n) \), i.e., amplification and phase-shift, and is assumed to be independent of \( n \). The expectation is further calculated conditioned on \( s(n) \) and \( H \), since the information symbols and the CSI are known at the transmitter side.

Using the independence of AWGN, it is straightforward to show that the received MSE reduces to

\[
\text{MSE}_R(n) := \| s(n) - G \beta \|_2^2 + \| G \|_F^2 \sigma^2
\]

with \( G = \text{diag} \{ G_1, \ldots, G_K \} \). From the transmitter architecture, we further have

\[
x(n) = A(n) \sqrt{P} T w(n),
\]

where \( T \) and \( w(n) \) are given by

\[
T = \text{diag} \{ T_1 \exp \{ j \omega_1 \}, \ldots, T_M \exp \{ j \omega_M \} \}, \quad (7a)
\]

\[
w(n) = [\exp \{ j \beta_1(n) \}, \ldots, \exp \{ j \beta_M(n) \}]^\top. \quad (7b)
\]

By substituting \( x(n) \) in \( 5 \), the received MSE reads

\[
\text{MSE}_R(n) := \| s(n) - A(n) \sqrt{P} G H T w(n) \|^2 + \| G \|^2 \sigma^2. \quad (8)
\]

Following the intuitive discussion at the beginning of the section, the design goal can be formulated via an optimization problem which minimizes the received MSE. This means that the optimal choice for the phase-shifts and amplification gain in transmission time interval \( n \) are given by

\[
[w^*(n), A^*(n)] = \arg\min_{w \in \mathcal{W}, A \in \mathbb{R}} \| s(n) - A \sqrt{P} G H T w \|^2, \quad (9)
\]

where \( \mathcal{W} \) is defined as

\[
\mathcal{W} = \{ w = \exp \{ j \beta \} : \beta \in \mathbb{B} \}. \quad (10)
\]

In other words, assuming that user \( k \) multiplies its received signal with \( G_k \) to calculate a soft estimate of its information symbol, the
sum of MSEs at the user terminals in time interval $n$ is minimized when we set $\beta_n = \angle w_{i_n}$ and $A(\cdot) = A^*(\cdot)$.

The tuning scheme in (9) can be observed as linear regression in which the regression coefficients in $w^*(\cdot)$ is determined via the LS method from regressor matrix $A^*(\cdot)$ $\sqrt{P_{\text{GHT}}}$ and the regressands in $s(\cdot)$. Such a problem in its basic form is trivially solved via zero-forcing. However, the fact that the regression coefficients are restricted to lie on the unit circle makes the problem intractable. To address this issue, we develop an iterative algorithm by modifying the gradient decent algorithm.

**Remark 1**: In general, $w^*(\cdot)$ and $A^*(\cdot)$ are functions of the user processing gains. Considering this dependency, one can write MSE at $n$ for $n \in [N]$ are functions of the user processing gains. Following this, the explicit form of $f_n(\cdot)$ is not known, this optimization is intractable to be addressed. One can however employ an alternating optimization approach to tune $G$ suboptimally.

### 3.2. Developing an Iterative Algorithm

When $B = [-\pi, \pi]$ and $A = 1$, the optimization problem in (9) reduces to the so-called unit-modulus LS; see [12,13] and the references therein. The benchmark approach for solving unit-modulus LS is to apply semidefinite relaxation [14]. The computational complexity in this case grows quadratically in $M$ which is relatively large for the given application. In [14], a fast iterative algorithm for unit-modulus LS based on gradient projection was proposed. The algorithm employs gradient decent approach to find a local minima. To keep the minima on the unit circle, it projects the updated point in each iteration on the unit circle. It was shown in [14] that this algorithm globally converges to a Karush-Kuhn-Tucker (KKT) point of the unit-modulus LS problem. The algorithm was further extended to auto-scaled unit-modulus LS, i.e., when the regressands are scaled with a tunable variable.

The key difference of (9) to an auto-scaled unit-modulus LS problem is that the support of regression coefficients, i.e., $\mathcal{W}$, is not necessarily the unit circle. Following the gradient projection approach in [14], this issue can be addressed by straightforward modifications. The result is shown in Algorithm 1. In this algorithm, $\text{Quant}_{\mathcal{W}}(\cdot)$ denotes a uniform phase quantizer which for $u \in \mathbb{C}^M$ is defined as

$$[\text{Quant}_{\mathcal{W}}(u)]_m = \arg\min_{w \in \mathcal{W}} |\angle w - \angle u_m|.$$  

By setting $\mathcal{W}$ to be the unit circle in the complex plane, this algorithm recovers the gradient projection algorithm given in (14).

### 4. NUMERICAL EXPERIMENTS

We now investigate performance of the proposed architecture through some numerical experiments. To this end, we consider the following scenario: The RFC is fed by an oscillator with power $P = 1$ and wavelength $\lambda = 8$ mm. The transmit antenna at the RFC has a horizontally omnidirectional radiation pattern whose beamwidth in the vertical plane is $120^\circ$. A reflecting surface of size $\sqrt{M} \lambda \times \sqrt{M} \lambda$ with $M$ elements is located in distance $R_d = \lambda \sqrt{M} / \pi$ from the

**Algorithm 1 Gradient Projection Algorithm**

**Initiate** Set initial step-size $\psi_0 \in (0, 1)$, and $\hat{H} = \sqrt{P_{\text{GHT}}}$

$$w_0(n) = \text{Quant}_{\mathcal{W}} \left( \frac{\hat{H}^* s(n)}{||\hat{H}s(n)||} \right)$$

Choose $E_{\text{Th}}$ and maximum number of iterations $T_{\text{max}}$. While $||w_{t+1} - w_t||^2 \geq E_{\text{Th}}$ and $t \leq T_{\text{max}}$

- Update the parameters of the RFC as
  $$A_{t+1}(n) = \frac{E \{ w^*_t(n) \hat{H}^* s(n) \}}{||\hat{H}w_t||^2}$$

- Update the step-size as
  $$\psi_{t+1} = \frac{\psi_0 A_{t+1}(n)}{\rho_{\text{max}}(A_{t+1}(n) \hat{H})}$$

- Update the parameters of the surface as
  $$v_{t+1}(n) = \hat{H}^* \left(s(n) - A_{t+1}(n) \hat{H}w_t(n)\right)$$
  $$w_{t+1}(n) = \text{Quant}_{\mathcal{W}}(w_t(n) + \psi_{t+1}v_{t+1}(n))$$

- Update $t \leftarrow t + 1$. 

**Output**: $A(n) = A_T(n)$ and $\beta_{m}(n) = \angle w_{T,m}(n)$ for $m \in [M]$, where $T$ is index of the last iteration.

RFC. The power efficiency of the elements is set to $\log \zeta = 0$ dB and the phase-shifts are quantized with $B$ bits, i.e.,

$$B = \left\{ -\pi + \frac{i \pi}{2^{B-1}} \text{ for } i = 0, \ldots, 2^{B-1} - 1 \right\}.$$  

(13)

The architecture is used for downlink transmission to $K$ users in a single cell. The users are assumed to be uniformly distributed in the network. We consider the standard Rayleigh model for the fading process and model the shadowing effects via the log-normal distribution. The entries of $\hat{H}$ are hence generated as

$$[\hat{H}]_{k,m} = \frac{\alpha_k}{\tilde{r}_k} h_{k,m},$$  

(14)

where

- $h_{k,m}$ for $k \in [K]$ and $m \in [M]$ are i.i.d. complex Gaussian random variables with zero mean and unit variance,

- $\alpha_k$ for $k \in [K]$ are zero-mean log-normal variables with standard deviation $\log \sigma_{\text{Shadow}} = 5$ dB,

- $\nu = 3.2$ is the attenuation exponent, and

- $\tilde{r}_k = \tau_k / r_0$ with $r_0$ being the distance of user $k$ to the BS and $r_0 = 100$ m denoting the minimal distance in the network.

Throughout numerical simulations, we consider encoded messages of length $N = 100$. The information symbols in each transmission interval are assumed to be i.i.d. zero-mean and unit-variance complex Gaussian random variables, i.e., $s(n) \sim \mathcal{CN}(0, I_K)$. The
post-processing gains at user terminals are further set such that path-loss and shadowing effects are compensated. Note that based on the discussion in Remark 1, this choice of $G$ is in general suboptimal.

To evaluate the performance of this setting, we define the per-user distortion as

$$D = \frac{1}{K} \left( \frac{1}{N} \sum_{n=1}^{N} \| s(n) - G H x(n) \|^{2} \right)$$

for a given realization of the channel. For this setting, the average transmit power per time interval is further calculated as

$$P_{\text{Out}} = \frac{1}{N} \sum_{n=1}^{N} |A(n)|^{2} P.$$  \hspace{1cm} (16)

Noting that the RFC transmits with power $A^{2}(n)P$ in interval $n$, we define the peak-to-average power ratio (PAPR) as

$$\text{PAPR} = \max_{n \in [1,N]} |A(n)|^{2} P_{\text{Out}}.$$  \hspace{1cm} (17)

These parameters are averaged over multiple channel realizations.

Fig. 2 shows the per-user distortion against the number of user for various values of $M$. Here, the phase-shifts are quantized with $B = 4$ bits. From the figure, it is observed that $D$ increases, as $K$ grows. This indicates that the distortion scales reversely with the number of antenna elements per user terminal. The variations in $D$ with respect to the surface size further agrees with this conclusion.

To compare the performance with a benchmark, we further plot the curve for a fully digital matched filtering considering the case with $M = 225$ antennas. In this scheme, a fully digital transmitter, i.e., a transmitter with $M = 225$ distinct RFCs, uses matched filtering for downlink transmission. In each transmit interval, the digital pre-coder is scaled, such that its transmit power is exactly the same as the one radiated in the proposed single-RF architecture. User $k$ further chose $G_{k}$, such that the impact of path-loss, shadowing and power scaling on its information symbol is compensated. From the figure, it is observed that the proposed architecture outperforms this fully digital scheme with even with $M = 121$ transmit antennas. Such an observation indicates the effectiveness of the proposed architecture with respect to the benchmark.

Although the growth in $K$ increases the per-user distortion, it can benefit in terms of PAPR of the transmit RFC. This is demonstrated in Fig. 3, where the transmit PAPR is plotted for the same settings against $K$. This behavior comes from the fact that as the number of users increases, the transmit signal couples higher number of information symbols which by the law of large numbers reduces the time variations.

Fig. 4 illustrates the impact of phase-shift quantization. Here, distortion is plotted against $K$ for various choices of $B$ when $M = 64$. As the figure demonstrates, by increasing the resolution of the phase-shifters, the performance improves. Comparing the results with the asymptotic case of $B \to \infty$, i.e., when $B = [-\pi, \pi]$, it is observed that the degradation due to the phase-shift quantization is not significant. This indicates that the proposed architecture is compatible with the available designs for reflecting surfaces [16,17].

5. CONCLUSIONS

A reflecting-surface-assisted single-RF MIMO architecture has been proposed. The transmitter consists of an illuminator and a passive re-
reflecting surface. This makes the architecture both cost-efficient and scalable. To tune the phase-shifters at the surface, a fast algorithm based on the method of LS has been developed. Simulations have demonstrated that desired per-user distortions are achievable via this architecture using available technologies of reflecting surfaces. The study in this work has considered a basic form of the proposed architecture and requires further investigations and extensions in various respects. The work in this direction is currently ongoing.

6. REFERENCES

[1] Antonis Kalis, Athanasios G Kanatas, and Constantinos B Papadias, “A novel approach to MIMO transmission using a single RF front end,” IEEE Journal on Selected Areas in Communications, vol. 26, no. 6, pp. 972–980, August 2008.

[2] Abbas Mohammadi and Fadhel M Ghannouchi, “Single RF front-end MIMO transceivers,” RF Transceiver Design for MIMO Wireless Communications, pp. 265–288, 2012.

[3] Mohammad A Sedaghat, Ralf R Mueller, and Georg Fischer, “A novel single-RF transmitter for massive MIMO,” Proc. 18th International ITG Workshop on Smart Antennas (WSA), pp. 1–8, March 2014, Erlangen, Germany.

[4] Mohammad A Sedaghat, Vlasis I Barousis, Ralf R Müller, and Constantinos B Papadias, “Load modulated arrays: a low-complexity antenna,” IEEE Communications Magazine, vol. 54, no. 3, pp. 46–52, March 2016.

[5] Vahid Jamali, Antonia M. Tulino, Georg Fischer, Ralf R. Müller, and Robert Schobert, “Scalable and energy-efficient millimeter massive MIMO architectures: Reflect-array and transmit-array antennas,” Proc. IEEE International Conference on Communications (ICC), May 2019, Shanghai, China.

[6] Ertugrul Basar, “Transmission through large intelligent surfaces: A new frontier in wireless communications,” arXiv preprint arXiv:1902.08463, 2019.

[7] Rang Liu, Hongyu Li, Ming Li, and Qian Liu, “Symbol-level precoding design for intelligent reflecting surface assisted multi-user MIMO systems,” arXiv preprint arXiv:1909.01015, 2019.

[8] Ahmed H Abdelrahman, Fan Yang, Atef Z Elsherbeni, and Payam Nayeri, Analysis and Design of Transmittarray Antennas, Morgan & Claypool Publishers, USA, 2017.

[9] Qingqing Wu and Rui 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, November 2019.

[10] Chongwen Huang, Alessio Zappone, George C Alexandropoulos, Mérav Debbah, and Chau Yuen, “Reconfigurable intelligent surfaces for energy efficiency in wireless communication,” IEEE Transactions on Wireless Communications, vol. 18, no. 8, pp. 4157–4170, August 2019.

[11] Ali Bereyhi, Vahid Jamali, Ralf R. Müller, Georg Fischer, Robert Schobert, and Antonia M. Tulino, “PAPR-limited precoding in massive MIMO systems with reflect- and transmit-array antennas,” Proc. Asilomar Conference on Signals, Systems, and Computers, November 2010, Asilomar Grounds in Pacific Grove, CA, USA.