# Switch-Based Hybrid Analog/Digital Channel Estimation for mmWave Massive MIMO

Alec Poulin
Dept. of Electrical and Computer Eng.
McGill University, Montreal, Canada
alec.poulin@mail.mcgill.ca

Alireza Morsali
Dept. of Electrical and Computer Eng.
McGill University, Montreal, Canada
alireza.morsali@mail.mcgill.ca

Benoit Champagne
Dept. of Electrical and Computer Eng.
McGill University, Montreal, Canada
benoit.champagne@mcgill.ca

2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall) | 978-1-7281-9484-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/VTC2020-Fall49728.2020.9348809

***Abstract*—This paper addresses the problem of channel estimation using pilots in hybrid analog/digital massive multiple-input multiple-output (MIMO) systems for future millimetre wave (mmWave) communications. To further reduce system cost and implementation complexity, we consider an alternative architecture derived from RF switches as opposed to the phase shifters in the conventional literature. The channel estimation is formulated as a combinatorial optimization problem where the aim is to minimize the mean square error (MSE) between the real and estimated channels over a finite set of allowed values for the precoder switches. A genetic algorithm (GA) is developed for solving this problem and obtaining the MIMO channel estimates. Simulations show that the proposed scheme can estimate channels as accurately, if not more, as an existing solution using phase shifters.**

***Index Terms*—Massive MIMO; channel estimation; hybrid architecture; RF switches; genetic algorithm.**

## I. Introduction

Massive multiple-input multiple-output (MIMO) is one of the key enabling technologies advanced to meet the exacting target performance requirements of the fifth generation (5G) of wireless systems [1]. Due to the use of millimetre-wave (mmWave) signals, the antenna arrays of massive MIMO systems will be smaller and include more antenna elements than previous technologies [2]. However, using more antennas requires having more radio-frequency (RF) chains, which in turn is problematic because the latter are expensive and power-hungry [2], [3]. One way to cope with this constraint is to employ a hybrid architecture, wherein an analog beamformer is used to reduce the effective number of RF signals prior to the application of baseband processing [4], [5].

Different approaches have been explored recently in order to develop channel estimation algorithms that are specifically tailored for this architecture. A method exploiting the angular sparsity of the channel was presented in [4], where the focus is on the estimation of a point-to-point narrow-band flat-fading mmWave channel with both transceivers having a hybrid architecture. A similar procedure was studied in [6] for the case of a wideband frequency-selective fading channel. In [7], a minimum mean square error (MSE) formulation leads to efficient algorithms for estimating the vector channel between single-antenna users and a multi-antenna BS. Moreover, subspace tracking is used as the core idea in [8], while in [9], deep convolutional neural networks are proposed to estimate mmWave massive MIMO channels with different correlation properties.

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and InterDigital Canada.

The aforementioned works all assume that the phase shifters in the analog units are fast enough to change configuration between each pilot symbol, which may not be possible in practice. Other types of hardware can be used for this task however, like RF switches, which are cheaper, faster, and more power-efficient than analog phase shifters [10]. Four different switching network configurations along with a compressed-sensing-based channel estimation method are presented in [10]. Besides, one of the configurations from [10] is used in [11], where a hybrid precoder is designed for a mmWave MIMO system, while another configuration is used with a matrix completion algorithm in [12] to estimate mmWave massive MIMO channels.

This paper explores pilot-based channel estimation in mmWave hybrid analog/digital massive MIMO systems using RF switches instead of phase shifters. The channel estimation is formulated as a combinatorial optimization where the aim is to minimize the MSE between the real and estimated channels over a finite set of allowed values for the precoder matrices. A genetic algorithm (GA) is developed for designing the discrete precoders and estimating the MIMO channel using this simpler hardware, and its performance is compared to that of an existing algorithm using phase shifters. It is shown that in spite of their inherent simplicity, RF switches can be employed to estimate channels as accurately as phase shifters, and thus hold potential for a future class of hybrid massive MIMO transceivers.

The following notation is employed: $\mathbf{A}_{[n_1,\ldots,n_K]}$ is the matrix made of the $n_1^{th},\ldots,n_K^{th}$ columns of $\mathbf{A}$, while $\mathbf{A}_{[m:n]}$ is the matrix made of its $m^{th}$ to $n^{th}$ columns, inclusively. $\mathcal{CN}(\mathbf{m},\mathbf{R})$ is a complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{R}$, and $\mathcal{U}(a,b)$ is a random variable uniformly distributed on the interval $[a,b]$. Finally, $\mathbb{R}_+ = \{x \in \mathbb{R} \,|\, x \geq 0\}$.

## II. System Model and Problem Formulation

The present work investigates pilot-based channel estimation in a single-cell massive MIMO system for the case of uplink transmission from single-antenna mobile stations (MSs) to a multi-antenna BS. Under the time-division duplexing (TDD) mode of operation and provided the coherence time of the radio channels is large enough, knowledge of the uplink channels then allows the downlink channels to be easily computed [13].

### A. System Model

As shown in Fig. 1, the BS is equipped with $N$ antennas and $L$ RF chains. An analog combiner is located between the antennas and the RF chains, and the signals output from the

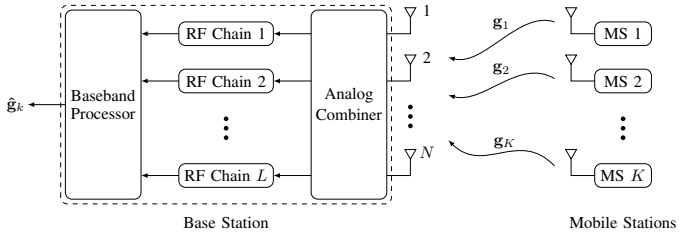



Fig. 1. Block diagram of the system model, with a hybrid mmWave massive MIMO base station and $K$ single-antenna mobile stations.

RF chains are fed into a digital baseband processor. The analog combiner can be implemented in different ways: if made of phase shifters, the narrow-band signals from the antennas are phase-shifted by different amounts, added together, and fed to the RF chains; if built from switches instead, the antenna signals can either go through different paths or be blocked, as will be explained in Section III-A. In both cases, the analog combiner is used to convert a signal vector with $N$ components into a lower-dimensional signal vector with $L$ components, where $L < N$. The lower-dimensional signal vector is then fed to the RF chains, where its components are down-converted and digitally sampled. The resulting digital signals are then fed to the baseband processor, which completes the channel estimation.

For the sake of simplicity, the pilots used in this work are assumed to be time-orthogonal: when one MS sends a pilot, all the other MSs remain mute. This design greatly simplifies the problem, allowing a single MS to be considered at a time, while providing the same performance as other pilot designs [7]. Hence, the user indices are not used in the rest of this paper. The uplink training process for channel estimation can be described as follows. The MS sends to the BS $T$ pilot symbols $\varphi_t \in \mathbb{C}$, $t \in \{1, \dots, T\}$, known beforehand and all satisfying $|\varphi_t| = 1$. At time $t$, the BS receives signal vector

$$\mathbf{r}_t = \sqrt{\rho}\mathbf{g}\varphi_t + \tilde{\mathbf{n}}_t, \tag{1}$$

where $\rho \in \mathbb{R}_+$ is the power of the pilot, $\mathbf{g} \in \mathbb{C}^{N\times 1}$ represents the uplink channel vector between the considered MS and the BS, and $\tilde{\mathbf{n}}_t \in \mathbb{C}^{N\times 1}$ is an additive white noise vector with zero mean, unit variance, and complex circular Gaussian distribution, i.e. $\tilde{\mathbf{n}}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. After being processed by the analog combiner and multiplied by the conjugate of the pilot symbol, the signal received at the baseband processor is

$$\mathbf{y}_t = \mathbf{F}_t(\sqrt{\rho}\mathbf{g}\varphi_t + \tilde{\mathbf{n}}_t)\bar{\varphi}_t = \mathbf{F}_t(\sqrt{\rho}\mathbf{g} + \mathbf{n}_t), \tag{2}$$

where $\mathbf{F}_t \in \mathbb{C}^{L\times N}$ represents the analog combiner at time $t$ and $\mathbf{n}_t = \bar{\varphi}_t\tilde{\mathbf{n}}_t \in \mathbb{C}^{N\times 1}$ is a white noise vector with the same statistics as $\tilde{\mathbf{n}}_t$, i.e. $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. The $T$ signals successively received at the baseband combiner can be combined together by defining the following vectors and matrices:

$$\mathbf{y}_\mathrm{c} = [\mathbf{y}_1^\mathrm{T}, \mathbf{y}_2^\mathrm{T}, \dots, \mathbf{y}_T^\mathrm{T}]^\mathrm{T}, \quad \mathbf{n}_\mathrm{c} = [\mathbf{n}_1^\mathrm{T}, \mathbf{n}_2^\mathrm{T}, \dots, \mathbf{n}_T^\mathrm{T}]^\mathrm{T},$$
$$\mathbf{F}_\mathrm{c} = [\mathbf{F}_1^\mathrm{T}, \mathbf{F}_2^\mathrm{T}, \dots, \mathbf{F}_T^\mathrm{T}]^\mathrm{T}, \ \mathbf{F}_\mathrm{d} = \mathrm{diag}(\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_T),$$

where $\mathbf{y}_\mathrm{c} \in \mathbb{C}^{LT\times 1}$, $\mathbf{n}_\mathrm{c} \in \mathbb{C}^{NT\times 1}$, $\mathbf{F}_\mathrm{c} \in \mathbb{C}^{LT\times N}$, and $\mathbf{F}_\mathrm{d} \in \mathbb{C}^{LT\times NT}$. It then follows from (2) that

$$\mathbf{y}_\mathrm{c} = \sqrt{\rho}\mathbf{F}_\mathrm{c}\mathbf{g} + \mathbf{F}_\mathrm{d}\mathbf{n}_\mathrm{c}. \tag{3}$$

We model the unknown channel $\mathbf{g}$ as a random vector with zero mean, i.e. $\mathbb{E}[\mathbf{g}] = \mathbf{0}$, and covariance matrix $\mathbf{R} = \mathbb{E}\big[\mathbf{g}\mathbf{g}^\mathrm{H}\big]$. Without loss of generality, $\mathbf{g}$ is normalized as $\mathbb{E}\big[\|\mathbf{g}\|^2\big] = \mathrm{tr}(\mathbf{R}) = N$. We assume that the channel vector $\mathbf{g}$ and the noise terms $\mathbf{n}_t$ are uncorrelated, i.e. $\mathbb{E}\big[\mathbf{g}\,\mathbf{n}_c^\mathrm{H}\big] = \mathbf{0}$. In our simulations (Section IV), the following mmWave channel model with limited scattering [4] is employed,

$$\mathbf{g} = \frac{1}{\sqrt{\sigma_\alpha^2 P}}\sum_{p=1}^{P}\alpha_p\mathbf{a}(\theta_p), \tag{4}$$

where $P$ is the number of propagation paths from the MS to the BS, $\alpha_p \sim \mathcal{CN}(0, \sigma_\alpha^2)$ and $\theta_p \sim \mathcal{U}(0, 2\pi)$ are respectively the complex gain and the angle of arrival (AoA) of the $p^{\text{th}}$ path, and $\mathbf{a}(\theta) \in \mathbb{C}^{N\times 1}$ is the response vector of the antenna array used at the BS in direction $\theta$. All the path gains $\alpha_p$ and AoAs $\theta_p$ are statistically independent from each other. While the channel estimation method developed herein can be applied to antenna arrays of arbitrary shape, only uniform linear arrays (ULAs) are considered for simplicity. For ULAs, the array response returns a column vector whose $n^{\text{th}}$ entry is expressed as $a_n(\theta) = e^{j(n-1)\frac{2\pi d}{\lambda}\sin\theta}$, where $\theta$ is measured from a line perpendicular to the array axis, $d$ is the distance between the antennas, and $\lambda$ is the signal wavelength. Using the array response of the ULA and the channel model defined in (4), it can be seen that matrix $\mathbf{R}$ has rank at most $P$, with entries given by

$$[\mathbf{R}]_{(m,n)} = \frac{1}{P}\sum_{p=1}^{P}e^{j(m-n)\frac{2\pi d}{\lambda}\sin\theta_p}. \tag{5}$$

### *B. Problem Formulation*

The goal of this paper is to estimate the channel vector $\mathbf{g}$ from the MS to the BS by producing an estimated channel $\hat{\mathbf{g}}$ that is as close as possible to the real channel in the minimum MSE sense. The problem is formulated as follows:

$$\begin{aligned}\underset{\mathbf{F}_1,\dots,\mathbf{F}_T,\mathbf{W}}{\text{minimize}} \ & \mathbb{E}\big[\|\mathbf{g} - \hat{\mathbf{g}}\|^2\big] \\ \text{subject to } & \hat{\mathbf{g}} = \mathbf{W}\mathbf{y}_\mathrm{c} \\ & \mathbf{y}_\mathrm{c} = \sqrt{\rho}\mathbf{F}_\mathrm{c}\mathbf{g} + \mathbf{F}_\mathrm{d}\mathbf{n}_\mathrm{c} \\ & \mathbf{F}_t \in \mathcal{F} \ \forall t \in \{1, \dots, T\},\end{aligned} \tag{6}$$

where $\mathbf{W} \in \mathbb{C}^{N\times TL}$ represents the baseband processor and $\mathcal{F}$ is the set of feasible matrices for the analog combiner, which depends on the exact architecture used at the BS. To minimize the MSE, it is possible to find an expression for the optimal $\mathbf{W}$ in terms of the $\mathbf{F}_t$, which is given by

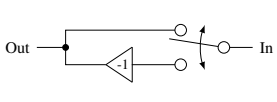
$$\mathbf{W}_\text{opt} = \sqrt{\rho}\mathbf{R}\mathbf{F}_\mathrm{c}^\mathrm{H}\left(\rho\mathbf{F}_\mathrm{c}\mathbf{R}\mathbf{F}_\mathrm{c}^\mathrm{H} + \mathbf{F}_\mathrm{d}\mathbf{F}_\mathrm{d}^\mathrm{H}\right)^{-1}. \tag{7}$$

Using this expression and defining $\mathcal{M}(\mathbf{F}) = \mathbb{E}\big[\|\mathbf{g} - \hat{\mathbf{g}}\|^2\big] = \mathbb{E}\big[\|\mathbf{g} - \mathbf{W}_\text{opt}\mathbf{y}_\mathrm{c}\|^2\big]$, where $\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_T]$, the objective function in (6) can be rewritten as

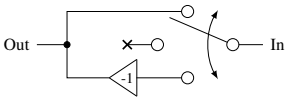
$$\mathcal{M}(\mathbf{F}) = \mathrm{tr}\Big(\mathbf{R} - \rho\mathbf{R}\mathbf{F}_\mathrm{c}^\mathrm{H}\left(\rho\mathbf{F}_\mathrm{c}\mathbf{R}\mathbf{F}_\mathrm{c}^\mathrm{H} + \mathbf{F}_\mathrm{d}\mathbf{F}_\mathrm{d}^\mathrm{H}\right)^{-1}\mathbf{F}_\mathrm{c}\mathbf{R}\Big). \tag{8}$$

### *C. Existing Solution with Phase Shifters*

The optimization problem (6) is addressed in [7] for a hybrid system with phase shifters as the analog components, that is, where the feasible set for the analog matrices is $\mathcal{F} = \left\{e^{j\theta} \,\middle|\, \theta \in [0, 2\pi[\right\}^{L\times N}$. The proposed approach begins by temporarily discarding the constant-magnitude constraint




(a) Switch configuration used for $\mathcal{B} = \{-1, 1\}$.




(b) Switch configuration used for $\mathcal{B} = \{0, -1, 1\}$.

Fig. 2. Basic switch configurations that can be used in an analog combiner to implement binary and ternary sets $\mathcal{B}$ of allowed values in the matrices $\mathbf{S}_t$.

imposed on the entries of $\mathbf{F}_t$, which is equivalent to replacing the constraint set $\mathcal{F}$ by $\mathbb{C}^{L\times N}$. Then, the optimal unconstrained combiners are computed under the assumption of known channel covariance matrix $\mathbf{R}$ using one of several proposed methods, including the so-called sequential optimization (SO) and alternating optimization (AO). Finally, these combiners are projected onto the original feasible set. That is, if the $(l,n)^{\text{th}}$ entry of the $t^{\text{th}}$ optimal unconstrained combiner is $[\mathbf{F}_t^{\text{opt}}]_{(l,n)} = r_{l,n}e^{j\theta_{l,n}}$, where $r_{l,n} \in \mathbb{R}_+$ and $\theta_{l,n} \in [0, 2\pi[$, then the entry at the same position in the projected combiner is $[\mathbf{F}_t^{\text{proj}}]_{(l,n)} = e^{j\theta_{l,n}}$.

In the case of phase shifters implemented with finite precision, the entries of the analog combiner matrices $\mathbf{F}_t$ must be quantized, which amounts to selecting the closest element from a finite subset of available points on the complex unit circle. Specifically, the quantized subset consists of $N_Q = 2^{N_B}$ points placed equidistantly along the circle, i.e. with angular separation of $\Delta\theta = 2\pi/N_Q$, and including the point $1 \in \mathbb{C}$, where $N_B$ is the number of quantization bits.

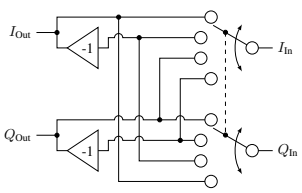
## III. Proposed Solution with Switching Networks

### A. Structural Elements

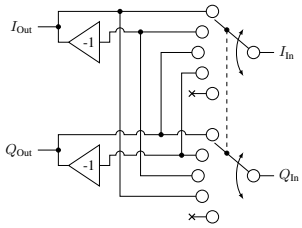
The solution proposed here to solve problem (6) uses analog combiners that are implemented with switches instead of phase shifters. The symbol $\mathbf{S}$ will thus be used instead of $\mathbf{F}$ for notational convenience. The set of allowed values for the entries of matrices $\mathbf{S}_t$ do not form a continuous set as it is the case for phase shifters. It is instead a discrete set $\mathcal{B}$, which can contain zero and points from the complex unit circle. The exact elements of $\mathcal{B}$ depend on the type of switches being used.

Two different types of switches are shown in Fig. 2. A configuration implementing the set $\mathcal{B} = \{-1, 1\}$ with a signal inverter is shown in Fig. 2a, while a configuration implementing the extended set $\mathcal{B} = \{0, -1, 1\}$ is shown in Fig. 2b. To include more possibilities, we propose equipping the antennas with signal splitters that output the in-phase and quadrature components separately. Using these devices with signal inverters allows a new architecture to be built that makes possible the inclusion of the imaginary unit and its inverse in the set of allowed values. Two switch configurations based on this architecture are depicted in Fig. 3. Fig. 3a shows a switch implementing $\mathcal{B} = \{-1, 1, -j, j\}$ while Fig. 3b shows a switch implementing $\mathcal{B} = \{0, -1, 1, -j, j\}$. It is important to note that the switching elements for the in-phase and quadrature branches in these switches must be in the same respective position, as emphasized by the vertical dashed line.

We finally note that in our approach, any antenna can be connected to any RF chain via a switching element. That is, the matrices $\mathbf{S}_t$ are not constrained to have a maximum number of non-zero entries per row.




(a) Switch configuration used for $\mathcal{B} = \{-1, 1, -j, j\}$.




(b) Switch configuration used for $\mathcal{B} = \{0, -1, 1, -j, j\}$.

Fig. 3. Switch configurations for the new architecture in Section III-A, which can be used to implement richer sets $\mathcal{B}$ of possible values for the matrices $\mathbf{S}_t$.

### B. Problem Formulation with Switches

Using the hybrid analog/digital architecture provided by the RF switches and defining the $T$-tuple of matrices $\mathbf{S} = [\mathbf{S}_1, \ldots, \mathbf{S}_T]$, the optimization problem (6) can be rewritten as

$$\begin{aligned} &\underset{\mathbf{S}}{\text{minimize}}\ \mathcal{M}(\mathbf{S}) \\ &\text{subject to } \mathbf{S} \in \mathcal{B}^{L\times N\times T}, \end{aligned} \tag{9}$$

where the objective function is

$$\mathcal{M}(\mathbf{S}) = \text{tr}\Big(\mathbf{R} - \rho\mathbf{R}\mathbf{S}_{\text{c}}^{\text{H}}\left(\rho\mathbf{S}_{\text{c}}\mathbf{R}\mathbf{S}_{\text{c}}^{\text{H}} + \mathbf{S}_{\text{d}}\mathbf{S}_{\text{d}}^{\text{H}}\right)^{-1}\mathbf{S}_{\text{c}}\mathbf{R}\Big) \tag{10}$$

and $\mathbf{S}_{\text{c}}$ and $\mathbf{S}_{\text{d}}$ are defined the same way as $\mathbf{F}_{\text{c}}$ and $\mathbf{F}_{\text{d}}$. This new problem is more challenging to solve than (6) because the feasible set is discrete rather than continuous. However, this complexity contrasts with the simplicity of the hardware used for implementing the solution. Switches are indeed simpler than the phase shifters considered in the original problem, and they are faster, more affordable, and consume less power [10].

Once the matrices $\mathbf{S}_1, \ldots, \mathbf{S}_T$ have been determined, they can be used to obtain the channel estimate in the same way as in the case with phase shifters. Specifically, $\hat{\mathbf{g}} = \mathbf{W}_{\text{opt}}\mathbf{y}_{\text{c}}$, where $\mathbf{W}_{\text{opt}}$ is the optimal baseband processor in (7), but with $\mathbf{F}_{\text{c}}$ and $\mathbf{F}_{\text{d}}$ replaced with $\mathbf{S}_{\text{c}}$ and $\mathbf{S}_{\text{d}}$, respectively

### C. Channel Estimation using Genetic Algorithm

The solution proposed here to solve problem (9) is a genetic algorithm (GA). As part of the larger class of evolutionary algorithms, a GA generates a set, called population, of feasible points, called individuals. At the $n^{\text{th}}$ iteration, the population $\mathcal{P}_n$ evolves into a new one, $\mathcal{P}_{n+1}$, in which the individuals are ideally closer to an optimum of the objective function [14]. GAs are interesting tools for optimization because they do not require any particular assumption on the solutions and can be adapted to solve a wide variety of problems, e.g., to configure seismic dampers [15], to reduce the risks of terrorism and piracy [16], and to optimize transportation [17], [18].

In GAs, a new population is produced from the previous one by means of selection, crossover, and mutation, as follows:

- Selection, or elitism, consists of taking individuals of the previous, or old, population and inserting them directly into the new population. The individuals that are selected are usually the best ones in terms of a selected cost function.
- Crossovers consist of selecting pairs of individuals from the old population, called parents, and for each pair selecting and swapping parts of them to create two new individuals.

- Mutations consist of choosing individuals from the old population, changing some of their features, called genes, and inserting the modified individuals into the new population.

In the GA presented here to solve the combinatorial problem (9), the population $\mathcal{P}_n$ consists of the collection of $M_n$ different $T$-tuples $\mathbf{S}_m^{(n)}$ taken from the feasible sets $\mathcal{B}^{L\times N\times T}$, where $m \in \{1, 2, \ldots, M_n\}$ and $M_n$ is the population size at the $n^{\text{th}}$ iteration. At each iteration, the current population $\mathcal{P}_n$ is updated to $\mathcal{P}_{n+1}$ by applying a combination of the mechanisms introduced above. Specifically, when a new population is produced, the individuals from the old population are first sorted according to their MSE, as defined in (10), and the best ones are used to produce the next population as follows:

- The $M_n^{\text{S}}$ individuals with the lowest MSE are selected without modification, where $M_n^{\text{S}} \in \{0, \ldots, M_n\}$ is an integer.
- The $M_n^{\text{C}}$ individuals with the lowest MSE are used to perform the crossover, where $M_n^{\text{C}} \in \{0, \ldots, M_n\}$. If $M_n^{\text{C}}$ is even, these individuals are paired sequentially to form parents. For each pair, two new individuals, called children, are created by random crossover. In the first child, each entry of the matrices $\mathbf{S}_t$ is taken from one of the parents with probability $p = \frac{1}{2}$. Then for the second child, the corresponding entry is selected from the other parent. If $M_n^{\text{C}}$ is odd instead, $M_n^{\text{C}}+1$ individuals are used to form parents, but the $(M_n^{\text{C}}+1)^{\text{th}}$ child produced is discarded.
- The $M_n^{\text{M}}$ individuals with the lowest MSE are chosen for the mutation process, where $M_n^{\text{M}} \in \{0, \ldots, M_n\}$. The mutation is carried on for each entry of the matrices $\mathbf{S}_t$ with probability $\mu$. If the mutation occurs, it consists of replacing the corresponding entry by an equiprobable random element taken from the set $\mathcal{B}$ of allowed values.

The iterative process continues until there is no improvement in the best or the average MSE for a given number $c_{\max}$ of iterations. The GA, presented in Algorithm 1, takes as input the dimensions $L$ and $N$ of the matrices $\mathbf{S}_t$ and their number $T$, and the set $\mathcal{B}$ of allowed values. It also depends on the following parameters: the initial size $M_0$ of the population; the shrinking parameter $\alpha \in ]0, 1]$, from which the population size at iteration $n$ is computed as $M_n = \lceil M_0\alpha^n \rceil$; the proportions $m_{\text{S}}, m_{\text{C}}, m_{\text{M}} \in [0, 1]$ of the population created by each of the three evolution mechanisms (satisfying $m_{\text{S}} + m_{\text{C}} + m_{\text{M}} = 1$), from which $M_n^{\text{S}} = \lfloor m_{\text{S}} M_n \rceil$, $M_n^{\text{C}} = \lfloor m_{\text{C}} M_n \rceil$, and $M_n^{\text{M}} = \lfloor m_{\text{M}} M_n \rceil$ are computed, where $\lfloor x \rceil$ is $x$ rounded to the nearest integer; the mutation probability $\mu \in [0, 1]$; and the number $c_{\max}$ of iterations without improvement before the algorithm stops.

The overall computational complexity of the GA can be obtained by evaluating the number of operations (ops) at each iteration. To simplify this analysis, we assume that $LT = N$, a practical condition needed to estimate all the parameters of the channels, and that the population size $M_n = M_0$ remains constant during the iterative process (i.e., $\alpha = 1$). At each iteration, computing the MSE (10) of each individual requires $10.3N^3 + \mathcal{O}(N^2)$ ops, implementing the selection, crossover and mutation procedures requires $M_0\mathcal{O}(N^2)$ ops, and sorting the population with respect to the MSE requires $2M_0 \log_2 M_0 + \mathcal{O}(M_0)$ ops. Hence, assuming that the algorithm stops at $n_{max}$ iterations, the total complexity of the GA is given by $n_{max}M_0(10.3N^3 + 2\log_2 M_0 + \mathcal{O}(N^2))$ ops.

**Algorithm 1:** Genetic algorithm applied to finding analog combiners for switching networks

**Input:** $L$, $N$, $T$, $\mathcal{B}$
**GA Parameters:** $M_0$, $\alpha$, $m_{\text{S}}$, $m_{\text{C}}$, $m_{\text{M}}$, $\mu$, $c_{\max}$

$n \leftarrow 0$
Generate initial population $\mathcal{P}_n = \{\mathbf{S}_1^{(n)}, \ldots, \mathbf{S}_{M_n}^{(n)}\}$ with each entry in each individual equiprobably picked from $\mathcal{B}$
Sort in increasing MSE order: $\mathcal{M}(\mathbf{S}_1^{(n)}) \leq \ldots \leq \mathcal{M}(\mathbf{S}_{M_n}^{(n)})$
Save $\mathcal{M}_{\text{best}}^{\min} \leftarrow \mathcal{M}(\mathbf{S}_1^{(n)})$, and $\overline{\mathcal{M}}_{\text{best}} \leftarrow \frac{1}{M_n}\sum_{m=1}^{M_n} \mathcal{M}(\mathbf{S}_m^{(n)})$
$c \leftarrow 0$
**while** $c < c_{\max}$ **do**
  $n \leftarrow n + 1$
  $M_n = \lceil M_0\alpha^n \rceil$
  $M_n^{\text{S}} = \lfloor m_{\text{S}} M_n \rceil$, $M_n^{\text{C}} = \lfloor m_{\text{C}} M_n \rceil$, $M_n^{\text{M}} = \lfloor m_{\text{M}} M_n \rceil$
  // Selections:
  Copy the $M_n^{\text{S}}$ best individuals from $\mathcal{P}_{n-1}$ to $\mathcal{P}_n$
  // Crossovers:
  Take the $M_n^{\text{C}}$ best individuals of $\mathcal{P}_{n-1}$ to form parents
  For each consecutive pair, produce children following the method in Subsection III-C, and store them in $\mathcal{P}_n$
  // Mutations:
  Copy the $M_n^{\text{M}}$ best individuals from $\mathcal{P}_{n-1}$ to $\mathcal{P}_n$ and randomly replace entries as described in Subsection III-C
  // Evaluation of the new population:
  Sort by MSE, so that $\mathcal{M}(\mathbf{S}_1^{(n)}) \leq \ldots \leq \mathcal{M}(\mathbf{S}_{M_n}^{(n)})$
  Save $\mathcal{M}_n^{\min} = \mathcal{M}(\mathbf{S}_1^{(n)})$, $\overline{\mathcal{M}}_n = \frac{1}{M_n}\sum_{m=1}^{M_n} \mathcal{M}(\mathbf{S}_m^{(n)})$
  **if** $\mathcal{M}_n^{\min} < \mathcal{M}_{\text{best}}^{\min}$ or $\overline{\mathcal{M}}_n < \overline{\mathcal{M}}_{\text{best}}$ **then**
    $c \leftarrow 0$
    **if** $\mathcal{M}_n^{\min} < \mathcal{M}_{\text{best}}^{\min}$ **then**
      $\mathcal{M}_{\text{best}}^{\min} \leftarrow \mathcal{M}_n^{\min}$
      $\mathbf{S}_{\text{best}} \leftarrow \mathbf{S}_1^{(n)}$
    **if** $\overline{\mathcal{M}}_n < \overline{\mathcal{M}}_{\text{best}}$ **then**
      $\overline{\mathcal{M}}_{\text{best}} \leftarrow \overline{\mathcal{M}}_n$
  **else**
    $c \leftarrow c + 1$

**Output:** Analog combiners for switching networks $\mathbf{S}_{\text{best}}$

## IV. Simulation Results

In this section, the performance of the proposed GA for precoder design and channel estimation in switch-based hybrid massive MIMO systems is evaluated by computer simulations. The BS is equipped with a ULA of $N$ omni-directional antenna elements with half-wavelength spacing, i.e. $d = \lambda/2$, where $N \in \{16, 32, 64\}$. The other system parameters are set as follows: $L = 8$ RF chains, $T \in \{1, 2, 4\}$ training pilots, and $P = 6$ propagation paths. In [7], it is demonstrated that SO achieves the same performance as AO but with lower complexity; hence it is used as a benchmark. SO minimizes the cost function (8) over the feasible set $\mathcal{F}^T = \{e^{\jmath\theta}\}^{L\times N\times T}$ and the analog combiner uses phase shifters; whereas the GA minimizes the function (10) over the set $\mathcal{B}^{L\times N\times T}$, where the analog combiner uses switches.

Our implementation of the GA uses an initial population $M_0 = 500$ and a shrinking parameter $\alpha = 0.98$, with $m_{\text{S}} = 0.1$, $m_{\text{C}} = 0.5$, and $m_{\text{M}} = 0.4$. The maximum number of strikes is set as $c_{\max} = 5$ and the mutation probability is $\mu = 2.5 \cdot 10^{-3}$. For the SO method, the analog combiners computed by the algorithm are normalized and quantized to the required number $N_B$ of bits, as described in Section II-C. Unless otherwise

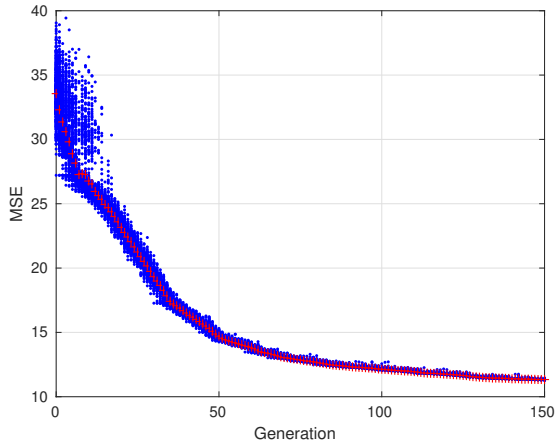



Fig. 4. MSE of the population versus iteration number of the GA. The blue dots are the MSEs of the individuals and the red crosses are the population averages at each iteration.

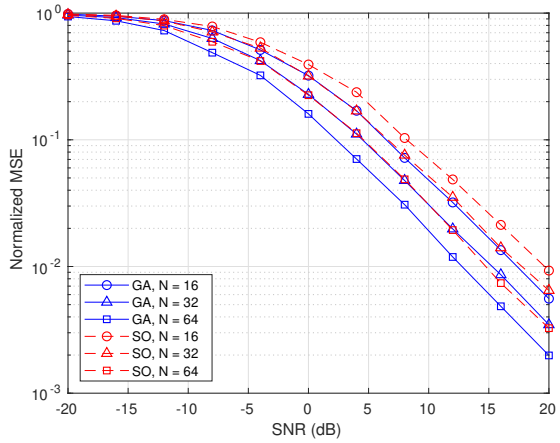



Fig. 5. NMSE versus SNR for the GA and the SO method with 1-bit quantization, for $T = 1$ pilot symbol and different numbers $N$ of antennas.

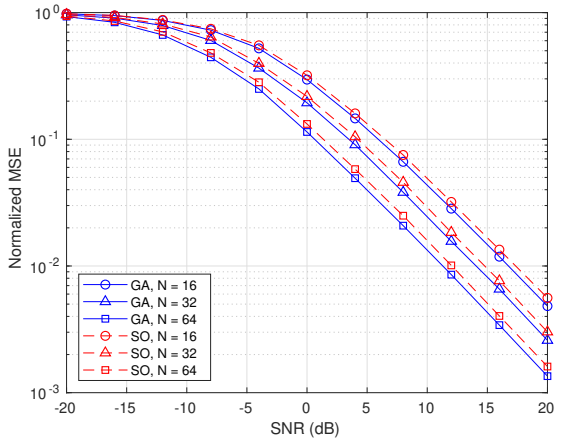



Fig. 6. NMSE versus SNR for the GA and the SO method with 2-bit quantization for $T = 1$ pilot symbol and different numbers $N$ of antennas.

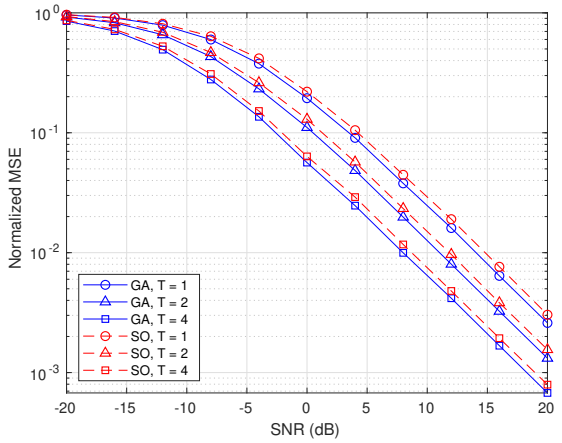



Fig. 7. NMSE versus SNR for the GA and the SO method with 2-bit quantization for different numbers $T$ of pilot symbols and $N = 32$ antennas.

specified, $N_B$ is chosen so that the number of possibilities for the phase angles matches the number of allowed values in the set $\mathcal{B}$ for the switches.

For both algorithms under comparison, the channel covariance matrix $\mathbf{R}$ is assumed to be known beforehand at the BS. For a discussion on how to estimate this matrix in practice, see [7]. In our simulations, $\mathbf{R}$ is computed using (5) with a fixed number $P = 6$ of paths and given sets $\{\theta_1, \ldots, \theta_P\}$ of AoAs independently generated with distribution $\mathcal{U}(0, 2\pi)$. Matrix $\mathbf{R}$ is used to generate synthetic data, as well as to implement the channel estimators under study. The true channel vectors in the simulations are generated as $\mathbf{g} = \mathbf{R}^{1/2}\mathbf{h}$, where $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is used to model the small-scale fading. The signal $\mathbf{y}_c$ received at the baseband processor is then obtained using (3), where the noise $\mathbf{n}_c \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{NT})$ and the pilot power $\rho$ provide the desired SNR level. Matrix $\mathbf{R}$ is also used by the SO and GA methods in their respective cost functions (8) and (10) to determine the optimal combiner matrices. Channel estimation is then achieved by applying the transformation $\hat{\mathbf{g}} = \mathbf{W}_{\text{opt}}\,\mathbf{y}_c$ to the received data $\mathbf{y}_c$, using the optimal $\mathbf{W}_{\text{opt}}$ in (7).

The accuracy of the channel estimates computed by the two algorithms is evaluated in terms of the normalized mean square error (NMSE) between the real and the estimated channel vectors. Specifically, for each point in the figures below (except Fig. 4), we first generate $N_1 = 20$ different sets $\{\theta_1^{(i)}, \ldots, \theta_P^{(i)}\}$ of AoAs and their corresponding covariance matrices $\mathbf{R}_i$, where $i \in \{1, \ldots, N_1\}$. For each $\mathbf{R}_i$, $N_2 = 5,000$ different channel realizations $\mathbf{g}_{i,j} = \mathbf{R}_i^{1/2}\mathbf{h}_j$ are generated, where $\mathbf{h}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $j \in \{1, \ldots, N_2\}$. Then, for each $\mathbf{g}_{i,j}$, the estimated channel $\hat{\mathbf{g}}_{i,j}$ is computed. Finally, the NMSE is evaluated as

$$\text{NMSE} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \frac{\|\mathbf{g}_{i,j} - \hat{\mathbf{g}}_{i,j}\|^2}{\|\mathbf{g}_{i,j}\|^2}. \tag{11}$$

Fig. 4 shows an example of the MSE performance of the population during the optimization process as the GA finds better individuals at each iteration, for a single realization of the algorithm. It can be seen that the performance improves until it stalls towards the end. The algorithm stops when the average MSE and the best MSE do not improve for $c_{\max}$ iterations. The typical run time of the GA exceeds that of the SO method. However, since it is based on switches, it lends itself to more flexible and less costly implementation. Also, as shown below, the GA offers more accuracy than the SO method.

Figures 5 and 6 show the NMSE versus SNR performance of the two methods for $T = 1$ pilot symbol and different numbers $N$ of antennas with 1-bit and 2-bit quantization in the analog combiner, respectively, corresponding to $\mathcal{B} = \{-1, 1\}$ and $\{-1, 1, -j, j\}$ in the case of the GA. It can be seen that

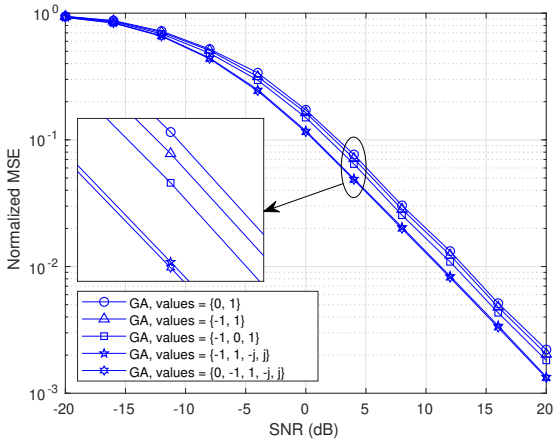



Fig. 8. NMSE versus SNR for the GA with different sets $\mathcal{B}$ of allowed values, for $T = 1$ pilot symbol, $N = 64$ antennas and SNR $= 0$ dB.

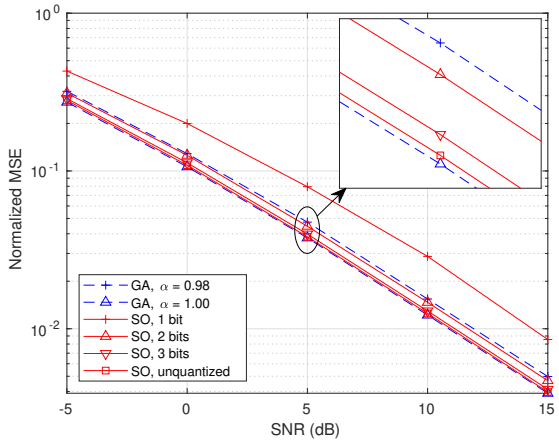



Fig. 9. NMSE versus SNR for the GA with $\mathcal{B} = \{0, -1, 1, -j, j\}$ and the SO method with different numbers of quantization bits, for $T = 2$ and $N = 32$.

the GA outperforms the SO method in both figures. In the 1-bit case, in particular, the GA requires half the number of antennas as the SO to offer a similar performance.

In Fig. 7, the comparison is carried out for different numbers $T$ of pilot symbols and $N = 32$ antennas with 2-bit quantization. The figure shows that regardless of the number of pilot symbols, the GA is consistently more accurate than the SO method, leading to an SNR gain of about 0.8 dB. Fig. 8 compares the performance of the GA for the different switch types presented in Section III-A, for the case of $T = 1$ pilot and $N = 64$ antennas. Using more complex switches, corresponding to larger sets $\mathcal{B}$, allows more possible values for the entries of the analog combiner, and thus provides better channel estimates.

Fig. 9 compares the performance of the GA to that of the SO method with different numbers of quantization bits in the phase shifters, for $T = 2$ and $N = 32$. The GA uses the switch configuration shown in Fig. 3b, corresponding to the set $\mathcal{B} = \{0, \pm 1, \pm j\}$, with two different values of the shrinking parameter, i.e. $\alpha = 0.98$ and 1. We note that the GA using the above 5-state switch ($\approx 2.3$ bits) and $\alpha = 1$ performs better than the SO method with infinite resolution phase shifters.

## V. Conclusion

We addressed the problem of uplink channel estimation in hybrid analog/digital massive MIMO systems for mmWave communications. An alternative architecture was introduced with RF switches as opposed to the phase shifters in the conventional literature, to reduce system cost and implementation complexity. The channel estimation was formulated as a combinatorial optimization problem, seeking to minimize the MSE between the real and estimated channels over a finite set of allowed values for the switches. A genetic algorithm (GA) was developed for solving this problem and obtaining the MIMO channel estimates. Simulations results confirmed that the proposed architecture can estimate channels as accurately, if not more, as existing hybrid methods using phase shifters.

## References


[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[5] A. Morsali, A. Haghighat, and B. Champagne, "Generalized framework for hybrid analog/digital signal processing in massive and ultra-massive-mimo systems," *IEEE Access*, vol. 8, pp. 100 262–100 279, 2020.

[6] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, June 2016.

[7] L. Pan, L. Liang, W. Xu, and X. Dong, "Framework of channel estimation for hybrid analog-and-digital processing enabled massive MIMO communications," *IEEE Trans. on Communications*, vol. 66, no. 9, pp. 3902–3915, Sep. 2018.

[8] S. Buzzi and C. D'Andrea, "Subspace tracking and least squares approaches to channel estimation in millimeter wave multiuser MIMO," *IEEE Trans. on Communications*, pp. 1–1, June 2019.

[9] P. Dong, H. Zhang, G. Y. Li, I. Gaspar, and N. Naderializadeh, "Deep CNN-based channel estimation for mmwave massive MIMO systems," *IEEE J. of Selected Topics in Signal Processing*, pp. 1–1, July 2019.

[10] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[11] F. Molina and J. Borràs, "Low-complexity switching network design for hybrid precoding in mmWave MIMO systems," in *European Signal Processing Conference (EUSIPCO)*, Sep. 2019, pp. 1–5.

[12] R. Hu, J. Tong, J. Xi, Q. Guo, and Y. Yu, "Robust channel estimation for switch-based mmwave MIMO systems," in *Int. Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2017, pp. 1–7.

[13] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[14] W. Banzhaf, P. Nordin, R. E. Keller, and F. D. Francone, *Genetic Programming: An Introduction On the Automatic Evolution of Computer Programs and Its Applications*, 1st ed. Morgan Kaufmann Publishers (San Francisco); Dpunkt-verlag (Heidelburg), Feb. 1998, ch. 4, pp. 87–104.

[15] M. Patrascu, "Genetically enhanced modal controller design for seismic vibration in nonlinear multi-damper configuration," *Proceedings of the Institution of Mechanical Engineers, Part I: J. of Systems and Control Engineering*, vol. 229, no. 2, pp. 158–168, Sep. 2015.

[16] J. Buurman, S. Zhang, and V. Babovic, "Reducing risk through real options in systems design: The case of architecting a maritime domain protection system," *Risk analysis : an official publication of the Society for Risk Analysis*, vol. 29, pp. 366–79, Mar. 2009.

[17] T. Vidal, T. G. Crainic, M. Gendreau, N. Lahrichi, and W. Rei, "A hybrid genetic algorithm for multidepot and periodic vehicle routing problems," *Operations Research*, vol. 60, no. 3, pp. 611–624, June 2012.

[18] A. George, B. R. Rajakumar, and D. Binu, "Genetic algorithm based airlines booking terminal open/close decision system," in *Proceedings of the Int. Conference on Advances in Computing, Communications and Informatics*, ser. ICACCI '12. New York, NY, USA: ACM, Aug. 2012, pp. 174–179.