A Stochastic Optimization Approach to Hybrid Processing in Massive MIMO Systems

Georgios K. Papageorgiou, *Member, IEEE*, Mathini Sellathurai, *Senior Member, IEEE*, Konstantinos Ntougias, Constantinos B. Papadias, *Fellow, IEEE*

***Abstract*—The high cost and energy consumption of fully digital massive multiple-input multiple-output (MIMO) systems has led to hybrid designs with fewer radio frequency (RF) chains than antennas. In this letter, we propose an efficient hybrid processing algorithm for point-to-point (P2P) massive MIMO systems that operate in either rich or poor scattering environments. The proposed scheme, i.e., hybrid processing via stochastic approximation with Gaussian smoothing (HPSAGS), alternates between a digital baseband and an analog RF precoder/combiner computation step. The method achieves state-of-the-art performance with low computational cost, which is essential for large MIMO systems.**

***Index Terms*—hybrid processing, massive MIMO, stochastic approximation, Gaussian smoothing, millimeter waves**

I. Introduction

Massive MIMO has been recognized as one of the key technologies for future generation communication networks, due to its high capacity gains [1]. Moreover, it is considered an integral component of millimeter wave (mmWave) communications, since the high array gain of massive MIMO systems can compensate for the large propagation losses in such frequencies [2], while the small wavelengths facilitate compact antenna designs [3]. However, implementing large fully digital transceivers is challenging, due to the requirement of feeding each antenna element (AE) by its own RF chain, which results in high cost and energy consumption [4]. In recent years there has been a vast amount of literature on hybrid analog-digital designs to this end. Some of the designs consider switches [5] and other consider load controlled passive elements [6]. The most popular one, though, employs phase shifters [4], demonstrating a good trade-off between performance and cost.

Some hybrid precoder/combiner designs for P2P systems consider mmWave channels. [7] provides an overview of these methods. Most of them are computationally demanding—for instance, in [8] a simplex 1-D iterative local search is performed *per element of the analog precoder*. The state-of-the-art is spatially sparse precoding with orthogonal matching pursuit (SSPOMP) [9]. This algorithm exploits the poor scattering nature of mmWave channels (sparsity) to achieve good performance with low complexity. Other schemes consider also rich scattering environments, which are typically found in systems that operate in sub-6 GHz. Nevertheless, their computational complexity is typically prohibitive. For example, the matrix decomposition precoding (MDP) scheme introduced in [10] demonstrates near-optimal performance in both rich and poor scattering environments at the cost of increased computational complexity, due to the requirement for *solving as many quadratic programming tasks as the number of antennas*.

In this letter, we propose an efficient algorithm (in terms of performance and computational cost) for hybrid processing that can be used in both rich and poor scattering environments to fill the aforementioned gaps in the related literature. The proposed alternating optimization algorithm employs a *convolution smoothing technique* [11] followed by a stochastic approximation scheme for the estimation of the phases (analog RF precoder/combiner) and demonstrates very good performance with low computational complexity, particularly in rich scattering environments.

II. System Model

We consider a P2P communication of a hybrid transmitter equipped with N_t antennas and M_t RF chains and a hybrid receiver with N_r antennas and M_r RF chains, where N_s data streams are supported. The architecture of both systems is a fully connected one with fewer RF chains than antennas, where $N_s \leq M_t \leq N_t$ and $N_s \leq M_r \leq N_r$. The transmitter applies a $M_t \times N_s$ baseband precoder $\mathbf{F}_B$ (enabling both amplitude and phase modifications) and a $N_t \times M_t$ analog precoder $\mathbf{F}_R$ (enabling phase changes only). Therefore, each (i,j)-th element of $\mathbf{F}_R$ satisfies $|(\mathbf{F}_R)_{i,j}| = 1/\sqrt{N_t}$. Finally, to meet the total transmit power constraint $\mathbf{F}_B$ is normalized to satisfy $\|\mathbf{F}_R\mathbf{F}_B\|_F^2 = N_s$. Assuming a narrowband block-fading propagation channel, as in [9], the received signal $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ before combining is $\mathbf{y} = \mathbf{H}\mathbf{F}_R\mathbf{F}_B\mathbf{s} + \mathbf{n}$, where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the normalized channel matrix with $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$, $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the transmitted signal with $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = (P/N_s)\mathbf{I}_{N_s}$, P is the average transmit power, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I}_{N_r})$ is the i.i.d. noise vector.

Furthermore, we assume that the channel is known at both the transmitter and the receiver, the processed received signal after combining is expressed as $\widetilde{\mathbf{y}} = \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{H}\mathbf{F}_R\mathbf{F}_B\mathbf{s} + \mathbf{z}$, where $\mathbf{z} = \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{n}$, while $\mathbf{W}_R$ and $\mathbf{W}_B$ denote the analog and digital combining matrices, respectively. A fully connected phase shifter design is considered for the combiner as well, hence, $|(\mathbf{W}_R)_{i,j}| = 1/\sqrt{N_r}$. Under Gaussian signaling the achieved instantaneous spectral efficiency is given by:

$$R(\mathbf{F}_R, \mathbf{F}_B, \mathbf{W}_R, \mathbf{W}_B) = \log_2(|\mathbf{I}_{N_s} + \frac{P}{N_s}\mathbf{R}_\mathbf{z}^{-1}\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H|), \quad (1)$$

where $\mathbf{R}_\mathbf{z} = \sigma_n^2\mathbf{W}_B^H\mathbf{W}_R^H\mathbf{W}_R\mathbf{W}_B$ is the noise covariance matrix after combining and $\widetilde{\mathbf{H}} = \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{H}\mathbf{F}_R\mathbf{F}_B$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2020.2969203, IEEE Wireless Communications Letters

Scattering Environment:
1) For rich scattering we consider i.i.d. Rayleigh fading channels. For this type of channels, we consider perfect/imperfect channel-state-information (PCSI/ICSI), which is modeled via $\hat{\mathbf{H}} = \sqrt{1-\alpha^2}\mathbf{H} + \alpha\mathbf{E}$, where $0 \leq \alpha \leq 1$ and $\mathbf{E} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.
2) For limited scattering (sparse) we consider the geometric clustered mmWave model [9] with N_c clusters and N_p paths per cluster:

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_c N_p}} \sum_{m=1}^{N_c} \sum_{n=1}^{N_p} \beta_{mn} \mathbf{a}_r(\phi_{mn}) \mathbf{a}_t(\theta_{mn})^H, \quad (2)$$

where $\beta_{mn} \sim \mathcal{CN}(0,1)$ is the complex channel gain of the (m,n)-th path, while $\mathbf{a}_r(\phi_{mn})$ denotes the receive array response vector at the azimuth angle of arrival (AoA) ϕ_{mn} and $\mathbf{a}_t(\theta_{mn})$ denotes the transmit array response vector at the azimuth angle of departure (AoD) θ_{mn}. The mean angles of each cluster (center) are uniformly distributed and the angles within each cluster are distributed according to the truncated Laplace distribution with angular spreads $\sigma_\phi, \sigma_\theta$, respectively. Finally, we considered uniform linear arrays (ULAs) with half-wavelength spacing of the N antenna elements (N_t for the transmitter and N_r for the receiver) in the numerical section with the array response vector given by $\mathbf{a}_k(\vartheta) = [1, e^{-\mathrm{j}\pi\sin\vartheta}, \ldots, e^{-\mathrm{j}\pi\sin\vartheta(N-1)}]^T/\sqrt{N},\ k \in \{r,t\}$.

III. Preliminaries

A. Hybrid Design

Assuming that $\mathrm{rank}(\mathbf{H}) \geq N_s$, the optimal precoder $\mathbf{F}_*$ and combiner $\mathbf{W}_*$ of a fully digital system can be found by the singular value decomposition (SVD) of the channel matrix $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are $N_r \times N_r$ and $N_t \times N_t$ unitary matrices, respectively, and $\mathbf{\Lambda}$ is the an $N_r \times N_t$ diagonal matrix with singular values in decreasing order on its diagonal. The optimal unconstrained precoder and combiner, for equal power allocation is given by $\{\mathbf{F}_*, \mathbf{W}_*\} = \{\mathbf{V}_1, \mathbf{U}_1\}$, where $\mathbf{V}_1$ and $\mathbf{U}_1$ are obtained from $\mathbf{V}$ and $\mathbf{U}$ by extracting their first N_s columns, respectively. Joint optimization of the hybrid precoders $\{\mathbf{F}_R, \mathbf{F}_B\}$ and combiners $\{\mathbf{W}_R, \mathbf{W}_B\}$ (global minimum of the joint design) is a difficult task due to the non-convex constraints of the analog RF precoder and combiner. The adopted approach is to first design the hybrid precoders, which are sufficiently close to the optimal ones, i.e., $\mathbf{F}_* = \mathbf{V}_1$, by solving the following optimization task:

$$\begin{aligned} &\min_{\mathbf{F}_R, \mathbf{F}_B} \left\| \mathbf{F}_* - \mathbf{F}_R\mathbf{F}_B \right\|_F^2, \\ &\text{s.t. } \mathbf{F}_R \in \mathcal{F}_R,\ \left\| \mathbf{F}_R\mathbf{F}_B \right\|_F^2 = N_s, \end{aligned} \quad (3)$$

where $\mathcal{F}_R = \{\mathbf{F}_R \in \mathbb{C}^{N_t \times M_t} : |(\mathbf{F}_R)_{i,j}| = 1/\sqrt{N_t}\}$ is the set of matrices with constant-magnitude entries. The fact that the error of the approximation in (3) is non zero makes $\mathbf{U}_1$ no further optimal. The linear MMSE combiner that will achieve the maximum spectral efficiency for linear and separate detection of each data stream is given by:

$$\mathbf{W}_* = \frac{\sqrt{P}}{N_s} \left(\frac{P}{N_s} \mathbf{H}\mathbf{F}_R\mathbf{F}_B\mathbf{F}_B^H\mathbf{F}_R^H\mathbf{H}^H + \sigma_n^2 \mathbf{I}_{N_r} \right)^{-1} \mathbf{H}\mathbf{F}_R\mathbf{F}_B. \quad (4)$$

Hence, given the set of optimized precoders and calculating $\mathbf{W}_*$ from (4), the hybrid combiner can be obtained in a similar manner as the solution to the following task:

$$\min_{\mathbf{W}_R, \mathbf{W}_B} \left\| \mathbf{W}_* - \mathbf{W}_R\mathbf{W}_B \right\|_F^2, \text{ s.t. } \mathbf{W}_R \in \mathcal{W}_R, \quad (5)$$

where $\mathcal{W}_R$ is the set of complex $N_r \times M_t$ matrices with constant-magnitude entries.

B. Gaussian Smoothing of Matrix Variable Functions

Definition 1: The random matrix $\mathbf{S} \in \mathbb{R}^{N \times M}$ follows a matrix variate normal distribution (MVND), denoted as $\mathbf{S} \sim \mathcal{MN}_{N\times M}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Psi})$, where $\mathbf{M} \in \mathbb{R}^{N \times M}$ is its mean, and $\mathbf{\Sigma} \in \mathbb{R}^{N \times N}$, $\mathbf{\Psi} \in \mathbb{R}^{M \times M}$ are positive definite matrices, if $\mathrm{vec}(\mathbf{S}) \sim \mathcal{N}_{NM}(\mathrm{vec}(\mathbf{M}), \mathbf{\Psi} \otimes \mathbf{\Sigma})$ [12]. The p.d.f. of $\mathbf{S}$ is given by:

$$p(\mathbf{S}|\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Psi}) = \frac{e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Psi}^{-1}(\mathbf{S}-\mathbf{M})\mathbf{\Sigma}^{-1}(\mathbf{S}-\mathbf{M})^T\right)}}{\sqrt{(2\pi)^{NM}\det(\mathbf{\Sigma})^M \det(\mathbf{\Psi})^N}}. \quad (6)$$

Let $\mathbf{M} = \mathbf{O}_{N\times M}$ (zero matrix) and $\mathbf{\Sigma} = \beta^2\mathbf{I}_N, \mathbf{\Psi} = \gamma^2\mathbf{I}_M$. Moreover, considering $\mu = \beta\gamma$ (6) is written as:

$$p(\mathbf{S}, \mu) = \frac{e^{-\frac{1}{2\mu^2}\|\mathbf{S}\|_F^2}}{\mu^{NM}\sqrt{(2\pi)^{NM}}}. \quad (7)$$

The smoothed approximation to the original function f with weighting Gaussian p.d.f., $p(\mathbf{S}, \mu)$, can be expressed via their convolution given by:

$$\begin{aligned} f_\mu(\mathbf{X}) = (p * f)(\mathbf{X}) &= \int_{\mathbb{R}^{N\times M}} p(\mathbf{S}, \mu) f(\mathbf{X} - \mathbf{S}) d\mathbf{S}, \\ &= \int_{\mathbb{R}^{N\times M}} p(\mathbf{S}) f(\mathbf{X} - \mu\mathbf{S}) d\mathbf{S}, \end{aligned} \quad (8)$$

where $p(\mathbf{S}) = p(\mathbf{S}, 1)$ is the standard MVND (using a change of variables). From (8), it is directly observed that $f_\mu(\mathbf{X}) = \mathbb{E}_\mathbf{S}\left[f(\mathbf{X} - \mu\mathbf{S})\right]$, which leads to:

$$\nabla_\mathbf{X} f_\mu(\mathbf{X}) = \mathbb{E}_\mathbf{S}\left[\nabla_\mathbf{X} f(\mathbf{X} - \mu\mathbf{S})\right], \quad (9)$$

where i.i.d. samples are obtained from the $\mathbb{R}^{N\times M}$ space with the p.d.f. $p(\mathbf{S})$. Therefore, the (one-sided) unbiased gradient estimator is expressed as $\nabla_\mathbf{X} f_\mu(\mathbf{X}) = \frac{1}{L}\sum_{\ell=1}^{L} \nabla_\mathbf{X} f(\mathbf{X} - \mu\mathbf{S}_{[\ell]})$. Using the change of variables $\mathbf{S} = -\mathbf{Y}$ in (9), summing and solving w.r.t. the gradient we obtain the two-sided estimate of the gradient, given by:

$$\nabla_\mathbf{X} f_\mu(\mathbf{X}) = \frac{1}{2L}\sum_{\ell=1}^{L} \left[\nabla_\mathbf{X} f(\mathbf{X} + \mu\mathbf{S}_{[\ell]}) + \nabla_\mathbf{X} f(\mathbf{X} - \mu\mathbf{S}_{[\ell]})\right]. \quad (10)$$

It should be noted that (10) suggests that L samples can be used for the gradient estimation, as in a mini-batch approach. However, in this work we only consider its stochastic flavor, i.e., $L = 1$ [11].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2020.2969203, IEEE Wireless Communications Letters

IV. Hybrid Precoding via Stochastic Approximation with Gaussian Smoothing

In this section we introduce an *iterative scheme* for hybrid precoding via Stochastic Approximation with Gaussian Smoothing (HPSAGS), which alternates between the optimization of the digital and the analog precoder, as a solution of (3). The solution of (5) for the combiners is similar and hence omitted due to space limitations (see Remark 1). Therefore, we drop the transmitter-receiver indices from the dimensions and use N, M instead. The scheme is summarized in Alg. 1.

A. Baseband Precoder Update

Given an initial solution $\mathbf{F}_R^{(0)}$, the set of hybrid precoders at the k-th iteration is $\left\{\mathbf{F}_R^{(k)}, \mathbf{F}_B^{(k)}\right\}$. Provided we have computed $\mathbf{F}_R^{(k)}$ the baseband precoder update, $\mathbf{F}_B^{(k)}$, is given by the solution of $\min_{\mathbf{F}_B} \left\|\mathbf{F}_* - \mathbf{F}_R^{(k)}\mathbf{F}_B\right\|_F^2$, given in close form:

$$\mathbf{F}_B^{(k)} = \left(\mathbf{F}_R^{(k)^H}\mathbf{F}_R^{(k)}\right)^{-1}\mathbf{F}_R^{(k)^H}\mathbf{F}_*. \tag{11}$$

B. Analog Precoder Update via Stochastic Approximation with Gaussian Smoothing

Next, follows the update of the analog precoder. To this end, we impose the constant-modulus structure on the matrix. Considering the non-linear mapping $g : \mathbb{R}^{N\times M} \to \mathbb{C}^{N\times M}$ with $g(\mathbf{\Theta}) = e^{\mathrm{j}\mathbf{\Theta}}/\sqrt{N}$, the precoder matrix is expressed via the element-wise function $\mathbf{F}_R = g(\mathbf{\Theta}), \mathbf{\Theta} \in \mathbb{R}^{N\times M}$. Hence, we seek for $\mathbf{\Theta}_{k+1}$ minimizing $f : \mathbb{R}^{N\times M} \to \mathbb{R}$ with $f(\mathbf{\Theta}) = \|\mathbf{F}_* - g(\mathbf{\Theta})\mathbf{F}_B^{(k)}\|_F^2$, leading to the update $\mathbf{F}_R^{(k+1)} = g(\mathbf{\Theta}_{k+1})$. Note that f defines a multiextremal mapping with respect to $\mathbf{\Theta}$. Thus, standard approaches to find a minimizer of f do not apply here. Nevertheless, the function is smooth with gradient:

$$\nabla_{\mathbf{\Theta}} f(\mathbf{\Theta}) = -2\,\mathrm{Re}\{\mathrm{j}g(\mathbf{\Theta}) \odot (\mathbf{F}_* - g(\mathbf{\Theta})\mathbf{F}_B^{(k)})^* \mathbf{F}_B^{(k)^T}\}, \tag{12}$$

where $\mathrm{Re}\{\cdot\}$ denotes the real part of the complex input, $\odot$ is Hadamard (element-wise) product and $(\cdot)^*$ the is the conjugate of the matrix. The objective of *convolution function smoothing* [11] is to represent f as a superposition of a uniextremal function and other multiextremal ones, which add some noise to the former, and perform minimization of the smoothed uniextremal function by filtering out the noise, eventually leading towards its global minimum. This is performed by generating a sequence of minimization runs, while reducing the amount of smoothing at the end of the cycle. For the smoothing of f we have followed the approach in Section III-B for the derivative, therefore, we attempt to solve the following stochastic optimization task at every k-th step (instead of f):

$$\min_{\mathbf{\Theta}} \left\{ f_{\mu_k}(\mathbf{\Theta}) = \mathbb{E}_{\mathbf{S}}\left[f(\mathbf{\Theta} - \mu_k\mathbf{S})\right]\right\}, \tag{13}$$

where $\mathbf{S}$ is sampled from the standard MVND in eq. (7) and the sequence $(\mu_k)_{k\in\mathbb{N}}$ is strictly decreasing with $\lim_{k\to\infty}\mu_k = 0$. However, in practice, a small finite number K is sufficient for the approximation. Finally, at the k-th iteration, Stochastic Gradient Descent (SGD) in Alg. 2 is employed for the phases' update. The computational efficiency of the algorithm is expressed in terms of the worst case complexity, which is $O(NM^2KT_{\max})$. The full code can be found online in [13].

Remark 1: Note that for the design of the combiner the same algorithm can be used with minor modifications, i.e., by replacing $\mathbf{F}_*$ with $\mathbf{W}_*$ in (4) and by neglecting the normalization of the baseband matrix in row 9 of Alg. 1.

Algorithm 1 Hybrid Precoding via Stochastic Approximation with Gaussian Smoothing

1: **procedure** HPSAGS($\mathbf{F}_*, \mathbf{\Theta}_0, (\mu_k)_{k=0}^{K-1}, \eta, T_{\max}, \epsilon$)
2: Set $k \leftarrow 0$
3: **while** $k < K$ **do**
4: Select $\mu \leftarrow \mu_k$
5: Compute $\mathbf{F}_B^{(k)}$ in (11) with $\mathbf{F}_R^{(k)} = g(\mathbf{\Theta}_k)$
6: $\mathbf{\Theta}_{k+1} \leftarrow$ SGD($\mathbf{F}_*, \mathbf{\Theta}_k, \mathbf{F}_B^{(k)}, \mu, \eta, T_{\max}, \epsilon$) using Alg. 2
7: Set $k \leftarrow k+1$
8: Compute $\mathbf{F}_R^{(K)} = g(\mathbf{\Theta}_K)$ and $\mathbf{F}_B^{(K)}$ from (11)
9: $\mathbf{F}_B^{(K)} \leftarrow \sqrt{N_s}\mathbf{F}_B^{(K)}/\|\mathbf{F}_R^{(K)}\mathbf{F}_B^{(K)}\|_F$
10: **Output:** $\mathbf{F}_R^{(K)}, \mathbf{F}_B^{(K)}$

Algorithm 2 Stochastic Gradient Descent

1: **procedure** SGD($\mathbf{F}_*, \mathbf{\Theta}_k, \mathbf{F}_B^{(k)}, \mu, \eta, T_{\max}, \epsilon$)
2: Set $t \leftarrow 0$, $\varepsilon_t \leftarrow \infty$ and $\mathbf{\Theta}_k^{(t)} \leftarrow \mathbf{\Theta}_k$.
3: **while** $t < T_{\max}$ and $\varepsilon_t > \epsilon$ **do**
4: Draw one sample from $p(\mathbf{S}, 1)$ in (7).
5: Compute the gradients at $\mathbf{\Theta}_k^{(t)} + \mu\mathbf{S}, \mathbf{\Theta}_k^{(t)} - \mu\mathbf{S}$ using (12)
6: Compute $\nabla_{\mathbf{\Theta}} f_\mu(\mathbf{\Theta}_k^{(t)})$ in (10) with $L = 1$
7: Update gradient $\mathbf{\Theta}_k^{(t+1)} \leftarrow \mathbf{\Theta}_k^{(t)} - \eta\nabla_{\mathbf{\Theta}} f_\mu(\mathbf{\Theta}_k^{(t)})$
8: Set $\varepsilon_t \leftarrow \|\mathbf{\Theta}_k^{(t+1)} - \mathbf{\Theta}_k^{(t)}\|_F/\|\mathbf{\Theta}_k^{(t)}\|_F$ and $t \leftarrow t+1$
9: **Output:** $\mathbf{\Theta}_{k+1}$.

V. Simulation Results

In this section we evaluate the performance of HPSAGS against a) MDP [10] in Rayleigh fading channels and b) MDP, SSPOMP [9] in mmWave channels (SSPOMP is not applicable to rich scattering environments). For the MDP we used the adaptive phase increment threshold implementation, which is faster, with parameters $\bar{\delta}^{(1)} = 0.1$ and $K_u = 100$. For all experiments HPSAGS' parameters were set to: $T_{\max} = 100$, $\epsilon = 10^{-4}$ and $\mu_{k+1} = \mu_k/2$, with $\mu_0 = 2.5$ and $K = 7$ using a random initialization $\mathbf{\Theta}_0 \sim \mathcal{U}[-\pi, \pi]$. For the precoder SGD's learning rate is set to $\eta = 10$ and for the combiner $\eta = 600$ and $\eta = 1600$ for the Rayleigh fading and mmWave channels, respectively. Results are averaged over 100 Monte Carlo runs.

In Fig. 1, we compare the average spectral efficiency (SE) of HPSAGS and MDP for a 256×64 MIMO system with $M_t = M_r = 10$ RF chains transmitting $N_s = 8$ data streams in i.i.d. Rayleigh fading channels for different signal-to-noise-ratio (SNR) values. We considered both PCSI ($\alpha = 0$) and ICSI with $\alpha = 0.6$. Fig. 2 demonstrates the performance of HPSAGS and MDP while we vary the number of antennas $N_t = N_r$ with $M_t = M_r = 8$ RF chains transmitting $N_s = 8$ data streams in i.i.d. Rayleigh fading channels at -10 dB SNR. It also illustrates the average processing time of these methods. Finally, in Fig. 3 we compare the average SE of HPSAGS, SSPOMP, and MDP for different SNR values considering a 64×64 MIMO system with $M_t = M_r = 10$ RF chains in mmWave channels with $N_c = 6, N_p = 3, \sigma_\phi = \sigma_\theta = 7.5^\circ$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2020.2969203, IEEE Wireless Communications Letters

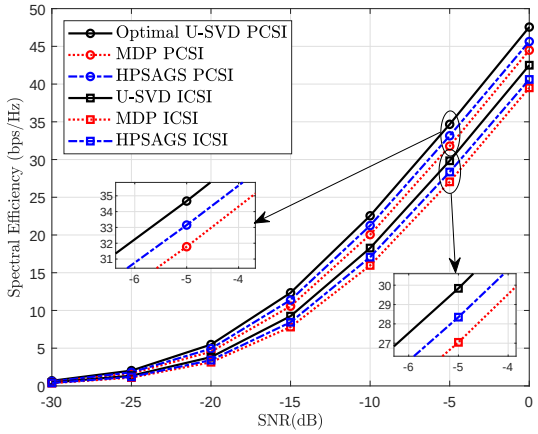

Fig. 1. Achieved SE of a 256×64 MIMO system with $M_t = M_r = 10$ RF chains transmitting $N_s = 8$ data streams in i.i.d. Rayleigh fading channels with PCSI and ICSI with $\alpha = 0.6$.

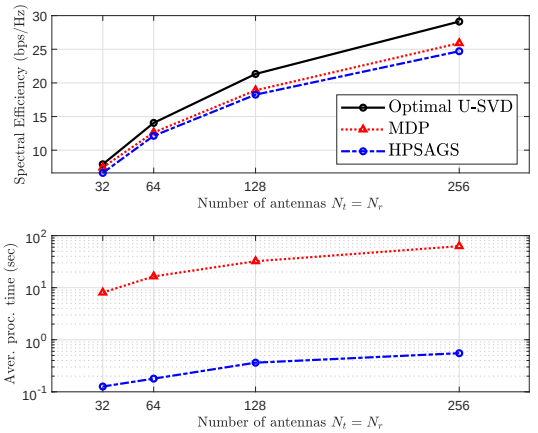

Fig. 2. SE (top) and average processing time (bottom) in log-scale vs the number of antennas in i.i.d. Rayleigh fading channels where $N_s = 8$ data streams are transmitted through $M_t = M_r = 8$ RF chains.

We observed that when the number of data streams is slightly smaller than the number of RF units, the proposed HPSAGS method outperforms MDP for both PCSI and ICSI in i.i.d. Rayleigh fading channels as well as SSPOMP and MDP in mmWave channels. On the other hand, when the number of data streams is equal to the number of RF units, we notice a small decrease in HPSAGS performance, particularly in mmWave channels. However, as observed in Fig. 2, HPSAGS demonstrates dramatically lower average processing time (two orders of magnitude) compared to MDP. It should be noted that the complexity of MDP is $O(NM^3Q)$, where Q is a convergence parameter, and therefore it cannot be directly compared to HPSAGS's complexity. It should also be mentioned that the complexity of SSPOMP is $O(NM^3)$, ranking this method first in terms of computational efficiency in limited scattering environments.

VI. Conclusions

We presented a novel approach to the task of hybrid processing in single-user massive MIMO systems. The derived algorithm for both rich and poor scattering channels demonstrates high spectral efficiency with low computational cost. Future research directions include extending the proposed algorithm to: a) multi-user setups and b) the frequency-selectivity case (exploiting the large available bandwidth at mmWave frequencies).

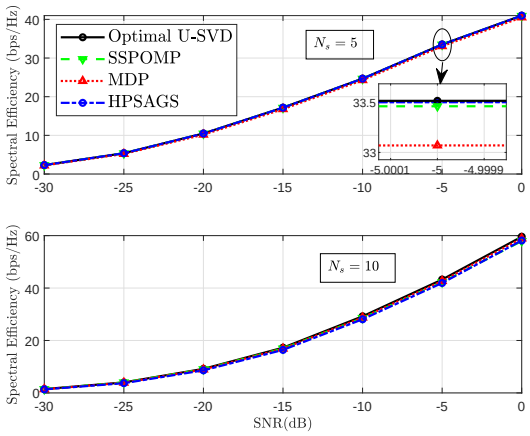

Fig. 3. Achieved SE of a 64×64 MIMO system with $M_t = M_r = 10$ RF chains in mmWave channels for $N_s = 5$ (top) and $N_s = 10$ (bottom) supported data streams.

References

[1] E. G. Larsson *et al.*, "Massive MIMO for Next Generation Wireless Systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[2] M. Shafi *et al.*, "Microwave vs. Millimeter-Wave Propagation Channels: Key Differences and Impact on 5G Cellular Systems," *IEEE Communications Magazine*, vol. 56, pp. 14–20, December 2018.

[3] A. Alkhateeb *et al.*, "MIMO Precoding and Combining Solutions for Millimeter-wave Systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, 2014.

[4] S. Han *et al.*, "Large-scale Antenna Systems with Hybrid Analog and Digital Beamforming for Millimeter Wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[5] S. Payami *et al.*, "Phase shifters versus switches: An energy efficiency perspective on hybrid beamforming," *IEEE Wireless Communications Letters*, vol. 8, pp. 13–16, Feb 2019.

[6] G. K. Papageorgiou, D. Ntaikos, and C. B. Papadias, "Efficient beamforming with multi-active multi-passive antenna arrays," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, June 2018.

[7] R. W. Heath and et al., "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 436–453, April 2016.

[8] C. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave mimo systems," *IEEE Wireless Communications Letters*, vol. 4, pp. 285–288, June 2015.

[9] O. El Ayach *et al.*, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE transactions on wireless communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[10] W. Ni, X. Dong, and W.-S. Lu, "Near-optimal Hybrid Processing for mMassive MIMO Systems via Matrix Decomposition," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3922–3933, 2017.

[11] M. Styblinski and T.-S. Tang, "Experiments in Nonconvex Optimization: Stochastic Approximation with Function Smoothing and Simulated Annealing," *Neural Networks*, vol. 3, no. 4, pp. 467–483, 1990.

[12] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman and Hall/CRC, 2018.

[13] http://github.com/ge99210/Hybrid-Precoding-Combining-. [Online; accessed 19 July 2019].