SNAP – POP for Massive MIMO Systems

Peter J. Smith*, Pawel A. Dmochowski†, Mark Hopper‡, and Harsh Tataria§
*School of Mathematics and Statistics, Victoria University of Wellington, Wellington, New Zealand
†School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand
‡School of Mathematics, University of Bristol, Bristol, U.K.
§Department of Electrical and Information Technology, Lund University, Lund, Sweden
e-mail: peter.smith@vuw.ac.nz, pawel.dmochowski@ecs.vuw.ac.nz, mark.hopper@bristol.ac.uk, and harsh.tataria@eit.lth.se

***Abstract*—Massive multiple-input multiple-output (MIMO) has gained a tremendous amount of attention as a key enabler of fifth-generation cellular systems. The majority of the vast literature on massive MIMO is based on fully *digital* processing, where each antenna element in the massive array is equipped with an up/down-conversion chain. Unlike previous works, we present a suite of greedy algorithms using pure *analog* processing which approach the performance of uplink *digital* zero-forcing (ZF) combining. Based on successive nulling and amplification processing (SNAP), the proposed techniques are evaluated in a multi-user context over a range of *heterogeneous* channel models. Rigorous comparisons are then made to other benchmark analog techniques, digital ZF and an upper bound on analog processing. The most promising approach is shown to have very *low-complexity* and *robustness* against a variety of channel conditions, while providing useful performance gains.**

I. Introduction

It is well understood by now that fifth-generation (5G) cellular systems should provide an order-of-magnitude increase in the average spectral efficiency (1 Gbps and above) relative to existing fourth-generation systems [1]. Phenomenal progress has been made in academia and industry to achieve the aforementioned targets, using concepts such as massive multiple-input multiple-output (MIMO) [2]. On the standardization front, the third generation partnership project have converged on the first global standard for 5G cellular, known as new-radio (NR) release 15 [3]. 5G-NR will operate on *two* frequency bands via the principle of *dual connectivity*. Namely, FR1, which contains bands below 6 GHz, and FR2 which contains higher bands approaching millimeter-wave (mmWave) frequencies. While it is desirable to have massive MIMO in the FR1 bands to provide uniform wide-area coverage, its inclusion is *essential* for FR2 bands to provide large array gain [4]. The potential of massive MIMO is fully exploited when *digital* processing is employed at the base station (BS), where each antenna element has a dedicated up/down-conversion radio-frequency (RF) chain. This facilitates coherent transmission and combining, such that the desired signal power is maximized and multi-user interference is nulled [5].

Nonetheless, fully digital processing is extremely difficult to realize in FR2 bands due to the exponentially increasing power consumption of mixed-signal components and RF interconnect complexity [4]. To this end, the use of *analog-digital* (a.k.a. *hybrid*) processing techniques have gained much attention to reduce the number of RF chains [6]–[8]. Hybrid techniques pioneered in [7] were later strongly motivated by the results in [8] showing that digital processing performance can be achieved with hybrid architectures if the number of up/down-conversion RF chains is *at least twice* the number of data streams. This popular approach is now the subject of numerous overviews in the related literature (see e.g., [4], [6] and references therein). In parallel with the work on hybrid processing schemes, there remains high interest in *purely analog*, phase-only processing (POP) techniques [9]–[13]. Investigations into *uplink* (UL) processing [9], [11], [12] and *downlink* processing [10]–[12] over both simple *Rayleigh fading* [9] and *mmWave channels* [9]–[12] have already been carried out. Due to the presence of only a single RF-chain, one may expect the average spectral efficiency of massive MIMO systems with pure analog processing to be *much less* than the digital counterpart. Nevertheless, the massive reduction in implementation complexity often makes analog processing a suitable choice. In fact, almost all of the current *commercial* 5G-NR street-macro and micro-cellular solutions within the FR2 bands are based on analog processing [14]. *To this end, a major challenge is to overcome the loss in achievable performance, while retaining the low-complexity structure of analog processors*. We are further motivated to consider POP techniques by recent results in [11] and [9]. In [11] it is shown that POP can sometimes result in a useful performance-complexity trade-off. Also, in the high performance region (small numbers of users at moderate to high SNR) it has been shown that purely analog processing can outperform some hybrid schemes [9].

Hence, in this paper, we consider the design of novel POP methods that provide further improvements in the high performance regime. In addition to the use of purely analog techniques, the designed algorithms, based on *successive nulling and amplification processing* (SNAP), are deliberately *sequential* and *greedy* in nature to further *reduce* complexity. We focus on the UL of a multi-user massive MIMO system and compare our novel schemes with typical analog and digital techniques, as well as a general upper bound on analog processing. As is typical in such studies, we also consider the performance loss relative to digital zero-forcing (ZF). Specifically, our contributions are as follows:

- A range of novel POP schemes, SNAP, are designed and presented with varying degrees of complexity, leading to the promising *reverse SNAP* method.
- The SNAP methods are shown to *outperform* typical

analog alternatives and achieve large fractions of the gains offered by optimal analog processing. To the best of our knowledge, this has been missing from the literature.
- The reverse SNAP algorithm is shown to be very low complexity and *robust* to a wide range of heterogeneous channel conditions, while providing useful gains in the high performance regime of massive MIMO systems.

II. System Model

We consider an UL massive MIMO system where K single-antenna users are served by a BS equipped with N antennas ($N \gg K$). The $N \times 1$ received signal at the BS is given by

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{s} = (s_1, s_2, \ldots, s_K)^T$ is the $K \times 1$ vector of transmitted signals, $\mathbf{H}$ is the $N \times K$ composite channel matrix and $\mathbf{n} = (n_1, n_2, \ldots, n_N)^T$ is the $N \times 1$ additive white Gaussian noise vector. Without loss of generality, we assume $\mathbb{E}[|s_i|^2] = 1$, $\mathbb{E}[|n_i|^2] = 1$ and $\mathbb{E}[|h_{ij}|^2] = \beta_j$ is the link gain of user j. Note that h_{ij} denotes the (i,j)-th entry of $\mathbf{H}$. We assume perfect channel knowledge at the BS. At first sight, this assumption may seem impractical. However, there are three reasons for making it: 1) The primary focus of the paper is on proposing analog processing techniques offering superior performance relative to conventional schemes, while retaining low-complexity. Here, in order to fundamentally understand the performance of the proposed method, perfect channel state information is required. 2) Obtaining a high quality estimate of the channel is a reasonable assumption in low mobility scenarios, where a large fraction of the NR subframe can be used for uplink training. 3) The results obtained from subsequent analysis can be regarded as a useful upper bound on what may be achieved in practice with imperfect channel estimates.

We assume the use of two heterogeneous propagation channel models in the paper. Namely, *semi-correlated Ricean fading* and a *ray-based* model. The correlated Ricean model is defined in the following: Let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$, where $\mathbf{h}_j$ is the channel vector for user j. Each user has its own K-factor, K_j, and correlation matrix, $\mathbf{R}_j$, so that [15]

$$\mathbf{h}_j = \sqrt{\frac{K_j}{K_j+1}}\mathbf{a}(\phi_j) + \sqrt{\frac{1}{1+K_j}}\mathbf{R}_j^{1/2}\mathbf{u}_j. \tag{2}$$

Note that $\mathbf{a}(\phi_j)$ is the dominant line-of-sight (LoS) component corresponding to the steering vector, $\mathbf{a}(\cdot)$, of the BS array and an azimuth AoA for the LoS path given by ϕ_j. Assuming the use of a uniform linear array (ULA) [15]

$$\mathbf{a}(\phi_j) = [1 \;\; e^{j2\pi d(\sin\phi_j)}, \ldots, \; e^{j2\pi(N-1)d\sin(\phi_j)}]^T. \tag{3}$$

Note that in (2), $\mathbf{R}_j$ is the correlation matrix and $\mathbf{u}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. We consider a uniformly distributed placement of scatterers. To this end, we consider a one-ring spatial correlation model for $\mathbf{R}_j$ with central angle ϕ_j and an azimuth angle spread of Δ_j^{or} for user j. Exact details on the one-ring model will be presented later in the paper. In addition to the semi-correlated Ricean fading model, we also consider the ray-based channel description, which is given in [16], and is a three-dimensional (3D) model based on C clusters of scatterers with S sub-paths per-cluster. In such a model, the propagation channel vector for user j is defined as

$$\mathbf{h}_j = \sum_{c=1}^{C}\sum_{s=1}^{S}\gamma_{cs}^j \, \mathbf{a}\,(\phi_{cs}^j, \theta_{cs}^j), \tag{4}$$

where γ_{cs}^j are the coefficients for subpath s in cluster c, ϕ_{cs}^j is the corresponding azimuth AoA and θ_{cs}^j is the elevation AoA. Here, we assume a *vertically orientated* uniform rectangular array (V-URA), where $\mathbf{a}(\cdot)$ is the corresponding steering vector. The ray coefficients are defined by $\gamma_{cs}^j = (\beta_{cs}^j)^{1/2}\exp(j\Phi_{cs}^j)$ where β_{cs}^j is the gain of the ray and Φ_{cs}^j is the phase offset modelled as $\Phi_{cs}^j \sim \mathcal{U}[0, 2\pi]$. Due to the clustered nature of the channel, the AoAs also take a clustered form in $\phi_{cs}^j = \phi_c^j + \Delta_{cs}^j$ and $\theta_{cs}^j = \theta_c^j + \delta_{cs}^j$. Here ϕ_c^j, θ_c^j are the central cluster angles and Δ_{cs}^j, δ_{cs}^j are the subpath offsets, respectively. Further details of the model parameters and distributions are given later on in the paper.

At the BS, linear processing is performed to detect the transmitted data in $s_1, s_2, \ldots, s_K$, using the combiner $\mathbf{w}_j$ for user j. Hence, the output of the combiner for user j is

$$\tilde{\mathbf{s}}_j = \mathbf{w}_j^H\mathbf{r} = \mathbf{w}_j^H\mathbf{h}_j s_j + \sum\nolimits_{\substack{i=1\\ i\neq j}}^{K}\mathbf{w}_j^H\mathbf{h}_i\, s_i + \mathbf{w}_j^H\mathbf{n}, \tag{5}$$

and the resulting signal-to-interference-plus-noise ratio (SINR) for user j can be written as

$$\text{SINR}_j = \frac{|\mathbf{w}_j^H\mathbf{h}_j|^2}{\sum_{\substack{i=1\\ i\neq j}}^{K}|\mathbf{w}_j^H\mathbf{h}_i|^2 + \mathbf{w}_j^H\mathbf{w}_j}. \tag{6}$$

In this paper, we consider POP methods which attempt to achieve SINRs which are close to those given by digital ZF combining. Hence, our target combiner is given by $\mathbf{W}^{\text{ZF}} = [\mathbf{w}_1^{\text{ZF}}, \ldots, \mathbf{w}_K^{\text{ZF}}]$, where $\mathbf{W}^{\text{ZF}} = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$, following the classical structure of the ZF combiner [16].

III. Processing Methods and Bounds

In this section, we define *three* SNAP algorithms, as well as digital ZF and a simple upper bound on POP. Some of the ideas are motivated by the *two-stage* structure of ZF combiner where the received signal is first multiplied by $\mathbf{H}^H$, i.e., performing maximum-ratio combining (MRC), followed by interference cancellation, achieved by $(\mathbf{H}^H\mathbf{H})^{-1}$. The proposed methods are described as follows:

A. Analog MRC

The simplest form of phase only processing is analog MRC [17], or equal-gain combining, where the processing weight vectors use the *phase information* in the channels. Here, $\mathbf{w}_j^{\text{A}} = \tilde{\mathbf{h}}_j$, where the tilde notation indicates keeping the phase component only. Hence, $\tilde{\mathbf{M}} = (\tilde{m}_{ij})$, where $\tilde{m}_{ij} = m_{ij}/|m_{ij}|$ for any matrix, vector or scalar, $\mathbf{M} = (m_{ij})$.

B. Analog ZF

Here, we simply take the phase information *in* the ZF combiner so that the combiner becomes $\mathbf{W}^{\text{A-ZF}} = \tilde{\mathbf{W}}_{\text{ZF}}$ and $\mathbf{w}_j^{\text{A-ZF}}$ is the j-th column of $\mathbf{W}^{\text{A-ZF}}$.

C. SNAP: The Forward Process

This algorithm builds on the interpretation of digital ZF as a *signal boosting process* followed by *interference reduction*. A greedy algorithm is proposed as a reduced complexity approach to implement this process as described below. In this exposition, user j is the *desired* user. For ease of notation, in this section we refer to the elements of $\mathbf{w}_j$ as $w_1, w_2, \ldots, w_N$.

Stage 1: Ordering. In stage 1, the antennas are ordered in *decreasing order of importance from the perspective of signal boosting.* We propose either signal power or per-antenna signal to interference ratio (SIR) as ordering metrics. The signal power at antenna k is defined by $|h_{kj}|^2$, and the SIR at antenna k is given by $|h_{kj}|^2/\sum_{i\neq j}|h_{ki}|^2$. Using these methods, the antennas are ordered from highest to lowest and for ease of notation, we assume that $\mathbf{H}$, is now ordered in this way.

Stage 2: Signal boosting. Here, analog MRC is performed for the first T antennas where T is a threshold which seperates the second and third stages. Hence, the antenna weights are given by $w_k = \tilde{h}_{kj}$ for $k = 1, 2, \ldots, T$. To this end, stage 2 *maximises* the signal power for the T *strongest* antennas. The choice of T is discussed later in the text.

Stage 3: Interference reduction. During stage 2, the signal power has been increased, but no attempt has been made to mitigate interference. Since the interferers will usually have quite different channels, there is no reason to expect that stage 2 will have a deliberate interference inflation effect. However, the interference is likely to grow in a random fashion during stage 2. In this stage, a sequential approach is taken and the remaining $N - T$ antenna weights are chosen to *cancel out* the *largest* interference terms.

After stage 1, the weights $w_1, w_2, \ldots, w_T$ are defined and the output from these antennas delivers the running summation of the signal terms (in column j) as

$$\mathrm{RS}_j = \sum\nolimits_{k=1}^{T} w_k^* h_{kj}. \tag{7}$$

These weights also create running sums of interference terms (in columns $i \neq j$) given by

$$\mathrm{RS}_i = \sum\nolimits_{k=1}^{T} w_k^* h_{ki}. \tag{8}$$

From these running sums we create the matrix $\mathbf{H}_T$, known as a *residual* matrix as it contains the running sums and the remainder of the $\mathbf{H}$ matrix after T weights have been decided. $\mathbf{H}_T$ is defined as

$$\mathbf{H}_T = \begin{bmatrix} \mathrm{RS}_1 & \mathrm{RS}_2 & \ldots & \mathrm{RS}_K \\ h_{T+1,1} & h_{T+1,2} & \ldots & h_{T+1,K} \\ \vdots & \vdots & \ddots & \\ h_{N,1} & h_{N,2} & \ldots & h_{N,K} \end{bmatrix}, \tag{9}$$

and the subscript T denotes that this is the residual matrix after T antenna weights have been allocated. Next, the largest magnitude interference term in $\mathbf{H}_T$ is found. This can be a running summation value, RS_i for $i \neq j$, as this is the interference term created by previous antenna weights. Alternatively, the highest interference term could be an individual channel value, some h_{ri} for $i \neq j$. The highest interference term is then partially cancelled by a phase rotation as described below for the three possible scenarios.

Scenario 1. If RS_i is the largest interference term then we also find the next largest interference term in column i, some h_{ri}. The weight for row r, w_r, is then chosen to cancel with the running sum for column i as below:

$$w_r = -\frac{h_{ri}}{|h_{ri}|}\frac{\mathrm{RS}_i^*}{|\mathrm{RS}_i|} = -\tilde{h}_{ri}\widetilde{\mathrm{RS}}_i^*. \tag{10}$$

The running sums are then updated such that[1]

$$\mathrm{RS}_i = \mathrm{RS}_i + w_r^* h_{ri}, \qquad \forall i \neq j, \tag{11}$$

and a new residual matrix $\mathbf{H}_{T+1}$ is made such that

$$\mathbf{H}_{T+1} = \begin{bmatrix} \mathrm{RS}_1 & \mathrm{RS}_2 & \ldots & \mathrm{RS}_K \\ h_{T+1,1} & h_{T+1,2} & \ldots & h_{T+1,K} \\ \vdots & \vdots & \ddots & \\ h_{r-1,1} & h_{r-1,2} & \ldots & h_{r-1,K} \\ h_{r+1,1} & h_{r+1,2} & \ldots & h_{r+1,K} \\ \vdots & \vdots & \ddots & \\ h_{N,1} & h_{N,2} & \ldots & h_{N,K} \end{bmatrix}. \tag{12}$$

Note that r may be equal to any value from $T + 1$ to N inclusive as $\mathbf{H}_T$ is not ordered in descending order for interference terms.

Scenario 2. If h_{ri} is the largest interference term and RS_i is the second largest term in column i then we proceed as in scenario 1. Again, row r is rotated so that the interference term partially cancels with the running sum for that column. Hence, the weight, w_r, is chosen by (10), the running sums updated using (11) and the residual matrix updated using (12).

Scenario 3. If h_{ri} is the largest interference term in the residual matrix and h_{si} is the second largest interference term in column i then rows r and s are both rotated to partially cancel both each other and the running sum for that column. Hence, the weight for row r is chosen as in scenario 1 as:

$$w_r = -\frac{h_{ri}}{|h_{ri}|}\frac{\mathrm{RS}_i^*}{|\mathrm{RS}_i|} = -\tilde{h}_{ri}\widetilde{\mathrm{RS}}_i^*. \tag{13}$$

This partially cancels the running sum. The weight for row s, w_s, is chosen to cancel the rotated version of h_{ri} such that

$$w_s = -\frac{h_{si}}{|h_{si}|}\frac{h_{ri}w_r^*}{|h_{ri}w_r^*|} = -\tilde{h}_{ri}\tilde{h}_{ri}\tilde{w}_r^*. \tag{14}$$

The running sums are then updated such that

$$\mathrm{RS}_i = \mathrm{RS}_i + w_r^* h_{ri} + w_s^* h_{si} \qquad \forall i \neq j, \tag{15}$$

[1]Note that w_r is chosen to reduce RS_i so there is no guarantee that RS_j for $j \neq i$ will be reduced.

and a new residual matrix $\mathbf{H}_{T+2}$ is made such that

$$\mathbf{H}_{T+2} = \begin{bmatrix} \mathrm{RS}_1 & \mathrm{RS}_2 & \dots & \mathrm{RS}_K \\ h_{T+1,1} & h_{T+1,2} & \dots & h_{T+1,K} \\ \vdots & \vdots & \ddots & \\ h_{r-1,1} & h_{r-1,2} & \dots & h_{r-1,K} \\ h_{r+1,1} & h_{r+1,2} & \dots & h_{r+1,K} \\ \vdots & \vdots & \ddots & \\ h_{s-1,1} & h_{s-1,2} & \dots & h_{s-1,K} \\ h_{s+1,1} & h_{s+1,2} & \dots & h_{s+1,K} \\ \vdots & \vdots & \ddots & \\ h_{N,1} & h_{N,2} & \dots & h_{N,K} \end{bmatrix}. \quad (16)$$

Stage 4: Iteration. The description above begins with T selected weights and a residual matrix of dimension $N-T+1 \times K$ denoted $\mathbf{H}_T$. However, the algorithm holds for any stage of the process where a residual matrix, $\mathbf{H}_t$ for $t \leq N$ is updated to produce $\mathbf{H}_{t+1}$ in scenarios 1 and 2 or $\mathbf{H}_{t+2}$ in scenario 3. Hence, stage 3 is repeated until N weights are selected and the residual matrix has only one row of running sums remaining, $\mathbf{H}_N$. This completes the weight calculation.

The SNAP algorithm requires the value of T, which is essentially the breakpoint between how many antennas to use for boosting the signal, and how many to use for interference cancellation. Two methods are explored here: a fixed value optimized to maximize the mean SINR according to the statistics of the channel (fixed) and the SINR optimal breakpoint achieved by an exhaustive search of T values in $\{1, 2, \dots N\}$.

We show in Sec. IV that the optimal choice of T delivers useful gains over the fixed value. Hence, to avoid the complexity of the exhaustive search, it is desirable to develop a sequential approach to setting T, which gives performance close to the optimal choice. Selecting T this way with the forward process is difficult, as at any stage of the iteration in stage 4, the future antenna weights are *unknown* and hence the potential for future nulling is hard to compute without running the full algorithm. Thus, we consider the reverse process, which is more amenable to an adaptive choice of T.

D. SNAP: The Reverse Process

It is trivial to set all N weights using analog MRC. This performs full signal boosting. Then, we start at the bottom of the ordered $\mathbf{H}$ matrix and work upwards, nulling interference at $N-T$ antennas. Here, it is easier to find a reasonable value of T in a sequential manner as the future weights are already set. The algorithm is below.

Stage 1: Ordering. Ordering is done as in stage 1 for the forward process.

Stage 2: Signal boosting. As for the forward process, analog MRC is used. Initial values of all weights are set in this way, so that $w_k = \tilde{h}_{kj}$ for $k = 1, 2, \dots, N$. The SINR of this weight selection is stored in SINR_1. Also, the running sums of interference values corresponding to the first $N-1$ weights are stored in the vector, $\mathbf{RS} = (\mathrm{RS}_1, \dots \mathrm{RS}_{j-1}, \mathrm{RS}_{j+1}, \dots \mathrm{RS}_K)$ where $\mathrm{RS}_i = \sum_{k=1}^{N-1} w_k^* h_{ki}$.

1: **procedure** REVERSE-SNAP($\mathbf{H}$)
2: $k = 0$
3: **while** stopping rule not satisfied **do**
4: $k = k + 1$
5: $c = \underset{i \neq j}{\operatorname{argmax}} \mathrm{RS}_i$ ▷ Find column with largest RS
6: $w_{N-k+1} = -\tilde{h}_{N-k+1,c} \widetilde{\mathrm{RS}}_c^*$ ▷ cancel column c
7: $\mathbf{RS} = \mathbf{RS} + w_{N-k+1}^* \mathbf{H}_{N-k+1,:} - w_{N-k}^* \mathbf{H}_{N-k,:}$
8: Compute SINR_{k+1} using (6) and updated weights
9: Store weights if SINR_{k+1} is the current maximum
10: Compute stopping rule
11: **end while**
12: **return** $w_1, w_2, \dots w_N$
13: **end procedure**

Fig. 1. SNAP algorithm using the reverse process

Stage 3: Interference reduction. The idea is to rotate rows of the $\mathbf{H}$ matrix to partially cancel the largest running sums of interference as in scenario 1 of the forward process. This process continues until some SINR condition is satisfied (the stopping rule). The iteration is described in Fig. 1 where $\mathbf{H}_{r,:}$ represents row r of $\mathbf{H}$.

The stopping rule. The stopping rule can be as simple as $k = N$ which allows the iteration to run N times and guarantees the maximum SINR from the reverse process. However, a large number of these iterations can be avoided if the iteration is stopped after a local maxima in the SINR is reached. We adopt the simple rule where the process stops after 5 successive SINRs have been below the previous peak value. Hence, the algorithm stops after k^* iterations where

$$k^* = \min\{k + 5 | \mathrm{SINR}_k > \mathrm{SNR}_{k+i},\, i = 1, \dots, 5\}, \quad (17)$$

returning the weights which give SINR_k.

Computational complexity. The reverse nulling procedure is low complexity as all operations, except for the sort in step 5, can be achieved via simple updates. Ignoring addition, subtraction and evaluating the amplitude and phase of a complex number, the computations required are itemized below. Step 5 requires a sort of $K-1$ numbers. Steps 6 and 7 require 3 complex multiplications. Step 8 requires $K+1$ complex multiplications, K real multiplications and one real division if the SINR calculation is performed recursively. For example, at stage 6 only a single weight, w_{N-k+1}, is being updated. Hence, in order to update the $\mathbf{w}_j^H \mathbf{h}_i$ terms in the numerator and denominator of (6) it suffices to remove the old term, $w_{N-k+1}^{\text{old}} h_{N-k+1,i}$, and replace it with the new term, $w_{N-k+1}^{\text{new}} h_{N-k+1,i}$. This involves one complex multiplication. K updates of this form are required for the $\mathbf{w}_j^H \mathbf{h}_i$ terms and one for the $\mathbf{w}_j^H \mathbf{w}_j$ term. Finally, K squares and one division are required and (6) is obtained.

E. POP Bound

Analog MRC is the optimal POP for boosting the desired signal power. If it is assumed that analog MRC also completely removes the interference then a simple upper bound on POP

follows. Hence, any POP method has an SINR less than or equal to $\text{SINR}_j^{\text{ub}}$, where

$$\text{SINR}_j^{\text{ub}} = \left|\tilde{\mathbf{h}}_j^H \mathbf{h}_j\right|^2 \left(\tilde{\mathbf{h}}_j^H \tilde{\mathbf{h}}_j\right)^{-1} = \frac{1}{N}\left(\sum_{r=1}^{N} |h_{jr}|\right)^2, \quad (18)$$

using $\mathbf{h}_j = (h_{j1}, h_{j2}, \ldots, h_{jN})^T$ and $\tilde{\mathbf{h}}_j^H \tilde{\mathbf{h}}_j = N$. Note that (18) follows simply from substituting $\mathbf{w}_j^{\text{A}} = \tilde{\mathbf{h}}_j$ into (6) and setting the interference to zero.

IV. Numerical Results

Figures 2-4 compare the SINR cumulative distribution functions (CDFs) for the processing methods presented in Sec. III for different types of channel models, namely iid Rayleigh, correlated Ricean and ray-based models, respectively. In all cases the system is parameterized by selecting the link gains or ray coefficients so that the digital ZF benchmark achieves a median SNR of 20 dB. The figures show the three types of forward SNAP processing discussed in Sec. III-C. These methods are identified by the type of breakpoint and ordering used. Specifically, the techniques are forward SNAP with a fixed breakpoint value with signal power and SIR ordering (denoted by *Forward SNAP (O1-fixed)* and *Forward SNAP (O2-fixed)*, respectively), and with an optimal breakpoint value (denoted by *Forward SNAP (optimal)*). The figures also plot the SINR CDFs of the reverse SNAP algorithm of Sec. III-D. These are compared to standard analog MRC and analog ZF, described in Secs. III-A and III-B and to the phase only bound (Sec. III-E). For simplicity, all users are assumed to have equal receive power. In all figures, we assume $K = 4$ users are present and $N = 100$. Fig. 2 shows a performance comparison for an iid Rayleigh fading channel. Considering the median values, we see that all of the proposed SNAP algorithms attain a 5 dB improvement over analog MRC, while a 6 dB improvement is a (usually non-achievable) theoretical upper bound. Out of the four SNAP variants, the optimum threshold naturally achieves the best forward SNAP performance. It is promising that the reverse SNAP approach is bracketed by forward SNAP (fixed) at low SNR and forward SNAP (optimal) for high SINR. Computationally, reverse SNAP is trivial so that it's effectiveness relative to forward SNAP and the upper bound make it an attractive implementation. We note that the upper bound may be loose as analog MRC does not null the interference, as is clearly shown in Fig. 2. Hence, the proximity of the SNAP techniques to the upper bound is suggestive of high performance. The small improvements offered by the forward SNAP (optimal) approach are due to the greater flexibility of the forward approach which is able to cancel individual interferers with each other as well as with running sums. Although this can also be implemented in the reverse scheme, the idea of the reverse SNAP method is to achieve low-complexity, hence this is not been considered.

In Fig. 3 a correlated Ricean channel is considered. The K-factor selected is $K_j = 0$ dB $\forall j$ such that half the channel power is LoS and half is scattered. The correlation matrices are chosen independently for each user using a one-ring model

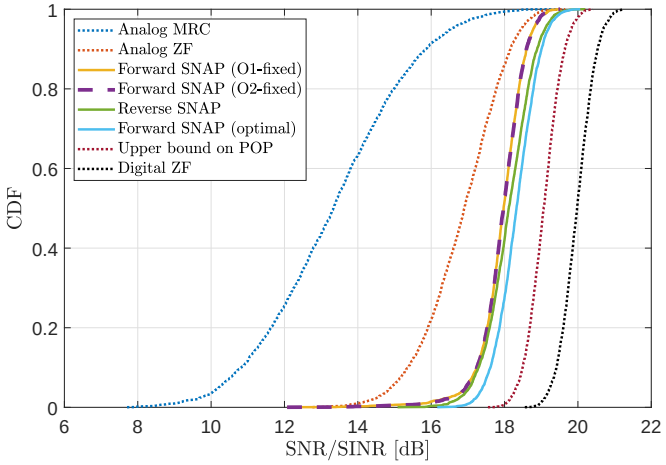

Fig. 2. SNAP performance comparison for i.i.d. Rayleigh fading.

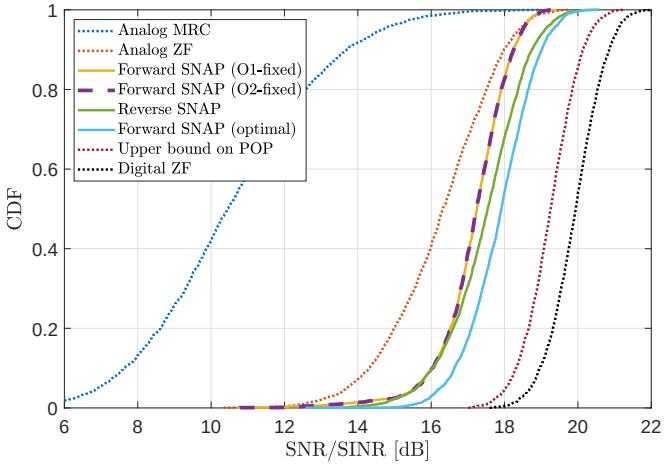

Fig. 3. SNAP performance comparison for correlated Ricean fading.

with a central angle $\phi_j \sim \mathcal{U}[-30°, 30°]$ and angle spread $\Delta_j^{\text{or}} = 30°$. Similar trends are observed as for iid Rayleigh, with a greater gap between the analog techniques and digital ZF. The SNAP techniques are spread out further behind the digital ZF as intuitively the SNAP methodology is enhanced by diversity, so that channels can be rotated to cancel out dominant interference terms. In a strong LoS environment, all channel coefficients have similar magnitudes, so there are no dominant interferers and this makes the nulling process more difficult. Nevertheless, the SNAP techniques are achieving over 80% of the gains offered by the upper bound over analog MRC and up to 50% of the gains offered by the upper bound over analog ZF. Fig. 4 shows the equivalent results for a ray-based channel (adopting the model in [16]). We assume a 10×10 V-URA with half-wavelength spacing along each axis. We consider $C = 20$ clusters, each with $S = 20$ subrays. In the azimuth domain, the central cluster angles satisfy $\phi_c^j \sim \mathcal{N}(0, (20°)^2)$ with Laplacian distributed subray offsets, $\Delta_{cs}^j \sim \mathcal{L}(0, (5°)^2)$. In the elevation domain, the central cluster angles, $\theta_c^j \sim \mathcal{L}(0, (5°)^2)$ with offsets $\delta_{cs}^j \sim \mathcal{L}(0, (2°)^2)$. The angle spreads used are in the range suggested by several measurement campaigns, eg. [18]. As expected from a ray-based model with a finite number of paths, the results are more similar to the correlated Ricean channel than the iid Rayleigh.

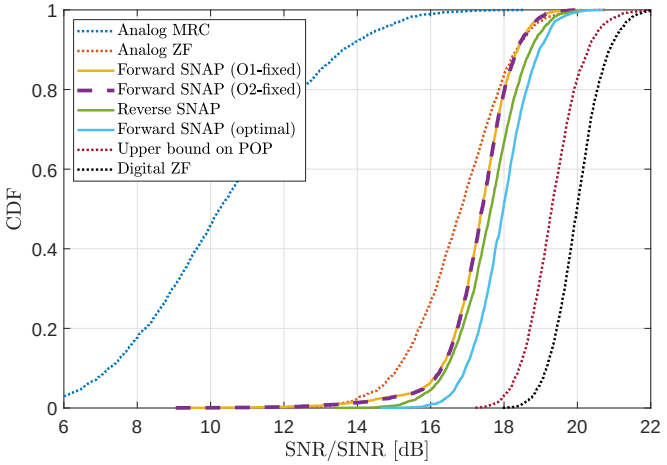

Fig. 4. SNAP performance comparison for a ray-based channel model and a vertically oriented uniform rectangular array.

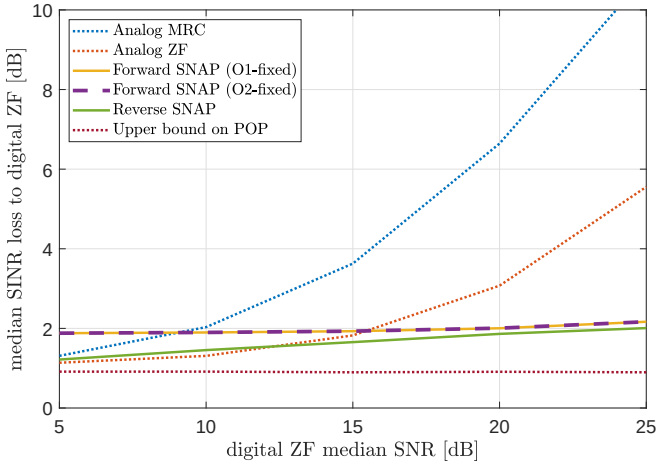

Fig. 5. Median performance loss of the SNAP techniques relative to digital ZF for a ramge of output SNR values.

The SNAP improvements over analog MRC and analog ZF for a V-URA are useful as recent work has shown that the V-URA is more prone to performance degradation with typical antenna spacing compared to other typical antenna topologies. Note that for all channels, the loss of reverse SNAP relative to the optimal forward SNAP is small, suggesting that reverse SNAP is close to optimal for these kind of sequential processing methods based on signal boosting and interference nulling.

Finally, Fig. 5 examines the performance for different levels of signal strength for iid Rayleigh fading. Specifically, we show the median SINR performance gaps of the different techniques from the digital ZF benchmark as a function of the median digital ZF SNR used to set the link gains in the system. We observe that while analog MRC and analog ZF result in a steeply increasing SINR penalty with increasing SNR, the SNAP algorithms (forward and reverse) exhibit an almost constant, small, performance drop. Fig. 5 suggests that the proposed algorithms are particularly well suited for moderate to high SNR. While the results are shown for iid Rayleigh fading, similar trends are observed for correlated Ricean and ray-based channel models. The key factor here is the overall performance. When the output SINRs are over 10 dB, the SNAP algorithms can offer good improvements. When the output SINRs are lower, then digital ZF is slightly better. This suggests a simple hybrid approach where either SNAP or analog ZF is used, and the algorithm switches between the two according to the higher SINR.

V. Conclusion

For medium-to-high SINRs, the purely analog SNAP techniques have been shown to offer useful performance improvements over the benchmark analog processing schemes, analog MRC and analog ZF. Furthermore, the results are close (within 1-2 dB for the simulated scenarios) to an upper bound on analog processing which is likely to be unachievable. In particular, the reverse SNAP method offers these performance gains with very low complexity. A comparison of reverse SNAP with several variants of forward SNAP suggests that reverse SNAP is close to optimal for similar analog techniques based on successive signal boosting and interference nulling.

References

[1] J. G. Andrews, *et al.*, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun., 2014.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[3] 3GPP TR 21.95 v1.1.0, "Technical specification group services and system aspects for release 15," Mar. 2019.

[4] E. Bjornson, *et al.*, "Massive MIMO in sub-6 GHz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100-108, Apr. 2019.

[5] E. G. Larsson, *et al.*, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[6] A. F. Molisch, *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134-141, Sep. 2017.

[7] X. Zhang, *et al.*, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091-4103, Nov. 2005.

[8] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Signal Process.*, vol. 10, no. 3, pp. 501-513, Apr. 2016.

[9] J. Du, *et al.*, "Rethinking uplink hybrid processing: When is pure analog processing suggested?," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5139-5144, May 2019.

[10] J. Wang, *et al.*, "On the performance of beam allocation based multi-user massive MIMO systems," in *Proc. IEEE ICC*, May 2019, pp. 1-6.

[11] S. Buzzi and C. D'Andrea, "Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmWave systems," *IEEE Trans. Green Commun. and Netw.*, vol. 2, no. 2, pp. 385-396, June 2018.

[12] Y. Sun and C. Qi, "Analog beamforming and combining based on codebooks in millimeter wave massive MIMO communications," in *Proc. IEEE GLOBECOM*, Dec. 2017.

[13] Y. Sun and C. Qi, "Weighted sum-rate maximization for analog beamforming andcombining in millimeter wave massive MIMO communications," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1883-1886, Aug. 2017.

[14] Ericsson, "Radio system portfolio for next generation access networks," Sep. 2019.

[15] H. Tataria, *et al.*, "Impact of line-of-sight and unequal spatial correlation on uplink MU-MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 634-637, Oct. 2017.

[16] C. L. Miller, *et al.*, "Analytical framework fo full-dimensional massive MIMO with ray-based channels," *IEEE J. Sel. Topics in Signal Proces.*, vol. 13, no. 5, pp. 1181-1195, Sep. 2019.

[17] S. Li, *et al.*, "Analysis of analog and digital MRC for distributed and centralized MU-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1948-1952, Feb. 2019.

[18] S. Sangodoyin, *et al.*, "Cluster characterization of 3-D MIMO propagation channel in an urban macro-cellular environment," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5076-5091, Aug. 2018.