ABSTRACT
We consider the problem of designing hybrid analog-digital beamformers in a downlink multi-user large-scale MIMO system. The objective is to minimize the total transmit power, while fulfilling SINR targets of all users. A dual virtual uplink problem is formulated for the original downlink problem based on the uplink-downlink duality theory, in order to decouple the digital beamformers in the constraints. Furthermore, an optimal method and a sub-optimal iterative method are devised to compute solutions of the hybrid beamforming problem. Simulation results demonstrate that the iterative method yields nearly optimal performance, despite its remarkably low complexity.

Index Terms—Hybrid beamforming, uplink-downlink duality, large-scale MIMO, power iteration method

1. INTRODUCTION
A large-scale multiple-input multiple-output (MIMO) system, with tens and hundreds of antennas at its base station (BS), is capable of producing extremely narrow and high-gain beams to simultaneously serve multiple users with very high data rate [1–4]. The new technology, however, comes with its own challenges. One of the them is the difficulty to design efficient digital baseband beamformers at the BS. Unlike in conventional digital baseband beamforming systems [5, 6], having a distinct radio frequency (RF) chain for each antenna in the large-scale MIMO systems is inefficient due to the high costs and the large power consumptions associated with it [7]. To overcome the problem, a hybrid analog-digital beamforming technique is proposed in the literature, which comprises both digital and analog beamforming [8–12]. In this technique, as illustrated in Fig. 1, digital beamforming is performed on the transmit signals, mixed and converted into analog domain using only a fraction of the RF chains of a fully-digital large-scale MIMO system. Subsequently, analog beamforming is carried out on the output of the RF chains using inexpensive phase shifters. Electronically switched phase shifters are employed in many systems to further reduce the power consumption of the RF electrical components [13, 14].

Another important aspect in mobile communication besides offering an immense data rate is to achieve green communication [15, 16], which is accomplished by maximizing the network’s energy efficiency in order to reduce undesirable impact of the technology on human health and the environment. To this end, it is judicious to incorporate the energy minimization aspects while engineering new hybrid beamforming techniques.

In this paper, we consider joint digital beamformer (DBF) and analog beamformer (ABF) design in a multi-user downlink (DL) system. An optimization problem is formulated in Sec. 3 to minimize the total transmit power, and simultaneously meet the signal-to-interference-plus-noise ratio (SINR) requirements of all users. To aid the employment of inexpensive and power-efficient discrete phase shifters, the ABFs are restricted to a predefined codebook. As the beamformers are coupled through co-channel interference, the problem is difficult to solve. In order to simplify the problem, we employ the well-known uplink-downlink duality theory [17–21] and develop a dual virtual uplink (VUL) problem, in which the DBFs are decoupled. Moreover, an optimal method and a sub-optimal iterative method are proposed to solve the VUL problem. Besides, performance of all variants of the proposed iterative method are compared with the optimal solution in Sec. 6.

2. SYSTEM MODEL
We consider a co-channel multi-user DL system with a BS equipped with \(N\) antennas, and \(K\) single antenna users. In each symbol period the BS transmits \(K\) mutually independent symbols, one for each user. Let \(s = [s_1, s_2, \ldots, s_K]^T\) denote the transmit symbol vector, \(p_k = |s_k|^2\) denote the power on the \(k\)th symbol, and \(p = [p_1, p_2, \ldots, p_K]^T\) be.
the transmit power vector. The BS is equipped with \( M \) RF chains, such that \( K \leq M \leq N \). The normalized DBF \( b_k = [b_{1k}, b_{2k}, \ldots, b_{MK}]^T \), with \( \|b_k\| = 1 \), is applied to the \( k \)th transmit symbol \( s_k, \forall k \in K \triangleq \{1, 2, \ldots, K\} \).

Each RF chain is connected to all \( N \) transmit antennas through discrete phase shifters. The predefined ABF codebook is denoted by \( D \). It consists of \( L \geq M \) unit-norm orthogonal beamformers with constant modulus elements, i.e., \( D \triangleq \{d_1, d_2, \ldots, d_L\}, \) with \( \|d_k\| = 1, \forall \ell \in L \triangleq \{1, 2, \ldots, L\}, \) and \( d_k^H d_m = 0 \) for \( \ell \neq m \). Let \( a_m \in D, \forall m \in M \triangleq \{1, 2, \ldots, M\} \) be the \( m \)th ABF. The ABFs are stacked in ABF matrix \( A = [a_1, a_2, \ldots, a_M] \).

The frequency-flat channel vector between the BS and the \( k \)th user is represented by \( b_k^{1 \times N} \). It is assumed that the complete channel state information is available at the BS [9, 10]. The received signal \( y_k \) at the \( k \)th user can be expressed as

\[
y_k = h_k A b_k s_k + \sum_{j=1, j \neq k}^K h_k A b_j s_j + n_k, \quad (1)
\]

where \( n_k \sim CN(0, \sigma^2) \) represents the additive white Gaussian noise (AWGN) at the \( k \)th user. The minimum SINR requirement of the \( k \)th user is denoted by \( \gamma_k \).

### 3. PROBLEM FORMULATION

We formulate, in this section, an optimization problem to minimize the total transmit power in the DL system. It is followed by, the formulation of a dual VUL problem.

Let \( P \) denote the total transmit power at the BS over all transmit symbols. Exploiting the fact that \( A^H A = I \) and \( \|b_k\| = 1 \), \( P \) can be written as

\[
P = \sum_{k=1}^K s_k^H b_k^H A^H A b_k s_k = \sum_{k=1}^K p_k \quad (2)
\]

Accordingly, the DL problem of minimizing the total transmit power while satisfying the SINR targets of users is given by

\[
\begin{align*}
\text{minimize} \quad & \sum_{k=1}^K p_k \\
\text{s.t.} \quad & p_k b_k^H A^H R_k A b_k \geq \gamma_k, \forall k \in K, \\
& p_k \geq 0, \forall k \in K, \\
& \|b_k\| = 1, \forall k \in K; \quad a_m \in D, \forall m \in M,
\end{align*}
\]

where \( R_k = h_k^H h_k / \sigma^2 \). The constrains in (3b) enforces the SINR at the \( k \)th user to the corresponding SINR target. The assumptions on DBFs and ABFs are administered by constrains in (3d).

However, problem (3) is a combinatorial problem. The coupling of the DBFs in constraints (3b) makes the problem even harder. To decouple the DBFs, we adopt uplink-downlink duality theory [19–23] and formulate a dual VUL problem as

\[
\begin{align*}
\begin{array}{l}
\text{minimize} \\
\text{s.t.}
\end{array}
\frac{q_k b_k^H A^H R_k A b_k}{q_k (\sum_{j=1, j \neq k}^K q_j A^H R_j A + I) b_k} \geq \gamma_k, \forall k \in K, \\
q_k \geq 0, \forall k \in K, \\
\|b_k\| = 1, \forall k \in K; \quad a_m \in D, \forall m \in M,
\end{align*}
\]

where \( q_k \) denotes the VUL transmit power of the \( k \)th user, and \( I \) denotes identity matrix. Accordingly, the original problem (3) and the derived problem (4) are empowered with the following properties [19–23]:

**(P1)** The DL problem (3) is feasible if and only if the VUL problem (4) is feasible.

**(P2)** The optimal ABF matrix \( A^* \) and the optimal DBFs \( \{b_k, \forall k \in K\} \) of the VUL problem (4) are optimal for the DL problem (3) as well.

**(P3)** The sum of the optimal transmit powers of the DL problem is same as that of the VUL problem, i.e., \( 1^T p^* = 1^T q^* \), where \( q^* \triangleq [q_1^*, q_2^*, \ldots, q_K^*]^T \) is the optimal VUL transmit power vector. Moreover \( q^* \) and \( p^* \) can be expressed as

\[
\begin{align*}
q^* &= (F - G^T)^{-1} 1, \\
p^* &= (F - G)^{-1} 1,
\end{align*}
\]

where \( 1 = [1, \ldots, 1]^T \) and

\[
\begin{align*}
[F]_{kj} &= \begin{cases} 
\frac{1}{\gamma_k} b_k^H A^* R_k A^* b_k^*, & \text{if } k = j, \\
0, & \text{otherwise,}
\end{cases} \\
[G]_{kj} &= \begin{cases} 
0, & \text{if } k = j, \\
b_j^H A^* R_k A^* b_j^*, & \text{otherwise,}
\end{cases}
\end{align*}
\]

As a consequence, we can obtain the optimal ABFs, DBFs, and transmit powers of the DL problem (3) by solving the VUL problem (4).

### 4. EXHAUSTIVE SEARCH METHOD

The exhaustive search method (ESM) computes the optimal solution of the problem (4) in two stages. In the first stage, all distinct ABF matrices \( A_i, \forall i \in I \triangleq \{1, 2, \ldots, L\}/(L - M)!/M! \), are obtained from the dictionary \( D \). In the second stage, for each \( A = A_i, \forall i \in I \), the power iteration (PI) method [19–23] is adopted to compute the corresponding optimal DBFs and transmit powers. Finally, the ABF matrix that reaps the smallest total transmit power, and the corresponding DBFs are selected as the optimal beamformers.

The PI method constitutes two steps, namely, 1) DBF update, 2) power update, which are performed sequentially and iteratively. The method starts with initialization of the VUL
transmit power vector \( q(0) \) to any positive random vector. The method runs until \( |1^T q(t - 1) - 1^T q(t)| \leq \delta \), where \( \delta \) denotes the prescribed numerical accuracy. The \( t \)th iteration of the PI method proceeds as follows:

**Step 1) DBF update:** Using the VUL transmit power vector of the \((t - 1)\)th iteration \( q(t - 1) \) and the selected ABF matrix \( A_i \), the largest eigenvalue \( \lambda_k \) and the corresponding principal eigenvector \( u_k \) are computed \( \forall k \in K \), by solving the following generalized eigenvalue (GEV) problem:

\[
q_k(t - 1)A_i^H R_k A_i u_k = \\
\lambda_k \gamma_k \left( \sum_{j=1, j \neq k}^K q_j(t - 1)A_i^H R_j A_j + 1 \right) u_k. \tag{9}
\]

Moreover, the DBF of the \( t \)th iteration \( b_k(t) \) is obtained by normalizing the principal eigenvector, i.e., \( b_k(t) = u_k/|u_k|_2 \).

**Step 2) Power update:** The VUL transmit power of the \( k \)th user is updated as

\[
q_k(t) = q_k(t - 1)/\lambda_k, \quad \forall k \in K. \tag{10}
\]

Furthermore, efficient iteration pruning techniques can be employed, based on the updated transmit power values, to speed-up the method without any loss of performance.

Nevertheless, the computational complexity of the ESM grows exponentially with \( L - M \) and thus, may not be suitable in practical scenarios.

### 5. LOW-COMPLEXITY HYBRID BEAMFORMING

In this section, we introduce a sub-optimal low-complexity hybrid beamforming (LOBE) method. Unlike the ESM, in which the ABF matrix is obtained in advance of the PI method, in the LOBE method the ABF matrix is updated within the PI method.

The initialization and termination of the LOBE method are carried out similar to the ESM. The \( t \)th iteration of the LOBE method can be summarized as follows:

**ABF update:** Using \( q(t - 1) \), an ABF matrix \( A(t) \) is obtained by employing one of the three analog beamformers selection algorithms proposed in Sec. 5.1. If the ABF matrices do not change for a predefined number of iterations, the ABFs are fixed for all future iterations.

**DBF update and power update:** Using \( q(t - 1) \) and \( A(t) \), the new DBFs \( b_k(t), \forall k \in K \), and VUL transmit power vector \( q(t) \) are computed using Eq. (9) and Eq. (10), as explained in the previous section.

#### 5.1. Analog Beamformers Selection Algorithms

The ABF matrix \( A(t) \) is obtained for the LOBE method employing one of the following three algorithms:

**(A1) Deflation algorithm:** This algorithm starts with all ABFs of the codebook \( D \). In every iteration an ABF is discarded from the codebook, whose removal causes the smallest increase in the total VUL transmit power \( 1^T q \) compared to the removal of any other remaining ABFs. The algorithm continues until \( L - M \) ABFs are discarded. Alg. 1 summarizes the algorithm, where \( D_{-\ell} \) denotes the ABF matrix obtained by discarding the \( \ell \)th ABF from \( D \), and \( |D| \) denotes the number of ABFs in \( D \).

**Algorithm 1: Deflation Algorithm**

1: Initialize \( \bar{D} \leftarrow D; \) \( \bar{q} \leftarrow q(t - 1) \)
2: for \( y \in \{1, 2, \ldots, L - M\} \) do
3: for \( \ell \in \{1, 2, \ldots, |D|\} \) do
4: \( \bar{A} \leftarrow \bar{D}_{-\ell} \)
5: \( \bar{q}(\ell) \leftarrow \) solve Eq. (9) & Eq. (10) using \( \bar{A} \) & \( \bar{q} \)
6: end for
7: \( i \leftarrow \arg\min 1^T q(z) \)
8: \( \bar{D} \leftarrow \bar{D}_{-i}; \) \( \bar{q} \leftarrow \bar{q}(i) \)
9: end for
10: \( A(t) \leftarrow \bar{D}; \) return \( A(t) \)

**(A2) Greedy correction algorithm:** In this algorithm, \( \bar{A} \) is initialized with any \( M \) random ABFs from the codebook \( D \). Then, each ABF of \( \bar{A} \) is exchanged with an ABF from \( D \setminus \bar{A} \), if the new ABF results in a smaller \( 1^T q \) compared to the current ABF and other ABFs of \( D \setminus \bar{A} \). Here, \( D \setminus \bar{A} \) represents the set of ABFs in \( D \) but not in \( \bar{A} \). The algorithm is summarized in Alg. 2, where \( \bar{E} \) represents the set of indices of ABFs in \( D \setminus \bar{A} \).

**Algorithm 2: Greedy Correction Algorithm**

1: Initialize \( \bar{A} \) with any \( M \) random ABFs from \( D \)
2: \( \bar{q} \leftarrow \) solve Eq. (9) & Eq. (10) using \( \bar{A} \) & \( q(t - 1) \)
3: for \( y \in \{1, 2, \ldots, M\} \) do
4: \( a_{curr} \leftarrow \bar{a}_y \)
5: for \( \ell \in \bar{E} \) do
6: \( a_y \leftarrow d_\ell \)
7: \( \bar{q}(\ell) \leftarrow \) solve Eq. (9) & Eq. (10) using \( \bar{A} \) & \( \bar{q} \)
8: end for
9: if \( 1^T \bar{q} \leq 1^T \bar{q}(z) \) then
10: \( \bar{a}_y \leftarrow a_{curr} \)
11: else
12: \( i \leftarrow \arg\min 1^T \bar{q}(z) \)
13: \( a_y \leftarrow d_i; \) \( \bar{q} \leftarrow \bar{q}(i) \)
14: end if
15: end for
16: \( A(t) \leftarrow \bar{A}; \) return \( A(t) \)

**(A3) M-best algorithm:** In this algorithm, the total VUL transmit power \( \bar{q}(\ell) \) is computed for \( \bar{A} = D_{-\ell}, \forall \ell \in \bar{E} \) and \( q(t - 1) \), by solving Eq. (9) and Eq. (10). Subsequently, \( L - M \) ABFs those correspond to the \( L - M \) smallest total VUL transmit power \( 1^T q(\ell) \) are identified and discarded.

3486
In the following the proposed LOBE method with deflation algorithm, greedy-correction algorithm, and M-best algorithm are referred to as LOBE-DA, LOBE-GCA, and LOBE-MBA, respectively.

6. NUMERICAL RESULTS

In this section, the performance of the variants of the LOBE method are compared with that of the optimal ESM. The performance of fully-digital beamforming is also presented for reference. In fully-digital beamforming, $K$ DBFs of length $N$ are constructed using the PI method, supposing $N$ RF chains are available at the BS.

For the simulations, a DL system with $N = 16$ transmit antennas, and $K = 4$ users is considered. Rayleigh fading channels are assumed with zero mean and unit variance. The noise variance at users are normalized to one. A codebook $D$ with $L = 16$ orthogonal ABFs is adopted. The SINR targets of all users are set to be identical. The simulations are carried out over 1000 Monte Carlo runs.

Fig. 2 shows that the total transmit power required by the proposed LOBE-DA and LOBE-GCA are almost identical to the optimal power achieved with the ESM. Even the performance of the LOBE-MBA, whose computational complexity is considerably lower than that of the ESM, is remarkably close to the optimal performance. Fig. 3 suggests that the gap, between the total transmit power required by the variants of LOBE method and the optimal transmit power reduces further as the number of RF chains $M$ increases.

In Table 1 it can be noticed that, just one ABF update is sufficient for LOBE-DA, whereas an average of four ABF updates are adequate for LOBE-GCA and LOBE-MBA. Moreover, the number of DBF updates and power updates required for the proposed iterative method to converge is less than ten. The average CPU time consumed by the various techniques, as listed in Table 2, clearly reveal the huge reduction in computational complexity of the proposed iterative method compared to that of the ESM.

Table 1. Average no. of iterations taken by the variants of LOBE for the convergence ($M = 6$, SINR target = 5 dB).

<table>
<thead>
<tr>
<th></th>
<th>LOBE-DA</th>
<th>LOBE-GCA</th>
<th>LOBE-MBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABF update</td>
<td>1.00</td>
<td>3.54</td>
<td>3.55</td>
</tr>
<tr>
<td>DBF and power update</td>
<td>6.76</td>
<td>8.52</td>
<td>9.60</td>
</tr>
</tbody>
</table>

Table 2. Average CPU time (in seconds) required by variants of LOBE method and ESM ($M = 6$ and SINR target = 5 dB).

<table>
<thead>
<tr>
<th></th>
<th>LOBE-DA</th>
<th>LOBE-GCA</th>
<th>LOBE-MBA</th>
<th>ESM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1539</td>
<td>0.1551</td>
<td>0.0427</td>
<td>50.06</td>
</tr>
</tbody>
</table>

7. CONCLUSION

In this paper, we considered a hybrid analog-digital beamformer design in a multi-user DL system. We formulated an optimization problem to minimize the total transmit power, while fulfilling the SINR targets of all users. The design of the ABFs and the DBFs are coupled due to the co-channel interference in the system, which makes the problem hard to solve directly. By employing the uplink-downlink duality theory, a dual VUL problem is formulated in which the DBFs are decoupled. Furthermore, the brute-force based optimal ESM method and the iterative low-complexity sub-optimal LOBE method are proposed to solve the problem. Simulation results suggest that the variants of the LOBE method achieve almost optimal performance with extremely low complexities.
8. REFERENCES


