ENERGY-EFFICIENT BIT ALLOCATION FOR RESOLUTION-ADAPTIVE ADC IN MULTIUSER LARGE-SCALE MIMO SYSTEMS: GLOBAL OPTIMALITY

Kien-Giang Nguyen*, Quang-Doanh Vu*, Le-Nam Tran†, and Markku Juntti*

* Centre for Wireless Communications, University of Oulu, P.O.Box 4500, FI-90014, Finland; Email: {giang.nguyen, doanh.vu, markku.juntti}@oulu.fi.
† School of Electrical and Electronic Engineering, University College Dublin, Ireland. Email: nam.tran@ucd.ie.

ABSTRACT

We consider uplink multiuser wireless communications systems, where the base station (BS) receiver is equipped with a large-scale antenna array and resolution adaptive analog-to-digital converters (ADCs). The aim is to maximize the energy efficiency (EE) at the BS subject to constraints on the users’ quality-of-service. The approach is to jointly optimize both the number of quantization bits at the ADCs and the on/off modes of the radio frequency (RF) processing chains. The considered problem is a discrete nonlinear program, the optimal solution of which is difficult to find. We develop an efficient algorithm based on the discrete branch-reduce-and-bound (DBRnB) framework. It finds the globally optimal solutions to the problem. In particular, we make some modifications, which significantly improve the convergence performance. The numerical results demonstrate that optimizing jointly the number of quantization bits and on/off mode can achieve remarkable EE gains compared to only optimizing the number of quantization bits.

Index Terms—Wireless communications, resolution-adaptive ADC, large-scale antenna systems, energy efficiency, discrete branch-reduce-and-bound.

1. INTRODUCTION

Large-scale multiple-input multiple-output (MIMO) technology is becoming one of the key enablers for the future wireless communications systems. Therein, a base station (BS) is equipped with a large number of antenna elements providing a great richness of spatial degrees of freedom so that the BS can serve effectively many users at the same time-frequency resource. It has been shown that large-scale MIMO systems can significantly improve spectral and energy efficiency even with simple receive signal processing techniques such as maximum-ratio combining (MRC) [1].

One practical concern in implementing a large-scale MIMO system is the amount of power consumed by analog-to-digital converters (ADCs). In theoretical works, ADCs are often modelled to have infinite resolution. However, it has been found that the consumed power at an ADC grows exponentially with the number of quantization bits [2, 3]. Consequently, in large-scale MIMO systems with a large number of ADCs, the operating power would be enormous with ideal high resolution ADCs. This has motivated to study the use of low resolution ADCs, i.e., ADCs with small numbers of quantization bits [4–12]. A proper use of non-uniform ADC resolutions can improve the performances compared to the uniform one [7,8]. However, optimizing the resolution accuracy often leads to very complex discrete optimization problems.

Energy efficiency (EE) has become an important design target in wireless systems during the last decade [13–15], see [16, 17] for recent overviews. It is known that using low resolution ADCs can improve the EE for large-scale MIMO systems [7,8,11]. However, the potential EE benefits of the non-uniform ADC resolutions has been explored in few special cases only. The EE maximization problem in a single-user system was considered in [7]. In [12], an algorithm determining the ADC resolutions for maximizing EE in point-to-point MIMO system was proposed. Selecting ADC resolutions for minimizing the mean square quantization error was the main focus in [8]. The ADC resolutions for maximizing the EE in multiuser MIMO systems was considered in [10]; all the radio frequency (RF) processing chains were always active. What is more, the proposed algorithms in the literature are heuristic (suboptimal) due to the difficulty of discrete programs.

In this paper, we aim at finding the potential ultimate benefits of the resolution-adaptive ADCs in terms of the EE performance in the uplink multiuser large-scale MIMO systems. We assume that channel state information is available at the BS only, and the BS uses the MRC at the digital combiner. We optimize both the on/off mode and the quantization resolution for each RF chain such that the EE at the BS is maximized, while the minimum data rate requirements for each user are satisfied. The problem is a discrete nonlinear program due to the discrete set of quantization bits. To this end, we propose an efficient globally optimal algorithm based on the discrete branch-reduce-and-bound (DBRnB) framework [18]. More explicitly, we propose some modifications on the branching, reduction, and bounding operations based on the specific structure of the considered problem in order to improve the algorithm’s efficiency. The numerical results demonstrate that optimizing the number of quantization bits and on/off mode of RF chains can improve remarkably the EE of the multiuser large-scale MIMO systems.

2. SYSTEM MODEL AND PROBLEM FORMULATION

Signal Model: We consider an uplink multiuser MIMO system in which a base station equipped with an array of $M$ antennas. It serves $K$ single-antenna users, and assume that $M \gg K$. The $K$ users simultaneously transmit independent data to the BS. Let $p_k$ denote the transmit power at user $k$. Also, let us denote by $h_k \in \mathbb{C}^{M \times 1}$ the channel vector between the BS and user $k$, and define matrix $H \triangleq [h_1, \ldots, h_K]$ which stacks all channel vectors. We assume that...
Table 1. The values of $\beta_i$ for $b_i \leq 5$ [19]

<table>
<thead>
<tr>
<th>$b_i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.3634</td>
<td>0.1175</td>
<td>0.03454</td>
<td>0.009497</td>
<td>0.002499</td>
</tr>
</tbody>
</table>

channel is flat, then the received signal vector at the BS is

$$r = HPs + n$$  \tag{1}$$

where $s \triangleq [s_1, ..., s_K]^T$ is the vector of transmitted symbols, $\mathbb{E}(|s_i|) = 1$; $P \triangleq \text{diag}(\sqrt{p_1}, ..., \sqrt{p_K})$; and $n \sim \mathcal{CN}(0, \sigma^2 I_M)$ is the additive white Gaussian noise (AWGN).

**Quantization Model:** We consider a fully digital receive array processing so that the number of RF processing chains is equal to that of the number of receive antennas $M$. We further assume that each RF chain is equipped with a resolution-adaptive ADC pair, that of the number of receive antennas $M$.

The signal-to-interference-plus-noise ratio (SINR) corresponding to user $k$ is

$$\gamma_k(b_i) = \frac{p_k |h_k^H(\Phi(b_i))|^2 |h_k|^2}{\sum_{j \neq k} |h_j^H(\Phi(b_i))|^2 h_j^2 + |\sigma|^2 |h_k^H(\Phi(b_i))|^2 |h_k|^2 + \Lambda(b_i) |h_k|^2}$$  \tag{5}$$

where $\Lambda(b_i) \triangleq \Phi(b_i) R(b_i) \Phi(b_i)$.

**Power Consumption Model:** We consider the power consumption model at the BS following those in [7, 8, 12], which include the amount of power spent on low-noise amplifiers (LNAs), RF processing functionalities, ADCs, and baseband signal processors. In particular, the amount of power for baseband signal processors denoted by $P_{BB}$ is assumed to be fixed. We suppose that a RF chain can be turn on or off. This is determined via the number of quantization bits, i.e., RF chain $i$ is turn off if $b_i = 0$. Let $g_i(b_i)$ be the amount of power for the LNA, RF process function, and ADC pair in RF chain $i$, then we have

$$g_i(b_i) = \begin{cases} \eta_{\text{rho}} & \text{if } b_i = 0 \\ P_{\text{RF}} + 2c_0 b_i & \text{otherwise} \end{cases}$$  \tag{6}$$

where $P_{\text{RF}} > 0$ is a constant, $c_0 \triangleq c_f$ is a constant depending on $c$, the Walden’s figure-of-merit which represents the energy consumption per conversion step, and $f_s$, the Nyquist sampling frequency. To sum up, the total consumed power at the BS is expressed as

$$P_{BS}(b) = P_{BB} + \sum_{i=1}^M g_i(b_i).$$  \tag{7}$$

**Problem Formulation:** From the above discussion, we can see that the number of quantization bits influences the total consumed power as well as the user data rate. We focus herein on finding the number of quantization bits at ADCs such that the energy efficiency at the BS is maximized under the constraints on the quality of service for each user, which is mathematically written as

$$\begin{array}{ll}
\text{maximize} & \sum_{k=1}^K B \log_2 \left(1 + \gamma_k(b_i)\right) \\
\text{subject to} & B \log_2 \left(1 + \gamma_k(b_i)\right) \geq Q_k, \forall k = 1, ..., K \\
& b_i \in \{0, ..., b_{\text{max}}\}, \forall i = 1, ..., M
\end{array}$$  \tag{8a}$$

where $B$ is the system bandwidth, $Q_k$ is the minimum data rate requirement of user $k$, and $b_{\text{max}}$ is the maximum number of quantization bits at an ADC. Problem (8) is NP-hard since variables $b$ are discrete. In order to explore the potential benefits of the resolution-adaptive ADCs, we develop an algorithm finding globally optimal solution to (8) presented next.

### 3. PROPOSED GLOBALLY OPTIMAL ALGORITHM

The proposed algorithm is developed based on the discrete monotonic optimization (DMO) framework [18]. Here we use the definitions of box, increasing function, and normal cone in [18].

We can see that the objective function of (8) is not monotonic with $b$, thus the current form of (8) is not suitable for applying DMO. As the first step, we equivalently reformulate (8) as

$$\begin{array}{ll}
\text{maximize} & \eta \\
\text{subject to} & \sum_{k=1}^K Br_k - \eta P_{BS}(b) \geq 0 \\
& r_k \leq \log_2 \left(1 + \gamma_k(b_i)\right), \forall k = 1, ..., K \\
& r_k \geq Q_k / B, \forall k = 1, ..., K \\
& b_i \in \{1, ..., b_{\text{max}}\}, \forall i = 1, ..., M
\end{array}$$  \tag{9a}$$

where $\eta$ and $r \triangleq \{r_k\}_{k=1}^K$ are newly introduced variables, which represent the energy efficiency at BS and data rate for each user, respectively. To expose the monotonic representation of (9), let us introduce variables $\vartheta_i = (1 - \beta_i)^2$, for all $i \in \{1, ..., M\}$, and matrix $\Psi(\vartheta) = \text{diag}(\vartheta) = \vartheta \Phi(b_i)^2$, where $\vartheta \triangleq \{\vartheta_i\}$. Note that $\vartheta_i$ is a discrete variable admitting the values in the set $V \triangleq \{0, 1, ..., b_{\text{max}}\}$. The matrix $A(b) \triangleq \Phi(b_i)^2 R(b_i) \Phi(b_i)$.
\{0, (1 - \beta_1)^2, \ldots, (1 - \beta_{\text{max}})^2\}. With these notations, we can write as
\[
\Lambda(b) = (\Phi(b))^3(I - \Phi(b))P = (\Psi(\vartheta))^{3/2}(I - (\Psi(\vartheta))^{1/2})P.
\]

In addition, since \( \vartheta_i \) is in fact a function of \( b_i \), we replace \( g(b_i) \) in (6) by \( g(\vartheta_i) \). i.e. \( g(\vartheta_i) \) belongs to discrete set \( \mathcal{B} \triangleq \{0, g_1, \ldots, g_{\text{max}}\} \) where \( g_i = P_{bb} + 2c_i^2 \). Now we can rewrite (9) as
\begin{align}
\text{maximize} & \quad f_0(x) \\
\text{subject to} & \quad f_k(x) - h_k(x) \leq 0, \quad \forall k = 1, \ldots, K \quad (10a) \\
& \quad f_{k+1}(x) - h_{k+1}(x) \leq 0 \quad (10b)
\end{align}

where \( x \triangleq [\eta, r, \vartheta] \in \mathbb{R}^{K+1+M} \) is the vector of variables and \( \Lambda \) is the corresponding feasible set interfered from (9), \( f_0(x) = \eta, f_k(x) = (2^k - 1) \sum_{i \neq k} p_i |h^i_0| \eta \|h^i_0\|_2^2 + \sigma^2 \|\Psi(\vartheta)\|_2^2 + h^i_0(\Psi(\vartheta))^2 F_h^i, h_k(x) = \left[(2^k - 1) h^i_0(\Psi(\vartheta))^2 F_h^i + p_k h^i_0(\Psi(\vartheta)) h_k^i\right], f_{k+1}(x) = \eta(P_{bb} + g(\vartheta_i)), \text{and } h_{k+1}(x) = \sum_{i=1}^{K} B_{r_i} \).

Note that \( x \) contains mixed continuous-discrete variables where \( x_i \) for \( i \leq K + 1 \) is continuous, and \( x_i \) for \( i > K + 1 \) is discrete.

We can see that all the functions in (10) are monotonic increasing with the variables. Thus, problem (10) is a discrete monotonic optimization program. From now on, we customize the DBRN algorithm introduced in [18] to optimally solve (10). Generally, a DBRN algorithm is an iterative procedure in which there are three basic operations performed at each iteration, namely, branching, reduction, and bounding [18]. The algorithm starts with an initial box which contains the feasible set of the problem. At each iteration, the branching operation divides a box into two smaller boxes; the reduction operation removes the portions of boxes that do not contain an optimal solution; and, the bounding operation improves the upper and lower bounds of objective as well as remove the boxes that do not contain an optimal solution. The procedure continues until a convergence criterion is met.

Let us start by determining the initial box. In particular, let \([x_i, x_i^\text{\#}]\) denote the box containing the feasible set of (10), i.e., \( x_i \in [x_i, x_i^\text{\#}] \Leftrightarrow x_i \leq x_i \leq x_i^\text{\#} \), where \( x_i \in [\eta, r, \vartheta] \) and \( x_i \in [\eta, \vartheta, \vartheta] \) are the vectors consisting of the lower and upper bound values of elements in \( x_i \), respectively. It is obvious that \( \vartheta_i = 0 \) and \( \vartheta_i = (1 - \beta_i(b_{\text{max}})) \); \( \vartheta_i = (Q_k/B_{\text{max}}) \), and \( F = \log_2(1 + \gamma_k(b_{\text{max}})) \); the bounds of \( \eta \) can be determined as
\[
\eta = \frac{\sum_{k=1}^{K} B_{r_k}}{P_{bb} + \sum_{i=1}^{M} g(\vartheta_i)}, \quad \text{and } \eta = \frac{\sum_{k=1}^{K} B_{r_k}}{P_{bb} + \sum_{i=1}^{M} g(\vartheta_i)}.
\]

In what follows, we describe the three basic operations. For the ease of exposition, we use \( B = [p, q] \) to present an arbitrary box which is processed in each iteration of the DBRN procedure, i.e., \( x_i \leq p_i \leq x_i \leq q_i \leq x_i^\text{\#} \) and \( f_0^\text{\#}(B) \) to denote the upper bound of \( \eta \) provided by the bounds of \( B \); \( R_{\text{max}} \) to denote the set of candidate boxes at iteration \( n \) of the algorithm.

**Branching**

At iteration \( n \), a box \( B \in R_{\text{max}} \) is selected to be branched into two new boxes of equal size. In particular, we choose \( B \) such that \( B = \arg \max_{B' \in R_{\text{max}}} f_0^\text{\#}(B') \) to ensure that the upper bound is monotonically decreasing, and then cut at one edge of \( B \) to create two new boxes \( B^{(n)}_1 = [p, q] \) and \( B^{(n)}_2 = [p', q'] \) as
\[
q_j' = \begin{cases} 
q_j - \frac{(q_j - p_j)}{2} & \text{if } j \leq K + 1, \\
\frac{(q_j - (q_j - p_j))}{2} & \text{if } j > K + 1.
\end{cases}
\]

where \( j \) is the edge index, and \( [x]_v \) and \( [x]_v \) denote the operators which, respectively, map \( x \) to the nearest-upper and nearest-lower discrete values in \( v \). Usually, the longest edge of a box is chosen for branching [18]. However, this standard rule is inefficient for (10). Thus we propose the following rule in order to improve the convergence efficiency.

**Proposed Branching Rule:** At first, the algorithm focuses on branching edges corresponding to \( r \) until the gap between \( F \) and \( r \) is sufficiently small. This is inspired by the observation that given the bounds of \( r, \vartheta \), the following inequality must hold for \( B \) to be feasible
\[
\mu(\sum_{i \neq k} p_i |h^i_0| \Psi(\vartheta)^2 h^i_0|_2^2 + \sigma^2 |\Psi(\vartheta)|_2^2 + h^i_0(\Psi(\vartheta))^2 F_h^i - \mu h^i_0(\Psi(\vartheta))^2 F_h^i + p_k h^i_0(\Psi(\vartheta)) h_k^i|_2^2 \leq 0
\]

which is due to the monotonicity of (10b), where \( \mu = 2^k - 1 \). Thus, branching \( r \) and using (14) might help removing the subsets of \( \vartheta \) and \( r \) which cannot be simultaneously satisfied.

After branching on \( r \), the algorithm focuses on \( \vartheta \). We recall that finding optimal \( \vartheta \) consists of two tasks: finding the set of active RF chains (i.e. \( \vartheta_i > 0 \)), and finding the number of quantization bits. We observe that, with standard branching rule, it takes many iterations to reach the optimal decision of turning off some RF chain. Thus, the idea is to first check the on/off mode of a RF chain. To this end, we modify the cuts in (12) and (13) for \( j > K + 1 \) as
\[
q_j = p_j = 0, \quad p_j' = v_1, \quad q_j' = q_j - \frac{p_j - p_j'}{2}, \quad p_j' = p_j + \frac{q_j - p_j}{2}, \quad \text{if } p_j = 0
\]

Finally, we note that we do not need to branch \( \eta \) since the bounds of \( \eta \) can be determined via the bounds of the other variables as in (11).

**Reduction**

Let us consider box \( \hat{B} = [p; q] \). The reduction operation finds smaller box \( B^\prime = [\hat{p}'; \hat{q}'] \subseteq B \) by cutting portions \( [p, \hat{p}'] \) and \( [\hat{q}', q] \) which do not contain optimal solutions. This can be done using [18, Lemma 16] by exploiting the monotonicity of the problem. For the considered problem, we propose the following efficient reduction cut.

**Proposed Reduction:** As discussed above, we can eliminate box \( \hat{B} \) by examining the inequality (14) inspired by the monotonic properties of the problem [18, Lemma 16]. Obviously, if (14) does not hold, then the box is infeasible. In addition, the monotonic properties can be used to update the new upper bound for \( r \). In particular, we recall the rate constraint \( r_k \leq \log_2 \left(1 + \frac{p_k h^i_0(\Psi(\vartheta)) h_k^i_2}{\psi(\vartheta) h^i_0(\Psi(\vartheta)) h_k^i_2 + \sigma^2 |\Psi(\vartheta)|_2^2 + h^i_0(\Psi(\vartheta))^2 F_h^i}\right) \). It is clear that
\[
r_k \leq r_k \leq \log_2 \left(1 + \frac{p_k h^i_0(\Psi(\vartheta)) h_k^i_2}{\psi(\vartheta) h^i_0(\Psi(\vartheta)) h_k^i_2 + \sigma^2 |\Psi(\vartheta)|_2^2 + h^i_0(\Psi(\vartheta))^2 F_h^i}\right)
\]
Algorithm 1 The proposed DBRnB-based algorithm solving (10)

1: **Initialization:** Determine $\mathbf{x}, \mathbf{r}$ and denote $\mathcal{R}_0 = [\mathbf{x}, \mathbf{r}]$.
2: repeat
3:  **Branching:** select $B = [p; q] \in \mathcal{R}_n$, determine cutting edge $j$ by $j = \arg\max_{2 \leq i \leq K + 2M} q_i - p_i$ then create $B_1^{(i)} = [p; q']$ and $B_2^{(i)} = [p'; q]$. Update $\mathcal{R}_n := \mathcal{R}_n \setminus \{B\}$.
4:  **Reduction:** remove $B_k^{(i)}, i = \{1, 2\}$ if violating (14), update $r$ in $B_k^{(i)}$ using (16)
5:  **Bounding:** For each box $B_k^{(i)}, i = \{1, 2\}$, calculate bounds of objective using (11).
6:  Calculate feasible objective $\eta_{\text{feas}}$ by (17), and update CBO $\hat{\eta} = \max\{\hat{\eta}, \eta_{\text{feas}}\}$.
7:  Update $\mathcal{R}_n := \mathcal{R}_n \setminus \{B_k^{(i)}, i = 1, 2 | \eta(B') \leq \hat{\eta}\}$
8: until Convergence

Bounding

We can find the bounds of $\eta$ in box $B'$ as in (11). The bounding is essential, since it declares the convergence, i.e., when $\eta - \hat{\eta} < \epsilon$ for small constant $\epsilon$. Another crucial task in bounding step is to prune boxes which do not contain optimal solutions. To be specific, assuming that we know some feasible points and the largest objective values, named current best objective (CBO) and denoted as $\hat{\eta}$, achieved from these points [18]. If the upper bound of $B'$, denoted by $\overline{\eta}(B')$, holds $\overline{\eta}(B') < \hat{\eta}$, $B'$ cannot provide an optimal solution, and, thus, should be ignored to save the computational resources. As a consequence, searching for a feasible point to update $\hat{\eta}$ can improve the algorithms’ efficiency. To this end, given the bounds of $r$ and $\theta$ in box $B'$, the feasible points can be simply calculated as

$$\eta_{\text{feas}} = \max\left\{ \frac{B_{\text{feas}}}{P_{\text{bb}}} + \sum_{i=1}^{M} g(\theta_i), \frac{B_{\theta_{\text{feas}}}}{P_{\text{bb}}} + \sum_{i=1}^{M} g(\xi) \right\} \quad (17)$$

where $B_{\text{feas}} = \log(1 + \gamma_k(\bar{\theta}))$, $B_{\text{feas}} \geq \xi_k$ and $B_{\theta_{\text{feas}}} = \log(1 + \gamma_k(\bar{\theta}))$, $B_{\theta_{\text{feas}}} \geq \xi_k$. The CBO is then updated as $\hat{\eta} = \max\{\hat{\eta}, \eta_{\text{feas}}\}$.

To sum up, the proposed algorithm optimally solving (8) is outlined in Algorithm 1.

4. NUMERICAL RESULTS

The simulation model is setup as follows. We consider a single cell with the radius of 500 meters. The BS is equipped with $M = 20$ antennas. There are $K = 2$ users randomly located in the cell. The path loss model is $PL_k[\text{dB}] = -128.1 - 37.6 \log_{10}(d_k/1\text{km}) + N(0, 5)$, where $d_k$ is the distance between user $k$ and the BS and $d_k > 100$ meters. The bandwidth is 10 MHz and the noise density is -174 dBm/Hz. The transmit power is set as $p_k = \rho_{\text{trans}}$, where $\rho_{\text{trans}} = -10$. The power consumption model, we assume that the BS $P_{\text{bb}} = 60 \text{mW}$, $P_{\text{bb}} = 200 \text{mW}$, $c_0 = 5 \times 10^{-5}$ [8]. The minimum data rate requirement is $Q_k = 1 \text{bit/s/Hz}$.

In Fig. 1, we compare the convergence performance of our proposed DBRnB procedure (i.e., Algorithm 1) with that of the standard one in [18]. In particular, the figure shows the values of upper and lower bounds of the considered algorithms over a random channel realization. As can be seen, the proposed approach can remarkably reduce the required number of iterations to arrive at the convergence criteria. In terms of average per-iteration run time, we have observed that the proposed method takes 0.008 second for each iteration, which is 10 times faster than the standard procedure.

Fig. 2 depicts the average EE performance versus the maximum number of quantization bits $b_{\text{max}}$. We can see that the proposed algorithm is superior to the baselines schemes. The results demonstrate the benefit of both using resolution-adaptive ADCs and properly turning off the RF chains. In addition, the results show that only considering non-uniform ADCs still achieves better EE performance than uniform ADCs. Finally, for the uniform ADC systems, using the large number of quantizations bits significantly degrades EE.

5. CONCLUSION

We have studied the potential EE performance of the resolution-adaptive ADCs in uplink multiuser large-scale MIMO systems. In particular, we have optimized the number of quantization bits at ADCs as well as the on/off modes of each RF chain with the objective of maximizing EE at the BS and the constraints on QoS for each user. The design problem is a discrete nonlinear program. Thus, we have developed an algorithm based on the generic DBRnB framework, which guarantees to provide a globally optimal solution. Numerical results confirms the effectiveness of the design approach in term of gaining EE for multiuser large-scale MIMO systems. In addition, the proposed algorithm can serve as benchmark in developing low-complexity suboptimal solutions, which will be considered in future work.
6. REFERENCES


