MIMO Receive Antenna Selection via Deep Learning and Greedy Adaptation

Cong Shen
University of Virginia
Charlottesville, VA 22904

Donghao Li

Jing Yang
The Pennsylvania State University
University Park, Pennsylvania 16802

***Abstract*—Computationally efficient optimal solutions for selecting a subset of antennas to maximize the mutual information of a MIMO channel have eluded the practitioners due to its combinatorial nature, and the performance gap is widened with massive MIMO. In this work, recent advances in deep learning are leveraged to develop a deep neural network (DNN) based receive antenna selection solution for a given problem dimension. We detail the neural network structure and evaluate several relevant figures of merit via numerical simulations. This data-driven solution is shown to achieve near optimal mutual information in simple settings, but does not scale naturally with the problem dimension. For the practical scenario where the number of selected antennas is unknown a priori, hybrid greedy solutions are proposed which build on the DNN-based solution for a given dimension and then greedily increase or decrease the number of antennas to approximate the optimal solution of the new problem dimension. Numerical simulations demonstrate the effectiveness of the hybrid solutions.**

I. Introduction

Deploying multiple antennas, especially at the base stations (BS) where the hardware size challenge is minimal compared with at the user equipments (UE) and as a result can accommodate potentially very large number of antennas, has become a reality in contemporary wireless standards and real-world systems [1]. However, the cost and complexity of the transceiver design of (massive) multi-input multi-output (MIMO) systems have become one of the key bottlenecks, which increase substantially with the system dimension [2].

One critical component in the MIMO receiver is the increased number of Radio Frequency (RF) chains – ideally each receive antenna needs a dedicated RF chain to process the analog and digital signals, but this can be prohibitive in terms of energy consumption and hardware cost. One direction in addressing this issue is to reduce the cost for each RF chain. This has led to different solutions such as hybrid analog-digital processing [3], low-resolution Analog-to-Digital Converter (ADC) [4], and low-noise amplifier (LNA) [5]. On the other hand, antenna selection (AS) aims at reducing the number of RF chains by only selecting a subset of antennas to process. This is a feasible alternative for reducing hardware energy and cost, and is the main focus of this work.

The work of C. Shen was partially supported by the US National Science Foundation (NSF) under Grant CNS-2002902. The work of J. Yang was partially supported by the US National Science Foundation (NSF) under Grant CNS-1956276.

The study of AS started almost at the same time as MIMO [6]–[8] but has gained significant interest in recent years due to the proliferation of massive MIMO [9]–[12]. Various performance measures and different algorithms have been proposed in the literature. In this work, we first approach the same AS problem by utilizing the recent advances in deep learning, in particular deep neural networks (DNN). The idea is to treat AS as a multi-class classification problem in machine learning, and then to train a DNN model that learns the nonlinear mapping from the channel conditions to the optimal subset of antennas that maximize the mutual information. This data-driven solution is principally different than the existing AS solutions, which are optimization-based and model-driven. Our approach, on the other hand, is data-driven and leverages the plethora of recent advances in DNN. We focus on the receive antenna selection (RAS) problem and evaluate the performance of DNN-based solutions via numerical experiments, which demonstrate their superior performance.

We then pivot to a practically important but largely overlooked aspect of RAS – the adaptability of the problem dimensions. Almost all of the existing solutions are non-adaptive with the target number of antennas, and require a pre-determined problem dimension in terms of the number of selected antennas. However, in practice we often desire the flexibility of changing the number of selected antennas without re-training the DNN. We propose two simple yet highly effective solutions for the dynamic RAS problem, called `Greedy+` and `Greedy-`, that combine the greedy principle with the proposed DNN solution for a pre-determined problem dimension. The proposed design is highly flexible in that it can support any division between the pre-determined dimension K and the dynamically changing dimension L (both increase and decrease). More importantly, it can accommodate any *arbitrary* solution for the pre-determined RAS problem, with no requirement of its performance. Numerical experiments validate the effectiveness of these solutions, which achieve near-optimal performance while maintaining the required flexibility with respect to L.

The rest of this paper is organized as follows. Section II presents the system model and describes the static receive antenna selection problem. The DNN-based solution is proposed in Section III, and the hybrid greedy solutions are introduced in Section IV. Finally, Section V concludes the paper.

II. System Model

A standard narrowband MIMO wireless channel is considered. The transmitter is equipped with M antennas while the receiver has a total of N antennas. The overall channel matrix $\mathbf{H} \in \mathbb{C}^{N\times M}$ captures the small-scale fading effect of the MIMO wireless channel, with the (i,j)th element $h_{i,j}$ denoting the complex channel coefficient between the jth transmit antenna and the ith receive antenna. A slow-fading channel model is assumed, in which the channel matrix $\mathbf{H}$ remains constant for a duration of T (the channel coherence time) and then changes independently to another value.

With all N receive antennas being utilized, the received signal is given by

$$\mathbf{y} = \sqrt{\frac{P}{M}}\mathbf{H}\mathbf{x} + \mathbf{z}, \tag{1}$$

where P denotes the total transmit power, $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the transmitted signal vector with normalized power, and $\mathbf{z}$ is the noise vector with independent and identically distributed (i.i.d.) complex circularly symmetric Gaussian entries with zero mean and unit variance. The fading distribution of $\mathbf{H}$ is not critical since the antenna selection utilizes the channel realizations, but for simplicity we assume that the elements in $\mathbf{H}$ are i.i.d. complex circularly symmetric Gaussian entries with zero mean and unit variance. In reality this holds when the transmit and receive antennas are sufficiently spaced and the channel is rich-scattering.

This work focuses on *receive antenna selection (RAS)*, where the task is to select K out of the total N antennas ($K \leq N$) at the receiver. The reason for AS could be due to the receiver only has K RF chains, or because of other considerations such as power and cost limitations. We use $\mathcal{S} = \{1, \cdots, N\} \doteq [N]$ to denote the set of all receive antennas, and S_K to denote the set of indices of the selected receive antennas with $|S_K| = K$.

AS depends critically on the information available to the algorithm and the performance metric. It is assumed that the overall channel matrix $\mathbf{H}$ is perfectly known at the receiver for AS. This seems to be a very strong assumption given that the receiver may only have K RF chains, and hence can at most perform channel estimation on K receive antennas. In reality, this can be handled using multiple pilot cycles and switching the RF chains on different set of receive antennas [10], [13], which is reasonable particularly in a slow-fading environment [7]. For the performance metric, we consider maximizing the mutual information for RAS [8], [10], [14]. Specifically, by assuming that elements in $\mathbf{x}$ follow i.i.d. standard Gaussian distribution, the mutual information between $\mathbf{x}$ and $\mathbf{y}$ with antenna subset S_K can be written as

$$C(S_K) = \log\det\left(\mathbf{I} + \frac{P}{M}\mathbf{H}_{S_K}\mathbf{H}_{S_K}^{\dagger}\right). \tag{2}$$

We also refer to this performance measure as *capacity* with the implicit understanding that this is for i.i.d. standard Gaussian input distribution. For a given K, the RAS problem seeks to find the optimal antenna set S_K^* that maximizes (2):

$$S_K^* = \arg\max_{S_K\subseteq\mathcal{S},|S_K|=K} C(S_K). \tag{3}$$

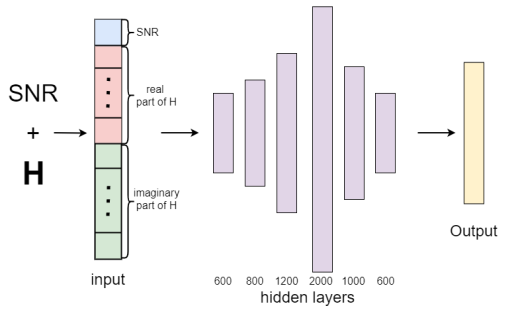


Fig. 1. Overview of the DNN structure.

III. Deep learning based RAS

The key component of the RAS problem in (3) is submatrix selection, which is known to be NP-hard [15]. An exhaustive search has to evaluate all $\binom{N}{K}$ subsets, which is computationally prohibitive when N is large (e.g., in massive MIMO). More efficient methods such as branch-and-bound (BnB) have been applied to RAS [10]. However, its complexity depends on the channel realizations and is limited by the hardness of the problem. As a result, it cannot eliminate the worst-case exponential complexity.

Given the significant advance of deep learning in solving classification problems, and the fact that AS can be viewed as a multi-class classification problem [16], it is natural to explore whether deep neural network (DNN) can be designed to provide a near-optimal alternative to the aforementioned methods. Similar attempts have been reported in radar antenna selection to minimize the Cramér-Rao lower bound [17], and joint multicast beamforming and antenna selection to maximize the minimum signal to noise ratio [18].

A. DNN design

In this section, we propose to solve Problem (3) with a DNN-based receive antenna selection method. The adopted DNN structure is simple: it has an input layer of size $2NM+1$, an output layer of size $\binom{N}{K}$, and J hidden layers with different number of nodes in each hidden layer. The hidden layers share a common pipeline of Rectified Linear Unit (ReLU) activation function and batch normalization (BN), while the output layer replaces ReLU with a Softmax function. The input size $2NM+1$ corresponds to separating the real and imaginary parts of complex-valued entries in the channel matrix $\mathbf{H}$ and the additional input value is the receive SNR P. A pictorial view of this simple DNN with $J = 6$ and hidden layer sizes of $600, 800, 1200, 2000, 1000, 600$ is given in Fig. 1.

We train the DNN above for various standard RAS tasks under the MIMO system model described in Section II. Data samples are generated with standard i.i.d. Raleigh block-fading channel for every $h_{i,j}$, and the labels are generated via brute-force search over all possible RAS candidates for each channel

realization. We use the standard cross entropy loss function for training, as well as mini-batch stochastic gradient descent with momentum. The learning rate usually is 1.0 and the momentum rate is always 0.9.

B. Figures of merit

The performance of DNN-based RAS is evaluated under three performance metrics. The first is the usual RAS accuracy, which is defined as the percentage of test data samples where the DNN output does not match the label (ground truth obtained from brute-force search). Note that as long as there exists at least one selected antenna that is not in the ground truth, this output is marked as erroneous. This objective, in a sense, is too restrictive because it is possible that two different DNN outputs, albeit both are erroneous, have different degrees of error (e.g., one has all selected antennas wrong, while the other only has one such error).

We thus further evaluate two other performance metrics. The first is called *partially correct accuracy*, which is defined as the percentage of correctly selected antennas in the RAS set. In comparison, we refer to the previous RAS accuracy as *best match accuracy*. For example, if the ground truth RAS subset is $\{1, 2, 4, 5\}$ and the DNN output is $\{1, 3, 4, 6\}$, then the best match accuracy is 0 while the partially correct accuracy is 50%.

The second derivative metric is the *capacity loss*, which is defined as

$$\mathsf{CL}_Q = \frac{c_{\max} - c_Q}{c_{\max}}$$

where $c_{\max}$ is the maximum achievable capacity under the optimal RAS, and c_Q is the achieved capacity under the DNN output antenna subset.

C. Performance evaluation

We train DNNs as described in Section III-A for three simple MIMO RAS problems: $(M, N, K) = (2, 4, 2)$, $(2, 8, 4)$ and $(4, 8, 4)$. The training data has $1e6$ samples and the test data has $1e5$ samples for each SNR. The three figures of merit described in Section III-B, which are evaluated by averaging over all test data samples at each SNR, are reported in Fig. 2. We can see that DNN-based RAS is effective for small dimensions – if we are interesting solely in finding the optimal subset of receive antennas, DNN RAS has an accuracy of over 90% throughout the range of SNRs for $(2, 4, 2)$ and $(2, 8, 4)$, but the accuracy decreases for $(4, 8, 4)$, which is a more difficult problem than the previous two. On the other hand, the partially correct accuracy characterizes the percentage of correctly selected antennas in the RAS set, and we note that across all problem dimensions the DNN-based method outputs accuracies between 91% and 99%. Lastly, the capacity loss directly measures the impact of RAS on the MIMO system utility, and we observe from Fig. 2(c) that the DNN-based solution is extremely effective – the largest percentage of capacity loss is less than 0.35%, across all SNR ranges and all problem dimensions.

D. Enhancement

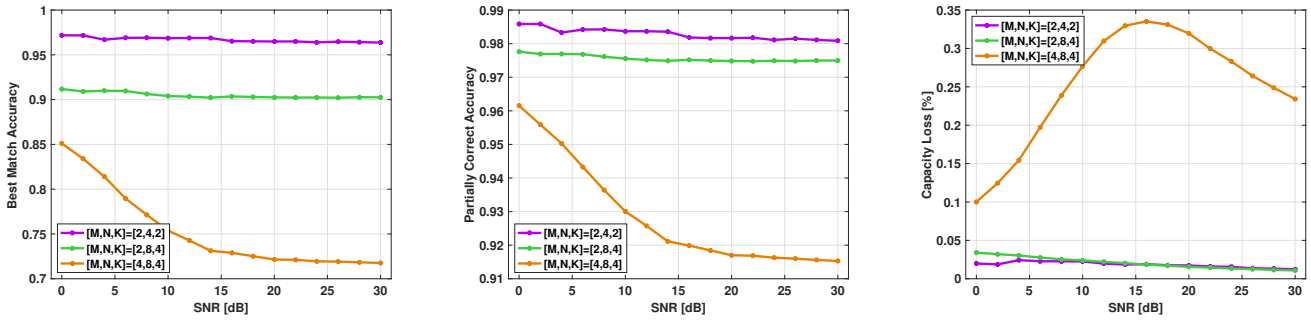
One interesting question that originates from the previous numerical results is that whether the excellent performance obtained by the DNN-based RAS design can be maintained when we reduce the DNN model size. This is a practically important aspect because smaller models require less storage and less computation. We have evaluated six different DNN model configurations as summarized in Table I, where model 1 corresponds to the previous model used in the simulation and models 2 to 6 are reduced-size ones. We note that models 1, 2, 5, 6 still have comparable sizes, while models 3 and 4 are of much smaller sizes. We carry out a numerical simulation where we fix the problem dimension to be $(4, 8, 4)$ and train each model across a set of SNRs the same way as in the previous experiment. We then report the average performance against the training epochs for all SNRs in Fig. 4 for the six models in Table I in order to evaluate their convergence behavior. Clearly, two of the smallest models perform poorly, suggesting that the model size does matter. However, models 2, 5 and 6 perform competitively with model 1, which also suggests that there is a reasonable tradeoff of reducing the model size and decreasing the RAS performance – we see that model 6 can have almost the same performance as model 1 but with only half of the model parameters. We also note that for models with similar number of parameters, the specific structure does not affect the performance noticeably.

TABLE I
STRUCTURE OF SIX DNN MODELS.

Model Index	No. Hidden Layers	No. Nodes	Floating-Point Model Size
1	6	600, 800, 1200, 2000, 1000, 600	26.3 MB
2	5	600, 800, 1200, 1000, 600	13.4 MB
3	3	100, 200, 200	347 KB
4	8	100, 200, 200, 200, 200, 200, 200, 100	1.1 MB
5	3	1000, 2000, 1000	16.6 MB
6	6	800, 900, 900, 900, 900, 800	16.0 MB

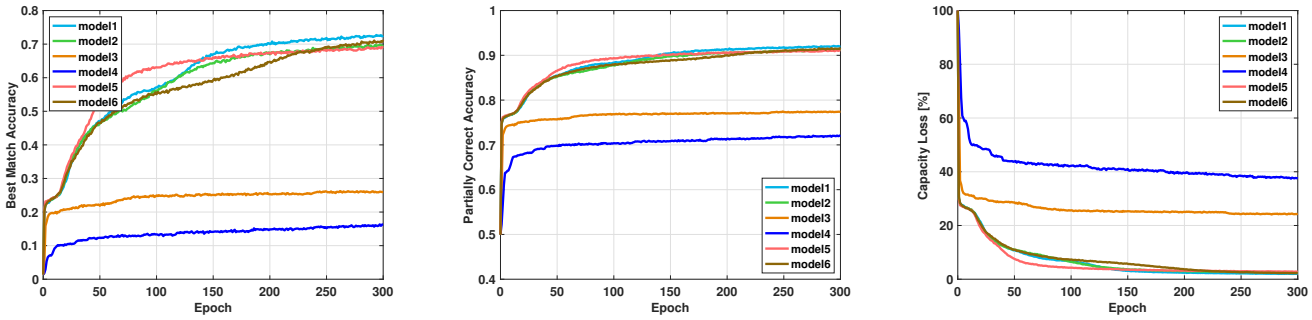
IV. RAS VIA GREEDY ADAPTATION

The previous section studies how to obtain a near-optimal RAS solution using DNN, and we have seen from the simulation results that this approach is generally effective for a given K. A significant limitation of the DNN-based solution, however, is that any trained DNN only works for a pre-determined K. It does not scale with K in a natural way – even if we are given a well-trained DNN that can approximate S_K^* very well, it is unclear how to find a good solution of S_{K+L}^* (the $(M, N, K+L)$ problem) or S_{K-L}^* (the $(M, N, K-L)$ problem) based on S_K that approximates S_K^*, in a computationally efficient way. This problem is practically relevant since the actual number of RF chains may not be known prior to the offline DNN training. State-of-the-art DNN is not adaptive with respect to the number of classes in the



(a) Best match accuracy (b) Partially correct accuracy (c) Capacity loss

Fig. 2. Average performance of DNN RAS under different metrics.



(a) Best match accuracy (b) Partially correct accuracy (c) Capacity loss

Fig. 3. Average performance of six different DNN models for RAS.

classification task. However, for the RAS problem, we need the flexibility of changing the number of selected receive antennas *after* the DNN has been trained offline for a fixed (M, N, K).

One possible solution for the adaptation problem is to train a set of DNNs, each covering one possible (M, N, K). This, however, requires much more offline computation as well as on-board storage. Ideally, we would like to have an *offline* well-trained DNN, which may have high offline complexity but can output near-optimal S_K^*, and then use a low-complexity *online* algorithm that builds on the offline K-solution to get a good $(M, N, K + L)$ or $(M, N, K - L)$ solution. Formally, we want to solve the following problem:

$$\begin{array}{ll} \text{Increase:} & S_{K+L}^* = \arg\max_{S\subseteq\mathcal{S},|S|=K+L} C(S), \\ \text{Decrease:} & S_{K-L}^* = \arg\max_{S\subseteq\mathcal{S},|S|=K-L} C(S), \end{array} \tag{4}$$

with a given (possibly approximate) solution S_K to Problem (3).

A. Greedy adaptation for $K + L$

We first propose `Greedy+` to solve the incrementally mismatched RAS problem, which is compactly described in Algorithm 1. It is not a surprise that the greedy principle is adopted for an approximation algorithm for $K + L$, as greedy search algorithms have been well studied in both RAS [6], [14] and transmit antenna selection [12], [19]. In particular, Vaze and Ganapathy [14] have proved that RAS with objective function (2) falls into the *monotone submodular function maximization (SFM)* category, which has been widely studied [20]. This is an important observation because there exists a rigorous performance guarantee for SFM: the simple greedy algorithm, which maximizes per-step mutual information gain, achieves $(1 - 1/e)$ fraction of the optimal RAS [21] with a polynomial-time complexity.

Algorithm 1: `Greedy+`

Input: $\mathbf{H}$, (M, N, K), P, $L : N - K > L > 0$, S_K
Initialize: $S_{K+L} = S_K$, $V = [N]$
for $t \leftarrow 1$ **to** L **do**
  $x \leftarrow \arg\max_{v\in V} \left(C(S_{K+L} \cup \{v\}) - C(S_{K+L})\right)$
  $S_{K+L} \leftarrow S_{K+L} \cup \{x\}$
end
Output: S_{K+L}

B. Greedy adaptation for $K - L$

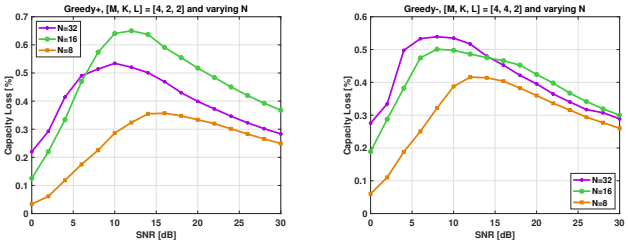
The dual problem of Section IV-A is that if we are given a solution of S_K that approximates S_K^*, how can we find a computationally efficient solution to approximate S_{K-L}^*. The proposed `Greedy-`, which is a dual version of Algorithm 1, is given in Algorithm 2. The idea of `Greedy-` is the opposite of `Greedy+`: starting from S_K, one always removes the element that has the least impact of capacity reduction. Such reverse-greedy operation is intuitive and easy to implement.

Algorithm 2: `Greedy-`

Input: $\mathbf{H}$, (M, N, K), P, $L : K > L > 0$, S_K
Initialize: $S_{K-L} = S_K$, $V = S_K$

for $t \leftarrow 1$ **to** L **do**
  $x \leftarrow \arg\min_{v \in V} \left(C(S_{K-L}) - C(S_{K-L} \backslash \{v\})\right)$
  $S_{K-L} \leftarrow S_{K-L} \backslash \{x\}$
end

Output: S_{K-L}



(a) Best match accuracy (b) Partially correct accuracy

Fig. 4. `Greedy+` (a) and `Greedy-` (b) for varying N.

C. Performance evaluation

We now evaluate the hybrid `Greedy+/-` designs in the same MIMO RAS problem, using the same system described in Section II. We focus on evaluating the capacity loss, which is with respect to the optimal solution of S^*_{K+L} or S^*_{K-L} by invoking the `Greedy+/-` algorithms with S^*_K. Figs. 4(a) and 4(b) present the numerical results for $K + L$ (`Greedy+`) and $K - L$ (`Greedy-`), respectively. We note that the RAS tasks here are more difficult than the ones in the DNN simulations of Section III-C due to a larger MIMO dimension N, so that we can evaluate the hybrid designs in a more practical, massive-MIMO-representative setting. The capacity losses reported in Fig. 4 are less than 0.7% for all SNRs and all receive antennas N, indicating that this hybrid method, which starts from an (M, N, K) solution and then greedily increases (or decreases) the antennas by maximizing (or minimizing) the per-step capacity gain (or loss) achieves almost the same performance as the optimal solution for $K + L$ (or $K - L$).

V. Conclusions

This paper made two contributions to the MIMO receive antenna selection problem. First, we have proposed a deep neural network based RAS solution that deviates from the existing model-driven, optimization-based approaches. The impact of model sizes to the RAS performance is also evaluated. Second, we have addressed the dynamic RAS problem where a good solution for a pre-determined K is given, but the actual problem may require selecting more (or fewer) than K receive antennas. The proposed `Greedy+/-` solutions are very flexible and can work with any pre-developed RAS solution for a given problem dimension (M, N, K) to achieve a low-complexity near-optimal RAS solution for $(M, N, K \pm L)$. Numerical simulations are carried out to validate the effectiveness of the proposed solutions.

References

[1] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, 2018.

[2] E. Bjornson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, 2016.

[3] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave OFDM large-scale antenna arrays," *IEEE J. Select. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, 2017.

[4] C. Mollen, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[5] C. Shen, P. Zhao, and X. Luo, "On energy efficient uplink multi-user mimo with shared lna control," *submitted to IEEE International Conference on Communications (ICC) 2021*.

[6] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 339–347, February 2004.

[7] A. F. Molisch, M. Z. Win, Yang-Seok Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1759–1772, July 2005.

[8] A. Dua, K. Medepalli, and A. J. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2353–2357, September 2006.

[9] J. Joung and S. Sun, "Two-step transmit antenna selection algorithms for massive MIMO," in *IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[10] Y. Gao, H. Vinck, and T. Kaiser, "Massive MIMO antenna selection: Switching architectures, capacity bounds, and optimal antenna selection algorithms," *IEEE Trans. Signal Processing*, vol. 66, no. 5, pp. 1346–1360, March 2018.

[11] S. Asaad, A. M. Rabiei, and R. R. Muller, "Massive MIMO with antenna selection: Fundamental limits and applications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8502–8516, December 2018.

[12] M. O. K. Mendonca, P. S. R. Diniz, T. N. Ferreira, and L. Lovisolo, "Antenna selection in massive MIMO based on greedy algorithms," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1868–1881, March 2020.

[13] A. Garcia-Rodriguez, C. Masouros, and P. Rulikowski, "Reduced switching connectivity for large scale antenna selection," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2250–2263, May 2017.

[14] R. Vaze and H. Ganapathy, "Sub-modularity and antenna selection in MIMO systems," *IEEE Commun. Letter*, vol. 16, no. 9, pp. 1446–1449, September 2012.

[15] A. Civril and M. Magdon-Ismail, "On selecting a maximum volume sub-matrix of a matrix and related problems," *Theoretical Computer Science*, vol. 410, no. 47, pp. 4801–4811, 2009.

[16] J. Joung, "Machine learning-based antenna selection in wireless communications," *IEEE Commun. Letter*, vol. 20, no. 11, pp. 2241–2244, November 2016.

[17] A. M. Elbir, K. V. Mishra, and Y. C. Eldar, "Cognitive radar antenna selection via deep learning," *IET Radar, Sonar Navigation*, vol. 13, no. 6, pp. 871–880, 2019.

[18] M. S. Ibrahim, A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Learning-based antenna selection for multicasting," in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[19] A. Konar and N. D. Sidiropoulos, "A simple and effective approach for transmit antenna selection in multiuser massive MIMO leveraging submodularity," *IEEE Trans. Signal Processing*, vol. 66, no. 18, pp. 4869–4883, September 2018.

[20] A. Krause and D. Golovin, "Submodular function maximization," in *Tractability: Practical Approaches to Hard Problems*. Cambridge University Press, February 2014.

[21] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions - I," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.