DynaPA: Dynamic Power Allocation for Improved Exploration-Exploitation in Active Sensing

Parthasarathi Khirwadkar, Mehmet Can Hücümenoğlu, Piya Pal
University of California San Diego
pkhirwad@ucsd.edu, mhucumen@ucsd.edu, pipal@ucsd.edu

2024 IEEE 13th Sensor Array and Multichannel Signal Processing Workshop (SAM) | 979-8-3503-4481-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/SAM60225.2024.10636376

***Abstract*—This paper addresses an adaptive active sensing problem of nonstationary sparse multi-target detection, by dynamically shaping the transmit beampattern of a MIMO radar. In order to strike a better balance between exploration and exploitation (two key ideas in any online decision making problem), a dynamic power allocation algorithm called DynaPA is developed that can redirect exploration energy for exploitation (and vice versa) depending on the environment. Numerical results demonstrate superior performance of DynaPA over existing algorithms with varying target numbers.**

***Index Terms*—Active Sensing, Sequential decision making, MIMO Radar, Transmit Beamforming, Reinforcement Learning.**

I. Introduction

Active and adaptive sensing using antenna arrays is a key enabler of many modern technologies ranging from autonomous systems to joint communication and sensing in massive multiple-input multiple-output (MIMO) wireless systems [1]–[4]. A main challenge in MIMO radar in a nonstationary environment with sparse number of targets is to detect new targets (exploration) while not losing the currently detected targets (exploitation). There is an inherent trade-off between these exploration and exploitation tasks since the total transmit power is limited [5]–[7].

Earlier works on multi-target detection and tracking [8], [9] typically assume the target motion model and environment dynamics are known in order to predict the future state of the targets and the environment. When such prior information is absent, online learning approaches based on reinforcement learning (RL) are recently being investigated as model-free approaches for sensing. In [10] RL was used for multi-target detection by extending the robust Wald-type statistic developed in [11] for single target. Building on [10], an improved RL-based method for MIMO radar was proposed in [12] to enhance the detection of targets with large dynamic ranges. However, [12] only focuses the transmit power on a limited number of directions with the largest test statistics based on the previous measurements at each pulse. To search for more targets, at each pulse the algorithm employs an ϵ-greedy policy. Specifically, with probablity ϵ, it chooses to focus the transmit power in directions with lower test statistics, and with probability $1-\epsilon$, it illuminates targets already detected. On the other hand, [13] divides the total power into two components for exploration and exploitation. The exploration power is fixed and utilized to create an omnidirectional beam, while the exploitation power is focused towards certain directions. The exploitation power is further divided based on the classification of targets as weak or strong in a *fixed ratio* between the two classes. This fixed ratio, and the fixed power dedicated to omnidirectional beams, make the algorithm less flexible and harder to adapt to dynamic scenarios where targets can appear and disappear in an unknown manner. Other works extend these methods to a joint sensing and communication framework [14], [15].

Contributions: We introduce a new dynamic power allocation scheme, named DynaPA, for transmit beamforming in MIMO radar with sparsely located targets. Our algorithm does not assume any environmental dynamics or target motion models. DynaPA utilizes the robust Wald-type statistics developed in [11] along with a novel "state transition-based" power allocation strategy which keeps track of the past state in addition to the current state. Our state-space also contains an "uncertainty state" which helps us decide when to explore and exploit in response to a dynamic target scene. As a result, DynaPA allocates adequate power to keep track of previously detected targets and utilizes the remaining power to search for new targets. It *dynamically* distributes the exploration power (as opposed to allocating a fixed budget, or maintain a fixed ratio of powers between strong and weak targets) based on the acquired target statistics. Further, DynaPA uses relatively less power for target tracking at first and only increases the allocated power when necessary to improve robustness of detection. Our numerical results demonstrate that DynaPA outperforms previous RL algorithms under a scenario with varying number of targets and SNRs.

II. Problem Formulation and Background

A. Measurement Model

We consider an active sensing model with a colocated[1] massive MIMO radar consisting of N_T transmit (Tx) antennas and N_R receive (Rx) antennas. The transmit and receive antennas are placed in a uniform linear array (ULA) with spacing equal to half the wavelength of the carrier frequency.

The transmitted signal at the n^{th} pulse $n = 1, \ldots, M$ with pulse width T is represented by $\mathbf{s}(t + (n-1)T) \in \mathbb{C}^{N_T}$ where $t \in [0, T]$ is the inter-pulse time (fast time) [18]. The transmitted signal is generated as a linear combination of

This work was supported in part by grants ONR N00014-19-1-2256, NSF 2124929, and DE-SC0022165

[1]Colocated (or monostatic) MIMO radar assumes phase coherence between the transmitters and receivers, as well as equal target directions of departure and arrival [16], [17].

orthonormal signals $\mathbf{s}_0(t) \in \mathbb{C}^{N_T}$ (i.e., $\int_0^T \mathbf{s}_0(t)\mathbf{s}_0^H(t)dt = \mathbf{I}$) given by

$$\mathbf{s}(t + (n-1)T) = \mathbf{W}^{(n)}\mathbf{s}_o(t)$$

where $\mathbf{W}^{(n)} = [\mathbf{w}_1^{(n)}, \mathbf{w}_2^{(n)}, \ldots, \mathbf{w}_{N_T}^{(n)}]^T \in \mathbb{C}^{N_T \times N_T}$ is the transmit beamforming weight matrix used at the n^{th} pulse. The vector $\mathbf{w}_k^{(n)}$ represents k^{th} row of $\mathbf{W}^{(n)}$ and corresponds to the weight vector of the k^{th} antenna. The per-antenna power at pulse n is constrained to $\|\mathbf{w}_k^{(n)}\|_2^2 = P_T,\ k = 1, \ldots, N_T$. Thus, the total transmit power $\overline{P}_T$ is given as $\overline{P}_T \triangleq \sum_{k=1}^{N_T} \|\mathbf{w}_k^{(n)}\|_2^2 = \mathrm{tr}(\mathbf{W}^{(n)}(\mathbf{W}^{(n)})^H) = P_T N_T$.

We consider a nonstationary setting where at pulse index n, there are $K^{(n)}$ targets in the directions $\{\theta_k^{(n)}\}_{k=1}^{K^{(n)}} \in [-\frac{\pi}{2}, \frac{\pi}{2})$. The received signal at each Rx antenna is processed by a bank of N_T matched filters tuned to a specific range-Doppler bin. Following [13], [19], we assume that all $K^{(n)}$ targets are in this same range-Doppler bin. In this case, using the fact that $\int_0^T \mathbf{s}_0(t)\mathbf{s}_0^H(t)dt = \mathbf{I}$, the measurements at the output of the matched filters, arranged in the form of a $N_R \times N_T$ matrix, is given by

$$\mathbf{Y}^{(n)} = \sum_{k=1}^{K^{(n)}} \alpha_k^{(n)} \mathbf{a}_R(f_k^{(n)})\mathbf{a}_T^T(f_k^{(n)})\mathbf{W}^{(n)} + \mathbf{C}^{(n)}. \quad (1)$$

Here $\alpha_k^{(n)}$ is an unknown time-varying coefficient associated with the k^{th} target that incorporates both the Radar Cross Section (RCS) and the two-way path loss. The vectors $\mathbf{a}_T(f) = [1, e^{j\pi f}, e^{j\pi 2f}, \ldots, e^{j\pi(N_T-1)f}]^T$ and $\mathbf{a}_R(f) = [1, e^{j\pi f}, e^{j\pi 2f}, \ldots, e^{j\pi(N_R-1)f}]^T$ represent the steering vectors of the transmit and receive arrays respectively, corresponding to the spatial frequency $f_k^{(n)} = \sin(\theta_k^{(n)}) \in [-1, 1)$. Finally, $\mathbf{C}^{(n)}$ denotes the combined effect of noise and other additive disturbances after matched filtering. By vectorizing $\mathbf{Y}^{(n)}$, we obtain $\mathbf{y}^{(n)} \in \mathbb{C}^{N_T N_R}$ as follows:

$$\mathbf{y}^{(n)} \triangleq \mathrm{vec}(\mathbf{Y}^{(n)}) = \sum_{k=1}^{K^{(n)}} \alpha_k^{(n)} \mathbf{h}^{(n)}(f_k^{(n)}) + \mathbf{c}^{(n)} \quad (2)$$

where $\mathbf{c}^{(n)} = \mathrm{vec}(\mathbf{C}^{(n)})$ is the disturbance vector and,

$$\mathbf{h}^{(n)}(f_k^{(n)}) \triangleq \left((\mathbf{W}^{(n)})^T \mathbf{a}_T(f_k^{(n)})\right) \otimes \mathbf{a}_R(f_k^{(n)}) \quad (3)$$

where $\otimes$ denotes the Kronecker product. The exact characterization of the disturbance vector $\mathbf{c}^{(n)}$ poses a difficult challenge. Thus, following [11], we only make the following mild assumption on $\mathbf{c}^{(n)}$:

[A1] Let $\{c_i^{(n)}\}\ \forall n$ be a stationary and circular complex-valued random process whose autocorrelation function satisfies $|E[c_i^{(n)}(c_{i-j}^{(n)})^*]| < cj^{-\rho}$ for $j \in \mathbb{Z}$ for some constants $c > 0$ and $\rho > 1$.

In other words, the autocorrelation values exhibit polynomial decay. It is important to highlight that this assumption is sufficiently flexible to encompass a wide range of practical disturbance models [11].

B. Robust Target Detection

We review the robust target detection framework from [11] which utilizes the aforementioned mild statistical assumption on the disturbance $\mathbf{c}^{(n)}$ without having to know its exact distribution. The range of spatial frequency $[-1, 1)$ is divided into a uniform grid $\mathcal{G} = \left\{-1 + \frac{2(l-1)}{L},\ l = 1, \ldots, L\right\}$ of size L. We assume there are sparse number of targets, i.e. $K^{(n)} \ll L$ located on the grid $\mathcal{G}$. It is further assumed that the measurements (2) are processed by a bank of L spatial filters, each of which tuned to a specific spatial frequency bin from $\mathcal{G}$ [10], [13] [2]. The measurement $\mathbf{y}_l^{(n)} \in \mathbb{C}^{N_T N_R}$ at the output of the l^{th} spatial filter (corresponding to angle bin l) is given by

$$\mathbf{y}_l^{(n)} = \begin{cases} \alpha_l^{(n)}\mathbf{h}_l^{(n)} + \mathbf{c}_l^{(n)} & (-1 + \frac{2(l-1)}{L}) \in \{f_k^{(n)}\}_{k=1}^{K^{(n)}} \\ \mathbf{c}_l^{(n)} & \text{otherwise} \end{cases} \quad (4)$$

where we denote $\mathbf{h}_l^{(n)} \triangleq \mathbf{h}^{(n)}(-1 + \frac{2(l-1)}{L})$ for simplicity. In order to perform target detection at each angle bin, a robust Wald-type statistic $\Lambda(\mathbf{y}_l^{(n)})$ is compared against a threshold λ selected according to a desired probability of false alarm $\mathcal{P}_f$

$$\Lambda(\mathbf{y}_l^{(n)}) \gtrless \lambda, \quad \Lambda(\mathbf{y}_l^{(n)}) \triangleq \frac{2|(\mathbf{h}_l^{(n)})^H \mathbf{y}_l^{(n)}|^2}{(\mathbf{h}_l^{(n)})^H \widehat{\mathbf{\Gamma}}_{l,n} \mathbf{h}_l^{(n)}}. \quad (5)$$

Here $\widehat{\mathbf{\Gamma}}_{l,n}$ is an estimate of the unknown covariance of disturbance $\mathbf{c}_l^{(n)}$ and calculated according to eq. (23) in [11]. For ease of notation, we will henceforth use $\Lambda_l^{(n)} = \Lambda(\mathbf{y}_l^{(n)})$.

III. A Novel Exploration-Exploitation Strategy for Dynamic Power Allocation

Since the dynamical model for targets is unknown, the authors in [10] propose to employ Reinforcement Learning (RL) to select a set of (possible) target angles at every pulse n, and solve a beamforming problem which aims to maximize the minimum gain over this set. Assuming that the angles are on a uniform grid of appropriate size, it can be shown that the optimum beamformers are of the following form

$$\mathbf{W}^{(n)} = \frac{1}{N_T}\mathbf{A}^* \mathrm{diag}\left((\mathbf{r}^{(n)})^{1/2}\right)\mathbf{A}^T. \quad (6)$$

Here $\mathbf{A} = [\mathbf{a}_T(f_1), \mathbf{a}_T(f_2), \ldots, \mathbf{a}_T(f_L)] \in \mathbb{C}^{N_T \times L}$ is the manifold matrix corresponding to the directions $f_l = -1 + \frac{2(l-1)}{L} \in \mathcal{G}$ and $\mathbf{A}^*$ is its complex conjugate. Note that $\overline{P}_T = \mathrm{trace}(\mathbf{W}^{(n)}(\mathbf{W}^{(n)})^H) = \sum_i [\mathbf{r}^{(n)}]_i$ for N_T divisible by L.[3] The vector $\mathbf{r}^{(n)} \in \mathbb{R}^L$ denotes how much transmit power will be allocated to each angle bin at pulse n. Indeed, the goal of RL-based approaches is to *ultimately choose the power allocation dynamically* and adaptively. Our contribution is to propose a novel power allocation scheme (called DynaPA), which is significantly different from recent RL-based methods. We briefly review the main idea behind existing RL-based

[2] Such spatial filters can be constructed under suitable assumptions on the grid size.

[3] Note that per antenna power $\|\mathbf{w}_i^{(n)}\|_2^2 = P_T = \overline{P}_T/N_T\ \forall i = 1, \ldots, N_T$ for every $\mathbf{r}^{(n)}$ satisfying $\|\mathbf{r}^{(n)}\|_1 = \overline{P}_T$

methods in order to elucidate the key difference(s) from our technique.

A. Reinforcement-learning based power allocation

Previous works [10], [12], [13] typically used RL to determine a number (say a_n) of angular bins to be sensed and then use an $\epsilon-$greedy approach to strike a balance between exploiting selected bins and (randomly) exploring bins believed to be unoccupied. In order to further improve the detection of weak targets, [13] proposed to divide the total power $\overline{P}_T$ into an omnidirectional beam with power P_1 and focus the rest of the power $\overline{P}_T - P_1$ towards the angle bins selected by the RL algorithm. The selected angle bins are further divided into "weak" and "strong" based on Wald statistic, and the power is allocated by maintaining a fixed ratio between the minimum gains of beamformer for weak and strong targets. The "exploration power" P_1 is typically kept fixed and not redirected dynamically for exploitation as needed. The ratio of power allocated for weak and strong targets is also kept fixed and it does not adapt to the changing scenario.

In order to address the above issues and further improve the performance of RL-based methods, we present a dynamic power allocation strategy named DynaPA, that aims to dynamically modify the transmit power available for "exploring" new/undetected targets, while also allocating sufficient power for tracking (or "exploiting") previously detected targets.

B. DynaPA: A new way to explore and exploit via Beamforming with Dynamic Power Allocation

We begin by introducing state variables $\zeta_l^{(n)} \in \{0,1,2\},\ l = 1,\ldots,L$ for each angle bin l and pulse index n. The state $\zeta_l^{(n)} = 0$ indicates that the l^{th} angle bin is believed to be empty while $\zeta_l^{(n)} = 1$ indicates that it is believed to be occupied. A novel contribution is to augment the state space with an "uncertainty state" $\zeta_l^{(n)} = 2$, which indicates uncertainty about the occupancy state of bin l. State transition $\zeta_l^{(n-1)} \rightarrow \zeta_l^{(n)},\ l = 1,\ldots,L$ is determined by $d_l^{(n)} \triangleq \mathbb{1}\{\Lambda_l^{(n)} > \lambda\}$, as illustrated in Figure 1.[4]

A key idea of DynaPA is that the power allocation rule is not just based on the current state $\zeta_l^{(n)}$ of bin l, but also on its past state $\zeta_l^{(n-1)}$ (in particular, the transition). To this end, we define sets that comprise of all angle bins undergoing same state transition

$$S_{i,j}^{(n)} \triangleq \{l \in [L] | \zeta_l^{(n-1)} = i, \zeta_l^{(n)} = j\}$$

where $i,j \in \{0,1,2\}$ and $i \rightarrow j$ is a valid state transition.

We initialize all states with $\zeta_l^{(0)} = 0$ and $r_l^{(1)} = \frac{\overline{P}_T}{L}, l = 1,\ldots,L$. First, we consider an intermediate power allocation for each bin

$$\widetilde{r}_l^{(n+1)} = \begin{cases} \gamma r_l^{(n-1)} & l \in S_{2,1}^{(n)} \\ r_l^{(n)} & l \in S_{0,1}^{(n)}, S_{1,1}^{(n)}, S_{1,2}^{(n)} \\ 0 & l \in S_{0,0}^{(n)}, S_{2,0}^{(n)} \end{cases} \tag{7}$$

[4]Note that we do not have transition $0 \rightarrow 2$ as the robust detection rule ensures probability of false alarm is low when choosing the threshold λ [11].

where $\gamma > 1$ is a fixed scaling factor ($\gamma = 1.3$ was used in simulations). Let $\widetilde{P}_T^{(n+1)} \triangleq \sum_{l=1}^{L} \widetilde{r}_l^{(n+1)}$. For angle bins transitioning $0 \rightarrow 1$ or $1 \rightarrow 1$, we allocate same power as the previous pulse since the allocated power is adequate for target detection. However, for angle bin $l \in S_{2,1}^{(n)}$ the transition $2 \rightarrow 1$ implies a transition $1 \rightarrow 2$ occurred in the previous pulse (i.e., $l \in S_{1,2}^{(n-1)}$). This suggests the allocated power $r_l^{(n-1)}$ was insufficient for reliable target tracking and hence we increase the allocated power to $\gamma r_l^{(n-1)}$ to prevent false negatives in the future. In order to determine the final power allocation, we consider two cases:

Case I: If $\widetilde{P}_T^{(n+1)} \geqslant \overline{P}_T$ then our final power allocation is

$$r_l^{(n+1)} = \frac{\overline{P}_T}{\widetilde{P}_T^{(n+1)}} \widetilde{r}_l^{(n+1)},\ l = 1,\ldots,L \tag{8}$$

and this meets the total power constraint. In this case, we do not allocate any extra power for exploration of unoccupied bins corresponding to $\zeta_l^{(n)} = 0$.

Case II: If $\widetilde{P}_T^{(n+1)} < \overline{P}_T$, define the exploration power $P_E^{(n+1)}$ available in $(n+1)^{th}$ pulse as

$$\begin{aligned} P_E^{(n+1)} &\triangleq \overline{P}_T - \sum_{l \in S_{0,1}^{(n)}, S_{1,1}^{(n)}, S_{2,1}^{(n)}} \widetilde{r}_l^{(n+1)} \\ &= \overline{P}_T - \widetilde{P}_T^{(n+1)} + \sum_{l \in S_{1,2}^{(n)}} \widetilde{r}_l^{(n+1)}. \end{aligned} \tag{9}$$

The strategy for the allocation of exploration power depends on whether or not the "uncertainty set" $S_{1,2}^{(n)}$ is empty, as follows

Case II a: If $S_{1,2}^{(n)} = \emptyset$, then the power allocation rules for the remaining bins $l \in S_{0,0}^{(n)}, S_{2,0}^{(n)}$ is given by:

$$r_l^{(n+1)} = \begin{cases} P_E^{(n+1)} \frac{\Lambda_l^{(n)}}{\sum_{i \in S_{0,0}^{(n)}, S_{2,0}^{(n)}} \Lambda_i^{(n)}} & l \in S_{0,0}^{(n)}, S_{2,0}^{(n)} \\ \widetilde{r}_l^{(n+1)} & \text{otherwise} \end{cases}. \tag{10}$$

Our exploration strategy is dynamic (not based on a fixed omnidirectional power allocation). The power allocation strategy from [10], [12] does not consider that targets with different reflectivity would require different amount of illumination to detect reliably. Our dynamic power allocation scheme assigns power based on the Wald statistic which is an indicator of the SNR level of each target.

Case II b: If $S_{1,2}^{(n)} \neq \emptyset$, we temporarily stop exploration for $l \in S_{0,0}^{(n)}, S_{2,0}^{(n)}$ and instead use $P_E^{(n+1)}$ to resolve/rectify our uncertainty about occupancy of angle bins which have lost their target (i.e., transitioned from state 1 to 2). We use the following update rule

$$r_l^{(n+1)} = \begin{cases} P_E^{(n+1)} \frac{\widetilde{r}_l^{(n+1)}}{\sum_{i \in S_{1,2}^{(n)}} \widetilde{r}_i^{(n+1)}} & l \in S_{1,2}^{(n)} \\ \widetilde{r}_l^{(n+1)} & \text{otherwise} \end{cases}. \tag{11}$$

Note that $P_E^{(n+1)} = \overline{P}_T - \widetilde{P}_T^{(n+1)} + \sum_{i \in S_{1,2}^{(n)}} \widetilde{r}_i^{(n+1)} \geqslant \sum_{i \in S_{1,2}^{(n)}} \widetilde{r}_i^{(n+1)}$, hence $r_l^{(n+1)} \geqslant \widetilde{r}_l^{(n+1)} = r_l^{(n)}$ for $l \in S_{1,2}^{(n)}$. Note that the power allocation strategy from [13] sets aside

a fixed power budget (P_1) for an omnidirectional beam for exploration, and this power budget cannot be redirected for other purposes. In contrast, our exploration strategy is dynamic (Case II a). Furthermore, as in Case II b, DynaPA can redirect the exploration power for rectifying uncertainty and reducing the chances of false negatives.

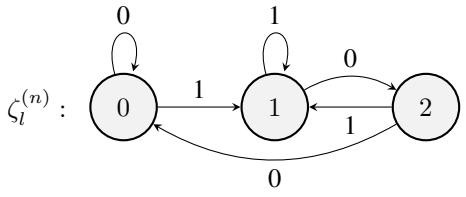


Fig. 1: State transition diagram for $\zeta_l^{(n)}$. The transition $\zeta_l^{(n-1)} \rightarrow \zeta_l^{(n)}$ depends on the value of $d_l^{(n)} = \mathbb{1}\{\Lambda_l^{(n)} > \lambda\}$ (indicated on the arrows)

IV. Simulations

We consider a MIMO radar setup with $N_T = 100$ and $N_R = 100$ antennas both arranged as a ULA. The per-antenna power is constrained to $P_T = \overline{P}_T/N_T = 1$. Following [10], [12] in order to ensure fair comparison, the angular region is divided into $L = 20$ angle bins and the probability of false alarm is set to $\mathcal{P}_f = 10^{-4}$ which results in the detection threshold $\lambda = -2\ln\mathcal{P}_f = 18.42$ [11]. We simulate a sparse, nonstationary scenario with a total of 4 radar targets and over 100 pulses, the details of which are given in Table I. The coefficients of the k^{th} target are i.i.d $\alpha_k^{(n)} \sim CN(0, \sigma_k^2)$ with SNR $= 10\log_{10}(\sigma_k^2)$. The results of all experiments are averaged over 1000 Monte Carlo simulations. We consider the noise is distributed as an SOS AR(6) process as given in [11] whose normalised power spectral density, shown in Figure 2a, is overlaid with the direction of each target indicated by red dashed lines.

In Figures 2b to 2d, we plot the average beam power[5] allocated to each angle bin at each pulse. For pulses $[1, 50]$, Figure 2b shows that DynaPA transmits most power towards Target 4 while Target 1 and 2 receive relatively less power. Figures 2c and 2d show that the approaches from [12] (denoted by "RL (enhanced)") and [13] (denoted by "RL (weak targets)" with $P_1 = \frac{\overline{P}_T}{4}$) illuminate Target 1 and 2 with more power than Target 4. For pulses $[51, 100]$, Figures 2c and 2d show that the power allocated by these approaches towards Target 4 increases once Targets 1 and 2 have left the scene. However, it is still less than the power allocated by DynaPA, as shown in Figure 2b.

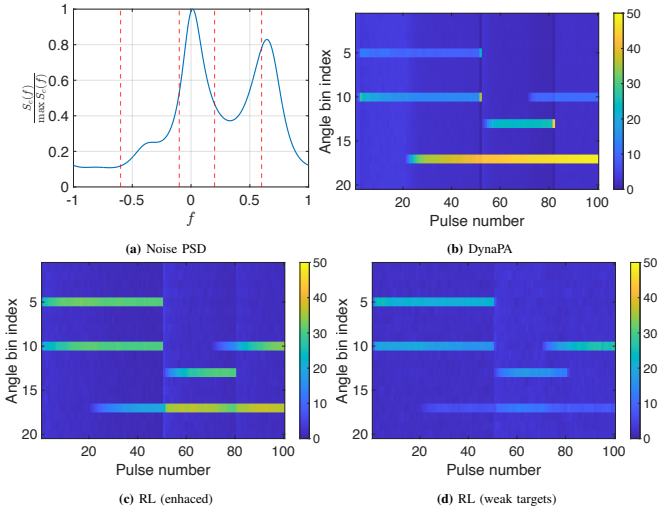
In Figure 3, we plot the dynamic probability of detection of each target as a function of the pulse number. We compare the performance of DynaPA against the two RL-based algorithms and an "Oracle" algorithm which knows the radar target locations and divides the total power equally towards the targets present at each pulse. For pulses $[1, 50]$, Figures 3a and 3b show that DynaPA does not sacrifice detection performance of Targets 1 and 2 even though it allocates less power compared to

[5]The beam power in bin l is given by $B_l^{(n)} = \|(\mathbf{W}^{(n)})^T \mathbf{a}_T(f_l)\|_2^2$ where $f_l = -1 + \frac{2(l-1)}{L} \in \mathcal{G}$.

Target	Bin Index	$\sin(\theta)$	SNR(dB)	Pulse interval
1	5	-0.3	-20	$[1, 50]$
2	10	-0.05	-15	$[1, 50] \cup [71, 100]$
3	13	0.1	-20	$[51, 80]$
4	17	0.3	-20	$[21, 100]$

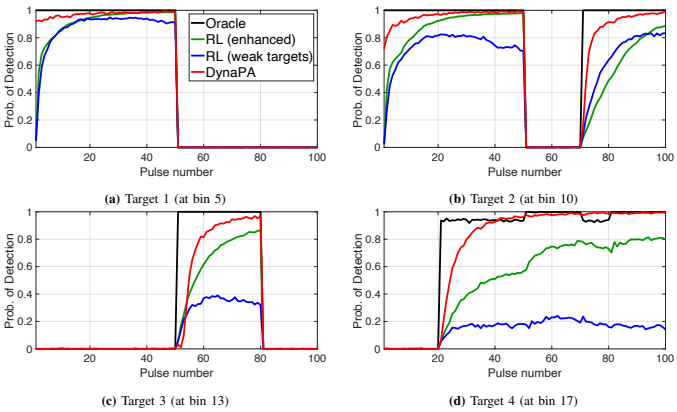
TABLE I: Simulation scenario

[12], [13]. Further, Figure 3d shows that DynaPA detects Target 4 faster than the RL-based approaches. In pulses $[51, 100]$, we observe a similar trend in Figures 3b and 3c with DynaPA detecting Targets 2 and 3 much faster than the RL-based approaches.



(a) Noise PSD **(b)** DynaPA **(c)** RL (enhaced) **(d)** RL (weak targets)

Fig. 2: (a) Noise power spectral density and (b)-(d) average beam pattern at each pulse



(a) Target 1 (at bin 5) **(b)** Target 2 (at bin 10) **(c)** Target 3 (at bin 13) **(d)** Target 4 (at bin 17)

Fig. 3: Time varying probability of detection for all targets as a function of pulse index

V. Conclusion

In this work we proposed a Dynamic Power Allocation (DynaPA) algorithm for transmit beamforming for sparse multi-target detection in MIMO radar. Our algorithm keeps track of past and current states (and the corresponding transitions) in order to judiciously allocate power for exploring new/undetected targets and exploiting the targets currently detected. In the future, it will be interesting to combine the main ideas behind DynaPA with other RL-based transmit beamforming methods and further enhance their performance.

REFERENCES

[1] P. Hügler, F. Roos, M. Schartel, M. Geiger, and C. Waldschmidt, "Radar taking off: New capabilities for uavs," *IEEE Microwave Magazine*, vol. 19, no. 7, pp. 43–53, 2018.

[2] D. Ma, N. Shlezinger, T. Huang, Y. Liu, and Y. C. Eldar, "Joint radar-communication strategies for autonomous vehicles: Combining two key automotive technologies," *IEEE signal processing magazine*, vol. 37, no. 4, pp. 85–97, 2020.

[3] K. V. Mishra, M. B. Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 100–114, 2019.

[4] S. Sun, A. P. Petropulu, and H. V. Poor, "Mimo radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98–117, 2020.

[5] A. Charlish, F. Hoffmann, R. Klemm, U. Nickel, and C. Gierull, "Cognitive radar management," *Novel Radar Techniques and Applications*, vol. 2, pp. 157–193, 2017.

[6] H. Godrich, A. P. Petropulu, and H. V. Poor, "Power allocation strategies for target localization in distributed multiple-radar architectures," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3226–3240, 2011.

[7] W. Yi, Y. Yuan, R. Hoseinnezhad, and L. Kong, "Resource Scheduling for Distributed Multi-Target Tracking in Netted Colocated MIMO Radar Systems," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1602–1617, 2019.

[8] S. Haykin, "Cognitive radar: a way of the future," *IEEE signal processing magazine*, vol. 23, no. 1, pp. 30–40, 2006.

[9] K. L. Bell, C. J. Baker, G. E. Smith, J. T. Johnson, and M. Rangaswamy, "Cognitive Radar Framework for Target Detection and Tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 8, pp. 1427–1439, 2015.

[10] A. M. Ahmed, A. A. Ahmad, S. Fortunati, A. Sezgin, M. S. Greco, and F. Gini, "A Reinforcement Learning Based Approach for Multitarget Detection in Massive MIMO Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 2622–2636, 2020.

[11] S. Fortunati, L. Sanguinetti, F. Gini, M. S. Greco, and B. Himed, "Massive MIMO Radar for Target Detection," *IEEE Transactions on Signal Processing*, vol. 68, pp. 859–871, 2019.

[12] F. Lisi, S. Fortunati, M. S. Greco, and F. Gini, "Enhancement of a State-of-the-Art RL-Based Detection Algorithm for Massive MIMO Radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 6, pp. 5925–5931, 2022.

[13] W. Zhai, X. Wang, M. S. Greco, and F. Gini, "Weak target detection in massive mimo radar via an improved reinforcement learning approach," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4993–4997, IEEE, 2022.

[14] W. Zhai, X. Wang, X. Cao, M. S. Greco, and F. Gini, "Reinforcement Learning Based Dual-Functional Massive MIMO Systems for Multi-Target Detection and Communications," *IEEE Transactions on Signal Processing*, vol. 71, pp. 741–755, 2023.

[15] W. Zhai, X. Wang, M. S. Greco, and F. Gini, "Reinforcement Learning based Integrated Sensing and Communication for Automotive MIMO Radar," *2023 IEEE Radar Conference (RadarConf23)*, vol. 00, pp. 1–6, 2023.

[16] R. Rajamäki and P. Pal, "Array-informed waveform design for active sensing: Diversity, redundancy, and identifiability," *arXiv preprint arXiv:2305.06478*, 2023.

[17] J. Li and P. Stoica, "Mimo radar with colocated antennas," *IEEE signal processing magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[18] C.-Y. Chen and P. P. Vaidyanathan, "Mimo radar space–time adaptive processing using prolate spheroidal wave functions," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 623–635, 2008.

[19] B. Friedlander, "On signal models for mimo radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3655–3660, 2012.