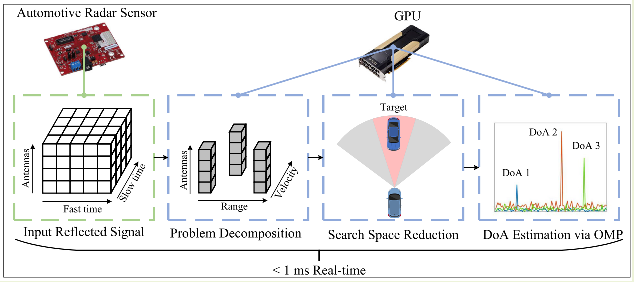
A Real-Time Super-Resolution DoA Estimation Algorithm for Automotive Radar Sensor

Yubo Wu, *Graduate Student Member, IEEE*, Chengzhang Li, *Member, IEEE*,
Y. Thomas Hou, *Fellow, IEEE*, and Wenjing Lou, *Fellow, IEEE*



***Abstract*—Direction-of-arrival (DoA) estimation is critical to obtaining precise information of targets for automotive radar sensors. To obtain fine radar imaging, super-resolution DoA estimation is needed to distinguish adjacent targets. Moreover, to adapt to the rapidly changing driving dynamics, the processing time for DoA estimation must meet very stringent timing requirement. Unfortunately, none of the existing methods can achieve *both* super-resolution *and* real-time requirement at the same time. In this article, we present the maximum-likelihood-based real-time super-resolution (MARS)—a real-time super-resolution DoA estimation algorithm. The main idea in MARS is to use maximum likelihood estimation (MLE) as the objective function. Unlike traditional MLE, MARS exploits the intrinsic correlation between the input data of adjacent time slots to substantially reduce the search space. Moreover, instead of using an exhaustive search to find the DoA solution, we employ the compressed sensing algorithm orthogonal matching pursuit (OMP) to efficiently discover an optimal or near-optimal solution for the MLE objective function based on the reduced search space. To further accelerate computation time, MARS decomposes the problems in each step into independent subproblems that can be efficiently executed on a GPU parallel computing platform. Simulation experiments show that MARS can achieve super-resolution in DoA estimation under 1 ms. Compared to state-of-the-art algorithms such as multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance techniques (ESPRIT), MARS outperforms both of them in DoA estimation while being the only known algorithm that can meet the stringent real-time requirement. Hardware experiments further illustrate that MARS outperforms state-of-the-art algorithms in target detection by achieving superior resolution.**

***Index Terms*— Automotive radar, direction-of-arrival (DoA) estimation, real time, super-resolution.**

I. INTRODUCTION

TECHNOLOGIES for autonomous driving have advanced rapidly in recent years and have the potential to revolutionize future transportation. A critical component to achieving a high level of automation (or full automation) is to obtain accurate information of targets on the road. Although camera and LiDAR can offer a fine resolution of imaging, they have their own limitations. For example, both camera and LiDAR are sensitive to weather conditions (e.g., fog, rain, and snow). For camera, although it can capture high-resolution images, it is not able to offer information on range, velocity, and direction-of-arrival (DoA) directly and needs other auxiliary algorithms to compute the information. In contrast, radar sensors are robust to weather conditions, inexpensive, and highly reliable. Due to these advantages, radar sensors are expected to remain indispensable in current advanced driver assistant systems (ADASs) and future autonomous driving systems [2].

An automotive radar sensor can simultaneously transmit and receive electromagnetic (EM) waves in the millimeter-wave (mmWave) frequency range of 76–81 GHz [3]. It can extract information about targets based on the EM waves reflected from the targets. For each target, an automotive radar sensor can be used to estimate range, velocity, and DoA [4], [5]. The range of the target can be determined by measuring the round-trip time delay that the EM waves take to propagate to and from a target. Estimation of the target velocity is based on the Doppler effect. The DoA estimation can be made by means

Received 5 September 2024; accepted 8 September 2024. Date of publication 23 September 2024; date of current version 14 November 2024. This work was supported in part by NSF under Grant CNS-2235232 and Grant CNS-1837519, Virginia Commonwealth Cyber Initiative (CCI), and Virginia Tech Institute for Critical Technology and Applied Science (ICTAS). An earlier version of this paper was presented at the Proceedings of the 18th European Radar Conference (EuRAD), London, U.K., April 5–7, 2022 [DOI: 10.23919/EuRAD50154.2022.9784501]. The associate editor coordinating the review of this article and approving it for publication was Dr. Sarbajit Paul. *(Corresponding author: Y. Thomas Hou.)*

Yubo Wu and Y. Thomas Hou are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: wuyubo@vt.edu; thou@vt.edu).

Chengzhang Li is with the Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: li.13488@osu.edu).

Wenjing Lou is with the Department of Computer Science, Virginia Tech, Arlington, VA 22203 USA (e-mail: wjlou@vt.edu).

Digital Object Identifier 10.1109/JSEN.2024.3462350

of an antenna array. By collecting reflected signals across different antenna elements, DoAs can be estimated from the phase offset between the received signals.

A fundamental problem in automotive radar sensor is how to achieve high-resolution estimation of all three parameters in real time [6]. In automotive systems, the target detection process, which spans from the reception of the reflected signal to the generation of detection results, should occur within 25–40 ms [7]. Considering additional processes such as down-chirp, passing through a low-pass filter (LPF), and possibly interference mitigation, the time available for parameter estimation is very limited. Although there is no explicit standard in the literature on the exact number, we believe that faster parameter estimation is preferable and our goal is to achieve ∼1-ms real-time estimation. The range and velocity estimation problems have already been well-solved by the hardware-based fast Fourier transform (FFT). Current automotive sensors can already achieve 30-cm range resolution and 2-m/s velocity resolution, respectively, in real time [3]. But for DoA estimation, it remains a challenging problem in the research community and is the focus of this article.

For DoA estimation, FFT-based algorithms require a prohibitively large number of antennas to achieve high resolution. For instance, to attain a resolution of 1° across a $[-50^\circ, 50^\circ]$ field of view (FOV), an FFT-based algorithm would require up to 100 antennas in one dimension. Hence, there is a critical need to achieve high resolution by using only a (reasonable) limited number of antennas [2], [3], [8], [9]. According to [10], *super-resolution* is defined as the ability of a DoA estimation method to surpass the natural resolution limit imposed by the FFT-based approach, which is fundamentally constrained by the number of antennas. It is important to note that the term—super-resolution—*does not* prescribe a specific degree of resolution. It merely refers to an algorithm's ability to achieve a finer DoA estimation resolution *beyond* what can be achieved by an FFT-based algorithm with practical constraints on its antennas.

The state-of-the-art super-resolution algorithms include two categories: maximum likelihood and subspace-based algorithms. The maximum likelihood estimation (MLE) algorithm [11], [12] can achieve super-resolution and enjoys high estimation accuracy but its computation time is prohibitively high, due to the large size of its search space and exhaustive search. Subspace-based algorithms such as multiple signal classification (MUSIC) [13] and estimation of signal parameters via rotational invariance techniques (ESPRIT) [14] can offer super-resolution more efficiently compared to MLE, but their computation time is still too high to meet the real-time requirement. They also suffer from poor performance under low SNR [15]. More discussions on related work are given in Section II.

In this article, we present a novel maximum-likelihood-based real-time super-resolution (MARS) algorithm. The main idea of MARS is to retain the MLE's capability of finding an optimal solution and to exploit its unique mathematical structure for parallel computation. The main contributions of MARS are the following.

1) We employ MLE as our objective function as MLE is most capable of finding an optimal DoA that meets our super-resolution requirement. Another benefit of choosing MLE over possible approaches is that MLE only requires a single snapshot (small amount of samples) and thus is more flexible and responsive than multisnapshot-based algorithms (e.g., MUSIC and ESPRIT). The challenge now is how to solve an MLE problem in ∼1-ms timescale.
2) To address this problem, we propose to exploit the correlation between the input data of adjacent time slots. Since the range, velocity, and DoA of each target only have modest change over two consecutive time slots, we propose to use the estimation results from the previous time slot to form a promising and reduced search space. Specifically, we use the DoA estimation result from the previous time slot as the center of the new search space and the possible deviation of DoA as the radius of the search space. This step effectively reduces the search space from the original entire FOV to a much narrower angle set.
3) After identifying the reduced search space, we utilize the compressed sensing algorithm orthogonal matching pursuit (OMP) [16] instead of an exhaustive search to find the optimal or near-optimal solution to MLE objective function. OMP significantly enhances time efficiency and substantially reduces processing time.
4) To reduce computation time and meet our real-time requirement, we propose to decompose the problems at each step into independent subproblems and employ commercial off-the-shelf (COTS) GPU to perform parallel computation. First, the DoA estimation problem is divided into many independent subproblems, which are executed in different blocks in parallel. For each subproblem, all the elements in the reduced search space are calculated in different threads in parallel, and the parallel reduction technique is used to find the optimal solution. Finally, by combining the solutions from all the subproblems, the original DoA estimation problem is solved. Throughout the process, we make meticulous engineering efforts to exploit the full power of GPU computing, including generating a sufficiently large number of threads, practicing a proper indexing method, and using shared memory wisely.
5) Through simulation experiments, we find that MARS can handle up to 300 reflection points (also known as targets) with super-resolution within 1 ms. The 300 reflection points correspond to 15–20 objects (as each object contains a dozen or so reflection points). By comparing MARS to MUSIC and ESPRIT, we find that MARS is the only algorithm that can achieve under 1-ms real-time processing. Further, MARS can offer better estimation accuracy than MUSIC and ESPRIT under the same SNR. To date, MARS is the only known DoA estimation algorithm that can achieve super-resolution in real time (sub-ms). Further hardware experiments show that MARS outperforms the state-of-the-art algorithms by achieving higher resolution and better identification of close objects.

The rest of this article is as follows. In Section II, we review the related work on DoA estimation for automotive radar sensor. Section III describes the background and problem statement. In Section IV, we present MARS. Section V elaborates on the GPU implementation. Section VI details the results of simulation experiments, while Section VII provides the outcomes of hardware experiments. Section VIII concludes this article.

II. RELATED WORK

Across the research landscape of DoA estimation for automotive radar sensor [2], [17], the most active branch is subspace-based methods (see [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]). Another branch is FFT-based methods based on novel antenna designs (see [41], [42], [54]). Apart from these two branches, there are some other miscellaneous methods reported in the literature (see [43], [44], [47]). In the rest of this section, we review the state of the art in each of these branches.

1) Subspace-Based Algorithms: Subspace-based algorithms exploit the phenomenon that when there are enough samples collected under a sufficiently high SNR, eigenvalues (corresponding to DoAs) and associated eigenvectors (corresponding to the incident signals) can be obtained from the sample covariance matrix. There are two families of algorithms in this class, namely, MUSIC-based algorithms [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30] and ESPRIT-based algorithms [31], [32], [33], [34], [35], [36], [37], [38].

In the original MUSIC algorithm [13], the DoAs are found by searching within the subspace formed by all the eigenvectors. There are many variations and extensions of MUSIC. References [18], [19], [20], [21], and [22] proposed to improve the performance of MUSIC under some special conditions by considering the mismatch of signal phase [18], the nonuniform antenna array [19], the coexistence of the far-field and near-field targets [20], [21], and impulsive noise [22]. However, the complexities of these algorithms are higher than the original MUSIC and the performance under the general conditions may not be improved. References [23], [24], [25], and [26] implemented MUSIC on different kinds of hardware platforms. Although the computation time was made shorter than before, the eigen decomposition of the large covariance matrix still remains and the final computational time is still over ~100 ms. References [27], [28], [29], and [30] modified the MUSIC algorithm to make it work efficiently under different multi-in multi-out (MIMO) antenna designs. However, the complexities of these algorithms are much higher than the 1-D array system.

ESPRIT [14] shares the same basic idea with MUSIC but utilizes the rotation invariance property to avoid the search process. There are some variations and extensions of ESPRIT. References [31], [32], [33], [34], and [35] improved the original ESPRIT algorithm in terms of estimation accuracy and robustness of different antenna settings. Zheng et al. [36], Li et al. [37], and Yang et al. [38] implemented ESPRIT in a bistatic MIMO system to support more signals and better DoA estimation performance.

Although some of these subspace-based algorithms can provide better performance, they still suffer from the fundamental deficiencies of MUSIC and ESPRIT: high complexity matrix operations on large covariance matrix, multiple snapshots requirement, and poor performance under low SNR. As a result, none of them can achieve super-resolution in real time on computing platforms typically used in automotive systems.[1]

2) FFT-Based Algorithms: By applying FFT to the received signals across the antenna domain, DoAs can be estimated based on the spectrum peaks. References [39], [40], [41], and [42] proposed hardware-based FFT algorithms. By using the custom-designed parallel structure and high-performance hardware [e.g., field-programmable gate array (FPGA)], the DoA estimation problem can be solved in real time. Traditional FFT has a low resolution for DoA estimation due to the limited number of antennas. Although the authors proposed to employ MIMO [39], [40] and virtual antenna techniques [41] to improve DoA resolution, the number of required antenna elements is prohibitively large if one wishes to achieve high resolution (e.g., 1°).

3) Other Methods: Zhang et al. [43], Guo et al. [44], and Sun and Zhang [45] exploited a novel antenna design, i.e., compressed sparse array, to achieve a high resolution for DoA estimation with a fewer number of antennas than the general uniform 1-D array. However, the compressed sparse array-based algorithm still cannot meet the super-resolution requirement unless it employs a prohibitively large number of antennas. Zheng et al. [46] proposed a deep neural network-based target detection algorithm to achieve high DoA resolution for automotive radar sensor. However, this algorithm requires a high-resolution imaging radar sensor equipped with 86 horizontal elements, which is prohibitively expensive and impractical for the mass commercial market. Engels et al. [47] used a multidimensional frequency model to estimate range, velocity, and DoA. It is an iterative method, with the estimation results being improved in each iteration. Its computation time cannot meet the timing requirement due to too many iterations. References [48], [49], [50], [51], and [52] utilize Newtonized orthogonal matching pursuit (NOMP) to estimate the DoA, taking advantage of its capability to locate DoAs in the continuous domain and its robustness against high noise and clutter power. However, the method's extensive search space and iterative process present significant challenges in meeting the ~1-ms real-time requirement. In summary, none of the existing DoA estimation algorithms can achieve *both* super-resolution *and* real-time requirement at the same time on computing platforms typically used in automotive systems.

III. BACKGROUND AND PROBLEM STATEMENT

In this section, we provide essential background on the signal model for a general DoA estimation problem. We assume the automotive radar system employs frequency-modulated continuous wave (FMCW), which is the most prevalent waveform used in the field [53]. Fig. 1 illustrates the standard reference model, where a moving target with a relative range R (in m) and a relative velocity V (in m/s) to an FMCW radar sensor that is equipped with an array of M antenna elements. Denote $m = 1, \ldots, M$ as the index of

---

[1] In this article, we consider only typical platforms including CPUs, FPGAs, and GPUs as discussed in [23], [24], [25], and [26]. High-end platforms such as the NVIDIA H100 are not considered due to their prohibitive cost, making them impractical for automotive systems.

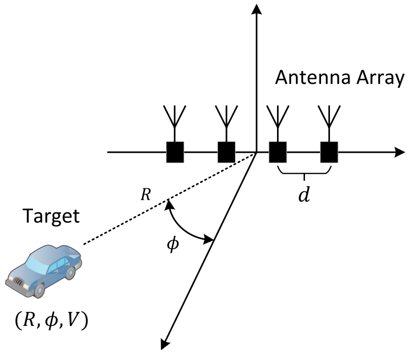


Fig. 1. Reference system model.

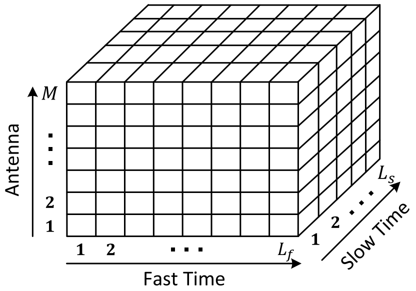


Fig. 2. Measurement $x[l_f, l_s, m]$ along the fast time (l_f), the slow time (l_s), and the space dimension (m).

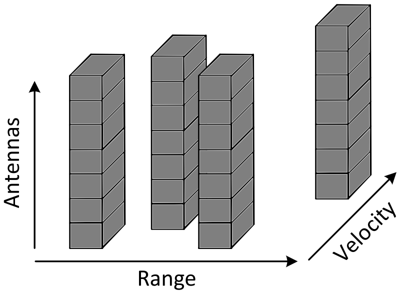


Fig. 3. Range–velocity bins with spectrum peaks after 2-D FFT.

the receiving antenna element. The spacing between adjacent antenna elements is d m. Denote ϕ as the azimuth angle of the target to the receiving antenna.

The carrier frequency of the FMCW radar sensor is f_c (in Hz). Then, the corresponding wavelength λ is $\lambda = C/f_c$, where C is the speed of light. Denote T_c (in s) as the chirp duration and S_c (in Hz/s) as the chirp slope of the radar sensor. Denote the sampling rate of the radar receiver as f_s (in Hz). Denote l_f as the sampling index within one chirp (also known as the fast-time index), i.e., $l_f = 1, \ldots, L_f$, where $L_f = f_s \cdot T_c$. Denote l_s as the index of chirps within one coherent processing interval (CPI) (also known as the slow-time index), i.e., $l_s = 1, \ldots, L_s$.

We assume additive white Gaussian noise (AWGN) and denote it as $\omega[l_f, l_s, m] \in \mathcal{CN}(0, \sigma^2)$. Denote $\xi[l_f, l_s, m]$ as the attenuation caused by the path loss, the antenna gain, and the radar cross section (RCS) of the target. The output of the radar receiver within the time length of one snapshot (corresponding to L_s consecutive chirps) $T_c L_s$ can be represented as

$$\begin{aligned} & x\left[l_f, l_s, m\right] \\ & \approx \xi\left[l_f, l_s, m\right] \exp\left\{ j2\pi \left[(f_b + f_d)\frac{l_f}{f_s} \right.\right. \\ & \qquad \left.\left. + \frac{f_c m d \sin(\phi)}{C} + f_d l_s T_c + \frac{2 f_c R}{C} \right] \right\} \\ & \quad + \omega\left[l_f, l_s, m\right] \\ & \qquad \text{for } l_f = 1, \ldots, L_f, \quad l_s = 1, \ldots, L_s \\ & \qquad m = 1, \ldots, M \end{aligned} \tag{1}$$

where $f_b = (2 S_c R / C)$ is the beat frequency and is dependent on the range of the target; and $f_d = (2V/\lambda)$ is the Doppler shift due to the movement of the target.

From (1), the measurement $x[l_f, l_s, m]$ spans over the fast time, the slow time, and the space dimension, as illustrated in Fig. 2. The basic radar operation is to extract range R, relative velocity V, and DoA ϕ of each target from (1) [2]. With today's hardware-based 2-D FFT [3], [55], [56], high-resolution range (30 cm) and velocity (2 m/s) can be readily obtained in real time. Specifically, for velocity V estimation, the Doppler shift f_d can be extracted by L_s-point FFT across the slow-time dimension. Similarly, f_b (and thus, range R) can be extracted by L_f-point FFT across the fast-time dimension. The spectrum after 2-D FFT spans over a 2-D range–velocity grid. When there are Q targets, the output of the radar receiver can be expressed as an extension of (1), represented by $\sum_{q=1}^{Q} x_q[l_f, l_s, m]$. In this scenario, multiple spectrum peaks may exist after 2-D FFT, which are identifiable by range and velocity, as illustrated in Fig. 3.

The output after 2-D FFT can be used as the input for DoA estimation. Note that each peak may be associated with multiple incident signals from a cluster of targets with the same range and velocity but different DoAs. Denote N_e as the total number of bins containing the spectrum peaks, and denote N_i as the number of DoAs to measure in the ith bin. Then, the peak in the ith bin corresponds to a linear combination of N_i incident signals. The input data for DoA estimation at the mth antenna element can be obtained as

$$x_i[m] = \sum_{n=1}^{N_i} \xi_n^i[m] e^{j2\pi \frac{md\sin(\phi_n^i)}{\lambda}} + \omega[m]. \tag{2}$$

For M antennas, denote set $\mathcal{X}_i = \{x_i[1], \ldots, x_i[M]\}$. For a total of N_e bins, denote $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_{N_e}\}$.

The goal of DoA estimation is to design a method (denoted as Ψ) to estimate ϕ_n^i via $\tilde{\phi}_n^i = \Psi(\mathcal{X})$. Among the various objectives employed in DoA estimation, we have opted to use the *least-squares method,* which is also utilized in both MUSIC and ESPRIT, due to its optimality at moderately high SNR. The objective function is defined as follows:

$$\min_{\Psi} \sum_{i=1}^{N_e} \sum_{n=1}^{N_i} \left| \phi_n^i - \tilde{\phi}_n^i \right|^2. \tag{3}$$

As discussed in Section II, none of the existing DoA estimation algorithms can solve (3) with super-resolution requirement in real-time on computing platforms typically used in automotive systems.

IV. MARS: A NOVEL MAXIMUM-LIKELIHOOD-BASED REAL-TIME SUPER-RESOLUTION ALGORITHM

The goal of this article is to develop an algorithm that can achieve super-resolution in DoA estimation in real time (under 1 ms). The main idea of MARS is as follows. First, MARS decomposes the DoA estimation problem into a large number of independent subproblems so that all subproblems can be solved in parallel simultaneously. Second, for each subproblem, one range–velocity bin could be processed with a simplified MLE objective function that can provide an optimal solution. Third, to address the prohibitive computation time associated with the MLE, MARS judiciously reduces the search space into a smaller but promising reduced search space by exploiting the intrinsic correlations in range, velocity, and DoA of a target between successive time slots. Fourth, leveraging the reduced search space, the compressed sensing algorithm OMP is employed to efficiently calculate the final estimated DoA. In each step of MARS, we strive to exploit parallel computing to accelerate computation time. Finally, we show how to speed up the initialization process with a limited number of iterations. We describe these ideas in detail in the rest of this section.

A. Problem Decomposition

We assume the estimation of range and velocity has already been handled by hardware-based FFT (see Fig. 3). The input data for our DoA estimation will be a large number of range–velocity bins with spectrum peaks, which contain the targets that need to be detected. Here, all the targets are divided into different clusters which have been already discriminated from others by range and velocity. So for DoA estimation, only the targets with the same range and velocity are coupled with each other, which means in (3), $\tilde{\phi}_n^i = \Psi(\mathcal{X}_i)$. Therefore, instead of solving (3) directly, we can decompose (3) into N_e independent subproblems as follows:

$$\min_{\Psi} \sum_{n=1}^{N_i} |\phi_n^i - \tilde{\phi}_n^i|^2, \quad i = 1, \ldots, N_e. \tag{4}$$

Since all the N_e subproblems can be solved in parallel, the computation time can be reduced significantly.

B. Simplifying MLE Objective Function

To obtain an optimal solution to (4), we choose to use the MLE for Ψ, assuming the complex Gaussian noise at each antenna is temporally and spatially independent. The rationale for choosing MLE includes several points. First, unlike FFT-based methods, MLE can achieve super-resolution without being limited by the number of antennas. Second, whereas super-resolution algorithms like MUSIC and ESPRIT require multiple snapshots, MLE only needs a single snapshot, thus providing enhanced flexibility. Third, both MUSIC and ESPRIT depend on the condition that the eigenvalues and eigenvectors of the sample covariance matrix accurately represent the MLE estimate, but this holds true only in scenarios with sufficiently high SNR; their performance deteriorates substantially in low SNR environments. In contrast, MLE consistently delivers more accurate DoA estimation even under high noise levels.

To solve the ith subproblem in (4), we can use the following equivalent expression by using simplified MLE:

$$\min_{\tilde{\phi}_n^i} \sum_{m=1}^{M} \left| x_i[m] - \sum_{n=1}^{N_i} \xi_n^i \exp\left[j2\pi \left(\frac{f_c m d \sin(\tilde{\phi}_n^i)}{c} \right) \right] \right|^2. \tag{5}$$

Denote the search space of the decision variable $\tilde{\phi}_n^i$ in the above subproblem as $\{\phi\}$. The size of $\{\phi\}$ is determined by the following three parameters.

1) *Field of View:* The field of view, denoted as FOV$[-\theta, \theta]$, is the angle through which a radar is sensitive to its reflected signal. FOV$[-\theta, \theta]$ differs among different types of radar sensors. For example, a short-range radar may have an azimuthal FOV of $[-80^\circ, 80^\circ]$, and a long-range radar may just have an FOV of $[-15^\circ, 15^\circ]$. A larger FOV means a larger search space.
2) *Granularity:* The granularity is determined by the required resolution and estimation accuracy. For example, if 1° resolution is needed, the granularity should be no greater than 1°.
3) *Number of Targets in Each Range–Velocity Bin:* The number of targets falling in the same range–velocity bin N_i corresponds to the number of reflection points (from the same or different objects). When N_i increases, the size of the search space will become larger.

For example, for a general short-range radar sensor with an FOV $[-80^\circ, 80^\circ]$, to achieve 1° resolution, the search space size of one bin can be up to 10^5. Due to this large space, exhaustive search cannot be used.

To meet the stringent time requirement, the size of the search space must be reduced. Among the three parameters that determine the size of the search space, granularity and the number of targets in each range–velocity bin cannot be reduced. This is because the granularity must be small enough to meet the super-resolution requirement and the number of targets in each range–velocity bin is determined by the traffic on the road that can be high in dense scenarios. Hence, the only parameter that can be used to reduce the search space is FOV. So instead of blindly using the FOV of the radar sensor, we propose to reduce its size by exploiting the correlation between estimation results from adjacent time slots.

Specifically, we explore data association to establish a connection between data of adjacent time slots. This is described in detail in Section IV-C. For data in the current time slot that has correlation with some data in the previous time slot, we propose to use the DoA estimation result from the previous time slot as the center of the reduced search space and possible DoA deviation as the radius to identify a much smaller but promising search space for (5). This is described in detail in Section IV-D. Based on the reduced search space for DoA, OMP is employed to obtain an optimal or near-optimal solution to (5). This is detailed in Section IV-E. For data in the current time slot that does not have correlation with any data in the previous time slot, we treat it as the initialization phase and propose to address it with a speed-up algorithm. This is described in detail in Section IV-F.

C. Data Association Between Successive Time Slots

In each time slot, after using 2-D FFT, we can find range–velocity bins corresponding to spectrum peaks, as shown in Fig. 2. To exploit the potential correlation between data from adjacent time slots, the estimation results from the previous time slot (including all the coordinates and DoAs) should be saved. For the current time slot, after we get the range and velocity of each bin, we can associate the current bins with the saved estimation results from the last time slot before we perform DoA estimation. Since the relative range and velocity only change slightly between adjacent time slots (even for a vehicle moving at a high speed), we propose a shortest-path method as follows.

Denote $E(t-1) = \{(R_1(t-1), V_1(t-1)), \ldots, (R_{N_e(t-1)}(t-1), V_{N_e(t-1)}(t-1))\}$ as the estimation results for range and velocity of all $N_e(t-1)$ bins at time slot $(t-1)$. Denote $E(t) = \{(R_1(t), V_1(t)), \ldots, (R_{N_e(t)}(t), V_{N_e(t)}(t))\}$ as the current estimation results of all $N_e(t)$ bins at time slot t. The data association problem can be formulated as follows:

$$\min_{f} \sum_{i=1}^{N_e(t)} d_{i,f(i)} \tag{6}$$

where $f : E(t) \to E(t-1)$ is a function that maps elements in $E(t)$ to elements in $E(t-1)$; $i \in [1, \ldots, N_e(t)]$ denotes index of the bin in the current time slot t and $f(i) \in [1, \ldots, N_e(t-1)]$ represents index of the bin in the previous time slot $t-1$. $d_{i,f(i)}$ is the Euclidean distance between the ith bin in the current time slot t [with coordinates $(R_i(t), V_i(t))$] and the $f(i)$th bin in the previous time slot $(t-1)$ [with coordinates $(R_{f(i)}(t-1), V_{f(i)}(t-1))$]. This Euclidean distance is calculated as

$$d_{i,f(i)} = \sqrt{\left(\frac{R_i(t)-R_{f(i)}(t-1)}{R_r}\right)^2 + \left(\frac{V_i(t)-V_{f(i)}(t-1)}{V_r}\right)^2} \tag{7}$$

where R_r and V_r are the range and velocity resolutions used for normalization. It should be noted that f is not necessarily a one-to-one mapping; it can be a many-to-one mapping, i.e., multiple current bins may be associated with the same previous bin. This implies that targets, which were previously located within a single bin, have now moved to two separate range–velocity bins.

When the total number of bins $N_e(t)$ becomes large, the time consumption of solving the data association problem in (6) via brute force can be quite long. To accelerate the computation time and meet the stringent real-time requirement, we will propose a better algorithm by exploiting the following observation.

We observe that the movement of each target on the road is continuous. Thus, range, velocity, and DoA of a target only change slightly over two successive slots. For example, suppose two vehicles traveling in opposite directions (i.e., toward each other) at 120 km/h (33.33 m/s), or 66.67 m/s in terms of relative velocity. With a range resolution of $R_r = 30$ cm and a maximum 10 ms time estimation interval,[2] according to (7), $((R_i(t) - R_{f(i)}(t-1))/R_r) = ((66.67 \times 10^{-2})/0.3) = 2.2$ and $((V_i(t) - V_{f(i)}(t-1))/V_r) = 0$ (assuming both vehicles are traveling at a constant speed). Hence, the distance $d_{i,f(i)} = 2.2$, which is a small number.

Since d in (7) is a small number in practice, for each current bin, we propose to calculate its associated bin in the previous time slot independently. By doing so, the original problem (6) becomes $N_e(t)$ number of independent subproblems with much lower complexity and can be solved in parallel. Specifically, for the ith bin at current time t, define the associated bin $f(i) = k_i$ in $(t-1)$ as

$$k_i = \arg\min_{j} d_{i,j} \tag{8}$$

where $j = 1, 2, \ldots, N_e(t-1)$ is the index of the bins in the last time slot $(t-1)$. To find k_i, the search space can be reduced to within a range of Δ (e.g., 2.2) that is centered around $[R_i(t), V_i(t)]$. There are two cases.

1) If we can find k_i within the reduced search space for (8), then bin i can be associated with the nearest neighbor bin k_i. With the previous estimation results including the range, velocity, and DoAs of k_i, we can form a reduced search space for (5). We will describe how to form this new search space in detail in Section IV-D.
2) Otherwise, there are two scenarios. One is the start of the engine; since there is no estimation before, no previous information can be utilized. The other is the first appearance of new targets. That is, there are new targets coming into the detection range of the radar, and thus, there is no correlated information in the previous time slot. Both scenarios can be treated as the initialization phase and the search space for the new bin cannot be reduced. How to speed up DoA estimation for the initialization phase will be addressed in Section IV-F.

D. Reducing Search Space

For the ith range–velocity bin (after 2-D FFT) at the current time slot t, if we can find k_i in (8) within a range Δ, then we can form a reduced search space for (5) where the real DoA ϕ_n^i is located. For each target n in k_ith bin in previous time slot [i.e., $(t-1)$], a subspace for DoA estimation in time t can be formed as follows:

$$s_n^{k_i} = \left\{\phi \mid \tilde{\phi}_n^{k_i} - e_n^{k_i} \le \phi \le \tilde{\phi}_n^{k_i} + e_n^{k_i}\right\} \tag{9}$$

for $n = 1, \ldots, N_{k_i}$, where n is the index of the targets in k_ith bin at $(t-1)$ and N_{k_i} is the total number of targets in the k_ith bin at $(t-1)$, $\tilde{\phi}_n^{k_i}$ is the last DoA estimation result of target n and is used as the center of the subspace, and $e_n^{k_i}$ is the expected DoA deviation between two adjacent time slots (i.e., the radius of the subspace).

To set the value of $e_n^{k_i}$, refer to Fig. 4. In Fig. 4, the DoA is the last time estimation result $\tilde{\phi}_n^{k_i}$. The range (distance) and relative speed between the target and radar are $((R_i(t) + R_{k_i}(t-1))/2)$ and $((V_i(t) + V_{k_i}(t-1))/2)$, respectively, both of which are averages between adjacent time slots. The tangential speed of the target with respect to range is $((V_i(t) + V_{k_i}(t-1))/2)\tan(\tilde{\phi}_n^{k_i})$. Assume T_s is the time interval between two consecutive estimations. Based on the fact that the DoA has a modest change between adjacent time slots, the DoA deviation (in degrees) caused by mobility can be calculated using the angular speed as

---

[2] T_s should not be confused with DoA estimation time, which is 1 ms in this article. Estimation interval refers to time interval between adjacent estimations and should not exceed 10 ms which is the time starting from receiving signals at the radar to eventually taking specific actions on the vehicle [7]. Since at least one snapshot is needed for DoA estimation, the estimation interval should be larger than the time length of one snapshot $T_c L_s$.

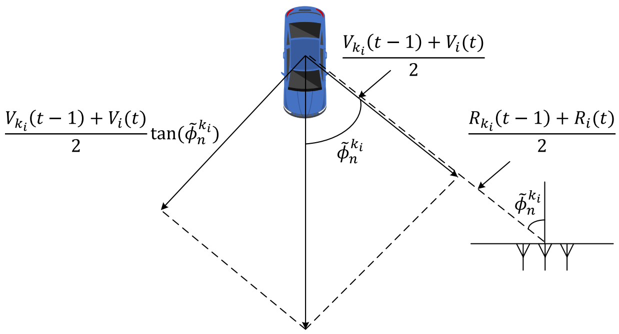


Fig. 4. Calculating DoA deviation $e_n^{k_i}$ from $\tilde{\phi}_n^{k_i}$ between time slots t and $(t-1)$.

$[180|(V_i(t)+V_{k_i}(t-1))\tan(\tilde{\phi}_n^{k_i})|T_s]/[\pi(R_i(t)+R_{k_i}(t-1))]$. Then, we can set $e_n^{k_i}$ (in degrees) to

$$e_n^{k_i} = \frac{180|V_i(t)+V_{k_i}(t-1)|\left|\tan\left(\tilde{\phi}_n^{k_i}\right)\right|T_s}{\pi\left(R_i(t)+R_{k_i}(t-1)\right)} + b \tag{10}$$

where the first term represents potential DoA deviation caused by mobility, and parameter b is used to account for potential estimation error due to noise. By specifying $e_n^{k_i}$ in (9), the size of the subspace $s_n^{k_i}$ for target n can be reduced substantially.

By combining all subspaces $s_1^{k_i}, \dots, s_{N_{k_i}}^{k_i}$, the new reduced search space S_i for the current ith subproblem is

$$S_i = \left(s_1^{k_i} \cup s_2^{k_i} \cdots \cup s_{N_{k_i}}^{k_i}\right) \cap \{\phi\} \tag{11}$$

where $\{\phi\}$ is the FOV of the radar. For each subproblem, (5) can be solved based on the reduced search space (11).

E. Estimating DoA Based on Compressed Sensing

After narrowing down the search space for DoA, solving MLE objective (5) remains challenging for several reasons. First, the number of targets in each range–velocity bin, denoted by N_i, is unknown, which increases both the search space and the time required for searching. Second, the attenuation factor ξ_n^i for each target n is also unknown, further expanding the search space. Third, given these issues, attempting to cover all possibilities through an exhaustive search would result in prohibitively long processing times, failing to meet real-time processing requirements. To address these challenges, we employ the compressed sensing algorithm OMP to compute the estimated DoA.

The DoA estimation problem can be reformulated as follows:

$$\begin{aligned} &\min_{\mathbf{a}} \ ||\mathbf{a}||_1 \\ &\text{s.t.} \ \ \mathbf{x}_i = \Omega \cdot \mathbf{a} \end{aligned} \tag{12}$$

where $||\cdot||1$ denotes the 1-norm, indicating a minimization objective focused on the sparsity of $\mathbf{a}$. The vector $\mathbf{x}_i$ represents the aggregated signal received across all antennas, expressed as $[x_i[1], \dots, x_i[M]]^{\mathrm{T}}$. The matrix Ω is the signal basis constructed from all possible DoAs in the reduced search space, expressed as $[D_1, \dots, D_{|S_i|}]$, where $|S_i|$ is the size of the reduced search space for the subproblem i. Each potential signal component D_j ($j = 1, \dots, |S_i|$) is defined as $[\exp[j2\pi((f_c \cdot 1 \cdot d\sin(\tilde{\phi}^i(j)))/c)], \dots, \exp[j2\pi((f_c \cdot M \cdot d\sin(\tilde{\phi}^i(j)))/c)]]^{\mathrm{T}}$, with $S_i = \{\tilde{\phi}^i(1), \dots, \tilde{\phi}^i(|S_i|)\}$. Given the observation that only a limited number of DoAs (or targets) are present within a single subproblem (a specific range–velocity bin), $\mathbf{a}$ is identified as a sparse vector. This vector is essentially a projection of the signals within the signal basis Ω, and (12) can be effectively resolved using a compressed sensing algorithm.

Among compressed sensing algorithms, we opt for OMP due to its time efficiency and suitability for implementation on parallel hardware platforms. The fundamental principle of OMP involves identifying the most plausible signals in each iteration and concluding the process once the signal's power threshold is met. After determining the sparse vector $\mathbf{x}_i$ by solving (12), the DoAs of the corresponding signals are selected as the final estimated DoA solution. During the implementation of OMP, several key considerations are as follows.

1) *Number of Iterations:* In OMP, each iteration identifies one DoA component, requiring multiple iterations to detect multiple targets. Our MARS restricts the total number of iterations for computational efficiency. Specifically, OMP is applied to each subproblem independently, so the number of iterations corresponds to the number of DoAs within each range–velocity bin. Given the fine range (30 cm) and velocity (2 m/s) resolution, the number of DoAs per bin is typically no more than three, as demonstrated in Section VI.
2) *DoA Spacing:* To achieve 1° super-resolution, the DoA spacing between $\tilde{\phi}^i(j)$ and $\tilde{\phi}^i(j+1)$ within the reduced search space Ω is set to 1°. However, the actual correlation between different signals in Ω depends on $\sin(\tilde{\phi}^i(j))$ and $\sin(\tilde{\phi}^i(j+1))$, as detailed in (5). Consequently, even with a consistent 1° DoA spacing, the correlation between signals varies as $\tilde{\phi}^i(j)$ changes. The impact of these varying correlations on the performance of OMP is evaluated in SectionVI.
3) *Stop Criterion:* In OMP, the iterations stop when either the number of iterations reaches the sparsity factor or the power of the residual signal falls below a certain threshold. Since the sparsity factor (the number of DoAs in each range–velocity bin) is unknown, we opt for the power threshold as the stopping criterion. In MARS, the power threshold of the residual signal is determined by the lowest power that the radar sensor can successfully detect.

F. Speeding Up Initialization

As discussed in Section IV-C, at the time of starting the engine or when there are new targets coming into the detection range of the radar, no previous information is available to exploit data correlation. Both scenarios will be treated as the initialization phase. Since we cannot find k_i in (8), a different DoA estimation method needs to be developed to solve (5).

By analyzing (5), we observe that it is an unimodal function with relatively high SNR within the whole search space FOV. Based on the observation, we propose an iterative intensification to solve (5).

In the first time slot, instead of calculating the objective function values of (5) for all possible DoAs (in the FOV $[-\theta, \theta]$) using OMP, we can just sample a small fixed number (say I) of DoAs (i.e., $\{-\theta, -\theta + (2\theta/(I-1)), \ldots, \theta - (2\theta/(I-1)), \theta\}$) to solve (5). After solving (5), the estimation result $\tilde{\phi}_n^i$ is among $\{-\theta, -\theta + (2\theta/(I-1)), \ldots, \theta - (2\theta/(I-1)), \theta\}$ and will have estimation errors due to sampling. Since the objective function is unimodal, among the I sampled DoAs, the samples within $[-\theta, \tilde{\phi}_n^i - (2\theta/(I-1))]$ have a decreasing objective value while the samples within $[\tilde{\phi}_n^i + (2\theta/(I-1)), \theta]$ have an increasing objective value. We can now narrow down the DoA search space to the most promising subspace of $[\tilde{\phi}_n^i - (2\theta/(I-1)), \tilde{\phi}_n^i + (2\theta/(I-1))]$ where the real DoA ϕ_n^i resides. We save both the estimation result $\tilde{\phi}_n^i$ and the subspace $[\tilde{\phi}_n^i - (2\theta/(I-1)), \tilde{\phi}_n^i + (2\theta/(I-1))]$ before we enter the next time slot.

In the next time slot, we form a reduced search space S_i with $s_n^{k_i}$ as follows:

$$s_n^{k_i} = \left\{ \phi \mid \tilde{\phi}_n^{k_i} - \frac{2\theta}{I-1} - e_n^{k_i} \le \phi \le \tilde{\phi}_n^{k_i} + \frac{2\theta}{I-1} + e_n^{k_i} \right\}. \tag{13}$$

Unlike (9) where $\tilde{\phi}_n^{k_i}$ is a close estimation of the real DoA $\phi_n^{k_i}$ at last time slot, in (13), since $\tilde{\phi}_n^{k_i}$ is not accurate, we use the stored promising subspace $[\tilde{\phi}_n^{k_i} - (2\theta/(I-1)), \tilde{\phi}_n^{k_i} + (2\theta/(I-1))]$ to form $s_n^{k_i}$. The expected DoA deviation $e_n^{k_i}$ can be calculated as follows:

$$e_n^{k_i} = \frac{180|V_i(t) + V_{k_i}(t-1)||\tan(\beta^*)|T_c L_s}{\pi(R_i(t) + R_{k_i}(t-1))} + b \tag{14}$$

where $|\tan(\beta^*)| = \max_\beta |\tan(\beta)|$ and $\beta \in [\tilde{\phi}_n^i - (2\theta/(I-1)), \tilde{\phi}_n^i + (2\theta/(I-1))]$. $|\tan(\beta^*)|$ is set as an upper bound to make sure the DoA deviation $e_n^{k_i}$ is large enough. Because $|\tan(\beta)|$ is a nonnegative convex function, $|\tan(\beta^*)|$ can be calculated as

$$\begin{aligned} &|\tan(\beta^*)| \\ &\quad = \max\left\{ \left|\tan\left(\tilde{\phi}_n^{k_i} - \frac{2\theta}{I-1}\right)\right|, \left|\tan\left(\tilde{\phi}_n^{k_i} + \frac{2\theta}{I-1}\right)\right| \right\}. \end{aligned} \tag{15}$$

In (14), instead of using T_s (e.g., 10 ms) as in (10), we use the time length of single snapshot $T_c L_s$ (e.g., 1 ms). Since the initialization process includes multiple estimations, the smaller the time interval between adjacent estimations is, the shorter the initialization process will be. Hence, we can use a single snapshot to accelerate the computation time.

After calculating all $s_n^{k_i}$'s in (13), we use (11) to form a reduced search space S_i. If the number of elements in S_i is no more than I, we can solve (5) by considering all the elements within S_i in parallel. After that, the initialization process is complete. Otherwise, we can continue sampling I DoAs within S_i and repeat the same process, until the number of elements in S_i is no more than I. After the initialization process is complete, the search space reduction method described in Section IV-D can be applied for further DoA estimation. The complete algorithm for the initialization process is shown in Algorithm 1.

During the initialization process, the parameter I is important as it affects both computation time and estimation accuracy. The larger the I is, the longer the computation time, and the more accurate the estimation result will be in each iteration. In our design, the value of I depends on the hardware of the parallel computing platform (see Section V for more details). This approach is justified by the fact that the number of targets with distinct DoAs (spaced greater than 1°) in each bin is relatively small——typically fewer than three—due to the high resolution of the range–velocity bin (30-cm range resolution and 2-m/s velocity resolution). As a result, the value of I is sufficiently large to ensure accuracy without compromising computational efficiency.

Algorithm 1 Initialization

```
if there is a new bin then
    Sample $I$ DoAs within the FOV$[-\theta, \theta]$ with the same interval $\frac{2\theta}{I-1}$ to solve (5) using OMP
    Save estimation result $\tilde{\phi}_n^i$ and sub-space $[\tilde{\phi}_n^i - \frac{2\theta}{I-1}, \tilde{\phi}_n^i + \frac{2\theta}{I-1}]$ for the next estimation time slot
    while receiving new snapshot do
        Form a reduced search space $S_i$ using (11) with $s_n^{k_i}$ (13) and $e_n^{k_i}$ (14).
        if the number of elements in $S_i$ is no more than $I$ then
            Solve (5) by considering all the elements within $S_i$ in parallel.
            Break
        else
            Sample $I$ DoAs within the $S_i$ with the same interval to solve (5)
            Save the estimation result $\tilde{\phi}_n^i$ and the most promising subspace for the next iteration.
        end if
    end while
end if
```

It is important to note that after the initialization process, the DoA estimation results may be suboptimal due to noise. However, by considering the noise factors as outlined in (9)–(11), MARS effectively mitigates errors from previous estimations. Consequently, any errors from the initialization process do not propagate across radar cycles.

G. Summary

The complete algorithm for MARS is given in Algorithm 2. In Algorithm 2, lines 1–4 show the start of the system. In the beginning, 2-D FFT is applied to estimate the range and velocity of all bins. Then, the initialization process in Section IV-F is used to estimate the DoAs of all targets since there is no previous information. Next, all the estimation results including range, velocity, and DoA will be saved for the next estimation. Lines 5–16 show the estimation process after receiving new input data at a time slot to do DoA estimation. Specifically, in line 6, 2-D FFT is applied to estimate the range and velocity of all bins. After that, for the new bins, in line 9, the initialization process in Section IV-F is used to estimate the DoAs. For the other bins, in line 11, the reduced search space in Section IV-D is used for DoA estimation. When the estimation is over, in line 15, all the estimation results will be saved for the next estimation.

Algorithm 2 MARS

```
Start of the radar:
Use 2-D FFT to estimate $R$ and $V$ for all range-velocity bins with spectrum peaks
Use initialization process in Section IV-F (Algorithm 1) to estimate DoAs for all bins
Save the estimation results including the coordinates of all range–velocity bins ($R$ and $V$) with spectrum peaks and the corresponding estimated DoA $\{\phi\}$
while at a time slot to do DoA estimation do
    Use 2-D FFT to estimate $R$ and $V$ for all range-velocity bins with spectrum peaks
    for each bin do
        if it is a new bin then
            Use initialization process in Section IV-F to estimate $\{\phi\}$.
        else
            Reduce search space based on Section IV-D
            Estimate DoA using OMP by considering all the elements in the reduced search space in parallel
        end if
    end for
    Save the current estimation results for next estimation
end while
```

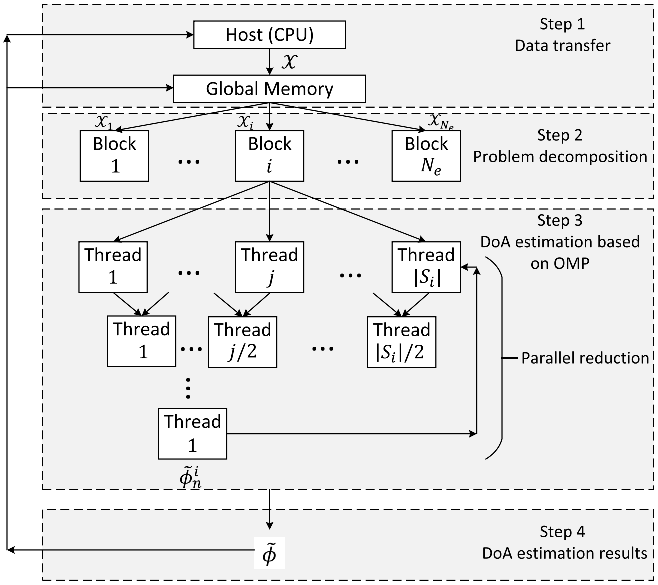


Fig. 5. Parallel implementation.

V. GPU IMPLEMENTATION

In this section, we present an implementation for MARS. To accelerate computation time, we exploit parallelism and implement our algorithm on NVIDIA DGX station—a COTS GPU computing platform. The DGX station has an Intel Xeon E5-2698 v4 2.2-GHz (20-core) CPU and four NVIDIA V100 GPU cards. The data communication between CPU and GPU is based on a PCIe V3.0 architecture [58]. We use only one V100 GPU card for our implementation which has 5120 CUDA cores. We use CUDA v10.2 [59], [60] as our programming platform. The parallel implementation is shown in Fig. 5.

Step 1 (Transferring Data): After 2-D FFT, we transfer measurement $\mathcal{X}$ for the range–velocity bins from CPU memory to the GPU global memory through PCIe V3.0—a relatively high-speed data transfer. This transfer time will be counted toward MARS's total execution time.

Step 2 (Decomposing Problem): The DoA estimation problem (3) is decomposed into N_e independent subproblems (5) that can be solved in parallel. On our GPU platform, N_e blocks are allocated to solve N_e subproblems, respectively. As a result of this process, the complexity is reduced from $\mathcal{O}(N_e)$ to $\mathcal{O}(1)$. Since the on-chip shared memory of each block is much faster than the global memory, the measurement $\mathcal{X}_i$ is transferred from the global memory to the shared memory of ith block (see Fig. 5).

Step 3 (DoA Estimation): Denote $|S_i|$ as the number of elements in the reduced search space S_i. Within the ith subproblem, after obtaining S_i in (11), OMP is used to solve each subproblem. In each iteration of OMP, $|S_i|$ threads are generated to calculate the correlation of each possible signal in the signal basis, reducing the complexity from $\mathcal{O}(|S_i|)$ to $\mathcal{O}(1)$. After that, parallel reduction technique is used to choose the DoA with the highest correlation factor as the estimated DoA, reducing the complexity from $\mathcal{O}(|S_i|)$ to $\mathcal{O}(\log|S_i|)$. OMP will stop once the detected signal power is smaller than a given threshold. Then, the least-squares calculation is conducted directly using the functions provided by cuSOLVER [61].

Step 4 (Finding Solution): By combining the estimation results from each iteration in OMP, we can get the solution of (3). We also save the estimation results in GPU memory for the next estimation and transfer them to the host memory that can be accessed by applications.

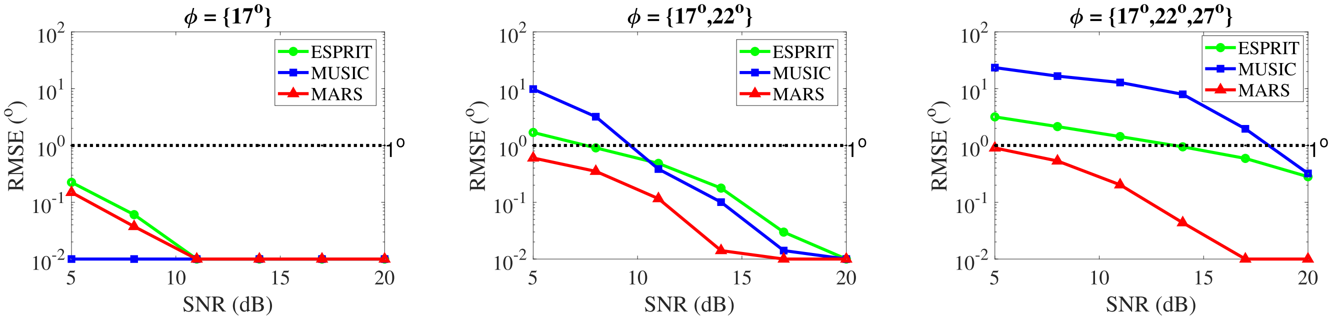
VI. SIMULATION EXPERIMENT

In this section, we evaluate the performance of MARS through simulation experiments. The objective of this section is twofold. First, we will check if MARS can achieve super-resolution and compare its DoA estimation accuracy with the state-of-the-art super-resolution algorithms (MUSIC and ESPRIT). Second, we will compare the computation time of MARS with MUSIC and ESPRIT to see if MARS can achieve 1-ms real-time processing.

A. Parameter Settings

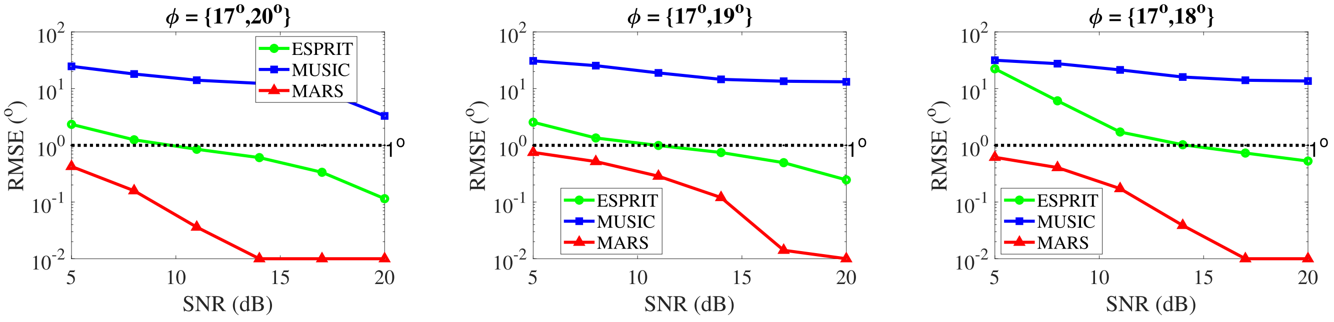
We configure the number of receiving antennas $M = 16$ in a 1-D antenna array with half-wavelength spacing. We assume a medium-range FMCW radar operating from 76 to 81 GHz. The range and velocity resolutions are 30 cm and 3 m/s, respectively [3]. The azimuthal FOV is $[-50^\circ, 50^\circ]$. The time length of a single snapshot is set to be 1 ms. T_s is set to be 10 ms and contains ten snapshots. This means for all three algorithms, the DoA estimation interval is 10 ms. For MARS, we set $\Delta = 2.2$ and $b = 1$. For MUSIC and ESPRIT, ten snapshots are needed to calculate the covariance matrix [62]. All three algorithms are implemented on the same platform [58].

For our evaluation purposes in the simulation experiment, we adopt a 1° resolution as our target and benchmark for all three algorithms. This criterion is selected because it can meet the requirements for most on-road scenarios [2], [8], [9]



(a) (b) (c)

Fig. 6. RMSE versus SNR for different numbers of DoA. (a) One DoA. (b) Two DoAs. (c) Three DoAs.



(a) (b) (c)

Fig. 7. RMSE versus SNR for different spacings between DoAs. (a) 3° spacing. (b) 2° spacing. (c) 1° spacing.

and it can be easily handled in a simulation environment. Based on the 1° resolution goal, the DoA granularity is set to 1°.

We set the total number of bins, N_e, to 100 (corresponding to 100 clusters of targets). We use SUMO [63] to generate a mobility model of all targets for an urban scenario. The velocity of each target follows a uniform distribution from 0 to 120 km/h.

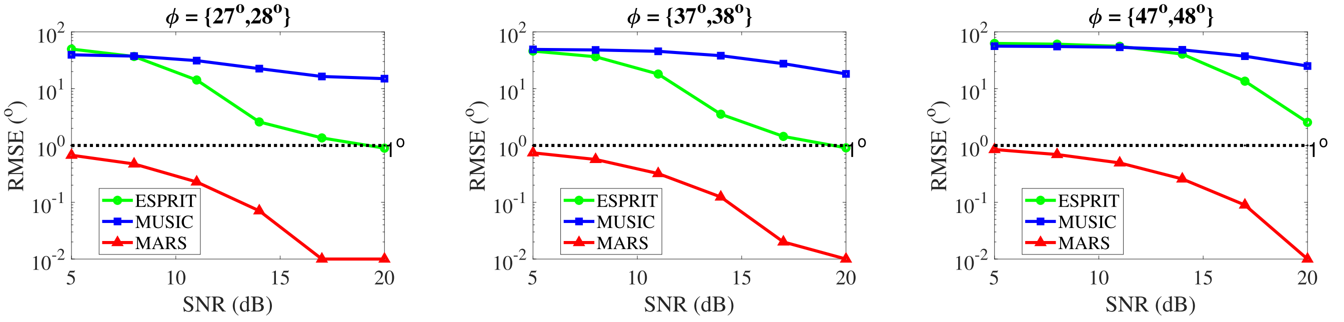
B. Performance of DoA Estimation

We first present the root-mean-square error (RMSE) of estimated DoA as a function of SNR for different DoA settings. Since DoA subproblems corresponding to different bins are equivalent and independent, we can only consider one subproblem within one bin while evaluating the estimation accuracy. For each point, we average over 10 000 simulations to minimize the impact of random noise (especially under low SNR).

In Fig. 6, RMSE versus SNR for different numbers of DOAs is shown. When there are multiple DoAs, the RMSE is averaged among the DoAs. The maximum number of DoAs is set to 3 as it is unlikely to have more than three targets (reflection points) within 30-cm interval in practice. In Fig. 6(a)–(c), RMSEs all decrease as SNR increases for all three algorithms (MARS, ESPRIT, and MUSIC), which is intuitive. From Fig. 6(a) to (c), the RMSE corresponding to each algorithm increases (for the same SNR) as the number of DoAs increases from 1 to 3. This is because when the number of DoAs in a bin increases, each incident signal will interfere with the others, which makes it harder to separate different DoAs. In Fig. 6(a), the performances of all three algorithms meet the 1° requirement when SNR $>$ 5 dB. As the SNR increases, all three curves become flat because the smallest resolution we take for accuracy is 0.01°. In Fig. 6(b), only MARS can meet the 1° requirement for SNR $>$ 5 dB while ESPRIT requires SNR $>$ 8 dB and MUSIC requires SNR $>$ 10 dB. In Fig. 6(c), only MARS can meet the 1° requirement for SNR $>$ 5 dB while ESPRIT requires SNR $>$ 14 dB and MUSIC requires SNR $>$ 18 dB.

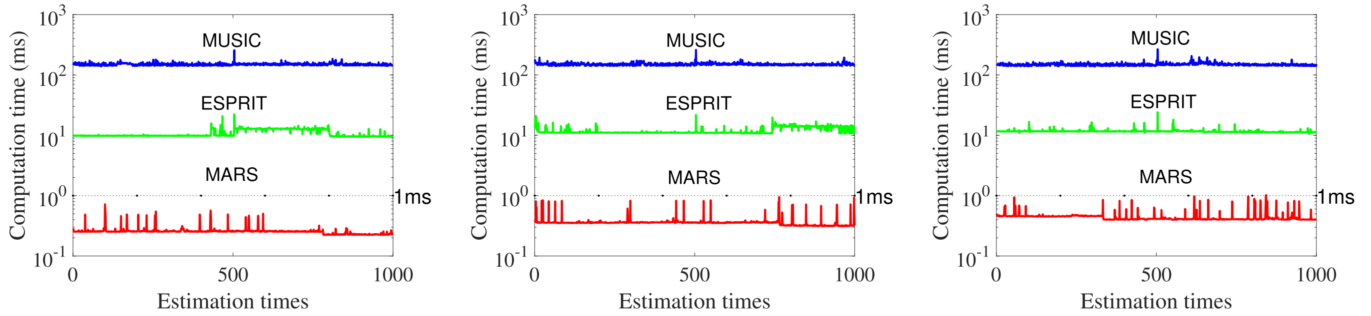
Fig. 7 shows RMSE versus SNR under two DoAs with different spacing between the two DoAs. From Fig. 7(a) to (c), only MARS can achieve 1° super-resolution with all DoA settings for SNR $>$ 5 dB. From Fig. 7(a) to (c), ESPRIT requires SNR $> 10, 11,$ and 14 dB for spacing $3°, 2°,$ and 1°, respectively, to achieve 1° super-resolution. MUSIC cannot achieve 1° super-resolution for any SNR within [5, 20] dB.

Fig. 8 shows RMSE versus SNR with the same DoA spacing but different DoA values. From Fig. 8(a) to (c), RMSE corresponding to every algorithm increases as the DoA value increases. This is because, as DoA increases, the gradient $d \cos\phi$ of the objective function decreases. As a result, the difference between objective function values of different incident signals with the same spacing becomes smaller, which makes it harder to separate different DoAs. From Fig. 8(a) to (c), only MARS can achieve 1° super-resolution with all DoA settings for SNR $>$ 5 dB. In Fig. 8(a) and (b), ESPRIT requires SNR $>$ 20 dB to achieve 1° super-resolution. In Fig. 8(c), ESPRIT fails to achieve 1° super-resolution over the range (SNR $\in [5, 20]$ dB). Note that MUSIC cannot achieve 1° super-resolution over the entire range (SNR $\in [5, 20]$ dB) for all three settings.



(a) (b) (c)

Fig. 8. RMSE versus SNR for the same spacing between DoAs but different values of DoAs.



(a) (b) (c)

Fig. 9. Computation time of MARS versus ESPRIT and MUSIC under different DoA settings. (a) One DoA/bin. (b) Two DoAs/bin. (c) Three DoAs/bin.

In summary, MARS is the only algorithm that can achieve 1° resolution in all experimental settings. It also outperforms both MUSIC and ESPRIT in all settings.

C. Computation Time

We now evaluate the processing time of MARS and compare it to ESPRIT and MUSIC. In Fig. 9(a) to (c), the numbers of DoAs in each bin are 1, 2, and 3. Assume the number of bins $N_e = 100$. Then, Fig. 9(a) to (c) corresponds to 100, 200, and 300 targets (reflection points), respectively. Unlike in Section VI-B where the accuracy of DoA estimation is independent of the number of bins, the computation time that we are going to measure depends on both the total number of bins and the number of DoAs in each bin. We set the simulation time to 10 s, which corresponds to 1000 estimations. The initialization process is not included in the simulation and will be studied separately in Section VI-D.

From Fig. 9(a) to (c), the computation times of all three algorithms increase to some extent as the number of DoAs/bins increases. Only MARS can achieve 1-ms real-time processing while MUSIC needs ∼100-ms computation time and ESPRIT ∼10-ms computation time.

In summary, MARS is the only algorithm that can achieve 1-ms real-time processing in all settings while MUSIC and ESPRIT algorithms require orders of magnitude more computation time.

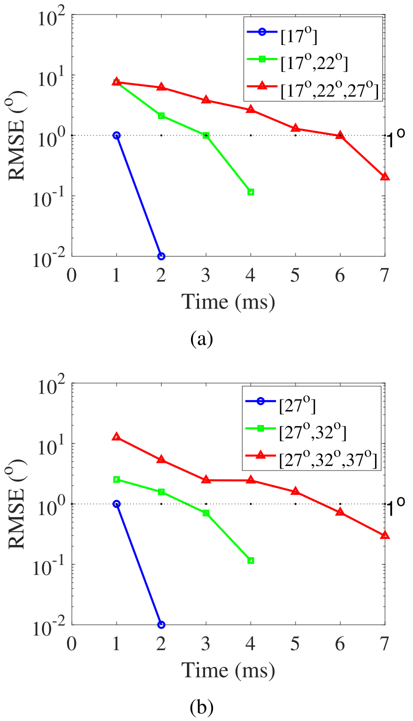


Fig. 10. Evolution of MARS' RMSE performance over time during initialization phase under different DoA settings.

D. Initialization Process

We now investigate MARS' performance during initialization phase. Fig. 10 shows MARS' RMSE performance over time during initialization phase under different DoA settings. The number of samples in the search space is set to $I = (N_T/N_e) = (2560/100) \approx 25$ and the SNR is set to 11 dB. In Fig. 10(a), we have the following observations.

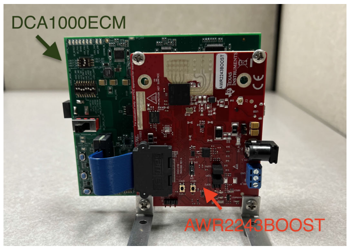


Fig. 11. TI radar kit used in our experiment.

TABLE I
PARAMETER SETTINGS

Parameter	Value
Center frequency	77 GHz
Bandwidth	1.8 GHz
ADC samples	256
Sample rate	10000 ksps
Frequency slope	30 MHz/μs
Ramp end time	60 μs
Idle time	100 μs
Chirp cycle time	160 μs
# chirps/frame (snapshot)	64

Fig. 12. Setup for detecting two targets in our experiment.

1) When there is only one DoA ($\{17^\circ\}$), after two estimations (2 ms), RMSE falls under 1°, which meets our required super-resolution. Therefore, the initialization process only takes 2 ms. From this time on, DoA estimation will follow the procedure in Section IV-D, with its computation time performance already shown in Section VI-C.
2) When the number of DoA increases to two (i.e., $\{17^\circ, 22^\circ\}$), it takes four estimations (4 ms) to bring RMSE down under 1° (our required super-resolution). This initialization process is longer than the previous case (with only one DoA) because there are more DoAs in the search space, and thus, more iterations are needed. After 4 ms, the initialization process is complete and DoA estimation will then follow the procedure in Section IV-D, with its computation time performance shown in Section VI-C.
3) When there are three DoAs (i.e., $\{17^\circ, 22^\circ, 27^\circ\}$), the initialization process requires seven estimations (i.e., 7 ms) to bring RMSE down to our required 1° super-resolution.

Fig. 10(b) shows our experimental results for a different set of DoA values. The results and observations are similar to those in Fig. 10(a). In summary, these results show that our initialization algorithm (Algorithm 1) is efficient and only takes a few ms to reach our required 1° super-resolution after a new target first appears in the radar detection range.

VII. HARDWARE EXPERIMENT

In addition to simulation study, we also conduct experiments using radar hardware in an indoor laboratory environment.

A. Radar Hardware and Experimental Setup

The hardware used in the experiment is depicted in Fig. 11. We use the AWR2243BOOST [3] automotive radar from Texas Instruments (TI), which operates in the 76–81-GHz frequency band and has a 5-GHz bandwidth. This radar is equipped with three transmit antennas and four receive antennas, with a maximum analog-to-digital (ADC) sampling rate of 45 Mb/s. The detailed parameters of the chirp configuration are presented in Table I. To facilitate the capture of raw data from AWR2243BOOST, we employ TI's DCA1000ECM real-time data capture card [64]. This card is designed to support both laboratory and mobile data collection scenarios.

The setup of our laboratory experiment is depicted in Fig. 12. In this setup, the two objects used for target detection are mounted on a slide to precisely control their trajectory and velocity. Both the radar and the objects are positioned at the same horizontal level so that we can focus exclusively on detecting variations in the azimuth angle. The estimation of the elevation angle has been disabled to prevent interference from reflective signals originating from Table I and surrounding objects.

It is important to note that laboratory-based hardware implementation shown in Fig. 12 cannot fully capture real-world conditions. There are mainly two limitations of laboratory-based hardware implementation. First, the velocity of targets in the real world could be much higher. However, as demonstrated in (10), where velocity is considered in forming the reduced search space, and as shown in the simulation results presented in Section VI, the estimation error remains within an acceptable range—typically under 1°. Second, the number of targets that need to be detected in real-world scenarios is larger. However, the performance of DoA estimation and timing will not change much (compared to the performance that we observed from our laboratory-based results). This is because our algorithm is designed to ensure that different clusters of targets (in separate range–velocity bins) are processed independently. When processed on a GPU parallel platform where all bins are handled simultaneously and in parallel, the DoA estimation performance and computation time are nearly identical to that in a laboratory-based setting (despite with fewer number of targets). In conclusion, although the laboratory-based hardware implementation cannot fully replicate real-world on-road scenarios, the performance differences are minimal, and the results can still be applied to address real-world scenarios.

In the experiment, we study three DoA estimation algorithms: 3D-FFT (which comes with the AWR2243BOOST card), MUSIC, and MARS. The latter two algorithms are executed on the raw data gathered by the DCA1000EVM on

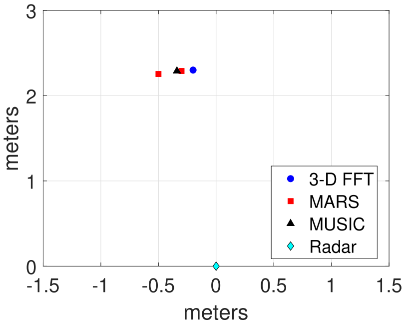


Fig. 13. Detection of object locations under three algorithms.

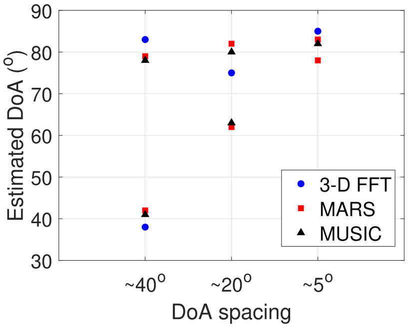


Fig. 14. Comparison of estimated DoAs at various positions under three algorithms.

the same computing platform used in Section VI (i.e., NVIDIA DGX station with one V100 GPU).

B. Experimental Results

With the experimental setup, we assess performance in two scenarios. First, we will examine the accuracy of DoA estimation for two fixed-location objects across three algorithms. Then, we will assess DoA estimation of the three algorithms when objects are placed at different positions with varying spacings.

The experimental results to detect the two fixed-location objects by the three algorithms are shown in Fig. 13, with radar sensor at the origin. Fig. 13 maps the locations of the two objects (~5° spacing with respect to the origin) onto a 2-D azimuth plane, with each position precisely defined by their range and DoAs. It should be noted that, in contrast to simulation experiments where the DoAs of reflected signals can be controlled precisely by the simulator, the actual DoAs in a hardware-based experiment cannot be controlled precisely. This is because the exact locations of the reflection points on the objects are unknown, and thus, the reflected signals can come from anywhere on the actual objects (e.g., the cardboards in our experiment).

As depicted in Fig. 13, employing 3-D FFT results in the detection of only one object (indicated by a blue marker). Likewise, the use of MUSIC identifies a single object (represented by a black marker). In contrast, under MARS, both of the two objects are successfully detected (highlighted by the two red markers). This experiment shows that MARS can offer a higher resolution in object detection than both 3-D FFT and MUSIC algorithms. The computation time of MUSIC is >100 ms and the computation time of MARS and 3D-FFT is <1 ms.

Fig. 14 shows the DoA estimation results under the three algorithms for the two objects with three different DoA spacings (~40°, ~20°, and ~5°).[3]

1) When DoA spacing of the two objects is approximately 40° apart, the two DoAs detected by the three algorithms are: 1) 3D-FFT: 83° and 38°; 2) MUSIC: 78° and 41°; and 3) MARS: 79° and 42°. Due to unknown reflection points on the objects (card boards), we cannot assess the accuracy of DoA estimations. Hence, our results can only be used to assess each algorithm's ability to detect both objects simultaneously, which is the case for this DoA spacing.
2) When DoA spacing of the two objects is approximately 20° apart: 1) 3D-FFT only detects a single DoA at 75°; 2) MUSIC detects two DoAs at 80° and 63°; and 3) MARS detects two DoAs at 82° and 62°. In other words, a DoA spacing of approximately 20° is already beyond the resolution of 3D-FFT in this experiment.
3) When DoA spacing of the two objects is approximately 5° apart: 1) 3D-FFT only detects a single DoA at 85°; 2) MUSIC only detects a single DoA at 82°; and 3) MARS can still detect two DoAs at 83° and 78°. In other words, MARS exhibits the highest resolution among the three algorithms.

The computation time of MUSIC is still >100 ms and the computation time of MARS and 3D-FFT is still <1 ms.

In summary, our hardware experimental results demonstrate that MARS can achieve higher DoA estimation resolution than the state-of-the-art DoA estimation algorithms and can achieve 1-ms real-time processing.

VIII. CONCLUSION

In this article, we presented MARS—a novel real-time super-resolution DoA estimation algorithm for automotive radar. The main idea of MARS is to employ maximum likelihood as our objective function as it is intrinsically designed to find an optimal solution. To address the prohibitively high complexity associated with traditional MLE, MARS exploited the correlation between the input data from adjacent time slots. Specifically, MARS uses the last estimation results to form a promising and reduced search space to solve the maximum likelihood problem. To further accelerate computation time, we proposed to exploit parallel computing whenever possible. To do this, MARS decomposes the problems at each step into independent subproblems and employs GPU to perform parallel computation. Special efforts were given to GPU engineering (in terms of block/thread management) that best matches the underlying hardware platform. Through extensive simulation experiments, we found that MARS can achieve 1° super-resolution within 1 ms and outperforms the state-of-the-art algorithms (MUSIC and ESPRIT) in both DoA estimation and timing performance. To the best of authors' knowledge, MARS remains the only known algorithm that can achieve super-resolution within

---

[3] Unlike in a simulation environment, a DoA spacing less than 5° is not feasible in our laboratory experiment (Fig. 12) because the reflection points on the two objects (card boards) cannot be controlled precisely.

1 ms. Further hardware experiment also demonstrates that MARS can achieve higher resolution than the other state-of-the-art algorithms.

References

[1] Y. Wu, C. Li, Y. T. Hou, and W. Lou, "Real-time DoA estimation for automotive radar," in *Proc. 18th Eur. Radar Conf. (EuRAD)*, London, U.K., Apr. 2022, pp. 437–440.

[2] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 22–35, Mar. 2017.

[3] *AWR2243: 76-GHz to 81-GHz Automotive Second-Generation High-Performance MMIC*. Accessed: Jan. 28, 2024. [Online]. Available: https://www.ti.com/product/AWR2243

[4] M. Richards, J. Scheer, and W. Holm, *Principles of Modern Radar: Basic Principles*. Raleigh, NC, USA: SciTech, 2010.

[5] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath Jr., "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016.

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

[7] S.-C. Lin et al., "The architectural implications of autonomous driving: Constraints and acceleration," in *Proc. 23rd Int. Conf. Architectural Support Program. Lang. Operating Syst.*, Williamsburg, VA, USA, Mar. 2018, pp. 751–766.

[8] F. Fölster and H. Rohling, "Signal processing structure for automotive radar," *Frequenz*, vol. 60, nos. 1–2, p. 1, Jan. 2006.

[9] F. Roos, D. Kellner, J. Dickmann, and C. Waldschmidt, "Reliable orientation estimation of vehicles in high-resolution radar images," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 9, pp. 2986–2993, Sep. 2016.

[10] Y. Xu, W. Feng Feng, J. Yan Hao, and H. K. Hwang, "Direction of arrival estimation using super-resolution algorithm," in *Proc. MTS/IEEE Oceans. Ocean Odyssey. Conf.*, vol. 2, Honolulu, HI, USA, Aug. 2001, pp. 749–755.

[11] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.

[12] M. P. Clark and L. L. Scharf, "Two-dimensional modal analysis based on maximum likelihood," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1443–1452, Jun. 1994.

[13] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[14] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[15] R. Roy, A. Paulraj, and T. Kailath, "Comparative performance of ESPRIT and MUSIC for direction-of-arrival estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 12, Dallas, TX, USA, Jun. 1987, pp. 2344–2347.

[16] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[17] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 829–846, Apr. 2018.

[18] P. Wenig, M. Schoor, O. Gunther, B. Yang, and R. Weigel, "System design of a 77 GHz automotive radar sensor with superresolution DOA estimation," in *Proc. Int. Symp. Signals, Syst. Electron.*, Montreal, QC, USA, Jul. 2007, pp. 537–540.

[19] M. Rubsamen and A. B. Gershman, "Direction-of-arrival estimation for nonuniform sensor arrays: From manifold separation to Fourier domain MUSIC methods," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 588–599, Feb. 2009.

[20] J. Liang and D. Liu, "Passive localization of mixed near-field and far-field sources using two-stage MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 108–120, Jan. 2010.

[21] J. He, M. N. S. Swamy, and M. O. Ahmad, "Efficient application of MUSIC algorithm under the coexistence of far-field and near-field sources," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2066–2070, Apr. 2012.

[22] W.-J. Zeng, H. C. So, and L. Huang, "ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, Sep. 2013.

[23] Z. Zou, W. Hongyuan, and Y. Guowen, "An improved MUSIC algorithm implemented with high-speed parallel optimization for FPGA," in *Proc. 7th Int. Symp. Antennas, Propag. EM Theory*, vol. 5985, Guilin, China, Oct. 2006, pp. 1–4.

[24] M. W. Majid, "Parallel implementation of the wideband DOA algorithm on single core, multicore, GPU and IBM cell be processor," *Sci. J. Circuits, Syst. Signal Process.*, vol. 2, no. 2, p. 29, 2013.

[25] J. Yan, Y. Huang, H. Xu, and G. A. E. Vandenbosch, "Hardware acceleration of MUSIC based DoA estimator in MUBTS," in *Proc. 8th Eur. Conf. Antennas Propag. (EuCAP)*, Hague, The Netherlands, Apr. 2014, pp. 2561–2565.

[26] N. Tayem, M. Omer, and A. A. Hussain, "Hardware implementation of MUSIC and ESPRIT on NI-PXI platform," in *Proc. IEEE Mil. Commun. Conf.*, Baltimore, MD, USA, Oct. 2014, pp. 329–332.

[27] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1161–1163, Dec. 2010.

[28] F. Shu et al., "Low-complexity and high-resolution DOA estimation for hybrid analog and digital massive MIMO receive array," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2487–2501, Jun. 2018.

[29] S. Chen, B. Sima, F. Xi, W. Wu, and Z. Liu, "Super-resolution DOA estimation using dynamic metasurface antenna," in *Proc. 14th Eur. Conf. Antennas Propag. (EuCAP)*, Copenhagen, Denmark, Mar. 2020, pp. 1–4.

[30] R. Cao, B. Liu, F. Gao, and X. Zhang, "A low-complex one-snapshot DOA estimation algorithm with massive ULA," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1071–1074, May 2017.

[31] F.-M. Han and X.-D. Zhang, "An ESPRIT-like algorithm for coherent DOA estimation," *IEEE Antennas Wireless Propag. Lett.*, vol. 4, pp. 443–446, 2005.

[32] F. Gao and A. B. Gershman, "A generalized ESPRIT approach to direction-of-arrival estimation," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 254–257, Mar. 2005.

[33] J. Hui and Y. Gang, "An improved algorithm of ESPRIT for signal DOA estimation," in *Proc. Int. Conf. Ind. Control Electron. Eng.*, Aug. 2012, pp. 317–320.

[34] J. Lin, X. Ma, S. Yan, and C. Hao, "Time-frequency multi-invariance ESPRIT for DOA estimation," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 770–773, 2016.

[35] W.-H. Fang and L.-D. Fang, "Joint angle and range estimation with signal clustering in FMCW radar," *IEEE Sensors J.*, vol. 20, no. 4, pp. 1882–1892, Feb. 2020.

[36] G. Zheng, J. Tang, and X. Yang, "ESPRIT and unitary ESPRIT algorithms for coexistence of circular and noncircular signals in bistatic MIMO radar," *IEEE Access*, vol. 4, pp. 7232–7240, 2016.

[37] J. Li, X. Zhang, and D. Jiang, "DOD and DOA estimation for bistatic coprime MIMO radar based on combined ESPRIT," in *Proc. CIE Int. Conf. Radar (RADAR)*, Guangzhou, China, Oct. 2016, pp. 1–4.

[38] X. Yang, G. Zheng, and J. Tang, "ESPRIT algorithm for coexistence of circular and noncircular signals in bistatic MIMO radar," in *Proc. IEEE Radar Conf. (RadarConf)*, Guangzhou, China, May 2016, pp. 1–4.

[39] M. Jamali, J. Downey, N. Wilikins, C. R. Rehm, and J. Tipping, "Development of a FPGA-based high speed FFT processor for wideband direction of arrival applications," in *Proc. IEEE Radar Conf.*, vol. 48, Bordeaux, France, Dec. 2009, pp. 1–4.

[40] Z. Ke, M. Peng, and Z. Jian-Yun, "DOA estimation algorithm based on FFT in switch antenna array," in *Proc. IEEE CIE Int. Conf. Radar*, vol. 2, Chengdu, China, Oct. 2011, pp. 1425–1428.

[41] B. Kim, S. Kim, and J. Lee, "A novel DFT-based DOA estimation by a virtual array extension using simple multiplications for FMCW radar," *Sensors*, vol. 18, no. 5, p. 1560, May 2018.

[42] Z. Zhang, C. Zhou, Y. Gu, and Z. Shi, "FFT-based DOA estimation for coprime MIMO radar: A hardware-friendly approach," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.

[43] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 3967–3971.

[44] M. Guo, Y. D. Zhang, and T. Chen, "DOA estimation using compressed sparse array," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4133–4146, Aug. 2018.

[45] S. Sun and Y. D. Zhang, "4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 4, pp. 879–891, Jun. 2021.

[46] R. Zheng, S. Sun, H. Liu, and T. Wu, "Deep neural networks-enabled vehicle detection using high-resolution automotive radar imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 4815–4830, Oct. 2023.

[47] F. Engels, P. Heidenreich, A. M. Zoubir, F. K. Jondral, and M. Wintermantel, "Advances in automotive radar: A framework on computationally efficient high-resolution frequency estimation," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 36–46, Mar. 2017.

[48] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5066–5081, Oct. 2016.

[49] M. Xu, J. Zhu, J. Fang, N. Zhang, and Z. Xu, "CFAR based NOMP for line spectral estimation and detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 6971–6990, Jun. 2023.

[50] L. Han, X. Liu, N. Zhang, S. Wu, J. Zhu, and Z. Xu, "Two-dimensional multi-snapshot newtonized orthogonal matching pursuit for DOA estimation," *Digit. Signal Process.*, vol. 121, Mar. 2022, Art. no. 103313.

[51] J. Zhu, Q. Zhang, P. Gerstoft, M.-A. Badiu, and Z. Xu, "Grid-less variational Bayesian line spectral estimation with multiple measurement vectors," *Signal Process.*, vol. 161, pp. 155–164, Aug. 2019.

[52] J. Zhu, L. Han, R. S. Blum, and Z. Xu, "Multi-snapshot newtonized orthogonal matching pursuit for line spectrum estimation with multiple measurement vectors," *Signal Process.*, vol. 165, pp. 175–185, Dec. 2019.

[53] C. Li et al., "A review on recent progress of portable short-range noncontact microwave radar systems," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1692–1706, May 2017.

[54] F. Meinl, E. Schubert, M. Kunert, and H. Blume, "Realtime FPGA-based processing unit for a high-resolution automotive MIMO radar platform," in *Proc. Eur. Radar Conf. (EuRAD)*, Paris, France, Sep. 2015, pp. 213–216.

[55] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 845–860, Mar. 2012.

[56] J. Lee, Y.-A. Li, M.-H. Hung, and S.-J. Huang, "A fully-integrated 77-GHz FMCW radar transceiver in 65-nm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 45, no. 12, pp. 2746–2756, Dec. 2010.

[57] Nvidia. *Optimizing Parallel Reduction in CUDA*. Accessed: Jan. 28, 2024. [Online]. Available: https://developer.download.nvidia.com/assets/cuda/files/reduction.pdf

[58] Nvidia. *NVIDIA TESLA V100 GPU Architecture*. Accessed: Jan. 28, 2024. [Online]. Available: https://images.nvidia.com/content/volta-architecture/pdf/volta-architecture-whitepaper.pdf

[59] Nvidia. *CUDA Toolkit Documentation: CuRAND*. Accessed: Jan. 28, 2024. [Online]. Available: https://docs.nvidia.com/cuda/curand/index.html

[60] Nvidia. *CUDA C Programming Guide V10.2.89*. Accessed: Jan. 28, 2024. [Online]. Available: https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html

[61] Nvidia. (2024). *CuSOLVER API Reference*. Accessed: May, 6, 2024. [Online]. Available: https://docs.nvidia.com/cuda/cusolver/index.html

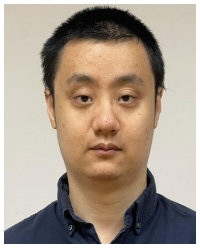
[62] W. Liao, "MUSIC for multidimensional spectral estimation: Stability and super-resolution," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6395–6406, Dec. 2015.

[63] Simulation of Urban MObility. *SUMO User Documentation*. Accessed: Jan. 28, 2024. [Online]. Available: https://sumo.dlr.de/docs/

[64] *DCA1000EVM: DCA1000 Evaluation Module for Real-time Data Capture and Streaming*. Accessed: Jan. 28, 2024. [Online]. Available: https://www.ti.com/tool/DCA1000EVM

Yubo Wu (Graduate Student Member, IEEE) received the B.E. and M.S. degrees in telecommunication engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2016 and 2019, respectively. He is pursuing the Ph.D. degree with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA.

His current research interests are automotive radar signal processing for autonomous vehicles, wireless networking, open radio access network (O-RAN) scheduling, and novel GPU-based design and implementation to accelerate computation for real-time applications.



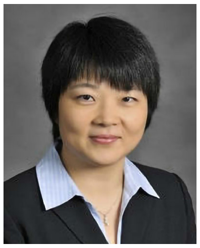
Chengzhang Li (Member, IEEE) received the B.S. degree in electronics engineering from Tsinghua University, Beijing, China, in 2017, and the M.S. and Ph.D. degrees in computer engineering from Virginia Tech, Blacksburg, VA, USA, in 2020 and 2022, respectively.

He is a Postdoctoral Researcher with the AI-EDGE Institute, The Ohio State University, Columbus, OH, USA. His current research interests are machine learning in edge networks, real-time scheduling in 5G, and age of information (AoI).

Y. Thomas Hou (Fellow, IEEE) received the Ph.D. degree from the NYU Tandon School of Engineering, Brooklyn, NY, USA, in 1998.

He is Bradley Distinguished Professor of Electrical and Computer Engineering at Virginia Tech, Blacksburg, VA, USA. He has authored or co-authored over 350 articles in IEEE/ACM JOURNALS and conferences. He has authored or co-authored two graduate textbooks: *Applied Optimization Methods for Wireless Networks* (Cambridge University Press, 2014) and *Cognitive Radio Communications and Networks: Principles and Practices* (Academic Press/Elsevier, 2009). He holds six U.S. patents. His current research focuses on developing real-time optimal solutions to complex science and engineering problems arising from wireless and mobile networks. He is also interested in wireless security.

Prof. Hou was on the editorial boards of a number of IEEE and ACM TRANSACTIONS and JOURNALS. He served as the Steering Committee Chair for IEEE INFOCOM Conference and was a member of the IEEE Communications Society Board of Governors. He was also a Distinguished Lecturer of the IEEE Communications Society. His articles were recognized by ten best paper awards from IEEE and ACM, including IEEE INFOCOM Test of Time Paper Award in 2023.



Wenjing Lou (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2003.

She is the W. C. English Endowed Professor of Computer Science with Virginia Tech, Arlington, VA, USA. Her research interests cover many topics in the cybersecurity field, with her current research interests focusing on wireless networks, privacy protection in machine learning systems, and security and privacy problems in the Internet of Things (IoT) systems.

Prof. Lou received the Virginia Tech Alumni Award for Research Excellence, the highest university-level faculty research award, in 2018, and the IEEE INFOCOM Test of Time Paper Award in 2020. She was a TPC Chair of IEEE INFOCOM 2019 and ACM WiSec 2020. She served as the Steering Committee Chair for IEEE CNS Conference from 2013 to 2020. She is currently a Steering Committee Member of IEEE INFOCOM. She served as the Program Director for the U.S. National Science Foundation (NSF) from 2014 to 2017. She is a Fellow of ACM. She is a highly cited researcher by the Web of Science Group.