Reinforcement Learning Based Power-Optimal Usage of Beamforming Antenna Array for Multi-Way Wireless Communication in Vehicular Traffic Environments

Suhasini Komarraju and Abhijit Chatterjee
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA
Email: skomarraju3@gatech.edu, abhijit.chatterjee@ece.gatech.edu

***Abstract*— There has been recent work on the design of antenna arrays for beamforming in dynamic evolving environments such as in vehicle-to-vehicle communication systems. A key problem is that of determining how to optimally use a large antenna array to communicate with multiple spatially located vehicles in dynamically changing channel conditions with minimal co-channel interference while minimizing overall power consumption of the wireless system. We envision disjoint subsets of antennas in the array being used to direct beams concurrently to different vehicles. The number of antennas, gain and phase of each RF-chain driving an antenna are optimized dynamically using a constrained quadratic cost formulation encompassing channel quality, interference and power consumption. This quadratic optimization problem is solved using behavior constrained bandit algorithm, a reinforcement learning based technique. A gaussian kernel is used to perform data clustering of vehicle environment and resulting solutions, allowing quick bootstrapping of the bandit solver to find optimal array configurations in real-time vehicle environments. Simulation studies prove the viability of the proposed scheme.**

***Keywords*— MIMO beamforming, antenna array**

I. Introduction

Vehicular communication technology has gained importance in recent years due to increased focus on traffic safety through use of intelligent traffic systems (ITS) [1]. The Dedicated Short Range Communication (DSRC) protocol is based on 5.9 GHz wireless systems with a transmission range of 1000m. However, noisy dynamically changing channel conditions in moving multi-vehicle environments necessitate the use of antenna systems that can be used to focus transmit energy in a particular direction [2] while minimizing the transmitted energy in other directions. Besides, future vehicular systems require high throughput, ultra-high reliability and ultra-low latency. Multiple Input Multiple Output (MIMO) antenna arrays operating in mm-wave spectrum can be used to satisfy these stringent requirements under dynamic channel conditions. But these systems pose several challenges such as high power consumption of circuitry involved, high path loss and low signal penetration [3].

Besides, unlike current vehicular communications that operate based on transmissions which are omnidirectional, mm-wave vehicular systems need to operate based on directional transmissions to compensate for the high path loss [4]. MIMO antenna arrays used for vehicular communication can be divided into fixed sub-arrays to cater to vehicles in multiple directions located around it or an entire array can be used to communicate in a given direction. Use of entire array or array divided into fixed sub-arrays poses challenges such as high power consumption (as all elements in a given fixed sub-array are used to communicate with a single vehicle irrespective of surrounding wireless environment) and limit on number of simultaneous vehicles an entire array can communicate with. Further, current research is focused on developing mm-wave MIMO antenna arrays operating on one mm-wave channel for all Vehicle-to-Vehicle communications [5]. As a result of this, the transmissions of pilot vehicle with vehicles other than the intended vehicle, cause interference at the intended vehicle. The key goal of this research is to develop learning-assisted algorithms for optimal usage of beamforming antenna arrays that operate at 60GHz, to enable a pilot vehicle to *communicate with multiple vehicles in its vicinity concurrently*, while minimizing power consumption and signal interference under dynamically changing channel conditions. We assume that a multiplexing architecture exists on baseband processing side that enables switching and processing of data streams across different radio frequency (RF) chains (each connected to an antenna element).

With regard to prior work, Use of Alamouti coding and MIMO techniques in V2V communications is presented in [6]. There has been significant prior work on energy efficiency of MIMO systems [7].In [2], a real-time antenna weight optimization framework for side-lobe, same frequency, interference cancellation using the signal in the main communication lobe of a linear antenna array as reference (pilot) is developed. In [8], a neural network is trained to produce antenna weight factors for a moving beam (number of antennas is fixed) and a genetic algorithm based beam direction control is presented in [9]. In contrast to the proposed research, the above algorithms focus on fixed antenna array configurations and limited numbers of users and are not suited to the automation of complex dynamic vehicle communication environments that are envisioned for the future.

II. Preliminaries and Approach

Consider the on-road scenario given in Figure 1. The pilot car (P) communicates with three other vehicles V_1, V_2 and V_3. In general, there could be more communicating vehicles. The communication channel between P and vehicle V_i is characterized by a 3-tuple [θ_i, d_i and SNR_i], where θ_i is the angular position of V_i relative to P, d_i is the distance and

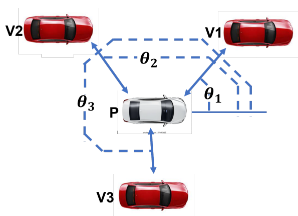


Fig. 1: Vehicle communication scenario

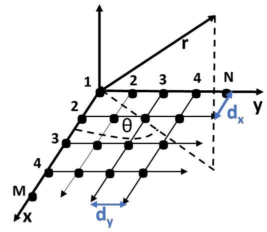


Fig. 2: Rectangular array

SNR_i is the SNR of the communication channel between P and V_i (only θ_i is shown in Figure 1 for brevity). We assume that P is equipped with an antenna array [10] to allow it to simultaneously communicate with V_1, V_2 and V_3 and that each antenna is associated with a unique RF chain where the gain and phase of each chain can be adjusted independently. The problem to be solved is defined as follows:

Optimal antenna array usage problem definition: Given N neighboring vehicles V_1, V_2, ..., V_N and the quality of the channel between P and V_i, find the number of antennas and phase and gain of corresponding antenna RF chains such that effective communication between P and V_i is established, power consumption is minimized and beam side lobes towards all other vehicles V_j, $j \neq i$, are minimized to specified levels to mitigate interference, *for all* $1 \leq i \leq N$. It is assumed that the power consumption metric is directly proportional and scales linearly with number of antenna elements chosen for communication with V_i

There are three aspects to the proposed approach in solving the above problem. At every time step of the evolving vehicle environment, we assume that the tuples X_i = [θ_i, d_i, SNR_i], $1 \leq i \leq N$ are regenerated and provided to our algorithm. The vector X = [X_1, X_2, ..., X_N] of dimension $3N$ is input to the procedure below.

Step 1. Determining numbers of antennas per V_i: First, the number of antenna elements needed for P to communicate with each vehicle is determined. Higher the number of antenna elements, better is the antenna array pattern and less is the interference caused at other vehicular locations [11]. But higher the number of antenna elements, higher is the power consumption. Hence, determining the number of antennas per V_i is a result of trade-off between power consumption of antenna array at P and interference caused at all other vehicles V_j, $j \neq i$. This relies on a Gaussian kernel based clustering algorithm that clusters the vector X and tags it with the corresponding solution to the antenna usage problem defined earlier. Incoming data X at the current time is mapped to a cluster and the number of antennas n_i per vehicle V_i, $1 \leq i \leq N$, corresponding to that cluster head is assigned for communication between the pilot car P and vehicle V_i. Accordingly, n_i antenna elements in the array are selected to enable the above communication.

Step 2. Finding antenna chain gains and phases: The solution to the vehicle environment given by the data cluster generated by the Gaussian kernel to which the given tuple X is mapped, is used as an initial condition (bootstrap) for solving quadratic optimization that produces the gains and phases of each antenna RF chain that solves the antenna usage problem for communication between vehicles, P and V_i. For N communicating vehicles, N such problems are solved concurrently. The quadratic optimization problem is solved using behavioral constrained bandit (BCB) algorithm [12] which is a reinforcement learning (RL) method. In our solution, we model the amplitudes and phases of antenna array elements along with number of elements per vehicle as actions. The learning environment is observed by observing the antenna array patterns generated by choosing any action. In this solution, there are two steps namely, the constraint learning step and the online recommendation step. In the first step, the constraints of quadratic optimization problem are learned by randomly choosing the actions. In the second step, the algorithm tries to minimize the power consumption while satisfying the constraints learnt in previous step. This is further explained in detail in Section IV.

Step 3. Updating clusters: With every incoming data X, either a new cluster is formed based on a given similarity metric or it is mapped to an existing data cluster. If X is mapped to an existing cluster, the aggregated solution for that cluster represented by the solution for the cluster head is updated. The process is repeated from Step 1 for successive time steps. Note that the objective of clustering is to *enable optimal beamforming solutions in near real-time* (fractions of a second) in dynamically evolving vehicle environments by providing a known initial solution for unknown vehicular environment.

III. Antenna Element Assignment

At each time step, Gaussian kernel is used to map the vector X to a cluster from prior iterations or to a new cluster. Each cluster is associated with configurations that require a given number of antennas (n_i) required for communication with V_i, $1 \leq i \leq N$. Prior to formation of clusters, initial number (n_i) of antenna elements required for P to communicate with each V_i are determined by solving the quadratic optimization problem using the BCB algorithm which is discussed in the next section.

At a given time step, the unassigned elements in the entire array are assigned in increasing order from left to right and from top to bottom in a greedy manner, with a wraparound at the end of each row. It is to be noted that this assignment results in a pseudo-random element assignment in future time steps due to the fact that the number of antenna elements n_i assigned for P-to-V_i communication depends on SNR_i. Since SNR_i changes across time steps, additional elements are assigned or de-assigned as required, allowing un-assigned elements to be allocated to other surrounding vehicles.

IV. Power-Optimal Beamforming Antenna Usage for Multi-Way Communication

Along with orientation of antenna elements, their input amplitudes and phases determine the overall beam radiation pattern. The field of the entire array is given by multiplying the field of a single element placed at origin with its *array factor* which depends on all the above said parameters [11]. For a rectangular array placed at the origin as shown in Figure 2, the array factor (AF) in the direction of θ (AF(θ)) is given by Equation 1, where the amplitude and phase values of corresponding RF chains attached to antenna elements located at (x,y) position are represented by A_{xy} and β_{xy} respectively. A_{xy} is taken to be zero when element at (x,y) is not being used. The values of A_{xy}, β_{xy} along with n_i are given as actions to the BCB algorithm. Each action is represented by a variable 'f' which takes natural numbers as values.

$$AF(\theta) = \sum_{x=1}^{x=M} \sum_{y=1}^{y=N} A_{xy} e^{j((x-1)kd_x cos\theta + (y-1)kd_y sin\theta + \beta_{xy})} \tag{1}$$

The received power (R_i) at vehicle V_i to detect a transmitted symbol depends on receiver sensitivity and channel SNR [11]. Friis free space loss equation [13] gives the transmit power (S_{Pi}) at vehicle (P) required to receive power (R_i) at vehicle V_i [13]. With increasing n_i, S_{Pi} increases but the power consumption of entire array increases linearly with n_i. Also, while communicating with vehicle V_i in the direction of θ_i (see Figure 1), beam side lobe power in the directions θ_j, $j \neq i$ must be minimized for interference mitigation. To solve the same, an optimization problem is defined as follows:

Minimize n_i subject to $|AF_{n_i}(\theta_i)|^2 \times K \geq S_{Pi}$, $|AF_{n_i}(\theta_j)|^2 \times K \leq S_{Pj} - 10dB \ \forall j \neq i$ In this, $AF_{n_i}(\theta_i), AF_{n_i}(\theta_j)$ is array factor of n_i antenna elements in θ_i, θ_j directions respectively. K is a multiplicative constant for transmitted power, n_i is number of elements for transmission between P and V_i, S_{Pj} is power transmitted by P in V_j direction.

The first constraint ensures that enough power is being transmitted in the direction of vehicle V_i while the rest of them ensure that the side lobe power in the direction of other vehicles ($V_j, j \neq i$) to be less than the main lobe power in those respective directions by at least 10dB, thus minimizing interference to a safe threshold value.

Given a vehicular environment X, in the BCB algorithm's first step, the agent randomly chooses actions and a *teaching agent* determines whether chosen actions are acceptable by computing the constraints. For an action f chosen at time t, if all the constraints are satisfied, reward metric $r_f^e(t)$ is 1 and otherwise, 0. Based on the reward, the agent updates its policy (the probability of choosing an action) during this phase as μ_f^e. During the second phase of the algorithm, the agent again randomly chooses an action (f) and gets a reward $r_f(t)$ which is given by $r_f(t) = -n_i$ in f. This reward is the negative of number of elements determined by action f to communicate with vehicle V_i and acts as a measure proportional to power consumption [14]. Based on this reward, the agent updates its policy of choosing action f during this phase as δ_f. Then a weighted combination of μ_f^e and δ_f is calculated for each action using a parameter κ as $(\kappa\mu_f^e)+(1-\kappa)\delta_f$. The combined reward is also calculated in similar way. Based on this value for all actions, given a vehicular configuration, the action that results in maximum weighted combination of policies is chosen and the actual reward is observed. The value of κ parameter is varied as shown in Section VI. The entire algorithm follows [12] and is skipped here.

V. Data clustering and information retrieval

The number of constraints in the optimization problem increases with increase in number(N) of surrounding vehicles as given by function $f(N) = N(N-1)$. With increase in number of surrounding vehicles, number of optimization variables increases leading to a very large dimensional search space. Hence, clustering based on Gaussian kernels [15] between vehicle configurations is performed to speed up reaching the optimal solution for quadratic optimization problem discussed above.

For clustering, 5000 vehicle configurations are generated such that, in each configuration, number of vehicles varies between 1 and 7 and SNR varies between 1 and 20dB. The number of clusters (C) is chosen to be 64 as the array considered at vehicle P is 8x8 and number of antenna elements required for communication between P and V_i can vary between 1 and 64. Given all vehicle configurations, the cluster heads are randomly assigned initially. A non-linear Gaussian kernel is applied to calculate the similarity metric between a given vehicular configuration and every cluster head. All the vehicular configurations (vectors denoted by X) are transformed into kernel space and the distance between a given data point and cluster head in transformed space is calculated as similarity metric. This similarity measure criterion is given in Equation 2, in which m_c represents cluster head of cluster c and $\psi(X)$ represents X in Kernel space. The cluster head is updated using Equation 3, every time a new point is added to the cluster. The Gaussian kernel used is shown in Equation 4 in which X and X' represent two different configurations.

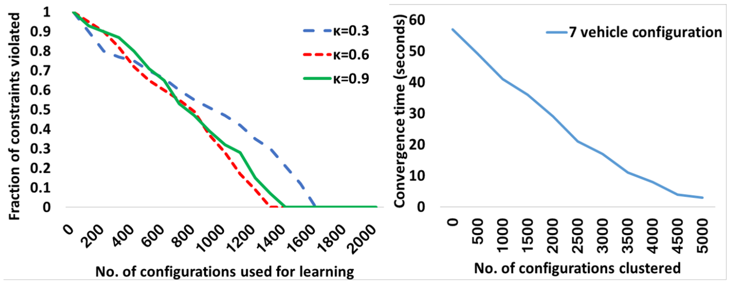
$$Similarity(C) = \sum_{c=1}^{c=C} \sum_{X \in c} \|\psi(X) - m_c\|^2 \tag{2}$$

$$m_c = \frac{\sum_{X \in c} X}{\text{No. of points in c}} \tag{3}$$

$$\psi(X)\dot{\psi}(X') = exp(\frac{-\|X - X'\|^2}{2\sigma^2}) \tag{4}$$

VI. Results

The optimization problem defined in Section IV is solved using the BCB algorithm and the fraction of number of constraints violated for a given vehicular configuration as a function of number of configurations used for learning is presented for varying values of κ parameter in Figure 3(a). The results of using clustering based initial solution is shown in Figure 3(b). In Figure 3(b), the time taken for convergence of the optimization of a vehicle configuration with N=7, as a function of number of vehicles clustered is presented. It can be seen that as the count of vehicles clustered increases, the initial solution for new configuration is improved, thereby improving the optimization time when the solution of cluster



(a) Behavioral error (Fraction of constraints violated) vs No. of configurations used for learning
(b) Convergence time of 7 vehicle configuration vs No. of configurations clustered

Fig. 3: Optimization time

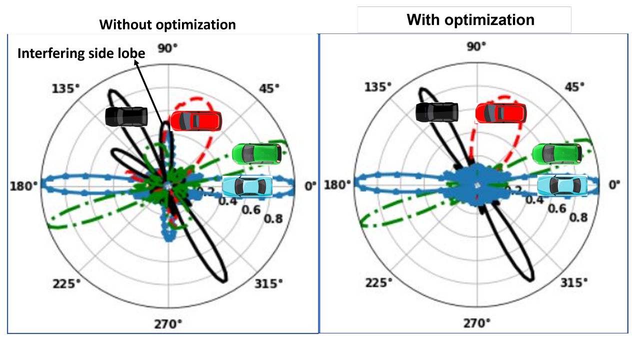


Fig. 4: Normalized array factor for configuration: V_1:(20°, SNR=12dB), V_2:(50°, SNR=18dB), V_3:(120°, SNR=14dB), V_4:(0°, SNR=10dB)

head is used as initial solution for unknown configuration. With continuous learning, more number of vehicular configurations and updating of clusters, the optimization time can be further reduced to fractions of a second.

The optimal and un-optimized normalized array factor patterns for a vehicular configuration with four surrounding vehicles obtained using proposed approach is given in Figure 4. In this, the antenna patterns are color-coded to represent the corresponding color-coded vehicle directions. The solution without optimization indicates the interference due to side lobes caused when 5 elements each are used to communicate with every vehicle. 8x8 antenna array assignment of a 4 vehicle scenario with high SNR values ($\geq$ 10dB) and a 7 vehicle scenario with low SNR values is shown in Figure 5 in which each color represents a different vehicle.

A comparison of our approach with an antenna array (8x8 size) divided into fixed sub-arrays (4x4 size each) is presented in Table 1. Each sub-array communicates with one vehicle. For power consumption comparison, it is assumed that each antenna element consumes 'x' mW power on average when the antenna element is switched ON. For example, in [14], the amount of power consumed per array element is 80mW. In our approach, the parameters shown in Table 1 vary depending on given input vehicular configuration.

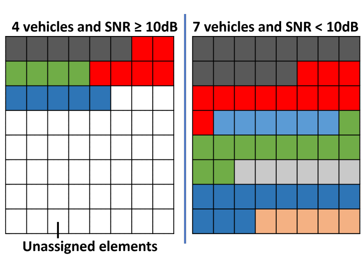


Fig. 5: Comparison of 8x8 antenna array assignment

Table 1: Comparison of this approach with antenna array with fixed sub-arrays

Parameter	Fixed sub-array	This
Maximum no. of vehicles	4	variable
Power consumption	64x	between 2x and 64x

VII. Conclusions

We present a novel approach of optimizing beamforming in vehicle-to-vehicle communication systems equipped with antenna arrays. The proposed approach shows the feasibility of extending to any number of vehicles and any antenna array size.

VIII. Acknowledgements

This research was supported by NSF Award ECCS 1815653.

References

[1] A. Demba and D. P. F. Möller, "Vehicle-to-vehicle communication technology," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, 2018, pp. 0459–0464.
[2] B. Widrow, P. Mantey, L. Griffiths, and B. Goode, "Adaptive antenna systems," *Proceedings of the IEEE*, vol. 55, pp. 2143–2159, 1967.
[3] M. Giordani, A. Zanella, and M. Zorzi, "Millimeter wave communication in vehicular networks: Challenges and opportunities," in *2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 2017, pp. 1–6.
[4] T. Zugno, M. Drago, M. Giordani, M. Polese, and M. Zorzi, "Toward standardization of millimeter-wave vehicle-to-vehicle networks: Open challenges and performance evaluation," *IEEE Communications Magazine*, vol. 58, no. 9, pp. 79–85, 2020.
[5] Y. Yin, H. Chen, Z. Li, R. Fukatsu, T. Yu, and K. Sakaguchi, "Design of antenna configuration for interference control in mmwave v2v communication systems," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–5.
[6] S. Turner, A. Al-Khalil, and A. Al-Sherbaz, "Enhancing the physical layer in v2v communication using ofdm – mimo techniques," 2013.
[7] A. Salh, N. Shah, L. Audah, Q. Abdullah, W. Jabbar, and M. Mohamad, "Energy-efficient power allocation and joint user association in multiuser-downlink massive mimo system," *IEEE Access*, 2020.
[8] A. Zooghby, C. Christodoulou, and M. Georgiopoulos, "Neural network-based adaptive beamforming for one- and two-dimensional antenna arrays," *IEEE Transactions on Antennas and Propagation*, vol. 46, no. 12, pp. 1891–1893, 1998.
[9] H. K. Kim, C. A. Azurdia-Meza, and C. Estévez, "Antenna array synthesis through genetic algorithms for urban v2v communications: Preliminary results," in *2020 Congreso Estudiantil de Electrónica y Electricidad (INGELECTRA)*, 2020, pp. 1–5.
[10] M. C. Tan, M. Li, Q. H. Abbasi, and M. A. Imran, "A wideband beamforming antenna array for 802.11ac and 4.9 ghz in modern transportation market," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 2659–2670, 2020.
[11] C. A. Balanis, "Antenna theory: analysis and design," 2005.
[12] A. Balakrishnan, D. Bouneffouf, N. Mattei, and F. Rossi, "Incorporating behavioral constraints in online ai systems," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, Jul. 2019.
[13] D. Hogg, "Fun with the friis free-space transmission formula," *IEEE Antennas and Propagation Magazine*, vol. 35, no. 4, pp. 33–35, 1993.
[14] L. Wu, H. F. Leung, A. Li, and H. C. Luong, "A 4-element 60-ghz cmos phased-array receiver with beamforming calibration," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 3, pp. 642–652, 2017.
[15] G. Tzortzis and A. Likas, "The global kernel k-means clustering algorithm," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008.