

1-1-2018

# Self-Organizing mmWave Networks: A Power Allocation Scheme Based on Machine Learning

Roohollah Amiri  
*Boise State University*

Hani Mehrpouyan  
*Boise State University*

# Self-Organizing mmWave Networks : A Power Allocation Scheme Based on Machine Learning

Roohollah Amiri\*, Hani Mehrpouyan\*

\*School of Electrical and Computer Engineering, Boise State University,  
{roohollahamiri,hanimehrpouyan}@boisestate.edu

**Abstract**—Millimeter-wave (mmWave) communication is anticipated to provide significant throughput gains in urban scenarios. To this end, network densification is a necessity to meet the high traffic volume generated by smart phones, tablets, and sensory devices while overcoming large pathloss and high blockages at mmWaves frequencies. These denser networks are created with users deploying small mmWave base stations (BSs) in a plug-and-play fashion. Although, this deployment method provides the required density, the amorphous deployment of BSs needs distributed management. To address this difficulty, we propose a self-organizing method to allocate power to mmWave BSs in an ultra dense network. The proposed method consists of two parts: clustering using fast local clustering and power allocation via *Q-learning*. The important features of the proposed method are its scalability and self-organizing capabilities, which are both important features of 5G. Our simulations demonstrate that the introduced method, provides required quality of service (QoS) for all the users independent of the size of the network.

## I. INTRODUCTION

Millimeter-wave (mmWave) communication is one of the main technologies of the next generation of cellular networks (5G). The large bandwidth at mmWave frequency has the potential to enhance network throughput by tenfolds [1]. However, large path loss and shadowing limit the performance of mmWave systems and need to be dealt with. One approach to overcome this problem is based on increasing the density of access points [2], [3]. However, as the number of access points increases, the complexity of network management increases. Keeping this in mind, one of the features of future mmWave base stations (BSs) is self-deployment by users. In other words access points can be deployed in a plug-and-play fashion, and the network architecture may change frequently. Considering the above points, 5G needs self-organizing methods to configure, adapt, or heal itself when necessary. In this paper, a self-organizing algorithm is proposed to maximize the sum capacity in a dense mmWave network while providing users with their required quality of service (QoS). The algorithm consists of clustering, based on fast local clustering (FLOC), and distributed power allocation, via *Q-learning*. Scalability and fast convergence of FLOC, adaptability and distributed nature of *Q-learning*, makes their combination a suitable tool to achieve self-organization in a dense network.

## II. SYSTEM MODEL

The System model considers a dense outdoor urban scenario as an important example of 5G, i.e., we consider the downlink of densely deployed mmWave BSs. To this end let us consider  $N$  mmWave BSs that are distributed based on the homogeneous spatial Poisson point process (SPPP) with

density  $\lambda_{BS}$  [4]. Each BS is associated to one user. BSs share a single frequency resource block (FRB) to support their associated users. We assume a time invariant channel model, i.e. slow fading. The channel vector between the BS  $i$  and user  $k$ , can be written as follows

$$H_{i,k} = (L_{i,k})^{-1} \times g_{i,k}, \quad (1)$$

where  $L_{i,k}$  and  $g_{i,k}$  denote the path loss and the path gain between the BS  $i$  and user  $k$ . The path loss between the BS  $i$  and its associated user  $i$ ,  $L_{i,i}$ , follows the free space propagation based on Frii's law [1]. Here, we consider that the majority of interferers have non-line-of-sight (NLOS) paths [5]. Hence, the path loss  $L_{i,k}$  ( $i \neq k$ ) can be written as [1]

$$L_{i,k} [dB] = \beta_1 + 10\beta_2 \log_{10}(d_{i,k}) + X_\zeta, \quad (2)$$

where  $\beta_1$  and  $\beta_2$  are factors used to achieve best fit to channel measurements,  $d_{i,k}$  is the distance between the BS  $i$  and the user  $k$ ,  $X_\zeta$  denotes the logarithmic shadowing factor, where  $X_\zeta \sim \mathcal{N}(0, \zeta^2)$ , and  $\zeta^2$  denotes the lognormal shadowing variance.

The received signal in the downlink at the  $k^{\text{th}}$  user includes the desired signal from its associated BS (BS  $k$ ), interference from neighboring BSs, and also thermal noise. Hence, the signal-to-interference-noise-ratio (SINR) at the  $k^{\text{th}}$  user is given by

$$\text{SINR}_k = \frac{P_k H_{k,k}}{\sum_{i \in D_k, i \neq k} P_i H_{i,k} + \sigma^2}, \quad (3)$$

where  $P_k$  denotes the power transmitted by the  $k^{\text{th}}$  BS,  $D_k$  is the set of interfering BSs, and  $\sigma^2$  denotes the variance of the additive white Gaussian noise. Accordingly, the normalized capacity at the  $k^{\text{th}}$  user is given by

$$C_k = \log_2(1 + \text{SINR}_k). \quad (4)$$

## III. PROBLEM FORMULATION

The goal of the optimization problem is to find the best power distribution between mmWave BSs ( $\vec{P}$ ) in order to maximize the sum capacity of the network, while supporting all users with their required QoS. The optimization problem ( $\mathbf{P}_1$ ) can be formulated as

$$\underset{\vec{P}}{\text{maximize}} \quad \sum_{k=1}^N \log_2(1 + \text{SINR}_k) \quad (5a)$$

$$\text{subject to} \quad P_k \leq P_{max}, \quad k = 1, \dots, N \quad (5b)$$

$$\text{SINR}_k \geq q_k, \quad k = 1, \dots, N. \quad (5c)$$

Here, the objective (5a) is to maximize the sum capacity of the network while providing all users with their required QoS in (5c). The first constraint, (5b), refers to the power limitation of every BS. The term  $q_k$  in (5c) refers to the minimum required SINR for the  $k^{\text{th}}$  user.

Eq. (5a) contains the interference term in the denominator of SINR term. In a dense network the interference term cannot be ignored [6]. Due to the presence of the interference term, the objective function (5a) is a non-concave function [7].

The solution to  $\mathbf{P}_1$  should have certain features. First, it should be distributed due to no central authority in this network. Second, the range of mmWave BSs is limited, so each user will receive interference from the BSs in its neighborhood. Therefore, the solution should consider local clustering to reduce the computation overhead. Third feature is self-healing. The number of BSs in the network changes sporadically, which means the solution should be adaptive to new possible architectures. Considering the above, in this paper, we propose a method which contains two parts : a fast local clustering method to locally cluster the BSs, and in each cluster, BSs will choose their transmitting power based on *Q-learning* [8]. *Q-learning* is model-free (adaptable) and gives the BSs the ability to learn from their environment by interacting with it (self-organization).

#### IV. CLUSTER BASED DISTRIBUTED POWER ALLOCATION USING Q-LEARNING (CDP-Q)

In our proposed method, mmWave BSs are considered as the agents of *Q-learning*, so the terms agent and mmWave BS are used interchangeably. CDP-Q is a distributed method in which multiple agents (mmWave BSs) find a sub-optimal policy (power allocation) to maximize the network capacity. CDP-Q consists of two parts: (1) clustering, and (2) power allocation. Clustering is based on a local clustering method, and power allocation is based on *Q-learning*. In the following each part is detailed.

##### A. MmWave BSs Clustering

Since mmWave signals suffer from high pathloss and shadowing, only neighboring BSs that are close in distance interfere with each other. Consequently, we propose to use a clustering mechanism to divide BSs into clusters in which the interference of one cluster is negligible on other clusters' users.

In this paper, we propose to use Fast local clustering (FLOC) [9] to divide mmWave BSs into clusters. FLOC is a distributed message-passing clustering method with  $\mathcal{O}(1)$  complexity, which guarantees scalability, and produces non-overlapping clusters. Another feature of FLOC is local self-healing, which means re-clustering, due to addition of a new node or removing a node, does not propagate through all clusters. In order to apply FLOC in a mmWave network, the following concepts are defined:

- *Cluster head (CH)*: The mmWave BS that is chosen as the head of the cluster. In our algorithm, there is no priority between a cluster head and other members of the cluster.
- *In-bound (IB), and out-band (OB) node*: In FLOC, a node is in-bound if it is a unit distance from a CH. A unit

distance is a set value, which in this case is the range of mmWave links, i.e., 100-200 m [1]. Accordingly, we define in-bound as 100 m, which is an indication of strong interference, and out-band as 200 m, which indicates the edge of the cluster around a CH. Finally, if a node  $j$  is in out-band distance of a cluster  $i$ , and not in an in-bound distance of any other clusters, then node  $j$  will join the cluster  $i$  as an OB node.

##### B. Distributed Power Allocation Using Q-Learning

The output of *Q-learning* is a decision policy (power allocation) which is represented as a function called Q-function. Here, the Q-function of agent  $k$  is represented as a table called a Q-table ( $Q_k$ ). The columns of a Q-table are the actions ( $a^k$ ), and the rows are the state ( $s^k$ ) of the agent  $k$ .

In multi-agent *Q-learning*, agents can act independently or cooperatively. In the independent learning, each agent interacts with the environment without communicating with other agents. In fact, it considers the other agents as part of the environment. Independent learning has shown good performance in many applications [10]. In independent learning, since the environment is not stationary, oscillation and longer convergence time might happen for the agents, but due to no communication overhead between agents compared to cooperative learning, we choose independent learning. Motivated by this fact, the agents will select their actions according to [11]

$$a_t^k = \arg \max_a Q_k(s_t^k, a), \quad (6)$$

in which, subscript  $t$  denotes time step  $t$  of *Q-learning*. The CDP-Q algorithm is presented in Algorithm 1.

---

#### Algorithm 1 The proposed CDP-Q algorithm

---

- 1: Cluster formation based on Sec. IV-A
  - 2: **for all** Clusters in Parallel **do**
  - 3:     **for all** Agents **do**
  - 4:         Initialize  $Q(s_t^k, a_t)$  arbitrarily
  - 5:         Initialize  $s_t^k$
  - 6:         **for all** episodes **do**
  - 7:             send  $Q_k(s_t^k, :)$  to other agents of the cluster
  - 8:             receive  $Q_j(s_t^j, :)$ ,  $j \in D_k, j \neq k$
  - 9:             Choose  $a_t^k$  according to Eq. 6
  - 10:             Take action  $a_t^k$ , observe  $R_t^k$
  - 11:              $Q_k(s_t^k, a_t^k) \leftarrow (1 - \alpha)Q_k(s_t^k, a_t^k) + \alpha(R_t^k + \gamma Q_k(s_{t+1}^k, a_t^k))$
  - 12:         **end for**
  - 13:     **end for**
  - 14: **end for**
- 

In the following the actions, states, and the reward function of the proposed *Q-learning* method are defined.

1) *Actions*: The set of actions (powers)  $A$  is defined as  $A = \{a_1, a_2, \dots, a_{N_{\text{power}}}\}$ , which uniformly covers the range between minimum ( $a_1 = P_{\text{min}}$ ) and maximum ( $a_{N_{\text{power}}} = P_{\text{max}}$ ) power.

2) *States*: We define  $N_r$  equally spaced concentric circles around the cluster head (CH) of each cluster. These circles, define  $N_r$  rings with  $r$  units of spacing, around the CH. The state of the agent  $k$  at time step  $t$  is defined as  $s_t^k = (n)$  which shows the ring number that the agent belongs to. Considering

the definition of the Q-table and the states at the beginning of this section, if the agents' location is fixed, each agent will choose just one row of its Q-table to search for the best action decision.

3) *Reward*:  $R_t^k$  is the immediate reward incurred due to selection of the action  $a_t^k$  at state  $s_t^k$  by the agent  $k$  at time step  $t$ . The constraint in (5c), can be represented as:  $C_t^k \geq \log_2(q_k)$ , for  $k = 1, \dots, N$ .  $C_t^k$  is the normalized capacity of agent  $k$  at time step  $t$ . Based on this, the normalized proposed reward function for the agent  $k$  at time step  $t$  is defined as

$$R_t^k = \underbrace{\frac{1}{2 \log_2(q_k)}}_{(a)} \left( \underbrace{C_t^k}_{(b)} - \underbrace{|C_t^k - 2 \log_2(q_k)|}_{(c)} \right). \quad (7)$$

The rationale behind the proposed reward function is as follows

- The term (a) normalizes the value of reward function.
- The objective of the optimization problem is to maximize the capacity of the network, so the term (b) results in a higher reward for higher capacity for an agent.
- To satisfy the QoS constraint for agent  $k$ , capacity deviation of its associated user from the required QoS, term (c), should result in a lower reward.
- There is a maximum reward (+1) for an agent to provide fairness between the agents which is shown in Fig. 1.
- The proposed reward function is a first order function of  $C_t^k$ , which reduces each iteration's complexity.

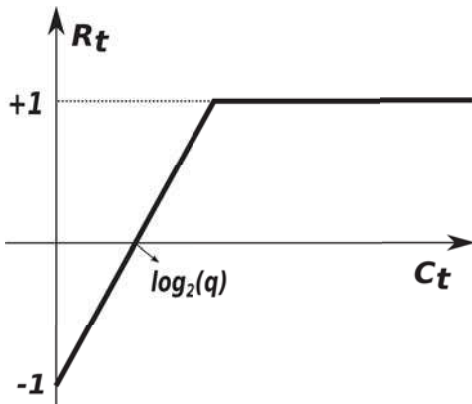


Fig. 1: Proposed reward function (RF).

## V. SIMULATION RESULTS

In this section the simulation setup is detailed and then the results of the simulations are presented.

### A. Simulation Setup

A dense mmWave BS network, with approximately 120 BSs in a  $1 \text{ km}^2$  area is considered. The BSs are distributed based on SPPP and operate independently in the network. Each BS, supports one user equipment (UE), which is located in a radius of 10m around the BS. The QoS for a user is defined as the required SINR to support the user's service. The value of  $q_k = 2.83$  is considered for all the users.

To perform *Q-learning*, the learning rate is considered as  $\alpha = 0.5$ , the discount factor as  $\gamma = 0.9$ ,  $N_{\text{power}} = 31$ ,  $r = 50\text{m}$ , and  $N_r = 4$ . The maximum number of iterations is set to 50,000. The remaining parameters of the simulation are represented in Table I.

TABLE I: Simulation Parameters

Param.	Value	Param.	Value
$f$	28 GHz	$P_{\min}$	-10 dBm
$\zeta$	8.7 dB	$P_{\max}$	35 dBm
$\beta_1$	72.0	$\beta_2$	2.92
$\sigma^2$	-120 dBm		

### B. Clustering Results

The implementation of clustering algorithm, is an event driven, message-passing distributed program in C++. Every BS is simulated as an independent thread, and is added to the network randomly in  $[0, 10]$  seconds. The clustering algorithm converges in less than 15 seconds for the assumed value for the  $\lambda_{BS}$ . The resulted clusters in two different distribution of BSs are shown in different colors in Fig. 2, and 3. Each cluster head (CH) is marked with a filled color.

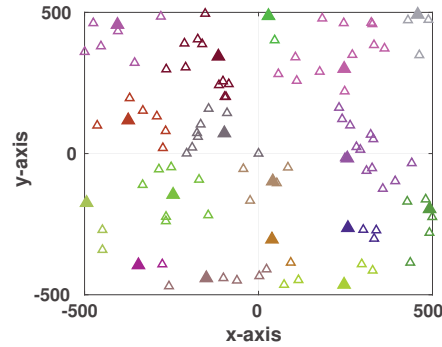


Fig. 2: 124 BSs in  $1 \text{ km}^2$ .

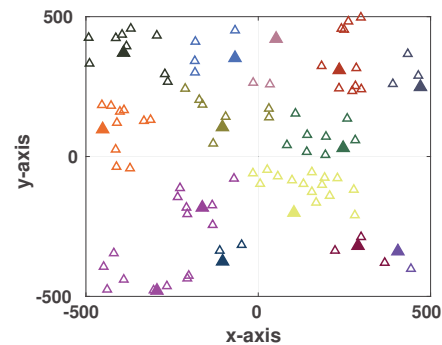


Fig. 3: 122 BSs in  $1 \text{ km}^2$ .

### C. Power Allocation Results

According to [1], [2], the coverage range of millimeter communication is in the range of 100 – 200 m, which means the maximum coverage of  $0.12 \text{ (km}^2\text{)}$  for each mmWave BS. Considering the interference-limited assumption and the value of  $\lambda_{BS} = 120 \text{ (BS/km}^2\text{)}$ , a cluster might have 14 mmWave BSs. Hence, the **CDP-Q** algorithm results in clusters that include 2 to 14 BSs.

The results of power allocation using the proposed reward function are compared to the exponential reward function proposed in [12], which are presented as EXP-Q in the simulations. For all possible cluster sizes, power allocation using the proposed reward function is simulated, and the normalized capacity of all BSs in the clusters are plotted in Fig. 4. The same simulations for EXP-Q are presented in Fig. 5.

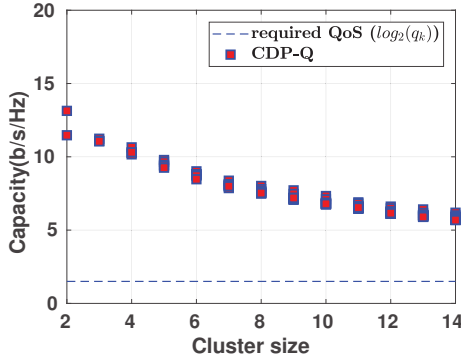


Fig. 4: Capacity of clusters' members.

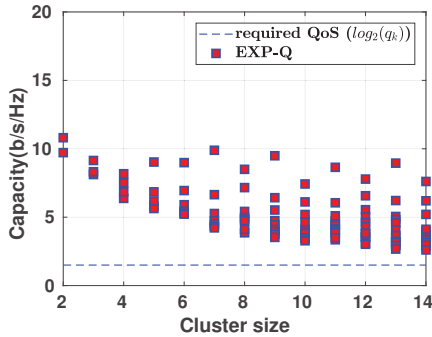


Fig. 5: Capacity of clusters' members.

As it is shown in these figures, while both reward functions satisfy all members for all sizes of clusters with their required QoS, the normalized capacity of users in the CDP-Q are close to each other, while in EXP-Q the normalized capacity of users are much diverse. The diversity of normalized capacity values in EXP-Q effects the fairness index. The fairness index in each cluster is measured using Jain's fairness index [13] and is shown in Fig. 6. In Fig. 6, the CDP-Q maintains fairness for all sizes of clusters, while EXP-Q fails to support users with fairness in large cluster sizes. On the other hand, total capacity of the clusters are shown in Fig. 7 with respect to the cluster size. According to Fig. 7, the CDP-Q provides higher capacity than the EXP-Q for all sizes of clusters.

## VI. CONCLUSION

In this paper, a self-organized distributed power allocation algorithm is presented. The proposed algorithm reduces the optimization complexity by using a distributed clustering method, and provides adaptability in power allocation by using *Q-learning*. The proposed reward function, satisfies required QoS for the users in all sizes of the resulted clusters, and outperforms the exponential based reward function.

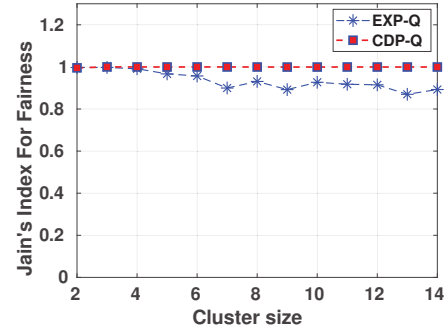


Fig. 6: Jain's fairness index.

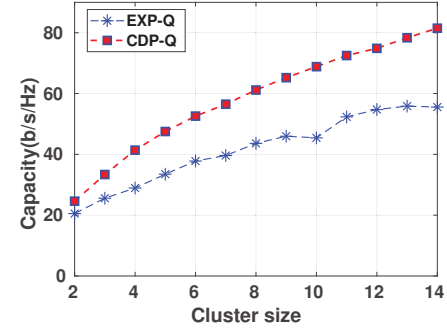


Fig. 7: Sum capacity of clusters.

## REFERENCES

- [1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, March 2014.
- [2] R. Baldemair, T. Irnich, K. Balachandran, E. Dahlman, G. Mildh, Y. Seln, S. Parkvall, M. Meyer, and A. Osseiran, "Ultra-dense networks in millimeter-wave frequencies," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 202–208, January 2015.
- [3] T. Bai and R. W. Heath, "Coverage in dense millimeter wave cellular networks," in *Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 2062–2066.
- [4] D. P. Kroese and Z. Botev, "Spatial process generation," Aug 2013.
- [5] M. Rebato, M. Mezzavilla, S. Rangan, F. Boccardi, and M. Zorzi, "Understanding noise and interference regimes in 5G millimeter-wave cellular networks," in *22th European Wireless Conference*, May 2016, pp. 1–5.
- [6] S. Niknam and B. Natarajan, "On the Regimes in Millimeter wave Networks: Noise-limited or Interference-limited?" *CoRR*, Apr. 2018. [Online]. Available: <https://arxiv.org/abs/1804.03618>
- [7] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug 2006.
- [8] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992. [Online]. Available: <http://dx.doi.org/10.1007/BF00992698>
- [9] M. Demirbas, A. Arora, V. Mittal, and V. Kulathumani, "A fault-local self-stabilizing clustering service for wireless ad hoc networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 9, pp. 912–922, Sept 2006.
- [10] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387–434, Nov 2005. [Online]. Available: <https://doi.org/10.1007/s10458-005-2631-2>
- [11] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.
- [12] H. Saad, A. Mohamed, and T. ElBatt, "Distributed cooperative Q-learning for power allocation in cognitive femtocell networks," in *Proc. IEEE Veh. Technol. Conf.*, Sept 2012, pp. 1–5.
- [13] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *CoRR*, 1998. [Online]. Available: <http://arxiv.org/abs/cs.NI/9809099>