A source coding perspective on node deployment in two-tier networks

Jun Guo, *Student Member, IEEE*, Erdem Koyuncu, *Member, IEEE*, and Hamid Jafarkhani, *Fellow, IEEE*

***Abstract*—Multi-tier networks have many applications in different fields. We define a novel two-tier quantizer that can be applied to different node deployment problems including the energy conservation in two-tier wireless sensor networks (WSNs) consisting of N access points (APs) and M fusion centers (FCs). We aim at finding an optimal deployment of APs and FCs to minimize the average weighted total, or Lagrangian, of sensor and AP powers. For one fusion center, $M = 1$, we show that the optimal deployment of APs is simply a linear transformation of the optimal N-level quantizer for density f, and the sole FC should be located at the geometric centroid of the sensing field. We also provide the exact expression of the AP-Sensor power function and prove its convexity. For more than one fusion center, $M > 1$, we provide a necessary condition for the optimal deployment. Furthermore, to numerically optimize the AP and FC deployment, we propose three Lloyd-like algorithms and analyze their convergence. Simulation results show that our algorithms outperform the existing algorithms.**

I. Introduction

A. One-Tier Quantization for Node Deployment

In many applications, one needs to provide service to and collect data from a geographical area of interest via multiple nodes, such as sensors or providers. Usually, the nodes are distributed such that each point or client in the area is only served by only one node, resulting in a partition of the target area to disjoint regions. The service cost for each node may depend on its characteristics and local service region. The fundamental goal in such a formulation is to jointly optimize the node locations and the corresponding service regions to optimize the overall performance, which is typically defined as an aggregate of node service costs.

As we also discuss in the sequel, the cost of providing service to a client is usually related to its distance to the local service node. Therefore, minimizing the total cost by optimally deploying the nodes and the service regions is identical to a spatial tessellation problem [2]. Such problems (which have also been referred to as facility location or node deployment problems by different research communities) are equivalent to the quantization problem of data compression and source coding. In fact, in the language of quantization theory, the service nodes and service regions correspond to the reproduction points and quantization regions, respectively. Minimizing the total service cost becomes equivalent to minimizing the corresponding average distortion. We now present several specific applications to highlight the above equivalency.

In sensor networks, sensors are deployed in a two-dimensional planar to collect data from the environment. The goal of sensor/node deployment is to maximize the sensing performance and quality. An example of the sensor deployment in a two-dimensional planar is illustrated in Fig. 1a, where each sensor is denoted by a dot that monitors its own region. In this scenario, the reproduction points correspond to the sensor nodes, and the quantization Voronoi regions correspond to the sensor partitions. The cost function (distortion measure) is usually a monotonically increasing function of the distance between the sensor and the event that is being sensed, and quantifies the inaccuracy of sensing. The most common cost function used in the quantization literature is the squared Euclidean distance [3]. It can be used directly in the sensor network example to measure the overall sensing inaccuracy in homogeneous wireless sensor networks (WSNs) [4]. For heterogeneous WSNs, a weighted Euclidean distance square measure can be used where the weights reflect the nodes' different sensing capabilities [5]. Other cost functions have also been utilized to formulate the sensing coverage or the sensing probability [5]–[8].

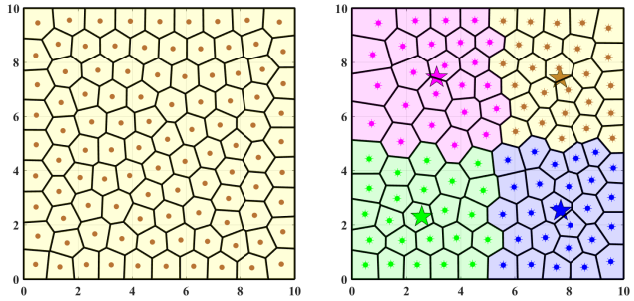


Fig. 1: Two example node deployments. (a) One-tier network. (b) Two-tier network. 100 first-tier nodes and 4 second-tier nodes are denoted by dots and stars, respectively. The cell partitions are denoted by polygons. The symbols associated with the same second-tier node are filled with the same color.

Another example is the heterogeneous base station (BS) deployment [9]–[12] in cellular networks where user equipments (UEs) are considered as the source, heterogeneous BSs are recognized by the reproduction points, and the BS cells are represented by the quantization Voronoi regions. Because of the path loss, the signal strength at the receiver is a non-increasing function of the communication distance [13]–[15]. Using such a non-increasing function as the cost function, the distortion measure can be defined as the expected signal to noise ratio (SNR) at the UEs, where the expectation is calculated for a given channel probability distribution.

B. Two-Tier Quantization for Node Deployment

The conventional spatial tessellation and node deployment problems in Section I-A ignore the hierarchical architecture that is inherent in many networks. In fact, to reduce the network burden, many practical networks have a two-tier structure, as an extension to the one-tier examples considered in Section I-A. For example, to provide an effective delivery service for the residents, the postal system uses a two-tier

This work was supported in part by the DARPA GRAPHS program Award N66001-14-1-4061. The work was presented in part at ICC-17. J. Guo and H. Jafarkhani are with Center for Pervasive Communications & Computing, University of California, Irvine. (e-mail: guoj4@uci.edu; hamidj@uci.edu) E. Koyuncu is with Department of Electrical and Computer Engineering, University of Illinois at Chicago. (e-mail: ekoyuncu@uic.edu)

network including the local post offices and the postal hubs. Then, there will be two costs associated with each delivery. A *first-tier-cost* that is the cost of delivering packages from clients to local post offices, determined by the population density and the distance from clients to their local post offices, and a *second-tier-cost* which is the total cost of delivering packages between local post offices and the corresponding postal hubs. The *second-tier-cost* is determined by the workload of the local post offices and the distance from the local post offices to their postal hubs. A similar two-tier network appears in the hospital system where local hospitals provide the basic medical treatment and the residents with severe disastrous issues are transferred to the hospital centers with more medical facilities.

Two-tier WSNs and two-tier cellular networks are also very common network architectures. A two-tier WSN [22]–[25] includes densely deployed sensors, multiple access points (APs), and fusion centers (FCs). In such a network, sensor nodes collect the data and send it to their APs for processing. Then, each AP transmits its aggregated data to its associated FC [22]–[25]. As depicted in Fig. 1b, the sensors, APs, and FCs correspond, respectively, to clients, first-tier nodes, and second-tier nodes in our two-tier network. One reasonable cost function in the two-tier WSN is the total energy consumption at sensors and APs. The objective is to optimize the trade-off between the sensor and AP energy consumptions.

The goal of this paper is to study node deployment problems in two-tier networks, where two-tier nodes are deployed to provide service for the clients in the target region. In such two-tier networks, as depicted in Fig. 1b, the second-tier nodes provide service for the first-tier nodes that serve the clients. Similar to the one-tier networks, the cost between two points is generally a non-decreasing function of the Euclidean distance. Let the first-tier-cost be the total cost between clients and first-tier nodes, and the second-tier-cost be the total cost between the first-tier nodes and the corresponding second-tier nodes. Moving the first-tier nodes towards the second-tier nodes, usually, will increase the average distance between the first-tier nodes and the local clients, resulting in the increase of the first-tier-cost. On the other hand, moving the first-tier nodes towards the local clients, usually, will increase their distance to the second-tier nodes and will result in an increase in the *second-tier-cost*. Therefore, there is a trade-off between the *first-tier-cost* and the *second-tier-cost*. Like the one-tier quantizer in Section I-A, we propose a two-tier quantizer to optimally manage such a trade-off.

Clearly, such a two-tier quantizer is fundamentally different from the one-tier quantizer discussed in the literature and Section I-A. Even though the two-tier quantizer has diverse applications, as discussed above, for the clarity of the presentation and as an important application, we focus on the example of energy consumption in two-tier WSNs in this paper.

C. Related Work and Applications

A significant body of literature exists on designing the one-tier quantizers. Gray et al. [3] summarize the theory and practice of quantization since its inception. The best possible quantization distortion in a high-resolution regime has been discussed in [26]–[28] and the application of the high-resolution theory to node deployment in heterogeneous sensor networks is provided in [29]. Clustering is a related method where the cluster heads and cluster regions are, respectively, the reproduction points and quantization regions. Many different hierarchical clustering methods, such as Agglomerative Clustering (AC) and Divisive Clustering (DC), are discussed in [30].

Furthermore, some existing quantization schemes in the literature are similar to our two-tiered quantizers. For example, successively refinable vector quantizers (SRVQs) [31]–[33] have multiple stages. However, while SRVQs progressively refine the quantization rate, our quantizers are fixed-rate quantizers and there is no rate refinement. As another example, hierarchical vector quantizers (HVQs) [34]–[36] are proposed to reduce the quantizer encoding complexity. HVQs employ quantizers of different dimensions in different hierarchical steps. In contrast, our method operates on the same dimension at different tiers. In addition, different from the existing SRVQs and HVQs of which the distortion is only determined by the homogeneous reproduction points, the distortion measure of the proposed two-tier quantization depends on both first-tier and second-tier reproduction points. Moreover, different stages of the HVQs are designed independently while we design the two tiers jointly.

Another related topic, facility location or node deployment, has been widely studied for different applications, such as service station distribution and sensing coverage. Jain et al. [37] study the cost optimization for connecting cities to open facilities. The authors assume that the facilities can be placed at finite candidate locations with different opening costs. A similar relaxation idea to solve the problem appears in [19] where the authors fix the facility location and only optimize the cell partitions. In order to balance the workload of the facilities, different cell partitions serve the same amount of area. Another work [38] focuses on the trade-off between sensing quality and communication cost in indoor wireless temperature sensor networks.

The above node deployment problems assume fixed node locations or finite location candidates. A more realistic assumption is the possibility of continuous node locations. In [6], the sensor nodes can be placed everywhere in the target region. Three virtual force algorithms, VECtor-Based Algorithm (VEC), VORonoi-Based Algorithm (VOR), and Minimax Algorithm, are proposed to maximize sensor coverage. Cortes et al. [4] consider multiple distortion measures, including the Euclidean distance square, in homogeneous WSNs. Voronoi Diagram [2] is shown to be the best cell partition for the one-tier homogeneous networks and Lloyd-like Algorithms are applied to minimize the distortions. A similar node deployment problem over a non-convex environment, where the nodes can be placed everywhere other than some obstacles, are considered in [39], [40]. The sensing coverage problem in heterogeneous WSNs where sensors have different sensing ranges is discussed in [41]–[43]. Mahboubi et al. [41] extend the three movement-assisted protocols in [6] to avoid coverage holes in heterogeneous WSNs. Wu et al. [42] utilize a scoring method to improve the sensor coverage rate. Yoon et al. [43] design a genetic algorithm to solve the node

deployment problem and compare the performance with other genetic algorithms. Our previous work [5] analyzes the sensing coverage in one-tier heterogeneous WSNs where the sensors have different sensing abilities and limited communication ranges.

Node deployment problems also appear in the context of cellular networks [9]–[11]. Coskun et al. [9] analyze the BS deployment in heterogeneous cellular networks including macro and micro BSs. The corresponding average distortion is the total channel capacity divided by the total power and the corresponding constraint requires a capacity threshold. In [9], a greedy algorithm is proposed to optimize the micro BS deployment. Zhang et al. [10] consider a multi-tier cellular network where BSs are placed in multiple tiers. The objective function is defined as the ratio of the minimum achievable throughput to the total power consumption. A survey of general BS deployment optimization in heterogeneous cellular networks can be found in [11].

Most existing node deployment algorithms and the corresponding optimizations only consider the nodes in the same tier. However, node deployment in two-tier networks requires joint deployments of nodes in both tiers. A simple extension of the existing algorithms is first to find the optimal first-tier node deployment and cell partition and, fixing the first-tier configuration, then solve the optimal second-tier node deployment and the connection between the two phases. It is not hard to find examples for which such a two-phase approach provides a poor performance. Therefore, a truly joint node deployment for two-tier networks is needed. To the best of our knowledge, the problem of node deployment for two-tier networks, which includes the facility deployment in two different tiers, has not been considered in the literatures.

D. Main Contributions

In this paper, we study the node deployment problem in two-tier networks and make the following contributions: (i) we design a two-tier quantizer with a distortion measure that can be used to formulate different optimization problems including the energy consumption in a two-tier WSN; (ii) we show that the optimal solution for the two-tier quantizer with one node in the second tier can be obtained by shrinking the optimal reproduction points of the optimal one-tier quantizer towards the sole second-tier node; (iii) we propose the necessary condition for the optimal node deployment in two-tier WSNs. The necessary condition helps us to design efficient node deployment algorithms; (iv) we find the optimal deployments and the corresponding minimum powers in two-tier WSNs for certain special cases; (v) we also propose numerical Lloyd-like algorithms to minimize energy consumption in general; (vi) we define a new formulation for the AP-Sensor power function, which shows the trade-off between the two kinds of power consumptions in WSNs, and prove its convexity.

E. Organization of the Paper

The rest of this paper is organized as follows: In Section II, we introduce the system model. In Section III, we find the optimal deployment of APs with one FC. In Section IV, we study the optimal deployment problem with multiple APs and FCs. In Section V, we provide AP-Sensor power function and analyze the trade-off between the AP power and the sensor power. In Section VI, we propose numerical algorithms to minimize the energy consumption. In Section VII, we present numerical simulations. Finally, in Section VIII, we draw our main conclusions. Some of the technical proofs are provided in the appendices.

II. System Model and Problem Formulation

We consider a two-tier WSN consisting of three kinds of nodes: sensors, APs, and FCs. As discussed in Section I, this is a canonical application of the two-tier quantizer. A similar network architecture has been studied in [22] where the authors ignore the energy consumption of the sensor nodes. We model the energy consumption by radio communication in two-tier WSNs.

Let Ω be a convex polygon in $\mathbb{R}^2$ including its interior. Given the target area Ω, N APs and M FCs are deployed to collect data. Without loss of generality, we assume that $N > M$. $\mathcal{I}_{\mathcal{A}} = \{1, \cdots, N\}$, and $\mathcal{I}_{\mathcal{B}} = \{1, \cdots, M\}$ denote the sets of AP indices and FC indices, respectively. AP deployment and FC deployment are, respectively, defined by $P = (p_1, \cdots, p_N)$ and $Q = (q_1, \cdots, q_M)$, where $p_n \in \mathbb{R}^2$ is the location of AP n and $q_m \in \mathbb{R}^2$ is the location of FC m. An AP partition $\mathbf{R}^{\mathcal{A}} = \{R_n^{\mathcal{A}}\}_{n \in \mathcal{I}_{\mathcal{A}}}$ is a collection of disjoint subsets of $\mathbb{R}^2$ whose union is Ω. Let $T : \mathcal{I}_{\mathcal{A}} \to \mathcal{I}_{\mathcal{B}}$ be an index map for which $T(n) = m$ if and only if AP n is connected to FC m. A continuous and differentiable function $f(\cdot) : \Omega^2 \to \mathbb{R}^+$ is used to denote the density of the data rate from the densely distributed sensors [22]. Then, $\int_R f(w)dw$ is the total amount of data generated from the sensors in region R in one time unit.

Usually, FCs have access to reliable energy sources and their energy consumption is not the main concern in this paper. Therefore, we focus on the energy consumption of sensors and APs. In fact, the energy consumed by sensors and APs comes from three parts: (i) Sensors transmit bit streams to APs; (ii) APs transmit bit streams to FCs; (iii) APs receive bit streams from sensors. The transmitting powers (Watt) of nodes, e.g., sensors and APs, mainly depend on two factors: (i) the *instant-transmission-power* (Joules/second); and (ii) the *channel-busy-ratio*: the percentage of time that the transmitter forwards data. The *average-transmitting-power* of AP n is defined as $\mathcal{P}_{t_n}^{\mathcal{A}} = E_{t_n}^{\mathcal{A}} \Gamma_n^{\mathcal{A}}, n \in \mathcal{I}_{\mathcal{A}}$, where $E_{t_n}^{\mathcal{A}}$ is the *instant-transmission-power* of AP n and $\Gamma_n^{\mathcal{A}}$ is the *channel-busy-ratio* for the channel from AP n to the corresponding FC. In order to achieve the required SNR thresholds at the receivers, the *instant-transmission-power* $E_{t_n}^{\mathcal{A}}$ should be set to a value that is determined by the distance, the antenna gain, and the SNR threshold [44]. More realistically, the transmission power is proportional to distance squared or another power of distance in the presence of obstacles [14]. Taking path-loss into consideration, it is reasonable to model the *instant-transmission-power* from AP n to FC $T(n)$ by $E_{t_n}^{\mathcal{A}} = \eta_{n,T(n)}^{\mathcal{A}} \|p_n - q_{T(n)}\|^{\alpha}$, where $\|\cdot\|$ denotes the Euclidean distance, $\eta_{n,T(n)}^{\mathcal{A}}$ is a constant determined by the antenna gain of AP n and the SNR threshold of FC $T(n)$, and $\alpha \in [2, 4]$ is the path-loss parameter. We consider an environment without obstacles, i.e., $\alpha = 2$. Let $\xi_{nT(n)}^{\mathcal{A}}$ be the AP n's *instant-transmitter-data-rate* which is determined by the SNR at the corresponding FC $T(n)$. In this paper, we focus on homogeneous sensors, APs,

and FCs. Hence, we consider identical sensor antenna gains, identical AP antenna gains, identical AP SNR thresholds, and identical FC SNR thresholds. Without loss of generality, we set $\eta^{\mathcal{A}}_{n,T(n)} = \eta^{\mathcal{A}}$ and $\xi^{\mathcal{A}}_{nT(n)} = \xi^{\mathcal{A}}$. In such a homogeneous scenario, the total amount of data is proportional to the number of sensors, and $f(\cdot)$ is just a scaled sensor density function. Note that AP n collects data from the sensors in $R^{\mathcal{A}}_n$, indicating that the *average-receiver-data-rate* - the amount of data received from sensors in the unit time - is $\int_{R^{\mathcal{A}}_n} f(w)dw$. It is reasonable to assume that the AP transmitters have idle time and forward data only when the collected sensing data comes into the buffer. Hence, the *channel-busy-ratio* is proportional to the *average-receiver-data-rate*, and can be written as $\Gamma^{\mathcal{A}}_n = \frac{\int_{R^{\mathcal{A}}_n} f(w)dw \cdot T/\xi^{\mathcal{A}}}{T} = \frac{\int_{R^{\mathcal{A}}_n} f(w)dw}{\xi^{\mathcal{A}}}$, where $\xi^{\mathcal{A}}$ is AP n's *instant-transmitter-data-rate*. Therefore, the *average-transmitter-power* of AP n can be rewritten as $\mathcal{P}^{\mathcal{A}}_{t_n} = E^{\mathcal{A}}_{t_n}\Gamma^{\mathcal{A}}_n = \frac{\eta^{\mathcal{A}}}{\xi^{\mathcal{A}}}\|p_n - q_{T(n)}\|^2 \int_{R^{\mathcal{A}}_n} f(w)dw$.

The power consumption at AP receivers is a constant and does not affect energy optimization [1]. Without loss of generality, we drop it from the total energy consumption. Therefore, to consider the total energy consumption, we use the sum of the APs' *average-transmitter-powers*, $P^{\mathcal{A}}_{t_n}$, which is calculated as

$$\overline{\mathcal{P}}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) = \sum_{n=1}^{N} \mathcal{P}^{\mathcal{A}}_{t_n} = \sum_{n=1}^{N} \int_{R^{\mathcal{A}}_n} \frac{\eta^{\mathcal{A}}}{\xi^{\mathcal{A}}} \|p_n - q_{T(n)}\|^2 f(w)dw. \quad (1)$$

Next, we discuss sensors' total power consumption. Similar to the power consumption at AP receivers, the power consumption at sensor receivers is also a constant and thus dropped from our objective function. In what follows, we focus on power consumption at sensor transmitters. Since f should be approximately uniform on an infinitesimal region $[w, w+dw]$, the total amount of data generated from $[w, w+dw]$ in one time unit is $f(w)dw$. As a result, the sum of *channel-busy-ratios* of the sensors in $[w, w+dw]$ is $\Gamma^{\mathcal{S}} = \frac{f(w)dw \cdot T/\xi^{\mathcal{S}}}{T} = \frac{f(w)dw}{\xi^{\mathcal{S}}}$, where $\xi^{\mathcal{S}}$ is sensors' *instant-transmitter-data-rate*. In addition, sensors within $[w, w+dw]$ are supposed to have approximately the same distance to the corresponding FC q, and then have the same *instant-transmission-power* $E^{\mathcal{S}}_t = \eta^{\mathcal{S}}\|w-q\|^2$, where $\eta^{\mathcal{S}}$ is a constant determined by the antenna gain of sensors and the SNR threshold of the corresponding AP. Hence, the sum of the sensors' *average-transmitter-powers* within $[w, w+dw]$ is $\frac{\eta^{\mathcal{S}}}{\xi^{\mathcal{S}}}\|q-w\|^2 f(w)dw$. Note that sensors in $\mathbf{R}^{\mathcal{A}}$ transfer data to AP n. Therefore, the sum *average-transmitter-powers* of the sensors in the whole region Ω is calculated as

$$\overline{\mathcal{P}}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) = \sum_{n=1}^{N} \int_{R^{\mathcal{A}}_n} \frac{\eta^{\mathcal{S}}}{\xi^{\mathcal{S}}} \|p_n - w\|^2 f(w)dw. \quad (2)$$

We define the scaled AP power and the scaled sensor power, respectively, as

$$\mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) \triangleq \frac{\xi^{\mathcal{A}}}{\eta^{\mathcal{A}}} \overline{\mathcal{P}}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) = \sum_{n=1}^{N} \int_{R^{\mathcal{A}}_n} \|p_n - q_{T(n)}\|^2 f(w)dw, \quad (3)$$

$$\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) \triangleq \frac{\xi^{\mathcal{S}}}{\eta^{\mathcal{S}}} \cdot \overline{\mathcal{P}}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) = \sum_{n=1}^{N} \int_{R^{\mathcal{A}}_n} \|p_n - w\|^2 f(w)dw. \quad (4)$$

In what follows, we focus on the scaled sensor power and the scaled AP power which are also, respectively, referred to as Sensor-power and AP-power for simplicity. On one hand, Sensors' densely deployment prevent the network from breaking by few sensor failures. On the other hand, the limited number of APs makes the recharging possible for APs. Under such scenario, saving the total energy consumption becomes an important requirement. Therefore, one objective could be minimizing the scaled total AP power consumption in (3) given a constraint on the total sensor power (4) or vice versa. Like rate-distortion function $R(D)$ and the distortion-rate function $D(R)$, we define the AP-sensor power function and Sensor-AP power function, respectively, as

$$A(S) \triangleq \inf_{(P,Q,\mathbf{R}^{\mathcal{A}},T):\mathcal{P}^{\mathcal{S}}(P,\mathbf{R}^{\mathcal{A}}) \le S} \mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) \quad (5)$$

$$S(A) \triangleq \inf_{(P,Q,\mathbf{R}^{\mathcal{A}},T):\mathcal{P}^{\mathcal{A}}(P,Q,\mathbf{R}^{\mathcal{A}},T) \le A} \mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}). \quad (6)$$

We can then define the Lagrangian objective function (two-tier distortion) to be minimized as

$$\begin{aligned} &D(P, Q, \mathbf{R}^{\mathcal{A}}, T) = \mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) + \beta \mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) \\ &= \sum_{n=1}^{N} \int_{R^{\mathcal{A}}_n} \left[\|p_n - w\|^2 + \beta \|p_n - q_{T(n)}\|^2\right] f(w)dw. \end{aligned} \quad (7)$$

Our main goal is to minimize the two-tier distortion over the following four variables: (i) AP deployment P; (ii) FC deployment Q; cell partition $\mathbf{R}^{\mathcal{A}}$; (iv) index map T. As we discussed in Section I, there are many applications that result in a two-tier quantization set-up and as long as the corresponding distortion is a function of the distance, our results will hold for those applications as well. Another interpretation is to consider (1) as the sensor energy consumption and (2) as the sensing quality measured by the mean square error (MSE) in a one-tier WSN scenario in which dots in Fig. 1b are sensors, stars are APs and the density function $f(\cdot)$ is replaced by the event probability density function. In other words, the proposed optimization problem can also provide the trade-off between energy efficiency and sensing quality in one-tier WSNs.

III. The Best Possible Distortion for the Two-tier WSNs with One FC

As discussed earlier, the node deployment problem in two-tier WSNs can be interpreted as a two-tier quantizer design problem whose reproduction points are APs and FCs. Similarly, if one only considers the energy consumption of sensors, the corresponding optimization problem becomes a "regular" (or one-tier) quantization problem with distortion

$$D_r(X, \mathbf{R}) = \sum_{n=1}^{N} \int_{R_n} \|x_n - w\|^2 f(w)dw. \quad (8)$$

Let $(X^*, \mathbf{R}^*) = \arg\min_{(X,\mathbf{R})} D_r(X, \mathbf{R})$ be the optimal one-tier quantizer. In some cases [22]–[25], only one FC or fusion center is deployed to collect data from the entire WSNs. In the case of one FC and multiple APs, the following proposition holds.

Proposition 1. *For a two-tier WSN with one FC, We have the following:*

(i) The optimal FC location is the centroid of the target region, i.e., $q^* = \frac{\int_\Omega w f(w) dw}{\int_\Omega f(w) dw}$.

(ii) The optimal AP locations of the two-tier WSNs are linear transformations of the optimal reproduction points $X^* = (x_1^*, \ldots, x_N^*)$*, i.e.,* $p_n^* = \frac{x_n^* + \beta q^*}{1+\beta}, n \in \mathcal{I}_\mathcal{A}$.

(iii) The optimal AP partition is the same as the optimal one-tier quantizer partition $\mathbf{R}^*=(R_1^*, \ldots, R_N^*)$*, i.e.,* $\mathbf{R}^{\mathcal{A}*}=\mathbf{R}^*$.

(iv) The best possible distortion is $\frac{1}{1+\beta} D_r(X^*, \mathbf{R}^*) + \frac{\beta}{1+\beta} \int_\Omega \|w - q^*\|^2 f(w) dw$.

The proof of the proposition is provided in Appendix A. By Proposition 1, one can obtain the optimal solution for the two-tier quantization by shrinking the optimal reproduction points of the one-tier quantizer towards the corresponding FCs. Note that the algorithms AC and DC in [30] identify the optimal reproduction of the N-level one-tier quantizer as the AP deployment and then determine the FC deployment in terms of the identified AP deployment, and vice versa. In what follows, we provide an example to compare the quantizer generated by AC/DC with our proposed optimal quantizer, respectively. Consider one FC, n APs, and a uniform distribution over the 1-dimensional target region $\Omega = [s, t]$. The reproduction points and the cells of an optimal one-tier quantizer are given by $x_n^* = s + \frac{(2n-1)(t-s)}{2N}, n \in \mathcal{I}_\mathcal{A}$, and $R_n^* = \left[s + \frac{(n-1)(t-s)}{N}, s + \frac{n(t-s)}{N}\right], n \in \mathcal{I}_\mathcal{A}$. In this case, the quantizers generated by AC and DC are identical. There is no surprise that the AP deployment and cell partition of AC/DC are exactly the reproduction points and partition of the optimal one-tier quantizer, i.e., $p_n' = x_n^*$, $R_n^{\mathcal{A}'} = R_n^*, n \in \mathcal{I}_\mathcal{A}$. FC location is the geometric centroid $q' = (t+s)/2$. However, according to Proposition 1, the optimal AP deployment and cell partition are, respectively, $p_n^* = s + \frac{\beta(t-s)}{2(1+\beta)} + \frac{(2n-1)(t-s)}{2N(1+\beta)}$ and $R_n^{\mathcal{A}*} = R_n^*, n \in \mathcal{I}_\mathcal{A}$. Similar to the solution of AC/DC, the optimal FC location is $q^* = q' = (t+s)/2$. The corresponding minimum distortion is $\frac{(t-s)^2}{12(1+\beta)N^2} + \frac{\beta(t-s)^2}{12(1+\beta)}$. In particular, for $\beta = 1$ and $\Omega = [-\frac{1}{2}, \frac{1}{2}]$ with 1 FC and 4 APs, the solution of AC/DC, shown in Fig. 2a, is $q' = 0$, $R_n^{\mathcal{A}'} = \left[\frac{n-3}{4}, \frac{n-2}{4}\right]$, and $p_n' = \frac{2n-5}{8}$, $n \in \mathcal{I}_\mathcal{A}$, indicting a distortion of $\frac{61}{768}$. But, the optimal FC location is $q^* = 0$, the optimal cells are $R_n^{\mathcal{A}*} = \left[\frac{n-3}{4}, \frac{n-2}{4}\right], n \in \mathcal{I}_\mathcal{A}$, the optimal AP locations are $p_n^* = \frac{2n-5}{16}, n \in \mathcal{I}_\mathcal{A}$, and the best possible distortion is $\frac{17}{384} < \frac{61}{768}$. The optimal two-tier quantizer is shown in Fig. 2b.

Furthermore, the best possible distortion can be exactly determined in the high resolution regime $N \to \infty$. In fact, the best possible distortion $D_r(X^*, R^*)$ of a one-tier quantizer is given by [26]

$$\kappa_d d N^{-\frac{2}{d}} \|f\|_{\frac{d}{d+2}} + o(N^{-\frac{2}{d}}), N \to \infty, \tag{9}$$

where $\|f\|_q \triangleq (\int_{\mathbb{R}^d} f^q(x) dx)^{\frac{1}{q}}$, and κ_d depends only on the dimension d. For example, we have $\kappa_1 = \frac{1}{12}$ and $\kappa_2 = \frac{5}{18\sqrt{3}}$ [26], [28]. By using Proposition 1, the best possible distortion in high resolution regime $N \to \infty$ with one FC is

$$\frac{1}{1+\beta} \kappa_d d N^{-\frac{2}{d}} \|f\|_{\frac{d}{d+2}} + \frac{\beta}{1+\beta} \int_\Omega \|w - c\|^2 f(w) dw + o(N^{-\frac{2}{d}}). \tag{10}$$

IV. The Optimal Deployment in Two-tier WSNs with Multiple FCs

In this section, we extend the analysis of the optimal deployment to WSNs with multiple FCs. Given multiple FCs, APs are divided into clusters in terms of the index map T, and the number of clusters is M. In particular, the m-th cluster is defined as $\mathcal{N}_m \triangleq \{n : T(n) = m\}$. Let N_m be the number of elements in $\mathcal{N}_m$, and $W_m = \bigcup_{n \in \mathcal{N}_m} R_n^\mathcal{A}$ be the m^{th} cluster region. Distortion in this general case is determined by (i) AP deployment, (ii) FC deployment, (iii) AP cell partition, and (iv) Clustering (or the index map from APs to FCs). Before we discuss the optimal AP and FC deployment, we need to know (a) the best index map T given P, Q, and $\mathbf{R}^\mathcal{A}$, and (b) the best AP partition $\mathbf{R}^\mathcal{A}$ given P, Q, and T.

The index map T only influences the second term in (7). To minimize the second term, each AP should transfer data to the closest FC. Thus, the best index map is $T_{[P,Q]}^E(n) = \arg\min_m \|p_n - q_m\|$. However, given P, Q, and $T = T_{[P,Q]}^E$, AP cell partition $\mathbf{R}^\mathcal{A}$ affects both terms in (7). The best AP cell partitions, called the energy Voronoi diagrams (EVDs), are

$$\begin{aligned} V_n^E(P,Q) = \{w \big| &\|p_n - w\|^2 + \beta \|p_n - q_{T_{[P,Q]}^E(n)}\|^2 \\ &\le \|p_l - w\|^2 + \beta \|p_l - q_{T_{[P,Q]}^E(l)}\|^2, \forall l \ne n, n \in \mathcal{I}_\mathcal{A}\}. \end{aligned} \tag{11}$$

Now, let $\mathbf{V}^E(P,Q) = \{V_n^E(P,Q)\}_{n \in \mathcal{I}_\mathcal{A}}$ be the energy Voronoi partition. Putting the best index map $T_{[P,Q]}^E(\cdot)$ and the best AP partition $\mathbf{V}^E(P,Q)$ into (7), the distortion is

$$\begin{aligned} &\widetilde{D}(P,Q) = D(P, Q, \mathbf{V}^E(P,Q), T_{[P,Q]}^E) \\ &= \sum_{n=1}^N \int_{V_n^E(P,Q)} \left(\|p_n - w\|^2 + \beta \min_m \|p_n - q_m\|^2\right) f(w) dw. \end{aligned} \tag{12}$$

Let $P^* = (p_1^*, \ldots, p_N^*)$ and $Q^* = (q_1^*, \ldots, q_M^*)$ be, respectively, the optimal AP and FC deployments. Let $v_n(P^*, Q^*) = \int_{V_n^E(P^*, Q^*)} f(w) dw$ be the Lebesgue measure (volume) of $V_n^E(P^*, Q^*)$. Without loss of generality, we may assume $v_n(P^*, Q^*) > 0$, as quantization cells with zero probability do not affect the overall distortion.

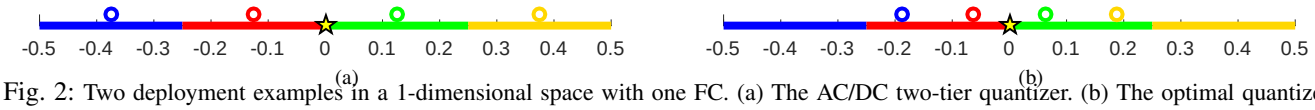
Proposition 2. *Let* $\alpha = 2$ *and* $N > M$*. The necessary conditions for the optimal deployments in the two-tier WSN with distortion defined by (7) are*

$$p_n^* = \frac{c_n(P^*, Q^*) + \beta q_{T_{[P^*,Q^*]}^E(n)}^*}{1+\beta}, n \in \mathcal{I}_\mathcal{A}, \tag{13}$$

$$q_m^* = \frac{\sum_{n: T_{[P^*,Q^*]}^E(n) = m} c_n(P^*, Q^*) v_n(P^*, Q^*)}{\sum_{n: T_{[P^*,Q^*]}^E(n) = m} v_n(P^*, Q^*)}, m \in \mathcal{I}_\mathcal{B}, \tag{14}$$

where p_n^* *is the optimal location for AP* n *and* q_m^* *is the optimal location for FC* m*,* $T_{[P^*,Q^*]}^E(n) = \arg\min_m \|p_n^* - q_m^*\|$ *is the best index map,* $v_n(P^*, Q^*) = \int_{V_n^E(P^*,Q^*)} f(w) dw$ *is the Lebesgue measure (volume) of* $V_n^E(P^*, Q^*)$ *and* $c_n(P^*, Q^*)$ *is the geometric centroid of* $V_n^E(P^*, Q^*)$.

The proof of the proposition is provided in Appendix B. According to (13), the optimal location of AP n, connected to FC m, should be on the segment $\overline{c_n(P,Q) q_m}$. According to (14), the best location of FC m should be the geometric centroid of the m^{th} cluster region $\bigcup_{n: T(n) = m} V_n^E(P^*, Q^*)$. Obviously, the optimal deployment and the optimal partition in Proposition 1 also satisfy the necessary conditions in Proposition 2. In the next section, using Proposition 2, we design



(a) (b)

Fig. 2: Two deployment examples in a 1-dimensional space with one FC. (a) The AC/DC two-tier quantizer. (b) The optimal quantizer. AP and FC locations are denoted by circles and star. The optimal partition cells are denoted by intervals. Each AP and its corresponding cell are illustrated by the same color.

Lloyd-like algorithms to determine the optimal deployment.

First, note that when the AP cell partition is fixed, the geometric centroid $c_n, n \in \mathcal{I}_{\mathcal{A}}$, and the volume of the cells $v_n, n \in \mathcal{I}_{\mathcal{A}}$, are fixed. Second, the index map $T^E_{[P,Q]}$ represents the best connection between APs and FCs (or clustering) if and only if P and Q are given. We now find the optimal deployment, the optimal partition, the optimal index map and the best possible distortion for a uniform density in one-dimensional space.

Theorem 1. *Let $\Omega = [s,t]$ with length $\mu(\Omega) = t-s$. Also, let*

$$\ell_a = \left(\beta + \lceil \tfrac{N}{M} \rceil^{-2}\right)^{-\frac{1}{2}}, \qquad \ell_b = \left(\beta + \lfloor \tfrac{N}{M} \rfloor^{-2}\right)^{-\frac{1}{2}} \tag{15}$$

$$M_a = (N \bmod M), \qquad M_b = M - (N \bmod M). \tag{16}$$

Then, given a uniform distribution on Ω with M FCs and N APs, the minimum distortion is

$$\frac{\mu^2(\Omega)}{12(1+\beta)} \left(M_a \ell_a + M_b \ell_b\right)^{-2}. \tag{17}$$

The minimum is achieved if and only if

- *(i) M_a of the clusters consist of $\lceil \frac{N}{M} \rceil$ APs each and have length $\ell_a \mu(\Omega)/(M_a\ell_a + M_b\ell_b)$,*
- *(ii) M_b of the clusters consist of $\lfloor \frac{N}{M} \rfloor$ APs each and have length $\ell_b \mu(\Omega)/(M_a\ell_a + M_b\ell_b)$,*
- *(iii) FCs are deployed at the centroids of the cluster regions,*
- *(iv) AP cells are uniform partitions of the cluster, and*
- *(v) AP n is deployed at $\frac{c_n + \beta q_{T(n)}}{1+\beta}$, $n \in \mathcal{I}_{\mathcal{A}}$, where c_n is the geometric centroid of AP n's cell.*

The proof of the theorem provided in Appendix C.

In particular, when $K = \frac{N}{M}$ is a positive integer, the optimal FC locations are $q^*_m = s + \frac{(2m-1)(t-s)}{2M}, m \in \mathcal{I}_{\mathcal{B}}$, and the optimal index map is $T^*(n) = \left\lceil \frac{n}{K} \right\rceil, n \in \mathcal{I}_{\mathcal{A}}$. The optimal AP locations are $p^*_n = s + \frac{(t-s)}{(1+\beta)} \left(\frac{(2n-1)}{2N} + \beta \frac{(2\lceil \frac{n}{K} \rceil - 1)}{2M} \right), n \in \mathcal{I}_{\mathcal{A}}$, and the optimal AP cell partitions are $R^{\mathcal{A}*}_i = \left[s + \frac{(n-1)(t-s)}{N}, s + \frac{n(t-s)}{N} \right], n \in \mathcal{I}_{\mathcal{A}}$. The corresponding minimum distortion is $\frac{(t-s)^2}{12(1+\beta)M^2} \left(\frac{1}{K^2} + \beta \right)$. However, in such a uniform distributed 1-dimensional scenario, the quantizers generated by AC [30] and DC [30] are identical but different from the above optimal solution. The corresponding FC and AP deployments of AC/DC are, respectively, the optimal M-level and N-level one-tier reproduction points, i.e., $p'_n = s + \frac{(2n-1)(t-s)}{2N}, q'_m = s + \frac{(2m-1)(t-s)}{2M}, n \in \mathcal{I}_{\mathcal{A}}, m \in \mathcal{I}_{\mathcal{B}}$. The corresponding AP cell partitions are $R^{\mathcal{A}'}_n = \left[s + \frac{(n-1)(t-s)}{N}, s + \frac{n(t-s)}{N} \right], n \in \mathcal{I}_{\mathcal{A}}$. In what follows, we provide an example to elucidate the gap between AC/DC quantizer and the optimal quantizer. For $\beta = 1$ and $\Omega = [-\frac{1}{2}, \frac{1}{2}]$ with 2 FC and 6 APs, the AC/DC quantizer, illustrated by Fig. 3a, is $Q' = \{-\frac{1}{4}, \frac{1}{4}\}$, $P' = \{-\frac{5}{12}, -\frac{3}{12}, -\frac{1}{12}, \frac{1}{12}, \frac{3}{12}, \frac{5}{12}\}$, and $R^{\mathcal{A}'}_n = \left[\frac{n-4}{6}, \frac{n-3}{6} \right]$, $n \in \mathcal{I}_{\mathcal{A}}$. The corresponding distortion is thus $\frac{49}{2592}$. Nonetheless, according to Theorem 1, the optimal FC deployment is $Q^* = \{-\frac{1}{4}, \frac{1}{4}\}$, the optimal AP deployment is $P^* = \{-\frac{1}{3}, -\frac{1}{4}, -\frac{1}{6}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}\}$, the optimal AP cells are $R^{\mathcal{A}*}_n = \left[\frac{n-4}{6}, \frac{n-3}{6} \right], n \in \mathcal{I}_{\mathcal{A}}$, and the optimal index map is $T^*(n) = \lceil \frac{n}{3} \rceil, n \in \mathcal{I}_{\mathcal{A}}$. The optimal quantizer, shown in Fig. 3b, gains the minimum distortion, $\frac{5}{432}$ which is much smaller than $\frac{49}{2592}$. The distortion gap holds for more complicated circumstances, e.g., non-uniform distributed 2-dimensional space. Therefore, in order to approach the optimal distortion, it is necessary to designed a better two-tier quantizer. This is the topic of the discussion in Section VI.

V. AP-Sensor Power Function

As explained before, minimizing the distortion in (7) is the unconstrained version of minimizing the sum of the AP powers when the sum of the sensor powers is constrained or vise versa. In this section, we study the minimization of the sum of the AP powers when the sum of the sensor powers is constrained via analyzing the AP-Sensor power function in (5). An AP-Sensor power pair (x, y) is said to be achievable if there exists a solution $(P, Q, \mathbf{R}^{\mathcal{A}}, T)$ with Sensor-power $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) \leq x$ and AP-power $\mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) = y$. $(P, Q, \mathbf{R}^{\mathcal{A}}, T)$ is said to be a feasible solution for the AP-Sensor power pair $(S, A(S))$ if and only if $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) \leq S$ and $\mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) = A(S)$. Let $F(S)$ be the set of the feasible solutions for the point $(S, A(S))$, and $\hat{F} = \bigcup_S F(S)$ be the set of all feasible solutions for all points on the curve $A(S)$. It is self-evident that every point above the curve $A(S)$ is achievable. We now discuss the convexity of the AP-Sensor power function. By parallel axis theorem, (2) can be rewritten as

$$\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) = \sum_{n=1}^{N} \int_{R^{\mathcal{A}}_n} \|c_n - w\|^2 f(w) dw + \sum_{n=1}^{N} \|p_n - c_n\|^2 v_n, \tag{18}$$

where $c_n = \frac{\int_{R^{\mathcal{A}}_n} w f(w) dw}{\int_{R^{\mathcal{A}}_n} f(w) dw}$ is the centroid of $R^{\mathcal{A}}_n$, $v_n = \int_{R^{\mathcal{A}}_n} f(w) dw$ is the volume of $R^{\mathcal{A}}_n$. Note that c_n and v_n are functions of $\mathbf{R}^{\mathcal{A}}$. c_n and v_n are constants if and only if the AP cell partition $\mathbf{R}^{\mathcal{A}}$ is fixed. Furthermore, AP-power, (1), can be rewritten as

$$\mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) = \sum_{n=1}^{N} \|p_n - q_{T(n)}\|^2 v_n. \tag{19}$$

The representations (18) and (19) are convenient for the following convexity analysis.

Lemma 1. *Let $D_R(k)$ be the minimum distortion of the k-level one-tier quantizer on space Ω. Let M and N be, respectively, the number of FCs and the number of APs, where $N > M$. The AP-Sensor power function is a non-increasing function with the domain $[D_R(N), +\infty)$ such that $A(S) > 0$ when $D_R(N) \leq x < D_R(M)$ and $A(S) = 0$ when $x \geq D_R(M)$.*

The proof of the lemma is provided in Appendix D.

(a) (b)

Fig. 3: Two deployment examples in a 1-dimensional space with two FCs. (a) The AC/DC two-tier quantizer. (b) The optimal two-tier quantizer. AP and FC locations are denoted by circles and stars. The optimal partition cells are denoted by intervals. Each AP and its corresponding cell are illustrated by the same color. The two clusters are denoted by solid and dashed lines.

A. Closed-form formulas and convexity for one FC

In this section, we assume only one FC and derive a closed-form analytical formula for the AP-Sensor power function. We also prove that the AP-sensor power function is convex.

Lemma 2. *When one FC $Q = (q)$ and multiple APs $P = (p_1, \ldots, p_N)$ are provided, and $\mathbf{R}^{\mathcal{A}}$ is fixed, the minimum $\mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T)$ with the constraint $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) \leq S$ is defined by*

$$\widehat{A}(S, \mathbf{R}^{\mathcal{A}}) \triangleq \inf_{(P,Q,T):\mathcal{P}^{\mathcal{S}}(P,\mathbf{R}^{\mathcal{A}})\leq S} \mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T). \quad (20)$$

We have the following results: (i) The domain of $\widehat{A}(S, \mathbf{R}^{\mathcal{A}})$ is $\{(S, \mathbf{R}^{\mathcal{A}}) | H(\mathbf{R}^{\mathcal{A}}) \leq S\}$;
(ii) When $(S, \mathbf{R}^{\mathcal{A}}) \in \{(S, \mathbf{R}^{\mathcal{A}}) | H(\mathbf{R}^{\mathcal{A}}) \leq S, D_R(N) \leq S < D_R(M)\}$, we have

$$\begin{cases} \widehat{A}(S, \mathbf{R}^{\mathcal{A}}) = \left[\sqrt{(S - H(\mathbf{R}^{\mathcal{A}}))} - \sqrt{D_R(1) - H(\mathbf{R}^{\mathcal{A}})}\right]^2, & S \in [D_R(N), D_R(1)) \\ 0, & S \in [D_R(1), +\infty) \end{cases} \quad (21)$$

where $H(\mathbf{R}^{\mathcal{A}}) = \sum_{n=1}^{N} \int_{R_n^{\mathcal{A}}} \|c_n - w\|^2 f(w) dw$, $c_n = \frac{\int_{R_n^{\mathcal{A}}} w f(w) dw}{\int_{R_n^{\mathcal{A}}} f(w) dw}$ is the centroid of $R_n^{\mathcal{A}}$, and $v_n = \int_{R_n^{\mathcal{A}}} f(w) dw$ is the volume of $R_n^{\mathcal{A}}$.

The proof is provided in Appendix E.

Theorem 2. *When one FC and multiple APs are provided, i.e., $M = 1$ and $N > 1$, the AP-Sensor power function $A(S)$ is a convex function with the expression*

$$A(S) = \begin{cases} \left[\sqrt{S - D_R(N)} - \sqrt{D_R(1) - D_R(N)}\right]^2, & S \in [D_R(N), D_R(1)) \\ 0, & S \in [D_R(1), +\infty) \end{cases} \quad (22)$$

The proof of the theorem is provided in Appendix F. When $M = 1$, as an inverse function of $A(S)$, the Sensor-AP power function is also a convex function with the expression

$$S(A) = \begin{cases} D_R(N) + \left[\sqrt{D_R(1) - D_R(N)} - \sqrt{A}\right]^2, & A \in [0, D_R(1) - D_R(N)) \\ D_R(N), & A \in [D_R(1) - D_R(N), +\infty) \end{cases} \quad (23)$$

Therefore, for $M = 1$, finding the solution of $A(S)$ on $S \in [D_R(N), D_R(1))$ is equivalent to (i) finding the minimizer of $\mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) + \lambda \mathcal{P}^{\mathcal{S}}(P, Q)$, where the Lagrange Multiplier $\lambda = -\left(\frac{\partial A(S)}{\partial S}\right) = \sqrt{\frac{D_R(1) - D_R(N)}{S - D_R(N)}} - 1$ is the negative derivative at $(S, A(S))$ and (ii) finding the minimizer of $D(P, Q, \mathbf{R}^{\mathcal{A}}, T) = \mathcal{P}^{\mathcal{S}}(P, Q) + \beta \mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T)$, where the Lagrange Multiplier $\beta = -\left(\frac{\partial A(S)}{\partial S}\right)^{-1} = \frac{1}{\sqrt{\frac{D_R(1) - D_R(N)}{S - D_R(N)}} - 1}$ is the negative reciprocal of the derivative at $(S, A(S))$.

In what follows, we provide an example for the 1-dimensional space $[s, t]$ and a uniform distribution. On one hand, due to the quantization theorem [3], we have $D_R(1) = \frac{(t-s)^2}{12}$ and $D_R(N) = \frac{(t-s)^2}{12N^2}$. Therefore, (22) becomes

$$A(S) = S - \frac{(t-s)^2}{6N^2} + \frac{(t-s)^2}{12} - 2\sqrt{\left(S - \frac{(t-s)^2}{12N^2}\right)\left(\frac{(t-s)^2}{12} - \frac{(t-s)^2}{12N^2}\right)}. \quad (24)$$

On the other hand, let $(P^*, Q^*, \mathbf{R}^{\mathcal{A}*}, T^*)$ be an optimal solution for $D(P, Q, \mathbf{R}^{\mathcal{A}}, T)$ with $\beta = \frac{1}{\sqrt{\frac{D_R(1) - D_R(N)}{S - D_R(N)}} - 1}$. By Proposition 1, we have $\mathcal{P}^{\mathcal{S}}(P^*, \mathbf{R}^{\mathcal{A}*}) = S$ and

$$\mathcal{P}^{\mathcal{A}}(P^*, Q^*, \mathbf{R}^{\mathcal{A}*}, T^*) = S - \frac{(t-s)^2}{6N^2} + \frac{(t-s)^2}{12} - 2\sqrt{\left(S - \frac{(t-s)^2}{12N^2}\right)\left(\frac{(t-s)^2}{12} - \frac{(t-s)^2}{12N^2}\right)}, \quad (25)$$

which is the same as (24).

VI. Node Deployment Algorithms

We introduce three algorithms, One-Tiered Lloyd (OTL), Two-Tiered Lloyd (TTL), and Combining Lloyd (CL), to minimize the distortion in two-tier WSNs. First, we quickly review the conventional Lloyd algorithm. Lloyd Algorithm has two basic steps in each iteration: (i) The node deployment is optimized while the partitioning is fixed; (ii) The partitioning is optimized while the node deployment is fixed. As shown in [5], Lloyd algorithm, which provides good performance and is simple enough to be implemented distributively, can be used to solve one-tier quantizers or one-tier node deployment problems. However, the conventional Lloyd Algorithm cannot be applied to two-tier WSNs where two kinds of nodes are deployed. We thus introduce three Lloyd-based algorithms to solve the optimal deployment problem in two-tier WSNs.

A. One-tier Lloyd Algorithm

OTL combines two independent Lloyd Algorithms. Using the Lloyd algorithm, an M-level one-tier quantizer is designed and its reproduction points are used as Q. Another N-level one-tier quantizer is designed and its partition is used as $\mathbf{R}^{\mathcal{A}}$. The index map is determined by $T(n) = \arg\min_m \|p'_n - q_m\|$ and the deployment P is determined by $p_n = \min_m \frac{p'_n + \beta q_m}{1+\beta}, n \in \mathcal{I}_{\mathcal{A}}$, where p'_n is the n^{th} reproduction point obtained by the N-level quantizer. Using Proposition 1, it is easy to show that, for the networks with one FC, the distortion of OTL converges to the minimum as long as the second Lloyd Algorithm provides the optimal N-level quantizer.

B. Two-tier Lloyd Algorithm

Before we introduce the details of TTL, we introduce two concepts: (i) AP local distortion and (ii) FC local distortion. The AP local distortion is defined as

$$D_n^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) = \int_{R_n^{\mathcal{A}}} \left[\|p_n - w\|^2 + \beta \|p_n - q_{T(n)}\|^2\right] f(w) dw.$$

The total distortion is the sum of the AP local distortions, i.e.,

$$D(P, Q, \mathbf{R}^{\mathcal{A}}, T) = \sum_{n=1}^{N} D_n^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T).$$

Similarly, for $\mathcal{N}_m \triangleq \{n : T(n) = m\}$, the FC local distortion is defined as

$$D_m^{\mathcal{B}}(P,Q,\mathbf{R}^{\mathcal{A}},T)=\sum_{n\in\mathcal{N}_m}\int_{R_n^{\mathcal{A}}}\left(\|p_n-w\|^2+\beta\|p_n-q_m\|^2\right)f(w)dw.$$

The total distortion is the summation of these FC local distortions, i.e., $D(P,Q,\mathbf{R}^{\mathcal{A}},T)=\sum_{m=1}^{M}D_m^{\mathcal{B}}(P,Q,\mathbf{R}^{\mathcal{A}},T)$. Let c_n and v_n be, respectively, the geometric centroid and the volume of the current AP cell partition. TTL iterates over four steps: (i) AP n moves to $\frac{c_n+\beta q_{T(n)}}{1+\beta}$; (ii) AP partitioning is done by assigning the corresponding EVD to each AP node; (iii) FC m moves to $\frac{\sum_{n\in\mathcal{N}_m}p_nv_n}{\sum_{n\in\mathcal{N}_m}v_n}$; (iv) Clustering is done by assigning the nearest FC to each AP. Furthermore, to avoid the APs with a zero-measure partition, we move such an AP towards a randomly selected FC q until the distance to the q is $min_n\|p_n-q\|$ after (ii).

In what follows, we show that the distortion with TTL converges. First, due to the parallel axis theorem and (28), the local distortion of AP n can be rewritten as

$$\begin{aligned}D_n^{\mathcal{A}}(P,Q,\mathbf{R}^{\mathcal{A}},T)&=\frac{1}{1+\beta}\int_{R_n^{\mathcal{A}}}\|c_n-w\|^2f(w)dw\\&+(1+\beta)\|p_n-\widehat{p}_n\|^2v_n+\frac{\beta}{1+\beta}\int_{R_n^{\mathcal{A}}}\|w-q_{T(n)}\|^2f(w)dw,\end{aligned}\tag{26}$$

where $\widehat{p}_n=\frac{c_n+\beta q_{T(n)}}{1+\beta}$. When Q, $\mathbf{R}^{\mathcal{A}}$, and T are given, the first term and the third term of (26) are constants. In other words, the AP local distortion becomes a function of $\|p_n-\widehat{p}_n\|$. Therefore, Step (i) does not increase the AP local distortions and then the total distortion. Second, given P, Q and T, EVDs minimize the total distortion, indicating that the total cost is not increased by Step (ii). Third, observe that for $\widehat{q}_m=\frac{\sum_{n\in\mathcal{N}_m}p_nv_n}{\sum_{n\in\mathcal{N}_m}v_n}$, we have $\sum_{n\in\mathcal{N}_m}v_n\|p_n-q_m\|^2=\sum_{n\in\mathcal{N}_m}v_n\left[\|p_n-\widehat{q}_m\|^2+\|q_m-\widehat{q}_m\|^2\right]$. Therefore, the local distortion of FC m can be rewritten as

$$\begin{aligned}D_n^{\mathcal{B}}(P,Q,\mathbf{R}^{\mathcal{A}},T)&=\sum_{n\in\mathcal{N}_m}\int_{R_n^{\mathcal{A}}}\|p_n-w\|^2f(w)dw\\&+\beta\left(\sum_{n\in\mathcal{N}_m}v_n\right)\|q_m-\widehat{q}_m\|^2+\beta\sum_{n\in\mathcal{N}_m}\left(v_n\|p_n-\widehat{q}_m\|^2\right).\end{aligned}\tag{27}$$

When P, $\mathbf{R}^{\mathcal{A}}$, and T are given, the first term and the third term in (27) are constants. In other words, the FC local distortion becomes a function of $\|q_m-\widehat{q}_m\|$. Therefore, Step (iii) does not increase the FC local distortions and then the total distortion. Last, given P, Q, and $\mathbf{R}^{\mathcal{A}}$, $T^E_{[P,Q]}(n)=\arg\min_m\|p_n-q_m\|$ minimizes the total distortion, indicating that the total distortion is not increased by Step (iv). In other words, the algorithm generates a positive non-increasing sequence of distortion values and therefore will converge.

C. Combining Lloyd Algorithm

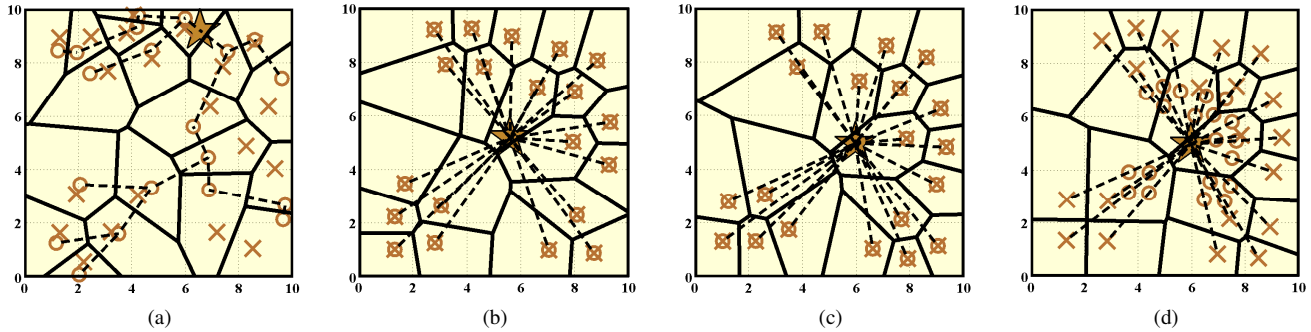
Note that there is no guarantee to achieve a minimum distortion with OTL when $M>1$. The distortion with TTL converges to the local minimum, however, depends on the initial deployments. Our simulations show that starting with some initial deployments, TTL ends with large minimum distortions. A natural idea to avoid these issues is to combine the two algorithms. First, OTL is applied to obtain a deployment with small distortion as the initial deployment for TTL. Second, TTL is applied to further decrease the distortion. We refer to such an algorithm as Combining Lloyd (CL) Algorithm.

VII. Performance Evaluation

We provide the simulation results in two two-tier WSNs: (i) WSN1: A two-tier WSN including one FC and 20 APs; (ii) WSN2: A two-tier WSN including 4 FCs and 20 APs. Similar to [4], [5], the target region is set to $\Omega=[0,10]^2$ and β is set to 1. The traffic density function is the sum of five Gaussian functions of the form $5\exp(0.5(-(x-x_c)^2-(y-y_c)^2))$, where centers (x_c,y_c) are (8,1), (4,9), (7.6,7.6), (9.4,5), and (2,2). We generate 50 initial AP and FC deployments on Ω randomly, i.e, every node location is generated with uniform distribution on Ω. For each initial AP and FC deployments, we connect every AP to its closest FC and then assign the corresponding EVD to the AP node. The maximum number of iterations is set to 100. FCs and APs are denoted, respectively, by colored five-pointed stars and colored circles. The corresponding geometric centroid of AP cells are denoted by colored crosses. Each FC and its connected APs form a cluster. To make clusters more visible, the symbols in the same cluster are filled with the same color. As discussed in Section I, the distortion is the weighted power. We compare the distortion of our three algorithms (OTL, TTL, and CL) with Minimum Energy Routing (MER), Agglomerative Clustering (AC) [30], and Divisive Clustering (DC) [30] algorithms. AC and DC are two clustering methods applied to multi-tier networks. MER combines Voronoi Partition [2] and Bellman-Ford algorithms. On one hand, Voronoi Partition is the optimal cell partition in the first tier (one-tier WSNs). On the other hand, when edge costs are set as the AP powers, Bellman-Ford Algorithm provides the flow with the minimum energy consumption [23], [45], indicating the optimal routing protocol in one tier network. Note that multi-hop communication among sensors is unnecessary because of the small distance between a sensor and its corresponding AP [22]. Therefore, the existing routing protocol, Bellman-Ford, is only applied to the second tier.

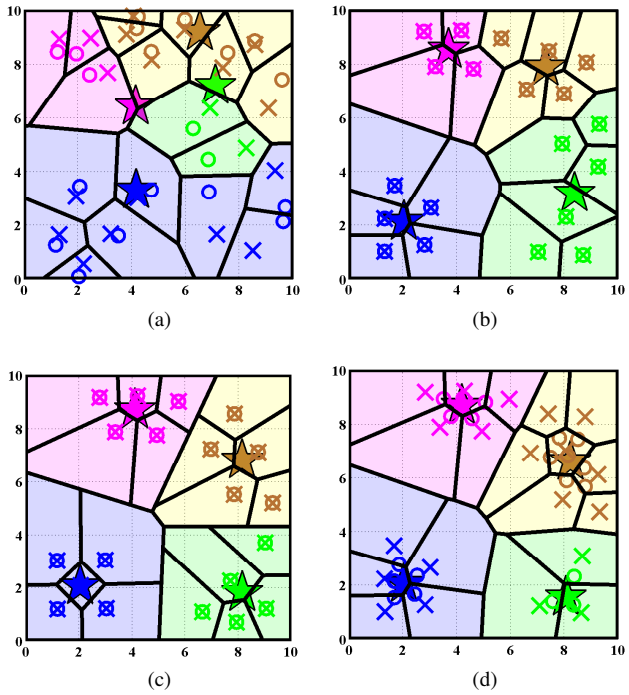
Figs. 4a, 4b, 4c, and 4d show one example of the final deployments of the four algorithms (MER, AC, DC, and CL) in WSN1. For MER, the multi-hop paths are denoted by black dotted lines in 4a. From Fig. 4d, we can find that APs are placed on the line between the corresponding FCs and cell centroids, as expected from Proposition 2. Compared to MER, CL saves 41.90% of the weighted total power in WSN1. However, AC and DC consume more energy compared to MER. Figs. 5a, 5b, 5c, and 5d show examples of the final deployments of MER, AC, DC, and CL, in WSN2, respectively. Compared to MER, the AC, DC, and CL save, respectively, 60.69%, 69.92%, and 81.55% of the weighted total power. Moreover, in Figs. 4 and 5, the FC deployments of the three algorithms, AC, DC, and CL, are approximately the same, but the AP deployments of AC and DC are more uniform than that of CL. Intuitively, compared to the MER, AC, and DC Algorithms, the APs are deployed closer to FCs and then produce smaller AP powers in CL Algorithm, which justifies our observation that CL saves more power compared to the other three algorithms.

Figs 6a and 6b illustrate the weighted total power (7) of different algorithms in WSN1 and WSN2. Our algorithms, OTL, TTL, and CL, outperform MER, AC, and DC in both



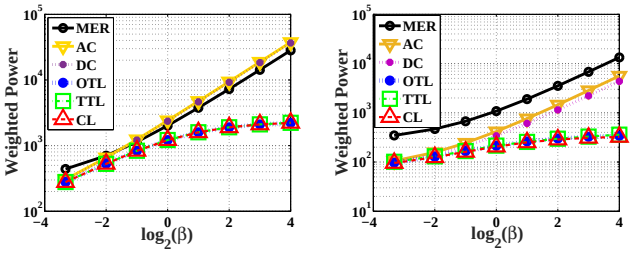
(a) (b) (c) (d)

Fig. 4: AP and FC deployments of different algorithms in WSN1. (a) MER. (b) AC (c) DC. (d) CL.



(a) (b) (c) (d)

Fig. 5: AP and FC deployments of different algorithms in WSN2. (a) MER. (b) AC (c) DC. (d) CL.

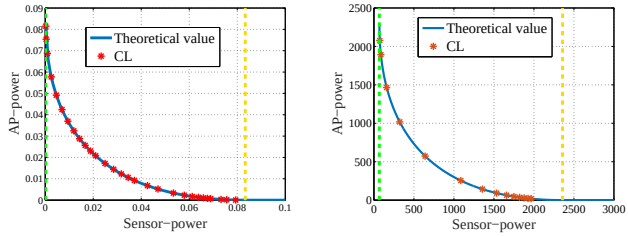


(a) (b)

Fig. 6: The weighted power comparison of different Algorithms. (a) WSN1. (b) WSN2.

networks, especially when β is large (the energy consumed on APs is significant). In WSN1, the three proposed algorithms have almost the same weighted power savings. Although it cannot be seen from the figure, for WSN2, TTL's power saving relative to MER is about 1% better than that of OTL. Comparing the two schemes directly, the power saving of TTL over that of OTL is about 1-3%. Besides, CL, whose performance is better than OTL and TTL in WSN2, requires the shortest running time among the three algorithms. When $\beta = 1$, the running time of repeating each algorithm 50 times is provided in Table I. Note that the running time is dominated by the calculation of the partition. OTL calculates Voronoi partitions for both FCs and APs. TTL calculates energy Voronoi partitions for APs only. Therefore, the running time of OTL increases as we increase the number of FCs. On the contrary, the running times of TTL are almost the same because WSN1 and WSN2 have the same number of APs. CL, as a combination of OTL and TTL, is associated with the running time that is related to the number of FCs. Besides, TTL spends a lot of time to relocate APs with zero-measure partitions because of the bad initial deployments. However, CL attains a good deployment by operating OTL before TTL which greatly reduces the time consumed by AP relocation. Consequently, as shown in Table I, CL requires the shortest running time among the three algorithms.

Figs. 7a and 7b illustrate the comparison between the AP Sensor power pair $(\mathcal{P}^{\mathcal{S}}, \mathcal{P}^{\mathcal{A}})$ using CL with different values of β and the AP-Sensor power function $A(S)$ in (5). One FC and twenty APs are provided, i.e., $M = 1$ and $N = 20$. Fig. 7a is based on a one-dimensional region uniformly distributed in $[0, 1]$ and Fig. 7b is based on WSN1. The green dotted line shows the value of $D_R(20)$ and the orange dotted line shows the value of $D_R(1)$. $D_R(1)$ and $D_R(20)$ in Fig. 7a are theoretical values derived from the quantization theory, i.e., $D_R(1) = \frac{1}{12}$ and $D_R(20) = \frac{1}{12*20^2}$. $D_R(1)$ and $D_R(20)$ in Fig. 7b are obtained from repeating Lloyd Algorithm and choosing the smallest distortion. From Figs. 7a and 7b, we find that the AP Sensor power pair $(\mathcal{P}^{\mathcal{S}}, \mathcal{P}^{\mathcal{A}})$ obtained from CL matches the theoretical AP-Sensor power function very well.



(a) (b)

Fig. 7: The comparison between the AP-Sensor power function and the performance of TTL: (a) One-dimensional region uniformly distributed in [0,1]. (b) WSN1.

TABLE I: Running times(s)

Algorithms	WSN1	WSN2
OTL	1931.41	2196.03
TTL	7078.12	7047.40
CL	1858.88	1968.30

VIII. Conclusion

A two-tier quantizer and its application, energy efficiency in a two-tier wireless sensor network (WSN), are studied in this paper. Different from one-tier WSNs, the two-tier WSN collects data from a large-scale wireless sensor network to fusion centers through access points. The necessary condition for optimal deployment implies that every AP location should be deployed between the centroid of its cell and its associated FC. In addition, the AP-Sensor power function is introduced and analyzed to provide the minimum AP power with a sensor power constraint. We also proposed Lloyd-like algorithms to minimize the distortion. Our simulation results show that our algorithms significantly save the weighted power or energy in two-tier WSNs. In this work, we do not consider collaboration among sensors; such a case will be considered in our future work.

Appendix A
Proof of Proposition 1

When there is only one FC, all APs transfer data to the unique FC located at q, and the index map is simply given by $T(n)=1, n\in\mathcal{I}_{\mathcal{A}}$. The corresponding distortion is

$$D(P,Q,\mathbf{R}^{\mathcal{A}},T)=\sum_{i=1}^{N}\int_{R_n^{\mathcal{A}}}\left[\|p_n-w\|^2+\beta\|p_n-q\|^2\right]f(w)dw.$$

Since

$$\|p_n-w\|^2+\beta\|p_n-q\|^2=(1+\beta)\left\|p_n-\tfrac{(w+\beta q)}{1+\beta}\right\|^2+\tfrac{\beta\|w-q\|^2}{1+\beta}, \quad (28)$$

we obtain

$$\begin{aligned}D(P,Q,\mathbf{R}^{\mathcal{A}},T)=&\frac{1}{1+\beta}\sum_{n=1}^{N}\int_{R_n^{\mathcal{A}}}\|((1+\beta)p_n-\beta q)-w\|^2f(w)dw\\&+\frac{\beta}{1+\beta}\int_{\Omega}\|w-q\|^2f(w)dw,\end{aligned}\quad (29)$$

The first term in (29) is the distortion of a one-tier quantizer with linear transformation of its reproduction points (AP locations). The minimum value of the first term is $D_r(X^*,\mathbf{R}^*)$ and can be achieved by choosing the optimal AP deployment P for any FC location q. On the other hand, the second term in (29) is the distortion of another quantizer whose reproduction point is the FC, and is independent of the choice of AP locations and partition cells. In other words, the second term only depends on the FC location q. As a result, one can optimize (29) by finding the optimal q^* to minimize the second term and then calculate the optimal AP deployment P^* for q^*. By parallel axis theorem, the second term achieves the minimum if and only if the FC is placed at the geometric centroid $c=\frac{\int_\Omega wf(w)dw}{\int_\Omega f(w)dw}$ of Ω, which proves (i). The best possible distortion is then the summation of $\frac{1}{1+\beta}D_r(X^*,\mathbf{R}^*)$ and $\frac{\beta}{1+\beta}\int_\Omega\|w-c\|f(w)dw$, which proves (iv). The two-tier quantizer achieves this minimum when $q^*=c$, $x_n^*=\left((1+\beta)p_n^*-\beta q^*\right), n\in\mathcal{I}_{\mathcal{A}}$, and $\mathbf{R}^{\mathcal{A}}=\mathbf{R}^*$, which proves (ii) and (iii).

Appendix B
Proof of Proposition 2

Let $(P^*,Q^*,\mathbf{R}^{\mathcal{A}*},T^*)$ be an optimal solution for (7). As we show at the beginning of Sec. IV, given the optimal deployment (P^*,Q^*), the optimal partition and the optimal index map are, respectively, $\mathbf{R}^{\mathcal{A}*}=V^E(P^*,Q^*)$ and $T^*=T^E_{[P^*,Q^*]}$. Thus, the optimal geometric centroid and the optimal Lebesgue measure (volume) of $\mathbf{R}^{\mathcal{A}*}$ can be represented as $c_n(P^*,Q^*)$ and $v_n(P^*,Q^*)$, where $c_n(P,Q)=\frac{\int_{V^E(P,Q)}wf(w)dw}{\int_{V^E(P,Q)}f(w)dw}$ and $v_n(P,Q)=\int_{V^E(P,Q)}f(w)dw$. According to the parallel axis theorem, given the optimal partition $V^E(P^*,Q^*)$ and the optimal index map $T^E_{[P^*,Q^*]}$, the objective function in (7) can be expressed as

$$\begin{aligned}D\Big(P,Q,V^E(P^*,Q^*),T^E_{[P^*Q^*]}\Big)=&\sum_{n=1}^{N}\int_{V_n^E(P^*,Q^*)}\|c_n(P^*,Q^*)-w\|^2f(w)dw\\&+\|p_n-c_n(P^*,Q^*)\|^2v_n(P^*,Q^*)+\sum_{n=1}^{N}\left[\beta\|p_n-q_{T^E_{[P^*,Q^*]}}\|^2v_n(P^*,Q^*)\right].\end{aligned}\quad (30)$$

The partial derivatives of (30) are

$$\begin{aligned}&\frac{\partial D(P,Q,V^E(P^*,Q^*),T^E_{[P^*,Q^*]})}{\partial p_n}\\&=2\left[(p_n-c_n(P^*,Q^*))+\beta\Big(p_n-q_{T^E_{[P^*,Q^*](n)}}\Big)\right]v_n(P^*,Q^*), n\in\mathcal{I}_{\mathcal{A}},\end{aligned}$$

and

$$\frac{\partial D(P,Q,V^E(P^*,Q^*),T^E_{[P^*,Q^*]})}{\partial q_m}=\sum_{n:T^E_{[P^*,Q^*]}(n)=m}2\beta(q_m-p_n)v_n(P^*,Q^*), m\in\mathcal{I}_{\mathcal{B}}.$$

Since (30) is a convex function of P and Q, the optimal deployment (P^*,Q^*) satisfies zero gradient. Solving for p_n^* and q_m^*, we obtain

$$p_n^*=\frac{c_n(P^*,Q^*)^*+\beta q^*_{T^E_{[P^*,Q^*](n)}}}{1+\beta}, n\in\mathcal{I}_{\mathcal{A}} \quad (31)$$

$$q_m^*=\frac{\sum_{n:T^E_{[P^*,Q^*]}(n)=m}p_n^*v_n(P^*,Q^*)}{\sum_{n:T^E_{[P^*,Q^*]}(n)=m}v_n(P^*,Q^*)}, m\in\mathcal{I}_{\mathcal{B}} \quad (32)$$

Substituting (31) to (32), we have

$$q_m^*=\frac{\sum_{n:T^E_{[P^*,Q^*]}(n)=m}c_n(P^*,Q^*)v_n(P^*,Q^*)}{\sum_{n:T^E_{[P^*,Q^*]}(n)=m}v_n(P^*,Q^*)}, m\in\mathcal{I}_{\mathcal{B}}. \quad (33)$$

Appendix C
Proof of Theorem 1

Before we discuss the best possible distortion in the uniformly distributed 1-dimensional space, we need to present the following concepts and lemmas. Let $\mu(W)$ be the (Lebesgue) measure of the set W. Let $d(N,\Omega)=\min_{x_1,\ldots,x_N}\int_\Omega\min_n\|x_n-w\|^2\frac{dw}{\mu(\Omega)}$ be the minimum distortion of the N-level one-tier quantizer for a uniform distribution on $\Omega\subset\mathbb{R}$.

Lemma 3. *We have $d(N,\Omega)\geq\frac{\mu(\Omega)^2}{12N^2}$ with equality if and only if Ω is the union of N disjoint intervals, each with measure $\frac{\mu(\Omega)}{N}$.*

Proof. Let $x_1,\ldots,x_N\in\Omega$, and $R_1,\ldots,R_N\subset\mathbb{R}$ respectively denote the reproduction points and the quantization cells of the optimal one-tier quantizer that achieves $d(N,\Omega)$. Note that x_n is the centroid of R_n. We have (34), where (a) follows since for any $x_n\in\mathbb{R}$, we have $d(1,R_n)=\int_{R_n}\|x_n-w\|^2\frac{dw}{\mu(R_n)}$

$$d(N,\Omega)\overset{(a)}{=}\sum_{n=1}^{N}\int_{R_n}\|x_n-w\|^2\frac{dw}{\mu(\Omega)}\overset{(b)}{=}\sum_{n=1}^{N}\frac{\mu(R_n)}{\mu(\Omega)}d(1,R_n)\overset{(c)}{\geq}\sum_{n=1}^{N}\frac{1}{\mu(\Omega)}\frac{\mu(R_n)^3}{12}\overset{(d)}{\geq}\frac{1}{12\mu(\Omega)}\left(\sum_{n=1}^{N}\mu(R_n)\right)^3N^{-2}\overset{(e)}{=}\frac{\mu(\Omega)^2}{12N^2}, \tag{34}$$

by definition, and (c) follows since for any $W\subset\mathbb{R}$, we have $d(1,W)\geq\frac{\mu(W)^2}{12}$ with equality if and only if W is an interval [46]. Also, (d) is the reverse Hölder's inequality, and (e) follows since $\sum_{n=1}^{N}\mu(R_n)=\mu(\Omega)$. Note that (c) is an equality if and only if Ω_ns are intervals, and (d) is an equality if and only if $\mu(\Omega_n)=\frac{\mu(\Omega)}{N},\forall n$. Therefore, (e) can be achieved if and only if Ω is the union of disjoint intervals with the same measure $\frac{\mu(\Omega)}{N}$. ∎

Lemma 4. *Let N and M be two positive integers such that $N\geq M$. We define a function $D^{LB}(e_1,\ldots,e_M)=\left(\sum_{m=1}^{M}\left(\beta+\frac{1}{e_m^2}\right)^{-\frac{1}{2}}\right)^{-2}$ with the domain $R_c=\{(e_1,\ldots,e_M)|\sum_{m=1}^{M}e_m=N,e_m\in\mathbb{N},\forall m\}$, where β is a non-negative constant. Let $M_a=(N \bmod M)$ and $M_b=M-M_a$. Then, $D^{LB}(e_1,\ldots,e_M)$ attains the unique minimum*

$$\left(M_a\left(\beta+\frac{1}{\lceil\frac{N}{M}\rceil^2}\right)^{-\frac{1}{2}}+M_b\left(\beta+\frac{1}{\lfloor\frac{N}{M}\rfloor^2}\right)^{-\frac{1}{2}}\right)^{-2} \tag{35}$$

where M_a of the e_ms are equal to $\lceil\frac{N}{M}\rceil$ and M_b of the e_ms are equal to $\lfloor\frac{N}{M}\rfloor$. In particular, when $M_a=0$, $D^{LB}(e_1,\ldots,e_M)$ attains the unique minimum $\left(\frac{\beta}{M^2}+\frac{1}{N^2}\right)$ at $\left(\frac{N}{M},\ldots,\frac{N}{M}\right)$.

Proof. Let $\mathbf{e}=(e_1,\ldots,e_M)\in R_c$, $D^{UB}(\mathbf{e})=\sum_{m=1}^{M}\left(\beta+\frac{1}{e_m^2}\right)^{-\frac{1}{2}}$. Minimizing D^{LB} is equivalent to maximizing D^{UB}. Let $i,j\in\{1,\ldots,M\}$ be two arbitrary indices. When $e_k,\forall k\neq i,j$ and $\gamma=e_i+e_j$ are fixed, D^{UB} only depends on the difference between e_i and e_j. Without loss of generality, suppose $e_i\geq e_j$. Let $\delta=e_i-e_j$, we have

$$\begin{aligned}\widetilde{D}_{ij}^{UB}(\delta)&=D^{UB}\left(e_1,\ldots,\frac{\zeta\gamma+\delta}{2},\ldots,\frac{\gamma-\delta}{2},\ldots,e_m\right)\\&=\sum_{k\neq i,j}\left(\beta+\frac{1}{e_k^2}\right)^{-\frac{1}{2}}+\left(\beta+\frac{1}{\left(\frac{\gamma+\delta}{2}\right)^2}\right)^{-\frac{1}{2}}+\left(\beta+\frac{1}{\left(\frac{\gamma-\delta}{2}\right)^2}\right)^{-\frac{1}{2}},\end{aligned} \tag{36}$$

and therefore,

$$\frac{\partial\widetilde{D}_{ij}^{UB}(\delta)}{\partial\delta}=\frac{1}{2}\left[\left(1+\beta\left(\frac{\gamma+\delta}{2}\right)^2\right)^{-\frac{3}{2}}-\left(1+\beta\left(\frac{\gamma-\delta}{2}\right)^2\right)^{-\frac{3}{2}}\right]. \tag{37}$$

Let $g(y)=\frac{1}{2}y^{-\frac{3}{2}},y\in(0,\infty)$, $y(x)=1+x^2$, $x\in[0,\infty)$. Since e_i and e_j are non-negative and $e_i\geq e_j$, we have $\delta\geq 0$, $\gamma\geq 0$, and then $x_1=\frac{\gamma+\delta}{2}\geq\frac{\gamma-\delta}{2}=x_2$. Consequently, $y(x_1)\geq y(x_2)>0$, and thus, $\frac{\partial\widetilde{D}^{UB}(\delta)}{\partial\delta}=g(y(x_1))-g(y(x_2))\leq 0$ with equality if and only if $\delta=0$. Therefore, $\widetilde{D}_{ij}^{UB}(\delta)$ is a decreasing function for non-negative continuous δ.

Let $\mathbf{e}^*=(e_1^*,\ldots,e_M^*)\triangleq\arg\min_{(e_1,\ldots,e_M)\in R_c}D^{LB}(e_1,\ldots,e_M)$ be a minimizer of D^{LB} on R_c, and $\hat{\delta}\triangleq\min_{i\neq j}|e_i^*-e_j^*|$ be the minimum difference among e_m^*s. Since e_i^*s are positive integers, we have $\hat{\delta}\in\mathbb{N}$. In what follows, we show that $\hat{\delta}\in\{0,1\}$. Suppose $\hat{\delta}\geq 2$. Then, we can find two indices $i,j\in\{1,\ldots,M\}$ such that $\delta=e_i^*-e_j^*\geq 2$. Let $\mathbf{e}'=(e_1^*,\ldots,e_i',\ldots,e_j',\ldots,e_M^*)$ be a new solution where $e_i'=e_i^*-1$, and $e_j'=e_j^*+1$. We have $\delta'=e_i'-e_j'=e_i^*-e_j^*-2=\delta-2<\delta$. Since $\widetilde{D}_{ij}^{UB}(\delta)$ is a monotonically decreasing function for non-negative continuous δ, we have $\widetilde{D}_{ij}^{UB}(\delta')>\widetilde{D}_{ij}^{UB}(\delta)$ where δ' and δ are non-negative integers. Thus, we have $D^{UB}(\mathbf{e}')>D^{UB}(\mathbf{e}^*)$ which contradicts the optimality of $\mathbf{e}^*$.

Therefore, $\hat{\delta}\in\{0,1\}$, and e_m^*s can thus assume at most 2 distinct values. Suppose M_1 of the e_m^*s are equal to h and M_2 of the e_m^*s are equal to $h+1$, where $h\geq 0$ is an integer. It is self-evident that at least one of M_1 or M_2 should be positive. Without loss of generality, suppose $M_1>0$ and $M_2\geq 0$. Since $M_1+M_2=M$ and $M_1>0$, we have $0<M_1\leq M$ and $0\leq M_2<M$. From the equalities $M_1+M_2=M$ and $M_1h+M_2(h+1)=N$, we obtain $Mh+M_2=N$. Solving the system $Mh+M_2=N$ and $0\leq M_2<M$, we have $h=\lfloor\frac{N}{M}\rfloor$ and $M_2=N \bmod M=M_a$. Finally, using the equality $M_1+M_2=M$, we can determine $M_1=M-(N \bmod M)=M_b$. ∎

Now, we have enough tools to derive the best possible distortion in the uniformly distributed 1-dimensional space. We have

$$\begin{aligned}D(P,Q,\mathbf{R}^{\mathcal{A}},T)&=\sum_{m=1}^{M}\sum_{n\in\mathcal{N}_m}\int_{R_n^{\mathcal{A}}}\left(\|p_n-w\|^2+\beta\|p_n-q_m\|^2\right)dw\\&\overset{(a)}{=}\sum_{m=1}^{M}\sum_{n\in\mathcal{N}_m}\int_{R_n^{\mathcal{A}}}\left(\frac{1}{1+\beta}\left\|(1+\beta)p_n-\beta q-w\right\|^2+\frac{\beta}{1+\beta}\|w-q\|^2\right)dw\\&\overset{(b)}{\geq}\sum_{m=1}^{M}\left[\frac{1}{1+\beta}d(N_m,W_m)+\frac{\beta}{1+\beta}d(1,W_m)\right]\frac{\mu(W_m)}{\mu(\Omega)}\\&\overset{(c)}{\geq}\frac{1}{12(1+\beta)\mu(\Omega)}\sum_{m=1}^{M}\mu^3(W_m)\left(\beta+\frac{1}{N_m^2}\right)\\&\overset{(d)}{\geq}\frac{1}{12(1+\beta)\mu(\Omega)}\left(\sum_{m=1}^{M}\mu(W_m)\right)^3\left(\sum_{m=1}^{M}\left(\beta+\frac{1}{N_m^2}\right)^{-\frac{1}{2}}\right)^{-2}\\&\overset{(e)}{=}\frac{\mu^2(\Omega)}{12(1+\beta)}\left(\sum_{m=1}^{M}\left(\beta+\frac{1}{N_m^2}\right)^{-\frac{1}{2}}\right)^{-2}\\&\overset{(f)}{\geq}\frac{\mu^2(\Omega)}{12(1+\beta)}\min_{\substack{N_1,\ldots,N_M\in\mathcal{N}\\\sum_{m=1}^{M}N_m=N}}\left(\sum_{m=1}^{M}\left(\beta+\frac{1}{N_m^2}\right)^{-\frac{1}{2}}\right)^{-2}\end{aligned} \tag{38}$$

where (a) follows from (28), the first inequality follows from the definition of $d(N,\Omega)$, the second inequality follows from Lemma 3, and the third inequality is the reverse Hölder's inequality. All these inequalities can be made tight with a specific choice of p_n, q_m, W_ms and N_ms. In fact, by Proposition 2, (b) is an equality if and only if $p_n=\frac{c_n+\beta q_{T(n)}}{1+\beta}$ and q_m is the centroid of W_m, indicating (iii) and (v) in Theorem 1. Also, according to Lemma 3, (c) is an equality if and only if $W_m,m\in\mathcal{I}_{\mathcal{B}}$, are intervals, and W_m is uniformly divided into N_m intervals. Therefore, (iv) in Theorem 1 is proved. According to the reverse Hölder's inequality, (d) is an equality

if and only if $\exists \tau > 0, \mu(W_m) = \tau \left(\beta + \frac{1}{N_m^2}\right)^{-\frac{1}{2}}, \forall m \in \mathcal{I}_{\mathcal{B}}$. Moreover, the sum of these measures is $\mu(\Omega)$, i.e., $\sum_{j=1}^{M} \mu(W_j) = \mu(\Omega)$. Therefore, the corresponding measure of the m^{th} cluster region is

$$\mu(W_m) = \frac{N_m \left(1 + \beta N_m^2\right)^{-\frac{1}{2}} \mu(\Omega)}{\sum_{j=1}^{M} N_j \left(1 + \beta N_j^2\right)^{-\frac{1}{2}}}. \tag{39}$$

Note that (e) is a function of $(N_1, \ldots, N_M)$ and (f) is just the minimum of (e). Therefore, the last inequality is an equality when we properly select the variables $N_1, \ldots, N_M$. Obviously, the above equality conditions are compatible, i.e., all equality conditions can be satisfied simultaneously. Therefore, (f) is an achievable lower bound, indicating the minimum distortion. The last thing is to determine the optimal $(N_1, \ldots, N_M)$ that attains the minimum of (e). By Lemma 4, (e) attains (f), if and only if M_a of the N_ms are equal to $\lceil \frac{N}{M} \rceil$ and M_b of the N_ms are equal to $\lfloor \frac{N}{M} \rfloor$. Substituting the optimal values for $N_m s$ to (e), we obtain the minimum distortion formula in (17). Substituting the optimal values for N_ms to (39), we get (i) and (ii) in Theorem 1.

Appendix D
Proof of Lemma 1

Since $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}})$ is the distortion of an N-level quantizer, we have $\mathcal{P}^{\mathcal{S}} \in [D_R(N), +\infty)$ which is then the domain of the function $A(S)$. Next, we justify the monotonicity of $A(S)$. Let $F(S)$ be the set of the feasible solutions for the point $(S, A(S))$. Thus, the AP-Sensor power function can be rewritten as $A(S) = \inf_{(P,Q,\mathbf{R}^{\mathcal{A}},T) \in F(S)} \mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T)$. For any two values S_1 and S_2 such that $D_R(N) \leq S_1 < S_2$, we have $F(S_1) \subseteq F(S_2)$ and then $A(S_1) \geq A(S_2)$. In other words, $A(S)$ is a non-increasing function. Note that $D_R(\cdot)$ is a mean-square-error distortion with continuous and differentiable source. Therefore, $D_R(\cdot)$ is a strictly decreasing function. Since $N > M$, we have $D_R(N) < D_R(M)$. Now, we discuss the values of $A(S)$ on $[D_R(M), +\infty)$. Let $(X^*, \mathbf{R}^*) = \arg\min_{(P,\mathbf{R})} \sum_{n=1}^{M} \int_{R_n} \|x - w\|^2 f(w) dw$ be the optimal solution for the M-level one-tier quantizer. $X^* = (x_1^*, \ldots, x_M^*)$ and $\mathbf{R}^* = (R_1^*, \ldots, R_M^*)$ are, respectively, the optimal reproduction points and quantization regions. Let $P' = (x_1^*, x_2^*, \ldots, x_M^*, x_1^*, \ldots, x_1^*)$ be a deployment including N points (APs), in which the last $N - M$ APs have the same location x_1^*. Afterwards, we define $\mathbf{R}'^{\mathcal{A}} = R^*$, $Q' = X^*$, and $T'(n) = T^E_{[P',Q']}(n) = \arg\min_m \|p'_n - q'_m\|$. Substituting $(P', Q', \mathbf{R}'^{\mathcal{A}}, T')$ to (4) and (3), we obtain $\mathcal{P}^{\mathcal{S}}(P', \mathbf{R}'^{\mathcal{A}}) = D_R(M)$ and $\mathcal{P}^{\mathcal{A}}(P', Q', \mathbf{R}'^{\mathcal{A}}, T') = 0$. In other words, $A(D_R(M)) \leq 0$. On one hand, since $A(S)$ is a non-negative non-increasing function, we have $A(S) = 0$ on $[D_R(M), +\infty)$. On the other hand, by quantization theorem, when N APs are placed at K distinct locations, where $K \leq M$, we have $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) \geq D_R(M)$. In other words, when $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) < D_R(M)$, N APs have at least $M + 1$ distinct locations. Under such circumstances, it is impossible to connect each AP to the FC with zero distance, indicating that $\mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) > 0$. Therefore, we have $A(S) > 0$ on $[D_R(N), D_R(M))$.

Appendix E
Proof of Lemma 4

At the beginning, we prove (i), the domain of $\widehat{A}(S, \mathbf{R}^{\mathcal{A}})$ shown in the lemma is correct. An input $(S, \mathbf{R}^{\mathcal{A}})$ belongs to the domain if and only if there exists a P such that $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) \leq S$. When $\mathbf{R}^{\mathcal{A}}$ is fixed, the range of $\mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}})$ is $[H(\mathbf{R}^{\mathcal{A}}), +\infty)$, where $H(\mathbf{R}^{\mathcal{A}})$ is the minimum distortion of a one-tier quantizer with partition $\mathbf{R}^{\mathcal{A}}$. Therefore, the domain of $\widehat{A}(S, \mathbf{R}^{\mathcal{A}})$ can be represented as $\{(S, \mathbf{R}^{\mathcal{A}}) | H(\mathbf{R}^{\mathcal{A}}) \leq S\}$. Next, we prove (ii), the value of $\widehat{A}(S, \mathbf{R}^{\mathcal{A}})$ on the domain shown in the lemma is correct. Note that one can simply achieve the minimum AP-power 0 with Sensor-power $D_R(1)$ by placing all APs at the centroid of the target area. Therefore, when $x \in [D_R(1), +\infty)$, $\widehat{A}(S, \mathbf{R}^{\mathcal{A}}) = 0$. In what follows, we focus on the case that $x \in [D_R(N), D_R(1))$. To calculate the value of $\widehat{A}(S)$ at $(S, \mathbf{R}^{\mathcal{A}})$, we assume that the AP cell partition is fixed as $\mathbf{R}^{\mathcal{A}}$. Therefore, the centroid, $c_n = \frac{\int_{R_n^{\mathcal{A}}} w f(w) dw}{\int_{R_n^{\mathcal{A}}} f(w) dw}$, and the volume of $R_n^{\mathcal{A}}$, $v_n = \int_{R_n^{\mathcal{A}}} f(w) dw$, are constants. Since $M = 1$, the index map $T(n) = 1, \forall n \in \mathcal{I}_{\mathcal{A}}$, is determined. Let q be the location of the unique FC. Therefore, the AP-power function becomes

$$\begin{aligned} \mathcal{P}^{\mathcal{A}}(P, Q, \mathbf{R}^{\mathcal{A}}, T) &= \sum_{n=1}^{N} \int_{R_n^{\mathcal{A}}} \|p_n - q\|^2 f(w) dw \\ &= \sum_{n=1}^{N} \|p_n - q\|^2 v_n = \sum_{n=1}^{N} \|p_n \sqrt{v_n} - q\sqrt{v_n}\|^2 = \|\widetilde{\mathbf{p}} - \widetilde{\mathbf{q}}\|^2, \end{aligned} \tag{40}$$

and the Sensor-power function becomes

$$\begin{aligned} \mathcal{P}^{\mathcal{S}}(P, \mathbf{R}^{\mathcal{A}}) &= \sum_{n=1}^{N} \int_{R_n^{\mathcal{A}}} \|p_n - w\|^2 f(w) dw = H(\mathbf{R}^{\mathcal{A}}) + \sum_{n=1}^{N} \|p_n - c_n\|^2 v_n \\ &= H(\mathbf{R}^{\mathcal{A}}) + \sum_{n=1}^{N} \|p_n \sqrt{v_n} - c_n \sqrt{v_n}\|^2 = H(\mathbf{R}^{\mathcal{A}}) + \|\widetilde{\mathbf{p}} - \widetilde{\mathbf{c}}\|^2, \end{aligned} \tag{41}$$

where $\widetilde{\mathbf{p}} \triangleq [p_1\sqrt{v_1}, \ldots, p_N\sqrt{v_N}]$, $\widetilde{\mathbf{q}} \triangleq [q\sqrt{v_1}, \ldots, q\sqrt{v_N}]$, and $\widetilde{\mathbf{c}} \triangleq [c_1\sqrt{v_1}, \ldots, c_N\sqrt{v_N}]$. Since $H(\mathbf{R}^{\mathcal{A}})$ is a constant, (20) can be rewritten as

$$\widehat{A}(S, \mathbf{R}^{\mathcal{A}}) = \inf_{(\widetilde{\mathbf{p}}, \widetilde{\mathbf{q}}) : \|\widetilde{\mathbf{p}} - \widetilde{\mathbf{c}}\|^2 \leq S - H(\mathbf{R}^{\mathcal{A}})} \|\widetilde{\mathbf{p}} - \widetilde{\mathbf{q}}\|^2. \tag{42}$$

When q is fixed, one should minimize the distance of $\widetilde{\mathbf{p}}$ to the fixed $\widetilde{\mathbf{q}}$ subject to the constraint that $\widetilde{\mathbf{p}}$ remain within a ball of radius-square $S - H(\mathbf{R}^{\mathcal{A}})$ centered at $\widetilde{\mathbf{c}}$. A simple geometric argument reveals that the optimal solution should then fall on a line between $\widetilde{\mathbf{p}}$ and $\widetilde{\mathbf{q}}$, i.e., $\widetilde{\mathbf{p}} = \frac{\widetilde{\mathbf{q}} + \lambda \widetilde{\mathbf{c}}}{1 + \lambda}$ for some $\lambda \geq 0$. Going back to the original variables, we have the optimal AP locations $p_n = \frac{q + \lambda c_n}{1 + \lambda}, n \in \mathcal{I}_{\mathcal{A}}$. Substituting the optimal AP locations to (41), the constraint in (42) becomes $(1 + \lambda)^2 \geq \frac{\sum_{n=1}^{N} \|q - c_n\|^2 v_n}{S - H(\mathbf{R}^{\mathcal{A}})}$. Specially, when $\lambda = 0$, we have $p_n = q, \forall n \in \mathcal{I}_{\mathcal{A}}$. In other words, all APs are placed at the same location which is equivalent to the scenario that only one AP is placed. Again, $D_R(1)$ is the minimum distortion for the one-tier quantizer and then the minimum sensor power when only one AP is placed. Thus, when $\lambda = 0$, we have $\mathcal{P}^{\mathcal{S}} \geq D_R(1)$. Since we only consider $S \in [D_R(N), D_R(1))$, to ensure the constraint $S \geq \mathcal{P}^{\mathcal{S}}$, λ cannot be 0. Thus, the possible range of λ is $(0, +\infty)$.

On the other hand, T is determined and the constraint in (20) is independent of $Q = \{q\}$ so that $\widehat{A}(S, \mathbf{R}^{\mathcal{A}})$ can be rewritten as

$$\begin{aligned}\widehat{A}(S,\mathbf{R}^{\mathcal{A}}) &= \inf_{(P,Q,T):\mathcal{P}^{\mathcal{S}}(P,\mathbf{R}^{\mathcal{A}})\le S} \mathcal{P}^{\mathcal{A}}(P,Q,\mathbf{R}^{\mathcal{A}},T)\\ &= \inf_{Q}\ \inf_{(P,T):\mathcal{P}^{\mathcal{S}}(P,\mathbf{R}^{\mathcal{A}})\le S} \mathcal{P}^{\mathcal{A}}(P,Q,\mathbf{R}^{\mathcal{A}},T)\\ &= \inf_{q}\ \inf_{\lambda:\lambda>0,(1+\lambda)^2\ge\frac{\sum_{n=1}^{N}\|q-c_n\|^2v_n}{S-H(\mathbf{R}^{\mathcal{A}})}} \sum_{n=1}^{N}\left\|\frac{q+\lambda c_n}{1+\lambda}-q\right\|^2 v_n \qquad (43)\\ &= \inf_{\lambda:\lambda>0,(1+\lambda)^2\ge\frac{\sum_{n=1}^{N}\|q-c_n\|^2v_n}{S-H(\mathbf{R}^{\mathcal{A}})}}\ \inf_{q} Z(q),\end{aligned}$$

where $Z(q)=\sum_{n=1}^{N}\|\frac{q+\lambda c_n}{1+\lambda}-q\|^2v_n$. In what follows, we attempt to find the optimal q to minimize $Z(q)$. Note that $Z(q)$ is a quadratic function and convex. Therefore, the unique minimum is associated with zero-gradient. Since $\frac{\partial Z(q)}{\partial q}=\frac{2\lambda^2}{(1+\lambda)^2}\left(q-\frac{\sum_{n=1}^{N}c_nv_n}{\sum_{n=1}^{N}v_n}\right)$, we obtain the optimal FC location $q=\frac{\sum_{n=1}^{N}c_nv_n}{\sum_{n=1}^{N}v_n}=\frac{\sum_{n=1}^{N}\int_{\Omega}wf(w)dw}{\sum_{n=1}^{N}v_n}=c$, where c is the centroid of the target region Ω. Substituting the optimal q to (43), we get $\widehat{A}(S,\mathbf{R}^{\mathcal{A}})=\inf_{\lambda:\lambda>0,(1+\lambda)^2\ge\frac{J(\mathbf{R}^{\mathcal{A}})}{x-H(\mathbf{R}^{\mathcal{A}})}}\frac{\lambda^2}{(1+\lambda)^2}J(\mathbf{R}^{\mathcal{A}})$, where $J(\mathbf{R}^{\mathcal{A}})=\sum_{n=1}^{N}\|c-c_n\|^2v_n$. According to the parallel axis theorem, we have

$$\begin{aligned}D_R(1)&=\sum_{n=1}^{N}\int_{R_n^{\mathcal{A}}}\|c-w\|^2f(w)dw=\sum_{n=1}^{N}\|c-c_n\|^2v_n\\&+\sum_{n=1}^{N}\int_{R_n^{\mathcal{A}}}\|w-c_n\|^2f(w)dw=J(\mathbf{R}^{\mathcal{A}})+H(\mathbf{R}^{\mathcal{A}}).\end{aligned}\qquad (44)$$

Therefore, $\widehat{A}(S,\mathbf{R}^{\mathcal{A}})$ can be rewritten as $\widehat{A}(S,\mathbf{R}^{\mathcal{A}})=\inf_{\lambda:\lambda>0,(1+\lambda)^2\ge\frac{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}{S-H(\mathbf{R}^{\mathcal{A}})}}\frac{\lambda^2}{(1+\lambda)^2}\left(D_R(1)-H(\mathbf{R}^{\mathcal{A}})\right)$. Solving the inequality $(1+\lambda)^2\ge\frac{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}{S-H(\mathbf{R}^{\mathcal{A}})}$ results in $\lambda\ge\sqrt{\frac{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}{S-H(\mathbf{R}^{\mathcal{A}})}}-1$ or $\lambda\le-\sqrt{\frac{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}{S-H(\mathbf{R}^{\mathcal{A}})}}-1$. Since $\lambda>0$, we have only one solution, i.e., $\lambda\ge\sqrt{\frac{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}{S-H(\mathbf{R}^{\mathcal{A}})}}-1$. Since $\frac{\lambda^2}{(1+\lambda)^2}$ is monotonically increasing, $\widehat{A}$ attains the minimum at $\lambda=\sqrt{\frac{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}{S-H(\mathbf{R}^{\mathcal{A}})}}-1$. Substituting $\lambda=\sqrt{\frac{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}{S-H(\mathbf{R}^{\mathcal{A}})}}-1$ to $\widehat{A}(S,\mathbf{R}^{\mathcal{A}})$, we prove (21).

Appendix F
Proof of Theorem 2

By Lemma 1, we know $A(S)=0$ on $[D_R(M),+\infty)$ and the domain of $A(S)$ is $[D_R(N),+\infty)$. In what follows, we study $A(S)$ on $[D_R(N),D_R(M))$. Let $\widehat{A}(S,\mathbf{R}^{\mathcal{A}})$ be the minimum AP-power with the Sensor-power constraint $\mathcal{P}^{\mathcal{S}}\le S$ for the fixed AP cell partition $\mathbf{R}^{\mathcal{A}}$ when $M=1$. Therefore, when $M=1$, minimizing the AP-power with the Sensor-power constraint $\mathcal{P}^{\mathcal{S}}(P,\mathbf{R}^{\mathcal{A}})\le S$ is equivalent to minimizing $\widehat{A}(S,\mathbf{R}^{\mathcal{A}})$ on its domain. According to Lemma 2, the domain of $\widehat{A}(S,\mathbf{R}^{\mathcal{A}})$ is $\{(S,\mathbf{R}^{\mathcal{A}})|H(\mathbf{R}^{\mathcal{A}})\le S\}$. Therefore, the AP-Sensor power function can be rewritten as

$$\begin{aligned}A(S)&=\inf_{(P,Q,\mathbf{R}^{\mathcal{A}},T):\mathcal{P}^{\mathcal{S}}(P,\mathbf{R}^{\mathcal{A}})\le S}\mathcal{P}^{\mathcal{A}}(P,Q,\mathbf{R}^{\mathcal{A}},T)\\&=\inf_{\mathbf{R}^{\mathcal{A}}:H(\mathbf{R}^{\mathcal{A}})\le S}\ \inf_{(P,Q,T):\mathcal{P}^{\mathcal{S}}(P,\mathbf{R}^{\mathcal{A}})\le S}\mathcal{P}^{\mathcal{A}}(P,Q,\mathbf{R}^{\mathcal{A}},T)\\&=\inf_{\mathbf{R}^{\mathcal{A}}:H(\mathbf{R}^{\mathcal{A}})\le S}\widehat{A}(S,\mathbf{R}^{\mathcal{A}})\\&=\inf_{\mathbf{R}^{\mathcal{A}}:H(\mathbf{R}^{\mathcal{A}})\le S}\left[\sqrt{(S-H(\mathbf{R}^{\mathcal{A}}))}-\sqrt{D_R(1)-H(\mathbf{R}^{\mathcal{A}})}\right]^2,\end{aligned}\qquad (45)$$

Note that $\mathbf{R}^{\mathcal{A}}$ affects $A(S)$ merely via $H(\mathbf{R}^{\mathcal{A}})$. Moreover, $H(\mathbf{R}^{\mathcal{A}})$ is the minimum distortion of the N-level quantizer with partition $\mathbf{R}^{\mathcal{A}}$, and $D_R(N)$ is the minimum distortion of the N-level quantizer over all possible partitions. Thus, $H(\mathbf{R}^{\mathcal{A}})\ge D_R(N)$. Given the value of S, to satisfy the inequality constraint $H(\mathbf{R}^{\mathcal{A}})\le S$, the range of $H(\mathbf{R}^{\mathcal{A}})$ is then $[D_R(N),S]$. Therefore, the AP-Sensor power function can be rewritten as

$$A(S)=\inf_{y:D_R(N)\le y\le S}\widetilde{A}(S,y),\qquad (46)$$

where $\widetilde{A}(S,y)=\left[\sqrt{(S-y)}-\sqrt{(D_R(1)-y)}\right]^2$. When $S\ne y$, we have

$$\frac{\partial\widetilde{A}(S,y)}{\partial y}=-2+\frac{S+D_R(1)-2y}{\sqrt{(S-y)(D_R(1)-y)}}.\qquad (47)$$

Extending $(S+D_R(1)-2y)^2$, we get

$$\begin{aligned}(S+D_R(1)-2y)^2&=((S-y)+(D_R(1)-y))^2\\&=(S-y)^2+(D_R(1)-y)^2+2(S-y)(D_R(1)-y)\\&>4(S-y)(D_R(1)-y)=\left[2\sqrt{(S-y)(D_R(1)-y)}\right]^2.\end{aligned}\qquad (48)$$

Again, we only consider the region $[D_R(N),D_R(1))$, indicating $D_R(N)\le S<D_R(1)$. Taking the constraint $D_R(N)\le y\le S$ into consideration, we have $y\le S<D_R(1)$. When $y\ne S$, the term $\frac{S+D_R(1)-2y}{\sqrt{(S-y)(D_R(1)-y)}}$ is positive. Therefore, we have $\frac{\partial\widetilde{A}(S,y)}{\partial y}>0,\forall S\in[D_R(N),D_R(1))$ and $S\ne y$. Therefore, when $y\in[D_R(N),S)$, $\widetilde{A}(S,y)$ is increasing and has its unique minimum at $y=D_R(N)$. Taking the case $S=y$ into account, we have $A(S)=\min(\widetilde{A}(S,D_R(N)),\widetilde{A}(S,S))$. After straightforward calculation, we get $\widetilde{A}(S,D_R(N))<\widetilde{A}(S,S)$ when $x\in[D_R(N),\hat{D}_R(1))$. Therefore, in order to minimize $\widetilde{A}(S,y)$, y should be minimized to $D_R(N)$. In other words, the AP-Sensor power function on $[D_R(N),D_R(1))$ can be rewritten as (22), which is a convex function. In sum, the AP-Sensor power function is convex on the domain $[D_R(N),+\infty)$.

References

[1] J. Guo, E. Koyuncu, and H. Jafarkhani, "Energy efficiency in two-tiered wireless sensor networks," *IEEE Int. Conf. Commun.*, May 2017.

[2] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd ed. New York: John Wiley & Sons, 2000.

[3] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[4] J. Cortes, S. Martinez and F. Bullo, "Spatially-distributed coverage optimization and control with limited-range interactions," *Mathematical Modelling and Numerical Analysis: Control, Optimisation and Calculus of Variations*, vol. 11, pp. 691-719, Oct. 2005.

[5] J. Guo and H. Jafarkhani, "Sensor deployment with limited communication range in homogeneous and heterogeneous wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6771-6884, Oct. 2016.

[6] G. Wang, G. Cao, and T. L. Porta, "Movement-assisted sensor deployment," *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 640–652, Jun. 2006.

[7] O. Arslan and D. E. Koditschek, "Voronoi-based coverage control of heterogeneous disk-shaped robots" *IEEE Int. Conf. Robot. Autom.*, May. 2016.

[8] N. Aitsaadi, N. Achir, K. Bousseta, and B. Gavish, "A gradient approach for differentiated wireless sensor network deployment", *1st IFIP Wireless Days*, Nov. 2008.

[9] C. C. Coskun and E. Ayanoglu, "Energy-efficient base station deployment in heterogeneous networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 593-596, Dec. 2014.

[10] T. Zhang, J. Zhao, L. An, and D. Liu, "Energy efficiency of base station deployment in ultra dense hetNets: A stochastic geometry analysis," *IEEE Wireless Commun. Lett.*, vol. 5, no. 2, pp. 184-187, Apr. 2016.

[11] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996-1019, Jul. 2013.

[12] E. Koyuncu, R. Khodabakhsh, N. Surya, and H. Seferoglu, "Deployment and trajectory optimization for UAVs: A quantization theory approach," *IEEE Wireless Commun. Networking Conf.*, Apr. 2018.

[13] H. Jafarkhani, *Space-Time Coding: Theory and Practice*, Cambridge, United Kingdom: *Cambridge Univ. Press*, 2005.

[14] K. Akkaya, M. Younis, "A survey on routing protocols in wireless sensor networks," *Ad Hoc Network (Elsevier)*, vol. 3, no. 3, pp. 325-349, 2005.

[15] L. Li and J. Y. Halpern, "Minimum-energy mobile wireless networks revisited," *IEEE Int. Conf. Commun.*, Jun. 2001.

[16] R. Stock, "Distance and the utilization of health facilities in rural Nigeria," *Social Sci. & Med.*, vol. 17, no. 9, pp. 563-570, 1983.

[17] D. R. Ingram, D. R. Clarke, and R. A. Murdie, "Distance and the decision to visit an emergency department," *Social Sci. & Med.*, vol. no. 1, pp. 55-62, Mar. 1978.

[18] Z. Drezner, and H. W. Hamacher, "Facility location: Applications and theory," *Operations Research & Decision Theory*, New York: Springer, 2001.

[19] J. G. Carlsson, E. Carlsson, and R. Devulapalli, "Balancing workloads of service vehicles over a geographic territory," *IEEE Int. Conf. Intell. Robots and Syst.*, Nov. 2013.

[20] J. G. Carlsson, E. Carlsson, and R. Devulapalli, "Shadow prices in territory division," *Netw. Spat. Econ.*, vol. 16, no. 3, pp. 893-931, 2016.

[21] J. P. Rodrigue, C. Comtois, and B. Slack, "The geography of transport systems," in *Social Science*, London, United Kingdom: Taylor & Francis Ltd, 2009.

[22] Y. T. Hou, Y. Shi, H. D. Sherali, and F. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 2579-2590, Sept. 2005.

[23] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," *IEEE Int. Conf. on Comput. Commun.*, Mar. 2000.

[24] M. Gani, "Optimal deployment control for a heterogeneous mobile sensor network," *9th Int. Conf. Control, Autom., Robot. and Vision*, Dec. 2006.

[25] W. Pawgasame, "A survey in adaptive hybrid wireless sensor network for military operations," *IEEE 2th Asian Conf. Defence Technol.*, Jan. 2016.

[26] W. R. Bennett, "Spectra of quantized signals," *The Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 446-472, Jul. 1948.

[27] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inf. Theory*, vol. 25, no. 4, pp. 373-380, Jul. 1979.

[28] P. L. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 139-148, Mar. 1982.

[29] E. Koyuncu and H. Jafarkhani, "On the minimum average distortion of quantizers with index-dependent distortion measures," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4655-4669, Sept. 2017.

[30] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, San Francisco, CA: *Morgan Kaufmann*, 2011.

[31] H. Jafarkhani and V. Tarokh, "Design of Successively Refinable Trellis Coded Quantizers," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1490-1497, Jul. 1999.

[32] H. Jafarkhani, H. Brunk, and N. Farvardin, "Entropy-Constrained Successively Refinable Scalar Quantization," *IEEE Data Compression Conf.*, Mar. 1997.

[33] D. Mukherjee and S. K. Mitra, "Successive refinement lattice vector quantization," *Trans. Image Process.*, vol. 11, no. 12, pp. 1337-1348, Dec. 2002.

[34] N. Chaddha, P. A. Chou, and R. M. Gray, "Constrained and recursive hierarchical table-lookup vector quantization," *IEEE Data Compression Conf.*, Apr. 1996.

[35] A. Gersho and Y. Shoham, "Hierarchical vector quantization of speech with dynamic codebook allocation," *IEEE Int. Conf. on Acoust., Speech, and Signal Process.*, Mar. 1984.

[36] H. Jafarkhani and N. Farvardin, "Channel-Matched Hierarchical Table-Lookup Vector Quantization," *IEEE Trans. Inf. Theory*, May 2000.

[37] K. Jain and V.V. Vazirani, "Primal-dual approximation algorithms for metric facility location and k-median problems," *40th Annu. Symp. Found. Comput. Sci.*, Oct. 1999.

[38] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communication costs," *5th Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2006.

[39] A. Breitenmoser, M. Schwager, J. C. Metzger, R. Siegwart, and D. Rus, "Voronoi coverage of non-convex environments with a group of networked robots," *IEEE Int. Conf. Robot. Autom.*, May 2010.

[40] A. Gusrialdi and L. Zeng, "Distributed deployment algorithms for robotic visual sensor networks in non-Convex environment," *Int. Conf. on Netw., Sens. and Control*, Apr. 2011.

[41] H. Mahboubi, K. Moezzi, A. G. Aghdam, K. Sayrafian-Pour, and V. Marbukh, "Self-deployment algorithms for coverage problem in a network of mobile sensors with unidentical sensing ranges," *IEEE Global Telecom. Conf.*, Dec. 2010.

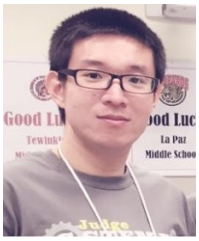
[42] C. Wu and Y. Chung, "Heterogeneous wireless sensor network deployment and topology control based on irregular sensor model," *2nd Int. Conf. Advances in Grid and Pervasive Comput.*, May 2007.

[43] Y. Yoon and Y.-H. Kim, "An efficient genetic algorithm for maximum coverage deployment in wireless sensor networks", *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1473-1483, Oct. 2013.

[44] M.I. Skolnik, *Radar Handbook*, New York: McGraw-Hill, 2008.

[45] D. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs, NJ: Prentice-Hall, 1992.

[46] J. H. Conway, and N. J. A. Sloane, Sphere packings, lattices and groups, New York: Springer Science & Business Media, 1993.



Jun Guo received the B.S. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2011 and the M.S. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2014. He is currently working towards the Ph.D. degree at University of California, Irvine.

His main research includes wireless sensor network, distributed computation, source coding and channel coding. Mr. Guo was the recipient of the Research Fellowship for the academic year 2014-2015.

Erdem Koyuncu is a Research Assistant Professor at the Department of Electrical and Computer Engineering of the University of Illinois at Chicago since 2016. He received the B.S. degree from the Department of Electrical and Electronics Engineering of Bilkent University in 2005. He received the M.S. and Ph.D. degrees in 2006 and 2010, respectively, both from the Department of Electrical Engineering and Computer Science of the University of California, Irvine (UCI). Between 2011 and 2016, he was a Postdoctoral Scholar at the Center for Pervasive Communications and Computing of UCI.

Hamid Jafarkhani is a Chancellor's Professor at the Department of Electrical Engineering and Computer Science, University of California, Irvine, where he is also the Director of Center for Pervasive Communications and Computing and the Conexant-Broadcom Endowed Chair. Among his awards are the IEEE Marconi Prize Paper Award in Wireless Communications, the IEEE Communications Society Award for Advances in Communication, and the IEEE Eric E. Sumner Award.

Dr. Jafarkhani is listed as a highly cited researcher in http://www.isihighlycited.com. According to the Thomson Scientific, he is one of the top 10 most-cited researchers in the field of "computer science" during 1997-2007. He is the 2017 Innovation Hall of Fame Inductee, School of Engineering, University of Maryland. He is a Fellow of AAAS, an IEEE Fellow, and the author of the book "Space-Time Coding: Theory and Practice."