

# Energy Scaling Laws for Distributed Inference in Random Fusion Networks

Animashree Anandkumar, *Student Member, IEEE*, Joseph E. Yukich, Lang Tong<sup>†</sup>, *Fellow, IEEE*, and Ananthram Swami, *Fellow, IEEE*

**Abstract**—The energy scaling laws of multihop data fusion networks for distributed inference are considered. The fusion network consists of randomly located sensors distributed i.i.d. according to a general spatial distribution in an expanding region. Among the class of data fusion schemes that enable optimal inference at the fusion center for Markov random field (MRF) hypotheses, the scheme with minimum average energy consumption is bounded below by average energy of fusion along the minimum spanning tree, and above by a suboptimal scheme, referred to as Data Fusion for Markov Random Fields (DFMRF). Scaling laws are derived for the optimal and suboptimal fusion policies. It is shown that the average asymptotic energy of the DFMRF scheme is finite for a class of MRF models.

**Index Terms**—Distributed detection, graphical models, random graphs, stochastic geometry and data fusion.

## I. INTRODUCTION

WE consider the problem of distributed statistical inference in a network of randomly located sensors, each taking a measurement and transporting the locally processed data to a designated fusion center. The fusion center then makes an inference about the underlying phenomenon based on the data collected from all the sensors.

For statistical inference using wireless sensor networks, energy consumption is an important design parameter. The transmission power required to reach a receiver distance  $d$  away with a certain signal-to-noise ratio (SNR) scales in the order of  $d^\nu$ , where  $2 \leq \nu \leq 6$  is the path loss [3]. Therefore, the cost of moving data from sensor locations to the fusion center, either through direct transmissions or multihop forwarding, significantly affects the lifetime of the network.

### A. Scalable data fusion

We investigate the cost of data fusion for inference, and its scaling behavior with the size of the network and the

area of deployment. In particular, for a network of  $n$  random sensors located at points  $\mathbf{V}_n = \{V_1, \dots, V_n\}$  in  $\mathbb{R}^2$ , a fusion policy  $\pi_n$  maps  $\mathbf{V}_n$  to a set of scheduled transmissions and computations. The average cost (e.g., energy) of a policy is given by

$$\bar{\mathcal{E}}(\pi_n(\mathbf{V}_n)) := \frac{1}{n} \sum_{i \in \mathbf{V}_n} \mathcal{E}_i(\pi_n(\mathbf{V}_n)), \quad (1)$$

where  $\mathcal{E}_i(\pi_n(\mathbf{V}_n))$  is the cost at node  $i$  under policy  $\pi_n$ . The above average cost is random, and we are interested in its scalability in random networks as  $n \rightarrow \infty$ .

*Definition 1 (Scalable Policy):* A sequence of policies  $\pi := (\pi_n)_{n \geq 1}$  is scalable on average if

$$\lim_{n \rightarrow \infty} \mathbb{E}(\bar{\mathcal{E}}(\pi_n(\mathbf{V}_n))) = \bar{\mathcal{E}}_\infty(\pi) < \infty$$

where  $\bar{\mathcal{E}}_\infty(\pi)$  is referred to as the *scaling constant*. A sequence of policies  $\pi_n$  is *weakly scalable* if

$$\text{p} \lim_{n \rightarrow \infty} \bar{\mathcal{E}}(\pi(\mathbf{V}_n)) = \bar{\mathcal{E}}_\infty(\pi) < \infty,$$

where p lim denotes convergence in probability. It is *strongly scalable* if the above average energy converges almost surely and is  $L^2$  (*mean squared*) *scalable* if the convergence is in mean square.

We focus mostly on the  $L^2$  scalability of the fusion policies, which implies weak and average scalability. We are interested in scalable data fusion policies that enable optimal statistical inference at the fusion center implying finite average energy expenditure as the network size increases.

To motivate this study, first consider two simple fusion policies: the direct transmission policy (DT) in which all sensors transmit directly to the fusion center (single hop), and the shortest path (SP) policy, where each node forwards its data to the fusion center using the shortest path route without any data combination at the intermediate nodes.

We assume, for now, that  $n$  sensor nodes are uniformly distributed in a square having area  $n$ . It is perhaps not surprising that neither of the two policies is scalable as  $n \rightarrow \infty$ . For the DT policy, intuitively, the average transmission from the sensors to the fusion center range scales as  $\sqrt{n}$ , thus  $\bar{\mathcal{E}}(\text{DT}(\mathbf{V}_n))$  scales as  $n^{\frac{\nu}{2}}$ . On the other hand, we expect the SP policy to have better scaling since it chooses the best multihop path to forward data from each node to the fusion center. However, even in this case, there is no finite scaling. Here, the average number of hops scales in the order of  $\sqrt{n}$ , and thus,  $\bar{\mathcal{E}}(\text{SP}(\mathbf{V}_n))$  scales in the order of  $\sqrt{n}$ . Rigorously establishing

<sup>†</sup>Corresponding author.

A. Anandkumar and L. Tong are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA. Email: {aa332@, ltong@ece.}cornell.edu

J.E. Yukich is with the Department of Mathematics, Lehigh University, Bethlehem, Pa. 18015. E-mail: joseph.yukich@lehigh.edu.

A. Swami is with the Army Research Laboratory, Adelphi, MD 20783 USA E-mail: a.swami@ieee.org.

This work was supported in part through collaborative participation in Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011 and by the Army Research Office under Grant ARO-W911NF-06-1-0346. The first author is supported by the IBM Ph.D Fellowship for the year 2008-09 and is currently a visiting student at MIT, Cambridge, MA 02139. The second author was partially supported by NSA grant H98230-06-1-0052 and NSF grant DMS-0805570. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

Parts of this paper are presented at [1], [2]

the scaling laws for these two non-scalable policies is not crucial at this point since the same scaling laws can be easily established for regular networks when sensor nodes are on two-dimensional lattice points. See [4].

Are there scalable policies for data fusion? Among all the fusion policies not performing data combination at the intermediate nodes, the shortest path (SP) policy minimizes the total energy. Thus, no scalable policy exists unless nodes cooperatively combine their information, a process known as *data aggregation*. Data aggregation, however, must be considered in conjunction with the performance requirements of specific applications. In this paper, we assume that optimal inference is made at the fusion center, and this places a constraint on data aggregation. It rules out sub-sampling of the sensor field, which is dealt in [5].

### B. Summary of results and contributions

In this paper, we allow data aggregation at intermediate nodes, but require that the fusion center achieves the same inference performance *as if* all raw observations were collected without any data combination. We assume that the underlying hypotheses can be modeled as Markov random fields (MRF) and investigate the energy scaling laws.

Given sensor locations  $\mathbf{V}_n$  and possibly correlated sensor measurements, finding the minimum energy fusion policy under the constraint of optimal inference is, in general, NP-hard [6], and hence, intractable. We will establish upper and lower bounds on the fusion energy of this optimal scheme and analyze their scaling behavior. The lower bound is obtained by a scheme conducting fusion along the Euclidean minimum spanning tree (MST), which is shown to be optimal when the observations are statistically independent under both hypotheses. The upper bound on the optimal fusion scheme is established through a specific suboptimal fusion scheme, referred to as Data Fusion over Markov Random Fields (DFMRF). DFMRF becomes optimal for independent observations where it reduces to fusion along the MST. For certain spatial dependencies among sensor measurements of practical significance, such as the Euclidean 1-nearest neighbor graph, DFMRF has an approximation ratio 2, i.e., it costs no more than twice the cost of the optimal fusion scheme, independent of the size and configuration of the network.

We then proceed to establish a number of asymptotic properties of the DFMRF scheme in Section IV, including the scalability of DFMRF, its performance bounds, and the approximation ratio with respect to the optimal fusion policy when the sensor measurements have dependencies described by a  $k$ -nearest neighbor graph or a disk graph (continuum percolation). Applying techniques developed in [7]–[9], we provide a precise characterization of the scaling bounds as a function of sensor density and sensor placement distribution. These asymptotic bounds for DFMRF, in turn, imply that the optimal fusion scheme is also scalable. Hence, we use the DFMRF scheme as a vehicle to establish scaling laws for optimal fusion. Additionally, we use the energy scaling constants to optimize the distribution of the sensor placements. For independent measurements, we show that the uniform

distribution minimizes the average energy consumption over all i.i.d spatial placements when the path-loss coefficient of transmission is greater than two ( $\nu > 2$ ). For  $\nu \in [0, 2)$ , we show that the uniform distribution is, in fact, the most expensive<sup>1</sup> node configuration in terms of routing costs. We further show the optimality of the uniform distribution applies for both the lower and upper bounds on the average energy consumption for correlated measurements with  $k$ -nearest neighbor dependency graph or disk dependency graph under certain conditions.

To the best of our knowledge, our results are the first to establish the scalability of data fusion for certain correlation structures of the sensor measurements. The use of energy scaling laws for the design of efficient sensor placement is new and has direct engineering implications. The heuristic policy DFMRF first appeared in [10], and is made precise here with detailed asymptotic analysis using the weak law of large numbers for stabilizing graph functionals. One should not expect that scalable data fusion is always possible, and at the end of Section IV, we will discuss examples of correlation structures where scalable data fusion does not exist.

### C. Prior and related work

The seminal work of Gupta and Kumar [11] on the capacity of wireless networks has stimulated extensive studies covering a broad range of networking problems with different performance metrics. See also [12]. Here, we restrict ourselves to related works on energy consumption and data fusion for statistical inference.

Results on scaling laws for energy consumption are limited. In [13], energy scaling laws for multihop wireless networks (without any data fusion) are derived under different routing strategies. The issue of node placement for desirable energy scaling has been considered in [14], [15], where it is argued that uniform node placement, routinely considered in the literature, has poor energy performance. It is interesting to note that, for fusion networks, uniform sensor distribution is in fact optimal among a general class of distributions. See Section IV-B.

Energy-efficient data fusion has received a great deal of attention over the past decade. See a few recent surveys in [16], [17]. It has been recognized that sensor observations tend to be correlated, and that correlations should be exploited through data fusion. One line of approach is the use of distributed compression with the aim of routing all the measurements to the fusion center. Examples of such approaches can be found in [18]–[20].

While sending data from all sensors to the fusion center certainly ensures optimal inference, it is not necessary for statistical inference. More relevant to our work is the idea of data aggregation, e.g., [21]–[23]. Finding aggregation policies for correlated data, however, is nontrivial; it depends on the specific applications for which the sensor network is designed. Perhaps a more precise notion of aggregation is in-network function computation where certain functions are computed by

<sup>1</sup>The path-loss coefficient for wireless transmissions, in general, satisfies  $\nu > 2$ .

passing intermediate values among nodes [24]–[27]. However, these works are mostly concerned with computing symmetric functions such as the sum function.

In the context of statistical inference using wireless sensor networks, the idea of aggregation and in-network processing has been explored by several authors. See, e.g., [28]–[34]. Most relevant to our work are [28]–[32] where the Markovian correlation structures of sensor measurements are exploited explicitly. These results, however, do not deal with randomly placed sensors, and energy scaling laws are not established.

The results presented in this paper extend some of our earlier work in the direction of scaling-law analysis in random fusion networks. In [6], [10], [35], for fixed network size and node placement, we analyzed the minimum energy fusion scheme for optimal inference and showed that it reduces to a Steiner tree under certain constraints. We also proposed a heuristic called the DFMRF<sup>2</sup>. In [36], we analyzed the optimal sensor density for uniform node placement which maximizes the inference error exponent under an average energy constraint, and in [37], [48], we derived the error exponent for MRF hypotheses. In [5], we analyzed optimal sensor selection (i.e., sub-sampling) policies for achieving tradeoff between fusion costs and inference performance.

The energy scaling laws derived in this paper rely heavily on several results on the law of large numbers on geometric random graphs. We have extensively borrowed the formulations and techniques of Penrose and Yukich [9], [38]. See Appendix A for a brief description and [7], [8], [39] for detailed expositions of these ideas.

## II. SYSTEM MODEL

In this paper, we will consider various graphs. Chief among these are (i) *dependency* graphs specifying the correlation structure of sensor measurements, (ii) *network* graphs denoting feasible links for communication, and (iii) *fusion* digraphs denoting the (directed) links used by a policy to route and aggregate data.

### A. Stochastic model of sensor locations

We assume that  $n$  sensor nodes (including the fusion center) are placed randomly with sensor  $i$  located at  $V_i \in \mathbb{R}^2$ . By convention, the location of the fusion center is denoted by  $V_1$ . We denote the set of locations of the  $n$  sensors by  $\mathbf{V}_n := \{V_1, \dots, V_n\}$ . For our scaling law analysis, we consider a sequence of sensor populations on expanding square regions  $Q_{\frac{n}{\lambda}}$  of area  $\frac{n}{\lambda}$  and centered at the origin, where we fix  $\lambda$  as the overall sensor density and let the number of sensors  $n \rightarrow \infty$ .

To generate sensor locations  $V_i$ , first let  $Q_1 := [-\frac{1}{2}, \frac{1}{2}]^2$  be the unit area square<sup>3</sup>, and  $X_i \stackrel{i.i.d.}{\sim} \kappa, 1 \leq i \leq n$ , be a set of  $n$  independent and identically distributed (i.i.d.) random variables distributed on support  $Q_1$  according to  $\kappa$ . Here,  $\kappa$  is a probability density function (pdf) on  $Q_1$  which is bounded away from zero and infinity. We then generate  $V_i$  by scaling  $X_i$  accordingly:  $V_i = \sqrt{\frac{n}{\lambda}} X_i \in Q_{\frac{n}{\lambda}}$ . A useful special case is

the uniform distribution ( $\kappa \equiv 1$ ). Let  $\mathcal{P}_\lambda$  be the homogeneous Poisson distribution on  $\mathbb{R}^2$  with density  $\lambda$ .

### B. Graphical inference model: dependency graphs

We consider the inference problem of simple binary hypothesis testing,  $\mathcal{H}_0$  vs.  $\mathcal{H}_1$ , on a pair of Markov random fields (MRF). Under regularity conditions [40], a MRF is defined by its (undirected) dependency graph  $\mathcal{G}$  and an associated pdf  $f(\cdot | \mathcal{G})$ .

Under hypothesis  $\mathcal{H}_k$ , we assume that the dependency graph  $\mathcal{G}_k := (\mathbf{V}_n, E_k)$  models the correlation among the sensor observations, where  $\mathbf{V}_n = \{V_1, \dots, V_n\}$  is the set of vertices corresponding to sensor locations, generated according to the stochastic model in Sec II-A. Note that the vertex sets under the two hypotheses are identical. Set  $E_k$  is the set of edges of the dependency graph  $\mathcal{G}_k$ , and it defines the correlations of the sensor observations, as described in the next section.

We restrict our attention to proximity-based dependency graphs. In particular, we consider two classes of dependency graphs<sup>4</sup>: the (undirected)  $k$ -nearest neighbor graph ( $k$ -NNG) and the disk graph, also known as continuum percolation. We expect that our results extend to other locally-defined dependency structures such as the Delaunay, Voronoi, the minimum spanning tree, the sphere of influence and the Gabriel graphs. An important property of the aforementioned graphs is a certain stabilization property (discussed in Appendix A) facilitating asymptotic scaling analysis.

### C. Graphical inference model: likelihood functions

We denote the (random) measurements from all the sensors in a vertex set  $\mathbf{V}$  by  $\mathbf{Y}_{\mathbf{V}}$ , and  $\mathbf{Y}_{\mathbf{U}}$  denotes the vector that contains observations on a vertex subset  $\mathbf{U} \subset \mathbf{V}$ . The inference problem can now be stated as the following hypothesis test:

$$\mathcal{H}_0 : \mathbf{Y}_{\mathbf{V}} \sim f(\mathbf{y} | \mathcal{G}_0, \mathcal{H}_0) \text{ vs. } \mathcal{H}_1 : \mathbf{Y}_{\mathbf{V}} \sim f(\mathbf{y} | \mathcal{G}_1, \mathcal{H}_1), \quad (2)$$

where  $f(\mathbf{y} | \mathcal{G}_k, \mathcal{H}_k)$  is the pdf of  $\mathbf{Y}_{\mathbf{V}}$  conditioned on the dependency graph  $\mathcal{G}_k$  under hypothesis  $\mathcal{H}_k$ . Note that the sensor locations  $\mathbf{V}_n$  have the same distribution under either hypothesis. Therefore, only the conditional distribution of  $\mathbf{Y}_{\mathbf{V}}$  under each hypothesis is relevant for inference.

Under each hypothesis, the dependency graph involves conditional-independence relations between the measurements [40]

$$Y_i \perp\!\!\!\perp \mathbf{Y}_{\mathbf{V} \setminus \mathcal{N}(i; \mathcal{G}_k)} | \{\mathbf{Y}_{\mathcal{N}(i; \mathcal{G}_k)}, \mathbf{V}\}, \quad \text{under } \mathcal{H}_k, \quad (3)$$

where  $\mathcal{N}(i; \mathcal{G}_k)$  is the set of neighbors of  $i$  in  $\mathcal{G}_k$ , and  $\perp\!\!\!\perp$  denotes conditional independence. In words, given the node locations and the measurements at neighbors of a node in the dependency graph, the measurement at a node is conditionally independent of the rest of the network.

<sup>4</sup>The  $k$ -nearest neighbor graph ( $k$ -NNG) has edges  $(i, j)$  if  $i$  is one of the  $k$  nearest neighbors of  $j$  or viceversa, and ties are arbitrarily broken. The disk graph has edges between any two points within a certain specified distance (radius).

<sup>2</sup>The DFMRF scheme is referred to as AggMST in [6], [35].

<sup>3</sup>The results in this paper hold for  $\kappa$  defined on any convex unit area.

The celebrated Hammersley-Clifford theorem states that, under the positivity condition [41], the log-likelihood function of a MRF with dependency graph  $\mathcal{G}_k$  can be expressed as

$$-\log f(\mathbf{y}_{\mathbf{V}} | \mathcal{G}_k, \mathcal{H}_k) = \sum_{c \in \mathcal{C}_k} \psi_{k,c}(\mathbf{y}_c), \quad k = 0, 1, \quad (4)$$

where  $\mathcal{C}_k$  is a collection of (maximal) cliques in  $\mathcal{G}_k$ , the functions  $\psi_{k,c}$ , known as *clique potentials*, are real valued, non-negative and not zero everywhere on the support of distribution of  $\mathbf{y}_c$ . We assume that the normalization constant is already incorporated in the potential functions to ensure that (4) indeed describes a pdf. In general, it is NP-hard to evaluate the normalization constant given arbitrary potential functions [42], but can be carried out at the fusion center without any need for communication of sensor measurements.

#### D. Graphical fusion model and energy consumption

The set of feasible communications links form the (directed) *network graph* denoted by  $\mathcal{N}_g(\mathbf{V})$ . We assume that it is connected but not necessarily fully connected, and that it contains the Euclidean minimum spanning tree over the node set  $\mathbf{V}_n$  and directed towards the fusion center  $V_1$ , denoted by DMST( $\mathbf{V}_n; V_1$ ). Transmissions are scheduled so as to not interfere with one other. Nodes are capable of adjusting their transmission power depending on the location of the receiver.

A fusion policy  $\pi$  consists of a transmission schedule with the transmitter-receiver pairs, the time of transmission, and the aggregation algorithm that allows a node to combine its own and received values to produce a new communicating value. We model a fusion policy  $\pi$  by a directed *fusion graph*,  $\mathcal{F}_\pi := (\mathbf{V}, \vec{E}_\pi)$ , where  $\mathbf{V}$  is the same set of vertices corresponding to sensor locations, and  $\vec{E}_\pi$  contains *directed links*. A directed<sup>5</sup> link  $\langle i, j \rangle$  denotes a direct transmission from  $i$  to  $j$  and is required to be contained in the network graph  $\mathcal{N}_g(\mathbf{V})$  for the transmissions to be feasible. If one node communicates with another node  $k$  times,  $k$  direct links will be added between these two nodes in the edge set  $\vec{E}_\pi$  of the fusion policy  $\pi$ . Since we are only interested in characterizing the overall energy expenditure, the order of transmissions is not important; we only need to consider the associated cost with each link in  $\vec{E}_\pi$  and calculate the sum cost for  $\pi$ .

Nodes communicate in the form of packets. Each packet contains bits for at most one (quantized) real variable and other overhead bits independent of the network size. We assume that all real variables<sup>6</sup> are quantized to  $K$  bits, and  $K$  is independent of network size and is sufficiently large that quantization errors can be ignored. Thus, for node  $i$  to transmit data to node  $j$  distance  $|i, j|$  away, we assume that node  $i$  spends energy<sup>7</sup>  $\gamma|i, j|^\nu$ . Without loss of generality, we assume  $\gamma = 1$ . Hence, given a fusion policy  $\mathcal{F}_\pi = (\mathbf{V}, \vec{E}_\pi)$  of network

<sup>5</sup>We denote a directed link by  $\langle i, j \rangle$  and an undirected link by  $(i, j)$ .

<sup>6</sup>In principle, the raw and aggregated data may require different amount of energy for communication, and this can be easily incorporated into our framework.

<sup>7</sup>Since nodes only communicate a finite number of bits, we use energy instead of power as the cost measure.

size  $n$ , the average energy consumption is given by

$$\bar{\mathcal{E}}(\pi(\mathbf{V})) = \frac{1}{n} \mathcal{E}(\pi(\mathbf{V})) = \frac{1}{n} \sum_{\langle i, j \rangle \in \vec{E}_\pi} |i, j|^\nu, \quad 2 \leq \nu \leq 6. \quad (5)$$

The model specification is now complete.

### III. MINIMUM ENERGY DATA FUSION

In this section, we present data fusion policies aimed at minimizing energy expenditure under the constraint of optimal inference at the fusion center. The scalability of these policies is deferred to Section IV.

#### A. Optimal data fusion: a reformulation

The inference problem, defined in (2), involves two different graphical models, each with its own dependency graph and associated likelihood function. They do share the same vertex set  $\mathbf{V}$  which allows us to join the two graphical models into one.

Define the joint dependency graph  $\mathcal{G} := (\mathbf{V}, E)$ , where  $E := E_0 \cup E_1$ , as the union of the two dependency graphs  $\mathcal{G}_0$  and  $\mathcal{G}_1$ . The minimal sufficient statistic<sup>8</sup> is given by the log-likelihood ratio (LLR) [43]. With the substitution of (4), it is given by

$$\begin{aligned} L_{\mathcal{G}}(\mathbf{Y}_{\mathbf{V}}) &:= \log \frac{f(\mathbf{Y}_{\mathbf{V}} | \mathcal{G}_0, \mathcal{H}_0)}{f(\mathbf{Y}_{\mathbf{V}} | \mathcal{G}_1, \mathcal{H}_1)} \\ &= \sum_{a \in \mathcal{C}_1} \psi_{1,a}(\mathbf{Y}_a) - \sum_{b \in \mathcal{C}_0} \psi_{0,b}(\mathbf{Y}_b) \\ &:= \sum_{c \in \mathcal{C}} \phi_c(\mathbf{Y}_c), \quad \mathcal{C} := \mathcal{C}_0 \cup \mathcal{C}_1, \end{aligned} \quad (6)$$

where the effective potential functions  $\phi_c$  are given by

$$\phi_c(\mathbf{Y}_c) := \sum_{a \in \mathcal{C}_1, a \subset c} \psi_{1,a}(\mathbf{Y}_a) - \sum_{b \in \mathcal{C}_0, b \subset c} \psi_{0,b}(\mathbf{Y}_b), \quad \forall c \in \mathcal{C}. \quad (7)$$

Hereafter, we will work with  $(\mathcal{G}, L_{\mathcal{G}}(\mathbf{Y}_{\mathbf{V}}))$ . Note that the LLR is minimally sufficient (i.e., maximum dimensionality reduction) implying maximum possible savings in routing costs under the constraint of optimal inference.

Given the node set  $\mathbf{V}$ , we can now reformulate the optimal data fusion problem as the following optimization

$$\mathcal{E}(\pi^*(\mathbf{V})) = \inf_{\pi \in \mathfrak{F}_{\mathcal{G}}} \sum_{i \in \mathbf{V}} \mathcal{E}_i(\pi(\mathbf{V})), \quad (8)$$

where  $\mathfrak{F}_{\mathcal{G}}$  is the set of valid data fusion policies

$$\mathfrak{F}_{\mathcal{G}} := \{ \pi : L_{\mathcal{G}}(\mathbf{Y}_{\mathbf{V}}) \text{ computable at the fusion center} \}.$$

Note that the optimization in (8) is a function of the dependency graph  $\mathcal{G}$ , and that the optimal solution is attained by some policy.

<sup>8</sup>A sufficient statistic is a well-behaved function of the data, which is as informative as the raw data for inference. It is minimal if it is a function of every other sufficient statistic.

### B. Minimum energy data fusion: a lower bound

The following theorem gives a lower bound on minimum energy in (8), given the joint dependency graph  $\mathcal{G}$  and the path-loss coefficient  $\nu$ .

*Theorem 1 (Lower bound on minimum energy expenditure):* Let  $\text{MST}(\mathbf{V})$  be the Euclidean minimum spanning tree over node set  $\mathbf{V}$ . Then,

- 1) the energy cost for the optimal fusion policy  $\pi^*$  in (8) satisfies

$$\mathcal{E}(\pi^*(\mathbf{V})) \geq \sum_{e \in \text{MST}(\mathbf{V})} |e|^\nu := \mathcal{E}(\text{MST}(\mathbf{V})), \quad (9)$$

- 2) the lower bound (9) is achieved (i.e., equality holds) when the observations are independent under both hypotheses. In this case, the optimal fusion policy  $\pi^*$  aggregates data along  $\text{DMST}(\mathbf{V}; V_1)$ , the directed minimum spanning tree, with all the edges directed toward the fusion center  $V_1$ . Hence, the optimal fusion digraph  $\mathcal{F}_{\pi^*}$  is the  $\text{DMST}(\mathbf{V}; V_1)$ .

*Proof:* We will first prove part 2), for which we consider the case when observations are independent, and the log-likelihood ratio is given by

$$L_{\mathcal{G}}(\mathbf{Y}_{\mathbf{V}}) = \sum_{i \in \mathbf{V}} L_i(Y_i), \quad L_i(Y_i) := \log \frac{f_{1,i}(Y_i)}{f_{0,i}(Y_i)}.$$

Consider  $\text{MST}(\mathbf{V})$ , whose links minimize the sum of the power weighted edges  $\sum_{e \in \text{Tree}(\mathbf{V})} |e|^\nu$ . It is easy to check that at the fusion center, the log-likelihood ratio can be computed using the following aggregation scheme along the  $\text{DMST}(\mathbf{V}; V_1)$  as illustrated in (1): each node  $i$  computes the aggregated variable  $q_i(\mathbf{Y}_{\mathbf{V}})$  from its predecessor and sends it to its immediate successor. The variable  $q_i$  is given by the summation

$$q_i(\mathbf{Y}_{\mathbf{V}}) := \sum_{j \in \mathcal{N}_p(i)} q_j(\mathbf{Y}_{\mathbf{V}}) + L_i(Y_i), \quad (10)$$

where  $\mathcal{N}_p(i)$  is the set of immediate predecessors of  $i$  in  $\text{DMST}(\mathbf{V}; V_1)$ .

To show part 1), we note that any data fusion policy must have each node transmit at least once and the transmission must ultimately reach the fusion center. This implies that the fusion digraph must be connected with the fusion center and the DMST with edge-weight  $|e|^\nu$  minimizes the total energy under the above constraints. Hence, we have (9).  $\square$

Note that the above lower bound is tight in the sense that the bound is achievable when the measurements are independent under both hypotheses. It is interesting to note that data correlations, in general, increase the fusion cost under the constraint of optimal inference performance.

### C. Minimum energy data fusion: an upper bound

We now consider the general case of correlated measurements and devise a suboptimal data fusion scheme which gives an upper bound on the optimal energy in (8). The suboptimal scheme, referred to as Data Fusion on Markov Random Fields

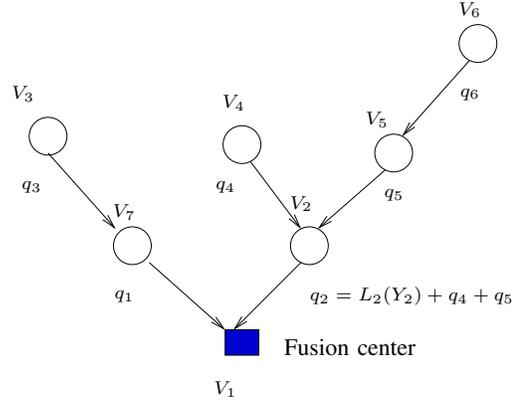


Fig. 1. The optimal fusion graph DMST for independent observations.

(DFMRF), is a natural generalization of the MST aggregation scheme described in Theorem 1.

Recall the form of the log-likelihood ratio for hypothesis testing of Markov random fields, given in (6)

$$L_{\mathcal{G}}(\mathbf{Y}_{\mathbf{V}}) = \sum_{c \in \mathcal{C}} \phi_c(\mathbf{Y}_c).$$

We shall use Fig. 2 to illustrate the idea behind DFMRF. It is made of two phases:

- 1) In the data forwarding phase, for each clique  $c$  in the set of maximal cliques  $\mathcal{C}$ , a *processor*, denoted by  $\text{Proc}(c)$ , is chosen randomly amongst the members of clique  $c$ . Each node in clique  $c$  then forwards its raw data to  $\text{Proc}(c)$  and  $\text{Proc}(c)$  computes the clique potential  $\phi_c(\mathbf{Y}_c)$ .
- 2) In the data aggregation phase, processors compute the sum of the clique potentials along  $\text{DMST}(\mathbf{V}; V_1)$ , the directed MST towards the fusion center.

Hence, the fusion digraph for the DFMRF scheme is the union of the two graphs in the above stages, viz., forwarding subgraph ( $\text{FG}(\mathbf{V})$ ) and aggregation subgraph ( $\text{AG}(\mathbf{V})$ ). The total energy consumption of DFMRF is given by

$$\mathcal{E}(\text{DFMRF}(\mathbf{V})) = \sum_{c \in \mathcal{C}(\mathbf{V})} \sum_{i \in c} \mathcal{E}^{\text{SP}}(i, \text{Proc}(c); \mathcal{N}_g) + \mathcal{E}(\text{MST}(\mathbf{V})), \quad (11)$$

where  $\mathcal{E}^{\text{SP}}(i, j; \mathcal{N}_g)$  denotes the energy consumption for the shortest path between  $i$  and  $j$  using the links in the network graph  $\mathcal{N}_g(\mathbf{V})$  (set of feasible links for direct transmission).

For independent measurements, the maximal clique set is trivially the set of vertices  $\mathbf{V}$  and hence, DFMRF reduces to aggregation along the  $\text{DMST}(\mathbf{V}; V_1)$ , which is optimal for independent observations. However, in general, DFMRF is not optimal. For the 1-nearest neighbor dependency graph, DFMRF has a constant approximation ratio with respect to the optimal data fusion scheme  $\pi^*$  in (8).

*Theorem 2 (Approximation under 1-NGG dependency [10]):* DFMRF is a 2-approximation algorithm when the dependency graph  $\mathcal{G}$  is the 1-nearest neighbor graph

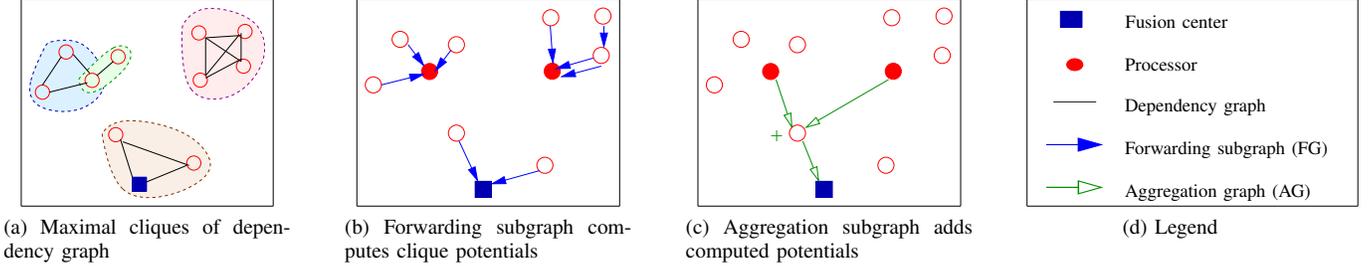


Fig. 2. Schematic of dependency graph of Markov random field and stages of data fusion.

$$\frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}))}{\mathcal{E}(\pi^*(\mathbf{V}))} \leq 2, \quad (12)$$

over all node sets  $\mathbf{V}$  in  $\mathbb{R}^2$ .

*Proof:* Since 1-NNG is acyclic, the maximum clique size is 2. Hence, for DFMRF, the forwarding subgraph (FG) is the 1-NNG with arbitrary directions on the edges. We have

$$\mathcal{E}(\text{FG}(\mathbf{V})) = \mathcal{E}(\text{1-NNG}(\mathbf{V})) \leq \mathcal{E}(\text{MST}(\mathbf{V})).$$

Thus,

$$\mathcal{E}(\text{DFMRF}(\mathbf{V})) = \mathcal{E}(\text{FG}(\mathbf{V})) + \mathcal{E}(\text{AG}(\mathbf{V})), \quad (13)$$

$$\leq 2 \mathcal{E}(\text{MST}(\mathbf{V})) \leq 2\mathcal{E}(\pi^*(\mathbf{V})), \quad (14)$$

where the last inequality comes from Theorem 1.  $\square$

Note that the above result does not extend to general  $k$ -NNG dependency graphs ( $k > 1$ ) for finite network size. However, as the network size goes to infinity, we will show in Section IV-B that a constant-factor approximation ratio is achieved.

#### IV. ENERGY SCALING LAWS

We now establish the scaling laws for optimal and suboptimal fusion policies. From the expression of average energy cost in (5), we see that the scaling laws will rely on the law of large numbers (LLN) for stabilizing graph functionals. An overview of the LLN is provided in Appendix A.

We recall some notations and definitions used in this section.  $X_i \stackrel{i.i.d.}{\sim} \kappa$ , where  $\kappa$  is defined on  $Q_1$ , the unit square centered at the origin. The node set is  $\mathbf{V}_n := \sqrt{\frac{n}{\lambda}}(X_i)_{i=1}^n$  and the limit is obtained by letting  $n \rightarrow \infty$  with fixed  $\lambda > 0$ .

##### A. Energy scaling for optimal fusion: independent case

We first provide the scaling result for the case when the measurements are independent under either hypothesis. From Theorem 1, the optimal fusion scheme minimizing total energy consumption is given by summation along the directed minimum spanning tree. Hence, the energy scaling is obtained by the analysis of the MST.

For node set  $\mathbf{V}_n$ , the average energy consumption of the optimal fusion scheme for independent measurements is

$$\bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) = \bar{\mathcal{E}}(\text{MST}(\mathbf{V}_n)) = \frac{1}{n} \sum_{e \in \text{MST}(\mathbf{V}_n)} |e|^\nu. \quad (15)$$

Let  $\zeta(\nu; \text{MST})$  be the constant arising in the asymptotic analysis of the MST edge lengths, that is

$$\zeta(\nu; \text{MST}) := \mathbb{E} \left[ \sum_{e \in E(\mathbf{0}; \text{MST}(\mathcal{P}_1 \cup \{\mathbf{0}\}))} \frac{1}{2} |e|^\nu \right], \quad (16)$$

where  $\mathbf{0}$  is a point at the origin of  $\mathbb{R}^2$ ,  $\mathcal{P}_\tau$  is the homogeneous Poisson process of intensity  $\tau$ , and  $E(\mathbf{0}; \text{MST}(\mathcal{P}_1 \cup \{\mathbf{0}\}))$  denotes the set of edges incident to the origin in  $\text{MST}(\mathcal{P}_1 \cup \{\mathbf{0}\})$ . The above constant is half the expectation of the power-weighted edges incident to the origin in the minimum spanning tree over a homogeneous unit intensity Poisson process, and is discussed in Appendix A in (41). Although  $\zeta(\nu; \text{MST})$  is not available in closed form, we evaluate it through simulations in Section V.

We now provide the scaling result for the optimal fusion scheme when the measurements are independent based on the LLN for the MST obtained in [9, Thm 2.3(ii)].

*Theorem 3 (Scaling for independent data [9]):* When the sensor measurements are independent under each hypothesis, the limit of the average energy consumption of the optimal fusion scheme in (15) is given by

$$\lim_{n \rightarrow \infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \stackrel{L^2}{=} \lambda^{-\frac{\nu}{2}} \zeta(\nu; \text{MST}) \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx. \quad (17)$$

Hence, asymptotically the average energy consumption of optimal fusion is a constant in the mean square sense for independent measurements. In contrast, forwarding all the raw data to the fusion center according to the shortest-path (SP) policy has an unbounded average energy growing in the order of  $\sqrt{n}$ .

The scaling constant for average energy in (17) brings out the influence of several factors on energy consumption. It is inversely proportional to the node density  $\lambda$ . This is intuitive since placing the nodes with a higher density (smaller area) decreases the average inter-node distances and hence, also the energy consumption.

The node-placement pdf  $\kappa$  influences the asymptotic energy consumption through the term

$$\int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx.$$

When the placement is uniform ( $\kappa \equiv 1$ ), the above term evaluates to unity. Hence, the scaling constant in (17) for

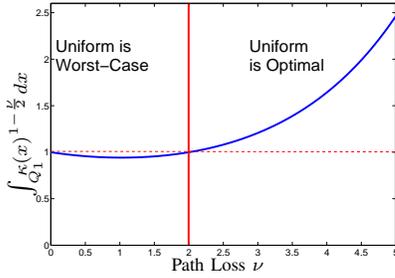


Fig. 3. Ratio of energy consumption under node placement distribution  $\kappa$  and uniform distribution as a function of path-loss  $\nu$ . See (18) and (19).

uniform placement simplifies to

$$\lambda^{-\frac{\nu}{2}} \zeta(\nu; \text{MST}).$$

The next theorem shows that the energy under uniform node placement ( $\kappa \equiv 1$ ) optimizes the scaling limit in (17) when the path loss  $\nu > 2$ . Also, see Fig.3.

*Theorem 4 (Minimum energy placement: independent case):*

For any pdf  $\kappa$  supported on the unit square  $Q_1$ , we have

$$\int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx \geq 1, \quad \forall \nu > 2, \quad (18)$$

$$\int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx \leq 1, \quad \forall \nu \in [0, 2). \quad (19)$$

*Proof:* We have the Hölder inequality

$$\|f_1 f_2\|_1 \leq \|f_1\|_p \|f_2\|_q, \quad \forall p > 1, q = \frac{p}{p-1}, \quad (20)$$

where for any positive function  $f$ ,

$$\|f\|_p := \left( \int_{Q_1} f(x)^p dx \right)^{\frac{1}{p}}.$$

When  $\nu > 2$ , in (20), substitute  $f_1(x)$  with  $\kappa(x)^{\frac{1}{p}}$ ,  $f_2(x)$  with  $\kappa(x)^{-\frac{1}{p}}$ , and  $p$  with  $\frac{\nu}{\nu-2} \geq 1$  which ensures that  $p > 1$ , to obtain (18).

For  $\nu \in (0, 2)$ , in (20), substitute  $f_1(x)$  with  $\kappa(x)^{\frac{1}{p}}$ ,  $f_2(x)$  with 1,  $p = \frac{2}{2-\nu} > 1$  to obtain (19).  $\square$

The above result implies that, in the context of i.i.d. node placements, it is asymptotically energy-optimal to place the nodes uniformly when the path-loss coefficient  $\nu > 2$ . The intuitive reason is as follows: without loss of generality, consider a clustered distribution in the unit square, where nodes are more likely to be placed near the origin. The MST over such a point set has many short edges, but a few very long edges, since some nodes are likely to occur near the boundary. On the other hand, for uniform point sets, the edges of the MST are more likely to be of similar lengths. Since for energy consumption, we have a power law on edge lengths with path loss  $\nu > 2$ , long edges are penalized harshly, leading to the result that the uniform distribution is optimal.

## B. Energy scaling for optimal fusion: MRF case

We now evaluate the scaling laws for energy consumption of the DFMRF scheme for a general Markov random field dependency among sensor measurements. The DFMRF aggregation scheme involves cliques of the dependency graph which arise from correlation between sensor measurements. Recall that the total energy consumption of DFMRF in (11) is given by

$$\begin{aligned} \mathcal{E}(\text{DFMRF}(\mathbf{V})) &= \sum_{c \in \mathcal{C}(\mathbf{V})} \sum_{i \subset c} \mathcal{E}^{\text{SP}}(i, \text{Proc}(c); \mathcal{N}_g) \\ &+ \mathcal{E}(\text{MST}(\mathbf{V})), \end{aligned} \quad (21)$$

where  $\mathcal{E}^{\text{SP}}(i, j; \mathcal{N}_g)$  denotes the energy consumption for the shortest path between  $i$  and  $j$  using the links in the network graph  $\mathcal{N}_g(\mathbf{V})$  (set of feasible links for direct transmission). We now assume that the network graph  $\mathcal{N}_g(\mathbf{V})$  is a *local u-energy spanner*. In the literature [44], a graph  $\mathcal{N}_g(\mathbf{V})$  is called a *u-energy spanner*, for some constant  $u > 0$  called its *energy stretch factor*, when it satisfies

$$\max_{i, j \in \mathbf{V}} \frac{\mathcal{E}^{\text{SP}}(i, j; \mathcal{N}_g)}{\mathcal{E}^{\text{SP}}(i, j; C_g)} \leq u, \quad (22)$$

where  $C_g(\mathbf{V})$  denotes the complete graph. In other words, the energy consumption between any two nodes is no worse than  $u$ -times the optimal value. Examples of energy spanners include the Gabriel graph<sup>9</sup> (with stretch factor  $u = 1$  when the path-loss  $\nu \geq 2$ ), the Yao graph, and its variations [44]. In this paper, we only require a weaker version<sup>10</sup> of the above property that there is at most  $u$ -energy stretch between the neighbors in the dependency graph

$$\max_{(i, j) \in \mathcal{G}} \frac{\mathcal{E}^{\text{SP}}(i, j; \mathcal{N}_g)}{\mathcal{E}^{\text{SP}}(i, j; C_g)} \leq u. \quad (23)$$

From (23), we have

$$\begin{aligned} \mathcal{E}(\text{FG}(\mathbf{V})) &\leq u \sum_{c \in \mathcal{C}(\mathbf{V})} \sum_{i \subset c} \mathcal{E}^{\text{SP}}(i, \text{Proc}(c); C_g), \\ &\leq u \sum_{c \in \mathcal{C}(\mathbf{V})} \sum_{i \subset c} |i, \text{Proc}(c)|^\nu. \end{aligned} \quad (24)$$

Recall that the processors are local within the clique, i.e.,  $\text{Proc}(c) \subset c$ , for each clique  $c$  in the dependency graph. Hence, in (24), only the edges of the processors of all the cliques are included in the summation. This is upper bounded by the sum of all the power-weighted edges of the dependency graph  $\mathcal{G}(\mathbf{V})$ . Hence, we have

$$\mathcal{E}(\text{FG}(\mathbf{V})) \leq u \sum_{e \in \mathcal{G}(\mathbf{V})} |e|^\nu. \quad (25)$$

<sup>9</sup>The longest edge in Gabriel graph is  $O(\sqrt{\log n})$ , the same order as that of the MST [45]. Hence, the maximum power at a node needed to ensure  $u$ -energy spanning property is of the same order as that needed for connectivity.

<sup>10</sup>In fact, it suffices to have  $\limsup_{n \rightarrow \infty} \max_{(i, j) \in \mathcal{G}(\mathbf{V}_n)} \frac{\mathcal{E}^{\text{SP}}(i, j; \mathcal{N}_g(\mathbf{V}_n))}{\mathcal{E}^{\text{SP}}(i, j; C_g(\mathbf{V}_n))} \leq u$ .

From (21), for the total energy consumption of the DFMRF scheme, we have the upper bound,

$$\mathcal{E}(\text{DFMRF}(\mathbf{V})) \leq u \sum_{e \in \mathcal{G}(\mathbf{V})} |e|^\nu + \mathcal{E}(\text{MST}(\mathbf{V})). \quad (26)$$

By (26), the total cost of this scheme  $\mathcal{E}(\text{DFMRF})$  is upper bounded by the sum of powers of edge lengths of the dependency graph, allowing us to draw upon the general methods of [9], [46].

From (26), the DFMRF scheme will scale whenever the right-hand side of (25) scales. By Theorem 3, the energy consumption along the MST scales. Hence, we only need to establish the scaling behavior of the first term in (25).

We now prove scaling laws governing the energy consumption of DFMRF and we also establish its approximation ratio with respect to the optimal fusion scheme. This in turn also establishes the scaling behavior of the optimal scheme.

*Theorem 5 (Scaling of DFMRF Scheme):* When the dependency graph  $\mathcal{G}$  is either the  $k$ -nearest neighbor or the disk graph, the average energy of DFMRF scheme satisfies

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)) \\ & \stackrel{a.s.}{\leq} \limsup_{n \rightarrow \infty} \left( \frac{1}{n} \sum_{e \in \mathcal{G}(\mathbf{V}_n)} u |e|^\nu + \bar{\mathcal{E}}(\text{MST}(\mathbf{V}_n)) \right) \\ & \stackrel{L^2}{=} \frac{u}{2} \int_{Q_1} \mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in \mathcal{G}(\mathcal{P}_{\lambda \kappa(x)} \cup \{\mathbf{0}\})} |\mathbf{0}, j|^\nu \right] \kappa(x) dx \\ & \quad + \lambda^{-\frac{\nu}{2}} \zeta(\nu; \text{MST}) \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx. \end{aligned} \quad (27)$$

*Proof:* See Appendix B.  $\square$

Hence, the above result establishes the scalability of the DFMRF scheme. In the theorem below, we use this result to prove the scalability of the optimal fusion scheme and establish asymptotic upper and lower bounds on its average energy.

*Theorem 6 (Scaling of Optimal Scheme):* When the dependency graph  $\mathcal{G}$  is either the  $k$ -nearest neighbor or the disk graph, the limit of the average energy consumption of the optimal scheme  $\pi^*$  satisfies the upper bound

$$\limsup_{n \rightarrow \infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \stackrel{a.s.}{\leq} \limsup_{n \rightarrow \infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)), \quad (28)$$

where the right-hand side satisfies the upper bound in (27). Also,  $\pi^*$  satisfies the lower bound given by the MST

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)) \stackrel{a.s.}{\geq} \liminf_{n \rightarrow \infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \\ & \stackrel{a.s.}{\geq} \lim_{n \rightarrow \infty} \bar{\mathcal{E}}(\text{MST}(\mathbf{V}_n)) \stackrel{L^2}{=} \lambda^{-\frac{\nu}{2}} \zeta(\nu; \text{MST}) \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx. \end{aligned} \quad (29)$$

*Proof:* From (9), the DFMRF and the optimal scheme satisfy the lower bound given by the MST.  $\square$

Hence, the limiting average energy consumption for both the DFMRF scheme and the optimal scheme is strictly finite, and is bounded by (27) and (29). These bounds also establish that the approximation ratio of the DFMRF scheme is asymptotically bounded by a constant, as stated below. Define the constant  $\rho := \rho(u, \lambda, \kappa, \nu)$ , given by

$$\rho := 1 + \frac{u \int_{Q_1} \frac{1}{2} \mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in \mathcal{G}(\mathcal{P}_{\lambda \kappa(x)} \cup \{\mathbf{0}\})} |\mathbf{0}, j|^\nu \right] \kappa(x) dx}{\lambda^{-\frac{\nu}{2}} \zeta(\nu; \text{MST}) \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx}. \quad (30)$$

*Lemma 1 (Approximation Ratio for DFMRF):* The approximation ratio of DFMRF is given by

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\pi^*(\mathbf{V}_n))} \\ & \stackrel{a.s.}{\leq} \limsup_{n \rightarrow \infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\text{MST}(\mathbf{V}_n))} \stackrel{L^2}{=} \rho, \end{aligned} \quad (31)$$

where  $\rho$  is given by (30).

*Proof:* Combine Theorem 5 and Theorem 6.  $\square$

We further simplify the above results for the  $k$ -nearest neighbor dependency graph in the corollary below by exploiting its scale invariance. The results are expected to hold for other *scale-invariant* stabilizing graphs as well. The edges of a scale-invariant graph are invariant under a change of scale, or put differently,  $\mathcal{G}$  is scale invariant if scalar multiplication by  $\alpha$  induces a graph isomorphism from  $\mathcal{G}(\mathbf{V})$  to  $\mathcal{G}(\alpha \mathbf{V})$  for all node sets  $\mathbf{V}$  and all  $\alpha > 0$ .

Along the lines of (16), let  $\zeta(\nu; k\text{-NNG})$  be the constant arising in the asymptotic analysis of the  $k$ -NNG edge lengths, that is

$$\zeta(\nu; k\text{-NNG}) := \mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in k\text{-NNG}(\mathcal{P}_1 \cup \{\mathbf{0}\})} \frac{1}{2} |\mathbf{0}, j|^\nu \right]. \quad (32)$$

*Corollary 1 ( $k$ -NNG Dependency Graph):* We obtain a simplification of Theorem 5 and 6 for average energy consumption, namely

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \stackrel{a.s.}{\leq} \limsup_{n \rightarrow \infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)) \\ & \stackrel{a.s.}{\leq} \limsup_{n \rightarrow \infty} \left( \frac{1}{n} \sum_{e \in \mathcal{G}(\mathbf{V}_n)} u |e|^\nu + \bar{\mathcal{E}}(\text{MST}(\mathbf{V}_n)) \right) \\ & \stackrel{L^2}{=} \lambda^{-\frac{\nu}{2}} [u \zeta(\nu; k\text{-NNG}) + \zeta(\nu; \text{MST})] \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx. \end{aligned} \quad (33)$$

The approximation ratio of DFMRF satisfies

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\pi^*(\mathbf{V}_n))} \stackrel{a.s.}{\leq} \limsup_{n \rightarrow \infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\text{MST}(\mathbf{V}_n))} \\ & \stackrel{L^2}{=} \left( 1 + u \frac{\zeta(\nu; k\text{-NNG})}{\zeta(\nu; \text{MST})} \right). \end{aligned} \quad (34)$$

*Proof:* This follows from [9, Thm 2.2].  $\square$

Hence, the expressions for scaling bounds and the approximation ratio are simplified when the dependency graph is the  $k$ -nearest neighbor graph. A special case of this scaling result for nearest-neighbor dependency under uniform placement was proven in [36, Thm 2].

It is interesting to note that the approximation factor for the  $k$ -NNG dependency graph in (34) is independent of the node placement pdf  $\kappa$  and node density  $\lambda$ . Hence, DFMRF has the same efficiency under different node placements. The results of Theorem 4 on the optimality of uniform placement are applicable here, but for the lower and upper bounds on energy consumption. We formally state it below.

*Theorem 7 (Minimum energy bounds for  $k$ -NNG):*

Uniform node placement minimizes the asymptotic lower and upper bounds on the average energy consumption in (29) and (33) for  $k$ -NNG dependency graph over all i.i.d. node placements  $\kappa$ .

*Proof:* From Theorem 4 and (33).  $\square$

We also prove the optimality of uniform distribution under disk-dependency graphs, but over a limited set of node placements  $\kappa$ .

*Theorem 8 (Minimum energy bound for disk graph):*

Uniform node placement minimizes the asymptotic lower and upper bounds on the average energy consumption in (29) and (33) for disk dependency graph over all i.i.d. node placements  $\kappa$  satisfying the lower bound

$$\kappa(x) > \frac{1}{\lambda}, \quad \forall x \in Q_1, \quad (35)$$

where  $\lambda > 1$  is the (fixed) node placement density.

*Proof:* We use the fact that for the disk graph with a fixed radius, more edges are added as we scale down the area. Hence, for Poisson processes with intensities  $\lambda_1 > \lambda_2 > 0$ ,

$$\mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in \mathcal{G}(\mathcal{P}_{\lambda_1} \cup \{\mathbf{0}\})} |\mathbf{0}, j|^\nu \right] \geq \mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in \mathcal{G}(\mathcal{P}_{\lambda_2} \cup \{\mathbf{0}\})} |\mathbf{0}, j|^\nu \right] \left[ \frac{\lambda_2}{\lambda_1} \right]^{\frac{\nu}{2}},$$

where the right-hand side is obtained by merely rescaling the edges present at intensity  $\lambda_2$ . Since, new edges are added at  $\lambda_1$ , this is an inequality, unlike the case of  $k$ -NNG where the edge set is invariant under scaling. Substituting  $\lambda_1$  with  $\lambda\kappa(x)$ , and  $\lambda_2$  by 1, we have

$$\begin{aligned} & \int_{Q_1} \mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in \mathcal{G}(\mathcal{P}_{\lambda\kappa(x)} \cup \{\mathbf{0}\})} |\mathbf{0}, j|^\nu \right] \kappa(x) dx \\ & \geq \lambda^{-\frac{\nu}{2}} \mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in \mathcal{G}(\mathcal{P}_1 \cup \{\mathbf{0}\})} |\mathbf{0}, j|^\nu \right] \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx, \\ & \geq \lambda^{-\frac{\nu}{2}} \mathbb{E} \left[ \sum_{j: (\mathbf{0}, j) \in \mathcal{G}(\mathcal{P}_1 \cup \{\mathbf{0}\})} |\mathbf{0}, j|^\nu \right], \quad \nu > 2. \end{aligned}$$

$\square$

Hence, uniform placement is optimal if we limit to distributions  $\kappa$  satisfying (35). We have so far established the finite scaling of the average energy when the dependency graph describing correlations among the sensor observations is either

the  $k$ -NNG or the disk graph. However, we cannot expect finite scaling for any general dependency graph. For instance, for the complete graph, the optimal fusion scheme reduces to a version of the shortest path (SP) routing, where the average energy consumption grows as  $\sqrt{n}$ . Since the LLR in (6) is now function over a single clique containing all the nodes, the optimal scheme consists of a unique processor chosen optimally, to which all the other nodes forward their raw data along shortest paths, and the processor then forwards the value of the LLR to the fusion center.

## V. NUMERICAL ILLUSTRATIONS

As described in Section II-A,  $n$  nodes are placed in area  $\frac{n}{\lambda}$  and one of them is randomly chosen as the fusion center. We conduct 500 independent simulation runs and average the results. We fix node density  $\lambda = 1$ . We plot results for two cases of dependency graph, viz., the  $k$ -nearest neighbor graph and the disk graph with radius  $\delta$ .

In Fig.4, we plot the simulation results for  $k$ -nearest neighbor dependency and uniform node placement. Corollary 1 establishes that the average energy consumption of the DFMRF scheme in (33) is finite and bounded for asymptotic networks under  $k$ -NNG dependency. The results in Fig.4a agree with theory and we note that the convergence to asymptotic values is quick, and occurs in networks with as little as 30 nodes. Moreover, the energy for DFMRF scheme increases with the number of neighbors  $k$  in the dependency graph since more edges are added. On the other hand, the average energy under no aggregation (SP policy) increases without bound, as predicted in Section I-A.

We plot the approximation ratio of the DFMRF scheme for  $k$ -NNG in (34) vs. the number of nodes in Fig.4b and vs. the path-loss coefficient  $\nu$  in Fig.4c. As predicted by Corollary 1, the approximation ratio is a constant for large networks, and find a quick convergence to this value in Fig.4b as we increase the network size. In Fig.4c, we also find that the approximation ratio is insensitive with respect to the path loss  $\nu$ . Hence, DFMRF scheme has nearly the same efficiency in the entire range of  $\nu \in [2, 6]$  under the  $k$ -NNG dependency.

In Fig.5a, we plot the average energy consumption of DFMRF in (27) under uniform node placement and disk dependency graph with radius  $\delta$ . The average energy is bounded, as predicted by Theorem 5. As in the  $k$ -NNG case, on increasing the network size, there is a quick convergence to the asymptotic values. Moreover, as expected, energy consumption increases with  $\delta$  since more edges are added to the dependency graph. Note that the energy consumption at  $\delta = 0$  and  $\delta = 0.3$  are nearly the same, since at  $\delta = 0.3$ , the disk graph is very sparse, and hence, energy consumed in the forwarding stage (FG) of LLR computation is small.

We now study the effect of node placement distribution on energy consumption. In Fig.5b and 5c, we compare the uniform node placement with i.i.d. placement according to pdf  $\kappa$  given by

$$\kappa(x) = \kappa_1(x(1))\kappa_1(x(2)), \quad x \in \mathbb{R}^2, \quad (36)$$

where, for some  $a \neq 0$ ,  $\kappa_1$  is given by the truncated exponential

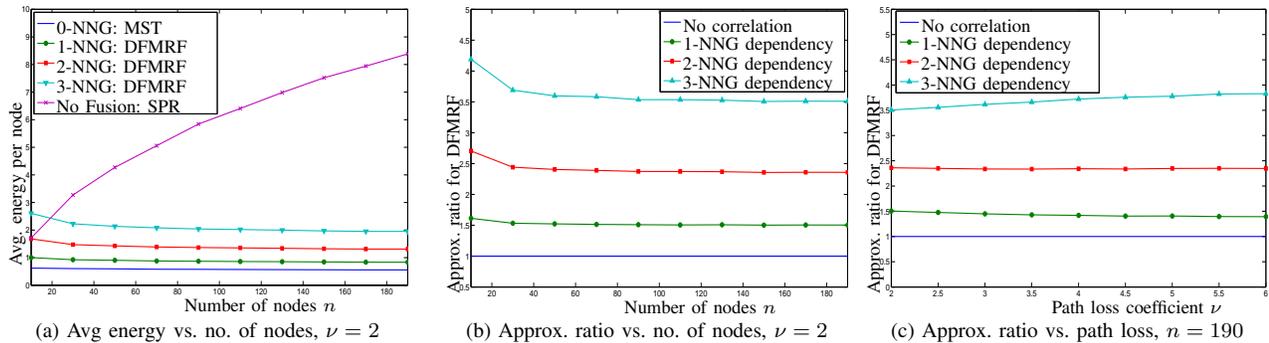


Fig. 4. Average energy consumption for DFMRF scheme and shortest-path routing for uniform distribution and  $k$ -NNG dependency. See Corollary 1.

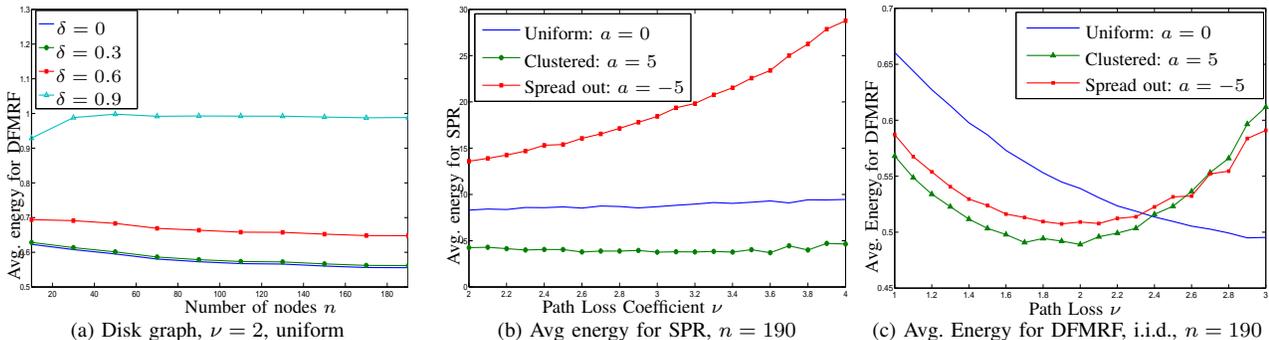


Fig. 5. Average energy consumption for DFMRF and shortest path (SPR) scheme. See Theorem 5.

$$\kappa_1(z) = \begin{cases} \frac{ae^{-a|z|}}{2(1 - e^{-\frac{a}{2}})}, & \text{if } z \in [-\frac{1}{2}, \frac{1}{2}], \\ 0, & \text{o.w.} \end{cases} \quad (37)$$

Note that as  $a \rightarrow 0$ , we obtain the uniform distribution in the limit. A positive (negative)  $a$  corresponds to clustering (spreading out) of the points with respect to the origin. In Fig.6, a sample realization for cases  $a = \pm 5$  and  $a = 0$  is shown.

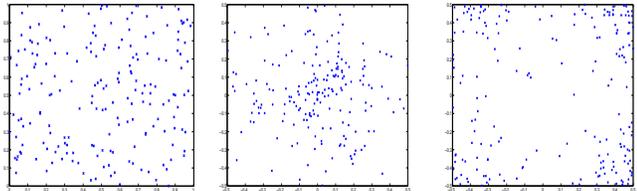


Fig. 6. Sample realization of  $n = 190$  points on unit square. See (36), (37).

Intuitively, for shortest-path (SP) policy, if we cluster the nodes close to one another, the average energy consumption decreases. On the other hand, spreading the nodes out towards the boundary increases the average energy. Indeed this behavior is validated by the results in Fig.5b, for  $\kappa$  defined above in (36) and (37). However, as we analyzed in the previous section, optimal node placement for the DFMRF scheme does not follow this simple intuition.

For i.i.d. data, from Theorem 4, the uniform node placement minimizes the asymptotic average energy consumption of the optimal scheme, which is aggregation along the MST, whenever the path-loss coefficient  $\nu \geq 2$ . For  $\nu \in [0, 2]$ , the uniform distribution has the worst-case energy. This is verified in Fig.5c, where for  $\nu \in [1, 3]$ , the uniform distribution initially has high energy consumption but decreases as we increase  $\nu$ . We see that at threshold of around  $\nu = 2.4$ , the uniform distribution starts having lower energy than the non-uniform placements (clustered and spread-out), while according to Theorem 4, the threshold is  $\nu = 2$ . Moreover, Theorem 4 also predicts that the clustered and spread-out distributions will have the same energy consumption since  $\int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx$  are equal for  $a = 5$  and  $a = -5$  for  $\kappa$  given by (36) and (37), and this approximately holds in Fig.5c.

## VI. CONCLUSION

We analyzed the scaling laws for energy consumption of data fusion schemes for optimal distributed inference. Forwarding all the raw data without fusion has an unbounded average energy as we increase the network size, and hence, is not a feasible strategy. We established finite average energy scaling for a fusion heuristic known as Data Fusion for Markov Random Fields (DFMRF) for certain class of spatial correlation models. We analyzed the influence of the correlation structure, node placement distribution, node density and the transmission environment on the energy consumption.

There are many issues that are not handled in this paper. Our fusion scheme DFMRF needs centralized topology information, and has to be extended to a distributed scheme, where

only local topology information is available. Our model currently only incorporates i.i.d. node placements. We expect our results to extend to the correlated node placement according to a Gibbs point process through the results in [47]. We have not considered here the scaling behavior of inference performance with network size, and is a topic of study in [37], [48]. We have not considered the time required for data fusion, and it will be interesting to establish bounds in this case. Our current correlation model assumes a discrete Markov random field. A more natural but difficult approach is to consider Markov field over continuous space [49] and then, sample it through node placements.

#### Acknowledgment

The authors thank A. Ephremides, T. He, D. Shah and the anonymous reviewers for helpful comments.

## APPENDIX

### A. Functionals on random points sets

In [9], [38], [50], Penrose and Yukich introduce the concept of stabilizing functionals to establish weak laws of large numbers for functionals on graphs with random vertex sets. As in this paper, the vertex sets may be marked (sensor measurements constituting one example of marks), but for simplicity of exposition we will work with unmarked vertices. We briefly describe the general weak law of large numbers after introducing the necessary definitions.

Graph functionals on a vertex set  $\mathbf{V}$  are often represented as sums of spatially dependent terms

$$\sum_{x \in \mathbf{V}} \xi(x, \mathbf{V}),$$

where  $\mathbf{V} \subset \mathbb{R}^2$  is locally finite (contains only finitely many points in any bounded region), and the measurable function  $\xi$ , defined on all pairs  $(x, \mathbf{V})$ , with  $x \in \mathbf{V}$ , represents the interaction of  $x$  with other points in  $\mathbf{V}$ . We see that the functionals corresponding to energy consumption can be cast in this framework.

When  $\mathbf{V}$  is random, the range of spatial dependence of  $\xi$  at node  $x \in \mathbf{V}$  is random, and the purpose of *stabilization* is to quantify this range in a way useful for asymptotic analysis. There are several similar notions of stabilization, but the essence is captured by the notion of stabilization of  $\xi$  with respect to homogeneous Poisson points on  $\mathbb{R}^2$ , defined as follows. Recall that  $\mathcal{P}_\tau$  is a homogeneous Poisson point process with intensity  $\tau$ .

We say that  $\xi$  is translation invariant if  $\xi(x, \mathbf{V}) = \xi(x + z, \mathbf{V} + z)$  for all  $z \in \mathbb{R}^2$ . Let  $\mathbf{0}$  denote the origin of  $\mathbb{R}^2$  and let  $B_r(x)$  denote the Euclidean ball centered at  $x$  with radius  $r$ . A translation-invariant  $\xi$  is *homogeneously stabilizing* if for all intensities  $\tau > 0$  there exists almost surely a finite random variable  $R := R(\tau)$  such that

$$\xi(\mathbf{0}, (\mathcal{P}_\tau \cap B_R(\mathbf{0})) \cup \mathcal{A}) = \xi(\mathbf{0}, \mathcal{P}_\tau \cap B_R(\mathbf{0}))$$

for all locally finite  $\mathcal{A} \subset \mathbb{R}^2 \setminus B_R(\mathbf{0})$ . Thus  $\xi$  stabilizes if the value of  $\xi$  at  $\mathbf{0}$  is unaffected by changes in point configurations outside  $B_R(\mathbf{0})$ .

$\xi$  satisfies the moment condition of order  $p > 0$  if

$$\sup_{n \in \mathbb{N}} \mathbb{E} [\xi(n^{\frac{1}{2}} X_1, n^{\frac{1}{2}} \{X_i\}_{i=1}^n)^p] < \infty. \quad (38)$$

We will use the following weak laws of large numbers throughout. Recall that  $X_i$  are i.i.d. with density  $\kappa$ .

*Theorem 9 (WLLN [9], [46]):* Put  $q = 1$  or  $q = 2$ . Let  $\xi$  be a homogeneously stabilizing translation-invariant functional satisfying the moment condition (38) for some  $p > q$ . Then

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi \left( \sqrt{\frac{n}{\lambda}} X_i, \sqrt{\frac{n}{\lambda}} \{X_j\}_{j=1}^n \right) \\ &= \int_{Q_1} \mathbb{E} [\xi(\mathbf{0}, \mathcal{P}_{\lambda \kappa(x)})] \kappa(x) dx \text{ in } L^q. \end{aligned} \quad (39)$$

We interpret the right-hand side of the above equation as a weighted average of the values of  $\xi$  on homogeneous Poisson point processes  $\mathcal{P}_{\lambda \kappa(x)}$ . When  $\xi$  satisfies scaling such as  $\mathbb{E} [\xi(\mathbf{0}, \mathcal{P}_\tau)] = \tau^{-\alpha} \mathbb{E} [\xi(\mathbf{0}, \mathcal{P}_1)]$ , then the limit on the right-hand side of (39) simplifies to

$$\lambda^{-\alpha} \mathbb{E} [\xi(\mathbf{0}, \mathcal{P}_1)] \int_{Q_1} (\kappa(x))^{1-\alpha} dx \text{ in } L^q, \quad (40)$$

a limit appearing regularly in problems in Euclidean combinatorial optimization. For uniform node placement ( $\kappa(x) \equiv 1$ ), the expression in (39) reduces to  $\mathbb{E} [\xi(\mathbf{0}, \mathcal{P}_\lambda)]$ , and the LLN result for this instance is pictorially depicted in Fig.7.

For example, if  $\xi(x, \mathbf{V})$  is one half the sum of the  $\nu$ -power weighted edges incident to  $x$  in the MST (or any scale-invariant stabilizing graph) on  $\mathbf{V}$ , i.e.,

$$\xi(x, \mathbf{V}) := \frac{1}{2} \sum_{e \in E(x, \text{MST}(\mathbf{V}))} |e|^\nu,$$

then substituting  $\alpha$  with  $\frac{\nu}{2}$  in (40),

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi \left( \sqrt{\frac{n}{\lambda}} X_i, \sqrt{\frac{n}{\lambda}} \{X_i\}_{i=1}^n \right) \\ &= \lambda^{-\frac{\nu}{2}} \mathbb{E} [\xi(\mathbf{0}, \mathcal{P}_1)] \int_{Q_1} (\kappa(x))^{1-\frac{\nu}{2}} dx \\ &= \lambda^{-\frac{\nu}{2}} \zeta(\nu; \text{MST}) \int_{Q_1} (\kappa(x))^{1-\frac{\nu}{2}} dx, \end{aligned} \quad (41)$$

where  $\zeta(\nu; \text{MST})$  is defined in (16).

### B. Proof of Theorem 5

The energy consumption of DFMRP satisfies the inequality in (27). For the MST we have the result in Theorem 3. We now use stabilizing functionals to show that

$$\frac{1}{n} \sum_{e \in \mathcal{G}(\mathbf{V}_n)} |e|^\nu$$

converges in  $L^2$  to a constant. For all locally finite vertex sets  $\mathcal{X} \subset \mathbb{R}^2$  supporting some dependency graph  $\mathcal{G}(\mathcal{X})$  and for all  $x \in \mathcal{X}$ , define the functional  $\eta(x, \mathcal{X})$  by

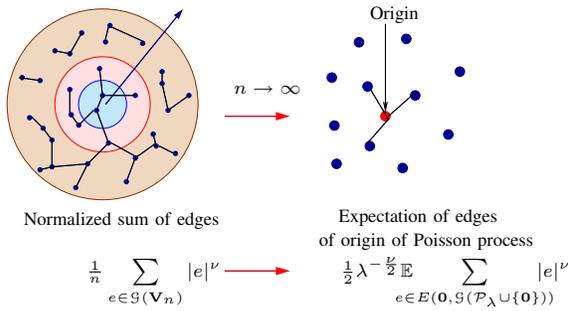


Fig. 7. LLN for sum graph edges on uniform point sets ( $\kappa \equiv 1$ ).

$$\eta(x, \mathcal{X}) := \sum_{y:(x,y) \in \mathcal{G}(\mathcal{X})} |x, y|^\nu. \quad (42)$$

Notice that  $\sum_{x \in \mathcal{X}} \eta(x, \mathcal{X}) = 2 \sum_{e \in \mathcal{G}(\mathcal{X})} |e|^\nu$ .

From [9, Thm 2.4], the sum of power-weighted edges of the  $k$ -nearest neighbors graph is a stabilizing functional and satisfies the bounded-moments condition (38). Hence, the limit in (39) holds when the dependency graph is the  $k$ -NNG.

We now show that the sum of power-weighted edges of the continuum percolation graph is a stabilizing functional which satisfies the bounded-moments condition (38), thus implying that the limit in (39) holds.

It is clear that  $\eta$  stabilizes with respect to  $\mathcal{P}_\tau$ ,  $\tau \in (0, \infty)$ , since points distant from  $x$  by more than the deterministic disc radius do not modify the value of  $\eta(x, \mathcal{P}_\tau)$ . Moreover,  $\eta$  satisfies the bounded moments condition (38) since each  $|x, y|$  is bounded by the deterministic disc radius and the number of nodes in  $n^{\frac{1}{2}} \{X_i\}_{i=1}^n$  which are joined to  $n^{\frac{1}{2}} X_1$  is a random variable with moments of all orders.  $\square$

## REFERENCES

- [1] A. Anandkumar, J.E. Yukich, A. Swami, and L. Tong, "Energy-Performance Scaling Laws for Statistical Inference in Large Random Networks," in *Proc. of ASA Joint Stat. Meet.*, Denver, USA, Aug. 2008.
- [2] A. Anandkumar, J. Yukich, L. Tong, and A. Swami, "Scaling Laws for Statistical Inference in Random Networks," in *Proc. of Allerton Conf. on Communication, Control and Computing*, Monticello, USA, Sept. 2008.
- [3] A. Ephremides, "Energy Concerns in Wireless Networks," *IEEE Wireless Communications*, no. 4, pp. 48–59, August 2002.
- [4] W. Li and H. Dai, "Energy-Efficient Distributed Detection Via Multihop Transmission in Sensor Networks," *IEEE Signal Processing Letters*, vol. 15, pp. 265–268, 2008.
- [5] A. Anandkumar, M. Wang, L. Tong, and A. Swami, "Prize-Collecting Data Fusion for Cost-Performance Tradeoff in Distributed Inference," in *Proc. of IEEE INFOCOM*, Rio De Janeiro, Brazil, April 2009.
- [6] A. Anandkumar, L. Tong, A. Swami, and A. Ephremides, "Minimum Cost Data Aggregation with Localized Processing for Statistical Inference," in *Proc. of INFOCOM*, Phoenix, USA, April 2008, pp. 780–788.
- [7] J. Steele, "Growth Rates of Euclidean Minimal Spanning Trees with Power Weighted Edges," *The Annals of Probability*, vol. 16, no. 4, pp. 1767–1787, 1988.
- [8] J. Yukich, "Asymptotics for Weighted Minimal Spanning Trees on Random Points," *Stochastic Processes and their Applications*, vol. 85, no. 1, pp. 123–138, 2000.
- [9] M. Penrose and J. Yukich, "Weak Laws Of Large Numbers In Geometric Probability," *Annals of Applied Probability*, vol. 13, no. 1, pp. 277–303, 2003.
- [10] A. Anandkumar, L. Tong, and A. Swami, "Energy Efficient Routing for Statistical Inference of Markov Random Fields," in *Proc. of CISS '07*, Baltimore, USA, March 2007, pp. 643–648.
- [11] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Tran. Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.
- [12] M. Franceschetti and R. Meester, *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2008.
- [13] Q. Zhao and L. Tong, "Energy Efficiency of Large-Scale Wireless Networks: Proactive vs. Reactive Networking," *IEEE JSAC Special Issue on Advances in Military Wireless Communications*, May 2005.
- [14] X. Liu and M. Haenggi, "Toward Quasiregular Sensor Networks: Topology Control Algorithms for Improved Energy Efficiency," *IEEE Tran. on Parallel and Distributed Systems*, pp. 975–986, 2006.
- [15] X. Wu, G. Chen, and S. Das, "Avoiding Energy Holes in Wireless Sensor Networks with Nonuniform Node Distribution," *IEEE Tran. on Parallel and Distributed Systems*, vol. 19, no. 5, pp. 710–720, May 2008.
- [16] Q. Zhao, A. Swami, and L. Tong, "The Interplay Between Signal Processing and Networking in Sensor Networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 84–93, 2006.
- [17] A. Giridhar and P. Kumar, "Toward a Theory of In-network Computation in Wireless Sensor Networks," *IEEE Comm. Mag.*, vol. 44, no. 4, pp. 98–107, 2006.
- [18] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, "Network Correlated Data Gathering with Explicit Communication: NP-Completeness and Algorithms," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, no. 1, pp. 41–54, 2006.
- [19] P. von Rickenbach and R. Wattenhofer, "Gathering Correlated Data in Sensor Networks," in *Joint workshop on Foundations of Mobile Computing*, 2004, pp. 60–66.
- [20] H. Gupta, V. Navda, S. Das, and V. Chowdhary, "Efficient gathering of correlated data in sensor networks," in *Proc. of ACM Intl. symposium on Mobile ad hoc networking and computing*, 2005, pp. 402–413.
- [21] S. Madden, M. Franklin, J. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on Database Systems*, vol. 30, no. 1, pp. 122–173, 2005.
- [22] C. Intanagonwiwat, R. Govindan, and D. Esterin, "Directed Diffusion: A Scalable and Robust Paradigm for Sensor Networks," in *Proc. 6th ACM/Mobicom Conference*, Boston, MA, 2000, pp. 56–67.
- [23] B. Krishnamachari, D. Estrin, and S. Wicker, "Modeling Data-centric Routing in Wireless Sensor Networks," in *IEEE INFOCOM*, New York, USA, 2002.
- [24] A. Giridhar and P. Kumar, "Maximizing the functional lifetime of sensor networks," in *Proc. of IPSN*, 2005.
- [25] —, "Computing and Communicating Functions over Sensor Networks," *IEEE JSAC*, vol. 23, no. 4, pp. 755–764, 2005.
- [26] S. Subramanian, P. Gupta, and S. Shakkottai, "Scaling Bounds for Function Computation over Large Networks," in *IEEE ISIT*, June 2007.
- [27] O. Ayaso, D. Shah, and M. Dahleh, "Counting Bits for Distributed Function Computation," in *Proc. ISIT*, Toronto, Canada, July 2008, pp. 652–656.
- [28] Y. Sung, S. Misra, L. Tong, and A. Ephremides, "Cooperative Routing for Signal Detection in Large Sensor Networks," *IEEE JSAC*, vol. 25, no. 2, pp. 471–483, 2007.
- [29] J. Chamberland and V. Veeravalli, "How Dense Should a Sensor Network Be for Detection With Correlated Observations?" *IEEE Tran. on Information Theory*, vol. 52, no. 11, pp. 5099–5106, 2006.
- [30] S. Misra and L. Tong, "Error Exponents for Bayesian Detection with Randomly Spaced Sensors," *IEEE Tran. on Signal Processing*, vol. 56, no. 8, 2008.
- [31] Y. Sung, X. Zhang, L. Tong, and H. Poor, "Sensor Configuration and Activation for Field Detection in Large Sensor Arrays," *IEEE Tran. on Signal Processing*, vol. 56, no. 2, pp. 447–463, 2008.
- [32] Y. Sung, H. Yu, and H. V. Poor, "Information, Energy and Density for Ad-hoc Sensor Networks over Correlated Random Fields: Large-deviation Analysis," in *IEEE ISIT*, July 2008, pp. 1592–1596.
- [33] N. Katenka, E. Levina, and G. Michailidis, "Local Vote Decision Fusion for Target Detection in Wireless Sensor Networks," in *Joint Research Conf. on Statistics in Quality Industry and Tech.*, Knoxville, USA, June 2006.
- [34] L. Yu, L. Yuan, G. Qu, and A. Ephremides, "Energy-driven Detection Scheme with Guaranteed Accuracy," in *Proc. IPSN*, 2006, pp. 284–291.
- [35] A. Anandkumar, A. Ephremides, A. Swami, and L. Tong, "Routing for Statistical Inference in Sensor Networks," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and R. Liu, Eds. John Wiley & Sons, 2009, ch. 25.
- [36] A. Anandkumar, L. Tong, and A. Swami, "Optimal Node Density for Detection in Energy Constrained Random Networks," *IEEE Tran. Signal Proc.*, vol. 56, no. 10, pp. 5232–5245, Oct. 2008.
- [37] —, "Detection of Gauss-Markov Random Fields with Nearest-neighbor Dependency," *IEEE Tran. Information Theory*, vol. 55, no. 2, Feb. 2009.

- [38] M. Penrose and J. Yukich, "Limit Theory For Random Sequential Packing And Deposition," *Annals of Applied probability*, vol. 12, no. 1, pp. 272–301, 2002.
- [39] M. Penrose, *Random Geometric Graphs*. Oxford University Press, 2003.
- [40] P. Brémaud, *Markov Chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer, 1999.
- [41] P. Clifford, "Markov Random Fields in Statistics," *Disorder in Physical Systems*, pp. 19–32, 1990.
- [42] M. Jerrum and A. Sinclair, "Polynomial Time Approximations for the Ising Model," *SIAM J. Computing*, vol. 22, no. 5, pp. 1087–1116, 1993.
- [43] E. Dynkin, "Necessary and Sufficient Statistics for a Family of Probability Distributions," *Tran. Math. Stat. and Prob.*, vol. 1, pp. 23–41, 1961.
- [44] X. Li, "Algorithmic, Geometric and Graphs Issues in Wireless Networks," *Wireless Comm. and Mobile Computing*, vol. 3, no. 2, March 2003.
- [45] P. Wan and C. Yi, "On the Longest Edge of Gabriel Graphs in Wireless Ad Hoc Networks," *IEEE Tran. on Parallel and Distributed Systems*, pp. 111–125, 2007.
- [46] M. Penrose, "Laws Of Large Numbers In Stochastic Geometry With Statistical Applications," *Bernoulli*, vol. 13, no. 4, pp. 1124–1150, 2007.
- [47] T. Schreiber and J. Yukich, "Stabilization and Limit Theorems for Geometric Functionals of Gibbs Point Processes," *Arxiv preprint arXiv:0802.0647*, 2008.
- [48] A. Anandkumar, J. Yukich, L. Tong, and A. Willsky, "Detection Error Exponent for Spatially Dependent Samples in Random Networks," in *submitted to Proc. of IEEE ISIT*, Seoul, S. Korea, July 2009.
- [49] H. Kunsch, "Gaussian Markov Random Fields," *J. Fac. Sci. Univ. of Tokyo*, no. 26, pp. 53–73, 1979.
- [50] M. Penrose and J. Yukich, "Central Limit Theorems For Some Graphs In Computational Geometry," *Annals of Applied Probability*, vol. 11, no. 4, pp. 1005–1041, 2001.