Energy Scaling Laws for Distributed Inference in Random Networks

Animashree Anandkumar*, Joseph E. Yukich†, Lang Tong*, and Ananthram Swami‡

*ECE Dept., Cornell University, Ithaca, NY 14853, Email: {aa332@,ltong@ece}.cornell.edu
†Dept., of Mathematics, Lehigh University, Bethlehem, Pa. 18015 Email: joseph.yukich@lehigh.edu
‡Army Research Laboratory, Adelphi, MD 20783, Email: a.swami@ieee.org

***Abstract*—The energy scaling laws of multihop data fusion networks for distributed inference are considered. The fusion network consists of randomly located sensors independently distributed 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 hypotheses, the minimum per-sensor energy cost is bounded below by a minimum spanning tree data fusion and above by a suboptimal scheme referred to as Data Fusion for Markov Random Field (DFMRF). Scaling laws are derived for the optimal and suboptimal fusion policies.**

***Index Terms*—Distributed detection, data fusion, graphical models, scaling laws, and wireless networks.**

I. Introduction

We consider the problem of distributed statistical inference via a network of randomly located sensors, each taking measurements and transporting the locally processed data to a fusion center. The fusion center then makes an inference about the underlying phenomenon based on data collected from individual sensors.

For statistical inference using wireless sensor networks, energy consumption is one of the most important design factors. The transmission power required for a receiver distance d away to have a certain signal-to-noise ratio (SNR) scales in the order of d^ν where $2 \leq \nu \leq 6$ is the path loss. 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, 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 $\mathbf{V}_n = \{V_1, \cdots, V_n\}$, a fusion policy π_n maps $\mathbf{V}_n$ to a set of transmissions. The per-sensor cost (e.g. energy) is given by

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

where $\mathcal{E}_i(\pi_n(\mathbf{V}_n))$ is the cost at node i under policy π_n. The per-sensor cost above is random, and we are interested in the energy scalability of such random networks as $n \to \infty$.

Definition 1 (Scalable Policy): A sequence of policies π_n is scalable on the average if

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

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

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

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

We focus mostly on the L^2 scalability of 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 with finite average per-sensor 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 and the shortest path (SP) policy where each node forwards its data to the fusion center using the shortest path route.

We assume for now that n sensor nodes are uniformly distributed with fixed density $\lambda > 0$ in a convex region having area $\frac{n}{\lambda}$. It is perhaps not surprising that neither of the two policies is scalable as $n \to \infty$. For the DT policy, intuitively, the average transmission range scales as $\sqrt{n}$, thus $\bar{\mathcal{E}}(\mathrm{DT}(\mathbf{V}_n))$ scales as $n^{\frac{\nu}{2}}$. For the SP policy, on the other hand, the average transmission distance does not scale with n, but the number of hops to the fusion center scales in the order of $\sqrt{n}$. Thus $\bar{\mathcal{E}}(\mathrm{SP}(\mathbf{V}_n))$ scales as $\sqrt{n}$. Rigorously establishing the scaling laws for these two non-scalable policies is not crucial at this point since the same scaling law can be easily established for regular networks when sensor nodes are on two-dimensional lattice points. See [2].

Are there scalable policies for data fusion? Among all the fusion policies not performing aggregation at the intermediate

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.

An extended version of this work has been submitted in [1].

nodes, the shortest path 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 performance requirements for specific applications. A simplistic approach may greatly reduce the amount of data transported but significantly affect the inference performance at the fusion center.

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. We assume that the underlying hypotheses can be modeled as Markov random fields and investigate energy scaling laws.

Given sensor locations $\mathbf{V}_n$ and possibly correlated sensor measurements, finding the minimum energy fusion policy is in general NP-hard and hence, intractable. We will establish upper and lower bounds on the fusion energy of the optimal scheme and analyze its scaling behavior. The lower bound is achieved by a minimum spanning tree fusion scheme, which is shown to be optimal when observations are statistically independent under both hypotheses. The upper bound is established through a specific suboptimal fusion scheme, referred to as Data Fusion over Markov Random Field (DFMRF). DFMRF becomes optimal for conditionally independent observations, and for certain spatial dependencies between sensor measurements of practical significance (e.g., nearest neighbor graph); it has an approximation ratio 2, i.e., it costs no more than twice the cost of the optimal fusion scheme, independent of the size of the network.

We then proceed to establish a number of asymptotic properties of DFMRF in Section IV-B, 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 [3], [4], we provide a precise characterization of the scaling bounds as a function of sensor node density and sensor distribution. These asymptotic bounds for DFMRF, in turn, are also applicable to the optimal fusion scheme. Hence, we use the DFMRF scheme as a vehicle to establish scaling laws for optimal fusion. Additionally, we use the expressions for scaling bounds to optimize the distribution of the sensor placements. For conditionally independent measurements and for correlated measurements with k-nearest neighbor dependency graph, we show that the uniform distribution minimizes the scaling bounds over all i.i.d placements.

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 an energy scaling law for the design of sensor placement is new and has direct engineering implications. The heuristic policy DFMRF first appeared in [5] and is made precise here with detailed asymptotic analysis using the weak law of large numbers for stabilizing graph functionals.

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* graphs denoting the 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 $\mathbf{V}_n$. For our scaling law analysis, we consider a sequence of sensor populations on an expanding square regions $Q_{\frac{n}{\lambda}}$ of area $\frac{n}{\lambda}$ centered at the origin where we fix λ as the overall sensor density and let the number of sensors $n \to \infty$.

To generate sensor locations V_i, first let $Q_1 := [-\frac{1}{2}, \frac{1}{2}]^2$ be the unit area square centered at the origin, and $X_i \overset{i.i.d.}{\sim} \kappa, 1 \leq i \leq n$ be a set of n independent and identically distributed (i.i.d.) random variables distributed on Q_1 according to κ. Here, κ is a probability density function (pdf) on Q_1 which is bounded away from zero and infinity. We next 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 λ.

B. Graphical inference model: dependency graphs

The inference problem we consider is the simple binary hypothesis testing $\mathcal{H}_0$ vs. $\mathcal{H}_1$ on a pair of Markov random fields (MRF). The theory of MRF is well established (see e.g., [6]). Under regularity conditions [6], an MRF is defined by its (undirected) dependency graph $\mathcal{G}$ and an associated density function $f(\cdot|\mathcal{G})$ on $\mathcal{G}$.

Under hypothesis $\mathcal{H}_k$, we assume dependency graphs $\mathcal{G}_k := (\mathbf{V}_n, E_k)$ which model the correlation structures of the sensor observations where $\mathbf{V}_n = \{V_1, \cdots, V_n\}$ 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 links or edges of the dependency graph $\mathcal{G}_k$ and it defines the correlations of the sensor observations. A precise definition of the dependency graph involves conditional-independence relations between the sensor measurements and can be found in [6].

We restrict our attention to proximity-based dependency graphs. In particular, we consider two classes of dependency graphs[1], the (undirected) k-nearest neighbor graph (k-NNG) and the disk graph also known as the continuum percolation

[1]The k-nearest neighbor graph (k-NNG) has edges (i, j) if i is one of the top 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).

graph. We expect our results to extend to other locally-defined dependency graphs such as the Delaunay or Voronoi graphs, the minimal spanning tree, the sphere of influence graph and the Gabriel graph. An important localization property of the aforementioned graphs is a certain stabilization property facilitating asymptotic scaling analysis.

C. Graphical inference model: likelihood functions

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

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

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

The celebrated Hammersley-Clifford theorem states that, under the positivity conditions [7], the log-likelihood function can be expressed as

$$-\log f(\mathbf{Y_V}|\mathcal{G}_k, \mathcal{H}_k) = \sum_{c \in \mathcal{C}_k} \psi_{k,c}(\mathbf{Y}_c), \tag{3}$$

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 $\mathbf{Y}_c$. We assume that the normalization constant is already incorporated in the potential functions.

D. Graphical fusion model and energy consumption

Nodes are capable of adjusting their transmission power and we assume that the sensor network is connected but not necessarily fully connected. The set of feasible communications links form the (directed) *network graph* denoted by $\mathcal{N}_g(\mathbf{V})$. Transmissions on feasible links are assumed perfect and they do not interfere with each other.

A fusion policy π consists of a transmission schedule that specifies 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 π by a directed *fusion graph* $\mathcal{F}_\pi := (\mathbf{V}, \overrightarrow{E}_\pi)$ where $\mathbf{V}$ is the same set of vertices corresponding to sensor locations, and $\overrightarrow{E}_\pi$ contains *directed links*. A directed link $< i, j >$ denotes a direct transmission from i to j and is contained in the network graph $\mathcal{N}_g(\mathbf{V})$. If one node communicates with another node k times, k direct links will be added between these two nodes. 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 $\overrightarrow{E}_\pi$ and calculate the sum cost for π.

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 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 $R_{i,j}$ away, we assume that node i will spend energy[2] $\gamma R_{i,j}^\nu$. Without loss of generality, we assume $\gamma = 1$. Hence, given a fusion policy $\mathcal{F}_\pi = (\mathbf{V}, \overrightarrow{E}_\pi)$ of network size n, the per-node energy consumption is given by

$$\bar{\mathcal{E}}(\pi(\mathbf{V})) = \frac{1}{n} \sum_{<i,j> \in \overrightarrow{E}_\pi} R_{i,j}^\nu, \quad 2 \le \nu \le 6. \tag{4}$$

The model specification is now complete.

III. Minimum Energy Data Fusion

We present in this section data fusion policies aimed at minimizing energy expenditure. The scalability of these optimal 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 graph and associated likelihood function. They do share the same vertex set $\mathbf{V}$, however, which allows us to join the two graphical models into one.

Define joint dependency graph $\mathcal{G}{:=}(\mathbf{V}, E), E{:=}E_0 \bigcup E_1$, as the union of two dependency graphs. The sufficient statistic is given by the log-likelihood ratio (LLR). With the substitution of (3), it is given by

$$\begin{aligned} L_\mathcal{G}(\mathbf{Y_V}) &:= \log \frac{f(\mathbf{Y_V}|\mathcal{G}_1, \mathcal{H}_0)}{f(\mathbf{Y_V}|\mathcal{G}_0, \mathcal{H}_1)} \\ &= \sum_{a \in \mathcal{C}_1} \psi_{1,a}(\mathbf{Y}_a) - \sum_{b \in \mathcal{C}_0} \psi_{0,b}(\mathbf{Y}_b) \\ &:= \sum_{c \in \mathcal{C}} \phi_c(\mathbf{Y}_c), \quad \mathcal{C}{:=}\mathcal{C}_0 \bigcup \mathcal{C}_1 \end{aligned} \tag{5}$$

Hereafter, we will work with $(\mathcal{G}, L_\mathcal{G}(\mathbf{Y_V}))$. Note that the LLR is minimally sufficient [8] implying maximum savings in routing costs due to fusion.

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

$$\mathcal{E}(\pi^*(\mathbf{V})) = \min_{\pi \in \mathfrak{F}_\mathcal{G}} \sum_{i \in \mathbf{V}} \mathcal{E}_i(\pi(\mathbf{V})) \tag{6}$$

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

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

Note that the optimization in (6) is a function of the dependency graph $\mathcal{G}$.

[2]Since nodes only communicate a finite number of bits, we use energy instead of power as the cost measure.

B. Minimum energy data fusion: a lower bound

The following theorem gives a lower bound on minimum energy given the joint dependency graph $\mathcal{G} = (\mathbf{V}, E)$ and path-loss coefficient ν.

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

1) the energy cost for the optimal fusion policy π^* satisfies

$$\mathcal{E}(\pi^*(\mathbf{V})) \geq \sum_{e \in \mathrm{MST}(\mathbf{V})} |e|^\nu := \mathcal{E}(\mathrm{MST}(\mathbf{V})) \qquad (7)$$

2) the lower bound (7) is achieved (i.e., equality holds) when the observations are conditionally independent under both hypotheses. In this case, the optimal data fusion policy π^* aggregates data along $\mathrm{DMST}(\mathbf{V}; V_1)$, the directed minimal spanning tree, with all the edges directed toward the fusion center V_1. Hence, the optimal fusion graph $\mathcal{F}_{\pi^*}$ is the $\mathrm{DMST}(\mathbf{V}; V_1)$.

Note that the above lower bound is tight in the sense that the bound is achievable when the measurements are conditionally independent. It is interesting to note that data correlations in general increase fusion cost.

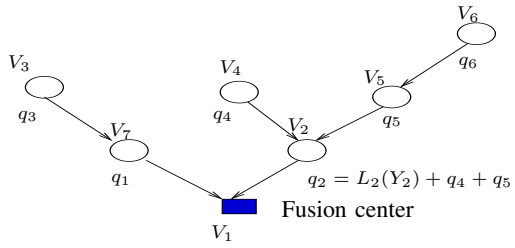


Fig. 1. The optimal fusion graph for conditionally independent observations.

C. Minimum energy data fusion: an upper bound

We now consider the general dependency graph and devise a suboptimal data fusion scheme which gives an upper bound on energy cost. The suboptimal scheme, referred to as Data Fusion on Markov Random Field (DFMRF), is a natural generalization of the MST aggregation scheme.

Recall the form of the log-likelihood ratio for a general Markov random field given in (5)

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

It should be now apparent that aggregation along the $\mathrm{DMST}(\mathbf{V}; V_1)$ does not deliver the LLR to the fusion center for a general MRF. This is because each function ϕ_c aggregates raw measurements $\mathbf{Y}_c$ at a common processor, and this in general is not possible along the DMST.

We shall use Fig. 2 to illustrate the idea behind DFMRF. The fusion graph of DFMRF policy is made of two phases corresponding to the union of two graphs: data forwarding graph ($\mathrm{FG}(\mathbf{V})$) and data aggregation graph ($\mathrm{AG}(\mathbf{V})$). See Fig 2 for an illustration.

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

For conditionally independent measurements, the maximum clique set is $\mathbf{V}$ and the DFMRF reduces to the $\mathrm{DMST}(\mathbf{V}; V_1)$, which is optimal for conditionally independent observations. In general, DFMRF is not optimal. For the nearest-neighbor dependency graph, DFMRF has a constant approximation ratio of 2 with respect to the optimal data fusion scheme [5].

IV. Energy Scaling Laws

We now establish the scaling laws for optimal and suboptimal fusion policies. From the expression of per-sensor energy cost, we see that the scaling laws will rely on the law of large numbers (LLN) for stabilizing graph functionals.

A. Energy scaling for optimal fusion: independent case

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

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

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

$$\bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) = \bar{\mathcal{E}}(\mathrm{MST}(\mathbf{V}_n)) = \frac{1}{n} \sum_{e \in \mathrm{MST}(\mathbf{V}_n)} |e|^\nu. \qquad (8)$$

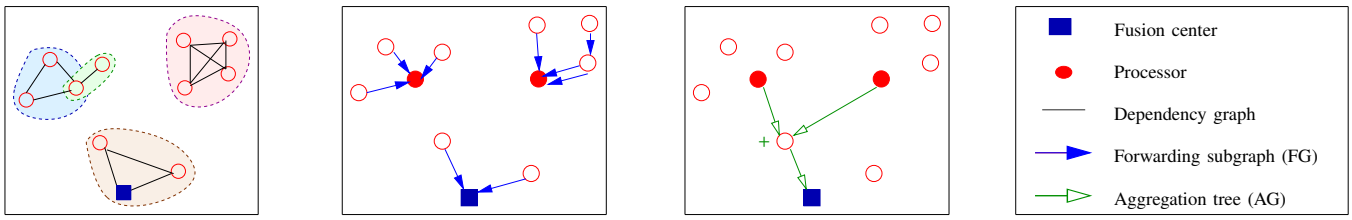
Let $\zeta(\nu; \mathrm{MST})$ be the constant arising in the asymptotic analysis of the MST edge lengths,

$$\zeta(\nu; \mathrm{MST}) := \mathbb{E}\Big[\sum_{e \in E(\mathbf{0}; \mathrm{MST}(\mathcal{P}_1 \cup \{\mathbf{0}\}))} \frac{1}{2} |e|^\nu \Big], \qquad (9)$$

where $\mathcal{P}_\tau$ is the homogeneous Poisson process of intensity τ. The above constant is half the expectation of the power-weighted edges belonging to the origin in the minimal spanning tree over a homogeneous unit intensity Poisson process.

We now provide the scaling result for the conditionally independent case based on the LLN for the MST obtained in [4, Thm 2.3(ii)].

Theorem 2 (Scaling for independent case [4]): When the sensor measurements are independent conditioned on each hypothesis, the limit of the average (per-node) energy consumption of the optimal fusion scheme in (8) is given by



(a) Maximal cliques of dependency graph (b) Forwarding subgraph computes clique potentials (c) Aggregation subgraph adds computed potentials (d) Legend

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

$$\lim_{n\to\infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \overset{L^2}{=} \lambda^{-\frac{\nu}{2}} \zeta(\nu; \mathrm{MST}) \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx. \quad (10)$$

Hence, asymptotically the average energy consumption of optimal fusion is a constant for independent measurements. In contrast, forwarding all the raw data to the fusion center has an unbounded average energy.

The scaling constant in (10) brings out the influence of several factors on energy consumption. The node density λ is inversely proportional to the limiting average energy. This is intuitive since placing the nodes with a higher density (smaller area) decreases the energy consumption. Although the constant $\zeta(\nu; \mathrm{MST})$ is not available in closed form, we evaluate it through simulations in Section V.

The node-placement pdf κ influences the limiting energy 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 limit for uniform placement equals

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

The next theorem shows that the energy under uniform node placement provides a lower bound on the limit for any general κ.

Theorem 3 (Minimum energy placement: independent case): For any pdf κ on the unit square Q_1, we have

$$\int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}} dx \geq 1, \quad \forall \nu \geq 2. \quad (11)$$

Proof: Using the convexity of the function $g(x) = x^{1-\frac{\nu}{2}}$ for $\nu \geq 2$ over the range of κ we obtain via Jensen's inequality

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

□

The above result implies that, in the context of i.i.d. node placements, from an energy point of view it is asymptotically optimal to place the nodes uniformly.

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 between sensor measurements. The total energy consumption of DFMRF is given by

$$\begin{aligned} \mathcal{E}(\mathrm{DFMRF}(\mathbf{V})) &= \sum_{c \in \mathcal{C}(\mathbf{V})} \sum_{i \subset c} \mathcal{E}^{\mathrm{SP}}(i, \mathrm{Proc}(c); \mathcal{N}_g) \\ &+ \mathcal{E}(\mathrm{MST}(\mathbf{V})), \end{aligned} \quad (12)$$

where $\mathcal{E}^{\mathrm{SP}}(i, j; \mathcal{N}_g)$ denotes the energy consumption for the shortest path between i and j using the links in $\mathcal{N}_g(\mathbf{V})$. We now assume that the network graph $\mathcal{N}_g(\mathbf{V})$ (set of feasible links) is a u-energy spanner [9], for some constant $u > 0$ called its energy stretch factor, and hence, satisfies

$$\max_{i,j \in \mathbf{V}} \frac{\mathcal{E}^{\mathrm{SP}}(i, j; \mathcal{N}_g)}{\mathcal{E}^{\mathrm{SP}}(i, j; C_g)} \leq u, \quad (13)$$

where $C_g(\mathbf{V})$ denotes the complete graph. Examples of energy spanners include the Gabriel graph (with stretch factor $u = 1$), the Yao graph, and its variations [9]. From (13), we have

$$\begin{aligned} \mathcal{E}(\mathrm{FG}(\mathbf{V})) &\leq u \sum_{c \in \mathcal{C}(\mathbf{V})} \sum_{i \subset c} \mathcal{E}^{\mathrm{SP}}(i, \mathrm{Proc}(c); C_g), \\ &\leq u \sum_{c \in \mathcal{C}(\mathbf{V})} \sum_{i \subset c} R^{\nu}_{i, \mathrm{Proc}(c)}. \end{aligned} \quad (14)$$

Recall that the processors are local: $\mathrm{Proc}(c) \subset c$. Hence, in (14), 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

$$\mathcal{E}(\mathrm{FG}(\mathbf{V})) \leq u \sum_{e \in \mathcal{G}(\mathbf{V})} |e|^{\nu}. \quad (15)$$

Hence, for the total energy consumption of the DFMRF scheme, we have the bound,

$$\mathcal{E}(\mathrm{DFMRF}(\mathbf{V})) \leq u \sum_{e \in \mathcal{G}(\mathbf{V})} |e|^{\nu} + \mathcal{E}(\mathrm{MST}(\mathbf{V})). \quad (16)$$

The DFMRF aggregation scheme involves cliques of the dependency graph which arise from correlation between sensor measurements. Nonetheless, by (16) 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 [4], [10].

By Theorem 3, the DFMRF scheme will scale whenever the right-hand side of (15) scales. We will establish that the LLN is applicable to the first term in (16) when the dependency graph is either the k-nearest neighbor or the disk graph.

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 4 (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 the upper bound

$$\begin{aligned}
&\limsup_{n\to\infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)) \\
&\overset{a.s.}{\leq} \limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{e\in\mathcal{G}(\mathbf{V}_n)} u\,|e|^{\nu} + \bar{\mathcal{E}}(\text{MST}(\mathbf{V}_n)) \Big) \\
&\overset{L^2}{=} \frac{u}{2} \int_{Q_1} \mathbb{E}\Big[\sum_{j:(\mathbf{0},j)\in\mathcal{G}(\mathcal{P}_{\lambda\kappa(x)}\cup\{\mathbf{0}\})} R^{\nu}_{\mathbf{0},j} \Big] \kappa(x) dx \\
&\qquad + \lambda^{-\frac{\nu}{2}} \zeta(\nu;\text{MST}) \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}}\, dx. \qquad (17)
\end{aligned}$$

Hence, the above result establishes scalability of the DFMRF scheme. 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 5 (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 π^* satisfies the upper bound

$$\limsup_{n\to\infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \overset{a.s.}{\leq} \limsup_{n\to\infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)), \qquad (18)$$

where the right-hand side satisfies the upper bound in (17). Also, π^* satisfies the lower bound given by the MST

$$\begin{aligned}
&\liminf_{n\to\infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)) \overset{a.s.}{\geq} \liminf_{n\to\infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \\
&\overset{a.s.}{\geq} \lim_{n\to\infty} \bar{\mathcal{E}}(\text{MST}(\mathbf{V}_n)) \overset{L^2}{=} \lambda^{-\frac{\nu}{2}} \zeta(\nu;\text{MST}) \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}}\, dx \qquad (19)
\end{aligned}$$

Proof: From (7), the DFMRF and the optimal scheme satisfy the lower bound given by the MST. □

Hence, the limiting average energy consumption under the DFMRF scheme and the optimal scheme is strictly finite, and is bounded by (17) and (19). 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

$$\rho := 1 + \frac{u \int\limits_{Q_1} \frac{1}{2}\mathbb{E}\Big[\sum\limits_{j:(\mathbf{0},j)\in\mathcal{G}(\mathcal{P}_{\lambda\kappa(x)}\cup\{\mathbf{0}\})} R^{\nu}_{\mathbf{0},j} \Big] \kappa(x) dx}{\lambda^{-\frac{\nu}{2}} \zeta(\nu;\text{MST}) \int\limits_{Q_1} \kappa(x)^{1-\frac{\nu}{2}}\, dx}. \qquad (20)$$

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

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\pi^*(\mathbf{V}_n))} \\
&\overset{a.s.}{\leq} \limsup_{n\to\infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\text{MST}(\mathbf{V}_n))} \overset{L^2}{=} \rho, \qquad (21)
\end{aligned}$$

where ρ is given by (20).

Proof: Combine Theorem 4 and Theorem 5. □

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 α 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 (9), let $\zeta(\nu; k\text{-NNG})$ be the constant arising in the asymptotic analysis of the k-NNG edge lengths,

$$\zeta(\nu; k\text{-NNG}) := \mathbb{E}\Big[\sum_{j:(\mathbf{0},j)\in k\text{-NNG}(\mathcal{P}_1\cup\{\mathbf{0}\})} \frac{1}{2} R^{\nu}_{\mathbf{0},j} \Big]. \qquad (22)$$

Corollary 1 (k-NNG Dependency Graph): We obtain a simplification of Theorem 4 and 5 for average energy consumption, namely

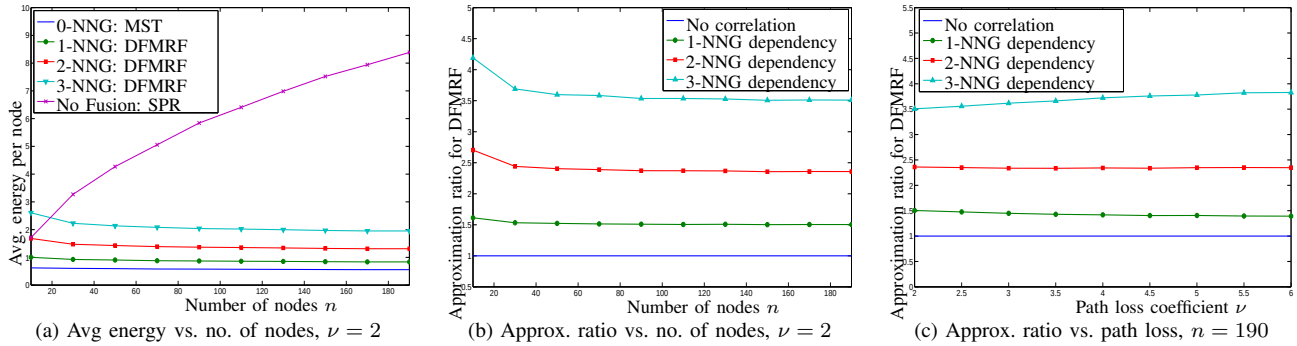
$$\begin{aligned}
&\limsup_{n\to\infty} \bar{\mathcal{E}}(\pi^*(\mathbf{V}_n)) \overset{a.s.}{\leq} \limsup_{n\to\infty} \bar{\mathcal{E}}(\text{DFMRF}(\mathbf{V}_n)) \\
&\overset{a.s.}{\leq} \limsup_{n\to\infty} \Big(\frac{1}{n} \sum_{e\in\mathcal{G}(\mathbf{V}_n)} u\,|e|^{\nu} + \bar{\mathcal{E}}(\text{MST}(\mathbf{V}_n)) \Big) \\
&\overset{L^2}{=} \lambda^{-\frac{\nu}{2}} \left[u\,\zeta(\nu; k\text{-NNG}) + \zeta(\nu;\text{MST}) \right] \int_{Q_1} \kappa(x)^{1-\frac{\nu}{2}}\, dx \qquad (23)
\end{aligned}$$

The approximation ratio of DFMRF satisfies

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\pi^*(\mathbf{V}_n))} &\overset{a.s.}{\leq} \limsup_{n\to\infty} \frac{\mathcal{E}(\text{DFMRF}(\mathbf{V}_n))}{\mathcal{E}(\text{MST}(\mathbf{V}_n))} \\
&\overset{L^2}{=} \Big(1 + u \frac{\zeta(\nu; k\text{-NNG})}{\zeta(\nu;\text{MST})} \Big). \qquad (24)
\end{aligned}$$

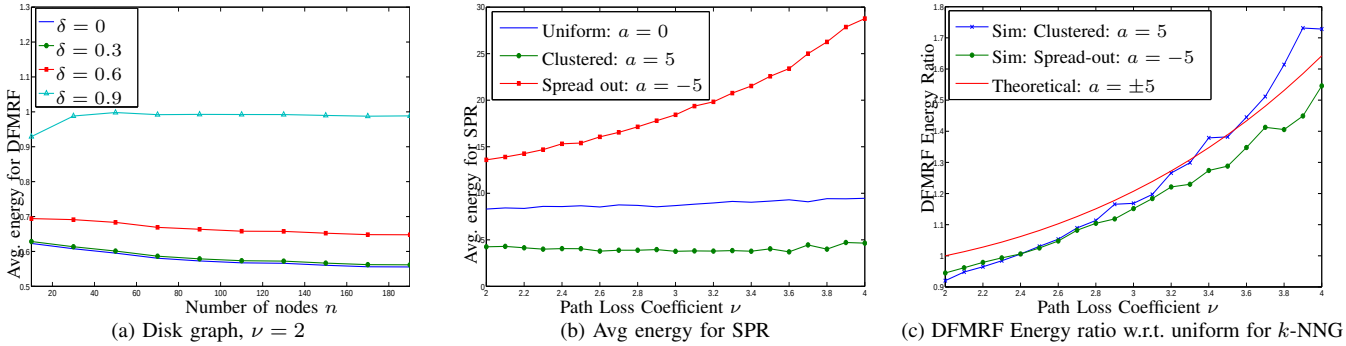
Proof: This follows from [4, Thm 2.2]. □

Hence, the expressions for scaling bounds and the approximation ratio are simplified when the dependency graph is the



(a) Avg energy vs. no. of nodes, $\nu = 2$ (b) Approx. ratio vs. no. of nodes, $\nu = 2$ (c) Approx. ratio vs. path loss, $n = 190$

Fig. 3. Average energy consumption for DFMRF scheme and shortest-path routing for uniform distribution and k-NNG dependency. See Corollary 1.



(a) Disk graph, $\nu = 2$ (b) Avg energy for SPR (c) DFMRF Energy ratio w.r.t. uniform for k-NNG

Fig. 4. Average energy consumption for DFMRF and shortest path (SPR) scheme. See Theorem 4.

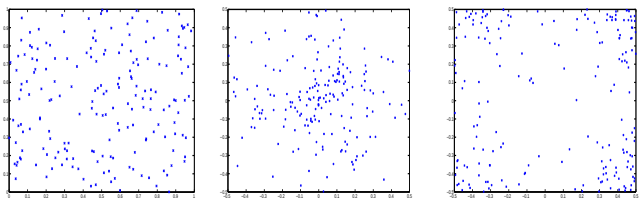
k-nearest neighbor graph. A special case of this scaling result for nearest-neighbor dependency under uniform placement was proven in [11, Thm 2].

It is interesting to note that the approximation factor for the k-NNG dependency graph in (24) is independent of the node placement pdf κ and node density λ. Indeed the actual energy consumption is governed by these parameters. The results of Theorem 3 on the optimality of uniform placement are also applicable here. We formally state it below.

Theorem 6 (Minimum energy placement for k-NNG): Uniform node placement minimizes the asymptotic upper bound 23) for average energy consumption under k-NNG dependency over all i.i.d. node placements κ.

Proof: This follows from Theorem 3. □

Hence, we have established the finite scaling of the average energy when the dependency graph describing the sensor observations is either the k-NNG or the disk graph.



(a) Uniform $a \to 0$ (b) Clustered $a = 5$ (c) Spread-out $a = -5$

Fig. 5. Sample realization of $n = 190$ points on unit square. See (25), (26).

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. Of the n nodes, we uniformly pick one of them as the fusion center. We fix node density $\lambda = 1$. We plot results for two cases of dependency graph, the k-nearest neighbor graph and the disk graph with radius δ.

In Fig.3, we plot the simulation results for k-nearest neighbor dependency and uniform node placement. In Fig.3a, the average energy consumption of DFMRF scheme converges quickly as the network size increases. On the other hand, the average energy under no aggregation increases without bound. Also, the energy for DFMRF scheme increases with the number of neighbors k in the dependency graph since more edges are added.

We plot the approximation ratio of the DFMRF scheme vs. the number of nodes in Fig.3b and vs. the path-loss coefficient ν in Fig.3c. We find that it is insensitive with respect to ν. Hence, DFMRF scheme is efficient for the entire range of $\nu \in [2, 6]$ under the k-NNG dependency.

In Fig.4a, we plot the average energy consumption of DFMRF under uniform node placement and disk dependency graph with radius δ. As expected, energy consumption increases with δ. As in the k-NNG case, on increasing the network size, there is a quick convergence.

We compare the i.i.d. uniform node placement with i.i.d.

placement according to pdf κ given by

$$\kappa(x) = \kappa_1(x(1))\kappa_1(x(2)), \quad x \in \mathbb{R}^2, \tag{25}$$

where for some $a \neq 0$, κ_1 is given by the truncated exponential

$$\kappa_1(z) = \begin{cases} \dfrac{ae^{-a|z|}}{2(1-e^{-\frac{a}{2}})}, & \text{if } z \in [-\frac{1}{2}, \frac{1}{2}], \\ 0, & \text{o.w.} \end{cases} \tag{26}$$

Note that as $a \to 0$, 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.5, a sample realization for cases $a = \pm 5$ is shown.

Intuitively, for shortest-path routing (SPR), if we cluster the nodes close to one another, the total energy consumption decreases. On the other hand, spreading the nodes out towards the boundary increases the total energy. Indeed this behavior is validated by the results in Fig.4b.

The behavior of the DFMRF scheme under different node placements, is however, not so straightforward. Recall that the asymptotic bound for average energy of DFMRF in (17) comprises two terms, one corresponding to edges of the dependency graph, and the other, to the edges of the MST. They may behave differently for different placement pdfs κ depending on the dependency graph model and path loss ν.

For the k-NNG dependency graph, from Theorem 6, uniform node placement minimizes the asymptotic bound on average energy. In Fig.4c, for the k-NNG dependency graph, we plot the ratio of energy of DFMRF under non-uniform placement with respect to the energy under uniform placement. We also plot the theoretical value of this ratio, given by Corollary 1 as

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

for κ given by (25) and (26), and find that the above expression is equal for $a = 5$ and $a = -5$. We observe that the simulation results are close to the theoretically predicted value.

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, and hence, is not a feasible strategy. We established constant average energy scaling for a fusion heuristic known as the 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 model currently only incorporates i.i.d. node placements. The behavior of the inference performance, along the lines of our preliminary results in [12] and its scaling laws is currently under investigation. We have not considered the time required for data fusion, and it will be interesting to establish its bounds.

Acknowledgment

The authors thank Anthony Ephremides and Ting He for comments.

References

[1] A. Anandkumar, J.E.Yukich, L. Tong, and A. Swami, "Energy Scaling Laws for Distributed Inference in Random Networks," *Submitted to IEEE J. Selec. Area Comm., available on Arxiv*, Aug. 2008.

[2] W. Li and H. Dai, "Energy-efficient distributed detection via multihop transmission in sensor networks," *Signal Processing Letters, IEEE*, vol. 15, pp. 265–268, 2008.

[3] J. E. Yukich, "Probability theory of classical Euclidean optimization problems," *Lecture Notes in Mathematics*, vol. 1675, 1998.

[4] M. D. Penrose and J. E. Yukich, "Weak laws of large numbers in geometric probability," *Annals of Applied probability*, vol. 13, no. 1, pp. 277–303, 2003.

[5] 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.

[6] P. Brémaud, *Markov Chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer, 1999.

[7] P. Clifford, "Markov random fields in statistics," *Disorder in Physical Systems*, pp. 19–32, 1990.

[8] E. Dynkin, "Necessary and sufficient statistics for a family of probability distributions," *Tran. Math, Stat. and Prob.*, vol. 1, pp. 23–41, 1961.

[9] X. Li, "Algorithmic, geometric and graphs issues in wireless networks," *Wireless Comm. and Mobile Computing*, vol. 3, no. 2, March 2003.

[10] M. Penrose, "Laws of large numbers in stochastic geometry with statistical applications," *Bernoulli*, vol. 13, no. 4, pp. 1124–1150, 2007.

[11] A. Anandkumar, L. Tong, and A. Swami, "Optimal Node Density for Detection in Energy Constrained Random Networks," *IEEE Trans. Signal Proc.*, vol. 56, no. 10, pp. 5232–5245, Oct. 2008.

[12] ——, "Detection of Gauss-Markov random fields with nearest-neighbor dependency," *Accepted to IEEE Tran. Information Theory*, Dec. 2006, available on Arxiv.