

Toward Resource-Optimal Consensus over the Wireless Medium

Matthew Nokleby, Waheed U. Bajwa, Robert Calderbank, and Behnaam Aazhang

Abstract—We carry out a comprehensive study of the resource cost of averaging consensus in wireless networks. Most previous approaches suppose a graphical network, which abstracts away crucial features of the wireless medium, and measure resource consumption only in terms of the total number of transmissions required to achieve consensus. Under a path-loss dominated model, we study the resource requirements of consensus with respect to three wireless-appropriate metrics: total transmit energy, elapsed time, and time-bandwidth product. First we characterize the performance of several popular gossip algorithms, showing that they may be order-optimal with respect to transmit energy but are strictly suboptimal with respect to elapsed time and time-bandwidth product. Further, we propose a new consensus scheme, termed *hierarchical averaging*, and show that it is nearly order-optimal with respect to all three metrics. Finally, we examine the effects of quantization, showing that hierarchical averaging provides a nearly order-optimal tradeoff between resource consumption and quantization error.

I. INTRODUCTION

Consider a network of N nodes, each of which possesses a scalar measurement $z_n(0) \in \mathbb{R}$. In *averaging consensus*, each node wishes to compute the average of these measurements: $z_{\text{ave}} = 1/N \sum_{n=1}^N z_n(0)$. Averaging consensus often is described as a simple, canonical example of distributed signal processing over sensor networks; a common narrative is that each node measures the local temperature and wants to compute the average temperature over the sensor field. Such simplicity, however, is deceptive, as consensus lies at the heart of an array of sophisticated problems, including load-balancing [1], distributed optimization [2]–[4], and distributed estimation and filtering [5], [6]. Advances in consensus therefore represent advances throughout distributed signal processing.

Due to their simplicity, flexibility, and robustness, *gossip algorithms* have emerged as a popular approach to consensus. In gossip, the network is modeled by a graph. Nodes iteratively pair with neighbors, exchange estimates, and average those estimates together, eventually converging on the true average. A large body of excellent work on gossip has been developed,

from the early *randomized gossip* of [7] to faster schemes such as *path averaging* [8] and *multi-scale gossip* [9]. Gossip is simple, requiring minimal processing and network knowledge, and it is robust, retaining performance even with failing links and dynamic topology.

However, the purpose of consensus strategies is often to facilitate processing over *wireless* networks, and wireless affords possibilities that existing strategies do not fully exploit. For example, in random geographic graphs, nodes are taken to be neighbors if they lie within a certain radius, usually chosen to be $\Theta(\sqrt{\log N/N})$ following [10]. Sophisticated consensus algorithms are then constructed in order to minimize the number of transmissions needed to achieve consensus. In wireless, however, transmit radius is adjustable via power control. Furthermore, given sufficient transmit power the network is fully connected, at which point consensus is trivial. This suggests both that wireless permits flexibility that may improve performance and that we must consider additional performance metrics—such as transmit power—that encompass more than just the number of transmissions.

While these issues have been taken up individually, in this work we consider them jointly to answer the following question: *What are the resource demands of consensus over wireless networks?* Our objective is to expand the framework in which consensus is studied in order to account for and exploit features of the wireless medium that previously have been overlooked. Given the ability to broadcast and to adjust connectivity dynamically, we seek fundamental limits on the wireless resources required to achieve consensus, as well as practical consensus strategies that attain those limits.

A. Contributions

First, in Section II we define a realistic but tractable framework in which to study the resource demands of consensus. We assume a path-loss dominated propagation model in which connectivity is determined by a signal-to-noise ratio (SNR) threshold. We define three resource metrics: the total energy expended in order to achieve consensus, the total time elapsed, and the time-bandwidth product consumed. In Section III we derive lower bounds on the required resources, and in Section IV we characterize several existing consensus strategies within our framework. We show that while path averaging is nearly order optimal with respect to energy expenditure, it remains suboptimal with respect to elapsed time and consumed time-bandwidth product.

Next, in Section V we propose a new consensus algorithm, termed *hierarchical averaging*, designed specifically for wireless networks. Instead of communicating with neighbors over

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org

M. Nokleby and R. Calderbank are with Duke University, Durham, NC (email: {matthew.nokleby, robert.calderbank}@duke.edu). W. U. Bajwa is with Rutgers University, Piscataway, NJ (email: waheed.bajwa@rutgers.edu). B. Aazhang is with Rice University, Houston, TX (email: aaz@rice.edu). This work was supported by the National Science Foundation, the Academy of Finland, the Tekes FiDiPro program, and the Air Force Office of Scientific Research under grant FA 9550-09-1-0643. Portions of this work were presented at the International Conference on Acoustics, Speech, and Signal Processing, Mar. 2012 and the Asilomar Conference on Signals, Systems, and Computers, Nov. 2012.

a graph, nodes broadcast estimates to geographically-defined clusters. These clusters expand as consensus proceeds, which is enabled by adjusting nodes' transmit powers. Similar to the hierarchical cooperation of [11] and the multiscale gossip of [9], small clusters cooperatively broadcast information to larger clusters, continuing until consensus is achieved. Depending on the particulars of the channel model, hierarchical averaging is nearly order optimal. In particular, when channel phases are fixed and identical, hierarchical averaging is order optimal with respect to all three metrics simultaneously, up to an arbitrarily small gap in the exponent, for path-loss exponents $2 \leq \alpha < 4$. In the more realistic case in which channel phases are random and independent, however, hierarchical averaging is no longer order optimal in transmit energy when $\alpha > 2$, although it remains order optimal with respect to the other two metrics.

Finally, in Section VI we incorporate quantization into our model. Since practical wireless links suffer from noise, achievable rates are finite and estimates must be quantized prior to transmission. This introduces a tradeoff: expending more energy increases the rate of the links, thereby reducing the quantization error inherent to each transmission and therefore the estimation error accrued during consensus. Therefore, in addition to the resource metrics of energy, time, and bandwidth, we introduce a fourth performance metric: mean-square error of nodes' estimates. Again we characterize existing consensus techniques. We also apply quantization to hierarchical averaging, showing that it permits an efficient tradeoff between energy and estimation error.

B. Prior Work

Consensus has been studied under various guises, including the early work of Tsitsiklis [2], which examined averaging in the context of distributed estimation. Recent interest in consensus was sparked by the introduction of randomized gossip [7], which defined the framework and developed the theoretical machinery in which most subsequent works have operated. Randomized gossip, however, has relatively slow convergence on random graphs, requiring roughly $\Theta(N^2)$ transmissions.¹ Since then, researchers have searched for faster consensus algorithms. In *geographic gossip* [12], nodes pair up with geographically distant nodes, exchanging estimates via multi-hop routing. The extra complexity garners faster convergence; geographic gossip requires roughly $\Theta(N^{3/2})$ transmissions. Geographic gossip was further refined by the introduction of *path averaging* [8], in which routing nodes contribute their own estimates “along the way.” Path averaging closes the gap to order optimality, requiring roughly $\Theta(N)$ transmissions, which is the minimum of any consensus algorithm.

A few works have addressed individually the wireless aspects we consider herein. The broadcast nature of wireless is considered in [13], [14]; however, in these works broadcast does not significantly improve performance over randomized

gossip. Multi-access interference is addressed—and in fact exploited—in [15], where lattice codes are used to compute sums of estimates “over the air.” In works that presage our own [16]–[18], the effects of power control are explored. In [16] the energy-minimal *fixed* power allocation for traditional consensus is derived. In [17], [18] the convergence speed of traditional consensus is studied as a function of nodes' fixed transmit radius, supposing that TDMA is used to mitigate interference among users.

Finally, many authors have studied the impact of noisy links on consensus. In [19], continuous-valued estimates are corrupted by zero-mean additive noise, and optimal linear consensus strategies are derived. For a similar model, the *bias-variance dilemma* is identified: Running consensus for longer reduces the bias of the resulting estimates, but it increases the variance. In [20], [21] *quantized consensus* algorithms are presented that achieve consensus over finite alphabets. In [22] standard gossip is augmented with *dithered* quantization and are shown to achieve consensus on the true average in expectation. In [23] the increasing correlation among estimates is exploited to construct a consensus algorithm employing Wyner-Ziv style coding with side information.

II. PRELIMINARIES

A. System Model

In defining the wireless model, we aim for a balance between tractability and practicality. To this end we make four critical assumptions, which we contend capture the salient features of wireless while maintaining simplicity: synchronous transmission, path-loss propagation, “protocol”-model connectivity, and orthogonalized interference management.

1) *Synchronous Transmission*: Let transmissions be broken up into time slots t , each slot having K channel uses. At time slot t , node n transmits a signal having average power $P_n(t)$. Nodes use capacity-achieving codes, and K is chosen sufficiently long that the probability of error is negligible.

2) *Path-loss Propagation*: Each node n has a geographic location $\mathbf{r}_n \in [0, 1] \times [0, 1]$ in the unit square, which we take to be independently drawn from a uniform distribution. Let $1 \leq k \leq K$ be a symbol time during block k . Under the path-loss model, the channel gain between any two nodes m, n is

$$h_{mn}(t, k) = \sqrt{G} \|\mathbf{r}_m - \mathbf{r}_n\|_2^{-\alpha/2} e^{j\theta_{mn}(t, k)}, \quad (1)$$

where $\alpha \geq 2$ is the path-loss exponent, $\theta_{mn}(t, k) \in [0, 2\pi)$ is a random phase with unspecified distribution, and G is a constant that depends on α and the transmit frequency. The choice of G does not affect scaling laws. In simulations, we choose G sufficiently small that channel gains are smaller than unity.

3) *Protocol Model*: Following the channel model in (1), the signal transmitted by node m arrives at node n with average power

$$R_{mn}(t) = \frac{1}{K} \sum_{k=1}^K |h_{mn}(t, k)|^2 P_n(t). \quad (2)$$

We suppose that a link exists between node n and m provided the received power is above a signal-to-noise ratio threshold

¹Throughout this paper we use the Landau notation: $f(n) = O(g(n))$ implies $f(n) \leq kg(n)$, $f(n) = \Omega(g(n))$ implies $f(n) \geq kg(n)$, and $f(n) = \Theta(g(n))$ implies $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$, all for constant k and sufficiently large n .

$\gamma > 0$. Define the *neighborhood* of node n as the set of nodes whose transmissions have sufficient received power:

$$\begin{aligned} \mathcal{N}_n(t) &= \{m : R_{mn}(t) \geq \gamma\} \\ &= \left\{m : P_m(t) \geq \frac{\gamma}{G} \|\mathbf{r}_m - \mathbf{r}_n\|_2^\alpha\right\}. \end{aligned} \quad (3)$$

For nodes $m \notin \mathcal{N}_n(t)$, we assume that node n suffers no interference from node m 's transmission. This assumption permits a tractable, geometric analysis of connectivity.

In hierarchical averaging, presented in Section V, we group nodes into clusters which transmit cooperatively. In this case we must expand the definition of neighborhoods to characterize the number of unique *signals* arriving at node n . Let $\mathcal{C}_i \subset \{1, \dots, N\}$, where $i \in \mathcal{I}$ for some index set \mathcal{I} , form a partition of the network. At time slot t , nodes within cluster \mathcal{C}_j transmit jointly. Define the received power at node n as

$$\begin{aligned} R_{\mathcal{C},n}(t) &= \frac{1}{K} \sum_{k=1}^K \left| \sum_{m \in \mathcal{C}_j} h_{mn}(t, k) \sqrt{P_m(t)} \right|^2 \\ &\approx E \left[\left| \sum_{m \in \mathcal{C}_j} h_{mn}(t, k) \sqrt{P_m(t)} \right|^2 \right], \end{aligned}$$

where the expectation is taken over the random phases, and the approximation is accurate for large K . Then, the neighborhood of n is the set of all clusters \mathcal{C}_j such that the received power exceeds γ :

$$\mathcal{N}_n(t) = \left\{ \mathcal{C}_j : E \left[\left| \sum_{m \in \mathcal{C}_j} h_{mn}(t, k) \sqrt{P_m(t)} \right|^2 \right] \geq \gamma, \right\}.$$

The connectivity of clusters depends on the distribution of the phases $\theta_{mn}(t, k)$. In the sequel we consider two possibilities. First, we consider the simple case in which the phases are equal and fixed. In this case, signals constructively combine at receivers, and the neighborhood of n can be written as

$$\mathcal{N}_n(t) = \left\{ \mathcal{C}_j : \left(\sum_{m \in \mathcal{C}_j} \|\mathbf{r}_m - \mathbf{r}_n\|_2^{-\alpha/2} P_m^{\frac{1}{2}}(t) \right)^2 \geq \frac{\gamma}{G} \right\}. \quad (4)$$

The second, and more realistic, case we consider is that each $\theta_{mn}(t, k)$ is independently and uniformly distributed across $[0, 2\pi)$. Then, signals do not combine coherently, and the neighborhood of n is

$$\mathcal{N}_n(t) = \left\{ \mathcal{C}_j : \sum_{m \in \mathcal{C}_j} \|\mathbf{r}_m - \mathbf{r}_n\|_2^{-\alpha} P_m(t) \geq \frac{\gamma}{G} \right\}. \quad (5)$$

Finally, we define the neighborhood of a *cluster* \mathcal{C}_i as those clusters in the neighborhood of any of its nodes:

$$\mathcal{N}_{\mathcal{C}_i}(t) = \{\mathcal{C}_j : \exists n \in \mathcal{C}_i \text{ with } \mathcal{C}_j \in \mathcal{N}_n(t)\} \quad (6)$$

4) *Frequency Orthogonalization*: At every time slot t , each node n transmits over a frequency slot $f_n(t) \in \mathbb{N}$. In order for transmissions to be successful, each incoming signal must arrive on a different frequency slot, which means that each

node must transmit on a frequency different not only from its neighbors, but also from its neighbors' neighbors. The number of required frequency slots can be described in terms of graph coloring. Let $\mathcal{G}_2(t)$ denote the graph of two-hop neighbors—of either nodes or clusters, as appropriate—at time slot t . Then, the number of frequency slots required at time t is

$$B(t) = \chi(\mathcal{G}_2(t)),$$

where $\chi(\cdot)$ is the (vertex) chromatic number. Clearly,

$$B(t) \geq \max_n |\mathcal{N}_n(t)| \quad (7)$$

Further, by Brooks' theorem [24],

$$B(t) \leq \max_n \text{Deg}_n(\mathcal{G}_2(t)) + 1, \quad (8)$$

where $\text{Deg}_n(\cdot)$ is the *degree* of node n over $\mathcal{G}_2(t)$, or the number of two-hop neighbors.

B. Performance Metrics: Ideal Links

We first consider the case in which the links between neighboring nodes are ideal, meaning that the quantization artifacts associated with finite-rate transmission are neglected. Therefore, at time t , node n decodes an infinite-precision, real-valued scalar from each $m \in \mathcal{N}_n(t)$.

At time t , each node updates its estimates by taking a linear combination of estimates received from neighboring nodes:

$$z_n(t) = \sum_{m \in \mathcal{N}_n(t)} a_{mn}(t) z_m(t), \quad (9)$$

where $a_{mn}(t)$ are arbitrary coefficients. Note that the coefficients, as well as the neighborhoods, may vary with time. Since connectivity may vary in time, the weights do not necessarily need to form a doubly-stochastic matrix, as is typically required in consensus with fixed topologies [7].

The first performance metric is the ϵ -*averaging time*, defined as the number of time slots required to achieve consensus to within a specified tolerance²:

$$T_\epsilon = \sup_{\mathbf{z}(0) \in \mathbb{R}^n} \inf \left\{ t : \Pr \left(\frac{\|\mathbf{z}(t) - z_{\text{ave}} \mathbf{1}\|}{\|\mathbf{z}(0)\|} \geq \epsilon \right) \leq \epsilon \right\}, \quad (10)$$

where $\mathbf{z}(t)$ is the vector of estimates $z_n(t)$. The scaling law of T_ϵ is the primary focus of study in consensus. However, it provides only a partial measure of resource consumption in wireless networks, so we consider other metrics as well.

The next metric is the *total transmit energy*, the energy required to achieve consensus to within the tolerance ϵ :

$$E_\epsilon = K \sum_{n=1}^N \sum_{t=1}^{T_\epsilon} P_n(t). \quad (11)$$

The final figure of merit is the *time-bandwidth product*, the total number of frequency slot uses required to achieve consensus to within the tolerance ϵ :

$$B_\epsilon = K \sum_{t=1}^{T_\epsilon} B(t), \quad (12)$$

²This definition is inspired by the well-known connection between Markov chain mixing and averaging consensus [25].

C. Performance Metrics: Finite-Rate Links

In practice, wireless links are noisy and therefore have finite rate, which precludes the infinite-precision exchange of scalars. Instead, nodes must quantize their estimates to a finite alphabet prior to each round of consensus. To simplify the discussion, we suppose that the measurements $z_n(0)$ are drawn from the finite interval $[0, 1)$. Throughout this paper, we employ dithered uniform quantization described in [22]. For alphabet size L , the quantization alphabet \mathcal{Z} is defined as the midpoints of L equally-sized quantization bins. The quantizer is defined as

$$\phi(z) = \min_{q \in \mathcal{Z}} |q - (z + u)|, \quad (13)$$

where u is a dither, drawn uniformly and randomly from $[-\Delta/2, \Delta/2)$ each time ϕ is called, and $\Delta = 1/L$ is the width of each quantization bin. Statistically, one can write the quantized value as

$$\phi(z) = z + v,$$

where v is uniformly distributed across $[-\Delta/2, \Delta/2)$ and independent of z .

The alphabet size $L = |\mathcal{Z}|$ depends on the quality of the wireless links. Since we define connectivity at signal-to-noise threshold γ , we take L to be determined by the Shannon capacity of a wireless link at SNR γ . Supposing unit bandwidth and slot duration, nodes successfully transmit $K \log_2(1 + \gamma)$ bits over the wireless links, which results in an alphabet size of $L = \lfloor 2^{K \log_2(1 + \gamma)} \rfloor = \lfloor (1 + \gamma)^K \rfloor$.

At time t , each node updates its estimate by taking a functional combination of quantized estimates received from neighboring nodes:

$$z_n(t) = g_n(\phi(\mathbf{z}(t)), t), \quad (14)$$

where the function g_n may only depend on estimates $\phi(z_m(t))$ for which $m \in \mathcal{N}_n(t)$. Again the connectivity and the function g_n may vary with time.

For a consensus algorithm with quantization, let T be the number of rounds for which consensus is run. Then let B and E be the time-bandwidth product and total transmit energy, defined as before but with T taking the role of T_ϵ . Finally, define the *mean squared error* as

$$\sigma^2 = \max_{\mathbf{z}(0) \in [0, 1]^N} E \left[\frac{1}{N} \sum_{n=1}^N (z_n(T) - z_{\text{ave}})^2 \right], \quad (15)$$

where the expectation is taken over any randomness in the quantization operator as well as in the consensus algorithm. There is an inherent tradeoff between the total transmit energy E and the mean-squared error σ^2 . One can always reduce the estimation error by injecting more transmit energy into the network and increasing the rate of the wireless links.

Finally, throughout this paper we rely on the following lemma, which shows that the number of nodes in any region is asymptotically proportional to its area to within an arbitrary tolerance.

Lemma 1 (Ozgur-Leveque-Tse, [11]): Let $A \subset [0, 1] \times [0, 1]$ be a region inside the unit square having area $|A|$, and let $\mathcal{C} = \{n : \mathbf{r}_n \in A\}$ be the nodes lying in A . Then, for any

$\delta > 0$,

$$(1 - \delta)|A|N \leq |\mathcal{C}| \leq (1 + \delta)|A|N, \quad (16)$$

with probability greater than $1 - 1/|A|e^{-\Gamma(\delta)|A|N}$, where $\Gamma(\delta) > 0$ is a quantity independent of N and $|A|$.

III. INNER BOUNDS

In this section we derive inner bounds on the resource costs for consensus over the proposed wireless model. We begin with the case of ideal links.

Theorem 1: For any consensus algorithm, and for any $\epsilon = O(1)$, we have, with probability approaching 1 as $N \rightarrow \infty$:

$$T_\epsilon = B_\epsilon = \Omega(1) \quad (17)$$

$$E_\epsilon = \Omega(N^{1-\alpha/2}). \quad (18)$$

Proof: The bounds on T_ϵ and B_ϵ are trivial. The bound on E_ϵ follows from the observation in [8] that consensus to within a constant tolerance requires at least $\Omega(N)$ transmissions. Each transmission from node n consumes at least enough energy to reach its nearest neighbor, which by (3) is equal to $\gamma/Gd_{\min}^\alpha(n)$, where $d_{\min}(n)$ is the distance between node n and its nearest neighbor. It is well-known (e.g., in [26]), that $d_{\min}(n) = \Theta(N^{-1/2})$ with high probability, so

$$E_\epsilon \geq \Omega(N) \frac{\gamma}{G} \Theta(N^{-\alpha/2}) = \Omega(N^{1-\alpha/2}). \quad \blacksquare$$

When links are rate limited, we derive an inner bound on the tradeoff between resources and estimation error.

Theorem 2: For any consensus algorithm with rate-limited links, any achievable tradeoff in performance metrics satisfies the following with high probability:

$$T = B = \Omega(K) \quad (19)$$

$$E = K \sum_{n=1}^N \Omega(N^{u_n - \alpha/2}) \quad (20)$$

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^N \Omega(N^{-2Ku_n}), \quad (21)$$

for $u_n > 0$ and $K > 0$. In particular, choosing each $u_n = u$ yields

$$E = \Omega(KN^{1+u-\alpha/2}) \quad (22)$$

$$\sigma^2 = \Omega(N^{-2Ku}). \quad (23)$$

Proof: As in the ideal-link case, the bounds on T and B are trivial; at least one transmission must be made, which requires K channel uses. To bound the tradeoff between energy and estimation error, momentarily consider a single node n . Suppose a genie supplies node n with z_{ave} , and further suppose that only node n 's nearest neighbor, denoted by m , needs to compute the average. In this case, the optimal strategy is for n to quantize z_{ave} and transmit it directly to node m . In principle, other nodes could transmit their measurements to m , but since they are no closer order-wise, and since they have only partial knowledge of the average, any energy they expend would be better used by node n .

Without loss of generality, let $P_n = N^{u_n - \alpha/2}$ denote the transmit power used by node n to transmit z_{ave} . Since again the distance between nearest neighbors is $\Theta(N^{-1/2})$, then with high probability the highest possible SNR is $\gamma = \Theta(N^{u_n})$. Therefore, $L = (1 + \gamma)^K = \Theta(N^{Ku_n})$. The square quantization error at node n on z_{ave} is $|e_n|^2 = \Theta(L^{-2}) = \Theta(N^{-2Ku_n})$. Repeating the argument for each n establishes the result. ■

By contrast with the ideal-link case, with quantization nodes can trade off between resources and quantization error by adjusting both K and the transmit power. In particular, observe that choosing high K reduces the exponent of the estimation error, but only increases the resource consumption by a constant factor. To highlight this tradeoff, we retain K in the scaling laws for quantization, whereas with ideal links we omit K .

Also note that the scaling laws are more favorable for higher path-loss exponents, which may give the false impression that increased path-loss *decreases* the resources required. So long as G is chosen such that channel gains are less than unity, the absolute resource consumption required for fixed N increases with α . The scaling laws merely indicate that increasing the number of nodes—and therefore decreasing the distance between nearest neighbors—is more beneficial for high path-loss. Of course, for sufficiently large N the far-field assumption on which path-loss is based breaks down, at which point the scaling laws no longer hold. For a discussion of this issue in capacity scaling laws, see [27], [28].

IV. GOSSIP ALGORITHMS

In this section we characterize several existing gossip algorithms with respect to the metrics defined in Section II. We focus on two variants that give a relatively comprehensive look at the state of the art: randomized gossip [7], which is probably the best-known approach to gossip, and path averaging [8], which is order optimal in terms of convergence speed³. Our first task is to adapt the graphical nature of gossip to our wireless model. In order to achieve consensus, it is necessary to choose the topology of the network such that the resulting graph is connected, which requires that each node be neighbors with every node within a radius of $\Theta(\sqrt{\log N/N})$ [10]. Therefore the neighborhoods must satisfy, for every node n that transmits during time slot t ,

$$\mathcal{N}_n(t) = \{m : \|\mathbf{x}_m - \mathbf{x}_n\|_2 < \Theta(\sqrt{\log N/N})\}.$$

By (3), the transmit power is

$$P_n(t) = \Theta((\log N/N)^{\alpha/2}), \quad (24)$$

for every node n transmitting during time slot t . This holds for both gossip algorithms considered in this section.

³Due to its similarity with hierarchical averaging, we might suspect that multiscale gossip [9] has superior performance to the gossip algorithms studied here with respect to our metrics. However, multiscale gossip differs from hierarchical averaging in the crucial sense that it operates using fixed connectivity. As a result, while we do not carry out the analysis here due to space constraints, one can show that the performance of multi-scale gossip is similar to that of path-averaging.

A. Randomized Gossip

We study the *synchronized* randomized gossip of [7]. At each time slot t , each node is randomly paired up with one of its neighbors. Paired nodes exchange estimates and average the estimates together, which results in the following dynamics:

$$\mathbf{z}(t) = \frac{1}{2}(\mathbf{W}(t) + \mathbf{I})\mathbf{z}(t-1),$$

where $\mathbf{W}(t)$ is a randomly-chosen permutation matrix such that $w_{mn} = 1$ only if nodes m and n are neighbors.

In [7, Theorem 9] the convergence of randomized gossip is characterized. It is shown that the averaging time satisfies

$$T_\epsilon = \Theta\left(N \frac{\log \epsilon^{-1}}{\log N}\right). \quad (25)$$

Using this fact, we derive bounds on the resource consumption of randomized gossip.

Theorem 3: For randomized gossip, the resource consumption scales as follows with high probability:

$$T_\epsilon = \Theta\left(N \frac{\log \epsilon^{-1}}{\log N}\right), \quad (26)$$

$$B_\epsilon = \Omega(N \log \epsilon^{-1}), \quad (27)$$

$$E_\epsilon = \Theta(N^{2-\alpha/2}(\log N)^{\alpha/2-1} \log \epsilon^{-1}). \quad (28)$$

Proof: The bound on T_ϵ follows from (25). Since every node transmits during every time slot t , $P_n(t) = \Theta((\log N/N)^{\alpha/2}) \forall n, t$. The transmit energy therefore follows

$$\begin{aligned} E_\epsilon &= K \sum_{t=1}^{T_\epsilon} \sum_{n=1}^N \Theta\left(\frac{\log N}{N}\right)^{\alpha/2} \\ &= T_\epsilon \Theta(N^{1-\alpha/2}(\log N)^{\alpha/2}) \\ &= \Theta(N^{2-\alpha/2}(\log N)^{\alpha/2-1} \log \epsilon^{-1}). \end{aligned}$$

Next, the required connectivity radius means that each neighborhood is defined by a region of area $\Theta(\log N/N)$. By Lemma 1, each neighborhood size satisfies

$$\begin{aligned} (1 - \delta)\pi \log N &\leq |\mathcal{N}_n(t)| \leq (1 + \delta)\pi \log N \\ |\mathcal{N}_n(t)| &= \Theta(\log N) \end{aligned}$$

with high probability. Plugging this into (12) yields

$$B_\epsilon = KT_\epsilon \Omega(\log N) = \Omega(N \log \epsilon^{-1}). \quad \blacksquare$$

B. Path Averaging

Next, we examine *path averaging* [8]. Instead of exchanging estimates with a neighbor, in path averaging each node chooses a geographically distant node with which to exchange its estimate; the exchange is facilitated by multi-hop rounding. In addition to facilitating the exchange, the routing nodes add their estimates “along the way,” allowing many nodes to average together in a single round. Once the average of all the nodes’ estimates is computed at the destination, the result is routed back to the source.

Path averaging is described in an asynchronous framework in which nodes independently “wake up,” initiate multi-hop

exchanges, and return to idle state sufficiently quickly that no two exchanges overlap in time. Placing path averaging into our synchronous framework, we suppose that at time t a pair of nodes n, m is randomly selected to engage in a multi-hop exchange. Letting $\mathcal{P}(t)$ be the set of nodes routing from n to m , we suppose that the $2(|\mathcal{P}(t)| - 1)$ transmissions required to route from n to m and back happen sequentially and thus require $2(|\mathcal{P}(t)| - 1)$ time slots. At time slot $t + 2(|\mathcal{P}(t)| - 1)$, a new pair is chosen. The dynamics for path averaging has the following form:

$$z_n(t + 2(|\mathcal{P}(t)| - 1)) = \begin{cases} \frac{1}{|\mathcal{P}(t)|} \sum_{m \in \mathcal{P}(t)} z_m(t), & n \in \mathcal{P}(t) \\ z_n(t), & \text{otherwise} \end{cases}. \quad (29)$$

In [8, Theorem 2] it is shown that, for a random uniform network,⁴ the expected path length is $E[|\mathcal{P}(t)|] = \Theta(\sqrt{N/\log N})$ and the number of exchanges required to achieve ϵ -consensus is $\Theta(\sqrt{N \log N} \log \epsilon^{-1})$. Combining these facts, the total number of required transmissions is $\Theta(N \log \epsilon^{-1})$.

In casting path averaging in our synchronous framework, we have retained the assumption that multi-hop exchanges do not overlap in time. In principle one could construct a synchronous path-averaging gossip in which multiple exchanges occur simultaneously, perhaps reducing the total amount of time required to achieve consensus. In the following theorem, we provide a rather optimistic bound on the resource consumption of any such synchronous formulation.

Theorem 4: For any synchronous path-averaging gossip, the resource consumption scales as follows with high probability:

$$T_\epsilon = B_\epsilon = \Omega\left(\sqrt{\frac{N}{\log N}}\right) \quad (30)$$

$$E_\epsilon = \Theta(N^{1-\alpha/2} (\log N)^{\alpha/2} \log \epsilon^{-1}). \quad (31)$$

Proof: We prove the bound on T_ϵ and B_ϵ by noting that each route has $\Theta(\sqrt{N/\log N})$ hops. Even in the ideal case in which every round of gossip occurs simultaneously, we still require $T_\epsilon = \Omega(\sqrt{N/\log N})$ sequential transmissions. Supposing that constant bandwidth is sufficient to accommodate the multiple exchanges, the same bound applies to B_ϵ .

To bound E_ϵ we point out that, as with randomized gossip, $P_n(t) = \Theta((\log N/N)^{\alpha/2})$ for every transmission. Since path-averaging requires $\Theta(N \log \epsilon^{-1})$ transmissions, the overall energy consumption scales as

$$E_\epsilon = \Theta(N^{1-\alpha/2} (\log N)^{\alpha/2} \log \epsilon^{-1}).$$

V. HIERARCHICAL AVERAGING

In this section we present *hierarchical averaging*. Much like multi-scale gossip [9] and the hierarchical cooperation of [11], in hierarchical averaging we recursively partition the network into geographically defined clusters. Each cluster achieves

⁴Technically, the convergence speed of path averaging is proven over a torus, so the results we prove in the sequel apply to the torus. Later we show numerical results that suggest that the similar results apply to a square network.

internal consensus by mutually broadcasting estimates. Nodes within a cluster then cooperatively broadcast their identical estimates to neighboring clusters at the next level. The process continues until the entire network achieves consensus. In the following subsection we describe the recursive partitioning, after which we describe the algorithm in detail and characterize its resource requirements.

A. Hierarchical Partitioning

We partition the network into T sub-network layers, one for each round of consensus, as depicted in Figure 1. At the top layer, which corresponds to the final round $t = T$ of consensus, there is a single cell. At the next-highest level $t = T - 1$, we divide the network into four equal-area square cells. Continuing, we recursively divide each cell into four smaller cells until the lowest layer $t = 1$, which corresponds to the first round of consensus. At each level t there are 4^{T-t} cells, formally defined as

$$\mathcal{C}_{jk}(t) = \{n : \mathbf{r}_n \in [(j-1)2^{t-T}, j2^{t-T}) \times ((k-1)2^{t-T}, k2^{t-T})\}, \quad (32)$$

where $1 \leq j, k \leq 2^{T-t}$ index the location of the cell.

Let $\mathcal{C}(n, t)$ denote the unique cluster at layer t containing n . Let $M(t) = \sqrt{2} \cdot 4^{\frac{t-T}{2}} = \Theta(4^{\frac{t-T}{2}})$ denote the maximum distance between two nodes in the same cluster at layer t .

We choose $T = \lceil \log_4(N^{1-\kappa}) \rceil$, where $0 < \kappa < 1$ is a small constant. In the following lemma we bound the cardinality of each cell.

Lemma 2: For every $1 \leq j, k \leq 2^{T-t}$ and $1 \leq t \leq T$, the cluster cardinalities simultaneously satisfy

$$|\mathcal{C}_{jk}(t)| = \Theta(4^t N^\kappa), \quad (33)$$

with probability greater than $1 - N^{2-2\kappa}/16 \cdot e^{-\Gamma(\delta)N^\kappa}$.

Proof: By construction, the area of each cell for $t = 1$ is

$$\begin{aligned} A &= 4^{1-T} \\ \implies 4^{1-\log_4(N^{1-\kappa})-1} &\leq A \leq 4^{1-\log_4(N^{1-\kappa})} \\ \frac{N^{\kappa-1}}{16} &\leq A \leq 4N^{\kappa-1}. \end{aligned}$$

By Lemma 1, the cardinality of each cell at layer $t = 1$ satisfies

$$(1 - \delta) \frac{N^\kappa}{16} \leq |\mathcal{C}_{jk}(t)| \leq (1 + \delta) 4N^\kappa, \quad (34)$$

with probability greater than $1 - N^{1-\kappa}/16 \cdot e^{-\Gamma(\delta)N^\kappa}$.

Define $E_{jk}(1)$ as the event in which $|\mathcal{C}_{jk}(1)|$ is outside the bounds specified in (34). Clearly $\Pr\{E_{jk}(1)\} \leq N^{1-\kappa}/16 \cdot e^{-\Gamma(\delta)N^\kappa}$. By the union bound, the total probability follows

$$\Pr\left(\bigcup_{1 \leq j, k \leq 2^{T-1}} E_{j,k}(1)\right) \leq \sum_{1 \leq j, k \leq 2^{T-1}} N^{1-\kappa}/16 \cdot e^{-\Gamma(\delta)N^\kappa} \leq N^{2-2\kappa}/16 \cdot e^{-\Gamma(\delta)N^\kappa} \rightarrow 0. \quad (35)$$

Therefore, every cell at $t = 1$ simultaneously satisfies $|\mathcal{C}_{jk}(1)| = \Theta(N^\kappa)$ with the desired probability. Finally, since each cell at layer t is composed of 4^{t-1} cells at layer 1,

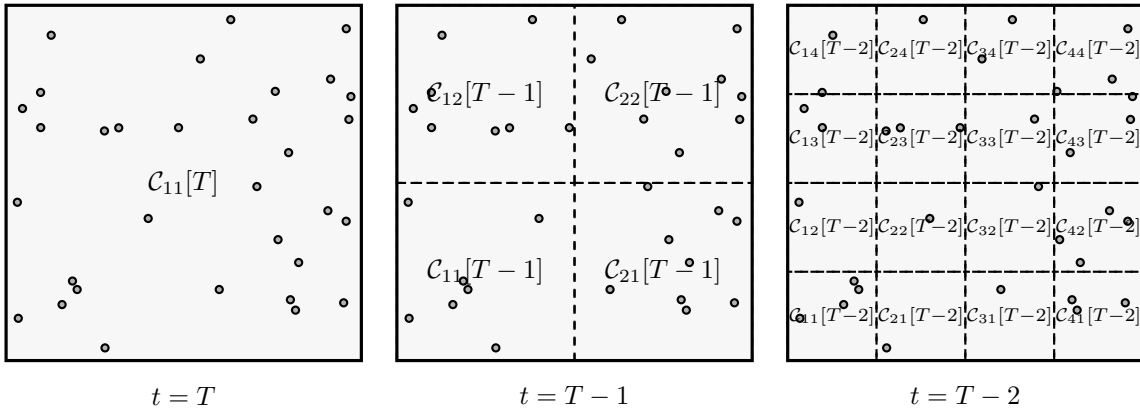


Fig. 1. Hierarchical partition of the network. Each square cell is divided into four smaller cells, which are each divided into four smaller cells, and so on.

the cardinalities at each layer satisfy the claim with high probability. ■

B. Algorithm Description

First, at time slot $t = 1$ each node broadcasts its initial estimate $z_n(0)$ to each member of its cluster $\mathcal{C}(n, t)$. In order to ensure that $n \in \mathcal{N}_m(t)$ for every $m \in \mathcal{C}(n, 1)$, each node transmits at power

$$P_n(1) = \frac{\gamma}{G} \max_{m \in \mathcal{C}(n, 1)} h_{nm}^\alpha \leq \frac{\gamma}{G} M(1)^\alpha = O(N^{(\kappa-1)\alpha/2}). \quad (36)$$

Each node averages the estimates in its cluster:

$$z_n(1) = \frac{1}{4^{1-T}N} \sum_{m \in \mathcal{C}(n, 1)} z_m(0). \quad (37)$$

We use the approximate normalization $1/4^{1-T}N$ instead of the exact factor $1/|\mathcal{C}(n, 1)|$ so that nodes at higher levels of the hierarchy need not know the cardinality of the cells. This approximation introduces no error into the final estimate.

After time slot $t = 1$, each node in each cluster $\mathcal{C}_{jk}(1)$ has the same estimate, which we denote by $z_{\mathcal{C}_{jk}(1)}(1)$. At each subsequent time slot $2 \leq t \leq T$, each cluster $\mathcal{C}(n, t-1)$ cooperatively transmits its estimate to its parent cluster at layer t . We take each $P_n(t)$ to be a constant. The transmit power required depends on the phase of the channel gains, as discussed in Section II-A. When the phases are fixed and identical, by (4) the transmit powers must satisfy

$$\left(\sum_{m \in \mathcal{C}(n, t-1)} h_{mn} P_m^{1/2}(t) \right)^2 = \frac{\gamma}{G},$$

which implies

$$\begin{aligned} P_m(t) &= \frac{\gamma}{G \left(\sum_{m \in \mathcal{C}(n, t-1)} h_{mn} \right)^2} \\ &\leq \frac{\gamma M(t)^\alpha}{G |\mathcal{C}(n, t-1)|^2} \\ &= O \left(4^{(\alpha/2-2)t} N^{-\alpha/2+\kappa(\alpha/2-2)} \right). \end{aligned} \quad (38)$$

When the phases are random and uniform, on the other hand, by (5) a similar argument shows

$$\begin{aligned} P_m(t) &= \frac{\gamma}{G \sum_{m \in \mathcal{C}(n, t-1)} h_{mn}^2} \\ &\leq \frac{\gamma M(t)^\alpha}{G |\mathcal{C}(n, t-1)|} \\ &= O \left(4^{(\alpha/2-1)t} N^{(\kappa-1)\alpha/2} \right). \end{aligned} \quad (39)$$

After receiving estimates from the other sub-clusters, each node updates its estimate by taking the sum:

$$\begin{aligned} z_n(t) &= \frac{1}{4} \sum_{\mathcal{C}(n, t-1) \subset \mathcal{C}(n, t)} z_{\mathcal{C}(n, t-1)} \\ &= \frac{1}{4^{t-T}N} \sum_{m \in \mathcal{C}(n, t)} z_m(0), \end{aligned} \quad (40)$$

where the second equality follows straightforwardly by induction. At time t , the identical estimate at each cluster is the average of the measurements from within that cluster.

Consensus is achieved at round T , where the four sub-clusters at level $t = T-1$ broadcast their estimates to the entire network. Evaluating (40) for $t = T$, we observe that hierarchical averaging achieves perfect consensus; there is no need for a tolerance parameter ϵ . This somewhat surprising result is the consequence of combining the flexibility of wireless, which allows us to adjust the network connectivity at will, with the assumption of ideal links. In the next section we will revisit this assumption.

In the following theorem we derive the resource requirements of hierarchical averaging.

Theorem 5: With high probability, the resource consumption of hierarchical averaging scales according to

$$T_\epsilon = B_\epsilon = O(N^\kappa), \quad (41)$$

$$E_\epsilon = \begin{cases} O(N^{1-\alpha/2+\kappa\alpha/2}), & \text{for fixed phase} \\ O(N^{\kappa\alpha/2}), & \text{for uniform phase} \end{cases}, \quad (42)$$

for any path-loss exponent $2 \leq \alpha < 4$, for any $\epsilon > 0$ and for any $0 < \kappa \leq 1$.

Proof: The bound on T_ϵ follows by construction; we chose $T = \lceil \log_4 N^{1-\kappa} \rceil = O(N^\kappa)$ layers of hierarchy and

constructed the algorithm such that consensus is achieved to within any tolerance $\epsilon > 0$,

We derive the bound on B_ϵ by examining the number of two-hop neighbors for each node. At time slot $t = 1$, by (36) each node transmits at power $P_n(1) = O(N^{(\kappa-1)\alpha/2})$. The neighbors are nodes within a radius of $O(N^{\kappa-1})$, and the two-hop neighbors are the nodes within twice the radius. Therefore, by Lemma 1, the number of two-hop neighbors for each node scales as $O(N^\kappa)$ with probability approaching 1 as $N \rightarrow \infty$. By (8), $B(1) = O(N^\kappa)$.

For rounds $2 \leq t \leq T$, we need to bound the number of clusters within two hops of each node. In (38) we chose the transmit powers such that the clusters transmit to each node in a circle of area $\pi M^2(t) = O(4^t N^{1-\kappa})$. By construction, each cluster $\mathcal{C}(n, t)$ covers an area of $O(4^t N^{1-\kappa})$. The number of clusters that can fit into the circle—and therefore the number of one-hop neighbors—is a constant. The number of two-hop neighbors, bounded above by the square of the number of one-hop neighbors, also remains constant, so $B(t) = O(1)$. Summing over all rounds, we get

$$B_\epsilon = K \sum_{t=1}^T B(t) = O(N^\kappa) + \sum_{t=2}^T O(1) = O(N^\kappa). \quad (43)$$

Finally, we derive the bounds on E_ϵ . For fixed phase, (36) and (38) imply that the transmit energy follows

$$\begin{aligned} E_\epsilon &= N \cdot O(N^{(\kappa-1)\alpha/2}) + N \sum_{t=2}^T O\left(4^{(\alpha/2-2)t} N^{-\alpha/2+\kappa(\alpha/2-2)}\right) \\ &\stackrel{(a)}{=} O(N^{1-\alpha/2+\kappa\alpha/2}) + O\left(N^{1-\alpha/2+\kappa(\alpha/2-2)} \frac{1-4^{(\alpha/2-2)T}}{1-4^{\alpha/2-2}}\right) \\ &\stackrel{(b)}{=} O(N^{1-\alpha/2+\kappa\alpha/2}) + O\left(N^{1-\alpha/2+\kappa(\alpha/2-2)} N^{(\alpha/2-2)(1-\kappa)}\right) \\ &= O(N^{1-\alpha/2+\kappa\alpha/2}), \end{aligned} \quad (44)$$

where (a) follows from the finite geometric sum identity, and (b) holds only when $\alpha < 4$. For uniform phase, first note that the condition in (39) is more strict than that of (36), so the transmit power satisfies

$$P_n(t) = O\left(4^{(\alpha/2-1)t} N^{(\kappa-1)\alpha/2}\right) \quad (45)$$

for all n, t . Substituting (45) into the definition of E_ϵ yields

$$\begin{aligned} E_\epsilon &= K \sum_{t=1}^T \sum_{n=1}^N P_n(t) \\ &= O\left(N^{1+(\kappa-1)\alpha/2} \sum_{t=0}^{T-1} 4^{(\alpha/2-1)t}\right) \\ &= O(N^{\kappa\alpha/2}), \end{aligned} \quad (46)$$

where again we have employed the finite geometric sum identity. ■

Hierarchical averaging achieves resource scaling arbitrarily close to the lower bound of Theorem 1 when phase is fixed. When phase is uniform, however, the energy consumption is strictly suboptimal for $\alpha > 2$. Note that resource scaling does not depend on the channel phases for $\alpha = 2$. In this case, hierarchical averaging is order optimal regardless of phase.

C. Numerical Results

We examine the empirical performance of the several consensus algorithms presented. We choose $\gamma = 10\text{dB}$, $\alpha = 4$, $\epsilon = 10^{-4}$, $\kappa = 10^{-4}$, $K = 10$, and $G = 10^{-3\alpha/2}$. We let N run from 10 to 1000, averaging performance over 50 random initializations for each value of N . Figure 2 shows the average transmit energy E_ϵ and time-bandwidth product B_ϵ . (Since the data for T_ϵ are rather similar to that of B_ϵ , we do not plot them.)

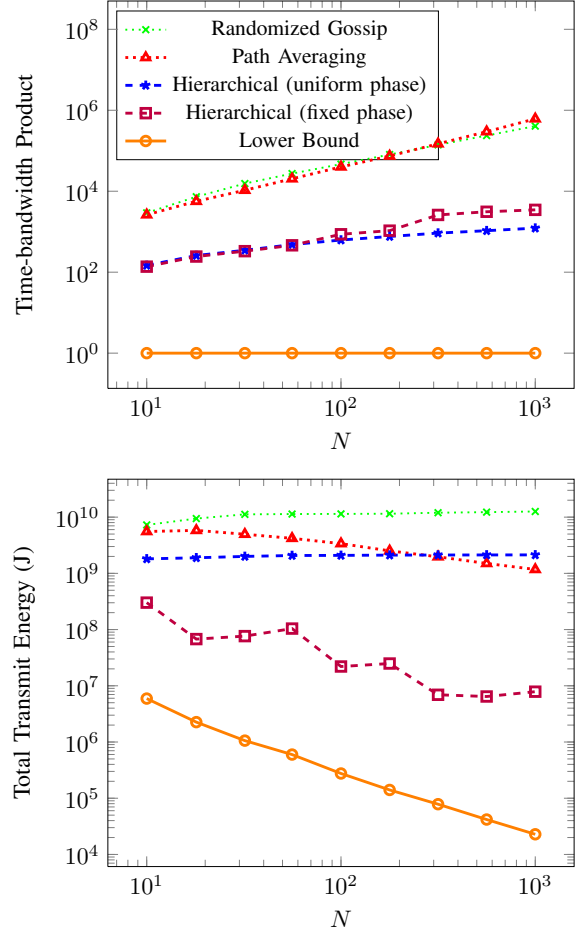


Fig. 2. Time-bandwidth product B_ϵ and transmit energy E_ϵ as a function of N .

With respect to time-bandwidth product, hierarchical averaging performs best, the required number of sub-channel uses growing slowly with N . The remaining two consensus schemes perform comparably, the required number of sub-channels growing approximately linearly in N . Note that, while we bounded the time-bandwidth product of path averaging with a strictly sub-linear term, this bound applied to hypothetical instantiations of the scheme in which multiple transmissions occur simultaneously. Our simulations used the ordinary algorithm, which requires $\Theta(N)$ sub-channel uses.

With respect to total transmit energy, hierarchical averaging performs best so long as the phases are fixed, in which case the performance is on par order-wise with the lower bound; the oscillations in energy are due to rounding $\log_4(N^{1-\kappa})$ in order to choose T . When phases are uniform, however,

performance depends on N . Even though path averaging has better scaling than hierarchical averaging under uniform phase, for small N hierarchical averaging requires less power. Finally, as expected, randomized gossip requires the most energy in any regime.

VI. QUANTIZATION

In this section we examine consensus with quantization. As in the case with ideal links, we first characterize the performance of existing quantized consensus algorithms with respect to the metrics specified in Section II-C. We cannot survey every approach in the literature, so we focus on the *quantized consensus* of [20], in which consensus is modified to preserve the average of quantized estimates each round. After deriving bounds on its performance, we turn to hierarchical averaging. We show that it achieves the lower bound of Theorem 2 when phases are fixed.

A. Quantized Consensus

In ordinary gossip, the primary difficulty of quantization is that quantizing estimates in general alters the average across the network. Thus, even if consensus is achieved, the dynamics will not in general converge on the true average of the (quantized) measurements. In *quantized consensus* [20], a family of consensus algorithms is proposed that preserves the average at each round; it converges to near-consensus around the true average.

Recall from Section II-C that \mathcal{Z} is the set of L points evenly distributed across $[0, 1)$, separated by quantization bin width $\Delta = 1/L$. Quantized consensus operates only on quantized values, so first we must quantize the real-valued measurements $z_n(0)$:

$$q_n(0) = \phi(z_n(0)), \quad (47)$$

where ϕ is the dithered quantizer described in Section II-C. Let $e_n(0) = \phi(z_n(0)) - z_n(0)$ denote the quantization error.

Much like in randomized gossip, at each round every node randomly selects a neighboring node and mutually averages, with the caveat that one node rounds “up” to the nearest member of \mathcal{Z} while the other rounds “down.” Letting i and j denote the two nodes in the exchange, we have⁵

$$q_i(t) = \left\lceil \frac{q_i(t-1) + q_j(t-1)}{2} \right\rceil_{\mathcal{Z}} \quad (48)$$

$$q_j(t) = \left\lfloor \frac{q_i(t-1) + q_j(t-1)}{2} \right\rfloor_{\mathcal{Z}}, \quad (49)$$

where $\lceil \cdot \rceil_{\mathcal{Z}}$ and $\lfloor \cdot \rfloor_{\mathcal{Z}}$ represent rounding up and down to the nearest element of \mathcal{Z} , respectively. In [20, Theorem 1] This algorithm is guaranteed to converge on near-consensus: in the limit, each $q_n(t)$ differs by at most a single bin, and the sum of the quantized measurements is preserved. It is difficult to bound the convergence speed of this process in general due to the non-linearity of the updates. However, for the case of a fully-connected graph, in [20, Lemma 6] it is shown

⁵In fact, [20] proposes a family of algorithms, and the one we use here is only one possibility. The convergence properties we exploit in the following are independent of the specific algorithm chosen.

that quantized consensus requires $\Omega(N^2)$ transmissions over $\Omega(N)$ consensus rounds. Using this fact, we bound the overall performance.

Theorem 6: The performance of quantized gossip scales, with high probability, as

$$T = \Omega(KN), \quad (50)$$

$$B = \Omega(KN \log N), \quad (51)$$

$$E = \Omega(KN^{2-\alpha/2+u}(\log N)^{\alpha/2}), \text{ and} \quad (52)$$

$$\sigma^2 = \Omega(N^{-2Ku-1}), \quad (53)$$

for any $u \geq 0$.

Proof: Choose a target SNR γ , which may vary with N . In order to maintain connectivity, links need to maintain γ at radius $\sqrt{\log N/N}$, which requires

$$P_n(t) = \Theta \left(\gamma \left(\frac{\log N}{N} \right)^{\alpha/2} \right). \quad (54)$$

By [20, Lemma 6], consensus requires $\Omega(N)$ rounds for fully-connected graphs, and the performance for random graphs cannot be any better. As in the proof of unquantized randomized gossip, the neighborhood size scales as $\Theta(\log N)$, so the time-bandwidth product scales as

$$B = \Omega(KN \log N). \quad (55)$$

Since consensus requires $\Omega(KN^2)$ total transmissions, the energy scales as

$$E = \Omega(K\gamma N^{2-\alpha/2}(\log N)^{\alpha/2}). \quad (56)$$

Finally, we examine the mean-squared error. In the best case, the dynamics converge on true consensus, meaning that $q_n(T)$ is the same for each n . In this case the final estimates are merely the average of the quantized measurements $z_n(0)$. Therefore the final estimates are

$$q_n(T) = \frac{1}{N} \sum_{n=1}^N q_n(0) = z_{\text{ave}} - \frac{1}{N} \sum_{n=1}^N e_n(0),$$

where $e_n(0)$ is the quantization error of the initial estimate. In the worst case, each $|e_n(0)| = \Delta/2 = L^{-1}/2$. Since the errors are uncorrelated, the squared error follows

$$\begin{aligned} \sigma^2 &= E \left[\left(\frac{1}{N} \sum_{n=1}^N e_n(0) \right)^2 \right] \\ &\geq \frac{1}{N^2} \sum_{n=1}^N E[|e_n(0)|^2] = \Omega(N^{-1}L^{-2}). \end{aligned}$$

Recall that $L = \lfloor (1 + \gamma)^K \rfloor$. Choosing $\gamma = N^u$ yields the result. ■

We hasten to point out that the bounds here are rather generous, since we supposed that convergence on a random graph is as fast as on a fully-connected graph. In practice, as we shall see in the numerical results presented later, the performance is somewhat worse.

B. Hierarchical Averaging

We characterize the performance of hierarchical averaging with quantization. As before, cells of nodes at lower layers achieve local consensus, after which they broadcast their estimates to nearby clusters, continuing the process until global consensus is achieved. Here, however, each estimate is quantized prior to transmission, which introduces error that accumulates during consensus.

We choose $T = \lceil \log_4 N^{1-\kappa} \rceil$ and define the cells $\mathcal{C}_{jk}(t)$ as before. At time slot $t = 1$, each node n quantizes its initial measurement $z_n(0)$ and broadcasts the quantized value to the nodes in $\mathcal{C}(n, 1)$. Following (36), this requires

$$P_n(1) = O(\gamma N^{(\kappa-1)\alpha/2}), \quad (57)$$

Each node n updates its estimate by averaging the quantized estimates in its cluster:

$$z_n(1) = \frac{1}{4^{1-T}N} \sum_{m \in \mathcal{C}(n,1)} z_m(0) + v_m(0). \quad (58)$$

As before we use the normalization factor $1/4^{1-T}N$ in order to avoid nodes' needing to know the cluster cardinalities. Next, at time slot $2 \leq t \leq T$, each cluster at layer $t-1$ quantizes its estimate and cooperatively broadcasts to the members of its parent cluster at layer t . Following (38) and (39), this requires

$$P_n(t) = \begin{cases} O(\gamma 4^{(\alpha/2-2)t} N^{-\alpha/2+\kappa(\alpha/2-2)}), & \text{fixed phases} \\ O(\gamma 4^{(\alpha/2-1)t} N^{(\kappa-1)\alpha/2}), & \text{uniform phases} \end{cases} \quad (59)$$

At time step $t > 1$, each node n averages together the estimates from each of the subclusters $\mathcal{C}(m, t-1) \subset \mathcal{C}(n, t)$. By induction, this estimate is

$$z_n(t) = \frac{1}{4^{t-T}N} \sum_{k \in \mathcal{C}(n,t)} (z_k(0) + v_k(0)) + \sum_{s=1}^{t-1} \sum_{M \in R_n(t,s)} 4^{s-t} v_M(s), \quad (60)$$

where $R_n(t, s)$ is the set of all clusters $\mathcal{C}(m, s)$ that are subsets of $\mathcal{C}(n, t)$. In other words, at round t we have the total average so far, corrupted by quantization noise from each of the rounds $s < t$. In the following theorem, we detail the resource-estimate tradeoff achieved by this scheme.

Theorem 7: Using dithered quantization, hierarchical averaging achieves the following tradeoff between resource consumption and estimation error with high probability:

$$T = B = O(KN^\kappa) \quad (61)$$

$$E = \begin{cases} O(KN^{1-\alpha/2+\kappa\alpha/2+u}), & \text{for fixed phases} \\ O(KN^{\kappa\alpha/2+u}), & \text{for uniform phases} \end{cases} \quad (62)$$

$$\sigma^2 = \Theta(N^{-2Ku}), \quad (63)$$

for any $u \geq 0$, $\kappa > 0$, and $2 \leq \alpha < 4$. In particular, for $u = 0$ the estimation error is constant in the network size using the same amount of energy as in the non-quantized case.

Proof: Choose an SNR γ , which may vary with N . Since the number of rounds and the cluster geometry is unchanged from the non-quantized case, we can repeat the argument from

Theorem 5, yielding $T = B = O(KN^\kappa)$. Repeating the arguments from Theorem 5, which yields

$$E = \begin{cases} O(\gamma N^{1-\alpha/2+\kappa\alpha/2}), & \text{for fixed phases} \\ O(\gamma N^{\kappa\alpha/2}), & \text{for uniform phases} \end{cases}.$$

All that remains is to bound the estimation error. Evaluating (60) for $t = T$, we get, for every n

$$\begin{aligned} z_n(T) &= \frac{1}{N} \sum_{k=1}^N (z_k(0) + v_k(0)) + \sum_{s=1}^{T-1} \sum_{M \in R_n(t,s)} 4^{s-T} v_M(s) \\ &= z_{\text{ave}} + \sum_{s=0}^{T-1} \sum_{M \in R_n(t,s)} 4^{s-T} v_M(s). \end{aligned}$$

The mean squared estimation error is therefore

$$\begin{aligned} \sigma^2 &= E \left[\sum_{s=0}^{T-1} \sum_{M \in R_n(t,s)} 4^{s-T} v_M(s) \right]^2 \\ &= \sum_{s=0}^{T-1} \sum_{M \in R_n(t,s)} 4^{2(s-T)} E[|v_M(s)|^2], \end{aligned}$$

where the equality is due to the independence of the quantization error terms. Since each $v_M(s)$ is uniformly distributed across $[-\Delta, \Delta)$, we have $E[|v_M(s)|^2] = \Theta(\Delta^2) = \Theta(L^{-2})$. Therefore, we have

$$\begin{aligned} \sigma^2 &= \Theta(L^{-2}) \sum_{s=0}^{T-1} \sum_{M \in R_n(t,s)} 4^{2(s-T)} \\ &= \Theta(L^{-2}) \sum_{s=0}^{T-1} 4^{T-s} 4^{2(s-T)} \\ &= \Theta(L^{-2} N^{-1}) \frac{1-4^T}{1-4} = \Theta(L^{-2}), \end{aligned}$$

since $4^T = \Theta(N)$. Recalling that $L = \Theta(\gamma^K)$, and choosing $\gamma = N^u$, yields the result. ■

C. Numerical Results

We examine the empirical performance of the quantized consensus discussed. We also run simulations for randomized gossip, employing dithered quantization to accommodate the finite-rate links. We again choose $\gamma = 10\text{dB}$, $\kappa = 10^{-4}$, $K = 10$, and $G = 10^{-3\alpha/2}$, and we again let N run from 10 to 1000 and average performance over 50 initializations, but here we choose $\alpha = 2$. Choosing γ constant means that the quantization error Δ is constant in N , and the minimum quantization error is itself constant. In Figure 3 we plot the energy E against the mean-square error σ^2 .

The energy expenditure for hierarchical averaging is consistent with theory, although we note that uniform phase results in higher expenditure than fixed phase, even though the scaling laws are the same. This is due to the coherence gain obtained when phases are fixed. The energy expenditure for randomized gossip increases roughly linearly in N , suggesting that the energy burden with fixed γ is similar to the non-quantized case. As expected, quantized consensus performs worse than

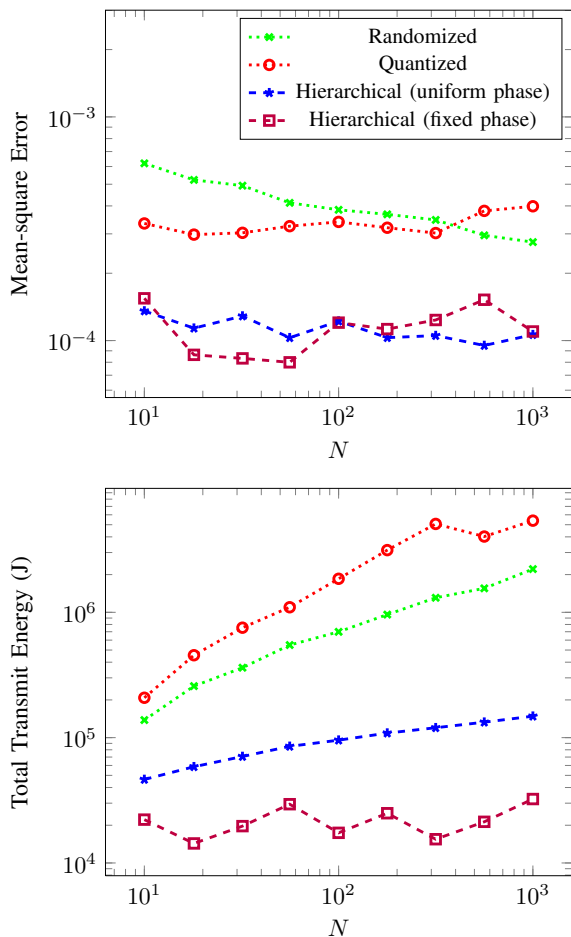


Fig. 3. Mean-square error σ^2 and total energy E as a function of N .

predicted by Theorem 6 in terms of energy consumption. Since γ is chosen as a constant, the mean-squared error remains approximately constant for each scheme.

VII. CONCLUSION

We have studied consensus from an explicitly wireless perspective, defining a realistic propagation environment, confronting interference, and characterizing the required resource consumption in terms of energy, time and bandwidth. For dense networks and under path-loss propagation, we have shown that consensus is *cheap*. In the worst case, the resource costs for consensus are nearly constant in the size of the network. Furthermore, we have shown that quantization is only a minor issue in consensus. By increasing the block length of transmissions, one can force the quantization error to decay as quickly as desired while paying only a constant penalty in energy, time, and bandwidth. We have shown that these gains can be realized with practical algorithms, such as hierarchical averaging, which is nearly order-optimal with respect to resource consumption for many environments.

Nevertheless, our wireless model involves several simplifying assumptions. We supposed synchronous transmission, that out-of-range nodes do not contribute interference, and that channel gains are determined entirely by path-loss. Future

work includes exploring the effects of relaxing these assumptions, incorporating medium-access techniques such as CSMA, or characterizing performance in the presence of channel fades. We expect that, since it is cooperative in nature, hierarchical averaging is particularly robust to fading.

Finally, improvements to hierarchical averaging are possible. In the case of uniform phase, hierarchical averaging was suboptimal with respect to transmit energy for $\alpha > 2$, since in this case transmissions do not coherently combine at receivers. In [11], distributed *receiver* cooperation is used to overcome this difficulty and secure order-optimal performance. We expect that a similar approach will suffice for consensus, regardless of channel phases or path-loss exponents.

REFERENCES

- [1] G. Cybenko, "Dynamic load balancing for distributed memory multiprocessors," *Journal of Parallel and Distributed Computing*, vol. 7, no. 2, pp. 279 – 301, 1989. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/074373158990021X>
- [2] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, Nov. 1984.
- [3] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Automatic Control*, vol. 31, no. 9, pp. 803 – 812, Sept. 1986.
- [4] J. Duchi, A. Agarwal, and M. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Automatic Control*, vol. 57, no. 3, pp. 592 – 606, Mar. 2012.
- [5] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection in sensor networks with packet loss and finite capacity links," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4118–4132, Nov. 2006.
- [6] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Approximate distributed Kalman filtering in sensor networks with quantifiable performance," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, 2005. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1147685.1147711>
- [7] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Info. Theory*, vol. 52, no. 6, pp. 2508 – 2503, June 2006.
- [8] F. Benezit, A. Dimakis, P. Thiran, and M. Vetterli, "Order-optimal consensus through randomized path averaging," *Information Theory, IEEE Transactions on*, vol. 56, no. 10, pp. 5150 – 5167, Oct. 2010.
- [9] K. I. Tsianos and M. G. Rabbat, "Multiscale gossip for efficient decentralized averaging in wireless packet networks," *submitted to IEEE Trans. Info. Theory*, 2010.
- [10] P. Gupta and P. Kumar, "Critical power for asymptotic connectivity," in *Proc. Conference on Decision and Control*, vol. 1, 1998, pp. 1106 – 1110.
- [11] A. Ozgur, O. Leveque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Info. Theory*, vol. 53, no. 10, pp. 3549 – 3572, Oct. 2007.
- [12] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1205–1216, March 2008.
- [13] D. Ustebay, B. Oreshkin, M. Coates, and M. Rabbat, "Rates of convergence for greedy gossip with eavesdropping," in *Proc. Allerton*, Monticello, IL, 2008.
- [14] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2748–2761, July 2009.
- [15] B. Nazer, A. G. Dimakis, and M. Gastpar, "Local interference can accelerate gossip algorithms," *IEEE J. Selected Topics Signal Proc.*, vol. 5, no. 4, pp. 876–887, Aug. 2011.
- [16] S. Sardellitti, S. Barbarossa, and A. Swami, "Optimal topology control and power allocation for minimum energy consumption in consensus networks," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 383–399, 2012.
- [17] S. Vanka, V. Gupta, and M. Haenggi, "Power-delay analysis of consensus algorithms on wireless networks with interference," *Int. J. Systems, Control, and Communications*, vol. 2, no. 3, pp. 256–274, 2010.

- [18] S. Vanka, M. Haenggi, and V. Gupta, "Convergence speed of the consensus algorithm with interference and sparse long-range connectivity," *IEEE J. Selected Topics Signal Proc.*, vol. 5, no. 4, pp. 855–865, Aug. 2011.
- [19] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33 – 46, 2007.
- [20] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, no. 7, pp. 1192 – 1203, 2007.
- [21] F. Benezit, P. Thiran, and M. Vetterli, "The distributed multiple voting problem," *IEEE J. Selected Topics Signal Proc.*, vol. 5, no. 4, pp. 791–804, Aug. 2011.
- [22] T. Aysal, M. Coates, and M. Rabbat, "Distributed average consensus with dithered quantization," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4905 –4918, oct. 2008.
- [23] M. Yildiz and A. Scaglione, "Coding with side information for rate-constrained consensus," *IEEE Trans. Signal Processing*, vol. 56, no. 8, pp. 3753 –3764, Aug. 2008.
- [24] R. L. Brooks, "On colouring the nodes of a network," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 37, pp. 194–197, 1941.
- [25] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, Sept. 2004.
- [26] M. Penrose, *Random geometric graphs*. Oxford, UK: Oxford University Press, 2003, vol. 5.
- [27] M. Franceschetti, M. Migliore, and P. Minero, "The capacity of wireless networks: Information-theoretic and physical limits," *IEEE Trans. Info. Theory*, vol. 55, no. 8, pp. 3413 –3424, Aug. 2009.
- [28] S. Lee and S. Chung, "Capacity scaling of wireless ad hoc networks: Shannon meets Maxwell," *IEEE Trans. Info. Theory*, vol. 58, no. 3, pp. 1702–1715, 2012.



Matthew Nokleby (S'04–M'13) received the B.S. (*cum laude*) and M.S. degrees from Brigham Young University, Provo, UT, in 2006 and 2008, and the Ph.D. degree from Rice University in 2012. He is currently a Research Scientist at Duke University, Durham, NC. He received a Texas Instruments Distinguished Fellowship from 2008–2012. His research interests include cooperative communications, network information theory, inference over wireless sensor networks, and information theory for high-dimensional signal processing.



Waheed U. Bajwa (S'98–M'09) received the B.E. degree (with Honors) in electrical engineering from the National University of Sciences and Technology, Pakistan, in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison in 2005 and 2009, respectively. He was a Postdoctoral Research Associate in the Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, from 2009 to 2010, and a Research Scientist in the Department of Electrical and Computer Engineering, Duke

University, Durham, NC, from 2010 to 2011. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering, Rutgers University, New Brunswick, NJ. His research interests include high-dimensional inference and inverse problems, compressed sensing, statistical signal processing, wireless communications, and applications in biological sciences, complex networked systems, and radar and image processing. He has more than three years of industry experience, including a summer position with GE Global Research in 2006.

Dr. Bajwa received the Best in Academics Gold Medal and Presidents Gold Medal in Electrical Engineering from the National University of Sciences and Technology (NUST) in 2001, and the Morgridge Distinguished Graduate Fellowship from the University of Wisconsin-Madison in 2003. He was Junior NUST Student of the Year (2000), Wisconsin Union Poker Series Champion (Spring 2008), President of the University of Wisconsin-Madison chapter of Golden Key International Honor Society (2009), and Rutgers University's nominee for the Packard Fellowship for Science and Engineering Award (2012). He served as a Guest Associate Editor for a special issue on "Compressive Sensing in Communications" of Elsevier's *Physical Communication Journal* (2010/2011). He is currently a member of Golden Key International Honor Society and a co-organizer of the CPSWeek 2013/IPSN Workshop on "Signal Processing Advances in Sensor Networks."



Robert Calderbank Robert Calderbank (M'89–SM'97–F'98) received the B.Sc. degree in 1975 from Warwick University, England, the M.Sc. degree in 1976 from Oxford University, England, and the Ph.D. degree in 1980 from the California Institute of Technology, Pasadena, all in mathematics. He is Dean of Natural Sciences at Duke University, Raleigh, NC. He was previously a Professor of Electrical Engineering and Mathematics, Princeton University, Princeton, NJ, where he directed the Program in Applied and Computational Mathematics.

Prior to joining Princeton University in 2004, he was Vice President for Research at AT&T, responsible for directing the first industrial research lab in the world where the primary focus is data at scale. At the start of his career at Bell Labs, his innovations were incorporated in a progression of voiceband modem standards that moved communications practice close to the Shannon limit. Together with Peter Shor and colleagues at AT&T Labs, he showed that good quantum error correcting codes exist and developed the group theoretic framework for quantum error correction. He is a co-inventor of space-time codes for wireless communication, where correlation of signals across different transmit antennas is the key to reliable transmission.

Dr. Calderbank served as Editor-in-Chief of the *IEEE TRANSACTIONS ON INFORMATION THEORY* from 1995 to 1998, and as Associate Editor for Coding Techniques from 1986 to 1989. He was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and from 2006 to 2008. He was honored by the IEEE Information Theory Prize Paper Award in 1995 for his work on the Z4 linearity of Kerdock and Preparata Codes (joint with A. R. Hammons Jr., P. V. Kumar, N. J. A. Sloane, and P. Sole), and again in 1999 for the invention of space-time codes (with V. Tarokh and N. Seshadri). He received the 2006 IEEE Donald G. Fink Prize Paper Award and the IEEE Millennium Medal, and was elected to the U.S. National Academy of Engineering in 2005.



Behnaam Aazhang (S'82–M'82–SM'91–F99) received his B.S. (with highest honors), M.S., and Ph.D. degrees in Electrical and Computer Engineering from University of Illinois at Urbana-Champaign in 1981, 1983, and 1986, respectively. From 1981 to 1985, he was a Research Assistant in the Coordinated Science Laboratory, University of Illinois. In August 1985, he joined the faculty of Rice University, Houston, Texas, where he is now the J.S. Abercrombie Professor, and Chair of the Department of Electrical and Computer Engineering. In addition,

he holds an Academy of Finland Distinguished Visiting Professorship appointment (FiDiPro) at the Center for Wireless Communication (CWC) in the University of Oulu, Oulu, Finland. He has served as the founding director of Rice's Center for Multimedia Communications from 1998 till 2006. His research interests are in the areas of communication theory, information theory, signal processing, and their applications to wireless communication, wireless networks, and neuro-engineering with emphasis on closed-loop real-time neuromodulation and brain stimulation.

Dr. Aazhang is a Fellow of IEEE and AAAS, a distinguished lecturer of IEEE Communication Society, and also a recipient of 2004 IEEE Communication Society's Stephen O. Rice best paper award for a paper with A. Sendonaris and E. Erkip. He has been listed in the Thomson-ISI Highly Cited Researchers and has been keynote and plenary speaker of several conferences. He has served as the co-chair of the Technical Program Committee of 2001 Multi-Dimensional and Mobile Communication (MDMC) Conference in Pori, Finland, the chair of the Technical Program Committee for 2005 Asilomar Conference, Monterey, CA, the co-chair of the Technical Program Committee of International Workshop on Convergent Technologies (IWCT), Oulu, Finland in 2005, guest editor for *IEEE Journal on Selected Areas of Communication* special issue on relay and cooperative communication in 2006, the general chair of the 2006 Communication Theory Workshop, Dorado, Puerto Rico, and the co-general chair of 2010 International Symposium on Information Theory (ISIT), in Austin, Texas.