A Scalable Wireless Communication Architecture for Average Consensus*

Shrut Kirti, Anna Scaglione, and Robert J. Thomas

School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
Email: sk482@cornell.edu, anna@ece.cornell.edu, rjt1@cornell.edu

***Abstract*— This paper introduces a multiple-access coding technique that is tailored to solve average consensus problems efficiently in wireless networks. We propose a novel data driven architecture which grants channel access to nodes based on their local data values. We analyze the performance of the scheme in the presence of quantization errors and noise. We show that our scheme is unbiased with respect to quantized consensus algorithms, it achieves good MSE performance, and it can be configured to provide a speedup in the convergence rate. The amount of speedup achieved is a function of $|\mathcal{Q}_k|$ which indicates the number of quantization bins used to represent the state variables exchanged during the computation.**

I. Introduction

Consensus problems have attracted significant research interest in recent years. This attention is not surprising given the applications of average consensus to a wide variety of problems of a distributed nature. Applications like information fusion in sensor networks [1], decision making and control among dynamic agents [2], and congestion control in computer and communication networks [3], [4], among others are good examples. Average consensus algorithms specify how each node i in a network of n nodes can learn the global average $\bar{\theta}(0) = \frac{1}{n}\sum_{i=1}^{n}\theta_i(0)$ by iteratively exchanging messages only with their neighboring nodes in the set $\mathcal{N}_i(k)$. $\theta_i(0)$ are observations known only locally at the start of the distributed computation. Motivated by the lack of scalability of average consensus algorithms implemented using traditional point-to-point links in the wireless medium, in this paper we discuss a new wireless communication architecture for average consensus.

A. Motivation

The computation speed of average consensus protocols is typically characterized as the number of iterations required to attain a certain error compared to the actual mean $\bar{\theta}(0)$, in a deterministic or statistical sense. There are two versions of the algorithm that make one or the other characterization more appropriate. (1) *Synchronous average consensus*: nodes exchange values with their neighbors on a constant or time-varying graph and there is a common clock that dictates the iterations and the advance of the algorithm for all nodes. (2) *Asynchronous or randomized average consensus*: its progress is driven by random pairwise exchanges between the nodes. Since any node is only allowed to transmit to one neighbor at a time it can be paired naturally with a random access policy.

Often times, to quantify the performance of these algorithms in a wireless communication environment, network connectivity is modeled by a planar random geometric graph $G(n, r)$ which consists of a unit area with uniformly distributed nodes that are considered adjacent if and only if they are separated by a distance smaller than r.

The speed of convergence of the synchronous algorithm with a time-invariant graph is described very elegantly in terms the algebraic connectivity of the graph [5]. Corollary 4.4 in [6] showed that for a fixed connectivity radius, both the energy and the averaging time required for convergence in synchronous consensus protocols scale as $O(n)$.

For randomized average consensus algorithms, which seem to fit the model of a wireless unstructured network more naturally, [7] showed that the geographic gossip protocol has an $\epsilon-$averaging time $O(n\log(1/\epsilon))$, and Theorem 6 in [8] showed that the averaging time of a gossiping algorithm on a random geometric graph with radius of connectivity r is $O(\frac{\log n}{r^2})$. Notably, [9] showed without restricting the computation to an average consensus framework, the average value of the initial states in a random planar network can be computed at a frequency $O(1/\log n)$.

A reason for the difference in the results of [7] and [8] is the way the channel access is structured in [8]: it assumes that each node will time its attempts to access the channel with a Poisson process with frequency 1, and that all these attempts will be successful. As the number of nodes grows, this will require a bandwidth expansion on the order of n. On the other hand, since ϵ needs to scale inversely with the maximum degree of the network and this degree tends to be constant like $O(nr^2)$ as $n \to \infty$, the $\log 1/\epsilon$ term in [7] brings back the $O(\frac{\log n}{r^2})$ in [8]. In summary, without a bandwidth expansion, the mixing time becomes increasingly slow as the number of nodes increases.

B. Contribution

In this paper we find physical layer codes that are a more efficient alternative to random scheduling. By allowing the nodes to form the network update directly at the physical layer, we present a solution that is fully scalable as the number of nodes increases. More importantly, it presents a natural trade-off between accuracy, bandwidth, and power that is insensitive or affected positively by an increasing number of nodes.

Specifically, in our scheme the time taken to complete one iteration is $\approx |\mathcal{Q}_k| \mathcal{B}^{-1}$ - independent of the network size n - where $\mathcal{B}$ is the bandwidth allocated for transmission and $|\mathcal{Q}_k|$ indicates the number of quantization bins used to represent the state variables exchanged during the computation.

Statistical analysis of the data driven consensus scheme provides the speed of convergence only in the limit as $|\mathcal{Q}_k| \to \infty$, in which case the algorithm has an $O(1)$ convergence speed. This does not raise much enthusiasm since having an infinite bandwidth available per iteration means that the scalability problems of the other consensus methods would vanish as well. However, if the same result of $O(1)$ number of iterations was observed for a finite $|\mathcal{Q}_k|$, then because each iteration would take a finite time or bandwidth to complete, scalability would be achieved. Interestingly, simulations show that decreasing $|\mathcal{Q}_k|$ increases the speed, an effect that requires complicated analysis to be captured in an exact formula, but that is quite evident numerically and fits with the intuition, as we will discuss later in the paper.

The paper is structured as follows. In Section II we present the system model. In Section III we formulate the physical layer communication design problem and present the data driven communication architecture as a solution. Section IV analyzes the performance of the data driven consensus estimate with respect to its bias, mean squared error, and convergence rate. We demonstrate a speedup in the convergence rate as n increases and as $|\mathcal{Q}_k|$ decreases.

II. System Model

We consider a network of n nodes where each node i has a local state variable $\theta_i(0)$[1] initially known only to itself. Traditional average consensus protocols [10], [11] delineate an iterative procedure in which the $(k+1)$st consensus update computation at node i is-

$$\theta_i(k+1) = \theta_i(k) + \epsilon u_i(k) \tag{1}$$

where ϵ is a step-size parameter and the update variable $u_i(k)$ is a function of the messages node i receives from its neighbors-

$$u_i(k) = \sum_{j \in \mathcal{N}_i(k)} a_{ij}(k)(\theta_j(k) - \theta_i(k)), \tag{2}$$

where $a_{ij}(k) \geq 0$ are the real valued entries of the adjacency matrix $A(k)$, which can be time varying and is defined to have $a_{ii}(k) = 0$, and $\mathcal{N}_i(k)$ is the neighbor set of node i.

Correspondingly, the state vector is updated as follows-

$$\boldsymbol{\theta}(k+1) = (I - \epsilon L(k))\boldsymbol{\theta}(k) = W(k)\boldsymbol{\theta}(k) \tag{3}$$

where $L(k)$ is the graph Laplacian matrix, defined by-

$$l_{ij}(k) = \begin{cases} \sum_{j \in \mathcal{N}_i} a_{ij}(k), & j = i; \\ -a_{ij}(k), & j \neq i, \text{ and } j \in \mathcal{N}_i. \\ 0 & \text{else} \end{cases} \tag{4}$$

and $W(k) \triangleq (I - \epsilon L(k))$.

[1]For convenience we assume $\theta_i(0) \in \mathbb{R}$.

The wireless channel model governing inter-node communications is specified by the following set of assumptions: $(a1)$ the channel is broadcast with flat Rayleigh fading. The gains for any node pair i, j at time slot l are indicated by spatially and temporally independent and identically distributed (*iid*) circularly symmetric complex Gaussian samples $h_{ij}[l] \sim^{iid} \mathcal{CN}(0, \sigma_{ij}^2)$ where $\sigma_{ij}^2 = \mathcal{K}(d^* + d_{ij})^{-\alpha}$, α is the path loss exponent, d_{ij} is the distance between the nodes, and d^* and $\mathcal{K}$, related by $\mathcal{K} = (\frac{1}{d^*})^{-\alpha}$, are modeling parameters that take into account the carrier frequency, the scattering environment and antennae gains. $(a2)$ The received signal at node i is distorted by additive noise $w_i[l] \sim^{iid} \mathcal{CN}(0, N_0)$. $(a3)$ The nodes are half-duplex. Denoting the samples of the discrete-time complex base-band equivalent transmit and receive signal models of node i by $s_i[l]$ and $r_i[l]$ respectively, due to the half-duplex constraint we have-

$$r_i[l] = \begin{cases} \sum_{j=1}^{n} h_{ij}[l] s_j[l] + w_i[l], & \text{if } s_i[l] = 0; \\ 0, & \text{if } s_i[l] \neq 0. \end{cases} \tag{5}$$

Finally, $(a4)$ we assume that the nodes are synchronized and (a5) all nodes have a fixed transmit power constraint P.

Remark 2.1: Note that because of the Shannon Theorem, N samples of the discrete time signal in (5) span a period of time not smaller than $1/\mathcal{B}$ where $\mathcal{B} = f_{\max} - f_{\min}$ is the width of the band allocated for transmission.

III. Data Driven Consensus Architecture

Instead of considering how to communicate $\theta_j(k),\ \forall j \in \mathcal{N}_i(k)$ to node i, we will look directly at an efficient communication method that allows node i to determine the *network information* $u_i(k)$ (2) which is not available locally. Our design problem is then to specify what the transmit samples $s_j[k],\ \forall j \in \mathcal{N}_i(k)$ should be and how $r_i[k]$ should be used to determine $u_i(k)$. The average consensus protocol offers a fundamental hint in this regard - the set of nodes $\{j : \theta_j(k) = \theta_i(k),\ j \in 1, \ldots, n\}$ contributes nothing to the update of node i at iteration k. This can be verified from (2).

A. Design of Multiple Access Channel Codes

Note that if the state values of the nodes could be represented with infinite precision, then under the assumption that they belong to some continuous distribution, the event that any two states are exactly equal has measure zero. However, it is impossible for nodes to exchange their state values exactly since measurement precision and the rates supported by communication channels are both finite. Let us denote the quantized value of $\theta_i(k)$ by $\tilde{\theta}_i(k) \in \mathcal{Q}_k = \{q_l^{(k)},\ l = 0, \ldots, |\mathcal{Q}_k| - 1\}$ where $q_l^{(k)}$ represent quantizer centroids that are possibly time varying. For a given set $\mathcal{Q}_k$, the quantization policy adopted is $\tilde{\theta}_i(k) = \arg\min_{q_l^{(k)}} |\theta_i(k) - q_l^{(k)}|$. Consequently, the update variable can be written as $\tilde{u}_i(k) = u_i(k) + \nu_i(k)$ where $\nu_i(k)$ is a quantization error term. Hence, in terms of communication costs, the quantized update $\tilde{u}_i(k)$ can be made arbitrarily close to the true update $u_i(k)$ by increasing the cardinality $|\mathcal{Q}_k|$.

The key design step is rewriting the update in (2) in terms of the quantized state variables as follows-

$$u_i(k) = \sum_{j=1}^{n}\int_{-\infty}^{\infty} a_{ij}(k)(q-\theta_i(k))\delta(q-\theta_j(k))dq \quad (6)$$

$$\tilde{u}_i(k) = \sum_{l=0}^{|\mathcal{Q}_k|-1} (q_l^{(k)}-\theta_i(k))V_{ki}(q_l^{(k)}) \quad (7)$$

where

$$V_{ki}(q_l^{(k)}) \triangleq \sum_{j=1}^{n} a_{ij}(k)\delta[q_l^{(k)}-\tilde{\theta}_j(k)] \quad (8)$$

and $\delta(\cdot)$ and $\delta[\cdot]$ are the Dirac and Kronecker delta functions respectively.

Equation (7) highlights that the terms $V_{ki}(q_l^{(k)})$, $l = 0,\ldots,|\mathcal{Q}_k|-1$, which are partial sums of the $a_{ij}(k)$ as shown in (8), convey the quantized network information. Therefore, if we can find coefficients $a_{ij}(k) > 0$ such that all nodes can collectively produce a signal corresponding to a single letter code $\hat{V}_{ki}(q_l^{(k)})$, which is a good approximation of $V_{ki}(q_l^{(k)})$, then we can solve both the data modulation and multi-access problems at once. Since the objective of the communication network is not to convey the states of individual nodes but to feed the consensus mechanism with an acceptable update variable $u_i(k)$, rather than contending for the medium to broadcast their own state, the nodes should generate signals cooperatively which will allow all their neighbors to retrieve $u_i(k)$ directly.

Consider a generalized poly-phase component of the transmit signal of node i, $s_i[m]$, at an arbitrary iteration $m \geq 0$. Parsing it into $|\mathcal{Q}_k|$ slots indexed by $l = 0,\ldots,|\mathcal{Q}_k|-1$, we propose the following multiple access code. The discrete-time complex base-band signal transmitted by any node i in channel access slot l within iteration k is-

$$s_i\left[l+\sum_{t=1}^{k-1}|\mathcal{Q}_t|\right] = e^{j\phi_i[l]}\delta\left[q_l^{(k)}-\tilde{\theta}_i(k)\right], l = 0,\ldots,|\mathcal{Q}_k|-1 \quad (9)$$

where $e^{j\phi_i[l]}$ introduces a uniform random phase offset in $[0, 2\pi]$ and is picked independently by node i[2]. Given the dependency of the transmit schedule on the data to be transmitted, we call the multiple access code *data driven*.

In order for this code to be functional, theoretically, we need to further impose the restriction that the fading process be reciprocal on average $E\{|h_{ij}[l]|^2\} = E\{|h_{ji}[l]|^2\}$. This condition of reciprocity in average power received is verified broadly in several media. It is true for wireless RF transmission in static networks and is a valid approximation in networks with a moderate degree of mobility. Note that in $(a1)$, $\mathcal{K}$ is such that $0 < E\{|h_{ji}[l]|^2\} < 1$. Hence, a possible choice for the weights of the adjacency matrix is-

$$a_{ij}(k) = \begin{cases} E\left\{\left|h_{ij}\left[l+\sum_{t=1}^{k-1}|\mathcal{Q}_t|\right]\right|^2\right\}, & \text{if } i \neq j, \\ 0, & \text{else} \end{cases} \quad (10)$$

[2]Consequently $E\{s_i[l]s_j[l']\} = 0,\ \forall\ l, l'$ and $\forall i \neq j$.

and with $0 < \epsilon < 1/\max(degree(A))$ this choice[3] satisfies the constraints for the stability of W given in [10]. In the general case where $A(k)$ is time varying, ϵ can be chosen small enough that the given constraint is satisfied for all $A(k)$. Note that our choice in (10) does not require that the channel gains be known in order to implement data driven consensus. The choice is dictated by theoretical considerations of algorithm stability.

Finally, accounting for the half-duplex constraint, we approximate $V_{ki}(q_l^{(k)})$ by-

$$\hat{V}_{ki}(q_l^{(k)}) = \begin{cases} \left|r_i\left[l+\sum_{t=1}^{k-1}|\mathcal{Q}_t|\right]\right|^2 - N_0, & \text{if} q_l^{(k)} \neq \tilde{\theta}_i(k) \\ 0, & \text{else.} \end{cases} \quad (11)$$

where $r_i\left[l+\sum_{t=1}^{k-1}|\mathcal{Q}_t|\right]$ is defined in (5). This gives us that-

$$\hat{u}_i(k) \triangleq \sum_{l=0}^{|\mathcal{Q}_k|-1}(q_l^{(k)}-\tilde{\theta}_i(k))\hat{V}_{ki}(q_l^{(k)}). \quad (12)$$

It is clear from equations (11) and (12) that the nodes do not need to decode the signals they receive. Instead, they use the magnitude of the received signal energy, after subtracting the noise power, to compute the consensus update directly.

Remark 3.1: Figs. 1(a) and 1(b) show the evolution of the states in a sample trial. From the division of each iteration period into $|\mathcal{Q}_k|$ slots and the multiple access strategy, it is clear that the time taken to complete one iteration must be $\approx |\mathcal{Q}_k|\mathcal{B}^{-1}$ - independent of the network size n - where $\mathcal{B}$ is the bandwidth required for transmission of signal $s_i[l]$. Therefore, while $n \gg |\mathcal{Q}_k|$, the data driven architecture will take a shorter amount of time than a perfect time division duplex strategy which schedules the nodes to transmit one at a time.

IV. Performance Analysis

We show that the data driven consensus architecture is unbiased with respect to quantized consensus and it achieves low MSE. We focus on the statistical properties of $\hat{u}_i(k)$ in our analysis because it governs the performance. We also show that on the average, when we have infinite precision, its speed of convergence becomes invariant with n.

A. Statistical Analysis of $\hat{u}_i(k)$

Lemma 4.1: Under assumptions $(a1)-(a5)$ and that the node states $\boldsymbol{\theta}(k)$ are known at the beginning of each iteration, the moment generating function (MGF) of $\hat{u}_i(k)|\boldsymbol{\theta}(k)$ is-

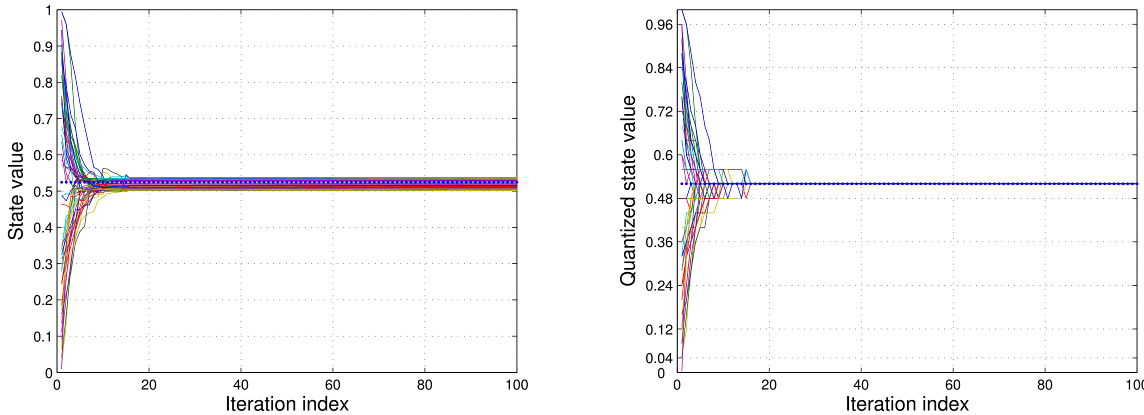
$$\phi_{\hat{u}_i(k)|\boldsymbol{\theta}(k)}(s) = \prod_{l=0}^{|Q_k|-1}\frac{e^{-s(q_l^{(k)}-\tilde{\theta}_i(k))N_o}}{1-s(q_l^{(k)}-\tilde{\theta}_i(k))\lambda_l} \quad (13)$$

where $\lambda_l = \sum_{j=1}^{n}\sigma_{ij}^2\delta[q_l^{(k)}-\tilde{\theta}_j(k)]+N_o$.

Proof: See Appendix I. ■

The following lemma shows that the data driven consensus algorithm is an unbiased implementation of any consensus

[3]Note that since $h_{ij}[l]$ are *iid* with respect to l, $\mathbb{E}\{|h_{ij}[l]|^2\}$ is independent of the slot index l.



(a) (b)

Fig. 1. *Data driven consensus*: we plot the evolution of the state vector $\boldsymbol{\theta}(k)$ for a network of 50 nodes with $|\mathcal{Q}_k|$ constant over the iterations. In (a) we show the unquantized state values that each node maintains and in (b) we show the actual quantized states exchanged with neighbors.

algorithm that uses the quantized consensus update $\tilde{u}_i(k)$, irrespective of the SNR.

Lemma 4.2: Let $\hat{\mathbf{u}}(k)$ and $\tilde{\mathbf{u}}(k)$ denote the update variables for data driven consensus and any other quantized consensus algorithm respectively. Then, under the assumptions $(a1)-(a5)$ we have that $\mathbb{E}\{\hat{\mathbf{u}}(k)|\boldsymbol{\theta}(k)\} = \tilde{\mathbf{u}}(k)$ where $\tilde{\boldsymbol{\theta}}(k)$ denotes the quantized state vector and $L \triangleq diag(A\mathbf{1}) - A$ with the entries of the adjacency matrix A selected as (10).

Proof: Using Lemma 4.1 we have that $\mathbb{E}\{\hat{u}_i(k)|\boldsymbol{\theta}(k)\} = \phi'_{\hat{u}_i(k)|\boldsymbol{\theta}(k)}(0) = \sum_{l=0}^{|\mathcal{Q}_k|-1}(q_l^{(k)} - \tilde{\theta}_i(k))(\lambda_l - N_o)$. This can also be verified directly from (12). From (7) and (8) it follows that to complete our proof we need to show $V_{ki}(q_l^{(k)}) = \lambda_l - N_o = \sum_{j=1}^{n}\sigma_{ij}^2\delta[q_l^{(k)} - \tilde{\theta}_j(k)]$. But $a_{ij} = \sigma_{ij}^2, \ \forall i \neq j$ and 0 otherwise by our choice in (10). Thus, $\mathbb{E}\{\hat{u}_i(k)|\boldsymbol{\theta}(k)\} = \tilde{u}_i(k)$ and the result follows. ■

The MSE is a function of $|\mathcal{Q}_k|$, network connectivity, and SNR as one might intuitively guess. An important consequence of this is that when the state space is quantized finely and when a well connected network is operating in a high SNR regime, the MSE will be small and the consensus value will be close to the true mean. This is numerically verified in Figs. 2(a) and 2(b). A detailed derivation of the analytical expression for the MSE is provided in [12].

Remark 4.3: (CONSERVATION OF SUM OF STATES) Define the average value of the state vector at the kth iteration as $g(k) = \frac{1}{n}\mathbf{1}^T\boldsymbol{\theta}(k)$. In traditional average consensus, the sum of the states is conserved at each iteration i.e. $g(k) = g(0) \ \forall \ k$. The presence of quantization, fading and noise prevents the conservation of the average of the states in the data driven algorithm, which in turn, is another measure of error and possible bias in the algorithm. We plot $\mathbb{E}\{(g(k) - g(0))^2\}$ in Figs. 3(a) and 3(b) with respect to different quantizers $|\mathcal{Q}_k|$ (that are invariant with k) and varying SNR regimes. In both cases the variance from the true mean of the states remains small.

B. Convergence speed in the limit $|\mathcal{Q}_k| \to \infty$

We use the second largest eigenvalue of W i.e. $\lambda_2(W)$ to provide a measure of the average number of iterations required for convergence [10]. We first show that in the limit $|\mathcal{Q}_k| \to \infty$ the effects of quantization and fading vanish and the consensus iterations are equivalently made with the average W whose entries depend on (10).

Lemma 4.4: In the limit $|\mathcal{Q}_k| \to \infty$, under $(a1)-(a5)$ we have that the data driven network update is the same as the unquantized update in (2)-

$$\hat{u}_i(k) \to \sum_{j=1}^{n}\sigma_{ij}^2(\theta_j(k) - \theta_i(k)). \tag{14}$$

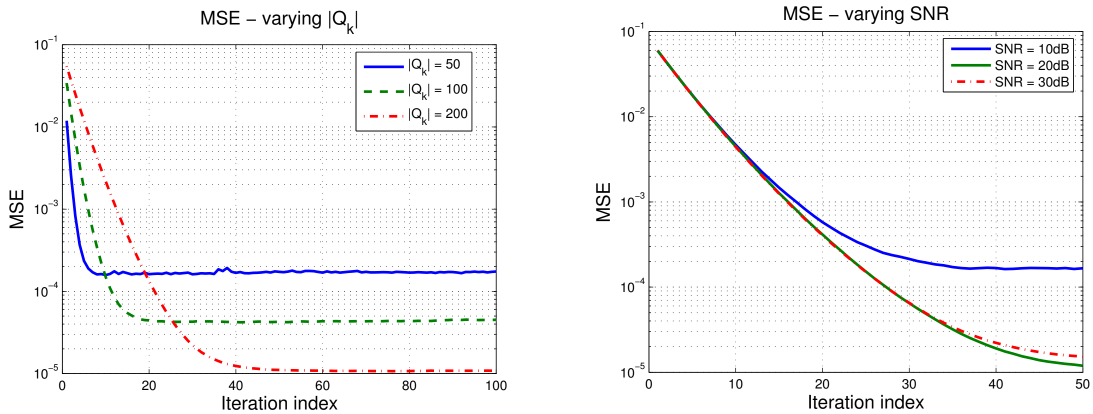
Proof: See Appendix II. ■

Next, we will use the fact that as the network becomes dense, the eigenvalues of the average W for the network whose nodes are deployed randomly in the network area can be approximated with those of a network in which the nodes are located on a regular grid. If we further ignore the edge effects in the 2-D network and approximate the network with wraparound in both the x and y directions, then the network is effectively on a 2-torus instead of the unit square. Using these approximations allows us to determine $\lambda_2(W)$ in closed form and we show that it grows $O(1)$ with n.

Consider n nodes arranged in a $\sqrt{n} \times \sqrt{n}$ regular grid on a 2-torus (assume that n is a perfect square). It can be verified that the resulting $n \times n$ distance matrix and the corresponding adjacency matrix are block circulant with each $\sqrt{n} \times \sqrt{n}$ block, circulant in itself. The adjacency matrix is $A = blockcirc(A_0, \ldots, A_{\sqrt{n}-1})$ where the kth block is $A_k = circ(a_0^k, a_1^k, \ldots, a_{\sqrt{n}-1}^k)$ (see [13]). The eigenvalue of A corresponding to position p in the qth block is-

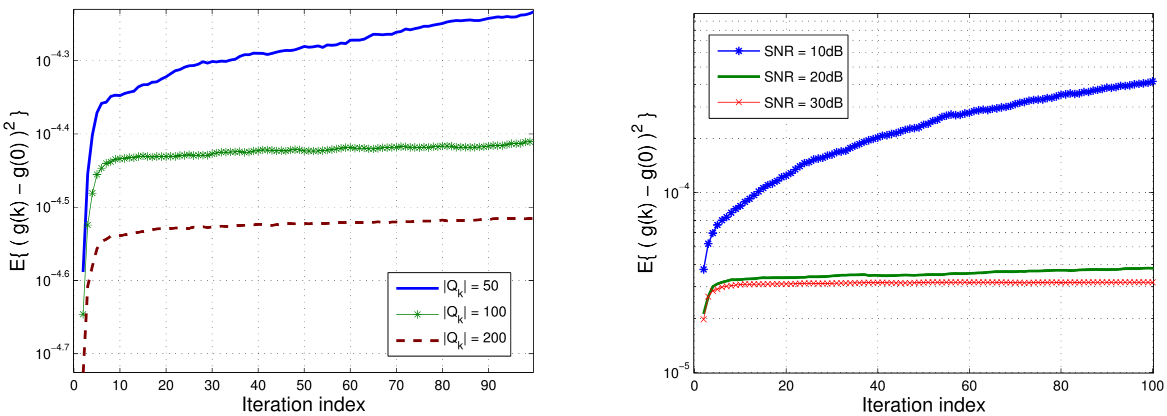
$$\lambda_{p,q}(A) = \sum_{k=0}^{\sqrt{n}-1}\left(\sum_{m=0}^{\sqrt{n}-1} a_m^k e^{j2\pi m\frac{p}{\sqrt{n}}}\right)e^{j2\pi k\frac{q}{\sqrt{n}}} \tag{15}$$

where $p, q = 0, \ldots, \sqrt{n} - 1$. An alternate expression is provided by Theorem 5.8.1 in [14]. The largest eigenvalue



(a) (b)

Fig. 2. MSE vs. iteration index for $n = 50$ - (a) $SNR = 30dB$ and varying $|\mathcal{Q}_k|$ and (b) for varying SNR with $|\mathcal{Q}_k| = 250$.



(a) (b)

Fig. 3. $\mathbb{E}\{(g(k) - g(0))^2\}$ *vs. iteration index*: (a) $SNR = 30dB$ and varying $|\mathcal{Q}_k|$ and (b) for varying SNR with $|\mathcal{Q}_k| = 200$, $\forall k$. Note that the average deviation from $g(0)$ is small in both plots and therefore the sum of the states is almost conserved. The plots were obtained using 1000 Monte Carlo trials on a network 50 nodes.

is $\lambda_1(A) = \sum_{k=0}^{\sqrt{n}-1} \sum_{m=0}^{\sqrt{n}-1} a_m^k$ which grows with n since all $a_m^k \in [0, 1]$. Using the definitions of L and W, we have that they are both block circulant matrices with circulant blocks and W has entries $w_0^0 = 1 - \epsilon \sum_{k=0}^{\sqrt{n}-1} \sum_{m=0}^{\sqrt{n}-1} a_m^k = 1 - \epsilon\lambda_1(A)$ and $w_m^k = \epsilon a_m^k$. This allows us to write $\lambda(W)$ in terms of $\lambda(A)$. Specifically,

$$\lambda_2(W) = 1 - \epsilon(\lambda_1(A) - \lambda_{P,Q}(A)) - \epsilon a_0^0 \qquad (16)$$

where P, Q are unknown indices and $a_0^0 = 0$ by definition. Therefore, $|\lambda_2(W)| \leq 1 + \epsilon(|\lambda_1(A)| + |\lambda_{P,Q}(A)|)$. Since $|\lambda_1(A)| + |\lambda_{P,Q}(A)|$ grows with n we have $|\lambda_2(W)| = 1 + \epsilon O(n)$. But, to satisfy the constraints on W, ϵ must scale inversely with the maximum degree of the network i.e. $\epsilon = O(1/n)$. Thus, $|\lambda_2(W)| = O(1)$ as $n \to \infty$ (Fig. 4). These arguments on the speed of convergence are valid only on average.

C. Convergence speed for finite $|\mathcal{Q}_k|$

It is intuitive that with coarser quantization we can converge faster although we may take an accuracy hit; assuming that the algorithm converges for these values of $|\mathcal{Q}_k|$. We provide heuristic arguments explaining this speedup, leaving a rigorous proof for future work.

Remark 4.5: Data driven consensus leads to a speedup in the convergence rate. The speedup is $\propto n/|\mathcal{Q}_k|$.

Let set $C_{k,l}$ denote the set of all nodes that are quantized to the same value $q_l^{(k)}$ in iteration k. We note that all nodes in $C_{k,l}$ achieve single-shot agreement regardless of the actual network topology. Based on the partitions $C_{k,l}$ we define a new $|Q_k| \times |Q_k|$ adjacency matrix A' with-

$$a'_{ij}(k) = \begin{cases} \sum_{j \in C_{k,l}} a_{ij}(k), & \text{if } i \in C_{k,l}, j \in C_{k,l'},\ l \neq l'; \\ 0, & \text{if } i = j \text{ or } i, j \in C_{k,l}. \end{cases} \qquad (17)$$

This allows us to recast the original consensus problem with at most n distinct node states to one with at most $|\mathcal{Q}_k|$ distinct states. Since all nodes in $C_{k,l}$ transmit the same signal synchronously (9), we obtain cooperative transmission gains [15] which implies that the connectivity of the graph defined by A' is greater than that of A. Consequently, it can be shown that both the largest and the smallest nonzero eigenvalues of L' will be larger than those of L leading to a faster convergence rate [16]. Since cooperative gains increase with $|C_{k,l}|$, the speedup is $\propto 1/|\mathcal{Q}_k|$. When $\theta_i(k)$ are assumed to be uniformly independently distributed, $|C_{k,l}|$ also increases with n leading to speedup.

V. Conclusion

We propose a novel data driven synchronous consensus algorithm. This is based on the principle that the channel resources should be shared based on the state values of the nodes rather than their identities. We show that the received signals can be used to extract the new update of the algorithm directly, even if it is impossible to discriminate which node has what state. In fact, part of the computation is made by the linear medium, turning interference into a resource which simplifies the computation. With this approach we attain an improved speed as the number of nodes grows.

Appendix I

Proof of Lemma 4.1: From $(a1)$, $(a2)$, and that cross-correlation $E\{s_i[l]s_j[l']\} = 0,\ \forall\ l, l'$ and $\forall i \neq j$, from (5) it follows that $r_i[l]|\boldsymbol{\theta}(k)$ are independently distributed zero mean complex gaussian RVs with variance $\lambda_l = \sum_{j=1}^{n} \sigma_{ij}^2 \delta[q_l^{(k)} - \tilde{\theta}_j(k)] + N_o$. Therefore, $|r_i[l]|^2\,|\boldsymbol{\theta}(k) \sim \exp(\frac{1}{\lambda_l})$. Let $X_l = |r_i[l]|^2|\boldsymbol{\theta}(k)$ and let $Y_l|\boldsymbol{\theta}(k) = a + bX_l$ be an affine transformation of X_l with $a = -(q_l^{(k)} - \tilde{\theta}_i(k))N_o$ and $b = q_l^{(k)} - \tilde{\theta}_i(k)$. It is easily shown that for $b \neq 0$, $Y_l|\boldsymbol{\theta}(k) \sim \frac{1}{b\lambda_l} e^{-\frac{1}{\lambda_l}\frac{y_l - a}{b}}$ with $y_l \in [a, \infty)$ if $b > 0$ and $y_l \in (-\infty, a]$ if $b < 0$. From (11) and (12) its clear that the update variable is a sum of independent random variables *i.e.* $\hat{u}_i(k)|\boldsymbol{\theta}(k) = \sum_{l=0}^{|\mathcal{Q}_k|-1} Y_l|\boldsymbol{\theta}(k)$. The MGF is $\phi_{\hat{u}_i(k)|\boldsymbol{\theta}(k)}(s) = \prod_{l=0}^{|\mathcal{Q}_k|-1} \mathbb{E}\{e^{sY_l|\boldsymbol{\theta}(k)}\}$. Simplifying this by using the pdf of $Y_l|\boldsymbol{\theta}(k)$ gives us the desired result.

Appendix II

Proof of Lemma 4.4: We provide only a brief sketch of the proof due to space constraints. Substitute (11) in (12).

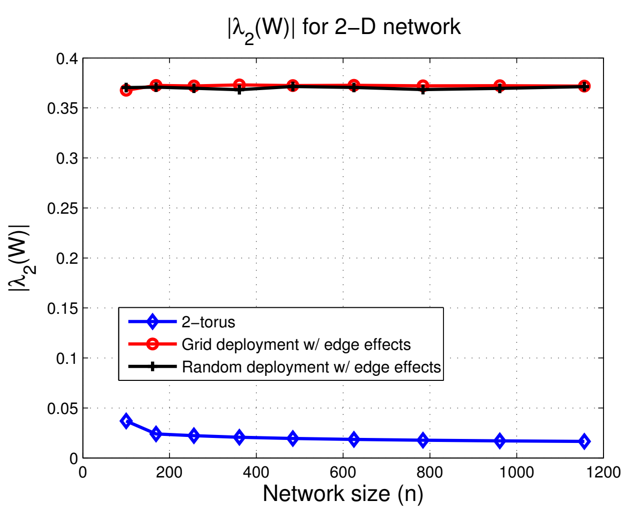


Fig. 4. $\lambda_2(W)$ saturates as n grows.

Now substitute (5) and expand the expression out by using $|r_i[l]|^2 = r_i[l]r_i[l]^*$ to get terms that are sums of $|\mathcal{Q}_k|$ components. Then in the limit $|\mathcal{Q}_k| \to \infty$, by the applying the WLLN term by term we have that-

$$\frac{|\mathcal{Q}_k|}{|\mathcal{Q}_k|} \sum_{l=0}^{|\mathcal{Q}_k|-1} (|w_i[l]|^2 - N_o)(q_l^{(k)} - \theta_i(k)) \to 0, \qquad (18)$$

since $\mathbb{E}\{|w_i[l]|^2\} = N_o$. Further, $h_{ij}[l]$ and $w_i[l]$ are zero mean RV's that are mutually independent with $s_j[l]$ so

$$\frac{|\mathcal{Q}_k|}{|\mathcal{Q}_k|} \sum_{j=1}^{n} \sum_{l=0}^{|\mathcal{Q}_k|-1} h_{ij}[l]s_j[l]w_i[l]^*(q_l^{(k)} - \theta_i(k)) \to 0. \qquad (19)$$

And finally as desired, we are left only with,

$$\begin{aligned}\hat{u}_i(k) \quad &\to \quad \sum_{j=1}^{n} \mathbb{E}\{|h_{ij}[l]|^2\}\delta[q_l^{(k)} - \tilde{\theta}_j(k)](q_l^{(k)} - \theta_i(k)) \\ &= \quad \sum_{j=1}^{n} \sigma_{ij}^2(\theta_i(k) - \theta_j(k)) \qquad (20)\end{aligned}$$

References

[1] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. International Conference on Information Processing in Sensor Networks*, Los Angeles, CA, April 2005.

[2] R. Olfati-Saber, "Distributed kalman filter with embedded consensus filters." Proc. 44th IEEE Conference on Decision and Control, Dec. 2005, pp. 8179 – 8184.

[3] J. Tsitsikilis, "Problems in decentralized decision making and computation," Ph.D. dissertation, MIT, 1984.

[4] G. Cybenko, "Load balancing for distributed memory multiprocessors," *Journal of Parallel and Distributed Computing*, vol. 7, pp. 279 – 301, 1989.

[5] R. Olfati-Saber and R. Murray, "Agreement problems in networks with directed graphs and switching topology." Proc. 42nd Conference on Decision and Control, 2003.

[6] A. Scaglione, "On the wireless communication architecture for consensus problems," in *Proc. Information Theory and Applications Workshop*, San Diego, CA, January 2007.

[7] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient aggregation for sensor networks," in *Proc. International Symposium on Information Processing in Sensor Networks (IPSN)*, Nashville, TN, April 2006.

[8] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," in *Proc. INFOCOM*, 2005.

[9] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 755 – 764, April 2005.

[10] 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.

[11] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. on Automatic Control*, vol. 49, no. 9, pp. 1520 – 1533, september 2004.

[12] S. Kirti, M. E. Yildiz, and A. Scaglione, "Scalable channel coding for average consensus," *submitted to IEEE Journal on Selected Areas in Communication sp. issue on Control and Communications, 2007.*

[13] R. Gray, *Toeplitz and circulant matrices: A review*. Technical Rept. No. 6504-1, Inform. Sys. Lab., Stanford Univ., Stanford, CA, April 1977.

[14] P. J. Davis, *Circulant Matrices*. New York: Wiley-Interscience, 1979.

[15] A. Scaglione and Y.Hong, "Opportunistic large arrays: Cooperative transmission in wireless multihop ad hoc networks to reach far distances," *IEEE Trans. SP '03*, vol. 51, no. 8, pp. 2082 – 2092, August 2003.

[16] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," in *Proc. IEEE Conf. on Decision and Control*, Mausi, Hawaii, Dec. 2003.