On the Problem of Adding Infinitely Many Values

(Invited Paper)

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

***Abstract*—Average consensus is a decentralized computation algorithm for calculating the average of the state variables of the nodes in a network. Over wireless networks, it is typically implemented using point-to-point random access scheduling. Wireless communications, however, are interference limited and when the available bandwidth is fixed, the expected delay in achieving a certain precision in the average value increases as the network size scales up. We show that this limitation in large networks disappears if we use specially structured codes. We analyze a combined source and channel coding strategy that uses an incoherent combination of power over orthogonal sub-channels for the average consensus protocol. We show that in spite of the bandwidth and power limitations, with our simple strategy the delay and precision can be kept bounded while increasing the number of participants.**

I. Introduction

Network architectures which utilize a fusion center are naturally constrained - the fusion center is a communication bottleneck and a single point of failure. Gossiping protocols [1], [2] on the other hand, offer an alternative by relying on repeated near-neighbor communications across the network. Such protocols are attractive in wireless sensor network applications for a number of reasons. For example, they alleviate the complexity of the network layer. The low control overhead of local communications is well suited to the low operating power restriction of sensors. In addition, gossiping protocols are robust to node failures and possible network topology changes since they can adapt to such modifications. Unfortunately, gossiping protocols are not a panacea. While they do not suffer from a single point of failure, they are not immune to communication bottlenecks. These bottlenecks can be described in different terms depending on the abstraction chosen to analyze the network. For example, possible bottleneck measures are: a limit on the number of messages exchanged, packet drop rate, the number of bits per message used, or delay.

In this paper we focus on a particular type of gossiping - *average consensus* - which is aimed at computing the average of the states of all nodes distributedly [1]–[3] (c.f. Section II for a review). In the case of average consensus, in order to achieve a certain precision in the consensus estimate in a given time, each node should successfully receive a certain average number of updates from its neighbors. The rate at which a node can receive these updates is a communication bottleneck. As the network size and the number of neighbors increases, the main source of this bottleneck is congestion since greater traffic in the channel is associated with increased packet drop and delay. Network congestion can be mitigated by increasing the available spectrum, but unfortunately, increasing transmission power yields no benefit. Therefore, even if we neglect the problem of transmitting under a rate constraint, computing the average of states in an infinitely dense network over a bandlimited wireless channel using standard scheduling is impossible due to congestion. In this paper we show that, by rethinking conventional assumptions on the communication architecture, it is possible to prove this conclusion invalid. We show that this limitation disappears if we combine the computation with source and channel coding.

This work is supported in part by the grant NSF CNS 05-24695.

The paper is organized as follows. In section II we present basic results on the average consensus algorithm and demonstrate its poor scalability. Section III introduces data driven consensus. In sections IV and V we examine the convergence and MSE properties of our algorithm. Section VI concludes the paper.

II. Average consensus and its performance

Each node is assumed to have a long sequence of data, denoted by vector $\boldsymbol{\theta}(0,i) \in \mathbb{R}^m$, to average where $i = 1,\ldots,n$ denote the node indices. Nodes initialize their local states with these values $\boldsymbol{\theta}(0,i)$ at iteration $t = 0$. At the $(t+1)$st iteration node i receives values $\boldsymbol{\theta}(t,j)$ from all the neighboring nodes $j \in \mathcal{N}_i$, where $\mathcal{N}_i$ denotes the set of neighbors of node i, and updates its state as follows-

$$\boldsymbol{\theta}(t+1,i) = \boldsymbol{\theta}(t,i) + \sum_{j=1}^{n} a_{ij}(t)(\boldsymbol{\theta}(t,j) - \boldsymbol{\theta}(t,i)), \qquad (1)$$

where $a_{ij}(t) \geq 0$ is the ij^{th} element of the non-negative network adjacency matrix $A(t)$ with $a_{ij}(t) = 0$ if $i = j$ or if the nodes are not neighbors. The update term-

$$\mathbf{u}(t,i) = \sum_{j=1}^{n} a_{ij}(t)(\boldsymbol{\theta}(t,j) - \boldsymbol{\theta}(t,i)) \qquad (2)$$

is obtained via near-neighbor communications. Stacking these vectors $\boldsymbol{\theta}(t,i)$ in a super vector $\boldsymbol{\theta}(t) \in \mathbb{R}^{mn}$:

$$\boldsymbol{\theta}(t) = [\boldsymbol{\theta}^T(t,1),\ldots,\boldsymbol{\theta}^T(t,n)]^T, \qquad (3)$$

the average consensus iterations can be written as $\boldsymbol{\theta}(t+1) = (W(t) \otimes I)\boldsymbol{\theta}(t)$ where $\otimes$ denotes the matrix Kronecker product and $W(t) = (I - diag(A(t)\mathbf{1}) + A(t))$.

Consensus protocols combine information diffusion with computation, which leads to a natural trade-off between the number of iterations required and the accuracy achievable in consensus. This trade-off has been studied extensively for both the synchronous [4] and asynchronous versions of the algorithm [1] and is summarized in the following lemma [1].

Lemma 1. *Let $W(t) = W$. The average Mean Squared Error (MSE) at iteration t for consensus without quantization is-*

$$MSE_{AC}(t) = \frac{1}{mn}\mathbb{E}\{||\boldsymbol{\theta}(t) - (J \otimes I)\boldsymbol{\theta}(0)||^2\} \quad (4)$$

$$\leq \lambda_2^{2t}(W)\frac{E\{\|\boldsymbol{\theta}(0)\|^2\}}{mn} \quad (5)$$

where $\lambda_2(W)$ denotes the second largest eigenvalue of W and $\|\cdot\|$ denotes the l_2 norm operator.

This single letter characterization of the network rate of convergence, denoted by $\lambda_2(W)$, holds if we assume that the communications between neighboring nodes are always successful and have infinite precision.

A. Scalability

We will now setup a comparison of standard average consensus with our structured codes. We illustrate heuristically that standard average consensus scales poorly with n because $\lambda_2(W) \overset{n\to\infty}{\to} 1$.

Consider the following extreme strategies for exchanging messages in the network: (1) keep the communications local or (2) try to reach distant neighbors. We examine what happens to $\lambda_2(W)$ for $n \gg 1$ for each of these strategies.

To illustrate the (lack of) scalability of the network in the first case, consider n sensors deployed in a circle of radius r and let a node communicate with k neighbors on either side of it. We let the set of neighbors $\mathcal{N}_i, \ \forall i$ be fixed while increasing n. The adjacency matrix $A = \epsilon \cdot circ(0, \mathbf{1}_k^T, \mathbf{0}_{n-2k-1}^T, \mathbf{1}_k^T)$, where $\mathbf{1}_k$ is the all-ones vector in $\mathbb{R}^k$ and ϵ some constant, is circulant. In this case the eigenvalues of W can be computed in closed form [5]. Given that $|\mathcal{N}_i| = 2k, \ \forall i$ is kept constant while n scales up, the following lemma holds.

Lemma 2. *For the given circulant network, and k fixed, the second largest eigenvalue is-*

$$\lambda_2(W) \overset{n\gg 1}{\to} 1 - O\left(\frac{1}{n^2}\right). \quad (6)$$

Proof: Using Theorem 3.2.2 from [5] and simplifying, we have $\lambda_2(W) = 1 - 2k\epsilon + 2\sum_{l=1}^{k}\epsilon\cos\left(\frac{2\pi l}{n}\right)$. Under $n \gg 1$, use $\cos x \approx 1 - x^2/2$ with $x = \frac{2\pi l}{n}$ to get $\lambda_2(W) = 1 - \frac{2\pi^2\epsilon}{n^2}\frac{k(k+1)(2k+1)}{3}$ which scales as $1 - O(\frac{1}{n^2})$. ■

This tendency is opposite to the well known requirement that for fast convergence the Fiedler eigenvalue ($\lambda_{n-1}(L) = 1 - \lambda_2(W)$) should be close to 1 (and $\lambda_2(W)$close to 0) [2]. This argument can be generalized to the case of a circulant network deployed on a torus in a straightforward manner.

To illustrate the (lack of) scalability of the protocol when nodes try to communicate to all peers, possibly far away, one can refer to the analysis of *geographic gossip* [3]. Intuitively, this approach seems to improve the algebraic connectivity and decreases the average mixing time. It was shown in [3] that in this scenario, $\lambda_2(W)$ scales as $O(1 - c/n)$ where c is some constant (the position of the nodes is irrelevant to obtain this result). Thus, also in this case, albeit with a better trend, $\lambda_2(W)$ approaches 1 as $n \gg 1$ and the convergence speed decreases. This result holds more generally for nodes deployed randomly in a unit square. The physical distance plays an even greater role if one considers the congestion related with multi-hop routing, as done in [6].

III. Collaborative MAC for Consensus

In this section we discuss our joint source channel computation scheme [7] and the assumptions that support it.

A. Assumptions

We first note that the state values $\boldsymbol{\theta}(t)$ cannot be exchanged with infinite precision because of communication constraints.

a.0 Quantization: We consider a simple suboptimal design where all nodes use the same quantizer, with codes $\mathcal{L} = \mathcal{Q}_1 \times \ldots \times \mathcal{Q}_m$ that belong to a square lattice. The lattice contains at most $L = \prod_{p=1}^{m} Q_p$ points, where $Q_p, \ p = 1, \ldots, m$, denotes the number of levels in the one dimensional uniform lattice $\mathcal{Q}_p$. Let $\boldsymbol{\theta}_p(0, i)$ denote the pth component of $\boldsymbol{\theta}(0, i)$. It is assumed that $\boldsymbol{\theta}_p(0, i)$ and $\boldsymbol{\theta}_{p'}(0, i)$ are independent $\forall p \neq p'$. We assume that the range of the pth quantizer at each node is $C\sigma_p$ where $\sigma_p = \max_i \sqrt{VAR[\boldsymbol{\theta}_p(0, i)]}$ and C is a positive constant that renders clipping errors statistically negligible. The state variables are quantized as follows: $\boldsymbol{\theta}(t, i) \mapsto \bar{\boldsymbol{\theta}}(t, i) \in \mathcal{L} = \{\mathbf{q}_1, \ldots, \mathbf{q}_L\}$, where vectors $\mathbf{q}_l \in \mathbb{R}^m$ comprise components which are quantizer centroids. The resulting quantization error $\mathbf{v}(t, i) = \boldsymbol{\theta}(t, i) - \bar{\boldsymbol{\theta}}(t, i)$ can be assumed to be uncorrelated from state to state, approximately uniform, with statistics:

$$VAR[\mathbf{v}_p(t, i)] = \frac{C^2}{12}\frac{\sigma_p^2}{Q_p^2}, \ \Sigma = \frac{C^2}{12}diag\left(\frac{\sigma_1^2}{Q_1^2}, \ldots, \frac{\sigma_m^2}{Q_m^2}\right) \quad (7)$$

$$E\{\mathbf{v}(t)\mathbf{v}^T(t)\} = (I \otimes \Sigma). \quad (8)$$

Accordingly, the quantized consensus update is-

$$\bar{\mathbf{u}}(t, i) = \sum_{j=1}^{n} a_{ij}(t)(\bar{\boldsymbol{\theta}}(t, j) - \bar{\boldsymbol{\theta}}(t, i)). \quad (9)$$

The assumption of uncorrelated quantization error, common in considering quantization noise, is unrealistic if the quantizer is too coarse. Hence, in all our conclusions, we assume that the quantizer has sufficient resolution for **a.0** to be valid in practice.

We make the following assumptions on the communication physical layer. Let $\tau \in \mathbb{R}$ indicate the continuous time variable. Let $Q = \sum_{p=1}^{m} Q_p$.

a.1 Signal space – Time is slotted in intervals of duration $T = 1$. The RF signals transmitted belong to a signal space of dimension $Q \propto BT$ complex dimensions where B is the allocated bandwidth around the carrier frequency. We denote by $\{c_l(\tau)\}_{l=1}^{Q}$, the base-band complex equivalent orthonormal basis chosen to span the signal space. The signal transmitted by node i in the tth iteration is: $S_i(\tau,t)=\sum_{l=1}^{Q} s_i(t,l)c_l(\tau-tT)$ where $s_i(t)=[s_i(t,1),\ldots,s_i(t,Q)]^T$ is the vector of coordinates of the transmit signal with respect to the basis $\{c_l(\tau)\}_{l=1}^{Q}$.

a.2 Node power constraint – Each node has a per iteration power constraint $P=\sum_{l=1}^{Q}|s_i(t,l)|^2$.

a.3 *Incoherent channel: Fading + AWGN* – Each received signal, whose complex envelope is $R_i(\tau,t)$ is affected by an independent additive white Gaussian noise process $\mathrm{W}_i(\tau)$ with noise spectral density N_0. The channel is broadcast. Its distortion on $c_i(\tau)$ due to asynchronism and multipath, can be captured by a single independent fading coefficient, changing over the slots as block fading, denoted by $h_{ij}(t)\sim\mathcal{CN}(0,\alpha_{ij})$, where α_{ij} is the average path-loss (large scale fading). Reciprocity holds on average, *i.e.* $\alpha_{ij}=\alpha_{ji}$.

a.4 *Half-duplex channel*: If node i has $s_i(t,l)\neq 0$ for $l\in\mathcal{R}\subseteq[1,Q]$, node i cannot sense any code transmitted in the sub-space spanned by $\{c_l(\tau)\}_{l\in\mathcal{R}}$.

Based on (**a.1**-**a.3**), a sufficient statistic for the received signal is $r_i(t,l)=n^{-1/2}<R_i(\tau,t),c_l(\tau-lT)>$. Given (**a.1**-**a.4**), for $l\in[1,Q]$ we have:

$$r_i(t,l)=\begin{cases}\frac{1}{\sqrt{n}}\left[\sum_{j=1}^{n}h_{ji}(t)s_j(t,l)+w_i(t,l)\right] & s_i(t,l)=0\\ 0 & \text{else.}\end{cases}\tag{10}$$

The received vector is $r_i(t)=[r_i(t,1),\ldots,r_i(t,Q)]^T$, $\forall i$.

B. Channel Codes

We consider channel coding and decoding strategies that do not have memory across blocks, i.e. is a code is a mapping: $\bar{\boldsymbol{\theta}}(t,i)\mapsto s_i(t),\ i=1,\ldots,n$. The receiver's objective is to retrieve the update estimate $\bar{\mathbf{u}}(t,i)$ in (9) from the received vector $r_i(t)$, i.e. the decoder is a mapping: $r_i(t)\mapsto\bar{\mathbf{u}}(t,i),\ i=1,\ldots,n$. First consider the encoder.

Lemma 3. *In quantized consensus, for any adjacency matrix $A(t)$, two nodes whose states fall in the same quantization bin at iteration t do not need to communicate during that iteration.*

Proof: Based on $\mathbf{a.0}$, the quantized update $\bar{\mathbf{u}}(t,i)=\sum_{j=1}^{n}a_{ij}(t)(\bar{\boldsymbol{\theta}}(t,j)-\bar{\boldsymbol{\theta}}(t,i))$ can be decomposed as:

$$\bar{\mathbf{u}}_p(t,i)=\sum_{k=1}^{Q_p}(q_k-\bar{\boldsymbol{\theta}}_p(t,i))\{\mathbf{m}(t,i)\}_{l'},p=1,\ldots,m,\tag{11}$$

where $l'=\sum_{p'=1}^{p-1}Q_{p'}+k$, $q_k\in\{\mathcal{Q}_p\}$, and

$$\{\mathbf{m}(t,i)\}_{l'}=\sum_{j=1}^{n}a_{ij}(t)\delta[q_{l'}-\bar{\boldsymbol{\theta}}_p(t,j)]\tag{12}$$

with $\delta[x]=\mathbb{I}_{\{x=0\}}$. Of all the terms $\{\mathbf{m}(t,i)\}_{l'}$, the one that corresponds to $q_k=\bar{\boldsymbol{\theta}}_p(t,i)$ is always weighted by zero. This proves our statement. ■

Based on this lemma, the state information can be embedded in a code that delivers at each node the message $\mathbf{m}(t,i)$ directly and collaboratively. $\mathbf{m}(t,i)$ denotes the network information that node i needs to know in order to compute its consensus update.

We now present the multiple access scheme and the codes for $\mathbf{m}(t,i)$. Let $l'=\sum_{p'=1}^{p-1}Q_{p'}+k$ where $k=1\ldots,Q_p$. Then we choose the coefficients $s_i(t,l')$ as follows:

$$\boldsymbol{\theta}_p(t,i)\mapsto\bar{\boldsymbol{\theta}}_p(t,i)\mapsto s_i(t,l')=e^{j\phi_{l'}}\delta[q_{l'}-\bar{\boldsymbol{\theta}}_p(t,i)]\tag{13}$$

where $q_{l'}\in\{\mathcal{Q}_p\}$ and $\phi_{l'}\sim\mathcal{U}[0,2\pi)$. Now consider the decoder.

Lemma 4. *With these codes, when the adjacency matrix coefficients are equal to $a_{ij}=\alpha_{ij}/n$, the Maximum Likelihood (ML) estimate of $\{\mathbf{m}(t,i)\}_l,\ l=1,\ldots,Q$ in* (12) *is:*

$$\{\hat{\mathbf{m}}(t,i)\}_l=|r_i(t,l)|^2-\frac{N_0}{n}.\tag{14}$$

Proof: Let $a_{ij}=\frac{\alpha_{ij}}{n}$ for $i\neq j$. Consider an arbitrary $p=1,\ldots,m$. Let $l'=\sum_{p'=1}^{p-1}Q_{p'}+k$ where $k=1\ldots,Q_p$. Then, the quantity we want to estimate given $r_i(t,l')$ is- $\{\mathbf{m}(t,i)\}_{l'}=\frac{1}{n}\sum_{j=1}^{n}\alpha_{ij}\delta[q_{l'}-\bar{\boldsymbol{\theta}}_p(t,j)]$. Given $s_j(t,l')$, from (10) we have that-

$$r_i(t,l')\sim\mathcal{CN}\left(0,\frac{1}{n}\sum_{j=1}^{n}\alpha_{ij}|s_j(t,l')|^2+\frac{N_0}{n}\right).\tag{15}$$

From the coding scheme (13), $|s_j(t,l')|^2=\delta[q_{l'}-\bar{\boldsymbol{\theta}}_p(t,j)]$. Substituting this in (15) we get that $r_i(t,l')\sim\mathcal{CN}(0,\{\mathbf{m}(t,i)\}_{l'}+\frac{N_0}{n})$. Then the ML estimate is given by:

$$\{\hat{\mathbf{m}}(t,i)\}_{l'}=\arg\max\frac{\exp\left\{-|r_i(t,l')|^2/(\{\mathbf{m}(t,i)\}_{l'}+\frac{N_0}{n})\right\}}{\pi(\{\mathbf{m}(t,i)\}_{l'}+\frac{N_0}{n})}$$

The result follows upon simplification. ■

The encoding strategy is given in (13). To decode, each node i constructs -

$$\hat{\mathbf{u}}(t,i)=\sum_{l=1}^{Q}(\mathbf{q}_l-\bar{\boldsymbol{\theta}}(t,i))\{\hat{\mathbf{m}}(t,i)\}_{l'}\tag{16}$$

from the received signal $r_i(t)$. Based on Lemma 4, $\hat{\mathbf{u}}(t,i)$ is the estimate of $\bar{\mathbf{u}}(t,i)$. Now the usual vector consensus update of the form (1) can be applied using quantized states.

In the above scheme, the ML estimate of $\{\mathbf{m}(t,i)\}_l$ is obtained by using only one received signal sample. A simple way to reduce the error in estimating the network information is to use repetition coding. Suppose a $(\Psi,1)$ repetition code is employed by all nodes. Then the code corresponding to the coefficient $s_i(t,l)$ is transmitted by each node Ψ times. The cost of using the repetition code is a bandwidth expansion on the order of Ψ per iteration.

Corollary 1. *When a $(\Psi,1)$ repetition code is used, the ML estimate of $\{\mathbf{m}(t,i)\}_{l'}$ is-*

$$\{\hat{\mathbf{m}}(t,i)\}_{l'}=\frac{1}{\Psi}\sum_{s=1}^{\Psi}|r_{is}(t,l')|^2-\frac{N_0}{n}\tag{17}$$

where $r_{is}(t,l')$ denotes the s^{th} sample of $r_i(t,l')$.

In the remainder of the paper we will analyze the performance of these data driven codes and verify that it scales with the network size.

IV. Convergence

A. Convergence in Expectation

In terms of the average behavior, our algorithm behaves exactly like the standard average consensus algorithm using *quantized* states and therefore converges in expectation.

Lemma 5. *For any given initial state $\boldsymbol{\theta}_k(0)$:*

$$E\{\{\hat{\mathbf{m}}(t,i)\}_{l'}\} = \{\mathbf{m}(t,i)\}_{l'},\ \forall l'.$$

Therefore, the multiple access coding method proposed on average tends to the same result as quantized consensus.

Proof: Let $l' = \sum_{p'=1}^{p-1} Q_{p'} + k$. Now $r_i(t,l')$ is circularly symmetric complex gaussian (15). So $|r_i(t,l')|^2$ is exponential with parameter $(\frac{1}{n}\sum_{j=1}^{n}\alpha_{ij}\delta[q_{l'} - \bar{\boldsymbol{\theta}}_p(t,j)] + N_0/n)^{-1}$. Noting $a_{ij} = \alpha_{ij}/n$, irrespective of the repetition code order Ψ, it follows that $E\{\{\hat{\mathbf{m}}(t,i)\}_{l'}\} = \sum_{j=1}^{n} a_{ij}\delta[q_{l'} - \bar{\boldsymbol{\theta}}_p(t,j)] = \{\mathbf{m}(t,i)\}_{l'}$ as desired. ■

In the next section we prove a stronger result. We prove that in the asymptote $n \to \infty$, the algorithm converges to the initial quantized mean almost surely.

B. Asymptotic Convergence

Let the state of the network at iteration t be denoted by the mn-length quantized state vector $\bar{\boldsymbol{\theta}}(t)$. When $Q < \infty$ the evolution of the quantized system states $\bar{\boldsymbol{\theta}}(t)$ can be modeled as a finite Markov chain governed by a transition probability matrix that is a function of the previous states and the statistics of the channel. Define a set of system states $\boldsymbol{\mathcal{S}} = \{\bar{\boldsymbol{\theta}} : \bar{\boldsymbol{\theta}} = (\mathbf{1}_n \otimes \mathbf{c})\}$ where $\mathbf{c} \in \mathbb{R}^m$ is a constant such that each $\mathbf{c}_p \in \mathcal{Q}_p$. The set $\boldsymbol{\mathcal{S}}$ contains all possible quantized consensus states. In the absence of noise, i.e. $N_0 = 0$, $\boldsymbol{\mathcal{S}}$ represents the set of all absorbing states of the finite Markov chain since the data driven transmission scheduling and **a.4** guarantee that no further updates will occur in the system. We can show that as $n \to \infty$, $\boldsymbol{\mathcal{S}}$ remains a closed class even when $N_0 \neq 0$.

Lemma 6. *In the limit $n \to \infty$, $\boldsymbol{\mathcal{S}}$ is a closed class-*

$$\lim_{n\to\infty} P(\boldsymbol{\mathcal{S}}\ is\ closed) = 1. \tag{18}$$

Proof: Let $\bar{\boldsymbol{\theta}}(t) \in \boldsymbol{\mathcal{S}}$ i.e. $\bar{\boldsymbol{\theta}}(t,i) = \mathbf{c},\ \forall i$. Then $Pr(\boldsymbol{\mathcal{S}}\ is\ closed)$

$$\begin{aligned}
&= P\left(\bigcap_{p=1}^{n}\bigcap_{i=1}^{n}\bar{\boldsymbol{\theta}}_p(t+1,i) = \mathbf{c}_p|\bar{\boldsymbol{\theta}}(t) \in \boldsymbol{\mathcal{S}}\right) && (19)\\
&\stackrel{a}{=} \prod_{p=1}^{m}\prod_{i=1}^{n} P(\bar{\boldsymbol{\theta}}_p(t+1,i) = \mathbf{c}_p|\bar{\boldsymbol{\theta}}(t) \in \boldsymbol{\mathcal{S}}) && (20)\\
&\stackrel{b}{=} \prod_{p=1}^{m}\prod_{i=1}^{n}\left[Q\left(\frac{-\frac{\Delta_p}{2}}{\sigma_{\hat{\mathbf{u}}_p(t,i)}}\right) - Q\left(\frac{\frac{\Delta_p}{2}}{\sigma_{\hat{\mathbf{u}}_p(t,i)}}\right)|\bar{\boldsymbol{\theta}}(t) \in \boldsymbol{\mathcal{S}}\right]\\
&\stackrel{n\to\infty}{\to} 1 && (21)
\end{aligned}$$

where Δ_p is the quantizer step size for the pth component and $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{t^2/2}dt$. (a) follows from the mutual independence of $\boldsymbol{\theta}_p(t+1,i)|\bar{\boldsymbol{\theta}}(t)$ over p, i and (b) from using the consensus update equation and approximating the distribution of $\hat{\mathbf{u}}_p(t,i)$ as gaussian. Refer to the appendix for an expression for $VAR(\hat{\mathbf{u}}_p(t,i)) = \sigma^2_{\hat{\mathbf{u}}_p(t,i)}$. Given $\bar{\boldsymbol{\theta}}(t) \in \boldsymbol{\mathcal{S}}$, in the limit $n \to \infty, VAR(\hat{\mathbf{u}}_p(t,i)) \to 0$. Hence, $Q\left(\frac{-\frac{\Delta_p}{2}}{\sigma_{\hat{\mathbf{u}}_p(t,i)}}\right) \to 1$ while $Q\left(\frac{\frac{\Delta_p}{2}}{\sigma_{\hat{\mathbf{u}}_p(t,i)}}\right) \to 0$, giving us the result. ■

It is well known that if $\boldsymbol{\mathcal{S}}$ is the only closed class and $\exists\ t^* > t$ such that $P(\bar{\boldsymbol{\theta}}(t^*) \in \boldsymbol{\mathcal{S}}|\bar{\boldsymbol{\theta}}(t) \notin \boldsymbol{\mathcal{S}}) > 0$[1], then $lim_{t\longrightarrow\infty}P(\bar{\boldsymbol{\theta}}(t) \in \boldsymbol{\mathcal{S}}) = 1$. Based on this fact we have the following theorem which we state without proof-

Theorem 1. *Let $\boldsymbol{\mathcal{S}}$ denote the set of all absorbing states. The data driven algorithm converges to consensus almost surely in the limit $n \to \infty$, $\lim_{t\to\infty} P(\bar{\boldsymbol{\theta}}(t) \in \boldsymbol{\mathcal{S}}) = 1$.*

V. Mean Squared Error Performance in the Limit

From our construction, one iteration entails a delay equal to the interval T that is the duration of our orthogonal basis $\{c_l(\tau)\}_{l=1}^{Q}$. Since the data driven consensus algorithm accumulates quantization errors and channel errors, the mean squared error (MSE) after a certain delay is an essential performance metric.

The MSE can be decomposed into contributions from three separate terms $\in \mathbb{R}^{mn}$: (1) the convergence error $\boldsymbol{\theta}(t) - (J \otimes I)\boldsymbol{\theta}(0)$, (2) the quantization error $\boldsymbol{\theta}(t) - \bar{\boldsymbol{\theta}}(t)$, and (3) the channel error $\mathbf{e}(t) = \hat{\mathbf{u}}(t) - \mathbb{E}\{\hat{\mathbf{u}}(t)\}$. Note that quantization error (by **a.0**) and channel error (by definition) are zero mean.

We examine the asymptotic behavior of the MSE for a specific topology. Consider n nodes deployed in a circle such that $a_{ij} = \alpha_{ij}/n$, as specified in Lemma 4, with $\alpha_{ij} = \mathcal{K}(d^* + d_{ij})^{-\gamma}$. γ is the path loss exponent, d_{ij} is the distance between the nodes, and d^* and $\mathcal{K}$, related by $\mathcal{K} = (\frac{1}{d^*})^{-\gamma}$, are modeling parameters that take into account the carrier frequency, the scattering environment and antennae gains. Due to symmetry in path loss, $A = \frac{1}{n}circ(0, \alpha_{1,2}, \ldots, \alpha_{1,(n+1)/2}, \alpha_{1,(n+1)/2}, \ldots, \alpha_{1,2})$ for n odd (the case for n even is similar). First we obtain an expression for $\lambda_2(W)$ for the given network topology.

Lemma 7. *For the given topology, $\lambda_2(W) \stackrel{n\gg1}{\to} 1 - \frac{c\pi^2}{6} < 1$.*

Proof: From [5], using $n \gg 1$, we get- $\lambda_2(W) = 1 - \frac{4\pi^2}{n^3}\sum_{r=1}^{k'} r^2\alpha_{1,r} \leq 1 - \frac{\pi^2(n^2-1)}{6n^2}\min_r \alpha_{1,r} = 1 - \frac{c\pi^2}{6}$ where $k' = (n-1)/2$ and $c = \min_r \alpha_{1,r} = \mathcal{K}(d^* + d_{1,\frac{n-1}{2}})^{-\gamma}$. ■

Because of our selection of entries a_{ij}, the behavior of $\lambda_2(W)$ is different from that given in Lemma 2 for a similar topology. The asymptotic average MSE characterizing the tradeoff between the allocated communication resources and precision in consensus is given by the following lemma.

[1] In our case this statement involves a somewhat tedious proof which is omitted here for lack of space.

Lemma 8. *For the given circulant network, as $n \to \infty$, the average MSE at a finite iteration t is-*

$$\lim_{n\to\infty} MSE(t) \leq \frac{O(f(\bar{\Delta}))(1-\lambda_2^2(W))}{1-\lambda_2^2(W)-\frac{2\xi}{\Psi}} \quad (22)$$
$$+\left(\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn}-\frac{\frac{2\xi}{\Psi}O(f(\bar{\Delta}))}{1-\lambda_2^2(W)-\frac{2\xi}{\Psi}}\right)$$
$$\cdot\left(\lambda_2^2(W)+\frac{2\xi}{\Psi}\right)^t$$

where $\lambda_2(W) \leq 1-\frac{c\pi^2}{6} < 1$ with c some positive constant, $f(\bar{\Delta}) = \frac{1}{m}\sum_{p=1}^m \Delta_p^2$ where Δ_p is the quantization interval of the pth quantizer, Ψ is the order of the repetition code, and constant ξ satisfies the following conditions-

1) $\xi < \frac{\Psi}{2}(1-\lambda_2^2(W))$
2) $\xi \leq \frac{\Psi}{2}\left(\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn}(1-\lambda_2^2(W)\right)$
$\cdot\left(\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn}+O(f(\bar{\Delta}))\right)^{-1}$.

Proof: See Appendix. ■

This allows us to conclude the main result of this section, i.e. it is possible to achieve a bounded average MSE using limited communication resources irrespective of the number of sensors. Finally, we note that by allocating more communication resources i.e. greater quantizer precision and increasing the order of the repetition code Ψ, the performance of the data driven codes for an infinitely large network approaches that of standard average consensus and is captured by the following corollary-

Corollary 2. *For the given circulant network, in the case of bandwidth expansion, the average MSE at a finite iteration t is-*

$$\lim_{n\to\infty} MSE(t) \leq \lambda_2^{2t}(W)\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn}. \quad (23)$$

VI. Conclusion

Packet-based average consensus gossiping does not scale in dense wireless networks due to network congestion. We proposed a solution to this problem by considering a joint source channel coding scheme which utilizes cooperative transmission through data driven scheduling. We showed that as the network becomes increasingly dense the algorithm converges almost surely. We characterized the MSE performance and further showed that adding infinitely many values with finite precision is possible in spite of the bandwidth limitations.

VII. Appendix

Proof Sketch of Lemma 8:

$$\begin{aligned}\boldsymbol{\theta}(t+1) &= \bar{\boldsymbol{\theta}}(t)+\hat{\mathbf{u}}(t)\\ &= \bar{\boldsymbol{\theta}}(t)+\mathbb{E}\{\hat{\mathbf{u}}(t)\}+\hat{\mathbf{u}}(t)-\mathbb{E}\{\hat{\mathbf{u}}(t)\} \quad (24)\\ &= (W\otimes I)\bar{\boldsymbol{\theta}}(t)+\mathbf{e}(t) \quad (25)\end{aligned}$$

where the last equality follows from Lemma 5 i.e. the average behavior of data driven consensus is the same as standard average consensus using quantized states. Writing the quantized state vector $\bar{\boldsymbol{\theta}}(t)$ in terms of the quantization error $\mathbf{v}(t)$, whose statistics were defined in equations (7)-(8), and expanding the above recursion we get-

$$\begin{aligned}\boldsymbol{\theta}(t) &= (W\otimes I)^t\boldsymbol{\theta}(0)+\sum_{i=1}^t(W\otimes I)^i\mathbf{v}(t-i)\\ &\quad +\sum_{i=1}^t(W\otimes I)^{i-1}\mathbf{e}(t-i). \quad (26)\end{aligned}$$

Recall that $\mathbf{v}(t_1)$ and $\mathbf{v}(t_2)$, $t_1 \neq t_2$ are uncorrelated. Given $\boldsymbol{\theta}(0)$, so are $\mathbf{e}(t_1)$ and $\mathbf{e}(t_2)$. Using the above expression for $\boldsymbol{\theta}(t)$ in the definition of MSE, with some algebra we can get-

$$\begin{aligned}MSE(t) &= \frac{1}{mn}\mathbb{E}\{||((W^t-J)\otimes I)\boldsymbol{\theta}(0)||^2\} \quad (27)\\ &\quad +\frac{1}{mn}\sum_{i=t}^1\mathbb{E}\{||(W^i\otimes I)\mathbf{v}(t-i)||^2\} \quad (28)\\ &\quad +\frac{1}{mn}\sum_{i=t}^1\mathbb{E}\{||(W^{i-1}\otimes I)\mathbf{e}(t-i)||^2\} \quad (29)\end{aligned}$$

We know the first term through Lemma 1. Consider the quantization error (28). Taking the trace of the scalar norm and reordering its argument, we get-

$$\begin{aligned}&\frac{1}{mn}\sum_{i=t}^1\mathbb{E}\{||(W^i\otimes I)\mathbf{v}(t-i)||^2\} \quad (30)\\ &= \frac{1}{mn}\sum_{i=t}^1 Tr\left(\mathbb{E}\{\mathbf{v}(t-i)\mathbf{v}^T(t-i)\}(W^{2i}\otimes I)\right) \quad (31)\\ &= \frac{1}{mn}\sum_{i=t}^1\sum_{p=1}^m\frac{\Delta_p^2}{12}\sum_{j=1}^n\lambda_j^{2i}(W) \quad (32)\\ &\leq f(\bar{\Delta})\left(\frac{t}{n}+\max_{2\leq j\leq n}\frac{\lambda_j^2(W)(1-\lambda_j^{2t})(W)}{1-\lambda_j^2(W)}\right) \quad (33)\\ &= f(\bar{\Delta})\left(\frac{t}{n}+\frac{\lambda_2^2(W)(1-\lambda_2^{2t}(W))}{1-\lambda_2^2(W)}\right) \quad (34)\end{aligned}$$

where $[\mathbb{E}\{\mathbf{v}(t-i)\mathbf{v}^T(t-i)\}]_{jp,jp} = VAR(\mathbf{v}_p(t-i,j)) = \frac{C^2\sigma_p^2}{12Q_p^2} = \frac{\Delta_p^2}{12}$ for $j = 1,\ldots,n$ and $p = 1,\ldots,m$, gives (32). Noting that $\lambda_1(W) = 1$ and defining $f(\bar{\Delta}) = \frac{1}{m}\sum_{p=1}^m\frac{\Delta_p^2}{12}$ gives (33). For any $t \geq 0$, the second largest eigenvalue of W, $\lambda_2(W)$ with $\lambda_2^2(W) \in (0,1)$, maximizes the term in (33). Finally, note that-

$$\lim_{n\to\infty}\frac{1}{mn}\sum_{i=t}^1\mathbb{E}\{||(W^i\otimes I)\mathbf{v}(t-i)||^2\} = O(f(\bar{\Delta})). \quad (35)$$

Now consider the last error term (29). By definition, $\mathbb{E}\{\mathbf{e}(t)\} = 0, \forall t$, and $[\mathbb{E}\{\mathbf{e}(t-i)\mathbf{e}^T(t-i)\}]_{jp,jp} = VAR(\hat{\mathbf{u}}_p(t-i,j))$. Then proceeding like before,

$$\begin{aligned}&\frac{1}{mn}\sum_{i=t}^1\mathbb{E}\{||(W^{i-1}\otimes I)\mathbf{e}(t-i)||^2\}\\ &= \frac{1}{mn}\sum_{i=t}^1\sum_{j=1}^n\sum_{p=1}^m VAR(\hat{\mathbf{u}}_p(t-i,j))\lambda_j^{2(i-1)}(W) \quad (36)\end{aligned}$$

Now, $VAR(\hat{\mathbf{u}}_p(t,j)|\bar{\boldsymbol{\theta}}(0)) = \frac{1}{\Psi}\sum_{l=1}^{Q_p}(q_l - \bar{\boldsymbol{\theta}}_p(t,j))^2\left(\frac{N_0}{n}\right)^2$

$$\begin{aligned}
&+ \quad \frac{2}{\Psi}\left(\frac{N_0}{n}\right)\sum_{s=1}^{n}\frac{\alpha_{js}^2}{n^2}(\bar{\boldsymbol{\theta}}_p(t,s) - \bar{\boldsymbol{\theta}}_p(t,j))^2 \\
&+ \quad \frac{1}{\Psi}\sum_{s=1}^{n}\frac{\alpha_{js}}{n}(\bar{\boldsymbol{\theta}}_p(t,s) - \bar{\boldsymbol{\theta}}_p(t,j))^2 \\
&\quad \cdot\sum_{s'=1}^{n}\frac{\alpha_{js'}}{n}I_{[\bar{\boldsymbol{\theta}}_p(t,s')=\bar{\boldsymbol{\theta}}_p(t,j)]}. \qquad (37)
\end{aligned}$$

where Ψ is the order of the repetition code. Define $\boldsymbol{\beta}(t) \triangleq \bar{\boldsymbol{\theta}}(t) - (J\otimes I)\boldsymbol{\theta}(t)$. We simplify the last term above as follows-

$$\frac{1}{\Psi}\sum_{s=1}^{n}\frac{\alpha_{js}}{n}(\bar{\boldsymbol{\theta}}_p(t,s) - \bar{\boldsymbol{\theta}}_p(t,j))^2\sum_{s'=1}^{n}\frac{\alpha_{js'}}{n}I_{[\bar{\boldsymbol{\theta}}_p(t,s')=\bar{\boldsymbol{\theta}}_p(t,j)]}$$

$$\begin{aligned}
&\leq \frac{\xi_j}{\Psi n}\sum_{s=1}^{n}(\bar{\boldsymbol{\theta}}_p(t,s) - \bar{\boldsymbol{\theta}}_p(t,j))^2 \qquad (38)\\
&= \frac{\xi_j}{\Psi n}\sum_{s=1}^{n}\left(||\boldsymbol{\beta}_p(t,s) - \boldsymbol{\beta}_p(t,j)\mathbf{1}||^2\right) \qquad (39)\\
&= \frac{\xi_j}{\Psi n}\left(||\boldsymbol{\beta}_p(t)||^2 + n|\boldsymbol{\beta}_p(t,j)|^2\right) \qquad (40)
\end{aligned}$$

where the cross-term $2|\boldsymbol{\beta}_p(t,j)|\boldsymbol{\beta}_p(t)^T\mathbf{1}$ can be neglected because $\boldsymbol{\beta}_p(t)^T\mathbf{1} = (\bar{\boldsymbol{\theta}}_p(t) - \boldsymbol{\theta}_p(t))^T\mathbf{1} = \mathbf{v}_p(t)^T\mathbf{1}$ using $J\mathbf{1} = \mathbf{1}$, and $\lim_{n\to\infty}\mathbf{v}_p(t)^T\mathbf{1} = \lim_{n\to\infty}\sum_{j=1}^{n}\mathbf{v}_p(t,j) = \mathbb{E}\{\mathbf{v}_p(t,j)\} = 0$ by WLLN. ξ_j is some constant that upper bounds $(\max_s\alpha_{js})\sum_{s'=1}^{n}\frac{\alpha_{js'}}{n}I_{[\bar{\boldsymbol{\theta}}_p(t,s')=\bar{\boldsymbol{\theta}}_p(t,j)]}$. By definition, $\alpha_{js'} < 1,\ \forall j,s'$, so $\xi_j < 1$. Thus, we can upper bound the variance as, $VAR(\hat{\mathbf{u}}_p(t,j)) \leq \frac{\xi_j}{\Psi n}\mathbb{E}\{||\boldsymbol{\beta}_p(t)||^2 + n|\boldsymbol{\beta}_p(t,j)|^2\} + O\left(\frac{N_0}{\Psi n}\right) + O\left(\frac{N_0^2}{\Psi n^2}\right)$ where the last two terms vanish either when the SNR is high, n is large, or a high order repetition code is used. Therefore,

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=t}^{1}\mathbb{E}\{||(W^{i-1}\otimes I)\mathbf{e}(t-i)||^2\} \qquad (41)\\
&\leq \frac{2\xi}{mn}\frac{1}{\Psi}\sum_{i=t}^{1}\lambda_2^{2(i-1)}(W)\mathbb{E}\{||\boldsymbol{\beta}(t-i)||^2\} + O\left(\frac{1}{n}\right)\\
&\quad + O\left(\frac{N_0}{\Psi n}\right) + O\left(\frac{N_0^2}{\Psi n^2}\right) \qquad (42)
\end{aligned}$$

where $\xi = \max_j \xi_j$. Now, using the recursion (26), $\boldsymbol{\beta}(k)$ can be written as follows-

$$\begin{aligned}
\boldsymbol{\beta}(k) &= \bar{\boldsymbol{\theta}}(k) - (J\otimes I)\boldsymbol{\theta}(k) = ((I-J)\otimes I)\boldsymbol{\theta}(k) + \mathbf{v}(k)\\
&= ((W^t - J)\otimes I)\boldsymbol{\theta}(0) + \sum_{l=k}^{1}((W-J)^l\otimes I)\mathbf{v}(k-l)\\
&\quad + \mathbf{v}(k) + \sum_{l=k}^{1}((W-J)^{l-1}\otimes I)\mathbf{e}(k-l). \qquad (43)
\end{aligned}$$

Noting that $W^kJ = J, \forall k$, with some algebra we get-

$$\frac{1}{mn}\mathbb{E}\{||\boldsymbol{\beta}(k)||^2\} \leq MSE(k) + \frac{\mathbb{E}\{||\mathbf{v}(k)||^2\}}{mn} \qquad (44)$$

where $\frac{\mathbb{E}\{||\mathbf{v}(k)||^2\}}{mn} = \frac{1}{m}\sum_{p=1}^{m}\frac{\Delta_p^2}{12} = f(\bar{\Delta})$. Therefore, we finally get-

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=t}^{1}\mathbb{E}\{||(W^{i-1}\otimes I)\mathbf{e}(t-i)||^2\} \qquad (45)\\
&\leq \frac{2\xi}{\Psi}\sum_{i=t}^{1}\lambda_2^{2(i-1)}(W)MSE(t-i) + O(f(\bar{\Delta})) \qquad (46)\\
&\quad + O\left(\frac{1}{n}\right) + O\left(\frac{N_0}{\Psi n}\right) + O\left(\frac{N_0^2}{\Psi n^2}\right) \qquad (47)
\end{aligned}$$

since $\frac{2\xi}{\Psi}\sum_{i=t}^{1}\lambda_2^{2(i-1)}(W)$ converges to some constant. Putting everything together and taking the limit $n\to\infty$,

$$\begin{aligned}
\lim_{n\to\infty}MSE(t) &\leq \lambda_2^{2t}\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn} + O(f(\bar{\Delta})) \qquad (48)\\
&\quad + \frac{2\xi}{\Psi}\lambda_2^{2(t-1)}(W)u(t-1) * MSE(t)
\end{aligned}$$

where $*$ denotes convolution and $u(t)$ is the discrete-time unit step function. We can simplify this expression further using frequency-domain techniques. Let $M(z) \triangleq \sum_{k=0}^{\infty}MSE(k)z^{-k}$. Then, taking the unilateral Z-transform of both sides of (49) and rearranging terms, we have that- $M(z) \leq (\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn}(1-z^{-1}) + O(f(\bar{\Delta}))(1-\lambda_2^2(W)z^{-1}))\times((1-(\lambda_2^2(W)+\frac{2\xi}{\Psi})z^{-1})(1-z^{-1}))^{-1}$.

The inverse Z-transform of this expression yields the solution-

$$\begin{aligned}
\lim_{n\to\infty}MSE(t) &\leq \frac{O(f(\bar{\Delta}))(1-\lambda_2^2(W))}{1-\lambda_2^2(W)-\frac{2\xi}{\Psi}}u(t) \qquad (49)\\
&+ \left(\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn} - \frac{\frac{2\xi}{\Psi}O(f(\bar{\Delta}))}{1-\lambda_2^2(W)-\frac{2\xi}{\Psi}}\right)\\
&\cdot\left(\lambda_2^2(W)+\frac{2\xi}{\Psi}\right)^t.
\end{aligned}$$

For the convergence of this solution, the following must hold - (1) $\lambda_2^2(W) + \frac{2\xi}{\Psi} < 1$ and (2) $\lim_{n\to\infty}\frac{\mathbb{E}\{||\boldsymbol{\theta}(0)||^2\}}{mn} - \frac{\frac{2\xi}{\Psi}O(f(\bar{\Delta}))}{1-\lambda_2^2(W)-\frac{2\xi}{\Psi}} \geq 0$. These conditions are equivalent to those stated in the lemma and they can be satisfied by choosing Ψ appropriately. This completes the proof.

References

[1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *Special Issue IEEE Trans. Information Theory and IEEE/ACM Transactions on Networking*, vol. 52, no. 6, pp. 2508 – 2530, June 2006.

[2] 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(9), pp. 1520 – 1533, Sept. 2004.

[3] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient aggregation for sensor networks," in *Proc. IPSN*, Nashville, TN, April 2006.

[4] 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.

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

[6] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388 – 404, Mar. 2000.

[7] S. Kirti, A. Scaglione, and R. J. Thomas, "A scalable wireless communication architecture for average consensus," in *Proc. 46th IEEE Conf. on Decision and Control*, New Orleans, LA, USA, Dec. 2007.