Fast Average Consensus in Clustered Wireless Sensor Networks by Superposition Gossiping

Meng Zheng[†‡], Mario Goldenbaum[⋆◇], Sławomir Stańczak[⋆◇] and Haibin Yu[†]

[†] Key Lab of Industrial Informatics, SIA, Chinese Academy of Sciences, Shenyang, 110016 China

[‡] Graduate School of the Chinese Academy of Sciences, Beijing, 100039 China

[⋆] Heinrich-Hertz-Lehrstuhl für Informationstheorie und theoretische Informationstechnik,
Technische Universität Berlin, Einsteinufer 25, 10587 Berlin, Germany

[◇]Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute,
Einsteinufer 37, 10587 Berlin, Germany

Abstract—**In this paper we propose a gossip algorithm for average consensus in clustered wireless sensor networks called superposition gossiping, where the nodes in each cluster exploit the natural superposition property of wireless multiple-access channels to significantly decrease local averaging times. More precisely, the considered network is organized into single-hop clusters and in each cluster average values are computed at a designated cluster head via the wireless channel and subsequently broadcasted to update the entire cluster. Since the clusters are activated randomly in a time division multiple-access fashion, we can apply well-established techniques for analyzing gossip algorithms to prove the convergence of the algorithm to the average consensus in the second moment and almost surely, provided that some connectivity condition between clusters is fulfilled. Finally, we follow a semidefinite programming approach to optimize wake up probabilities of cluster heads that further accelerates convergence.**

I. INTRODUCTION

Fast-convergent average consensus algorithms that efficiently exploit wireless resources (e.g., bandwidth, energy) are highly desired in wireless sensor networks (WSN) as building blocks for more advanced distributed signal processing and optimization approaches [1], [2]. In particular, gossip algorithms have received a great deal of attention in recent years because they allow sensor nodes to distributively achieve a network-wide consensus without a routing protocol by letting nodes only locally exchange data with their nearest neighbors (see for example [1], [3]–[6] and references therein).

With few exceptions (e.g., [7]–[9]), when designing gossip algorithms for average consensus, the inherent broadcast property of the wireless channel is usually seen to be obstructive so that the design objective is to avoid interference caused by simultaneously transmitting nodes via standard protocols that orthogonalize transmissions in time or frequency. However, since averaging is essentially superposition and the exchange of raw sensor readings between nodes is in this context not necessary, the broadcast property of wireless channels can beneficially be exploited [9], [10]. Therefore, we propose in this paper an iterative *superposition gossiping* for any WSN that is organized into a fixed number of single-hop clusters. In a randomly chosen/activated cluster, the local average-iteration is computed at the cluster head in a single step by

letting the nodes transmit simultaneously their current states. Subsequently, the cluster head updates the entire cluster by broadcasting the computed local average. The algorithm leads to significant improvements with respect to convergence speed, provided that some connectivity condition between clusters is fulfilled.

Boyd et al. presented in [3] a framework for the design and the analysis of randomized gossip algorithms for average consensus in arbitrary connected networks, where pairs of nodes are chosen randomly to exchange the data. They found that the convergence time of such a random pairwise gossiping (RPG) depends on the second largest eigenvalue of the iterative matrix that characterizes the algorithm. To analyze the principal convergence behavior of superposition gossiping, we consider in this paper the averaging over ideal multiple-access channels and find out that the underlying weight matrix exhibits properties that allow us to apply well-known techniques from [3] to prove convergence in the second moment and almost surely. Since the convergence speed of superposition gossiping depends on the second largest eigenvalue of the weight matrix as well, we follow a semidefinite programming approach to optimize the probabilities with which cluster heads are activated. This results in a considerable improvement in convergence to the global average.

A. Related Work

It seems that [7]–[9] are the first publications in which the authors recognized that the broadcast property as well as the superposition property of wireless channels can be useful in average consensus problems. The broadcast gossip algorithm presented in [8] relies for example on the fact that all nodes within a certain connectivity radius can benefit from the current state of a single randomly chosen transmitting node. On the other hand, Kirti et al. exploit in [7] the multiple-access aspect of the physical layer, while similar to our work in this paper, Nazer, Dimakis and Gastpar take in [9] advantage of both, the superposition and the broadcast property of wireless channels.

The main difference to our work lies in the fact that the considerations in [9] were primarily focused on information

theoretical aspects by using computation coding [11], while our work exploits the idea of analog computation via multiple-access channels [10].

Gossip algorithms are generally designed for WSNs with planar topology. In contrast to this paper, [9] considers only implicitly a simple clustering structure that changes randomly over time. Another publication that incorporates clustering techniques for average consensus can be found in [12]. However, the applicability of the results may be limited, since (i) the proposed cluster-based fixed iterations require that all nodes are globally synchronized, which is illusive to ensure in large-scale networks, and (ii) it is assumed that averages within clusters impact the remaining clusters only by a costly information exchange between cluster heads.

B. Paper Organization

The paper is organized as follows. Section II presents the system model, while our superposition gossip algorithm is outlined in Section III. Section IV is devoted to an extensive convergence analysis. Afterwards, in Section V, the convergence rate of our proposed algorithm is further be improved by an optimization of the wake up probabilities of cluster heads. Some numerical examples in Section VI demonstrate the performance of superposition gossiping in comparison to some alternative algorithms, and finally, Section VII concludes the paper.

C. Notational Remarks

Vectors are denoted by bold lowercase letters and matrices by bold uppercase letters, respectively. The sets of real, nonnegative real, nonnegative integer and natural numbers are described by $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{Z}_+$ and $\mathbb{N}$. The length n vector of all ones is denoted by $\mathbf{1}_n$ and the $n \times n$ identity matrix and matrix of all zeros by $\boldsymbol{I}_n$ and $\mathbf{0}_{n \times n}$, respectively. " $\preceq$ " denotes matrix inequality, that is $\boldsymbol{A} \preceq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is negative semidefinite.

II. SYSTEM MODEL

A. Network Model

Consider a wireless sensor network consisting of $N \in \mathbb{N}$ spatially distributed nodes that are organized into a fixed set of clusters $\mathcal{C} := \{\mathcal{C}_1, \ldots, \mathcal{C}_{|\mathcal{C}|}\}$, where $\mathcal{C}_i$ denotes the set of nodes belonging to cluster i, $i = 1, \ldots, |\mathcal{C}|$, and $|\mathcal{C}_i| = N_i \in \mathbb{N}$ the corresponding number of nodes, respectively. The nodes in a cluster are arbitrarily numbered such that $\mathcal{C}_i = \{0_i, \ldots, N_i - 1\}$, with 0_i the label of the ith designated cluster head (see Fig. 1).[1]

Definition 1 (Connected Clusters). Two clusters in a clustered WSN, say $\mathcal{C}_i$ and $\mathcal{C}_j$, $i, j = 1, 2, \ldots, |\mathcal{C}|$ and $i \neq j$, are called *connected* if $\mathcal{C}_i$ and $\mathcal{C}_j$ share at least one common node (i.e., $\mathcal{C}_i \bigcap \mathcal{C}_j \neq \varnothing$).

---

[1]Note that the numbering of nodes does not mean that they have unique identities which could be used for computations.
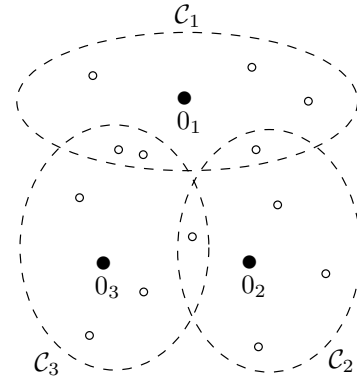

Fig. 1. Qualitative representation of a connected clustered WSN $\mathcal{C}$ consisting of $N = 16$ nodes and clusters $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$, where 0_i denotes the cluster head of cluster i, $i = 1, 2, 3$.

Definition 2 (Connected Clustered WSN). A clustered WSN is called *connected* if for any two clusters $\mathcal{C}_i$ and $\mathcal{C}_j$, $i \neq j$, there exists a sequence of connected clusters from $\mathcal{C}_i$ to $\mathcal{C}_j$.

According to Definitions 1 and 2 we are able to describe a clustered WSN by an undirected graph $\mathcal{G} = (\mathcal{C}, \mathcal{E})$, with $\mathcal{C}$ the set of clusters as above and $\mathcal{E} = \{(i, j)\}$ the set of edges.

Remark 1. Note that $(k, \ell) \in \mathcal{E}$ means that clusters $\mathcal{C}_k$ and $\mathcal{C}_\ell$ are connected. If the clustered WSN is connected, of course the corresponding graph $\mathcal{G}$ is connected as well.

See Fig. 1 for an example of a connected clustered WSN.

B. Time Model

Based on the asynchronous time model in [3], we assume that cluster heads wake up according to a clock that ticks independently at a rate $\mu_i \in \mathbb{R}_+$ Poisson process, $i = 1, \ldots, |\mathcal{C}|$. We adjust μ_i such that in a sufficiently small time interval and with high probability two cluster heads do not wake up simultaneously.

C. Intra-Cluster Communication

Let the nodes jointly observe a physical phenomenon resulting in sensor readings $x_i \in \mathcal{X} \subset \mathbb{R}$, $i = 1, \ldots, N$, such that the vector $\boldsymbol{x}(0) := [x_1(0) = x_1, \ldots, x_N(0) = x_N]^T \in \mathcal{X}^N$ denotes the initial state of the network. Then, intra-cluster communication can usually be described by the standard affine model of a wireless multiple-access channel (MAC). That is, if cluster $\mathcal{C}_i$, $i = 1, \ldots, |\mathcal{C}|$, is active at time $t \in \mathbb{Z}_+$, the real signal received by cluster head 0_i can be written as

$$y_{0_i}(t) = \sum_{k \in \mathcal{C}_i \setminus \{0_i\}} h_{ik}(t) s_k\big(x_k(t)\big) + n_i(t), \quad (1)$$

where $s_k : \mathbb{R} \to \mathbb{R}$ denotes a transmit signal of node $k \in \mathcal{C}_i$ depending on state $x_k(t)$, $h_{ik} \in \mathbb{R}$ a flat-fading coefficient between node k and cluster head 0_i and $n_i \in \mathbb{R}$ receiver noise at 0_i, respectively. If we ignore fading and noise in combination with transmit signals $s_k(x) \equiv x$, for all $k \in \mathcal{C}_i$,

Algorithm 1 Superposition Gossiping

1: Initialization: $t = 0$, $x_k(0) = x_k \in \mathcal{X}$, $k = 1, \ldots, N$
2: At $t \in \mathbb{Z}_+$ the clock of the cluster head of cluster $i \in \{1, \ldots, |\mathcal{C}|\}$ ticks randomly
3: Cluster head 0_i wakes up all nodes in cluster $\mathcal{C}_i \in \mathcal{C}$
4: Nodes $1_i, \ldots, N_i - 1$ of $\mathcal{C}_i$ transmit their values simultaneously to cluster head 0_i, resulting in receive signal

$$y_{0_i}(t) = \sum_{k \in \mathcal{C}_i \backslash \{0_i\}} x_k(t) \qquad (3)$$

5: Cluster head 0_i computes the local average

$$x_{0_i}(t+1) = \frac{1}{N_i}\big(x_{0_i}(t) + y_{0_i}(t)\big) \qquad (4)$$

6: Cluster head 0_i broadcasts the local average $x_{0_i}(t+1)$ resulting in updates

$$x_k(t+1) = \begin{cases} x_{0_i}(t+1) & , k \in \mathcal{C}_i \backslash \{0_i\} \\ x_k(t) & , k \notin \mathcal{C}_i \end{cases} , \qquad (5)$$

$k = 1, \ldots, N$
7: Back to step 2 until convergence (in some sense)

---

(1) reduces to an *ideal MAC*

$$y_{0_i}(t) = \sum_{k \in \mathcal{C}_i \backslash \{0_i\}} x_k(t) , \qquad (2)$$

which reveals that the natural mathematical operation of a wireless MAC is simply *superposition*.

III. Superposition Gossiping for Clustered Wireless Sensor Networks

The distributed average-consensus problem is to design adequate interaction rules between the nodes of the network (i.e., a protocol) such that the state vector $\boldsymbol{x}(t) \to \bar{x}\mathbf{1}_N$ as fast as possible, with $\bar{x} := \frac{1}{N}\sum_{i=1}^{N} x_i(0)$ the average over the initial values. To achieve this, most of the cited work try to avoid the superposition (2) caused by simultaneous transmissions by using standard protocols such as time-division multiple access. However, since *averaging is essentially superposition*, the interference property as well as the broadcast property of wireless channels can profitably be exploited to significantly accelerate the averaging in clusters [10], [11].

This observation leads us to the following gossiping procedure. Let randomly wake up any cluster head according to the time model in Section II-B that subsequently wakes up all remaining nodes in the cluster. Then, the nodes transmit simultaneously their current state values to the cluster head, resulting in a receive signal (2) that enables the cluster head to compute the average of the entire cluster in a single step since it is not interested in raw individual state values. Finally, if it broadcasts the average to all nodes in the cluster, a local average-consensus is achieved in merely two steps. The resulting algorithm, which we denote as *superposition gossiping*, is formally stated in Algorithm 1.

Remark 2. Although practical gossip algorithms suffer from limitations such as power constraints, fading, receiver noise, synchronization issues, we focus in this paper on the principal behavior of superposition gossiping and consider therefore averaging over ideal MACs (2) only. Extensions to realistic MACs (1) follow along similar lines such as in [10], [13] and will be part of future work.

IV. Convergence Analysis

In this section, we study the convergence behavior of superposition gossiping for connected clustered WSNs. If we denote by $\boldsymbol{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ the state vector at time $t \in \mathbb{Z}_+$, evolving from initial state $\boldsymbol{x}(0)$, Algorithm 1 can be concisely summarized by the linear equation

$$\boldsymbol{x}(t+1) = \boldsymbol{W}(t)\boldsymbol{x}(t) , \qquad (6)$$

where for every $t \in \mathbb{Z}_+$, matrix $\boldsymbol{W}(t) \in \mathbb{R}^{N \times N}$ is randomly chosen from the set $\{\boldsymbol{W}_i\}_{i=1}^{|\mathcal{C}|}$ of weight matrices that consist of elements

$$W_{jk}^{(i)} := \begin{cases} 1 & , \text{if } j \notin \mathcal{C}_i, k = j \\ \frac{1}{N_i} & , \text{if } j, k \in \mathcal{C}_i \\ 0 & , \text{else} \end{cases} . \qquad (7)$$

In other words, $\boldsymbol{W}(t) = \boldsymbol{W}_i$ if the clock of cluster head 0_i ticks at time t. The following useful property of the symmetric matrices $\boldsymbol{W}_i$ can be easily proven.

Lemma 1. The weight matrices $\boldsymbol{W}_i$, $i = 1, \ldots, |\mathcal{C}|$, are doubly stochastic.

Now, let us denote the expectation of weight matrices $\boldsymbol{W}_i$ as $\boldsymbol{W} = (W_{jk}) := \sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{W}_i$, where $p_i \geq 0$, $\sum_{i=1}^{|\mathcal{C}|} p_i = 1$, denotes the probability that $\boldsymbol{W}_i$ is chosen independently out of $\{\boldsymbol{W}_i\}_{i=1}^{|\mathcal{C}|}$. Because the expected matrix $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ is important for analyzing the convergence of Algorithm 1, we start by proving some properties.

Let $\mathcal{G}(\boldsymbol{W})$ be defined as the directed graph of node set $\mathcal{V} := \{1, \ldots, N\}$, where a directed edge leading from node $j \in \mathcal{V}$ to node $i \in \mathcal{V}$ exists, if and only if $W_{ij} \neq 0$.

Lemma 2. If $\mathcal{G}$ is connected, then $\mathcal{G}(\boldsymbol{W})$ is strongly connected.

Proof: For any two nodes $m \in \mathcal{V}$ and $n \in \mathcal{V}$, it suffices to consider the following two cases:

(i) Let m and n be distinct nodes in the same cluster $\mathcal{C}_i$. Then, from (7) follows immediately $W_{mn}^{(i)} = \frac{1}{N_i} > 0$ and according to the definition of the expectation $\boldsymbol{W}$ that

$$W_{mn} = \sum_{i=1}^{|\mathcal{C}|} p_i W_{mn}^{(i)} > 0 . \qquad (8)$$

(ii) Now, let nodes m and n belong to two different clusters $\mathcal{C}_i$ and $\mathcal{C}_j$, respectively. Since $\mathcal{G}$ is connected, there exists a sequence of connected cluster pairs from $\mathcal{C}_i$ to $\mathcal{C}_j$. Consider therefore the sequence of length ℓ, $1 \leq \ell \leq N - 2$, of clusters $(\mathcal{C}_i, \mathcal{C}_{i_1}), (\mathcal{C}_{i_1}, \mathcal{C}_{i_2}), \ldots, (\mathcal{C}_{i_\ell}, \mathcal{C}_j)$, and representatives of the common nodes $m_1 \in \mathcal{C}_i \bigcap \mathcal{C}_{i_1}, m_2 \in$

$\mathcal{C}_{i_1} \bigcap \mathcal{C}_{i_2}, \ldots, m_{\ell+1} \in \mathcal{C}_{i_\ell} \bigcap \mathcal{C}_j$. Then, we get a sequence of weight matrices $\boldsymbol{W}_i, \boldsymbol{W}_{i_1}, \ldots, \boldsymbol{W}_{i_\ell}, \boldsymbol{W}_j$, where the elements $W_{mm_1}^{(i)}, W_{m_1 m_2}^{(i_1)}, \ldots, W_{m_\ell m_{\ell+1}}^{(i_\ell)}, W_{m_{\ell+1} n}^{(j)}$ are all positive. Thus, according to (8), the corresponding elements $W_{mm_1}, W_{m_1 m_2}, \ldots, W_{m_\ell m_{\ell+1}}, W_{m_{\ell+1} n}$ of $\boldsymbol{W}$ are positive as well, which means that there is a path from node $m \in \mathcal{V}$ to node $n \in \mathcal{V}$ and vice versa. ∎

In order to achieve an average consensus in expectation (i.e., $\boldsymbol{x}(t) \to \bar{x}\mathbf{1}_N$ in expectation for $t \to \infty$), it can be easily concluded that the tth power $\boldsymbol{W}^t$ of $\boldsymbol{W}$ has to converge in expectation to $\boldsymbol{J} := \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T$, as t tends to infinity. From [3] we already know that to ensure this, $\boldsymbol{W}$ has to fulfill the following two properties:

(P.1) $\boldsymbol{W}$ is doubly stochastic.

(P.2)   The spectral radius $\rho(\boldsymbol{W} - \boldsymbol{J})$ of matrix $\boldsymbol{W} - \boldsymbol{J}$ is smaller than 1.

Lemma 3. The average weight matrix $\boldsymbol{W}$ of superposition gossiping satisfies properties (P.1) and (P.2).

    Proof: (P.1): Lemma 1 in conjunction with (8) leads to

$$\boldsymbol{W}\mathbf{1}_N = \left(\sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{W}_i\right)\mathbf{1}_N = \sum_{i=1}^{|\mathcal{C}|} p_i \mathbf{1}_N = \mathbf{1}_N,$$

and

$$\mathbf{1}_N^T \boldsymbol{W} = \mathbf{1}_N^T \left(\sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{W}_i\right) = \sum_{i=1}^{|\mathcal{C}|} p_i \mathbf{1}_N^T = \mathbf{1}_N^T .$$

(P.2): We start by stating some results on the eigenvalues of $\boldsymbol{W}$. According to (7), there exists for each $i = 1, \ldots, |\mathcal{C}|$ an orthogonal matrix $\boldsymbol{C}_i \in \mathbb{R}^{N \times N}$ such that

$$\boldsymbol{W}_i = \boldsymbol{C}_i \begin{pmatrix} \boldsymbol{I}_{r_i} & \\ & \mathbf{0}_{N_i \times N_i} \end{pmatrix} \boldsymbol{C}_i^T ,$$

with $r_i := \text{rank}(\boldsymbol{W}_i) = N - N_i$. Then, we have

$$0 \leq \boldsymbol{z}^T \boldsymbol{W} \boldsymbol{z} = \sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{z}^T \boldsymbol{W}_i \boldsymbol{z}$$
$$= \sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{z}^T \boldsymbol{C}_i \begin{pmatrix} \boldsymbol{I}_{r_i} & \\ & \mathbf{0}_{N_i \times N_i} \end{pmatrix} \boldsymbol{C}_i^T \boldsymbol{z}$$
$$\leq \sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{z}^T \boldsymbol{z} = \boldsymbol{z}^T \boldsymbol{z} ,$$

for all $\boldsymbol{z} \in \mathbb{R}^N$, which means that $\boldsymbol{W}$ is positive semidefinite and $\boldsymbol{W} - \boldsymbol{I}_N$ is negative semidefinite, that is

$$0 \leq \lambda_i(\boldsymbol{W}) \leq 1$$

$i = 1, \ldots, N$, where $\lambda_i(\boldsymbol{W})$ denotes the ith largest eigenvalue of $\boldsymbol{W}$. Further, we notice that

$$(\boldsymbol{W} - \boldsymbol{J})\mathbf{1}_N = \boldsymbol{W}\mathbf{1}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T\mathbf{1}_N = \mathbf{0}_N , \quad (9)$$

with $\mathbf{0}_N$ the length N vector of all zeros. Eq. (9) offers that $\mathbf{1}_N$ is an eigenvector which corresponds to the eigenvalue 0

and to the only nonzero eigenvalue of $\boldsymbol{J}$. Thus there exists an orthogonal matrix $\boldsymbol{M}$ that diagonalizes $\boldsymbol{W} - \boldsymbol{J}$

$$\boldsymbol{M}(\boldsymbol{W} - \boldsymbol{J})\boldsymbol{M}^T = \text{diag}\big(0, \lambda_2(\boldsymbol{W}), \ldots, \lambda_N(\boldsymbol{W})\big) .$$

Now, according to Lemma 2, $\mathcal{G}(\boldsymbol{W})$ is strongly connected and matrix $\boldsymbol{W}$ therefore irreducible (see [14, Observation A.30]) such that from the Perron-Frobenius theorem for irreducible matrices [14, Theorem A.32] follows that $\lambda_1(\boldsymbol{W}) = 1$ is simple and $\lambda_i(\boldsymbol{W}) < 1$, for all $i = 2, \ldots, N$. Hence, $\rho(\boldsymbol{W} - \boldsymbol{J}) = \max_i \lambda_i(\boldsymbol{W}) < 1$ which completes the proof. ∎

Consider now any initial state $\boldsymbol{x}(0) \in \mathcal{X}^N$. If we define the error vector at time $t \in \mathbb{Z}_+$ as $\boldsymbol{\varepsilon}(t) := \boldsymbol{x}(t) - \boldsymbol{J}\boldsymbol{x}(0) \in \mathbb{R}^N$, we obtain in addition to the convergence in expectation of superposition gossiping in connected clustered WSNs the following upper bound for the mean squared error [3]

$$\mathbb{E}\big\{\boldsymbol{\varepsilon}(t)^T \boldsymbol{\varepsilon}(t)\big\} \leq \lambda_2(\boldsymbol{W}) \boldsymbol{\varepsilon}(t-1)^T \boldsymbol{\varepsilon}(t-1)$$
$$\leq \lambda_2^t(\boldsymbol{W}) \boldsymbol{\varepsilon}(0)^T \boldsymbol{\varepsilon}(0) , \quad (10)$$

from which we conclude with $\lambda_2(\boldsymbol{W}) < 1$ the following theorem.

Theorem 1. The mean squared error of Algorithm 1 converges to zero as $t \in \mathbb{Z}_+$ tends to infinity, that is

$$\lim_{t \to \infty} \mathbb{E}\big\{\boldsymbol{\varepsilon}(t)^T \boldsymbol{\varepsilon}(t)\big\} = 0 . \quad (11)$$

In addition, we state in Theorem 2 a somewhat stronger convergence property of superposition gossiping.

Theorem 2. Algorithm 1 converges to the average consensus almost surely, that is

$$\mathbb{P}\left(\lim_{t \to \infty} \boldsymbol{x}(t) = \boldsymbol{J}\boldsymbol{x}(0)\right) = 1 . \quad (12)$$

    Proof: The proof follows immediately from Theorem 1, from Robbins & Siegmunds theorem and from Fatou's Lemma [15]. ∎

V. IMPROVING THE CONVERGENCE SPEED

The right hand side of (10) shows that the convergence speed of superposition gossiping is determined by the second largest eigenvalue of $\boldsymbol{W}$. Therefore, to improve the convergence speed we want to minimize $\lambda_2(\boldsymbol{W})$ in this section.

Obviously, the eigenstructure of $\boldsymbol{W}$ is determined by the following factors:

- The number $|\mathcal{C}|$ of clusters,
- the distribution of nodes among different clusters,
- the number of common nodes between clusters,
- the wake up probabilities p_i, $i = 1, \ldots, |\mathcal{C}|$.

Since the first three bullets are a matter of topology control, which we want to consider in our future work, we address in the following the last bullet by optimizing the vector of wake up probabilities $\boldsymbol{p} := [p_1, \ldots, p_{|\mathcal{C}|}]^T$ (i.e., in some sense the

order of cluster activation). The corresponding optimization problem can therefore be written as

$$\min_{\boldsymbol{p} \in \mathbb{R}_+^{|\mathcal{C}|}} \lambda_2(\boldsymbol{W})$$
$$\text{s.t.} \ \ \boldsymbol{W} = \sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{W}_i \qquad (13)$$
$$\sum_{i=1}^{|\mathcal{C}|} p_i = 1$$

which can be reformulated to the convex program

$$\min \xi$$
$$\text{s.t.} \ \ \boldsymbol{W} - \boldsymbol{J} \preceq \xi \boldsymbol{I}_N$$
$$\boldsymbol{W} = \sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{W}_i \qquad (14)$$
$$\sum_{i=1}^{|\mathcal{C}|} p_i = 1$$
$$p_i \geq 0, i = 1, \ldots, |\mathcal{C}| \ .$$

Such problems (14) are readily transformed to a standard semidefinite program by defining $\boldsymbol{v} := [\boldsymbol{p}^T, \xi]^T \in \mathbb{R}^{|\mathcal{C}|+1}$, that is

$$\min \boldsymbol{c}^T \boldsymbol{v}$$
$$\text{s.t.} \ \ \text{diag}\big(F(\boldsymbol{v}), \boldsymbol{a}_1^T \boldsymbol{v}, \ldots, \boldsymbol{a}_{|\mathcal{C}|}^T \boldsymbol{v}\big) \preceq \boldsymbol{0}_{(|\mathcal{C}|+1) \times (|\mathcal{C}|+1)} \quad (15)$$
$$\boldsymbol{a}_0^T \boldsymbol{v} = 1 \ ,$$

where $\boldsymbol{c} := [0, \ldots, 0, 1]^T \in \mathbb{R}^{|\mathcal{C}|+1}$, $F(\boldsymbol{v}) := \sum_{i=1}^{|\mathcal{C}|} p_i \boldsymbol{W}_i - \xi \boldsymbol{I}_N - \boldsymbol{J}$, $\boldsymbol{a}_0 := [\boldsymbol{1}_{|\mathcal{C}|}^T, 0]^T$, $\boldsymbol{a}_i := [-\boldsymbol{e}_i^T, 0]^T \in \mathbb{R}^{|\mathcal{C}|+1}$, $i = 1, \ldots, |\mathcal{C}|$, and $\boldsymbol{e}_i \in \mathbb{R}^{|\mathcal{C}|+1}$ the ith unit vector, respectively.

The optimal solution $\boldsymbol{p}^\star$ of problem (15) as well as of problems (13) and (14) can then be found by standard algorithms from semidefinite programming.

VI. NUMERICAL EXAMPLES

To demonstrate the huge potential of the proposed superposition gossiping in clustered WSNs, we compare in this section Algorithm 1 with the RPG protocol from [3] as well as with the broadcast gossiping approach (BG) from [8]. To this end, we consider a sensor network example consisting of $N = 16$ nodes that are distributed among 1, 2 or 3 clusters and we assume that the corresponding clustered WSN is connected if the number of clusters is greater one (i.e., $|\mathcal{C}| > 1$). The nodes are deployed to monitor temperature values and we suppose that the initial state $x_i(0)$ of each node is drawn uniformly from $[0\,^\circ\text{C}, 10\,^\circ\text{C}]$.

According to the considerations in Section V, we distinguish in the three cluster scenario between two cases. First, we consider cluster heads that wake up equally likely (i.e., $\boldsymbol{p} = \frac{1}{3}\boldsymbol{1}_3$) and second, we consider cluster heads that wake up due to the optimal probabilities $\boldsymbol{p}^\star = (0.2771, 0.4053, 0.3177)^T$, resulting from solving problem (15).

The corresponding simulations are depicted in Fig. 2, where the "average error" is defined as

$$\bar{\varepsilon}(t) := \frac{1}{M} \sum_{m=1}^{M} \|\boldsymbol{\varepsilon}_m(t)\|_2 \ ,$$

with $\|\boldsymbol{\varepsilon}_m(t)\|_2$ the error in the mth simulation round, $m = 1, \ldots, M$, and M chosen to be 10^3. The plots verify the convergence of superposition gossiping in connected clustered sensor networks and indicate huge performance gains
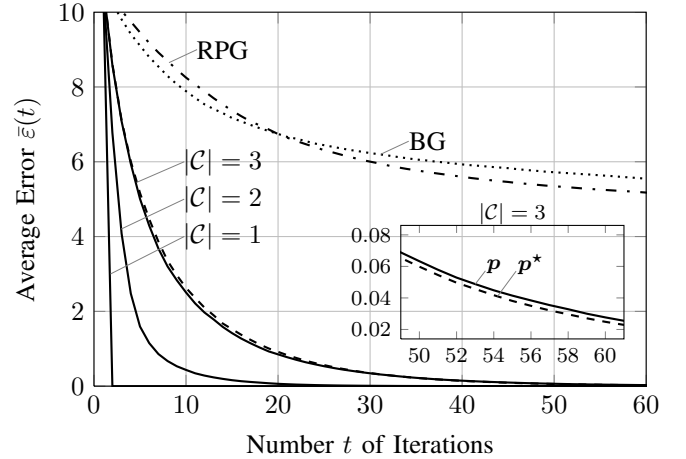

Fig. 2. Superposition gossiping vs. random pairwise gossiping (RPG) [3] and vs. broadcast gossiping (BG) [8] for a wireless sensor network with $N = 16$ nodes and $|\mathcal{C}| \in \{1, 2, 3\}$ connected clusters. The small figure represents a zoom to illustrate the acceleration due to optimized wake up probabilities as a solution of the optimization problem (15).

in comparison to RPG and BG, also for uniform wake up probabilities.

Note that the difference between $\boldsymbol{p}$ and $\boldsymbol{p}^\star$ has only a marginal impact in the considered example (see the zoom in Fig. 2) since the number of clusters in the considered numerical example is relatively small.

Remark 3. The plots in Fig. 2 indicate that the number of clusters crucially impacts the convergence speed of superposition gossiping. Since we considered error-free computation over ideal MACs as well as error-free broadcasting of updated states within clusters, the case $|\mathcal{C}| = 1$ is of course optimal and requires only 2 iterations to converge. In practice, however, adequate cluster design/topology control have to take into account the trade-off between convergence rate, energy efficiency and computation accuracy, since due to path losses and fading, steps 4–6 in Algorithm 1 are superimposed by noise. The consideration of these practical issues will therefore be part of our future work.

VII. CONCLUSIONS

In this paper, we proposed a randomized iterative gossip algorithm for wireless sensor networks that are organized into clusters, called "superposition gossiping", that rapidly reaches an average consensus. The benefit of the algorithm lies in the exploitation of the superposition and the broadcast property of wireless channels, which significantly increases the averaging speed in each cluster. Under the assumption of noiseless computation in each cluster, we have shown that the algorithm converges to the average consensus in different stochastic senses, if the clustered network fulfills connectivity between clusters due to common nodes. Since the convergence speed depends on the second largest eigenvalue of a network describing mean weight matrix, we formulated a corresponding eigenvalue-minimization problem to calculate optimal wake

up probabilities that further increase the convergence rate. Since the paper was devoted to deliver the idea and principal behavior of superposition gossiping, we incorporate realistic channels in our future work.

REFERENCES

[1] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[2] M. Çetin, L. Chen, J. W. Fisher III, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 42–55, Jul. 2006.

[3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[5] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1205–1216, Mar. 2008.

[6] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.

[7] S. Kirti, A. Scaglione, and R. J. Thomas, "A scalable wireless communication architecture for average consensus," in *Proc. 46th IEEE Conference on Decision and Control (CDC)*, New Orleans, USA, Dec. 2007, pp. 32–37.

[8] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, Jul. 2009.

[9] B. Nazer, A. G. Dimakis, and M. Gastpar, "Local interference can accelerate gossip algorithms," *J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 876–887, Aug. 2011.

[10] M. Goldenbaum, S. Stańczak, and M. Kaliszan, "On function computation via wireless sensor multiple-access channels," in *Proc. IEEE Wireless Communications & Networking Conference (WCNC)*, Budapest, Hungary, Apr. 2009.

[11] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[12] W. Li and H. Dai, "Cluster-based distributed consensus," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 28–31, Jan. 2009.

[13] M. Goldenbaum and S. Stańczak, "Computing functions via simo multiple-access channels: How much channel knowledge is needed?" in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, USA, Mar. 2010, pp. 3394–3397.

[14] S. Stańczak, M. Wiczanowski, and H. Boche, *Fundamentals of Resource Allocation in Wireless Networks*, ser. Foundations in Signal Processing, Communications and Networking. Berlin: Springer, 2009, vol. 3.

[15] B. T. Polyak, *Introducion to Optimization*. Optimization Software, 1987.