On the Communication Architecture for Wide-Area Real-Time Monitoring in Power Networks

Shrut Kirti, Zhifang Wang, Anna Scaglione, Robert Thomas
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
Email: {sk482, zw18}@cornell.edu, anna@ece.cornell.edu, rjt1@cornell.edu

Abstract— **Reliable and efficient operation of power networks is of paramount importance. In this paper we explore communication architectures that leverage Phasor Measurement Units (PMUs) and fusion centers to enable robust real-time monitoring of power networks over wide areas. We examine scalability issues in data aggregation for a power network in which PMUs are attached to every bus. We argue that in power networks, the data recorded by the PMUs are spatially correlated and therefore the data admit a sparse representation in a given basis. Furthermore, the number of bits required to reconstruct the PMU data to within a given accuracy at a remote location grows sublinearly with the density of the power network. Our results are: i) the PMUs should not transmit their data to the fusion center independently or asynchronously - an intermediate data aggregation step is beneficial, and ii) if we perform data aggregation then we can add more PMUs to the network to achieve finer network monitoring without causing network congestion.**

I. INTRODUCTION

Power networks are one of the most essential infrastructure components in the world today. While the legacy SCADA architecture provides health information about the steady-state network, its utility in detecting transient changes in the power network and preventing the kind blackouts that struck parts of the United States in 2003 is limited. Yet so far, there has only been a limited effort to incorporate the latest innovations in signal processing to improve the efficiency and reliability of power networks. For instance, there have been significant developments in autonomous real-time monitoring and/or control in recent years. The application of these techniques can improve the performance of power networks especially with respect to the two aforementioned metrics - efficiency and reliability.

Another recent development that may have a special relevance to power systems is sensor networking [1]. One way sensor networks may be used to aid the monitoring and control of power networks is by allowing the real time acquisition of the data collected by the several Phasor Measurement Units (PMU) attached to the power network at various locations in the power network. The PMUs periodically record the voltage and current in the power line to which they are locally attached. These phasor measurements can then be sent to a central control unit, referred to as a *fusion center*, whose task will be - a) to decide in real-time whether the power network is about to enter an unstable state and b) if the power network is about to become unstable, issue control commands to rectify the problem.

The key contributions of this paper are in two complementary parts: 1) in the first part we review the relevant work in sensor control and monitoring and discuss which options are the most likely to meet the system requirements and 2) in the second part we examine the PMU network as a source of data and demonstrate that the PMU data admit a sparse representation. This allows us to argue that the type of data aggregation schemes proposed in the first part of the paper are appropriate for the type of data source under consideration.

The paper is organized as follows. In Section II we present the overall control and monitoring problem for power networks by means of an example. Section III focuses on the communication architecture aspect of the problem. It covers alternative approaches being considered by other researchers and reviews communication themes that we think are especially suited to power networks. Sections IV and V investigate whether the structure of the data is sparse over the FFT basis and how the number of bits needed to represent the whole evolution of the data scales as the network density increases. We numerically obtain the number of bits needed to reconstruct the PMU data as the density of the power network increases and show that increasing correlation in the PMU data favors data aggregation rather than independent PMU transmissions. Section VI concludes the paper.

II. PROBLEM OVERVIEW

To get an idea of how a sensor network can help in the control and monitoring of the power network and to motivate the issues addressed in the remainder of the paper let us consider a toy example. Consider a one machine to ∞-bus power network. This simple network is governed by the following swing equation-

$$M\ddot{\theta}(t) + D\dot{\theta}(t) = P_m - \frac{V_1 V_2}{X}\sin\theta(t) + \mu(t) \qquad (1)$$

where M, D, P_m, and jX are the moment of inertia, damping co-efficient, mechanical input power of the generator, and the impedance of the transmission line respectively. The voltage at the machine bus and the ∞-bus is denoted as $V_1\angle\theta(t)$ and $V_2\angle 0$ respectively. $\mu(t)$ is a term that comes from the presence of the control box as discussed below.

It is possible that occasionally the machine may experience a disturbance which may cause the system to lose stability if
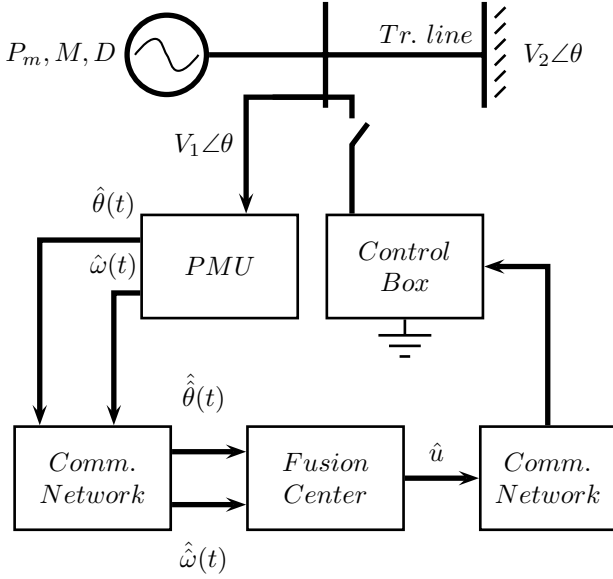
Fig. 1. (A one bus power network is connected to the control mechanism as shown. The fusion center might use a switch slope strategy. If $f(\theta(t), \omega(t)) > slope$ then $u(t + \Delta) = 1$, otherwise $u(t + \Delta) = 0$.

corrective action is not taken. For this reason, an automated mechanism which can monitor the power network and take corrective action when required is of significant interest. Depending on the type of disturbance that occurs, the control that is required will be different.

A simple mechanism for the control and monitoring function may use the PMU data to obtain estimates of the phase and the frequency (which is derivative of the phase) of the bus signal $\hat{\theta}(t)$ and $\hat{\omega}(t)$ respectively which are not exact due to quantization error:

$$\hat{\theta}(t) = \theta(t) + \epsilon_\theta(t) \quad (2)$$
$$\hat{\omega}(t) = \omega(t) + \epsilon_\omega(t) \quad (3)$$

where $\epsilon_\theta(t)$ and $\epsilon_\omega(t)$ denote the errors due to finite bit resolution.

These estimates are transported over the communication network to the fusion center which applies some function $f(\theta(t), \omega(t))$ to compute a control parameter $u(t + \Delta) = (f(\theta(t), \omega(t)) \lessgtr 0)$. This control parameter is then sent back over the same communication network to the local control box to arrive at time $t + \Delta$. Note that Δ is the round-trip communication delay from the PMU sensors to the fusion center. As an example of a control signal, $u(t + \Delta)$ can be a simple on-off switch signal for the local control box which generates $\mu(t+\Delta) = -\xi u(t)$, where ξ is some constant. Since the fusion center only has access to estimates of $\theta(t)$ and $\omega(t)$, $u(t)$ may have errors as well.

Two remarks are in order: first, note that in this toy example, since we have only one bus, the control strategy can actually be implemented locally at the PMU; second, the cost of forwarding each additional bit from the bus to the fusion center leads to an exponential decrease in the distortion i.e. we get a better estimate of $\theta(t)$ and $\omega(t)$. In fact, any admissible quantizer of $\theta(nT_s)$ or $\omega(nT_s)$ (from a simple scalar quantizer

to an optimized k-dimensional vector quantizer) can be shown to have an operational distortion rate function such that:

$$D_k(R) \approx \mathcal{O}(2^{-2R}), \quad (4)$$

where the choice of the quantizer affects only the constant terms hidden in the big-O notation [2]. In equation (4), D is the distortion that can be achieved by using R bits to represent the data using a kth order vector quantizer. For example, for a simple scalar quantizer ($k = 1$) over the range $[1, -1]$, $D_1(R) = \frac{1}{12}2^{-2R}$.

In a complex system with multiple interconnected buses, however, it is not a good idea to let the PMUs forward bits to the fusion center independently. The traffic generated by the sampling and quantization function of each PMU when they are operating independently, grows in the order of the number of buses in the system N. When the number of PMUs in the network becomes large the amount of traffic generated will cause congestion. Instead, the control operation should account for the state of multiple buses simultaneously.

This simple example is helpful in understanding the scope of the problem we are ultimately trying to address - real-time control of transient instabilities in a wide-area power network. Real-time network monitoring, the focus of this work, is an important first step in that direction. The example also highlights a crucial inherent tradeoff - between the fidelity of the reconstruction of $\theta(t)$ and $\omega(t)$ at the fusion center and the delay incurred in transporting these quantities over the communication network.

If the PMU sensors and the fusion center, which acts as the control unit, were co-located, then the different parts of the network could use dedicated interconnections that would guarantee the timely delivery of the PMU measurements to the fusion center and the control parameters from the fusion center to the PMUs. However, in an actual power network the PMUs and the fusion center can be separated by significant distances since the grid may span a large area. Therefore, we require a reliable communication network to shuttle information around the system. What are the criteria for the best transport mechanism? Since the fusion center must make decisions to implement corrective action in real-time, it is vital that the round-trip delay Δ from sensors to the fusion center be sufficiently short. The control is used to mitigate transient instabilities, therefore, the strict delay constraint is based on the *critical clearing time*, which is system dependent. In serious conditions this can be as short as $\Delta < 300\ ms$, while in ordinary contingencies it can be on the order of $1\ s$. The round-trip delay of the sensor communication depends on the underlying physical layer communication technique, the network topology, the amount of data that needs to be exchanged, processing times at sensors and at the fusion center, and the response time of the mechanical devices in the power network. The design parameters in our problem are the physical layer communication technique and the amount of data that needs to be exchanged. The solution of the joint control and communication problem is not trivial and has attracted a lot of interest from the research community as is discussed in the section below.

III. COMMUNICATION ARCHITECTURES FOR CONTROL AND MONITORING

To the best of our knowledge, there are two roughly orthogonal approaches that researchers have considered with respect to the control and monitoring problem. The approach the control community has adopted is largely based on utilizing the Internet as the backbone for intra-power network communication. The research on sensor networks, instead, has focused on the design of scalable dedicated communication networks for solving the data gathering problem. Much work in this area has attempted to take advantage of the structure of the data being observed to achieve the best possible reconstruction while transmitting the least possible amount of information over a dedicated network, typically an ad-hoc network of some sort. These two approaches are discussed below.

A. The Internet as the Communication Network

A communication network that already exists on a grand scale and that can be integrated with the power network is the Internet. One must, however, consider three issues before proceeding with such a solution. *The first issue* is whether the Internet can guarantee the security of the PMU measurements. Relying on the Internet to perform basic control operations over the power network would create an interdependency between two critical infrastructures, exposing the latter to the vulnerability of the former. Unencrypted PMU data transmitted over the Internet can be intercepted by prying eyes at several stages of routing; to avoid this, encryption schemes may be used at the cost of increasing the amount of information to be sent. A more serious problem is that such transmission will be susceptible to all kinds of spoofing, DoS, and other attacks prevalent on the Internet which can be used to modify the data the fusion center receives and ultimately cause catastrophic damage to the operation of the power network. *The second issue* is whether the Internet can provide a reliable and delay constrained communication channel. While this is an active research area, current Internet protocols like TCP, UDP, or RTP either randomly drop or delay data packets. This randomness poses a significant challenge to the methods of traditional control theory. In particular, it has recently been shown in [3] that the separation principle of optimal LQG control does not hold in general for networked control systems. A networked control system does not have dedicated interconnections between its components and must rely on a non-deterministic communication network to enable the components to communicate. While [3], [4] take on the challenge of updating classical control theory results for networked control systems, there has been previous work which has contributed to this development. For example, in [5] the authors propose modeling the plant and the controller as a deterministic time invariant discrete-time system connected to a stochastic uncertainty, and in [6] the authors examine the connection between the bit rate that the communication channel can support and the control objectives that the controller can fulfill. Similar ideas are explored in [7], [8].

The third issue is the scalability of the solution. It is obvious that providing Internet connectivity to PMUs at every bus in the power network has a considerable cost, especially because of the provision of bandwidth that supporting all these transmissions would entail. When thinking of the Internet as the medium to transport the PMU samples, our considerations in Section II regarding the precision of the data representation are frivolous: the cost of transporting packets has a minimal growth with the size of their payload; adding bits to get a more precise sample reconstructions adds a virtually negligible cost. This makes packet losses, rather than losses in measurement precision, a more relevant problem for the papers cited above. However, packet losses are caused precisely by the congestion that results from delivering information independently and asynchronously from a large number of PMUs to the control location. Ironically, not only are the observed data generated synchronously, they are actually governed by coupled equations, so they are anything but independent. *This leads us to speculate that a dedicated sensor network that aggregates data over extended portions of the network prior to forwarding the data to a fusion center would reduce the complexity and enable timely control, unlike the brute force approach of hooking up each PMU to the Internet.* What medium would then be suitable to use for the sensor network transmissions? Most of the recent work on sensor networks considers wireless connectivity because of the ease of deploying wireless networks [1]. It should not go unnoticed, however, that there have been new developments in the area of power-line communications [9], [10] which may provide a suitable dedicated network that can allow us to opportunistically exploit the infrastructure itself to transport information that is relevant to its control.

B. Background on Data Retrieval in Sensor Networks

Sensor systems sample data fields that are often either strongly correlated or mostly idle both spatially and temporally. When thinking of a communication network to support the gathering of data from the sensors, this fact has two major consequences: 1) the data should be digitized in a way that exploits the fact that the aggregate data are correlated or idle; 2) the separation theorem, one of the corner-stones of communication systems design, does not hold. For example, it is simple to understand that correlated data traveling through a joint route in distinct packets could be obviously combined in one smaller-sized packet via standard compression and over a large scale this could lead to significant reduction in the amount of traffic generated [11]. This implies that using a general purpose network for the power network application at hand will be particularly inefficient and detrimental to the rapid and differential delivery of the PMU data to the fusion center. Next we highlight a few approaches that have emerged in the coding and signal processing research that are applicable to our power network application.

1) Distributed source coding: Consider two scalar discrete sources of data X and Y with joint statistics $p(x, y) \neq p(x)p(y)$. Information theoretical results established in the 70's by Slepian and Wolf [12] indicate that the two sources can be compressed *separately* so as to achieve an aggregate rate that can be equal to the rate achievable by having *access to both* sets of data at the same location (see Figure 2). In short,
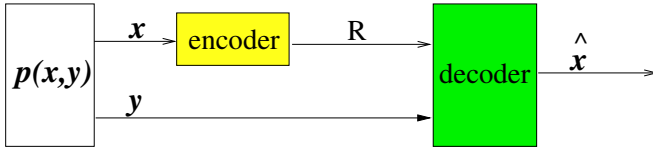
Fig. 2. Coding with side information: y is available only at the decoder, and $\hat{x}$ is the reconstructed source

distributed source coding can be as efficient as a centralized compression algorithm having access to the entire data set. Wyner and Ziv [13], [14] extended the analysis to analog sources and showed that it was possible to achieve similar compression results on x, when y was available at the decoder. The literature on multi-terminal source coding, e.g. [15]–[17], extended the Slepian-Wolf and Wiener-Ziv framework to multiple noisy sensor observations. Duplets of constructive distributed source coding designs for analog sources have been developed recently [18]–[24]. In the context of sensor networks, this implies that much fewer bits per sensor should be sent, primarily because of the ability of the decoder to extrapolate missing information from other measurement data. Essentially, the reduction in the amount of information that needs to be sent to allow the fusion center to reconstruct the PMU data occurs directly at the source. However, a weakness of distributed source coding lies in requiring knowledge of the prior statistics of the data. Often times this assumption is unrealistic, because the prior distribution is a function of the underlying state of the system, which is precisely what it is unknown.

2) Sparse representations and random projections: Designing data gathering architectures poses communication problems that using distributed source coding alone cannot address, especially when one is faced with strict time delay constraints. Because of the inherent cost of multiplexing a channel among sources, for the timely delivery of the important data it is critical that the predictable part of the data does not congest the network. This can be avoided by using sparse representations of the data. At the same time, this data aggregation scheme makes very general assumptions on the correlation structure of the data, which allows for a robust and scalable system.

Let $\boldsymbol{\theta}$ be a data vector of size $N \gg 1$ and $\boldsymbol{\Psi}$ a matrix associated with a certain basis for $\mathbb{R}^N$. A typical approach taken in compression is to assume that most of the energy of the data vectors generated by a vector source lies in few modes that belong to a set of basis functions $\boldsymbol{\Psi}$, i.e. it is such that:

$$\boldsymbol{\theta} = \boldsymbol{\Psi} \boldsymbol{c} \qquad (5)$$

where $\boldsymbol{c}$ has a very small number of non-negligible entries. If the statistics of $\boldsymbol{\theta}$ are known, the best $\boldsymbol{\Psi}$ is the so called Karúnen Löeve Transform, i.e. the set of eigenvectors of the covariance matrix $\boldsymbol{R}_\theta = E\{\boldsymbol{\theta}\boldsymbol{\theta}^T\}$. In this case, the most significant coefficients can be sorted out from the statistics - they correspond to the largest eigenvalues of $\boldsymbol{R}_\theta$. If this knowledge is not available, however, there are other means to achieve compression.

One in particular, which is a new result in compressive sampling theory and distributed computing [25]–[27], inspired

a promising approach [28] for data aggregation in sensor networks. Compressive sampling theory states that if $\boldsymbol{\theta}$ has a *sparse representation* as in (5) it is possible to reconstruct $\boldsymbol{c}$ (and therefore interpolate $\boldsymbol{\theta}$ via (5)) from a small number of projections onto a second basis $\boldsymbol{P}$ that is incoherent[1] with the first one, i.e. from $\boldsymbol{P}\boldsymbol{\theta}$, where $\boldsymbol{P}$ is $L \times N$ with $L \ll N$. This fact leads to a very simple and general data aggregation procedure where there destination node of the information needs only to compute L projections rather than having access to the entire data record of dimension N. In a network with multiple intermediate relays, this can be achieved by using random gossiping algorithms [29]–[33]. Over a wireless/broadcast multiple access channel [34], [35], instead, transmitting the analog data implies that they will be naturally added by the medium weighted by the respective link-gain coefficients, say $\boldsymbol{h} = (h_1, \ldots, h_N)$. Assuming that the link gains are constant over the short period of time in which the data are forwarded, if each node broadcasts over L channel uses the vector $\boldsymbol{p}_i\theta_i(nT_s)$, where $\boldsymbol{p}_i = (p_1, \ldots, p_L)$ is a set of random coefficients, the receiver samples $\boldsymbol{y} = (y_1, \ldots, y_L)$ will be:

$$\boldsymbol{y} = \boldsymbol{P} diag(\boldsymbol{h})\boldsymbol{\theta} + \boldsymbol{w} = \boldsymbol{P}'\boldsymbol{\theta} + \boldsymbol{w}, \qquad (6)$$

where $\boldsymbol{P}' = \boldsymbol{P} diag(\boldsymbol{h})$ is the composite random projection matrix and $\boldsymbol{w}$ is the receiver additive noise vector. The channel coefficients $\boldsymbol{h} = (h_1, \ldots, h_N)$ can be estimated and tracked by sending a training sequence periodically, with a period that is comparable to the so called *channel coherence time*. In this scheme the delay for transmitting each new sample, $\boldsymbol{\theta}(nT_s)$, is $\mathcal{O}(LW^{-1})$, where W is the bandwidth dedicated to the transmission.

3) Data driven channel access methods: Alternative approaches that challenge the classical separation theorem, design joint source-channel coding strategies that lead to efficient estimation of all the sensors' measurements at the cluster-head or fusion center. Examples of this methodology are the works on type-based multiple-access (TBMA) [36], [37] and group-testing multiple-access [38], [39].

In type-based multi-access (TBMA), the observed data $\boldsymbol{\theta}(nT_s)$ are quantized and, therefore, belong to a discrete distribution. The knowledge of the measurements' statistics is not required at the sensors. Each sensor collects several samples $\theta_i(nT_s)$, $n = 1, \ldots, K$ and computes its *type*, i.e. the relative frequency of each symbol alphabet in the input sequence. Let us assume that $\mathcal{T}$ is the set of possible types indexed by $j = 1, \ldots, |\mathcal{T}|$. Depending on its measured type, each sensor transmits a code, say $\boldsymbol{p}_j$, over a shared communication channel to the cluster-head. Let $\boldsymbol{P} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{|\mathcal{T}|})$. In this scheme the channel output, of dimension $|\mathcal{T}| \times 1$, is

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{w} \qquad (7)$$

where, using the indicator function $1(\cdot)$ such that $1(true) =$

---

[1]Incoherence means that the second basis cannot admit a sparse representation of the elements of the first basis. Independent and identically distributed (i.i.d) Gaussian vectors provide a universal basis that is incoherent with any given basis with high probability.

$1, 1(false) = 0$, the jth entry of the input $\boldsymbol{x}$ is

$$\{\boldsymbol{x}\}_j = \sum_{i=1}^{N} h_i 1(\text{sensor-}i \text{ type} = j) \qquad (8)$$

where h_i are the channel link gains and $\boldsymbol{w}$ is additive noise. Thus, the channel output conveys sufficient statistics of the measured data $\theta_i(nT_s)$, $n = 1, \ldots, K$. In particular, as the number of buses N goes to infinity, assuming the h_i have identical distribution, $\{\boldsymbol{x}\}_j \rightarrow NE\{h\}P(\text{sensor-}i \text{ type} = j|\theta_i(nT_s))$, $n = 1, \ldots, K$. The estimation performance asymptotically approaches that of the scheme where the cluster-head has direct access to the measured data. However, for the case where the wireless channel has fading, a large performance loss incurs [36], [37]. A natural drawback of this scheme is the latency required to code the data.

Interestingly, TBMA can be seen as the feed-forward version of another approach called group-testing multiaccess (GTMA) [38], [40], [41]. In GTMA, the classification of the quantized vector value is done by imposing a sequence of queries on groups of sensors simultaneously, such as, *"are your data all in the range $[a, b]$?"*. In [38], [40] it is assumed that the replies are sent over a wireless multiple access channel and a sensor sends a signal only when it disagrees with the query/guess of the cluster-head. Considering the case of P different ranges tested in parallel $[a_j, b_j], j = 1, \ldots, P$ the received signal is equivalent to that in (8) where now:

$$\{\boldsymbol{x}\}_j = \sum_{i=1}^{N} h_i 1(\theta_i \notin [a_j, b_j]) \qquad (9)$$

which provides a way to determine if the sensors are or not all in a specific range.

If the query inquires the type of a long sequence of data per sensor, without any further query/feedback, then the GTMA scheme becomes equivalent to the TBMA scheme. It is possible, as argued in [41], to exploit the sparsity of the data in the sense of (5) to speed up the acquisition of the data. This is illustrated in Figure 3. Clearly in the control problem, the intervals corresponding to the queries that we indicate in Figure 3, should be optimized to achieve the maximum accuracy on the control function rather than on the data themselves. Similarly, in this scheme the delay for transmitting each new sample is $\mathcal{O}(LW^{-1})$, where W is the bandwidth dedicated to the transmission and L is the number of rows of the matrix $\boldsymbol{P}$ in (7). The potential advantage of $GTMA$ is that the queries could be done sequentially rather than in parallel, each providing a refinement on the previous query. This makes the number of queries and, therefore, the delay, variable.

The first step towards applying any of these methodologies is, however, to get a sense of how the source of data in a power network scales by showing the impact of the coupling in the power network equations. This paper investigates this topic further, trying to answer the question of how many bits are really needed to represent the state of a section of the power network. We specifically investigate whether the structure of the data is sparse over the FFT basis and how the number of
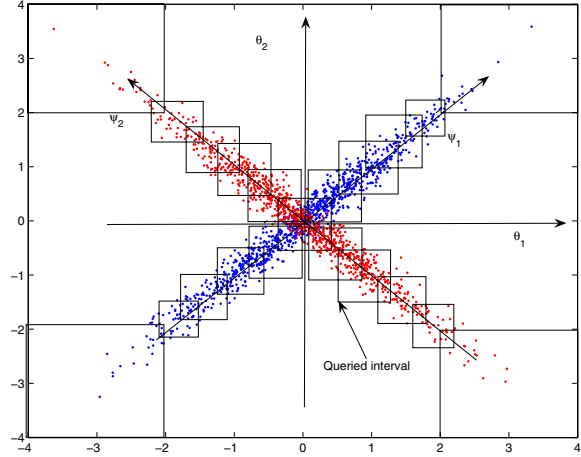

Fig. 3. The data generated cluster either around the direction of $\boldsymbol{\psi}_1$ or around the vector $\boldsymbol{\psi}_2$, which means that the vector of measurements lies typically in a one-dimensional space $\boldsymbol{\theta} = (\theta_1, \theta_2) \approx c_i \boldsymbol{\psi}_i$. This means that rather than exploring all possible ranges of values, a precise reconstruction of the vector can be obtained querying jointly the sensors on the intervals whose borders are represented in the plot.

bits needed to represent the whole evolution of the data scales as the network density increases.

IV. SOURCE CHARACTERIZATION

The vector $\boldsymbol{\theta}(t)$ of phasor measurements from all the PMUs is our data source. In order to design the best mechanism to transport this data source to the fusion center, we rely upon the source characteristics. In this paper we consider the most severe fault that can occur in a power network, referred to as a contingency. A contingency is defined as the event when some bus in the network suffers a three-phase short circuit fault.

At any time instant t the power network can be in one of $2N + 1$ states. The first state is the *pre-contingency state* in which the network is in a steady state and there is not much variation in the measurements made by the PMUs. The network spends most of its time in this state. Then, there are N *contingency states* which correspond to the event that a contingency has occurred on bus i. Contingencies occur very rarely and the network is highly unlikely to record two contingencies at the same time; but when a contingency does occur there is a lot of fluctuation in the source data. Finally, the network can be in one of N *post-contingency states* which correspond to the event that the contingency on bus i has been cleared but the PMU outputs evolve according to the dynamical system equations presented next, in Section IV-A. The fusion center can still send a control signal and prevent the overall system from losing stability when the network is in the post-contingency states.

It is intuitive from the structure of the power network that the components of the source vector cannot be independent in time or in space. Temporally, the value of a source element at a time t depends on its past values. Spatially, the value of a source element is affected by the values of its neighbors. This behavior is also exhibited by wide-area power networks.
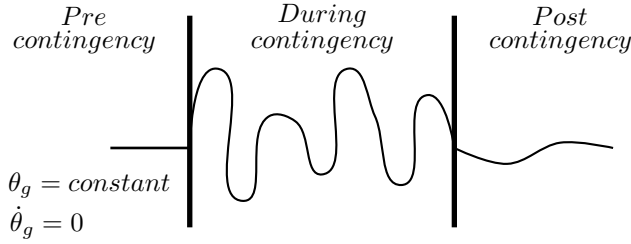
Fig. 4. The network can be in either of the depicted states.

For example, [42] proposes a continuum model for large-scale power networks and finds that the disturbances in the phase angles of the power system propagate slowly and exhibit dispersion phenomena. We examine the spatial correlation structure of the PMU data in a power network in the post-contingency state next to make these heuristical remarks more concrete.

A. The dynamical model for the PMU measurements

We consider a network in which the loads and the generators are alternatingly arranged in a loop, see Figure 5. All transmission lines between adjacent buses have the same impedance $(R+jX)/N$, where $(R+jX)$ is the aggregate impedance of the entire ring network. This ensures that the system density increases as the number of nodes increases. Note that since PMUs are assumed to be attached to every bus, the densities of the power network and the PMU network are the same. The network comprises m generators and loads each such that the total number of buses is $N = 2m$. It is assumed that all the generators, except for the reference-bus generator (located at bus 1), have the same inertia, damping, mechanical power input, and source impedance. This assumption is crucial in analyzing the changes in the spatial correlation structure and consequently the number of bits required to represent the data as the density of the network changes.
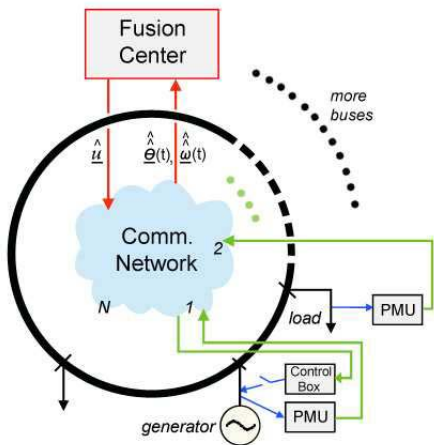

Fig. 5. Power network in a loop topology. A generator is connected to bus 1 and a load is connected to bus 2. Generators and loads are located alternately till bus N. Each bus has a locally attached PMU and in addition the generator buses have a control box that is connected to the communication network. The fusion center receives the network measurement vectors $\hat{\boldsymbol{\theta}}(t)$ and $\tilde{\boldsymbol{\omega}}(t)$, and outputs $\hat{\boldsymbol{u}}(t + \Delta)$.

In Section III we presented three broad communication strategies that might facilitate real-time monitoring of large power networks. In the following sections we consider the scalability of one particular strategy - sparse representations - in further detail. The objective of our analysis restricts the class of network topologies that we can tractably consider. A key advantage of considering the loop topology is that because of the straightforward ordering of the buses, it is relatively easy to find a suitable basis in which the phase data are sparse and thus compressible (c.f. Section V). Finding such a basis is relatively harder for a general IEEE test network, though the existence of a basis is by no means precluded - for example, the eigenvectors of a particular network matrix usually form a good basis to compress the data in that network. Thus, the ideas and observations made for the loop topology can be generalized to other power network topologies.

The generator swing equations and the network load flow equations of the loop system are as follows:

$$\frac{d\boldsymbol{\phi}(t)}{dt} = \boldsymbol{\omega}(t) \tag{10}$$

$$diag(\boldsymbol{m})\frac{d\boldsymbol{\omega}(t)}{dt} + diag(\boldsymbol{d})\boldsymbol{\omega}(t) = \boldsymbol{p}_m - \boldsymbol{p}_g(t) \tag{11}$$

$$\boldsymbol{p}(t) + j\boldsymbol{q}(t) = diag(\boldsymbol{u}(t))\boldsymbol{Y}\boldsymbol{u}^H(t) \tag{12}$$

$$p_{g_i}(t) + jq_{g_i}(t) = -j\,\frac{E_{g_i}e^{j\phi_i(t)}}{X_{g_i}}u_i^*(t) \quad i = 1, ..., \frac{N}{2} \tag{13}$$

where (10), (11), and (13) only apply to generator buses and $\boldsymbol{\theta}(t)$, $\boldsymbol{\omega}(t)$, $\boldsymbol{p}_m$, $\boldsymbol{p}_g(t)$ represent the vectors of internal angles, frequency, mechanical input power, and electrical output power of generators and control variable respectively; while $diag(\boldsymbol{m})$ and $diag(\boldsymbol{d})$ are diagonal matrices containing the generator inertia and damping coefficients, listed in the vectors (m) and $\boldsymbol{d}$ respectively.

Equation (12) applies to all the buses. $\boldsymbol{p}+j\boldsymbol{q}$ is the vector of injected complex power into the buses. For the generator buses, $\{\boldsymbol{p}+j\boldsymbol{q}\}_i = \{\boldsymbol{p}_g+j\boldsymbol{q}_g\}_i$ and for the load buses, $\{\boldsymbol{p}+j\boldsymbol{q}\}_i = -\{\boldsymbol{p}_l + j\boldsymbol{q}_l\}_i$ The complex bus voltage vector is given by the vector $\boldsymbol{u}(t) = (V_1(t)e^{j\theta_1(t)}, V_2(t)e^{j\theta_2(t)}, \ldots, V_N(t)e^{j\theta_N(t)})$ where $\{\boldsymbol{v}(t)\}_i = V_i(t)$ is the voltage magnitude and $\{\boldsymbol{\theta}(t)\}_i = \theta_i(t)$ is the bus voltage angle. $\boldsymbol{Y}$ is admittance matrix of the system. For the ring configuration, $\boldsymbol{Y}$ has a special structure

$$\boldsymbol{Y} = \begin{bmatrix} 2y & -y & 0 & \ldots & 0 & -y \\ -y & 2y & -y & 0 & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & -y & 2y & -y \\ -y & 0 & \ldots & 0 & -y & 2y \end{bmatrix} \tag{14}$$

where $y = \frac{N}{R+jX}$ is the admittance parameter of the transmission line. In our simulations we set the parameters as follows. The power base is $S_B = 100(MVA)$, the voltage base is $V_B = 400(kV)$, the frequency base $f = 60(Hz)$; the transmission line has a total impedance $R + jX = 1.7298 + j10.3782(p.u.)$, each load bus has initial loads as $\{\boldsymbol{p}_l(t) + j\boldsymbol{q}_l(t)\}_{i=1}^{N/2} = 0.1 + j0.02$, each generator bus except the reference bus 1 has an initial bus voltage magnitude $\{\boldsymbol{v}(0)\}_{i=2}^{N/2} = 1.0$ and initial real power $\{\boldsymbol{p}_g(0)\}_{i=2}^{N/2} = 0.0937$,
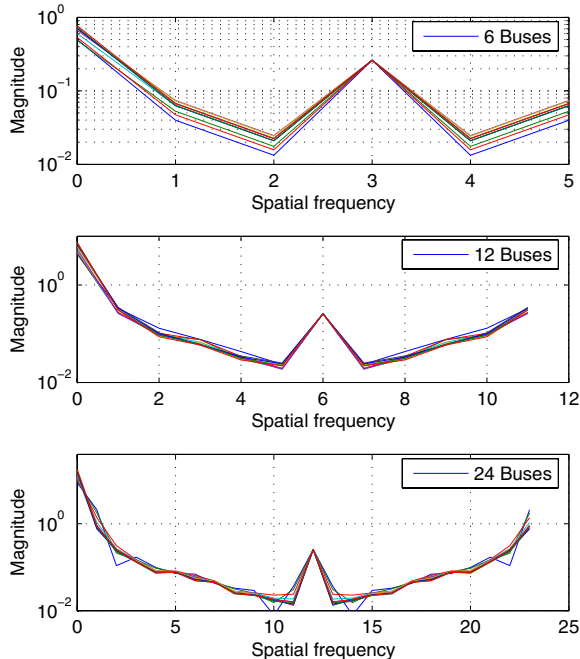
Fig. 6. The spatial correlation in the PMU phase data is significant as can be seen from the dominant peaks. The data has predominantly low frequency content which grows as the density of the network increases.
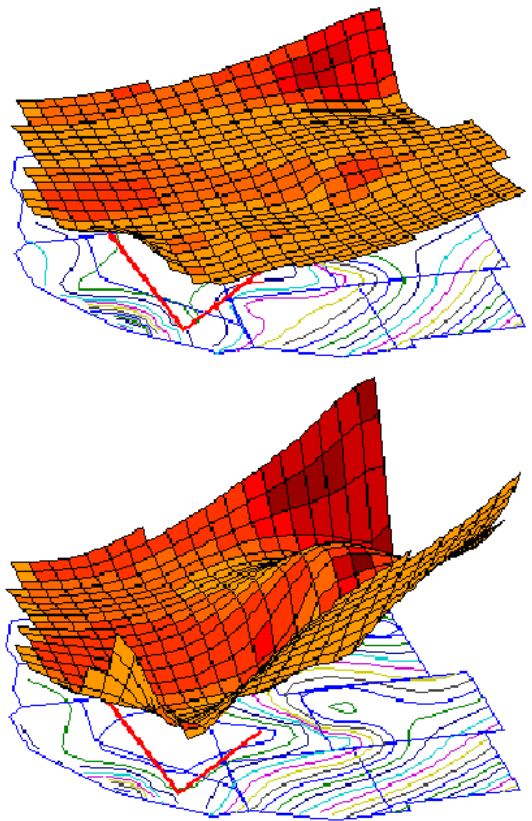

Fig. 7. Simulation of the 1994 disturbance in the Western Systems Coordinating Council (WSCC) power system: (top) shows a snapshot of voltage angles during the contingency; (bottom) shows the evolution of the contingency so that finally the system becomes unstable.

for the reference bus $m_1 = 10.0$, $d_1 = 10.0$, and for each of other generators, $\{m\}_{i=2}^{N/2} = 1.0$, $\{p\}_{i=2}^{N/2} = 1.0$. With these initial generation and load settings, AC power flows are computed to get the initial equilibrium. During the dynamic simulation, the loads take constant-impedance model to simplify and speed up the computation, whose impedance are set to

$$\frac{\{p_l(0) - jq_l(0)\}_i}{(\{u(0)\}_i)^2}. \tag{15}$$

To obtain the statistics of the PMU data, we considered multiple contingency scenarios in each of which bus i was selected randomly using a uniform distribution and short circuited for a random duration t_d which is exponentially distributed with mean $\mu = 0.1s$. After t_d seconds the contingency was cleared. These experiments were repeated for different values of N, in effect varying the network density. While the probability of the occurrence of these events in an actual power network is small, these cases provide useful information on how the network output will behave on the average in the worst scenarios.

We found that the PMU phase data $\theta(t)$ (see Figure 8) only has a few dominant frequency components as shown in Figure 6. This is an indication of the spatial correlation and of the fact that $\theta(t)$ has a spatially sparse representation. Each sub-plot in Figure 6 contains several closely matching curves. Each curve corresponds to the FFT of the PMU data averaged over longer and longer time periods, ranging from [1, 100] samples.

Though we have considered the ring topology for ease of analysis, some measure of spatial correlation exists in power networks regardless of the network topology. Researchers have worked with this idea before. In [43], the authors identified

groups of generators in large power networks that swing together in response to a contingency - a phenomena formally known as coherancy. [44] and [42] model large-scale dense networks with a continuum model which highlights spatial information absent in conventional models. In Figure 7 we reproduce Figure 1.1 from [44] which shows a "visualization of the voltage phase angles and associated dynamics as a continuous surface over the entire system." Figure 9 shows the IEEE 118 Bus Test Case representing a portion of the American Electric Power System (in the Midwestern US). The figure plots the bus voltage angles with respect to the spatial coordinates of a steady-state power flow of this system. The voltage angles are interpolated where no bus is present. It shows that even in extreme contingency conditions, spatial correlation exists in voltage angles [44].

While the system equations provided here allow us to interpolate the evolution of the system from the initial conditions, in general the precise parameters of the network are not available to the fusion center. Hence, a robust approach to data monitoring must assume little about the underlying structure of the network. Assuming knowledge of whether or not a contingency has occurred and where it occurred, can obviously speed up the acquisition, just as the knowledge of the underlying state helps reaching high compression efficiency in sources that are Hidden Markov Models (HMM) [45]; the
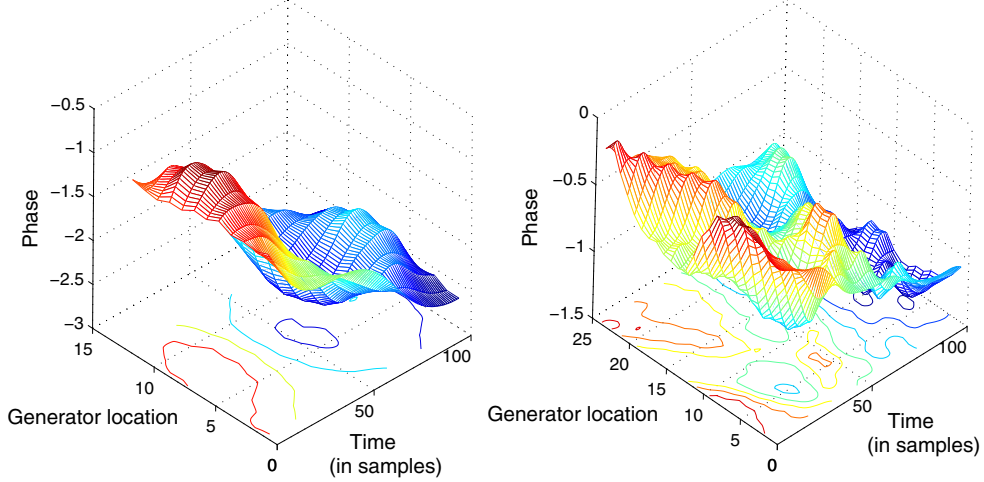
Fig. 8. The plots on the left and the right show how the phase data evolves over time for $N = 12$ and $N = 24$ respectively.

exploitation of HMM in a distributed environment requires communication feedback from the cluster-head to the sensors, which is an interesting possibility to explore, especially in the context of methodologies such as GTMA. Next we will use simple heuristics to demonstrate the *sparsity* of the PMU data.

V. ON THE SPARSITY OF THE PHASOR DATA

Denote the value of the data source at time t by the vector $\boldsymbol{\theta}(t)$. Since the voltage variations are modest, in the following we consider only the phase information. The ith element of $\boldsymbol{\theta}(t)$ represents the phase measurement of the ith PMU in the power network at time t. The source components be sampled every T_s seconds to give $\boldsymbol{\theta}[n] = \boldsymbol{\theta}(nT_s)$, $n \in \mathcal{Z}$.

As we have discussed in the previous section, a robust approach towards compression consists in relying on a very general concept of *sparsity* of the data representation, in the sense of (5). In fact, (5) generalizes the idea that the measurements are smooth, by highlighting the fact that the data in general lie in a small dimensional linear manifold. In this section our goal is to show via numerical results that $\boldsymbol{\theta}[n]$ admits a sparse representation and that the number of

Fig. 9. The voltage angles in the IEEE 118 Bus Test Case in the steady state.

bits required tends to grow sub-linearly as the network size N grows while its total impedance is kept constant.

To exploit the high temporal correlation of the data without adding latency in the encoding system, we encode the data differentially in time, hence the data actually quantized are

$$\boldsymbol{e}[n] = \boldsymbol{\theta}[n] - \alpha\boldsymbol{\theta}[n-1] \qquad (16)$$

where $\alpha < 1$ is a forgetting factor.

For the ring topology the fact that the admittance matrix $\boldsymbol{Y}$ in (14) is circulant, suggests that a good candidate for the basis functions $\boldsymbol{\Psi}$ is the FFT matrix, which we used to produce our results. The coefficients

$$\boldsymbol{c}[n] = \boldsymbol{\Psi}^H \boldsymbol{e}[n] \qquad (17)$$

have conjugate symmetry, since the data are real. Hence we can quantize and transmit only half of them for the reconstruction. For a sparse representation, most of the entries $\boldsymbol{c}[n]$ will have a negligible magnitude, which is exactly what we have observed for our data, see Figure 6. We denote by $[-\sqrt{\lambda_i}, \sqrt{\lambda_i}]$ the range of the i-th coefficient entry $\{\boldsymbol{c}[n]\}_i = c_i[n]$ and consider as our objective that of obtaining a certain per sample mean-squared error distortion in the representation of $\boldsymbol{e}[n]$:

$$\frac{1}{N}E\{\|\hat{\boldsymbol{e}}[n] - \boldsymbol{e}[n]\|^2\} = \frac{1}{N}E\{\|\hat{\boldsymbol{c}}[n] - \boldsymbol{c}[n]\|^2\} = D. \quad (18)$$

Considering a simple scalar quantizer, the optimal rate allocation solves the following minimization problem:

$$\min \sum_{i=1}^{N} R_i \ \text{ subject to } \ \sum_{i=1}^{N} \frac{\lambda_i}{12} 2^{-R_i} = D, \qquad (19)$$

whose solution, apart from scaling constants is analogous to the so called 'reverse' water-filling algorithm (see [46], Theorem 13.3.3) where,

$$D_i = \begin{cases} K & \text{if } K < \lambda_i/12, \\ \lambda_i^2 & \text{if } K \geq \lambda_i/12 \end{cases}. \qquad (20)$$
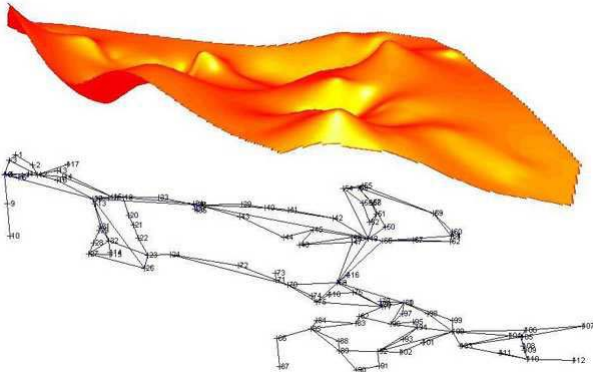
The rate of each quantizer is therefore

$$R_i = \log_2\left(\frac{\lambda_i}{12 D_i}\right)$$

The constant K is calculated recursively, using the constraint $\sum_{i=1}^{N} \frac{\lambda_i}{12} 2^{-R_i} = D$, setting initially $K = D/N$ and progressively setting $R_i = 0$ for the indexes such that $K \geq \lambda_i/12$, as prescribed in (20). Each entry of the vector $c[n]$, $\{c[n]\}_i = c_i[n]$ is therefore quantized with a different number of bits R_i and over a different range width $2\sqrt{\lambda_i}$, i.e. $\hat{c}_i[n] = Q_{R_i, \sqrt{\lambda_i}}(c_i[n])$. In the next section we characterized numerically the mean squared error (MSE) distortion $\|\boldsymbol{\theta}(nT_s) - \widehat{\boldsymbol{\theta}}(nT_s)\|^2$ attained on the data using this simple transform coding method.

A. Numerical results

Using the simulation data for the loop configuration discussed in Section IV-A, we implemented simple compression scheme as expressed in Section V. Since the scheme is suboptimum, it provides a bound on how many bits are required to achieve the desired distortion on the set of data we analyzed. The study of these trends can help establish whether or not the approaches discussed in Section III-B are relevant to the data aggregation of PMU measurements.

The PMU phase data are reconstructed recursively

$$\widehat{\boldsymbol{\theta}}[n] = \alpha\widehat{\boldsymbol{\theta}}[n-1] + \widehat{\boldsymbol{e}}[n] = \alpha\widehat{\boldsymbol{\theta}}[n-1] + \boldsymbol{\Psi}\widehat{\boldsymbol{c}}[n].$$

Using this reconstruction we computed the mean square error (MSE) per sensor corresponding to each step size. The MSE/sensor was obtained by averaging $\|\boldsymbol{\theta}(nT_s) - \widehat{\boldsymbol{\theta}}(nT_s)\|^2$ over the different contingency cases for which the data were simulated and over the time and then normalizing it by the number of sensors in the network (assuming of course, that there is a sensor on each bus).

Figure 10 shows how the the required rate increases as we request lower MSE/sensor. This follows the established trend from rate distortion theory. It can be seen that as the network density increases, the growth in the required rate tends to saturate. Figure 11 answers the question posed in the introduction of this paper - how many bits are needed to represent the state of the network which has a strong correlation structure and how does this scale with the density of the network. Figure 11 shows two curves - the original curve corresponds to the data obtained from the simulations and the fitted curve, obtained using Matlab poly-fitting tools, highlights the observed trend. It shows that if we fix the metric of MSE/sensor, then the required rate grows *sublinearly* with the network density. As mentioned elsewhere in this paper, this has important consequences for the communication architecture - 1) the PMUs should not transmit their data to the fusion center independently or asynchronously - an intermediate data aggregation step is beneficial and 2) we can add more PMUs to the network to achieve finer network monitoring without causing network congestion.
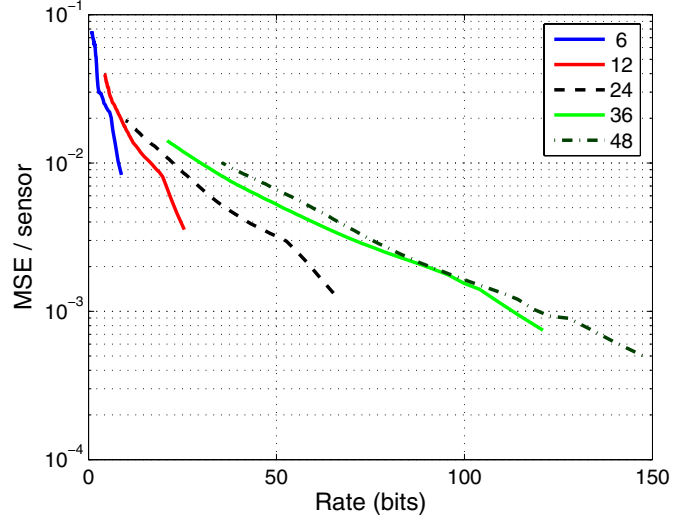

Fig. 10. The MSE/sensor decreases as the number of bits spent is increased. Data are plotted for increasing network density as indicated in the legend.
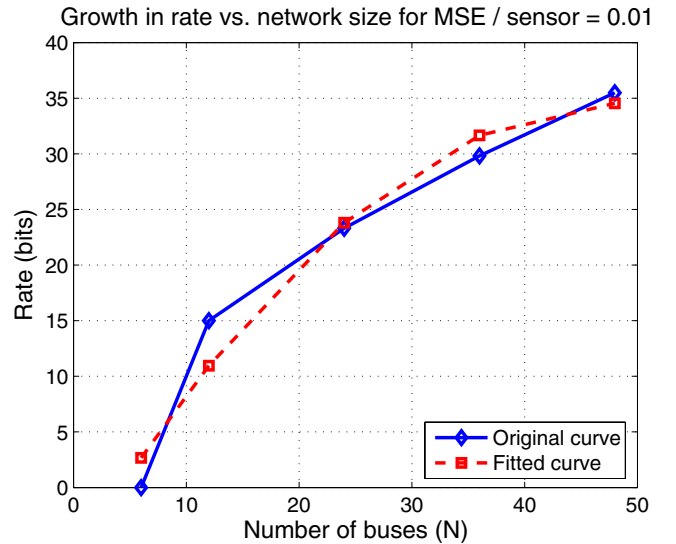

Fig. 11. For a pre-set MSE/sensor, the number of bits required to reconstruct the phase data from the PMUs grows sublinearly as the network density increases.

VI. CONCLUSION

The coupling in the dynamical equations of the power network leads to spatio-temporal correlation in the PMU measurements. This correlation impacts the number of bits that are needed to represent the state of the power network. Based on this phenomenon, in Section III-B we explored several relevant communication architectures for implementing a monitoring mechanism in power networks. To summarize, the broad communication themes are: distributed source coding, sparse representations and random projections, and data driven channel access methods. Finding good basis functions for obtaining *sparse* representations of the PMU data could open up a series of interesting choices from the data aggregation themes that have been discussed. Future work will address the selection of the theme most suited for the control application

in power networks and the development of a practical control mechanism. Numerical results demonstrate that the number of bits required to reconstruct the PMU data to within a given accuracy grows sub-linearly with the density of the power network. This shows that the scheme scales favorably as the number of PMUs in the power network is increased.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," in *IEEE Communication Magazine*, vol. 40, Aug. 2002, pp. 102–114.

[2] A.Gersho and R. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[3] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, "Foundations of control and estimation over lossy networks," *Proc. IEEE, Special Issue on Networked Control Systems, to appear in 2006*, 2006.

[4] O. C. Imer, S. Yuksel, and T. Basar, "Optimal control of dynamical systems over unreliable communication links."

[5] N. Elia and J. Eisembeis, "Limitation of linear control over packet drop networks," *Proc. IEEE Conference on Decision and Control*, vol. 5, p. 5152 5157, Dec. 2004.

[6] S. Tatikonda and S. Mitter, "Control under communication constraints," *Proc. IEEE Transactions on Automatic Control*, vol. 49, p. 1056 1068, 2004.

[7] N. Elia and S. Mitter, "Stabilization of linear systems with limited information," *Proc. IEEE Transactions on Automatic Control*, vol. 46, p. 1384 1400, 2001.

[8] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," in *Proc. IEEE Transactions on Automatomatic Control*, vol. 50, June 2005, p. 910 915.

[9] M. Lobashov, G. Pratl, and T. Sauter, "Implications of power-line communication on distributed data acquisition and control system," in *Proc. IEEE Conference on Emerging Technologies and Factory Automation*, vol. 2, 2003, pp. 607 – 613.

[10] S. Galli, A. Scaglione, and K. Dostert, "Broadband is power: internet access through the power line network," in *IEEE Communication Magazine*, vol. 41, May 2003.

[11] A. Scaglione and S. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *Proc. International Conference on Mobile Computing and Networking*, Atlanta, GA, September 2002.

[12] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. IT-19, pp. 471 – 480, July 1973.

[13] A. D. Wyner, "Recent results in the shannon theory," *IEEE Trans. Information Theory*, vol. IT-20, pp. 2 – 10, January 1974.

[14] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. IT-22, pp. 1 – 10, January 1976.

[15] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem [multiterminal source coding]," *IEEE Trans. Information Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[16] H. Viswanathan and T. Berger, "The quadratic gaussian ceo problem," *IEEE Trans. Information Theory*, vol. 43, no. 5, pp. 1549–1559, September 1997.

[17] Y. Oohama, "The rate-distortion function for the quadratic gaussian ceo problem," *IEEE Trans. Information Thoery*, vol. 44, no. 3, pp. 1057 – 1070, May 1998.

[18] S. Servetto, "Lattice quantization with side information," in *Proc. Data Compression Conference (DCC)*, Snowbird, Utah, March 2000.

[19] J. Bajcsy and P. Mitran, "Coding for the slepian-wolf problem using turbo codes," in *Proc. IEEE Global Telecom. Conference (GLOBECOM)*, November 2001, pp. 25 – 29.

[20] M. Sartipi and F. Fekri, "Distributed source coding in wireless sensor networks using ldpc coding: the entire slepian-wolf rate region," in *Proc. IEEE Wireless Communications and Networking Conference*, March 2005.

[21] R. Zamir, S. Shamai, and U. Erez, "Nested linear lattice codes for structured multiterminal binning," *IEEE Trans. Information Theory*, vol. 48, no. 6, pp. 1250 – 1276, June 2002.

[22] S.Pradhan and K. Ramchandran, "Distributed source coding using syndroms (discus): design and construction," *IEEE Trans. Information Theory*, vol. 49, no. 3, pp. 626 – 643, March 2003.

[23] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71 – 83, January 2005.

[24] D. Rebollo-Monedero, A. Aaron, and B. Girod, "Transforms for high rate distributed source coding,," in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, CA, November 2003.

[25] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," 2004. [Online]. Available: http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:math/0410542

[26] ——, *The dantzig selector: statistical selection when p is much larger than n, Preprint*, May 2005.

[27] D. L. Donoho, *Compressed sensing, Manuscript*, September 2004.

[28] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Proc. International Symposium on Information Processing in Sensor Networks (IPSN)*, Nashville, TN, April 2006.

[29] 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.

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

[31] ——, "Mixing times for random walks on geometric random graphs," in *Proc. SIAM Workshop on Analytic Algorithms and Combinatorics*, Vancouver, Canada, January 2005.

[32] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. IEEE Symposium on Fundations of Computer Science*, 2003, p. 482.

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

[34] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projection," *Submitted to IEEE Trans. Information Theory*, March 2005.

[35] M. Duarte, M. Wakin, D. Baron, and R. Baraniuk, "Universal distributed sensing via random projection," in *Proc. International Conference on Information Processing in Sensor Networks*, Nashville, TN, April 2006.

[36] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 613 – 626, February 2006.

[37] K. Liu and A. Sayeed, "Type-based decentralized detection in wireless sensor networks," *Submitted to IEEE Trans. Information Theory*, October 2005.

[38] Y. W. Hong and A. Scaglione, "On multiple access for correlated: a content-based group testing approach," in *Proc. IEEE Information Theory Workshop*, San Antonio, TX, October 2004.

[39] ——, "On multiple access for correlated sources: a content-based group testing approach," *submitted to IEEE Tran. Information Theory*, February 2005.

[40] ——, "Content-based multiple access: combining source and multiple access coding for sensor networks," in *Proc. IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, Sepetmber 2004.

[41] Y. W. Hong, A. Scaglione, and R. Manohar, "Dense sensor networks that are also power efficient: When more is less," *Milcom*, 2005.

[42] M. Parashar, J. S. Thorp, and C. E. Seyler, "Continuum modeling of electromechanical dynamics in large-scale power systems," *Proc. IEEE Trans. on Circuits and Systems*, vol. 51, pp. 1848 – 1858, Sept. 2004.

[43] A. Gallai and R. J. Thomas, "Coherency identification in large scale electric power system," *Proc. IEEE Trans. on Circuits and Systems, Special Issue on Power Systems*, vol. CAS-29, pp. 777 – 781, Nov. 1982.

[44] M. Parashar, "Continuum modeling of power networks," *Ph.D. Thesis, Cornell University*, 2003.

[45] Y. Dong, S. Chang, and L. Carin, "Rate-distortion bound for joint compression and classification with application to multiaspect scattering," *IEEE Sensors Journal*, vol. 5, pp. 481–492, June 2005.

[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.