Lossless Compression of Synchronized Phasor Measurements

Ray Klump, *Member, IEEE*, Pooja Agarwal, *Student Member, IEEE*, Joseph Euzebe Tate, *Member, IEEE*, Himanshu Khurana, *Senior Member, IEEE*

***Abstract*—By reporting time-synchronized phasor magnitudes and phase angles at rates at or above the system frequency, phasor measurement units promise to dramatically increase our ability to understand both historical and real-time power system conditions. This new information does not come without a cost, however, and one potential barrier to the effective utilization of this new data source is the increased amount of information transmission and storage capability these devices require. One way to mitigate the increased storage requirements of synchrophasor data is to compress the data, although this compression should not come at the cost of reduced accuracy. This paper proposes a new method for the lossless compression of voltage magnitude and phase data in which known characteristics of the power system are used to improve upon common, off-the-shelf compression techniques. The method is evaluated with real and simulated PMU data to show its effectiveness.**

***Index Terms*— phasor measurement units, data compression**

I. Introduction

The North American power grid is undergoing a major modernization effort as it transforms into what is commonly known as the "smart grid". At the heart of this modernization effort are cyber infrastructure components for advanced computing and networking. Key smart grid applications and drivers such as demand response and wide-area situational awareness will need to make effective use of these components in realizing the envisioned capabilities of the next-generation power grid. Such applications and capabilities will naturally involve the creation, transmission, storage and processing of large amounts of data. Addressing this data explosion will require research and development of data management and processing tools that are custom tailored to power grid applications and data flows. This paper focuses on an important aspect of the problem: synchrophasor data compression.

Wide area measurement systems (WAMS) for the power grid are recognized as one of the key functionalities of the emerging smart grid. The need for such systems was highlighted in the August 2003 blackout report [1], which included WAMS development as one of its major recommendations. Since the report was published, there has been a significant increase in the deployment of phasor measurement units (PMUs) [2], which provide measurements that are GPS-synchronized in time and at data rates several orders of magnitude higher than traditional SCADA measurement systems. Because the synchrophasor measurements obtained from every PMU in the system are referenced to the same GPS-based time signal, they can provide an accurate snapshot of the entire interconnection and significantly increase the wide-area awareness of grid operators. Recognizing this potential, the U.S. Department of Energy (DOE), the North American Electric Reliability Corporation (NERC), and a wide array of North American electric utilities, vendors, consultants, and academics have been collaborating over the past several years on the North American SynchroPhasor Initiative (NASPI), whose vision is to improve power system reliability through wide-area measurement, monitoring and control [3].

One of the main challenges being considered by NASPI is how to store and exchange the large amounts of data generated by PMUs. In the North American Eastern Interconnection (EI), the key architectural component is the SuperPDC [4], a centralized phasor data collection system located at the Tennessee Valley Authority (TVA) that collects, archives, and transmits synchrophasor data over the entire interconnection. Although the deployment of PMUs is still quite modest (approximately 120 within the EI), the SuperPDC at TVA is already collecting 3.6 billion measurements per day and requires data storage capabilities of 36 GB per day [5]. This number is expected to grow significantly over the next few years due in part to the number of Smart Grid Investment Grants [6] that include funding for synchrophasor deployments. NERC requirements for improved disturbance monitoring standards [7],[8] are also likely to drive an increased need for more synchrophasor data storage capability due to the ability of PMUs to help utilities meet these new standards [9]. The NERC dynamic measurement standard [8] requires retention of voltage, current, frequency, and power measurements for at least 3 years at the request of NERC, NPCC, or the Reliability Coordinator. In addition, recorded data for each disturbance must be retrievable in its native form for 10 calendar days regardless of whether or not it has been

This work was supported in part by the U.S. National Science Foundation (NSF) award 0524695. Data was provided by the Tennessee Valley Authority for evaluation of our algorithm.

R. Klump is with Lewis University, Romeoville, IL 60446 USA, and with the Information Trust Institute at the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (email: klumpra@lewisu.edu).

P. Agarwal and H. Khurana are with the Information Trust Institute at the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: pagarwl@illinois.edu, hkhurana@illinois.edu).

J. E. Tate is with the University of Toronto, Toronto, ON M5S 3G4 Canada (e-mail: zeb.tate@utoronto.ca).

explicitly requested. To meet these new requirements, utilities will have to account for the large amounts of data storage that are built into these standards.

One way to reduce the increased expenditures associated with increased data storage capability is to find ways to compress the data. Designing effective data compression techniques for phasor data faces interesting challenges. First, the techniques need to be lossless, because many applications being developed are based on access to high-precision accurate data and in some cases the standard requires the storage of data in its native format (i.e., without any loss of information). Secondly, floating point measurements such as those obtained from PMUs are typically harder to compress than textual information and integers, because the mantissas of consecutive floating point numbers based on IEEE standard 754 are separated by sign bits and exponent bits. Thirdly, observed synchrophasor data has shown significant variation over several different time scales. Although much of the variation in PMU measurements is correlated to events on the grid, it is nontrivial to determine precisely how events on the grid manifest themselves within the various PMU data streams. Fourthly, researchers working on the analysis of PMU data have had limited access to real PMU data, and the bulk of the work in PMU signal characterization has proceeded from a theoretical basis as in [10]. Further discussion of the differences in real and simulated PMU data, along with the impact on compression, is provided below in section IV.

With all of these challenges in mind, this work explores the design of lossless data compression techniques applied to phasor data archives such as the SuperPDC. We look specifically at the compression of voltage phasor data. Specifically, we adapt techniques used with great success in the lossless compression of images [11] to develop a two-stage synchrophasor measurement compression technique. The first stage consists of encoding based on expected power system behavior to increase redundancy in the data. The second stage involves applying off-the-shelf compression techniques to the more redundant data set. Since these off-the-shelf approaches improve with the level of redundancy, they yield a compressed data set that is smaller than would be achieved by naively applying standard compression techniques to raw synchrophasor data.

We demonstrate techniques suitable for the first stage of our proposed compression strategy. We propose a variation on the delta-encoding approach described in [12] that we call slack-referenced encoding (SRE). The pre-processing technique attempts to capitalize on the tendency of power system quantities to vary smoothly during non-transient periods and to respond as self-correlating clusters during transients. Using four different data sets, which include simulated transient stability data and actual data during disturbance and quiescent operating periods, we test the effectiveness of the preprocessing approach in combination with the following standard compression tools: Deflate [13], BZip2 [14], and LZMA [15].

This is the first work we know of that proposes and demonstrates an effective, easily computed lossless data compression technique for synchrophasor data archives. Furthermore, because it utilizes off-the-shelf compression algorithms, it can be implemented with minimal effort using widely available libraries. Our experimental results demonstrate practical value by analyzing the performance of the compression technique on both simulated and actual PMU data. These results show that SRE, combined with LZMA, compresses voltage phasor data effectively for all data sets, reducing data set size by a factor greater than 2 to 1 for the cases studied.

The rest of this paper is organized as follows. Section II describes common lossless data compression techniques and concludes by explaining the benefits of two-stage compression. Section III describes our pre-processing approach, SRE. Section IV presents the results of using our approach to compress four different data sets. Finally, Section V concludes by identifying opportunities for further refinement of the methods presented in this paper.

II. Lossless Data Compression Techniques

If PMU data archives are to be compressed, they must be compressed without loss, at least for now. There are both legal and technical justifications for this requirement. One of the hopes of deploying PMUs more widely is that traditional nonlinear state estimation [16], the process whereby unmeasured system quantities are fit using a nonlinear model to measured quantities, will evolve into a linear state estimator [17] in which unmeasured quantities can be directly calculated. It is not yet known exactly what level of accuracy is needed to ensure that a synchrophasor-based linear estimator works properly. However, if the state estimator results are to be used for other analyses (e.g., contingency analysis or online transient stability analysis), the introduction of error could cause unpredictable downstream errors. From a legal perspective, should an event occur that results in litigation, participants need to be able to argue that the data on which they based their decisions was faithfully captured, so that they can create a *post-mortem* picture of system conditions that faithfully represents the actual system at the time of the event. Also, the NERC dynamic recording standard [8] requires all dynamic recording data to be available for 10 days without modification, which requires lossless compression. Therefore, compressing data archives should not result in a loss of fidelity.

The encoding of measurement data can be described as lossless if the encoded data is represented at the level of precision at which the measuring devices record them, which will vary by device. A commonly used standard for specifying synchrophasor data is IEEE 754 [18] with single precision floating point values. With respect to this standard, our interpretation of “lossless” is the storage of all 23 bits of the mantissa. We will demonstrate later in this paper how different levels of precision impact the compression results and afford an opportunity for alternative encoding schemes.

Much work has been done in the area of lossless data compression, and [19] provides an excellent background in

this area. The purpose of lossless data compression is to encode data in such a way that, when it is later decompressed, the original data set results. Lossless data compression techniques generally perform two tasks. First, they develop a statistical model of the data. Then, they take advantage of the statistical model to transform the data such that the number of stored bits is decreased without any loss of information.

There are many different ways to transform the data to take advantage of its statistical properties. One of the simplest approaches, run-length encoding (RLE), records a run of repeating symbols as one symbol and the number of times it repeats [19]. Another technique called Swinging Door (SD) [20], discards repetitions, where a repetition is judged by whether consecutive values lie within some tolerance. Although traditionally a lossy technique, SD can be made lossless by storing some reference value in full and using a tolerance that is less than measurement error. A dictionary encoder such as any of the Lempel-Ziv variants [21],[22],[23] builds a dictionary of symbol sequences; when it encounters a symbol sequence in the data set that is in the dictionary, it replaces the symbol sequence with its position in the dictionary. An entropy encoder like Huffman coding [24] replaces each symbol with a sequence whose length depends on how frequently the symbol appears in the original data set. It therefore encodes the most frequently occurring symbols using the shortest sequences and less frequently occurring symbols using longer ones. Arithmetic coding [19] is very similar to Huffman coding, but it works on sequences of symbols rather than on individual ones, replacing them with a representative fraction computed from the individual symbol frequencies.

There are also many techniques that pre-process the data to maximize the effectiveness of these approaches. The Burrows-Wheeler Transform (BWT) [25] rearranges the data to increase the length or frequency of repetitive runs, so that dictionary and entropy encoders can more effectively compress the data. Delta Encoding (DE) [26], which computes and stores the actual changes in the quantities to be compressed from one stage to the next, is effective when the changes in the quantities recur frequently, so that the other techniques we have listed have more redundancy with which to work. Rather than store the actual differences in the quantities themselves, first-order linear prediction [27] predicts the next value in a sequence of data points and records the difference between the predicted value and the actual value. If the predictor is good, the errors will fall within a tight range near zero, yielding a highly repetitive and, therefore, compressible data set.

A number of well-known compression tools have emerged that combine these various transformations. Deflate [13], first implemented in PKZip, uses a combination of LZW (Lempel-Ziv-Welch) to eliminate duplicate sequences and Huffman coding to represent sequences using smaller codes proportional to their frequencies. The compression tool bzip2 [14] uses BWT to rearrange the data into a more repetitive form and then applies Huffman coding. LZMA [15] combines Lempel-Ziv's dictionary-based approach with an arithmetic coding step that rewards the most frequently occurring sequences with short dictionary indices.

Image compression tools use a variety of these techniques to compress bitmaps. Lossy approaches like JPEG judge adjacent pixels to be identically colored if their color difference lies within a tolerance. It then uses run-length encoding to replace runs of "identical" pixels. Like swinging door, JPEGs can be made less lossy by reducing the tolerance, but the limit is the original bitmap. The lossless image compression technique PNG [11] improves upon this by using a two-stage approach. The first stage uses linear prediction to predict the color of each pixel based upon the color of its neighbors and records the difference between the predicted and actual colors. It then uses Deflate to encode these differences, which should exhibit more redundancy than the original image data. We use a similar two-stage approach to compress PMU data in this work with the first stage using SRE inspired by an analysis of synchrophasor data.

III. Compression of PMU Data

In this section, we present the SRE approach to compressing voltage phasor data. SRE uses a modified form of delta encoding to preprocess the data so that it can then be more effectively compressed by a more generic, statistical compression technique. It is effective because of the spatial and temporal correlation of PMU data.

Suppose a data archive stores data from p different measurements, each taken over q cycles. If we let $\boldsymbol{M}$ represent the archive, then

$$\boldsymbol{M} = \left[\boldsymbol{M_1}\ \boldsymbol{M_2}\ \ldots \boldsymbol{M_p}\right] \tag{1}$$

where each $\boldsymbol{M_i}$ is a vector of q measurements,

$$\boldsymbol{M_i} = \left[M_{i,1}\ \ldots\ M_{i,q}\right] \tag{2}$$

The task is to preprocess $\boldsymbol{M}$ to make it more compressible. Here we propose a preprocessing strategy that takes advantage of two characteristics of power system quantities.

First, PMUs record measurements at some multiple of nominal system frequency, most commonly at 30 measurements per second. If a PMU recording at this rate monitors a 60 Hz sinusoidal signal, then every measurement will be the same, and the measurement-to-measurement difference will be a constant zero. Although real power system signals are neither perfectly sinusoidal nor operating continuously at 60Hz, these ideals can be used to approximate system operation reasonably well, particularly during quiescent periods. This suggests that delta encoding would be an effective preprocessing step for power system quantities, since the measurement-to-measurement differences will tend to be small. With delta encoding, each value is replaced by its change from the last measurement. In other words, $M_{i,j}$, the value of the ith measured quantity at time j, will be replaced by

$$M'_{i,j} = M_{i,j} - M_{i,j-1} \tag{3}$$

The system does not usually operate at a constant 60Hz frequency, so the measurement-to-measurement differences will not be zero as specified in the prevailing PMU standard [28]. We therefore observe a second characteristic of power system quantities, one that pertains particularly to phase angles and also to magnitude measurements at nodes that are electrically close. If the load, generation, and topology of the system remain fixed, then the frequency will be constant at all

the nodes, and so the angles will be fixed relative to one another. If we choose one node to be the reference, a technique commonly done in power flow studies with the selection of a slack bus, then the measurement-to-measurement differences will again tend toward zero. To recreate each quantity's data stream, then, we would need only the initial value of each quantity and the full stream for the reference node. Real systems do experience disturbances, however, and the system frequency will not be constant at all nodes. However, in the absence of disturbances, we expect the measurement-to-measurement differences after referring quantities to the slack bus to be quite small. Moreover, [12] demonstrates that this spatial correlation of angles exists even in the wake of stressful contingencies. This is the basis for the Slack-Referenced Encoding (SRE) scheme.

Specifically, if $M_{i,j}$ th represents the *i*th signal measured at time *j*, and if $M_{ref,j}$ represents the reference signal measured at time *j*, then SRE replaces $M_{i,j}$ by the following:

$$M'_{i,j} = \left(M_{i,j} - M_{ref,j}\right) - \left(M_{i,j-1} - M_{ref,j-1}\right) \qquad (4)$$

The voltage angles of nodes vary coherently as long as the frequencies remain close. For voltage magnitudes, nodes which are electrically close to each other because they are connected through low-impedance paths will exhibit similar voltages, and so referring voltages to a slack will be beneficial as long as the chosen slack is electrically close to the signal being compressed. More will be said about the choice of slack in Section IV.

To summarize, the SRE algorithm is applied to a time-indexed stream of measurement signals *s*, one of which is chosen as the reference signal s_{ref}. The algorithm is expressed succinctly as follows.

```
For each time t
  For each signal s
    If t = 0,
      record s(t)
    Else
      If s is s_ref,
        record s(t)-s(t-1)
      Else
        record (s(t)-s(t-1))-
                    (s_ref(t)-s_ref(t-1))
      Endif
    Endif
  Next s
Next t
```

Note that each record operation stores the value as an integer rather than a float. Integers compress more efficiently than float values, because the mantissas in a stream of float values are separated by sign and exponent bits, whereas no such separations exist in streams of integer values. Care must be taken to ensure that the conversion of values from floating point to integer representations occurs without loss.

IV. Results

We now evaluate the effectiveness of the proposed approaches in compressing data from a variety of different sources. Specifically, we test the effectiveness of the following preprocessing techniques

- Delta Encoding (DE)
- Swinging Door (SD)
- Slack-Referenced Encoding (SRE)

in combination with the following off-the-shelf compression tools

- Deflate (LZ77 + Huffman)
- Bzip2 (BWT + Huffman)
- LZMA

applied to voltage magnitude and angle data from the following data sets:

- Small Sim: 10-minute transient stability simulation of a 37-bus system with a line outage event [29] (432.7 Kbyte each of voltage magnitude and angle data)
- Quiescent: 10 minutes of real PMU data from the Tennesee Valley Authority (TVA) system measured during a quiescent period (351.6 Kbyte each of voltage magnitude and angle data)
- Event: 1 hour of PMU data recorded during and after a large line outage on the TVA system [29] (5.356 Mbyte each of voltage magnitude and angle data)
- Large Sim: 10 second transient stability simulation of a system with approximately 7,400 nodes during a 3-phase-to-ground fault event (33.84 Mbyte each of voltage magnitude and angle data)

Tables 1 through 9 show the effectiveness of the candidate compression techniques on the various data sets.

A. Small Sim

The 37-bus model shown in Fig. 1 was simulated under the outage of the JO138 – SAVOY138 line shown circled near the bottom right side of the diagram. Generators were modeled as round rotor synchronous machines with quadratic saturation and an IEEE Type I exciter. The slack bus, SLACK345, is located near the top right of the diagram.

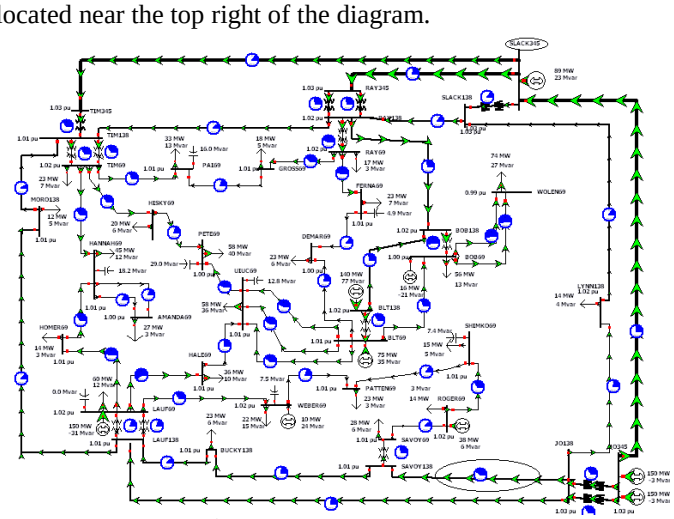

Fig. 1. One-line diagram for Small Sim system [29]

Table 1 shows the effectiveness of the various approaches to compressing 10 minutes worth of voltage magnitude data from the simulated outage. Table 2 shows the compression ratios for the various techniques applied to voltage phase angles. Table 1 shows that all techniques do a good job of compressing the voltage magnitudes, with SRE and DE doing a particularly effective job preprocessing the data so that a compression technique such as LZMA can compress it more effectively. The benefits of SRE and DE become quite

pronounced when they are applied to compressing phase angles, as Table 2 shows.

Fig. 2, which shows the variation of three of the phase angles during the simulation, helps highlight the benefits of preprocessing using SRE and DE for this case. After the disturbance, the system frequency quickly settles at the same value across the system, and the phase angles begin to change in synchronism with each other. Therefore, the phase angles at all nodes remain constant relative to the slack, yielding a highly compressible data set. In the case of DE, even though the phase angles are not referenced to the slack, they are changing at a fixed rate that results in a strong temporal correlation. Because of the fixed per-sample change in angles, the delta values settle to a constant, again yielding a highly compressible data stream.

Table 1
Compression ratios of voltage magnitudes, Small Sim

	Deflate	Bzip2	LZMA
Unprocessed	4.0878	3.9306	5.6383
SD	4.5241	4.9123	6.3615
SRE	10.1172	14.3504	13.8924
DE	8.7577	12.5092	12.6933

Table 2
Compression ratios of phase angles, Small Sim

	Deflate	Bzip2	LZMA
Unprocessed	1.1362	1.2059	2.3262
SD	1.5484	1.2058	2.1047
SRE	7.4212	10.2076	10.3717
DE	5.3794	9.2089	9.8546

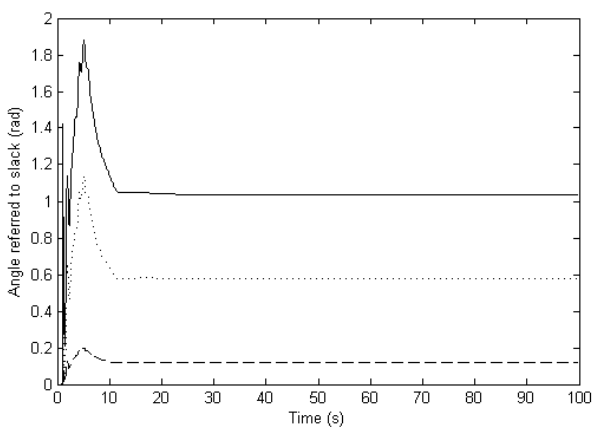

Fig. 2. Three angles from the Small Sim study that have been referenced to the slack bus

B. *Quiescent*

Real signals tend to be noisier than the results of simulation, and a usable a compression algorithm must work well in spite of this noise. To test the effectiveness of the approaches on real measurements, we first applied the techniques to ten minutes of voltage phasor measurements collected at seven buses on the TVA system during a period in which no significant disturbances occurred. Table 3 and Table 4 show the compression ratios. Again, SRE and DE prove particularly effective as preprocessing steps for both voltage magnitudes and phase angles, especially for Bzip2 and LZMA, with SRE having a slight edge. The compression ratios are not as high as in the Small Sim case due in part to the presence of noise in actual PMU data. For example, Fig. 3 shows the phase angle at one node relative to the phase angle at a chosen slack node. Even though this is a quiescent period, there is noise in the signal, and this reduces the compressibility of the data set.

Table 3
Compression ratios of voltage magnitudes, Quiescent

	Deflate	Bzip2	LZMA
Unprocessed	1.7565	2.2547	2.2131
SD	1.8028	2.3055	2.2516
SRE	2.0895	2.5401	2.5407
DE	2.0491	2.4899	2.4915

Table 4
Compression ratios of voltage angles, Quiescent

	Deflate	Bzip2	LZMA
Unprocessed	1.2681	1.2150	1.9523
SD	1.3324	1.2136	1.9356
SRE	2.3324	2.6234	2.9661
DE	2.0889	2.5284	2.6528

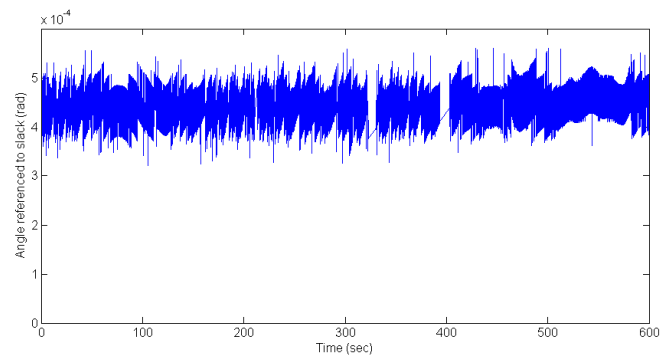

Fig. 3. A phase angle from the Quiescent measurement set that has been referenced to the slack bus

It is interesting to note the impact of filtering the noise on the compression results. One way to filter the noise is to store the floating point values at lower precision. Table 5 demonstrates for the Quiescent case how the compression ratio using LZMA in the compression stage improves when fewer bits of precision are stored. Float values at full precision store 23 mantissa bits, which represents approximately 7 significant digits. Table 5 suggests that, when meter accuracies are known, or when agreements are reached regarding the precision at which data must be stored to ensure the accuracy of studies that use the data, the compression techniques presented here will become more effective. In fact, one of the findings of [12] is that the number of bits needed to reconstruct PMU data exhibiting temporal and spatial correlation grows sublinearly as network density increases. Therefore, as PMUs become more widely deployed, there may be opportunities to store the data with fewer bits, which will result in better compression.

Table 5
Compression ratios of the Quiescent measurement set voltage magnitudes for various precisions using LZMA for the compression stage

	23 bits (full prec)	19 bits	16 bits
Unprocessed	2.2131	2.6428	4.3740
SRE	2.2516	3.6901	6.0392
DE	2.5407	3.6097	6.1008

C. *Event*

Next, the compression method was tested with real PMU data recorded at seven buses on the TVA system during a large line switching event [29]. Fig. 4 shows the per unit voltage magnitude recorded at one PMU in the system, and Fig. 5 shows the phase angle recorded at the same PMU.

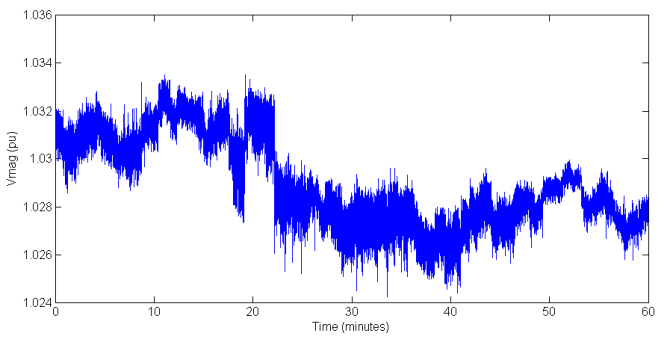

Fig. 4. A bus's voltage magnitude measurements from the Event data set

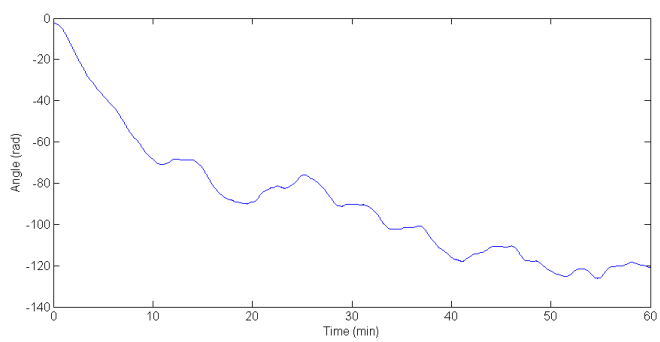

Fig. 5. Voltage phase angle measurements from the same bus as Fig. 4

Table 6 and Table 7 show how the various techniques compress this data set. Values were recorded at full floating-point precision. The DE and SRE approaches again prove to be effective preprocessors, especially for the phase angle data. The results in this case benefit both from the size of the data set and the gradual damping of the transient. As Fig. 5 suggests, the voltage and phase angle signals begin to settle down as the hour continues, so the latter part of this very long archive becomes highly compressible.

Table 6
Compression ratios of voltage magnitudes, Event

	Deflate	Bzip2	LZMA
Unprocessed	1.7192	2.2119	2.1452
SD	2.0733	2.4692	2.4251
SRE	1.9953	2.4581	2.4510
DE	2.0321	2.4965	2.5015

Table 7
Compression ratios of voltage angles, Event

	Deflate	Bzip2	LZMA
Unprocessed	1.5480	1.3388	2.4703
SD	1.7190	1.6445	2.2892
SRE	3.0158	3.5641	3.9377
DE	2.7784	3.3588	3.6802

D. *Large Sim*

Our final test case involves a simulation of a large system that experiences a three-phase-to-ground fault. This is a loosely coupled system, with many long transmission lines separating many discrete control areas. Ten seconds of voltage magnitude and phase angle calculations were recorded at approximately 7,400 nodes. The point of this example is to determine how the compression techniques would perform at a large "super PDC" that collects data from thousands of PMUs deployed throughout the grid.

Table 8 shows the compression ratios for the 7,400 voltage magnitude signals, and Table 9 shows the compression ratios for the corresponding phase angles. The slack used for SRE in this example is a single large generator. As one would expect for a large, loosely coupled system, the use of a single slack as the reference for the SRE does not provide the same benefits as it did for smaller systems. In fact, SRE performs significantly worse at helping compress voltage signals than the other approaches, although its performance with phase angles is comparable to that of DE. As node voltage magnitudes behave similarly only when the nodes are electrically close, this result is not surprising. It suggests that future work should investigate how best to choose reference signals to maximize compression for data archived in large data concentrators.

Table 8
Compression of voltage magnitudes, Large Sim

	Deflate	Bzip2	LZMA
Unprocessed	1.4202	1.5947	2.0611
SD	1.7093	2.2328	2.2513
SRE	1.2622	1.4461	1.4779
DE	1.5573	1.7452	3.0326

Table 9
Compression of phase angles, Large Sim

	Deflate	Bzip2	LZMA
Unprocessed	1.2353	1.2242	1.8534
SD	1.2867	1.3590	1.5705
SRE	1.6208	2.2318	2.8184
DE	1.6257	2.2158	2.9946

V. Conclusion and Future Work

This paper has proposed a two-stage compression technique for synchrophasor measurement unit data. The first stage capitalizes on characteristics of power systems, such as the interconnectedness of its nodes and that periodicity of its signals, to encode the data so that it can be compressed more effectively by standard compression techniques. We have

demonstrated the effectiveness of delta-encoding and a related approach called slack-referenced encoding that are particularly effective pre-processing techniques, with the SRE technique yielding best results for compressing phase angle data. Future work will look to increase the performance of these compression techniques for large-scale synchrophasor data archives. This will involve exploring how to choose references that maximize the compression ratios of SRE and investigating how to define the size and measurement set of each archive so that they may be more efficiently compressed.

VI. References

[1] U.S.-Canada Power System Outage Task Force. (2004) Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations. [Online]. Available: https://reports.energy.gov.

[2] A. G. Phadke, J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*, New York: Springer, 2008.

[3] North American Synchrophasor Initiative, "Synchrophasor Technology Roadmap", March 13, 2009. [Online]. Available: http://www.naspi.org/resources/2009_march/phasortechnologyroadmap.pdf.

[4] J. Zuo, R. Carroll, P. Trachian, J. Dong, S. Affare, B. Rogers, L. Beard, Y. Liu, "Development of TVA SuperPDC: Phasor applications, tools, and event replay," in *Proc. 2008 IEEE Power and Energy Society General Meeting*, 20-24 July 2008.

[5] openPDC, "Background". [Online]. Available: http://openpdc.codeplex.com/#Background.

[6] Recovery Act Selections for Smart Grid Investment Grant Awards, U.S. Department of Energy, October 27, 2009. [Online]. Available: http://www.energy.gov/8216.htm.

[7] *Define Regional Disturbance Monitoring and Reporting Requirements*, NERC Standard PRC-002-1, Aug. 2006.

[8] *Disturbance Monitoring Equipment Installation and Data Reporting*, NERC Standard PRC-018-1, Aug. 2006.

[9] M. Adamiak, R. Hunt, "Application of Phasor Measurement Units for Disturbance Recording", presented at the 10th Annual Georgia Tech Fault and Disturbance Analysis Conference, Atlanta, GA, USA. [Online]. Available: http://www.truc.org/papers.htm.

[10] S. Chakrabarti, E. Kyriakides, "PMU measurement uncertainty considerations in WLS state estimators", *IEEE Trans. Power Sys.*, vol. 24, no. 2, May 2009, pp. 1062-1071.

[11] G. Roelofs, "Compression and filtering" in *PNG: The Definitive Guide*. Sebastopol, CA: O'Reilly and Associates, 2003. [E-book]. Available: http://jamsb.austms.org.au/courses/CSC2406/semester3/resources/misc/pngbook/chapter09.html.

[12] S. Kirti, Z. Wang, A. Scaglione, R. Thomas, "On the Communication Architecture for Wide-Area Real-Time Monitoring in Power Networks," in *Proc. 40th Annual Hawaii International Conference on System Sciences*, 2007.

[13] P. Deutsch, "Deflate Compressed Data Format Specification version 1.3", RFC 1951, Aladdin Enterprises, May 1996. [Online]. http://www.ietf.org/rfc/rfc1951.txt

[14] J. Seward, "The bzip2 and libbzip2 official home page", [Online]. http://www.bzip.org/.

[15] I. Pavlov, "7z Format", [Online]. http://www.7-zip.org/7z.html.

[16] A. Monticelli, "Electric power system state estimation," Proc. IEEE, vol. 88, no. 2, pp. 262–282, Feb. 2000.

[17] A. G. Phadke, J. S. Thorp, K. J. Karimi, "State estimation with phasor measurements", *IEEE Trans. Power Sys.*, vol. 1, no. 1, Feb. 1986. pp. 233-238.

[18] *IEEE Standard for Floating-Point Arithmetic*, IEEE Standard 754-2008, Aug. 2008.

[19] D. Salomon, *Data Compression, The Complete Reference, Fourth Edition*, London: Springer, 2007.

[20] E. H. Bristol, "Data compression for display and storage", U.S. Patent 4 669 097, Oct. 21, 1985.

[21] J. Ziv, A. Lempel, "A universal algorithm for sequential data compression", *IEEE Trans. Info. Theory*, vol. IT-23, no. 3, May 1977,. pp. 337-342.

[22] J. Ziv, A. Lempel, "Compression of individual sequences via variable-rate coding", *IEEE Trans. Info. Theory*, vol. IT-24, no. 5, Sep. 1978, pp. 530-536.

[23] T. A. Welch, "A technique for high-performance data compression", *Computer*, vol. 17, no. 6, June 1984, pp. 8-19.

[24] D. A. Huffman, "A method for the construction of minimum-redundancy codes", *Proc. of the I.R.E.*, Sept. 1952, pp. 1098-1101.

[25] M. Burrows, D. J. Wheeler, "A block-sorting lossless data compression algorithm", Digital Equipment Corporation, Palo Alto, CA, SRC-RR-124, May 1994.

[26] D. Gottlieb, S. A. Hagerth, P. G. Lehot, H. S. Rabinowitz, "A classification of compression methods and their usefulness for a large data processing center", in *Proceedings of the May 19-22, 1975, National Computer Conference and Exposition*, Anaheim, CA, May 1975. pp. 453-458.

[27] G.-O. Glentis, K. Berberidis, S. Theodoridis, "Efficient least squares adaptive algorithms for FIR filtering", IEEE Signal Processing Magazine, July 1999.

[28] *IEEE Standard for Synchrophasors for Power Systems*, IEEE Standard C37.118-2005, Mar. 2006.

[29] J. E. Tate, T. J. Overbye, "Line outage detection using phasor angle measurements", *IEEE Trans. Power Sys.*, vol. 23, no. 4, Nov. 2008, pp. 1644-1652.

Ray Klump received his BS, MS, and PhD in electrical engineering from the University of Illinois at Urbana-Champaign. He is currently Associate Professor of Mathematics and Computer Science at Lewis University and a Visiting Research Associate Professor at the Information Trust Institute at Illinois. His current research interests include power system stability and smart grid security.

Pooja Agarwal received her BE degree in Computer Science and Engineering from Manipal Institute of Technology, India. She is currently pursuing MS at University of Illinois, Urbana Champaign. Her research interests lie in the field of network security and applied cryptography.

Joseph Euzebe Tate (S'03-M'08) received the BS degree in electrical engineering from Louisiana Tech University, Ruston, LA, in 2003 and the MS and PhD degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 2005 and 2008. He is currently an Assistant Professor in the Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto. His research focuses on the utilization of advanced metering and processor technologies to improve the reliability and efficiency of power system operations.

Himanshu Khurana received his MS and PhD from the University of Maryland, College Park in 1999 and 2002 respectively. He is currently a Principal Research Scientist at the Information Trust Institute and a Research Assistant Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. His research interests lie in the area of distributed system security, especially as applied to large-scale distributed systems and critical infrastructures.

Acknowledgements

The authors would like to thank TVA for providing data for these experiments. They would also like to thank Parth Gupta for his help in evaluating related compression techniques. This material is based upon work at University of Illinois supported by the National Science Foundation under Grant No. CNS-0524695 and Department of Energy under Award Number DE-OE0000097, and at the University of Toronto by the National Science and Engineering Research Center under Grant No. 372055.