Efficient Compression of PMU Data in WAMS

Phani Harsha Gadde, *Student Member, IEEE*, Milan Biswal, *Member, IEEE*,
Sukumar Brahma, *Senior Member, IEEE*, and Huiping Cao

***Abstract*—Widespread placement and high data sampling rate of current generation of phasor measurement units (PMUs) in wide area monitoring systems result in huge amount of data to be analyzed and stored, making efficient storage of such data a priority. This paper presents a generalized compression technique that utilizes the inherent correlation within PMU data by exploiting both spatial and temporal redundancies. A two stage compression algorithm is proposed using principal component analysis in the first stage and discrete cosine transform in the second. Since compression parameters need to be adjusted to compress critical disturbance information with high fidelity, an automated but simple statistical change detection technique is proposed to identify disturbance data. Extensive verifications are performed using field data, as well as simulated data to establish generality and superior performance of the method.**

***Index Terms*—Data compression, phasor data concentrator, phasor measurement units, wide area monitoring systems.**

I. Introduction

As WIDE area monitoring systems (WAMS) grow to monitor large interconnected power systems, the number of Phasor Measurement Units (PMUs) has also been growing rapidly. For example, in North America, after the initiation of the Western Interconnection Synchrophasor Project (WISP), the number of PMUs installed in the Western Electricity Coordination Council (WECC) system has increased from 56 in 2005 to 584 till April 2015 [1], [2]. The WISP shares operational data with 97 participants over a secured wide area network spanning the entire Western Interconnection [2] and has over 77 Phasor Data Concentrators (PDCs). In WISP, almost every PMU feeds data to a PDC. The detailed PMU datas-flow and interconnections in North America are mentioned in [3]. The rapid increase in the number of PMUs, higher sampling rate of the modern PMUs, and more stringent criteria for system integrity violations are expected to result in several fold expansion in the already large volumes of PMU data. As pointed out in [4], 100 PMUs having 20 measurements at 30 Hz sampling rate generate over 50 GB of data in one day. Higher sampling rate of 60 Hz or 120 Hz in modern PMUs will increase this data volume to 100 GB or 200 GB respectively. This means even larger volume of PMU data will need to be stored in PDCs, or transmitted to Super PDCs. The NERC PRC 002-2 [5] requires that the actual recorded disturbance data be preserved for 10 calendar days. However, all PMU data are finally archived by utilities. Archived data can be used for purposes like model validation, testing new wide area protection and control applications that use PMU data, or training/testing disturbance classifiers typically used in system visualization applications that are being commercialized.

Clearly, compression methods are necessary to archive these data. Such methods essentially seek to maximize the Compression Ratio (CR), and minimize the loss of data, which is measured by certain error metrics. The parameters chosen for any compression technique heavily depend on the nature of the data. In case of PMU data, most of the data will be relatively constant or slow-varying with overriding noise, but in case of a disturbance the data will vary. The speed and the nature of variation again depend on the type of a disturbance. Therefore, any compression technique chosen to compress PMU data with high fidelity should be able to detect a disturbance and change the compression parameters for efficiently compressing *any* disturbance data, and be robust enough not to be affected by noise in the data. Validation of any approach should therefore encompass real world data, including disturbance data corresponding to major power system disturbances.

A lossless compression approach with slack referenced encoding (SRE) is suggested for PMU data in [4], where the voltage data are compressed with a maximum CR of 10.12. This approach is heavily biased towards fidelity at the expense of CR, and is not comprehensively tested. In [6] authors have used wavelet packet transform (WPT) to compress the PMU data resulting in a low CR of 2 with root mean square error (RMSE) of 3.68×10^{-6}. A comparative assessment of standard compression algorithms has been presented in [7], where *szip* algorithm is used to achieve a CR of 2.77 for voltage data and 3.77 for frequency data, which are low values. In [8] an embedded zerotree wavelet based denoising and compression method is introduced, but it lacks proper analysis, and doesn't perform well for real PMU data. Recently, a dimensionality reduction algorithm has been proposed by Xie *et al.* [9] for PMU data utilizing principal component analysis (PCA) subspace for detection of disturbances, but the issue of compression is not adequately addressed. Another real

Manuscript received August 28, 2015; revised December 3, 2015 and January 20, 2016; accepted February 26, 2016. Date of publication March 15, 2016; date of current version August 19, 2016. This work was supported in part by the National Science Foundation, and in part by the Centers of Research Excellence in Science and Technology (CREST) under Grant HRD-1345232. Paper no. TSG-01016-2015.

P. H. Gadde, M. Biswal, and S. Brahma are with the Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM 88003 USA (e-mail: phgadde@nmsu.edu; milanb@nmsu.edu; sbrahma@nmsu.edu).

H. Cao is with the Computer Science Department, New Mexico State University, Las Cruces, NM 88003 USA (e-mail: hcao1@nmsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2016.2536718

time event detection and data archival scheme has been proposed in [10], where PCA is used for event detection, and least square curve fitting is used for compression. It reports a relatively better CR of 63.3 for voltage data, but with a higher normalized RMSE of 0.176 for steady state data. Moreover, the CR reduces to 4 for the voltage data during a disturbance event. This paper does not consider current and frequency data which are also recorded by PMUs, and will have different variations during disturbances. In addition, it uses data from PMUs in a campus microgrid, and hence their method is not tested extensively for large scale data from transmission systems, which have quite different disturbance patterns than a microgrid. The aim of this paper is to propose a generalized method that works with any number of PMUs and significantly improves the CR while maintaining excellent fidelity, and show the robustness of the method on extensive data from field as well as from accurate simulations.

With the occurrence of a power system disturbance, multiple PMUs at different geographical locations are triggered simultaneously, capturing various snapshots of the *same event*. Therefore, it stands to reason that the parametric signatures of all snapshots would be similar. Thus, correlation among multiple local PMU data streams at the PDC can be exploited for the compression of data. Exploiting this crucial property of PMU data, in this work we use PCA to minimize the spatial redundancy among data captured by multiple PMUs. We further reduce temporal redundancies in the data by applying Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) on the principal coefficients (PCs). A method to choose adaptive window length is developed to attain efficient compression for both, normal as well as disturbance data. We finally apply a classical lossless data compression technique on the data that are already compressed using the above approach to make the storage even more efficient.

II. Methodology

Let $P_i(n)$ denote the measured voltage, frequency or current data sample expressed in per unit at time index n recorded by the i^{th} PMU with a reporting rate of f_s samples per sec (sps). Let there be N number of PMUs in a geographical region, connected to one PDC. The block diagram describing the proposed approach for compression is shown in Fig. 1. The proposed method implements Statistical Change Detection (SCD) to detect and isolate disturbance data. This technique will be described in Section II-A. If no disturbance is detected, a fixed data window W of 10 s is used. This value is based on studies presented in Section II-B, which shows that the 10 s window provides a good combination of data-fidelity and CR for normal PMU data. If, however, a disturbance is detected, continuously computed Statistical Variance (SV) is used to select an appropriate W, which is used to capture the disturbance details with higher fidelity, although with some reduction in CR. Then, data compression is achieved as described in the previous paragraph of Section I. The following subsections describe the processes within the block diagram of Fig. 1. The performance indices for compression used in this paper are CR (represents reduction in data) as defined in (1), RMSE (represents error, or dissimilarity between original and reconstructed data averaged over the length of dataset) as defined in (2), and Maximum Absolute Deviation Error - MADE (represents the worst case error at any sample point) as defined by (3). These are widely accepted measures, and together form a demanding compression metric.

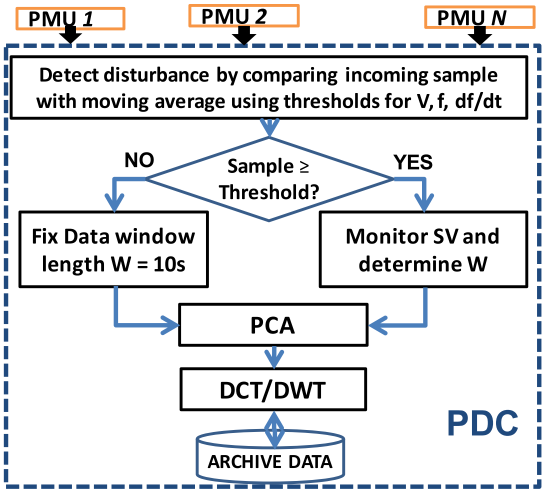

PMU 1
PMU 2
PMU N
Detect disturbance by comparing incoming sample with moving average using thresholds for V, f, df/dt
NO
Sample ≥ Threshold?
YES
Fix Data window length W = 10s
Monitor SV and determine W
PCA
DCT/DWT
ARCHIVE DATA
PDC

Fig. 1. Block Diagram of proposed PMU Data Compression Technique.

$$CR = \frac{N_o}{N_c}, \tag{1}$$

where, N_c is the number of data points after compression, and N_o is the number of data points in the original data.

$$RMSE = \sqrt{\frac{\sum_{n=1}^{L}\left(P_i(n) - P_i^{'}(n)\right)^2}{L}}, \tag{2}$$

$$MADE = max\left\{\left|\left(P_i(n) - P_i^{'}(n)\right)\right|\right\} \; ; n = 1, 2, \ldots L, \tag{3}$$

where $P_i(n)$ and $P_i^{'}(n)$ are the original and reconstructed data samples, respectively, from the i^{th} PMU, and L is the length of the data sequence. Note that Higher CR and lower errors mean higher quality compression.

A. Statistical Change Detection

In order to preserve critical changes in PMU measurements due to disturbances, a higher fidelity is needed. However, in an automated environment, it is required to detect the inception and duration of disturbances, so the fidelity can be enhanced for that period. A statistical change detection (SCD) algorithm is used for this purpose. Instead of using the standard 10 s window length, a variable window length that envelops the full disturbance is used for compression of disturbance data. The NERC PRC 002 disturbance triggering criteria [5] recommends either of the following to detect a disturbance i) frequency $<$ 59.55 Hz or $>$ 61 Hz, ii) rate of change of frequency $df/dt > 0.124$ Hz/s, iii) Undervoltage trigger set no lower than 85% of the normal operating voltage for a duration of 5 seconds. These criteria are too conservative and capture only major disturbances. Since we would like even the smaller variations in PMU data to be preserved with higher fidelity, we select less conservative (more aggressive) trigger criteria for voltage and frequency - voltage $\leq$ 99% of nominal voltage, frequency $\leq$ 59.94 Hz, or $\geq$ 60.06 Hz (meaning Δf of $\pm 0.1\%$). Disturbances contributing smaller variations in our study were mainly capacitor switchings. These thresholds can be relaxed for larger number of PMUs, since there is more likelihood of a PMU being "electrically close" to any given disturbance (discussed in Section III-B). It should be noted that this choice can be left to the user; the method is not affected by it.

Thus, $P_i(n)$, the incoming sample n generated by PMU i is declared a disturbance sample if the deviation $\Delta(n)$=$|P_i(n) - \mu(n)|$ falls outside our selected thresholds, where $\mu(n) = \frac{1}{K}\sum_{k=n-K}^{n-1} P_i(k)$, which is the average value of the measured parameter over the 10 s window before disturbance. For PMUs with reporting rate of f_s sps, $K = f_s * 10$.

Once $P_i(n)$ is declared a disturbance sample, we start measuring the statistical variance (SV) of the disturbance samples over a moving window of 3 cycles as per (4). SV is tracked for the PMU that contributed data resulting in the largest deviations from thresholds, in other words, for the PMU that captured the strongest disturbance signatures.

$$SV(n) = \frac{1}{s}\sum_{k=0}^{s-1}(P_i(n+k) - \mu(n))^2, \qquad (4)$$

where s is the number of samples inside the window over which variance is calculated. For $f_s = 60$ sps, $s = 3$; for $f_s = 30$ sps, the number is rounded to 2. The calculation of variance starts when the window fills up with s samples once the disturbance is detected, and the value of n in (4) keeps increasing by one as the window slides one sample at a time. The window length is based on the assumption that the disturbance will last at least for 3 cycles. A user can choose to reduce this based on the PMU reporting rate. For $f_s = 30$ sps, which is the reporting rate of field data in our study, there is just one sample every two cycles; hence this choice. It should be noted that in case a disturbance lasts less than 3 cycles, the choice will be conservative, and will still capture the disturbance with higher fidelity.

Disturbance is considered over when all samples (s) in the moving window fall below the trigger threshold. Thus, the SV threshold which will indicate the termination of disturbance window would be $\frac{s \times 0.01^2}{s} = 10^{-4}$ pu for voltage, and $\frac{s \times 0.001^2}{s} = 10^{-6}$ pu for frequency. The voltage and frequency data of a recorded event and the selection of disturbance windows based on this logic are shown in Figs. 2 and 3. The light vertical lines are the fixed 10 s window, whereas the thick vertical lines show the data window selected by SCD.

B. Principal Component Analysis

PCA transforms the coordinates of the multidimensional data in R^d such that the data are expressed in terms of d new principal axes, arranged in the order of decreasing variance. PCA has the ability to reveal hidden dynamics underlying a complex dataset by identifying the data as a linear combination of new basis vectors or PCs. The first PC of a given set of correlated variables encodes the maximum variance of the data, and is a linear combination of the variables having the maximum variability among all linear combinations. The second PC defines the next largest amount of variation not accommodated by the first PC, and is orthogonal to the first PC. Thus, d^{th} PC has the least contribution to the data. Some of these PCs contributing to small variation of the data are neglected to achieve dimensionality reduction.

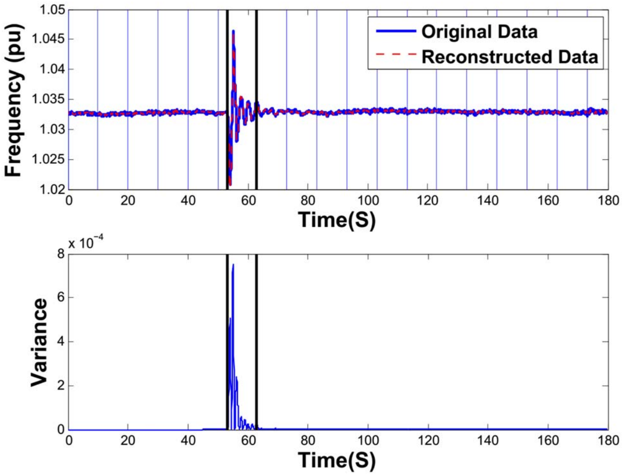


Fig. 2. The light vertical lines represent the fixed 10 s window and the thick vertical lines show the data window selected by SCD.

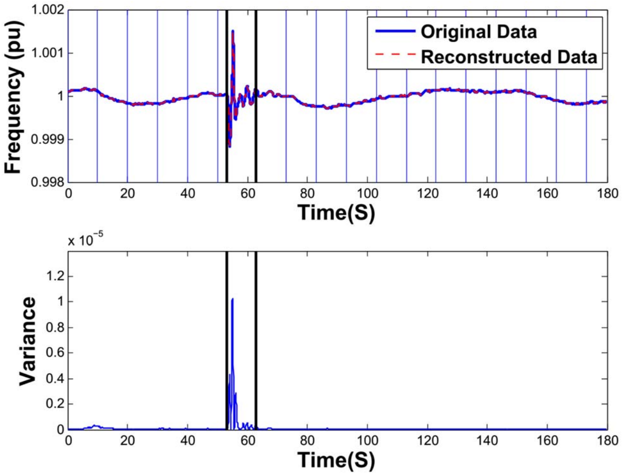


Fig. 3. The light vertical lines represent the fixed 10 s window and the thick vertical lines show the data window selected by SCD.

PCs can be calculated by determining the orthogonal eigenvectors of the data covariance matrix [9]. Each PMU records ($M = T \times F_s$) data samples for each variable in a duration of T seconds with a sampling rate of F_s samples per seconds. Then the streaming data from N PMUs for each variable can be expressed as a data matrix $D_{M\times N}$. The typical variables measured by PMUs are voltage phasors, current phasors, frequency, and rate of change of frequency (ROCF). For any measurement variable, the data matrix is expressed as:

$$D_{M\times N} = \begin{pmatrix} P_1(1) & P_2(1) & \cdots & P_N(1) \\ P_1(2) & P_2(2) & \cdots & P_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ P_1(M) & P_2(M) & \cdots & P_N(M) \end{pmatrix}$$

The data matrix $D_{M\times N}$ is normalized according to (5).

$$X_{M\times N} = \frac{D_{M\times N} - DM_{M\times N}}{DS_{M\times N}} \tag{5}$$

Where $DM_{M\times N}$ and $DS_{M\times N}$ are the sample mean and standard deviation matrix whose elements are calculated according to (6) and (7).

$$DM(n) = \frac{\sum_{m=1}^{M} D(m, n)}{M} \tag{6}$$

$$DS(m, n) = \sqrt{\frac{\sum_{i=1}^{M} (D(i, n) - DM(n))^2}{M}} \tag{7}$$

where $n = 1, 2, \dots, N$, and $m = 1, 2, \dots, M$.

The next step in PC calculation is to compute a sample covariance matrix (CM) V of dimension $(N \times N)$ according to (8).

$$V_{N\times N} = X^T X \tag{8}$$

The eigenvectors of the sample CM arranged in decreasing order of their eigen values or variances are the PCs. Eigenvalues and eigenvectors can be computed using Singular Value Decomposition (SVD) of the normalized data matrix X or eigen value decomposition (EVD) of the covariance standardized data matrix V. The EVD of V is expressed as:

$$V_{N\times N} = EYE^T, \tag{9}$$

where E is a $(N \times N)$ matrix consisting of the eigen vectors or PCs and Y is a $(N \times N)$ diagonal matrix with the diagonal values representing the variances of the respective PCs.

The goal is to discard PCs with relatively less variance without loss of critical information. There are two major challenges to be addressed here: (i) What should be the criteria for discarding the insignificant PCs? (ii) What should be the window length or duration of PMU data stream to be processed at a time?

The first question is answered by selecting minimum number of PCs such that the reconstruction errors are significantly low. One approach is to fix a threshold value for PC and reject the PCs falling below it, but the choice of such a threshold strongly depends on the nature of the data, and the number of PMUs. Therefore, we choose the first Q number of PCs such that the root mean square error (RMSE) between the reconstructed and original sequences remains less than 10^{-4} and MADE remains less than 10^{-2} pu, which are quite stringent limits. For example, Fig. 2(a) and 3(a) show the original and the reconstructed data with these error thresholds for a recorded event. Clearly, the performance is excellent for normal as well as disturbance data. RMSE and MADE are tied to the normalized cumulative variance of the preserved PCs. Let λ_i represent the eigen value or variance corresponding to the i^{th} PC, the normalized cumulative variance (Λ) is expressed as: $\Lambda = \frac{\sum_{i=1}^{Q} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$. We observed that for RMSE to be less than 10^{-4} and MADE to be less than 10^{-2} pu, $\Lambda \geq 0.8$ for normal data, and $\Lambda \geq 0.95$ for the disturbance data.

Answer to the second question involves a trade off between window length and compression ratio, because a longer data window length will require higher computations, more memory, and longer execution time, thus affecting the near real-time processing capabilities. In contrast, a small data window does not result in a good compression ratio. A comparison of the average execution time and the CR for various lengths of data windows for our field data is presented in Table I. Based on this comparison, a data window length (W) of 10 *s* appears as a good trade-off. However, the fixed window width is employed for the signal segment under normal conditions. While processing disturbance data we propose an adaptive window length based on statistical change detection, explained in Section II-A.

TABLE I
EXECUTION TIME AND CR

Window Length (Sec)	Execution Time (Sec)	CR
180	24.65	4.0652
20	0.72	4.0498
10	0.213	3.9921
3	0.104	3.8704
2	0.0897	3.7489

C. Temporal Redundancy Minimization

PCA carries out a linear transform on the PMU data on to a reduced dimension subspace. The resulting significant PCs can be interpreted as equivalent PMU data. Therefore, temporal redundancy among time samples in individual PMU data is also expected to be reflected in PCs. Motivated by this, we propose to apply compression algorithms meant for minimizing redundancy among samples in a time series on the PCs. In this work we evaluated the performance of two widely accepted algorithms used in data compression literature; DCT [11], [12] and DWT [13], [14]. DCT is an integer transform which converts the data into a transform domain in terms of cosinusoid basis functions. The transform domain is generally sparse with some coefficients having relatively small magnitude and carrying negligible information. These insignificant coefficients can be neglected based on a threshold value without affecting fidelity. The DWT maps the data onto a transform domain, but offers a large number of basis functions or wavelets to choose from. The sparsity of the data in the transform domain is decided by the chosen basis function. We evaluated a large number (90 types) of wavelet basis, including members of basic Haar, Daubechies, and Symlets wavelet families, and found that the *db*1 wavelet to be the most suitable.

As in the case of PCA, cumulative energy based thresholds were used for discarding insignificant DWT and DCT coefficients. If $C_k(i)$ denotes the i^{th} DWT or DCT coefficient, the cumulative energy contained in the first L_c coefficients (starting from the largest coefficient) is calculated as $\sum_{i=1}^{L_c} |C_k(i)|^2$. The normalized cumulative energy is expressed as $\nu = \frac{\sum_{i=1}^{L_c} |C_k(i)|^2}{\sum_{\forall i} |C_k(i)|^2}$. In order to keep RMSE and MADE within the limits, the threshold values of $\nu \geq 0.9$ for normal data and $\nu \geq 0.95$ for disturbance data had to be selected.

TABLE II
COMPRESSION OF FIELD PMU DATA WITH SCD AND PCA

Data type		CR	RMSE	MADE
Voltage	Magnitude	2.76	2.228×10^{-6}	1.423×10^{-3}
	Angle	3.43	3.435×10^{-6}	0.841×10^{-4}
Frequency		4.57	0.515×10^{-6}	0.503×10^{-3}
Current	Magnitude	3.71	4.613×10^{-6}	1.324×10^{-3}
	Angle	3.79	3.79×10^{-6}	1.421×10^{-3}

TABLE III
COMPRESSION OF FIELD PMU DATA WITH SCD, PCA AND DWT

Data type		CR	RMSE	MADE
Voltage	Magnitude	5.43	5.663×10^{-6}	5.965×10^{-3}
	Angle	9.44	10.966×10^{-6}	2.028×10^{-3}
Frequency		12.24	7.959×10^{-6}	1.618×10^{-3}
Current	Magnitude	3.65	7.147×10^{-6}	2.247×10^{-3}
	Angle	5.86	9.705×10^{-6}	1.631×10^{-3}

III. PERFORMANCE EVALUATION

A. Performance on Field Data

The compression performance of the proposed approach was first evaluated with field data obtained from four PMUs placed at 345 kV buses, and owned by a utility in the U.S. A total of 1582 disturbance records were available spanning the period from 2007 to 2010. Each record contained data-streams for voltage magnitude & angle, current magnitude & angle, and frequency, all recorded at 30 frames per second (fps). Each record is of 3-minute duration, containing 55 *s* of pre disturbance and 125 *s* of post disturbance data. The physical locations of the PMUs, data acquisition, and pre-processing to remove bad data is discussed in [15]. These data were compressed using the SCD based adaptive windowing strategy and thresholds for PCA, DWT, and DCT as explained in Section II.

Compression performance after the first stage - PCA with SCD - is shown in Table II for data sequences corresponding to different measurements. Each data sequence is a concatenation of 1582 3-minute long field-recorded files as mentioned earlier. It can be seen that the overall compression of these sequences have moderate CR with errors significantly below the targeted value. This is because the number of PCs is an integer, so the choice results in Λ greater than the targeted thresholds.

To illustrate the improved performance after the second stage of compression, the compression performance obtained with combined SCD, PCA, and DWT is summarized in Table III, and the performance with SCD, PCA, and DCT is summarized in Table IV. Clearly, the compression with DCT outperforms the compression with DWT, and is hence recommended. While the RMSE and MADE remain in more or less the same low range, the increase in CR is substantial. An important observation from the results is that compression ratio is not the same for all variables. This makes sense because the variations in different parameters are different in magnitude and duration due to the physics, resulting in different volumes of disturbance data to be compressed. This aspect has not been highlighted in published literature.

TABLE IV
COMPRESSION OF FIELD PMU DATA WITH SCD, PCA AND DCT

Data type		CR	RMSE	MADE
Voltage	Magnitude	8.27	3.579×10^{-6}	4.573×10^{-3}
	Angle	16.15	2.7073×10^{-6}	3.766×10^{-3}
Frequency		24.17	1.1949×10^{-6}	1.675×10^{-3}
Current	Magnitude	6.55	9.087×10^{-6}	5.820×10^{-3}
	Angle	12.41	11.739×10^{-6}	2.228×10^{-3}

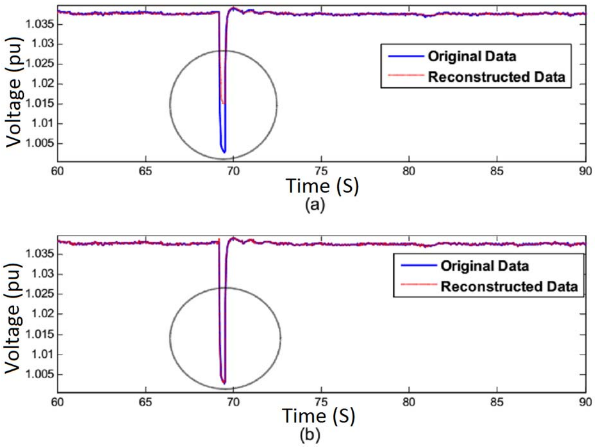


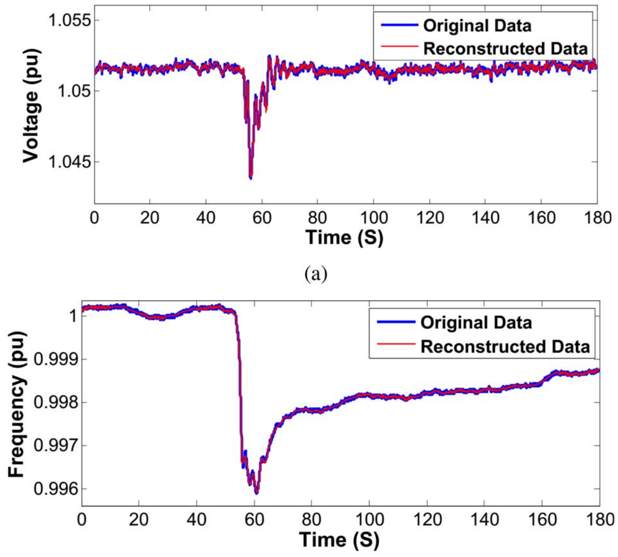
Fig. 4. The original and the reconstructed voltage waveforms with (a) fixed window, (b) SCD.

In order to illustrate the effectiveness of the proposed SV based disturbance detection and adaptive window (SCD), we illustrate in Fig. 4 the reconstructed waveforms when a) data were compressed by PCA and DCT with a fixed 10 s window, and b) data were compressed by SCD, PCA, and DCT. Clearly, the compression performance is comparable for steady state data, where the variations are minimal, but the fixed window approach loses crucial information during the sharp disturbance. The waveform shown here is taken from field data.

Fig. 5 demonstrates the reconstruction performance for a sample voltage and frequency dataset recorded by a field PMU during an event, when SCD-PCA-DCT is used for compression. It shows that not only the crucial variations in the disturbance data are preserved with excellent fidelity, but the noise in the waveforms is also reduced (not eliminated). The PCA maps the spectrally correlated observations into uncorrelated PCs, so discarding some insignificant PCs results in limited noise suppression. The filtering properties of DCT and DWT are well known.

B. Performance on Simulated Data

Since the proposed method exploits spatial correlation of multiple PMUs capturing the same event, it should perform better with a larger number of PMUs. The field PMU data has only four PMUs connected to a PDC. However, the number of PMUs and PDCs have significantly increased over the last few years as quantified in Section I. Some commercial PDCs offer capacity for integrating up to 40 PMUs. Therefore, it is important to examine how the proposed method performs with



(a)

(b)

Fig. 5. The original and the reconstructed disturbance waveforms recorded by a PMU corresponding to (a) voltage and (b) frequency.

a larger number of PMUs. As can be seen from the description of the proposed methodology in Section II, the method is general, and can accommodate any number of PMUs for choosing the significant PCs.

In addition, the field data might not have included all the possible major types of disturbances, as they were generated by the NERC PRC 002 disturbance triggering criteria as described in Section II-A. The field data had confirmed labels corresponding to only three types of disturbances - fault, generation loss, and line tripping. It is worthwhile to verify the compression performance for other major disturbance types encountered in power systems.

To address these factors, we simulated equivalent PMU data in General Electric's positive sequence loaf flow (PSLF) software using the WECC power system model corresponding to summer 2009. Guided by the actual PMU location map [3], we simulated 10 PMU-measurements at different geographical locations around New Mexico and Arizona to record voltage & current phasors, and frequency. In total, 385 disturbance files were created comprising of equal number of faults (FLT), generation loss (GNL), load switching (LS) on/off, shunt capacitor switching (SHC) on/off, shunt reactor switching (SHR) on/off, synchronous motor switching (SMS) off, and series capacitor switching (SRC) on/off. Typical waveforms corresponding to most of these disturbances can be found in [16]. Each file, like the field data, was 3-minute long, with 55 s of pre-disturbance data. But in this case, the sampling rate was kept 60 fps, consistent with the new generation of PMUs. Based on reference [9], we added 92 dB additive white Gaussian noise to the simulated data to mimic real world PMU data. Since the method has been shown to reduce noise in field data, a different noise value should not affect the results. These files were subjected to compression using the SCD-PCA-DCT technique with same implementation as that for field data. While using the SCD to detect disturbances, it was found that capacitor switchings had the smallest variations. However, if they occurred "electrically nearer" to a PMU, the variations in that PMU data were larger. This means that the disturbance detection thresholds can be relaxed for larger number of PMUs. Our choice has been made *conservatively* for the actual PMU-placements in the New Mexico and Arizona area of the WECC system.

TABLE V
COMPRESSION OF SIMULATED PMU DATA WITH SCD AND PCA

Data type		CR	RMSE	MADE
Voltage	Magnitude	8.93	1.091×10^{-6}	0.456×10^{-3}
	Angle	10.33	0.132×10^{-6}	0.102×10^{-3}
Frequency		13.27	0.291×10^{-6}	0.162×10^{-3}
Current	Magnitude	6.80	4.567×10^{-6}	1.220×10^{-3}
	Angle	8.11	2.705×10^{-6}	1.041×10^{-3}

The compression performance for simulated data for different disturbance types is summarized in TABLE V and TABLE VI. Representative original (black) and reconstructed (red) waveforms of voltage and frequency for some of the disturbances are shown in Fig. 6 – 9 to visually demonstrate the compression and filtering performance.

C. Discussion of Compression Performance

Comparison of TABLE II and TABLE V clearly shows the performance of PCA improves significantly with larger spatial correlation that comes from larger number of PMUs (10 for simulated data, and 4 for field data). TABLE VI shows that DCT enhances this performance further. Thus, the proposed method promises better compression performance as the number of PMUs increase. Note again that the CR is different for different parameters, because they register different variations for the same disturbance event.

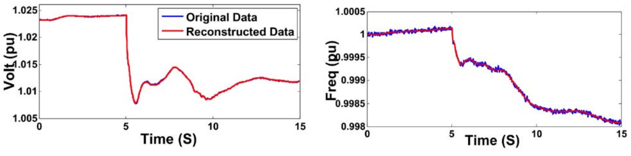
D. Applying Standard Lossless Compression Algorithms

Performance comparison of some standard lossless compression algorithms on PMU data is presented in [7], which indicated very low CR. We chose to apply some widely used lossless compression algorithms described in [17] that target floating point data. These algorithms mainly use entropy coding for compressed binary representation of the data. We considered 7-zip compressor which has the option for a dictionary compression scheme named Lempel-Ziv-Markov Chain Algorithm (LZMA), DEFLATE based on LZ77 & Huffman coding, bzip2 based on Burrow-Wheeler's block sorting algorithm & Huffman Coding, and a version of non-dictionary prediction by partial matching (PPM) scheme named PPMd algorithm [18]. To evaluate the compression performance, the actual number of bytes required for storing the data in a hard disk was calculated. The compression ratio considering physical memory requirements is expressed as:

$$CR_{PM} = \frac{\text{Physical storage required for raw data}}{\text{Physical storage required for compressed data}} \tag{10}$$

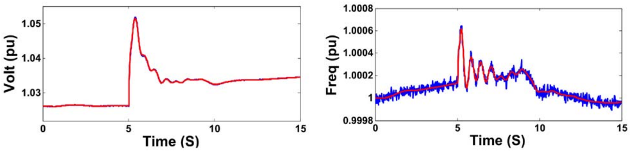
TABLE VI
PERFORMANCE ON COMPRESSING SIMULATED PMU DATA

			FLT	GNL	LS on/off	SHC on/off	SHR on/off	SMS off	SRC on/off	ALL
Voltage	Magnitude	CR	14.56	16.82	20.50	18.83	18.02	21.25	19.19	18.45
		RMSE	1.82×10^{-6}	1.80×10^{-6}	1.02×10^{-6}	0.73×10^{-6}	0.85×10^{-6}	0.80×10^{-6}	0.88×10^{-6}	1.13×10^{-6}
		MADE	1.21×10^{-3}	0.48×10^{-3}	0.50×10^{-3}	0.98×10^{-3}	1.60×10^{-3}	0.73×10^{-3}	0.73×10^{-3}	0.82×10^{-3}
	Angle	CR	20.58	21.39	21.33	20.94	20.21	21.76	22.76	21.29
		RMSE	0.71×10^{-6}	0.66×10^{-6}	0.15×10^{-6}	0.17×10^{-6}	0.17×10^{-6}	0.21×10^{-6}	0.25×10^{-6}	0.33×10^{-6}
		MADE	0.92×10^{-3}	0.94×10^{-3}	0.12×10^{-3}	0.13×10^{-3}	0.15×10^{-3}	0.83×10^{-3}	0.28×10^{-3}	0.48×10^{-3}
Frequency		CR	32.63	38.49	34.32	32.78	35.66	36.28	38.29	31.207
		RMSE	0.63×10^{-6}	0.31×10^{-6}	0.80×10^{-6}	0.78×10^{-6}	0.92×10^{-6}	0.19×10^{-6}	0.20×10^{-6}	0.55×10^{-6}
		MADE	0.99×10^{-3}	0.16×10^{-3}	0.94×10^{-3}	0.80×10^{-3}	0.80×10^{-3}	0.10×10^{-3}	0.35×10^{-3}	0.59×10^{-3}
Current	Magnitude	CR	11.72	11.85	11.75	12.27	12.63	10.96	12.23	11.91
		RMSE	5.81×10^{-6}	7.39×10^{-6}	3.27×10^{-6}	4.48×10^{-6}	8.18×10^{-6}	6.62×10^{-6}	3.66×10^{-6}	5.63×10^{-6}
		MADE	1.02×10^{-3}	1.10×10^{-3}	1.71×10^{-3}	1.45×10^{-3}	1.86×10^{-3}	1.03×10^{-3}	1.30×10^{-3}	1.35×10^{-3}
	Angle	CR	16.88	18.69	17.42	18.11	18.62	16.41	17.77	17.07
		RMSE	6.11×10^{-6}	8.13×10^{-6}	7.33×10^{-6}	7.14×10^{-6}	8.21×10^{-6}	8.11×10^{-6}	6.32×10^{-6}	7.33×10^{-6}
		MADE	1.33×10^{-3}	1.14×10^{-3}	1.07×10^{-3}	1.63×10^{-3}	1.44×10^{-3}	1.13×10^{-3}	1.34×10^{-3}	1.30×10^{-3}



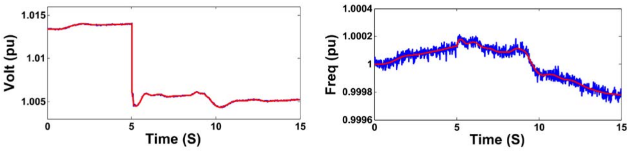
(a) (b)

Fig. 6. The original and reconstructed (a) voltage and (b) frequency waveforms during loss of generation.



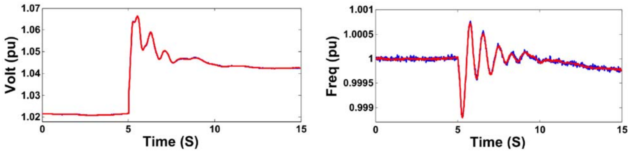
(a) (b)

Fig. 7. The original and the reconstructed (a) voltage and (b) frequency waveforms during switching off of load.



(a) (b) Frequency

Fig. 8. The original and the reconstructed (a) voltage and (b) frequency waveforms during Shunt capacitor switching off.



(a) Voltage (b)

Fig. 9. The original and the reconstructed (a) voltage and (b) frequency waveforms during series capacitor switching on.

TABLE VII
CR_{PM} OF FIELD PMU DATA WITH STANDARD ALGORITHMS

		LZMA	PPMd	bzip2	DEFLATE
Voltage	Magnitude	4.38	3.98	5.97	4.38
	Phase	12.12	3.03	7.57	3.03
Frequency		13.72	9.63	16.09	12.27
Current	Magnitude	1.78	0.91	1.79	1.79
	Phase	3.31	1.65	2.48	2.47

TABLE VIII
CR_{PM} OF FIELD PMU DATA WITH STANDARD ALGORITHMS APPLIED ON THE COMPRESSED DATA

		LZMA	PPMd	bzip2	DEFLATE
Voltage	Magnitude	**27.91**	26.22	26.86	22.88
	Phase	**38.77**	32.62	32.81	33.17
Frequency		38.91	38.55	**40.12**	38.56
Current	magnitude	18.13	16.42	**18.46**	18.11
	Phase	33.26	32.24	31.14	**33.56**

The field PMU data files were stored in comma separated variable (CSV) format, in a Personal Computer with NTFS File System under Windows 7 operating system. Physical storage requirement was recorded from the stored folder properties. After applying compression, the storage required by the compressed files was also recorded from the folder properties. The comparison of CR_{PM} on raw field PMU data with standard lossless compression algorithms is presented in TABLE VII. We further applied these standard algorithms on the compressed coefficients obtained using our proposed SCD-PCA-DCT approach. The resulting CR_{PM} are presented in TABLE VIII. It can be concluded from these tables that the standard lossless compression techniques in addition with the proposed compression approach significantly reduces the physical storage requirements. TABLE VIII shows that LZMA and bzip2 perform well overall, with LZMA having a slight edge. The overall reduction in physical storage requirement while storing all the PMU data (voltage, current, and frequency) was observed to be $1 - \frac{1}{31.39} = 96.8\%$ with $CR_{PM} = 31.39$, which shows notable compression performance.

IV. Implementability

The average running times of the compression and decompression techniques when implemented in MATLAB for a 3-minute data file were observed as 0.33 *s* and 0.41 *s* respectively. As most of the disturbances last for less than 10 *s*, the maximum length of data stored in the buffer is 10 *s*, which takes an average of 0.018 *s* for compression and storage. However, a dedicated and optimized executable implementation can result in much smaller running time. Therefore the proposed algorithm does not consume excessive time.

V. Conclusion

This paper develops a data compression method based on physics-based assumptions of spatial and temporal redundancies existing in WAMS data recorded by multiple PMUs. First a criterion to detect and capture data corresponding to a power system disturbance is laid out based on the NERC PRC 002 standard, which can be applied to system dynamics resulting from any disturbance. The compression methods - PCA to exploit spatial redundancy and DCT to exploit temporal redundancy - themselves are independent of the operating conditions, in the sense that their performance is dictated by two error bounds that are widely accepted and used in data compression across many domains. However, thresholds need to be set to preserve the energy (reflected in PCs or DCT-coefficients) in the compression methods such that the reconstruction errors remain within bounds. These thresholds are set by comprehensive testing over a large set of data. The data consist of field data and high fidelity simulation data generated using PSLF - a state of the art software used in industry. Therefore, the suggested thresholds have a high degree of credibility and generality across different operating conditions. The compressed data using PCA and DCT are further subjected to lossless floating point compression via LZMA algorithm to reduce physical memory requirement. The proposed compression technique is general, and hence can be applied to a PDC fed from any number of PMUs, and does not need any modifications if this number changes during operation. It is shown that the current trend of increase in the number of PMUs will make the compression even better. The compression results are significantly better than those reported in literature. The low computation time of the compression and decompression algorithms illustrate the near real-time adaptability of the proposed approach.

References

[1] K. E. Martin, “Phasor measurement systems in the WECC,” in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Montreal, QC, Canada, 2006, pp. 1–7.

[2] A. Silverstein. (Apr. 2015). *PMUs and You: An Update on Synchrophasor Tech Across America*. [Online]. Available: http://www.energybiz.com/article/15/04/pmus-and-you-update-synchrophasor-tech-across-america/.

[3] North American Synchrophasor Initiative. (Feb. 2014). *PMUs and Synchrophasor Data Flows in North America*. [Online]. Available: https://www.smartgrid.gov/document/pmus_and_synchrophasor_data_flows_north_america/.

[4] R. Klump, P. Agarwal, J. E. Tate, and H. Khurana, “Lossless compression of synchronized phasor measurements,” in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Minneapolis, MN, USA, Jul. 2010, pp. 1–7.

[5] “Disturbance Monitoring and Reporting Requirements,” North Amer. Electr. Rel. Corp., Atlanta, GA, USA, Tech. Rep. PRC-002-2, Nov. 2014.

[6] J. Khan, S. Bhuiyan, G. Murphy, and J. Williams, “PMU data analysis in smart grid using WPD,” in *Proc. IEEE PES T D Conf. Exp.*, Apr. 2014, pp. 1–5.

[7] P. Top and J. Breneman, “Compressing phasor measurement data,” in *Proc. IEEE Power Energy Soc. Gen. Meeting (PES)*, Jul. 2013, pp. 1–4.

[8] J. Khan, S. M. A. Bhuiyan, G. Murphy, and M. Arline, “Embedded-zerotree-wavelet-based data denoising and compression for smart grid,” *IEEE Trans. Ind. Appl.*, vol. 51, no. 5, pp. 4190–4200, Sep./Oct. 2015.

[9] L. Xie, Y. Chen, and P. R. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.

[10] Y. Ge *et al.*, “Power system real-time event detection and associated data archival reduction based on synchrophasors,” *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2088–2097, Jul. 2015.

[11] D. Salomon, *Data Compression: The Complete Reference*. New York, NY, USA: Springer, 2004.

[12] K. R. Rao and P. C. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Boston, MA, USA: Academic Press, 2014.

[13] J. Ning, J. Wang, W. Gao, and C. Liu, “A wavelet-based data compression technique for smart grid,” *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 212–218, Mar. 2011.

[14] S. Santoso, E. J. Powers, and W. M. Grady, “Power quality disturbance data compression using wavelet transform methods,” *IEEE Trans. Power Del.*, vol. 12, no. 3, pp. 1250–1257, Jul. 1997.

[15] O. P. Dahal and S. M. Brahma, “Preliminary work to classify the disturbance events recorded by phasor measurement units,” in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, Jul. 2012, pp. 1–8.

[16] O. P. Dahal, S. M. Brahma, and H. Cao, “Comprehensive clustering of disturbance events recorded by phasor measurement units,” *IEEE Trans. Power Del.*, vol. 29, no. 3, pp. 1390–1397, Jun. 2014.

[17] P. Ratanaworabhan, J. Ke, and M. Burtscher, “Fast lossless compression of scientific floating-point data,” in *Proc. Data Compress. Conf. (DCC)*, 2006, pp. 133–142.

[18] I. Pavlov. (Apr. 2015). *7-Zip File Archiver*. [Online]. Available: http://www.7-zip.org/.

Phani Harsha Gadde (S’15) received the B.Tech. degree from Jawaharlal Nehru Technological University, India, in 2013. He is currently pursuing the Master’s degree in power and energy systems with New Mexico State University, USA.

Milan Biswal (M’09) received the Ph.D. degree from Siksha ‘O’ Anusandhan University, India, in 2013. He is currently a Postdoctoral Researcher with the CREST Interdisciplinary Center of Research Excellence in Design of Intelligent Technologies for Smart Grids, New Mexico State University, USA. His current research interests include signal processing applications in smart grid, specifically for wide area monitoring systems.

Sukumar Brahma (M’04–SM’07) is the William Kersting Endowed Chair Associate Professor and the Associate Director of the Electric Utility Management Program, New Mexico State University, USA. He is an Editor of the IEEE Transactions on Power Delivery. He is the Past Chair of the IEEE PES’s Life Long Learning Subcommittee and Distribution System Analysis Subcommittee, the Chair of Power and Energy Education Committee, and a Member of Power System Relaying Committee.

Huiping Cao is an Assistant Professor of Computer Science with New Mexico State University. Her research interests are in the areas of data mining and databases. She has published data management and data mining articles in highly competitive venues. She has served on the editorial board of the *Journal on Data Semantics* as a Reviewer for peer-reviewed journals, and as a Program Committee Member for many international conferences and workshops.