Signal-Preserving Compression for Large Synchrophasor Measurement Data Sets

Zeb Tate
University of Toronto
zeb.tate@utoronto.ca

Ray Klump
University of Illinois
klump@illinois.edu

Project Overview

- Goal
  - Signal-preserving compression of synchrophasor data archives
- Approach
  - Pre-compression steps tailored to power system domain
- Results
  - Using simulated and measured data
  - Compression ratios up to 20:1
- Future work

Why Bother Compressing Data?

- 1 PMU taking 20 measurements at 30 sps
  - 4,800 bytes / second
- 40 PMUs reporting to a PDC
  - 15.5 GB / day
- 10 PDCs reporting to a centralized data reposity (e.g., SuperPDC)
  - 56 TB / year
- Less storage space means
  - **Less hardware** (reduced budget)
  - **More redundancy** (reduced data loss)

Key Principles for Synchrophasor Compression

- Knowledge of grid behavior should inform the compression of synchrophasor data
  - For example, under normal operating conditions, buses exhibit coherency in frequency
    - Similar to the coherencies in neighboring pixels of an image, which is the basis for PNG, JPG, etc.
- Compression of data archives must not reduce signal (information) content

Initial Approach

Two step process combining domain knowledge with high-performance, generalized compression tools

1. Pre-compression using knowledge of grid behavior
   – Exploit temporal and spatial correlation of signals
2. General compression tools
   – Apply state-of-the-art, common off-the-shelf (COTS) compressors such as bzip2, LZMA

Pre-Compression for PMU Data

- Idea: Use two types of signal coherency
  - Temporal
    - Cycle-to-cycle changes are usually small
  - Spatial
    - Signals from proximate buses evolve together
- Implementation: record sample-to-sample differences in each signal referenced to the sample-to-sample difference of a reference signal

Slack-Referenced Encoding

Input data (p data sets, with q data points per set)

$$\mathcal{M} = \{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p\}$$

$$\mathbf{m}_i = \{m_i(\Delta T), m_i(2\Delta T), m_i(3\Delta T), \ldots, m_i(q\Delta T)\}$$

Difference Encoding (DE)

$$\mathbf{m}_i^{DE} = \{m_i^{DE}(\Delta T), m_i^{DE}(2\Delta T), \ldots, m_i^{DE}(q\Delta T)\}$$

$$m_i^{DE}(k\Delta T) = \begin{cases} m_i(\Delta T) & \text{if } k = 1 \\ m_i(k\Delta T) - m_i((k-1)\Delta T) & \text{if } k \in [2, q] \end{cases}$$

Slack-Referenced Encoding (SRE)

$$\mathbf{m}_i^{SRE} = \{m_i^{SRE}(\Delta T), m_i^{SRE}(2\Delta T), \ldots, m_i^{SRE}(q\Delta T)\}$$

$$m_i^{SRE}(k\Delta T) = m_i^{DE}(k\Delta T) - m_{REF}^{DE}(k\Delta T)$$

Compression Stage

- Use well-known, high-performance lossless techniques
- Chosen implementations:
  - DEFLATE
  - BZip2
  - LZMA

Test Cases

Name	Type	Description	Duration	Size
Small Sim	Simulation	37-bus system with line outage	10 minutes	432.7 KByte
Quiescent	TVA	Data from 8 PMUs during quiet period	10 minutes	351.6 Kbyte
Event	TVA	Data from 13 PMUs during an interval with a sizable line outage	60 minutes	5.36 Mbyte
Large Sim	Simulation	7,400 PMUs during a 3-phase-to-ground fault	10 seconds	67.9 Mbyte

Compression Results
Voltage Magnitudes

COMPRESSION RATIOS OF VOLTAGE MAGNITUDES FOR DIFFERENT TEST CASES USING VARIOUS COMPRESSION TECHNIQUES

Case	Max Compression with Pre-Processing (SRE, DE, or SD)	Max Compression w/o Pre-Processing	Improvement by Pre-Processing
Small Sim	14.35 (Bzip2 + SRE)	5.64 (LZMA)	2.54
Quiescient	2.54 (LZMA + SRE)	2.25 (BZip2)	1.13
Event	2.50 (LZMA + DE)	2.21 (BZip2)	1.13
Large Sim	3.03 (LZMA + DE)	2.06 (LZMA)	1.47

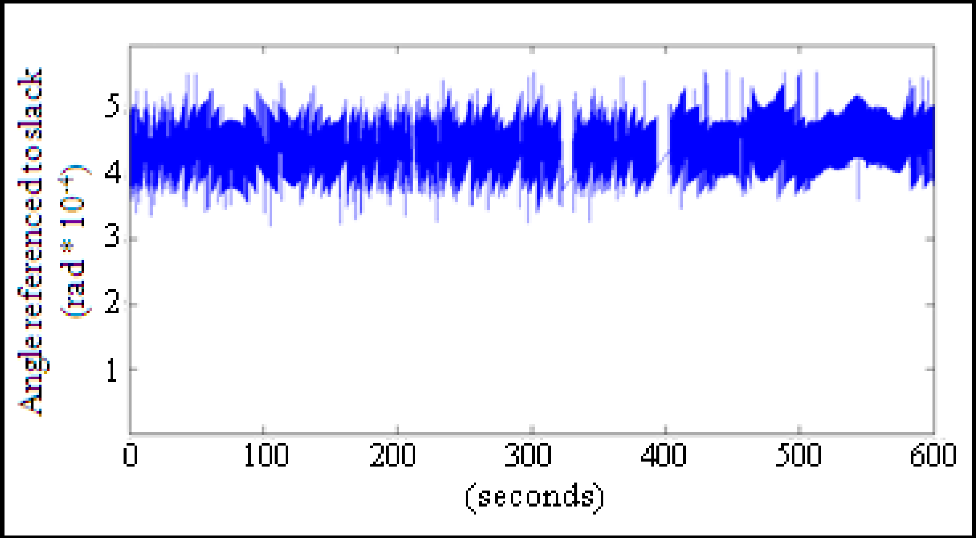
Compression Results
Phase Angles

COMPRESSION RATIOS OF VOLTAGE PHASE ANGLES FOR DIFFERENT TEST CASES USING VARIOUS COMPRESSION TECHNIQUES

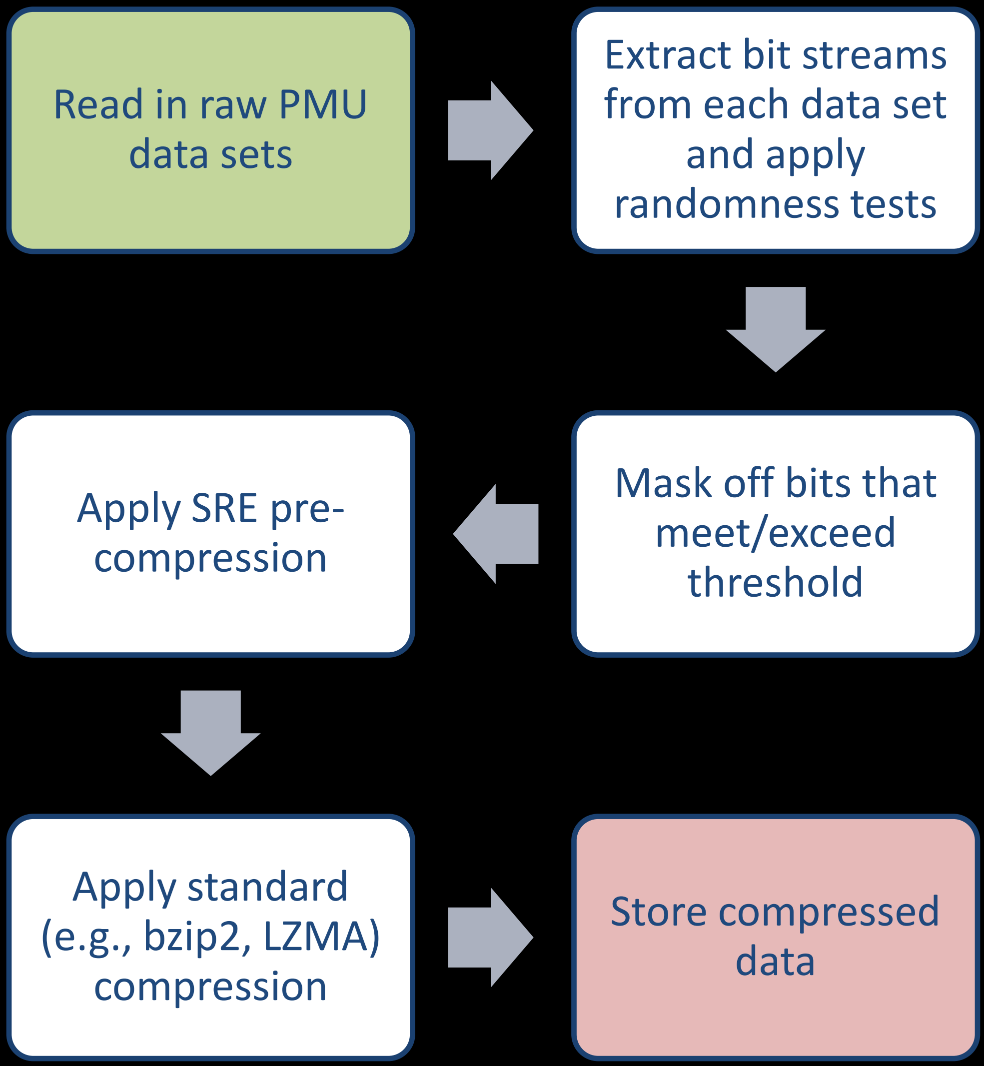
Case	Max Compression with Pre-Processing (SRE, DE, or SD)	Max Compression w/o Pre-Processing	Improvement by Pre-Processing
Small Sim	10.37 (LZMA + SRE)	2.33 (LZMA)	4.45
Quiescient	2.97 (LZMA + SRE)	1.95 (LZMA)	1.52
Event	3.94 (LZMA + SRE)	2.47 (LZMA)	1.60
Large Sim	2.99 (LZMA + DE)	1.85 (LZMA)	1.62

Impacts of Random Noise on PMU Data Compression

- Bad news: random noise in signals not compressible
- Good news: we don't care about storing/retrieving random noise
- Solution: Identify and remove noise before (pre-)compression

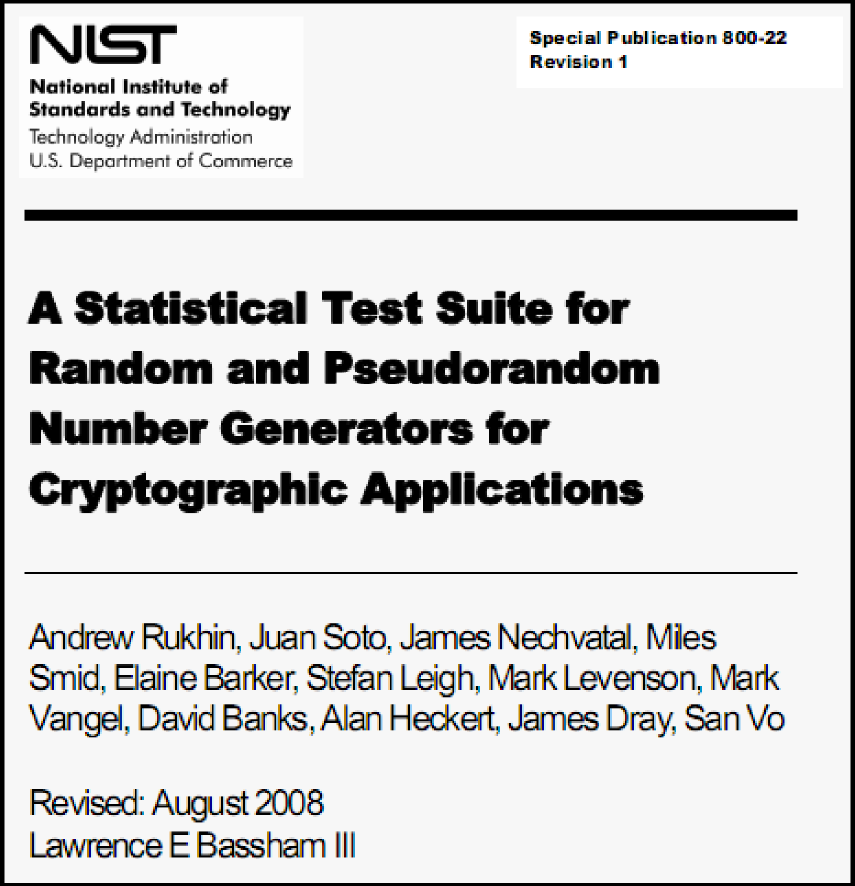


Signal-Preserving Data Compression



Identification of Random Bitstreams

- Frequency
- Block frequency
- Cumulative sums
- Runs
- Longest run of ones
- Spectral test


NIST
National Institute of Standards and Technology
Technology Administration
U.S. Department of Commerce

Special Publication 800-22
Revision 1

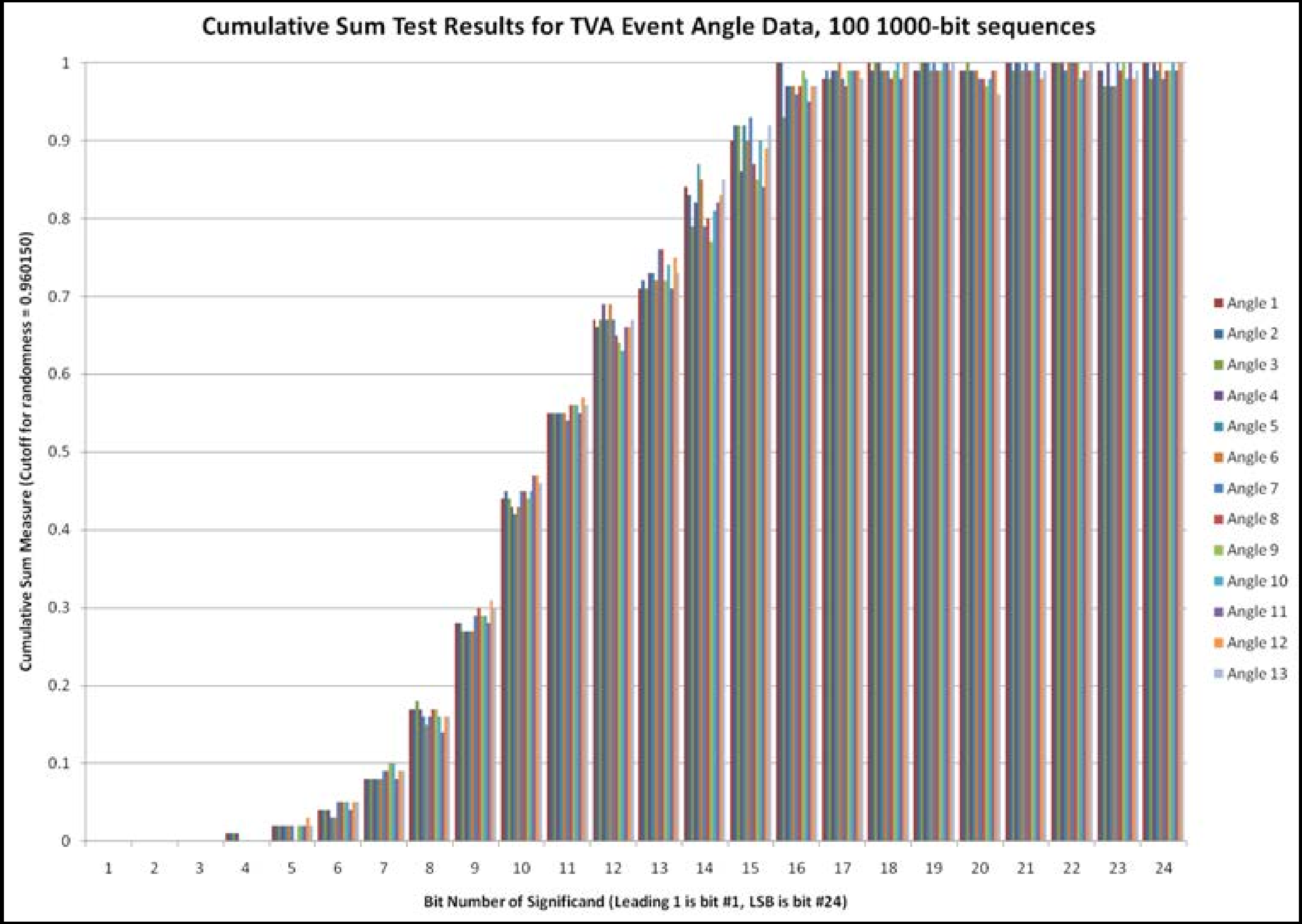
A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications

Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, Elaine Barker, Stefan Leigh, Mark Levenson, Mark Vangel, David Banks, Alan Heckert, James Dray, San Vo

Revised: August 2008
Lawrence E Bassham III

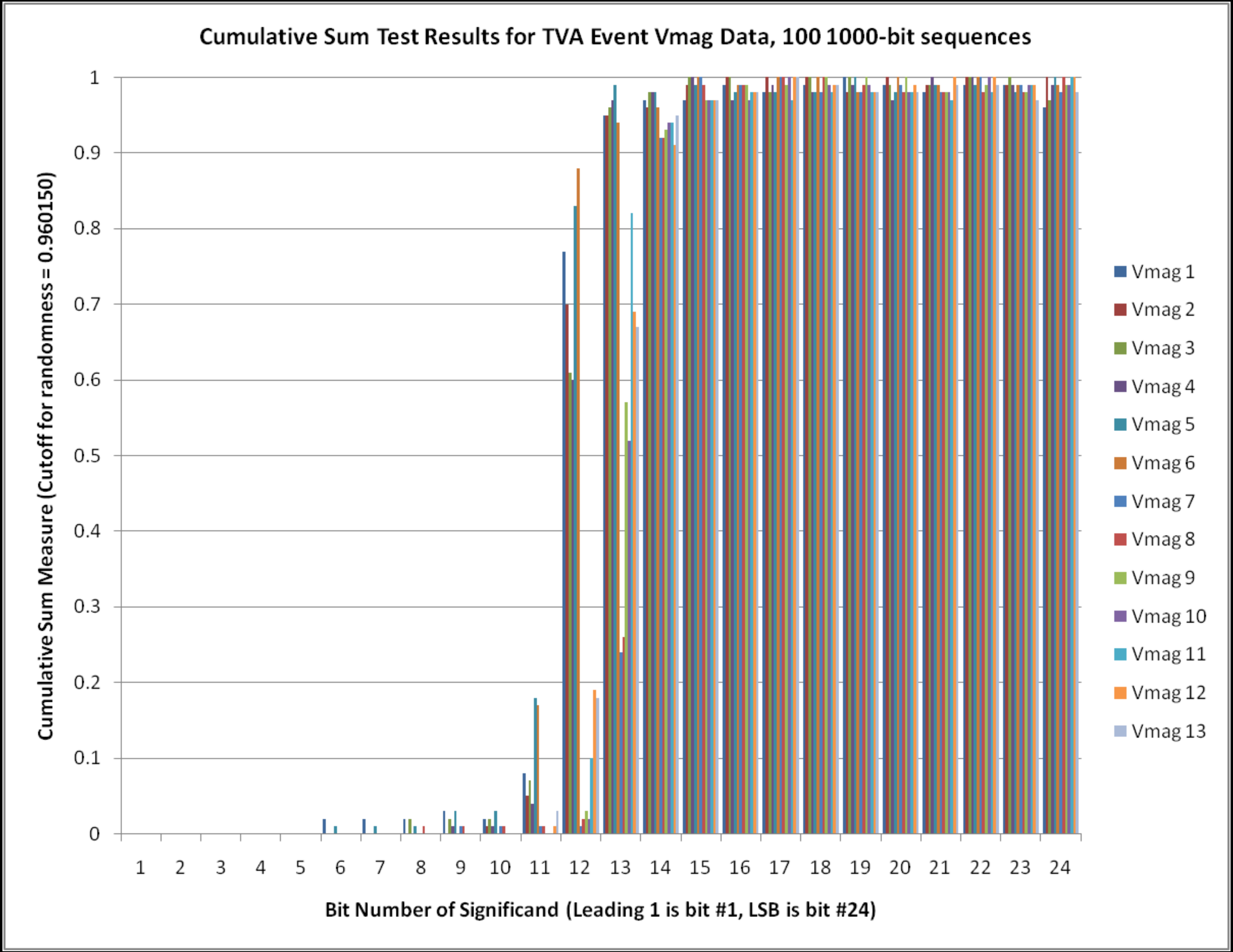
Results of Randomness Tests

Phase Angles



Results of Randomness Tests

Voltage Magnitudes



Effect of Masking Noise Bits

Phase Angles

COMPRESSION RATIOS OF VOLTAGE ANGLES FOR THE EVENT DATA SET AFTER MASKING THE 4 LEAST SIGNIFICANT BITS

	Deflate	Bzip2	LZMA
Unprocessed	2.02	1.75	3.58
SD	2.09	2.10	3.04
SRE	5.02	6.05	6.92
DE	4.49	5.71	6.38

COMPRESSION RATIOS OF VOLTAGE ANGLES FOR THE EVENT DATA SET AFTER MASKING THE 8 LEAST SIGNIFICANT BITS

	Deflate	Bzip2	LZMA
Unprocessed	3.84	4.72	11.00
SD	3.87	4.37	6.13
SRE	12.45	18.14	17.36
DE	13.38	20.81	18.82

Effect of Masking Noise Bits

Voltage Magnitudes

COMPRESSION RATIOS OF VOLTAGE MAGNITUDES FOR THE EVENT DATA SET AFTER MASKING THE 4 LEAST SIGNIFICANT BITS

	Deflate	Bzip2	LZMA
Unprocessed	2.49	3.41	3.16
SD	2.02	2.91	2.82
SRE	2.58	3.41	3.38
DE	2.54	3.39	3.34

COMPRESSION RATIOS OF VOLTAGE MAGNITUDES FOR THE EVENT DATA SET AFTER MASKING THE 8 LEAST SIGNIFICANT BITS

	Deflate	Bzip2	LZMA
Unprocessed	4.67	6.40	5.93
SD	3.63	5.17	4.93
SRE	4.99	6.21	6.35
DE	4.95	6.15	6.31

Best Compression Ratios for Event Scenario Data

Quantity	Without Pre-Compression	With Pre-Compression	With Pre-Compression and 4-bit Noise Removal	With Pre-Compression and 8-bit Noise Removal
Voltage Mags	2.21	2.50	3.39	6.31
Phase Angles	2.47	3.94	6.92	20.81

Future Work

- Testing with more data sets
- Improved selection of reference signals
- Integration with commercial PDC

Contact Information

Zeb Tate
zeb.tate@utoronto.ca

Ray Klump
klump@illinois.edu

Acknowledgements

Funding

U.S. Department of Energy
Natural Sciences and Engineering Research Council of Canada

Sample data

Tennessee Valley Authority