New Horizons in Time-Domain Astronomy
Proceedings IAU Symposium No. 285, 2011
R.E.M. Griffin, R.J. Hanish & R. Seaman, eds.

doi:10.1017/S1743921312000993

Fast Transient Detection as a Prototypical "Big Data" Problem

Dayton L. Jones, Kiri Wagstaff, David Thompson, Larry D'Addario, Robert Navarro, Chris Mattmann, Walid Majid, Umaa Rebbapragada, Joseph Lazio, and Robert Preston

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
email: dayton.jones@jpl.nasa.gov

Abstract. The detection of fast (< 1 second) transient signals requires a challenging balance between the need to examine vast quantities of high time-resolution data and the impracticality of storing all the data for later analysis. This is the epitome of a "big data" issue—far more data will be produced by next generation-astronomy facilities than can be analyzed, distributed, or archived using traditional methods. JPL is developing technologies to deal with "big data" problems from initial data generation through real-time data triage algorithms to large-scale data archiving and mining. Although most current work is focused on the needs of large radio arrays, the technologies involved are widely applicable in other areas.

Keywords. methods: data analysis, astronomical data bases: miscellaneous

Fast transient signals are predicted from a number of astronomical objects and processes, and are expected to be associated with extreme physical conditions. Several technologies are critical for large-scale dedicated or commensal searches for fast transients.

Power consumption by digital electronics is likely to be a major operating cost for future radio facilities in searches for fast transient signals, particularly for wide-band digital beam-forming and correlation. We must be able to afford the cost of generating high-rate data before the analysis of high-rate data becomes relevant. JPL has investigated ASIC architectures that can reduce power consumption in cross-correlation dramatically for arrays with large numbers of antennæ (D'Addario 2011). This has obvious application to the SKA and other large instruments.

A second area of development at JPL is adaptive algorithms to perform real-time processing in high-volume data flows, when storage of data for later analysis is not an option (data triage). Machine-learning algorithms that provide transient triggers and automated processing of buffered high-time-resolution data are now being tested by the VLBA V-FASTR project (Brisken *et al.* 2011; Thompson *et al.* 2011; Wayth *et al.* 2011). Similar techniques can be used to detect time-varying interference (self-induced or external), anomalies in array performance-monitoring data, and some aspects of time-varying calibration. Data-adaptive algorithms could also control front-end data collection based on the statistical properties of event detections, allowing optimized sampling.

We are also working on highly scalable data archive frameworks for astronomy. Archive design will determine how much real-time data triage will be needed, and how much analysis can be done off-line. JPL has developed a Process Control System based on Object Oriented Data Technology (OODT), components of which are being evaluated for use at several observatories. OODT is open source software, and is the first NASA software to become a top-level project at the Apache Software Foundation.

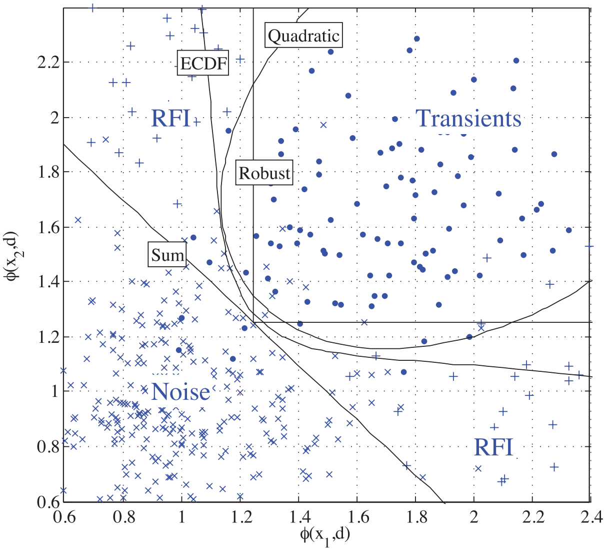

Figure 1. Ability of different multi-station transient detection methods to separate noise, RFI, and true transients. The axes show signal strength from each of two separate VLBA antennæ.

The JPL-funded technology development described here is part of an evolving end-to-end approach to "big data" problems. Fast transient searches provide an excellent test case for that work, in addition to the possibility of exciting near-term scientific results.

Low-Power Digital Signal Processing. JPL has completed a systematic study of the effect of architecture choices on the power consumption of large cross-correlators used for all aperture synthesis radio arrays. Architectures that minimize memory use and I/O data rates can significantly reduce power consumption, sometimes making the difference between a particular facility or instrument being feasible to operate or not.

Real-Time Machine-Learning Algorithms. As an example of the value of machine-learning algorithms for real-time data triage, Fig. 1 shows an application in fast (< 1 s) transient signal detection. The quadratic discriminant algorithm provides greater robustness against false alarms, and greater sensitivity.

Scalable Data Archives and Data Mining. The JPL Process Control System is a set of reusable components from the Object Oriented Data Technology (OODT) framework developed by Dan Crichton. It is scalable, hardware- and database-independent, and is interoperable, with a flexible plug-in capabilities for user data processing tools and algorithms.

Acknowledgements

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the U.S. National Aeronautics and Space Administration.

References

D'Addario, L. 2011 *Square Kilometre Array Memo*, 133, `http://www.skatelescope.org/`
Brisken, W., *et al.* 2011, *Proc. National Radio Science Meeting*, Boulder, CO, 2011 January 6
Thompson, D. R., *et al.* 2011, *ApJ*, 735, 98
Wayth, R. B., *et al.* 2011, *ApJ*, 735, 97

https://doi.org/10.1017/S1743921312000993 Published online by Cambridge University Press