

# Fast Algorithms for Time Series: algorithms and applications to Finance, Physics, Music, Biology, and other Suspects \*

Dennis Shasha

Courant Institute  
New York University  
New York, NY 10012  
shasha@cs.nyu.edu  
U.S.A.

Yunyue Zhu

Courant Institute  
New York University  
New York, NY 10012  
yunyue@cs.nyu.edu  
U.S.A.

Alberto Lerner

Courant Institute  
New York University  
New York, NY 10012  
lerner@cs.nyu.edu  
U.S.A.

## Abstract

Financial time series streams are watched closely by millions of traders. What exactly do they look for and what do they do with that information? Physicists study the time series emerging from their sensors. The same question holds for them. Musicians produce time series. Consumers may want to compare them. This tutorial presents techniques for correlating pairs of time series using Discrete Fourier

---

\*Work supported in part by U.S. NSF grants IIS-9988636 and N2010-0115586.

and Discrete Wavelet transforms, sketches, and Dynamic Time Warping. The tutorial goes on to discuss burst detection over many window sizes at once. It explains these techniques in an intuitive way through applications that use them: pairs trading for stocks, gamma ray detection, in physics, and query by humming in music. Finally it reviews the work in database language support for ordered data and presents some examples of a simple model.

This tutorial is largely based on the upcoming book *High Performance Discovery in Time Series: techniques and case studies*, Springer-Verlag 2004. A variant of this tutorial was presented at the Brazilian Database Symposium in Manaus, Amazonas in October. Also, parts were presented at Cornell in September. You can find many of the power point slides for this tutorial at <http://cs.nyu.edu/cs/faculty/shasha/papers/burstcorr.ppt> and <http://cs.nyu.edu/cs/faculty/shasha/papers/aquery.ppt>.

## 1 Outline

### 1. Finance and Correlation

- Financial markets, number of securities, volume of trades, trader behavior, introduction to pairs trading.
- Types of correlation: whole sequence, landmark, and sliding window.
- The practical importance of sliding window correlation.
- A historical survey of techniques for high speed correlation using Fourier or Wavelet transforms when the power spectrum is localized to a few coefficients.
- Sketch based computation when the power spectrum is poorly behaved.
- Synchronous vs. lagged correlation.
- Empirical results
- Applications to date.
- Open problems.

### 2. Astrophysics and Burst Detection

- Motivating application: Gamma ray detection based on particle scatter. Data size is large.
- Importance of bursts.

- Burst as a function of window size. However, window size is unknown *a priori*.
- Algorithmic techniques for detecting bursts over many window sizes simultaneously in one pass.
- Applicability to spatial detection.
- Open Problem: Extending to many time series, e.g. for market basket analysis.

### 3. Music and Approximate Time Series Recognition

- Music recognition of one recording with another.
- Digital Rights Management applications.
- Human music recognition: query by humming.
- Nature of the noise: no absolute pitch, poor tracking of intervals, variable tempo.
- A historical survey of note segmentation approaches. String edit distance.
- A historical survey of dynamic time warping techniques.
- Envelope optimizations.
- Rhythm filters.
- Open problem: scaling to millions.

### 4. Data Management for Order

- Applications of order: finance, networks, biology.
- The case for order in a query language vs. sequences as abstract data types
- Language desiderata: declarative, simple, dynamic.
- A historical survey of SQL extensions for order
- Examples of queries including burst detection and correlation discovery
- Optimization opportunities for static data.
- Alternative architectures for streaming data: on demand operators (viz. the data stream languages) vs. a batch and query paradigm.
- Open Problems.

## 2 Summary

The tutorial is aimed at researchers in streams, data mining, and scientific computing. Its applications should interest anyone who works with scientists or financial “quants.” The emphasis will be on recent results and open problems, so this is a ripe area for further advance.

1.5 hours is sufficient times to cover the major ideas and to discuss the open problems.

Dennis Shasha will give the entire presentation.