

IBM Research



TRISTAN: Real-Time Analytics on Massive Time Series

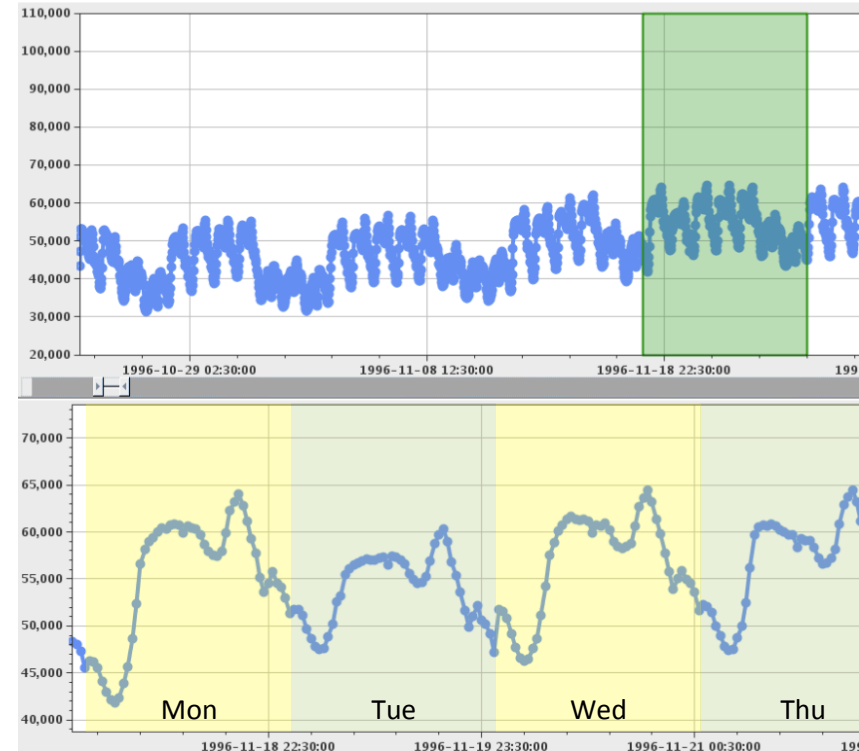
Philippe Cudré-Mauroux

[eXascale Infolab](#), University of Fribourg
Switzerland

Joint work w/ [IBM Research Smarter Cities Tech. Lab](#)

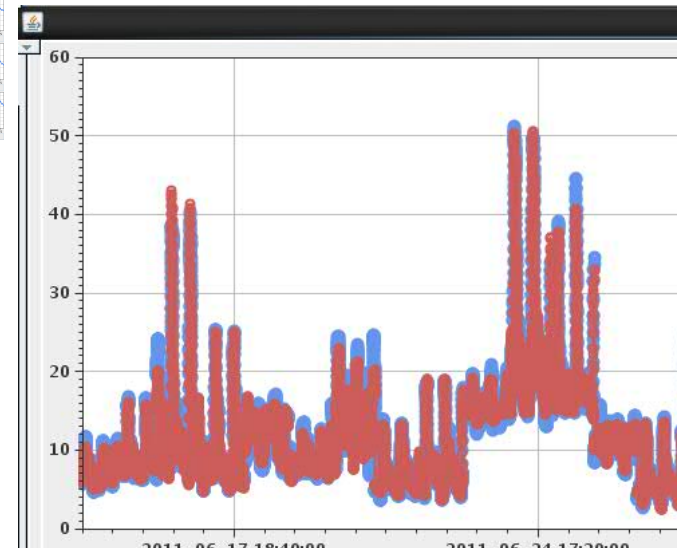
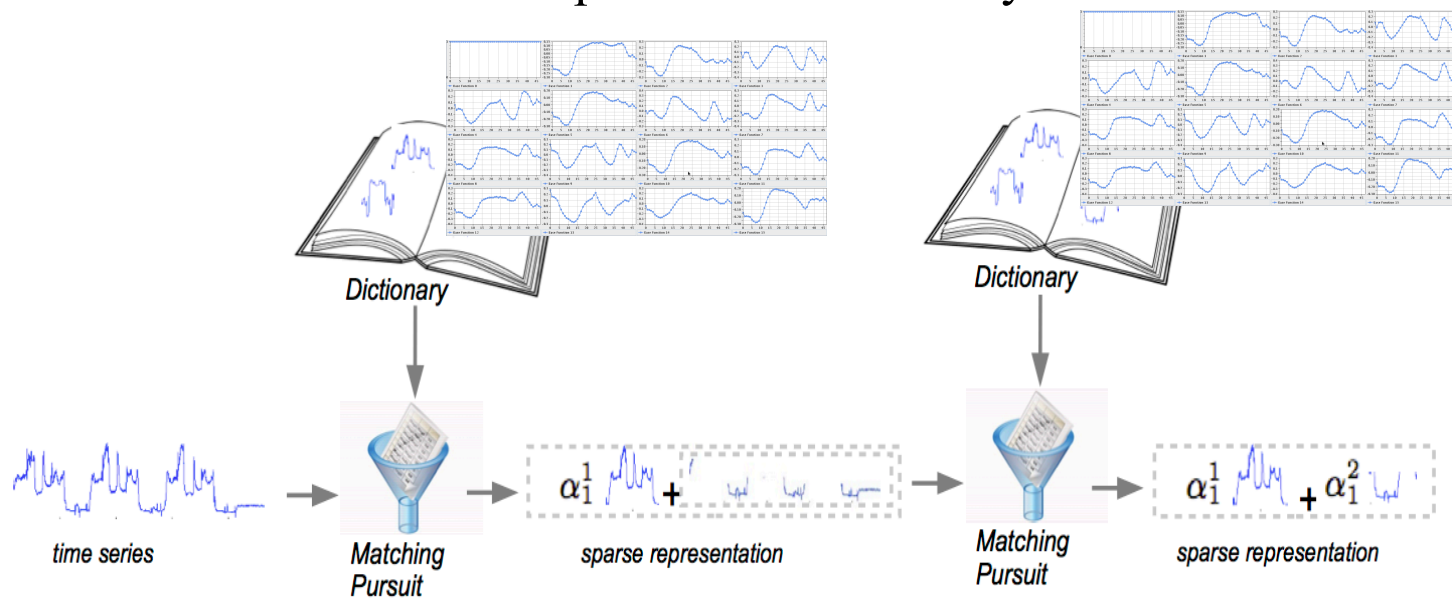
Smarter Cities Data

- *Tricklets!*
 - Noisy, misaligned, redundant **time series** from sensors
- Massive amounts of data
 - Compression?
 - Smoothing, re-sampling, interpolation?
- ... and: often exhibit **recurring** patterns

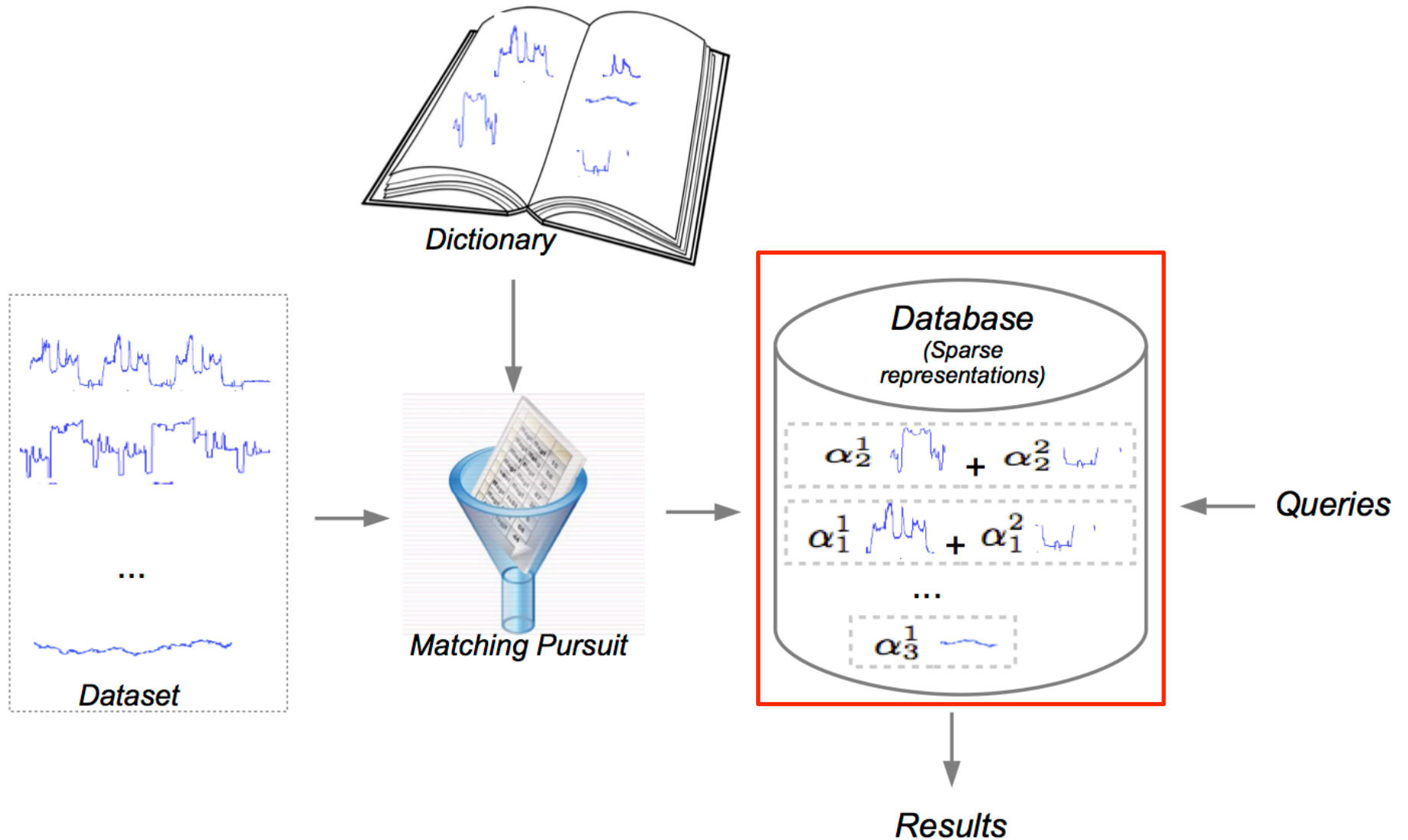


TRISTAN

- Long-term archival & **interactive querying** of massive amounts of tricklets
 - Encodes time-series using **Matching Pursuit**
 - Universal (choose your dictionary *atoms*)
 - Semantic (?) encoding
 - Efficient, effective
 - Exponential error decay

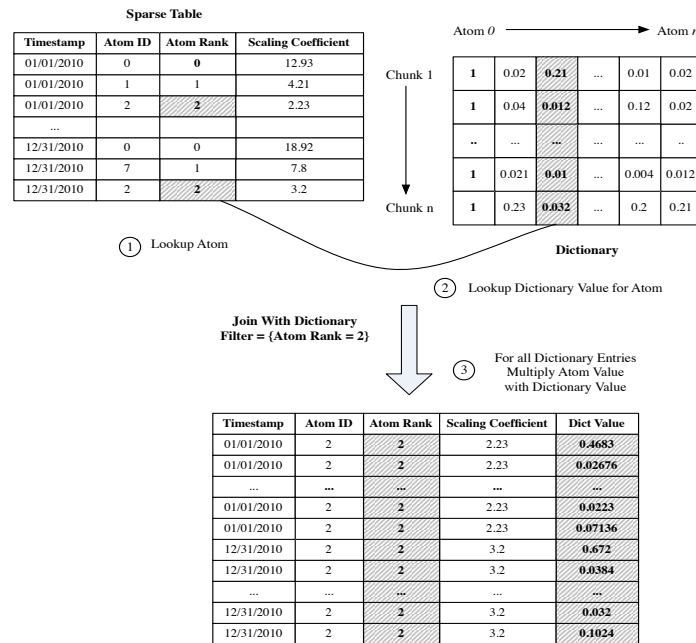


Sparse Data Representation



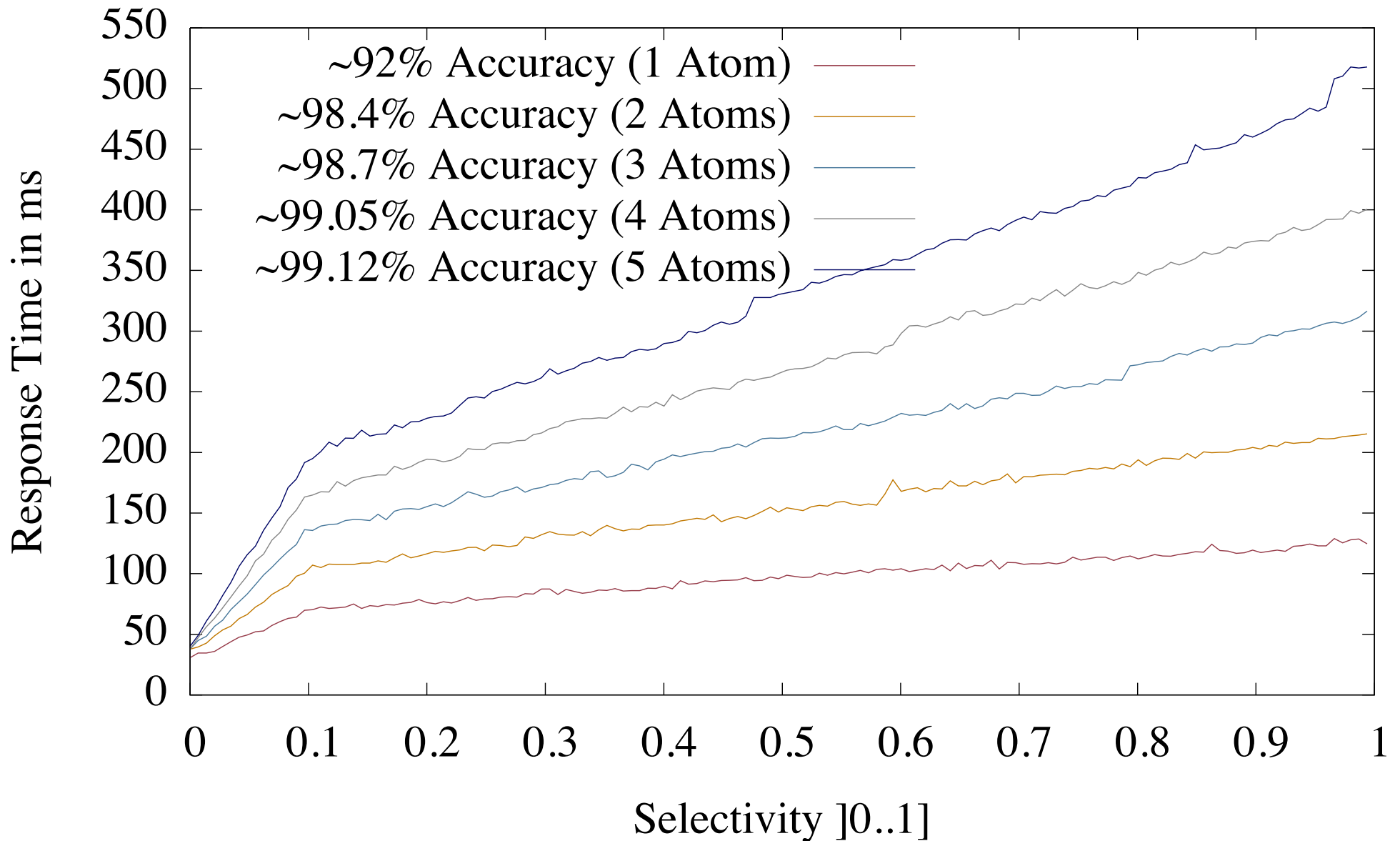
Data Processing

- (Optimized) signal reconstruction



- Computation **on compressed data!**
 - $\sum_{l=1}^L s_l(t) = \sum_{l=1}^L \sum_{j=1}^K w_{j,l} \varphi_j(t) = \sum_{j=1}^K \varphi_j(t) \sum_{l=1}^L w_{j,l}$
 - Aggregates, correlations, matching etc.
- **1:100** compression ration, **250x** speedup compared to state-of-the-art system

Latency VS Accuracy Trade-off



TRISTAN DBMS

- Now available in **Hyrise** (open-source, hybrid DBMS)
 - Being tested in **IBM DB2**
- Thanks to all partners
 - Alice Marascu, Pascal Pompey, Eric Bouillet, Olivier Verscheure
IBM Research—Ireland Smarter Cities Tech. Centre
 - Michael Wurst
IBM Business Analytics Boeblingen—Germany
 - Martin Grund, Philippe Cudré-Mauroux
eXascale Infolab, University of Fribourg—Switzerland



= > phil@exascale.info