

Detector/Data Characterization Robot

Towards Data Mining

Soumya D. Mohanty

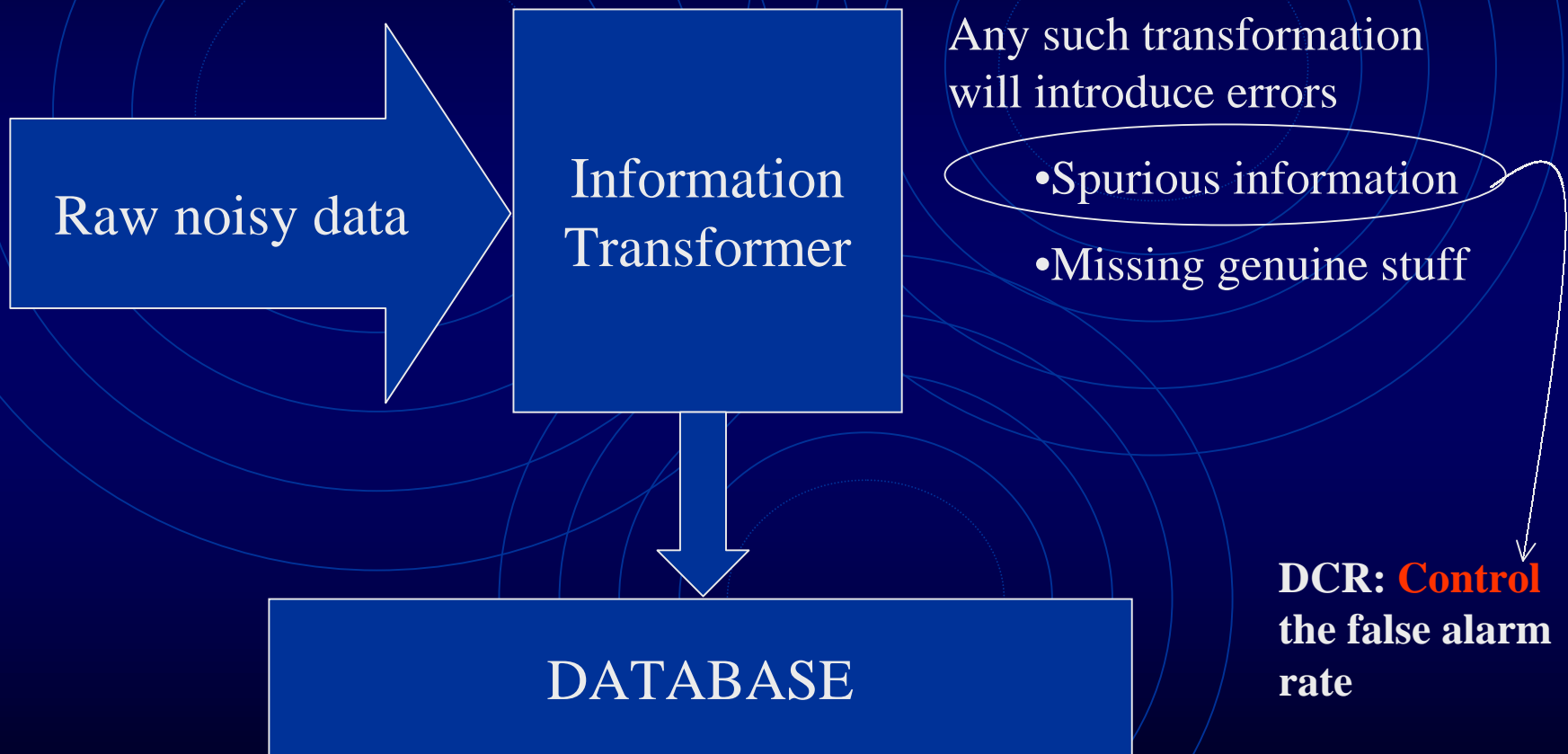
Max Planck Institut für Gravitationsphysik

LIGO-G020036-00-Z

Using a database: Data Mining & Data Exploration

- Different but complementary approaches.
- **Data exploration:**
 - I want to see the time series corresponding to a bunch of triggers that I selected from a database. (Then do more analysis on this selected data.)
 - Typically, Follow up data is short, Quick look environment needed, no specific queries
- **Data Mining:**
 - Can the transients seen over a month be classified into groups? What was the rate of transients in each group as a function of time (Maybe some types occur in the day, some occur in the night). (Then use this information to quantify the quality of long data stretches).
 - Purely database based; Re-analysis of raw data may be impractical

Raw Data to Database



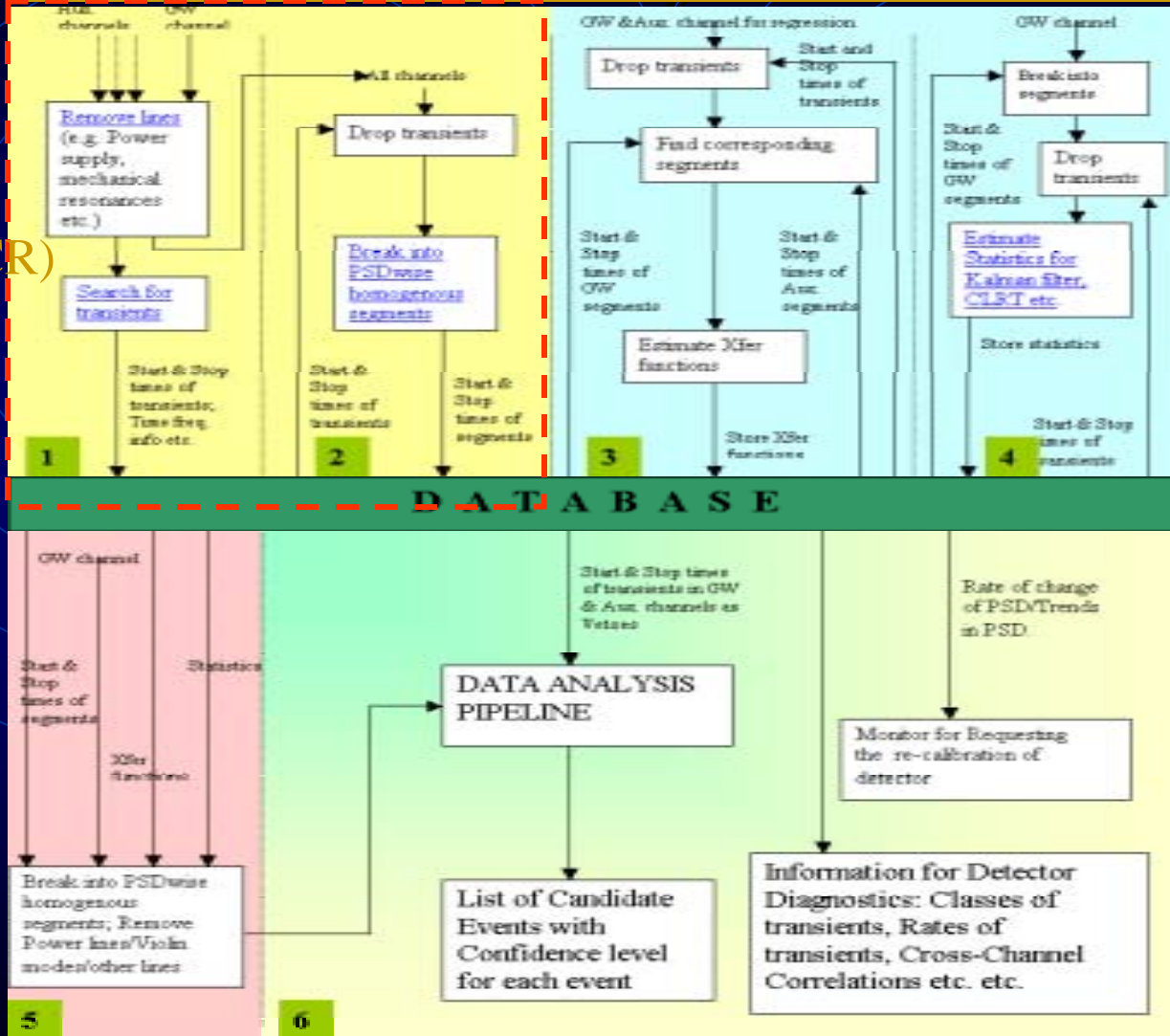
Control on False Alarm Rate

- Important for Data mining
 - Statistical analysis done on database itself since reanalysis of long stretch of data expensive
 - Need to put error bars
- Not so important for Data exploration
 - Looking for information about specific events
 - Each explorer will work with his/her own short data stretch

Initial Design of DCR

Soumya Mohanty, Soma Mukherjee, CQG, 2002.

Restricted
DCR (rDCR)

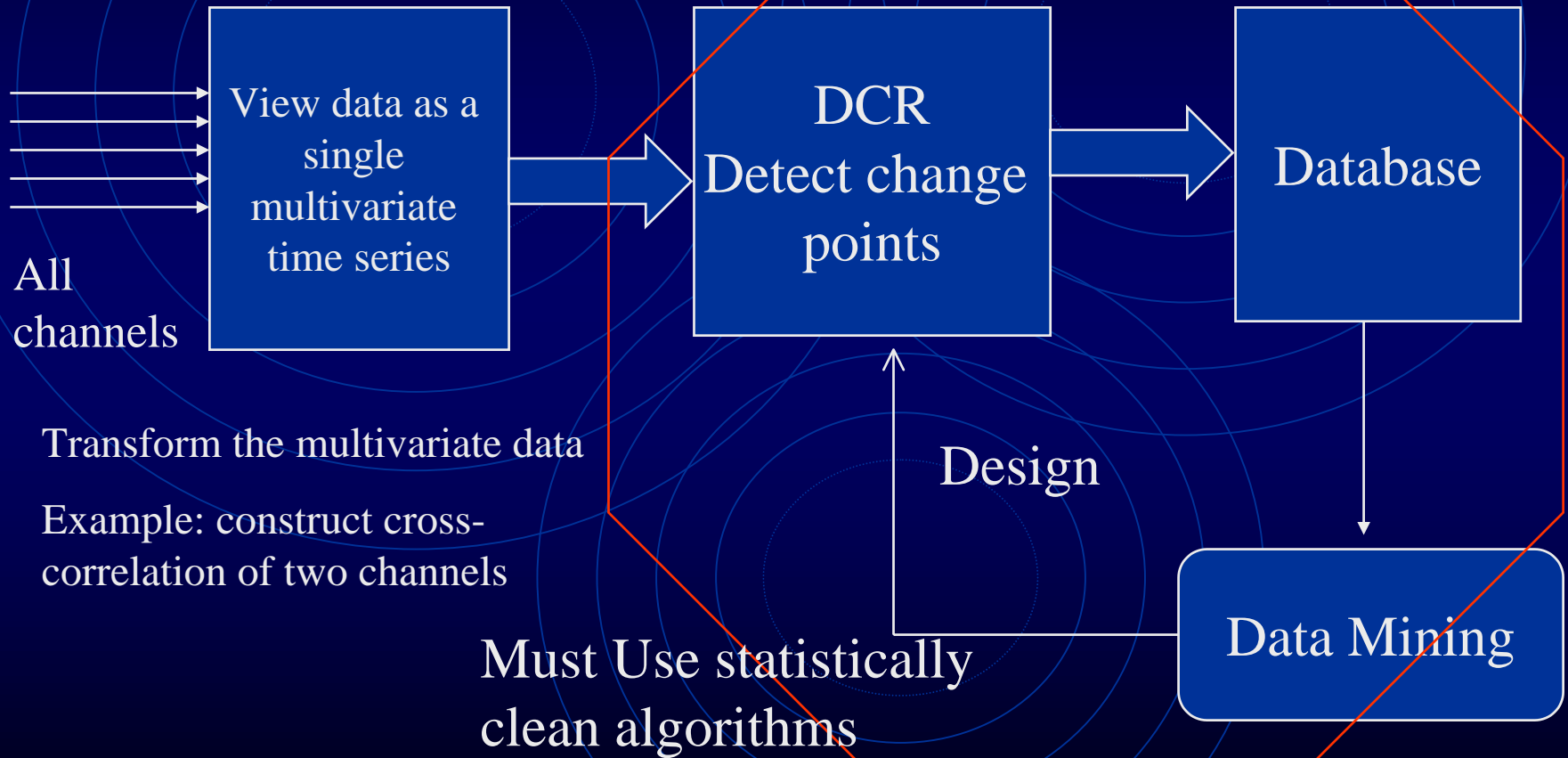


Implementation

- C++ code for restricted DCR algorithm ready
- It will run in GEO using the GEO++ library
- Data mining issues under active investigation
 - Earlier talks on Automated classification, non-stationarity, line removal transient test
 - Data mining will be done using Matlab's database toolbox

<http://www.aei.mpg.de/~mohanty/DCR/DCRindex.html>

Future



DCR on the Web