LIGO-G-1200500

Improving Gravitational Wave Searches by Using Multivariate Classifiers to Incorporate Auxiliary Channel Information

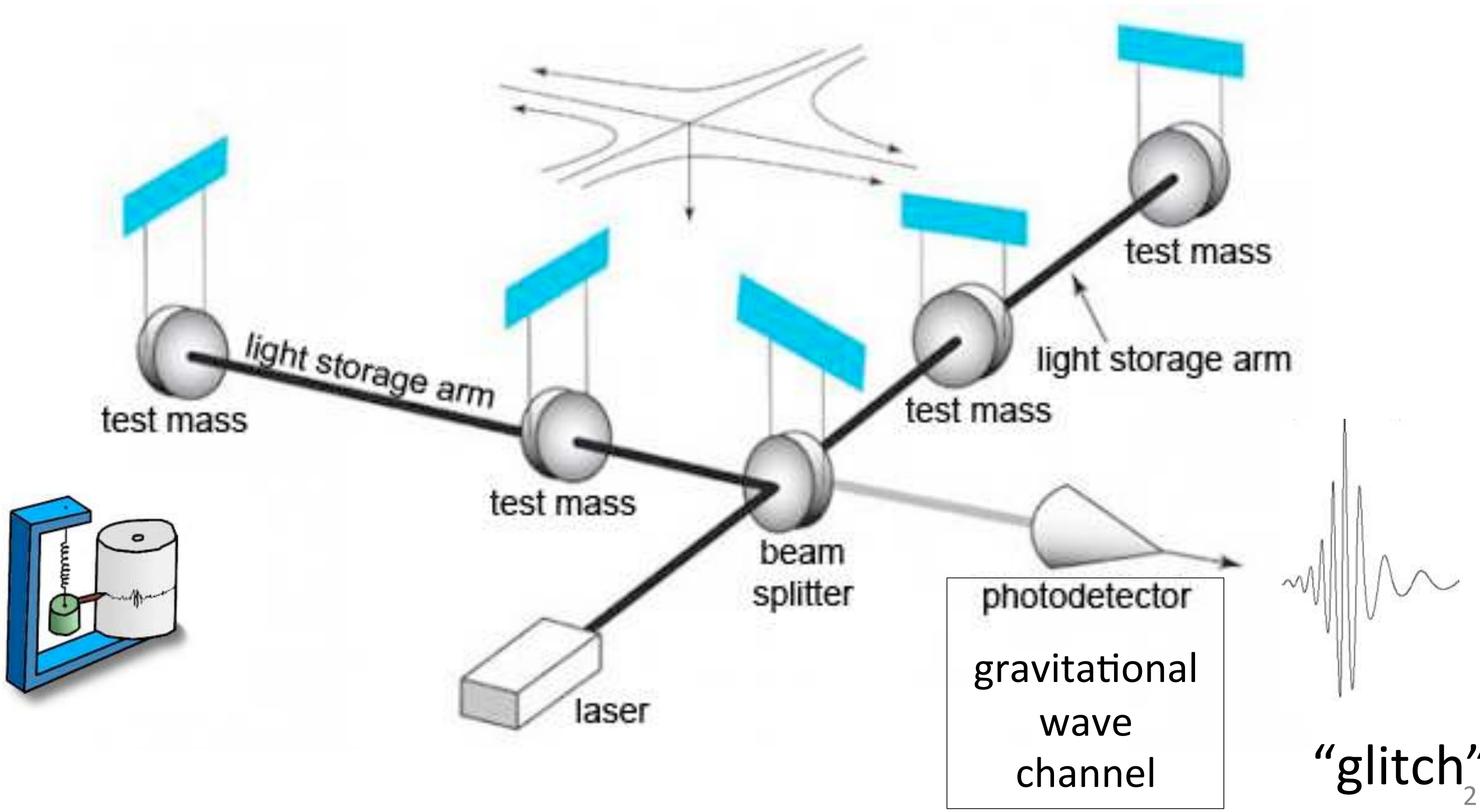
Kari Hodge

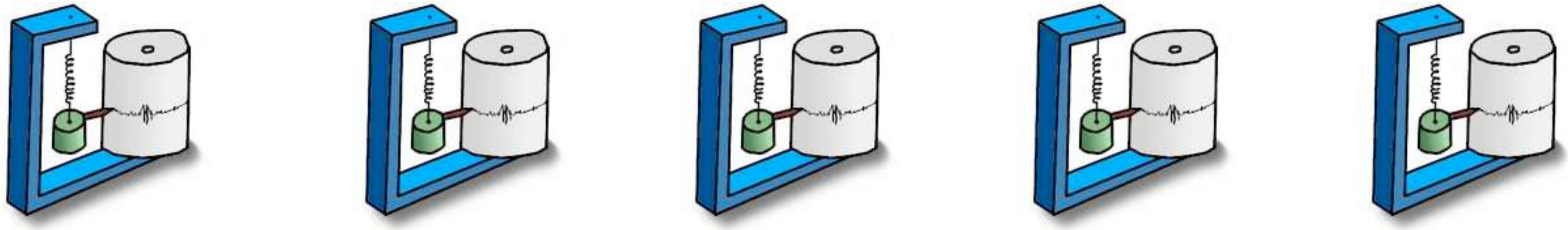
California Institute of Technology

Rahul Biswas[7], Lindy Blackburn[2], Junwei Cao[6], Reed Essick[1], Kari Hodge[4], Andrew Lundgren[8], Erik Katsavounidis[1], Young-Min Kim[5,3], Chan-Hwan Lee[5], John J. Oh[3], Sang Hoon Oh[3], Ye Tao[6], Ruslan Vaulin[1], Xiaoge Wang[6]

[1]Massachusetts Institute of Technology, [2]NASA Goddard Space Flight Center, [3]National Institute for Mathematical Sciences, [4]California Institute of Technology, [5]Pusan National University, [6]Tsinghua University, [7]University of Texas at Brownsville, [8]AEI Hanover

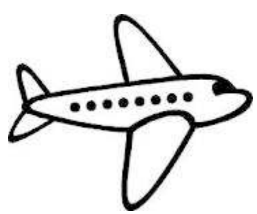
Laser Interferometer Gravitational-Wave Observatory (LIGO)



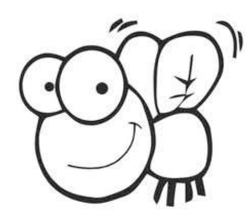


The detectors have hundreds of such **auxiliary channels** monitoring the instrument and its environment

In practice, a multitude of them witness "something" and the cause of the glitch is not clear



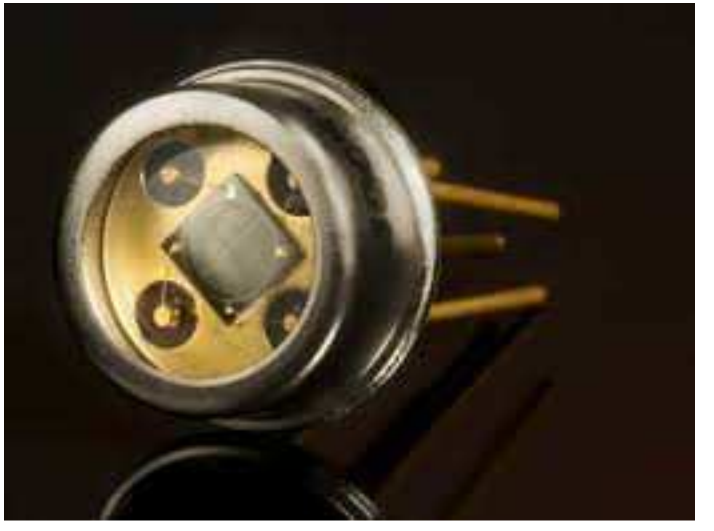
?



?

Auxiliary Channels

- Environmental:
  - seismometers
  - microphones
  - magnetometers
- Instrumental:
  - beam splitter motion
  - mirrors' angular alignment
  - common arm

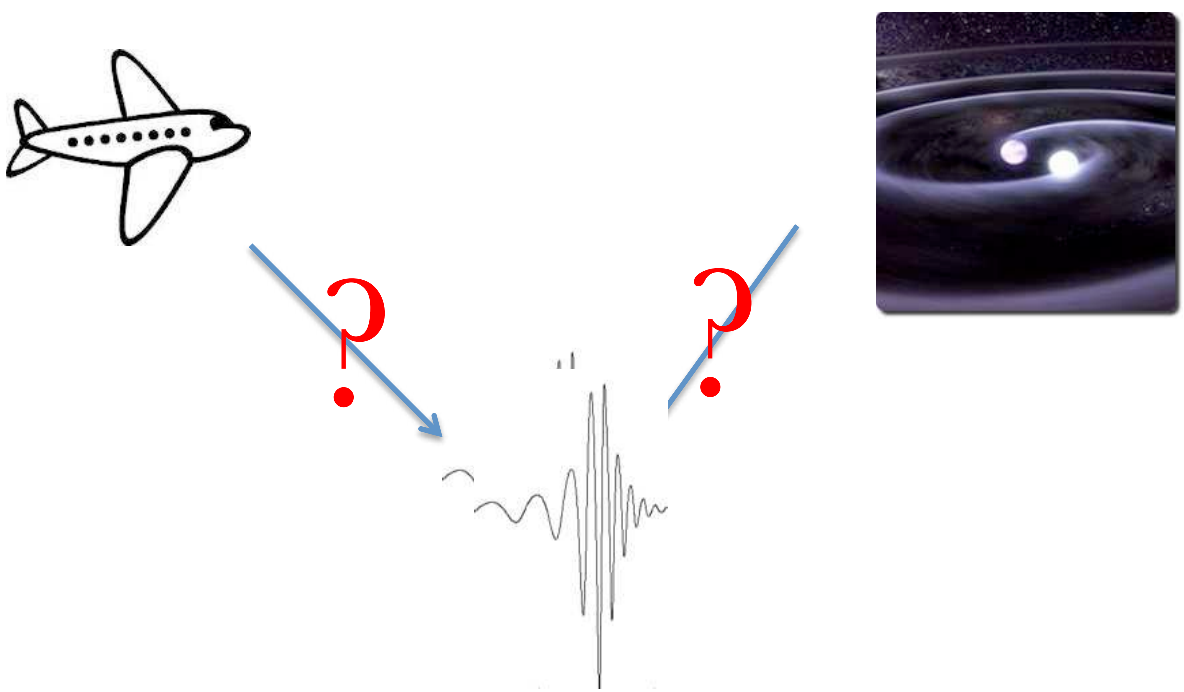


Problem:

Glitches increase false alarm rates, obscuring detections and raising upper limits

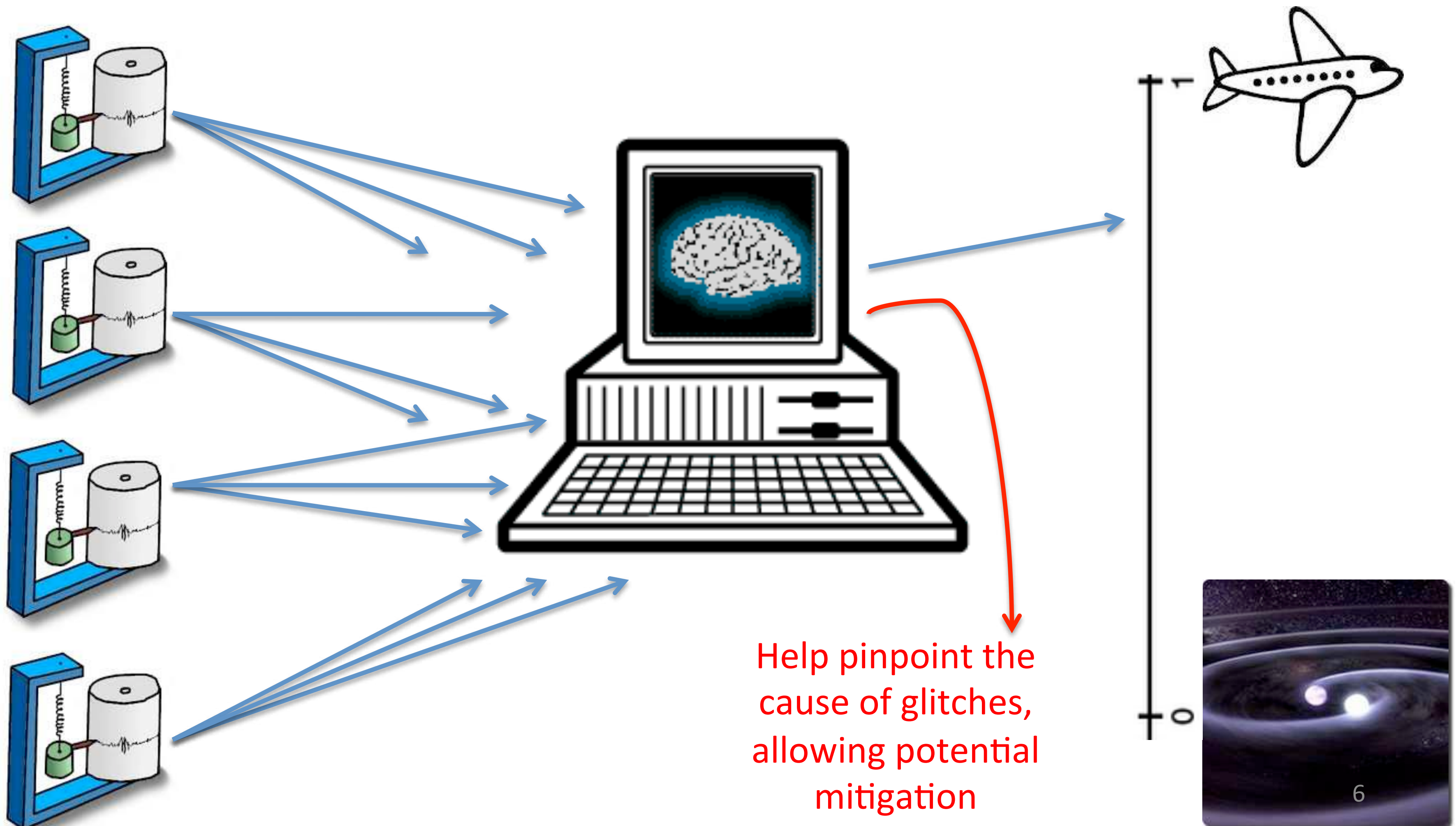
So we'd like to be able to either:

- veto these times prior to analysis
  (historical method, based on at most a few channels per veto)
- re-rank events based on the result of auxiliary channel analysis
  (difficult analytically because of the sizable number of channels)



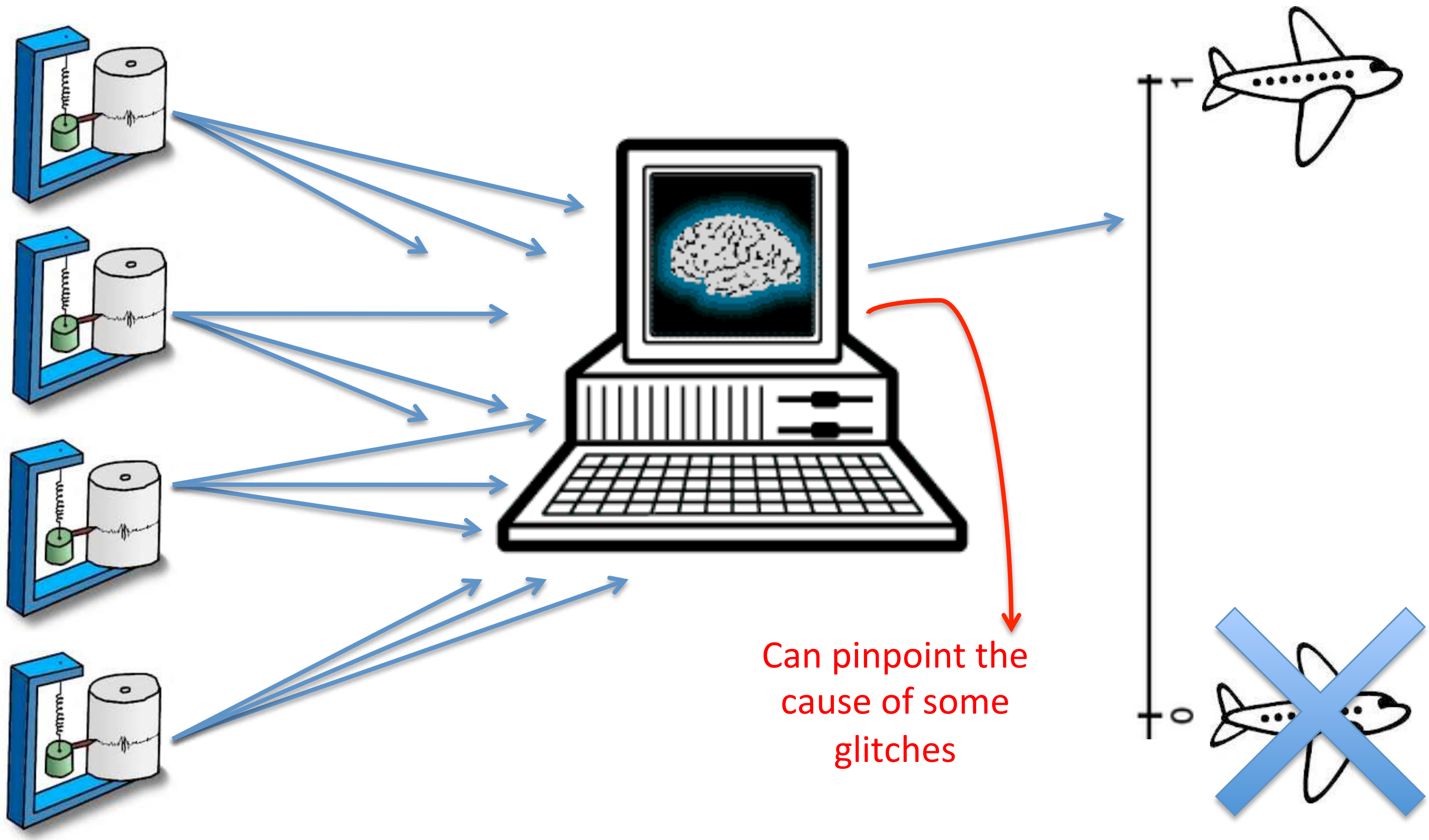
Ideal Solution:

Combine information from these auxiliary channels into a Multivariate Classifier (MVC) with existing search algorithms, let it determine how likely it was that what was seen in the GW channel was a glitch of environmental/instrumental origin as opposed to an actual gravitational wave



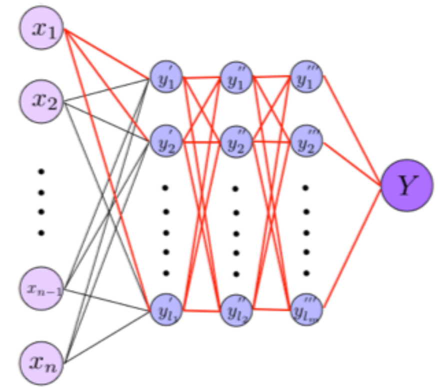
So Far:

Completed half of this – separating generic artifacts from “clean” times.
Future work will couple this to specific searches.

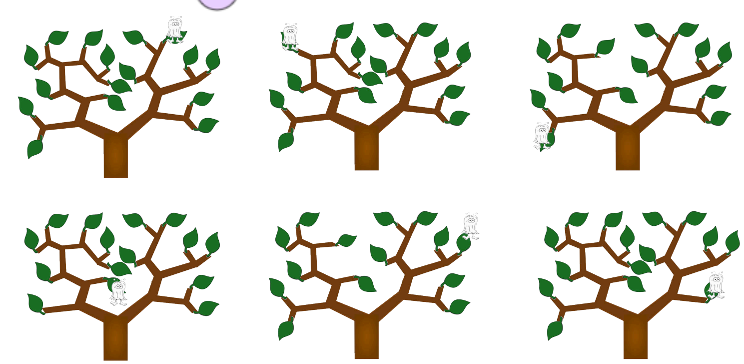


Multivariate Classifiers (MVCs)

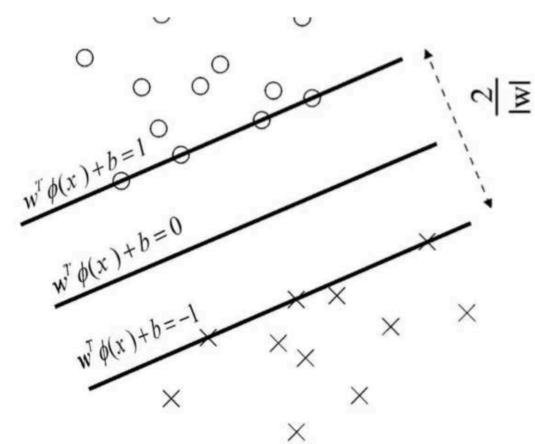
- Artificial Neural Networks (ANN)
  - multilayer network using back-propagation algorithm iRPROP from the FANN library



- Random Forests of Bagged Decision Trees ("Forest")
  - StatPatternRecognition library



- Support Vector Machines (SVM)
  - LibSVM package is used to transform non-linearly separable parameters into a higher-dimensional space where they might be linearly separable



Input to the MVCs

- classifiers are constructed with a training set – events with known class
- IMPORTANT: the training set must be unbiased
  - can't contain any information from the gravitational wave channel
  - glitch set shouldn't contain gravitational waves

Input to the MVCs

- Training set:
  - Class 1: the glitches (what we are looking for)
    - identified by a Kleine-Welle trigger in the gravitational-wave channel above a nominal threshold
  - Class 0: “clean” times
    - at least 100 ms away from a Class 1 sample

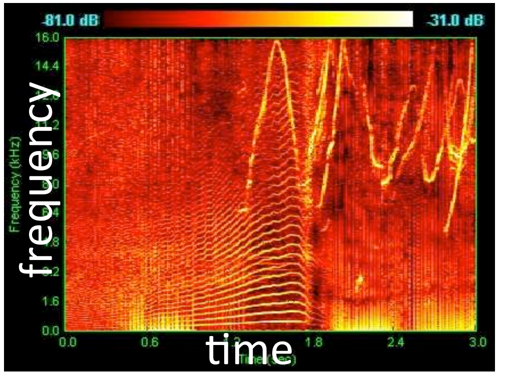


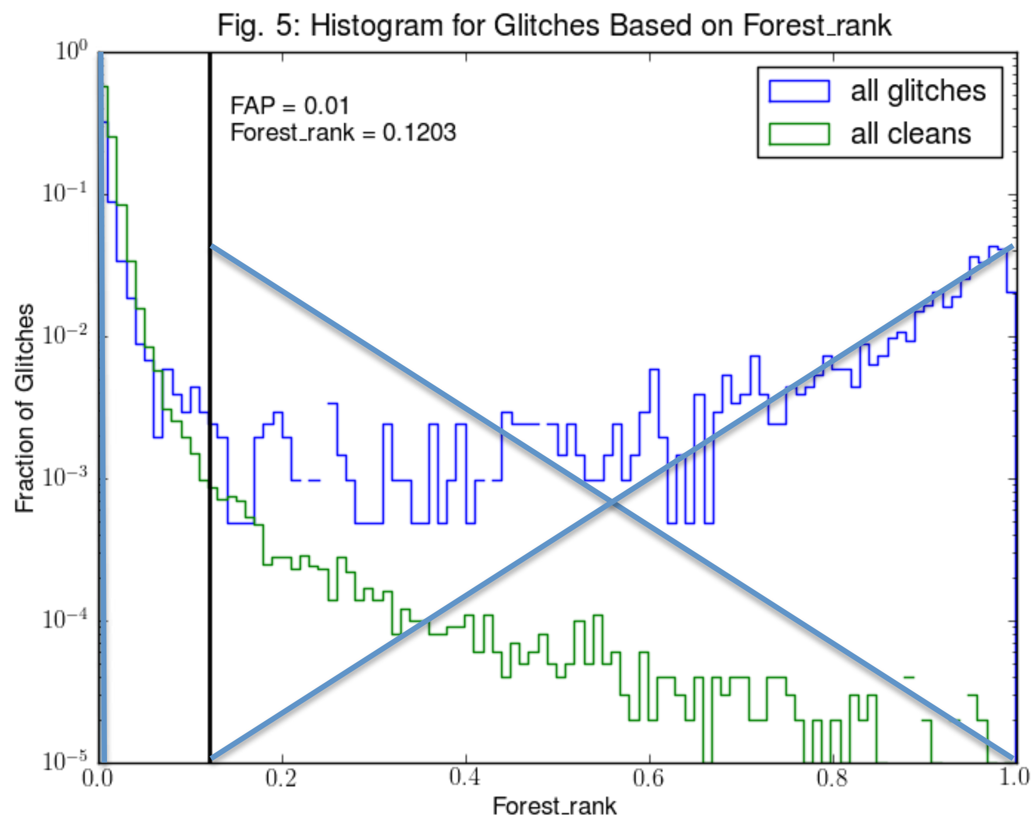
figure: generic spectrogram

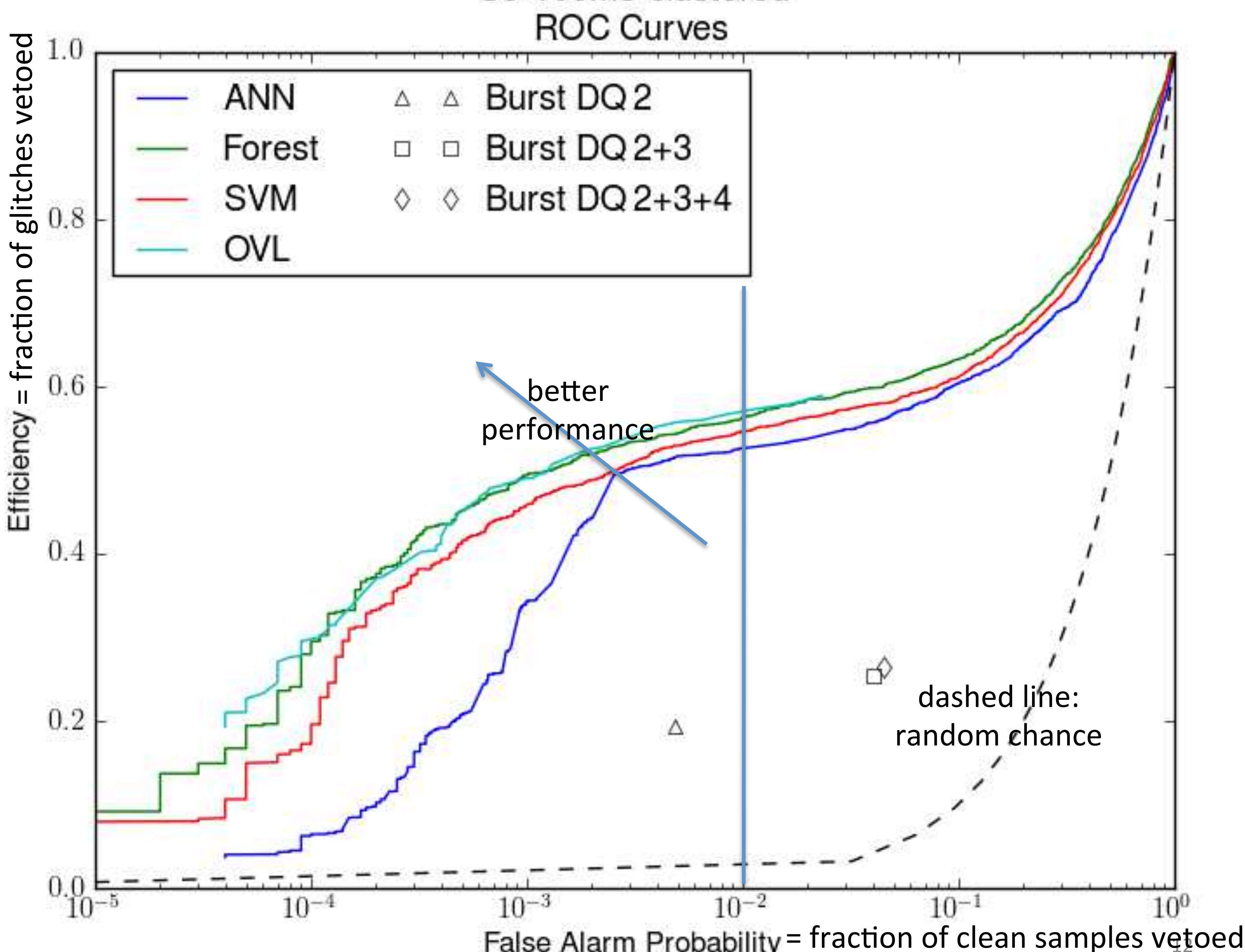
Kleine-Welle captures excess power in the time-frequency plane via wavelet analysis

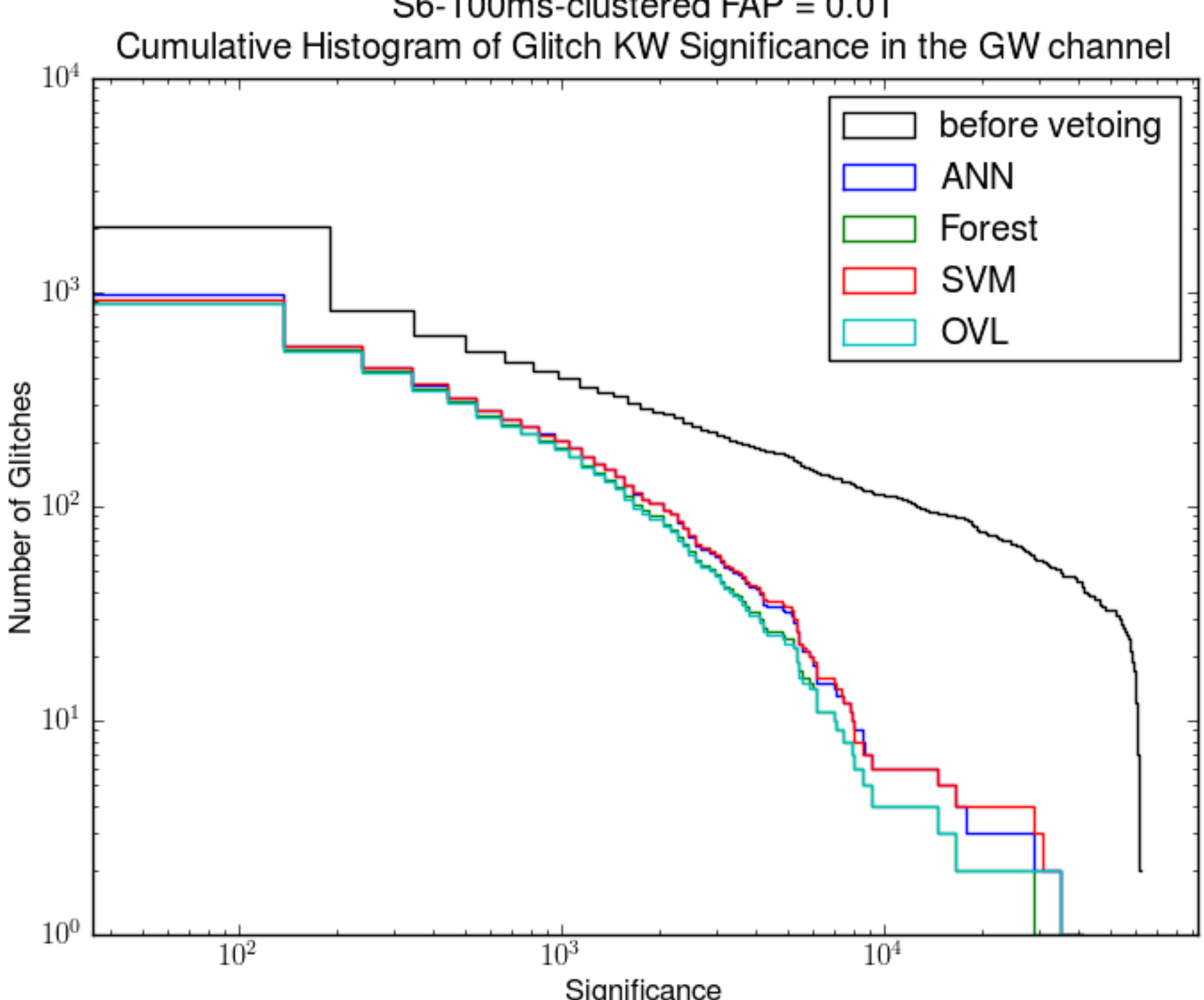
Each event is parameterized by 5 numbers per auxiliary channel. These numbers are extracted from the auxiliary channels with the same Kleine-Welle wavelet analysis.

Output of the MVCs

- Once the MVC is trained, you can feed it an event and it will spit out a number between 0 and 1, where 1 indicates it is more glitch-like.
- By picking a threshold on this continuous rank, one can veto any GPS time where the rank exceeds this threshold

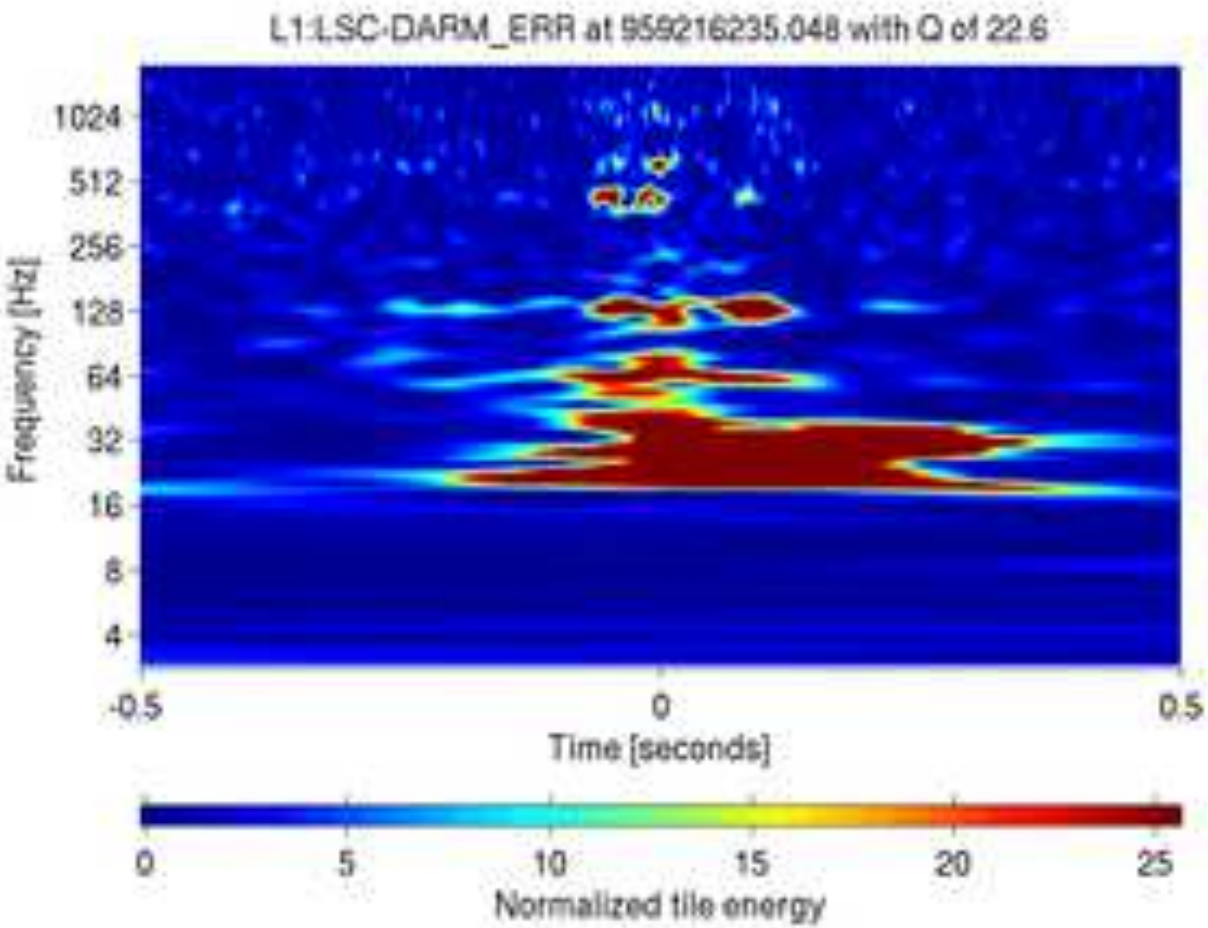



ROC Curves
Efficiency = fraction of glitches vetoed
False Alarm Probability = fraction of clean samples vetoed
ANN
Forest
SVM
OVL
Burst DQ 2
Burst DQ 2+3
Burst DQ 2+3+4
better performance
dashed line: random chance
0.0
0.2
0.4
0.6
0.8
1.0
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}


S6-100ms-clustered FAP = 0.01
Cumulative Histogram of Glitch KW Significance in the GW channel
before vetoing
ANN
Forest
SVM
OVL
Number of Glitches
Significance
10^0
10^1
10^2
10^3
10^4
10^2
10^3
10^4

Detector Characterization

- We can identify the channels witness to specific classes of glitches and feed this information back to instrumentalists



Train only on events that have this character in GW channel

The trained forest has the most splits on a specific set of channels

This "jitter" is caused by a non-linear coupling of slow motion of the optics with noise at higher frequencies in the output mode cleaner

Conclusions

- The key to distinguishing a gravitational wave from an instrumental or environmental glitch is in the auxiliary channels
- Once trained, the MVCs can classify new events rather quickly
- The different classifiers find mostly the same glitches
- MVCs are not just a black box!

Future Work

- expansion of input variables will lead to better performance
  - include other time-frequency analyses (e.g. Omega) to capture more information
  - direct channel readout for slow channels
- classification of different glitch types
- diagnosis of the physical causes of more classes of glitches (upconversion noise)
- thorough check of veto safety (hardware injections)
- interfacing with CBC and Burst searches
  - define a detection statistic for a search which will include MVC output