The tfclusters package

Julien Sylvestre
LIGO-MIT

LIGO-G010344-00-D

LSC Meeting - LIGO Hanford Observatory
13 August 2001

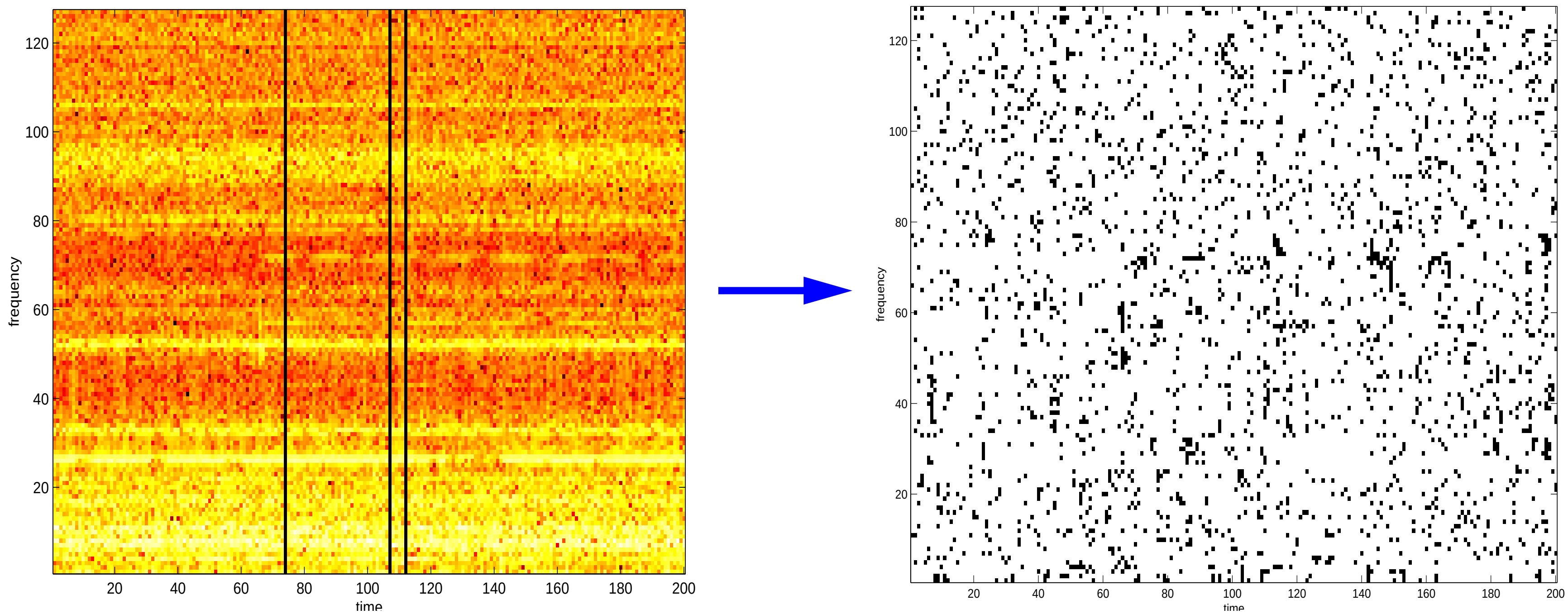
tfclusters

- Efficient algorithm for detecting unmodeled bursts in Gaussian noise
- Based on time-frequency power thresholding and clustering analysis. Uses short-time Fourier decomposition.
- lal implementation completed and tested on E2 data
- lalwrapper implementation completed (tested only with stand-alone wrapperAPI on fake data)
- Will work best on white noise, but can handle lines and colored noise (but must be Gaussian)
- Parallel implementation in DMT for real-time triggers generation

LIGO

Noise model and first threshold

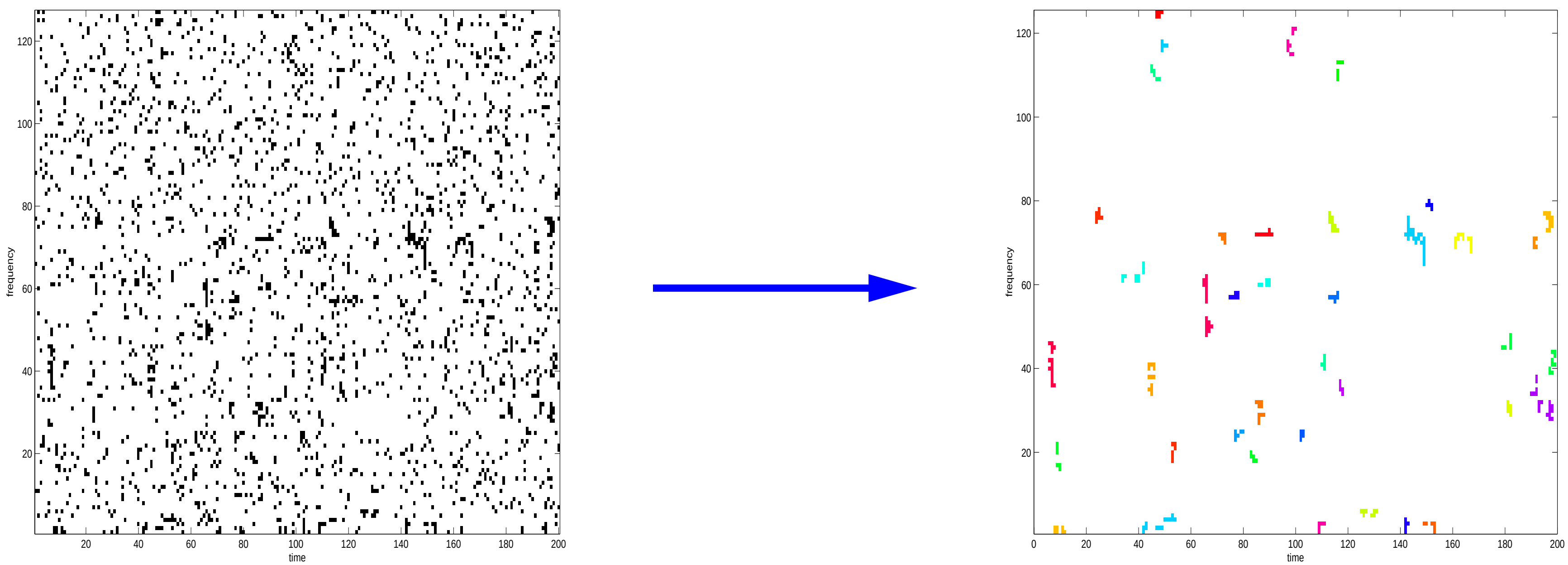
- Threshold spectrogram to get uniform black pixel probability: power threshold from exponential or non-central χ^2




```
void LALComputeSpectrogram(LALStatus*, Spectrogram*, TFPlaneParams*, REAL4TimeSeries*);
void LALTFCRiceThreshold(LALStatus *, REAL4 *, RiceThresholdParams *);
void LALGetClusters(LALStatus*, CList*, Spectrogram*, CListDir*);
```

Clustering Analysis
(second threshold)

- In white noise, large clusters are exponentially unlikely: threshold on cluster size (number of pixels)
- To increase efficiency, allow close pairs of small clusters




```
void LALGetClusters(LALStatus*, CList*, Spectrogram*, CListDir*);
```

Integrated power (third threshold)

- Reject a fraction of the clusters that would make the first two cuts with just steady-state noise: threshold on

  Prob(observed integrated power | cluster size, passed 1^{st} & 2^{nd} cuts)

```
void LALClusterPowerThreshold(LALStatus *, CList*, CList*, CListDir*);
```

- For θ_{ij} the representation of the signal in the short-time Fourier basis, the test is

$$\sum_{\text{cluster}} |\theta_{ij}|^2 \Theta(|\theta_{ij}|^2 - \lambda) > \Lambda(\text{cluster size})$$

lalwrapper so

- split time series in n overlapping segments
- get event list for every segment
- merge lists on master

Optimality

Problem	Optimal Test	Optimal Estimator
Binary Hypothesis $W = \{\theta\}$	Likelihood ratio	N/A
Prior on signal $p(\theta)$	Averaged likelihood ratio	Bayes estimation
Filter bank $W = \{\theta_i : \mathrm{i}=1,2,...\}$	Maximum likelihood ratio	Maximum likelihood
Smooth (sparse) signal	Power after thresholding	Hard thresholding

$$y = s + n, s \in W$$

LIGO

Smooth (sparse) signals

- Model the signal subspace as a L_p ball minus the (L_2) ball of signals with SNR < ε

$$W = U_p(C) \setminus U_2(\varepsilon)$$

for the balls

$$U_p(C) = \left\{ \theta \in R^N : \sum_i |\theta_i|^p < C^p \right\}$$

- L_p balls are made of sparse vectors if p<2. Sparse vectors in the wavelet domain (or STFT) are smooth functions in the time domain.
- If p>2, optimal detectors are "quadratic forms" in y.
- If p<2, optimal detectors involve "coordinate-wise truncations" on y.

LIGO

Remarks

- Detectors with coordinate-wise truncations are optimal over a wide range of measures of smoothness (Besov,Triebel,...)
- Clustering analysis is added “by hand”. No proof of optimality (yet).
- Of course, must choose a basis where the signal is sparse. Wavelets good for that (unconditional bases of smooth functional spaces).