Noise Identification in Gravitational wave search using Artificial Neural Networks

Young-Min Kim (Pusan Nat'l Univ. & NIMS)

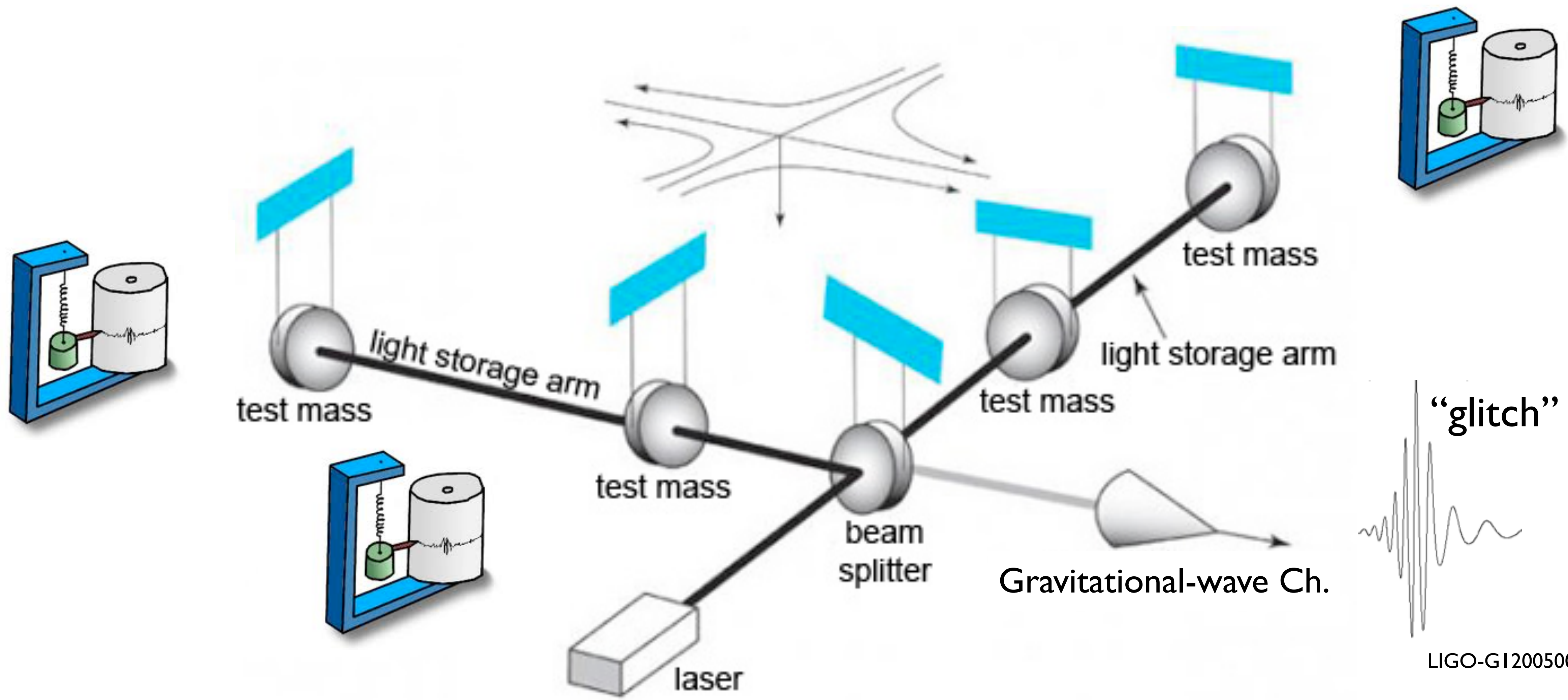
R. Biswas[1], L. Blackburn[2], J. Cao[3], R. Essick[4], K. A. Hodge[5], E. Katsavounidis[4], K. Kim[6,7], Y.-M. Kim[8,7], E.-O. Le Bigot[3], C.-H. Lee[8], J. J. Oh[7], S. H. Oh[7], E. J. Son[7], R. Vaulin[4], X. Wang[9], and T. Ye[9].

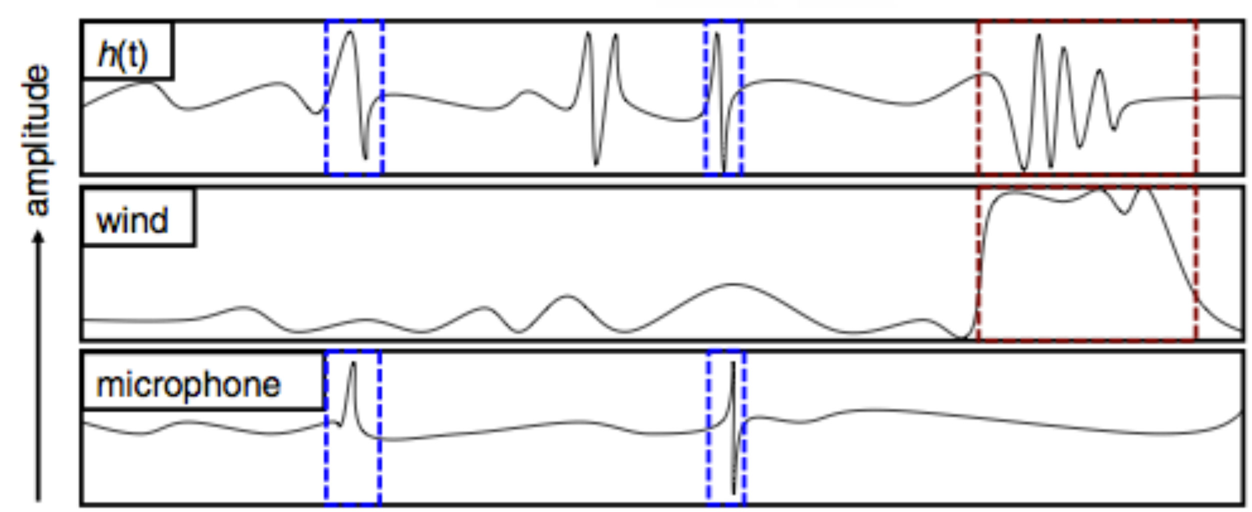
[1]UTB, [2]NASA/GSFC, [3]TNLIST, Tsinghua Univ., [4]LIGO-MIT, [5]LIGO-Caltech, [6]Hanyang Univ., [7]NIMS, [8]Pusan Nat'l Univ. [9]Tsinghua Univ.

K. Hayama, T. Yamamoto (Osaka City Univ.)

4th K-J workshop on KAGRA @ Osaka Univ.

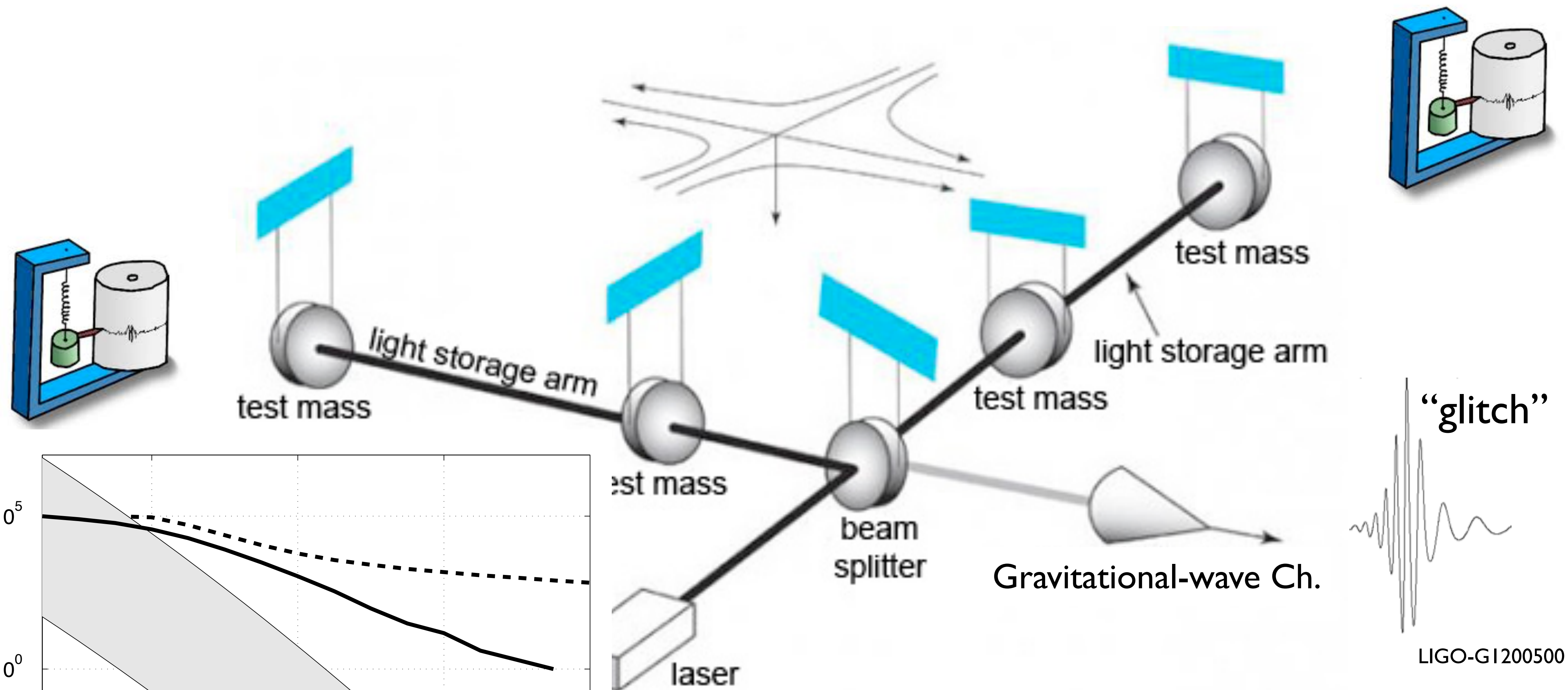
Laser Interferometer for GW Observation

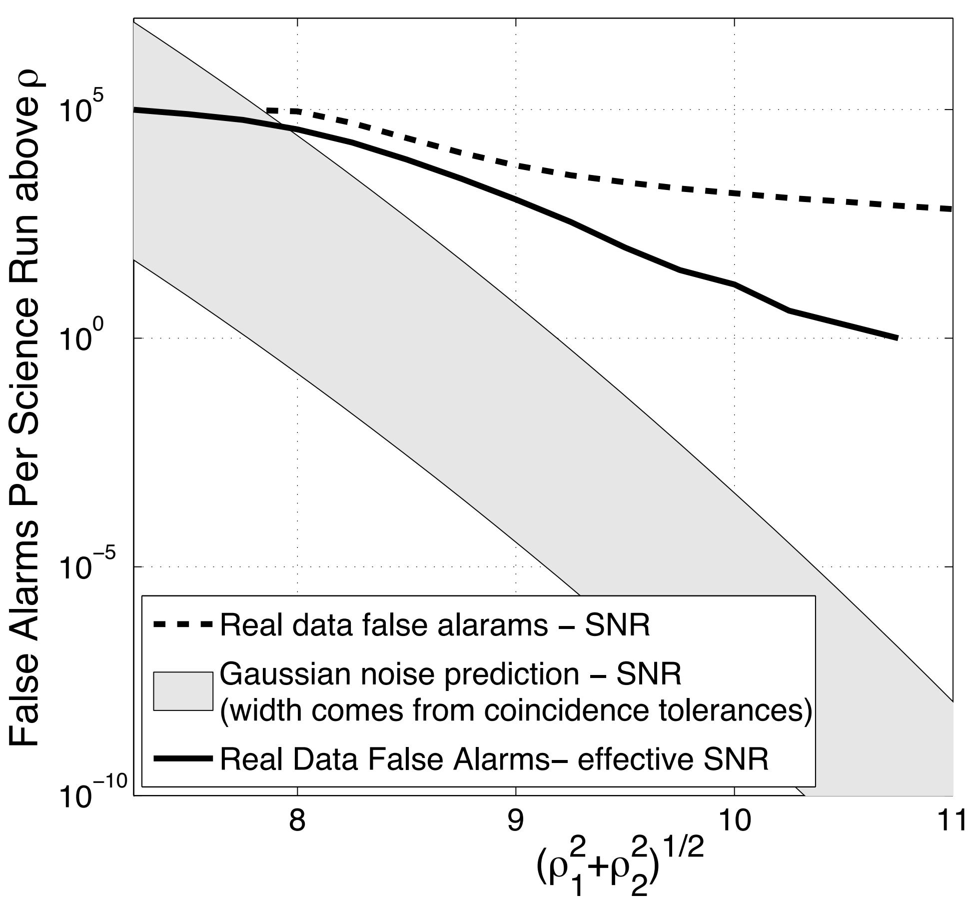


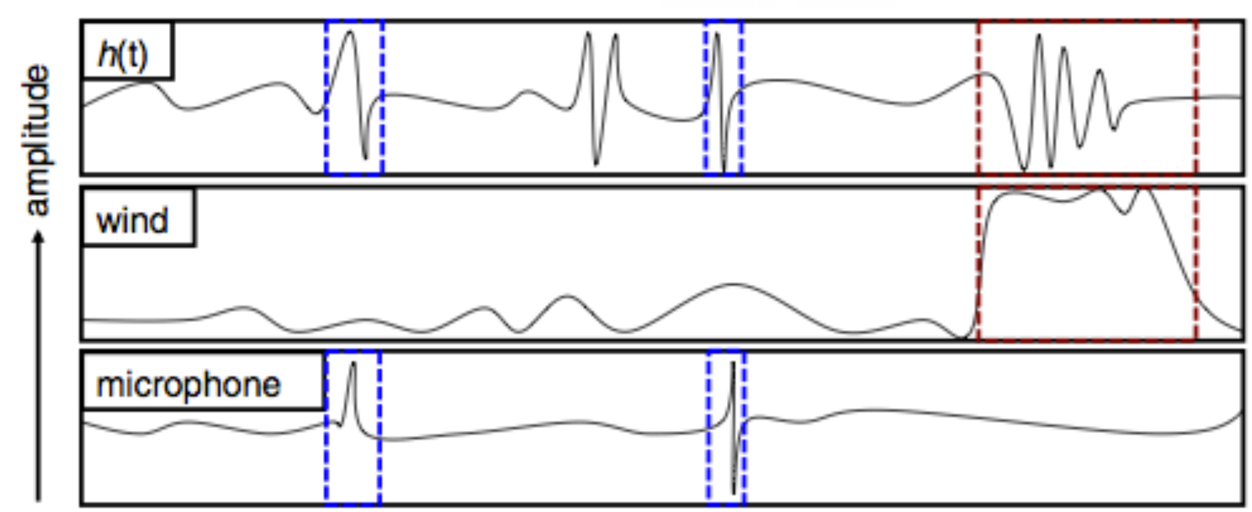


Univ.

Laser Interferometer for GW Observation





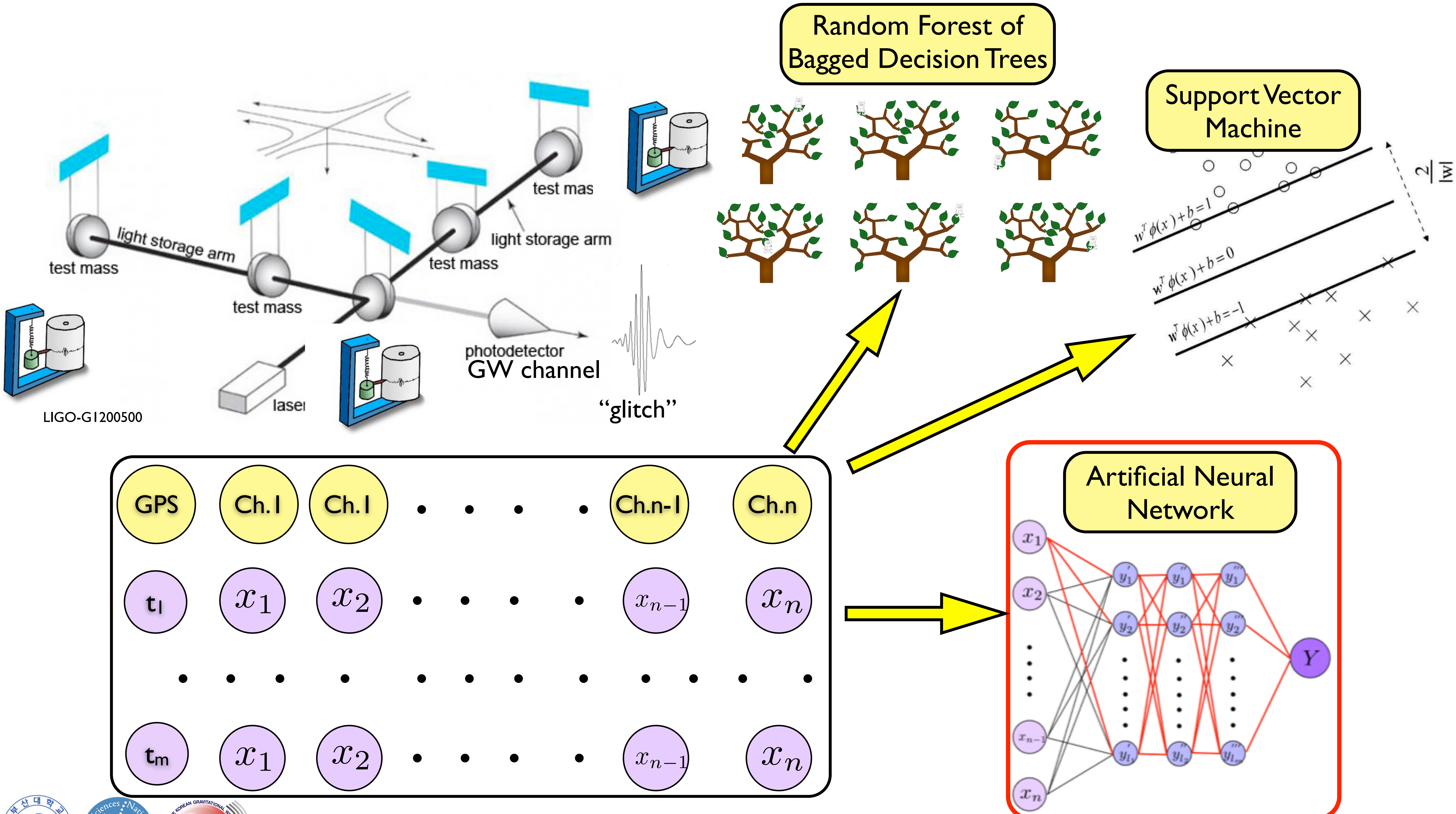


Univ.

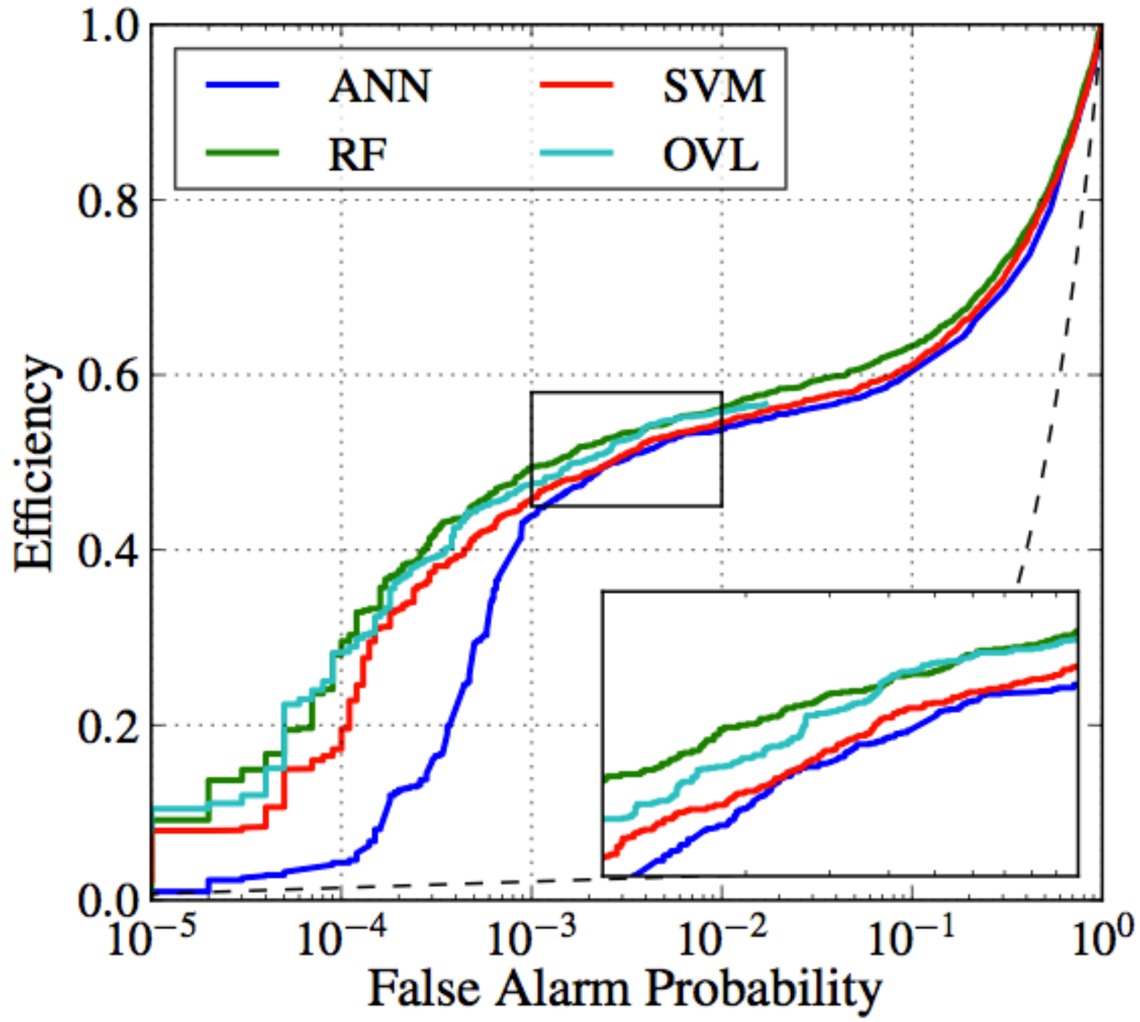
Goal

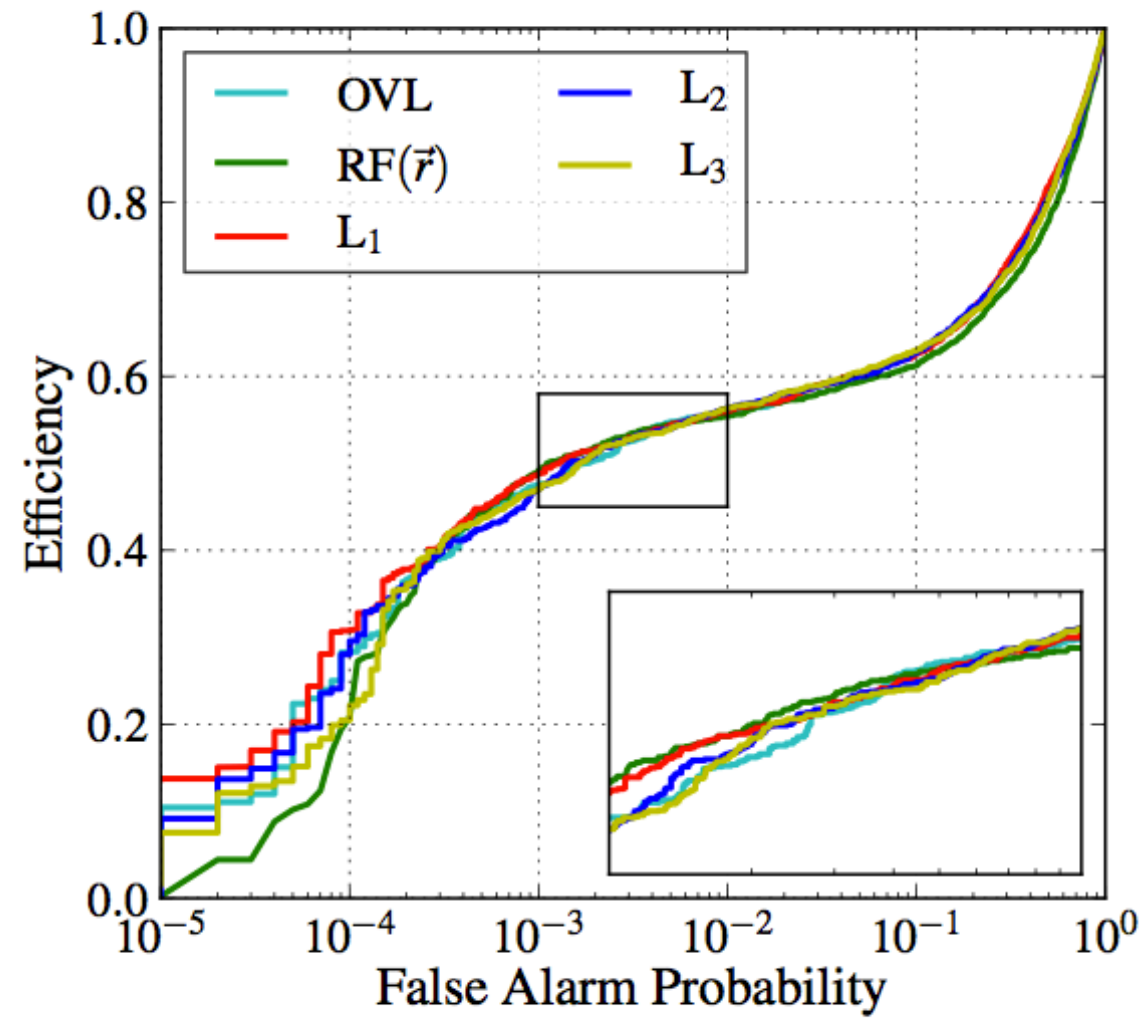
- Automate process of Glitch (non-gaussian noise transient) Identification in data preparation stage
  - Cleaning the data, monitoring, and feedback for commissioning or tuning

Multi-Variate Classifiers



AuxMVC - first draft



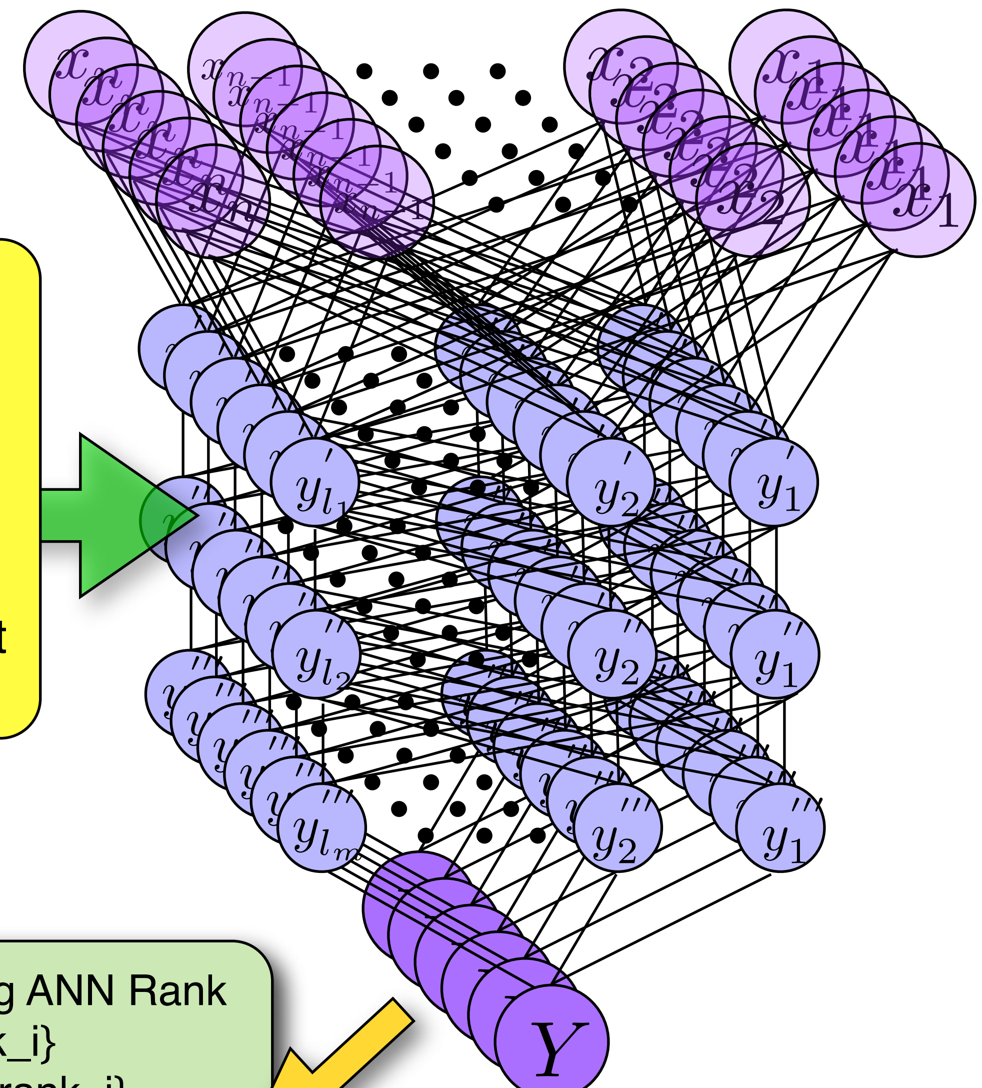


arxiv:1303.6984

Artificial Neural Networks Ensemble

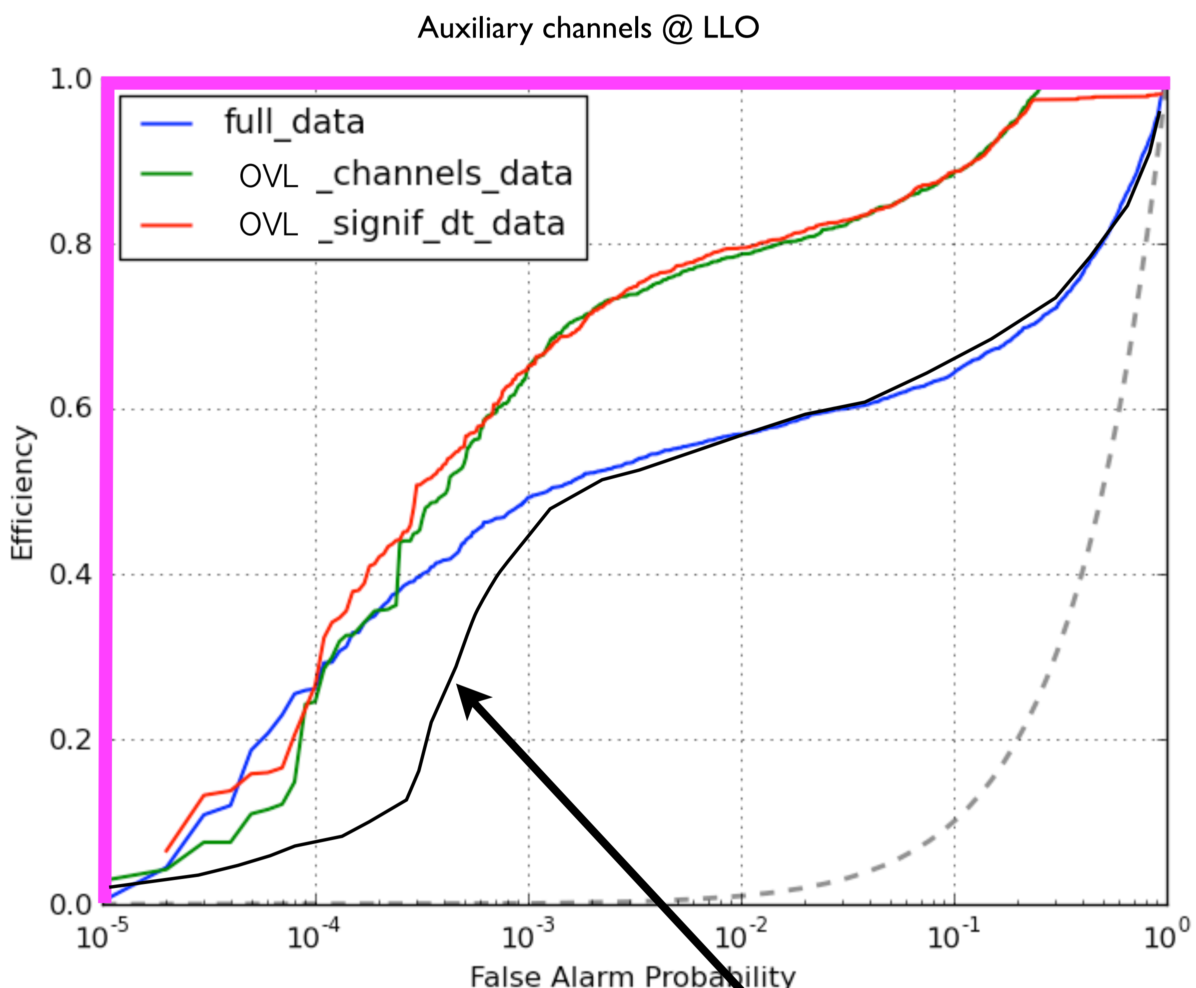
- Training Algorithm : GA (Montana & Davis, 2001), iRPROP(Igel & Hüsken, 2000)
- Evaluation : Direct application to trained network
  - ➡ tuning training parameters by hand, low complexity due to restricted network
  - ➡ over-trained ??
  - ➡ topological degeneracy from full connection
  - ➡ partial connections but random generation

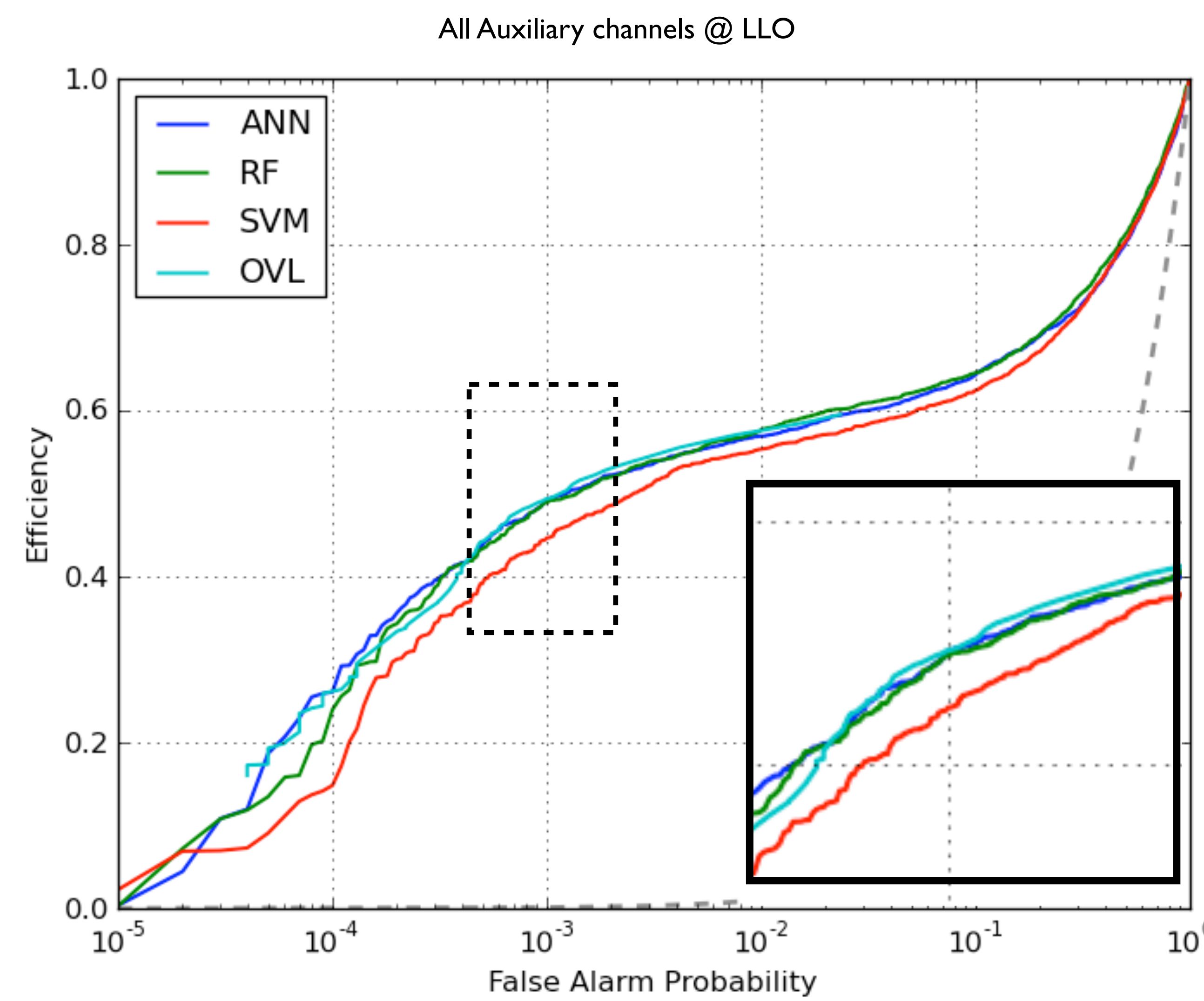
Use lots of trained neural networks for evaluation : Benchmarking Random Forest



Combining ANN Rank
- max{rank_i}
- average{rank_i}
- max{eff_i/fap_i}
- likelihood ratio
- something new ?

Application Results



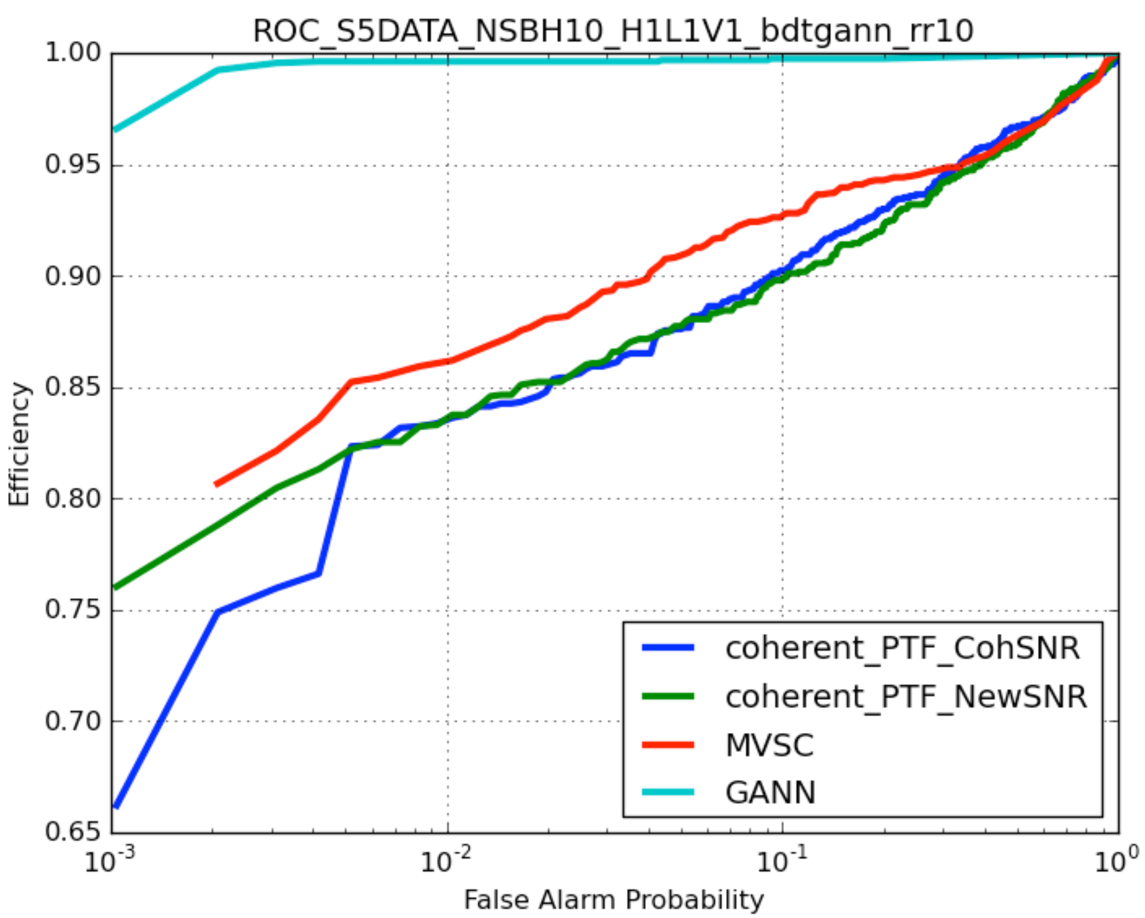


single network with full connection of limited number of neurons

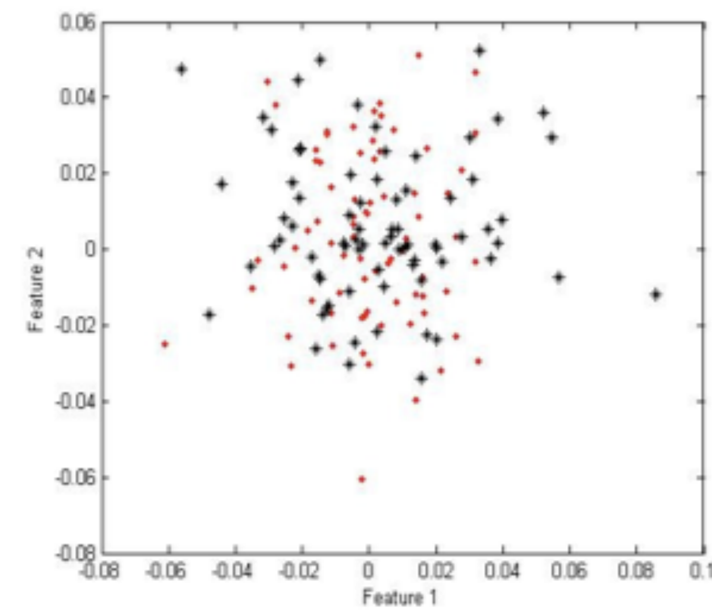
LIGO-G1201110

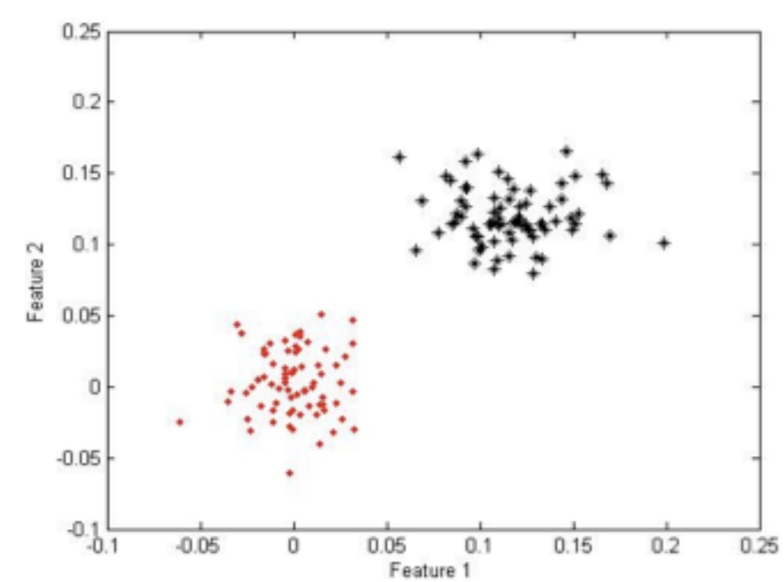
LIGO

Q: Why AuxChannel Data So Different with GRB Data in the viewpoint of Multivariate Classifiers?

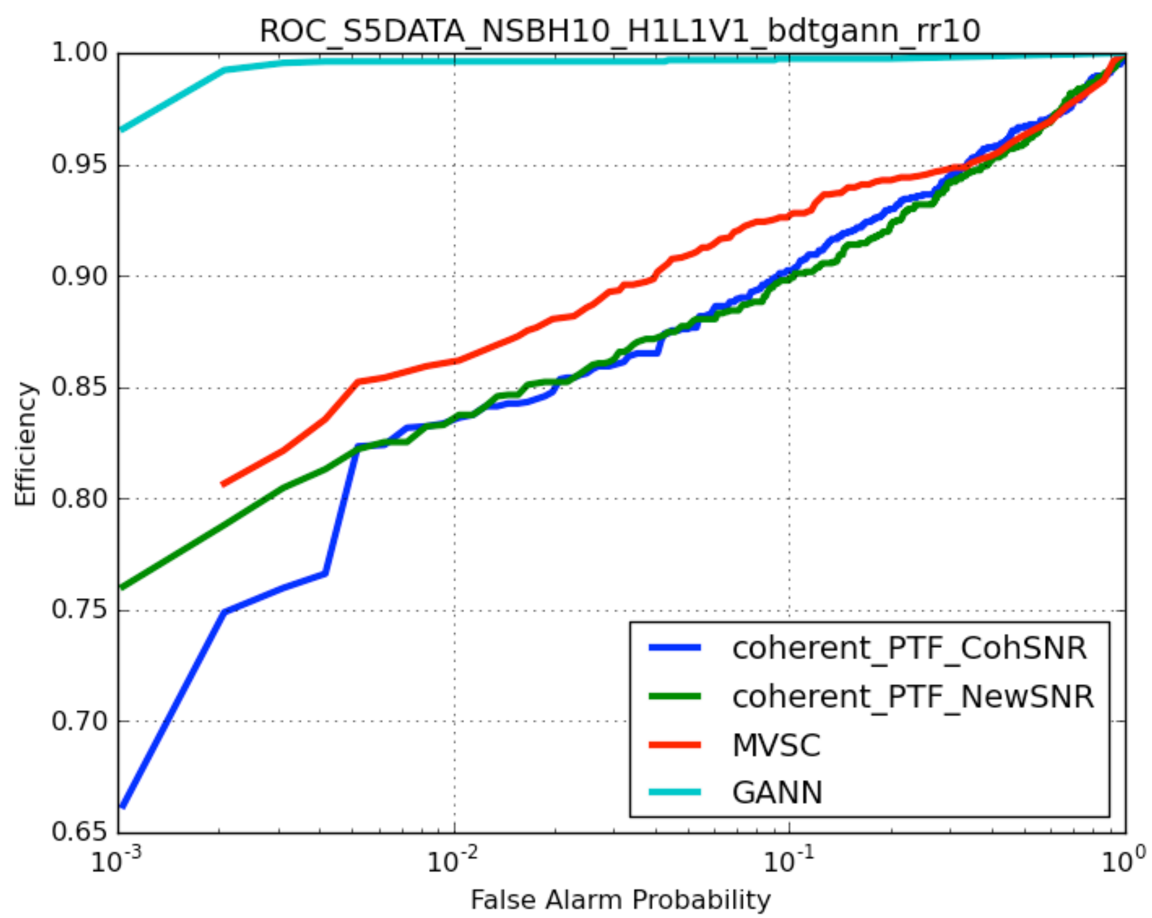


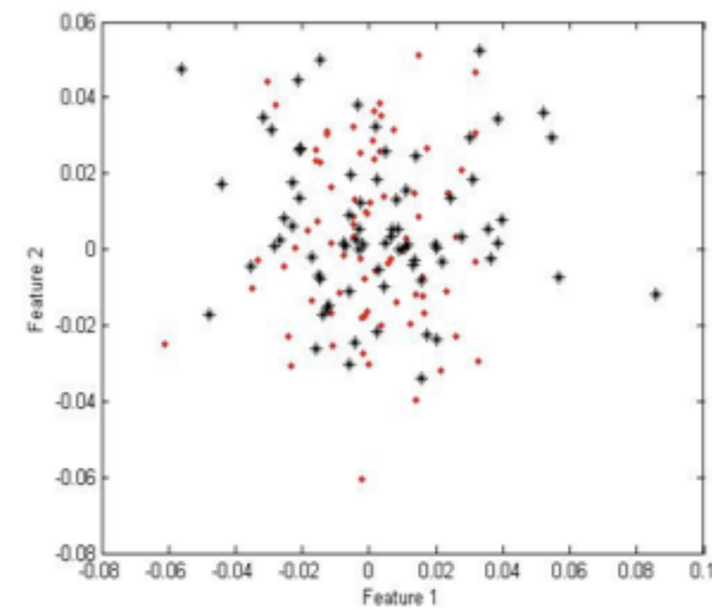
from Kyungmin Kim's presentation

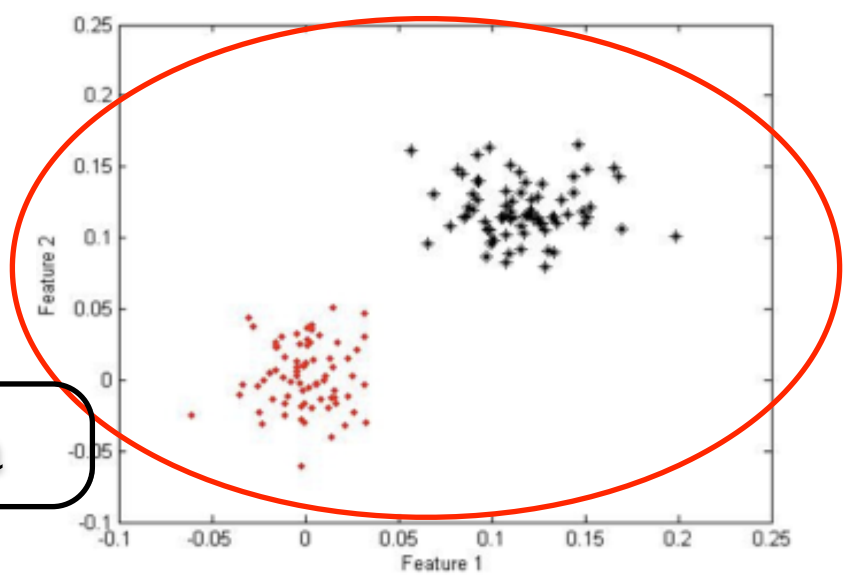




Q: Why AuxChannel Data So Different with GRB Data in the viewpoint of Multivariate Classifiers?



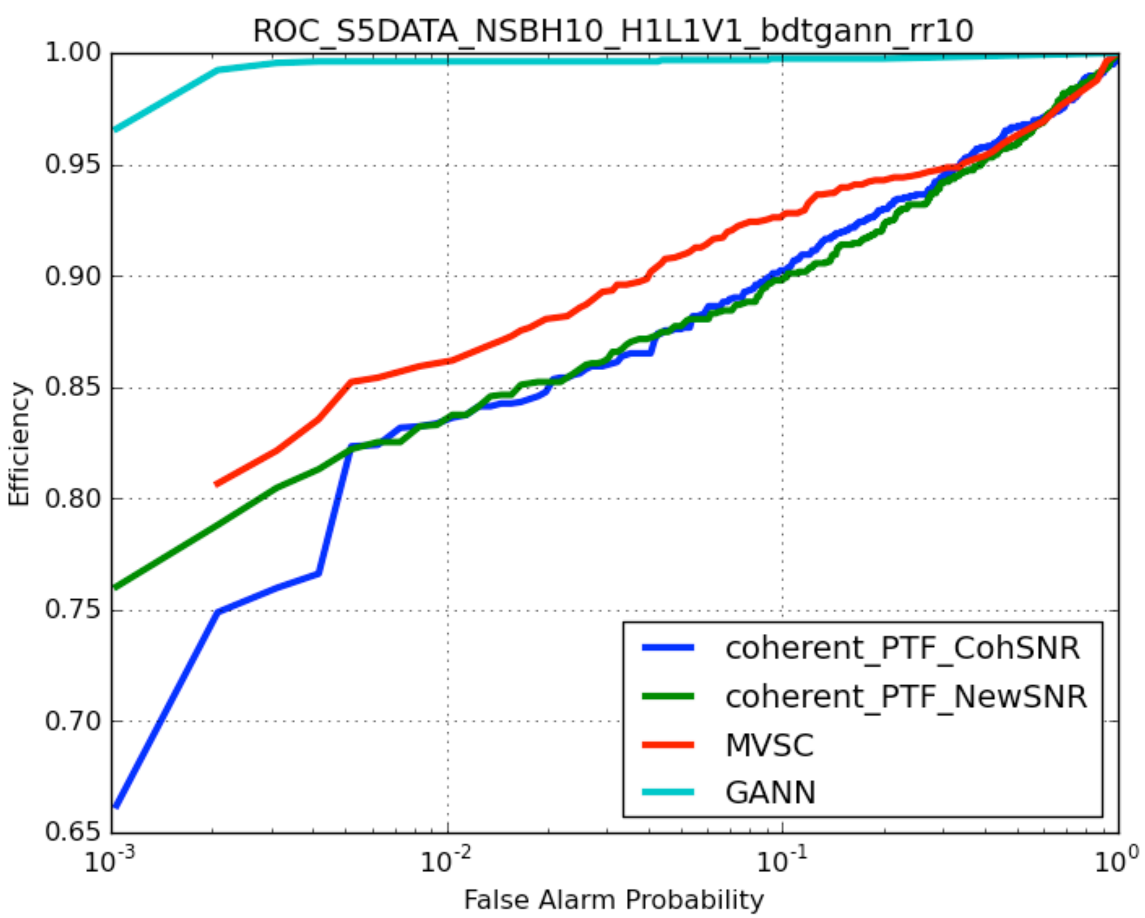




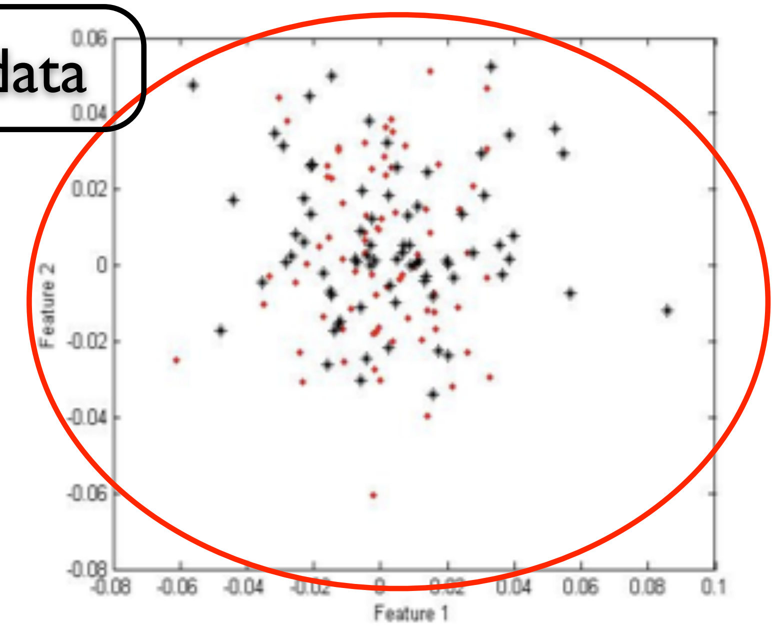
from Kyungmin Kim's presentation

CBC-GRB data

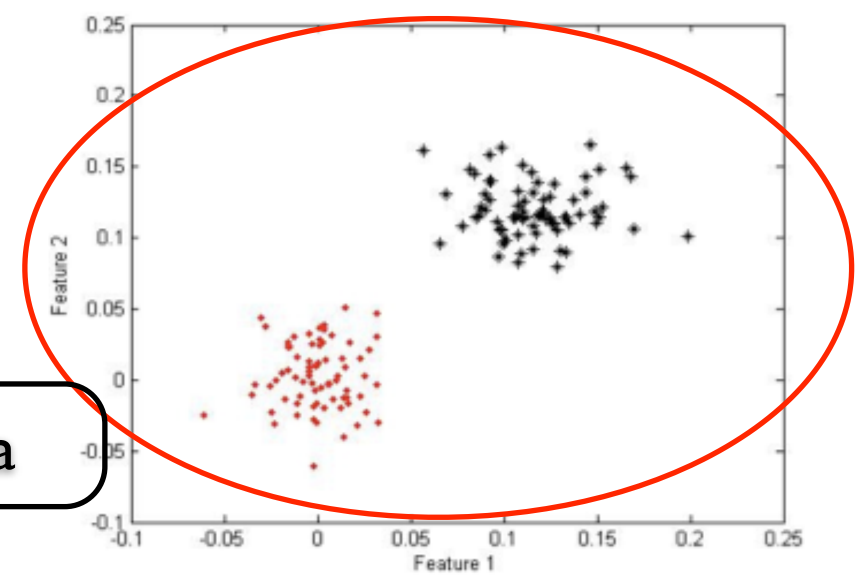
Q: Why AuxChannel Data So Different with GRB Data in the viewpoint of Multivariate Classifiers?



Auxiliary Ch. data



CBC-GRB data



from Kyungmin Kim's presentation

Feature Selection : t - statistic

1. one-sample t-test

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

$$\begin{aligned} \bar{x} &= \text{sample mean} \\ s &= \text{sample standard deviation} \\ n &= \text{sample size} \\ \mu_0 &= \text{specific value} \end{aligned}$$

2. two-sample t-test

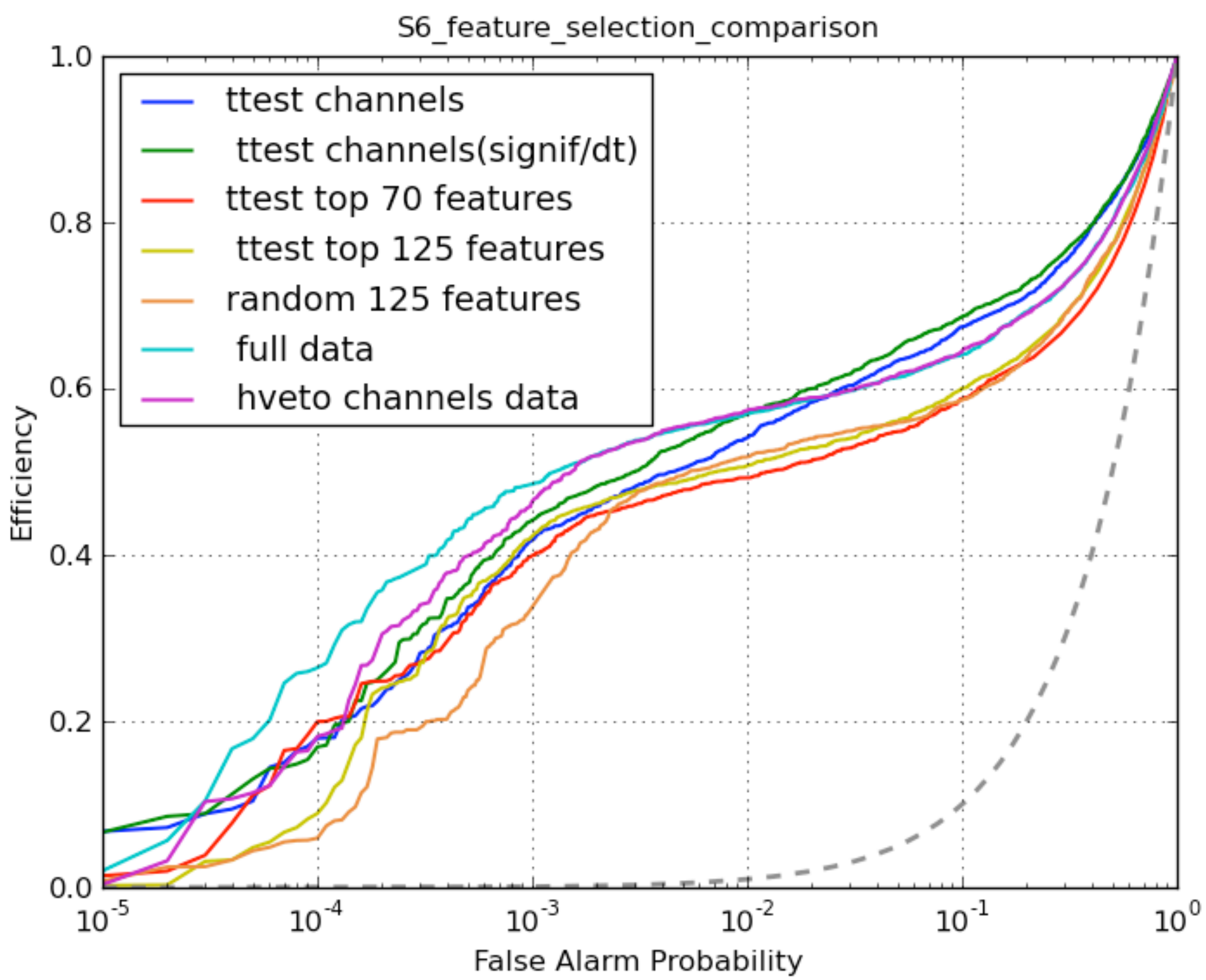
$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

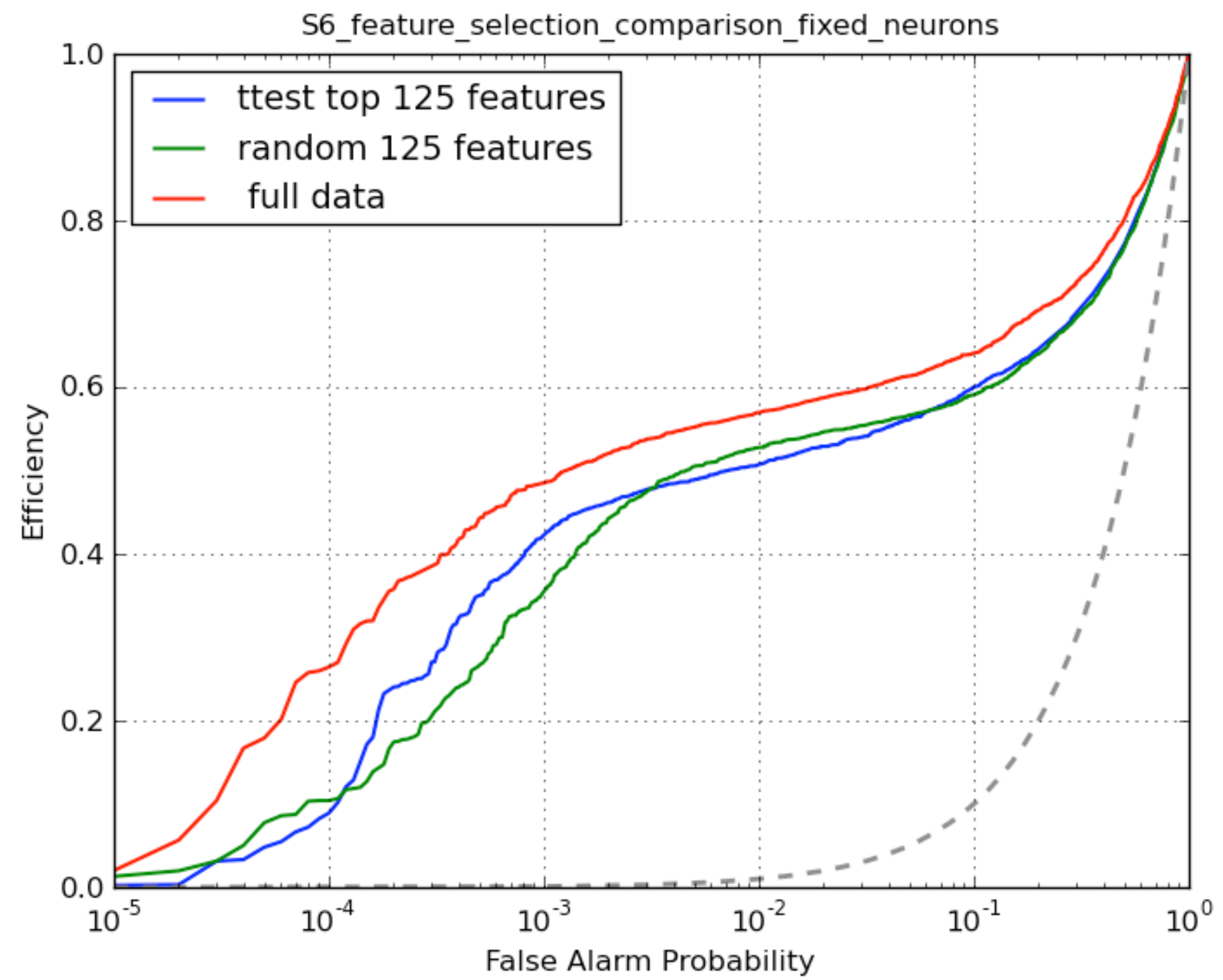
$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

$$\begin{aligned} \bar{X}_i &= \text{sample mean of class i} \\ s_i &= \text{sample standard deviation of class i} \\ n_i &= \text{sample size of class i} \end{aligned}$$

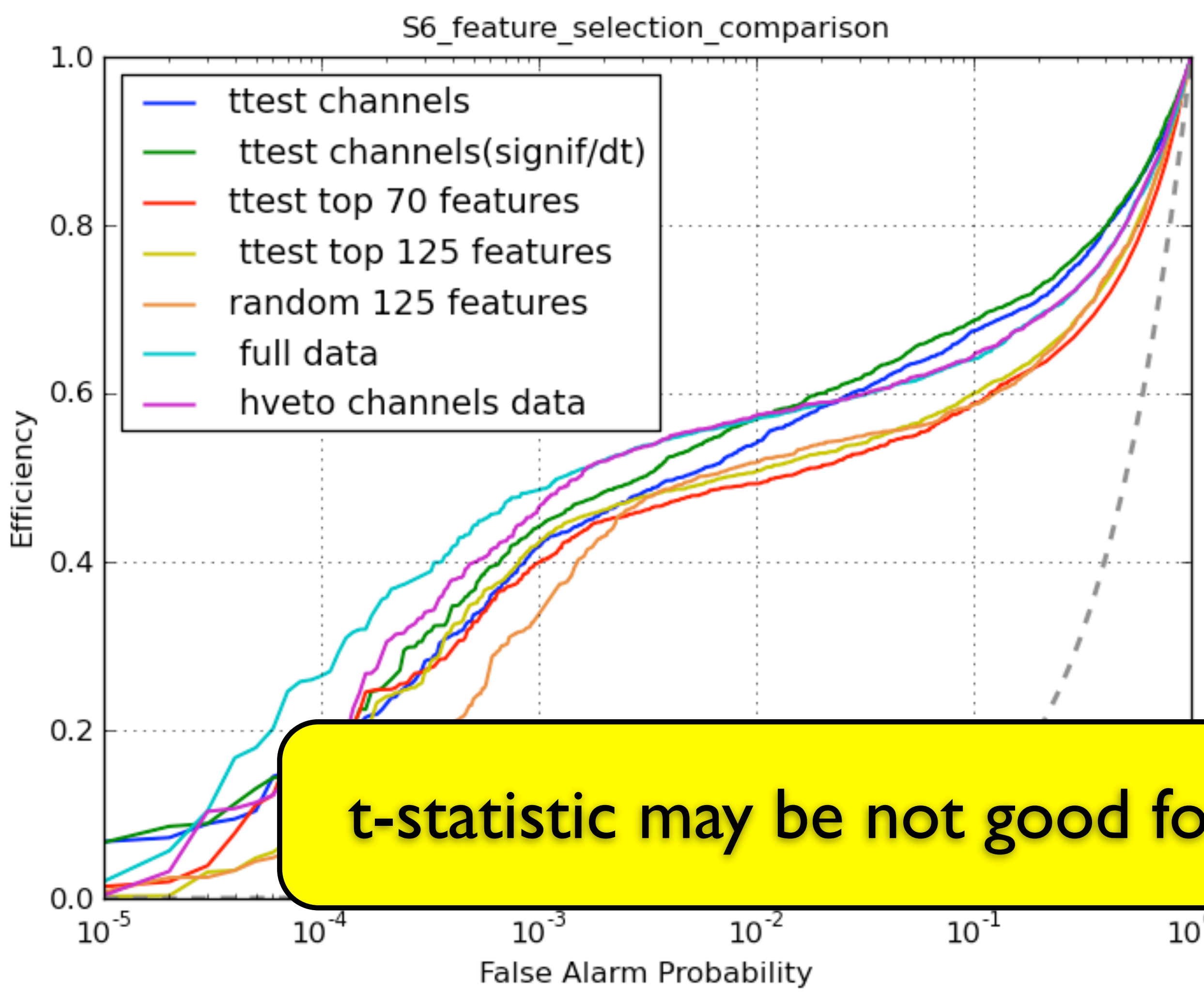
test on each auxiliary channel

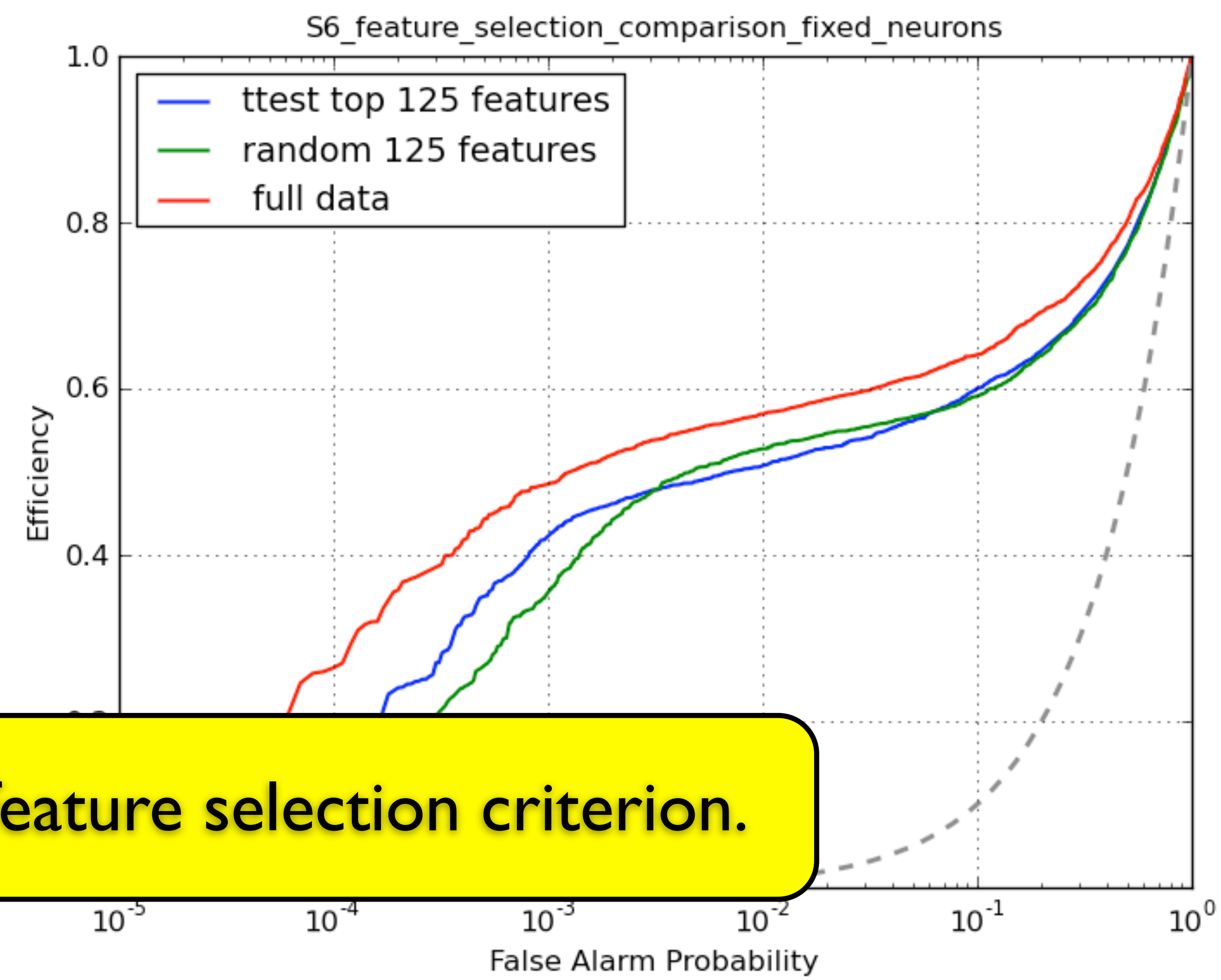
ROC performances of t-test features





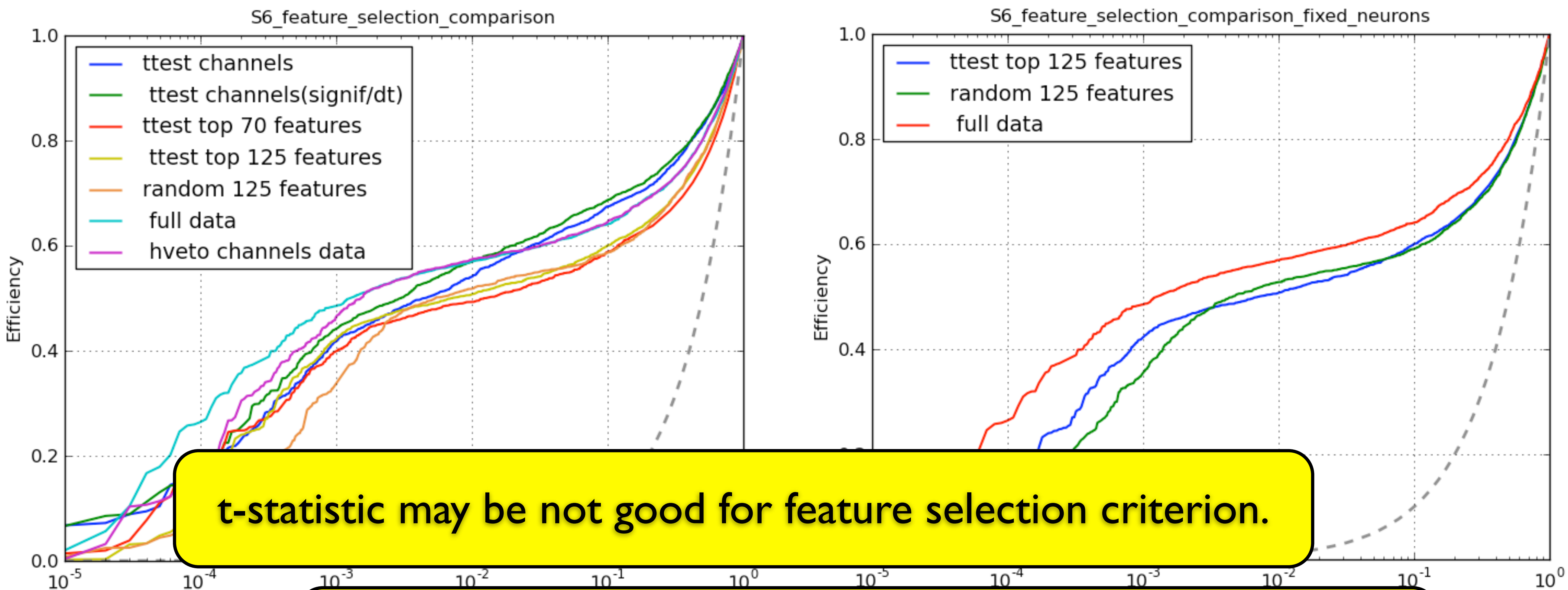
ROC performances of t-test features





t-statistic may be not good for feature selection criterion.

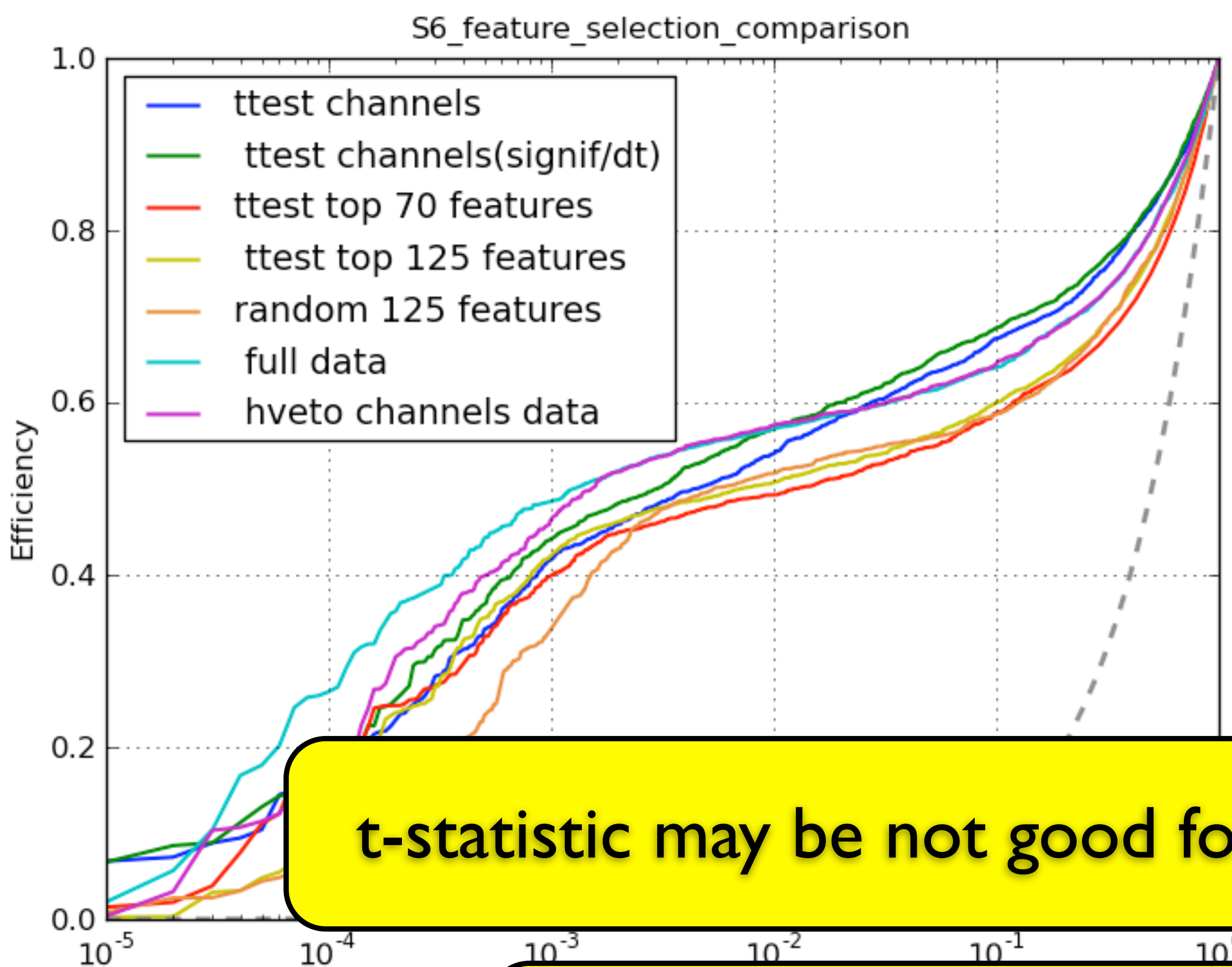
ROC performances of t-test features

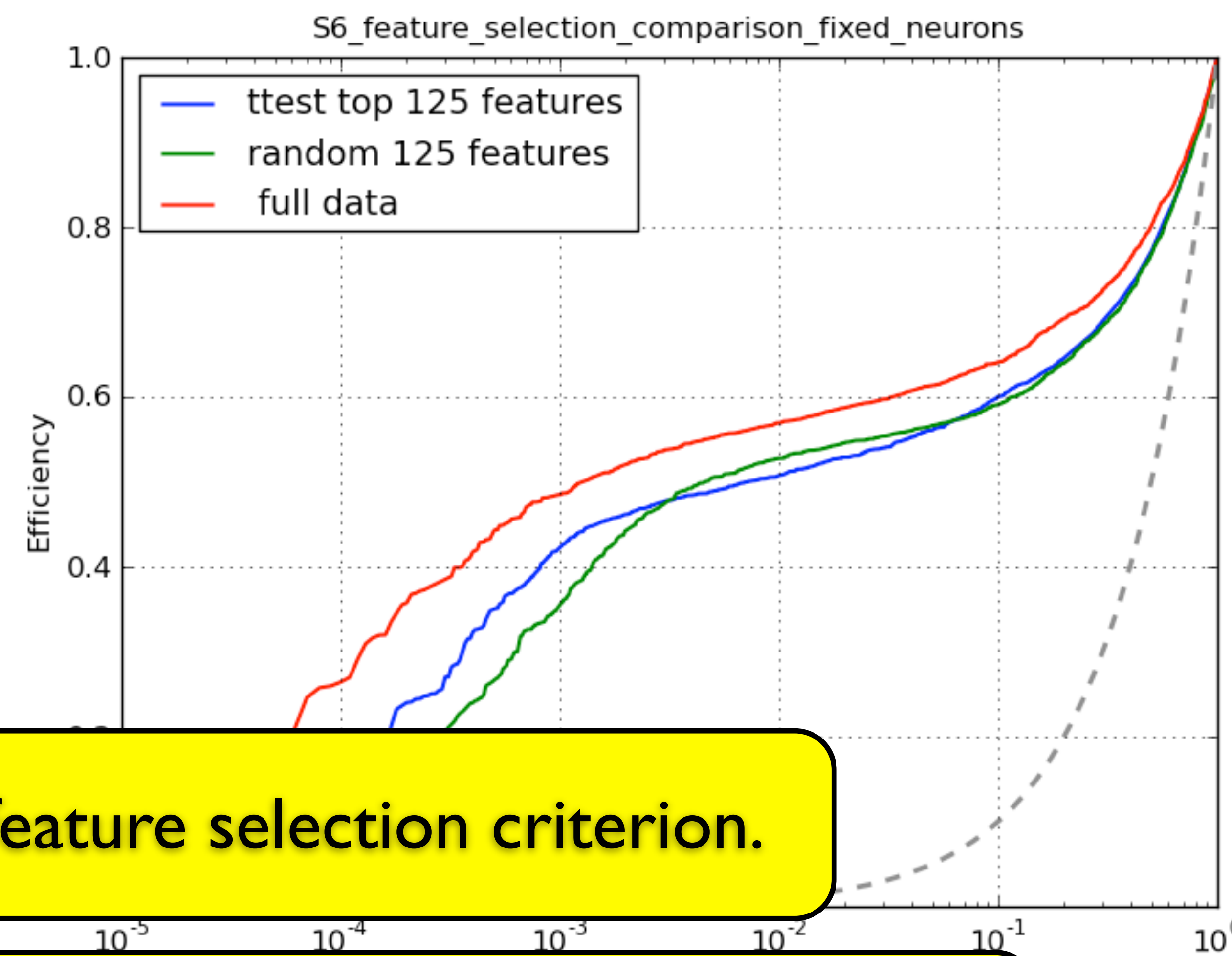


t-statistic may be not good for feature selection criterion.

Feature selection is still needed for better performance.

ROC performances of t-test features





t-statistic may be not good for feature selection criterion.

Feature selection is still needed for better performance.

Deferent input information is needed.

Goal

- Automate process of Glitch (non-gaussian noise transient) Identification in data preparation stage
  - Cleaning the data, monitoring, and feedback for commissioning or tuning
- Inclusion of different information which current input data (KW triggers) may not contain.
  - consideration of different trigger generation algorithm

Trigger Generation methods

1. KleineWelle algorithm (Current trigger generation)
2. Omega pipeline (Q-transformation)
3. Excess Power
4. Omicron pipeline
5. GSTLAL pipeline
6. Or Something NEW ???

The Hilbert-Huang Transformation

- Hilbert Transformation + Empirical Mode Decomposition
  - Decomposition of complicated data into a collection of Intrinsic Mode Function (IMF)
- It works well for non-stationary and non-linear data.
- Preliminary study of HHT on GW data analysis.
  - J. B. Camp et al, "Application of the Hilbert-Huang transformation to the search for gravitational waves" PRD 75, 061101 (2007)
  - A. Stroeer et al, "Methods for detection and characterization of signals in noisy data with the Hilbert-Huang transformation", PRD 79, 124022 (2009)

Hilbert transform

[wikipedia.org]

- In mathematics and in signal processing, the Hilbert transform is a linear operator which takes a function, u(t), and produces a function, H[u](t), with the same domain.

$$\mathcal{H}[u](t) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{u(\tau)}{t-\tau}d\tau$$

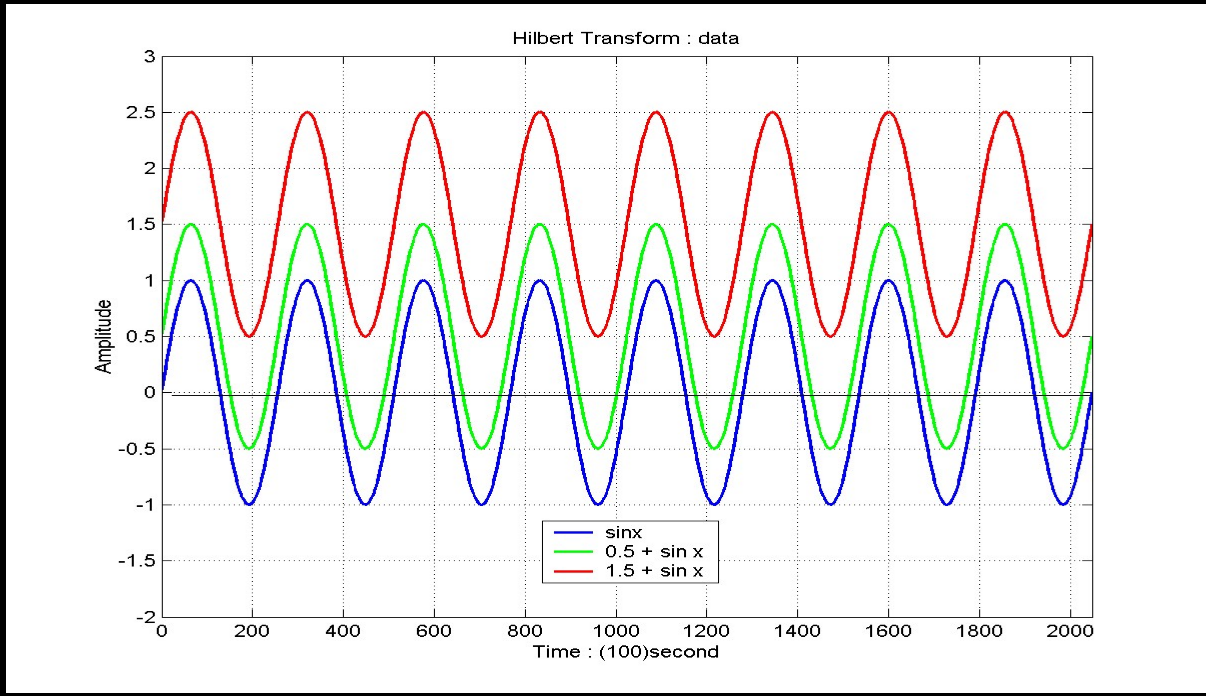
- Complexify:

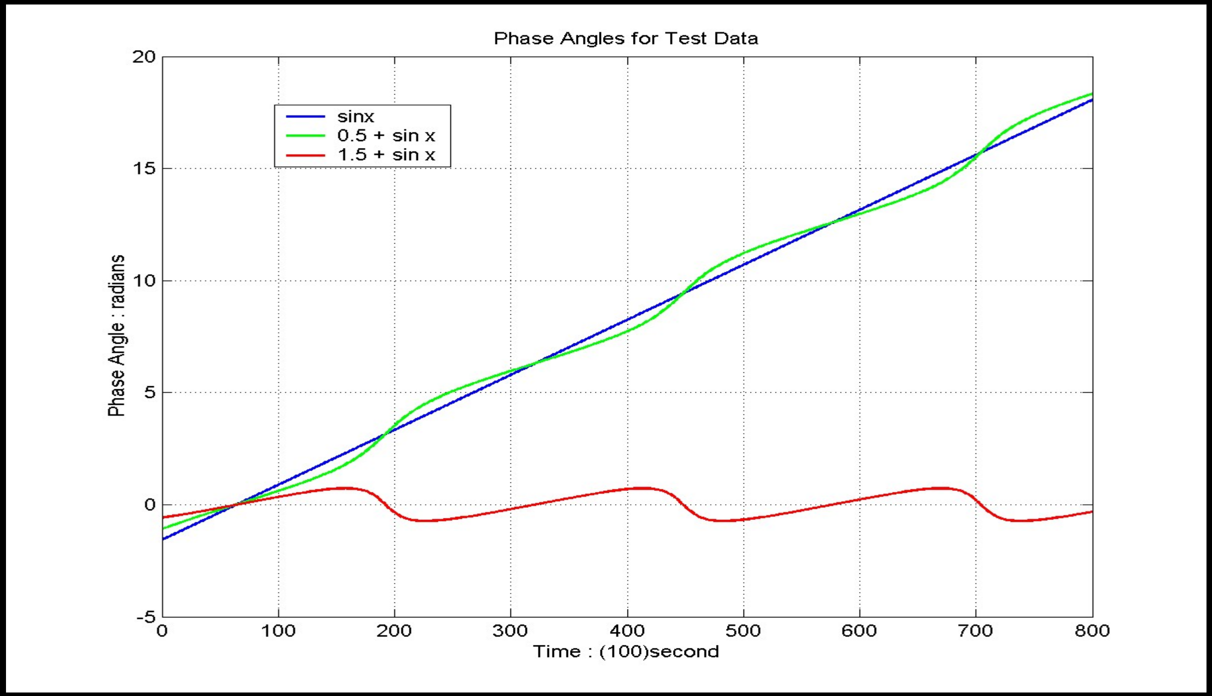
$$z(t) = u(t) + i\ \mathcal{H}[u](t) = a(t)e^{i\theta(t)}$$

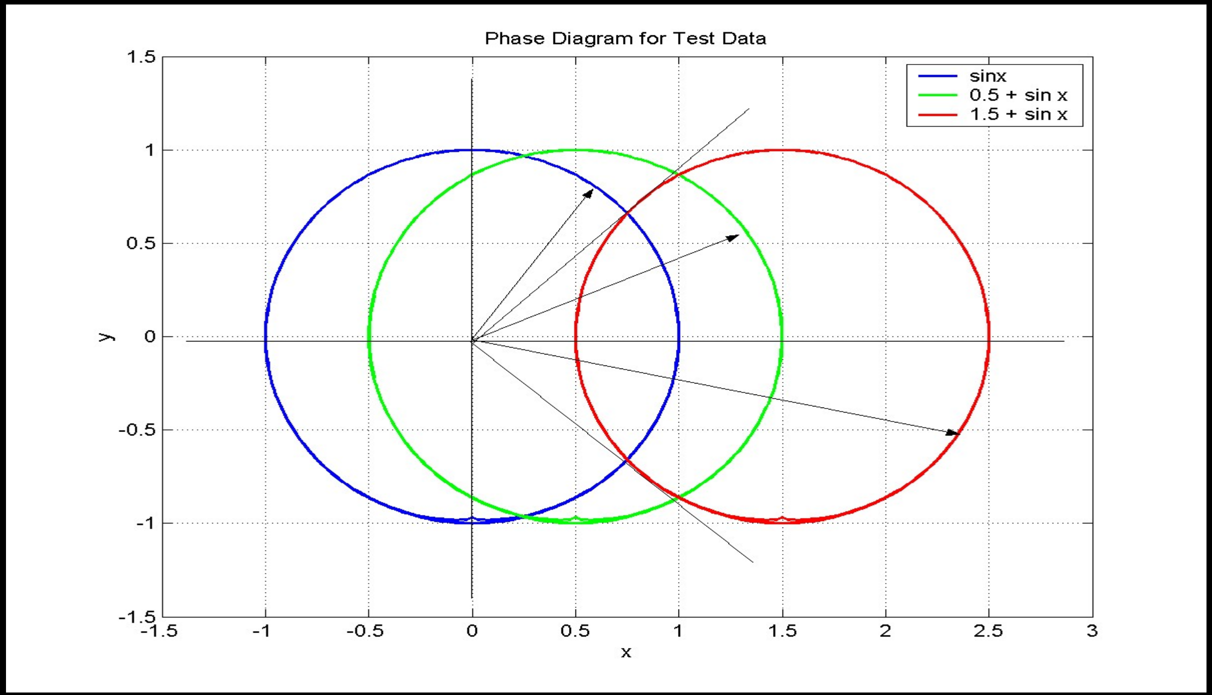
- Relationship with the Fourier transform:

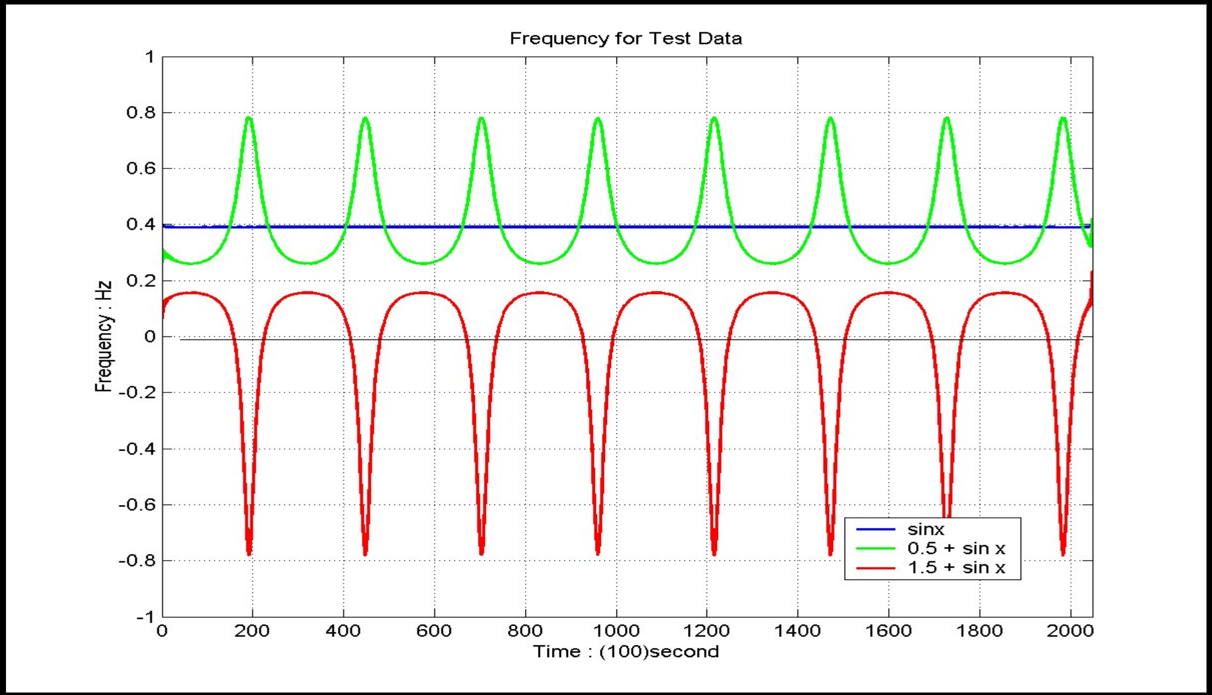
$$\mathcal{F}\left[\mathcal{H}[u]\right](\omega) = -i\ \mathrm{sgn}(\omega)\mathcal{F}[u](\omega)$$

A problem of traditional Hilbert transform



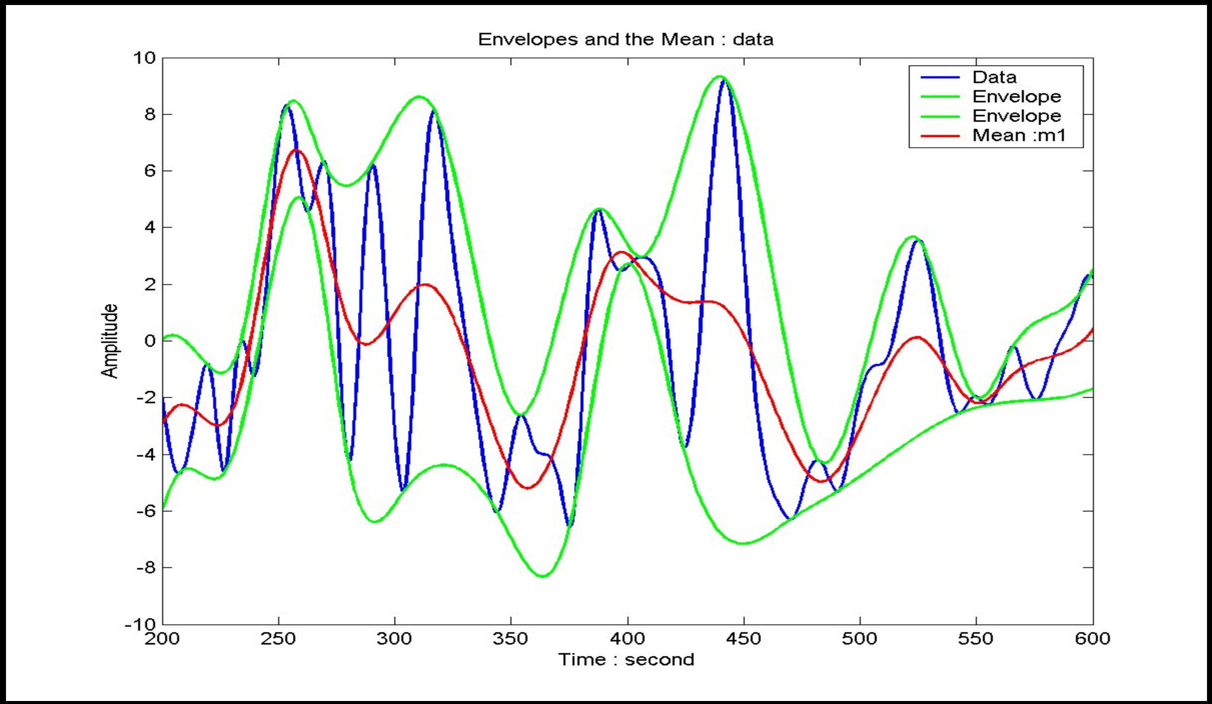


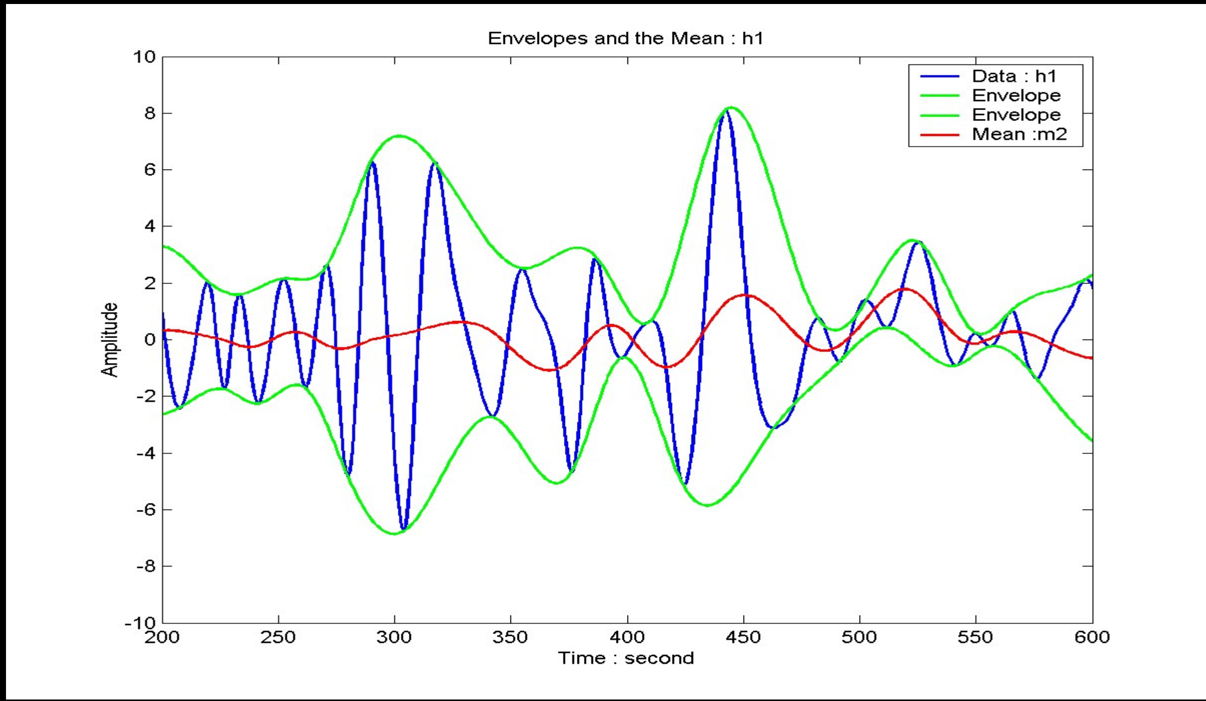




[Courtesy: Norden E. Huang]

Empirical Mode Decomposition (EMD)





"Sifting"

$$u(t) - m_1 = h_1,$$

$$h_1 - m_2 = h_2,$$

$$\vdots$$

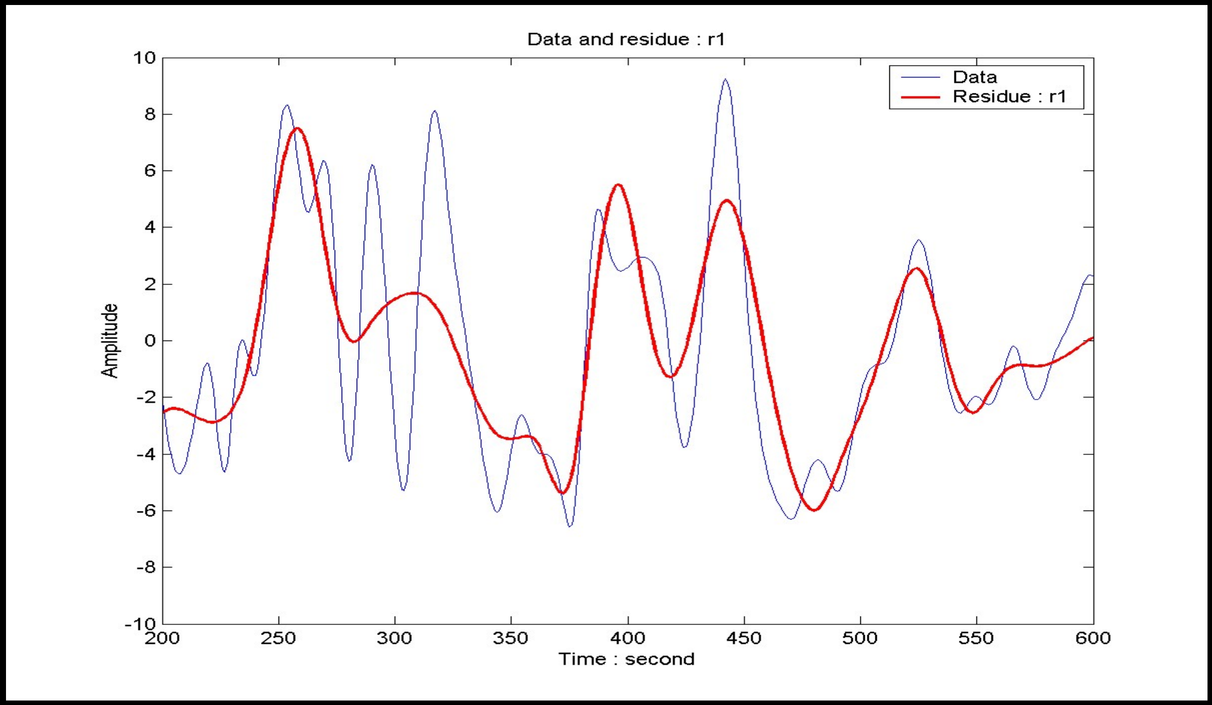
$$h_{k-1} - m_k = h_k \equiv c_1$$

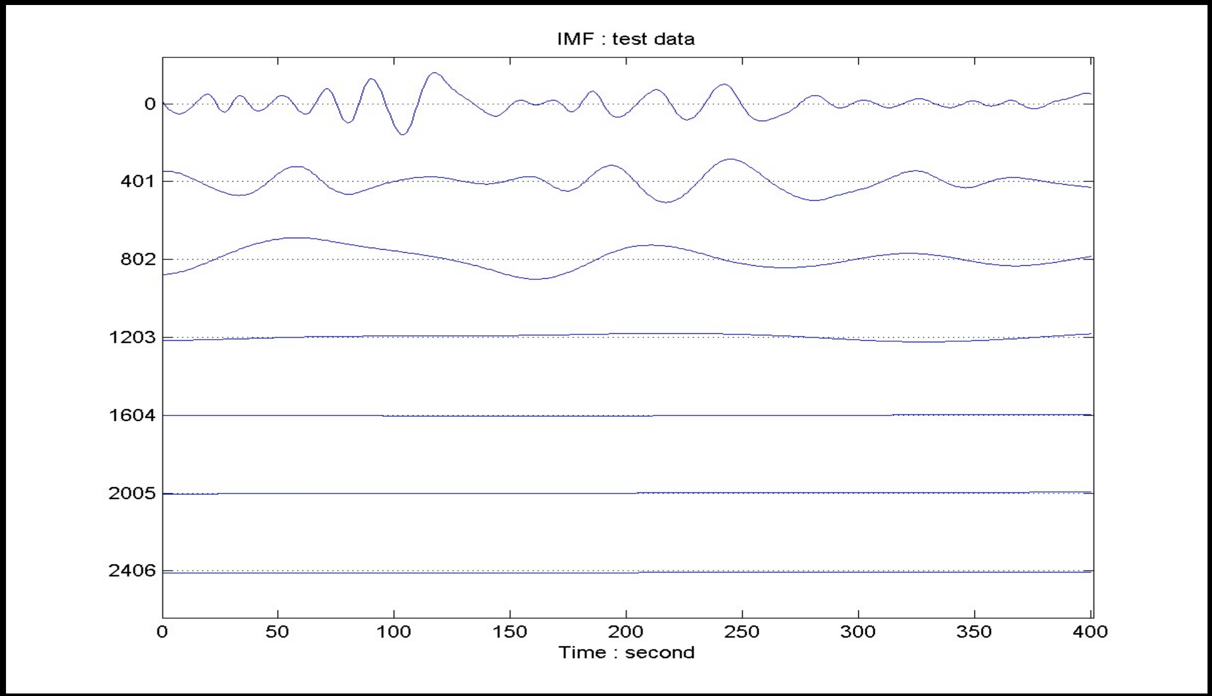
$$\Rightarrow u(t) - c_1 \equiv r_1 = \sum_i m_i$$

$$SD = \frac{\sum_t |h_{k-1}(t) - h_k(t)|^2}{\sum_t h_{k-1}^2(t)}$$

[Courtesy: Norden E. Huang]

Intrinsic Mode Function (IMF)





$$u(t) - c_1 = r_1,$$
$$r_1 - c_2 = r_2,$$
$$\vdots$$
$$r_{n-1} - c_n = r_n$$
$$\Rightarrow u(t) - \sum_j c_j = r_n$$

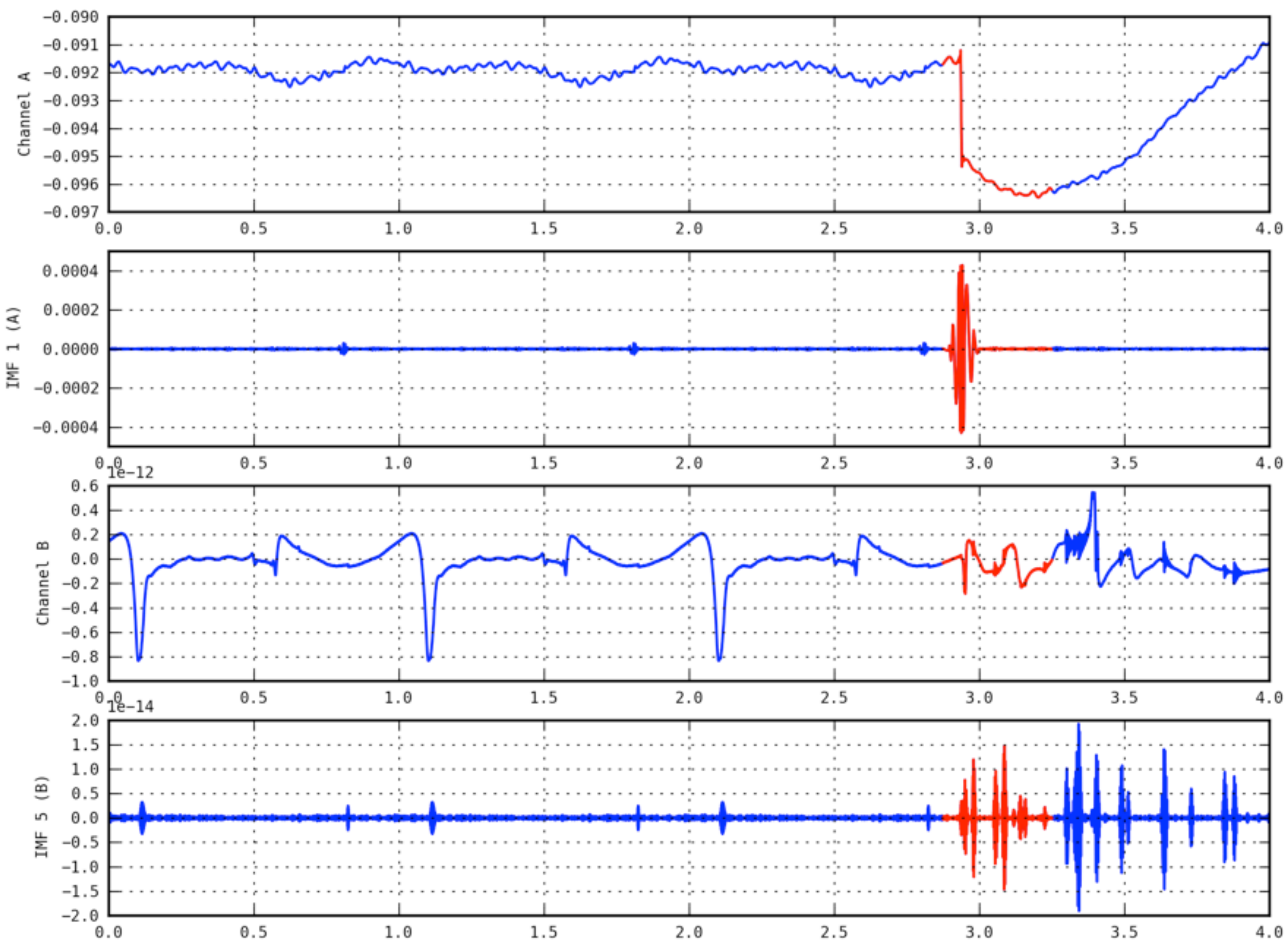
$$c_j(t) + i\,\mathcal{H}[c_j](t) = a_j(t)e^{i\theta_j(t)}$$

[Courtesy: Norden E. Huang]

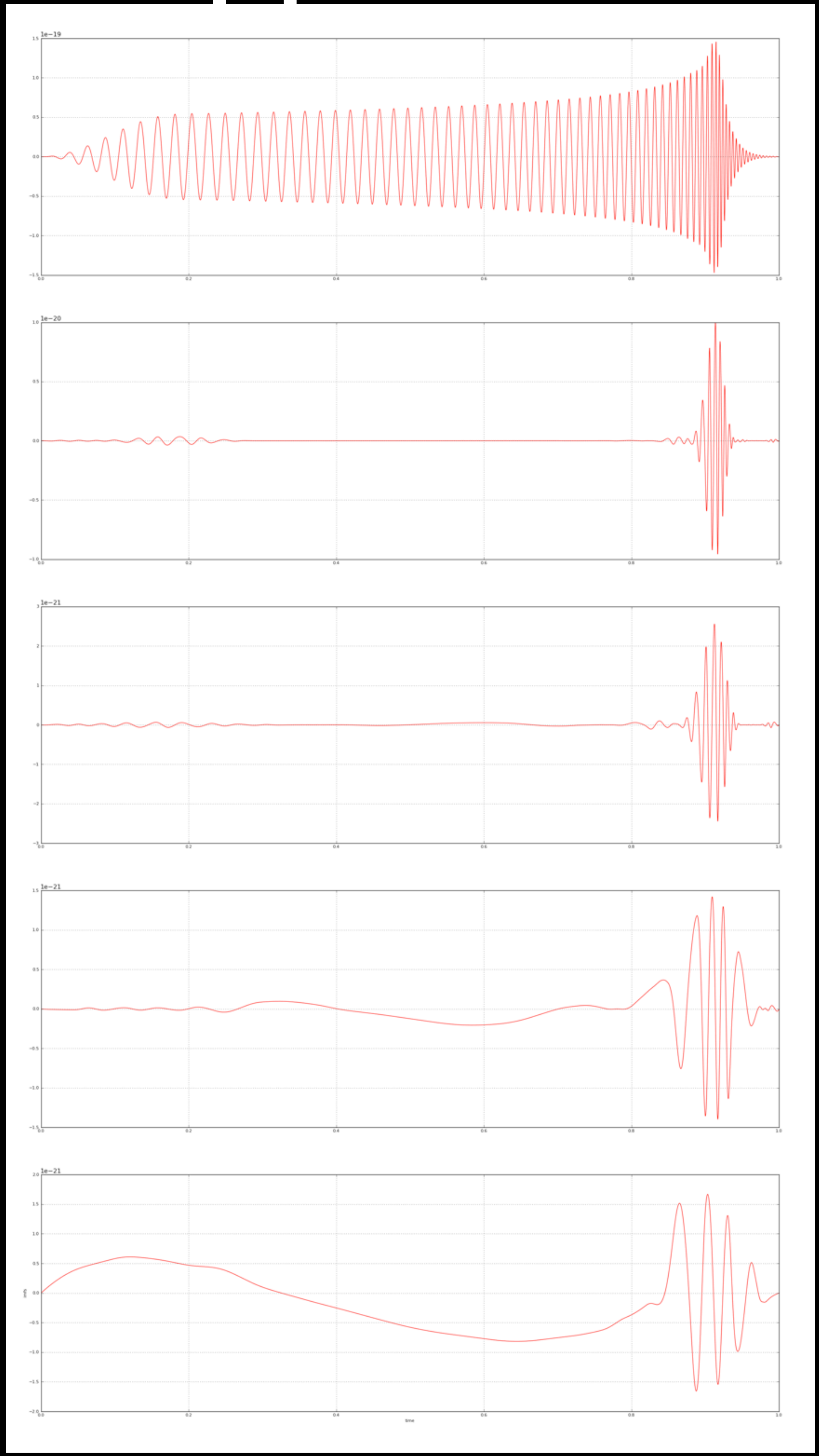
	Fourier	Wavelet	Hilbert
Basis	**a priori**	**a priori**	**Adaptive**
Frequency	Integral transform: Global	Integral transform: Regional	Differentiation: Local
Presentation	Energy-frequency	Energy-time-frequency	Energy-time-frequency
Nonlinear	**no**	**no**	**yes**
Non-stationary	**no**	**yes**	**yes**
Uncertainty	yes	yes	no
Harmonics	yes	yes	no

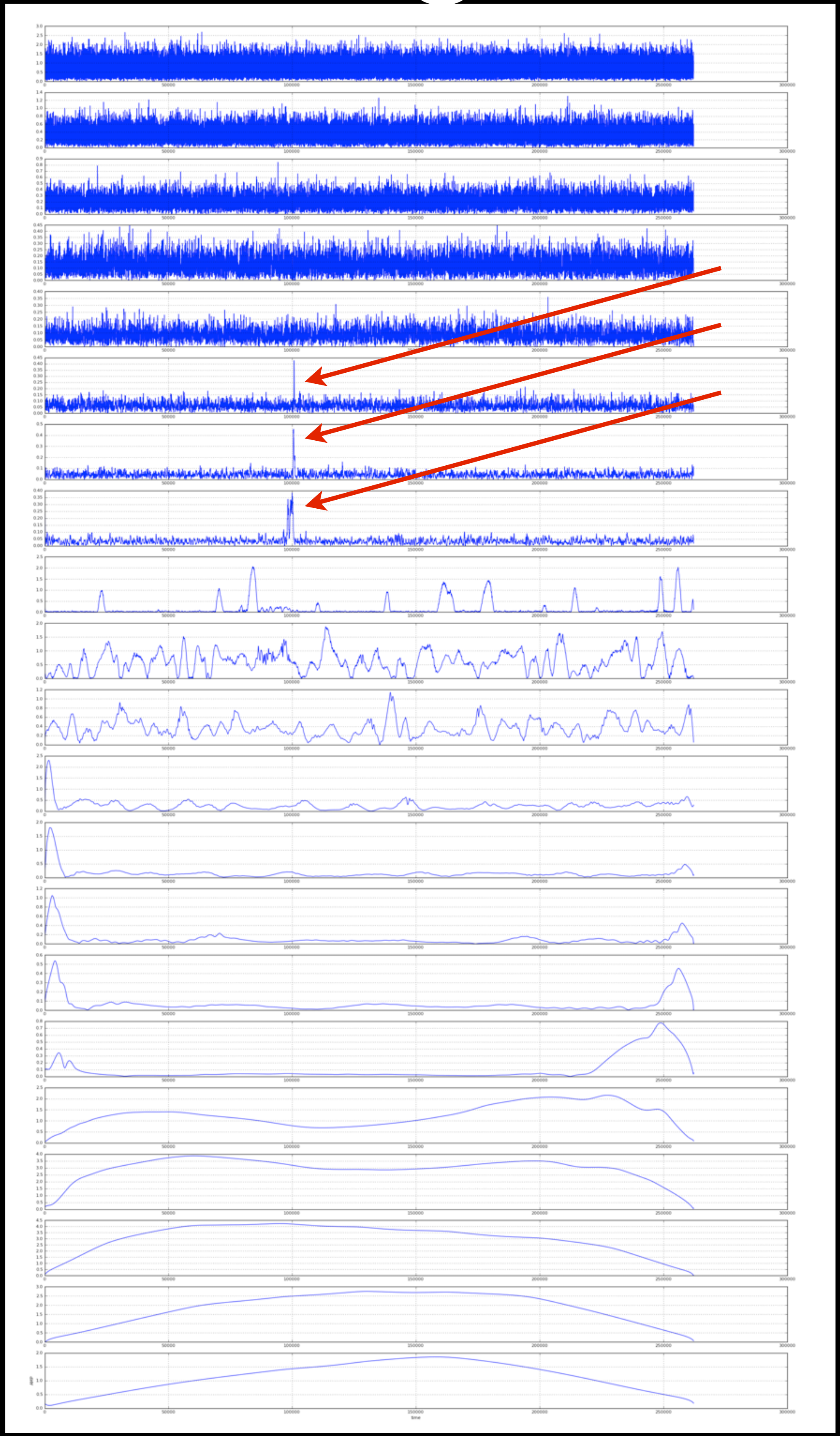
[Courtesy: Norden E. Huang]

Application to Aux. Ch.



Application to CBC signals



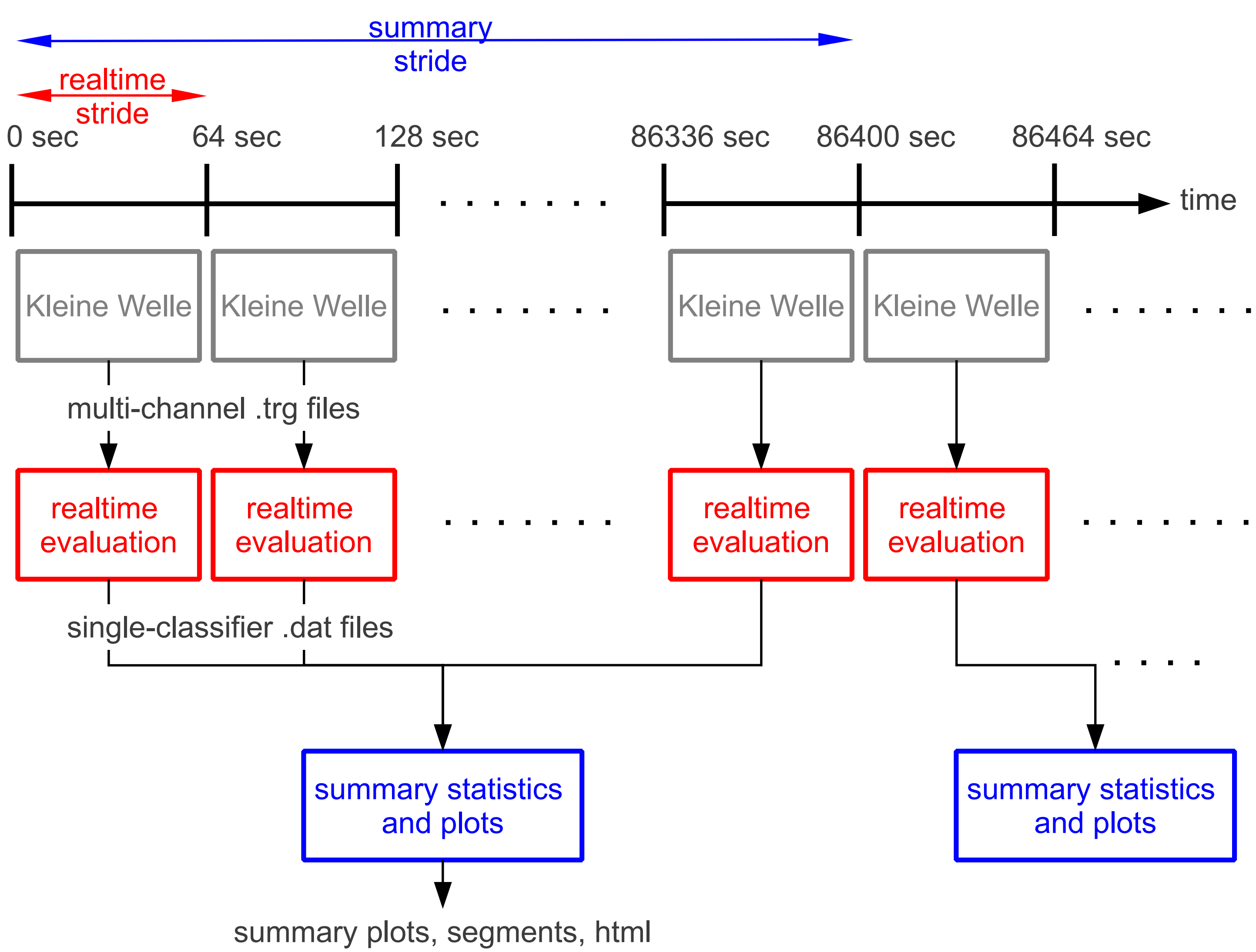


Goal

- Automate process of Glitch (non-gaussian noise transient) Identification in data preparation stage
  - Cleaning the data, monitoring, and feedback for commissioning or tuning
- Inclusion of different information which KW algorithm may not extract from raw data.
  - consideration of different trigger generation algorithm
- Real time (or low latency) analysis on Glitch Identification

Low latency Glitch Identification

- On-going project : iDQ pipeline
  - real-time (low latency) glitch identification and data quality (DQ) analysis using Machine Learning Algorithm

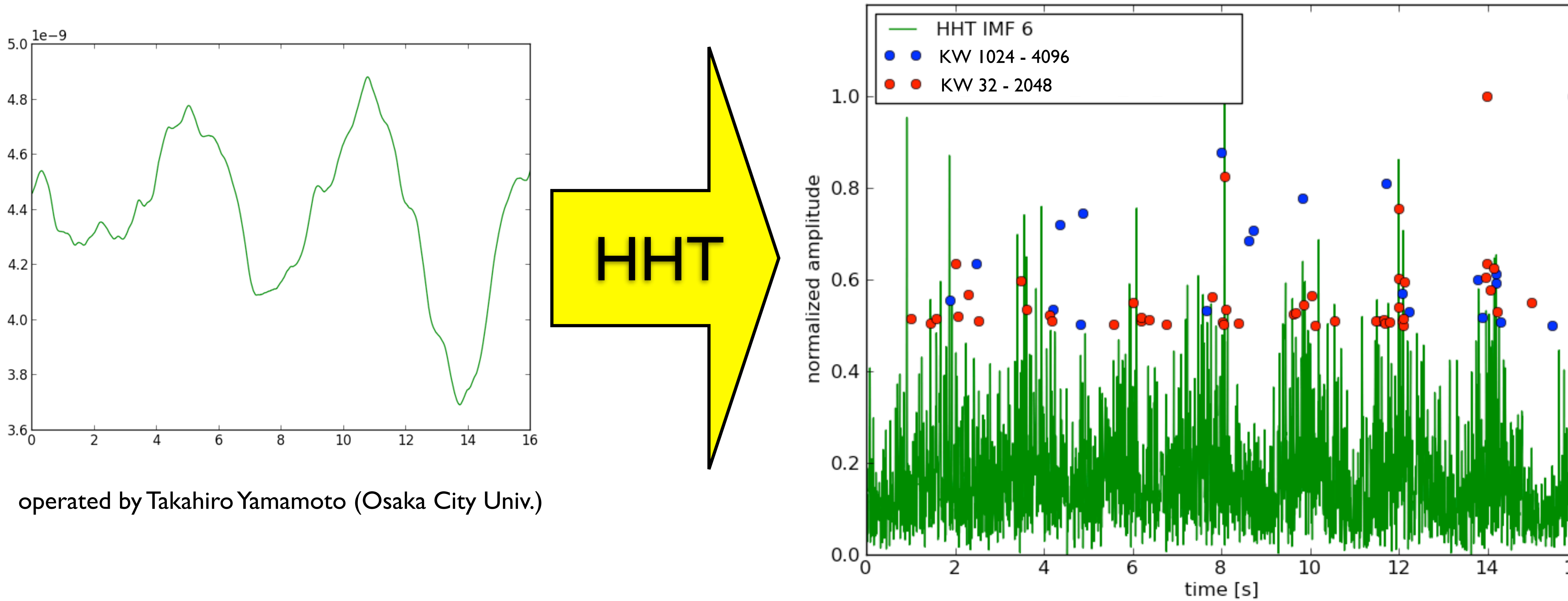


Collaboration with KAGRA

- Hayama's visit to NIMS (Feb. 2013)
  - Discussion on Research subjects for publication
  - Application of current methods to CLIO data. (if not, LIGO data as plan B.)
  - Multi-class MVC development
  - Develop a measure that measures channel's responsibility for glitches
  - Application of SOM to LIGO data
- Application of HHT to CLIO data

HHT on CLIO h(t) data

- HHT on CLIO h(t) --> 12 IMF



operated by Takahiro Yamamoto (Osaka City Univ.)

Summary

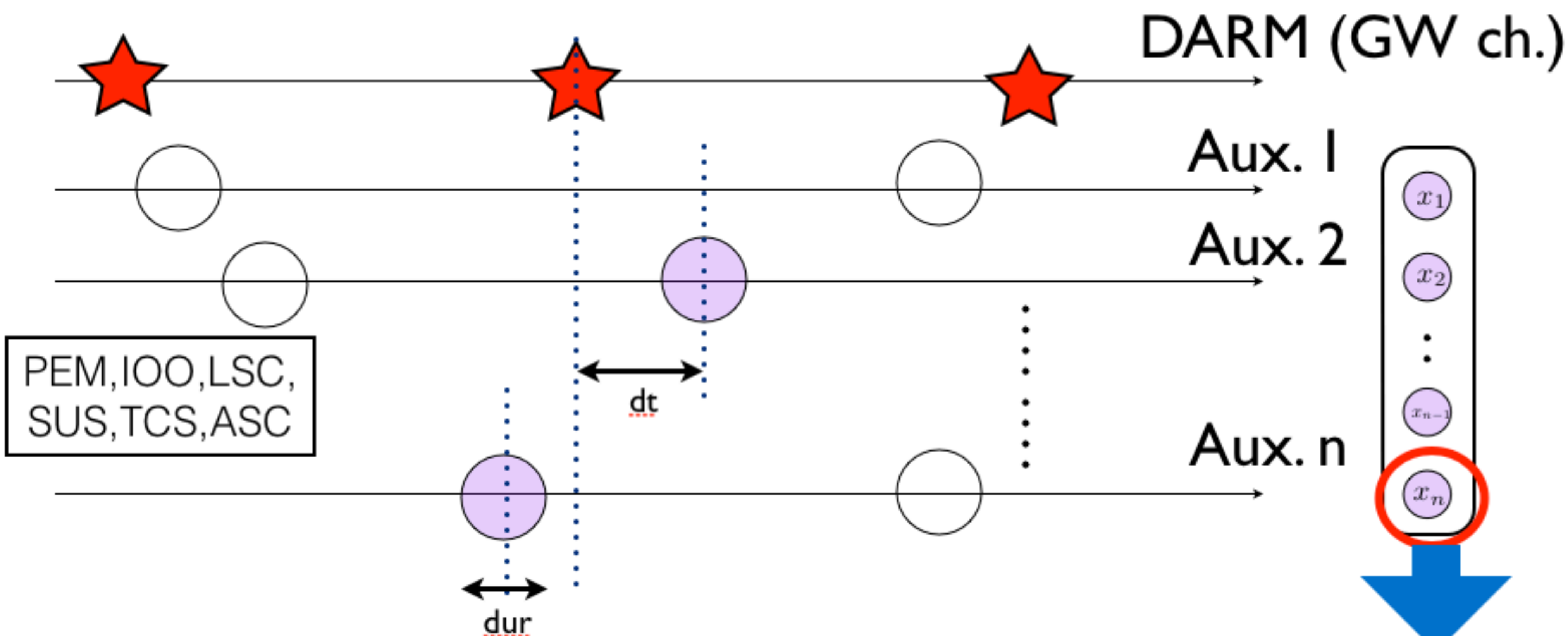
- Application of Machine Learning Algorithm (MLA) to Auxiliary channels data for glitch identification
  - auxmvc paper was submitted (arxiv:1303.6984)
- Artificial Neural Networks Ensemble was developed and applied to auxiliary channel data
  - At FAP 0.1%, efficiency is more than 50% for LIGO Livingston data.
  - Feature selection using t-statistic was studied. More investigation is needed.
- Trigger generation methods became more important.
  - We are studying HHT, KW algorithm, and other trigger methods.
- HHT application to CLIO data

Future works

- Genetic algorithm aided ANN (GANN) module optimization.
  - Network Topology Optimization
  - Investigation on a quantitative indicator or auxiliary channels to ANN rank
  - Implementation of GANN on computing accelerators (GPU, FPGA, etc.)
- Development of low latency analysis pipeline : iDQ pipeline, etc.
  - Implementation of GANN into iDQ
  - Access to online segments
  - Runs for Engineering Runs
- Application of MLA/GANN to different Trigger Generation Algorithm
  - Trigger generation scheme development for HHT
  - Application to Omega and/or GSTLAL triggers
- Application of HHT and MLA/GANN to CLIO data

Extra Slides

Data Preparation

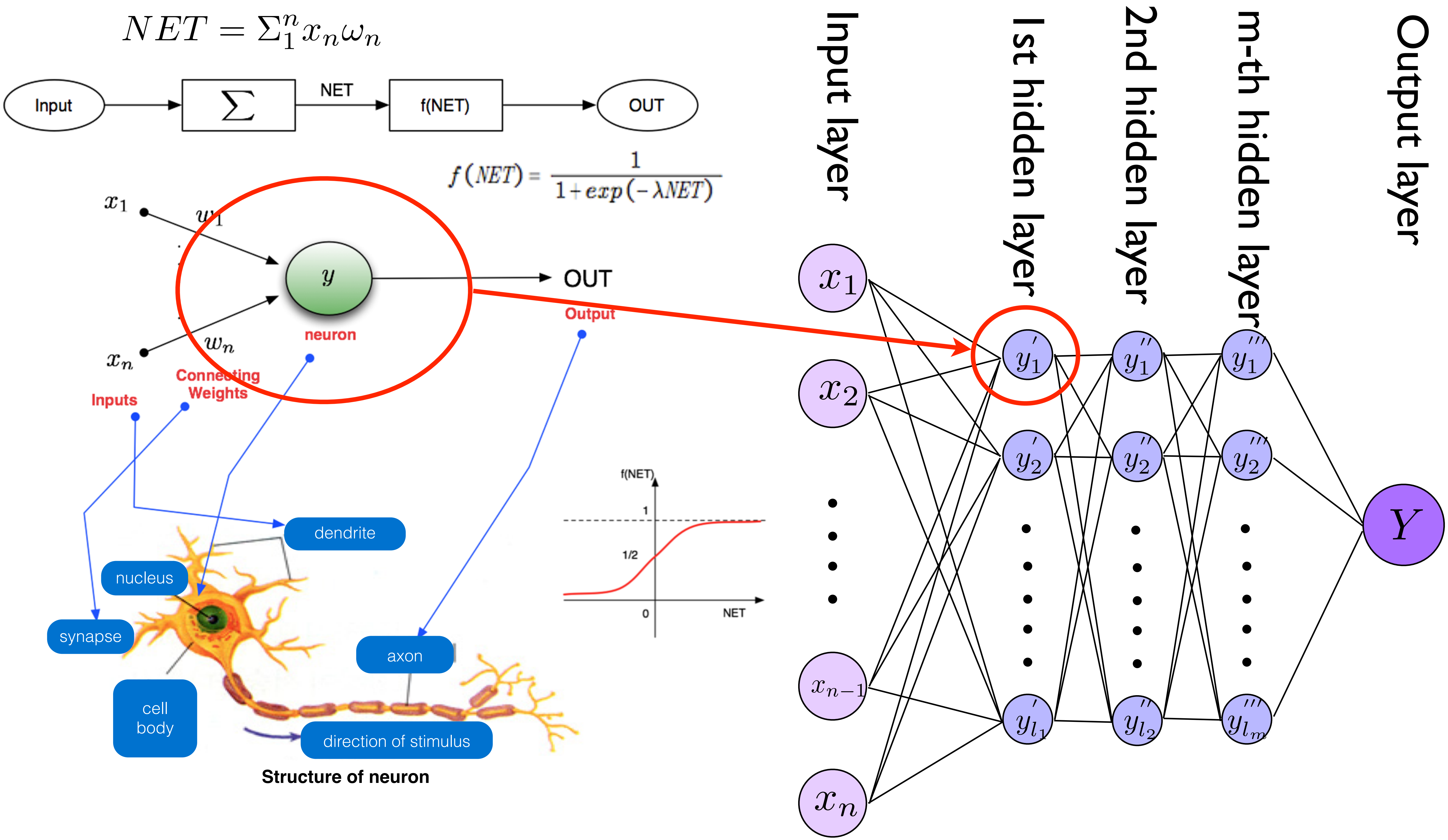


- Noise artifacts (called glitches) : generated by running KleineWelle (KW) over GW channel from the LIGO Livingston detector (LLO).
- Clean Samples : 10^5 randomly distributed times t_i within the live-time of detector operation.

(a) cluster significance of loudest trigger within +/-100ms, or significance of nearest trigger if none exists (significance is related to the SNR)
(b) dt between DARM trigger and AUX trigger in (a)
(c) duration of AUX trigger in (a)
(d) frequency of AUX trigger in (a)
(e) npts (cluster size) of AUX trigger in (a)

ANN - Basic Structure

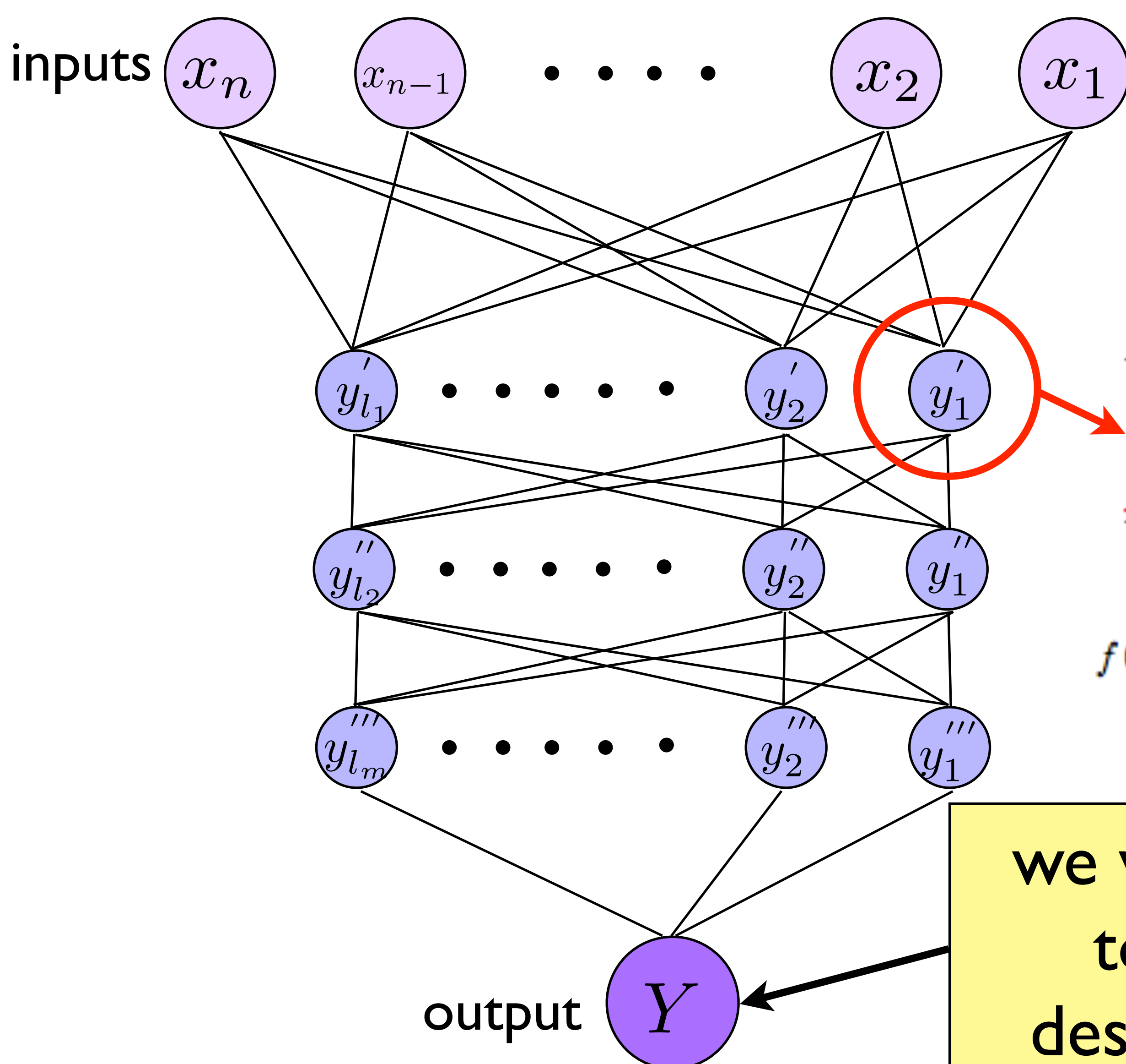
$$NET = \Sigma_1^n x_n \omega_n$$

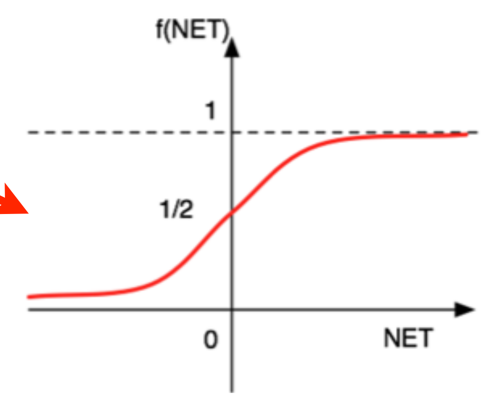


ANN - BackPropagation Algorithm

- How to get the good results in ANN ?
  - Make good network !! This means.......

All we have to do is to construct optimal sets of connection weights for each neuron from training.





$$f(NET) = \frac{1}{1+exp(-\lambda NET)}$$

connection weights

$$NET = \Sigma_1^n x_n \omega_n$$

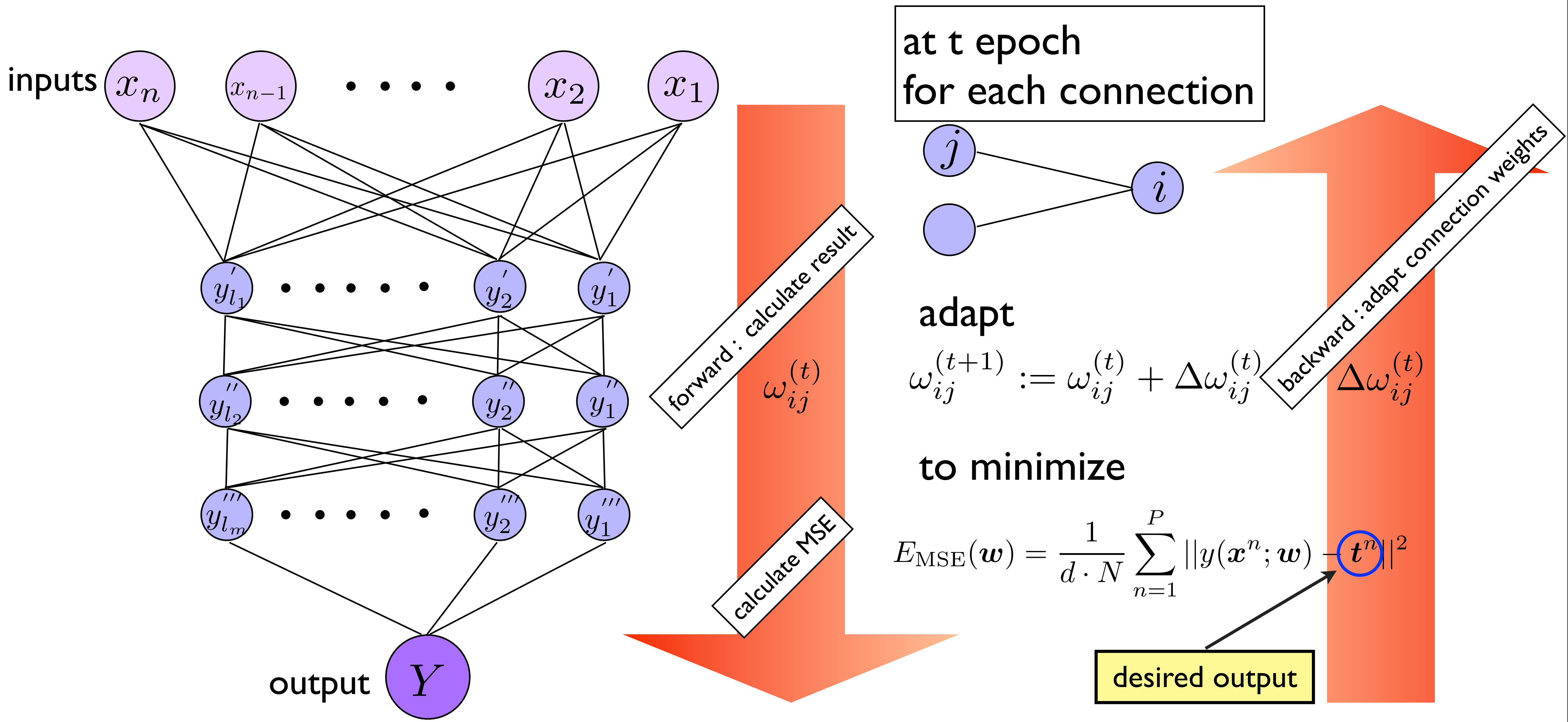
how to find optimal set ?

we want output to be close desired output

1 for glitches
0 for clean samples

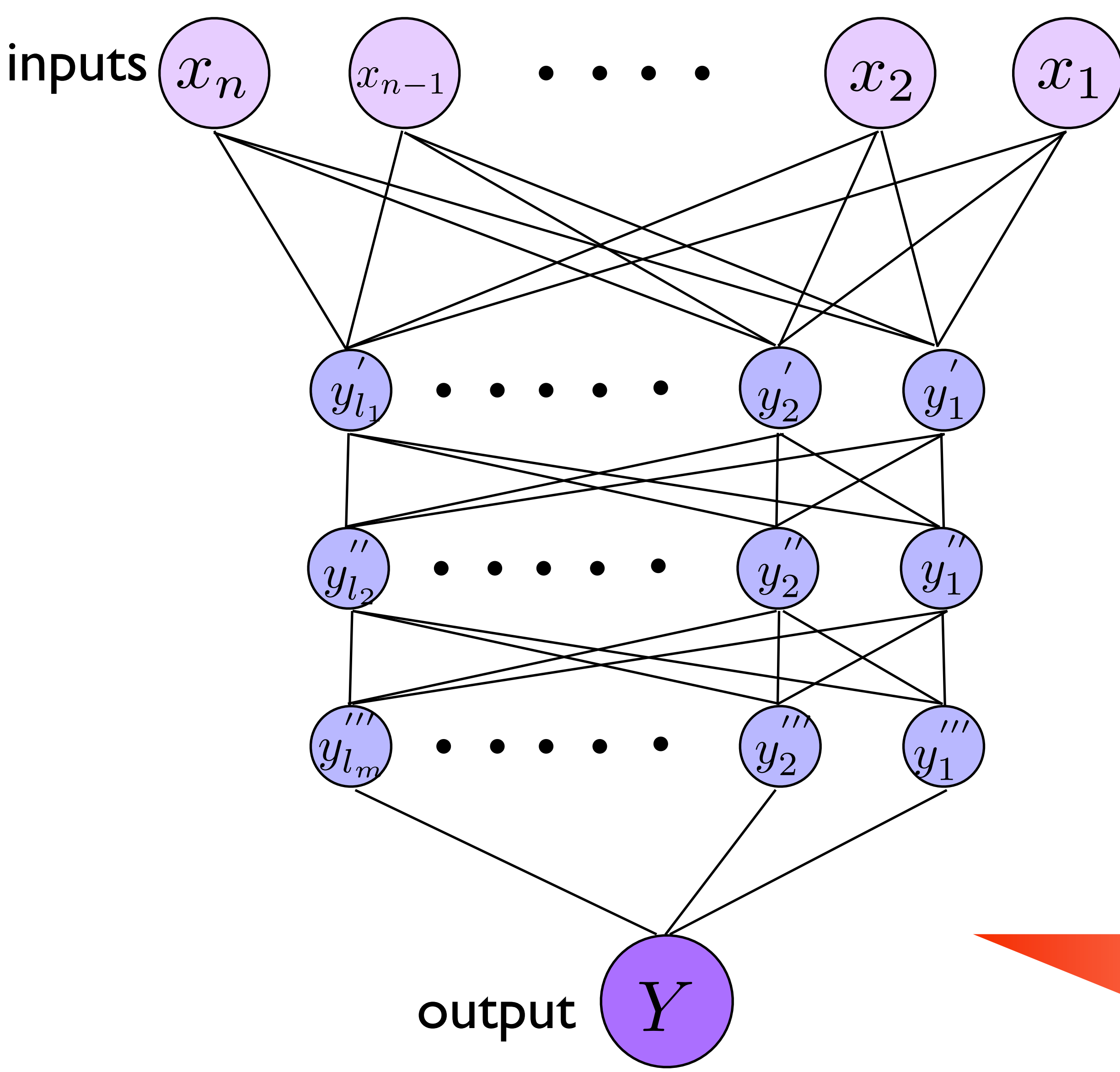
ANN - BackPropagation Algorithm

- In each epoch, the connection weights are calculated to reduce error between desired output and calculated output.



ANN - RPROP Algorithm

- Resilient backPROPagation algorithm (Igel & Hüsken, 2000)
  - improved version of original RPROP (Riedmiller&Braun,1992)



FANN library - iPROP

$$\Delta w_{ij}^{(t)} := -\operatorname{sign}\left(\frac{\partial E}{\partial w_{ij}}^{(t)}\right)\cdot \Delta_{ij}^{(t)}$$

$$\Delta_{ij}^{(t)} := \begin{cases} \min\left(\eta^{+}\cdot\Delta_{ij}^{(t-1)}, \Delta_{\max}\right) & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t-1)}\cdot\frac{\partial E}{\partial w_{ij}}^{(t)} > 0 \\ \max\left(\eta^{-}\cdot\Delta_{ij}^{(t-1)}, \Delta_{\min}\right) & , \text{ if } \frac{\partial E}{\partial w_{ij}}^{(t-1)}\cdot\frac{\partial E}{\partial w_{ij}}^{(t)} < 0 \\ \Delta_{ij}^{(t-1)} & , \text{ otherwise }, \end{cases}$$

$$\frac{\partial E^{(t)}}{\partial \Delta_{ij}^{(t-1)}} = \frac{\partial E^{(t)}}{\partial w_{ij}^{(t)}}\,\frac{\partial w_{ij}^{(t)}}{\partial \Delta_{ij}^{(t-1)}} = -\frac{\partial E}{\partial w_{ij}}^{(t)}\operatorname{sign}\left(\frac{\partial E}{\partial w_{ij}}^{(t-1)}\right)$$