Glitchology: an Elementary Perspective

Innocenzo M. Pinto

February 17, 2015
Outlook

- Glitch Busting
- Assembling Glitch Databases
- Glitch Enthomology
- Glitchy Noise Modeling
- A Simple, Physically Driven Model
- Proto Glitches, PCA, and Beyond
- Ongoing Work
Transient disturbances of environmental (exogenous, e.g. lightning) and/or instrumental (endogenous, e.g. laser) origin;

Appear *ubiquitously* in the data gathered by interferometric GW detectors, with a *wide range* of energies;

*Idiosyncratic* signals: exhibit a *wide variety of shapes*; still mostly visually *similar* and erratically recurrent.

Glitch *rate* roughly inversely proportional to glitch *strength*;

Important impact on instrument’s noise (*non-stationarity, heavy-tails*) -> glitches *spoil* naïve (Gaussian) detectors.

... word comes from Yiddish term גליטש
Glitch Sources

“External”
- Seismic activity
- Acoustic noise of various origin
- Electric/magnetic surge

Structural
- Vacuum pipe expansion/contraction (stepwise)

“Sub systems”
- Laser control system
- Thermal compensation system
- Cavity alignment

“sensors/actuators”
- Piezo driver malfunction
- Sensor/actuator malfunction
- Digital circuitry noise

Residual (unknown origin)

[J. Aasi et al., Class. Quantum Grav. 29 (2012) 155002]
Glitchology Goals

- Tracing out the origin of typical glitches families, and tweaking the machine design so as to suppress or mitigate them;

- Identifying surviving glitches in the GW channel, capitalizing on information from the instrumental / monitoring channels, and tagging/vetoing the data appropriately;

- Characterizing statistically the residual glitchy noise, and devising robust detection algorithms (noise modeling);

- Subtracting identified glitches from the GW channel (noise cancellation);
A. DiCredico et al., “Gravitational Wave Burst Vetoes in the LIGO S2 and S3 Data Analysis,” Class. Quantum Grav. 22 (2005) S1051.


N. Christensen et al., “LIGO S6 Detector Characterization Studies,” Class. Quantum Grav. 27 (2010) 194010.

F. Acernese et al., "Noise Studies During the First Virgo Science Run and After," Class. Quantum Grav. 25 (2008) 184003.

M. Del Prete et al., "Characterization of a Subset of Large Amplitude Noise Events in VIRGO Science Run 1 (VSR1)," Class. Quantum Grav. 26 (2009) 204022.


etc.
Old Style Vetoing

Check whether a trigger in the GW channel is *coincident in time* with a trigger in one (or more) AUX channel(s), *at a given significance level*.

(characterize preliminarily *accidental coincidence statistics via time slide experiments*).

[K.C. Cannon, LIGO P070085]
Use knowledge of the couplings (transfer functions) between the AUX channels and the GW one to check consistency between transients occurring in the GW and the instrumental channels.

Better efficiency, lower rate of accidental vetos.

[P. Ajith et al., PRD 76 (2007) 042004].
## Data Quality

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Prescription for analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT1</td>
<td>Flags obvious and <strong>severe malfunctions</strong> of the detector</td>
<td>Science data are redefined when removing CAT1 segments. Offline analysis pipeline should run on data only after removing CAT1 time periods.</td>
</tr>
<tr>
<td>CAT2</td>
<td>Flags noisy periods where the <strong>coupling between the noise source and the GW channel is well established</strong></td>
<td>Triggers should be removed if flagged by a CAT2 veto. A significant trigger surviving the CAT2 vetoes is a good candidate for detection follow-up.</td>
</tr>
<tr>
<td>CAT3</td>
<td>Flags noisy periods where the <strong>coupling between the noise source and the GW channel is not well established</strong></td>
<td>CAT3 vetoes should not be applied blindly. Triggers flagged by a CAT3 veto should be followed up carefully. CAT3 vetoes are applied to compute upper-limits.</td>
</tr>
<tr>
<td>CAT4</td>
<td>Flags <strong>time periods where hardware injections were performed</strong></td>
<td>Periods flagged by a CAT4 veto are used for specific studies. This category is applied for any search analysis.</td>
</tr>
<tr>
<td>CAT5</td>
<td>Advisory flag to <strong>keep track of problems for which no (or very low) impact was seen in the GW channel</strong></td>
<td>CAT5 are only used at the last stage of the detection follow-up. The validity of the flagging must be carefully checked by experts.</td>
</tr>
</tbody>
</table>

[F. Robinet et al., Class. Quantum Grav. 29 (2011) 155002]
DQ Figures of Merit

\[ d = \text{Dead-time (DQ-flagged fraction of science time)} \]

\[ UP = \text{Usable-percentage (fraction of DQ segs used to flag at least one trigger)} \]

\[ \varepsilon = \text{Efficiency (fraction of triggers which are flagged)} \]

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{BURST} & \text{CAT1} & \text{CAT2} & \text{CAT3} & \text{KW vetoes} \\
\hline
\text{UP, SNR > 5} & 75.5 \% & 87.5 \% & 73.1 \% & 91.2 \% \\
\hline
\hline
\text{d} & \text{CAT1} & \text{CAT1+2} & \text{CAT1+2+3} & \text{CAT1+2+3+KW vetoes} \\
\hline
\text{d, SNR > 5} & 0.8 \% & 5.0 \% & 8.1 \% & 8.2 \% \\
\hline
\text{\varepsilon, SNR > 5} & 1.2 \% & 16.5 \% & 22.6 \% & 24.2 \% \\
\hline
\text{\varepsilon, SNR > 8} & 5.4 \% & 61.8 \% & 74.6 \% & 76.3 \% \\
\hline
\text{\varepsilon, SNR > 15} & 23.8 \% & 86.4 \% & 88.6 \% & 89.2 \% \\
\hline
\end{array}
\]

\[ \varepsilon/d = 1 \leftrightarrow \text{flag is basically random; } \varepsilon/d > 1 \leftrightarrow \text{flag deemed as effective;} \]

A (Poissonian) probability of rejecting a true GW signal is attached to each DQ.

[F. Robinet et al., Class. Quantum Grav. 27 (2010) 194012]
We should refrain from throwing the baby out with the bath water...
Glitch Busting
Assembling Glitch Databases
Glitch Enthomology
Glitchy Noise Modeling
A Simple, Physically Driven Model
Proto Glitches, PCA, and Beyond
Ongoing Work
GW Interferometer Noise

Appears to consist of three basic components:

\[ n(t) = n_{NB}(t) + n_g(t) + n_{res}(t) \]

Narrowband component

“strong” (discernible) glitches

residual (floor) component

Besides the “good” expected Gaussian component, it may contain a locally Gaussian but breathing.
(slowly fluctuating variance)

None of these terms is described by a Gaussian - stationary distribution!
Naïve detection estimation algorithms based on a Gaussian Stationary noise-model will underperform!!
Simulating Glitch Noise

Let:

\( \Theta = [\tau, \tau + T] \) the time window;

\( \lambda_\Theta \) the glitch firing-rate; may fluctuate adiabatically (Cox process)

\( K_\Theta \) the random, (Poisson distributed, Hurwitz-Kač th.) no. of glitches in \( \Theta \):

\[
prob[K_\Theta = K] = \frac{(\lambda_\Theta)^K \exp[-\lambda_\Theta]}{K!}
\]

\( \{t_\Theta^{(k)} \mid k = 1 \ldots K_\Theta\} \) a set of \( K_\Theta \) random i.i.d. (uniform) firing times in \( \Theta \);

\( \psi(t, \tilde{a}) \) an elementary generic transient waveform \((atom)\) whose shape is set by the (vector) parameter \( \tilde{a} = \{a_1, a_2, \ldots, a_N\} \);

\( \{\psi(t - t_\Theta^{(k)}; \tilde{a}^{(k)} \mid k = 1 \ldots K_\Theta\} \) a set of atoms, with \( \{\tilde{a}^{(k)} \mid k = 1 \ldots K_\Theta\} \) a set of random parameters (with known distributions)

[M. Principe and I. Pinto, Class. Quantum Grav. 25 (2008) 075013]
A Glitchy Noise Model

The resulting stochastic process (generalized shot noise) has been introduced by David Middleton [D. Middleton, IEEE T-EMC-21 (1979) 209].

\[
n_g(t \in \Theta) = \sum_{k=1}^{K_{\Theta}(\Theta)} \psi(t - t^{(k)}_{\Theta}; \tilde{a}^{(k)}_{\Theta})
\]

Its characteristic functions in additive \( N(0,\sigma) \) noise can be computed to any order [D. Middleton, J. Appl. Phys. 22 (1951) 1143], e.g.

\[
F^{(1)}(\xi) = \exp \left\{ \lambda \Theta \left\{ \langle \exp[\i \xi \psi(t - \tau; \tilde{a})] \rangle_{(\tau; \tilde{a})} - 1 \right\} - \frac{\sigma^2 \xi^2}{2} \right\}
\]

Its PDF is also well approximated by a mixture of (a few) Gaussians with zero mean and different standard deviations.

[M. Principe and I. Pinto, Class. Quantum Grav. 25 (2008) 075013]
A Glitchy Noise Model, contd.

The statistical properties of the above glitchy noise model depend mainly on *a few gross parameters*: the product between the (local) glitch-rate \( \bar{\lambda}_T \) and the typical glitch time-width \( \Delta \psi \), and the *glitch contrast* against the residual (Gaussian) noise.

Strong glitches (which are *detectable* against the residual noise floor) occur with typical rates such that \( \bar{\lambda}_T \Delta \psi \leq 1 \). Their effect is making the noise distribution *heavy-tailed*.

Weak (undetectable) glitches occur at relatively *high rates*. In the limit where \( \bar{\lambda}_T \Delta \psi \gg 1 \), one can show that the glitch noise distribution *becomes Gaussian, irrespective of the individual glitch shapes*. High-rate weak glitches may originate the observed (slowly) non-stationary locally-Gaussian component, as an effect of (slow) fluctuations in \( \bar{\lambda}_T \) (Cox process).

[M. Principe and I. Pinto, Class. Quantum Grav. 25 (2008) 075013]
Robust algorithms for the detection of unmodeled GW transients in a network of (non-colocated) interferometers affected by glitchy noise have been discussed in [M. Principe and I. Pinto, LIGO-P1000134; expanded version subm. to PRD, 2014] Based on robust implementations of the locally optimum network likelihood ratio.
Improvement over LC

Theoretical ROC curve, Gaussian assumption, using actual 1st and 2nd order moments

\( \delta_h = 20, \quad \lambda = 0.5 \)

LOD (filtered data, filtered pseudo-template)

LOD (filtered data, raw pseudo-template)

LC (raw data, raw pseudo-template)

(LHO, best DOA)
Roadmap

DONE

Validate a quick reliable procedure for extracting training sets of “clean” glitches from (long) time series. **Key problem: robustness.**

Investigate “fine structure” of glitches (look for evidences of the fact that, in general, glitches are “multi-component”) → **skeleton representations.**

Investigate glitch (linear) decompositions using a *pre-determined* set of elementary “glitches”. Simplest choice: SG-atoms → **Matching-Pursuit representation of glitches using a Gabor-atom dictionary.**

Investigate glitch (linear) decompositions using an *adaptively-determined* set of elementary glitches → **Basis-Pursuit, K-means-SVD dictionary construction.**

**Estimate-and-subtract “strong” glitches from time series (BSS) → reduce vetoed fraction**

**Improve generalized shot-noise model of (residual) glitchy noise component → improve noise characterization.**
Outlook

Glitch Busting
Assembling Glitch Databases
Glitch Enthomology
Glitchy Noise Modeling
A Simple, Physically Driven Model
Proto Glitches, PCA, and Beyond
Ongoing Work
Enthomology: the study and understanding (Greek: λογία) of the insects (Greek: ἔντομοι), literally “entities that consist of (several) parts”
Glitch Enthomology Goals

Shedding light on the *fine-structure* of noise glitches, in preparation of their *classification*.

Identifying *multi-component* glitches, and isolating their *constituents*, so as to (hopefully) check back their physical origin;

etc (see roadmap) ...1
Natural choice for analyzing non-stationary (transient) signals;

All limited by Gabor resolution bound (frequency · time uncertainty) - to different extents, e.g.,

Short-Time Fourier Transform (worst of them all in terms of TF resolution);

Non-uniform TF-tiling linear transforms, including wavelet (e.g., Kleine-Welle, L. Blackburn, LIGO-T060221 etc), and constant-Q (e.g., Q-pipeline, S. Chatterji, LIGO-G060044 etc) transforms;

Wigner-Ville transform, best (in energy-preserving transform class) in terms of resolution ...

... but, being bi-linear, exhibiting signal-signal and signal-noise intermodulation artifacts which may affect its readability ...
Atomic Decompositions

[A. Altheimer, LIGO G0900045; M. Princip, CQG 26 (2009) 045003; N. Cornish and T. Littenbergh LIGO-P1000084]

Hilbert-Huang Transform

[A. Stroer et al., Class. Quantum Grav. 28 (2011) 155001; Phys. Rev. D79 (2009) 124022]
Signals can be “decomposed”, in principle, in an infinite number of ways into a finite or infinite number of “components”.

*Meaningful* definitions should identify multi-component signals such that their components have *different physical origins*.

Time frequency distributions suggest an *operational, intrinsic definition* of a $N$-component signal as a signal whose *effective* TF support consists of $N$ disjoint compact sets.

Note: by *effective* TF support here, we mean (following Bedrosian) the TF region where the signal energy exceeds the floor level due to measurement noise or representation accuracy.
QT became a std tool in GW data analysis (Qscan, Omega and Omicron LIGO-Virgo pipelines). QT can be regarded as template bank (in the whitened signal manifold), consisting of sine-Gaussians with different center time $t_0$, center frequency $f_0$, and quality factors Q. Typically, $Q \in (4, 64)$ and $f_0 \in (10^2, 10^3)$ Hz.
The Q-Transform

Project signal into time-shifted, time-windowed-sinusoids, whose time-widths are *inversely proportional* to their center frequencies.

\[
X_Q[m, k] = \sum_{n=0}^{N-1} x[n]e^{-i2\pi nk/N}w[m - n, k]
\]

Efficient computation in terms of D(W)FT,

\[
\tilde{X}[l] = \sum_{n=0}^{N-1} x[n]e^{-i2\pi nl/N}, \quad \tilde{W}[l] = \sum_{n=0}^{N-1} w[n, k]e^{-i2\pi nl/N}
\]

\[
X_Q[m, k] = \sum_{l=0}^{N-1} \tilde{X}[l + k] \tilde{W}[l, k]e^{-i2\pi ml/N}
\]


[S. Chatterji et al., Class. Quantum Grav. 21 (2004) S1809]
Features the *uniformly highest TF localization* among all unitary (energy-preserving, aka Cohen-class) transforms

\[ W_x(t, f) = \int_{-\infty}^{+\infty} \tilde{x}(t + \frac{\tau}{2}) \tilde{x}^*(t - \frac{\tau}{2}) e^{-i2\pi ft} d\tau, \text{ where: } \tilde{x}(t) = x(t) + i\mathcal{H}[x(t)] \]

In view of its bilinear nature, the WVT is plagued by *intermodulation artifacts*, except for two very special cases:

- **linear chirps**, \( f = f_0 + \beta t; \)
- **Gabor (SG) atoms**: \( s(t) = C \exp \left[ \left(\frac{t - t_0}{\delta t}\right)^2 \right] \cos[\Omega(t - t_0)] \)
WVT Artifacts

WVT of 2-atom Gabor molecule

WV of 0-PN Chirp, \( f = f(0)(1 - \frac{t}{t_c})^{-3/8} \)

Intermodulation artifact

Actual skeleton (IFL)

Intermodulation artifacts, single signal
Getting Rid of WVT Artifacts

Smoothing by use of an ambiguity-function (AF) adapted kernel [R. Baraniuk and D. Jones, Signal Proc. 32 (1993) 263] - Husimi / Choi-Williams can be seen as special cases; Suppress intermodulation terms (which map far from the origin in the AF plane) at the expense of a worse TF resolution

Re-assigment (re-squeezing) [P. Flandrin et al., IEEE T-SP 43 (1995) 1968] re-allocate value of smoothed WV at each TF point P to barycenter of the WV smoothed by kernel at P

Retrieving the TF skeleton of the WV from a reduced cardinality AF-data subset, using a sparsity constraint [P. Flandrin and P. Borgnat, IEEE T-SP58 (2010) 2974 ; P. Addesso et al., LIGO-P1200170]
TF Skeleton from WVT

Compute Wigner-Ville TF-representation of data \( W(t,f) \);
Compute related ambiguity function (2D Fourier transform of WV), call it \( A(t,f) \);
Construct Beraniuk-Jones optimal smoothing kernel for \( A(t,f) \);
Identify Beraniuk-Jones optimal smoothing kernel level curves;
Select the contour encircling \( \sim N \) samples (Heisenberg cardinality constraint);
Find the TF skeleton \( \Sigma(t,f) \) by solving the constrained optimization problem (sparse synthesis problem)

\[
\min_{\Sigma} \| \Sigma \|_0 \text{ subject to } \| \mathcal{F} (\Sigma) - A \|_2 \leq \varepsilon
\]

[P. Addesso et al., LIGO-P1200170; expanded version subm. to PRD (2015)]
Atomic Decomposition

Attempts to decompose a signal into TF “blobs” (TF atoms)

Simplest choice: Gabor (sine-Gaussian) atoms

\[ x(t \mid A_0, t_0, f_0, \tau_0, \psi_0) = A_0 \exp \left( -\frac{(t-t_0)^2}{\tau_0^2} \right) \cos(2\pi f_0 t + \psi_0) \]

(5 free parameters)

Matching Pursuit (MP) approximation (*greedy* algorithm)

5 param optimization

Atomic Decomposition, contd.

[A. Fusco, work in progress, 2014]
Naïve atomic decomposition shows atoms which are strikingly outside the time support of the glitch (dark-energy atoms). Must include a penalty functional in the greedy optimization algorithm.
Hilbert-Huang Transform

**Step 1** - “decompose” signal into Huang’s *intrinsic modes*

\[ s(t) = \sum_{k=1}^{N} a_k(t) \cos[\phi_k(t)] \]

- Signal has at least one max & one min (otherwise must be differentiated first);
- “time scale” defined by time lapse between extrema;
- At any point, mean value of the (local) max-envelope and (local) min-envelope is zero for each IMF;
- In each IMF the number of min, max and zero crossings does not differ by more than 1.


**Step 2** - Compute Hilbert–transform \((H)\) of Huang modes \(\rightarrow\) (sparse) TF coding

\[ A_k(t) \exp[i\psi(t)] = a_k(t) \cos[\phi_k(t)] + iH\{a_k(t) \cos[\phi_k(t)]\} \]

\[ \omega_k(t) = \frac{d\psi_k}{dt} \quad \text{(instantaneous frequencies)} \]

\[ A_k(t) \quad \text{(instantaneous amplitudes)} \]

**Step 3** - Encode signal into a set of ridges (*1D-TF features*):

\[ HHT[s] = \sum_{k=1}^{N} A_k(t) \delta[\omega - \omega_k(t)] \]
Huang empirical modes (original waveform in cyan)

Several “modes” exist (mostly!) outside the time support of the original waveform (e.g., mode #3). They give only a (destructive) interferential contribution (dark energy, similar to MP-based AD).
i) The IMF decomposition is not unique: there is an infinite number of IMF choices which will reproduce the same $s(t)$, differing in the way the non-stationarity is embedded in the AM, FM or both.

ii) The claim frequently made in the GW literature that "the HH method achieves infinite frequency resolution" is unsubstantial


iii) The numerical differentiation used to retrieve the instantaneous frequency is quite sensitive to noise. Boashash and coworkers compared the STFT, WD, RGK-smoothed WV, BD, HHT in the presence of noise for a number of representative waveforms. They found that the RGK-smoothed WV and the HHT gave the best and worst performance, respectively.

[N. Stevenson et al., IET Signal Proc. 4 (2010) 447]
QT vs WVT vs CS-WVT vs HHT, contd.

Gabor Molecule (beat)

HHT is fooled by the beat!!
A Multi Component Glitch (LIGO, S5)

more in [V. Pierro et al., Ligo-T1300598]
A Nonlinear (GEO 600) Glitch
Outlook

Glitch Busting
Assembling Glitch Databases
Glitch Enthomology
Glitchy Noise Modeling
A Simple, Physically Driven Model
Proto Glitches, PCA, and Beyond
Ongoing Work
Original formulation (CUSUM change detection test)

Let \( \{x_n| n = 1, 2, \ldots, N\} \) a time series; \( x_n \) has density

\[
\begin{align*}
    & f_0(x_n), \; n \leq n_0 \\
    & f_1(x_n), \; n > n_0
\end{align*}
\]

**Initialize**

\[
Z_0 = 0 \quad n = 1
\]

**Read** \( x_n \)

**Compute**

\[
Z_n = \max \left[ 0, Z_{n-1} + g(x_n) \right]
\]

**NO**

\[
n = n + 1
\]

**YES**

\( Z_n \geq \gamma \)

*declare change*

**How to choose \( \gamma \)?**

**optimal choice is log-likelihood**

\[
g(x_n) = \ln \left[ \frac{f_1(x_n)}{f_0(x_n)} \right]
\]
Page Test - from $N(0, \sigma_0)$ to $N(0, \sigma_1)$

$$f_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left( -\frac{x^2}{2\sigma_i^2} \right), \quad i = 0, 1 \quad \implies \quad g(x_n) = \frac{x_n^2}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) - \ln \left( \frac{\sigma_1}{\sigma_0} \right)$$

Define: $\frac{\sigma_1^2}{\sigma_0^2} = 1 + \rho$,

$$g(x_n) = \frac{1}{2} \left[ \left( \frac{\rho}{1+\rho} \right) \frac{x_n^2}{\sigma_0^2} - \ln(1+\rho) \right]$$

this is a sort of “contrast” (aka SNR)

Set maximum false alarm rate, $T_f^{-1} \quad \implies \quad T_f f_s = N_f = \text{min.} \# \text{ samples between f.a.}$

Compute threshold: $\gamma = \ln \left\{ -N_f E[g(x)|H_0] \right\} = \ln \left\{ \frac{N_f}{2} \left[ \ln(1+\rho) - \frac{\rho}{1+\rho} \right] \right\}$

Detector performance measured in terms of average delay in change-detection

$$N_d = \frac{\gamma}{E[g(x)|H_1]} = \frac{2\gamma}{\rho - \ln(1+\rho)}$$

... transients shorter than this will not be detected, on average ...

Note: $N_f$ is exponential in $\gamma$

$N_d$ is linear in $\gamma$

For fixed $\gamma$, length of shortest detectable transient scales inversely with SNR

$$N_d(\rho') < N_d(\rho), \quad \forall \rho' > \rho,$$
Example: change of variance occurring in finite interval

Page's statistic (aka, CUSUM)

... see discussion in:

Wang and Willett,

Wang and Willett,
IEEE AES Conf Proc 2005
Paper # 1536

Wang and Willett,

Easily robustified if $\sigma_0$ fluctuates:

$$\sigma_{0,r} = c\hat{\sigma}_0, c > 1$$

$$T_f(\sigma_{0,r}) < T_f(\sigma_{0,r})$$

Collaborations with Peter Willett’s group active since long  [V. Matta]...
**Fast**: 4 sec CPU
to sieve 4000 sec of data on a i7!

Note: *different* vertical scales, set to accommodate largest glitch in chunk

- $c=1.2$
- $F_{\text{inf}}=10$
- $F_{\text{sup}}=256$
- $N_C=4000$ sec
- $f_s=4096\text{Hz}$
- $f'_s=512\text{Hz}$
Outlook

Glitch Busting
Assembling Glitch Databases
Glitch Enthomology
Glitchy Noise Modeling
A Simple, Physically Driven Model
Proto Glitches, PCA, and Beyond
Ongoing Work
Disturbances
GW Interferometer as MIMO

inputs | outputs

GW

Disturbance entry ports

IFO

“GW” channel

“AUX” channels
In frequency (Fourier) domain:

\[
\begin{pmatrix}
  y_1(\omega) \\
  y_2(\omega) \\
  \vdots \\
  y_N(\omega)
\end{pmatrix} =
\begin{pmatrix}
  H_{11}(\omega) & H_{12}(\omega) & \cdots & H_{1M}(\omega) \\
  H_{21}(\omega) & H_{22}(\omega) & \cdots & H_{2M}(\omega) \\
  \vdots & \vdots & \ddots & \vdots \\
  H_{N1}(\omega) & H_{N2}(\omega) & \cdots & H_{NM}(\omega)
\end{pmatrix}
\begin{pmatrix}
  x_1(\omega) \\
  x_2(\omega) \\
  \vdots \\
  x_M(\omega)
\end{pmatrix}
\]

\[H_{pq}(\omega) = \text{transfer function between input-port } \#q \text{ and output-port } \#p\]
**Nonlinear Input/Output**


\[
y(t) = \int_{-\infty}^{\infty} d\tau h_1(t; \tau) x(\tau) + \\
+ \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 h_2(t; \tau_1, \tau_2) x(\tau_1) x(\tau_2) + \\
+ \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \int_{-\infty}^{\infty} d\tau_3 h_3(t, \tau_1, \tau_2, \tau_3) x(\tau_1) x(\tau_2) x(\tau_3) + \ldots
\]

- linear response
  - linear kernel, \( h(t) = \mathcal{F}^{-1} [H(\omega)] \)
- nonlinear response (quadratic term)
- nonlinear response (cubic term)
- nonlinear (quadratic, cubic, etc.) Volterra-Wiener kernels

... straightforward in principle...
Nonlinear Input/Output, contd.

Spectral forms (time-invariant systems)

\[ y(t) = \int_{\infty}^{\infty} df \ H_1(\omega)X(\omega) \exp(\omega t) + \]
\[ + \int_{\infty}^{\infty} \int_{\infty}^{\infty} df_1 df_2 H_2(\omega_1, \omega_2)X(\omega_1)X(\omega_2)\exp[i(\omega_1 + \omega_2)t] + \]
\[ + \int_{\infty}^{\infty} \int_{\infty}^{\infty} \int_{\infty}^{\infty} df_1 df_2 df_3 H_3(\omega_1, \omega_2, \omega_3)X(\omega_1)X(\omega_2)X(\omega_3)\exp[i(\omega_1 + \omega_2 + \omega_3)t] + \cdots \]

2D Fourier transform of \( h_2(\cdot) \)

3D Fourier transform of \( h_3(\cdot) \)

Spectral Volterra-Wiener kernels

\[ Y(\omega) = \sum_{n=1}^{\infty} Y^{(n)}(\omega) = \int_{\infty}^{\infty} df_1 \int_{\infty}^{\infty} df_2 \cdots \int_{\infty}^{\infty} df_n H_n(\omega_1, \ldots, \omega_n) \prod_{p=1}^{n} X(\omega_p) \delta \left( \omega - \sum_{q=1}^{n} \omega_q \right) \]
Nonlinear Glitches

Bi-linear glitch zoo table (LIGO)

A GEO600 “arch” glitch in TF plot (courtesy M. Was)
**Mickey-Mouse (Linear) Model**

- **Primary disturbance (spectrum)**
  \[ \mathcal{W}(\omega) \]
  - You can estimate with PEM

- **Disturbance as it reaches IFO entry port \( p \)**
  \[ X_p(\omega) = A_p \mathcal{W}(\omega) \exp(i \omega t_p) \]
  - Coupling amplitude and delay, both essentially random

- **Disturbance as it appears at IFO output port \( k \)**
  \[ Y_k(\omega) = \sum_{p=1}^{M} H_{kp}(\omega) X_p(\omega) \]
Disturbance as it appears at IFO output port # \( k \)

You can estimate from PEM

\[
Y_k(\omega) = \sum_{p=1}^{M} H_{kp}(\omega) A_p |\mathcal{W}(\omega)| \exp(i \omega \tau_p)
\]

Generally, the \( H_{kp}(\omega) \) are not available, because the IFO disturbance entry-ports are not accessible (and may be even un-identified).
Simplifying Assumptions

Linearity $\rightarrow$ Volterra-Wiener series (some formal complications)

Time-invariance $\rightarrow$ wide sense is OK (but update needed)

“Simple” (amplitude/delay) transfer functions from primary disturbance to IFO input ports $\rightarrow$ actual transfer functions can be incorporated in the $H_{pq}$

Many disturbances $\rightarrow$ straightforward
Wideband (Short) Disturbances

Disturbance as it appears at IFO output port \# k

\[
Y_k(\omega) = \sum_{p=1}^{M} H_{kp}(\omega)A_p \exp(i\omega \tau_p) \quad \leftrightarrow \quad y_k(t) = \sum_{p=1}^{M} A_p h_{kp}(t - \tau_p)
\]

Disturbances appear at IFO output ports as linear combinations (with random amplitudes and delays) of a finite (M) number of (wide sense) invariant waveforms, the \( h_{kp}(t) \).
All short glitches in the GW data channel are linear superpositions of the same $M$ functions

$$y_0(t) = \sum_{p=1}^{M} A_p h_{0p}(t - \tau_p) \quad \leftrightarrow \quad Y_0(\omega) = \sum_{p=1}^{M} H_{0p}(\omega) A_p \exp(i\omega \tau_p)$$

only $\{A_p, \tau_p \mid p = 1 \ldots M\}$ (amplitudes and delays) are different for each glitch.

If $\{h_{0p}(t) \mid p = 1 \ldots M\}$ and/or $\{H_{0p}(\omega) \mid p = 1 \ldots M\}$ were known, we could easily reconstruct, and then cancel by subtraction any/all glitches occurring in the GW data channel.

We call $\{h_{0p}(t) \mid p = 1 \ldots M\}$ the proto (aka, basic) glitches.
Glitchy GW Data

- focus on glitchy GW channel \((y_0)\) data;

- GW data containing (only) short glitches can be written

\[
Y_0(\omega) = \sum_{p=1}^{M} H_{0p}(\omega) \sum_{g} A_{pg} \exp(i\omega \tau_{pg}) = \sum_{p=1}^{M} \tilde{y}_p H_{0p}(\omega)
\]

sum over glitches

- Many independent glitchy data (realizations) are available containing putatively (provided the instrument is stationary during their collection time) the same \(h_{0p}(t)\)

\[
Y_0^{(i)}(\omega) = \sum_{p=1}^{M} \tilde{y}_{p}^{(i)} H_{0p}(\omega), \quad i = 1,2, \ldots, Q, \quad Q \gg M
\]
Glitchy GW Data, contd.

- Many (independent) glitchy data (realizations) available

\[ Y_0^{(i)}(\omega) = \sum_{p=1}^{M} \tilde{y}_p^{(i)} H_{0p}(\omega), \quad i = 1, 2, \ldots, Q, \quad Q \gg M \]

Known, measured \hspace{2cm} Unknown, sought \hspace{2cm} Unknown \ a priori

Unknown (different i correspond to different realization of the random array \( \{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_M\} \))

Questions

- Do the \( h_{0p}(t) \) proto-glitches form a kind of minimum-redundant dictionary for representing generic glitches?
- How many are they/how many do we need?
- How to extract the \( \{H_{0p}(\omega) \mid p = 1 \ldots M\} \) from glitchy datasets?

...Technically, a BSS - like problem...
Glitch Busting
Assembling Glitch Databases
Glitch Enthomology
Glitchy Noise Modeling
A Simple, Physically Driven Model
Proto Glitches, PCA, and Beyond
Ongoing Work
Principal Component Analysis (PCA) may answer the question “how many proto-glitches are there/how many do we need”.

Specifically, PCA may estimate the dimension of the manifold to which glitches belong.

Independent PCA implementations on different (LIGO) glitch datasets indicate that glitches live in a ~ 20 dimensional manifold, to within a < 5% energy ($L^2$) error.

Glitches are transient waveforms with time-limited support $T$, sampled at some frequency $f_s$. A glitch $g$ is thus represented by a point $g$ in Euclidean space $\mathbb{R}^N$, $N = f_s T$.

Let $D_g = \{g_i | i = 1, 2, \ldots, G\}$ our glitch dataset. Let $D_g$ the $N \times G$ matrix whose columns are the $g_i$ vectors;

1) “Standardize” matrix $D_g$, subtracting from each column its average:

$$\tilde{g}_{ij} = g_{ij} - \left(1/G\right) \sum_j g_{ij}$$

2) Compute covariance matrix $\Sigma$ among standardized column vectors:

$$\Sigma_{hk} = (\tilde{g}_h, \tilde{g}_k)$$

3) Diagonalize $\Sigma$, and enumerate the eigenvalues in order of decreasing (absolute) value. The corresponding ordered eigenvectors yield the PCA basis $\{\pi_i | i = 1, 2, \ldots, G\}$.

The magnitude of the PCA eigenvalues drops steeply beyond a certain order $N^*$. Correspondingly, the glitch energy is fully accounted by using only the first $N^*$ vectors from the PCA basis (addition of further terms does not improve $L^2$ accuracy to any sensible extent).

Hence, $N^*$ represents a sort of effective dimension of the manifold spanned by the glitch dataset.

PCA appears as a compressive coding where the vectors (glitches) $g_i = \{g_{im} | m = 1 \ldots N\} \in R^N$, $i = 1 \ldots G$, are encoded by the vectors $\alpha_h = \{ (g_h, \pi_k) | k = 1 \ldots N^* \} \in R^{N^*}$, $h = 1 \ldots G$, with $N^* < N$. 
Glitch Clustering in PC Space

After a glitch dataset \( \{g_1, \ldots, g_n\} \) has been represented in a nonredundant (e.g., PCA) “basis”, clustering algorithms can be used to identify families of similar glitches.

Several glitch clustering algorithms have been proposed/used, based, e.g., on proximity measures in (coarse)-feature space [S. Mukherjee et al., CQG 24 (2007) S701], longest-common subsequences [S. Mukherjee et al., J. Phys. Conf. Ser. 243 (2010) 012006], Kohonen self-organizing maps [S. Rampone et al., Int. J. Mod. Phys. C24 (2013) 1350085], and ANN [S.M. Kim et al., LIGO-G1201110]

Clustering goodness can be gauged using suitable metrics, e.g., the Davies-Bouldin (DB) index [IEEE T-PAMI, 1 (1979) 224]
For each cluster, define a measure of *concentration* (average distance of cluster members from cluster centroid)

\[
S_i = \frac{1}{T_i} \sum_{j=1}^{T_i} \| X_j - A_i \|_p , \text{ where } \\
i = 1 \ldots N_c \ (N_c = \text{number of clusters}) \\
A_i = \text{centroid of cluster-}i \\
T_i = \text{number of members in cluster-}i \\
p = 2 \ (\text{Euclidean distance}) \text{ usually}
\]

Also, define a measure of *cluster-to-cluster separation*,

\[
M_{ij} = \| A_i - A_j \|_p , \text{ where } \\
i, j = 1 \ldots N_c, \ i \neq j \ (N_c = \text{number of clusters}) \\
A_{i,j} = \text{centroid of cluster-}i,j
\]

Davies-Bouldin index of clustering goodness is:

\[
\rho_{DB} = \frac{1}{N_c} \sum_{i=1}^{N_c} \max_{j: j \neq i} \frac{S_i + S_j}{M_{ij}}
\]
Hints from Glitch Clustering

The number of clusters $N_c$ is usually found to be comparable to the embedding dimension $N^*$. In many cases, the cluster-centroid waveforms are found to be almost coincident with glitches whose origin is known, obtained by “hammering” some specific instrument input-port.

This is suggestive that glitch cluster centroids may correspond to the sought proto-glitches $h_i^{(0)}(t)$.

[I.M. Pinto, L. Troiano et al., Int. J. Mod. Phys. C24 (2013) 1350084]
Beyond PCA

• Arguably, the key requirement of a “natural” glitch dictionary $D_{\pi} = \{\pi_i | i = 1,2, \ldots, P\}$ would be to allow to represent each glitch in the dataset $D_{g} = \{g_i | i = 1,2, \ldots, G\}$ using a minimum number of elements from $D_{\pi}$.

• Let $D_{g}$ the $N \times G$ matrix whose columns are the $g_i$ vectors, $D_{\pi}$ the $N \times P$ matrix whose columns are the $\pi_i$ vectors, and $D_{\alpha}$ the $G \times P$ matrix whose elements are the representation coefficients $\alpha_{mn} = (g_{m}, \pi_{n})$. Then, the above requirement can be written:

$$\min \|D_{\alpha}\|_0 \text{ subject to } \|D_{g}^T - D_{\alpha} \cdot D_{\pi}^T\|_2 < \varepsilon$$

which is, technically, a sparse(st) coding problem.

• The above $L_0$ (constrained) optimization problem is $NP$-hard, but under broad assumptions, it can be transformed into a convex $L_1$ optimization problem [D.L. Donoho et al., PNAS 100 (2003) 2197].
Sparse Coding

• When the dictionary is *given in advance*, sparse coding can be implemented via “pursuit” algorithms, of which several flavors exist [D.L. Donoho et al., PNAS 100 (2003) 2197].

• In our case, we would like to find out the “best” or natural dictionary as part of our sparse(st) representation problem.

• This can be implemented efficiently by iterative algorithms that switch between sparse-coding (using a given dictionary), and adaptive dictionary-updating (based on clustering).

• Such algorithms (known as k-SVD) can be regarded as implementing a BSS on the given glitch dataset.

[Kreutz et al., Neural Comp. 15 (2003) 349;
Aharon et al., IEEE T-SP 54 (2006) 4311]
The Best Model of a Cat?

• It is further arguable that the sought “natural” dictionaries are closest to the speculated \( \{h_{0p}(t) \mid p = 1 \ldots M\} \) proto glitch sets.

In the words of Norbert Wiener, “the best model for a cat is another cat, or preferably, the same cat.”
Outlook

Glitch Busting
Assembling Glitch Databases
Glitch Enthomology
Glitchy Noise Modeling
A Simple, Physically Driven Model
Proto Glitches, PCA, and Beyond
Ongoing Work
... the ingredients of the impulse response of LTI systems

$$\psi(t; A, \theta, q, \alpha, \omega, \varphi) =$$

$$= AU(t - \theta)(t - \theta)^q \cdot$$

$$\cdot \exp[-\alpha(t - \theta)] \cdot$$

$$\cdot \cos[\omega t + \phi] \cdot$$

(6 free parameters, \( q \in \mathbb{N} \))
WiP - Mock Data Tests

• *Linear modes* (and linear combinations thereof) are the *simplest meaningful choice* for simulating *proto-glitches*;

• *Generate a (finite) dictionary of waveforms* $\mathcal{D}_g = \{ g_i | i = 1, 2, \ldots, G \}$ from random realizations of the *shape* parameters $(q, \alpha, \omega, \varphi)$ in the above linear modes,

• Generate $N > G$ time series $\{ s_k | k = 1, 2, \ldots, N \}$ featuring the above dictionary elements, *with random amplitudes* $(A)$ and *time-locations* $(\theta)$ (mock glitchy data).

• Add (stationary, white) Gaussian noise with given variance(s) to the above time series.

• Test BSS algorithms’ performance in retrieving the dictionary $\mathcal{D}_g = \{ g_i | i = 1, 2, \ldots, G \}$ from the mock data $\{ s_k | k = 1, 2, \ldots, N \}$

• Next step will be to use real (Virgo VSR4, LIGO S5) data.
ありがとうございました

Thank you for your kind attention and patience!
Next Move is up to You!

Intuition

Modeling

Experiment

 Formalization

you are here