A Method and Algorithm for Detecting GW Signals in Strong Noise

Meagan C. Patrick
New College of Florida

This is an exploration of the efficacy of single channel ICA (SCICA) when applied to a gravitational signal buried in strong noise. The main motivation for the research is for analyzing waves incoming from the Virgo ground-based interferometer in Cascina, Italy, where the signal to noise ratio is extremely low. The student research was done as part of an NSF International Research Experience for Undergraduates (IREU) through the University of Florida, in collaboration with the INFN and the University of Naples, "Federico II."

I. GRAVITATIONAL WAVEFORMS

Models of two astrophysical sources of GW are considered in this paper. Inspiraling binaries are modeled for simplicity as chirps, while pulsars are modeled for simplicity as sine waves.

While there have been efforts to describe the waveforms corresponding to gravitational sources, matched filters were shown to overestimate the signal-to-noise ratio (SNR) of stellar mass binaries [3]. When templates used for matched filters do not correspond accurately to the incoming signal, the signal may not be detected. Therefore it is of great importance to implement blind methods which do not rely on previous knowledge of the signal.

II. BLIND SOURCE SEPARATION

In order to search for GW signals without knowing the waveform, we use blind source separation (BSS). The method of BSS considered is Independent Component Analysis (ICA). ICA assumes only that the sources are statistically independent and that no more than one of the sources is Gaussian. ICA arose out of neural networks and information theory, and can be done either by negentropy maximization (ME) or minimization of mutual information (MMI). With blind source separation is that the probability density function (pdf) of the sources must be estimated without any information of the sources.

FastICA is one algorithm to implement ICA that uses a non-linearity g in place of the pdf and uses a iterative learning algorithm to maximize negentropy J.

$$J(y) = [E < G(y) > -E < G(v) >] \qquad (1)$$

where y and v are both of zero mean and unit variance. g is the derivative of G. The observed signals are represented as $X = AS$, where S is a matrix of the unknown source signals, and A is an unknown mixing matrix. The source signals are recovered by $S = A^{-1}X$, or $S = WX$, where W is the inverse of A. W is randomly chosen and then updated in iterations so that $w^T x$ maximizes the negentropy J.

The following is the FastICA one-unit algorithm:

1. Choose random weight vector w.
2. Let $w+ = Exg(w^T x) - Eg'(w^T x)$
3. Let $w = w + /||w + ||$
4. If not converged, go back to 2.

Decorrelation must be performed after each iteration to ensure that $w^T x$ converges to different maxima. Convergence is achieved when old and new values are parallel, and can be done through symmetric decorrelation or deflation.

EFICA (Efficient FastICA) is an improvement on the FastICA algorithm that uses a different nonlinearity for each signal in an attempt to reach the Cramer Rao bound. The EFICA algorithm has been charted with a flowchart in Figure 1.

Here is an example of how EFICA is able to better recover four deterministic signals. The original figures can be seen in Figure 2. While neither recovered perfectly all four signals, EFICA in Figure 3 far outperforms FastICA in Figure 4.

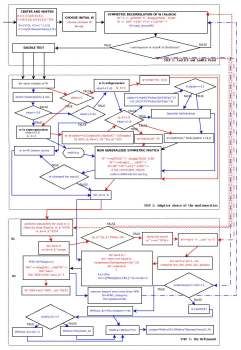

FIG. 1: EFICA Flow Chart

III. TIME DELAY EMBEDDING

Taken's embedding theorem states that if you are able to observe a scalar quantity of dynamic variables then the geometric structure may be unfolded in phase space composed of vectors that are components of the scalar observation applied to powers of the dynamic random

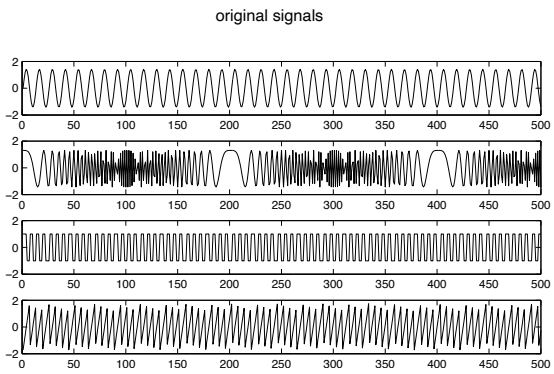

FIG. 2: original signals

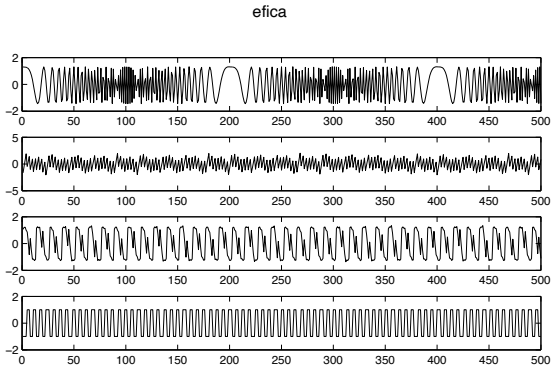

FIG. 3: efica signals

variable. The observation and the random variable must both be sufficiently smooth and the euclidean dimension is sufficiently large [4].

$$\Phi_T(x) = (\alpha(x), \alpha(f(x)), ...\alpha(f^{k-1}(x))) \qquad (2)$$

In time delay embedding, the dimensions are formed by time delays of the random variable.

$$\Phi_T(x) = (\alpha(x), \alpha(x+\tau), \alpha(x+2\tau), ...\alpha(x+N\tau)) \quad (3)$$

When a signal is properly embedded it can be seen clearly without overlaps.

A. Time Delay τ

Taken's embedding theorem does not specify how to choose τ, though it must be a multiple of the sampling frequency. A change in $/tau$ is very important as it is amounts to a change in coordinates, and so it must be chosen before the minimum embedding dimension. If τ is too large there is not information supplied by the time delay, while if τ is too small then not enough information is

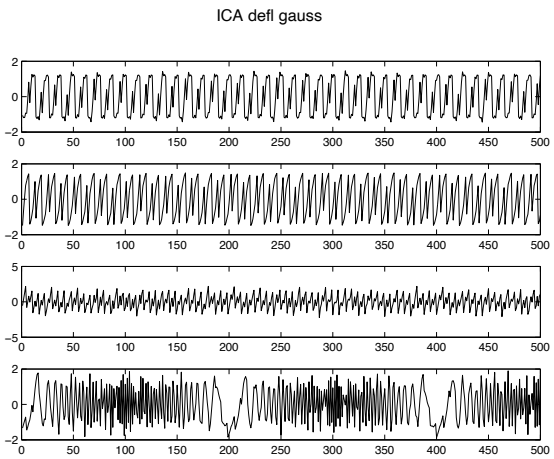

FIG. 4: fastica signals

supplied. Treating the average mutual information (MI) of the data as an improvement on the non-linear autocorrelation function, τ can be found as the first minimum of the MI [4]. In Figure 6 can be seen the graph of the mutual information of a univariate chirp using the `amutual` function provided with TSTOOL. The chirp is plotted in euclidean 2-d space in Figure 5.

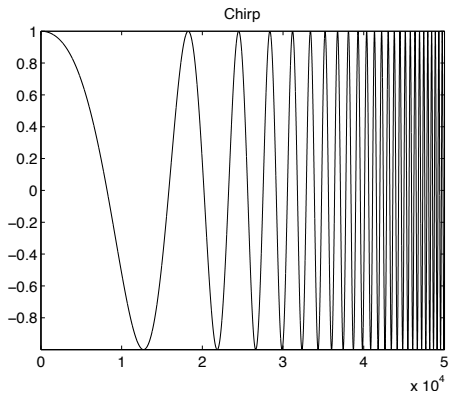

FIG. 5: chirp in euclidean space

B. Embedding Dimension

If the data is clean then it can be embedded in infinitely high dimensions. However, the data from ground-based interferometers is highly contaminated by formally infinitely dimensional noise. Therefore, it is important to embed the signal in the lowest dimension in which the signal is properly embedded. The minimum embedding dimension is found by using the method of false near-

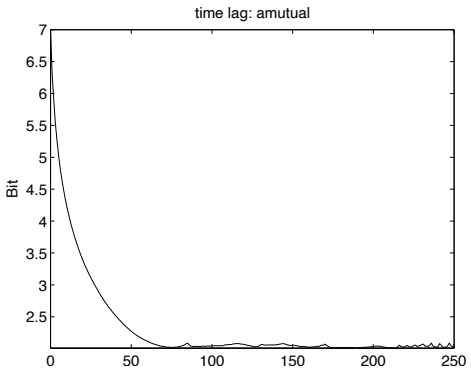

FIG. 6: amutual: time lag τ

est neighbors. An improvement on this method is the Cao method, which is included in the TSTOOL package. Figure 7 can be used to find the minimum embedding dimension, which is subjectively chosen as the point at which the false nearest neighbors levels off.

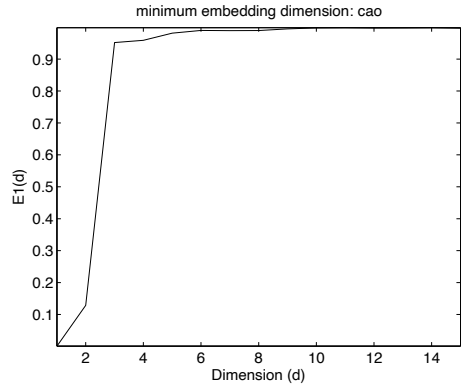

FIG. 7: cao: minimum embedding dimension

In Figure 8 it can be seen that 2-dimensions is not enough to embed, while in Figure 9, 3-dimensions is shown to be large enough to prevent overlaps and properly embed.

IV. LIMITS OF SCICA

A. Noise Limits of Sine and Chirp Signals

For the sine wave,the recovery threshold found in these trials for the SNR was -29 decibels, though it could be lower. The limit was tested using EFICA, and the signal was mixed with noise using a random mixing matrix. At -29 decibel SNR, the signal was not successfully embedded in 3-d, meaning it had overlaps, but the mutual information plot was periodic.

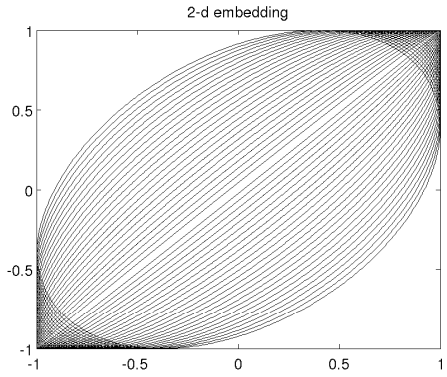

FIG. 8: chirp embedded in 2-d phase space

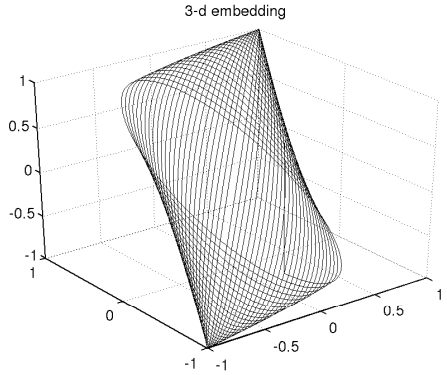

FIG. 9: chirp embedded in 3-d phase space

For the Chirp wave, the recovery threshold found in these trails for the SNR was -12 decibels for both FastICA and EFICA, though it could be lower. More trials must be done.

The recovery for FastICA and EFICA was the same for chirps, and EFICA did not demonstrate an advantage over FastICA. For the sine wave, however, more signals were outputted using EFICA than FastICA. This is most likely because FastICA selectively disregards certain sig-

TABLE I: Sine Wave mixed with Noise with Random Mixing Matrix: Signal Recovery Rates with EFICA

Coeff of noise #	$SNR(decibels)$	$RecoveryRates$
.01	37	9/9
.1	17	10/10
1	-3	6/8
10	-23	11/32
20	-29	2/9
30	-32.5	0/9
50	-37	0/9

TABLE II: EFICA vs. FastICA Recovery Rates for Chirp

Coeff. of noise #	$SNR(decibels)$	$Fastica$	$EFICA$
.01	27.9	6/6	6/6
.1	7.9	6/6	6/6
1	-12	6/6	6/6
2.5	-20.1	0/2	0/2
5	-26.1	0/2	0/2
7	-29.6	0/2	0/2
10	-32	0/5	0/5
20	-38.1	0/5	0/5

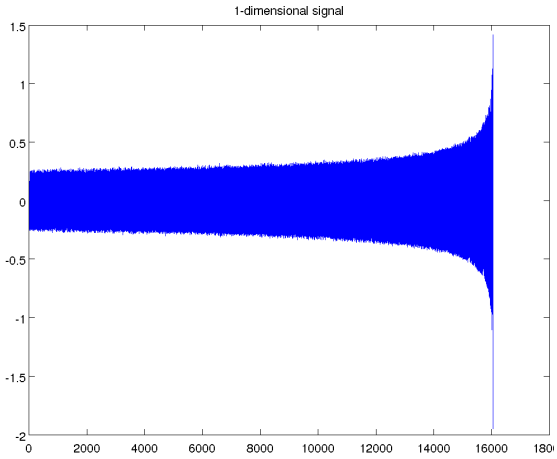

FIG. 10: Dirty Chirp (coefficient of noise is .01)

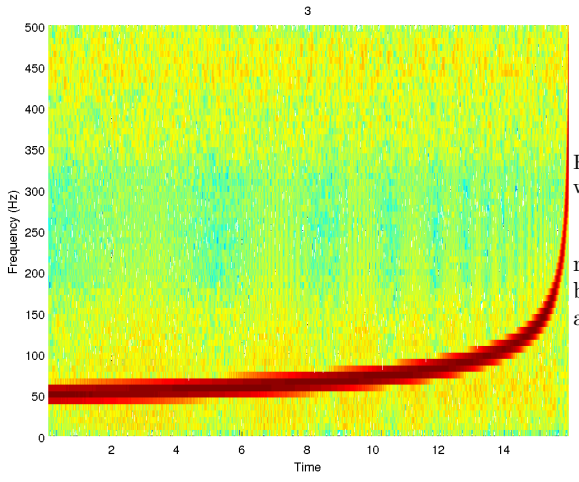

FIG. 11: Spectrogram of Chirp recovered by FastICA (coeff. of noise of .01)

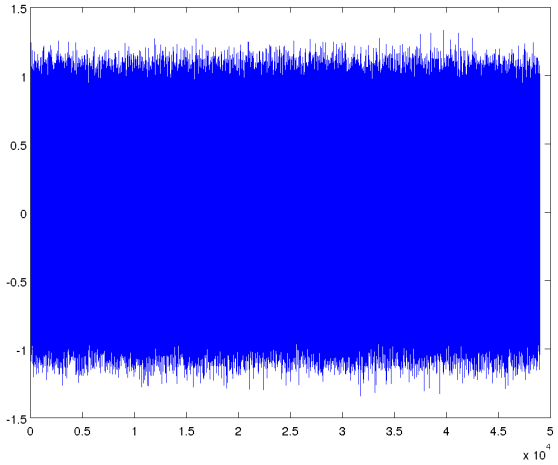

FIG. 12: Dirty Sine (coefficient of noise is .01)

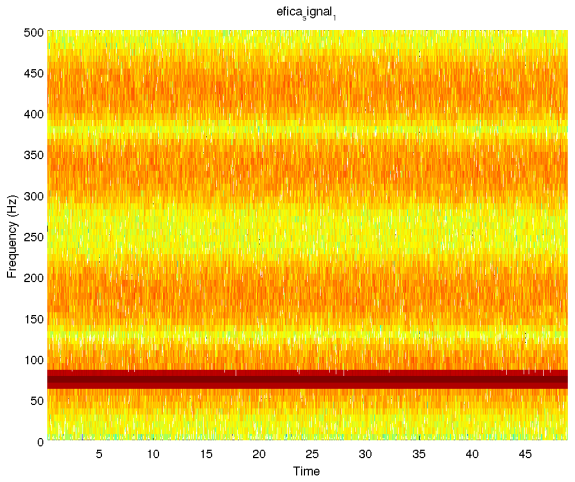

FIG. 13: EFICA Spectrogram of Sine Wave linearly mixed with Noise

nals beforehand, while EFICA does not. EFICA may be better for Chirp signals because of its efficiency, as it is approximately 3 times faster than FastICA.

B. 3-d embedding

Contamination by noise can prevent the signal from being correctly embedded in the appropriate dimensions, as seen in the following figures. To better the chances of recovery of a signal, one can view the signal in 3-d and then decide whether or not to choose a different time lag

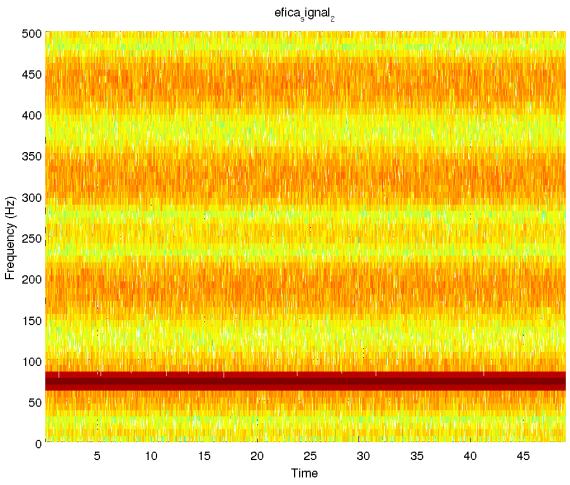

FIG. 14: EFICA Spectrogram of Sine Wave linearly mixed with Noise

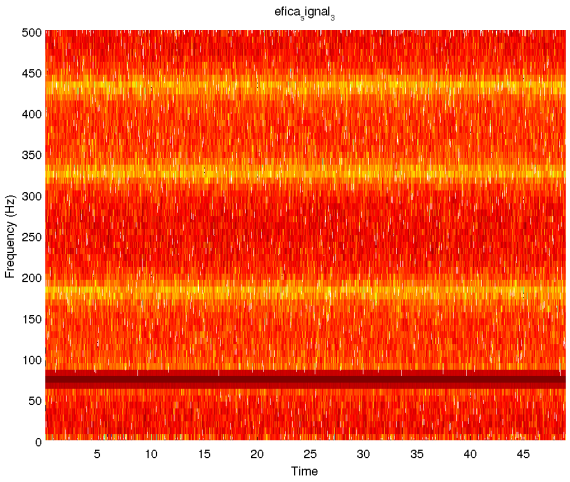

FIG. 15: EFICA Spectrogram of Sine Wave linearly mixed with Noise

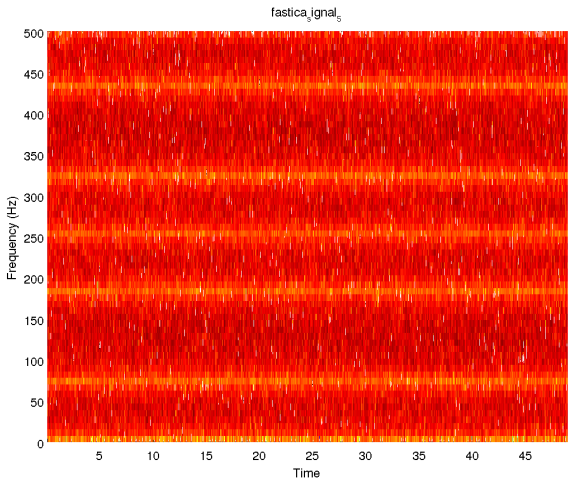

FIG. 16: FastICA Spectrogram of Sine Wave linearly mixed with Noise

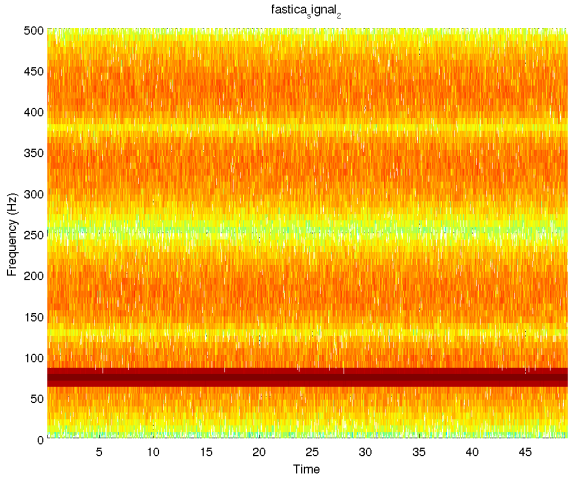

FIG. 17: FastICA Spectrogram of Sine Wave linearly mixed with Noise

and dimension.

C. Signal-to-Interference Ratio (SIR)

The SIR is a feature of EFICA in which it measures how much of one signal is contained in another. In theory it is a good measure of recovery of the signals, in the matrix representing SIR in grey tone, but in practice lighter values, which should correspond to less contaminated signals, did always not correlate to recovered signals.

D. Ways of Determining SNR

To better predict the recovery rate of a signal, we experimented with using a matched filter to determine the SNR.

E. Classifying the Signals

The signals in this report were classified by analysis of their spectrograms, normplots, histograms, and empiri-

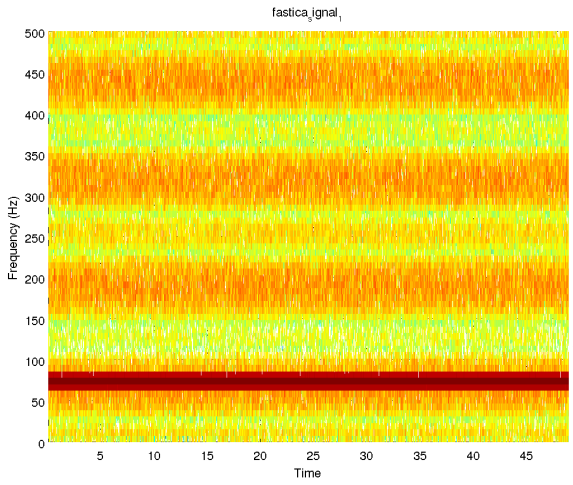

FIG. 18: FastICA Spectrogram of Sine Wave linearly mixed with Noise

cal cumulative distributions. In order to go to the next step, which is to find the probability of signal recovery given a specific SNR, it is necessary to both automate the procedure and make it more objective. One way of determining the presence of a signal without human help is with kurtosis. Since a Gaussian signal has a kurtosis of 3 the presence of a deterministic signal can be uncovered by checking whether the kurtosis falls outside of a certain margin away from 3.

V. CONCLUSION

SCICA alone has proven to be insufficient for a signal embedded in strong noise. Other methods must also be applied, including Principal Component Analysis. Future work should look for ways to automate the SCICA procedure so that the probabilities of signal recovery can be accurately determined.

VI. ACKNOWLEDGMENT

Manyy thanks to Professor Leopoldo Milano, Dr.Luca Forte and assistant professor Fabio Garufi for their assistance to my work, during my internship in Naples at the University of Naples, "Federico II." I would also like to acknowledge Professor Bernard Whiting and Professor Guido Mueller, who were very helpful as organizers of the IREU.

[1] M. Maggiore, "Stochastic backgrounds of gravitational waves," arXiv:gr-qc/00080271, (2000)
[2] Scott A. Hughes, "Gravitational waves from merging compact binaries," arXiv:0903.4877v2, (2009)
[3] C. Van Den Broeck. "Binary black hole detection rates in inspiral gravitational wave sources." Class.Quantum Grav. 23 L51-L58: arXiv:gr-qc/0604032v2 (2006)
[4] Henry D.I. Abarbanel. *AnalysisofObservedChaoticData*, Q172.5.C45A23, 1997.

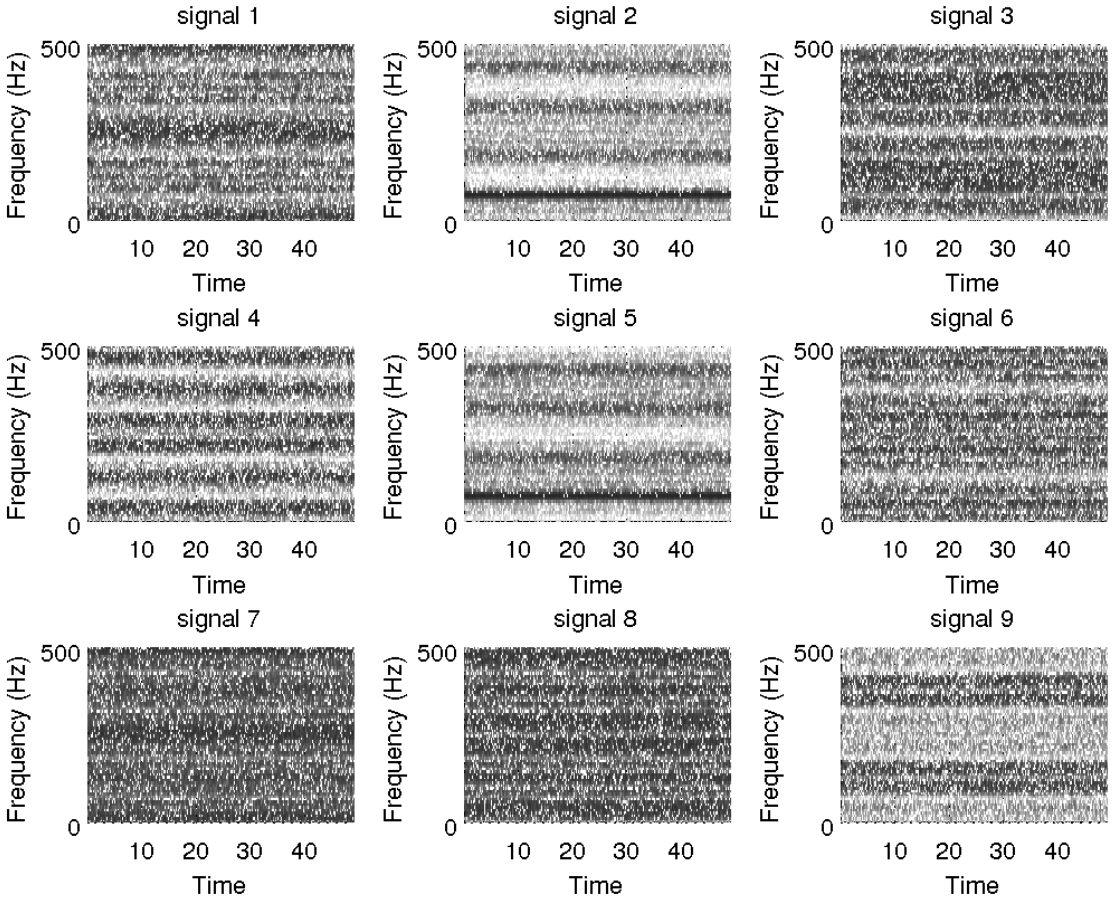

FIG. 19: -29 decibel SNR, sine recovery at 70 hz

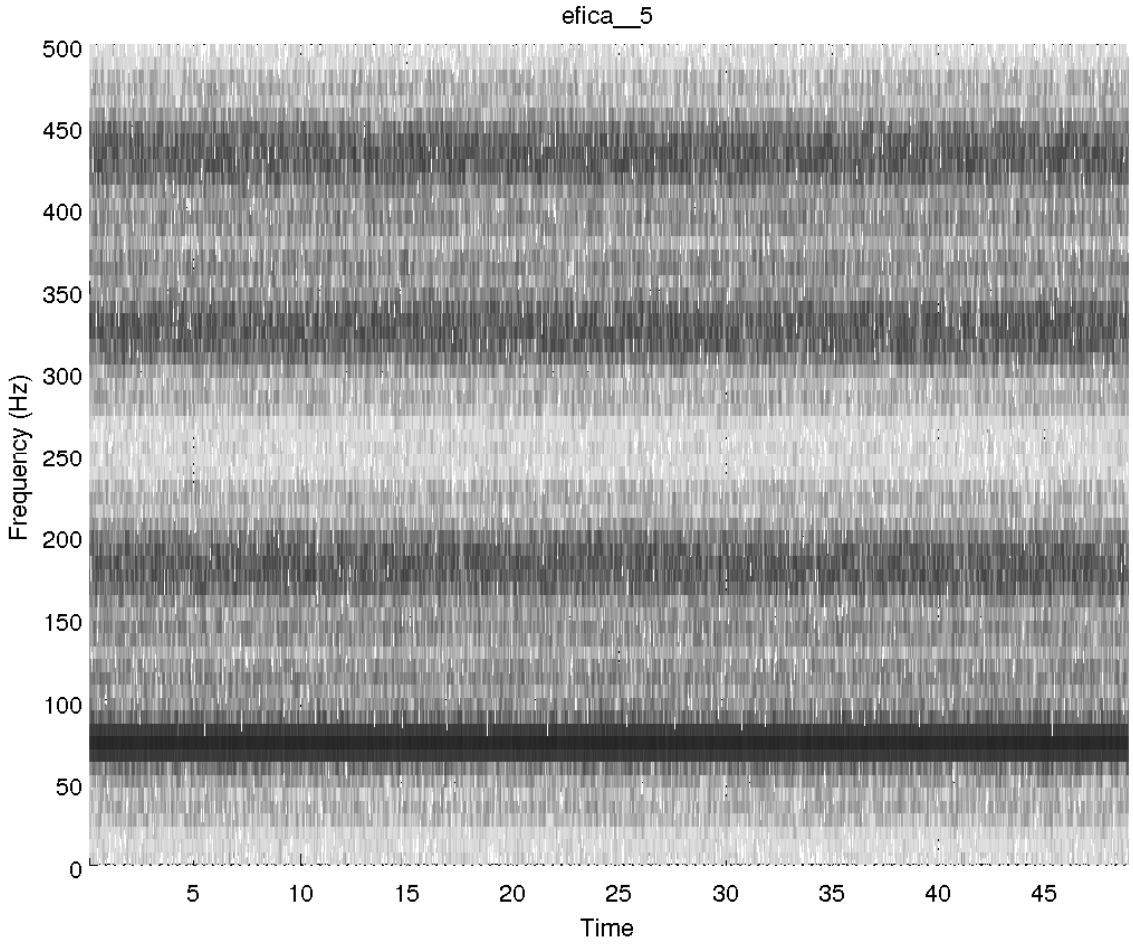

FIG. 20: -29 decibel SNR, sine recovery at 70 hz

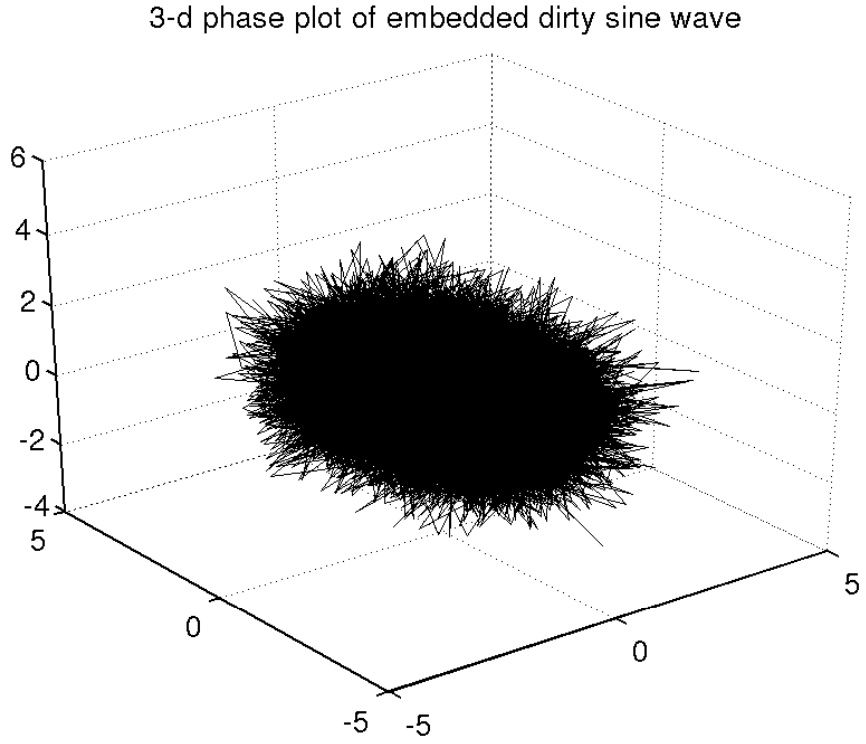

FIG. 21: -29 decibel SNR, not embedded in 3d

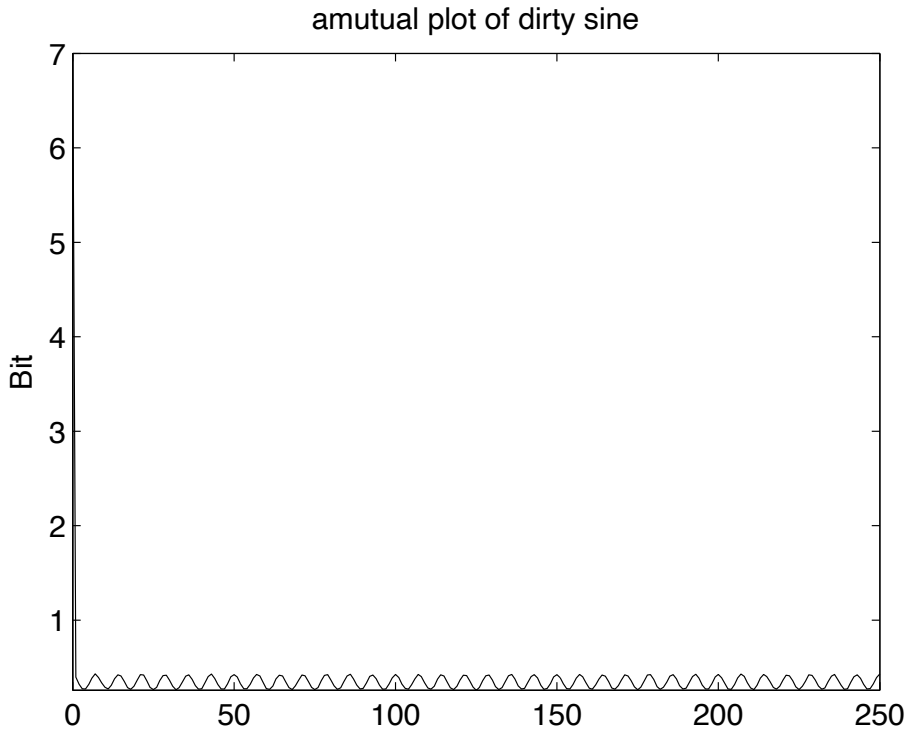

FIG. 22: -29 decibel SNR, amutual graph is periodic

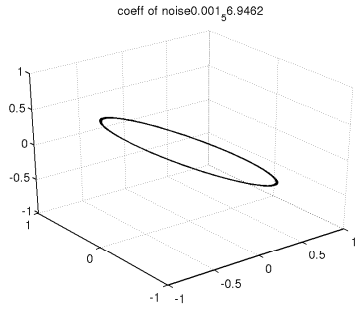

FIG. 23: 3-d embedding of sine wave with 56.95 decibel SNR

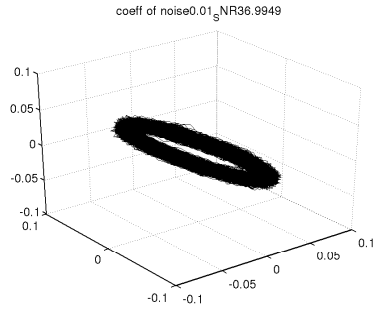

FIG. 24: 3-d embedding of sine wave with 36.99 decibel SNR

FIG. 25: 3-d embedding of sine wave with 16.97 decibel SNR