Chapter 11

Future Improvements for Compact Binary Coalescence Searches

In previous chapters we focused on the procedures used for the Search for Low Mass CBCs in the First Year of LIGO's Fifth Science Run (S5) Data, referred to as "this search," including data analysis improvements that were made based on the collective experience of analyzing previous science runs' data. In this chapter, we focus on such improvements that are yet to be made, or in the process of being implemented within the LSC CBC Group that will further benefit future searches. These include improvements in separating signals from noise with better detection statistics (section 11.1), better estimating the background of a particular search (section 11.2), coherently combining the data from multiple detectors (section 11.3), cleaning the data before analysis to remove detector glitches (section 11.4), and lowering the latency of the searches to improve the scientific output (section 11.5).

11.1 Detection Statistic

The goal of developing new detection statistics is to get to the *ideal detector response to Gaussian noise*. We are already close to this ideal for the low mass search, and in particular the low mass portion of the low mass search. This is because the low mass waveforms a more broadband than the high mass waveforms, which makes the standard SNR calculation relatively insensitive to detector glitches. As was seen in this search's waveforms, including information about the background noise of the detectors allowed us to develop a detection statistic that prevents the contamination of the low mass portion of parameter space by the inclusion of the high mass portion.

The addition of the background information can be seen as a step toward defining a likelihood ratio for the detection statistic. The likelihood ratio Λ is the ratio of the probability that a particular trigger c was associated with a true signal $P(c|h)$ and the probability that a particular trigger was due to background noise alone $P(c|0)$ given by

$$\Lambda = \frac{P(c|h)}{P(c|0)} \ . \tag{11.1}$$

As we saw in section 9.3.2, the FARc detection statistic can be converted to a FAP (i.e., the probability of getting any background triggers louder than that trigger) using equation (7.6a).

The final step toward a likelihood ratio detection statistic is calculating $P(c|h)$ using the injected signals. This procedure is being demonstrated to gain additional sensitivity in the triggered search for CBC signals associated with short GRBs [142]. The reason for this is that including the $P(c|h)$ term introduces more information that we can use to distinguish noise triggers from signal triggers. However, there is a downside of introducing this term. As in calculating $P(c|0)$ where included information about how the pipeline responded to noise in the GW channel, for $P(c|h)$ we include information about how the pipeline responds to signals in the GW channel. However, there is a difference between the two calculations. For the former, there was a well defined way to characterize the noise in the detectors (i.e., time shifting the data before analysis). Whereas for the latter, since we do not know the astrophysical distribution of signals we end up making a choice for the distributions of signals. The resulting likelihood ratio detection statistic then depends more strongly on those assumptions of the signal distributions than the previous detection statistic.

A different approach toward an improved detection statistic is to use a multivariate classifier to rank order the triggers. Such a classifier develops a classification algorithm by taking in a large number of parameters (such as the many recovered parameters of the coincident triggers we record and information from auxiliary channels if the detectors). It does this for both background triggers and injection triggers and then looks for how the two classes of triggers separate in the multidimensional space. Different classifications of this sort are currently being tested using triggers from the analysis of this search. This technique can be powerful because it allows easy inclusion even more information beyond that included in the present detection statistic, ρ_{eff} (equation (5.40)), and the associated background and signal probability factors $P(c|0)$ and $P(c|h)$.

11.2 Background Estimation

One of the main questions we ask at the end of a search is "what is the probability that a particular trigger came from the background noise?" As described above, the FARc detection statistic has a simple answer to this question, however the current procedure of calculating the FARc is limited in the smallest FAP it can estimate, when there is only one background trigger with significance equal to or greater than the trigger in question. Naïvely, since we perform 100 time-shifted analyses to estimate the background, we would expect to be able to estimate a minimum FAP of 1%. Unfortunately this was under the assumption that we had a single experiment, while in fact the FARc calculation combines the results of nine different trigger categories, which are effectively different experiments. We then have to multiply the minimum FAP by this trials factor giving us a minimum FAP that we can estimate of 9%.

One way to get around this problem and lower the minimum FAP is to increase the number of time-shifts we perform. However there are two problems associated with this. First of all, we want to estimate the background for triggers with roughly the same combination of detector noise conditions. This means that we cannot shift the data by too large an amount, otherwise the noise characteristics from multiple detectors will no longer be overlapping. Since there are time correlations in the data from a single detector, we need to shift the data of multiple detectors by a minimum amount so we do not include those correlations in our background estimation. The combination of the two effects gives us a maximum number of time-shifts we can perform. Second, on taking into account practical considerations, as we increase the number of shifts, the amount of time it takes to analyze the data increases roughly by the same factor.

A different way to estimate the background is to build a model of the background that we could then use to extrapolate to low FAPs. One way to do this is to produce a joint probability distribution function (PDF) for the background starting from the PDFs from the original single-detector triggers. As stated in chapter 5, for each detector the PDF of the SNR squared can be modelled as a χ^2 distribution with two degrees of freedom for the Gaussian noise component. Since real data also contains a nonGaussian component that dominates in the tail of the distribution, we would add a Poisson distribution to capture these tails. To factor in the effect of signal-based vetoes, we would need to run those portions of the pipeline with and without the vetoes to see their effect in each detector. We would then multiply the PDF of each detector by the ratio of the two to get the PDF including vetoes. Finally, since we would want to rank order the triggers according to effective SNR, we would construct the theoretical distribution for the value of the χ^2 veto (i.e., χ^2 distribution with $2p-2$ degrees of freedom and noncentrality parameter given by that for glitches), and combine it with the SNR PDF to obtain an effective SNR PDF for single detectors. The joint PDF for coincident triggers could then be taken as the outer product of the single-detector distributions, normalized such that its integral gives a rate that matches the number of triggers in the time-shifted analyses divided by the sum of the time-shifted analyses observation times.

11.3 Coherent Analysis

Some of the more powerful vetoes we use in analyzing GW-detector data are the amplitude consistency vetoes between interferometers. These vetoes ensure that the strength of a trigger is consistent between different detectors. However, there is also phase information that can be compared. This additional information can be included by matched filtering a coherent data stream, as described in section 3.6 and references [143, 144].

11.4 Data Cleaning and Noise Regression

Currently, all of the loudest triggers produced by analyzing LIGO data are the result of glitches in the detectors. As described in section 6.8.1, we remove times during which we expect the detectors to be glitching. Currently, as the signal processing we use for templated CBC searches introduces artifacts that can last for a few seconds around a glitch, there are wide windows we must add to these times to ensure that triggers resulting from such glitches are removed. These windows can make up a significant portion of each time we wish to veto, so trying to reduce or eliminate these windows could help save additional analysis time. One way to do this is to switch over from vetoing such times to eliminating the glitches from the data processing. This could happen if instead of vetoing the required times, plus their windows, the required times' data was excised and replaced with the appropriately colored Gaussian noise such that the transients from the excision boundaries were minimized. These times could then be flagged as not involving real data but would allow the analysis of adjacent times.

Another option, which would be much better when possible, would be the regression of noise using information from the auxiliary data channels. For situations where the transfer function between the auxiliary data channel and the strain data channel are known, this would allow excess noise from these sources to be virtually eliminated from strain data and improving sensitivity. For channels with a time-varying coupling to the strain data that is recorded, adaptive filtering could be used to eliminate both excess noise as well as transients caused by an increase and then decrease of the coupling.

11.5 Low Latency Searches

Improving the latency of CBC searches for GWs would have a significant scientific impact. Currently, there is an inherent latency in the trigger generation of roughly 30 minutes, since the analysis proceeds in batch mode analyzing 2048 seconds of data at a time (chapter 6). This is a long enough time that any electromagnetic counterparts could have faded into the background by the time telescopes received an announcement of a likely GW. There are several techniques that are being utilized to reduce this latency. References [145, 146] take advantage of the chirp-like nature of CBC GW signals to break the signal up into different bands, which can be combined to generate a trigger with an SNR that is roughly the same as the current full band analysis. These techniques can reduce the latency of the search to a few seconds. Additionally, for BNS signals in Adv. LIGO, if only the lowest frequency bands are used, a trigger for a GW could be produced a few minutes *before* the arrival of the signal from the neutron stars' collision, alerting the LIGO detectors to remain online for the next few minutes, potentially even changing the response of the detectors to the high frequency portion of the inspiral by dynamically detuning the signal extraction optical cavity, and alerting electromagnetic telescopes where and when to look, resulting in multimessenger observations of the same event.