

On the Strategy for Sending EM Alerts in O1

Peter Shawhan, with input from Marica Branchesi and Larry Price – March 30, 2015

Heading into the O1 run this fall, we will have signed MOUs with about 74 groups of EM astronomers. A sample MOU is here: <https://dcc.ligo.org/LIGO-M1400071>. Some groups have control of their own facilities (or are conducting ongoing transient surveys) or have obtained telescope time during O1; others are still building telescopes/instruments or have not yet obtained telescope time. By the terms of the MOUs, we will only share GW event candidate alerts with the groups who will actually have observing capabilities during the run, which will probably be around 60 groups.

Recall that no group is *obligated* to follow up any given event candidate. An unfavorable sky location, bad weather, scheduling priorities or technical issues can also prevent a group from observing a given event candidate. We will provide the estimated false alarm rate (FAR) for each candidate, so the observers will have a handle on its significance and can decide for themselves how much effort to make to observe it. Some facilities are dedicated to transient astronomy and there is very little cost to follow up every GW event candidate, while others (such as satellites in orbit and large telescopes on the ground) are more precious and may be used more selectively, e.g. only if a plausible EM counterpart has already been seen by another observer.

I think everyone agrees that we should issue alerts for sufficiently significant event candidates, i.e. no one is suggesting that we skip the EM follow-up program for O1. The key question we have to settle ASAP is: **What FAR threshold should we set for issuing alerts during O1?** At the Stanford LVC meeting in August 2104, we discussed three possible thresholds, ranging from 1 per month to 1 per 100 years, with their pros and cons; see slides 9 and 10 of <https://dcc.ligo.org/LIGO-G1401016>. We did not reach a collaboration decision at that time, though (despite repeated questions from some astronomers). At the recent Pasadena LVC meeting, the EM follow-up team leaders proposed to the group (on slides 8 and 9 of <https://dcc.ligo.org/LIGO-G1500323>) to adopt a target rate of 1 alert per month, for each of CBC and Burst, and corrected for livetime, in order to yield ~ 3 CBC and ~ 3 Burst alerts over the course of the O1 run. The rationale for this choice is discussed below, but first we should ask and answer some basic questions.

What are our criteria for publishing events?

Well, most of our upper limit papers have briefly described the most significant event or events from the analysis, even if they were consistent with background. So a better question is, what are our criteria for interpreting the significance of the events we find? I'd say there are four qualitatively different ways we might write about an event candidate (or a set of candidates, if there are multiple outliers considered collectively):

1. We interpret this as a definitive detection of a GW event(s)—nominally “5 sigma”, or false probability $\lesssim 10^{-6}$. Write a “detection” paper focusing on this event (or set of events).
2. This is evidence for a GW signal—nominally “3 sigma”, or false probability $\lesssim 0.003$. Write an “evidence for” paper focusing on this event candidate (or set of candidates).

3. We have an outlier which was unlikely to come from the background distribution—nominally “2 sigma”, or false probability $\lesssim 0.05$ —which is noteworthy, but not significant enough to be claimed as an event candidate. Just describe it in an upper limits paper.
4. Our loudest trigger is consistent (less than 2 sigma) with what we expect from background. Maybe describe it briefly in the upper limits paper for completeness.

What is our nominal threshold for considering an event candidate scientifically interesting?

Probably the 3-sigma level, i.e. the level at which we’d write a paper presenting it as “evidence” instead of just writing an upper-limit paper. For the 3-month O1 run, that implies that the event candidate must have a FAR of less than 1 in $(0.25/0.003) \approx 83$ years, which is conveniently about the same as the 10^{-2} yr^{-1} (i.e., 1 in 100 years) used to do rate calculations in the “Prospects for Localization” document (<http://arxiv.org/abs/1304.0670>). Assuming that we are evaluating the background properly, it is very unlikely for a noise fluctuation to appear so significant, by construction.

Note that it is the FAR assigned to the event candidate which indicates its significance, not the FAR threshold used to select triggers. That is, if we loosen the FAR threshold, that does not detract from any high-significance (very low FAR) event candidates in our sample; it just may select some additional candidates with more modest FAR.

What events might we get as a function of the threshold?

Our strategy for setting a threshold for selecting candidates can and should be run-dependent, even ignoring the possibility of EM counterparts. For O1, assuming that LLO and LHO both have ranges of 60 Mpc and working from the table in the “Prospects for Localization” document, the very most optimistic estimate is that we can expect ~ 1 binary neutron star CBC event (having a FAR of 10^{-2} yr^{-1} or better) during O1. The “realistic” rate is a factor of 10 lower, i.e. ~ 0.1 event during O1. Taking that to be exactly 0.100 for purposes of comparison, and using the exponential background model in the document, we can calculate as a function of FAR threshold:

| | | |
|-----------------|---------------------|-----------------------|
| 1 per 100 years | $\rho_c \geq 12$ | 0.100 events selected |
| 1 per year | $\rho_c \geq 11$ | 0.130 events selected |
| 1 per month | $\rho_c \geq 10.46$ | 0.151 events selected |

Since all of these numbers are much less than 1, it is fair to say that lowering the threshold has only a very small additive chance (3.0% or 5.1% if the assumptions are accurate) of selecting an additional real event during O1. On the other hand, it’s a 30% or 51% relative increase in number of events, which will become more valuable in later observing runs. For example, in 2019 when the “realistic” rate is supposed to be 20 per year, lowering the FAR threshold to 1 per year would give ~ 6 more real events per year in the selected sample while introducing only ~ 1 background event. That would still be a high “purity” of $26/(26+1)=96\%$, and likely a more optimal strategy overall due to giving smaller statistical uncertainties on rate calculations and population studies. (Note that most of the events in the sample will still have $\rho_c \geq 12$, and FAR $< 10^{-2} \text{ yr}^{-1}$.)

The previous three questions and answers have had nothing to do with EM alerts or follow-up observations. Now we discuss those aspects.

Could an apparent EM counterpart make an O1 event candidate more interesting?

Yes, in some cases, assuming that the observer has done the study to establish that the claimed counterpart is unlikely to have been a chance coincidence. It depends on what the base FAR of the GW event candidate is, and just how small the chance coincidence probability of the apparent EM counterpart is.

For publication cases 1 or 2, where we're already claiming a confident or likely GW signal and writing a paper about it, having any believable EM counterpart adds to the information about the event, making it even more compelling and giving us data to examine the astrophysics of the source. Hooray!

If we have a candidate that is “case 3” based on the GW data alone, we are already saying that it was pretty unlikely for a noise fluctuation to produce such an event, but not quite up to our standards for identifying it as evidence for a real GW event. In this case, finding a physically consistent EM counterpart with a reasonably low false coincidence probability – say, 5% or less (“2 sigma”) – really would make this event more convincing, and would basically move the event into the “case 2” category in terms of how we'd want to publish it. It would have to be a jointly authored paper in this case because neither the GW nor the EM data alone would be conclusive, but the combination would be more than 3 sigma. For O1, “case 3” corresponds to a FAR threshold of about 1 per 5 years.

What about “case 4” events, i.e. the loudest triggers which are outliers but consistent with the background? These would be selected by a FAR threshold of 1 per month, for instance. For O1, we expect the purity of this sample to be low but nonzero. For example, using the proposed threshold to yield ~6 background events over the course of the run, and assuming the “realistic” CBC rate is correct, the nominal purity is $(0.151)/(0.151+6)=2.5\%$. For a GW trigger at that threshold, it would take a remarkable EM counterpart to pick out a real event in the sample; it would need to have a false coincidence probability of less than $\sim 7 \times 10^{-5}$ for the joint probability to be significant at the 3-sigma level. A short GRB coincidence easily achieves that, thanks to combining the spatial (~ 0.01) and temporal (6 s time window every ~ 6 days $\rightarrow \sim 10^{-5}$) random coincidence probabilities, but it's not clear what other kind of EM signature can; maybe an X-ray afterglow.

Some GW candidates in the “case 4” sample will have somewhat lower FARs and thus would not require such a spectacular EM counterpart. For instance, an O1 candidate with a FAR of 1 per year has a nominal purity of $(0.130)/(0.130+0.25)=34\%$, and an EM counterpart with a random coincidence probability of $\sim 10^{-3}$ is then sufficient to give a 3-sigma joint significance. It might be possible to find an optical transient signature that is distinct enough at this level, but this has not yet been demonstrated. This is the kind of thing that astronomers need to work on.

So why are we proposing to use a 1-per-month FAR threshold for O1 alerts?

In a word: practice. It is crucial that astronomers start confronting the challenges and techniques of seeking EM counterparts under realistic conditions, and see what they can do. We developed our data analyses during early science runs (when there was very little expectation of detecting a GW signal), and when confronted with the Big Dog, we had to improve our background estimation techniques to get a good handle on the significance of such a loud outlier. The astronomers are going to have to go through a similar evolution to understand how to work with possible counterparts in our large sky areas. The handful we were working with during the S6/VSR2+3 run already learned some things from the EM follow-ups done then, and we certainly learned a lot about optical and X-ray image data and how to interpret it together with the observers.

What papers might be written?

We should plan for the publication scenarios outlined in the four cases early in this document, based on the GW data alone, unless there is a compelling EM counterpart for a “case 3” candidate that motivates us to write a joint paper. In all cases, the terms in the MOU hold: “Any apparent counterpart to the GW event candidate, that was identified due to the GW candidate alert, is strictly embargoed: it may not be published or presented prior to the public announcement or publication of the GW event candidate by LIGO and VIRGO.” That means the astronomers cannot publish *anything* about an event candidate, even non-proprietary information such as its existence, until (unless) we publish it.

If we have written a detection or “evidence for” paper (i.e., case 1 or 2), then our partners will be free to write companion or follow-up papers presenting their EM observations and interpretation of their data. If they need more details about the GW event than what we have published, e.g. to do joint parameter estimation, then they may inquire about co-authoring a paper with us. Good either way!

As mentioned above, there is a chance that we will write a joint paper as a result of having a marginally significant GW event candidate (initially case 3) which becomes more interesting due to a significant EM counterpart. If this occurs, we will need to be comfortable with the statistical analysis of the jointly observed event, and the conclusions drawn from it in the paper we’re co-authoring.

If we have written an upper-limit paper (case 3 or 4), then our partners will be free to write papers about the follow-up observations they have done. The important thing here is that our paper will be out first and will clearly present the (marginal or minimal) significance of the GW event candidates. **I propose that we commit to publish at least a brief description of all alerts we send out to partners, so that they will be able to refer to them after our paper is out;** in other words, we should not keep any alerts in permanent embargo status. If the strongest outliers are consistent with background (case 4), then we will state that. Any follow-up paper written by astronomers will have to acknowledge that context, as readers will know and judge what they say.

Could any astronomer's paper damage LIGO/Virgo?

We do not believe so. Since our paper will come out first with the description of the GW event candidate(s), our position will be clear. Any follow-up papers are the responsibility of the authors of those papers. If they write anything inappropriate, e.g. claiming to have found a counterpart without a good statistical analysis to back up that claim, then the scientific community may react negatively toward them, but not toward us. Anyway, all (or at least most) of our EM partners are competent professional scientists who should be aware of the statistical issues, and should write reasonable things if they report their observations, unless they simply make a mistake.

It is possible that the astronomers may discover a moderately significant EM counterpart, one that they present as evidence for a real event, even though it falls below our statistical standards for claiming a detection. There is nothing wrong with that, as long as the analysis is solid. A hint of a signal is a hint of a signal, and if we are clear about our (conservative) standards for evidence, people should respect us for that. As we wrote in the Open Call document (<https://dcc.ligo.org/LIGO-M1300550/public>): “LVC and EM observers acknowledge that the threshold for generating a GW alert is deliberately being set low, so that most (hopefully not all!) GW candidates reported at least in early advanced science runs will likely be noise fluctuations.... LVC will only publish as ‘detections’ or ‘strong evidence’ GW candidates which have a sufficiently low estimated false-alarm rate based on the GW data.”

Why six alerts in O1?

Because it is enough to practice with, without being excessive. Even with different sky locations, time of day/night, weather, and operational constraints, each observer group can probably follow up a few of the alerts. It may take a few tries to get things working smoothly. Having several alerts during O1 is also important to us (LVC) since we need to practice our procedures under real-data conditions, including monitoring what observations are made by our partners and what information they report back to us. Each alert will take some effort, but six alerts spread out over the course of the run seems manageable and gives us some time to adjust between them if necessary.

What about just scheduling ~6 alerts during a designated practice period such as ER8 and then using a tighter threshold during O1?

Marica, Larry and I think that is a bad idea, on three levels:

First, on a practical level, compressing the practice into a short period of time – especially, right around the beginning of the science run when everyone will be busy – would be stressful for us and for the astronomers. The astronomers may need to take multiple nights to follow up a given candidate, especially if they are following light curves of apparently variable objects, so they may not be able to respond to alerts for different events in artificially close succession, even if they have the telescope time in the right window.

Second, we see no benefit of a compressed practice period versus practicing over the course of O1. If there are any concerns about astronomers making inappropriate claims from the data

(though we think those are minimal, and would not reflect poorly on LIGO/Virgo anyway), those concerns should be about the same for ER8 as for O1 since it is all real data anyway.

Third, and perhaps most importantly, it would show bad faith toward our partners. We acknowledge that loosening the threshold is for purposes of practice, but it is practice with real data that could potentially yield a real event (and that is important). Under these boundary conditions, we should provide the best candidate events, i.e. the ones most likely to be real. Otherwise, we would effectively be saying “we are deliberately giving you sample alerts which are not our best candidates, because we do not trust you to be responsible scientists”. They signed up to receive alerts during our observing runs, so that is what we should give them.

Marica’s remarks:

“I also don't think that it is a good solution the one of sending alerts with a different threshold during ER8 and O1. I don't see any differences between the two periods if we send out alerts that they can follow with EM observations.

I don't think that the astronomers will write papers of counterpart detection with a high FAR if they know deeply the meaning of our threshold. Also for them, it would be very bad to declare a fake detection on a so important new science. On the other hand it is extremely important for them to show (mainly to obtain time on expensive facilities) that all the steps, starting from us are ready to make this new science possible.”

Larry’s remarks:

“I agree - Practice runs during ER8 are a good thing, but certainly no substitute for ~monthly alerts in O1. I also don't buy the argument that we'll be faced with inappropriate claims from our MOU partners. Sure, it's a possibility, but I think they have much more to lose (e.g. ToO time on expensive telescopes) than to gain from such claims. Personally, I think it's more likely we'll find ourselves in a situation where very few partners have followed anything up.”