



**UNIVERSITÉ  
DE GENÈVE**

---

**FACULTÉ DES SCIENCES**  
Section de physique

Laboratoire de physique III ; Astronomie et astrophysique

---

## **Hunting for the host galaxies of gravitational-wave sources**

---

David BEKRI

22th January 2022

**Table des matières**

<b>1</b>	<b>Introduction and Theory</b>	<b>3</b>
<b>2</b>	<b>Methods</b>	<b>4</b>
2.1	Data . . . . .	4
2.2	Equations . . . . .	5
2.3	Implementation . . . . .	6
<b>3</b>	<b>Results and discussion</b>	<b>7</b>
<b>4</b>	<b>Conclusion</b>	<b>10</b>

## 1 Introduction and Theory

Ever since the first predictions of gravitational waves (GWs) at the beginning of the 20th century (Einstein, 1916), physicists have been investigating what we could learn from observing these. Lots of efforts have been made in order to detect them, and eventually led to the recent first direct detection of a GW signal in 2015 (Abbott et al., 2016).

GWs can be seen as analogous to electromagnetic waves. They both propagate at the speed of light. Electromagnetic waves are the carriers of electromagnetic energy, which is a consequence of changing electric and magnetic fields. On the other hand, GWs are the carriers of gravitational energy, which is a consequence of gravitational field change (Einstein and Rosen, 1937). There are different types of sources of GWs, that we are going to discuss further in this introduction.

GWs are ripples in the space-time, created from accelerated mass. Their amplitude increases with the mass involved as well as the acceleration. Therefore, one of the most powerful, and easily-detectable class of GW sources are binary compact objects, such as black-hole binaries, neutron-star binaries, or black-hole-neutron-star binaries, in the process of merging. Another promising class is supernovae explosions where a lot of mass, typically a few times the mass of the sun, quickly changes state of motion.

Despite being accurately predicted by general relativity, GWs are very hard to detect because their small amplitudes. The strain of a GW is typically at the order of magnitude of  $10^{-21}$  m (Abbott, 2016). As a comparison, the size of elemental particles such as quarks is  $10^{-16}$  m. This is the reason why the first GWs have been detected only recently, thanks to very sensitive detectors such as the Laser Interferometer Gravitational-Wave Observatory (Abbott, 2016). This long-awaited experimental result led to a revolution in astronomy.

In order to understand the implications of such a discovery, we can specifically look at the detection of the GW signal GW 170817 in 2017, which was caused by the merging of two neutron stars (Abbott et al., 2019). The particularity of this event is that, while being detected by two different GW detectors, LIGO and Virgo, it was also observed thanks to the detection of electromagnetic waves by many observatories. This experiment as well as the study of other GWs allow to confirm various results and make new discoveries in astrophysics and cosmology, for example testing the theory of general relativity for stronger gravity regime than other usual experiments, determine the origin of certain types of gamma-ray bursts, or even measuring the value of the Hubble constant in an independent way (Abbott et al., 2021).

The study of the GW signal GW 170817, with follow-up observations using different observational facilities, is a perfect example of multi-messenger astronomy, a new field in which physicists use different types of signals from different particles, for example electromagnetic waves, cosmic rays, neutrinos, and recently GWs, in order to better understand a unique event and its physical meaning.

Multi-messenger observations depend on our ability to identify the location of the GW source, so that observational facilities around the world can join the efforts. Essentially, optical telescopes were scanning towards the positions of host galaxy candidates. From that signal we got the probability distribution of the distance at each direction.

The traditional way of searching for the host is to use the 3D localization map (probability of source being at a given direction in the sky and a given distance) that is produced by analyzing the GW signal, and cross-match it with a catalogue providing positions and distances of galaxies. This results into a list of host galaxy candidates and the relative probability for being the true host. However, this does not account for the stellar population content of a galaxy. If the highest probability is found towards

the direction of a system of two galaxy, one being very massive (containing many stars) and the other being a dwarf galaxy, which galaxy should be given priority for follow-up observations ? In principle, the full star-formation and metallicity history of the galaxies is needed to quantify the probability of hosting a GW event similar to the detected one. However, this information is unknown for the majority of the known galaxies, and we can rely on measures that are relatively easy to constrain : stellar mass, star-formation rate, and metallicity. Using binary population synthesis models (Artale et al., 2020), it is possible to simulate the production of compact-object mergers in simulated galaxies, and create prescriptions for the rates of GW sources as a function of the aforementioned stellar population parameters.

Long before the experimental confirmation of GWs, astrophysicists have elaborated several galaxy catalogues (Dálya et al., 2018; Ducoin et al., 2020) which contain that kind of information as well as other properties of a consequent amount of galaxies in a given area.

In this report, we will look at the data on the GW signal GW 170817 gathered by LIGO, and cross-match this data with the Heraklion Extragalactic Catalogue (HECATE ; Kovlakas et al. 2021). HECATE is a galaxy catalogue tailored for local universe studies, providing a wealth of relevant information for the majority of the galaxies, which is an improvement over previous catalogues. It contains 204,733 galaxies in the local Universe, and provides estimates for the stellar mass, the star-formation rate, and the metallicity for a large portion of it. The goal of this project is to see if we can use this catalogue for targeted follow-up searches for EM counterparts, and identify the correct galaxy candidate as the most likely host of a signal, and how taking in account different parameters of galaxies affect our probability results. We use the only such event to date, GW 170817, to cross-validate our methodology.

## 2 Methods

### 2.1 Data

The analysis of GW signals results into a localization map. This way we can quantify the probability of the source to be located at a certain direction in the sky.

In our experiment, we will use the localization map of the GW signal GW 170817. A sample of its content can be seen in Table 2.

UNIQ	PROBDENSITY	DISTMU	DISTSIGMA	DISTNORM
1	0.0	inf	1.0	0.0
2	0.0	inf	1.0	0.0
...	...	...	...	...
15461352	165.8	37.17	7.08	0.000698
15461353	148.8	37.2	7.14	0.000696

TABLE 1 – Sample of the GW sky localization HEALPix map in tabular format (Górski et al., 2005). An infinity value for the distance mean is a ‘dummy’ value meaning that the probability density at this direction is zero.

This file contains different types of values attributed to indexes. Those indexes represent regions in the sky, that we can imagine as pixels, all part of the celestial sphere 3D sphere. The division of sky in pixels is a Hierarchical Equal Area isoLatitude Pixelization (HEALPix ; Górska et al. (2005)).

The values of the column PROBDENSITY represent the probability density per solid angle at the given position (we note that all pixels have the same area or solid angle, and therefore the absolute scale is irrelevant). The other columns from left to right represent the mean distance value, the standard deviation, and the normalization of the Gaussian probability density distribution of the distance at the given pixel.

The HECATE catalogue consists of diverse properties for each of the 204733 galaxies in it. The most important things for our project are the equatorial sky coordinates, the right ascension (RA) and the declination, that allow us to locate those galaxies in the sky, and stellar population parameters such as the stellar mass, the star-formation rate, and the metallicity. A sample of its content can be seen in the following table :

PGC	OBJNAME	RA	DEC	logSFR_HEC	logM_HEC	METAL
2	UGC12889	0.069	47.27	NaN	11.078	NaN
3	PGC000003	0.013	-18.0082	0.42	1356756	NaN
...	...	...	...	...	...	...
16	PGC000020	0.0	0.047	15.18	0.33	0.203
...	...	...	...	...	...	...

TABLE 2 – Sample of the HECATE catalogue data

It is worth noting that the stellar mass, the star-formation rate and metallicity data are not known for all the galaxies in the HECATE. It is possible that one or more of these parameters is missing, which will have a significance later on.

The idea of this project is to create a script that crossmatches the HECATE with a given skymap of a GW signal, in order to create a priority list, i.e. which galaxies are most likely to host the GW source. The crossmatch allows us to consider only the regions of the sky corresponding to actual galaxies of the catalogue. Using the unique case of GW 170817 where the true host galaxy is confirmed, we will see if our methodology is successful in giving high priority to this galaxy (NGC 4993).

## 2.2 Equations

In this section we define the quantities of interest that we need in order to generate our priority lists.

We define  $P_{2D}$  as the probability of a galaxy to host the GW source given its position in the sky,  $(\alpha, \delta)$  where  $\alpha$  and  $\delta$  are the equatorial coordinates of the galaxy :

$$P_{2D} = P_i, \quad \text{such that } (\alpha, \delta) \in A_i, \quad (1)$$

where  $P_i$  is the probability density (see column PROBDENSITY of Table 2) of the  $i$ -th pixel,  $A_i$ , of the HEALPix map which includes the galaxy coordinates.

In order to take into account the third dimension in the probability,  $P_{3D}$ , we multiply with the probability density at the distance of the galaxy  $D_{\text{galaxy}}$ , using the signal's distance probability distribution that corresponds to the given celestial position. We use the following relation (Ducoin et al., 2020) :

$$P_{3D} = \frac{P_{2D}}{\text{Pixel Area}} \times N_{\text{pixel}} \times e^{-\frac{(D_{\text{galaxy}} - \mu_{\text{pixel}})^2}{2\sigma_{\text{pixel}}^2}}, \quad (2)$$

where  $\mu_{\text{pixel}}$ ,  $\sigma_{\text{pixel}}$  and  $N_{\text{pixel}}$  refer respectively to the mean distance value, the standard deviation, and the normalization factor at a given pixel.

We then want to take into account the different galaxy parameters provided by the HECATE catalogue, namely the stellar mass, the star-formation rate, and the metallicity, in that order of priority. We can now introduce three astrophysical terms from (Artale et al., 2020) :

$$\log n_M = 1.038 \times \log M - 6.09, \quad (3)$$

$$\log n_{M,\text{SFR}} = 0.8 \times \log M + 0.323 \times \log \text{SFR} - 3.555 \quad (4)$$

$$\log n_{M,\text{SFR},Z} = 0.701 \times \log M + 0.356 \times \log \text{SFR} + 0.411 \times \log(Z) - 1.968 \quad (5)$$

where the  $\log M$  and  $\log \text{SFR}$  are the decimal logarithms of the stellar mass (in units of  $M_\odot$ ; column `logM_HEC` in the HECATE) and star-formation rate (in units of  $M_\odot \text{ yr}^{-1}$ ; column `logSFR_HEC`), respectively. The metallicity  $Z$  is computed by us using the equation :

$$Z = Z_\odot \times 10^{\log(O/H) - \log(O/H)_\odot} \quad (6)$$

where  $Z_\odot$  is the solar metallicity,  $\log(O/H)_\odot$  is the solar oxygen abundance relative to the hydrogen, and  $\log(O/H)$  is the gas-phase metallicity of the galaxy in the HECATE (column `METAL` in the HECATE; see Kovlakas et al. 2021 for more details).

To compute the probability taking into account one or more of the galaxy parameters, we multiply  $P_{3D}$  by the corresponding astrophysical term above. For example, the probability that takes into account the stellar mass and the star-formation rate is  $P_n(M, \text{SFR}) = P_{3D} \times \log n_{M,\text{SFR}}$ .

Finally, since not all the galaxy parameters are known for all galaxies, we define a probability that applies the above formulas accordingly for each galaxy, depending on which many of its parameters are known. We call that probability  $P_{\text{final}}$  (see next section).

## 2.3 Implementation

All of this project has been done using Python on a Jupyter notebook. In order to do the correspondence between the angle value attributed to galaxies, and their position in the sky, we used a particular function of the `healpy` package.

The code is structured as follows. We first extract the paths of the data files we are going to use. We then define a function that loads the galaxy catalogue and the skymap and does the crossmatching.

Since data on star-formation rate was missing for several galaxies in the direction of NGC 4993, for illustrative purposes, we use a supplementary catalogue with star-formation rates and incorporated them into our data. These star-formation rates are computed using supplementary photometry from the *Wide-field Infrared Survey Explorer* (WISE) point-source photometry catalogue and star-formation rate calibrations from Kovlakas et al. (2021).

From there we use of the formulas described in the previous section we compute the various probabilities and proceed to make our priority maps and lists.

In particular, the  $P_{\text{final}}$  probability is coded with a function that takes into account different galaxy parameters depending on the data available for each galaxy, and will use the corresponding equation among equations (3), (4) and (5), accordingly. For example, if the metallicity parameter is missing, then the function returns the probability taking into account only the stellar mass and the star-formation rate. If the star-formation rate is also missing, it will only take stellar mass into account.

We display the top 5 results of each priority list at the end of the script.

In order to better understand the results, we used the `matplotlib` and the `ligo.skymap` libraries to visualize them. We proceed by generating a sky map of all the galaxies of the catalogue. Each of the galaxies will then appear as points with a color according to its likeliness of being the right candidate in the priority scheme.

### 3 Results and discussion

We display in the following figures the priority lists we have generated, displaying candidates from the most probable galaxies at the top to the least probable ones at the bottom. Each list is taking into account an additional galaxy parameter. Let's recall that the true host of the GW we are looking at is the galaxy NGC 4993.

Each list is followed by the correspondent plot in order to understand those results better and make the comparison between the different priority lists easier. As visible on the color scale on the right, the more the color of the dot representing a galaxy tends to purple, the higher is its probability value and therefore its rank in the priority list. The true host is highlighted by a green square, and the candidate with the highest probability for the current setting in highlighted in red.

In Table 3 and Figure 1 we show the priority list and skymap for the 2D probability :

Galaxy name	$P_{2D}$
PGC169672	0.00667
PGC831458	0.00656
PGC158506	0.00652
PGC837800	0.00638
2MASXJ13042517	0.00626

TABLE 3 – Priority list based on the 2D probability.

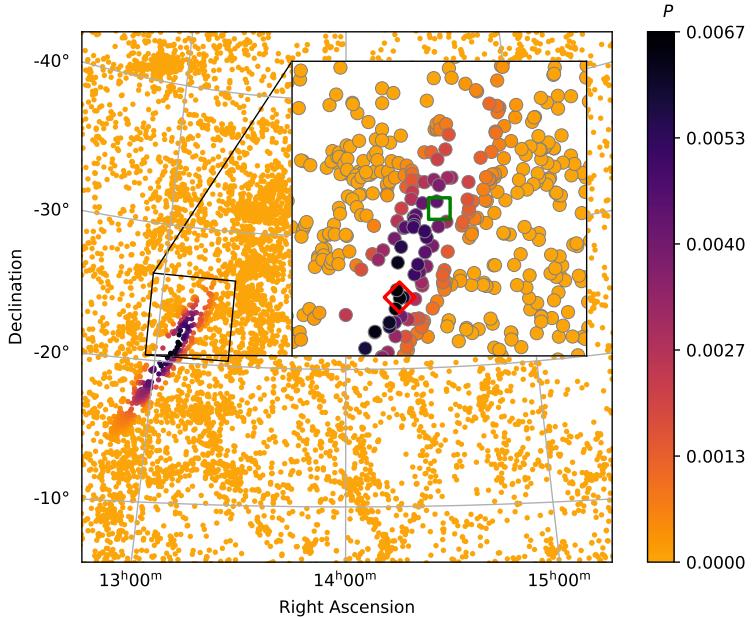


FIGURE 1 – Skymap showing the probability in the 2D scheme.

We see that when we only consider the  $P_{2D}$  probability obtained from the catalogue, the true candidate we expect does not even appear in our list. We now show the  $P_{3D}$  probability (see Equation 2).

Galaxy name	$P_{3D}$
ESO508-004	0.097
ESO575-053	0.0853
NGC4993	0.084
ESO575-055	0.0821
ESO508-024	0.0677

TABLE 4 – Priority list based on the 3D probability.

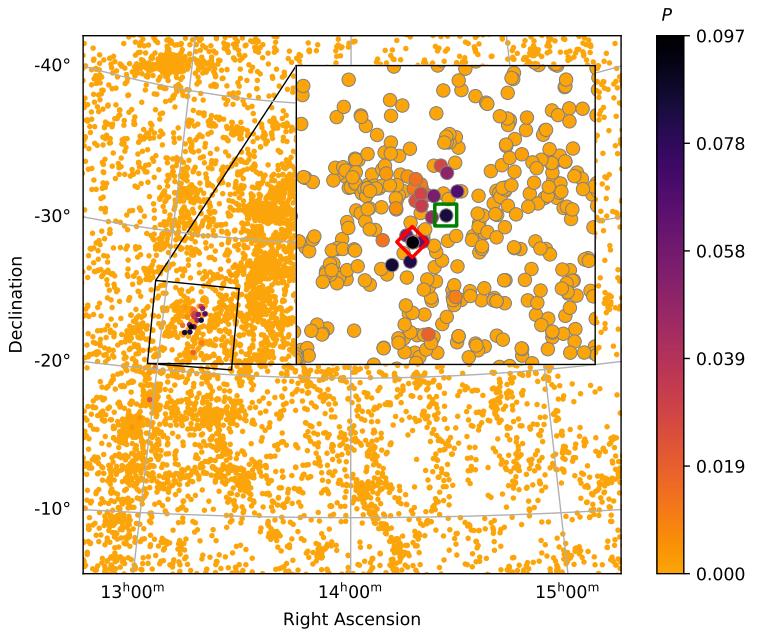


FIGURE 2 – Skymap showing the probability in the 3D scheme.

We observe a clear improvement with the P3D probability since the expected galaxy, NGC 4993, is in the third position in this list, whereas it was not even present in the previous one.

Galaxy name	$Pn(M)$
NGC4993	0.279
ESO575-053	0.108
ESO508-024	0.102
ESO575-055	0.080
NGC4968	0.0741

TABLE 5 – Priority list accounting for the stellar mass.

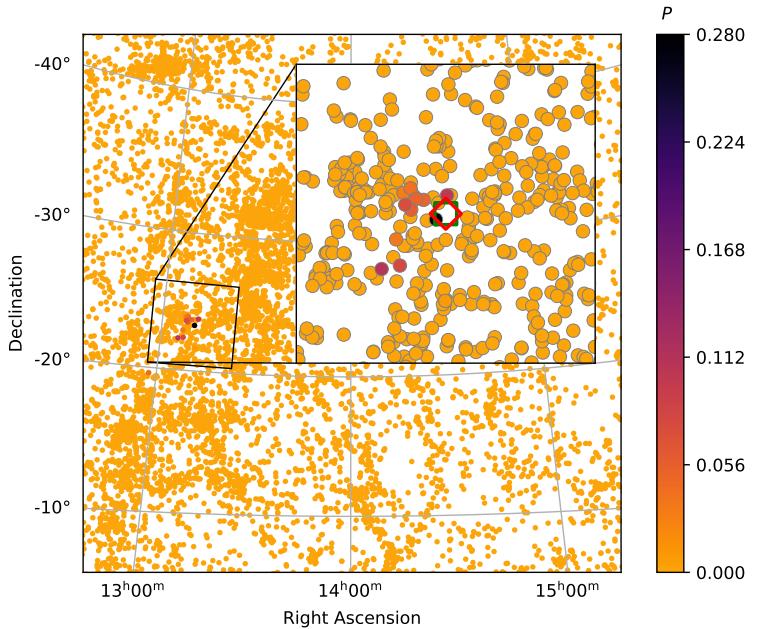


FIGURE 3 – Skymap showing the probability in the 3D scheme.

When including the stellar mass parameter, we already obtain a satisfying result, since the true host galaxy is at the top of the priority list (see Table 5). We can also see much more contrast between

nearby galaxies (see Figure 3).

From there, we observe that for the last two lists, we still have roughly the same results as the previous one. We can see minor changes in the probabilities values, and a shift between the galaxy at the 4th and the 5th position when we take into account the star formation rate, but overall it suggests that the star formation rate and the metallicity are not as important parameter in our specific case, and we manage to obtain the correct candidate at the top of the list.

Galaxy name	$P_n(M, SFR)$
NGC4993	0.249
ESO575-053	0.115
ESO508-024	0.111
NGC4968	0.092
ESO575-055	0.086

TABLE 6 – Priority list including stellar mass and star formation rate

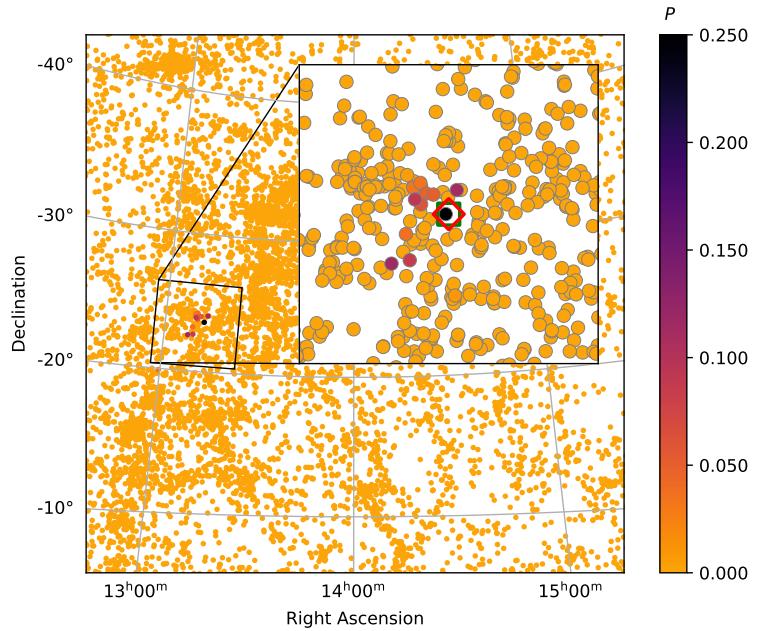


FIGURE 4 – Skymap showing the probability in the 3D scheme.

Galaxy name	$P_{\text{final}}$
NGC4993	0.247
ESO575-053	0.113
ESO508-024	0.110
NGC4968	0.091
ESO575-055	0.085

TABLE 7 – Priority list including all parameters

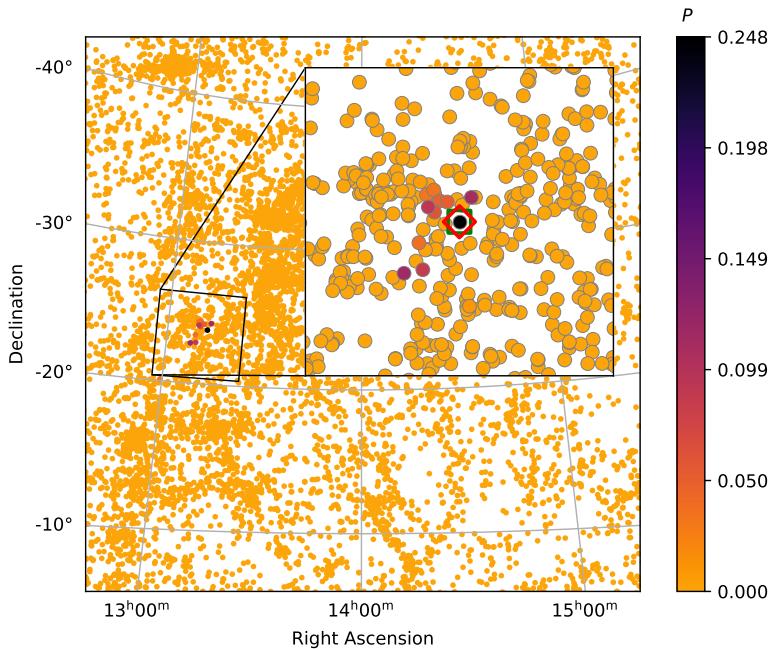


FIGURE 5 – Skymap showing the probability in the 3D scheme.

## 4 Conclusion

In this project we have managed to generate 5 priority lists that give us the galaxies most likely to be the host of the GW signal GW 170817. Each of those priority lists are taking into account different galaxy parameters that contain crucial information related to the frequency of events susceptible to be at the origin of the GW signal.

Despite some missing information for some galaxies of the HECATE catalogue, we managed to successfully find the expected host for the signal, NGC 4993.

Furthermore, comparing those priority lists gives us precious insights on how important are galaxy parameters such as stellar mass, star-formation rate, or metallicity. It even allows to compare those parameters between them and their importance, as clearly illustrated in our visualization plots. We can definitely conclude that the more we know about those parameters and include them in our analysis, the better get the results to what we are expecting.

In the process of this project, we found that we improved on the previous effort of using the HECATE for identifying hosts of GW sources (see Applications in Kovlakas et al., 2021). The LIGO-Virgo skymap contains ‘illegal values’ (e.g., negative probability densities) which were not correctly accounted for when normalizing the 2D and 3D probabilities. In addition, the  $P_{3D}$  was not calculated properly (doubly accounting for the normalization). While these corrections have a small effect on the  $P_{2D}$  and distance term, which do not give high priority to the true host galaxy, since they are multiplied in the astrophysically-informed priority lists, they significantly improved the final results (high rank of NGC 4993, and high contrast between nearby galaxies).

Given the success of this project, we plan to incorporate this script in the HECATE portal (see

<https://hecate.ia.forth.gr/>). Triggers from the LIGO-Virgo collaboration will be processed automatically (downloading the localization sky map, crossmatching, etc.), and prioritization lists will be accessible to the astronomical community to aid in the search for electromagnetic counterparts.

## Références

- B. P. Abbott. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6) :061102, February 2016. doi : 10.1103/PhysRevLett.116.061102.
- B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, and Acernese. Properties of the Binary Black Hole Merger GW150914. *Phys. Rev. Lett.*, 116(24) :241102, June 2016. doi : 10.1103/PhysRevLett.116.241102.
- B. P. Abbott, R. Abbott, T. D. Abbott, and Acernese. Properties of the Binary Neutron Star Merger GW170817. *Physical Review X*, 9(1) :011001, January 2019. doi : 10.1103/PhysRevX.9.011001.
- B. P. Abbott, R. Abbott, T. D. Abbott, and Abraham. A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *ApJ*, 909(2) :218, March 2021. doi : 10.3847/1538-4357/abdcbb.
- M. Celeste Artale, Michela Mapelli, Yann Bouffanais, Nicola Giacobbo, Mario Pasquato, and Mario Spera. Mass and star formation rate of the host galaxies of compact binary mergers across cosmic time. *MNRAS*, 491(3) :3419–3434, January 2020. doi : 10.1093/mnras/stz3190.
- G. Dálya, G. Galgóczi, L. Dobos, Z. Frei, I. S. Heng, R. Macas, C. Messenger, P. Raffai, and R. S. de Souza. GLADE : A galaxy catalogue for multimessenger searches in the advanced gravitational-wave detector era. *MNRAS*, 479(2) :2374–2381, September 2018. doi : 10.1093/mnras/sty1703.
- J. G. Ducoin, D. Corre, N. Leroy, and E. Le Floch. Optimizing gravitational waves follow-up using galaxies stellar mass. *MNRAS*, 492(4) :4768–4779, March 2020. doi : 10.1093/mnras/staa114.
- A. Einstein and N. Rosen. On Gravitational Waves. *Journal of The Franklin Institute*, 223 :43–54, January 1937. doi : 10.1016/S0016-0032(37)90583-0.
- Albert Einstein. Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, pages 688–696, January 1916.
- K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix : A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622(2) :759–771, April 2005. doi : 10.1086/427976.
- K. Kovlakas, A. Zezas, J. J. Andrews, A. Basu-Zych, T. Fragos, A. Hornschemeier, K. Kouroumpatzakis, B. Lehmer, and A. Ptak. The Heraklion Extragalactic Catalogue (HECATE) : a value-added galaxy catalogue for multimessenger astrophysics. *MNRAS*, 506(2) :1896–1915, September 2021. doi : 10.1093/mnras/stab1799.