A needle in (many) haystacks: Using the false alarm rate to sift gravitational waves from noise

Yanyan Zheng, **Marco Cavaglià**, **Ryan Quitzow-James**, and **Kentaro Mogushi** explain how statistics helps scientists spot ripples in the fabric of space and time

What does a weather forecast have in common with the hunt for gravitational waves? Not a lot, you might think. One concerns meteorological conditions here on Earth. The other is about identifying ripples in the fabric of space and time. And yet both activities involve a statistical quantity called the false alarm rate, or FAR for short. You, the reader, also probably use the FAR in your daily life, much more frequently than you might realise. In fact, every time you face a decision that depends on the probability an event may occur, the FAR comes into play.

Suppose you have to travel tomorrow to a faraway place. How do you decide whether to pack an umbrella or a bottle of sunscreen? A good idea would be to check the weather forecast for your destination. If the chance of rain is 90%, you will probably pack an umbrella. Even if there is still a 10% chance that it will be sunny, you will feel pretty confident it will rain. You make this decision by unconsciously estimating how frequently a sunny day may occur at that location when the forecast predicts a 90% chance of rain. If there are only a small number of sunny days when the chance of rain is 90%, you may correctly guess that tomorrow's forecast is reliable – and the smaller the number of sunny days, the more confident you should be. Moreover, when the predicted chance of rain is higher, say 99%, your decision should be more likely to be the right one.

The fraction of sunny days with rain forecasts at or above a given percentage defines the FAR for that prediction level. In technical terms, we say that the FAR is a function of a ranking statistic (in this case, the chance of rain) that defines the likelihood of an experiment's outcome.

Left: Artist's impression of binary black holes about to collide. It is not known if there were any electromagnetic emissions associated with GW190521. Image credit: Mark Myers, ARC Centre of Excellence for Gravitational Wave Discovery (OzGrav).

So, what has this got to do with ripples in space-time? Scientists from the Laser Interferometer Gravitational-Wave Observatory (LIGO) Scientific Collaboration[1] and the Virgo Collaboration[2] use the concept of FAR to determine the likelihood that a signal seen in their detectors is a gravitational wave from a cosmic collision of massive objects in space rather than a terrestrial or instrumental data artefact.

The hunt for gravitational waves

Just as a spoon stirs a cup of coffee, accelerating massive objects stir space and time, generating outward-propagating waves in the geometry of the universe. These waves travel at the speed of light, stretching and compressing the space dimensions as they go.

In the early morning of 14 September 2015, almost a hundred years after Albert Einstein's discovery of general relativity, the twin LIGO detectors in Livingston, Louisiana, and Hanford, Washington, recorded for the first time a gravitational-wave signal from space.[3] The event, called GW150914, originated 1.3 billion years ago from the merger of a pair of stellar-mass black holes into a single, more massive black hole. As the first telescopic observations of Galileo in 1609 marked the beginning of modern astronomy, so the GW150914 detection gave rise to a completely new way of exploring our universe. Since that first detection, LIGO and Virgo scientists have observed tens of these cosmic cataclysmic collisions of black holes and neutron stars,[4] and gravitational-wave astronomy has established itself as a powerful new branch of science to study the dark side of the cosmos.[5] More than 1,500 researchers from over 100 institutions in over 20 countries operate, develop and analyse the data from a world-wide network of gravitational-wave observatories that includes the two LIGO detectors in the USA, the European Virgo detector in Italy, the KAGRA detector in Japan and the GEO600 detector in Germany.[6]

The basic common design of these detectors is that of a modified Michelson interferometer.[7] The LIGO detectors consist of two arms, each 4 kilometres long and orthogonal to one another. They operate by splitting a laser beam at the point where the arms meet (the vertex), with a beam then sent down each arm. Mirrors located at the end of each arm reflect these beams back to the vertex where they interfere and recombine. Time variations in the light of the recombined beam are measured with a photodetector. Figure 1 shows an aerial view of the Louisiana LIGO observatory and a simplified layout of the detector (not to scale).

The lengths of LIGO's arms are tuned relative to each other such that the beams destructively interfere at the vertex, that is, no light reaches the photodetector. When a gravitational wave passes through the interferometer, its arms are rhythmically stretched and compressed. This causes a time-dependent difference in the arm lengths and a variation in the light measured by the photodetector. If a gravitational-wave signal is present in the data, the photodetector output contains information about the amplitude and the phase of the gravitational wave.

The effect of a gravitational wave on the LIGO detector is very small. Typical waves from astrophysical sources warp space-time by a distance less than one ten-thousandth of the diameter of a proton over the length of LIGO's interferometer arms! This amplitude is much smaller than the detector output in the absence of signals, the so-called detector instrumental *background noise*.[8] Therefore, detection of gravitational-wave signals requires extremely sensitive detectors and sophisticated analysis techniques.

A needle in a haystack

Looking for gravitational waves in the detector data is like trying to recognise a song at a very noisy concert. Just as the singer's voice may be covered by the chatter and cheers of the crowd, gravitational-wave signals are typically buried in the detector's background noise. One way to increase the confidence of detecting a signal is to use multiple detectors. If a consistent signal is seen in multiple detectors, there is a higher chance that it comes from space instead of being due to terrestrial noise. For this reason, LIGO and its partners typically implement *time-coincident searches* between different detectors to reject false positives. Since gravitational waves travel at the speed of light, a gravitational-wave signal must be recorded in separate detectors within their light time of flight.

After the detection candidates pass the time-coincident check, they are ranked by a statistic. The ranking statistic used depends on the kind of signal being sought. If the shape of the signal is known from theory, such as in searches for mergers of black holes and neutron stars, the main ranking statistic is the signal-to-noise ratio (SNR).[9] Figure 2 (page 28) shows the theoretical waveform that originates from a binary black hole merger embedded in the detector ▶

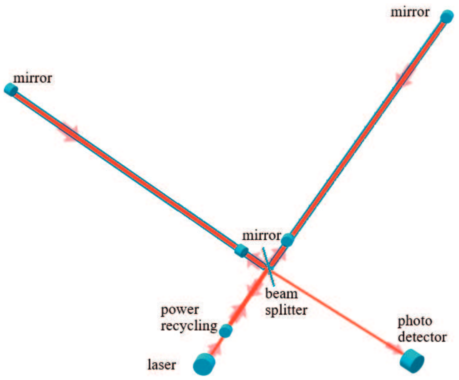


Figure 1: Top: Aerial photograph of the LIGO site in Livingston, Louisiana. Bottom: Simplified diagram of a LIGO detector.

Yanyan Zheng is a physics PhD student at the Missouri University of Science and Technology and a member of the LIGO Scientific Collaboration.

Marco Cavaglià is professor of physics and director of the Institute for Multi-messenger Astrophysics and Cosmology at the Missouri University of Science and Technology. He is a senior member of the LIGO Scientific Collaboration.

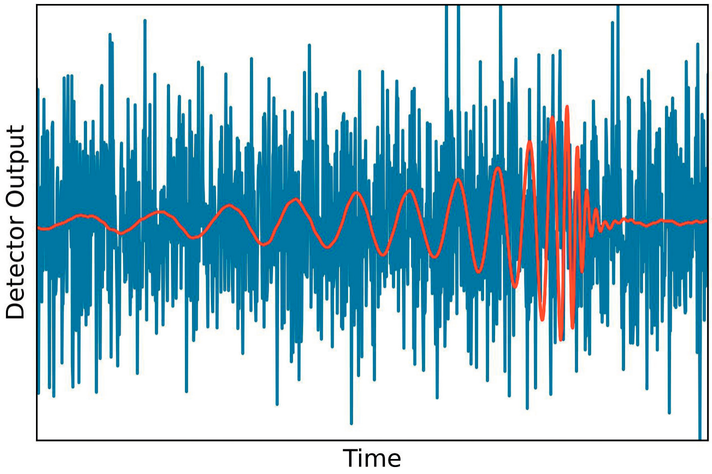


Figure 2: A typical gravitational-wave signal (red) buried in the background noise of the detector (blue). This simulated signal corresponds to the merger of a binary system of two black holes each of mass equal to 40 times the mass of the Sun.

noise. The SNR is roughly proportional to the amplitude of the signal divided by the typical amplitude of the noise. The higher the SNR, the stronger the signal compared to the noise and the more likely it is that the signal can be detected. Thus, the SNR is a good candidate for a ranking statistic to define a FAR. Just as the probability of a sunny day should decrease when the chance of rain becomes higher, the probability that a time-coincident signal in multiple detectors is not a gravitational wave decreases for higher SNR. By setting a threshold on the SNR, we can determine the FAR of the signal candidate and provide a measure of how confident we are that it is real.

Computing the FAR

How do we compute the FAR of a candidate signal with a given SNR? In simple terms, we count the number of background noise events with SNRs equal to or above the SNR of the candidate and then divide by the total analysed time.

The box "False alarm rate and false alarm probability" contains the technical details, but to understand it more intuitively, imagine a gravitational-wave detector as a weather forecaster in a particularly sunny place. Most of the time the forecaster predicts a small chance of rain for the next day, and her prediction turns out to be accurate. However, on some rare days, she gets a strong indication that rain may be on its way and so she predicts a much higher chance of rain. Suppose that tomorrow's predicted chance of rain is 90%. This is equivalent to our SNR in the hunt for gravitational waves. How would we calculate the FAR and the "false alarm" probability (FAP) that tomorrow will nevertheless be sunny despite her predicted 90% chance of rain?

To calculate the FAR and the FAP we need to examine past data. Imagine that in the past 300 sunny days at that location the weather forecaster predicted a chance of rain at or above tomorrow's prediction only three times, and on those three days it was 90%, 95% and 99%. The 300 days constitute our "background" data. To get the FAR of tomorrow's rain forecast, we divide the number of past "false alarms" (3) by the number of background days (300). This gives us a FAR of 3/300 = 0.01 per day (or 3.65 per year), which translates to a 1% FAP of tomorrow being sunny. The FAP would of course be lower (0.3%) if tomorrow's predicted chance of rain were 99% as there was only 1 background event in which the predicted chance of rain was at or above that level. The higher the predicted chance of rain (or SNR, in gravitational-wave detection), the lower the probability of a false forecast (detection).

The more background data we collect, the more accurately we can calculate the FAR. Thus, increasing the amount of background data to analyse is a crucial step of all physical experiments. This is relatively straightforward to do for weather data, for which we have decades of forecasts and actual measurements. When it comes to gravitational waves, the data collected by a detector is limited by the time it has been operating. So gravitational-wave scientists have devised a standard technique, called *time-shifting*,[10] to increase the duration of the background data.

The time-shift technique consists of generating fake coincident data by selecting the data from one detector and shifting the data from all other detectors in time by some arbitrary amount larger than the light time of flight between the detectors. This procedure provides scientists with a set of data that can be used to measure the

False alarm rate and false alarm probability

Mathematically, the FAR of a gravitational-wave signal candidate is defined as

$$\mathrm{FAR} = \frac{N}{T_{\mathrm{BKG}}}$$

where N is the number of detector background noise events with ranking statistic equal to or above that of the candidate event, and T_{BKG} is the total duration of the background data. Under the assumption that the background noise follows the Poisson distribution,

$$P(k) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$

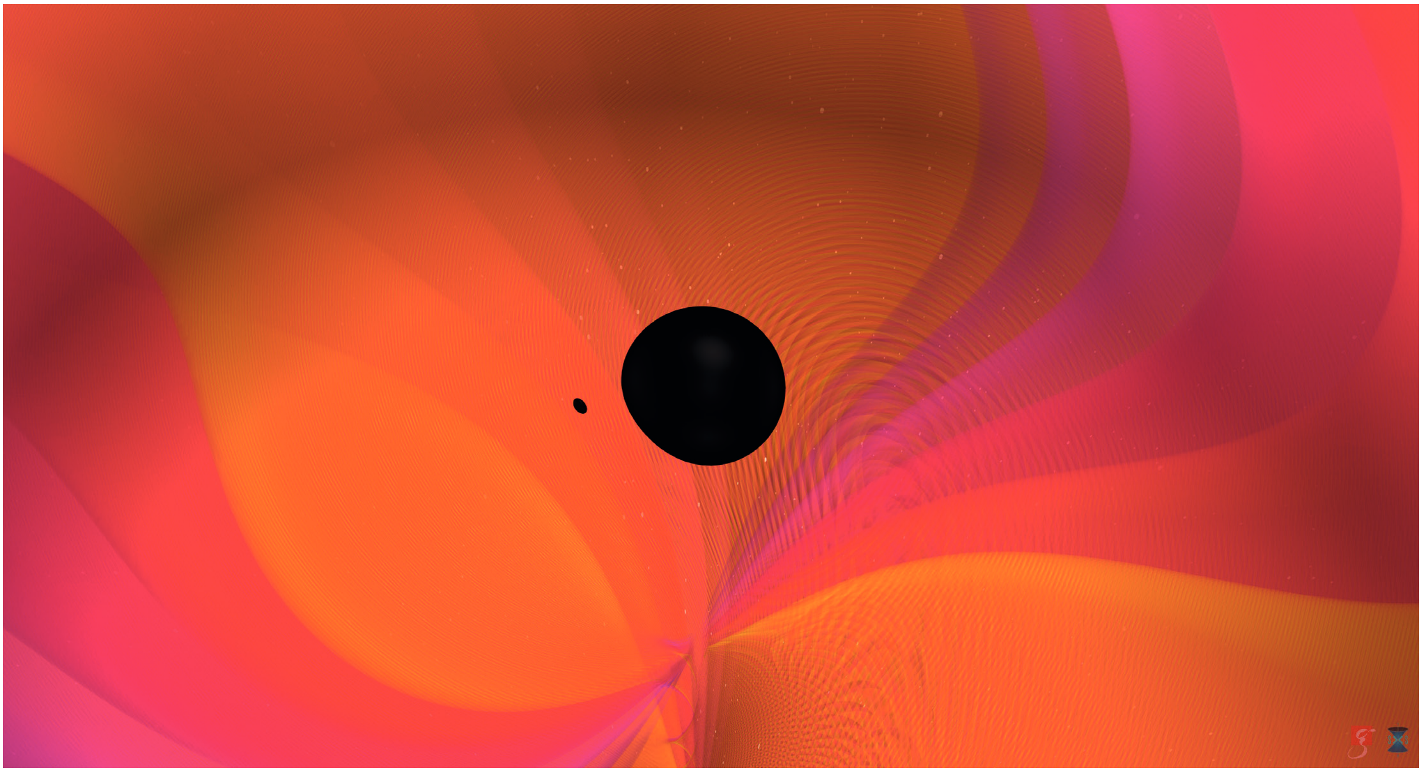
where k is the number of times an event occurs and λ is the average number of events, the (false alarm) probability that a non-astrophysical event with the same ranking statistic of a gravitational-wave candidate occurs at least once in the search time period T_0 is

$$\mathrm{FAP} = 1 - e^{-N(T_0/T_{\mathrm{BKG}})}$$

The FAP provides an alternative way to estimate the significance of a gravitational-wave candidate event. For example, the first gravitational-wave detection, GW150914, has an estimated FAR of less than 1 in 203,000 years, corresponding to a probability of less than 1 in 5,000,000 that the signal was due to terrestrial noise.[3]

Ryan Quitzow-James is a postdoctoral research associate at the Missouri University of Science and Technology. He has been working on gravitational-wave detection and LIGO data analysis since 2009.

Kentaro Mogushi is a PhD student in physics at the Missouri University of Science and Technology. He contributes to the LIGO Scientific Collaboration in various areas.



Above: This image is a still from a video visualization of the coalescence of two black holes that inspiral and merge, emitting gravitational waves. One black hole is 9.2 times more massive than the other and both objects are non-spinning. The high mass-ratio amplifies gravitational wave overtones in the emitted signal. The gravitational-wave signal produced is consistent with the observation made by the LIGO and Virgo gravitational-wave detectors on 14 August 2019 (GW190814).
Credit: N. Fischer, S. Ossokine, H. Pfeiffer, A. Buonanno (Max Planck Institute for Gravitational Physics), Simulating eXtreme Spacetimes (SXS) Collaboration.

number of accidental (false) events which naturally happen because of the background noise. In our weather forecast example, this would be equivalent to looking at forecasts from different meteorologists. If, say, two different weather forecasts predict rain for tomorrow, we could estimate whether this is just an accidental coincidence by shifting all the daily forecasts of one of them by an arbitrary number of days (greater than the typical duration of a storm, say) and measure the likelihood that their forecasts accidentally match.

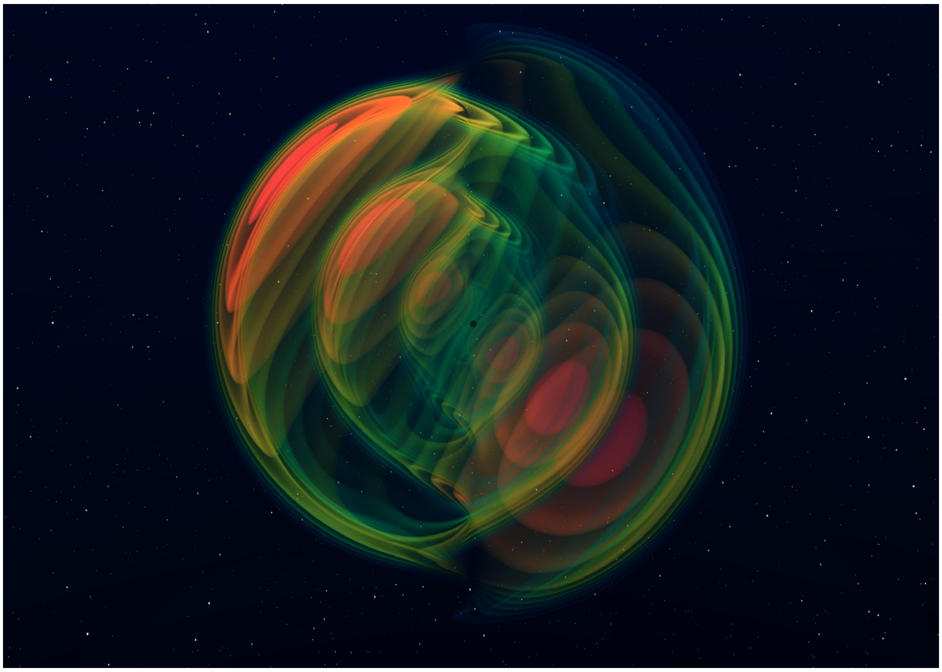


Figure 3: Numerical simulation of gravitational waves emitted by a black-hole binary merger. This event, denoted by GW190412, was discovered by LIGO and Virgo on 12 April 2019.[13] The two merging black holes had masses of about 30 and 8 times the mass of the Sun. The signal has a FAR ranging from less than 1 in 100,000 years to less than 1 in 1,000 years depending on the technique used to recover the signal.
Image credit: N. Fischer, H. Pfeiffer, A. Buonanno (Max Planck Institute for Gravitational Physics), Simulating eXtreme Spacetimes (SXS) Collaboration.

Glossary

Background noise. Fluctuations in the output of an instrument in the absence of a signal due to instrumental and environmental effects.

Black hole. A compact object so dense that even light cannot escape its gravitational pull.

General relativity. The theory of gravity proposed by Albert Einstein in 1915. Space and time form a single entity that warps in the presence of matter or energy. The motion of objects is determined by the curvature of space-time.

Gravitational wave. The dynamic warping of space-time caused by the accelerated motion of massive objects such as a binary system of orbiting black holes.

Interference. Superposition of two or more waves to form a resultant wave. Constructive and destructive interference result from the interaction of coherent waves with the same frequency but different phases.

Michelson interferometer. A device that utilises the interference of light waves to perform precise measurements of distance or wavelength.

Neutron star. The collapsed core of a massive star. The matter in neutron stars can be more than 10^{14} times denser than water.

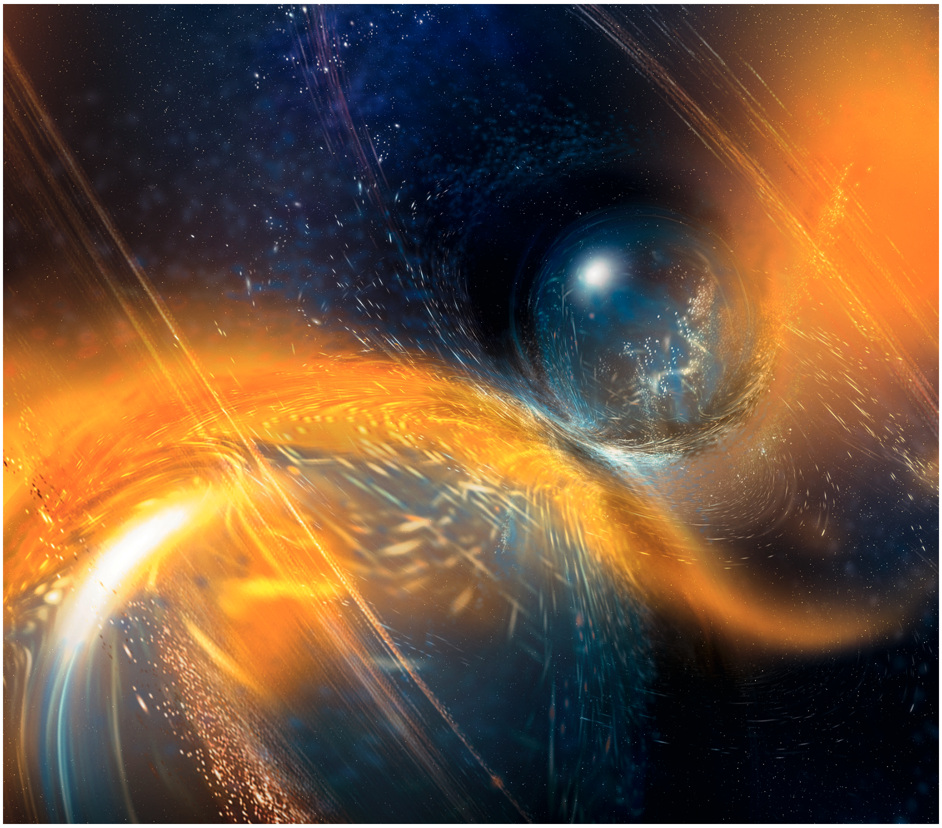
Photodetector. A sensor that converts light into electrical current.

Proton. One of the subatomic particles forming the nucleus of atoms. The estimated radius of a proton is of the order of 10^{-15} metres.

Sensitivity. A measure of the smallest signal that a physical instrument is able to detect. The sensitivity of a detector is limited by the background noise.

Signal-to-noise ratio. A measure of the level of a signal with respect to the level of background noise.

Waveform. A theoretical gravitational-wave signal as predicted by Einstein's theory of general relativity.

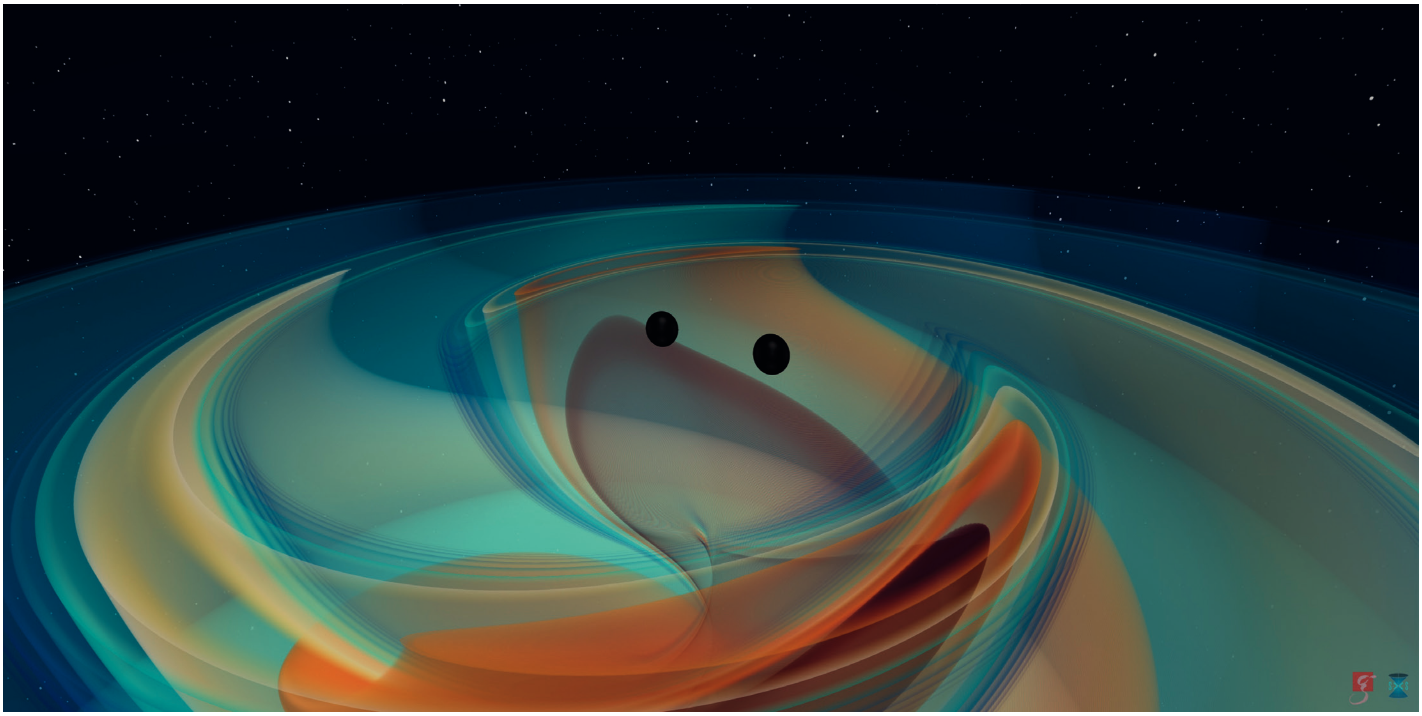


In searches for gravitational waves, any candidate signal found in time-shifted data must be caused by random coincidences of instrumental or environmental noise events. Under the assumption that the detector noise does not change too much over time, the background is representative of the time-coincident detector data and can be used to estimate the FAR of a gravitational-wave candidate. The smaller the FAR of an event at a given value of the ranking statistic, the less likely it is that this event is due to the detector's background. In the case of no detections, the FAR allows scientists to set upper limits on the rate of gravitational-wave events. Therefore, the concept of FAR is crucial to investigate gravitational waves from any kind of transient gravitational-wave sources, even from as yet undetected sources such as nearby supernovae.

As we mentioned earlier, the FAR of a gravitational-wave candidate, such as the one depicted in Figure 3 (page 29), depends on the background noise. Thus, it can be made more accurate by improving the detector,[6] mitigating environmental and instrumental noise sources,[11] and improving data analysis algorithms.[12] Many scientists and students in LIGO, Virgo, GEO600 and KAGRA are currently working to make detection techniques increasingly efficient, bring the detectors to design sensitivity, and develop the next generation of gravitational-wave interferometric detectors.

Gravitational-wave scientists also employ a plethora of other statistical tools which could not be covered in this brief article.[8] Statistics played a fundamental role in the detection of gravitational waves and the birth of multi-messenger astrophysics, enabling scientists to look deeply into our universe and understand some of its most fascinating mysteries. As we continue into the future, come rain or shine, this emergent branch of science will continue to rely upon, and benefit from, statistical science. ■

Left: Artist's impression of the binary neutron star merger observed by LIGO Livingston on 25 April 2019 (GW190425). Image credit: National Science Foundation/LIGO/Sonoma State University/A. Simonnet.



Above: This image is a still from a video of a numerical simulation of a heavy black-hole merger (GW190521). The two black holes inspiral and merge, emitting gravitational waves. The black holes have large and nearly equal masses, with one only 3% more massive than the other. The simulated gravitational wave signal is consistent with the observation made by the LIGO and Virgo gravitational wave detectors on 21 May 2019 (GW190521).
Image credit: N. Fischer, H. Pfeiffer, A. Buonanno (Max Planck Institute for Gravitational Physics), Simulating eXtreme Spacetimes (SXS) Collaboration.

Note

More information on gravitational-wave research can be found by visiting ligo.org, www.virgo-gw.eu, and gwcenter.icrr.u-tokyo.ac.jp/en. Publicly released LIGO data can be found at the Gravitational Wave Open Science Center: gw-openscience.org.

Disclosure statement

The authors are members of the LIGO Scientific Collaboration. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the US National Science Foundation under grant PHY-0757058. Computational resources are provided by the LIGO Laboratory and supported by the US National Science Foundation Grants PHY-0757058 and PHY-0823459, as well as resources from the Gravitational Wave Open Science Center (gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. The authors are partially supported by the National Science Foundation under grants PHY-1921006 and PHY-2011334. The authors declare no conflicts of interest or affiliations beyond their academic appointments.

References

1. The LIGO Scientific Collaboration *et al.* (2015) Advanced LIGO. *Classical and Quantum Gravity*, **32**, 074001.
2. Acernese, F. *et al.* (2015) Advanced Virgo: A second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, **32**, 024001.
3. Abbott, B. P. *et al.* (2016) Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, **116**, 061102.
4. Abbott, B. P. *et al.* (2019) GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Physical Review X*, **9**, 031040.
5. Abbott, B. P. *et al.* (2017) GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters*, **119**, 161101.
6. Abbott, B. P. *et al.* (2018) Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, **21**, 3.
7. Michelson, A. A. and Morley, E. W. (1887) On the relative motion of the Earth and the luminiferous ether. *American Journal of Science* (third series), **34**(203), 333–345.
8. Abbott, B. P. *et al.* (2020) A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Classical and Quantum Gravity*, **37**, 055002.
9. Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A. and Creighton, J. D. E. (2012) FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Physical Review D*, **85**, 122006.
10. Abbott, B. *et al.* (2005) Search for gravitational waves from galactic and extra-galactic binary neutron stars. *Physical Review D*, **72**, 082001.
11. Abbott, B. P. *et al.* (2018) Effects of data quality vetoes on a search for compact binary coalescences in Advanced LIGO's first observing run. *Classical and Quantum Gravity*, **35**, 065010.
12. Cuoco, E., Powell, J., Cavaglià, M. *et al.* (2020) Enhancing gravitational-wave science with machine learning. Preprint, arXiv:2005.03745 [astro-ph.HE].
13. Abbott, R. *et al.* (2020) GW190412: Observation of a binary-black-hole coalescence with asymmetric masses. *Physical Review D*, **102**, 043015.