Comparing Chirp Mass and SNR Properties of Gravitational Wave Events in LIGO's Third and Fourth Observing Run

Andrew Valentini

May 2024

Abstract

The Laser Interferometer Gravitational Wave Observatory (LIGO) has recently received major experimental upgrades in the form of frequency-dependent light squeezing, marking the beginning of its current observing run, O4. These upgrades are expected to improve the sensitivity of the detector, allowing it to see lighter mass compact object merger events and merger events of greater distances than the previous observing run, O3, was able to detect. To test these expectations, we carry out a collection of basic statistical tests on the publicly-available O3 and O4 data in this study. We confirm the expectations that LIGO's O4 data has a higher variance in the chirp mass and SNR distributions than in O3 and confirm the expectation that O4's chirp mass distribution has a greater mean than that of O3, which we explain to be a proxy for the distances of merger events. We additionally come to the unexpected result that O4's SNR distribution has a very similar value to that of O3. We provide the reasoning for these expectations and close with further explanation of this study's results.

1 Introduction

The existence of gravitational waves was first theorized by Albert Einstein in the year that followed the publishing of his general theory of relativity [2]. Because his theory of gravity replaced the Newtonian conception of space as a static stage on which the universe's events play out with a more dynamic picture that is best captured by John Archibald Wheeler's famous statement "matter tells spacetime how to curve and curved spacetime tells matter how to move", the idea that movements in this space would cause vibrations in it is not a completely startling consequence. Gravitational waves were first experimentally inferred to exist by Russell Alan Hulse and Joseph Hooton Taylor Jr in 1974 [9] from the orbital decay of a neutron star and pulsar merger which was characteristic of the emission of gravitational waves of a binary system, an observation which was later awarded the Nobel Prize in 1993. There are a great deal of (somewhat controversial) historical events that played out between the observation made by Hulse and Taylor and the origin of the plans to build an instrument sensitive enough to directly detect the incredibly small spatial fluctuations caused by the passing of gravitational waves but, for the purposes of this paper, it's enough to just state that such a detector, called the Laser Interferometer Gravitational Wave Observatory (LIGO), began to be built in 1994 and was operational by 2002.

The initial experimental iteration of LIGO lasted until 2010 with no detection of gravitational waves and, during this time, plans to improve the instrument's already absurd sensitivity were underway. These experimental upgrades were finished in 2015, marking the beginning of advanced LIGO and eventually earning the detector LIGO the designation of being the "single largest enterprise undertaken by the NSF" [1]. The National Science Foundation's most costly investment finally brought forth a detection on September 14th, 2015 which was later determined to have come from the merging of stellar-mass black holes[1], which finally satisfied the decades-long anticipation of being able to directly detect gravitational waves from the most energetic events in our universe.

---

[1] The existence of which had not been previously observed and was a major astrophysical discovery in its own right.

Since the initial event, LIGO has continued to detect gravitational waves from the merging of black holes and neutron stars. The detector has not been continuously operational, though; it has gone through four separate observing runs, each of which have been separated in time by experimental upgrades to the detector. Marking the beginning of its fourth observing run, O4 (May of 2023), LIGO has recently received its most significant upgrade since its transition from the initial and advanced stage of the detector in the form of frequency-dependent light squeezing [5]. This significant upgrade is expected to increase the detector's responsiveness to gravitational waves; being able to detect events with properties similar to those detected in the previous observing run, O3, with greater confidence and detecting events that O3 was not able to. The improved sensitivity of O4 in comparison to O3 is clear by just noting the number of events detected over the span of each run. O3 detected 56 events over a span of nearly a year of observation while O4 has already detected 81 events over eight months.

Given the significant experimental upgrades to LIGO and the clear increase in the number of detected events, this paper aims to determine if a collection of basic statistical tests can be used to determine if the actual events that LIGO's current observing run (O4) is detecting are physically different than those of the previous observing run (O3). We know that LIGO is now able to detect events more frequently in O4, but we would like to know if the properties of these events are any different than those detected in O3. Doing so would confirm the expectations that LIGO's experimental upgrades have allowed it to detect events in O4 that it wasn't able to in O3. Making such a confirmation would be of importance to the collaboration because doing so would verify the expected increase in performance of the detector, which could possibly identify exotic merger events and, maybe most importantly, possibly provide a better understanding of black holes in the early universe.

To approach the question of whether or not the events detected in O4 are statistically different from those in O3, we must first identify parameters of these events which are available for both O3 and O4 data so that a valid comparison may be carried out. This brings us to the problem that every gravitational wave physicist must deal with. A now notorious property of black holes was first discovered in 1967 by Werner Israel [3] that tells us that black holes can only *ever* be described by three physical properties to an external observer: its mass, charge, and spin. All other information carried by objects that fall into black holes is thought to be forever lost to the external universe. John Archibald Wheeler expressed this curiosity with the statement "black holes have no hair", which led physicists to refer to this result as the no-hair theorem[2]. Even worse, the charge of black holes is typically neglected since observations indicate that black holes are nearly always neutrally-charged and the spin of black holes has been shown to have little effect on the emission of gravitational waves. We are therefore limited to just the masses of the merging black holes for a parameter able to describe a physical property of the event.

Thankfully, the component masses of a merging system are almost fully responsible for the emission of gravitational waves for technical reasons that extend beyond the scope of this paper, but which are physically intuitive since one would expect that higher mass mergers would interact more violently with the gravitational field, causing gravitational waves of higher amplitude than lower mass mergers. This makes the masses of the merging system ideal for comparing the populations of black hole mergers in O3 and O4. We would therefore expect O4 to occasionally detect events with lower component masses than O3 was capable of due to the detector's increased sensitivity and so will compare the statistical properties of O3 and O4's mass distributions to quantify how the experimental upgrades have affected LIGO's capabilities.

In order to easily compare the mass properties of the events of O3 and O4, we would like a single parameter that accounts for the component masses of the merger system instead of having to deal with individual mass components and having to apply a statistical test to each. This could be done by just considering a direct sum of the component masses, but we will instead consider a measure called *chirp mass* in this study. The chirp mass of an event

---

[2] The use of the word "theorem" is slightly misleading in this context. There still does not exist a rigorous proof of the no-hair theorem and is therefore often referred to as the no-hair conjecture by mathematicians. Despite this fact, the no-hair theorem is almost universally accepted to be true by gravitational physicists.

is defined as

$$\mathcal{M} := \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{1}$$

where m_1 and m_2 are the masses of the merger's components. The chirp mass of an event is actually what is first inferred when a detection is made since it is more directly related to the frequency evolution of gravitational waves, which is a property that is physically observable in the data. Estimates for m_1 and m_2 are *then* made from the chirp mass. Choosing to compare the distributions of $\mathcal{M}$ between O3 and O4 will thus not only allow us to look at a single distribution for each observing run, simplifying our analysis, but also adds an additional layer of confidence in the results of this work since $\mathcal{M}$ is able to be more accurately determined than m_1 and m_2 individually, since the latter are determined from the former.

We know that the component masses of a merging system are a good measure for the inherent energy of a merger event and that comparing the chirp mass distributions of O3 and O4 data will allow us to make conclusions about how the recent experimental upgrades to LIGO have affected the events the detector is able to observe. We cannot only consider this parameter to analyze LIGO's improved sensitivity, though, because the mass alone says nothing about how the energy of the gravitational waves will appear to *us* once they pass through the Earth. Similar to a sound wave, a gravitational wave will lose energy at a single point as it propagates since its spreading requires that its energy be distributed over increasingly larger volumes of space. To provide a more complete perspective on how LIGO's increased sensitivity has affected its ability to detect gravitational waves, we must therefore also consider a parameter that is not inherent to the source of these gravitational waves but instead captures how this energy will appear to us after the wave has traveled some distance to pass through the Earth. The obvious choice for such a parameter would just be the distance from the location of the merger event to Earth, but for reasons unknown to the author, LIGO only publicly provides such data for O3 events. Work has been done to construct a model able to make distance predictions for events in O4 [6], but we will not rely on these predictions to proceed in this study for reasons having to do with the confidence of such predictions.

Alternatively, we can consider the signal to noise ratio (SNR) of each event as an indicator of how energetic the gravitational waves appear to us on Earth, which is publicly provided for O3 and O4 data by the LIGO collaboration. As the name indicates, this parameter is a measure for how clearly the signal of each gravitational wave event appears in the data with respect to LIGO's instrumental noise. We would expect that an event detected in O4 with the same component masses as an event detected in O3 would have a higher SNR value, since the gravitational wave signal is expected to have been made more noticeable in O4 due to its experimental upgrades.

In this study, we therefore consider properties of chirp mass and SNR distributions of events across O3 and O4 to draw conclusions about the nature of the improved detector and consider how these improvements have changed the sort of black hole merger events that LIGO capable of detecting.

2 Methods

This study analyzes the properties of gravitational wave events caused by binary black hole merger events. LIGO is also capable of detecting and does detect gravitational waves from neutron star-neutron star and black hole-neutron star merger events, but we only expect there to be a few of these exceptional events in both O3 and O4 and so will proceed by assuming all of the data comes from black hole-black hole merger events to consider how LIGO's capabilities have changed. A more complete analysis would locate and remove these non-black hole-black hole merger events in O3 and O4 data so that the conclusions made in this paper could be strengthened. This point will be mentioned again in a later section and we will make comments on how the dividing of our data could serve as future work.

The population under study in this work is all black hole-black hole merger events in the universe. It should be clarified that every object with mass in the universe creates gravitational waves due to its interaction with the gravitational field, but only gravitational waves from the most energetic events in the universe, black hole and neutron star mergers, are able to be detected by LIGO due to how small these field perturbations are[3]. The sample considered for this study is all of the gravitational wave events in O3 and O4, which is comprised of 56 and 81 events, respectively. Again, even given LIGO's incredible sensitivity, we are fortunate for each gravitational wave detection that is made since gravitational wave astronomy is a relatively young field and still has very little experimental data to provide to data scientists and theorists, meaning we cannot be selective with the data that is confirmed to have come from gravitational waves at this time.

The O3 and O4 gravitational wave events are independent of each other since it's obvious that two black holes can only ever merge once. We assume that the population distribution from which these event samples are detected is unchanged from O3 to O4. This is not an unreasonable assumption but it is not without error for a similar reason as before. We know that the number of black holes capable of merging with others in the universe will decrease over time as these black holes actually merge with each other, but additionally, over time, more black holes will be created from stellar collapse events, thus positively contributing to the distribution of black holes able to merge. We assume that this rate of merging, which naturally decreases number of black holes available for merging, is equivalent to the rate of black hole production from stellar collapse to support our study's assumption that the background population of possible black hole-black hole merger events remains constant from O3 to O4.

The discussion of how LIGO is actually able to extract the data used in this study from gravitational wave events is very long. In fact, entire textbooks have already been written on this subject [8]. It's enough for the purposes of this paper to state that LIGO uses a process of matched filtering to detect gravitational waves due to how faint the signals of even the most powerful events can be relative to LIGO's instrumental noise. This process assumes a

---

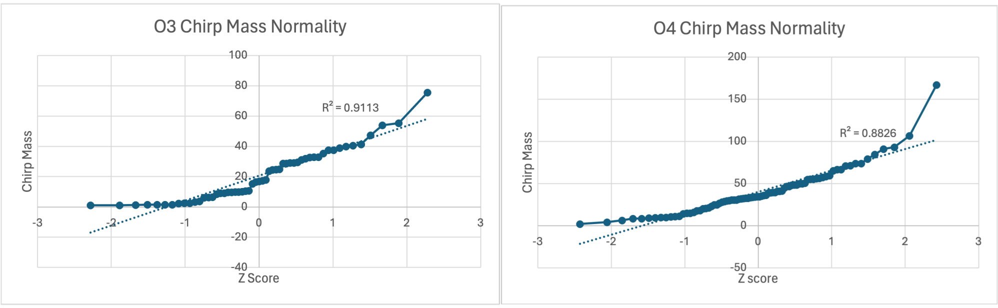
[3] The gravitational waves detected by LIGO carry an average spatial strain of 10^{-21} meters. For scale, this is the size of our galaxy changing in its diameter by the size of your fingernail. LIGO's ability to detect changes in length of this magnitude is one of humanity's most incredible accomplishments.

bank of template waveforms of what the emission of gravitational waves have been theorized to look like from the merging of compact objects such as black holes and neutron stars and when these waveform templates are fitted against LIGO's observed data and there is a high correlation between one of these templates and LIGO's data, an almost immediate alert is given to scientists to then determine if this signal has truly come from a gravitational wave or is just the result of high levels of noise. If the event is confirmed to have come from a gravitational wave, a notification to the collaboration and to the public is made that a gravitational wave signal has been detected from a merger event with properties given by the waveform template that caused the alert. At this point, the event data is then compiled for the collaboration on the website GraceDB [4]. This is where the data for this study was collected.

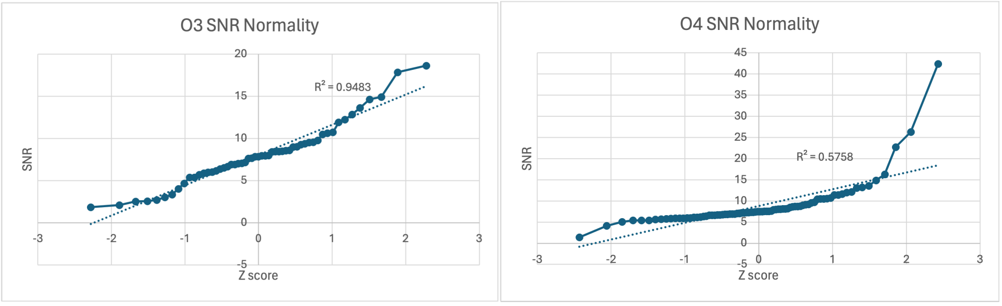
3 Results

Before attempting to compare the statistical properties of the chirp mass and SNR distributions made up by events detected in O3 and O4, we must first determine whether or not this data is normally distributed so that we know which statistical tests would be most appropriate to apply. We do this by producing quantile plots. A parameter that follows a normal distribution should have a linear relation in these plots. We quantify the linearity of each plot by computing a coefficient of determination, called the R-squared value. Note that an R-squared value of $R^2 = 1$ indicates an absolute linear relationship in these plots while an $R^2 = 0$ indicates no linear relationship of the data at all. We present the results in Fig 1.

Before proceeding to the statistical tests of this paper, we must also compute the variance and mean of the chirp mass and SNR distributions in O3 and O4 to get a scale for the sort of numbers we'll be dealing with in this study and to begin motivating what results we may expect from our tests:



(a) O3 chirp mass quantile plot (b) O4 chirp mass quantile plot



(c) O3 SNR quantile plot (d) O4 SNR quantile plot

Figure 1: Quantile plots and their associated R^2 values for the for the chirp mass and SNR distributions of O3 and O4 data. Linearity of these lines, quantified by the R^2 value, indicates a normal distribution in the given parameter.

Table 1: The variance and mean of the chirp mass and SNR distributions for **O3 data**

O3	s^2	μ
$\mathcal{M}$	286.836	20.577
SNR	13.089	8.035

Table 2: The variance and mean of the chirp mass and SNR distributions for **O4 data**

O4	s^2	μ
$\mathcal{M}$	710.411	40.062
SNR	26.674	8.810

Note that SNR is a ratio of the signal detected to the LIGO's noise, so it is a unitless quantity, and chirp mass is expressed here in terms of solar units, where one solar unit (1 $M_\odot$) corresponds to the mass of our sun, 1.989×10^{30} kg.

3.1 Chirp Mass Analysis

By visually inspecting (a) and (b) from Fig. 1 and taking note of their R^2 values, proceed with the assumption that the O3 and O4 chirp mass distributions are normally-distributed, which will influence the type of statistical test we will use in the analysis. It's clear that neither plot is perfectly linear, though, which indicates that the distributions are not perfectly normal and this will impact the confidence of our findings, which will be discussed later. Proceeding with the assumption that the distributions of chirp mass of O3 and O4 events are normally-distributed and the knowledge that each merger event is independent and randomly-occurring, we will use an F-test to first determine if the variance of these chirp mass distributions differ between O3 and O4.

In constructing our null and alternative hypotheses for this F-test, we must first recall that lower mass events will emit gravitational waves in lower energies than larger mass events. We also need to consider the expectation O4 will be able to detect classes of black holes that O3 would not be able to and the fact that greater cosmological distances correspond to earlier periods of the universe. If O4 is able to detect merger events at greater distances than O3 was able to, this means that O4 will be capable of detecting black holes that merged earlier in the universe than O3 could detect, due to the constancy of the speed of light. For the purposes of this paper, it's enough to state that the amount of metal present in the universe has increased over time, beginning with the complete absence of it at the origin of the universe, and that stars with greater concentrations of metal will die sooner than those with lower metallicities. This informs us that stars of the early universe were able to grow much larger and more massive than stars today. Because the eventual collapse of massive stars is the process by which black holes are created, meaning there is a direct correspondence between the mass of the progenitor star and the resulting black hole, we are therefore led to the hypothesis that the earliest black holes will be the most massive, especially considering the obvious fact that they also will have existed for longer periods of time than younger black holes, allowing them to accumulate mass over more time. With the expectation that

LIGO's O4 will be able to detect fainter events, caused by lighter mass events *and* events of greater distance, and therefore larger masses, than O3 was able to, these results lead us to the expectation that the distribution of chirp mass in O4 should have a *larger variance* than that of O3. We expect that O4 should be able to detect a wider range of mass values than O3 was capable of due to the detector's experimental upgrades. Written out more compactly, we have constructed the null and alternative hypotheses

$$H_0 : s^2_{\text{O4}} \leq s^2_{\text{O3}},$$

$$H_a : s^2_{\text{O4}} > s^2_{\text{O3}}$$

where s^2 stands for the variance of the O3 and O4 chirp mass distributions.

With these hypotheses in mind, we carry out an F-test, detailed in [7], on the O3 and O4 chirp mass distributions with a type 1 error probability α of 5% and are given the following results:

Table 3: $\mathcal{M}$ F-Test Results

Critical Value	Test Statistic	p-value
1.5244	2.4767	0.00027

With this F-test, we additionally construct a 95% confidence interval around the test statistic value of $F = \frac{s^2_{\text{O4}}}{s^2_{\text{O3}}}$. From this, we find that

$$\frac{s^2_{\text{O4}}}{s^2_{\text{O3}}} \in (1.6248, 3.7021)$$

with 95% confidence. This tells us that the variance of O4's chirp mass distribution is expected to be as little as 1.6 times greater and as large as 3.7 times greater than that of O3's with 95% confidence.

We now move on to comparing the means of O3 and O4's chirp mass distributions. To restate what has been explained above, we expect LIGO will be able to detect lower *and* higher mass events in O4 than O3 was able to due to the experimental upgrades of the detector and what we know about the evolution of stars. If O4 was only able to detect lower mass events that O3 was not able to, we would expect the mean of O4's chirp mass distribution to be smaller than O3's. But because O4 is *also* expected to detect higher mass events due to the expectation that it will detect events at greater distances and thus earlier in cosmological time, we expect that these high mass events will have a much larger impact on the movement of O4's mean than the addition of lighter mass events. We thus expect O4's mean chirp mass to be greater than O3's. Formally, we construct the hypothesis for the means of these distributions to be

$$\begin{aligned} &H_0 : \mu_{\mathrm{O4}} - \mu_{\mathrm{O3}} \leq 0, \\ &H_a : \mu_{\mathrm{O4}} - \mu_{\mathrm{O3}} > 0 \end{aligned} \tag{2}$$

since we want to test if the mean of O4's chirp mass distribution really is greater than that of O3.

From these hypotheses, we carry out a t' test with unequal variances, detailed in [7], since we have clearly demonstrated that $s^2_{\mathrm{O4}} \neq s^2_{\mathrm{O3}}$ and again because the samples are independent and we have shown the discussions to be relatively normal in (a) and (b) of Fig. 1, which are all conditions that must all be satisfied to apply such a test. From this test, we find:

Table 4: $\mathcal{M}$ t' Results

Critical Value	Test Statistic	p-value
1.6562	5.2303	3.1491×10^{-7}

With this t' test, we additionally construct a 95% confidence interval and we find that

$$\mu_{\mathrm{O4}} - \mu_{\mathrm{O3}} \in (17.6963, 21.2735)$$

with 95% confidence. This tells us that the mean of O4's chirp mass distribution is very likley to be 17.69 to 21.27 solar mass units larger than the mean of O3's chirp mass distribution, confirming our original hypothesis.

3.2 SNR Analysis

Similar to the reasoning that went into the construction of hypothesis for the variance in chirp mass of the O3 and O4 distributions, we expect that the distribution of SNR values for O4 should have a higher variance than that for O3. We expect that O4 is able to detect events with lower SNR values than O3 was able to due to LIGO's experimental upgrades and we expect that the SNR of an an event observed in O3 would have a higher value if also observed in O4 (which will also contribute to the construction of our hypotheses for testing the means of the SNR distributions). A more detailed literature review would be necessary to determine if this ability to detect events with lower SNR in O4 is due to the noise of the detector itself being lowered, due to the actual gravitational wave signals becoming amplified over the same noise level as in O3, or a combination of both the noise level being lowered and the signals being amplified. We will offer further commentary on the interpretation of these results in the following section but it is enough now to just state that we expect the variance of O4's SNR distribution to be greater than that of O3's since we expect the detection of events with lower *and* higher SNR values in O4 than in O3. Written out more compactly as before:

$$H_0 : s^2_{\text{O4}} \leq s^2_{\text{O3}},$$

$$H_a : s^2_{\text{O4}} > s^2_{\text{O3}}.$$

In order to test this hypothesis, we must conduct a test for comparing differences in variances for data that assumes independence of samples and non-normality of the distributions since (d) of Fig. 1 and its associated R^2 value clearly indicates that the distribution of SNR values for O4 data is not normally distributed. Because I know of no such comparative variance test that assumes a non-normal distribution, we proceed with using the F-test again. Proceeding with the F-test while knowing that O4's distribution does not satisfy the normality assumption for using this test will not completely invalidate this study's results but what follows should be accepted with caution. More discussion on the interpretation of these results will be provided in the proceeding section. We therefore proceed to test our hypothesis for the O3 and O4 SNR distributions by using an F-test again with a type 1 error probability, α, of 5% and are given the following results:

Table 5: SNR F-Test Results

Critical Value	Test Statistic	p-value
1.5244	2.0379	0.0030

From this F-test, we again construct a 95% confidence interval around the test statistic value of $F = \frac{s^2_{\text{O4}}}{s^2_{\text{O3}}}$. From this, we find that

$$\frac{s^2_{\text{O4}}}{s^2_{\text{O3}}} \in (1.3369, 2.7244)$$

with 95% confidence. This tells us that the variance of O4's SNR distribution could be as little as 1.3 times greater and as large as 2.7 times greater than that of O3's with 95% confidence.

We now move on to comparing the mean SNR value of the O3 and O4 data. Because we can be fairly confident that O4's SNR distribution is non-normal from (d) of Fig. 1, we will use the Wilcoxon Rank Sum Test, detailed in [7], to compare these distributions since it assumes our independent distributions are non-normal. Because we have large sample sizes for both O3 and O4, we specifically will compute a z-score from the Wilcoxon Rank Sum test. In order to construct our hypotheses around this test, we must again consider that O4 is expected to detect events that O3 was not capable of detecting and that O4 would be expected to detect O3 events with a higher significance. Again, we are unsure if the experimental improvements in O4 have resulted in the actual lowering of the detector's noise level, the amplification of signals over that noise level, or a combination both, but we expect that the mean of O4's SNR distribution to be greater than that of O3 regardless. If it's the case that only the noise level of the detector has been lowered, we would expect the events in O4 to have higher SNR values than those in O3, increasing the mean of O4's SNR distribution. If it's the case that the signals are being amplified over the same noise level, we again would expect O4's events to be detected with higher SNR values than in O3, which would also imply the increase in mean of O4's SNR distribution. It's obvious that if it's the case that both the noise level of the detector has been lowered *and* the gravitational wave signals themselves have been amplified over the noise, we would expect an even greater difference between O3 and O4's mean SNR. Regardless of how the experimental upgrades have effected LIGO's ability to detect gravitational waves, we therefore expect the mean of O4's SNR distribution to be greater than that of O3. More formally;

$$
\begin{aligned}
H_0 &: \mu_{\mathrm{O4}} - \mu_{\mathrm{O3}} \leq 0, \\
H_a &: \mu_{\mathrm{O4}} - \mu_{\mathrm{O3}} > 0.
\end{aligned}
\tag{3}
$$

After carrying the Wilcoxon Rank Sum test out with a type 1 error probability of 5%, we find:

Table 6: SNR Wilcoxon Rank Sum Test Results

Critical Value	Test Statistic	p-value
0.20199	1.95996	.41200

So we are led to not reject our null hypothesis in this case, which tells us that it's likely the case that the mean of O4's SNR distribution is less than or equal to that of O3's. Discussion on this result is necessary and follows in the next sections.

4 Discussion

4.1 Summary of Results

In this paper, we have analyzed statistical properties of O3 and O4 chirp mass and SNR distributions to consider how the experimental upgrades of O4 have changed LIGO's capabilities and the type of gravitational wave events able to be detected by it. We first demonstrated with relatively high statistical significance that the variance of the chirp mass distribution in O4 is greater than O3's by 1.6 to 3.7 times. Note that the probability of a type two error occurring in this study and therefore accepting a false null hypothesis, β, represents the chance of incorrectly claiming that the variance of O4's chirp mass distribution is less than or equal to that of O4 in this context. We then showed that the mean of O4's chirp mass distribution is greater than O3's by 17.69 to 21.27 solar mass units with very high statistical significance. A type two error would represent the chance of incorrectly claiming that the mean of O3's chirp mass distribution really is greater than or equal to that of O4's. We moved on to O3 and O4's SNR distributions to first demonstrate that O4's variance in SNR is 1.3 to 2.7 times greater than O3's with relatively high statistical significance. The type two error would represent the chance of incorrectly stating that the variance of O3's SNR distribution really is greater than or equal to that of O4's. We then finally demonstrated that the mean of O4's SNR distribution is surprisingly equal to that of O3's distribution, again with relatively high statistical significance. The type two error probability, which is important in this context, is the probability of accepting our finding when it is actually false (the probability of stating the mean of O4's SNR distribution is equal to that of O3's when it really is different). In summary, the findings of this paper can be written compactly in the following table. With 95% confidence, we have demonstrated:

Table 7: A summary of what we have demonstrated in this paper with 95% confidence

	s^2	μ
$\mathcal{M}$	O4 > O3	O4 > O3
SNR	O4 > O3	O4 = O3

4.2 Interpretation

The results of this paper, summarized in Table 7, clearly indicate both a difference between O3 and O4 in the black hole merger events as captured by the differing variance of the chirp mass distributions, and a difference in the detector's response to these events, as captured by the differing SNR distributions. It was expected that there would be noticeable differences between the data collected in O3 and O4 due to LIGO's major experimental upgrade, frequency-dependent light squeezing, so the results of this paper confirm these expectations. We did not expect the degree to which the chirp mass distributions would differ nor did we expect the result of the mean SNR analysis, however.

First, the results of this chirp mass analysis inform us that LIGO is in fact observing lower and higher mass events than O3 was able to when considering the *very* large difference in the variance of the two distributions. The larger variance of the O4 distribution indicates that LIGO is able to detect gravitational waves from a much wider range of mass values with its experimental upgrades. The fact that O4 is observing lower mass events than O3 should not be surprising even without much knowledge of the physics of gravitational wave emission. Lower mass events will radiate energy in gravitational waves at lower amplitudes than large mass events, which means the lower mass events will always be more difficult to detect. Why LIGO is also able to detect higher mass events than it was previously capable of with these upgrades may be less clear to the reader without a background in cosmology or astrophysics. As detailed in the analysis section, the black holes with the highest masses in our universe will also be some of its oldest since they have had longer spans of time to accumulate mass over and because the progenitor stars of these earlier black holes had much less metallicity since metal was very scare in the early universe, meaning they could grow much larger. The large difference in the mean of the O3 and O4 distributions found in this study additionally supports the interpretation that O4 is now able to see more high-mass merger events and is therefore very likely able to see further back into the universe. Further analysis would need to be carried out to determine if this large difference in means is due to a few very massive events detected in O4 or is due to a high volume of events with relatively high mass. In either case, we can be fairly confident that the large differences in the variance and the mean of the chirp mass distributions. It wouldn't be expected to have been caused by O4 just happening to observe more high-mass events that O3 would have also been capable of observing since we have shown that these differences have a high statistical significance. A similar statistical analysis would need to be carried out on the actual distance distributions of O3 and O4 when the data is publicly available to more confidently state this result, though. The equally long spans of observing time of O3 and O4 provides additional assurance that these differences are very likely not just due to the random chance that O4 happened to be operating during the time gravitational waves from high-mass merger events reached the Earth.

Lastly, the results of the SNR analysis inform us that, with less likelihood than the chirp mass analyses, the variance of the SNR of events detected has increased in O4 but the mean of this metric has not changed. This tells us that the distribution in SNR of the events detected in O4 has not moved by a statistically significant amount relative to O3 but the width of this distribution has been shown to have increased in O4. The difference in the variance of these distributions tells us that LIGO is able to detect events with lower SNR values than O3 was capable of *and* is detecting the events that O3 would have detected with greater confidence. We believe that the fact that the mean of these distributions did not change by a statistically significant amount but the variance did still confirms our original predictions in an unexpected way. Since the SNR is a metric for how energetic a signal appears to us on Earth, this result tells us that the average observed strength of these signals has not changed. We do still believe that we can state with some confidence that LIGO is able to see events of greater distances and therefore larger masses from the mass analysis. My assumption is that if an event were somehow able to be observed in both O3 and O4, it would have a higher SNR in O4, thus shifting to the right on the SNR space. This would explain why the variance of the SNR distribution has increased; quiet events in O3 are now being observed with greater confidence in O4, which increases LIGO's chances of detecting them and therefore increases the number of these events that were quiet in O3 increases. To keep the mean SNR fixed, this tells me is that LIGO is detecting events with much lower SNR than was possible in O3; meaning the noise floor as discussed earlier has been lowered in O4. We believe that the likely unchanged mean SNR between O3 and O4 is a combination of *both* events that were able to be detected in O3 being detected in O4 with greater confidence *and* events with too small a signal to be detected in O3 due to the detector's noise can now be detected in O4, albeit with a very low SNR value. This result seems to imply that LIGO's recent experimental upgrades have in fact lowered the noise level of the detector, rather than amplifying the strength of gravitational wave signals, but a more full literature review on these upgrades is necessary to better inform these findings.

The largest flaw of this study, as briefly discussed previously, is due to the fact that we assumed throughout that all of the data was coming from black hole-black hole (BBH) merger events, when we know this is not true. The number of neutron star-neutron star (NSNS) and neutron star-black hole (NSBH) merger events that have been recorded in LIGO's history is small but this is an incorrect assumption that will slightly misrepresent the true black hole-black hole population statistics of O3 and O4, especially since the mass properties of the BNS events will be very different than the BBH events and because these events are known to interact with the detector differently. A more complete analysis would require the removal of the known BNS and NSBH events from the O3 and O4 data so inferences could be drawn about the BBH populations in O3 and O4 in isolation. An interesting extension of this work would also come out of this removal, for one would then have three distinct populations of O3 and O4 data. It would make for an interesting project to see if the large differences we have shown between the full O3 and O4 distributions would also hold for the separated data. We also know that the results of our SNR variance analysis should not be accepted with much confidence since we know that at least one of the distributions is non-normal but we applied a statistical test that assumes both distributions are normal, the F-test. Applying a similar test that doesn't make such an assumption wouldn't be expected to change the finding of the analysis, though but only be expected to change the confidence of our finding.

Our finding that the mean of O4's SNR distribution is less than or equal to that of O3 would have been a completely unexpected result given the original reasoning that went into the construction of this hypothesis no matter what the type of positive effect of LIGO's experimental upgrades had on the detector. By just comparing the computed the means for O3 and O4's SNR distribution from Table 1 and Table 2, however, we see that this should not have been as shocking of a result.

5 Future Work

For additional future work, the most obvious improvement of this study would be to apply a non-parametric test for testing the difference in variances of the O3 and O4 distributions since we know that the O4 SNR distribution is not normally-distributed from (d) of Fig. 1 but we have have applied a test that assumes the underlying distributions are normal. This would likely lead to more confident findings than presented in this work. Because O4 is currently in progress as of the time of this writing, future work may repeat this paper's analyses with additional data from O4 as it continues to detect gravitational waves to determine if the results of this study change with the addition of more O4 data, especially considering the variance of the SNR distribution. Finally, we have completely ignored the data from LIGO's previous observing runs, O1 and O2, in this study. Though there were not as many detections made in these observing runs as in O3 and O4, applying the tests of this study to the data that we have not considered could reveal an interesting trend. Just as it was expected that O4 would be capable of detecting events that O3 was not able to and detecting events O3 *was* able to with more significance, the same results would also be expected when comparing O1 to O2 data and O2 to O3 data. Having a chronological study of all of LIGO's data in this way would be a great demonstration of its improved sensitive and capabilities over time and could possibly inform the collaboration of which upgrades have led to the most dramatic improvements.

A more full literature review is also necessary to understand the improved nature of the detector. Such knowledge would inform the results of this study and would be very useful for interpreting the mean SNR result, as discussed previously. The SNR results seem to imply that it is the actual noise level of the detector that has been lowered rather than the amplification of gravitational wave signals over this noise. If this is in fact what LIGO's experimental upgrades have allowed O4, this would greatly increase the confidence of these findings and the interpretation of them.

References

[1] Diane E. Banegas. “LIGO: The Search for Gravitational Waves”. `https://web.archive.org/web/20160915004424/https://www.nsf.gov/news/news_summ.jsp?cntn_id=103042`. 2008.

[2] Albert Einstein. “Näherungsweise Integration der Feldgleichungen der Gravitation”. In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (Jan. 1916), pp. 688–696.

[3] Werner Israel. “Event Horizons in Static Vacuum Space-Times”. In: *Phys. Rev.* 164 (5 Dec. 1967), pp. 1776–1779.

[4] LIGO Scientific Collaboration. *Gravitational Wave Candidate Event Database (GraceDB).* `https://gracedb.ligo.org/`. 2024.

[5] LIGO Scientific Collaboration. *LIGO Surpasses the Quantum Limit.* `https://www.ligo.caltech.edu/news/ligo20231023`. 2023.

[6] Lucas Peterson, Andrew Valentini, Kaitlyn Prokup, Hiroki Imura. *Estimating the Luminosity Distance and Mass Properties of BBH Merger Events in LIGO O4 Data.* 2024.

[7] Micheal T. Longnecker R. Lyman Ott. *An Introduction to Statistical Methods and Data Analysis.* Cengage Learning, 2015.

[8] Peter R. Saulson. *Fundamentals of Interferometric Gravitational Wave Detectors.* World Scientific Pub Co Inc, 2017.

[9] J. H. Taylor and J. M. Weisberg. “A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16”. In: 253 (Feb. 1982), pp. 908–920.