Response to referees' comments for manuscript

Dear Editor,

Please find enclosed a revised version of our manuscript with the title *"Comprehensive analysis of time-domain overlapping gravitational wave transients: A Lensing Study"*. The authors would like to thank the referee for their substantial comments and suggestions, which significantly contributed to improving the quality of this manuscript. We have revised the manuscript carefully and thoroughly, the major changes in the manuscript are labeled in blue. In the following, we first summarize the main changes in the revised manuscript, and then present itemized detailed responses/corrections to all the referees' comments (all authors' responses here are in blue).

Major Changes

1. We have expanded **Sec. II (Gravitational Lensing)** and **Sec IV. (Setup of the Analyses)** to provide intuition and clarity for the lensing models.

2. We have added a discussion in **Sec. V (Results)** regarding the physical constraints of the point-mass lens model, which enforces positive time delays.

3. Following the referee's suggestion, we have added a discussion in **Sec. VI (Conclusions)** emphasizing that spatial localization serves as a crucial first-step discriminator before attributing lensing features.

4. We have added units in the plots and the main text for the impact parameter defined in terms of the Einstein radius.

Response to Referee

1. *In Fig 1. After the ringdown in the unlensed signals, there appears to be a low frequency modulation. I am not sure if this is physical, and could present issues in adding another modulation to the lensed signals in the case of overlapping image.*

   Response: We thank the referee for careful observation. To clarify our setup, we inject overlapping pairs of gravitational wave signals and infer them using a single

unlensed image. Therefore, the modulation you mentioned pertains to both the over-lapping signals and lensing, rather than a lensed overlapping image. The modulation in the signal after the ringdown pertains to an unphysical issue that would not contribute to the biases observed. The pertaining low-frequency behavior is due to the minimum frequency set at $f_{\min} = 20$ Hz for the waveform generation of $\mathrm{SINGLES}_{\mathrm{A/B}}$, which results in the Gibbs phenomenon as a step-function in the frequency domain, corresponding to a decayed waveform at that frequency. When examining the matches by windowing this segment, we observe a mismatch difference of $\mathcal{O}(10^{-5})$ between the original waveform and the gated waveform, and hence it would not affect our results. A corresponding notebook is available at this link for clarity.

2. *It might be good to have some discussion of the choices of* $\mathrm{SNR_B}/\mathrm{SNR_A}$ *and time delays that are being sampled from, because they could seem a bit arbitrary. While the SNR ratios seem fine given the limits for a point caustic you have in your lens model, however, the arrival time differences for these caustics should always be positive. While you can generate these configurations in structure-rich lenses (see* `https://arxiv.org/abs/2510.14953` *for instance), it would be good to have some discussion as to why these choices of parameters are made.*

   Response: We agree with the referee that for a standard single-lens system, the time delays between images are physically constrained to be positive relative to the primary image. However, the primary focus of this work is to inject *randomly overlapping* independent binaries (where Δt_{c} can be arbitrarily positive or negative) and test if they can *mimic* a lensed signal. We purposely explore negative Δt_{c} (where the quieter signal arrives first) to see if the lens models (which enforce specific time orderings) yield false positives in these unphysical regimes. Also, by exploring $\Delta t_{\mathrm{c}} < 0$, we explore the possibility where the weaker SNR signal coalesces before the higher SNR signal, which can be (falsely) interpreted as magnification of the image in the geometric optics limit of lensing.

3. *Related to the previous point, there should probably be some mention of the specifics of the lens models and how they are being used in the study to recover the lens parameters. The current text relies heavily on the reader's previous knowledge of lens models. Having some of the basics of the lens model spelled out would greatly help readers gain*

intuition on the relation of the searches back to the lensing parameters being recovered, and their various implications.

Response: We thank the referee for this valuable suggestion. We agree that explicitly defining the lens models aids in interpreting the parameter recovery results. We have revised **Sec. II (Gravitational Lensing)** and **Sec IV. (Setup of the Analyses)** to include the explicit mathematical formalism used in our analysis.

4. *In page 4 paragraph 2, you state that a point lens can create image pairs with a $\pi/2$ phase shift between them. This is true, but the point lens cannot create image pairs where the type II image arrives before the type I image. This, along with the implications of the validity of using a point lens model to recover lensed signals in a regime in which it cannot produce these image configurations should be mentioned in the text. In other words, you might be trying to infer lens parameters with a lens configuration that is not allowed by the lens model you have chosen (namely the $\Delta t_{\rm c} < 0$). You discuss this in the microlensing section, where you say that microlensing can only produce $\Delta t_{\rm c} > 0$, and thus fits lensed signals with this property well. However, the same is true in geometric optics.*

Response: We have added a discussion in **Sec. V (Results)** explicitly acknowledging that the lens model imposes this time ordering, and clarified in **Sec. II (Gravitational Lensing)**.

5. *In Fig. 3 (left panel), it is not obvious to me why there is a much higher bias for positive time delays than negative time delays, and why this is mostly seen for high values of $\mathscr{M}_B/\mathscr{M}_A$ and $\rm SNR_B/SNR_A$. This could just be my lack of understanding, but it might be good to directly address this in the text.*

Response: We have clarified this behavior in the revised manuscript in **Sec. V.A.2 (Microlensed Model)**. When we inject an overlapping signal where the secondary (weaker) signal arrives later, the morphology is consistent with the expected characteristics of the lens model. The lensed model can better fit the superposition, leading to higher Bayes factors, and biased parameter inference is observed. High mass ratios and comparable SNRs produce a distinct, loud secondary signal with a specific chirp profile. When this distinct secondary signal arrives *after* the primary ($\Delta t_{\rm c} > 0$), it pro-

vides strong features that the microlensing template can degenerate with, mimicking a second image with comparable SNR to the primary.

6. *I think that one topic that really should be brought up in the paper is comparing the sky localization maps of these events as a means of rejecting or supporting lensing claims, especially in future detectors where the localizations will be greatly improved. While I agree that it should be expected that isolated quasi-circular events arriving at the detector nearby in time with comparable chirp masses and SNRs will appear to be lensed in many cases, but access to the location on the sky from which they are coming from seems like a natural first step to begin to break this degeneracy. While providing quantitative information of how well you could use this to break the degeneracies shown in this work is most likely outside of the scope of this work, I believe that it would improve the final messaging of the conclusions, which understandably ends on a bit of a somber note due to the findings of the analysis without considering the sky localizations.*

Response: We agree with the referee. While this study isolates the time-domain degeneracy in the presence of strong overlaps, spatial localization is indeed a powerful veto for widely separated signals. True lensed images must originate from the same sky location, whereas random overlaps will likely have disjoint sky posteriors. However, this would primarily be necessary as a tool before attributing it to lensing. In our study, we inject overlapping pairs and infer using a single image template; thereby, we would not be able to infer two different sky localizations, attributing them to two signals. This would be particularly relevant in the strong overlap regime ($\Delta t_\mathrm{c} \lesssim 0.1$ s) that we considered, and search pipelines would report on a single signal due to the trigger window. Furthermore, when the second signal is sub-threshold, the detection mechanisms would fail, and we might still observe false degeneracies due to lensing. We have added a dedicated paragraph in the **Conclusions (Sec. VI)** emphasizing that sky localization will serve as a crucial discriminator to break the degeneracies identified in this work, particularly with the improved localization of future detector networks.