A GIS-based workflow for the remote evaluation of potential Cosmic Explorer sites

Tooba Ansar[1], Robert Schofield[2], Michael Landry[2]
[1]*Amherst College, Amherst, MA 01002*
[2]*LIGO Hanford Observatory, Richland, WA 99352*
(Dated: October 21, 2025)

Cosmic Explorer is a proposed third-generation gravitational wave observatory that will offer an order-of-magnitude improvement in broadband sensitivity over current gravitational-wave detectors. This leap in sensitivity results primarily from upscaling the LIGO arm length by a factor of ten. With its 40 km-long arms, Cosmic Explorer will detect gravitational wave sources from across the universe that remain unresolved by existing detectors. Selecting an optimal site for the observatory is both critical and challenging. Cosmic Explorer requires significant land clearance to host a detector of its scale, as well as minimal sources of ambient noise to achieve its high sensitivity. We present a Python-based workflow that integrates Geographic Information Systems (GIS) with publicly available national datasets on land ownership, transportation lines, and anthropogenic noise sources such as wind turbines and active mines to automate remote analysis of potential sites. Preliminary runs of our workflow evaluated nearly 6,000 sites in the contiguous United States that could physically host a 40 km interferometer, eliminated thousands of those due to limited land availability or proximity to noise sources, and identified a few hundred most promising for eventual on-site testing. This approach is scalable and cost-efficient and can be further developed to include additional site evaluation criteria, ultimately accelerating the site selection process for Cosmic Explorer.

I. INTRODUCTION

Current gravitational wave detectors are limited in their sensitivity by the length of their interferometer arms. Since the amplitude of a passing gravitational wave is characterized by the strain or the fractional change in arm length it induces, increasing the arm length increases the strength of the signal relative to noise sources [1]. The sensitivity of a gravitational wave detector is therefore positively correlated with its arm length.

In spite of the constraints on their sensitivity, current ground-based gravitational wave detectors have been widely successful, enabling the first observation of a binary black hole merger and a coalescing neutron star and establishing gravitational wave astronomy as a novel way of observing the universe [2, 3]. The next step in the advancement of ground-based gravitational wave astronomy would then be to improve the sensitivity of this current, proven technology. With an order-of-magnitude increase in arm length over the Advanced LIGO detectors, the proposed third-generation Cosmic Explorer observatory will be ten times as sensitive as Advanced LIGO, offering this improvement in sensitivity [4–7].

Much like Advanced LIGO, the Cosmic Explorer concept consists of two L-shaped observatories at two widely separated locations in the United States, where each of the observatories will house a Michelson interferometer. However, while the LIGO interferometer consists of two 4 km-long arms, Cosmic Explorer will consist of one detector with 40 km-long arms and another detector with 20 km-long arms [4].

Although this extension of the arm length is the main reason for its order-of-magnitude improvement in sensitivity over current detectors, choosing a site most suitable for an observatory of its scale is also uniquely challenging. While many locations in the United States can topologically and geologically host 20 and 40 km detectors, it is important to choose a site with minimal ambient noise. Gravitational waves are detected by measuring displacements on the order of 10^{-19} m in the interferometer arms, and ambient noise can therefore obscure these minute displacements and limit the sensitivity of Cosmic Explorer [7, 8]. There are several sources of ambient noise:

(1) Ground motion such as seismic activity can cause the detector parts to vibrate and result in detector noise. It can also lead to fluctuations in the local gravitational field, generating what is known as local gravity noise. Seismic noise in the 1-30 Hz range can potentially couple into the detector mechanically. In general, the best sites for Cosmic Explorer are those with low ground motion across the 10 mHz to 30 Hz frequency band.

(2) Acoustic noise couples into the detector through direct modulation of the vacuum envelope and the associated motion of the detector components, which in turn modulates scattered light within the detector. In addition, acoustic waves can change the air pressure near the detector and cause fluctuations in the local gravitational field to introduce local gravity noise. The current sensitivity model for Cosmic Explorer assumes typical ambient acoustic levels of 1 mPa/$\sqrt{\text{Hz}}$ [4]. To properly assess the suitability of a potential location, we must evaluate the ambient acoustics of the site.

(3) Electromagnetic noise can interfere with Cosmic Explorer's sensitive frequency band; radio frequency interference can disrupt the modulation and demodulation techniques used for controlling Cosmic Explorer's interferometer.

(4) Environmental factors include weather conditions such as wind and thunderstorms and anthropogenic sources such as vehicles, air traffic, and rail lines, both

of which can induce noise. It is critical to study seasonal weather changes over year-long timescales, as well as to investigate the susceptibility of potential sites to natural disasters such as earthquakes and hurricanes.

Preliminary evaluation of Cosmic Explorer sites began in the mid-2020s [5]. Prior work has involved a national-scale suitability analysis using GIS. Thousands of sites in the United States have been identified that can physically fit a 40 km L-shaped interferometer using a geometric algorithm that integrates 91 different social, cultural, physical, and scientific parameters. The various potential sites are then scored based on a cost function that incorporates the geology, geography, and topography of the land. The results of this analysis produce clusters of potential detector configurations within promising areas of interest in the United States [9]. Fig. 1 shows such a cluster.

However, a national-scale suitability analysis does not account for data at the site-specific level [9]. For example, a 40 km observatory requires extensive land clearance, and many of the identified configurations intercept buildings or transportation lines at various points along the beam tubes.

Given the large number of candidate sites that have been identified in prior site selection stages, conducting field surveys at each of the sites to evaluate noise levels and ensure significant stretches of land clearance would be both resource-intensive and time-consuming. Initial attempts at a remote approach involved visual identification of noise sources and land availability using satellite imagery. However, a significant challenge with this approach was the manual, repetitive, and cumbersome nature of the work and the great potential for human error. Therefore, it is critical to develop a scalable, automated approach to remotely identify noise sources and survey 40 km of land clearance at the site-specific level. This will allow for the identification of a few sites that seem most promising in meeting or exceeding Cosmic Explorer's target sensitivity. These candidates will then be evaluated on-site using seismometers, magnetometers, and microphones.

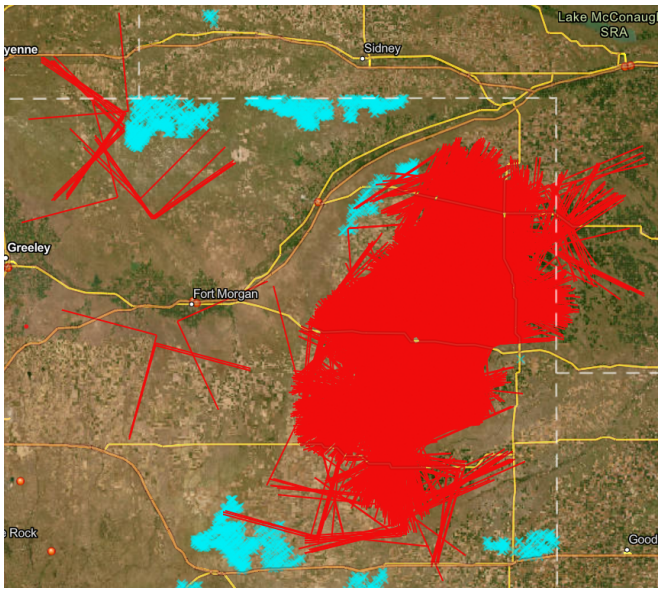


FIG. 1. Potential configurations of the 40-km L-shaped detector at a candidate location in Colorado shown in red. Many such locations have been identified around the United States, with hundreds of similarly clustered configurations.

II. METHODS

A. Parameters for Site-Level Analysis

In our analysis, we focus on sites for the 40 km Cosmic Explorer observatory. Given the scale of a 40 km facility, sufficient land availability is both a major consideration and challenge. Having more land owners in the vicinity of a potential site would require greater negotiation [9]. Buildings are therefore best avoided. However, a 40 km L-shaped configuration in most parts of the United States is likely to intercept a building along its 40 km length.

Vehicles, rails, active mines, and wind turbines can introduce noise both seismically and acoustically, while power lines can introduce electromagnetic noise. It is therefore important to consider the prevalence of these anthropogenic sources of noise around a site, particularly close to the three end stations (namely the X-end, Y-end, and corner stations) which are most sensitive to ambient noise.

B. Approach

For our analysis, we were provided with Keyhole Markup Language (KML) files from the national suitability study. KML is a geospatial data format used to encode geographic features and their coordinates and attributes, allowing visualization and analysis on GIS platforms. Each KML contained a collection of 40 km polylines representing potential detector configurations within an area of interest, with the two detector arms of each site defined as distinct line features. The files also included extended metadata fields such as arm length, opening and rotation angles, and site-specific suitability metrics including elevation, tilt, and total composite score.

Since the inputs for our analysis were KML files, employing a GIS platform for site-level evaluation was the most natural choice. GIS platforms are well suited for remote site evaluation owing to their integration of remotely-sensed imagery with spatial datasets. We chose ArcGIS as our platform for site evaluation due to its geospatial analysis capabilities, native Python libraries for scripting and automation, and the availability of diverse spatial datasets.

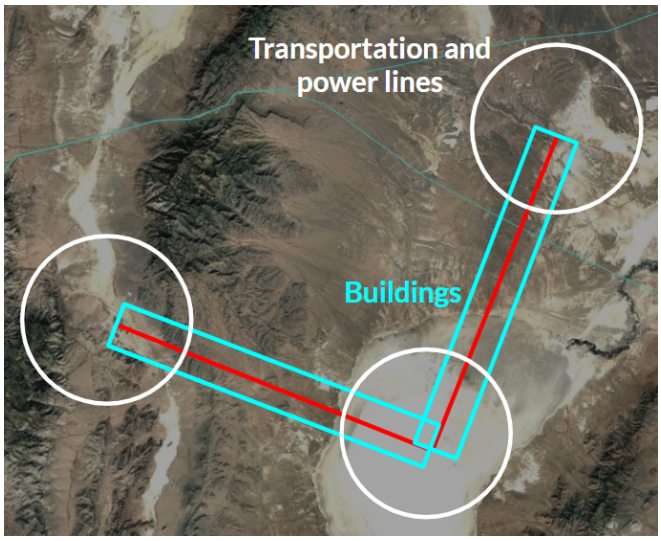


FIG. 2. Depiction of a potential site (in red) showing the regions for the identification of site evaluation parameters. The circular regions around the three stations represent the area that must be inspected for the prevalence of noise sources. The rectangular extents enclosing the beam tubes show the regions that must be surveyed for potential obstructions. Note that the L shape shown is not a real potential site and the depiction of the arm and extents is not to scale.

To implement our site evaluation workflow, we developed an Application Programming Interface (API) using the ArcGIS API for Python. We chose to develop an ArcGIS API mainly because it allowed for web-based sharing of maps within and between organizations via ArcGIS Online, in addition to being substantially more cost-effective than alternative proprietary GIS software. Web-based sharing is a crucial feature given the size of the Cosmic Explorer consortium.

Fig. 2 depicts the general approach to the identification of parameters for site-level analysis. Looking for sufficient land clearance involves specifying a clearance zone around the 40 km beam tubes to identify any buildings that are close to or intercept the arms of a potential site. This clearance zone can also inspect for anthropogenic noise sources such as active mines and wind turbines. To estimate vehicular traffic and determine the prevalence of power lines near the highly sensitive end stations, the total kilometers of transportation and power lines within a certain distance of the stations can be computed and compared across sites.

C. ArcGIS API Workflow

Before explaining the pipeline for our API, we must introduce and define certain terminology specific to GIS platforms.

- A *layer* is a collection of spatial data that represents geographic features.
- *Vector data* is a way to represent spatial data using geometric shapes such as points, lines, and polygons. Thus, in the case of discrete geographic features such as buildings and roads, a layer stores vector data in the form of points and lines, respectively.
- A grouping of similar geographic features is referred to as a *feature layer*. For instance, all the buildings in the United States would be stored in a feature layer as vector data of points. A building feature layer would then be called a *point feature layer*.
- A *buffer* is a geospatial analysis tool that defines a polygon of a specified distance around a geographic feature, allowing for analysis of features that are within the buffer zone.

We designed a two-stage API workflow. A schematic of the workflow is depicted in Fig. 3. The inputs to the API are KML files containing several hundreds to thousands of polylines representing various detector configurations within a cluster. The KML files are first converted to shapefiles (a file format for storing geographic data) and then stored as line feature layers in ArcGIS Online.

The first stage of screening involves identifying and eliminating configurations that intersect buildings within the buffer zones. The API creates a rectangular buffer of configurable width along each line feature to characterize the required land clearance zone, storing these buffers

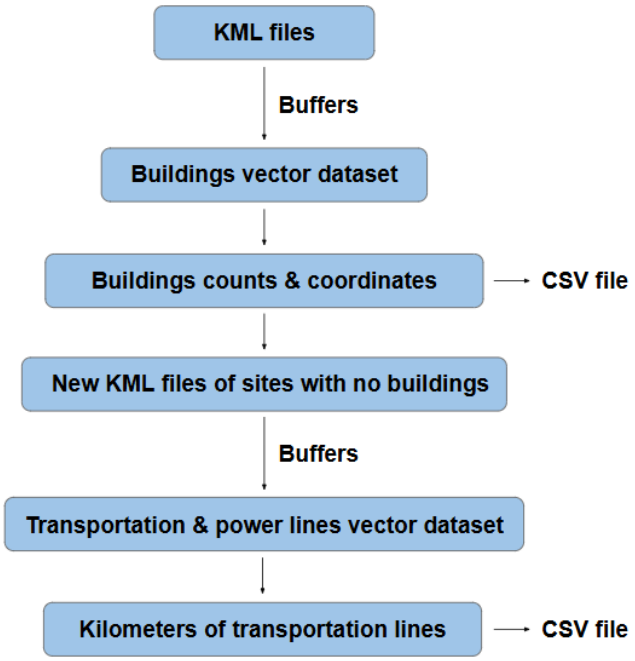


FIG. 3. Schematic defining the workflow for the ArcGIS API. KML files are taken in as inputs. Buffers are defined to characterize clearance zones around each arm. New KML files are exported after the first screening stage. Detailed results are appended to CSV files.

as hosted feature layers. Then, a national-scale vector dataset of all building footprints larger than 450 feet is loaded as a point feature layer and spatially indexed for rapid querying. A spatial intersection is then performed between each arm's buffer and the building layer to identify any overlaps.

For each site with an intersecting building, the script records the total number of buildings and their centroid coordinates and writes them to a CSV file. Sites intersecting one or more buildings within the buffer are excluded from further analysis. The original KML files are therefore "filtered" to only include those sites with no buildings within the buffer zones, and these new KML files are exported for subsequent screening.

For the second stage, circular buffers of configurable radii are created around each of the three end stations to define zones that are particularly sensitive to noise. To allow for distance calculations within the buffer, the latitudes and longitudes of each buffer are reprojected using an azimuthal equidistant projection centered on the centroid of the buffer. National datasets for roads, highways, rail lines and power lines are loaded as line feature layers, and the API then queries the intersection of the buffer with the datasets to delineate all the power and transportation lines enclosed within the buffer.

For each candidate site, we then compute the total length of transportation lines within the circular buffer of the detector. These results are exported to a CSV file for comparison across sites.

It is important to note that the two-stage API presented here need not be implemented rigidly in that order. The script for the first stage, which in our implementation identifies buildings, can be used to identify and retrieve counts and coordinates for any dataset represented as a point feature layer. Similarly, the second stage can be implemented to compute lengths of any line feature layer. In our case, we ran the API on point feature layers for wind turbines and active mines across the sites that contained no buildings retrieved from the KML files used in this study, but the results were trivial because there were no turbines or mines in close proximity to those sites.

III. INITIAL RUNS AND RESULTS

To validate the workflow, we ran the API on 7 different KML files of two different clusters in Idaho and Colorado. Each cluster contained several hundred to several thousand potential 40 km detector configurations. The goal was to test whether the API could successfully identify configurations with sufficient land availability and compute distances on line feature layers.

The first stage of the analysis identified and excluded sites that intersected existing buildings within a specified clearance zone. A successful run with an identified building is shown in Fig. 4. Of 5,931 total configurations, approximately 90% of the sites were eliminated during this stage due to building intersections, significantly reducing the number of candidate sites for further evaluation. A detailed summary of the results of the first stage is given in Table I.

TABLE I. Table detailing the number of buildings identified within the multiple KML files associated with each of the two clusters in Idaho and Colorado.

	Number of sites	Sites with no buildings
Idaho	2183	598
Colorado	3748	12
Total	5931	610

To verify that the identification was correct, some sites were visually inspected for buildings at random. To validate that the extracted coordinates accurately corresponded to the right building within the buffer of the right configuration, we used the Map Viewer feature in ArcGIS Online. The number of buildings counted by the API within certain buffers made it evident as to why visual inspection of satellite imagery needed to be replaced with an automated framework. A specific configuration in Idaho, for instance, intersected 207 buildings, and visually inspecting and counting these buildings would have been extremely challenging.

Following building identification, the original KML files were filtered to only include configurations with no buildings. Fig. 5 shows the results of filtering on the Idaho cluster.

The second stage computed the distance of transportation networks and power lines close to the end stations. Fig. 6 shows an example of all the gravel roads identified within a buffer. To validate these distance calcu-

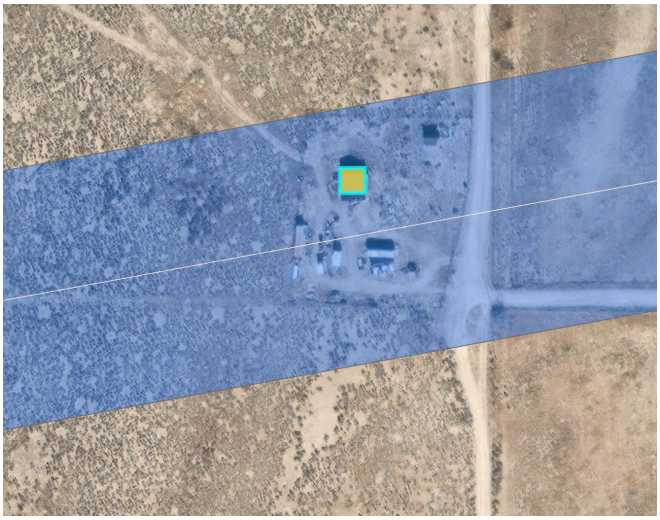
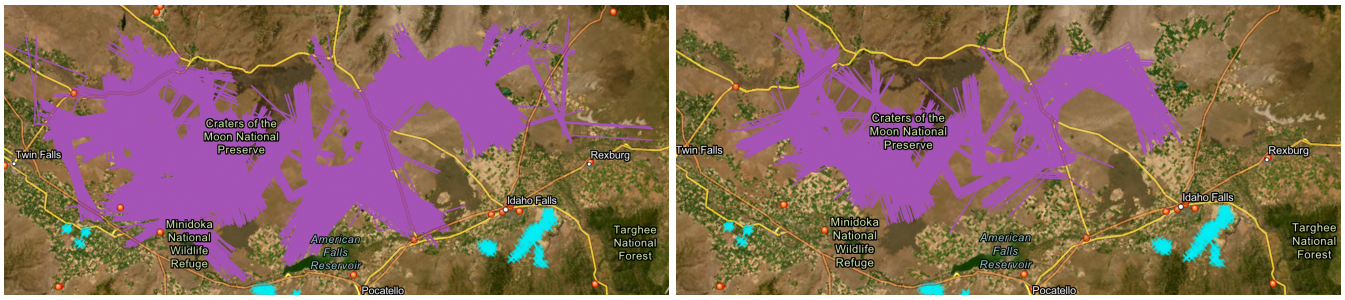


FIG. 4. A building near a detector arm successfully identified on the vector dataset of buildings in the United States. The white line feature represents the detector arm while the blue delineates our buffer, which is, in this case, 200 m on either side of the arm.



(a) Idaho cluster with all 2,183 site configurations.

(b) Idaho cluster "filtered" to 598 sites.

FIG. 5. Idaho cluster before and after all the configurations that intersected at least one building were eliminated from the original KML file. The original cluster had 2,183 sites and 1,585 of these sites intersected a building, leaving the cluster with 598 sites.

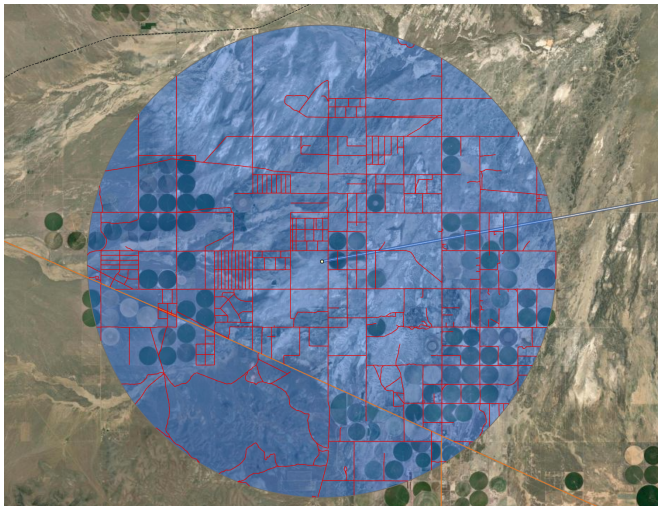


FIG. 6. The gravel roads within a circular buffer of radius 10 km around an end station mapped in red. The total distances of these roads are calculated after identification.

lations, we corroborated the results with the results retrieved from the Summarize Attributes feature available in ArcGIS Pro.

IV. OUTLOOK

The current implementation of the API can be extended by incorporating additional site evaluation criteria beyond those included in this preliminary study. Certain geographic areas and features count as exemption zones and must be avoided altogether. For example, wild and scenic rivers are safeguarded by the law, and sites in the vicinity can be identified and eliminated using our script that handles line feature layers. Similarly, vector datasets for oil fields can be integrated into the script that assesses point feature layers.

We used distance calculations on transportation lines in lieu of vehicular traffic. A path forward in this case is to incorporate a deep learning model that can identify and count vehicles on satellite images. While we deployed pre-trained deep learning models for vehicle identification and counting, we ran into challenges with the runtime, input raster resolution, and accuracy of the models. Fine-tuning a pre-trained model and integrating it into the API could be a promising step forward.

The current implementation uses a more binary system for site evaluation, such as eliminating all sites that intersect one or more buildings. A more robust approach could assign relative weights to each site evaluation criterion to determine a suitability score for each site. This would allow for a more quantitative comparison of all the sites.

V. CONCLUSION

We developed a Python-based API for ArcGIS to automate the remote evaluation of candidate sites for the proposed 40 km Cosmic Explorer gravitational-wave observatory. The API identifies potential sites with sufficient land clearance and minimal proximity to anthropogenic noise sources. Preliminary tests on clusters in Idaho and Colorado successfully filtered thousands of candidate configurations down to a few hundred, demonstrating that this approach significantly reduces the number of sites requiring on-site testing.

VI. ACKNOWLEDGMENTS

I would like to thank my mentors Robert Schofield and Michael Landry for their guidance throughout the project. This work was supported by the National Science Foundation as part of the Caltech LIGO Summer Undergraduate Research Fellowship Program.

[1] S. Dwyer, D. Sigg, S. W. Ballmer, L. Barsotti, N. Mavalvala, and M. Evans, Gravitational wave detector with cosmological reach, Phys. Rev. D **91**, 082001 (2015).

[2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116**, 061102 (2016).

[3] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, Phys. Rev. Lett. **119**, 161101 (2017).

[4] M. Evans, R. X. Adhikari, C. Afle, S. W. Ballmer, S. Biscoveanu, S. Borhanian, D. A. Brown, Y. Chen, R. Eisenstein, *et al.*, A Horizon Study for Cosmic Explorer: Science, Observatories, and Community (2021), arXiv:2109.09882 [astro-ph.IM].

[5] M. Evans, A. Corsi, C. Afle, A. Ananyeva, K. G. Arun, S. Ballmer, A. Bandopadhyay, L. Barsotti, M. Baryakhtar, E. Berger, E. Berti, S. Biscoveanu, S. Borhanian, F. Broekgaarden, D. A. Brown, C. Cahillane, L. Campbell, H.-Y. Chen, K. J. Daniel, A. Dhani, J. C. Driggers, *et al.*, Cosmic Explorer: A Submission to the NSF MPSAC ngGW Subcommittee (2023), arXiv:2306.13745 [astro-ph.IM].

[6] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, *et al.*, Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO (2019), arXiv:1907.04833 [astro-ph.IM].

[7] K. J. Daniel, J. R. Smith, S. Ballmer, W. Bristol, J. C. Driggers, A. Effler, M. Evans, J. Hoover, K. Kuns, M. Landry, *et al.*, Criteria for identifying and evaluating locations that could potentially host the Cosmic Explorer observatories, Rev. Sci. Instrum. **96**, 014502 (2025).

[8] E. D. Hall, K. Kuns, J. R. Smith, Y. Bai, C. Wipf, S. Biscans, R. X. Adhikari, K. Arai, S. Ballmer, L. Barsotti, Y. Chen, M. Evans, P. Fritschel, J. Harms, B. Kamai, J. G. Rollins, D. Shoemaker, B. J. J. Slagmolen, R. Weiss, and H. Yamamoto, Gravitational-wave physics with Cosmic Explorer: Limits to low-frequency sensitivity, Phys. Rev. D **103**, 122004 (2021).

[9] W. Bristol, C. Lukinbeal, J. Hoover, K. Daniel;, P. Sledge, J. Russell, and M. Evans, An Interdisciplinary Site Suitability Analysis for Cosmic Explorer, in *24th International Conference on General Relativity and Gravitation (GR24) and 16th Edoardo Amaldi Conference on Gravitational (Amaldi16) Waves* (2025).