JOURNAL of
UNDERGRADUATE
RESEARCH IN PHYSICS
AND ASTRONOMY

VOLUME XXXV
ISSUE 1

2026

JURPA

A PUBLICATION OF THE SOCIETY OF PHYSICS STUDENTS
Available online at **students.aip.org/jurpa**

JURPA

SOCIETY OF PHYSICS STUDENTS

1 Physics Ellipse
College Park, MD 20740-3841

301.209.3007 (tel)
301.209.0839 (fax)
sps@aip.org
www.spsnational.org

Find SPS on:

Table of Contents

2026 • VOLUME XXXV, ISSUE 1

Journal of Undergraduate Research in Physics and Astronomy (ISSN 0731-3764) is a publication of the Society of Physics Students, published annually by the American Institute of Physics. Printed in the USA. POSTMASTER: Send address changes to JURPA, One Physics Ellipse, College Park, MD 20740-3841.

AIP

Letter from the Outgoing Editor

by Will Slaton, University of Central Arkansas

Will Slaton.

When I became editor of the *Journal for Undergraduate Reports in Physics* in 2018, I inherited a publication with a clear purpose: to give undergraduate researchers a venue worthy of their work. Nine years and 97 published articles later, I am stepping down as editor of what is now the *Journal for Undergraduate Research in Physics & Astronomy*, and I am leaving with no reservations about the health of this journal or the future of undergraduate research in our field.

The most visible change during this period came in June 2023, when the SPS and Sigma Pi Sigma Executive Committee voted unanimously to expand the journal's scope to include astronomy, formalizing something the submission record was already showing us. Scan the past three volumes and you will find studies of fast radio bursts and brown dwarf variability, machine-learning approaches to galaxy merger classification, exoplanet detection through Kepler data, and coronal mass ejection analysis. Astronomy belongs here, and its presence will only grow.

The breadth of work across all nine volumes has been genuinely remarkable. Three papers capture the range well: "Analyzing Kepler Data Through Machine Learning: A Study of Neural Networks for Exoplanet Detection" (2026) bridges computational methods and observational astronomy; "Analysis of Mechanical Analog for Understanding ACL Injuries" (2025) shows how physics illuminates everyday biomechanics; and "The Rotation Curve of the Milky Way Galaxy as Evidence for Dark Matter" (2022) takes on one of the field's most compelling open questions. What connects these papers is not subject matter but standard—original research, conducted with a faculty mentor, communicated with the rigor the scientific community expects.

I have long believed that undergraduate research is among the most valuable educational experiences we offer. A student who has worked through an original problem, encountered difficulties along the way, revised their methods, and written up their findings for peer review has learned something no course can fully teach. They have practiced science rather than studied it. These experiences strengthen graduate and professional school applications, build genuine confidence, and develop the computational, theoretical, and experimental fluency that careers in physics and astronomy demand.

The students who published in JURP and JURPA went on to doctoral programs, professional schools, and careers built on the habits of mind that research instills. That trajectory is not coincidental. Research experience changes what students believe they are capable of, and publishing raises the ceiling further. Watching that happen, repeatedly, over nine years has been the most satisfying part of this role.

I want to close with acknowledgments that are long overdue. Kendra Redmond, managing editor of the *SPS Observer* and *Radiations*, provided tireless editing of JURPA manuscripts, making us all look better than we perhaps deserved. None of this would have happened without Brad Conrad, former director of SPS and Sigma Pi Sigma, who took me up on the idea of resurrecting JURP in 2018 and secured its articles for publication with the American Institute of Physics. My thanks go as well to the faculty mentors who made each project possible and to the reviewers who gave careful attention to student work.

In a pleasing bookend, the incoming editor, Brittney Hauke, published an article in the first issue of JURP that I helped edit. I am grateful she is taking the reins, and I leave knowing JURPA is in good hands.

Meet the New Editor

Brittney Hauke is a longtime SPS member and volunteer. She is an alum of the Coe College SPS chapter and has served on the Physics and Astronomy Planning Committee since 2016. As journals managing editor for the American Ceramic Society and a former student researcher who published in JURPA, Hauke comes to this position with highly relevant experience. She holds BAs in physics and art and an MS and PhD in materials science and engineering.

The Relationship Between Coronal Mass Ejection Intensity and Cosmic Ray Flux

Kristian Qirko,[1,a)] Rex Paster,[2] Evangeline Selking,[3] Zoya Siddiqui,[3] Sydney Stapleton,[3] Maya Zacks,[1] Marybeth Senser,[3] Nathan A. Unterman,[1] and Mark R. Adams [4]

[1]*New Trier High School, 305 Winnetka Avenue, Winnetka, Illinois 60093, USA*
[2]*Rochelle Zell Jewish High School, 1095 Lake Cook Road, Deerfield, Illinois 60015, USA*
[3]*Downers Grove South High School, 1436 Norfolk Street, Downers Grove, Illinois 60516, USA*
[4]*University of Illinois Chicago, 845 W Taylor Street, Chicago, Illinois 60607, USA*

[a)]*Corresponding author: kristianqirko@gmail.com*

Abstract. This investigation examines the relationship between coronal mass ejection (CME) intensity and cosmic ray muon flux in North America using ground-level detectors. CMEs carry charged particles that interact with Earth's magnetic field, which deflects incoming galactic cosmic rays. As these cosmic rays reach the atmosphere, they generate showers of secondary particles, including muons, which can be measured by ground detectors. However, atmospheric pressure also influences muon flux. In this work, a method of correction was developed and tested that significantly reduced this barometric shielding effect. This correction enabled clearer characterization of muon flux changes due to the May 2024 CME. Analysis of corrected data using multiple detectors in the United States found an inverse relationship between CME intensity and cosmic ray muon flux consistent with a Forbush decrease pattern. Additionally, the impact of statistical error on the visibility of CME-induced effects was assessed.

INTRODUCTION

Coronal mass ejections (CMEs) are plasma from the Sun's corona that carry large magnetic fields which disrupt Earth's magnetic field.[1] These disturbances alter the total magnetic field surrounding Earth, affecting how incoming galactic cosmic rays are deflected. Cosmic rays are charged, high-energy particles likely originating from supernovae.[1,2] When cosmic rays collide with particles in our atmosphere, they produce showers of secondary particles. One such particle is the muon, which can penetrate deep into the atmosphere and be detected at ground levels.[3] As a result, the muon flux measured by ground-level QuarkNet detectors is used as a proxy for cosmic ray flux. CME intensity can be measured through changes to Earth's magnetic field using the K-index scale (Kp), which ranges from 1 to 9.[4] The CME of May 10, 2024, reached a Kp rating of 9, making it a highly significant disruption and the focus of this paper. Such events can disturb the usual path of cosmic rays. Typically, this results in a sharp decrease in cosmic ray flux with a gradual recovery back to normal levels, a phenomenon known as a Forbush decrease.[5] This study aims to refine a pressure correction method, assess how statistical error impacts the visibility of CME-induced effects, and provide additional evidence on the existence of the Forbush decrease and its reproducibility using multiple detectors in various locations. Studying CMEs and cosmic rays provides insight into two critical areas: the effects of cosmic rays on human health and the impact of CMEs on both space- and Earth-based technology.[6]

METHODOLOGY

This investigation utilized QuarkNet cosmic ray muon detectors consisting of scintillating plastic, photomultiplier tubes, and data acquisition cards.[7] Employing a twofold coincidence rate for each detector reduced background noise and improved signal reliability. The data were grouped in 3-h bins to balance sample size and statistical error with temporal resolution, allowing for clearer identification of onset and recovery patterns.[7]

First, data were corrected for barometric pressure fluctuations. Pressure variations affect muon flux because an increase in atmospheric density shields muons from detectors.[3] The muon rate changes by about 0.2% per millibar;

multiplied by typical daily pressure variations, this effect can be on the order of the size of the Forbush decrease. The resulting variation in muon rate can obscure data patterns, so barometric corrections were applied to each data point. A control month with no disruption in Earth's magnetic field was examined for each detector. A linear regression was fitted to the uncorrected flux versus pressure plot for the control month. From that graph, the slope m along with median flux F_{median} and pressure P_{median} were recorded. These values were then applied with Eq. (1) to each point on the uncorrected flux versus time graph (Fig. 1) to create a corrected graph (Fig. 2).

$$F_{corrected} = F_{original}\left(1 + \frac{m}{F_{median}}\left(P_{original} - P_{median}\right)\right) \quad (1)$$

The barometric correction revealed a clearer Forbush decrease and a more linear recovery (Fig. 2), demonstrating the method's effectiveness on real-world data.

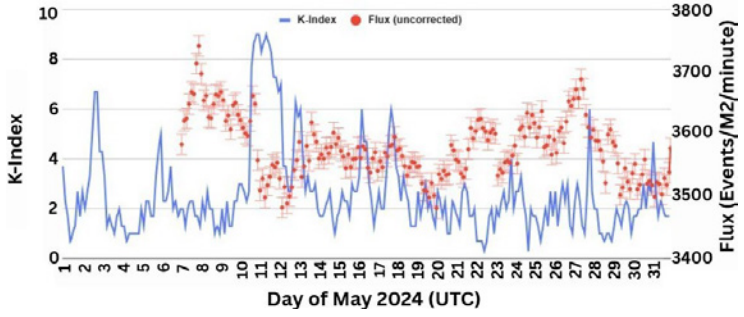


FIGURE 1. Uncorrected flux (red) and K-index (blue) versus time recorded in Skokie, Illinois, over the month of May 2024. Flux is in number of muons per m^2 per min. Data include short-term fluctuations obscuring the expected decrease and subsequent recovery, making the overall Forbush decrease pattern difficult to identify. The error bars represent $1/\sqrt{N}$, with N being the number of events.

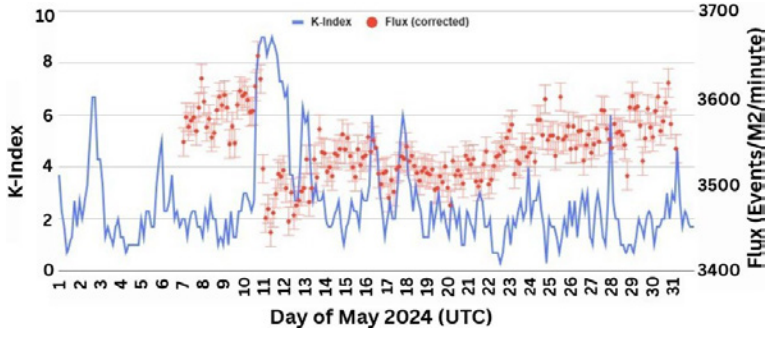


FIGURE 2. Corrected flux (red) and K-index (blue) versus time recorded in Skokie, Illinois, over the month of May 2024. Flux is in number of muons per m^2 per min. After pressure correction, the data reveal a clear Forbush decrease pattern with a gradual recovery toward baseline levels and minimal fluctuations that coincide with additional geomagnetic disturbances. The error bars represent $1/\sqrt{N}$, with N being the number of events.

RESULTS

In Table 1 the percent change in muon flux was calculated. First, we took the difference between the average flux of the period leading up to the CME event and the estimated flux value of the initial recovery phase (from the line of best fit), evaluated at the time of the initial flux drop caused by the CME. This difference was divided by the average and multiplied by 100 to find the percent change. The z-value was calculated by dividing the difference by the standard deviation of the calm period. This resulting z-value was used to calculate the p-value, the probability of randomly measuring this change from the mean. The recovery time measures how long after the initial flux drop the linear fit returned to the average flux value prior to the event.

TABLE 1. The percent change, percent error, *p*-value, average flux preceding the CME event, and recovery time in days for each cosmic ray detector during the May 10, 2024, CME event.

Location	% Change	% Error	Drop *p*-Value	Average Flux Prior to Event (counts/m²/min)	Recovery Time (days)	Recovery *p*-Value
Batavia, IL	-1.43	1.12	0.22	579	7	<0.0001
Batavia, IL	-2.25	0.32	<0.0001	7614	7	<0.0001
Northfield, IL	-2.52	0.30	<0.0001	7904	8	<0.0001
Skokie, IL	-2.65	0.46	<0.0001	3586	9	<0.0001
Skokie, IL	-4.30	0.46	<0.0001	3491	6	<0.0001
Northbrook, IL	-2.68	1.15	0.023	544	9	<0.0001
Winnetka, IL	-2.96	0.30	<0.0001	8223	11	<0.0001
Honolulu, HI	-2.22	1.90	0.10	207	17	0.0002
Honolulu, HI	-1.27	1.31	0.39	422	7	0.0021

Measured flux depends on the unique configuration of each detector. All detectors with sufficient sensitivity recorded signal patterns like the ones seen in Fig. 2, with an initial drop followed by an average projected linear recovery of 7.3 days and a second and third CME event occurring on May 16 and 18, respectively, lengthening the overall average recovery to 13.6 days. However, this pattern was less visible in detectors with low flux, such as those in Hawaii (Fig. 3). A characterization of the flux recovery pattern is provided in Fig. 4.

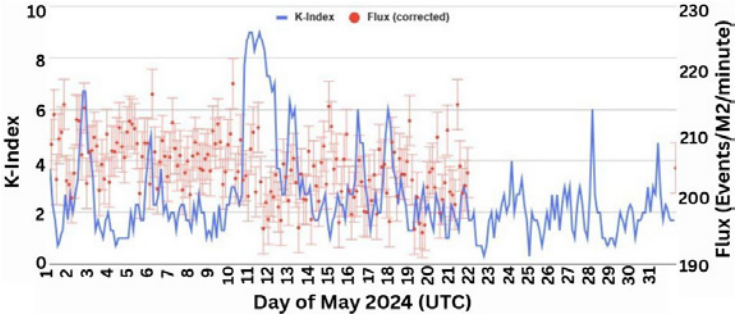


FIGURE 3. Pressure-corrected flux (red) and K-index (blue) versus time recorded in Honolulu, Hawaii, over the month of May 2024. Although a Forbush decrease is present, the pattern is less visually apparent than in higher-flux detectors because the lower count rate produces larger statistical fluctuations. The error bars represent $1/\sqrt{N}$ with N being the number of events.

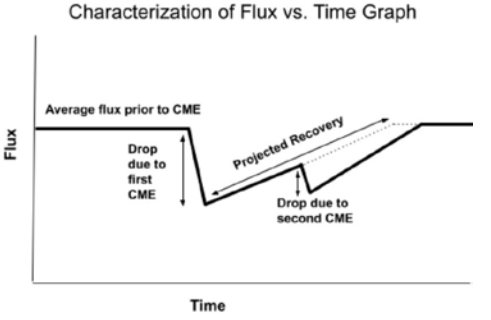


FIGURE 4. A schematic of the idealized pattern for CME onset and recovery. A rapid decline in muon flux coinciding with CME-driven geomagnetic disturbance is followed by a gradual recovery toward baseline levels as the interplanetary magnetic field stabilizes. A second sharp decrease corresponds to a subsequent CME event, demonstrating how multiple solar disturbances can produce repeated decreases that interrupt the recovery phase.

DISCUSSION

During May 10–12, 2024, a 9-Kp intensity CME, one of the strongest of the decade, occurred. Nine detectors recorded a decrease in muon flux on May 10, 2024. In all detectors with sufficient sensitivity, the drop was followed by a gradual recovery to baseline flux by May 30, 2024 (Fig. 2). This event demonstrated an inverse relationship between CME intensity and muon flux: as the K-index rose to 9, flux dropped; as the index returned to about 3, flux increased. This pattern, known as a Forbush decrease, has been supported by prior research.[8,9] The consistency of the decrease across all detectors with sufficient sensitivity, regardless of their location, strengthens the conclusion that the results were caused by a large-scale event. This multisite validation also supports the idea that CME events can encompass large areas while demonstrating the ability of ground-level detectors to measure space weather.

CONCLUSION

This study aimed to assess the impacts of CME events by measuring muon flux at locations across the United States. A pressure correction method was created and validated using real-world data. A Forbush decrease pattern was observed by all cosmic ray detectors with sufficient sensitivity, which demonstrates the pattern's reproducibility. Detectors with low flux measured a statistically insignificant decrease at CME onset due to the small sample size and large error. However, seven of the nine detectors recorded statistically significant decreases in muon flux consistent with a Forbush decrease, demonstrating reproducibility of the pattern across multiple locations.

ACKNOWLEDGMENTS

The authors thank Downers Grove South High School in Downers Grove, Illinois, and New Trier High School in Northfield, Illinois. This work was supported by QuarkNet National Science Foundation Grant No. PHY-2309272.

REFERENCES

1. N. Gopalswamy, National Solar Physics Workshop **20**, 108–130 (2020).
2. A. Arámburo-García, K. Bondarenko, A. Boyarsky, D. Nelson, A. Pillepich, and A. Sokolenko, Phys. Rev. D **104**, 83017 (2021).
3. P. G. Isar, *AIP Conf. Proc.* **1645**, 349–352 (2015).
4. NOAA Space Weather Prediction Center, "NOAA Space Weather Prediction Center," NOAA, https://www.swpc.noaa.gov/ (accessed May 31, 2025).
5. S. E. Forbush, PNAS **43(1)**, 28–41 (1957).
6. K. M. Omatola and K. M. Okeme, Arch. Appl. Sci. Res. 1825–1832 (2012).
7. J. Rylander, T. Jordan, J. Paschke, and H.-G. Berns, "QuarkNet Cosmic Ray Muon Detector User's Manual, Series '6000' DAQ," QuarkNet, https://quarknet.org/sites/default/files/cf_6000crmdusermanual-small.pdf (accessed May 31, 2025).
8. K. P. Arunbabu, H. M. Antia, S. R. Dugad, S. K. Gupta, Y. Hayashi, S. Kawakami, P. K. Mohanty, A. Oshima, and P. Subramanian, Astron. Astrophys. **580**, 44 (2015).
9. E. Barouch and L. F. Burlaga, J. Geophys. Res. **80**, 449–456 (1975).

Analyzing Kepler Data Through Machine Learning: A Study of Neural Networks for Exoplanet Detection

Tyler Carlson[a] and Renuka Rajapakse[b]

Department of Physics, University of Massachusetts Dartmouth, Dartmouth, Massachusetts 02747, USA

[a]Corresponding author: tcarlson1@umassd.edu

[b]rrajapakse@umassd.edu

Abstract. As modern space observatories generate increasingly large photometric datasets, the need for scalable, automated tools to interpret transit signals has become critical. To address this scale problem, we examined which combinations of light-curve input representations are most informative and computationally efficient for a 1D convolutional neural network (CNN) to separate real transits from false-positive signals. Across eight systematically compared input configurations, local multiscale transit views achieved the best performance (validation AUC = 0.942, minimum validation loss = 0.30), outperforming global phase-curve inputs in validation AUC and minimum validation loss. Building on that result, we propose applying the trained CNN to unconfirmed KOI candidates to produce an independent, confidence-tiered ranking that augments NASA's existing candidate disposition-confidence framework. The same approach could later be extended to archival catalogs from the K2 mission (the repurposed continuation of Kepler) and future transit-survey candidate lists, improving the efficiency with which limited telescope time is allocated.

INTRODUCTION

Space missions such as the Kepler Space Telescope [1] and the Transiting Exoplanet Survey Satellite (TESS) monitor millions of stars simultaneously, producing photometric time series at a rate that far exceeds the capacity for manual scientific review. When a planet passes in front of its host star, it causes a brief, periodic dimming of the measured stellar flux (a transit signal) whose depth, duration, and morphology encode the planet's physical and orbital properties [2]. Kepler's automated pipeline flags these events as threshold crossing events (TCEs), a subset of which are promoted to Kepler objects of interest (KOIs). The KOI catalog, maintained by the NASA Exoplanet Science Institute, currently lists thousands of systems sorted into confirmed exoplanets, confirmed false positives, and unconfirmed candidates.

False positives, which can be attributed to signals from eclipsing binaries, instrumental artifacts, or stellar variability mimicking planetary transits, are the main challenge in this classification task [3, 4]. Follow-up spectroscopic observations required to confirm a candidate are both expensive and time limited; misallocating observatory resources to likely false positives carries a direct scientific cost. NASA maintains an existing automated vetting pipeline for candidate classification that reports Robovetter disposition scores. According to the DR25 catalog description, "the disposition score is simply the fraction of Monte Carlo iterations that result in a disposition of PC" [5], where PC denotes a *planetary candidate*. In this paper, we refer specifically to that Robovetter *disposition-confidence* score. The convolutional neural network (CNN) developed in the present work is ultimately intended to be applied to candidate systems to produce an independent confidence score, complementing NASA's Robovetter disposition-confidence score. The CNN could later be extended to archival candidate catalogs from the K2 mission (the repurposed continuation of the Kepler spacecraft after its reaction-wheel failures) and to future transit-survey candidate lists.

A major obstacle in this work is that raw Kepler light curves are heterogeneous, varying in length and cadence from star to star, and so must be reduced to a fixed-length numerical form before any CNN can ingest them. This is done by defining a *light-curve input configuration*: settings that include binning resolution, truncation of the light curve, and data smoothing. This work focuses on the first stage of that effort, identifying which configurations are most informative and computationally efficient. It is part of a larger project to train a high-accuracy supervised CNN on selected inputs and apply it to unconfirmed candidates to produce a prioritized, confidence-tiered observing list.

INPUT REPRESENTATION AND DATA PIPELINE

Dataset and Quality Control

Light-curve data were retrieved from the cumulative KOI table of the NASA Exoplanet Archive [6] using the `Lightkurve` Python package [7]. The raw catalog contained 2,746 confirmed exoplanets, 4,839 confirmed false positives, and 1,979 unconfirmed candidates. A quality-control pipeline filtered out entries showing corrupted files, insufficient data coverage, local flux anomalies within the transit window, secondary-eclipse signatures near phase 0.5, significant transit asymmetry, out-of-range transit durations, and flux noise exceeding signal-to-noise thresholds. Refining the dataset to keep only clean, intact systems improves neural network performance metrics (see Appendix A for full criteria).

Multiscale Light-Curve Representation

Every example of a given view must contain an identical number of data points across all stars, so the variable-length, variable-cadence Kepler light curves must be normalized by phase-folding and binning before training [2]. Using `Astropy` [8] for the underlying time-series and unit handling, each light curve was flattened to remove stellar variability trends, phase-folded at the known orbital period, and centered on the midtransit time T_0. From this, two classes of view were produced: a *global view* spanning the full phase profile (capturing out-of-transit stellar behavior) and a *local view* truncated to a window of $\pm 1.2 \times (T_{14}/2)$ around midtransit, capturing ingress, depth, and egress. Each view was resampled at multiple bin counts (200, 100, and 50 for local; 5,000 for global), and a Gaussian-convolved counterpart of each binned view was also created to provide a noise-suppressed representation of transit shape. Figure 2 in Appendix A illustrates these representations for a confirmed exoplanet, KIC 11241912.

Data Format: CSV to NPZ

The pipeline originally stored processed light curves as CSV files. This introduced substantial overhead: Python was forced to parse text, infer types, and convert to NumPy arrays on every training run, with floating-point precision loss. To eliminate this, all data were reprocessed and stored as compressed NumPy binary (`.npz`) archives. Each `.npz` file contained two `.npy` arrays per view, phase and flux stored as correlated arrays, and stellar metadata—Kepler ID, orbital period, transit epoch, transit duration, stellar radius, effective temperature, and surface gravity. This format is natively compatible with TensorFlow [9], loads in a single function call, preserves full floating-point precision, and allows for training without any needed conversion. Metadata embedded in each file will serve as further filtering and feature inputs in future model iterations.

Input Comparison Results

Eight CNN configurations were trained under identical hyperparameters, differing only in which combination of views was supplied as input. Figure 1 summarizes training AUC, validation AUC, training accuracy, validation accuracy, minimum validation loss, and best epoch for each variant. Here AUC denotes the area under the receiver operating characteristic curve, a threshold-independent measure of how reliably the network's confidence ranking separates genuine transits from false positives (1.0 being perfect separation and 0.5 no better than chance). The strongest overall configuration is Local Data, which loads all local transit views simultaneously and ranks best on both validation AUC and minimum validation loss. This supports the value of multiscale local binning: lower-bin views capture the broad transit profile, while higher-bin views preserve finer characteristics such as ingress and egress shape.

By contrast, Global Data performs weakest across most validation metrics, showing that the current network does not use global phase information well on its own. Some false-positive types can contain useful out-of-transit signatures, but extracting them likely requires more targeted preprocessing, filtering, and pooling; global inputs thus appear most valuable when merged with local branches rather than used in isolation.

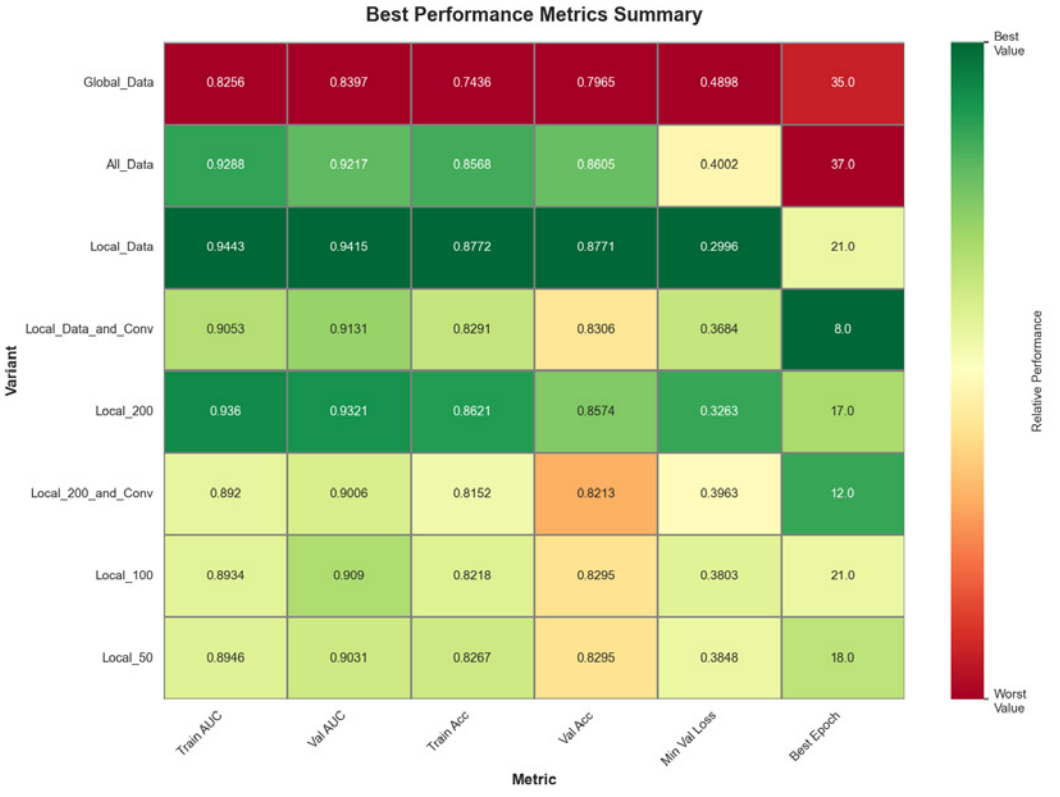


FIGURE 1. Best-performance summary for all eight CNN input configurations. Cells show raw metric values, including best epoch as a raw epoch count. Color reflects relative performance within each metric column. The worst value in a column maps to red and the best to green. For minimum validation loss and best epoch, lower values are treated as better in the color mapping. Local Data performs best overall, whereas Global Data performs worst across most validation metrics.

SUPERVISED CNN TRAINING AND PERFORMANCE

The best-performing configuration (Local Data, all three local scales without convolved counterparts) trained the final supervised model. The parallel 1D CNN processes each view through an independent convolution–pooling branch before hierarchical merging into fully connected layers and a sigmoid output node. This produces a planet-likeness score $p \in [0, 1]$ whose higher values indicate closer resemblance to a genuine exoplanet transit (see Appendix B). All models were trained on GPU hardware using TensorFlow with binary cross-entropy loss, an Adam optimizer at base learning rate 10^{-4} with dynamic adjustment, early stopping at patience 15 over 100 epochs, and class weights of 1.64 (exoplanet) and 0.72 (false positive) to compensate for class imbalance. Architecture details are given in Appendix B; full training curves are shown in Appendix C.

The current supervised model achieves a validation AUC of 0.942 and a minimum validation loss of 0.30, a strong baseline for a first-pass vetting classifier. Several pathways could improve these metrics further: augmenting the training pool with NASA's catalog of *injected* simulated transits [10], refining the data-cleaning thresholds to reduce residual label noise, and tuning the convolutional filter widths and pooling strides, particularly on the global branch. These directions are detailed in Appendix C.

INFERENCE ON UNCONFIRMED CANDIDATES AND PRIORITIZATION

The final stage of this project will apply the trained supervised network to the 1,851 unconfirmed KOI candidates. Once the model demonstrates sufficient validation performance (including high AUC, stable precision–recall, and acceptable calibration on a held-out test set), it can be applied to the unlabeled candidates. Because the network

has learned to separate planetary from nonplanetary light-curve shapes, its output score p provides a planet-likeness ranking rather than a calibrated probability; with further calibration, it could later be interpreted as an estimated probability.

Candidates will be sorted in descending order of this score and organized into three confidence tiers: **Tier 1** (high confidence, p above a precision-optimized threshold) — systems whose transit shape most closely resembles confirmed exoplanets, recommended for immediate spectroscopic or radial velocity follow-up; **Tier 2** (intermediate confidence) — systems warranting additional statistical vetting such as centroid shift analysis or transit timing variation searches before allocating telescope time; and **Tier 3** (low confidence) — systems whose signals most closely resemble known false-positive shapes. After sorting, each entry's Kepler ID and stored stellar metadata will be reattached so the output is a fully annotated, observatory-ready prioritized list.

Although NASA provides Robovetter disposition scores, the proposed pipeline will produce an independent CNN-derived ranking based only on learned light-curve characteristics. This ranking will not replace existing vetting products, but it will function as a practical prioritization metric. Candidates with high CNN scores can be inspected first, while low-scoring systems can be deprioritized or flagged for additional diagnostic checks before telescope time is allocated.

CONCLUSION

This paper aims to establish how light-curve data should be represented for machine-learning-assisted exoplanet candidate vetting. We find that multiscale local transit views, stored in compressed `.npz` format, provide the most efficient and informative input for a 1D CNN classifier, achieving a validation AUC of 0.942 on the Kepler KOI catalog. This input-representation study is the first stage of a larger project that will progress through supervised classification and inference on unconfirmed candidates toward a confidence-tiered candidate list designed to complement NASA's Robovetter-based vetting outputs with a distinct CNN confidence value. With continued improvements to data cleaning, training-pool augmentation, and hyperparameter optimization, the broader pipeline could, after further validation, grow into a scalable vetting tool for Kepler and future transit surveys such as TESS.

ACKNOWLEDGMENTS

We wish to acknowledge Dr. Sarah Caudill of the University of Massachusetts Dartmouth Department of Physics for her guidance and support. We would also like to thank Mr. Mark Munkacsy for his advice and feedback. This work was supported in part by the U.S. Office of Naval Research and the Massachusetts Space Grant Consortium (NASA MSGC).

REFERENCES

1. W. J. Borucki, D. Koch, G. Basri, N. Batalha, T. Brown, D. Caldwell, J. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. DeVore, et al., Science **327**, 977 (2010).
2. C. J. Shallue and A. Vanderburg, Astron. J. **155**, 94 (2018).
3. T. D. Morton, Astrophys. J. **761**, 6 (2012).
4. C. J. Burke and J. Catanzarite, "Planet detection metrics: Per-target detection contours for data release 25," Tech. Rep. KSCI-19109-002 (NASA Ames Research Center, Moffett Field, CA, 2017).
5. S. E. Thompson, J. L. Coughlin, K. Hoffman, F. Mullally, J. L. Christiansen, C. J. Burke, S. Bryson, N. Batalha, M. R. Haas, J. Catanzarite, et al., Astrophys. J. Suppl. Ser. **235**, 38 (2018).
6. R. L. Akeson, X. Chen, D. Ciardi, M. Crane, J. Good, M. Harbut, E. Jackson, S. R. Kane, A. C. Laity, S. Leifer, et al., Publ. Astron. Soc. Pac. **125**, 989 (2013).
7. Lightkurve Collaboration, "Lightkurve: Kepler and TESS time series analysis in Python," Astrophysics Source Code Library, ascl:1812.013 (2018).
8. Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, et al., Astron. Astrophys. **558**, A33 (2013).
9. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., "TensorFlow: Large-scale machine learning on heterogeneous systems" (2015), software available from https://www.tensorflow.org.
10. J. L. Christiansen, "Planet detection metrics: Pixel-level transit injection tests of pipeline detection efficiency for data release 25," Tech. Rep. KSCI-19110-001 (NASA Ames Research Center, Moffett Field, CA, 2017).

Appendices are available at https://pubs.aip.org/aip/jurp.

A Convolutional Neural Network (CNN) Approach to Galaxy Merger Classification

Sohan Subudhi,[a] Alex Xu,[b] and Darshan Desai[c]

The University of Texas at Austin, Austin, Texas 78712, USA

[a]Corresponding author: sohan.subudhi@utexas.edu
[b]alex.xu@utexas.edu
[c]dashkad2004@utexas.edu

Abstract. Galaxy mergers provide critical insights into cosmic evolution, yet identification in survey data remains challenging. We introduce a rigorously verified dataset of 228 hand-classified mergers and 292 nonmergers, curated from Galaxy Zoo DR2 and DR5, VV objects, and Toomre merger candidates. Using this data, we trained a convolutional neural network with four convolutional blocks, utilizing L2 regularization, dropout, and augmentations like random rotations and flips. Our model achieved 74% test accuracy, demonstrating the dataset's utility. The findings indicate that improvements through synthetic data or domain adaptation would enhance machine learning performance. We present this dataset as a resource to advance automated merger classification.

INTRODUCTION

Past Work

Classifying galaxy mergers would shed light on the theorized hierarchical formation of the universe by enabling the reconstruction of galaxy assembly histories over cosmic time. Furthermore, identifying and categorizing merger events would help constrain the distribution and evolution of dark matter halos, improve estimates of merger rates that produce gravitational wave sources, and clarify the role of mergers in triggering active galactic nuclei activity [1].

However, large-scale manual classification of mergers proves difficult, given the time scale of galaxy merger events, the poor visibility of certain features of galaxies such as tails and bars, and uncertainty as to whether or not a merger has occurred. Modern hydrodynamic simulations effectively model clusters and galaxies en masse [2], spurring the development of numerous machine learning (ML) models for galaxy merger classification. Observationally based approaches in particular have applied convolutional neural networks (CNNs) to real survey imagery, using transfer learning to detect mergers [3] and to assemble large observational merger catalogs [4].

Past work by Pearson et al. [5] and Margalef-Bentabol et al. [6] provides context and support for our research. The first study addresses the challenges of temporal classification in mergers, using the IllustrisTNG simulation to train ML models capable of predicting whether a galaxy is in a premerger, ongoing, or postmerger stage. The second study evaluates the performance of different machine learning techniques for merger detection, emphasizing the need for models that generalize effectively across diverse datasets. It is important to note that our study differs by addressing a broad range of limitations identified in both of the above works. While the first study highlights the power of simulations to improve temporal accuracy, it underscores the need for more examples and better integration of intermediate stages, which our work looks to address. Similarly, the second study reveals significant drops in classification performance when models are applied to data from different domains (sources of merger data), highlighting the limitations of current methods in transferring knowledge across datasets.

Research Question

There is a gap between simulation and real-data performance among classifiers today [6] that can be addressed by gathering more real merger data to build stronger models. With new observations from the James Webb Space Telescope (JWST) and the Simonyi Survey Telescope capturing high-redshift images of potential galaxy mergers, we aim to support classification by constructing a verifiable dataset and subsequently developing an effective classifier trained upon it.

From a data collection standpoint, we define a galaxy merger as the result of multiple galaxies passing through each other, disturbing the structure of the involved galaxies and resulting in long-lasting changes to the shape and

appearance of a galaxy, such as observable tidal effects and distortion in the shape of a galaxy. A nonmerger for this study is defined as a galaxy without obvious perturbation. It is important to note the lack of any tidal tails, bridges, or other types of visual distortion that are typically seen within a merger. The nonmergers typically exhibit a well-defined structure, such as the arms of a spiral galaxy or smooth elliptical morphology. We also chose to exclude any galaxies that appear irregular to avoid a potential confounding variable.

DATA SOURCES

The data distribution and our sources are depicted in Table 1. These sources were selected based on their demonstrated reliability in previous studies and their established prominence in the field [7]. Our sources comprise Galaxy Zoo DR5 [8], Galaxy Zoo Merger Project [9], VV objects [10], and Toomre [11] merger candidates. A merger candidate is defined as galaxies marked as interacting based on their respective catalogs. We randomly sampled these from each catalog to do manual classification.

TABLE 1. Merger image data sources.

Merger Galaxy Distribution			**Nonmerger Galaxy Distribution**		
Catalog (Survey)	**Candidates**	**Validated**	**Catalog (Survey)**	**Morphology**	**Number**
Galaxy Zoo Merger Project (SDSS)	54	54	Galaxy Zoo DR2 (SDSS)	Spirals	50
				Ellipticals	50
Galaxy Zoo DR5 (SDSS)	248	93	Galaxy Zoo DR5 (SDSS)	Round	46
				In-Between	75
VV-Objects (SDSS/NED)	219	86		Cigar-shaped	12
Toomre Mergers (PAN-STAARS)	9	9		Edge-on	8
				Spirals	51
Total mergers	530	242	**Total nonmergers**		**292**
After deduplication	**228**				

We classified these candidates by visual inspection in a double-blind setting, including images in our merger set if the majority of three authors determined it was a merger. When it appeared that a smaller galaxy was either far away from or cannibalized by a larger galaxy, we evaluated whether there were enough structural disturbances and distortions to indicate a merger interaction. Unclear cases where a majority of authors did not believe the candidate was a merger were rejected from both the merger and nonmerger sets to prevent potential misclassifications. Note that the nonmergers were obtained separately from Galaxy Zoo, where the galaxy morphology labels were more reliably crowd-sourced.

Since our final dataset concatenated objects from multiple catalogs, we performed a deduplication step to catch overlapping images both within and across surveys. Using the catalog centroids, we implemented a coordinate filter ($|\Delta\text{right ascension (RA)}| + |\Delta\text{declination (DEC)}| < 0.1^\circ$). Although this overestimates true angular distance at higher declinations, it successfully isolated a subset of potential duplicates that we then manually verified by eye. The redshift data obtained through Galaxy Zoo indicated mergers were all low-z and less than 0.15. Redshift was excluded from this matching criteria since it was not universally available across our catalogs.

This process does not exclude a merger until manually verified by comparing duplicates and, in the worst case, leads to mergers that were not duplicated being excluded. After this process, the number of mergers in our dataset decreased from 242 to 228. In the case of duplicates, we kept the more centered copy and prioritized images from less-represented surveys to improve image-source balance. Images from the VV and Toomre catalogs were not centered on galaxies due to outdated RA and DEC values.

ANALYSIS

For our model, we utilized a CNN [12] with four convolutional blocks of 32, 64, 96, and 128 filters with a 3×3 kernel, followed by global average pooling and a 256-unit dense classifier. The source images were 512×512; to select the input resolution, we trained and evaluated three otherwise-identical models at 128×128, 256×256, and 512×512. The 256×256 model was chosen since it gave the best test accuracy, balancing spatial detail with computation complexity. Input sizes were kept to powers of two to down-sample cleanly through the four pooling stages. The remaining choices (block depth, filter progression, and regularization) follow standard design principles for a dataset of this size. We applied L2 regularization with $\lambda = 0.01$, batch normalization, and dropout rates between

30% and 50%. We used Adam as our optimizer with a learning rate of $\eta = 0.0001$ and binary cross-entropy as our loss function, which are conventional parameter settings. A fuller hyperparameter search is left for future work. We trained a model on the set of 228 mergers and 292 nonmergers detailed above. In training the model, we used a split of 0.65 train and 0.20 validation and 0.15 test. Additionally, the following data augmentations were applied while training: normalizing pixel values to the $[0, 1]$ range, random rotations up to 20°, horizontal and vertical shifts of $\pm 15\%$ of the image width, and horizontal and vertical image flipping.

RESULTS

TABLE 2: Model performance metrics.

Class	Precision	Recall	F1 Score
Merger	0.65	0.92	0.76
Nonmerger	0.90	0.60	0.72
Balanced accuracy: 0.7583		*AUC: 0.8852*	

TABLE 3: Test-set confusion matrix.

		Predicted	
		Merger	Nonmerger
True	Merger	33	3
	Nonmerger	18	27

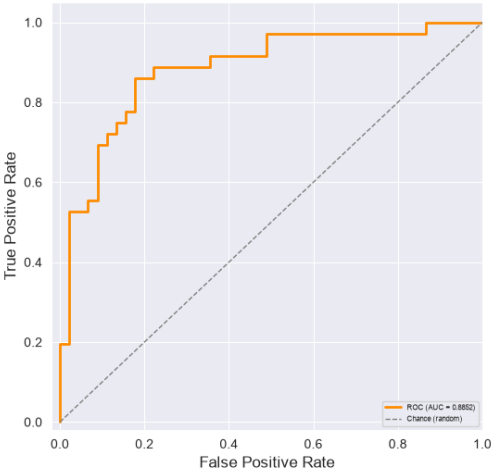


FIGURE 1: Test-set ROC (positive = merger).

The model achieved a test accuracy of ∼74%, a weighted recall of 74%, and a weighted precision of 79% (Table 2). For comparison, a baseline classifier that always predicts the majority class (in this case, the nonmerger class) has an accuracy of $\frac{45}{36+45} \times 100 = 55.6\%$ on the test set, so the model represents a substantial improvement over chance. This separability is reflected in the receiver operating characteristic curve (Fig. 1), which yields an area under the curve of 0.8852. As shown in the confusion matrix (Table 3), the model recovers 92% of true mergers but misclassifies 18 of the 45 nonmergers as mergers, indicating a mild bias toward predicting the *merger* class.

As per the misclassifications of the model (Fig. 2), images commonly thought to be of a nonmerger appear to be primarily black (empty) with a very small subject at the center of the image. The image composition is similar to that of higher redshift images belonging to nonmergers. Such images could potentially be better handled through a variable scale at which images are extracted depending on redshift. Images that were misclassified as mergers were typically spiral galaxies that have hazy and more separated spiral arms or somewhat elongated galaxies, indicating that those might be features the CNN responds strongly to; sharpening the model's ability to separate genuine interaction from lookalike structures is a promising direction for reducing them. Correct classifications associate mergers with clear interaction signatures such as tidal bridges, tails, or closely paired nuclei linking the merging components. Nonmergers are associated with isolated single galaxies, often compact and set against a sparse field, with no companion or interaction signature, indicating the model keys on the presence or absence of the morphological cues it associates with interaction.

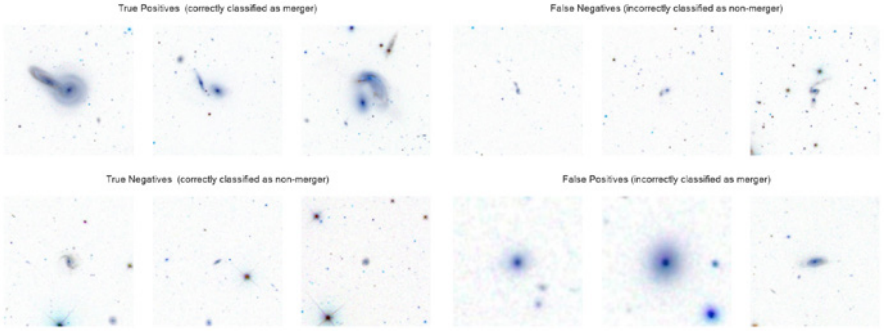


FIGURE 2: Select test set classifications. The inverse images are from SDSS and use SDSS-*g*, SDSS-*r*, and SDSS-*i* bandpasses.

CONCLUSION

The accuracy achieved by Margalef-Bentabol et al. [6] when using a CNN trained on Hyper Suprime-Cam survey images was lower than our final results, however, we believe that the low quantity of data may have impacted this. We support further testing with additional data; data gathered from different sources may also lead to different results. Our study advances galaxy merger classification by producing data for the implementation of domain adaptation to mitigate these biases, ensuring machine learning models are not constrained by patterns specific to any single survey or simulation. Within our experiment, we examined recall separately by source. Across the full dataset the model recovered mergers at 100% recall for VV objects, 96.3% for the Galaxy Merger Project sample, 93.5% for Galaxy Zoo DR5, and 88.9% for the Toomre candidates, with the mean predicted merger probability following the same ordering. The variation is modest, suggesting that catalog-level selection and labeling, rather than the imaging itself, drives the spread. We note, however, that our imaging is predominantly SDSS, so this dataset cannot by itself characterize the larger domain shift expected for surveys of markedly different depth and resolution, such as the Hyper Suprime-Cam imaging used by [6].

Our model demonstrates that mitigating class bias and taking certain data preprocessing steps, such as variable scaling based on red shift, are important factors to consider in a merger classifier. Future studies looking to build upon these results will have a well-defined list of mergers to draw from to assist in advancing this field beyond our work. In the future, we would like to augment our results by incorporating (manually verified) synthetic data obtained from IllustrisTNG. We also propose the use of domain adaptation [3] when training a better model on synthetic data. We would also like to customize the scale at which images are captured using redshift or another parameter, to standardize the size of the galaxy in the image. An interesting extension would be to predict the stage of merger evolution on a continuous scale from 0 (relaxed) to 1 (disturbed and strongly interacting). This may also allow for the classification of mergers as minor or major interactions based on interaction strength. We would also like to investigate distinguishing between wet and dry mergers through the incorporation of multi-band photometry, which may provide additional information regarding star formation, gas content, dust, and stellar mass.

ACKNOWLEDGMENTS

The authors would like to express sincere gratitude to Dr. Shyamal Mitra, Dr. Karl Gebhardt, the peer mentors, and other members of the Geometry of Space research stream for their invaluable guidance and mentorship throughout this research project.

REFERENCES

1. J. E. Barnes and L. Hernquist, Annu. Rev. Astron. Astrophys. **30**, 705–742 (1992).
2. D. Nelson, V. Springel, A. Pillepich, V. Rodriguez-Gomez, P. Torrey, S. Genel, M. Vogelsberger, R. Pakmor, F. Marinacci, R. Weinberger, et al., Comput. Astrophys. Cosmol. **6**, 2 (2019).
3. S. Ackermann, K. Schawinski, C. Zhang, A. K. Weigel, and M. D. Turp, Mon. Not. R. Astron. Soc. **479**, 415–425 (2018).
4. W. J. Pearson, L. E. Suelves, S. C.-C. Ho, N. Oi, S. Brough, B. W. Holwerda, A. M. Hopkins, T.-C. Huang, H. S. Hwang, L. S. Kelvin, et al., Astron. Astrophys. **661**, A52 (2022).
5. W. J. Pearson, V. Rodriguez-Gomez, S. Kruk, and B. Margalef-Bentabol, Astron. Astrophys. **687**, A45 (2024).
6. B. Margalef-Bentabol, L. Wang, A. La Marca, C. Blanco-Prieto, D. Chudy, H. Domínguez-Sánchez, A. D. Goulding, A. Guzmán-Ortega, M. Huertas-Company, G. Martin, et al., Astron. Astrophys. **687**, A24 (2024).
7. M. Walmsley, A. M. M. Scaife, C. Lintott, M. Lochner, V. Etsebeth, T. Géron, H. Dickinson, L. Fortson, S. Kruk, K. L. Masters, et al., Mon. Not. R. Astron. Soc. **513**, 1581–1599 (2022).
8. M. Walmsley, C. Lintott, T. Géron, S. Kruk, C. Krawczyk, K. W. Willett, S. Bamford, L. S. Kelvin, L. Fortson, Y. Gal, et al., Mon. Not. R. Astron. Soc. **509**, 3966–3988 (2022).
9. A. J. Holincheck, J. F. Wallin, K. Borne, L. Fortson, C. Lintott, A. M. Smith, S. Bamford, W. C.. Keel, and M. Parrish, Mon. Not. R. Astron. Soc. **459**, 720–745 (2016).
10. B. A. Vorontsov-Velyaminov, *Catalogue of Interacting Galaxies*. Available at http://www.sai.msu.su/sn/vv/.
11. A. Toomre, "Mergers and Some Consequences," in *The Evolution of Galaxies and Stellar Populations*, edited by B. M. Tinsley and R. B. Larson (Yale University Observatory, New Haven, CT, 1977), p. 401.
12. Y. LeCun, in *Connectionism in Perspective*, edited by R. Pfeifer, Z. Schreter, F. Fogelman, and L. Steels (Elsevier, Zurich, 1989), pp. 143–150.

QUIET: Ensuring the Efficacy of Damping Systems for Cryogenic Dark Matter and Rare Event Searches

Sara Earnest

Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, USA

searnes1@alumni.jh.edu

Abstract. It is a common problem in detector physics that reducing noise of one type introduces noise from another source, lowering the practical sensitivity of detector instruments. This investigation is an analysis of vibrational noise from around a pulse-tube-cooled dilution refrigerator (DR) intended to test instruments for particle dark matter and rare event searches. Analysis of the frequencies of vibrational noise near the DR finds that the active vibrational damping used in the lab lowers the noise by more than 87.36% in the frequency band between 6 and 19 Hz, which is where the cooling system contributes the most significant mechanical noise.

INTRODUCTION

This work addresses a complex question in two parts. First, for sensitive, cryogenic rare event searches, how can we characterize vibrational noise from the cooling system? And second, how can we determine whether existing systems to mitigate such noise are effective? These questions are important, as ground-based cryogenic experiments for low energy measurements require extremely sensitive detectors to resolve very weak signals [1, 2]. However, in being so sensitive, these detectors also pick up a significant amount of noise from their environment. For example, in cryogenically cooling cavities to reduce thermal noise, vibrational noise is introduced through the instruments that comprise the cooling system. As a result, the accuracy of such a system's signal readout is heavily impacted by the efficacy of vibrational damping systems, and it becomes necessary to confirm and qualify the efficacy of that damping system.

The Johns Hopkins laboratory where we conducted this work uses an LD400 BlueFors dilution refrigeration (DR) system [3] to test microwave cavities along with semiconductor and superconducting material samples at millikelvin temperatures in support of dark matter and rare event searches conducted in other laboratories, including the HAYSTAC, ALPHA, CUORE, and CUPID experiments. It is also set up to do detector research and design. The ability to accurately identify low-frequency vibrational noise in the lab environment supports the HAYSTAC experiment in particular. The presence of mechanical noise at frequencies under 200 Hz [4] can directly interfere with instruments like Josephson parametric amplifiers (JPAs) [5] that are used to enhance the sensitivity of detectors looking for axion signals by mitigating the noise from fundamental quantum uncertainty [5].

The cooling process for the DR introduces consistent vibrational noise at mid-to-low frequencies through the use of pulse tubes to precool the 3-He before it is circulated through the system [6]. A more detailed discussion of dry dilution refrigerators can be found in Pobell's *Matter and Methods at Low Temperatures* [7]. In addition to the DR, the laboratory uses a heavy frame for passive vibrational damping, as well as a Table Stable AVI-200XL/LP active damping system [8]. The combination of active and passive damping was intended to reduce vibrational noise in the experimental payload primarily over the 6–20 Hz frequency range, and thus in this work we sought to not only identify the DR's noise environment but also to ensure the damping systems worked as intended.

THE EXPERIMENT

As mentioned, this lab uses an active vibrational damping system to mitigate mechanical noise. For this experiment, we used an accelerometer to record this noise and wrote a script in MATLAB to take its power spectral density to characterize which frequencies were most prominently introduced by the DR cooling system. The accelerometer is a PCB Piezoelectronics model 393B04 and is sensitive to vibrations between 0.06 and 450 Hz [9]. Both the damping system and the accelerometer are mounted on the scaffolding that holds up the dilution refrigerator (see Fig. 1). This placement of the accelerometer was chosen because from there it can easily detect noise from the pulse tubes without having to be inside the cavity itself. See the Table Stable website for more detail on the damping system specifications

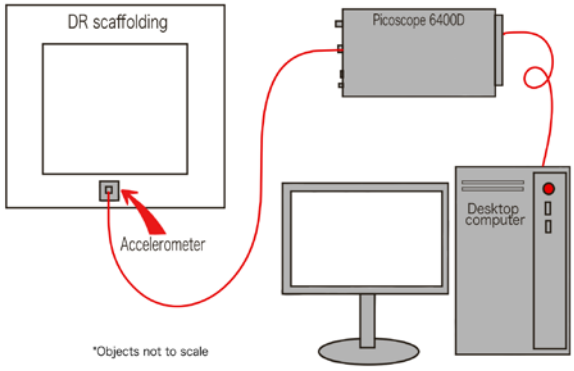


FIGURE 1: A schematic showing the connection path of the accelerometer to the picoscope and desktop computer used to export vibration data for processing.

[8]. The heavy passive frame of the DR acts as a stable base on the laboratory floor, while the active damping units sit directly between this frame and the scaffolding that holds up the dilution refrigerator. This configuration places the active system right in the path of the pulse tube vibrations, allowing it to absorb and cancel the mechanical noise before it can shake the experimental setup. The accelerometer data were digitized using a PicoScope 6404D PC oscilloscope and exported as time-series voltage vectors for spectral processing [10]. We then used the MATLAB code we wrote to generate a power spectral density from the data. The script was coded to use the accelerometer signal output (which is exported as segmented waveforms of a continuous dataset) to stitch the data together into a continuous time history and use user-selected, power-of-two window segments to balance frequency resolution against background noise variance. It also calculates the mean sampling frequency and from these items computes the final power spectral density curves. The accuracy of the MATLAB processing script was successfully verified prior to data analysis by resolving a clean, single peak from a 100-Hz calibration signal generator.

THE PSD MODEL

The power spectral density (PSD) is a tool that allows the user to clearly see the intensity of recurring frequencies of a continuous signal—in this case, that signal is vibrational noise. While a standard power spectrum measures discrete power at specific frequencies, a power spectral density characterizes how the power of a random, continuous signal is distributed over a continuous frequency spectrum per unit of bandwidth [11]. Using a spectral density rather than a raw spectrum allows us to make accurate, normalized comparisons of random vibration signals with different total recording lengths [11]. The power spectrum of a time-dependent function $x(t)$ is defined from the Fourier transform:

$$S_{xx}(f) = \lim_{T\to\infty} \frac{1}{T}|\hat{x}_T(f)|^2, \tag{1}$$

where $\hat{x}_T(t) = \int_{-\infty}^{\infty} e^{-i2\pi ft}x(t)$.

Here, $x(t)$ is defined in terms of a voltage with time history units of mV/s. Because there is no unique electrical impedance associated with the open-circuit sensor layout, physical "power" is conventionally reckoned in terms of the square of the voltage signal, which remains directly proportional to the true mechanical energy of the vibrations [11]. Generally, the units of a PSD will be the ratio of units of variance per unit of frequency. For example, the power spectrum of this series of voltages (in milliVolts) over time (in seconds) has units of mV^2/Hz.

To quantify the total vibrational energy contained within a specific frequency band (such as the target 6–19 Hz window), the PSD of the voltage signal, denoted as $S_{xx}(f)$, must be integrated over that frequency interval,

$$V^2_{\text{band}} = \int_{f_1}^{f_2} S_{xx}(f)df. \tag{2}$$

Multiplying the spectral density (mV^2/Hz) by the differential frequency width (df, in hertz) cancels the frequency dimension, yielding the total mean-square voltage (V^2_{band}) in units of mV^2. Taking the square root of this integrated

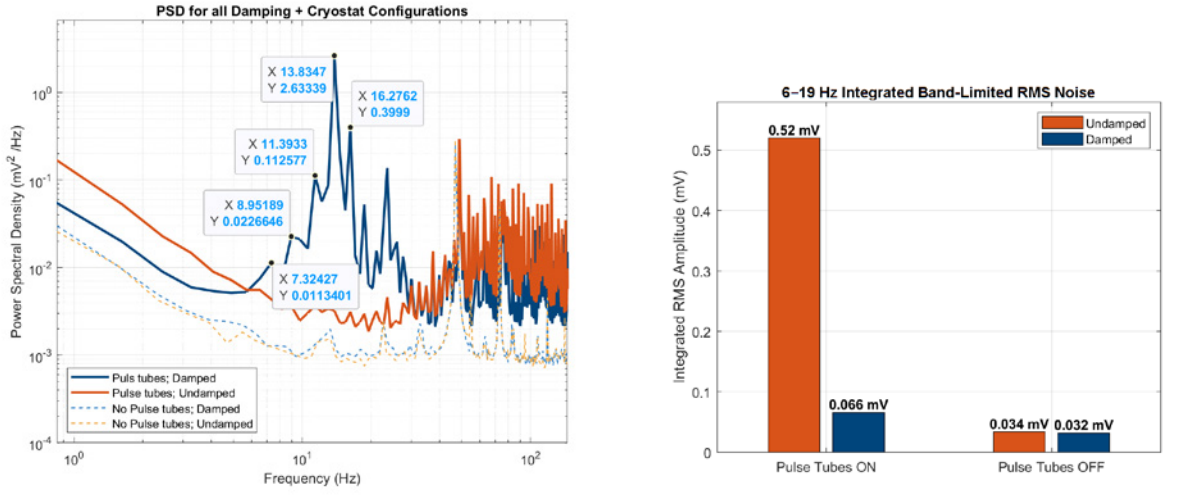
value yields the band-limited root-mean-square (RMS) amplitude [11]:

$$V_{\mathrm{RMS}} = \sqrt{\int_{f_1}^{f_2} S_{xx}(f)\,df}. \tag{3}$$

This operation returns the final metric to a linear physical dimension of millivolts and provides the dimensionality used to construct the comparative baseline analyses presented in Fig. 2(b).

DATA AND RESULTS

Figure 2 shows the PSD for all four operational configurations. Without the pulse tubes (PT) or damping system running, the baseline noise level is low and generally flat. Engaging the active damping system introduces a minor increase in noise below 6 Hz compared to the undamped signal, with a slight bulge in amplitude between 4 and 5 Hz with the the PT off. This is a side effect of the table's feedback loop known as the "waterbed effect" [8, 12] (a full description of which can be found in [12]). However, this is not considered disruptive for experiments like HAYSTAC, as it can be handled by the detector's internal tuning systems [5]. In contrast, the 6–20 Hz vibrations from the cooling system directly shake the cavity. Thus, the damping system is considered especially successful if it reduces these sharp peaks up to 20 Hz down toward the laboratory's baseline noise level.



(a) (b)

FIGURE 2: (a) Log-scale power spectral density (PSD) of environmental vibrational noise measured at the dilution refrigerator scaffolding across all four permutations of pulse tube (PT) and active damping operation. (b) Visual representation of the 87.36% total reduction in band-limited RMS acceleration amplitude achieved within the critical 6–19 Hz frequency band, compared with the baseline RMS amplitude when the pulse tubes are off.

The data compared here were acquired during the same testing phase under identical signal processing configurations. Because the DR needs time to thermally stabilize when turning cooling subsystems on or off, data isolating the active pulse tube states (PT On) and the unexcited environmental baselines (PT Off) were recorded on separate days. However, within each respective day's observation window, conditions were separated chronologically only by the time required to toggle the active damping table (less than 5 min). In all cases, vibration data were recorded continuously at a sampling rate of 2000 Hz across five sequential blocks, but spectral processing was performed over the full continuous run. Consequently, all configurations of noise environment conditions shown in F ig. 2 utilize an identical total integration time of 163.84 s (1.64×10^5 ms). Utilizing Welch's method with a rectangular window, a segment length of 4096 samples, and 50% overlap, this expanded dataset yielded 159 independent segments for ensemble averaging to lower the statistical variance of the baseline noise floor.

Beyond a threshold of about 6 Hz, the undamped, active cryocooler spectrum (PT On, Damping Off) is heavily dominated by a dense cluster of intermediate peaks separated by about 2.5 Hz, labeled in Fig. 2(a) at 7.3, 8.9, 11.4, 13.8, and 16.3 Hz. Comparing this profile to the baseline state (PT Off, Damping Off) confirms that these distinct peaks are not ambient facility noise but rather the result of the natural structural resonances of the cryostat's metal frame and the rhythmic pulsing of the helium compressor. As demonstrated by the integrated band analysis in Fig. 2(b), the active damping system successfully suppresses this entire cluster of induced modes, reducing the total band-limited RMS noise within the 6–19-Hz region by 87.36% (from 0.5201 mV to 0.0657 mV) and returns the vibrational noise level near the cryostat remarkably close to its fundamental unexcited baseline of 0.0338 mV within that frequency band. Additionally, if we take into account frequencies from 1 to 200 Hz (the band which is of particular concern for the use of JPAs in the HAYSTAC experiment [4]), we continue to see noise attenuation with the pulse tubes on up to about 22%.

CONCLUSION

The power spectra obtained from this investigation are useful for understanding the vibrational noise environment near the DR in this cryogenic test laboratory and how well the current damping system reduces vibrational noise. We found a clear reduction in noise at the harmonics of the natural forcing frequency of the pulse tubes when using the damping system. With the vibrational noise environment characterized, the combination of active and passive damping employed in this laboratory is confirmed to be effective in mitigating this noise at the higher harmonics of the pulse tube signal, suppressing almost entirely peak resonances in the undamped signal, specifically reducing noise in the 6–19 Hz frequency band by 87.36%. Verifying that noise in this specific frequency band can be heavily mitigated using the current combined isolation methodology provides a critical, quantified benchmark for ongoing detector optimization and signal readout accuracy. Additionally, as the MATLAB script used to analyze the accelerometer data is saved within the laboratory computers, similar analyses can be completed easily with future noise measurements.

ACKNOWLEDGMENTS

Many thanks to Johns Hopkins associate physics professor Danielle Speller for her advice and mentoring, and to the members of her Speller Lab for providing a supportive environment for me to conduct this work as an undergraduate.

REFERENCES

1. P. Sikivie, "Invisible axion search methods," Rev. Mod. Phys. **93**, 015004 (2021).
2. A. J. AMillar, S. M. Anlage, R. Balafendiev, P. Belov, K. van Bibber, J. Conrad, M. Demarteau, A. Droster, K. Dunne, A. G. Rosso, et al., "Searching for dark matter with plasma haloscopes," Phys. Rev. D **107**, 055013 (2023).
3. BF-LD-SERIES, *BF-LD-SERIES CRYOGEN-FREE DILUTION REFRIGERATOR SYSTEM: User Manual*.
4. X. Bai, M. Jewell, J. Echevers, K. van Bibber, A. Droster, M. H. Esmat, S. Ghosh, E. Graham, H. Jackson, C. Laffan, et al., "Dark matter axion search with HAYSTAC phase II," Phys. Rev. Lett. **134** (2025).
5. K. M. Backes, D. A. Palken, S. A. Kenany, B. M. Brubaker, S. B. Cahn, A. Droster, G. C. Hilton, S. Ghosh, H. Jackson, S. K. Lamoreaux, et al., "A quantum enhanced search for dark matter axions," Nat. **590**, 238–242 (2021).
6. Bluefors, "How does a dilution refrigerator work?" Bluefors Newsletter (2023).
7. F. Pobell, *Matter and Methods at Low Temperatures, 3rd ed.* (Springer-Verlag, Berlin Heidelberg, 2007).
8. Table Stable, ABI-200LP Manual, *Active Vibration Isolation System AVI-200 LP: Instruction Manual*.
9. Model 393B04, *PCB Piezotronics, Inc. Model 393B04 Seismic, miniature (50 gm), ceramic flexural ICP® accel., 1 V/g, 0.06 to 450 Installation and Operating Manual*.
10. Picoscope 6000 Series, *Picoscope 6000 Series PC Oscilloscopes (A, B, C, and D Models): User's Guide*.
11. A. V. Oppenheim and G. Verghese, *Signals, Systems & Inference* (Pearson, London, 2016).
12. A. Megretski, "The waterbed effect," MIT OpenCourseWare, Lecture Notes for 6.245: Multivariable Control Systems (2004), MIT.

Maximum Nuclear-Recoil Energy in Elastic Dark Matter–Nucleus Scattering

Belkacem E. Khodja

Boston University, 700 Commonwealth Avenue, Boston, Massachusetts 02215, USA

khodja@bu.edu

Abstract. Multiple independent observations indicate that most of the matter in the universe is nonluminous and nonbaryonic. A leading experimental strategy is direct detection, in which a dark matter particle scatters elastically from a target nucleus, producing a nuclear recoil. In this paper, we derive the kinematic maximum recoil energy expected from elastic scattering as a function of the dark matter mass, the target-nucleus mass, and the incident speed (taking a representative galactic speed $v \sim 10^{-3}c$). Using representative target materials found in current experiments (cryogenic crystals, scintillators, and liquid noble gases), we evaluate how the maximum recoil energy scales across the low-mass and high-mass regimes. The results show a rapid rise in recoil energy when the dark matter mass is below the target-nucleus mass, followed by saturation in the heavy dark matter limit; for ($v \sim 10^{-3}c$), the saturation values are $\sim$245 keV for xenon and $\sim$52.3 keV for silicon. This simple kinematic analysis helps clarify how target choice constrains energy thresholds and, as a result, influences experimental sensitivity to different dark matter mass ranges without computing event rates or experimental limits.

INTRODUCTION

Understanding the composition of the universe remains a central goal of modern physics. Astronomical measurements consistently imply that the mass inferred from gravity is much larger than the mass that can be accounted for by luminous stars, gas, and dust [1–3]. Early evidence came from Zwicky's analysis of galaxy clusters, where velocity dispersions suggested a large "missing mass," and later from Rubin and Ford's analysis of galaxy rotation curves as well as studies on gravitational lensing and precision cosmology [1–5]. Collectively, these observations motivate the hypothesis of dark matter: a stable, nonluminous component that interacts predominantly through gravity and constitutes most of the matter content of the universe [1–3]. Establishing the particle nature of dark matter is therefore a major experimental objective [1–3,6].

Following the original dynamical anomalies reported in clusters and galaxies, the field has developed several complementary search strategies [1–3]. Indirect-detection experiments look for dark matter annihilation or decay products (such as gamma rays, antimatter, or neutrinos), while collider searches attempt to produce dark matter in high-energy particle collisions and infer its presence through missing momentum [2,3,6]. Direct-detection experiments, the focus here, aim to observe the tiny energy deposition from a dark matter particle scattering inside a terrestrial detector [6–8]. Because each approach probes different interaction models and systematics, consistency across multiple techniques will be essential for a convincing discovery [2,3,6].

Direct-detection methods aim to observe dark matter through its interaction with ordinary matter, most commonly via elastic scattering from nuclei [6-9]. In a typical event, a dark matter particle transfers a small fraction of its kinetic energy to a target nucleus, generating a recoil that can be measured through scintillation light, ionization charge, phonons, or a combination of channels, depending on the detector technology [6–9]. The measurable recoil spectrum depends on both particle physics (interaction type and mass) and astrophysics (the local dark matter velocity distribution) [6–8,10]. A useful, detector-independent quantity is the maximum recoil energy permitted by kinematics. In the following, we derive the maximum possible nuclear-recoil energy and use it to interpret how detector thresholds and target masses jointly shape sensitivity across candidate dark matter masses [6,7,12]. This sets the upper edge of the recoil range and provides intuition for which target materials are best matched to a given dark matter mass scale [7,8,12].

RECOIL-ENERGY KINEMATICS

In the case of direct detection, we consider elastic scattering of a dark matter particle of mass m_χ from a target nucleus of mass M_T, as shown in Fig. 1. As described in [12], for non-relativistic scattering, the nuclear-recoil energy E_R is related to the momentum transfer magnitude $q = \|\mathbf{q}\|$ through

$$q = (2\, M_T\, E_R)^{\frac{1}{2}}. \tag{1}$$

Writing q in terms of the reduced mass $\mu = m_\chi M_T/(m_\chi + M_T)$, the incident relative speed v, and the center-of-mass (CM) frame angle θ between the incoming and outgoing relative momenta, one obtains

$$q = 2\mu v \sin\frac{\theta}{2}. \tag{2}$$

Equating the right-hand sides of Eqs. (1) and (2) and then solving for E_R, one obtains

$$E_R = \frac{2\mu^2 v^2}{M_T} \sin^2\frac{\theta}{2}. \tag{3}$$

To maximize E_R for backward scattering, set $\theta = \pi$. Hence, the maximum recoil energy $E_{R,max}$ is

$$E_{R,max} = \frac{2\mu^2 v^2}{M_T}. \tag{4}$$

This maximum recoil energy is the key quantity used in the calculations and comparisons that follow.

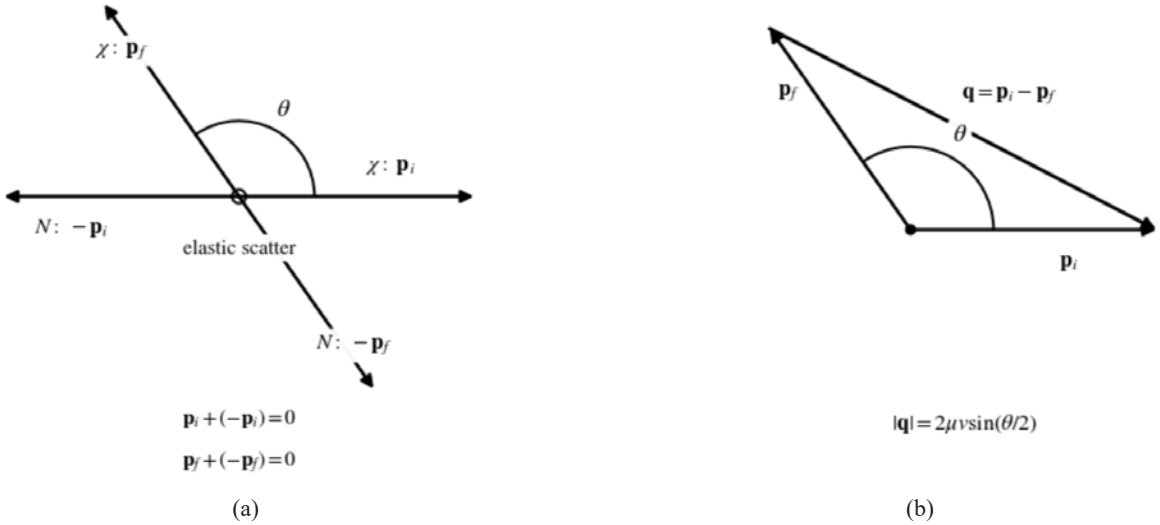


FIGURE 1. (a) Schematic of elastic scattering in the center-of-mass (CM) frame. The target nucleus is labeled *N* and the dark matter particle is labeled χ. The total momentum of *N* and χ before and after the scattering is 0. Furthermore, the momentum of *N* is equal in magnitude but opposite in direction to χ before and after the scattering. (b) Momentum-transfer geometry in the CM frame. The CM frame angle θ between the incoming and outgoing relative momenta is shown alongside the momentum transfer. By setting $\theta = \pi$, we obtain backscattering.

NUMERICAL RESULTS

We plot Eq. (4) as a function of the dark matter mass m_x for several target materials used in direct-detection technologies: xenon in liquid noble gas chambers, silicon in cryogenic crystals, and sodium and iodine in scintillating crystals. We set $c = 1$ and M_T as provided in Table 1 in Appendix A. We also set $v \sim 10^{-3}c$ as a representative galactic dark matter speed. This value is not intended to model the full local velocity distribution; rather, it provides a simple benchmark of the same order as the velocity that is commonly used, which is $v \sim 220$ km/s [11].

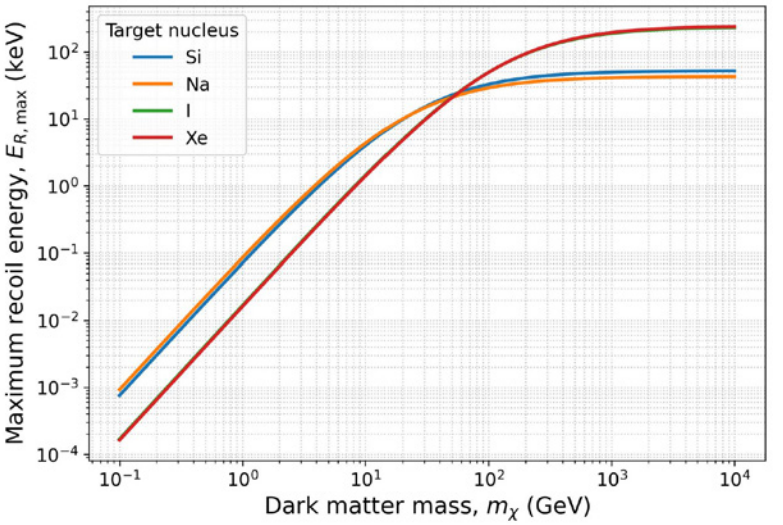


FIGURE 2. Plot of maximum recoil energy $E_{R,max}$ against the dark matter mass m_x for the target materials in Table 1.

The dark matter mass range is chosen to span a broad range, from sub-GeV-scale candidates to multi-TeV weakly interacting massive particle (WIMP)-scale candidates. It is not meant to include ultralight candidates such as axions. The elastic, nonrelativistic treatment used here is appropriate as a first-pass kinematic model for the mass ranges we consider as well as our choice of the galactic dark matter speed. From Fig. 2 we observe that the maximum recoil energy from elastic scattering plateaus for all target materials. This follows because in the large dark matter mass limit ($m_\chi \gg M_T$), the reduced mass μ approaches the target mass M_T, so the kinematic maximum recoil energy becomes approximately target dependent and independent of m_χ. In this limit, the plateau values are $E_{R,max} \sim 52.3, 42.8, 236,$ and 245 keV for silicon, sodium, iodine, and xenon, respectively. For an experiment to be sensitive to such recoils, the detector's effective threshold must lie below the relevant recoil energies; at the kinematic level this can be expressed as

$$E_{R,max} \geq E_T. \tag{5}$$

However, E_T is experiment dependent because detectors measure observable quantities such as ionization and scintillation light rather than measuring the true nuclear-recoil energy E_R directly. Calibration is then used to convert these observables into recoil-energy estimates. Thresholds are often quoted either in keVnr (nuclear-recoil energy) or in keVee (electron-equivalent energy), and converting between them requires a detector-response model and calibration.

For example, in scintillating crystals composed of sodium iodide, the experimental threshold is often quoted in electron-equivalent energy, and nuclear recoils appear at reduced visible energies due to quenching. This makes the mapping between a detector's quoted threshold and the underlying nuclear-recoil energy nontrivial, and it reinforces why kinematic limits alone are not sufficient to claim sensitivity without a detector-response model. This point is especially relevant for canonical WIMP searches, where nuclear recoils are the primary signal channel. The trend in Fig. 2 illustrates an important design principle: at fixed incident speed, heavier targets tend to produce larger maximum recoil energies for a given dark matter mass, while lighter targets can provide improved kinematic matching to low-mass dark matter. In the heavy dark matter limit ($m_\chi \gg M_T$), $E_{R,max}$ approaches a target-dependent constant proportional to $M_T v^2$, whereas in the light dark matter regime ($m_\chi \ll M_T$), the maximum recoil energy falls rapidly

with m_χ. Explicitly, the reduced mass becomes $\mu \sim m_\chi$, so Eq. (4) becomes $E_{R,max} \sim 2{m_\chi}^2 v^2/M_T$. Hence, the maximum recoil energy scales quadratically with dark matter mass: decreasing m_χ by a factor of 10 decreases $E_{R,max}$ by a factor of 100.

These scaling behaviors help explain why modern experimental programs deploy multiple target materials to cover a broad dark matter mass range, particularly across the WIMP-motivated mass scales. Scintillating detectors such as COSINE-100 use sodium and iodine targets to search for WIMPs [13]. On the other hand, cryogenic experiments such as SuperCDMS use silicon crystals to explore smaller dark matter masses within the sub-GeV range [14]. Liquid Xenon detectors such as LUX and XENON1T measure scintillation and ionization in dual-phase time-projection chambers in their search for WIMPs [15,16].

CONCLUSION

We used elastic-scattering kinematics to derive and visualize the maximum nuclear-recoil energy expected in direct-detection experiments as a function of dark matter mass and target nucleus mass. By comparing representative detector materials, we highlighted the transition from the low-mass regime, where recoil energies are strongly suppressed, to the high-mass regime, where the maximum recoil energy saturates. Although a full sensitivity projection requires detector-specific efficiencies, backgrounds, and response models, the maximum-recoil framework provides a clear first-pass guide for interpreting energy thresholds and for understanding why a portfolio of target materials is valuable. Continued improvements in low-threshold technologies, background rejection, and exposure will be critical for advancing discovery potential across the remaining viable dark matter parameter space.

ACKNOWLEDGEMENTS

We thank Prof. Zeynep Demiragli for her advice and help.

REFERENCES

1. G. Bertone and D. Hooper, "A history of dark matter," Rev. Mod. Phys. **90**, 045002 (2018).
2. L. Bergström, "Dark matter evidence, particle physics candidates and detection methods," Ann. Phys. (Berlin) **524**, 479–496 (2012).
3. S. Navas, C. Amsler, T. Gutsche, C. Hanhart, J. J. Hernández-Rey, C. Lourenço, A. Masoni, M. Mikhasenko, R. E. Mitchell, C. Patrignani, et al., "Review of particle physics–dark matter," Phys. Rev. D **110**, 030001 (2024).
4. F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln," Helv. Phys. Acta **6**, 110–127 (1933).
5. V. C. Rubin and W. K. Ford, Jr., "Rotation of the Andromeda Nebula from a spectroscopic survey of emission regions," Astrophys. J. **159**, 379–403 (1970).
6. J. Billard, E. Figueroa-Feliciano, and L. Strigari, "Direct detection of dark matter—APPEC committee report," Rep. Prog. Phys. **85**, 056201 (2022).
7. M. Schumann, "Direct detection of WIMP dark matter: Concepts and status," J. Phys. G **46**, 103003 (2019).
8. T. M. Undagoitia and L. Rauch, "Dark matter direct-detection experiments," J. Phys. G **43**, 013001 (2016).
9. L. Baudis, "Dark matter direct detection: status, results and future plans," J. Phys.: Conf. Ser. 3162 012007, 2025.
10. K. Freese, M. Lisanti, and C. Savage, "Colloquium: Annual modulation of dark matter," Rev. Mod. Phys. **85**, 1561 (2013).
11. N. W. Evans, C. A. J. O'Hare, and C. McCabe, "Refinement of the standard halo model for dark matter searches in light of the Gaia Sausage," Phys. Rev. D **99**, 023012 (2019).
12. J. D. Lewin and P. F. Smith, "Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil," Astropart. Phys. **6**, 87–112 (1996).
13. G. Adhikari, P. Adhikari, E. Barbosa de Souza, N. Carlin, S. Choi, M. Djamal, A. C. Ezeribe, C. Ha, I. S. Hahn, E. J. Jeon, et al., "Search for a dark matter-induced annual modulation signal in NaI(Tl) with the COSINE-100 experiment," Phys. Rev. Lett. **123**, 031302 (2019).
14. R. Agnese, A. J. Anderson, T. Aramaki, I. Arnquist, W. Baker, D. Barker, R. Basu Thakur, D. A. Bauer, M. A. Bowles, P. L. Bring, et al., "Projected sensitivity of the SuperCDMS SNOLAB experiment," Phys. Rev. D **95**, 082002 (2017).
15. D. S. Akerib, H. M. Araújo, X. Bai, A. J. Bailey, J. Balajthy, S. Bedikian, E. Bernard, A. Bernstein, A. Bolozdynyaet, A. Bradley, D. Bryam, et al., "First results from the LUX dark matter experiment at the Sanford Underground Research Facility," Phys. Rev. Lett. **112**, 091303 (2014).
16. E. Aprile, J. Aalbers, F. Agostini, M. Alfonsi, L. Althueser, F. D. Amaro, M. Anthony, F. Arneodo, L. Baudis, B. Bauermeister, et al., "Dark matter search results from a one ton-year exposure of XENON1T," Phys. Rev. Lett. **121**, 111302 (2018).

The appendix is available at https://pubs.aip.org/aip/jurp.

Magnetic Field Amplification Through Magnetic Dipole Vortices Merging in the Upstream of Relativistic Electron/Ion Collisionless Shocks

Mariane Diby[a)] and Neda Naseri[b)]

Department of Physics, Adelphi University, Garden City, New York 11530, USA

[a)]Corresponding author: marianediby@mail.adelphi.edu
[b)]nnaseri@adelphi.edu

Abstract. Relativistic collisionless shocks generate magnetic fields and accelerate particles in high-energy astrophysical environments. Using fully 2D kinetic particle-in-cell simulations, we investigate the physical mechanisms responsible for magnetic field amplification due to magnetic dipole vortices (MDVs) merging in the upstream region of the relativistic collisionless electron/ion shocks. We demonstrate that magnetic growth continues even after the nonlinear stage of MDVs. We show that the merging of MDVs increases magnetic field amplitude by 50% and enhances transverse size. This process provides a mechanism for large-scale magnetic field amplification in relativistic electron/ion collisionless shocks, relevant for understanding particle acceleration in astrophysical systems.

INTRODUCTION

Understanding the origin and energization mechanisms of cosmic rays remains a central problem in high-energy astrophysics. Observations indicate that cosmic rays are strongly associated with collisionless shocks in astrophysical environments such as supernova remnants, relativistic jets, and gamma-ray bursts [1]. In these plasma environments, particle–particle collisions are negligible, and energy dissipation is instead mediated through collective electromagnetic interactions. Studying collisionless shock formation and evolution is therefore essential for understanding how particles are accelerated to high energies in astrophysical systems.

A key mechanism mediating relativistic collisionless shocks is the Weibel instability [2, 3]. This instability produces transverse current filaments and rapidly amplifies small magnetic perturbations. As these filaments evolve nonlinearly, they merge and reorganize, forming larger magnetic structures (magnetic dipole vortices, MDVs) and enhancing field strength. Numerical and theoretical studies have shown that the Weibel instability process is crucial for shock formation, particle isotropization, and magnetic field amplification in relativistic plasmas [2, 4].

Particle-in-cell simulations of relativistic collisionless shocks reveal that the nonlinear evolution of filamentary structures can generate magnetic structures upstream of the shock [5]. In particular, localized MDVs develop from the merging and pinching of current filaments produced after the nonlinear stage of the ion Weibel instability [5]. These MDVs appear after the nonlinear stage of filament evolution and propagate opposite to the shock propagation direction, creating transient regions of intense magnetic fields.

Although previous studies have demonstrated the formation of MDVs [5], their merging remains less explored. Despite their potential impact on the large-scale magnetic structure of collisionless shocks, the dynamics of MDVs merging and the resulting amplification of magnetic fields have not been studied [5, 6, 7]. The present work investigates the behavior of MDVs merging in the upstream region of the relativistic collisionless electron/ion shocks and analyzes the evolution of their merging processes and associated magnetic field amplification.

NUMERICAL METHOD

To investigate these processes, we analyzed data obtained from kinetic plasma simulations based on the particle-in-cell (PIC) method [5].

The simulations analyzed in this work were performed using the Virtual Laser Plasma Lab (VLPL) code [8, 9], a fully electromagnetic relativistic PIC framework. The code was modified to minimize noise properties of numerical instabilities by using third-order shaped particles and current smoothing [5, 8]. The code solves the relativistic equations of motion for charged particles together with Maxwell's equations. Parallel computational techniques allow large-scale simulations of relativistic collisionless shock dynamics [5, 9]. The simulations used a rectangular

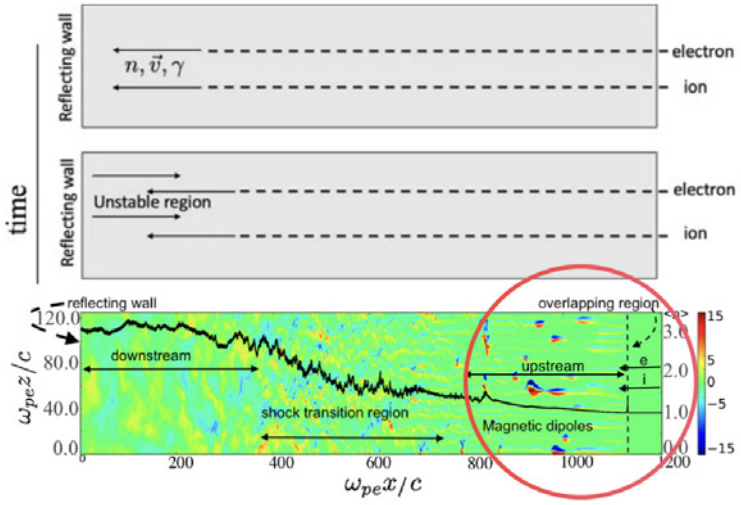


FIGURE 1. Schematic of the particle-in-cell simulation setup used to model relativistic collisionless shock formation. Top panel: Plasma streams propagate toward a reflecting boundary, producing counterstreaming electron and ion populations. The interaction between these streams triggers the Weibel instability, generating current filaments and magnetic field structures that mediate the formation of the collisionless shock. Bottom panel: Color plot of the out-of-plane magnetic field B_y in $x-z$ plane at $\omega_{pe}t = 1140$. The solid black curve (axis on right) shows transversely averaged density normalized to upstream unperturbed density. The dashed line is where the incoming e/i beams meet the counterstreams. All quantities are in normalized units [5].

simulation box in the x–z plane with dimensions $L_x = 1300\,\lambda_{pe}$ in the longitudinal direction and $L_z = 130\,\lambda_{pe}$ in the transverse direction and using the grid size $\Delta x = \Delta z = \lambda_{pe}/10$. Each computational cell contained 16 macroparticles (eight electrons and eight ions).

Electron/ion plasma flows with the mass ratio $m_i/m_e = 32$ and equal densities n_0 reflecting from the left boundary produce counterstreaming electron and ion populations. After reflection, the incoming and reflected plasmas interpenetrated and interacted, triggering the Weibel instability. This instability generates current filaments and magnetic field structures that mediated the formation of a collisionless shock [5]. The shock forms at $t \sim 420\ 1/\omega_{pe}$, where the density close to the wall reaches $3.1n_0$, in agreement with the jump condition $n = (1 + \frac{\gamma+1}{\gamma(\Gamma-1)})n_0$, where Γ is the adiabatic index in 2D geometry [4, 5]. A schematic representation of shock formation is shown in Fig. 1.

Physical quantities were expressed in dimensionless units normalized to characteristic plasma parameters. Time was normalized to the inverse electron plasma frequency ω_{pe}. Spatial scales were normalized to the electron skin depth $\lambda_e = c/\omega_{pe}$. Electromagnetic fields were expressed in normalized form as $\tilde{B}_y = \frac{eB_y}{m_e c\omega_{pe}}$, $\tilde{E}_{x,z} = \frac{eE_{x,z}}{m_e c\omega_{pe}}$ so that all quantities are dimensionless and consistent with relativistic PIC conventions. Current densities were also normalized as $\tilde{J} = J/n_e ec$. The incoming plasma flows are relativistic, characterized by the Lorentz factor $\gamma = 1/\sqrt{1-\frac{v^2}{c^2}} = 15$, where v is the plasma flow velocity. Previous simulations have shown that the nonlinear evolution of Weibel-generated current filaments can produce MDVs in the upstream region of the developing shock [5]. In this work, we analyzed magnetic-field data extracted from these simulations to identify MDVs merging and study magnetic field amplification due to merging process. By tracking these structures over time, we characterized the dynamics of MDVs merging and evaluated its role in amplifying magnetic fields in upstream regions of collisionless shocks.

SIMULATION RESULTS

Magnetic Vortex Formation

As the filamentary structures evolve, MDVs appear in the upstream region as discussed in [5]. Figure 1 shows a schematic of the interaction of two interpenetrating electron/ion beams. The bottom panel shows the color plot of the out-of-plane magnetic field B_y in the $x-z$ plane at $t\omega_{pe} = 1140$, as well as the MDVs in the upstream region (as

marked) in the simulation. The dashed line is where the incoming e/i beams meet the counterstreams. Transversely averaged plasma density normalized to unperturbed upstream density (black) is overlaid on the color plot of the magnetic field. MDVs propagate opposite to the shock propagation direction (+x) and interact with neighboring MDVs over time, as can be seen in successive B_y snapshots (Fig. 2).

MDV Merging and Magnetic Field Amplification

Figure 2 shows the merging process of the MDVs in the upstream region. The three columns display the charge density, the longitudinal current density J_x, and the transverse magnetic field B_y, respectively. Each row corresponds to a simulation time.

At early times, filamentary structures are visible in both the charge density and current density distributions. These filaments become increasingly pronounced as the merging occurs, and the magnetic field grows by the merging process. At later times, neighboring MDVs approach each other and merge. This behavior is visible in both the current density and magnetic field distributions. Figure 3 shows cross-sectional profiles of $B_y(z)$ extracted across the vortex structures before and after merging. The plots of charge density and longitudinal current density show that MDV merging happens due to the incoming stream's ion flow. Plots of charge density show ion currents in the middle of MDVs with electrons in the periphery. The plots of longitudinal current density show the direction of flow, which is in the $-x$ direction, opposite to the shock propagation direction.

In Fig. 3, we notice that before merging, localized peaks in B_y are observed at around 10 in normalized units. Similar behavior is observed by the transverse induced electric field ($E_z \approx v_x B_y/c$), as can be seen in Fig. 3. After merging, the peaks increase to approximately 15, and the spatial extent of the resulting MDVs expands. Multiple merging events follow a similar pattern, with consistent amplification of the peak B_y and growth of larger coherent magnetic structures.

DISCUSSION

The simulation results indicate that magnetic field amplification in the upstream region of relativistic collisionless shocks is primarily driven by the interplay of ion dynamics, filament interactions, and MDVs merging. Initially, the Weibel instability generates transverse current filaments accompanied by small-scale magnetic fields, as observed in Fig. 1. These filaments are subsequently reinforced by the motion of incoming ions, which sustain and strengthen the longitudinal current channels. This ion-driven reinforcement increases the stability and amplitude of the filaments, allowing magnetic structures to persist and grow beyond the electron-scale saturation.

Parallel current filaments attract one another due to Ampère's force law, leading to the coalescence of neighboring filaments. This attraction drives the merging of MDVs, as evident in the magnetic field evolution shown in Fig. 3. During merging events, the resulting MDV is transferred from smaller-scale filaments to larger, more coherent MDV structures, producing an increase in both the peak magnetic field amplitude and the spatial extent of the magnetic structures. In our simulations, the peak magnetic field increases from approximately 10 to 15 in normalized PIC units during vortex merging events, corresponding to an amplification of roughly 50%. This behavior is analogous to an inverse cascade in turbulent plasma, where chaotic, small-scale motions can generate massive, long-lived magnetic fields. Compared with previous studies [2, 4, 5], our work emphasizes the role of MDVs merging as a distinct mechanism for magnetic amplification of the MDVs. While earlier research highlighted MDVs generation and growth, we explicitly show that successive MDV coalescence contributes to amplifying the magnetic field and spatial coherence. The consistent amplification observed across multiple merging events suggests that this process may play a significant role in shaping the large-scale magnetic structure upstream of relativistic shocks. MDVs in the upstream region of relativistic electron/ion collisionless shocks play an important role in the preacceleration of both electrons and ions which then participate in the Fermi mechanism [8]. The energy gain of particles in the upstream region happens during the interaction of particles with evolving fields of MDVs. As MDVs evolve, their electromagnetic fields increase. Particles can gain a significant amount of energy during a short interaction with MDVs. The preaccelerated particles moving toward a shock can participate in the Fermi acceleration process in the shock transition region.

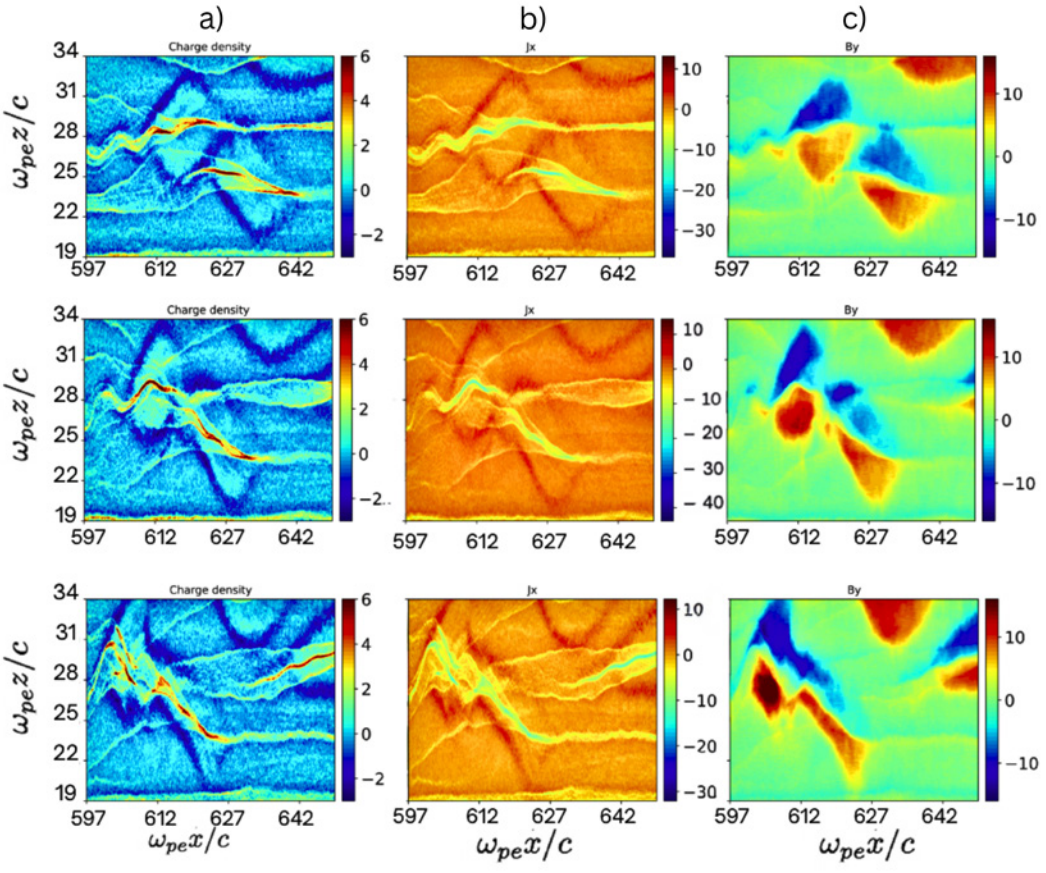


FIGURE 2. Temporal evolution of plasma structures showing the development of current filaments and early dipole-like magnetic structures at at $\omega_{pe}t = 726,\ 745,\ 760$, respectively. Columns correspond to (a) charge density, (b) longitudinal current density J_x, and (c) transverse magnetic field B_y. Rows show successive times from early to intermediate stages of the simulation. All quantities are in normalized PIC units.

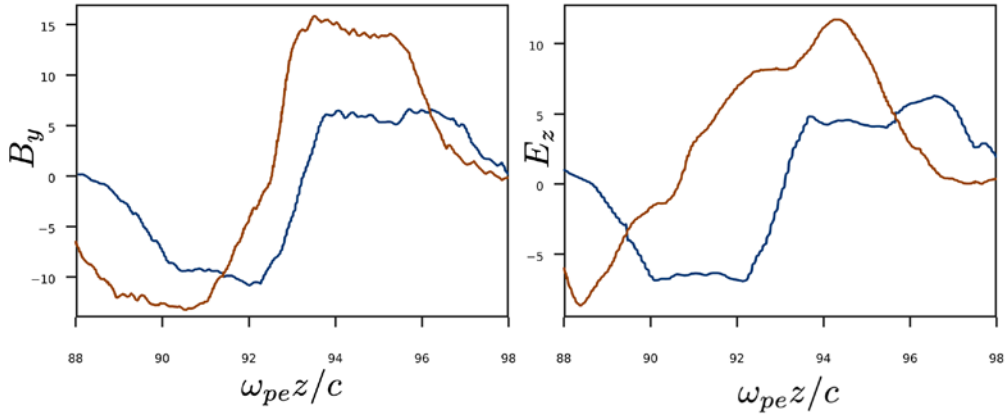


FIGURE 3. Evolution of the out-of-plane magnetic field B_y and transverse electric field E_z during MDV merging event at $\omega_{pe}t = 726, 760$. The panels compare profiles before (blue) and after (orange) the merging event. Following the interaction, the peak B_y increases from approximately 10 to about 15 in normalized units, and the transverse spatial extent of the magnetic structure broadens. E_z follows the same behavior.

JURPA only publishes the first four pages of papers in print. Continue reading at https://pubs.aip.org/aip/jurp.

Fabrication and Characterization of Plastic Scintillators with PPO and Coumarin in an Epoxy Matrix

Vyara Tsvetkova,[a] Keya Amundsen,[b] and Lila Goldenberg[c]

Department of Physics and Astronomy, Wellesley College, Wellesley, Massachusetts 02481, USA

[a]Corresponding author: vt101@wellesley.edu
[b]ka106@wellesley.edu
[c]lg119@wellesley.edu

Abstract. We fabricate plastic scintillators with varying concentrations of 2,5-diphenyloxazole (PPO) as a primary fluor and 7-diethylamino-4-methylcoumarin (coumarin) as a wavelength shifter and characterize their light yield. We demonstrate that effective scintillators can be made, and in one sample we achieve 27% higher light yield than in a commercially available scintillator. However, we also show that spatial variability has a significant impact on measured light yield. We show that higher concentrations of PPO and coumarin generally lead to higher light yields. We also demonstrate that the fabricated plastic scintillators can successfully detect muons in a CosmicWatch detector.

INTRODUCTION

Scintillators are widely used in particle detection and radiation measurement to convert the energy of ionizing radiation into visible light. For example, high-energy photons interact with a scintillator via the photoelectric effect, Compton scattering, or electron-positron pair creation, transferring their energy to electrons in the material. Molecules in the medium are excited by these electrons and subsequently emit deexcitation photons. Charged particles such as muons excite and ionize the molecules directly, triggering the same excitation/deexcitation processes, producing photons with a predictable, material-specific wavelength [1].

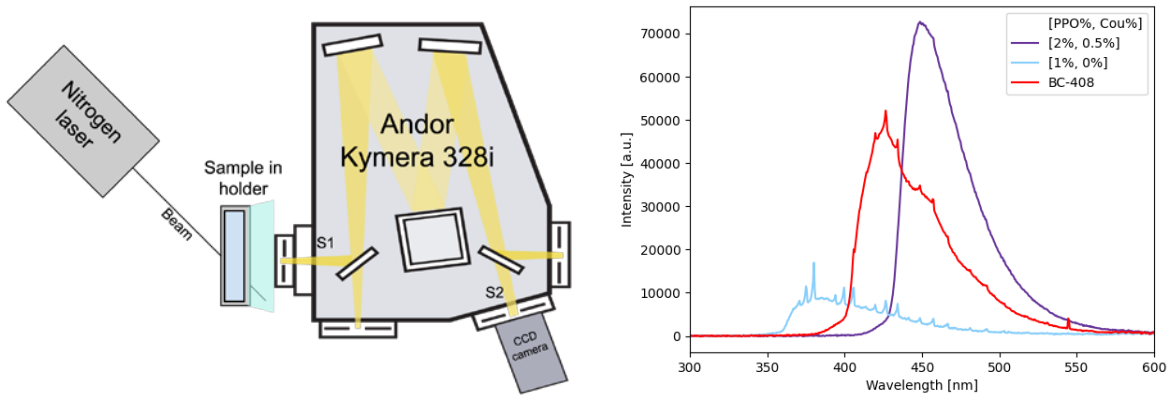
Plastic scintillators are transparent solids usually made from a polymer base such as polystyrene or polyvinyltoluene [1]. An organic fluorescent compound (fluor) is added to the base to make it scintillate [2]. In our case, the primary fluor, 2,5-diphenyloxazole (PPO), emits in the near-UV range, so a secondary wavelength shifter, 7-diethylamino-4-methylcoumarin (coumarin), is added to shift the emission to the visible spectrum since that is where most photon detectors are sensitive [1]. Having two wavelength shifters is a good way to limit the self-absorption within the scintillator by providing a larger Stokes shift, which is the separation between the wavelengths at which light is emitted and absorbed.

Commercial plastic scintillators such as Bicron BC-408 are readily available. BC-408 has a well-characterized emission spectra, peaking around 425 nm, and is frequently used in particle detectors. It has a high light output, fast response time, and good optical transparency [3]. However, we were interested in making our own scintillator to enable greater control of shape, size, and resulting photon emission spectra, since different photon detection systems may have different ranges of light collection capabilities. We particularly wanted a final emission wavelength well-matched to the SiPM chips used for CosmicWatch detectors [4]. We were also interested in demonstrating that it is feasible and inexpensive to make homemade scintillators in an undergraduate laboratory setting. Our highest concentration scintillator had a materials cost of 2.42 USD, compared to 21-25 USD for a similarly sized BC-408 scintillator [3]. Because we had previously used BC-408 with a CosmicWatch detector and were familiar with its characteristics, we used it as a reference for comparisons with the scintillators we fabricated.

This project originated in an upper-level undergraduate experimental techniques course at Wellesley College focused on particle physics detector technology. Our goals were to fabricate epoxy-based plastic scintillators using PPO as the primary scintillator and coumarin as the wavelength shifter and characterize their wavelength-shifting behavior and performance in a CosmicWatch muon detector [4].

METHODS AND MATERIALS

We chose to make our scintillators $5 \times 5 \times 1\,\mathrm{cm}^3$ so that we could use them in a CosmicWatch muon detector. To achieve the necessary volume, we used epoxy and hardener in a ratio of $1 : 0.83$ as specified by the manufacturer (the Epoxy Resin Store), resulting in a calculated weight of 22.51 g. We used concentrations of 0.5%–2% PPO and



(a) (b)

FIGURE 1: (a) Measurement setup. The laser beam is incident on the center of the sample but not the entrance slit of the spectrometer (S1). The dispersed light exits the spectrometer at slit S2, where it is imaged by a CCD camera. (b) Emission spectra of homemade (purple) and commercial (red) scintillators with a PPO-only sample (blue) shown for comparison. The spectra peak at 449 nm and 427 nm for the homemade and commercial scintillators, respectively. The roughness and additional peaks correspond to peaks in the nitrogen laser spectra that are above the absorption limit for PPO. We reason that coumarin in the purple sample absorbs additional light between 350 and 430 nm, resulting in a smoother spectrum.

0.01%–0.5% coumarin, as done in similar epoxy-based scintillator projects [5, 6]. We additionally made two reference samples, one with no PPO scintillator and another with PPO but no wavelength shifter. We measured and combined the amounts of epoxy and scintillators, stirring each sample by hand for 5 min until the mixture appeared visually homogeneous, and then added hardener. Samples were then cured on a hot plate for 2 h at around 60°C. Samples were not polished after fabrication.

To excite the PPO but not the coumarin, and therefore better isolate absorption by the primary fluor from absorption and emission by the wavelength-shifter, we ideally would have used a light source with a wavelength between 240 nm and 319 nm [7, 8]. We did not have such a source, so we instead used a nitrogen laser (337 nm). This wavelength still falls within the absorption range of the PPO, but it is long enough to directly excite the coumarin. A comparison of a sample containing only 1% PPO with a sample containing an additional 0.02% coumarin revealed a 72% increase in light yield for the coumarin-containing sample, indicating that direct absorption and emission by coumarin significantly contributes to light yield. We use two wavelength shifters, despite these complications, to give an overall larger shift in wavelength; the gap between PPO's emission and absorption peaks is only about 50 nm, but by adding coumarin, we can increase the gap to approximately 150 nm [7, 8]. The resulting emission wavelength of 499 nm is also better suited to our SiPM than that of just PPO at around 350 nm. As seen in Fig. 1(a), the laser was angled such that the beam was incident on the center of the sample without directly entering the spectrometer (S1). Spectra were averaged over 30 1-s measurements. To find the integrated light yield (ILY), we integrated the spectra over a 350–600 nm range. We defined relative light yield (RLY) as the ratio of each sample's ILY to that of the commercial sample. To test the spatial uniformity of our samples and quantify the effects of uneven mixing during the fabrication process, we recorded emission spectra at three locations (left, middle, right) on selected samples.

To test our scintillators in a particle detection context, we integrated one of our samples into a CosmicWatch muon detector [4].

RESULTS AND DISCUSSION

As seen in Fig. 1(b), the emission from the PPO-only sample peaked at 384 nm, while samples with both PPO and coumarin had a peak at 446 nm. Because there was no residual 384-nm peak in the PPO with coumarin sample spectrum, we conclude that the coumarin efficiently converted the PPO emission photons. Our scintillators have a higher peak emission wavelength than the commercial scintillator, although the emission intensity varies across samples, with RLYs ranging from 0.29 to 1.27.

TABLE 1: Relative light yield of our samples normalized to the commercial scintillator.

PPO (%)	Coumarin (%)	RLY	Name
	0.5	1.27	Sample 5
2.0	0.25	0.824	
	0.02	0.292	
	0.5	0.483	
1.0	0.25	0.455	
	0.02	0.423	
	0.5	0.458	
0.5	0.25	0.485	
	0.02	0.339	Sample 3

TABLE 2: Relative light yield for the different locations on the samples and the resulting spatial uncertainty, compared to the commercial scintillator measured at the center. The spatial uncertainty is defined as $\sigma_{\text{spatial}} = \max\left[\frac{L_{\text{left}}-L_{\text{center}}}{L_{\text{center}}}, \frac{L_{\text{right}}-L_{\text{center}}}{L_{\text{center}}}\right]$.

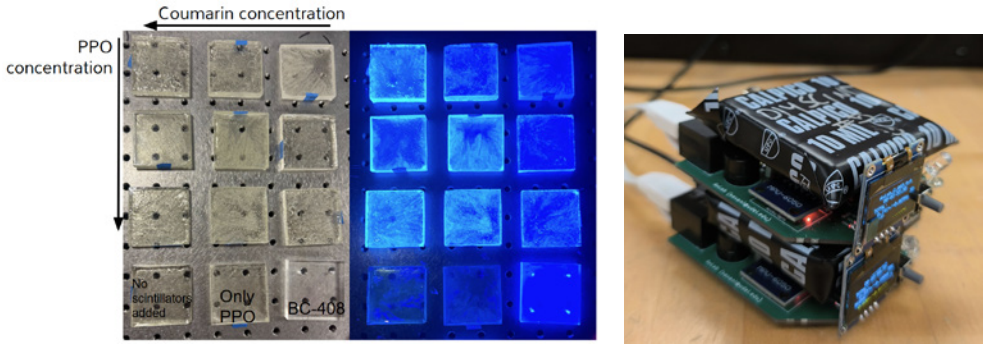
Sample	RLY_{left}	$\text{RLY}_{\text{center}}$	$\text{RLY}_{\text{right}}$	σ_{spatial} (%)
Sample 3	0.43	0.38	0.58	55.7%
Sample 5	0.62	0.64	0.69	8.2%
Commercial	0.98	1.0	1.09	8.8%

To identify a trend between PPO and coumarin concentrations and RLY, we created samples across a range of primary scintillator and wavelength shifter concentrations. The RLY values for these samples are shown in Table 1. These values were found with the laser beam incident on the approximate center of each sample. For a given PPO concentration, RLY increases with coumarin concentration. However, we do not find a monotonic increase of RLY with PPO concentration for a fixed coumarin concentration. This is particularly noticeable for the 0.02% coumarin samples. A likely explanation is that for the samples with the least coumarin (0.02%, corresponding to 5 mg), the 1-mg error associated with measurement and incomplete powder transfer has a larger proportional impact than for samples with higher concentrations of scintillator.

We selected the least and most visually uniform scintillators from our samples, samples 3 and 5, respectively [see Fig. 2(a)], to study the spatial variability in the RLY. We measured RLY with the scintillator illuminated at three different locations to see if the spatial variability impacts our light yield. We found no variation in the emission spectra shape or peak wavelength, but we observe significant variability in the RLY. The results are summarized in Table 2, and the spectra can be seen in Fig. 3. Sample 3 shows the largest spatial variability (56%), consistent with its appearance, while sample 5 and the commercial scintillator show comparable spatial uncertainties (8%), indicating similar levels of uniformity. The spatial variability in light yield suggests our mixing process could be improved. For future work, spatial variability studies could be done with samples of the same concentrations of PPO and coumarin. That would create a controlled comparison with all variables constant except the mixing quality.

These spatial nonuniformity tests had a slightly different setup than the PPO/coumarin concentration tests in Table 1. The differences in the setup were minor: small changes in the angle and distance to the laser. Our unpolished samples, made in a mold, have one flat side and one irregular side, though we did not keep track of which faced the spectrometer for the concentration light yield study. Although the RLY values show high variability resulting from spatial inhomogeneity in the samples, we are confident in the measured trend of RLY with PPO/coumarin concentration. Additional work that results in more homogeneous samples could help quantify the relationships between concentration and light yield. Similarly, we see a strong correlation between the amount of visual spatial nonuniformity of samples 3 and 5 and the measured spatial light yield differences. We note that the PPO and coumarin concentrations differ between the two samples, so their concentration could be a confounding variable.

To confirm that our scintillators were sensitive to ionizing radiation, we integrated sample 5 into a CosmicWatch muon detector. Coincidence data were first collected using two commercial scintillators integrated into two detectors in a stacked configuration. Next, we replaced the top BC-408 scintillator with our sample, otherwise maintaining the same configuration [see Fig. 2(b)]. Coincidence was recorded when both detectors in the stack reported an event at the same time, indicating a throughgoing muon. The coincidence rate is defined as the total number of coincidence events divided by the measurement time. The coincidence rate for the commercial-commercial configuration was $0.19 \pm 0.02\ \text{s}^{-1}$, while that for the sample 5-commercial configuration was $0.030 \pm 0.008\ \text{s}^{-1}$. This means that the muon detection efficiency with sample 5 is 16% of that with two commercial scintillators. This demonstrates that our sample is capable of detecting muons. It is important to note that the smoothness of the surface of the scintillator greatly impacts the photon detection efficiency: a polished surface provides a more uniform optical interface and increases the probability that photons are transmitted toward the detector [4]. The commercial samples were polished before integration into the muon detector; our homemade sample was not. We expect that polishing our scintillators would enhance the muon detection efficiency, though that study is beyond the scope of this work.



(a) (b)

FIGURE 2: (a) Photographs of scintillator samples under room light (left) and under UV illumination (right). Sample 3 is in the top right corner, and sample 5 is in the left column of the third row. (b) CosmicWatch muon detector configuration. Two detectors are stacked vertically, resulting in nearly 2π solid angle coverage on the sky.

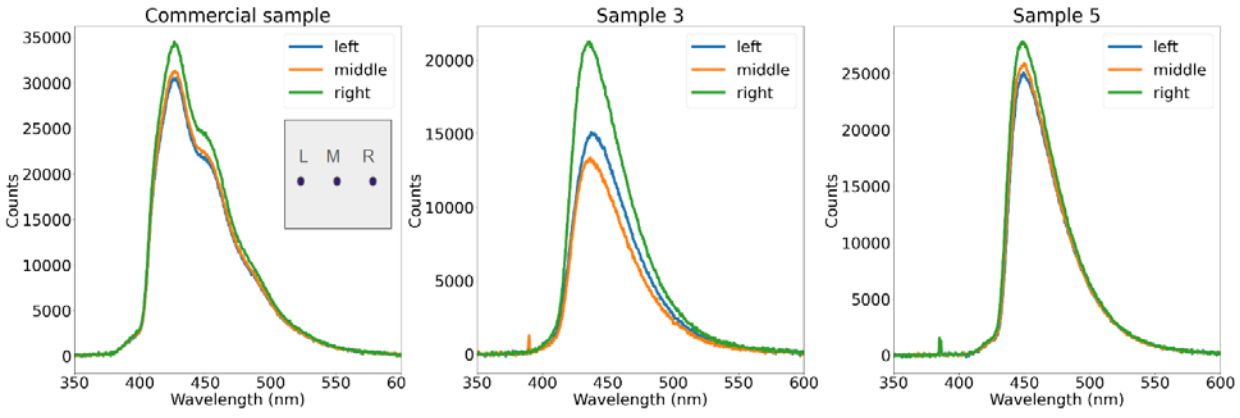


FIGURE 3: Emission spectra for the commercial sample, sample 3, and sample 5. The blue curve shows spectra from the left location, the orange curve the middle, and the green curve the right. The locations on the sample (L = left, M = middle, and R = right) can be seen in an inset in the first plot.

CONCLUSION

We created and characterized epoxy-based scintillators with a range of concentrations of PPO (fluor) and coumarin (wavelength shifter). In doing so, we demonstrated that using PPO and coumarin as primary and secondary scintillators works to convert UV light into visible light. We found that higher concentrations of both PPO and coumarin generally increase light yield, and we produced a sample with a higher light yield than a commercial scintillator. Many of our samples showed high spatial variability, so there is room for further investigation of synthesis methods to ensure more uniform distribution of the fluor and wavelength shifter powders in the epoxy. By integrating one of our samples into a CosmicWatch detector, we successfully measured muon candidates in coincidence, though at a lower detection efficiency relative to a commercial scintillator. This work shows that it is viable to make and characterize epoxy-based scintillators in an undergraduate laboratory.

ACKNOWLEDGMENTS

We wish to acknowledge our course instructors, James Battat and Jonathan Kemp, and our PHYS304 classmates at Wellesley College for their support in this project. We would additionally like to thank ABBA for the music.

JURPA only publishes the first four pages of papers in print. For references, visit https://pubs.aip.org/aip/jurp.

SPS JOBS

Where Will Your Physical Science Degree Take You?

Find REUs, internships, and research opportunities from leading organizations that help you:

- ✓ Build your résumé
- ✓ Gain research experience
- ✓ Prepare for graduate school

Applications are open now.
Start your search at SPS Jobs today.

Get started now!

Don't Miss Out!

Leading employers are posting student and graduate positions now.

Peak REU season is October–March.
Be sure to sign into your account and set up job alerts now to get notified when postings are added! The most sought-after roles fill early, apply now!

SPS JOBS

Explore exciting early career opportunities in the physical sciences with SPS Jobs, including:

- REUs, internships and entry level STEM jobs from leading employers in every sector
- Hiring trends and salary data
- Real-world stories from physicists working in a variety of roles across varied industries
- Tailored job alerts so you never miss an ideal opportunity
- *...and more!*

SCAN TO GET YOUR **COMPETITIVE EDGE!**

AIP

An Optical Emulation of the BB84 Quantum Key Distribution Protocol

Alec L. Riso,[1, a)] Karthik Thyagarajan,[2, b)] Connor Whiting,[3, c)] and Katherine Jimenez[4, d)]

[1] *Department of Physics, Cornell University, Ithaca, New York 14853, USA*
[2] *Department of Computer Science, Purdue University, West Lafayette, Indiana 47907, USA*
[3] *Department of Physics, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA*
[4] *Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*

[a)] Corresponding author: alr328@cornell.edu
[b)] kthyagar@purdue.edu
[c)] gcw3@illinois.edu
[d)] katiejim@umich.edu

Abstract. Quantum key distribution (QKD) is an approach to secure communication that uses the principles of quantum mechanics to generate a shared secret key which can be used in encrypting messages. Unlike traditional cryptography, which relies on mathematical problems too computationally difficult for a classical computer to solve, an ideal QKD protocol can be secure even against the capabilities of quantum algorithms. The BB84 protocol, a famous implementation of QKD, uses the polarization states of photons to carry information. We constructed an accessible, fiber-optic-based apparatus that emulates the generation of a key through the BB84 protocol. Although a true demonstration of BB84 would require a single-photon regime, our many-photon system shows the information encoding, transmission, and basis-sifting steps of the protocol. Due to its accessibility and visual clarity, our apparatus is intended as an educational tool that makes each logical step of the protocol visible.

INTRODUCTION

Quantum computing has the potential to completely reshape our technological landscape. With applications in areas such as AI, finance, and healthcare, efficient quantum algorithms can give us the ability to solve problems that are currently intractable. However, future advancements in quantum computing can present some alarming consequences. One of the most concerning possibilities is the breaking of asymmetric data encryption methods which keep our money safe and our personal data private. This leaves us vulnerable to attacks on financial data, medical data, and other personally identifying information.

Many of these concerns arise from *Shor's algorithm* [1], a quantum algorithm developed by Peter Shor in 1994. It allows for the factorization of numbers to be done in polynomial time with respect to the number of input bits, rather than the subexponential time required by the best-known classical algorithms, such as the general number field sieve [2]. As the numbers get larger, the gap between Shor's algorithm and its classical alternatives widens, potentially shortening solution times from centuries to minutes. Through its application to factorization and discrete logarithm computing problems [3, 4], Shor's algorithm can effectively break much of our current data encryption schemes, such as RSA, DSA, and ECC. Fortunately, implementing Shor's algorithm with large numbers would require a quantum computer with significantly more qubits than we currently possess. Although today's largest quantum systems are surpassing 1,000 physical qubits, breaking RSA encryption at current qubit error rates would likely require hundreds of thousands [5]. However, the pace of development in quantum hardware and quantum error correction suggests this number could keep decreasing.

Even though current quantum systems are too small to accurately run Shor's algorithm at scale, the possibility still poses a threat. A strategy known as "harvest now, decrypt later" can be employed by bad actors: they collect encrypted data now and store it until a sufficiently powerful quantum computer becomes available. While some recovered information may become obsolete, long-term government operations, financial history, or personal medical records could still be highly damaging if exposed.

One promising direction to mitigate this risk is quantum key distribution (QKD). Like traditional methods, QKD assumes key-based encryption, but it leverages the principles of quantum mechanics to achieve security [6]. Two core principles enable QKD's effectiveness. First, quantum states can exist in superposition, a probabilistic mix of possible measurable outcomes. Observation of a quantum state in superposition collapses the state onto one of these possible outcomes, which has the potential to expose the presence of an eavesdropper. In QKD, an eavesdropper is not certain of the quantum states they intercept, which means they cannot be sure if their measurements are collapsing

a superposition and disturbing the state. Second, the no-cloning theorem states that it is impossible to create perfect copies of an arbitrary quantum state [7]. This means the eavesdropper cannot copy the quantum states they intercept and simply measure the cloned states.

These properties make QKD well suited for applications requiring high levels of security, including the secure transmission of classified government and military information, the protection of sensitive financial data and transactions, and the safeguarding of confidential healthcare records. Despite its promise, QKD faces practical challenges. Current QKD systems are expensive and require specialized equipment and expertise, quantum signals degrade over long distances, and integrating QKD with existing infrastructure remains difficult. Nevertheless, research and development in QKD are rapidly advancing. As threats to classical encryption schemes grow, QKD stands out as a promising approach to secure communication.

Algorithm Overview

The fundamental goal of QKD is for two parties to establish a shared secret key, represented by a string of bits. Rather than sending a preexisting key through the channel, QKD distills the key from correlated measurement records that Alice and Bob build up and then compare with each other. The procedure must either keep this key private or, if someone is eavesdropping on the transmission, make that eavesdropping detectable.

This algorithm is done between a transmitter, denoted Alice, and a receiver, denoted Bob. Once the key is securely established and verified, and an encryption scheme is established with this key, Alice can publicly send messages to Bob, which Bob can easily decrypt without worrying that someone else has stolen their data.

Our optical emulation of QKD encodes information in the polarization of light carried by polarization-maintaining fiber-optic cables. Eavesdropping on such a channel is similar to how direct eavesdropping on classical fibers takes place. Much of it occurs at the entry, exit, and transition periods where the encrypted information gets processed or, in the case of the internet, web servers. Directly eavesdropping on the intermediate cable is more difficult and would noticeably disrupt the channel. Other implementations of QKD may be wireless, in which case eavesdropping can come in many other forms. In the case that an eavesdropper is found, another channel may be selected, or the transmission may be attempted at another time.

We focus on the implementation of QKD over fiber optics, modeled after the well-known BB84 protocol [8]. We built a communication network that allows real-time emulation of the protocol to viewers. The system showcases how the BB84 protocol works and serves as a valuable educational tool in optics and the potential future of quantum security. It is important to note that this setup uses a large number of photons, and therefore does not exactly replicate the BB84 protocol, which manipulates the quantum states of single photons [8]. The setup is modular and combines optics, electronics, and software in a way that makes every step of the algorithm visible and understandable. As the quantum era unfolds, teaching tools like this will be crucial to prepare the next generation of scientists and engineers.

Photonic Implementation and Measurement Bases

In order to emulate quantum key distribution using photonics, we must transport our information using photons. We store this information in the form of the polarization of light emitted from our nanosecond pulse laser. By first polarizing the unpolarized light streaming out from the laser, we can set all photons to a known initial state to modify later.

If two polarizers are oriented exactly perpendicular to each other, a photon that passes through the first polarizer will not pass through the second one. If we think of polarization as a quantum state vector, the vectors are orthogonal in this case. Additionally, two polarizers oriented in the same direction will cause 100% of the photons that pass through the first polarizer to pass through the second one. These can be thought of as parallel vectors. In our many-photon regime with approximately linear polarization, this interaction can be described classically using Malus's law:

$$I = I_0 \cos^2 \theta. \tag{1}$$

In this equation, θ is the angle between the polarization direction of the incident light and the axis of the polarizer, while I_0 and I are the incident and transmitted light intensities, respectively. Thus, Malus's law follows the same rules regarding orthogonal and parallel polarization directions. Our system will be in this many-photon regime, but a true

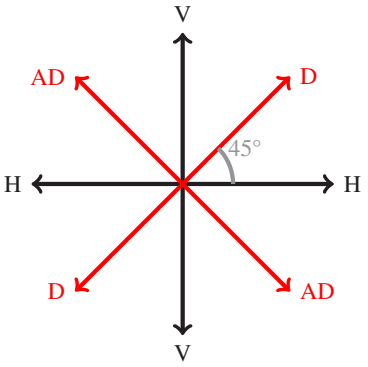


FIGURE 1. A diagram showing the orientation of polarization states corresponding to two bases.

BB84 protocol operating with single photons cannot be described classically. Regardless of the regime, we can think of polarization directions as vectors that can form a measurement basis.

Any two orthogonal polarization vectors can form a two-dimensional orthogonal basis. In this context, a basis refers to how we choose to measure a given vector in our two-dimensional polarization space. We can choose any two orthogonal polarization vectors by orienting our polarizer. This forms an orthogonal basis, which can handle any input polarization vector in our two-dimensional space but constrains the measurement output to one of the two basis states, which determines whether or not the light passes through.

Horizontally polarized light is orthogonal to vertically polarized light, and diagonally polarized light is orthogonal to antidiagonally polarized light, thus forming two separate bases as shown in Fig. 1. We can choose to measure light in either basis, but the output must be one of the two specified basis vectors of the basis that we chose. For example, if we measure one photon in the horizontal/vertical (H/V) basis by using a horizontal polarizer, it is forced to choose either the horizontal vector (pass through) or the vertical vector (do not pass through). Thus, if our aim is to get certain results, we would feed in light that is previously polarized to a basis vector (H or V in this example) so that we can fully predict the path the photon will take.

If one measures light polarized as a H/V basis vector in the diagonal/antidiagonal (D/AD) basis (and vice versa), the measurement projects the state onto one of the two D/AD basis vectors rather than reading out a definite preexisting value. Since a given H/V basis vector is 45 degrees from both D/AD basis vectors, a single photon has a 50% chance of being measured as diagonal and a 50% chance of being measured as antidiagonal. For the bright, many-photon pulses used in this apparatus, this appears as the optical power dividing approximately equally between the two D/AD outputs, as described by Eq. (1).

A step-by-step explanation of a BB84-style QKD protocol implemented using the previously described polarization bases is provided in Appendix A.

METHODS

To generate the light, we fed the output of a 980-nm laser directly into a polarization-maintaining (PM) optical fiber. This fiber then passed through an inline polarizer to regularize the polarization of all photons. We use the PM fiber in order to keep the polarization constant after we set the initial state with our inline polarizer. The fiber is then connected to a single-mode optical fiber, which is more vulnerable to polarization change. This fiber first passes through a motorized polarization controller that represents Alice. The Alice controller is manipulated by a servo motor, which changes the polarization state of the photons in one of four ways, depending on the desired bit and the randomly selected basis (H, V, D, or AD). This light travels through another PM fiber, which represents the communication channel. The light then passes into the second polarization controller, which represents Bob. This controller rotates the polarization state of the photons in one of two ways, depending solely on Bob's selected basis (H/V or D/AD). The servo motor settings are determined by an Arduino UNO, coded with specific degree measures as instructions.

Finally, the output is passed into an inline polarizing beamsplitter, which splits the light into two channels based on the polarization state. One channel represents a final result of a 0 bit, and the other represents a 1 bit. The channels

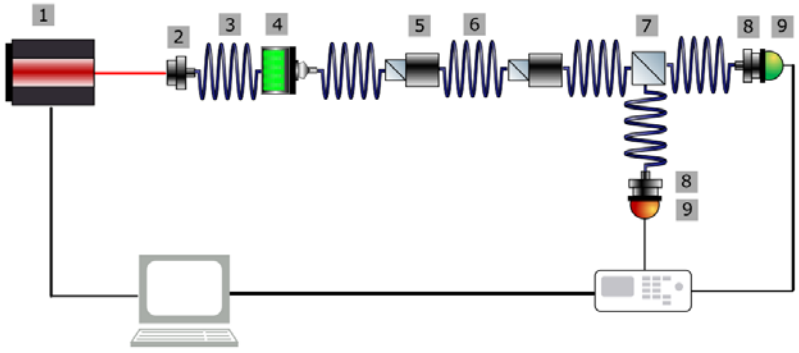


FIGURE 2. A diagram displaying the optomechanical setup of the physical QKD implementation. Light emerges from a nanosecond pulse laser [1] and is coupled [2] into a polarization-maintaining fiber-optic cable [3] with an initialized polarization [4]. Alice manipulates the polarization state [5,6] and sends the information to Bob, who manipulates the polarization again. The polarization states are divided by the beamsplitter [7] and collected into one of two photon detectors [8,9]. All part numbers and specifications are detailed in Appendix B.

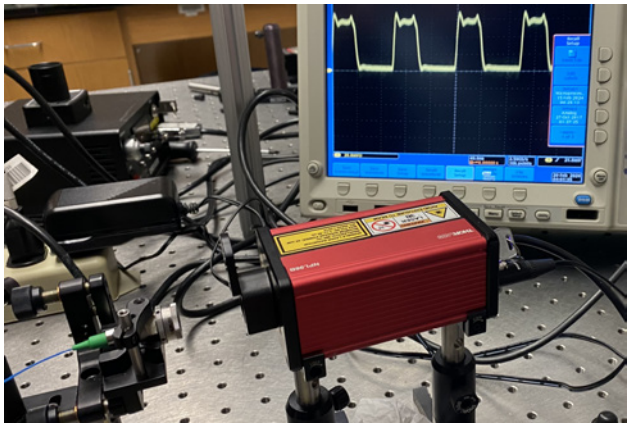


FIGURE 3. A photo of our nanosecond pulse laser, the fiber coupling interface, and the oscilloscope readout from our detector.

are then fed to two separate photon detectors and connected to an oscilloscope, which displays the power output from each channel. This output shows either a large signal in the 0-bit channel (detector 0), a large signal in the 1-bit channel (detector 1), or the optical power divided approximately equally between the two channels. Bob's detector results are displayed on an oscilloscope, where the readings for each channel can be displayed and uploaded to our computer for analysis. A diagram of our setup is shown in Fig. 2, and a photo of our laser and subsequent fiber-optic coupling is shown in Fig. 3.

In our many-photon regime, where we maintain linear polarization, the polarization interactions can be described by Eq. (1). If Bob chooses the same basis as Alice, he will rotate the polarization of light to the same measurement basis as the polarizing beamsplitter, leading to a large signal in one of the detectors. If he chooses incorrectly, the light will be rotated to a measurement basis 45 degrees offset from the polarizing beamsplitter, causing an even split between the signal in each detector. Although different, this is meant to represent the probabilistic 50/50 result that would happen to a single photon in a true BB84 implementation. Unlike the true single-photon implementation, the equal-power split directly reveals mismatched bases; this is a limitation of our demonstration.

The motorized polarization controllers work by applying a mechanical stress to the optical fiber, which induces birefringence, where there is a difference in refractive index based on the polarization axis. This causes the polarization components of the light to shift relative to one another, inducing a change in the overall polarization state. A specific servo-motor configuration corresponds to a certain applied stress and thus a certain shift in polarization,

Optimal Divergent and Elliptical Angle of Gaussian Beams in Trap Relying on Optical Pumping

Chanyeong Seo

Trapped Ion Quantum Information Group, ETH Zürich, Otto-Stern-Weg 1, 8093 Zürich, Switzerland
Georgia Institute of Technology, 225 North Avenue NW, Atlanta, Georgia 30332, USA

cseo33@gatech.edu

Abstract. Cooling and trapping neutral atoms is essential for quantum technologies and precision measurement. To cool atoms, lasers can be used to exert forces that slow the motion of the atoms, as if they move inside a dense fluid. Inside a magneto-optical trap (MOT), atoms can be conveniently cooled down to 1 mK by superimposing six laser beams in combination with a magnetic field gradient. Recently, a MOT has been realized by delivering divergent light directly from optical fibers to atoms, providing a compact, miniaturizable configuration. It has been shown that this system can further be simplified by removing the magnetic field gradient in a so-called trap relying on optical pumping (TROOP), which achieves confinement using only divergent laser light. To extend the fiber approach to TROOP effectively, we investigate the influence of divergence and polarization angles on trapping behavior and identify the optimal angles. We perform Monte Carlo trajectory simulations of cesium atoms on the $J_g = 4 \rightarrow J_e = 5$ transition to identify optimal divergent and elliptical angles at fixed initial beam power $I_{\text{initial}} = 1.11 \times 10^9$ W/m^2 and find that a beam divergence near 8° and perfectly circular polarization maximize trapping stability, whereas for angles smaller than 5°, the system reflects the behavior of an optical molasses with vanishing confinement.

INTRODUCTION

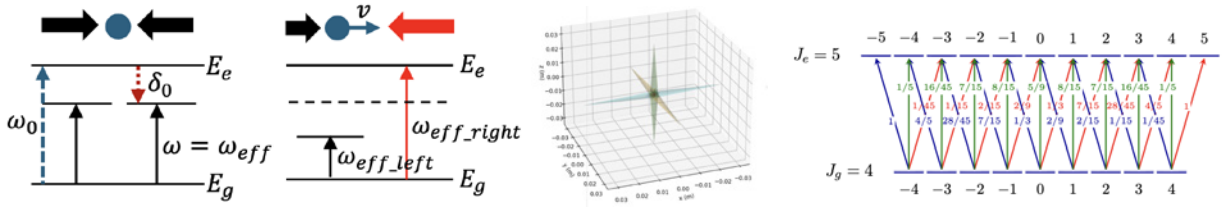
Trapping and cooling neutral atoms to ultracold temperatures, near absolute zero, has offered advances in quantum technology and precision measurement for cutting-edge applications, such as ultraprecise atomic clocks and qubits in emerging quantum computers. Cooling atoms refers to reducing their speed because the temperature is proportional to the average kinetic energy of the atoms. This can be achieved by illuminating the atoms with laser beams whose frequency is below an atomic resonance, called red-detuned beams, which exert a process known as Doppler cooling.

Consider a two-level atom with ground state E_g, excited state E_e, and resonance frequency ω_0 interacting with two counterpropagating laser beams of frequency ω, red-detuned by $\delta_0 < 0$, as illustrated in Fig. 1(a). For a stationary atom, both beams are equally likely to be absorbed because absorption is more probable when the beam frequency seen by the atom ω_{eff} is closer to the atomic resonance ω_0 and the beams have the same frequency ω. However, if the atom has a velocity, the Doppler effect shifts the beam frequencies in the perspective of the atom. As shown in Fig. 1(b), an atom moving to the right perceives a higher effective frequency $\omega_{\text{eff,right}}$ from the left-going beam and a lower effective frequency $\omega_{\text{eff,left}}$ from the opposite beam. Since absorption is more probable when the effective frequency is closer to the atomic resonance ω_0, the photons are preferentially absorbed from the beam opposing the atomic motion as $\omega_{\text{eff,right}} \approx \omega_0$ and $\delta_{\text{eff,right}} \approx 0$. Each absorption event transfers photon momentum, which in the red-detuned case, together with the Doppler shift, produces a velocity-dependent force that reduces the speed of the atom, called a Doppler cooling process.

To dampen the velocity of atoms in all directions, an optical molasses (OM) is yielded by arranging six counterpropagating beams along the x, y and z axes. As if the atoms move inside dense honey, they are slowed and cooled down to their theoretical lowest temperature, the Doppler temperature, typically on the order of hundreds of microkelvins. At this temperature, the motion of the atoms is too slow to produce a significant enough Doppler shift. Once the atoms are at the near-stationary state, they lose preferential absorbance of photons as in Fig. 1(a) with $\omega_{\text{eff,left}} \approx \omega_{\text{eff,right}}$, absorb a photon in any direction, and scatter in a random direction by reemitting the photon. Thus, although the OM cools and thus slows down the atoms, they eventually diffuse out of the interaction region with the laser beams and are lost. To overcome this lack of positional confinement in OM, a MOT combines Doppler cooling and magnetic fields. A magnetic field changes the atomic resonance transition similar to the Doppler effect. A position-dependent magnetic field combined with the OM laser beam configuration provides both the Doppler cooling and the restoring force toward the trap center simultaneously, ensuring that the cold atom remains at the center with a true confinement of ultracold atoms. In this work, we explore a different approach that provides a position-dependent restoring force without the need for a magnetic field, a trap relying on optical pumping (TROOP) [1]. The TROOP completely replaces the role of the magnetic fields with an optical pumping process using six divergent polarized beams to achieve

confinement, shown in Fig. 1(c). Carefully chosen polarizations among σ^+, π, and σ^- pump the electron into a specific energy level, shown in Fig. 1(d), which absorbs more photons from the lasers that offer position-dependent force; this process is named optical pumping. The optical pumping process will be explored thoroughly in the section on the TROOP Trapping Mechanism.

To create divergent beams in conventional free space, macroscopic lenses and six beams should be carefully aligned, which is experimentally demanding and sensitive to misalignment. Recently, a compact MOT design has been demonstrated by achieving divergent light directly from optical fibers to atoms [2]. This fiber-based approach fixes the beam geometry by construction, reducing the optical alignment complexity and providing a compact, miniaturizable configuration. To extend this approach to TROOP, which inherently requires divergent beams, we theoretically investigated how their divergence and polarization angles influence the resulting trapping behavior and identify optimal angles. We modeled the three-dimensional motion of cesium atoms in optical pumping between the energy level transition $J_g = 4 \rightarrow J_e = 5$ using a Monte Carlo trajectory simulation. Throughout the simulation, the beams have the power $I_{\text{initial}} = 1.11 \times 10^9$ W/m^2 to set a Rabi frequency Ω at the center similar to the natural decay rate of cesium Γ, following the choice of these parameters from the previous investigation [1]. Here, Ω measures how strongly the laser drives transitions between ground and excited states, while Γ describes how quickly electrons return to ground states from excited states. Operating with $\Omega \approx \Gamma$ keeps electrons in ground states, so they can continue to preferentially absorb photons from the beams. We expect that this theoretical exploration introduces and optimizes the optical fiber-based TROOP experiment, facilitating integration into compact quantum device-scale platforms.



(a) (b) (c) (d)

FIGURE 1. (a) Energy diagram for a stationary two-level atom with ground state E_g and excited state E_e. An atom (blue circle) is shown between two counterpropagating laser beams, whose arrow lengths are drawn proportional to the frequencies seen in the perspective of the atom. The atom has a resonance frequency ω_0, while the laser beams have frequency ω that is red-detuned from resonance by $\delta_0 = \omega - \omega_0 < 0$. (b) When the atom moves to the right with velocity $\mathbf{v}$, the Doppler effect shifts the frequencies of the two beams in the atomic frame. For an atom moving to the right, the left-going beam is shifted closer to resonance $\omega_{\text{eff,right}}$ (red arrows) and a decreased effective frequency $\omega_{\text{eff,left}}$. The atom then preferentially absorbs photons from the left-going beam. (c) Example TROOP geometry with six divergent beam. (d) Energy transitions $J_g = 4 \rightarrow J_e = 5$ with $\sigma^{\pm}$ and π couplings labeled by the squared Clebsch–Gordan (CG) coefficients. J_g has 9 ground states from $m_g = -4$ to $m_g = 4$ and J_e has 11 excited states from $m_e = -5$ to $m_e = 5$. The red, green, and blue lines and numbers show the electron's movement for the polarization components σ^+, π, and σ^-, respectively, and the corresponding $(C_q(m_g \rightarrow m_e))^2$ for each $m_g \rightarrow m_e$ transition pair.

THEORY

TROOP Trapping Mechanism

In addition to the Doppler effect, the traditional TROOP setup uses six divergent Gaussian laser beams, red-detuned from atomic resonance, arranged along the x, y, and z axes, as shown in Fig. 1(c). All beams are circularly polarized, but significantly, their helicities are arranged in a mixed configuration which must satisfy condition $h_z = -h_x = -h_y$ [1]. Off-center, the atom feels an intensity imbalance in the three possible polarization components of light (σ^+, π, σ^-), and this imbalance contributes a net restorative force. Furthermore, nonuniform transition probabilities associated with the various polarizations amplify biased preference in the polarity of the light. This is governed by the squared Clebsch–Gordan (CG) coefficient for an atom's ground-state magnetic quantum number m_g to an excited-state magnetic quantum number m_e pair, shown in Fig. 1(d). The magnitude of the squared Clebsch–Gordan coefficient $C_q^2(m_g \rightarrow m_e)$ reflects the probability of the polarization q to excite the atom from m_g.

For example, suppose that $h_z = -1$ and a cesium atom is in $z > 0$ and $m_g = 0$. Since the Clebsch–Gordan coefficients

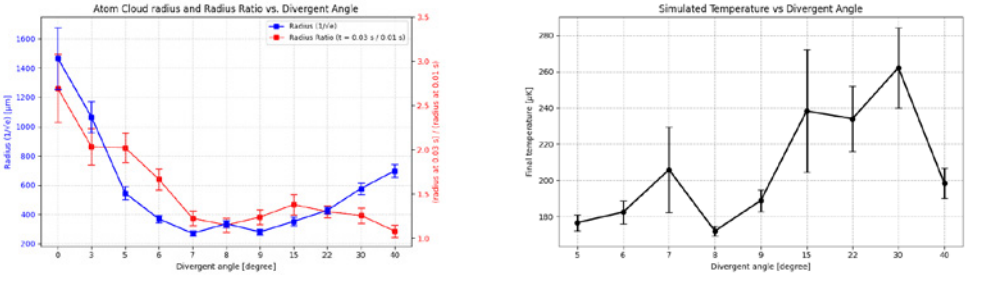
for all polarizations are equal to $1/3$ as shown in Fig. 1(d), σ^- polarized light, the highest intensity, is likely to be absorbed. Once an atom begins to absorb preferentially, optical pumping drives its internal state to reinforce that bias. From $m_g = 0$, the absorption of a σ^- photon excites to $m_e = -1$, with spontaneous decay likely to $m_g = -1$. From there, σ^- absorption is favored over σ^+ because $C_q^2(-1 \to -2) = 7/15$ exceeds $C_q^2(-1 \to 0) = 2/15$. Repeated cycles pump the atom to $m_g = -4$, where the σ^- transition to $m_e = -5$ is strongly preferred, suppressing the σ^+ transition from $m_g = -4$ to $m_e = -3$ by a factor of 45. Thus, both the intensity imbalance in the polarizations and the optical pumping contribute to the biased photon absorption of σ^- that forces the atom to $-\hat{z}$ from $z > 0$, toward the center. After crossing the origin ($z < 0$) the atom remains in $m_g = -4$ and continues to absorb σ^-, so the force still points towards $-\hat{z}$. Reversing the direction of the force requires pumping to $m_g = +4$ so that σ^+ dominates. Direct σ^+ excitation from $m_g = -4 \to m_e = -3$ is still strongly suppressed, and awaiting a large intensity imbalance leads to deviation from the trapping radius of the experiment. Instead, a π component offers an alternative path by exciting the electron from $m_g = -4$ to $m_e = -4$, followed by spontaneous decay to $m_g = -3$ preferentially, eventually reaching $m_g = +4$ and restoring a force towards $+\hat{z}$. The formation of π-polarized light is shown in Appendix A, and the force fields for different m_g states, divergent angles, and elliptical angles are presented in Appendices B, C, and D, respectively.

Monte Carlo Simulation Algorithm

When a ground-state sublevel m_g absorbs a photon, the excited magnetic quantum number is $m_e = m_g + 1$ for σ^+, $m_e = m_g$ for π, and $m_e = m_g - 1$ for σ^-, as shown in Fig. 1(d). For each beam j, we calculate the corresponding scattering rate using the effective detuning of the beam, including the Doppler shift and the local intensity of each spherical component shown in Eqs. (16a)–(16d). The scattering rate for that channel is proportional to the squared Rabi frequency Ω_{jq}^2 and the squared CG coefficient $C_q^2(m_g \to m_e)$ [3]. Summing over all beams and polarizations gives the total scattering rate R_{tot} [4]. For a small time step Δt, a scattering event is accepted with probability $P = R_{\text{tot}}\Delta t \ll 1$ [5]. If it scatters, we choose the specific beam and and update the atomic velocity according to the photon momentum. If no scattering occurs, the atom propagates ballistically. The detailed Monte Carlo simulation algorithm is explained in Appendix E. For this study, the initial beam intensity was fixed at $I_{\text{initial}} = 1.11 \times 10^9\,\text{W/m}^2$, and all six beams were positioned 3.5 cm from the origin, directed toward the center. We used a natural linewidth of $\Gamma = 2\pi \times 5.2$ MHz and a red-detuning of $\delta_0 = -\Gamma$ for all beams, with helicities satisfying $h_z = -h_x = -h_y = 1$.

RESULTS AND DISCUSSION

Optimal Divergent Angle of Beams



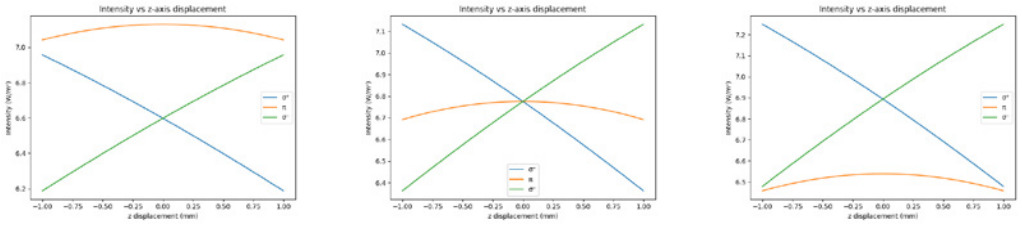
(a) (b)

FIGURE 2. Data is for a fixed peak intensity $I_{\text{initial}} = 1.11 \times 10^9\,\text{W/m}^2$ and elliptical angle 45°. (a) The blue curve shows variation of the largest $1/\sqrt{e}$ radius of the atomic cloud with beam divergence angle, and the red curve shows the ratio of the largest radius at $t = 0.03$ s to that at $t = 0.01$ s. (b) Simulated mean atomic temperature vs beam divergence angle. The error bars indicate standard error (SE) of the mean; radius SEs are computed across trajectories, and ratio SEs use first-order error propagation.

The blue curve in Fig. 2(a) shows that the trapping radius reaches the minimum level at 7° and 9°. Lower and higher

angles lead to higher trapping radii. To assess trapping stability independent of absolute size, the red curve shows how the trapping radius dispersed over time. For divergent angles less than 5°, this ratio exceeds 2.0, and the cloud remains relatively large despite temperatures near the Doppler limit, indicating a crossover to optical molasses. For angles larger than 15°, the radius increases even as the ratio approaches unity, consistent with a reduced intensity on the axis at large divergence and therefore weaker scattering and trapping. Figure. 2(b) shows that the mean temperature exhibits a minimum near 8° and then rises as the divergence increases, reflecting the decreased cooling efficiency.

Optimal Elliptical Angle of Beams



(a) (b) (c)

FIGURE 3. Intensity distributions of σ^+ (blue), π (orange), and σ^- (green) along the z quantization axis with an elliptical angle of (a) 44.5°, (b) 45° (perfect circular polarization), and (c) 45.5° when all six beams were divergent by 22°.

The simulations indicate that TROOP performance is highly sensitive to polarization ellipticity. We identified an optimal value of 45°, perfect circular polarization, and the small deviation of the elliptical angle causes diffusion within 0.01 s. As shown in Fig. (3), deviations from 45° increase or decrease the π component along the quantization axis, breaking the intensity symmetry between π and $\sigma^{\pm}$ and suppressing the required biased absorption. Consequently, the trap becomes almost identical to optical molasses, losing confinement.

CONCLUSION

We have performed Monte Carlo trajectory simulations to investigate the optimal beam configuration for a TROOP. The results demonstrate that a divergent angle near 8° yields the most favorable balance between the trapping radius, stability, and temperature, optimizing the system. Smaller angles than 5° lead to optical molasses, while larger angles reduce the efficiency of scattering. Likewise, the simulations confirm that TROOP is highly sensitive to polarization circularity, with the optimal elliptical angle of 45° corresponding to perfect circular polarization. Any deviation from this condition enhances the π component and suppresses the necessary imbalance between the σ^+ and σ^- transitions, thus weakening the trap. These findings establish the optical fiber-based design criteria for optimizing the performance of TROOP, guiding the development of simpler and more portable configurations for future quantum technologies. We suggest investigating the effect of initial beam intensities on TROOP in the future work.

ACKNOWLEDGMENTS

We thank Dr. Henrik Hirzler and Dr. Jonathan Home for their supervision and this invaluable opportunity.

REFERENCES

1. P. Bouyer, P. Lemonde, M. Ben Dahan, A. Michaud, C. Salomon, and J. Dalibard, Europhys. Lett. **27**, 569–574 (1994).
2. C. S. Nichols, L. M. Nofs, M. A. Viray, L. Ma, E. Paradis, and G. Raithel, Phys. Rev. Appl. **14**, 044013 (2020).
3. R. D. Glover, J. E. Calvert, and R. T. Sang, Phys. Rev. A **87**, 023415 (2013).
4. F. Atoneche and A. Kastberg, Eur. J. Phys. **38**, 045703 (2017).
5. R. K. Hanley, P. Huillery, N. C. Keegan, A. D. Bounds, D. Boddy, R. Faoro, and M. P. A. Jones, J. Mod. Opt. **65**, 667–676 (2018).

Appendices are available at https://pubs.aip.org/aip/jurp.

Suppression of Higher-Order Transverse Modes in a Nd:YVO_4 Laser Using Hard and Soft Aperturing

Benjamin Holt[a] and York Young[b]

Department of Physics, Utah Valley University, 800 W University Parkway, Orem, Utah 84058, USA

[a]Corresponding author: benjamin.holt@uvu.edu
[b]york.young@uvu.edu

Abstract. Laser resonators possess higher-order transverse modes (HOTMs) corresponding to solutions of Maxwell's equations within cavity boundary conditions. Thermal lensing at high pump powers in the gain medium can facilitate HOTMs, creating nonuniform intensity profiles that are detrimental if the laser drives nonlinear processes. This study investigates two methods of inhibiting HOTM onset: an intracavity iris ("hard" aperture) and a reduced pump laser radius ("soft" aperture). Data show that for 808-nm input power between 0 and 7.145 W, a hard aperture eliminates HOTMs, decreasing slope efficiency by 65.3%, whereas a soft aperture delays HOTM appearance, increasing slope efficiency by 11.0%.

INTRODUCTION

The Nd:YVO_4 1064-nm laser discussed in this paper will be used to pump a tunable, long-wave infrared optical parametric oscillator (OPO) for scanning breast cancer tissue for microcalcifications.[1] Calculations performed using the SNLO software package (AS Photonics, Albuquerque, NM) indicate that the OPO in this system will require ~6.5 W of average power to have sufficient intensity to drive the frequency conversion process when Q-switched (producing short, high-energy pulses). This conversion requires the Nd:YVO_4's output beam, which is powered by an 808-nm laser, to have a smooth, Gaussian phase front—a transverse electromagnetic (TEM) mode signified by TEM_{00}. TEM modes (propagating beam shapes) are represented by solutions to Maxwell's equations in a lasing cavity based on the cavity's boundary conditions, which are dictated by its physical geometry.[2] At low pump powers corresponding to weak thermal lenses, higher-order transverse modes (HOTMs) diverge too rapidly to be supported by a cavity designed for only TEM_{00}. However, as the 808-nm pump power approaches the level required to achieve 1064-nm output power greater than 6 W, the thermal lens focal length shortens. HOTMs which previously had high diffractive losses may now be supported by the cavity and impact the output beam.

A thermal lens forms when pump light absorption deposits heat into the gain crystal, leading to temperature gradients. One gradient is in the radial direction, and since the index of refraction is a function of temperature, the radial temperature gradient gives rise to an index of refraction gradient, creating a lensing effect. However, an abrupt axial gradient on the input and output faces of the crystal and the corresponding "bulging" of those ends due to thermal expansion dominates the overall thermal lens established by the pump beam.[3] Thermal lensing increases with laser pump power, therefore HOTMs become more supported by the cavity as pump power increases. These HOTMs have radii larger than TEM_{00}, as shown in Fig. 1.

One way to reduce this thermal lensing effect is to use a Nd:YVO_4 crystal that has undoped regions on one or both axial faces. This substantially reduces the temperature gradient at the physical end faces and reduces the thermal lens effect arising from bulging.[5] The laser crystal used in this experiment has an undoped front face, free of Nd ions.

Common practices for eliminating HOTMs include (1) placing a physical ("hard") aperture within the cavity to create losses for the larger-in-diameter HOTMs[2] and (2) using a "soft" aperture—a reduction in diameter of the incident 808-nm pump beam, creating an aperture that provides gain preferentially within a radius that is too small to support HOTMs but large enough to generate the TEM_{00} beam.[6]

While hard apertures impose diffractive losses on HOTMs because of the latter's larger diameter, they also cause diffractive losses for the TEM_{00}, reducing output power.[2] Soft apertures introduce no such losses, using a gain bias that favors the TEM_{00} mode while providing a gain distribution insufficient to support HOTMs.

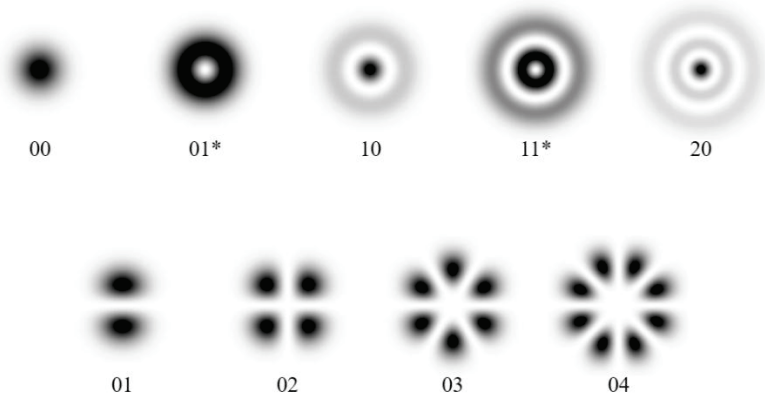


FIGURE 1. Theoretical spatial intensity profiles of various higher-order transverse modes (HOTMs).

MODEL

To model the TEM_{00} mode, the open-source software tool reZonator was used.[7] reZonator is designed for modeling, analyzing, and optimizing laser resonators and TEM_{00} beam propagation using the ABCD matrix method. Initial use of this tool for a flat input coupler (IC) and flat output coupler (OC) laser cavity design reported an unstable resonator. Only after including the thermal lens effect in the gain crystal was the laser resonator stable and reZonator able to calculate the TEM_{00} mode size at any location in the laser cavity.

This laser's geometry and crystal possessed the same physical dimensions and dopant concentration as studied by Zhuo et al.[5] That study calculated and measured the focal length of the thermal lens within the crystal at different pump powers for a pump spot size of 320 μm ($1/e^2$ diameter in intensity). In this study, the model reported by Zhuo et al. was used to estimate the thermal lens for slightly smaller pump spot sizes over a range of pump powers.

In the reZonator model, the doped region of the crystal was split in half and a thin lens, whose focal length represented the thermal lens for a given pump power, was inserted. Doing so enabled calculation of the TEM_{00} mode size within the cavity for any pump power and guided the choice of diameter and location of a hard aperture.

According to this model, the TEM_{00} beam had a radius of 193.7 μm at 3.15-W pump power at a location downstream from the Nd crystal. The ideal location for mode discrimination would be close to the IC, but a 4-mm gap between the IC and the crystal made aperture placement logistically impossible at that location.

An equation for HOTM sizes in cavities with cylindrical symmetry is[4]

$$w_{pl} = (2p + l + 1)^{1/2} w_0, \tag{1}$$

where l is the number of angular nodes, p is the number of radial nodes in the intensity transverse to the propagation axis, and w_0 is the TEM_{00} beam's radius (1/e in the electric field amplitude, or $1/e^2$ in intensity). Based on the TEM_{00} calculation from reZonator, the two lowest HOTMs (TEM_{01} and TEM_{10}) have 1/e amplitude radii of 268.7 and 329.1 μm, respectively.

EXPERIMENT

This experiment began with characterizing a baseline laser configuration which involved a flat IC; Nd:YVO_4 crystal; 200-mm focal length, UV-fused, silica AR-coated intracavity lens; and an 85% reflectivity output coupler. The crystal had 0.5% neodymium doping concentration. Its cross section was 4 mm × 4 mm and its total length was

11 mm. The front 4 mm in the axial dimension were undoped to reduce thermal lensing. An 808-nm, fiber-coupled diode laser was used to pump the Nd:YVO_4 crystal. For optical power input and output data collection, a Thorlabs S425C thermal power sensor was used. Its optical power range was from 2 mW to 10 W, with a 100-μW resolution. A diagram of the system is shown in Fig. 2. The physical cavity length was 147 mm, long enough for a Q-switch to be added at a later stage. Input 808-nm pump powers ranged from 0 to 7.15 W. The output beam's TEM modes were observed with the aid of a Kentek 1.37-in VIEW-IT infrared laser detector disc, which was placed in the far field near the power sensor. HOTM presence was judged visually via the lack of radial symmetry or the presence of radial nulls. Typically, both types of deviations from TEM_{00} were observed simultaneously, displaying a mix of HOTMs.

Following the characterization, a physical or "hard" aperture with a 320-μm radius was introduced into the cavity. While nonoptimal, the aperture was placed 42 mm downstream from the IC. This is where the TEM_{00}, TEM_{01}, and TEM_{10} modes' radii were calculated to be 193.7, 268.7, and 329.1 μm, respectively. Power data were again taken.

The final set of laser data was taken after the hard aperture was removed and the relative distance of the lens pair upstream from the IC was adjusted to shorten their effective focal length. This reduced the 808-nm pump radius from 274 to 265 μm. The pair was translated as a group to ensure the pump waist location was centered longitudinally (along the laser propagation axis) in the crystal.

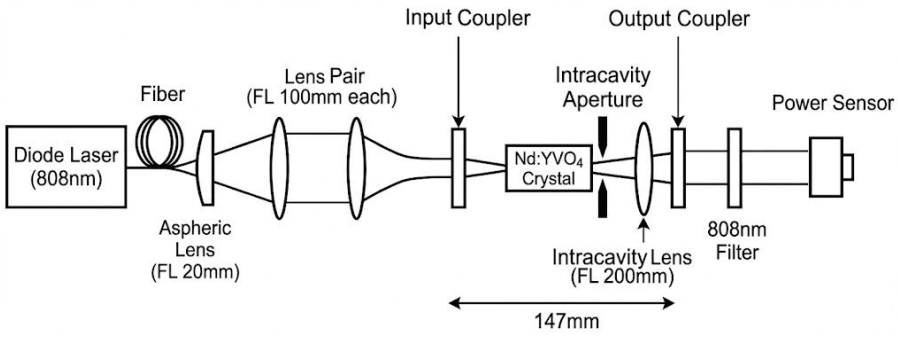


FIGURE 2. Experimental laser setup.

RESULTS AND DISCUSSION

Figure 3 presents the data sets gathered for the three laser configurations. Linear fits of the experimental data reveal a significant difference in performance between the original cavity and the hard-aperture configuration. Without the aperture, the system had a slope efficiency of 43.8% and HOTMs appeared at 3.12 W of input power. When the 320-μm intracavity aperture was placed in the system, the slope efficiency dropped to 15.3% but no HOTMs appeared over the entire pump power range. In theory, such a large loss in slope efficiency is not expected for this aperture size; less than 1% round trip loss for the TEM_{00} mode is expected.[2] This loss may be due to the hard aperture not being precisely centered radially about the TEM_{00} beam. Mounting the aperture to a finely adjustable *x*-*y* stage is the next step in testing this hypothesis.

In the soft-aperture configuration, the slope efficiency increased from 43.8% to 48.6% and the laser threshold pump power was reduced from ~1.1 to ~1.0 W. Such efficiency improvements are expected in a four-level laser, wherein small signal gain is proportional to pump intensity and the smaller pump radius leads to a higher intensity for the same pump power.[4] However, and more relevant to this study, the smaller gain aperture delayed the onset of HOTMs until 5.82 W of input pump power—a marked improvement.

Decreasing pump spot size to deter HOTM onset may initially seem counterintuitive, as a smaller pump spot increases the strength of the crystal's thermal lens, promoting HOTM appearance. These data, however, suggest that for slight reductions in spot size, soft-aperture HOTM discouragement outweighs the stronger thermal lens support. Future work on this system may include performing thermal lens focal length and mode size calculations to predict at what pump spot size this trend changes. A set of parametric measurements of HOTM onset and pump spot size could then be made to assess the validity of those predictions and find the optimal pump spot size.

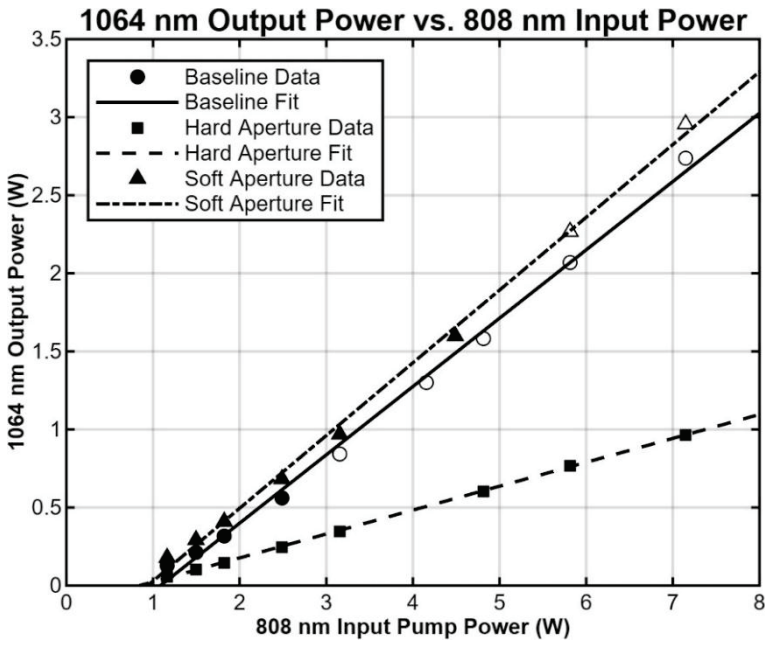


FIGURE 3. Output power of 1064-nm light vs. 808-nm pump power. The lines represent linear least-squares fits. Solid data points represent TEM_{00} power and hollow data points indicate HOTMs. The standard deviation of each data point was ~0.01 W, which is smaller than the data markers in this plot.

CONCLUSION

The inclusion of a 320-μm-radius intracavity hard aperture in the Nd:YVO_4 laser system proved effective in enforcing TEM_{00}-mode-only operation in the 1064-nm laser output. However, this mode purity came at a substantial cost to the laser's optical efficiency. While much of that efficiency loss is likely due to hard-aperture misalignment, the increased efficiency of the soft-aperture approach makes it more desirable for our 6.5-W TEM_{00} goal. With a pump beam area reduction of only 6.5% yielding successful HOTM suppression, exploring smaller pump spot configurations will comprise future studies of this system.

ACKNOWLEDGMENTS

The authors are grateful to those who have helped fund this project: the UVU College of Science Scholarly Arts Committee and Dr. Vincent M. Rossi for a share of a Grant for Engaged Learning. Additionally, they are thankful for the work of former research group members Alex Gibb, Brantson Wayman, and Isabell Hudnall. Special thanks to Dr. Paul Weber for the use of his microscope to measure the hard-aperture diameter. Lastly, the authors are appreciative of the efforts of anonymous reviewers for clarifying feedback and ideas for future investigation.

REFERENCES

1. Y.-P. Tseng, P. Bouzy, C. Pedersen, N. Stone, and P. Tidemand-Lichtenberg, Biomed. Opt. Express **9**(10), 4979–4987 (2018).
2. K. Ait-Ameur, Appl. Opt. **32**(36), 7366–7372 (1993).
3. B. Neuenschwander, R. Weber, and H. P. Weber, IEEE J. Quantum Electron. **31**(6), 1082–1087 (1995).
4. W. Koechner, *Solid-State Laser Engineering* (Springer, New York, 2006).
5. Z. Zhuo, T. Li, X. Li, and H. Yang, Opt. Commun. **274**, 176–181 (2007).
6. L. Y. Wang and G. Stephan, Appl. Opt. **30**(15), 1899–1910 (1991).
7. N. I. Chunosov, reZonator (Software), available at http://rezonator.orion-project.org.

Visualizing the Dispersion of Internal Wave Beams in Density-Varying Fluid

Vohn Jacquez,[1, a)] Zachary Phan,[2] Zachary Taebel,[3] Dylan Bruney,[4] Pierre-Yves Passaggia,[5] and Alberto Scotti[6]

[1)]*Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 28223, USA*
[2)]*Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina 27695, USA*
[3)]*Scripps Institution of Oceanography, University of California at San Diego, San Diego, California 92037, USA*
[4)]*Wake Forest University, Wake Forest, North Carolina 27109, USA*
[5)]*Laboratoire PRISME, Universite d'Orléans, Orléans, UPRES 4229, France*
[6)]*School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, Arizona 86287, USA*

[a)]Corresponding author: vbjacq8@unc.edu

Abstract. Physics curricula typically focus on electromagnetic, quantum, and standard mechanical waves propagating through media like air or taut strings. We investigated internal gravity waves, a peculiar type of mechanical wave that propagates through oceans and has significant geophysical implications. These dispersive waves obey a partial differential equation that accounts for the fluid density profile and buoyant force, resulting in a unique dispersion relation that couples their frequency directly to their angle of propagation. We explored this intricacy with a laboratory experiment to generate and observe internal waves in a more accessible environment than the ocean depths. In this experiment, we created and verified a linear saltwater stratification and applied an oscillatory topographic forcing. Because internal waves are invisible to the naked eye, we used background-oriented Schlieren optical techniques to track the density gradient and visualize wave propagation. Finally, we extracted a power spectrum of the density gradient to obtain the system's wave vector modes, quantitatively demonstrating the unique dispersive properties of internal waves.

INTRODUCTION

At the beach, onlookers may observe surface waves several meters high. Once they break, their energy shapes the shoreline. However, there also exist energetic waves that propagate below the surface. In this underwater abyss, these peculiar *internal waves* may span hundreds of meters, and their breaking instead shapes the distribution of nutrients, heat, and dissolved gases through vertical mixing. Their behavior has significant regional effects, including mitigating coral bleaching [1], nourishing ecosystems, and regulating melting rates of Arctic glaciers [2].

Beyond their environmental implications, internal waves are also peculiar mathematically. For example, they require stratified fluids (where density varies with height) to propagate. Their dispersion relation couples their frequencies directly to both the direction of propagation and the gradient of density, and their energy propagation is perpendicular to the phase propagation. Although the linear governing equation may resemble the classical wave equation, the buoyancy-restoring force associated with stable stratification produces this distinctive anisotropic dispersion relation [3, 4]. Our experiment specifically looked at the largest internal waves in our ocean, which are called *internal tides*. These waves arise when periodic tidal current flows over underwater ridges, converting barotropic tidal wave motion into baroclinic internal wave motion [5]. To emulate these waves and observe their unique behavior, we conducted a laboratory experiment that used a stratified tank and oscillatory forcing for wave generation, and optical analysis methods. These quick, inexpensive methods provide an accessible alternative to the costly field measurements and long research cruises typically required to study this phenomenon.

INTERNAL WAVE EQUATION SOLUTION

The internal gravity wave solution for stratified fluids is derived from the Euler equations, which describe momentum and mass conservation in an inviscid fluid. We simplified the equations to a solvable form with the following assumptions: First, the equations can be linearized by assuming small wave amplitudes. Second, since our stratification is weak, we invoked the Boussinesq approximation to characterize the density $\rho(z)$ as a reference density ρ_0 alongside a background density profile $\bar{\rho}(z)$ with small perturbations ρ', such that $\rho(z) = \rho_0 + \bar{\rho}(z) + \rho'$. Lastly, we assumed

motion is restricted to a 2D plane. With these approximations, the following equation was derived for the velocity vector $\mathbf{u}(x,z,t)$, where (x,z) denote spatial coordinates and t is time:

$$\frac{\partial^2}{\partial t^2}\nabla^2\mathbf{u} + N^2\frac{\partial^2}{\partial x^2}\mathbf{u} = 0. \tag{1}$$

Here, N refers to the Brunt-Väisälä frequency: $N = \sqrt{-\frac{g}{\rho_0}\frac{d\bar{\rho}(z)}{dz}}$ [6]. This is the frequency at which a fluid parcel oscillates when vertically displaced from its equilibrium (neutral buoyancy) position and thus measures the strength of our stratification. To seek solutions to Eq. (1), we substitute in a simple traveling wave of the form $\mathbf{u}(x,z,t) = \hat{\mathbf{u}}e^{i(kx+mz-\omega t)}+\text{c.c.}$ to Eq. (1). Here, k and m are the wavenumber components of the trial wave solution we verify, and ω is the frequency. This leads to

$$\omega^2(k^2+m^2)\mathbf{u} - N^2k^2\mathbf{u} = 0. \tag{2}$$

From Eq. (2), traveling waves are solutions to Eq. (1) provided their wavenumbers and frequency satisfy the *dispersion relation*:

$$\omega = N\sqrt{\frac{k^2}{k^2+m^2}} = N\sin\theta, \tag{3}$$

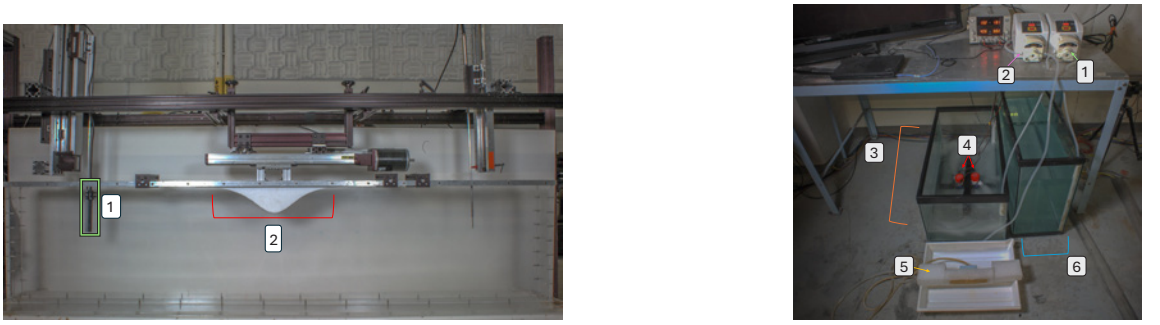
where θ is the angle from the horizontal at which the superposition of internal waves (wave group) propagates. One key consequence of Eq. (3) is that the wave group propagates perpendicularly to the phase velocity of its constituent modes (see Appendix B). Another key feature of this dispersion relation is that the frequency depends only on the ratio m/k and the Brunt-Väisälä frequency, not the wave vector's magnitude. Thus, waves of a given frequency must propagate at a fixed angle to the horizontal, independent of their wavelength. Conversely, a single frequency admits infinitely many wavelength modes (infinite degeneracy) corresponding to different magnitudes of the wave vector at the same angle. Because k and m are also sign-free, the waves are horizontally and vertically symmetric about the forcing. When $\theta = 90°$, the maximum allowable frequency is reached: $\omega = N$. Interestingly, although the dispersion relation was derived for the velocity, it applies to all wave variables including the density perturbation ρ' (when plugging the solution for $\mathbf{u}$ into the linearized Euler equations, the other variables also follow complex exponential solutions). The consequence is that we can characterize wave motion through ρ', a quantity that we optically determined in our experiment.

METHODS

We first used a two-bucket method with pure rock salt to create the necessary linear stratification [7]. This required three tanks: a freshwater holding tank, a saltwater tank, and a wave tank (203 cm long, 46 cm tall, 8 cm wide). Setting a 1:2 ratio for the saltwater tank's inflow to outflow resulted in a linearly stratified density profile, which corresponded to a constant Brunt-Väisälä frequency and a constant wave angle θ. We verified our stratification by lowering a conductivity probe [see Fig. 1(b)] down one edge of the tank with a stepper motor. We mapped the readings using a conductivity-to-density spline fit, calibrated beforehand with saline solutions of known densities. We measured a Brunt-Väisälä frequency of $N = 1.394$ rad s^{-1}.

To emulate internal tides, we needed an oscillatory forcing. However, replicating the reference frame of internal tides involves flowing water past a topography (typically done with an intensive oscillating wall or paddle [8]). Instead, we moved a partially submerged ridge in a periodic loop with a stepper motor to simulate flow over a bottom topography in the fluid's frame of reference [9]. Our model ridge took the form $h(x) = h_0\text{sech}^2\left(\frac{x}{h_0}\right) - h_0\text{sech}^2\left(\frac{l_0}{h_0}\right)$, with h_0 being the maximum height and l_0 being the half-length. With an acrylic stencil, we used a hot wire to trace the shape out of nonporous styrofoam, yielding $l_0 = 17.2$ cm and $h_0 = 3.1$ cm. We oscillated the topography at a constant frequency of $\omega_0 = 1.047$ rad s^{-1} with a stroke amplitude of 1.34 cm. More details of the setup can be found in [10].

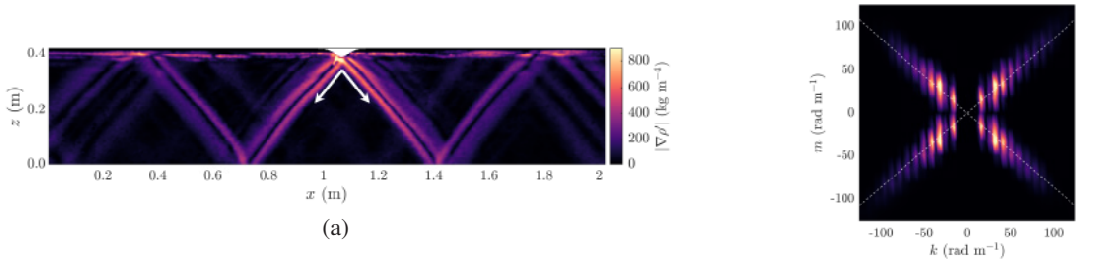
We visualized the dispersive behavior of internal waves with background oriented Schlieren (BOS) imaging. BOS relies on the differing refractive indices of fluids with varied densities, a relationship that also causes desert mirages. As with mirages, observers viewing the tank see a distorted image of the background. To utilize this phenomena, we projected a random dot pattern on a flat surface 280 cm behind the tank. We used an array of high-resolution cameras (3312×2488 pixels) positioned 146 cm in front of the tank to capture photos of the full dot pattern at 2 Hz. By comparing the distortion across pictures, we estimated the gradient in density [11]. For further details on our setup and analysis, supplemental material is available on our GitHub repository [12].



(a) (b)

FIGURE 1. (a) Wave tank apparatus for internal wave generation and measurement. Casting a conductivity probe (1) yields a preexperiment Brunt-Väisälä frequency. Waves are formed by oscillating the topography (2) with a stepper motor. (b) Experimental setup utilizing the "two-bucket" method. Pumps 1 and 2 (1,2) move water through the saltwater and freshwater tanks (3,6) as bilge pumps (4) mix water in the saltwater tank. The diffuser (5) floats on the mixing tank surface as the mixture pumps in.

RESULTS AND DISCUSSION



(a) (b)

FIGURE 2. (a) Magnitude of the density perturbation gradient $|\nabla\rho'|$ (in $\mathrm{kg}\cdot\mathrm{m}^{-4}$) obtained from BOS. The superimposed vector indicates the expected angle $\theta = 48.7^\circ$ from the internal wave dispersion relation. Data were collected 7.5 min after the onset of forcing and smoothed using splines [13]. (b) The 2D power spectrum of $|\nabla\rho'|$ obtained using BOS as a function of wavenumbers (k,m). Dashed lines have a slope corresponding to the ratio of m/k set by Eq. (3) for linear internal waves at a single frequency. A Hann window of size 38.0 cm $\times$ 202.1 cm reduced spectral leakage. The final FFT resolution was 3.11 rad m^{-1} in x and 16.53 rad m^{-1} in z.

Our BOS visualization is shown in Fig. 2(a). Here, we observed that our forcing constructed not only a single sinusoidal wave, but rather a strong localized beam on the diagonals. To investigate this beam, we carried out a 2D FFT (fast Fourier transform) on the density perturbation fields and depicted the most prominent wave vectors in the system [Fig. 2(b)]. We found that there was not just one mode but multiple modes in each quadrant. Analogous to internal tide generation in the ocean (where forcing is at a tidal frequency but unspecified wavenumber), these modes corresponded to individual sinusoids generated by our forcing frequency, with the X pattern arising from sign freedom in both k and m and the constant ratio m/k set by ω_0 for all sinusoids. Thus, numerous sinusoidal modes with different wavelengths and amplitudes, yet with the same frequency, created a stable wavepacket with crests and troughs corresponding to the bright stripes present in 2(a). Because the modes shared the same frequency, the same angle must also be shared; the whole wave group propagated in the same direction with wave packets "stacked" down the diagonals. At the onset of forcing, the beam evolved by propagating throughout the tank, but the structure of the beam envelope remained in place. While wave packets in other areas of physics often disperse and lose their form (e.g., quantum free particles, photons in media), the wave group envelope is conserved with time. This stability is characteristic of linear internal wave regimes, confirmed through verification of the propagation angle predicted by Eq. (3), $\theta = 48.7^\circ$ (shown by white arrows in Fig. 2(a)).

Although linear wave theory predicts that the wave packet's amplitude remains constant, we observed that the pattern was brightest nearest to the forcing, gradually dimming as it propagated away and reflected off the bottom.

This was due to shear instability and dissipation. Similarly, dimming can be observed in the corners of the X in the FFT, where higher modes (higher m and k) faded away.

Thus, by virtue of stratification and buoyancy, we observed how a single forcing mechanism created a wave beam: a wavepacket formed from the superposition of sinusoidal modes at one frequency. Unlike other waves in physics that unravel due to dispersion, we demonstrated how the internal wave beam propagated with a preserved shape and unified direction due to sharing a common frequency and angle via Eq. (3). The uniqueness of these waves is further emphasized when realistic physical conditions (i.e. boundary conditions and shear/viscosity) that real fluids are subjected to are considered.

CONCLUSION

Motivated by their peculiar mathematical and geophysical properties, we simulated internal tides in a laboratory setting with stratification and oscillatory forcing. Using BOS imaging, we demonstrated that the internal waves' dispersive behavior gave rise to a localized wave beam. There were numerous constituent modes at the same frequency that allowed for a stable, in-place superposition to form, as found in our 2D FFT results. This stable wave beam is characteristic of linear internal waves. We verified our operation in this regime by confirming our results agreed with the dispersion relation curves. In addition to stratification and buoyancy, we found that boundary conditions and non-conservative forces produced discretization and decoherence of the wave beam modes, emphasizing the uniqueness of an internal wave system.

More complex physical properties are exhibited in turbulent regimes, an object of further study. We predict that increasing the forcing's length scales or driving frequency could move the experimental system toward a *nonlinear* regime. One could also use BOS to extract where energies lie, as the emergence of nonlinearity disperses energy to new frequencies. These frequencies would form linearly independent wave vectors, leading to decoherence in the initial wave beam. With some effort, our setup could also be extended to include additional oceanographic dynamics such as double diffusion or rotation. The former could be achieved using sugar and salt as stratifying agents, while the latter requires mounting our setup on a rotating platform.

ACKNOWLEDGMENTS

This research is supported by NSF Award No. OCE-2405801, MOMS: A Minimal Ocean Mixing System. We thank Jim Mahaney and the UNC-CH JFL for experiment facilitation, as well as peers who helped review our paper.

REFERENCES

1. A. S. Wyatt, J. J. Leichter, L. T. Toth, T. Miyajima, R. B. Aronson, and T. Nagata, "Heat accumulation on coral reefs mitigated by internal waves," Nat. Geosci. **13**, 28–34 (2020).
2. P. Washam, K. W. Nicholls, A. Münchow, and L. Padman, "Tidal modulation of buoyant flow and basal melt beneath Petermann Gletscher ice shelf, Greenland," J. Geophys. Res. Oceans **125**, e2020JC016427 (2020).
3. A. Gill, *Atmosphere-Ocean Dynamics*, Atmosphere-Ocean Dynamics, Vol. 30 (Elsevier Science, 1982) p. 131.
4. J. Lighthill, *Waves in Fluids*, Vol. 13 (Cambridge University Press, Cambridge, England, 2002) p. 1501.
5. C. B. Whalen, C. De Lavergne, A. C. Naveira Garabato, J. M. Klymak, J. A. MacKinnon, and K. L. Sheen, "Internal wave-driven mixing: Governing processes and consequences for climate," Nat. Rev. Earth Environ. **1**, 606–621 (2020).
6. P. Kundu and I. Cohen, *Fluid Mechanics* (Academic Press, New York, 2010) p. 265.
7. G. Oster and M. Yamamoto, "Density gradient techniques,," Chem. Rev. **63**, 257–268 (1963).
8. L. Gostiaux and T. Dauxois, "Laboratory experiments on the generation of internal tidal beams over steep slopes," Phys. Fluids **19**, 028102 (2007).
9. P. Echeverri, M. Flynn, K. B. Winters, and T. Peacock, "Low-mode internal tide generation by topography: An experimental and numerical investigation," J. Fluid Mech. **636**, 91–108 (2009).
10. J. Vohn, Z. Phan, Z. Taebel, D. Bruney, P.-Y. Passaggia, and A. Scotti, "A versatile laboratory approach to reproduce and analyze internal ocean wave dynamics," arXiv:2603.13512.
11. P.-Y. Passaggia, V. K. Chalamalla, M. W. Hurley, A. Scotti, and E. Santilli, "Estimating pressure and internal-wave flux from laboratory experiments in focusing internal waves," Exp. Fluids **61**, 1–29 (2020).
12. V. Jacquez, Z. Phan, Z. Taebel, D. Bruney, P.-Y. Passaggia, and A. Scotti, "InternalWavesSetup: Laboratory setup and analysis for internal-wave experiments," `https://github.com/vbjacq8/InternalWavesSetup` (2026).
13. D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," Comput. Stat. Data Anal. **54**, 1167 (2010).

Appendices are available at https://pubs.aip.org/aip/jurp.

First Principle Analysis of the Magnetism and Electronic Structure of Fe_2XSi (X=Ti, V)

Hanieh Kachooee,[1, a)] Cindy Kim,[1, b)] Jason Carbajal,[1, c)]
K. Hettiarachchilage,[2, d)] and N. Haldolaarachchige [1, e)]

[1]*Department of Physical Science, Bergen Community College, Paramus, New Jersey 07652, USA*
[2]*Department of Physics and Astronomy, College of Staten Island, 2800 Victory Blvd., Staten Island, New York 10314, USA*

[a)]Corresponding author: hanieh.kachooee@gmail.com
[b)]mkim159526@me.bergen.edu
[c)]jcarbajal152832@me.bergen.edu
[d)]kalani.hettiarachchilage@csi.cuny.edu
[e)]nhaldolaarachchige@bergen.edu

Abstract. The electronic, magnetic, and mechanical properties of Fe_2XSi, where X is titanium (Ti) and vanadium (V), are investigated using computational methods. Volume optimization reveals that the ground state of Fe_2TiSi is a nonmagnetic narrow-band-gap semiconductor, and that of Fe_2VSi is a ferrimagnetic metal. The negative formation energies of both materials confirm the stability of their crystal structures. Mechanical properties confirm the static stability of the crystal structure and suggest that the titanium compound is ductile; however, the vanadium material is brittle. Density of states and band structure studies confirm the nonmagnetic semiconducting nature with an indirect band gap at Γ and X for the titanium material and the ferrimagnetic-metallic nature of the vanadium material. Electronic and magnetic properties of the materials were investigated with applied tension (negative pressure) and compression (positive pressure) to the crystal structure. The pressure study on Fe_2TiSi shows a tunable band gap and a possible semiconductor-metal transition, and Fe_2VSi shows a tunable magnetic moment and a possible low-spin/high-spin magnetic transition.

INTRODUCTION

Heusler compounds are a class of intermetallic materials with diverse electronic and magnetic properties. The common form of Heusler compounds is a ternary intermetallic compound. They have a cubic structure composed of either a half-Heusler (XYZ), a system limited to one magnetic sublattice, or a full-Heusler (X_2YZ), which can have two magnetic sublattices [1]. The X and Y variables represent transition metals, and Z represents a *p*-group element. Heusler compounds compose a large family with large diversity; there are thousands of known variants with different elemental compositions [2]. Due to the variety, the electronic, magnetic, and structural properties of the compound differ, which also allows a diverse range of physical properties to be observed within one of its members [3]. Heusler compounds permit great flexibility of tuning physical properties via electronic band engineering [4]. Their highly tunable crystal and electronic structures enable control of functionalities such as magnetism, semiconductivity, half-metallicity, and superconductivity [2, 3].

In general, Heusler compounds with magnetic elements such as Fe, Co, and Ni tend to be metallic because of high electron density near the Fermi energy and *d*-electron delocalization, but some show interesting half-metallic magnetism, such as Fe_2YSi [5]. However, in certain Heusler systems, small changes in their composition can lead to different electronic and magnetic properties, enabling rare combinations of semiconducting and ferromagnetic behavior [6, 7]. Full-Heusler compounds are particularly well suited for this purpose. Unlike half-Heuslers, which contain a vacant lattice site and often exhibit weaker magnetism, full-Heuslers have all lattice sites fully occupied. This occupancy allows stronger magnetic interactions, enabling ferromagnetic alignment and higher total magnetic moments [1]. For instance, the full-Heusler compound Fe_2CrAl can maintain ferromagnetic ordering of its Fe and Cr atoms while preserving a semiconducting gap [8, 9, 10]. Chemical substitution can tune the balance between conduction and magnetism, enabling access to electronic states that are normally incompatible in conventional semiconductors. By adjusting electron count, band structure, and magnetic interactions, researchers can design materials with tailored electronic and magnetic properties. Because semiconducting ferromagnets are rare, they attract interest for potential applications in spintronics and thermoelectrics [11, 12]. In particular, Fe-based full-Heusler compounds are promising due to their composition of abundant, low-cost elements and high thermal stability. For example, Fe_2TaAl and Fe_2TaGa exhibit semiconducting band gaps while satisfying mechanical stability criteria, indicating both electronic

and mechanical functionality [13]. These findings motivate further investigation into Fe-based Heusler compounds as candidates for multifunctional materials that combine semiconducting, magnetic, and thermoelectric properties within a single, tunable crystal framework.

Despite studies on Fe-based Heuslers, the ternary compounds Fe_2TiSi and Fe_2VSi remain underexplored [14, 15, 16]. This system contains *d*-group elements, which introduce moderate electron correlations and hybridization effects. A Fe_2TiSi pure bulk phase has not yet been reported experimentally, and theoretical studies of the ground state of Fe_2XSi, where X=Ti and V, seem to lack the use of different exchange-correlation functionals [17, 18, 19]. Therefore, here we provide a comparative study of Fe_2TiSi and Fe_2VSi full-Heusler compounds. In this study, we perform a density functional theory (DFT) study of Fe_2XSi where X=Ti and V. Using volume optimization and self-consistent field calculations, we investigate their ground-state properties, electronic band structures, density of states, magnetic behavior, and mechanical properties. Then the effect of pressure on the electronic structure is investigated.

COMPUTATIONAL METHOD

All simulations were performed using the Expanse and Bridges high-performance computer clusters [20, 21]. Quantum ESPRESSO (QE) was utilized to analyze the Fe-(Ti/V)-Si systems, and the Material Project and Materials Cloud platforms were used to extract structural data and potential functions [22, 23, 24]. Volume optimization was done by using the variable-cell relaxation (vc-relax) method with QE. Three spin configurations, nonmagnetic (NM, no spin orientations), ferromagnetic (FM, atomic spins align in parallel), and antiferromagnetic (AFM, nearest-neighbor atomic spins align in antiparallel directions), were taken into account for each compound. The convergence was performed with the plane-wave cutoff of 60 Ry for kinetic energy and 600 Ry for the charge density. Additionally, the self-consistent-field (SCF) convergence threshold was set at 10^{-9} Ry. The lowest total energy of the three spin configurations (NM, FM, AFM) was used to determine in which case the compound was at the ground state. For the magnetic ground state, the Hubbard potential was applied within the generalized gradient approximation (GGA+U). Hubbard potential U values are set to U=2.00 eV for Fe atoms and U=1.50 eV for V atoms. These U values are in good agreement with the literature for transition-metal compounds with moderate electron correlations [25]. If the ground state showed a band gap, then the modified Becke-Johnson (mBJ) method was used to validate the band gap. The $9 \times 9 \times 9$ k-point matrix was used. To get the density of states (DOS), a non-self-consistent field (NSCF) calculation was performed, followed by a DOS calculation. Formation energy was calculated using the ground-state energies of the compounds and constituent elements. Mechanical properties were calculated with the ElaStic software package [26], which cooperated with QE.

RESULTS AND DISCUSSION

Ground State with Relaxed Volume Optimization

Figure 1(a) shows the standard cubic unit-cell structure of full-Heusler compounds Fe_2XSi, which crystallizes into the $L2_1$ structure (space group of 225, $Fm\bar{3}m$). Ti and V atoms (light blue) are positioned at (0, 0, 0), and Si atoms (dark blue) are positioned at (0.5, 0.5, 0.5). The arrangement of X and Si atoms forms a rock-salt-like sublattice. The Fe atoms (bronze) take the tetrahedral voids at (0.25, 0.25, 0.25) and (0.75, 0.75, 0.75). Figure 1(b) shows the Fe atom cubic sublattice inside the full-Heusler cubic unit cell. Figure 1(c) shows the first Brillouin zone in the reciprocal space of the face-centered cubic (FCC) lattice. The red dotted arrows in Fig. 1(c) represent the high-symmetry path $\Gamma \rightarrow X \rightarrow W \rightarrow K \rightarrow \Gamma \rightarrow L$ that is used for the electronic band structure calculations. The selection of a high-symmetry path plays an important role when identifying the nature of the band gap, whether direct or indirect, of a semiconducting compound, and generally novel electronic properties.

Table 1 shows a summary of volume optimization to investigate the ground state of Fe_2XSi. Three spin configurations, NM: nonmagnetic (no spin alignments), FM: ferromagnetic (all spins in parallel), and AFM: antiferromagnetic (nearest spins align in antiparallel), are used to identify the ground state. The table shows the lattice parameter, unit-cell volume, total energy (E_0), and energy difference (i.e., ΔE between the lowest energy state and other states). When total energy is compared between different configurations, it is evident that the ground state of the Fe_2TiSi is nonmagnetic and Fe_2VSi is ferrimagnetic (antiparallel spin alignment between atoms with a net positive magnetic moment). The ground-state lattice constants of Fe_2TiSi and Fe_2VSi are in good agreement with previously reported values in

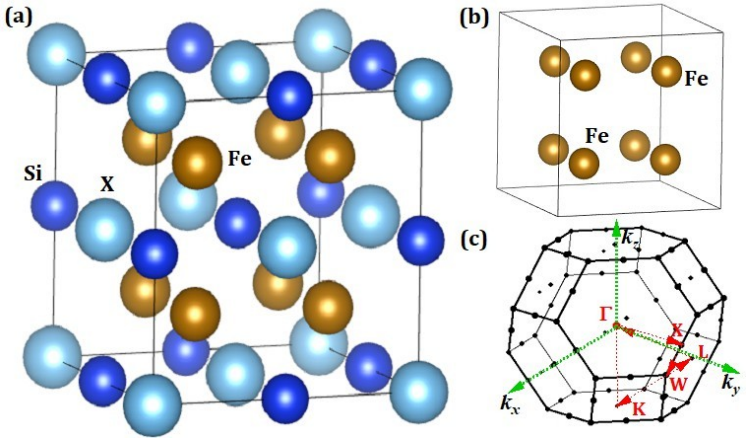


FIGURE 1. (a) Cubic full-Heusler structure of Fe_2XSi (X=Ti, and V) showing Si (dark blue), Ti and V (light blue), and Fe atoms (bronze). (b) Fe atoms in a cubic sublattice. (c) The first Brillouin zone and high-symmetry points. The green arrows represent reciprocal lattice axes.

TABLE 1. Volume optimization results for Fe_2TiSi and VFe_2Si using VC-relax.

Method	a (Å)	V (Å^3)	E_0 (eV)	ΔE (eV)
Fe_2TiSi				
NM	5.688	184.05	−48389.657	0.000
FM	5.691	184.32	−48389.634	0.023
AFM	5.688	184.06	−48389.634	0.023
Fe_2VSi				
NM	5.652	180.54	−49976.480	16.100
FM	5.659	181.28	−49992.580	0.000
AFM	5.887	203.98	−49989.294	3.286

the literature [15, 17].

These ground states can be investigated within the valence electron counting (VEC) based Slater-Pauling (SP) rule that gives the total magnetic moment of a full-Heusler, X_2YZ, where X and Y are d group elements, and Z is a main group element:

$$M_t = Z_t - 24, \tag{1}$$

where M_t is the total magnetic moment per formula unit and Z_t is the total valence electrons per (f.u.). According to the SP rule the Fe_2TiSi compound should have zero total magnetic moment because the total VEC is Z_t=24. This agrees well with our calculations. VEC for the Ti compound is as follows: four electrons from Ti ($3d^2 4s^2$), sixteen from the two Fe atoms ($2 \times 3d^6 4s^2$), and four from Si ($3s^2 3p^2$).

TABLE 2. Magnetic moments of the Fe_2VSi system.

	Magnetic moment (μ_B/f.u.)			
Method	Total	Fe	V	Si
GGA	0.9301	0.5238	-0.0978	-0.0197
GGA+U	0.9450	1.0508	-0.8271	-0.0528

Table 2 shows calculated magnetic moment results of Fe_2VSi with GGA and GGA+U. The SP rule predicts that the vanadium compound should have a magnetic ground state with M_t=1.0 μ_B because the VEC of Fe_2VSi is Z_t=25. The VEC of the vanadium compound is as follows: five electrons from V ($3d^3 4s^2$), sixteen from the two Fe atoms ($2 \times 3d^6 4s^2$), and four from *Si* ($3s^2 3p^2$). Our calculation shows that GGA results in a total magnetic moment of 0.93 μ_B per f.u., which is close to the predicted value of the SP rule and recent literature [17]. However, the GGA method

underestimates electron correlations and therefore results in a lower magnetic moment on correlated electron systems [27, 28]. This is particularly evident with the magnetic moment per Fe atom (0.5238 μ_B per f.u.), which is a very low value for the localized magnetic moment per Fe atom. Our GGA+U calculation results in a slightly higher total magnetic moment (1.05 μ_B per f.u.), which is in great agreement with the SP rule and with reported values in literature [17]. Additionally, the GGA+U method resulted in higher magnetic moments per atom for the Fe_2VSi compound, with a magnetic moment per Fe atom M(Fe/f.u.)=1.0508 μ_B, a magnetic moment per V atom M(V/f.u.)=-0.8271 μ_B, and a magnetic moment per Si atom M(Si/f.u.)=-0.0528 μ_B. This leads to a total magnetic moment M(total/f.u.)=0.9450 μ_B. There are not reported magnetic moments with GGA+U in the literature for this material. The GGA+U results from this study agree well with the experimentally measured total magnetic moment of this compound and other similar full-Heusler materials, such as Co_2RuSi [17, 29]. The higher total magnetic moment with the Hubbard potential (U) confirms the effect of moderate electron correlations. Generally, the GGA method is not able to localize *d*-electrons in V and Fe atoms, which leads to a reduced or quenched total magnetic moment. When the Hubbard potential U is applied (GGA+U), it incorporates localized *d*-electrons in V and Fe atoms, which leads to a higher total magnetic moment. When comparing the sign of the magnetic moment of each atom, the Fe moment is positive and the V and Si moments are negative. This suggests that V and Si spins are antiparallel to the Fe spin. Fe_2VSi has a positive total magnetic moment per formula unit, which means it should be ferrimagnetic (FiM, antiparallel spin between Fe and V). However, the ground state has all Fe spins aligned in the same direction (parallel), which means that Fe atoms have ferromagnetic (FM) spin coupling within the Fe atom sublattice.

Formation Energy

Formation energy is calculated by using the total energy of the compounds Fe_2XSi (X=Ti and V) and the total energies of each element, Ti, V, Fe, and Si. Formation energy is an important parameter to investigate the thermodynamic stability of the crystal structure. Generally, the formation energy of a compound phase must be negative at thermal equilibrium with respect to the elemental phases that constitute the material. Formation energy was calculated using the following equation,

$$\begin{aligned} E_f &= \frac{E_{Fe_2XSi} - N_X E_X - N_{Fe} E_{Fe} - N_{Si} E_{Si}}{N} \\ &= \frac{E_{Fe_2XSi} - E_X - 2E_{Fe} - E_{Si}}{4}, \end{aligned} \tag{2}$$

where E_f is the formation energy, E_{Fe_2XSi} is the total energy of the full Huesler compounds (Fe_2TiSi and Fe_2VSi), and the energy of elemental phases is given by E_X (X= Ti or V), E_{Fe}, and E_{Si}. $N_X = 1$ represents one Ti or V atom per f.u., $N_{Fe} = 2$ represents two Fe atoms per f.u., and $N_{Si} = 1$ represents one Si atom per f.u.

Total energy calculations were done with an optimized cubic unit cell for Fe_2TiSi and Fe_2VSi Heusler compounds with space group 225 ($Fm\bar{3}m$). Total energy calculations of elemental phases were done with the unit cells of vanadium with space group 229 ($Im\bar{3}m$), titanium with space group 194 ($P6_3/mmc$), iron with space group 229 ($Im\bar{3}m$), and silicon with space group 227 ($Fd\bar{3}m$). Formation energy was calculated by using Eq. (2) and resulted in $E_f(Fe_2TiSi) = -1.04$ eV/atom and $E_f(Fe_2VSi) = -1.01$ eV/atom. These negative formation energies for both compounds suggest that these full-Heusler compounds are stable at thermal equilibrium and fall within the acceptable values reported in the literature [30, 31].

Mechanical Properties and Static Stability

Table 3 shows the analysis and comparison of elastic and mechanical properties of full-Heusler Fe_2XSi (X=Ti and V). Both compounds show mechanical stability as they obey the following criteria: ($C_{11} - C_{12} > 0$, $C_{11} - 2C_{12} > 0$, and $C_{44} > 0$), which confirms the stable structures [30, 32]. Our calculation shows that Fe_2TiSi is significantly stiffer when compared with Fe_2VSi because Young's modulus and the shear modulus of Fe_2TiSi are higher than those of Fe_2VSi. The materials' ductility vs. brittleness can be checked with the bulk-to-shear modulus ratio (B/G) according to Pugh's criterion (ductile when B/G $>$ 1.75 and brittle when B/G $<$ 1.75). Our calculation suggests that Fe_2TiSi is a brittle (B/G = 1.539) compound and Fe_2VSi is a ductile (B/G = 3.786) compound.

Development of an Improved Probe Design and Analysis Method for Reliable Thermopower Measurements

Christian RJ Castillo,[a] Navya Sampathi, Joseph White, and Pei-Chun Ho[b]

Department of Physics, California State University, Fresno, 2345 E. San Ramon Ave., Fresno, California 93740, USA

[a]Corresponding author: christiancastillo@mail.fresnostate.edu
[b]pcho@csufresno.edu

Abstract. Thermopower is a fundamental thermodynamic property that provides insight into the behavior of strongly correlated electron materials, such as their scattering time and density of state. Quantified by the Seebeck coefficient, this value represents the ratio of voltage generated over the temperature gradient applied across a sample. In our lab, previous thermopower measurements were hindered by lengthy experiments and complicated analysis. To improve efficiency, we implemented a faster analysis technique and began directly regulating temperature on the cold-side thermometer. Testing with 99.95% platinum showed consistent results between the slope and steady-state methods under a temperature gradient ~3% above the regulated cold-side temperature, as well as agreement with literature above 40 K. Deviations below 40 K were attributed to the differential type-T thermocouple, which has limited resolution for low-temperature measurements. These results confirm the slope method's validity, the use of the thermocouple as a reference sample, and our probe's overall effectiveness above 40 K. To enhance sub-40-K accuracy, we are developing a new probe incorporating a hot-side thermometer and simultaneous measurement of target and reference samples, enabling more precise removal of background contributions.

INTRODUCTION

Accurately measuring the properties of advanced materials is necessary to characterize and understand their behaviors. One key measurement is the Seebeck coefficient, a quantity that describes the thermoelectric ability of a material. Determining this is not only essential for evaluating a material's potential in thermoelectric applications, such as waste heat recovery and thermocouples, but also for gaining insight into its fundamental properties, including scattering time and density of states, due to its relationship with electrical conductivity.[1]

The Seebeck coefficient S is typically obtained by establishing a temperature gradient across a sample and measuring both the resulting temperature difference and the voltage between its two ends. It is defined as the negative ratio of the generated voltage ΔV over the applied temperature difference ΔT, shown in Eq. (1):

$$S = -\frac{\Delta V}{\Delta T}. \tag{1}$$

When experimentally determining the Seebeck coefficient of a sample, it is important to acknowledge challenges such as the contributions of background wiring to the measured voltage.[2] Figure 1 demonstrates a thermopower experiment with wires connecting to the hot and cold ends of a sample leading to a voltmeter. Here, the $+z$ direction is defined from cold to hot along the sample. Analyzing the electric potential ΔV across the sample and the wire, we integrate two definitions of the electric field E and set them equal, as shown in Eq. (2). The first definition of E is the negative change in voltage dV over dz such that $E = -\frac{dV}{dz}$. The second is the foundation of thermopower, relating E to the temperature gradient $\frac{dT}{dz}$ via the Seebeck coefficient S, defined as $E = S\frac{dT}{dz}$. Solving Eq. (2) for the measured Seebeck coefficient $S_{meas,sample}$ shows how wiring affects results, as seen in Eq. (3). Because of this, it is important to remove background contributions in the analysis to determine the true Seebeck coefficient values of a sample.

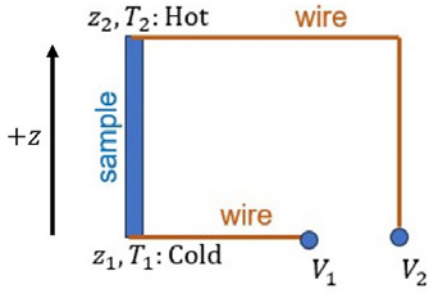


FIGURE 1. Schematic of the basic wiring to read the voltage generated across a sample when doing thermopower experiments. The $+z$ direction is defined from cold to hot along the sample.

$$\int_{z_1}^{z_2} E_{sample,z}\, dz + \int_{z_2}^{z_1} E_{wire,z}\, dz = \int_{z_1}^{z_2} S_{sample} \frac{dT}{dz}\, dz + \int_{z_2}^{z_1} S_{wire} \frac{dT}{dz}\, dz \tag{2}$$

$$S_{meas,\,sample} = -\frac{\Delta V}{\Delta T} = (S_{sample} - S_{wire}) \tag{3}$$

In the lab, we are developing a probe that can reliably measure the Seebeck coefficient of various samples. With our current design, we measured and analyzed the Seebeck coefficient of a 99.95% platinum (Pt) wire sample and compared it to known literature values.[3,4] The analysis was done using a trick that treats our differential type-T thermocouple as a reference sample, taking advantage of the probe's design and analyzing it as a constantan sample to remove the background contributions in the following method, shown in Eq. (4):[5]

$$S_{Pt} = S_{Pt,\,measured} - S_{Constantan,measured} + S_{Constantan}. \tag{4}$$

EXPERIMENTAL METHOD

Physical Instrumentation

The thermopower probe uses a heater mounted on a hot platform to establish controlled temperature gradients across a Pt wire sample suspended between two platforms, as seen in Fig 2. A Cernox thermometer measures temperature on the cold platform and a differential type-T thermocouple (a constantan wire) spans the platforms to read the temperature difference across them. Because the constantan wire experiences the same temperature gradient as the Pt sample, it can serve dual purposes as a reference to subtract background contributions.

(a) (b)

FIGURE 2. (a) Top view of a thermopower probe with the hot platform on the left and the cold platform on the right. (b) Close-up view of a thermopower probe focused on the 99.95% Pt wire sample suspended across the platforms.

At a set regulated cold temperature, a temperature gradient within 1%–3% is established across the platforms using the heater, which is eventually turned off to allow cooling back to equilibrium. This is repeated to scan the range of our cryocooler from 300 to 8 K. We analyze data using the slope method, which requires a much shorter experimentation time than the traditional steady-state method because we do not need to reach thermal equilibrium during the heating period. Previous studies in the lab have shown that the two methods yield nearly identical values.[6]

Analysis Method

To determine the Seebeck coefficient using the slope method, the rise and decay of temperature and voltage are directly compared. As shown in Fig. 3(a), the cold-platform temperature is kept constant whilst the hot-platform temperature increases, creating a temperature gradient across the sample. This generates a voltage, shown in Fig. 3(b), which we measure for the Pt sample and then subtract that of the constantan wire. The voltage's clear response to temperature indicates very good thermal contact between the sample and the thermometer.

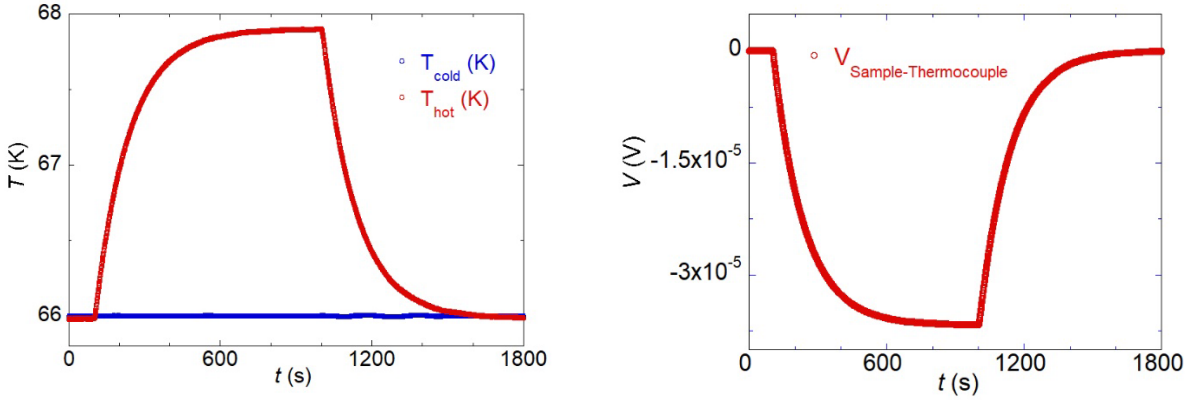


FIGURE 3. (a) Temperature versus time and (b) voltage versus time for a mean temperature $T = 66.9$ K.

To determine the exact relationship between the changing voltage and temperature gradient, we graph voltage against temperature and perform a linear fit, demonstrated in Figs. 4(a) and 4(b). The resulting slope represents the Seebeck coefficient. By averaging the values from both the heating and cooling periods, we determine the relative Seebeck coefficient of the sample with respect to the constantan wire. This is determined for the mean temperature, defined as $T_{cold} + \frac{\Delta T}{2}$, where ΔT is the largest measured temperature difference between the platforms. We then add Seebeck coefficient values of constantan from literature to get the final values of the Pt sample, as in Eq. (4). It is important to note that because this analysis method is highly dependent on a constant cold-side temperature, as shown in Fig 3(a), temperature fluctuations on the cold side can lead to significant errors in determining Seebeck values.

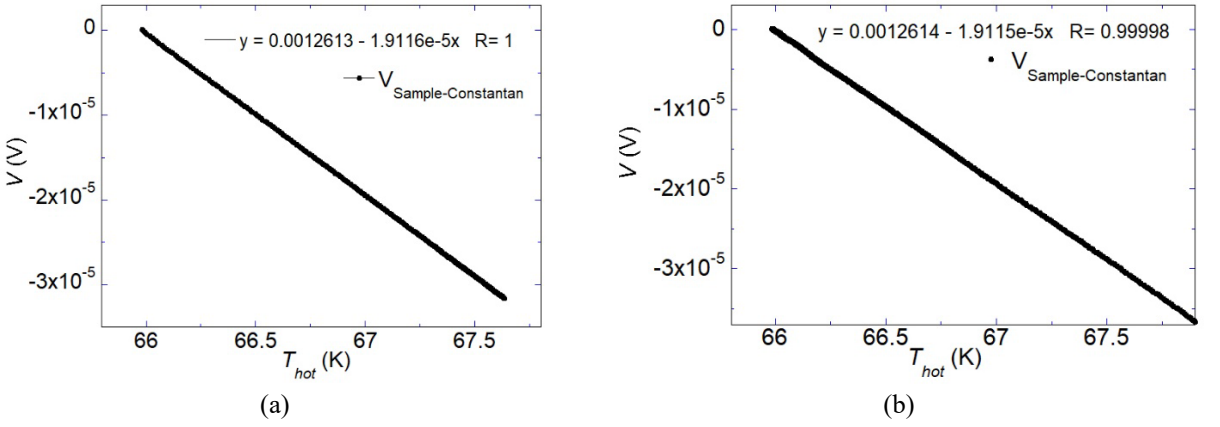


FIGURE 4. (a) Voltage versus hot-side temperature of the heating period and (b) voltage versus hot-side temperature of the cooling period for a mean temperature $T = 66.9$ K.

RESULTS AND FUTURE WORK

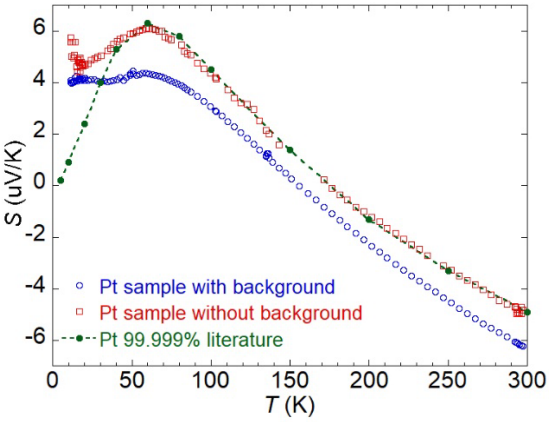


FIGURE 5. Seebeck coefficient versus temperature for a 99.999% platinum sample in the literature (green) and analyses of our 99.95% platinum sample without background removal (red) and with background removal (blue).

As shown in Fig. 5, our corrected measurements are in good agreement with literature values above 40 K, with an average absolute error of about 5%. This validates our use of both the slope method and the differential type-T thermocouple as a reference for removing background contributions in the 40–300-K range. Below 40 K, however, deviations arise from the limited resolution of the thermocouple. At very low temperatures, the Seebeck coefficient of constantan approaches zero. This creates an increasingly smaller voltage across the thermocouple we use to determine the temperature difference across the sample. As such, our resolution drops dramatically, prompting the need for a new probe design to improve low-temperature measurements.

Multiple attributes aimed at improving overall accuracy are still to be incorporated into the new design. There will be Cernox thermometers on both the hot and cold platforms, alongside the type-T thermocouple, to eliminate the reliance on the thermocouple at low temperatures. In addition, the new design will feature two sample positions to allow simultaneous measurement of a target and reference sample for more accurate background removal. It will also have a movable Delrin hot platform to accommodate various sized samples, topped with a sapphire window to ensure high thermal conductivity at low temperatures. Delrin is specifically chosen for the hot platform, as opposed to the copper cold platform, to create thermal insulation between the sample's hot side and copper puck.[7]

ACKNOWLEDGMENTS

This undergraduate research at California State University, Fresno, was supported by NSF DMR-1905636, Fresno State Smittcamp Family Honors College, and the Frank Sutton Research Fund.

REFERENCES

1. G. Prunet, F. Pawula, G. Fleury, E. Cloutet, A. J. Robinson, G. Hadziioannou, and A. Pakdel, Mater. Today Phys. **18**, 100402 (2021).
2. J. Martin, T. Tritt, and C. Uher, Jour. Appl. Phys. **108**, 121101 (2010).
3. R. E. Bentley, *Theory and Practice of Thermoelectric Thermometry* (Springer, Singapore, 1998), pp. 72, 214.
4. R. L. Powell, W. J. Hall, C. H. Hyink Jr., and L. L. Sparks, *Thermopower Reference Tables Based on the IPTS-68* (NBS Monogr. 125, Natl. Bur. Stand., USA, 1974), pp. 38–42.
5. A. L. Pope, R. T. Littleton IV, and T. Tritt, Rev. Sci. Instrum. **72**, 3129–3131 (2001).
6. A. Capa Salinas, J. Velasquez, and P.-C. Ho, Proceedings of the National Conference of Undergraduate Research (NCUR) 2019, pp. 171–177.
7. N. Sampathi, "Advancing Thermopower Measurement: Probe Design and Seebeck Coefficient Analysis," Master's thesis, California State University, Fresno, 2024.

How to Address Nonlinearity and Estimate Uncertainty with Polynomial Fitting in Hubbard U Calculations for DFT+U

Alan Antony,[1, a)] David D. O'Regan,[1, b)] and Lórien MacEnulty[1, 2, c)]

[1]*School of Physics, CRANN Institute and AMBER Research Centre, Trinity College Dublin, The University of Dublin, Ireland*
[2]*Modeling Division, Institute of Condensed Matter and Nanosciences, Université Catholique de Louvain, Chemin des Étoiles 8, Louvain-la-Neuve 1348, Belgium*

[a)]Corresponding author: antonyal@tcd.ie
[b)]david.o.regan@tcd.ie
[c)]lorien.macenulty@uclouvain.be

Abstract. Finite-difference linear response calculations for the Hubbard U and Hund's J parameters in DFT+U—a corrective method aiming to improve the accuracy of DFT calculations of strongly correlated electron systems like Mott insulators—typically involve linear regression of small response datasets. Many subspaces of material systems, however, are found to exhibit nonlinear response (e.g., curvature) that may disrupt the extraction of the required linear component. For these systems, low-order polynomial regressions can capture curvature in linear response datasets, permitting the quantification of uncertainty on the Hubbard parameters through the unbiased root-mean-square error. What remains unclear in this context is the optimal strategy practitioners can employ to avoid the overfitting that occurs at higher polynomial degrees. In this work, we determine uncertainties on the Hubbard parameters and investigate polynomial degree selection strategies based on leave-one-out cross validation (LOOCV) and repeated K-fold cross validation (RKFCV). We show that LOOCV is a practical technique, not least due to its simple closed-form formula, given the small perturbation-occupancy datasets typically used in DFT+U. For larger datasets, LOOCV exhibits high variance in its estimated errors, whereas RKFCV provides more stable estimates. In practical calculations, we have found that fitting to a quadratic or cubic polynomial before taking the slope at zero perturbation of the resulting fit satisfactorily addresses the issue of nonlinear response.

INTRODUCTION

Density functional theory (DFT) is a computational framework for determining the properties of materials. In DFT the total energy is reformulated as a functional of the electron density, which includes an approximated exchange-correlation (XC) term. Common XC functionals are the local density approximation (LDA) and generalized gradient approximation (GGA), and although more accurate functionals exist, they have higher computational cost. DFT using such (semi-)local functionals tends to incorrectly predict the electronic structure of strongly correlated materials, where strong electron interactions restrict electronic motion [1]. The DFT+U(+J) method introduces subspace-tailored corrective terms, the strengths of which are set by the Hubbard U and Hund's J parameters. DFT+U was originally designed to better capture correlation effects such as on-site Coulomb interactions and intra-atomic exchange coupling. U and J are often determined semiempirically, although this precludes the application of DFT+U+J to materials that have not been discovered or synthesized. This semiempiricism is limiting within high-throughput discovery-oriented workflows, where structures are randomly generated and screened for certain properties.

Alternatively, U and J can be determined ab initio through various techniques, among them minimum-tracking linear response [2] (see Appendix A), cRPA [3], and the focus of this work: the self-consistent field (SCF) linear response method [4], wherein U is determined for a chosen subspace by evaluating its response to a range of small subspace potential perturbations (α). In this formalism, suppressing matrix inversion for simplicity, U is defined as

$$\mathrm{U} = \chi_0^{-1} - \chi^{-1} = \underbrace{\left(\frac{d(n_0^{\uparrow} + n_0^{\downarrow})}{d\alpha}\bigg|_{\alpha=0}\right)^{-1}}_{\chi_0} - \underbrace{\left(\frac{d(n^{\uparrow} + n^{\downarrow})}{d\alpha}\bigg|_{\alpha=0}\right)^{-1}}_{\chi}, \tag{1}$$

where χ_0 (χ) is the response of the unscreened (screened) total occupancy n_0^{σ} (n^{σ}) on spin channel σ due to a minor subspace potential perturbation α in the absence (presence) of electronic screening. J is determined analogously [5, 6, 7], although we relegate all details of this analogy to Appendix A to focus the ensuing discussion on the Hubbard U. In practice, χ_0 and χ are extracted as the slopes of ordinary least-squares (OLS) linear regression, typically conducted on 5-9 perturbation-total-occupancy data points. However, many system-functional-subspace combinations exhibit non-negligible nonlinear response [7, 8, 9]. Nonlinearity in the response often occurs near to empty or full subspace filling, and hence it appears to be a drawback of overly optimized subspace choices. It also arises generally for larger perturbation strengths, which are sometimes necessary to improve the response's signal-to-noise ratio.

How to address nonlinearity in this context is of contemporary practical interest. Here, we develop a straightforward nonlinear response postprocessing method, designed expressly for uncomplicated integration into linear response (LR) practitioners' existing Hubbard U workflows. Leveraging the simplicity of low-order polynomial regressions, we apply well-attested statistical techniques to real perturbation-occupancy datasets, providing suggestions on how to regress response and avoid under- or overfitting. Our inquiries fit into a broader discussion on obtaining uncertainty estimates for the Hubbard parameters, which quantify the noisiness of the underlying response even when it is linear.

METHODOLOGY

Consider a screened or unscreened LR dataset comprising $5 \leq k \leq 9$ (α_i, $N_i = n_i^{\uparrow} + n_i^{\downarrow}$) pairs, indexed from $i = 1$ to k. Assuming no uncertainty in α_i, one may use OLS to fit a polynomial function $f^{(p)}(\alpha) = \sum_{q=0}^{p} c_q \alpha^q$ of degree p and coefficients $\{c_q\}$ to the data, about which the data points are normally distributed with zero mean and constant variance. The behavior of this regression may be characterized by the unbiased mean squared error (MSE) of the curve—the sum of squared residuals, differences between the data point and the curve, divided by degrees of freedom of the curve ($k-(p+1)$), the square root of which yields the unbiased root-mean-square (RMS) error. The uncertainty on the Hubbard U and its constituent response functions are then obtained as [10]

$$\sigma_{\mathrm{U}} = \sqrt{\left(\frac{\sigma_\chi}{\chi^2}\right)^2 + \left(\frac{\sigma_{\chi_0}}{\chi_0^2}\right)^2}, \qquad \sigma_\chi = \sqrt{\underbrace{\frac{\sum_{i=1}^{k}[N_i - f^{(p)}(\alpha_i)]^2}{k-p-1}}_{\text{unbiased MSE}} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}_{22}}, \qquad \boldsymbol{X} = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^p \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^p \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \alpha_k & \alpha_k^2 & \dots & \alpha_k^p \end{pmatrix}, \tag{2}$$

where $\boldsymbol{X}$ is the design matrix, and the $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}_{22}$ term encodes the effect of the distribution of α_i on c_1 (i.e., the first-order coefficient of the polynomial function), and so on the value of χ or χ_0. The $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}_{22}$ term, which stems from the formulation of the variance-covariance matrix for OLS, in the simplest case provides the physical units and spacing of the perturbations α.

While Eq. (2) effectively quantifies the uncertainty on the Hubbard U, it does not permit the identification of the most appropriate polynomial degree to fit the LR data. OLS minimizes the sum of the squared residuals, but as p increases, the function gains additional degrees of freedom, allowing it to bend closer to data points and necessarily shrinking the (biased) RMS error. For noisy datasets, then, increasing p causes the curve to better fit the noise, minimizing the RMS error but overfitting the curve [11]. The unbiased RMS error, which features in the definition of σ_χ, introduces a penalty for higher p to address this, although it is unclear whether minimizing the unbiased RMS or σ_χ corresponds to the optimal p. Given the typically small datasets aggregated for the first-principles determination of U, we test the efficacy of leave-one-out cross validation (LOOCV), a diagnostic for overfitting that quantifies a regression's ability to predict unseen data. This property may be minimized to identify the optimal p regressing a particular dataset.

In the above dataset of k (α_i, N_i) points, LOOCV removes a point and uses the remaining $k-1$ points to compute the parameters of a polynomial regression of degree p. The difference between the omitted point and the point as predicted by the curve is squared, then recorded as the validation error of the curve. This process is repeated iteratively, changing the omitted point until all k points have been used for validation. The mean of the validation errors is then calculated, generating a validation error quantifier for the curve. By repeating this process for different p, one may map the relation between validation error and polynomial degree; the smallest validation error is identified as the optimal curve. Fortunately, an analytical formula encapsulates this process entirely and is given by

$$\sigma_{\mathrm{LOOCV}} = \frac{1}{k}\sum_{i=1}^{\mathrm{k}} \frac{(N_i - f^{(p)}(\alpha_{\mathrm{i}}))^2}{(1-\boldsymbol{H}_{ii})^2}, \tag{3}$$

where $\boldsymbol{H}_{ii}$ is the ith diagonal element of the hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ [11]. $\boldsymbol{H}_{ii}$ weights the residual of the ith data point by its contribution to the regression parameters. Overfitting due to high polynomial degree models can lead to instability in the diagonals of the hat matrix, resulting in deviations in σ_{LOOCV}.

If more LR datapoints are available, practitioners can consider using the typically more stable generalization of LOOCV called repeated K-fold cross validation (RKFCV). RKFCV randomly partitions the dataset into K equally sized folds. The p-degree polynomial regressions are trained on $K-1$ folds and validation tested on the final fold, which produces a validation error defined as the sum of the square of residuals across all data points in the fold. This

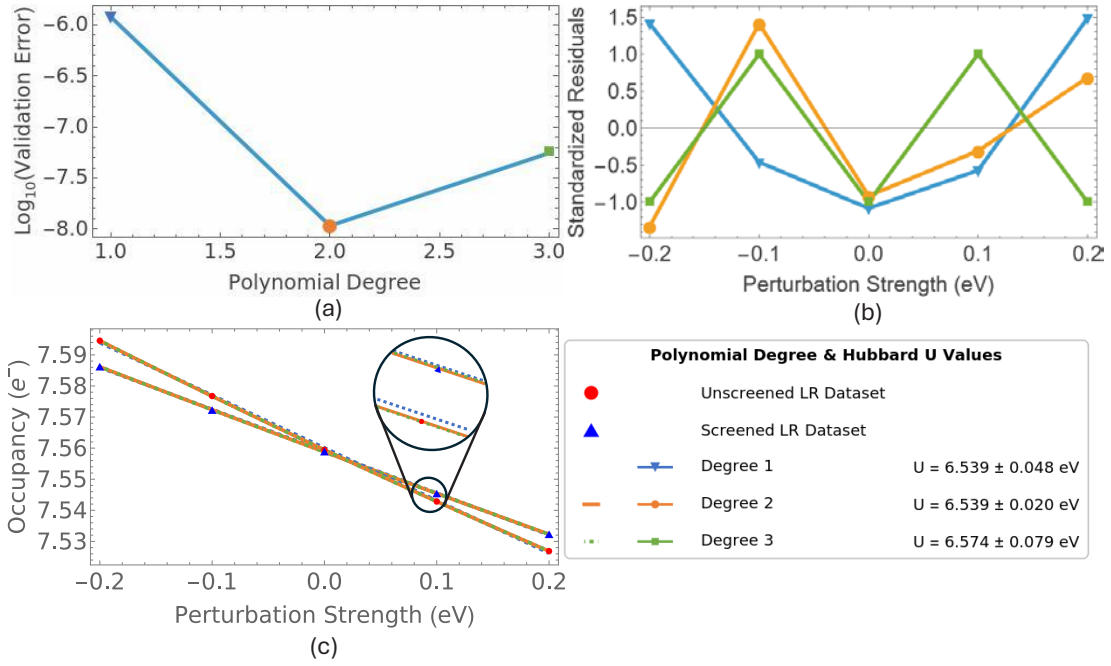


FIGURE 1: For the five-point LR dataset, perturbations conducted on Ni 3*d* sites in AF_I NiO: (a) LOOCV validation error (quantity of Eq. (3) shown on a log scale) of the unscreened data and (see bottom right panel) associated Hubbard U parameters as a function of polynomial degree p using the LOOCV procedure; (b) standardized residuals as a function of perturbation strength α for $p = 1,\ 2$, and 3 [13]; and (c) unscreened and screened occupancies versus perturbation overlaid with regressed curves for $p = 1,\ 2$, and 3.

process is repeated iteratively, changing out the validation fold each time. The mean of the validation errors across all folds is determined. The dataset is then randomly partitioned again into K folds, and the process repeats. The mean of the validation error across all repeats is calculated and mapped to and optimized with respect to degree p.

To test the effectiveness of the LOOCV method in an LR context, we analyze regressions up to degree $k-2$—the maximum polynomial degree possible, as each training fold contains at most $k-1$ points—for two ABINIT-generated [12] perturbation-total-occupancy LR datasets. The first is a typically sized LR set of five data points, and the second is much larger, with 39 data points. We repurpose DFT calculations performed on Ni 3*d* subspaces in antiferromagnetic (AF) NiO, two of which were initially published in [7]. In addition to the calculation details in Appendix B, we refer the reader to this article for more precise computational details. The two cross-validation methods are then assessed using a residual analysis based on the principle that residuals should be randomly distributed about the regression curve, not imbued with polynomial symmetry coming from the regression. To this end, we first investigate the randomness of the residuals produced by the optimal degree as predicted by LOOCV [13]. For the large perturbation-occupancy response dataset, the LOOCV-predicted optimal degree is then compared to the optimal degrees predicted by RKFCV with different validation set (fold) sizes to assess the variability of the LOOCV technique. In the RKFCV applications, if the dataset was not evenly divisible, some points were left out. Details of the linear response datasets used in this investigation can be seen in Appendix B.

RESULTS

As shown in Fig. 1(a), LOOCV demonstrated minimized validation error at $p = 2$ for both the unscreened and screened values of the small dataset. Indeed, the standardized residuals in Fig. 1(b) demonstrate noticeable randomness, in particular, asymmetry across the $\alpha = 0$ axis for $p = 2$ only, corroborating the LOOCV prediction. Standardized residuals are plotted to facilitate comparison between the residuals of the different curves, as they were often of different magnitude, and to identify potential outliers in the data. Please consult Sec. 2.2 of [13] for more details and a description of the standardization process. The associated Hubbard U—where $p = 2$ was used to regress both the

unscreened and screened response data—was identical to that at $p = 1$; however, the $p = 2$ uncertainty is much lower, even minimized compared to the other degrees tested. It is unclear whether there is a correlation between minimized σ_U and minimized validation error, particularly, as based on our method, minimized σ_U does not necessarily mean that σ_χ and σ_{χ_0} were simultaneously minimized via $p = 2$ regression. For example, for a nearly identical dataset to the above, differing in its magnetic ordering, lattice parameter, and k-point sampling (AF_{II} NiO instead of AF_I), LOOCV identified $p = 1$ for the screened dataset but $p_{max} = 3$ for the unscreened one. Systematic testing across a broad range of LR datasets and reevaluation of assumptions underlying the behavior of χ_0 and χ are necessary to solidify the efficacy of taking σ_U as an indicator of optimal p. Although the value of U is shown in Fig. 1 to vary only subtly with p, an example of large nonlinear response and hence p dependence is shown in Fig. (2) of [7].

Concurrence of the LOOCV validation error minimum at the maximum testable polynomial degree is an indication that the dataset is perhaps too small, as higher-order polynomial degrees would need to be tested to verify it as a minimum. This, in addition to the standard statistical benefits of additional points, leads us to recommend performing 6+ perturbations for rigorous LR demonstrating nonlinearity. Not too many data points should be used, however. For the 39-point dataset, LOOCV predicted a minimum validation error at degree 14 [Fig. 2(a)]. This behavior is attributable to the high variance of LOOCV. Each model, for a given degree, is trained on $k - 1$ data points, making the models highly correlated, as $k - 2$ datapoints are shared between any model. This results in high correlation between the predicted validation errors and hence a high variance in the final averaged validation error quantifier. While LOOCV is a nearly unbiased estimator of the generalization error (i.e., with an infinite dataset, the LOOCV validation error would closely align with the generalization error of the curve), this property has a tradeoff with high variance in its estimates. It can yield unstable estimates of the optimal degree, particularly when testing high-order polynomial degrees. RKFCV can mitigate this issue, trading variance for bias by increasing the validation set size [11]. As a result, in Fig. 2(a) we observe that the larger validation set sizes afforded by RKFCV provided more stable validation error curves, which predicted an optimal degree of 5, even as we changed the number of folds K. As shown in Fig. 2(b), convergence to a reliable optimal degree required validation set sizes with around 15%-20% of the total dataset. While the residuals [Fig. 2(c)] clearly demonstrate polynomial symmetry for $p = 2$, $p = 5$ exhibits more noticeable randomness, albeit to a degree that is not a priori distinguishable from the randomness of $p = 14$ residuals.

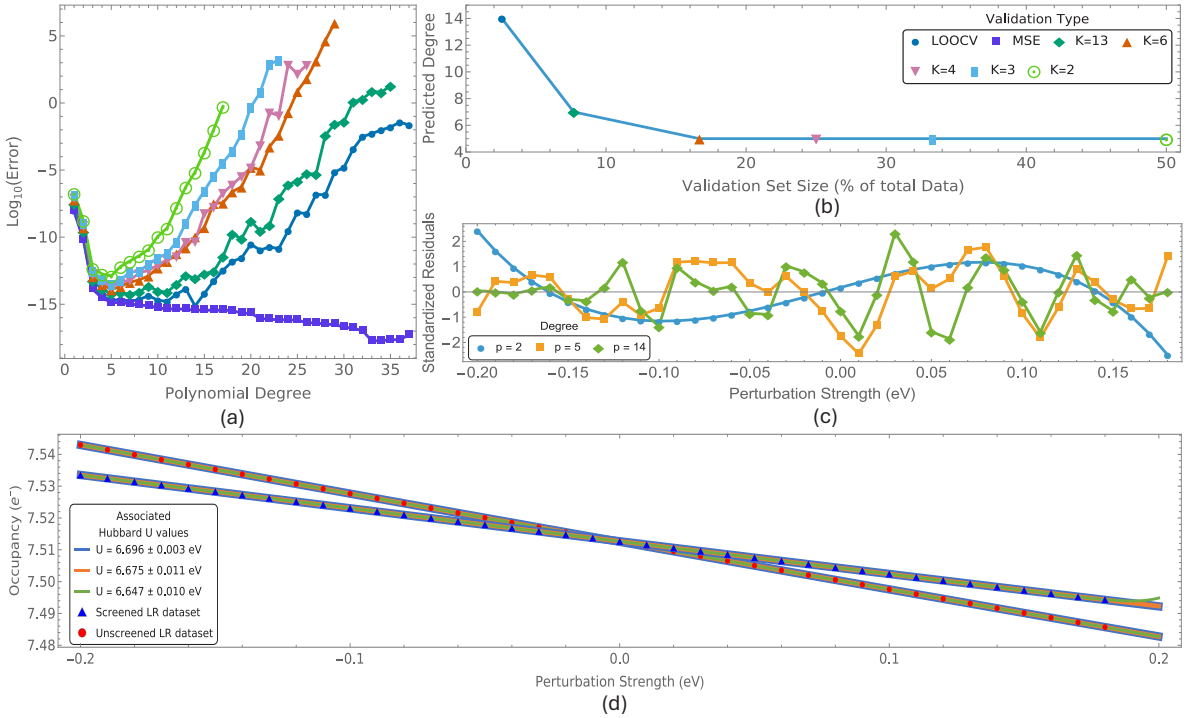


FIGURE 2: For the 39-point, unscreened LR dataset: (a) LOOCV error using the LOOCV procedure, RKFCV error using the RKFCV procedure using K validation sets, and MSE [shown in Eq. (2)] as a function of polynomial degree; (b) RKFCV-predicted optimal degree versus validation set size; (c) standardized residuals of fitted polynomials [13], which are shown in (d) for unscreened and screened response data with $p = 2$, $p = 5$, and $p = 14$.

JURPA only publishes the first four pages of papers in print. Continue reading at https://pubs.aip.org/aip/jurp.

Preparing a Manuscript for Publication

Brad Conrad,[1] Rex Adelberger,[2] and Will Slaton[3, a)]

[1]*National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899, USA*
[2]*Department of Physics, Guilford College, Greensboro, North Carolina 27410, USA*
[3]*Department of Physics and Astronomy, University of Central Arkansas, Conway, Arkansas 72035, USA*

[a)] Corresponding author: wvslaton@uca.edu

Abstract. Scientific manuscripts focus on providing a scientific argument to a specific group. In fact, audience selection is potentially the most important decision a science communicator needs to make before preparing a manuscript for publication. This document will outline a process to draft a manuscript for the *Journal of Undergraduate Research in Physics and Astronomy* (JURPA) but can also be used for most publications. In this specific case, junior and senior physics majors are your primary audience. They are knowledgeable about physics, but unlike you, they have not spent much time trying to understand the specific problem discussed in your report.

INTRODUCTION

There is a big difference between the comments you write in the margin of your lab notebook and what you might write in a paper for a scientific journal. Your laboratory data book is a chronological, definitive record of everything you did. It contains all the data and what you did, even if it was ultimately wrong, as well as comments reflecting your thoughts at that time. A journal article is a focused summary and discussion of the research question, your processes, and your conclusions. Authors should avoid discussing experimental dead ends and, instead, present a clear scientific argument. The reader does not have to be able to reproduce the work from the journal article. Instead, the reader should be able to understand the physics, techniques, and rationale behind why you did what you did.

Make sure that you present the material in a concise and logical way. Overly complicated or long sentences make the logic of an argument difficult to follow. Choose a paragraph structure that focuses the attention of the reader on the development of the ideas. Paragraphs should connect to each other, as the manuscript is a focused, logical argument.

SECTIONS OF THE PAPER

Abstract

An abstract is a self-contained paragraph that concisely explains what you did and presents key results. The abstract is often published separately from the body of the paper, so you cannot assume the reader of the abstract also has a copy of the rest of the paper. You cannot refer to figures or data that are presented in the body of the paper. Since abstracts are used in literature searches, all key words that describe the paper should be included in the abstract. Be quantitative with results, and keep it to less than 100 words.

Introduction

This section outlines the background necessary to introduce your results; it is not an abbreviated review of what you are going to discuss in detail later. This section should present the necessary theoretical and experimental context such that a colleague who is not an expert in the field can understand the data presentation and discussion. If you are going to use a particular theoretical model to extract some information from your data, this model should be discussed in the Introduction.

Where appropriate, factual information should be referenced. If you know where there is a good discussion of some item, you do not need to repeat it. When presenting background information, you may guide the reader to a detailed description of a particular item with a statement such as, “A more detailed discussion of laminar flow can be found elsewhere [1].”

How one proceeds from this point depends upon whether the paper is about a theoretical study or an experiment. We first suggest a format for experimental papers and then one that describes a theoretical derivation.

Experimental Investigations

The Experiment

This section guides the reader through the techniques and apparatus used to generate the data. Schematic diagrams of equipment and circuits are often easier to understand than prose descriptions. A statement such as, "A diagram of the circuit used to measure the stopping potential is shown in Fig. 6," is better than a long description. It is not necessary to describe what is shown in a diagram unless the average reader would not be able to follow the diagram.

If special experimental techniques were developed as part of this work, they should be discussed here. You should separate the discussion of the equipment used to measure something from your results. This section should not include data presentations or discussions of error analysis.

Data and Results

The data are the truths of your work. This section should lead the reader through the data and how errors were measured or assigned. The numerical data values should be presented in tables and figures, each with its own number and caption, e.g., "The results of the conductivity measurements are shown in Table 3." It is difficult to follow narratives when the numerical results are included as part of the narrative. Raw, unanalyzed data should not be presented in the paper. All figures and tables should be referred to by their number and discussed in the narrative.

Theoretical Studies

The Model

This part of your paper should consist of a theoretical development of the constructs used to model the physical system under investigation. Equations should be on separate lines and numbered consecutively. The letters or symbols used in the equations should be identified in the narrative, e.g., "The potential W can be approximated as

$$W \approx Z - \sigma(\rho), \tag{1}$$

where Z is the number of protons, and σ is the screening constant that is dependent on the charge density ρ of the inner electrons of the K and L shells." If you wish to use this equation at a later time in the narrative, refer to it by its number, e.g., "The straight line fit shown in Fig. 3 indicates that Eq. (1) can be used to extract a value of...."

Calculations

This section presents a summary and discussion of the numerical results calculated from the model. The results should be presented in tables or graphs, each with a caption. Data that are not interpreted by the writer have no place in a paper. Reference numerical results that are used in the calculations and come from previous work done by others.

CONCLUSION

In this section, briefly summarize the key result and supporting argument. Be sure to list important quantities and, if appropriate, where this research could lead in the future.

ACKNOWLEDGMENTS

This short section should acknowledge help received from others. This is where you give credit to research advisors, colleagues, and the person in the machine shop who helped you build a piece of equipment. You may also include your funding source, if appropriate.

REFERENCES

In JURPA manuscripts, reference materials should be cited in text using brackets [1] or superscripts. The references, numbered in order of appearance, are collected together at the end of the paper. JURPA references should follow a modified version of the Vancouver system. References should be numbered using Arabic numerals followed by a period, as shown below, and should follow the format in the examples.

Note that you do not need to include article titles in citations for JURPA papers, and journal titles should be given in their ISO 4 standard abbreviations. Refer to authors by their first initial and last name as shown. Include all authors unless there are more than ten. In that case, list the first ten followed by "et al." (If you use the Overleaf template for your references, include all authors in the source code and defer to the default number listed.)

1. L. C. McDermott, Am. J. Phys. **58**, 734–742 (1990).
2. P. M. Sadler and R. H. Tai, Sci. Educ. **85**, 111–136 (2001).
3. C. D. Hoyle, D. J. Kapner, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, U. Schmidt, and H. E. Swanson, Phys. Rev. D **70(4)**, 042004 (2004).
4. M. P. Brown and K. Austin, Appl. Phys. Lett. **85**, 2503–2504 (2004).

Book titles should be in italics, and both authors and editors should be acknowledged as shown. Do not forget to include the location where the book was published.

5. P. Brown and K. Austin, *The New Physique* (Publisher Name, City, State, 2005), pp. 25–30.
6. J. B. Marion, *Classical Dynamics of Particles and Systems*, 2nd ed. (Academic Press, New York, 1970).
7. J. M. Martín-Martínez, in *Adhesion Science and Engineering,* Vol. 2, edited by D. A. Dillard and A. V. Pocius (Elsevier, New York, 2002), p. 573.
8. J. Bishop, in *The Gale Encyclopedia of Science*, Vol. 2, 6th ed., edited by K. H. Nemeh and J. L. Longe (Gale, Farmington Hills, MI, 2021), pp. 1182–1184.

Additional examples are shown for reports (10–12), websites (13–15), and other sources (16–18).

9. E. R. Banilower, "2018 NSSME+: Status of High School Physics" (Horizon Research, Inc., Chapel Hill, NC, 2019).
10. M. Plisch, R. M. Goertzen, and R. Scherr, "Sustaining Programs in Physics Teacher Education: A Study of PhysTEC Supported Sites," in *Noyce Conference* (Washington, DC, 2014).
11. Texas A&M University Education Research Center, "Evaluation of 2017 Mitchell Institute Physics Enhancement Program (MIPEP) Summer Institute" (2017).
12. SpecialChem SA, *Glass transition temperature* (Omnexus, 2022). Available at https://omnexus.specialchem.com/polymer-properties/properties/glass-transition-temperature.
13. CosmicWatch, *Catch Yourself a Muon* (2023). Available at www.cosmicwatch.lns.mit.edu/about.
14. A video demonstration of the prototype is available at https://youtu.be/GcB1Kzv6Cxo.
15. C. D. Smith and E. F. Jones, "Load-cycling in cubic press," in *Shock Compression of Condensed Matter*, AIP Conference Proceedings 620, edited by M. D. Furnish et al. (AIP Publishing, Melville, NY, 2002), pp. 651–654.
16. D. L. Davids, "Recovery effects in binary aluminum alloys," PhD thesis, Harvard University, 1998.
17. A. Cox, developer, Tracking the Coriolis Force [computer software] (Open Source Physics 2012). Available at www.compadre.org/Repository/document/ServeFile.cfm?ID=12042&DocID=2935.

FORMATTING AND STYLE

Tables and Figures

Readers often scan papers by looking at the figures and data tables before they read the narrative of the work. Each table or figure should be numbered and have a descriptive caption. Take care to put enough information in the caption for the reader to get a feel for the meaning of the data presentation. In some journals, tables and figures are placed by the layout editors at the corners of the page to make the format attractive and easy to read, so a figure may not even be on the same page as the discussion of that figure. All lines shown on graphs should be identified, e.g., "The dashed line is drawn to guide the eye," or "The solid line is a fit to the data using the Ising model."

An example graph is shown in Fig. 1. The graph is sized by the range of data points. A graph with all the data points clustered in one small corner does not help the reader get a feel for the dependence of your data. Error bars must be shown with data points.

Figures should be centered. To help stay within the space requirement, consider having two figures next to each other. If figures have more than one part, each part should be labeled (a), (b), etc. Be careful that the figures you present are not too busy; too much information makes it difficult to pick out the important parts. Remember that figures often appear much smaller in print, so make sure labels are about the same font size as the narrative. Consider how best to convey the information clearly. Filled vs empty symbols, or solid vs dashed lines, offer high contrast. Figures should have high resolution or they may appear blurry.

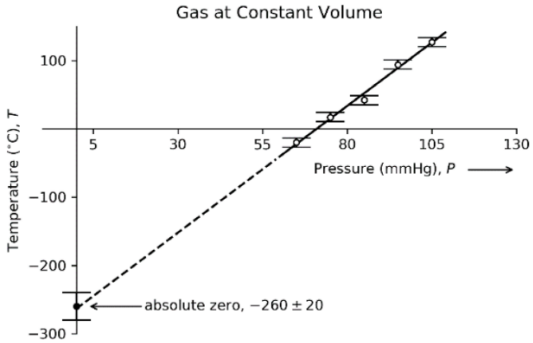


FIGURE 1. A graph of gas temperature versus pressure for an ideal gas at constant volume. The solid line drawn is the least-squares fit straight line to the data. The dashed line extrapolates to the intercept, with uncertainty, denoting an estimate of absolute zero. This figure is adapted from John Taylor's *An Introduction to Error Analysis*, 2nd edition.

Numbers and Units

Include uncertainties in any experimentally measured data presented in tables, such as in Table 1. Use scientific notation when presenting numbers, $(7.34 \pm 0.03) \times 10^7$ eV. Take care that you have the correct number of significant digits in your results; just because the device shows six digits does not mean that they are significant. Use the MKS system of units and the formatting below.

TABLE 1. Energy states found in the numerical search. The accepted values for these states are also listed.

State	Experimental, eV	Theoretical, eV
3S	5.15 ± 0.01	5.13
4S	1.89 ± 0.02	1.93
3P	2.96 ± 0.01	3.02

Style

It is often helpful to outline your paper, with figures, before you write it. In this way, you can be sure that the logical development is linear and does not resemble two octopuses fighting. One generally writes the report in the past tense. Even though you might have done the work by yourself, use "we," e.g., "We calculated the transition probability."

There are words or phrases you should be careful of using: *Fact*—This is a legal word generally avoided in the physics literature. *Proof or prove*—These words are meaningful in mathematics, but you cannot prove something in physics, especially experimental physics. *The purpose of this experiment is*—Through background information we outline the issue we aim to solve. *One can easily show that*, *It is obvious that*, or *One clearly can see*—Such statements only intimidate the reader who does not find your work trivial. *The data was* or *the data shows*—Data is the plural form of the noun datum: "The data are" or "The data show that." *Human error*—errors must be quantified, not swept under a rug. *In order to*—Just say "to" instead.

We look forward to your submission!

Visa and Immigration Policy: What's Changing?

- Global talent has long powered U.S. science and engineering
- Shifting policies are reshaping that foundation
- Access our policy guide for the latest developments in this changing landscape

Scan for more in-depth information

AIP

PT PHYSICS TODAY

NOVEMBER 2026

MARK YOUR CALENDAR

8TH ANNUAL CAREERS ISSUE

Enhanced visibility opportunities for recruiters and exclusive careers-focused content for job seekers across the physical sciences.

For a sneak preview of the issue and more information on special recruitment advertising options, scan the QR code.

AIP

Society of Physics Students
1 Physics Ellipse
College Park, MD 20740-3841

JURPA

JOURNAL *of*
UNDERGRADUATE
RESEARCH IN PHYSICS
AND ASTRONOMY

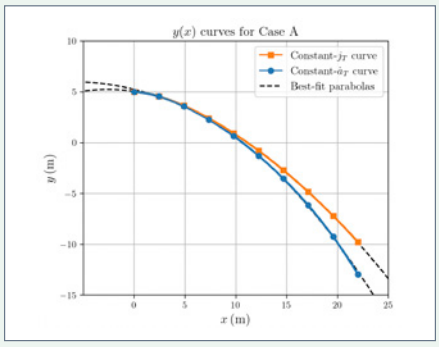
What is JURPA?

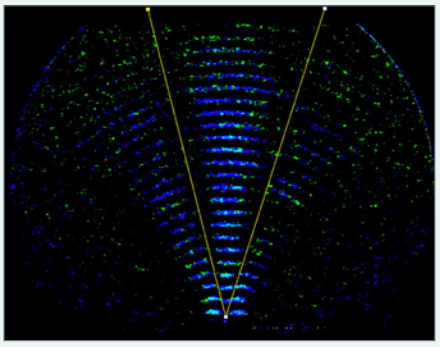
The *Journal of Undergraduate Research in Physics and Astronomy* (JURPA) is a peer-reviewed publication of the Society of Physics Students comprised of research, outreach, and scholarly reporting.

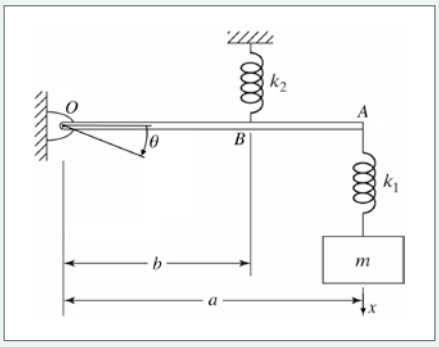
JURPA provides exposure for SPS members conducting physics and astronomy research while also supporting faculty and promoting scholarship through peer-reviewed publication.

Research & Scholarship

JURPA publishes independent research from across the physical sciences, including physics, astronomy, quantum, engineering physics, astrophysics, optics, mathematical physics, nanoscience, acoustics, meteorology, physics and astronomy education, particle physics, electronics, medical physics, and more. Papers are peer reviewed, and accepted submissions are indexed and searchable.







Submit Your Research!

All current SPS members are eligible to submit to JURPA. The author(s) must have performed all the work reported in the paper as an undergraduate student.

JURPA submissions are due on March 15 each year.

To submit your research and review previous JURPA publications, please visit students.aip.org/jurpa.