

MURJ

Massachusetts Institute of Technology
Undergraduate Research Journal



News p. 6

**MIT SERC: Social and
Ethical Responsibilities of
Computing**

Features p. 10

**Modern Astronomy & the
James Webb Telescope**



Develop your
scientific career
with our support

Search for your new role quickly by discipline,
country, salary and more on naturecareers.com

nature careers



Contents

INTRODUCTORY LETTER

- 2 From the Editor

NEWS

- 6 A look at the latest MIT Science News

FEATURES

- 10 **Peering into the Past: How the James Webb Telescope Revolutionizes Modern Astronomy**
Lia Bu

UROP SUMMARIES

- 14 **Rational Design of Highly Selective and Plasticization Resistant Polymers of Intrinsic Microporosity (PIMs) Inspired by Competitive Sorption**
Naksha Roy, Katherine Mizrahi Rodriguez, Zachary P. Smith

REPORTS

- 20 **Improved Binary Classification Accuracy in Natural Language Processing via Test-Time Augmentation**
Helen Lu, Divya Shanmugam, Harini Suresh, John Guttag
- 27 **Determining M52's Cluster Age from Its Color-Magnitude Diagram**
Codrin P. Oneci, Richard Binzel
- 31 **On Symmetric Rank Decompositions of the 3x3 Matrix Multiplication Tensor**
Jason Yang, Virginia Vassilevska Williams



UNDERGRADUATE
RESEARCH JOURNAL
Volume 43, Spring 2022

Editor-In-Chief

Gabrielle Kaili-May Liu

Content Editor

Catherine Griffin

Layout Chief

Faith Choe

Research Editor

Anusha Puri

Content Staff

Lia Bu

Hillel Dei

Sydney Chun

Shinjini Ghosh

Arbri Koplaku

Prathysha Kothare

Audrey Lorvo

Divya Nori

Lily Sensen

Ananth Shyamal

Ezra Yaador

Grace Zhang

Celina Zhao

MURJ Staff
MIT Undergraduate Research Journal

**Massachusetts
Institute of Technology**

May 2022

Dear MIT Community,

As our first fully in-person academic year amidst the COVID-19 pandemic comes to a close, we are excited to present the 43rd issue of the MIT Undergraduate Research Journal (MURJ). MURJ is a biannual student-run publication that features diverse and groundbreaking undergraduate research happening across campus. In publishing this issue, we would like to highlight the importance of active consideration and collaborative addressing of the social and ethical impacts of any type of work, and especially in computing research. While some may view ethics in computing primarily as a tool to put guardrails on and limit technology, it is important to recognize that this is also about shaping scientific innovation for the continued benefit and well-being of humanity. Doing so requires working across disciplines and beyond academia at all levels of development—design, implementation, impacts—and it is essential that we do so with an informed, well-rounded, and precise understanding of the social, ethical, and legal frameworks in which these developments will occur. MURJ is proud to help advance progress in this ongoing journey as we showcase the hard work, creativity, and innovation of our peers. We are grateful to passionate and talented students that make up our MIT community and their continued work to make a difference, as well as the mentors who dedicate their time to training the next generation of researchers, industry practitioners, and global leaders.

In this issue, you will find a series of research articles covering a diverse array of topics, ranging from the dating and analysis of open stellar cluster M52, to the development of a novel divide-and-conquer algorithm for practical use in square matrix multiplication, to the application of test-time augmentation techniques from

No material appearing in this publication may be reproduced without written permission of the publisher. The opinions expressed in this magazine are those of the contributors and are not necessarily shared by the editors. All editorial rights are reserved.

Massachusetts
Institute of Technology

MURJ Staff
MIT Undergraduate Research Journal



UNDERGRADUATE
RESEARCH JOURNAL
Volume 43, Spring 2022

image classification to advance state-of-the-art natural language processing. You will also read a short summary of current UROP research studying the structure and properties of advanced polymers for use in energy-efficient industrial gas separation techniques.

In addition to our research articles, we present insightful features that spotlight exciting scientific topics and happenings around the Institute. We feature reporting on how the James Webb Telescope is revolutionizing modern astronomy, as well as an in-depth introduction to the pioneering programs and efforts of the Social and Ethical Responsibilities of Computing at the MIT Schwarzman College of Computing.

Semiannual publication of this journal is the product of hard work, collaboration, and commitment by dedicated MURJ staff members, and often a product of years of hard work and investment by undergraduate researchers and their mentors. We would like to thank our editorial board and contributors for their time and hard work this semester, as well as all the undergraduates who shared their research with us and the greater MIT community.

For previous issues of the MIT Undergraduate Research Journal, please visit our website at murj.mit.edu. If you are interested in contributing to future issues of the MIT Undergraduate Research Journal, we invite you to join our team of authors and editors or submit your research for our Fall 2022 issue. Please contact murj-officers@mit.edu if you have any questions or comments.

Sincerely,

Gabrielle Kaili-May Liu
Editor-in-Chief

Layout Staff

Gina Choi
Gabrielle Kaili-May Liu

Research Staff

Willow Carretero
Chavez
Alexis Cho
Evan Ewing
Shinjini Ghosh
Prathysha Kothare
Seo Yeong Kwag
Kimberly Liao
Audrey Lorvo
Divya Nori
Brooke Schmelz
Lily Sensen
Kristen Si

No material appearing in this publication may be reproduced without written permission of the publisher. The opinions expressed in this magazine are those of the contributors and are not necessarily shared by the editors. All editorial rights are reserved.



abbvie

Would you like to contribute
and improve people's lives?

We Offer That.

We leverage more than 130 years of innovation with therapies in 33 disease areas, which means we're creating limitless ways to make an impact. We're extraordinarily passionate about our work but we also know how deeply meaningful it is to cultivate a fulfilling life outside of the lab.

That's what makes AbbVie such a perfect career fit.

People. Passion. Possibilities.®

Explore opportunities
and find your fit at
abbvie.com/careers



ences. During 2021 alone, original SERC materials were incorporated in MIT courses with more than 2100 students enrolled across all 5 schools. They have proven critical in supporting new courses such as 24.133 Experiential Ethics, which is offered annually during the summer and now in its third iteration. Learn more about Experiential Ethics at <https://elo.mit.edu/experiential-ethics/>.

In addition, SERC regularly commissions and publishes case studies based on original research for use in undergraduate instruction. These cases examine social, ethical, and policy challenges of present-day efforts in computing ranging from privacy and surveillance to inequality and human rights to algorithms and robotics. They are peer-reviewed by multidisciplinary teams of senior researchers at MIT and checked for balance and accessibility by undergraduate volunteers. Check out the SERC Case Studies, available for free public access, at <https://mit-serc.pubpub.org/>.

Complementary to the case studies are active learning projects, original homework assignments and demonstrations that facilitate the embedding of SERC material in existing classes. Check out active learning projects, newly hosted for public access on MIT OpenCourseWare, at <https://ocw.mit.edu/courses/res-tll-008-social-and-ethical-responsibilities-of-computing-serc-fall-2021/>.

More recently, SERC is working to incorporate the Ethical Computing Protocol (ECP) for educational and practical use. The ECP was originally developed by philosophers Abby Jaques and Milo Phillips-Brown and consists of a structured series of exercises to guide consideration of the benefits and consequences of a proposed research project. While it was initially designed for use with in-person facilitation, SERC's ECP Team is currently working to create a free, interactive, online platform for self-guided use of the ECP in research and development.

Research

SERC Research seeks to develop responsible research and development practices by building community, connections, and infrastructure beginning at the undergraduate level and extending to senior faculty. Fall 2021 witnessed the launching of the SERC Scholars program — open to MIT undergraduates, graduate students and postdocs — which serves as a platform through which students and postdocs can deepen their engagement with SERC and advance SERC projects while entering into a broader community of like-minded SERC advocates and leaders. The first cohort of SERC Scholars was named in Fall 2021, consisting of 9 undergraduate students recognized for their contributions to SERC. The success of the program in its inaugural year is a testament to the MIT community's commitment to advancing socially and ethically responsible development of computing.

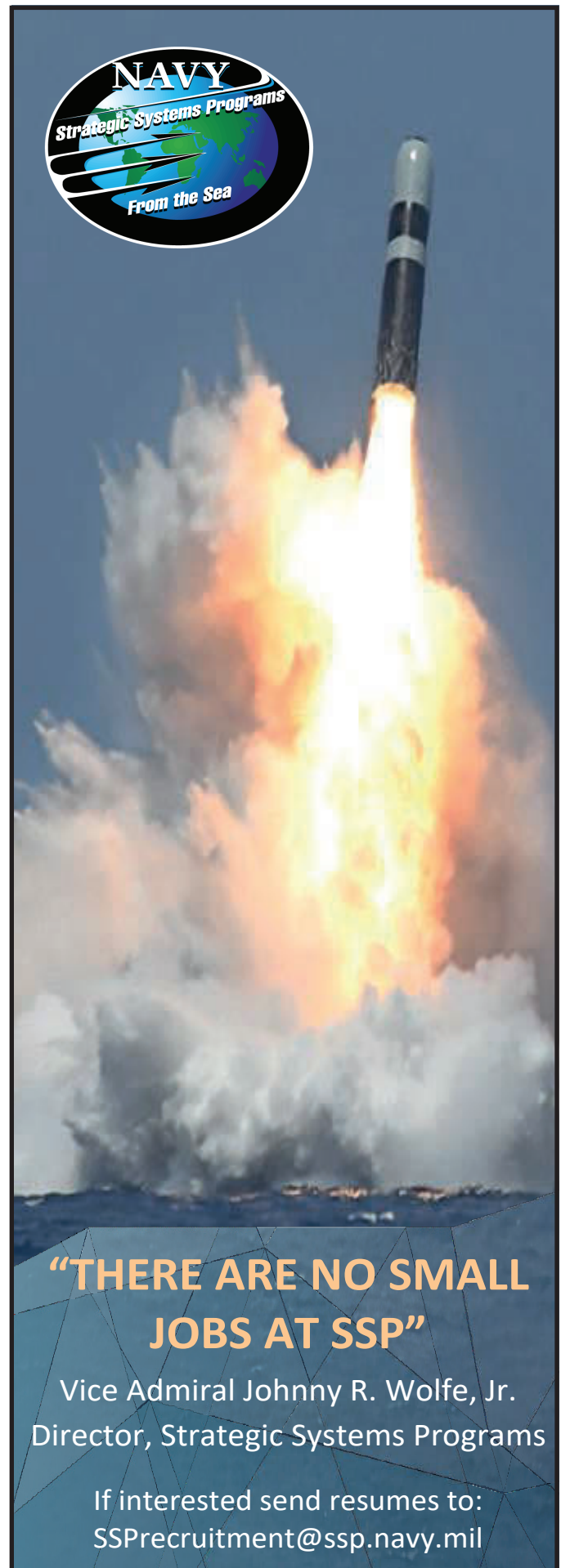
Beyond the MIT student body, SERC Research connects researchers and staff across MIT to advance the development and implementation of a cross-cutting platform for SERC study and practice. This effort includes researchers whose work can benefit from combining methods, insights, and approaches across fields and various aspects of SERC, as well as faculty drawn from 19 different departments, labs, and centers who serve as advisors and participants on SERC boards. Building upon recommendations from the SERC Action Group on Computing, Data, and Anti-racism, along with insights from the Legal, Ethical, and Equity Committee for MIT Campus Planning, these teams are developing guided resources and tools for responsible computing aimed at identifying and mitigating potential harms and informing the design and use of computing tools in civic and industry sectors for better outcomes.

Broader Engagements

Through its Broader Engagements, SERC aims to catalyze partnerships with external stakeholders to shape the future of SERC in application and practice. This is currently centered around SERC's Computing and Policy Task Forces, public forums, and the Legal, Ethical, and Equity Committee for MIT Campus Planning (LEE). SERC task forces bring together leading researchers, industry professionals, and policymakers to develop concrete recommendations for regulatory frameworks and policy-aware technological development, with the goal of guiding "safe, equitable, and innovative advances" in various disciplines. These efforts coordinate with existing programs including the Technology and Policy Program, Internet Policy Research Initiative, and MIT Governance Lab. Insights are additionally sourced from scholars, practitioners, and civic groups through SERC-sponsored public forums. On campus, these insights are put into practice through the LEE, established in April 2020 and co-chaired by Kaiser, Shah, and MIT Vice President and General Counsel Mark DiVincenzo, which provides oversight to ensure proposed operations, research, and protocols adhere to safety, privacy, and equity and, more broadly, MIT's ethical standards and values. SERC additionally operates through international and domestic student and research exchange programs.

More broadly, different SERC frameworks inform different legal and regulatory decisions and consequences. Without careful navigation, this could lead to greater disparity in access to technologies, amplify inherent biases, and escalate gaps between "developed" and "developing" nations. How can we attempt to strike a balance between divergent ideals in the development of computing ethics? Can we achieve a shared framework of SERC principles applicable in cross-disciplinary, cross-cultural, and cross-national settings? MIT SERC is taking critical steps toward uniting researchers, industry practitioners, policymakers, and civic groups in answering these questions.

— Gabrielle Kaili-May Liu



“THERE ARE NO SMALL JOBS AT SSP”

Vice Admiral Johnny R. Wolfe, Jr.
Director, Strategic Systems Programs

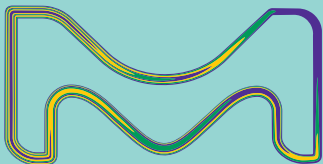
If interested send resumes to:
SSPrecruitment@ssp.navy.mil

Make your science run like clockwork with Supelco[®] Analytical Products

The critical nature of your work means you need products you can rely on for the most demanding applications. That's why the Supelco[®] portfolio provides a comprehensive range of analytical products and services designed to meet the breadth of your specific needs.

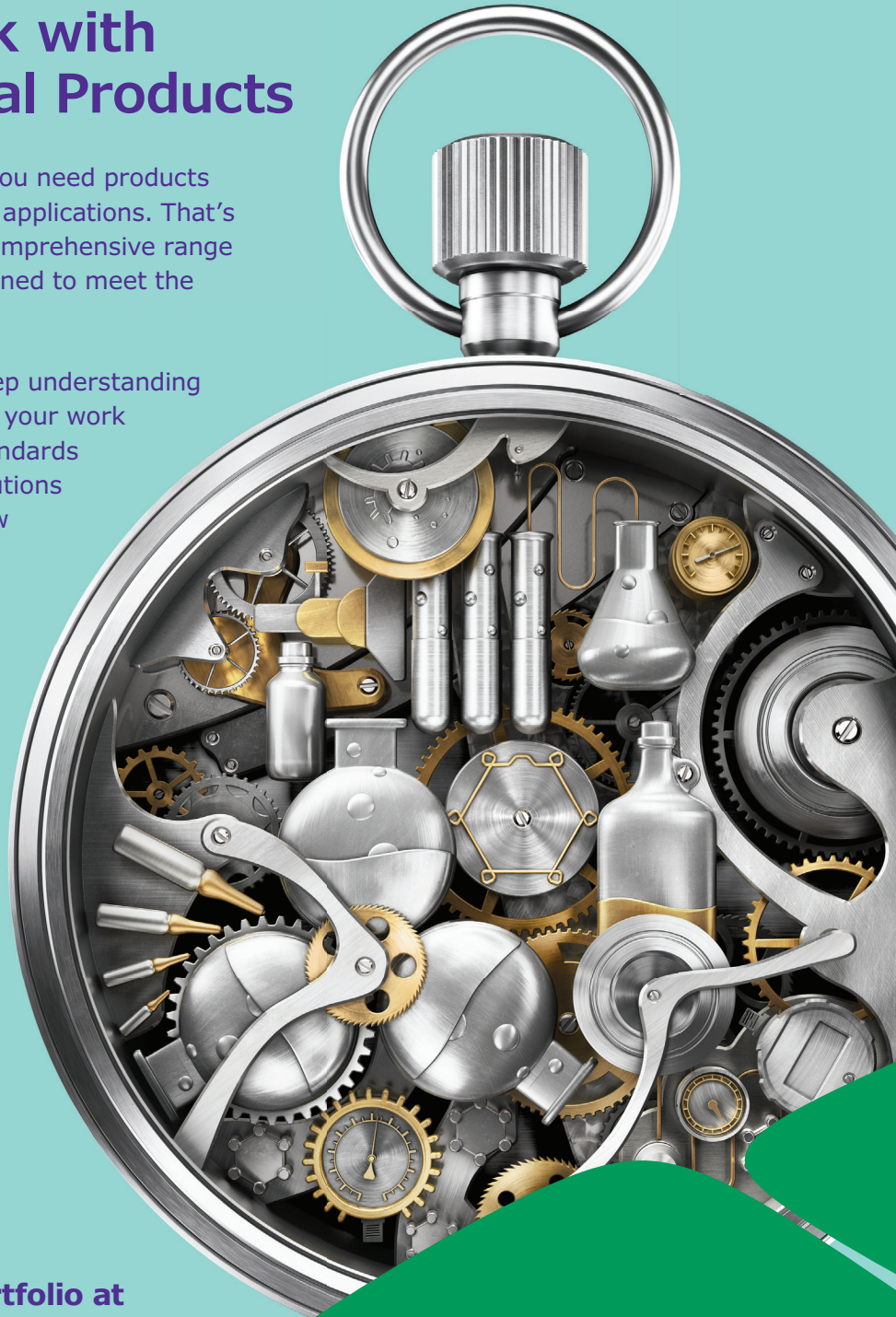
Everything we develop stems from a deep understanding of the regulations and quality standards your work must adhere to. We ensure our high standards match yours so that we can provide solutions that support every step of your workflow for guaranteed success.

For carefully crafted precision and accuracy every time, choose Supelco[®] Analytical Products.



© 2022 Merck KGaA, Darmstadt, Germany and/or its affiliates. All Rights Reserved. MilliporeSigma and the vibrant M are trademarks of Merck KGaA, Darmstadt, Germany or its affiliates. All other trademarks are the property of their respective owners. Detailed information on trademarks is available via publicly accessible resources.

40990 04/2022



Explore the comprehensive
Supelco[®] analytical product portfolio at
SigmaAldrich.com/supelco

MilliporeSigma is the U.S.
and Canada Life Science
business of Merck KGaA,
Darmstadt, Germany.

Supelco[®]
Analytical Products

MURJ Features

Peering into the Past: How the James Webb Telescope Revolutionizes Modern Astronomy

FEATURING MIT PROFESSOR MICHAEL McDONALD

By Lia Bu

Imagine you're an astrophysicist on a train.

A person sits down next to you and asks what you do for a living. The joke goes that if you don't want to engage in conversation, tell them you're a physicist. If you do, tell them you're an astronomer.

One of the most fascinating aspects of astronomy and astrophysics is the amount of passion present. Many people, myself included, grew up loving space, astronauts, and everything to do with the galaxies.

With the James Webb Telescope, the astrophysics field is currently in a period of revival and renewal. To better understand how the James Webb Telescope contributes and pushes new discoveries in science (and why everyone is so excited about it), I turn to Professor Michael McDonald, an assistant professor of physics at MIT.

We start off the conversation by discussing his general work and interest in astrophysics. Contrary to many, Professor McDonald discovered his love for astrophysics later in his career for his PhD. He credits an enthusiastic astrophysics professor he studied under for three years.

Professor McDonald's research focuses on galaxy formation, their evolution, and the influence of black holes in restricting or aiding their growth. With one of the coolest job descriptions, Professor McDonald focuses on galaxies anywhere from 40 million light years away to billions of light years away. Looking at

both close and far galaxies is key for his research.

For reference, light has a finite travel time of 300,000 kilometers/second (around 200,000 miles/second). For example, the sun will probably last for another seven to eight billion years. If something were to happen to the sun, those future humans wouldn't notice until 8 minutes later when light from the incident reached them. Reading the sky, then, is like reading a history book made by stars. The farther the galaxy, the more into the past Professor McDonald and others can witness old galaxy formation. A scientist then can cross-reference that galaxy with a closer one to hypothesize how galaxies evolve over time.

The fascinating technology used for this research spans both land and space-based telescopes. Professor McDonald has used both types of telescopes to capture amazing images of space. On land, the professor's favorite telescope is the Magellan, a classic observatory in Chile. Unlike other telescopes, where the scientist sends the code for a person working next to the telescope to run (and remotely collects data), the scientist must travel to use the Magellan telescope in real-time. Professor McDonald describes this process as inefficient, but also "romantic and inspiring." A scientist, with this method, can make adjustments in real time based on their data.

The James Webb Telescope is the next in line of this amazing technology. In comparison to the Hubble Space Telescope, its optical imaging quality

is very similar. Where it makes advancements is its ability to see infrared light.

Professor McDonald uses a sound analogy to help illustrate why the James Webb Telescope is extremely impactful. When an ambulance siren screams by, it initially sounds high-pitched when coming towards you, but then drops in tone as it drives away. This is due to the Doppler effect, which is when the frequency of sound (or light) changes as the source moves with respect to the observer. When galaxies far away move, they become red-shifted. The light that they emit becomes redder the farther they move away, which means that telescopes like the Hubble can't see them. However, the James Webb Telescope can, opening (literally) astronomer's eyes to new possible galaxies or other objects in space, especially in the deep field.

With the James Webb Telescope, Professor McDonald seeks to understand Neon -6 gas behavior in these galaxies. Currently, research shows that there are volumes of gas around 100 million degrees Kelvin. Over time, if this gas isn't stimulated, it should cool down to approximately 100 degrees. However, that cool gas hasn't popped up- it's missing from the data. Professor McDonald turns to another comparison: "You can count the skiers getting off the ski lift at the top of a hill, and then count them as they reach the bottom of the hill. If those numbers agree, everything makes sense. If you have fewer skiers at the bottom of the hill than at the top of the hill, something is going wrong." With the James Webb Telescope, Professor McDonald hopes to investigate how much gas is at an intermediate temperature to map out how the gas cools.

With every new piece of technology, the field has consistently grown and expanded in leaps. For example, LIGO confirmed gravitational waves when two black holes crashed into each other, sparking a lot more discoveries in that field. While the James Webb Telescope will excel at mid-infrared to infrared discoveries, it could generate an entire new field, a kind of "unknown."

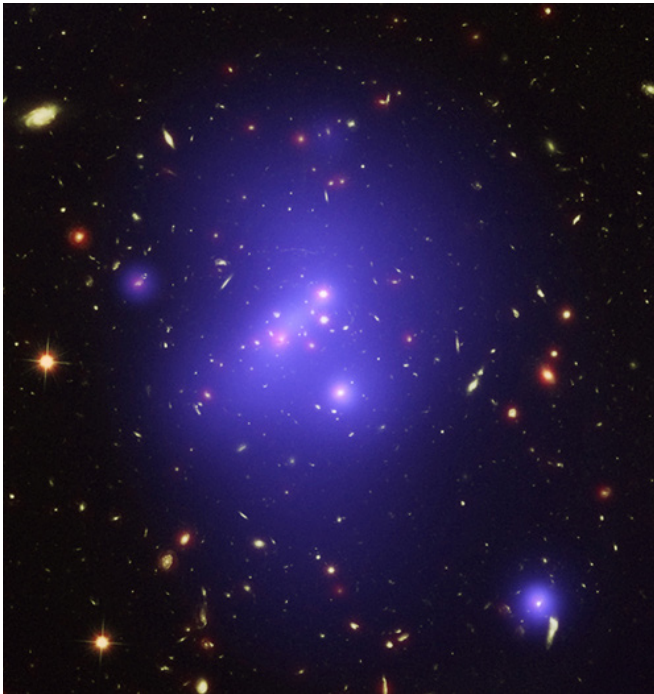
Professor McDonald describes the next year as an "upmanship." So many new things will be discovered as the James Webb Telescope gets used, and this will generate hit news. Already, the five-second calibration image from the James Webb Telescope stirred public interest in the project. The research images themselves will be from hours of exposure, highlighting the telescope's capabilities in propelling research forward.

To finish the conversation, I ask about Professor McDonald's scientific lineage and advice he offers for young people entering academia and the scientific fields. For his PhD, he was advised by Sylvain Veilleux, who in turn was advised by Donald Osterbrock, and in turn by Subrahmanyan Chandrasekhar. For his Master's, he was advised by Stephane Courteau, who was advised by Sandra

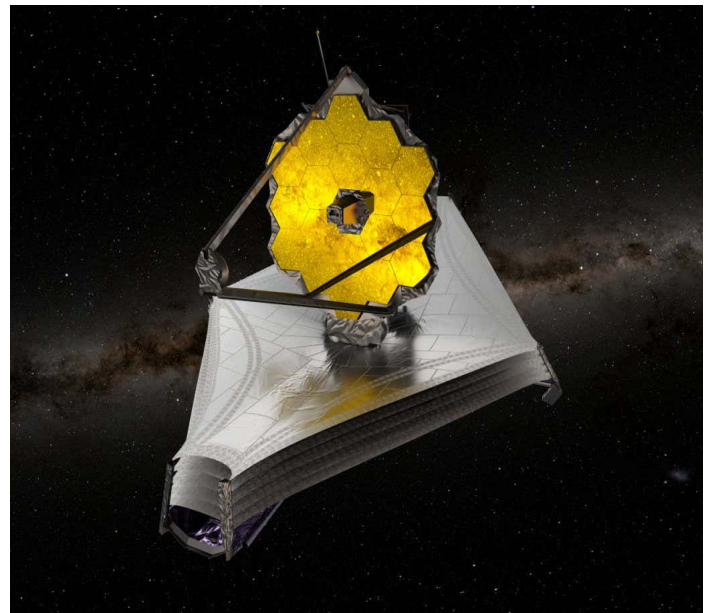
Faber, and in turn by Vera Rubin. And as for his advice? "Try a bunch of things. Trial and error is a valid way to do science." Academia, he states, is very much a self-motivated career. His excitement and enthusiasm drive his research. As a result, he advises students to find a career that they enjoy.

The James Webb Telescope is an exciting time for those in the astronomy and astrophysics field like Professor McDonald. From just a short conversation with the professor, I was amazed with the field of astrophysics, and the new heights it can achieve with the James Webb Telescope. The next year will be thrilling, brimming with amazing discoveries that can potentially redefine how we think about space and our place in the universe.

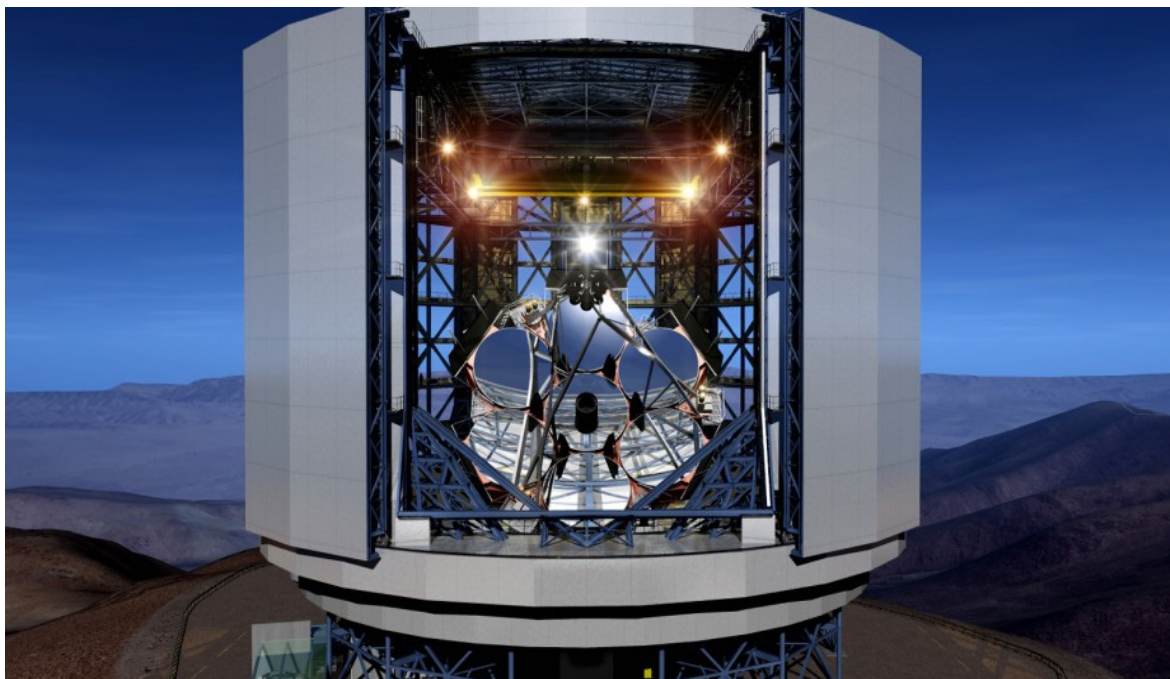
*"Try a bunch of things.
Trial and error is a valid
way to do science."*



McDonald, Michael. "Michael McDonald - Associate Professor - MIT Kavli Institute for Astrophysics and Space Research." Accessed April 18, 2022. <https://space.mit.edu/~mcdonald/research.html>.



Crane, Leah. "JWST: The James Webb Space Telescope Has Arrived at Its Final Destination | New Scientist." *NewScientist - The James Webb Telescope Has Arrived at its Final Destination*, January 2022. <https://www.newscientist.com/article/2305705-the-james-webb-space-telescope-has-arrived-at-its-final-destination/>.



ANU Communications and Engagement. "New President for World's Largest Telescope." ANU. *The Australian National University*, January 25, 2017. <https://www.anu.edu.au/news/all-news/new-president-for-worlds-largest-telescope>.

MURJ

UROP

Summaries

Rational Design of Highly Selective and Plasticization Resistant Polymers of Intrinsic Microporosity (PIMs) Inspired by Competitive Sorption¹

Naksha Roy², Katherine Mizrahi Rodriguez³, Professor Zachary P. Smith⁴

1 Department of Chemical Engineering, MIT

2 Student Contributor, Class of 2022, Department of Chemical Engineering, MIT

3 Supervisor, Department of Materials Science and Engineering, MIT

4 Principal Investigator, Department of Chemical Engineering, MIT

In the last few decades, the membrane gas separations industry has grown significantly (Baker et al., 2014). However, major chemical separations like natural gas purification have been largely carried out via energy-intensive processes, such as distillation and amine absorption, which account for 10-15% of the world's energy consumption (Sanders et al., 2013; Sholl et al., 2016). Development of new energy-efficient separation techniques can help prevent 100 million tons of CO₂ emissions and save \$4 billion in annual US energy consumption costs (Sholl et al., 2016). Unlike traditional processes, polymer membranes have facile operation and a small, modular footprint. However, polymer performance is constrained by a trade-off between throughput (permeability) and separation ability (selectivity). Polymers of intrinsic microporosity (PIMs) show excellent pure-gas separation performance due to their rigid backbones and high free volume (gaps or pores between polymer chains) structure (McKeown et al., 2006), but they are susceptible to physical aging and plasticization, where selectivity is significantly

reduced for gas mixtures at high-pressure (Swaiden et al., 2014). Here, we elucidated direct structure-property relationships between polymer chemistry, sorption affinity, and mixed-gas competition effects for six functional PIM-1 derivatives through mixed-gas and high-pressure transport studies.

Polymeric membranes of six PIM-1 analogues with chemical moieties were prepared using steps detailed in our published work (Fig. 1) (Mizrahi et al., 2021). Films were first evaluated with CO₂/CH₄ binary mixed-gas feeds at a low total pressure of 2 atm (Fig. 2a). A comparison of pure- and mixed-gas results reveal that all polymers, with the exception of PIM-tBOC, demonstrate an increase in CO₂/CH₄ permselectivity and minimal changes in CO₂ permeability. Specifically, the amine functionalized PIM, PIM-NH₂, exhibited an exceptional 140% increase in permselectivity from pure-gas to mixed-gas scenario. A further look at PIM-NH₂ performance at incremental CO₂/CH₄ mixture compositions shows that the CO₂/CH₄ mixed-gas permselectivity slightly increased with increasing content of CO₂, illustrating how competition (i.e., CH₄ exclusion from the polymer matrix) is proportional to the molar ratio of CO₂ over CH₄ in the mixture (Fig. 2b). Importantly, when comparing pure- and mixed-gas tests at the same CO₂ pressure, CO₂ permeabilities were roughly equivalent to each other, whereas, mixed-gas CH₄ permeability was significantly lower compared to pure-gas, due to competitive sorption and exclusion of CH₄.

Gas transport through a polymer is described using the sorption-diffusion model, where permeability is defined as the product of diffusion and sorption coefficients (Wijmans et al., 1995); therefore, sorption and diffusion behaviors were investigated. Additionally, to provide deeper insight into the distinct CO₂ sorption behavior, the infinite-dilution sorption coefficient S^∞ , which describes sorption of the first gas molecule into the polymer, was calculated for all polymers considered (Fig. 1). Notably, comparing pure-gas sorption isotherms and mixed-gas sorption predictions for PIM-NH₂ reveals that CH₄ sorption decreased more than that of CO₂ in mixed-gas predictions (Fig. 3a), while little change was observed between pure- and mixed-gas diffusion behavior. CH₄ experiences a significant reduction in sorption in mixed-gas case due to the presence of highly soluble CO₂, suggesting that mixed-gas

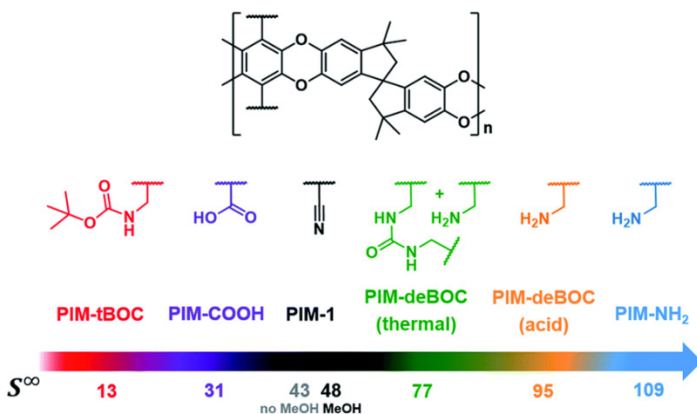


Fig. 1. Generalized PIM-1 backbone and functionalities for the six PIM-1 analogues investigated in this study, along with their corresponding sorption coefficients at infinite dilution (S^∞) in $\text{cm}_{\text{STP}}^3 \text{cm}_{\text{polymer}}^{-3} \text{atm}^{-1}$

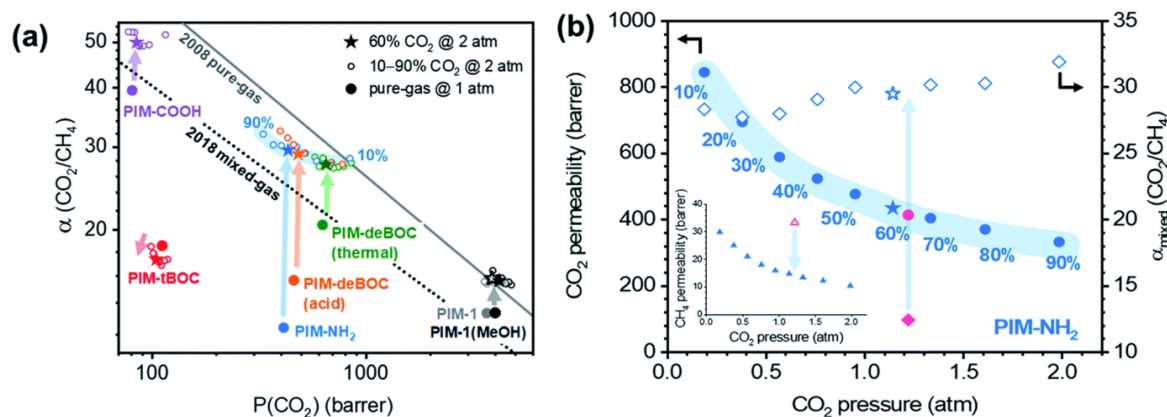


Fig. 2 (a) Permeability-selectivity plot for PIMs containing 2008 pure-gas and 2018 mixed-gas upper-bound relationships. Pure-gas tests were performed at 35 °C and 1.2 atm (filled circles) and mixed-gas tests at 35 °C and 2 atm at nine compositions (open circles). Stars denote mixed-gas tests performed with a 60/40 CO₂/CH₄ composition and a CO₂ partial pressure similar to that of pure-gas. (b) CO₂ mixed-gas permeabilities (blue filled circles) and CO₂/CH₄ permselectivities (open diamonds) for incremental CO₂ compositions at a total pressure of approximately 2 atm for PIM-NH₂. Inset corresponds to mixed-gas CH₄ permeabilities (filled triangles). Pink points indicate pure-gas CO₂ (circle) and CH₄ (open triangle) permeabilities, and pure-gas CO₂/CH₄ permselectivity (diamond). Stars denote permselectivity (open) and CO₂ permeability (filled) tested at a 60/40 CO₂/CH₄ composition.

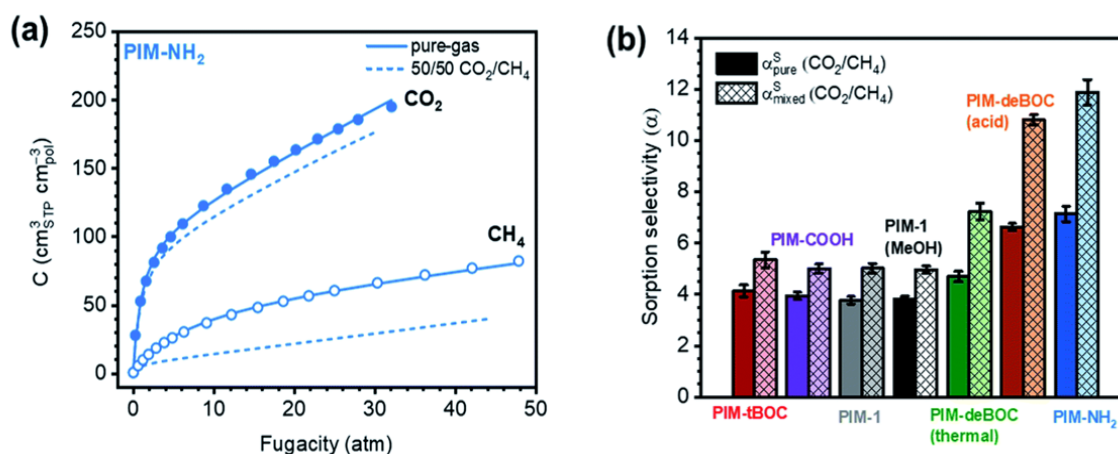


Fig. 3 (a) Pure-gas CO₂ and CH₄ sorption isotherms (circles) fitted with the dual-mode sorption (DMS) model (solid lines), and CO₂ and CH₄ 50/50 mixed-gas sorption isotherms predicted with the DMS model (dashed lines) for PIM-NH₂. The method of using DMS model for mixed-gas sorption predictions is discussed in detail in Mizrahi et al. (2021). Open (CH₄) and filled (CO₂) symbols denote experimentally collected data. (b) Pure- and predicted mixed-gas sorption selectivities at a CO₂ partial fugacity of 1 atm.

enhancements were sorption-driven. The overall increases in CO₂/CH₄ mixed-gas sorption selectivity follow trends of S_{∞} , with PIM-NH₂ demonstrating the highest increase (Fig. 3b).

Plasticization is a significant challenge for the deployment of membranes at industrially relevant high pressures, when penetrants like CO₂ swell the polymer matrix and result in detrimental losses in selectivity. PIM-NH₂ was evaluated for plasticization stability using a 50/50 CO₂/CH₄ mixture and retained high mixed-gas selectivity (>20) up to total feed fugacity (measure of the real partial pressure of a gas in comparison to ideal gas) of 24 atm (Fig. 4a). Moreover, as presented in Fig. 4b, PIM-1 shows an immediate up-turn in CH₄ permeability with increasing CO₂ fugacity, while PIM-NH₂ retains stable CH₄ permeability. Typically, when co-permeated with CO₂ at high pressures, CH₄ experiences an increase in diffusion, resulting in increased CH₄ permeabilities (Ricci et al., 2020). The stable CH₄ permeability exhibited by PIM-NH₂ may be attributed to amine hydrogen bonding effects, which enhance interchain rigidity of the polymer matrix and lead to stronger plasticization resistance compared to PIM-1.

Results from these mixed-gas and high-pressure tests collectively demonstrate the promise of primary amine functionalized PIMs for developing highly sorption-selective and plasticization-resistant membranes for CO₂-based industrial gas separations. Importantly, increasing solubility-selectivity through amine functionalization of the polymer allowed us to leverage competitive sorption and significantly improve permselectivity for CO₂-based gas mixtures. Additional parts of this study, including transport performance analysis in other binary mixtures and comparisons to other microporous polymers, are covered in-depth in our published work (Mizrahi et al., 2021). These findings will be important when designing polymeric membranes for separations like natural gas purification, which requires efficient CO₂ removal and a high CH₄ purity. Many industrial separations usually involve more complex gas mixtures, so our next steps include gas transport studies in ternary mixtures, such as CO₂/CH₄/H₂S, to evaluate membrane performance under more industrially-relevant conditions.

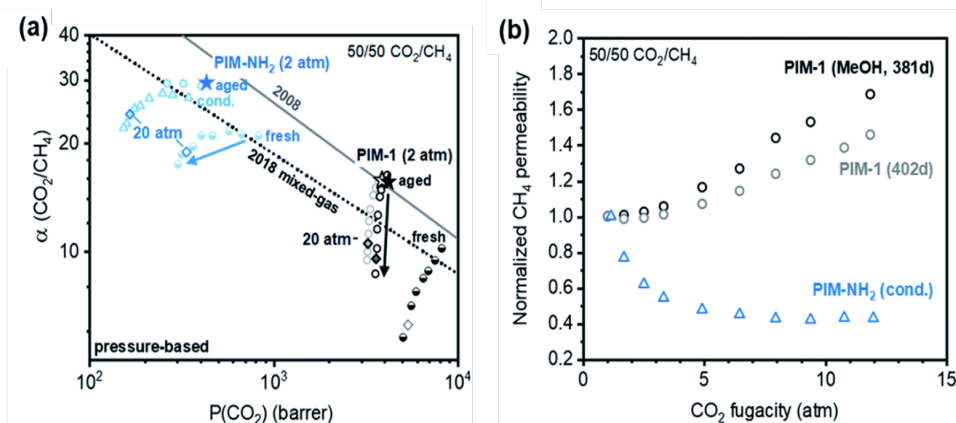


Fig. 4 (a) Mixed-gas CO₂/CH₄ upper-bound and plasticization performance for conditioned PIM-NH₂ (open blue triangles), fresh PIM-NH₂ (half-filled blue circles), aged PIM-1 films which underwent methanol treatment (381 d, open black circles) or no methanol treatment (402 d, open gray circles), and a freshly methanol-treated PIM-1 film (half-filled black circles). Mixed-gas tests performed with a 60/40 CO₂/CH₄ mixture at 2 atm total pressure are denoted as stars. High-pressure mixed-gas points collected at a total pressure of 20 atm (e.g., 10 atm of CO₂) are highlighted as diamonds for comparison with the pressure-based 2018 mixed-gas upper bound. (b) Normalized mixed-gas CH₄ permeability versus CO₂ partial pressure for methanol treated PIM-1 (open black circles), untreated PIM-1 (open gray circles), and conditioned PIM-NH₂ (open blue triangles).

References

- Baker, R. W., & Low, B. T. (2014). Gas separation membrane materials: a perspective. *Macromolecules*, 47(20), 6999-7013. Retrieved on 10 April 2022 from <https://pubs.acs.org/doi/abs/10.1021/ma501488s>.
- McKeown, N. B., & Budd, P. M. (2006). Polymers of intrinsic microporosity (PIMs): organic materials for membrane separations, heterogeneous catalysis and hydrogen storage. *Chemical Society Reviews*, 35(8), 675-683. Retrieved on 10 April 2022 from <https://pubs.rsc.org/en/content/articlelanding/2006/cs/b600349d/unauth>.
- Mizrahi Rodriguez, K., Benedetti, F. M., Roy, N., Wu, A. X., & Smith, Z. P. (2021). Sorption-enhanced mixed-gas transport in amine functionalized polymers of intrinsic microporosity (PIMs). *Journal of Materials Chemistry A*, 9(41), 23631-23642. Retrieved on 10 April 2022 from <https://pubs.rsc.org/en/content/articlehtml/2021/ta/d1ta06530k>.
- Ricci, E., Benedetti, F. M., Dose, M. E., De Angelis, M. G., Freeman, B. D., & Paul, D. R. (2020). Competitive sorption in CO₂/CH₄ separations: the case of HAB-6FDA polyimide and its TR derivative and a general analysis of its impact on the selectivity of glassy polymers at multicomponent conditions. *Journal of Membrane Science*, 612, 118374. Retrieved on 22 April 2022 from <https://www.sciencedirect.com/science/article/abs/pii/S0376738820309522>.
- Sanders, D. F., Smith, Z. P., Guo, R., Robeson, L. M., McGrath, J. E., Paul, D. R., & Freeman, B. D. (2013). Energy-efficient polymeric gas separation membranes for a sustainable future: A review. *Polymer*, 54(18), 4729-4761. Retrieved on 22 April 2022 from <https://www.sciencedirect.com/science/article/pii/S0032386113005399>.
- Sholl, D. S., & Lively, R. P. (2016). Seven chemical separations to change the world. *Nature*, 533(7603), 316-316. Retrieved on 10 April 2022 from <https://www.nature.com/articles/532435a>.
- Swaidan, R., Ghanem, B., Al-Saeedi, M., Litwiller, E., & Pinnau, I. (2014). Role of intrachain rigidity in the plasticization of intrinsically microporous triptycene-based polyimide membranes in mixed-gas CO₂/CH₄ separations. *Macromolecules*, 47(21), 7453-7462. Retrieved on 10 April 2022 from <https://pubs.acs.org/doi/abs/10.1021/ma501798v>.
- Wijmans, J. G., & Baker, R. W. (1995). The solution-diffusion model: a review. *Journal of membrane science*, 107(1-2), 1-21. Retrieved on 10 April 2022 from <https://www.sciencedirect.com/science/article/abs/pii/037673889500102I>.

Chase the miracles that *change* lives

“I have had the opportunity to branch out of the traditional roles of a pharmacist and pursue a different career path that still supports patients in an impactful way. I have been able to apply my scientific background in a new and exciting way, which pushes me to grow every single day. I’m so thankful to be surrounded by and learn from such passionate and talented colleagues.”

-*Belle Estee Dante*
Senior Product Manager, Rare Disease



Join our team in Cambridge, MA!

Dubbed Sanofi at Cambridge Crossing, our new state-of-the-art facility in Cambridge, MA will create an innovation hub promoting close collaboration and integration among business units.

Find out how you too can become part of a team dedicated to chasing the miracles of science that improve people’s lives.

Apply today at sanofiaticambridgecrossing.com

Pursue *Progress*. Discover *Extraordinary*.

sanofi

MURJ Reports

Improved Binary Classification Accuracy in Natural Language Processing via Test-Time Augmentation

Helen Lu¹, Divya Shanmugam², Harini Suresh², John Guttag³

¹ Student Contributor, MIT Class of 2022, Department of Electrical Engineering and Computer Science, Cambridge, MA 02139

² Supervisor, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139

³ Principal Investigator, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139

This research was conducted in affiliation with the Data Driven Inference Group at MIT CSAIL.

TEST-TIME AUGMENTATION—THE AGGREGATION OF PREDICTIONS ACROSS TRANSFORMED EXAMPLES OF TEST INPUTS—IS AN ESTABLISHED TECHNIQUE TO IMPROVE THE PERFORMANCE OF IMAGE CLASSIFICATION MODELS. ALTHOUGH TEST-TIME AUGMENTATION (TTA) MAY BE APPLIED TO ANY DATA MODALITY, IT HAS SEEN LIMITED ADOPTION IN NLP DUE IN PART TO THE DIFFICULTY OF IDENTIFYING LABEL-PRESERVING TRANSFORMATIONS. IN THIS PAPER, WE PRESENT AUGMENTATION POLICIES THAT YIELD SIGNIFICANT ACCURACY IMPROVEMENTS WITH LANGUAGE MODELS. A KEY FINDING IS THAT AUGMENTATION POLICY DESIGN—FOR INSTANCE, THE NUMBER OF SAMPLES GENERATED FROM A SINGLE, NON-DETERMINISTIC AUGMENTATION—HAS A CONSIDERABLE IMPACT ON THE BENEFIT OF TTA. OUR EXPERIMENTS SHOW THAT TTA POLICIES WHICH DO NOT GENERATE MORE THAN ONE SAMPLE FROM A GIVEN AUGMENTATION YIELD LITTLE TO NO IMPROVEMENT IN CLASSIFICATION ACCURACY. EXPERIMENTS ACROSS BINARY CLASSIFICATION TASKS AND DATASETS SHOW THAT TEST-TIME AUGMENTATION CAN DELIVER CONSISTENT IMPROVEMENTS OVER CURRENT STATE-OF-THE-ART APPROACHES.

1. Introduction

Designing accurate and robust machine learning models for classification tasks typically requires large labeled datasets. One standard practice for artificially expanding a given dataset is data augmentation, in which new data are generated by transforming existing examples. While data augmentation is often performed during model training, recent work has shown that the use of data augmentation at inference, or test-time augmentation (TTA), can improve model accuracy (Krizhevsky et. al 2012, Szegedy et. al 2015, Simonyan et. al 2014, Jin et. al 2017, Matsunaga et. al 2017), robustness (Prakash et. al 2018, Song et. al 2017, Cohen et. al 2019), and uncertainty estimates (Ayhan et. al 2018, Matsunaga et. al 2018, Smith et. al 2018, Wang et. al 2019). TTA produces a final prediction by aggregating model predictions (usually by averaging) made from multiple transformed versions of a given test input. In addition to improved model performance, TTA is popular because of its low implementation burden—predictions use a pre-trained model, so no hyperparameter tuning is necessary, and inference on multiple examples can be easily parallelized. Leveraged successfully, TTA can improve model performance without requiring additional data or changes to the underlying model.

While TTA is commonly used in computer vision, there is relatively little research on the potential of TTA in natural language processing (NLP). One reason for this is that it is not clear which augmentations to apply to text inputs at test-time. Computer vision enjoys a set of “standard” test-time augmentations, including flips,

crops, and scales, which are often assumed to be label preserving. In contrast, label-preserving transformations to text are frequently task-dependent, and as a result, the design of performant test-time augmentation policies is not straightforward. A number of design choices, including which augmentations to choose and how to aggregate the resulting predictions, are involved in TTA. We focus on the first step—understanding which augmentations to choose—in the context of natural language processing.

The goal of our work is to explore the use of TTA for improving NLP-based classification. To do this, first we establish four types of TTA policies, apply these to text, and use a simple average to aggregate the resulting predictions. We then use this method to improve accuracy on binary classification of the WILDS CivilComments dataset. We empirically analyze the four different configurations and propose optimal augmentation policies that most significantly improve classification accuracy.

Our results establish that TTA can provide significant improvement to language model classification accuracy. We find that the scale of TTA accuracy increase is heavily dependent on augmentation policy design. Exploring four augmentation policy configurations, we determine that configurations aggregating across multiple samples from non-deterministic augmentations yields the greatest benefit. We also show that augmentations modifying text on a word-level significantly outperform augmentations modifying text on a character-level. We conclude with practical recommendations for the application of TTA in NLP.

2. Methods

We present a method that applies TTA to NLP. We borrow terminology from Shanmugam et al. and assume two inputs to our method:

1. A pre-trained black-box classifier $f: X \rightarrow R^C$ that maps text inputs to a vector of class logit predictions. We use X to denote the space of inputs on which the classifier can operate and C to denote the number of classes. We assume that f is not fully invariant with respect to augmentations (i.e., predictions on transformations of a test input are not guaranteed to be the same).
2. A list of M augmentation transforms, $[a_1, \dots, a_m]$ which we deem an augmentation policy A . Each transform $a_m: X \rightarrow X$ is a stochastic transform that modifies the given input at a character- or word-level based on parameters defining the probability the transform is applied and the magnitude of transformation (e.g. percent of words changed, percent of characters changed), where larger magnitudes translate to larger changes to the example.

Given a text input t , we apply an augmentation policy A containing M augmentation transforms to generate M transformed inputs. All $M + 1$ inputs (transformed and the original) are then passed into the pre-trained classifier f to generate $(M + 1) \cdot R^C$ vectors containing the class logit predictions. We then generate a single prediction by applying a simple average to the $M + 1$ logit predictions. While there are more complex means of aggregating predictions that may maximize prediction performance, we choose to use averaging because it is the simplest version of TTA and suited our goal of understanding whether TTA is useful in any form in NLP tasks.

3. Experimental Setup

We evaluate the performance of our method across datasets and architectures laid out in the following sections.

3.1 Datasets

The WILDS CivilComments dataset (Koh et. al 2021) a modification of the original CivilComments dataset, consists of 448,000 comments made on online articles (Figure 1). Ranging from word lengths of 1 to 315, the comments are labeled as either toxic or non-toxic using a binary indicator obtained from crowdsourcing and majority voting among at least 10 crowdworkers. Each comment is also annotated with binary indicators regarding content mention of nine demographic identities (male, female, LGBTQ, Christian, Muslim, other religions, black, white, other) and six characteristics (severely toxic, obscene, threatening, insulting, identity attacking, and sexually explicit).

3.2 Dataset Splits

Toxic	Comment	Male	Female	...	Identity Attack
0	No, he was accused of being a racist white man.	1	0	...	0
1	The truth hurts so you have to reply with your idiotic comments. You really do not realize how ignorant you are.	0	0	...	1

Figure 1: The WILDS CivilComments dataset contains online comments labeled as either toxic or non-toxic and annotated with 13 binary indicators characterizing comment content.

We divide the given datasets into training (70%) and test (30%) sets. We report only the accuracy performance of the test set because of two reasons: 1) our method architecture leverages models already trained on the training set, and 2) the default TTA approach to aggregation (i.e. averaging) requires no training data.

3.3 Augmentation Policies

We consider two classes of augmentation policies from the *nlpaug* library covering a total of fifteen different augmentations (Figure 3). Insertion, deletion, and substitution of randomly-selected characters are categorized as *character-based* and of randomly-selected words as *word-based*.

For each class of augmentation policies, we chose parameters that maximized accuracy on a held-out validation set. Character-based augmentation policies transformed a random character from 10% of randomly-selected words of an input, with the constraint that any selected word less than 4 characters long would remain unmodified. Word-based augmentation policies modified one random word from a given input, and sentence-based augmentation policies leveraged the default parameters of the *nlpaug* library.

3.4 Models

For all experiments, we used the state-of-the-art classifiers developed for the WILDS CivilComments. We downloaded the pre-trained models, which were fine-tuned DistilBERT-based-uncased models using implementation from Wolf et al. (2019) and hyperparameter settings from Koh et al (2021): batch size 16; learning rate 10^{-5} using the AdamW optimizer (Loshchilov and Hutter, 2019) for 5 epochs with early stopping; an L2-regularization strength of 10^{-2} ; a maximum number of tokens of 300; a learning rate chosen through a grid search over $\{10^{-6}, 2 \times 10^{-6}, 10^{-5}, 2 \times 10^{-5}\}$; and standard/default values for all other hyperparameters. The predictions generated from these models on unaugmented test inputs served as our baseline accuracy.

3.5 Augmentation Configurations

Each augmentation policy (e.g., word deletion) we employed was stochastic and generated numerous variations of a given input. We compared the effect of aggregating a single augmented example generated by an augmentation policy against aggregating multiple augmented examples generated by the same augmentation policy in obtaining a prediction. We generated four augmented samples for multi-sample configurations to maximize accuracy, as determined using a held-out validation.

We also explored the effects of aggregating predictions from different augmentation policies applied to a given input. We chose to evaluate aggregations of four augmentation policies to complement the optimal four sample configurations outlined above.

Original Input: No, he was accused of being a racist white man.

Character-Based Augmentation		
Augmentation	Augmented Output	Description
Optical Character Recognition (OCR)	No, he was accu 8 ed of being a racist white man.	Modify characters in text by simulating OCR errors as defined by mapping.
Keyboard	No, he was accused of being a racist w U ite man.	Modify characters in text by simulating typo errors via keyboard distance.
Random Character Insertion	No, he was accus nd of being a racist white man.	Modify text by injecting characters randomly.
Random Character Swap	No, he was accused of being a racist wh tie man.	Modify text by swapping adjacent characters randomly.
Random Character Deletion	No, he was accus e of being a racist white man.	Modify text by deleting characters randomly.
Random Character Substitution	No, he was accus nd of being a racist white man.	Modify text by substituting characters randomly.
Word-Based Augmentation		
Augmentation	Augmented Output	Description
Random Word Swap	No, he was accused of being a white racist man.	Modify text by swapping adjacent words randomly.
Random Word Deletion	■ , he was accused of being a racist white man.	Modify text by deleting words randomly.
Random Token Split	No, he was accused of being a racist wh l ite man.	Modify text by splitting up words randomly.
Spelling Substitution	No, he was accused of being a racist wh th man.	Modify words in text by leveraging a predefined spelling mistake dictionary to simulate spelling errors.
Wordnet Synonym	No, he was charge of being a racist white man.	Modify text by leveraging semantic meaning and WordNet to substitute words with a random synonym.
PPDB Synonym	No, he was accused of being a racist white servicemen .	Modify text by leveraging semantic meaning and the Paraphrase Database (PPDB) to substitute words with a random synonym.
Antonym	No, he was accused of being a racist black man.	Modify text by leveraging semantic meanings to substitute words with a random antonym.
Contextual Word Embedding Insertion	No, he was accused of being a racist indian white man.	Modify text by leveraging contextual word embeddings to insert words at random positions.
Random Contextual Word Embedding Substitution	No, he was accused of being a murdered white man.	Modify text by leveraging contextual word embeddings to substitute words at random positions.

Figure 2: Fifteen augmentations were used from the nlpaug library to modify a given text and generate a new data point. Each augmentation is stochastic, generating a result with varying degrees of modification based on given parameters.

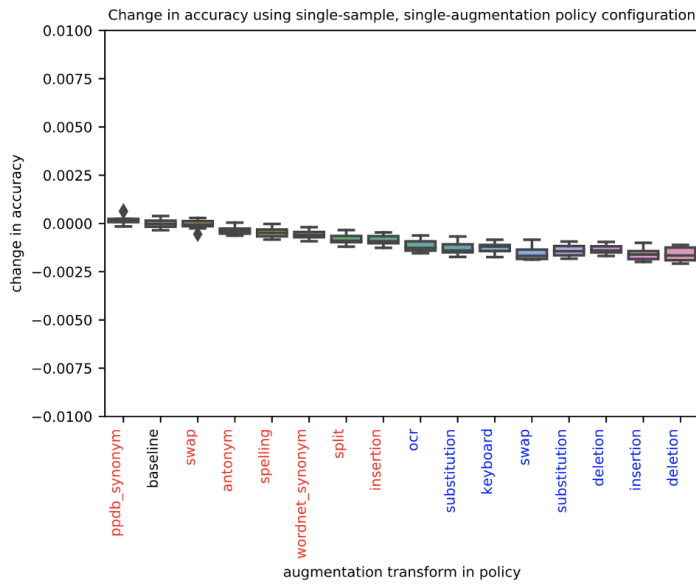


Figure 3: Single-sample, single-augmentation configuration yields no significant improvement in accuracy. Accuracy calculations were averaged across 10 samples, each containing a random 80% of the test inputs. Augmentation transforms in red are word-based, blue are character-based, and black are baseline.

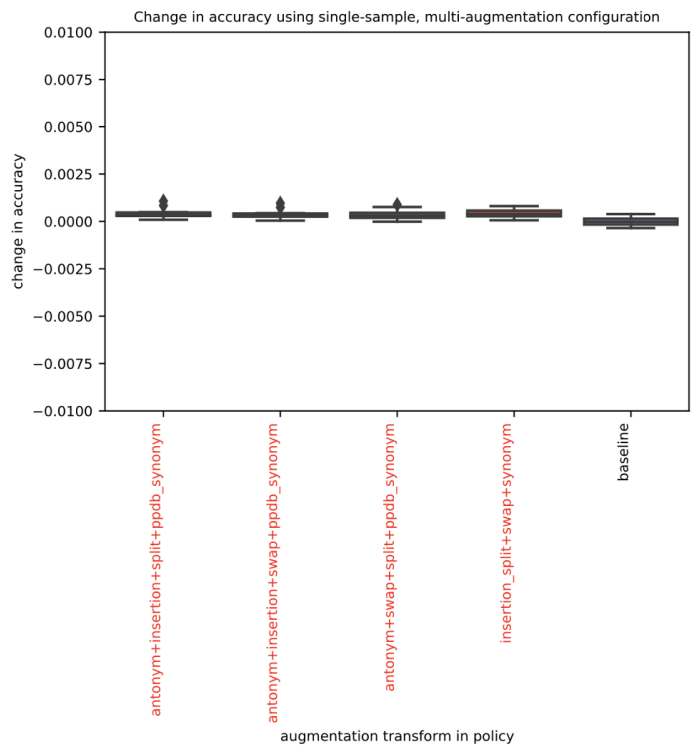


Figure 5: Single-sample, multi-augmentation configuration yields statistically insignificant improvement in accuracy for word-based augmentation transforms. Each of the four word-based augmentation transforms in a policy was applied to a text input to yield four augmented outputs, whose class logit predictions were averaged with the baseline predictions to yield a final classification prediction. Accuracy calculations were averaged across 10 samples, each containing a random 80% of the test inputs.

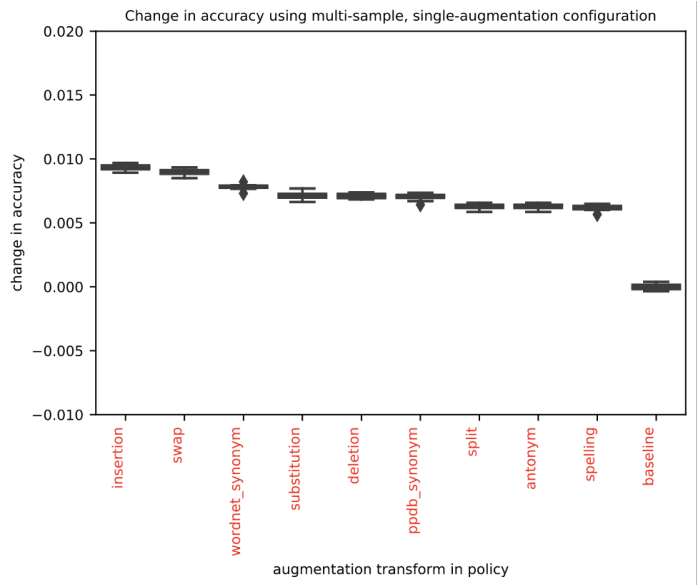


Figure 4: Multi-sample, single-augmentation configuration yields significant improvement in accuracy for word-based augmentation transforms. Each word-based augmentation transform was applied to a text input to yield four augmented outputs, whose class logit predictions were averaged with the baseline predictions to yield a final classification prediction. Accuracy calculations were averaged across 10 samples, each containing a random 80% of the test inputs.

Leveraging these two dimensions, we generated four total augmentation configurations:

- *Single-sample, single-augmentation:* Averaging of the class logit prediction vector of the original, unaugmented input and one class logit prediction vector generated following augmentation policy $A = [a_m]$.
- *Single-sample, multi-augmentation:* Averaging of the class logit prediction vector of the original, unaugmented input and four class logit prediction vectors, each generated following a different augmentation transform in policy $A = [a_m, a_n, a_p, a_q]$. All transforms were of the same class (i.e. either all word-based or all character-based).
- *Multi-sample, single-augmentation:* Averaging of the class logit prediction vector of the original, unaugmented input and four class logit prediction vectors, each generated following the same augmentation transform a_m .
- *Multi-sample, multi-augmentation:* Averaging of the class logit prediction vector of the original, unaugmented input and sixteen class logit prediction vectors: four samples generated by each of four different augmentation transforms $a_m, a_n, a_p,$ and a_q . All transforms were of the same class (i.e. either all word-based or all character-based).

3.6 Statistical Significance

We use a pairwise t-test to measure the statistical significance (with respect to the baseline model) of our accuracy results, which were obtained by averaging 10 random subsamples containing 80% of the test set.

4. Results

4.1 Single-Sample, Single-Augmentation Configuration

For the single-sample, single-augmentation configuration, only the augmentation policy using the transform `ppdb_synonym` (substitution of randomly chosen words with a synonym determined by the PPDB database) outperformed the baseline model's classifications (Figure 3). However, the increase in accuracy was found to be statistically insignificant. We observe here that character-based augmentation transforms worsened classification accuracy to a greater extent than word-based augmentation.

4.2 Multi-Sample, Single-Augmentation Configuration

For the multi-sample, single-augmentation configuration, all nine word-based policies yielded statistically-significant improvements to classification accuracy (Figure 4). The same effect, however, was not observed for character-based augmentation transforms, none of which improved classification accuracy with respect to the baseline model.

4.3 Single-Sample, Multi-Augmentation Configuration

For the single-sample, multi-augmentation configuration, four augmentation policies slightly outperformed the baseline model's classifications, but not significantly (Figure 5). All four augmentation policies consisted of four distinct word-based transforms. We observed that none of the augmentation policies consisting of four distinct character-based transforms improved classification accuracy with respect to the baseline model.

4.4 Multi-Sample, Multi-Augmentation Configuration

For the multi-sample, multi-augmentation configuration, all 126 augmentation policies—combinations of four out of nine word-based augmentation transforms—significantly outperformed the baseline model's classifications. We observed that none of the augmentation policies consisting of four distinct character-based transforms statistically improved classification accuracy with respect to the baseline model.

4.5 Analysis

Our results show that policy design significantly influences accuracy improvements of TTA.

Fixing the number of augmentations in the policy, we compare results using two approaches: one which produces multiple samples from a single augmentation, and another which uses single samples from different augmentations. We find that drawing multiple samples from a single transform produces significantly larger improvements in accuracy: from 0.006 to 0.100.

A natural extension to the multi-sample single-augmentation approach is to replicate it with multiple augmentations. This produces the multi-sample multi-augmentation configuration and we find that it yields a marginal, often insignificant benefit (at most a .04% improvement in accuracy). Given how the multi-augmentation multi-sample configuration comes with the computational cost of many more predictions at test-time, this suggests that multi-sample single-augmentation approaches offer a better tradeoff between computational efficiency and accuracy.

Another trend we observed in our results is that word-based transforms consistently outperformed character-based transforms. This is likely due to how word-based transforms modify inputs. The best-performing word-based transforms incorporated semantic meaning via predefined libraries (WordNet Synonym,

PPDB Synonym, Antonym) or embedding models (Contextual Word Embedding Insertion and Substitution). It is likely that the alterations made by character-based transforms, on the other hand, are altering fundamental word stems. These stems are often used by natural language models to encode words (and their meanings). Modifying a single character may generate a new stem with unrelated (or opposite) meaning, resulting in misclassification of the transformed input.

The accuracy improvements generated from applying TTA are significant for three key reasons. First, TTA is able to outperform a model fine-tuned to perform well on test data. Despite the fact that the original accuracy is 92.3%, TTA can offer improvements of up to 93.7%. Secondly, the baseline model was trained without training-time augmentation (Koh et al., 2021). This indicates that TTA may have even greater potential in NLP contexts than in computer vision, where augmentations used in test-time often parallel those used by the model in train-time. Thirdly, our approach yielded improvements even when using a simple aggregation scheme of averaging. It is likely that, similar to computer vision, greater improvements are possible by weighting augmentations differently.

5. Discussion

In this paper, we investigate accuracy improvements when TTA is used in tandem with language model classification. Through an analysis of the WILDS CivilComments dataset, we show that the appropriate augmentation policy design enables TTA to significantly improve binary classification accuracy of language models.

The insights shared in this study can improve the field's understanding of TTA applications in NLP. This work opens promising areas for future work:

- *TTA in multi-label classification:* The analysis conducted in this paper on the WILDS-CivilComments dataset examines the benefits of TTA in binary classification. Additional work on TTA in multi-label classification is worth considering, especially since state-of-the-art models for this task have greater room for improvement.
- *Better aggregation techniques:* While this work established the application of TTA in NLP, further work to optimize accuracy improvements via more complex aggregation techniques could be conducted. Past work on TTA in computer vision has shown that learned weighting of augmented inputs further improves classification accuracy (Shanmugam et al., 2021) and application of a similar technique in NLP is worth considering.

References

- Abdar, M., Pourpanah, F., Hussain, S., Rezadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., & Makarek, V. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76: 243-297.
- Ayhan, M.S., & Berens, P. (2018). Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks.

Cohen, J., Rosenfeld, E. & Kolter, Z. (2019). Certified adversarial robustness via randomized smoothing. *International conference on machine learning*. PMLR 97: 1310-1320.

Jin, H., Li, Z., Tong, R. & Lin, L. (2018). A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. *Medical physics* 45(5): 2097-2107.

Koh, P.W., Sagawa, S., Marklund, H., Zie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R.L., Gao, I., Lee, T., David, E., Stavness, I., Guo, W., Earnshaw, B., Haque, I., Beery, S.M., Leskovec, J., Kundaje, A., Pierson, E., ... Liang, P. (2021). Wilds: A benchmark of in-the-wild distribution shifts. *International conference on machine learning*. PMLR 139: 5637-5664.

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25.

Liu, J.Z. (2019). Variable selection with rigorous uncertainty quantification using Bayesian deep neural networks. *Bayesian Deep Learning Workshop At NeurIPS*.

Prakash, A., Moran, N., Garber, S., DiLillo, A., & Storer, J. (2018). Deflecting adversarial attacks with pixel deflection. *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Matsunaga, K., Hamada, A., Minagawa, A., & Koga, H. (2017). Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. *arXiv preprint arXiv:1703.03108*.

Shanmugam, D., Blalock, D., Balakrishnan, G., Guttag, J. (2021). Better aggregation in test-time augmentation.

Proceedings of the IEEE/CVF International Conference on Computer Vision.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Smith, L., & Gal, Y. (2018). Understanding measures of uncertainty for adversarial example detection. *arXiv preprint arXiv:1803.08533*.

Song, Y., Kim, T., Nowozin, S., Ermon, S., & Kushman, N. (2017). Pixeldefend: Leveraging generative models to understand and defend against adversarial examples. *arXiv preprint arXiv:1710.10766*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T. (2019). Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing* 338: 34-45.

Zhang, J., Kailkhura, B., Han, T.Y. (2020). Mix-n-match: Ensemble and compositional methods for uncertainty calibration in deep learning. *International conference on machine learning*. PMLR 119: 11117-11128.



The logo for HebeCell features a stylized circular icon on the left containing a green and blue figure resembling a person or a cell. To the right of the icon, the word "HebeCell" is written in a bold, green, sans-serif font.

Incurable no more

The future of cancer treatment

Our mission is to develop and commercialize cell-based therapeutics to treat incurable diseases and contribute to the fields of regenerative medicine and immuno-oncology.

- Groundbreaking, patented technology with proven results
- Leading biologists and researchers
- Collaborative, clinical approach

21 Strathmore Rd, Natick, MA, 01760, United States
 info@hebecell.com • www.hebecellcorp.com
 (508) 545-2190

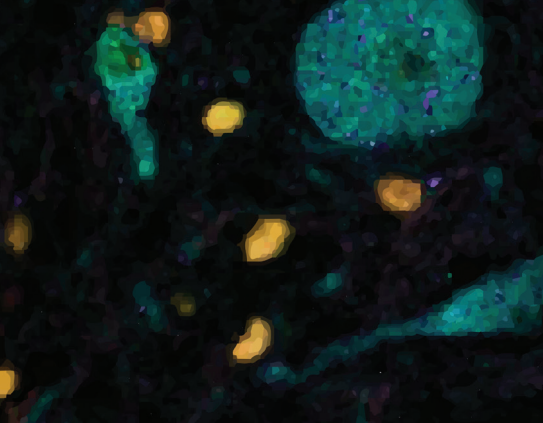


A woman with long dark hair, wearing safety glasses and a white lab coat, is looking intently at something off-camera. She is in a laboratory setting with other people in the background. The lab coat has the Alexion logo on the sleeve.

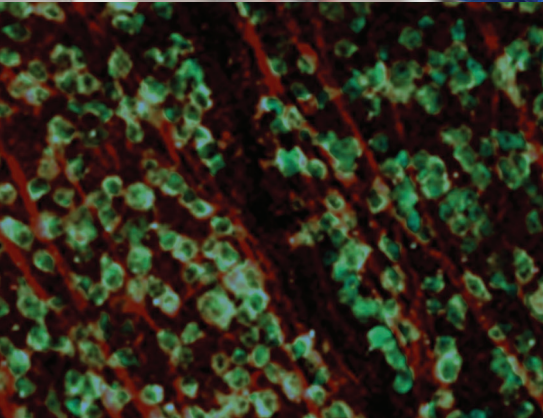
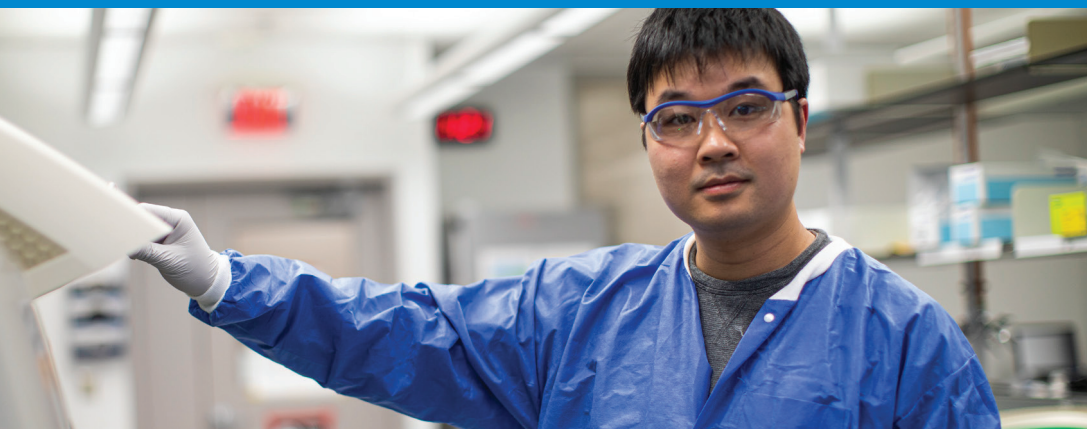
Rare Inspiration. Changing Lives.

At Alexion, our mission is to transform the lives of people affected by rare diseases and devastating conditions through the development and delivery of innovative medicines, as well as through supportive technologies and healthcare services.

Our innovation is driven by a sense of urgency and a relentless quest for answers, because lives are at stake. Through research and development we're building a better tomorrow, every day. [alexion.com](https://www.alexion.com)



where
science
meets **humanity**™



Innovation inspired by the passion of our people

As pioneers in neuroscience, Biogen discovers, develops and delivers worldwide innovative therapies for people living with serious neurological diseases as well as related therapeutic adjacencies. We are focused on advancing the industry's most diversified pipeline in neuroscience that will transform the standard of care for patients in several areas of high unmet need.

biogen.com



Determining M52's Cluster Age from Its Color-Magnitude Diagram

Codrin P. Oneci¹, Richard Binzel²

¹ Student Contributor, MIT Class of 2021, Department of Physics, Cambridge, MA 02139

² Supervisor, Department of Earth, Atmospheric and Planetary Sciences, Cambridge, MA 02139

THIS ARTICLE PRESENTS OBSERVATIONS AND ANALYSIS THROUGH WHICH THE AGE OF THE OPEN CLUSTER M52 WAS DETERMINED. LITERATURE PRESENTS A VARIETY OF CLUSTER AGES FOR THE MESSIER OBJECT M52. OBSERVATIONS IN R', I' AND CALIBRATION CCD FRAMES WERE DONE AT WALLACE ASTRONOMICAL OBSERVATORY. BY FITTING PADOVA ISOCHRONES THROUGH THE COLOR MAGNITUDE DIAGRAM OF M52 I WAS ABLE TO DETERMINE THE CLUSTER AGE OF M52 AS 80MYR+60MYR/-20MYR. THIS PROJECT SHOWS THAT THE VARIOUS AGES FROM THE INTERVAL 60-140 MYR ARE PLAUSIBLE AND THAT STARS OF M52 FORMED RECENTLY AND AT SLIGHTLY DIFFERENT TIMES.

1. Introduction

Designing accurate and robust machine learning models for clastellar clusters are groups of stars located in the same physical region of the Galaxy. Globular clusters are gravitationally bound, while open clusters are not. Colormagnitude diagrams (CMD) for stellar clusters are used for determining the age of clusters. CMDs are created by plotting the color index of stars from the cluster against the absolute magnitude of stars from a cluster. In such a CMD one can observe stars that are on the path of becoming red giants (stars that depleted their hydrogen fuel and cannot start a more energetic thermonuclear cycle) in a region called "turnoff point". Since for a cluster most stars are formed from the same matter, it is useful to associate ages to clusters with the purpose of studying stellar evolution models. Stellar isochrones are curves in CMD and Hertzsprung-Russell diagrams representing the population of stars that have different masses but the same composition and age. In the past isochrones were fitted by hand by astronomers in order to determine a cluster age using models of stars from the main sequence and doing fitting at the turnoff point. Modern computational tools allow us to automate this process, and in the last years significant steps forward have been made [1].

M52 is a young open cluster from the Messier catalog that is visible to the naked eye. M52 was chosen as an object of study because it is not very sensible to atmospheric noise (it has a low magnitude) and is very well suited for observations from MIT Wallace Astronomical Observatory (WAO) (its equatorial coordinates are included in the window of objects that can be observed during the fall nights at WAO). Since M52 contains a large variety of stars, its plausible isochrones have a high spread in age. That may imply that different fitting and extraction algorithms would be biased toward specific age values [2].

2. Background

The cluster age of M52 is 60 ± 10 million years (Myr) when Padova isochrones are fitted through its color magnitude diagram in the (J, J-H) coordinates [3]. This result is in accordance to previous knowledge of the age of M52 situated between 35 and 135 Myr. Older results [4] suggest an age of 95 Myr

from ubvy photometry of chosen stars from the cluster, with a distance modulus of 11.3 ± 0.1 (corresponding to approximately 1.7 kiloparsecs). In the past M52 attracted attention of astronomers due to the F7Ib supergiant that is part of it [5]. Isochrone fitting is most frequently based on simple stellar populations, for which metallicity and mass distributions are tried until the data from the CMD is fitted by an evolutionary program [6].

3. Astronomical Observations and Data Collection

I have gathered data using a remote-controlled 14-in Celestron with equatorial mount, a STL1001E CCD sensor and a SLOAN filter wheel. The observations were made at the Wallace Astronomical Observatory in the evenings of 7 and 14 October 2020. For the frames with nonzero exposure time, the exposure time of choice was 40s. Choosing the exposure time of 40 seconds was such that the pixels don't oversaturate and such that the usual assumptions related to CCD operation in the linear regime hold. Specifically, during the measurements the CCD functional regime at which the charge corresponding to each pixel in an image can be linearly translated into an incident photon count. The r' and i' filters were chosen in order to obtain high signal to noise ratios in the measurements (FIG. 1). By using large wavelength filters I avoided having additional noise in the star CCD observations. Moreover the seeing due to atmospheric refraction is lower and thus the source extraction algorithm gives a lower flux estimation uncertainty. This improved the estimation of the magnitudes and resulted in a more precise CMD.

Table 1 summarizes the type and quantity of data collected at the Wallace Observatory during the observational sessions. Filters i' and r', corresponding to infrared wavelengths, respectively red wavelengths were selected during different measurements through mechanical movements of the filter wheel of the Celestron telescope. All the data collected from the CCD was in the form of .FIT files, which required data type conversion in the analysis steps of this project. The second column of the table shows the total number of sequential repeated measurements for each frame type (aggregated from two observational sessions).

TABLE I. Data Table

CCD Frame Type	Repeats	Observation
r'	20	
i'	20	
Dark	20	14-in Celestron + STL1001E
Bias	20	
Flat Field r'	10	
Flat Field i'	10	

While the r' and i' frames are specific to the object of study (including data of the celestial objects), all the other frames taken (last four rows of Table 1) correspond to CCD sensor calibration.

Dark frames are taken with the aperture closed (no incoming photons from celestial objects) for a given exposure time and are used to estimate the thermal agitation noise causing non-zero readings in the CCD frames. Multiple measurement repeats are needed for increasing estimation accuracy (reducing estimation uncertainty) and for eliminating non-reoccurring events such as cosmic rays causing electron excitation in the Silicon structure of the CCD. In order to minimize thermal agitation and temperature variation noise in measurements the CCD sensor was cooled (thermostat temperature -100C according to the manufacturer).

Bias frames are another type of calibration frames and are taken with zero exposure time, by quickly registering data with pixel values in a sequence in order to estimate the bias error due to the reading process. Typically this is due to manufacturing tolerances, and the physical process causing such errors is the charge leaking and gain due to excitation during their manipulation throughout the sensor using time varying electric fields.

Finally the flat frames are taken with the CCD sensor being pointed towards a diffuse translucent surface, with the CCD filter wheel on and some relatively distant object (light bulb a few meters away from the telescope) producing light behind the diffuse and thin material sheet. The flat frames were taken with a few seconds

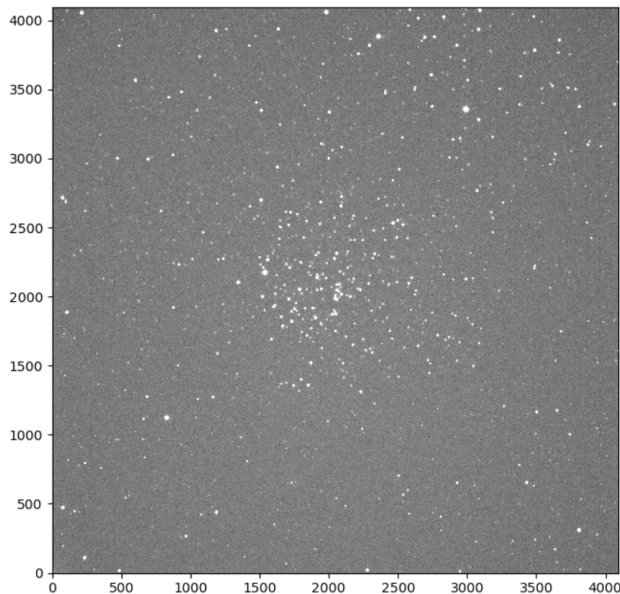


FIG. 1. M52 raw image in the r' filter.

exposure time to prevent oversaturation of pixels. Since all pixels have slightly different physical properties, their sensitivities differ, and flat fields serve as a means to estimate differences in individual pixels responses to incoming photons.

4. Data Reduction and Analysis

Using the repeated frames, I computed the arithmetic average of the arrays falling in specific categories (Darks, Biases, r', i' and flat fields). Afterwards I computed the calibrated frames according to the formula:

$$CCD_{calibrated} = \frac{CCD_{raw} - D - B}{FF}$$

where D is the averaged Dark signal, B is the average Bias and FF is the flat field normalization array for the corresponding filter.

Further in the 30'x30' FOV frames I defined a virtual circular aperture (FIG. 2) that is centered on M52 and occupies 0.6 of the linear FOV. The size of this aperture was chosen empirically: large enough to have a continuity of stars in the main sequence of M52 in the CMD, such that the distribution of stars and the turnoff point is evident, but small enough to minimize star-noise caused by stars that are in the aperture but are not included in M52, so they appear as being scattered in the CMD.

Since I worked with r' and i' data I then extracted the sources in the calibrated frame using the Python SEP library (which finds isoluminant contours and fits ellipsoid sources with gaussian distributions of signal). Thus I obtained flux levels in both r' and i' filters. I transformed this fluxes into magnitudes using the photometric formula

$$m_0 - m_1 = -2.5 \log \frac{F_1}{F_0}$$

where m_0 and F_0 are the magnitude and flux of the reference star (Vega for instance) while m_1 and F_1 are the magnitude and flux of the star that is under analysis.

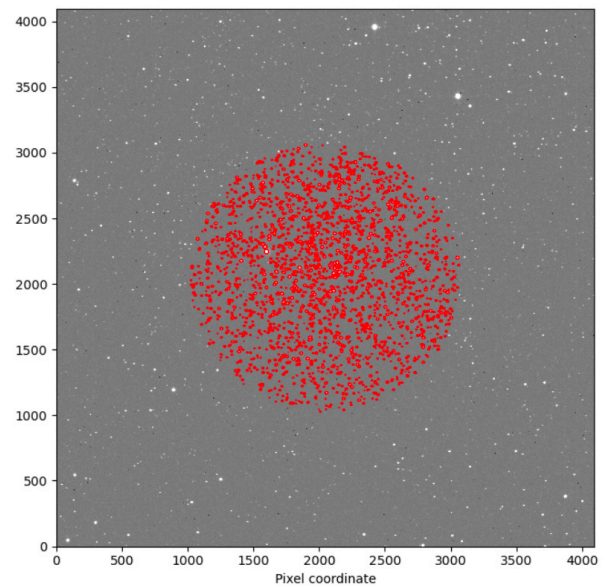


FIG. 2. M52 calibrated image in the i' filter, with extracted sources shown as red ellipsoids. The circular aperture is centered on M52 and has diameter 0.60xFOV.

All written source code and astronomical data gathered was made available at the following repository: <https://github.com/codr-oneci/Stellar-cluster-age-from-color-magnitude-diagram>.

5. Results and Analysis

The obtained i' and r' pair magnitudes had to be then translated to a color index $r'-i'$ and an absolute magnitude i' . I identified GAIA NGC 7654 766 (maximum apparent luminosity in the chose aperture) for and then used a correction factor of 22.38 mag for finding absolute i' . For this calibration I observed that NGC 7654 766 has apparent magnitude in the i' filter of 8.38 mag, while initially that would correspond to -14 mag in my raw result. In principle I could have obtained the same result using the known distance modulus for M52 11.1 mag .

I obtained the CMD for M52 by plotting absolute i' against the $r'-i'$ color index (FIG. 3). Subsequently I had to retrieve already computed Padova isochrones from the research group server. As an observation, isochrones are curves in the Hertzsprung-Russel diagram which correspond to stars of the same age from the same cluster (having the same composition and metallicity) but having different initial stellar mass, and are used in stellar evolution models to evaluate the observational parameters of stellar clusters and their evolution tracks. I further wrote software for transforming their relevant columns from a .dat file into the Numpy arrays needed for this project. Data retrieved was in the SDSS ugriz photometric system. By plotting together the CMD of M52 and the Padova isochrones of ages 20-140 Myr (FIG. 4), I observed that translating the CMD (FIG. 5) using an extinction coefficient of $E(r'-i')=0.2$ was necessary to match the main sequence of Padova isochrones for stars that are having still much hydrogen to burn (stars that are not close to the turn-off point). The sum of the extinction coefficient aforementioned and the raw color index gave the intrinsic color indexes of M52 stars. I used $[Fe/H]=0.11$ in the retrieved Padova isochrones, because I knew from literature that the M52 cluster is young. As I did not find iron-line metallicity analysis results for M52, I have made the assumption that these stars should have slightly sub-solar metallicity.

This analysis resulted in the cluster age $t_{M52,new} = 80Myr + 60Myr/ - 20Myr$.

In the language of hypothesis testing, we would be curious to see whether the null hypothesis

$$H_0 : t_{M52,new} = t_{M52,literature}$$

passes the t-test with 95 percent confidence or if we should consider the alternate hypothesis to be more lilyly.

The t-value is

$$t = \frac{t_{M52,new} - t_{M52,literature}}{(\sigma(t_{M52,new})^2 + \sigma(t_{M52,literature})^2)^{1/2}}$$

Observe that $|t| = 0.25 < 1.96$ so we conclude that the null hypothesis is the most likely hypothesis (from the ones considered). The results of this project reconfirm previous literature findings. Obviously the high measurement variance (in my experiments and in literature) relative to the estimated parameter value (cluster age) also contribute towards reconfirming contemporary literature results on M52's age. A significant result would be obtained with a completely new and detailed method, such as incorporating spatial star distribution in the cluster itself (inferred from dynamic models and Doppler spectroscopy).

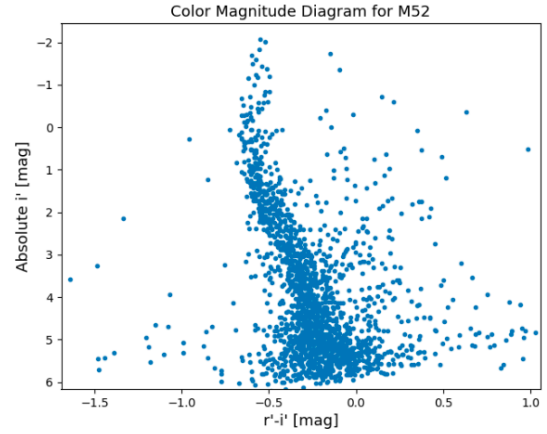


FIG. 3. Color-magnitude plot of M52. The vertical axis shows absolute i' magnitudes, while the horizontal axis has $r'-i'$ color indexes for individual stars. The values are not corrected for extinction and absorption of light from the open cluster.

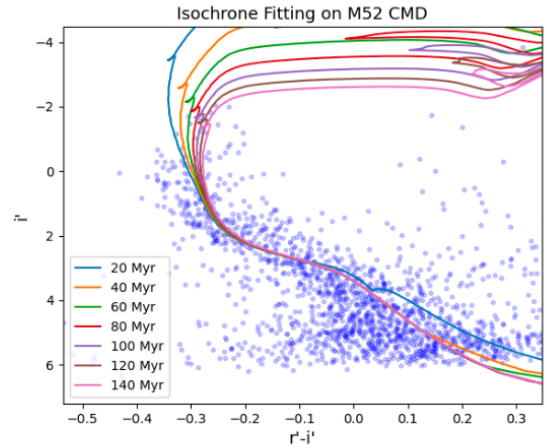


FIG. 4. Relevant age Padova Isochrones fitted on the CMD of M52. You can see that all the isochrones fit well most of the CMD, the differences becoming evident only at the turnoff point.

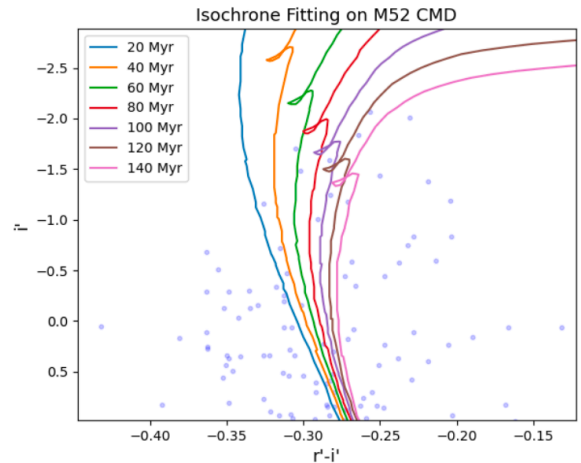


FIG. 5. Detail of the turnoff point of M52.

6. Conclusion

Based on the Padova isochrone fitting in the CMD of M52 I determined the cluster age to be $t_{M52,new} = 80M \text{ yr} + 60M \text{ yr} - 20M \text{ yr}$. This is in accordance to previous literature results of $t_{M52,literature} = 60Myr \pm 10Myr$ [3].

This project shows that the various ages from the interval 60-140 Myr mentioned in astrophysics literature are all plausible. A potential explanation of this fact is that stars of M52 formed at slightly different times. This would be a consequence of a non-uniform mass distribution of gas during the time when this cluster formed: dispersed gas matter (mostly hydrogen and helium, and heavier elements generated in population I stars) converged into dense entities such as stars at different moments due to gravitational attraction between matter constituents. As the cluster is definitely young, different star formation times may account for the high dispersion in its age estimated through isochrone fitting. In order to investigate this hypothesis in the future one would need to create a spatial model of M52 stars and analyse separately stars from different regions of the clusters, after correcting for the star dynamics from the formation of the cluster: one would expect stars that were in proximity during the star formation time to originate from the same gas matter and to have the same age.

References

[1] Monteiro, H., Dias, W.S., and Caetano, T.C. (2010). Fitting Isochrones to photometric data – a new global optimization tool. *Astronomy & Astrophysics*, 516.

[2] Pandey, A. K., Nilakshi, O., Sagar, R., and Tarusawa, K. (2001). NGC 7654: An interesting cluster to study star formation history. *Astronomy & Astrophysics*, 374, 504-522.

[3] Bonatto, C., Bica, E. (2018). Methods for improving open cluster fundamental parameters applied to M 52 and NGC 3960 *Astronomy & Astrophysics*, manuscript aa5315, revision 1.

[4] Kaltcheva, N.T. (1990). Photographic UVBY Photometry of the Open Cluster M52. *Astrophysics and Space Science*, 173(1), 69-76.

[5] Persch, P. (1960). The Galactic Cluster NGC 7654 (M 52). *Astrophysical Journal*, 132, 689

[6] Bruzual, G.A. (2010). Star clusters as simple stellar populations. *Philosophical transactions of the Royal Society A*, 368(1913).



Join our team!

At Biocytogen, passionate people work together to accelerate progress in biomedical science. We're serious about our work, but also fun to work with—and we're looking to add new members to our rapidly-growing team of talented science and service professionals.

We're proud of our world-class technology, collaborative work environment, abundant opportunities for advancement, and competitive benefits.

Come join us!



ANTIBODY
DRUG
DISCOVERY



HUMANIZED
MOUSE
MODELS



PRECLINICAL
SERVICES

Current Open Positions:

- Business Development Manager - Product & Services
- Business Development Manager - Licensing
- Sr. Scientist/ Product manager for Gene Targeting
- Scientist for Content Marketing
- In vitro Pharmacology Research Associate
- In vivo Pharmacology Research Associate
- Lab Operations Associate
- Operations Associate

Scan to learn more or visit
biocytogen.com/career
and apply today.



BIOCYTOGEN is an Equal Opportunity Employer. Employment opportunities at BIOCYTOGEN BOSTON CORP are based upon one's qualifications and capabilities to perform the essential functions of a particular job. All employment opportunities are provided without regard to race, color, religion, sex, national origin, ancestry, age, sexual orientation, gender identity and expression, veteran status, disability, marital status, genetic information, or any other characteristic protected by law. This Equal Employment Opportunity policy governs all aspects of employment, including, but not limited to, recruitment, hiring, selection, job assignment, promotions, transfers, compensation, discipline, termination, layoff, access to benefits and training, and all other conditions and privileges of employment.



INSTRON®

THE LEADING FORCE

IN MECHANICAL TESTING



www.instron.com

On Symmetric Rank Decompositions of the 3x3 Matrix Multiplication Tensor

Jason Yang¹, Virginia Vassilevska Williams²

¹ Student Contributor, MIT Class of 2025, Department of Electrical Engineering and Computer Science, Cambridge, MA 02139

² Supervisor, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139

FAST DIVIDE-AND-CONQUER ALGORITHMS FOR MATRIX MULTIPLICATION CAN BE MADE BY FINDING LOW-RANK DECOMPOSITIONS OF MATRIX MULTIPLICATION TENSORS. THE TENSOR OF 3 X 3 MATRIX MULTIPLICATION IS THE SMALLEST MATRIX MULTIPLICATION TENSOR IN WHICH THE LOWEST POSSIBLE RANK OF ANY DECOMPOSITION OF IT IS UNKNOWN. WE SEARCH THROUGH SETS OF POSSIBLE DECOMPOSITIONS OF RANK AT MOST 23, THE CURRENT BEST-KNOWN UPPER BOUND ON THE LOWEST POSSIBLE RANK. TO MAKE THIS SEARCH FEASIBLE, WE WORK OVER THE FIELD OF INTEGERS MOD 2 FOR THIS TENSOR AND IMPOSE SPARSITY AND SYMMETRY CONSTRAINTS. WE SHOW THAT FOR CERTAIN TYPES OF SYMMETRIES, THERE DO NOT EXIST ANY SUCH DECOMPOSITIONS. THIS MEANS THAT IN ORDER TO FIND A LOW-RANK DECOMPOSITION OF THIS TENSOR THAT MATCHES OR SURPASSES THE CURRENT RECORD, ONE MUST USE LOOSER SYMMETRY REQUIREMENTS OR CONSIDER NON-INTEGER DECOMPOSITIONS.

1. Introduction

Matrix multiplication is a fundamental operation in linear algebra, making it useful in fields such as applied mathematics, physics, and engineering. One of the most important problems in computational linear algebra is finding the asymptotically most efficient algorithm for matrix multiplication. To multiply two $n \times n$ matrices, the naive algorithm takes $O(n^3)$ time. In 1969, Strassen found a clever divide-and-conquer algorithm for this problem that takes $O(n^{\log_2 7}) = O(n^{2.807})$ time (Strassen, 1969). His work started the field of fast matrix multiplication, and a series of improvements to his divide-and-conquer technique has led to the current fastest algorithm running in $O(n^{2.3728596})$ time, found by Alman and Williams (Alman & Williams, 2021).

However, most fast matrix multiplication algorithms have extremely high constant factors in their running time, rendering them impractical. This is partially because, in order to achieve their low asymptotic timecomplexity, they rely on the limiting behavior of certain divide-and-conquer procedures as their "order" approaches infinity. Thus, some researchers have attempted to discover more moderately-sized divide-and conquer algorithms in the hopes of finding a practical matrix multiplication algorithm that is faster than Strassen's algorithm for square matrices, and this is the subject of our research.

The problem of fast matrix multiplication comes down to finding a low-rank tensor decomposition of the $\langle n, k, m \rangle$ matrix multiplication tensor (Bläser, 2013)

$$\langle n, k, m \rangle_{a,b,c,d,e,f} = \begin{cases} 1 & a = b \text{ and } c = d \text{ and } e = f \\ 0 & \text{else} \end{cases},$$

for chosen constants n, k, m .¹

A R -rank decomposition of this tensor is a set of R triplets of coefficient matrices

$$\{(A_r, B_r, C_r) \mid 0 \leq r < R, A_r \in \mathbb{R}^{n \times k}, B_r \in \mathbb{R}^{k \times m}, C_r \in \mathbb{R}^{m \times n}\}$$

such that

$$\langle n, k, m \rangle = \sum_{r=0}^{R-1} A_r \otimes B_r \otimes C_r,$$

where \otimes denotes the tensor product; or, written in another way,

$$\langle n, k, m \rangle_{a,b,c,d,e,f} = \sum_{r=0}^{R-1} A_{r,a,b} B_{r,c,d} C_{r,e,f}.$$

It can be shown that a R -rank decomposition of the $\langle n, k, m \rangle$ tensor yields a $O(N^{\log_{nkm} R})$ -time divide-and-conquer algorithm for multiplying two $N \times N$ matrices: the matrices are split into block matrices containing $n \times k$ and $k \times m$ matrix elements, respectively, and R recursive multiplications between these matrix elements are used to make the product matrix (Bläser, 2013). Thus, the goal of essentially every fast matrix multiplication algorithm, including Strassen's algorithm, is to find a R -rank decomposition of a certain $\langle n, k, m \rangle$ tensor for the smallest possible R , and obtain a new lowest exponent $\log_{nkm} R$ for running time.²

¹ Our definition of this tensor is slightly different from the conventional definition in that the last two indices are flipped, but any decomposition of the conventional matrix multiplication tensor can be trivially converted to a decomposition of this tensor, and vice versa. We use our definition in this report because the tensor will have better symmetry, which will soon become important.

² Technically, more recent algorithms use very complicated techniques, and their exponents cannot be easily expressed in the form $\log_a b$, but their ultimate goal is still to find a low-rank decomposition of a $\langle n, k, m \rangle$ tensor.

A table of known upper and lower bounds on the lowest possible rank R for small n, k, m is given in (Smirnov, 2013). In particular, $R = 7$ is optimal for $\langle 2, 2, 2 \rangle$, and the lowest known rank for $\langle 3, 3, 3 \rangle$ is $R = 23$, first achieved in 1976 by Laderman (Laderman, 1976) and matched by several other people since (Courtois, Bard, & Hulme, 2011). Our primary research focus is the $\langle 3, 3, 3 \rangle$ tensor.

2. Motivation for our work

Strassen’s algorithm uses the following 7-rank decomposition of $\langle 2, 2, 2 \rangle$:

$$\langle 2, 2, 2 \rangle = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \\ + \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \\ + \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Ballard (Ballard, 2017) noticed that this tensor decomposition is cyclically symmetric: if we replace every triplet of matrices $A \otimes B \otimes C$ with $B \otimes C \otimes A$, or every triplet with $C \otimes A \otimes B$, the decomposition stays identical except possibly the order of the triplets. To make this symmetry more clear, we can rewrite Strassen’s decomposition as

$$\langle 2, 2, 2 \rangle = (I \otimes I \otimes I)$$

$$+ (A_0 \otimes A_1 \otimes A_2 + A_1 \otimes A_2 \otimes A_0 + A_2 \otimes A_0 \otimes A_1)$$

$$+ (B_0 \otimes B_1 \otimes B_2 + B_1 \otimes B_2 \otimes B_0 + B_2 \otimes B_0 \otimes B_1);$$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, A_0 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}$$

$$B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, B_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, B_2 = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Ballard also noticed that $\langle n, n, n \rangle$ is cyclically symmetric for all n :

$$\langle n, n, n \rangle_{a, b, c, d, e, f} = \langle n, n, n \rangle_{c, d, e, f, a, b} = \langle n, n, n \rangle_{e, f, a, b, c, d}$$

This fact suggests that cyclically symmetric decompositions may exist for $\langle n, n, n \rangle$ tensors where $n > 2$, which can be useful since the set of all cyclically symmetric lists of matrix triplets is far smaller than the set of all arbitrary lists of matrix triplets. Indeed, Ballard found a 23-rank cyclically symmetric decomposition of the $\langle 3, 3, 3 \rangle$ tensor, matching the current lowest-known rank; his decomposition works over the integers, although all coefficients are balanced ternary (i.e. $\{1, 0, -1\}$) (Ballard, 2017).

Although Ballard did not find a cyclically symmetric 22-rank decomposition, the number of such decompositions is enormous (about $3^{(22 \cdot 9 \cdot 3)/3} \approx 3 \times 10^{94}$ for balanced ternary, or about $2^{(22 \cdot 9 \cdot 3)/3} \approx 4 \times 10^{59}$ for mod 2), so it seems unlikely that all of them have been exhausted. Therefore, we have hope that such a decomposition may be out there that successfully decomposes the $\langle 3, 3, 3 \rangle$ tensor.

3. Our work

Our work consists of searching over all possible integer decompositions of the $\langle 3, 3, 3 \rangle$ tensor over mod 2, while enforcing additional symmetry on the searched decompositions beyond cyclic symmetry. Both the additional symmetry and the mod 2 restriction greatly reduce the size of the search space, and another reason we work over mod 2 is that we suspect valid integer decompositions of the $\langle 3, 3, 3 \rangle$ tensor of rank ≤ 23 to be extremely rare, so working over mod 2 can rule out many potential tensor decompositions over the integers.

3.1 Symmetries

If we examine Strassen’s decomposition more closely, we can write $B_0 = d(A_0)$, $B_1 = d(A_1)$, $B_2 = d(A_2)$, where

i.e. the matrices B_i are simply 180-degree rotated versions of

$$d\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} d & c \\ b & a \end{pmatrix},$$

A_i . Also notice that $d(I) = I$, so Strassen’s decomposition satisfies 180-degree-rotation symmetry in addition to cyclic symmetry.

How can we generalize this to $n \times n$ matrices for the $\langle n, n, n \rangle$ tensor? For our research, we have chosen three kinds of symmetries:

- S_n : $P_\sigma(M) = M'$, where $M'_{i, j} = M_{\sigma(i), \sigma(j)}$ and σ is any permutation over n elements
- Shift: above, but $\sigma = ((i+k) \bmod n)$ for some parameter k
- Flip: $\sigma = (0, \dots, n-1), (n-1, \dots, 0)$.

To ensure that these symmetry restrictions are reasonable, notice that $\langle n, n, n \rangle$ satisfies S_n symmetry (and thus also the shift and flip symmetries), since $\langle n, n, n \rangle_{a, b, c, d, e, f} = \langle n, n, n \rangle_{\sigma(a), \sigma(b), \sigma(c), \sigma(d), \sigma(e), \sigma(f)}$ for any permutation σ .

To transform matrix triplets, we denote $\mathcal{O}(A, B, C) = (B, C, A)$ and the shorthand $P_\sigma(A, B, C) = (P_\sigma(A), P_\sigma(B), P_\sigma(C))$. Then $\mathcal{F} = \{P_\sigma \circ C^c \mid 0 \leq c < 3, \sigma \in \mathcal{S}\}$ set of all transformations over matrix triplets that we will use to enforce symmetry, where \mathcal{S} is the set of allowed permutations σ depending on what symmetry is used.

Finally, it can be shown that an \mathcal{F} -symmetric tensor decomposition must be of the form

$$\mathcal{G}(M) = \parallel_{(A, B, C) \in M} \text{list}(\{f((A, B, C)) \mid f \in \mathcal{F}\}),$$

for some set of matrix triplets M . The curly braces $\{\}$ denote a set, where duplicate elements in the set are ignored; the list() function represents putting the elements of the input set into a list in arbitrary order; and \parallel denotes list concatenation.

In this sense, the matrix triplets in M are “generators” that, when transformed with functions in \mathcal{F} , generate the actual tensor decomposition.

3.2 The search problem

From this point onward, all arithmetic between tensors is done mod 2. Let $T := \langle n, n, n \rangle$, and, borrowing notation from (Tichavsky, 2021), let $[[D]] := \sum_{(A, B, C) \in D} A \otimes B \otimes C$.

For each possible r , let $L_r := \{(A, B, C) \text{ s.t. } |\mathcal{G}(\{(A, B, C)\})| = r\}$ be the set of all triplets that produce r -sized groups, where if two triplets t_1, t_2 yield the same tensor $[[\mathcal{G}(t_1)]] = [[\mathcal{G}(t_2)]]$ only one of t_1, t_2 is stored in L_r .

We restrict maximum total rank to be R , and to reduce search space, we add an additional restriction that depends on a manually adjusted parameter Z . Our search problem is then:

$$\text{Find } M \subseteq \bigcup_{1 \leq r \leq |\mathcal{F}|} L_r \text{ such that } [[\mathcal{G}(M)]] = \mathbb{T},$$

$$\text{subject to } |\mathcal{G}(M)| \leq R,$$

$$\#(A \otimes B \otimes C) \leq Z \quad \forall (A, B, C) \in M,$$

where $\#(T) := (\# \text{ of nonzero elements in } T)$.

3.3 Algorithm

Instead of iterating over all possible M , our algorithm uses a generalized “meet-in-the-middle” technique:

- $D_{i,0}, D_{i,1}$ such that $\cup\{a|b : a \in D_{i,0}, b \in D_{i,1}\} \supseteq D$, where D is the sets of all valid M ;
- for each i , iterate over each $M_1 \in D_{i,1}$ and use some data structure to efficiently query whether there exists a $M_0 \in D_{i,0}$ s.t. $[[\mathcal{G}(M_0)]] = \mathbb{T} - [[\mathcal{G}(M_1)]]$.

In this way, the time complexity of searching will be $O(\Sigma_i (|D_{i,0}| + |D_{i,1}|Q))$, where Q is the query time complexity, rather than $O(|D|)$. In our case, $Q = O(\log|D_{i,0}|)$, but since we will later force all $|D_{i,0}|$ to be below some large constant in order to restrict memory consumption, we can assume $Q = O(1)$. In this way, the time complexity of the searching will be $O(\Sigma_i (|D_{i,0}| + |D_{i,1}|))$.

To construct such sets $D_{i,0}, D_{i,1}$, we use what we call the *profile* of a set of matrix triplets M :

$$\mathbb{P}(M) := [\#\{(A, B, C) \in M \text{ s.t. } |\mathcal{G}(\{(A, B, C)\})| = r\}]_{0 \leq r \leq |\mathcal{F}|},$$

i.e. a histogram of $|\mathcal{G}(\{(A, B, C)\})|$ over all $(A, B, C) \in M$.

Given the profile $P = \mathbb{P}(M)$ of M , we can immediately evaluate the total rank of $\mathcal{G}(M)$ as $R(P) = \sum_{0 \leq r \leq |\mathcal{F}|} rP_r$. Thus, the set of all possible profiles of valid M under our search constraints is:

$$\mathcal{P} = \{P \mid \sum_{1 \leq r \leq |\mathcal{F}|} rP_r \leq R, P_r \leq |L_r| \quad \forall r\}.$$

We then solve the following optimization problem:

$$\text{Find integer } S, \text{ subsets } \mathcal{P}_{i,0}, \mathcal{P}_{i,1} \subseteq \mathcal{P} \text{ for } 0 \leq i < S,$$

$$\text{minimizing } \sum_i c(\mathcal{P}_{i,0}) + c(\mathcal{P}_{i,1}),$$

$$\text{subject to } \cup_i \{P_0 + P_1 \mid P_0 \in \mathcal{P}_{i,0}, P_1 \in \mathcal{P}_{i,1}\} \supseteq \mathcal{P},$$

where $+$ denotes element-wise addition between two lists,

and

$$c(Q) = \sum_{P \in Q} \prod_{0 \leq r \leq |\mathcal{F}|} \binom{|L_r|}{P_r}$$

is the number of possible matrix triplet lists with profile $\in Q$;

We can then let $D_{i,k} := \{\text{all possible sets of matrix triplets } M \mid \mathbb{P}(M) \in \mathcal{P}_{i,k}\}$, $k = 0, 1$, which can be constructed via depth-first search. Since $\mathbb{P}(M) \in \mathcal{P} \quad \forall M \in D$, and $\mathbb{P}(M_0 \cup M_1) = \mathbb{P}(M_0) + \mathbb{P}(M_1)$, we have that $\forall M \in D, M_0 \in D_0, M_1 \in D_1$ s.t. $M = M_0 \cup M_1$, so we cover the entire search space D .

In practice, we add an additional constraint, $c(\mathcal{P}_{i,0}) \leq 10^9 \quad \forall i$, in order for the program to fit within limited memory.

To obtain an optimal solution, notice that we can designate eachv set $\{P_0 + P_1 \mid P_0 \in \mathcal{P}_{i,0}, P_1 \in \mathcal{P}_{i,1}\}$ to contain a certain subset Q_i of profiles \mathcal{P} ; we can then find some $V_{j,0}, V_{j,1}$ such that $Q_{i,j} = V_{j,0} + V_{j,1}$, for each $Q_{i,j} \in Q_i$, and then set $\mathcal{P}_{i,k} = \{V_{j,k}\}$ for $k = 0, 1$. Essentially, we fix which profiles of \mathcal{P} will be covered by which sum-sets and also fix how each profile is covered, and then generate a solution $\mathcal{P}_{i,0}, \mathcal{P}_{i,1}$ satisfying those constraints.

Since the size of this problem can be quite large (e.g. for cyclic+ S_n symmetry and $n = 3, R = 23, Z = 729$, we have $|\mathcal{P}| = 308$), we must settle for approximate optimal solutions. For this we fix S and use beam search with a beam width of 10,000: at each iteration, we consider the next profile $\mathcal{P} \in \mathcal{P}$, iterate over each $V_0, V_1 \in \mathcal{P}$ such that $P = V_0 + V_1$, iterate over each $0 \leq s \leq S$, and augment each solution $\{(\mathcal{P}_{s,0}, \mathcal{P}_{s,1})\}_i$ in the beam by setting $(\mathcal{P}_{s,0}, \mathcal{P}_{s,1})$ to $(\mathcal{P}_{s,0} \cup \{V_0\}, \mathcal{P}_{s,1} \cup \{V_1\})$. We start with $S = 1$ and repeatedly increase S by 1, until the objective function $\sum_i c(\mathcal{P}_{i,0}) + c(\mathcal{P}_{i,1})$ stops decreasing, and keep the best found solution.

Memory Optimization

Every set of matrix triplets M in $D_{i,0}$ is generated from a certain profile $P = (\mathcal{P}_{i,0})_p$ and set of indices $i_{r,j}$ for $1 \leq r \leq |\mathcal{F}|, 0 \leq j \leq P_r$, where the resulting M is $[[\cup_r \{(L_r)_{i_{r,j}}\}]]$. We encode M as

$$\left(\sum_{q=0}^{p-1} c(\{(\mathcal{P}_{i,0})_q\}) \right) + \sum_{r=1}^{|\mathcal{F}|} \left[C_{|L_r|, (\mathcal{P}_{i,0})_{p,r}}([i_{r,j}]_j) \prod_{r'=r+1}^{|\mathcal{F}|} \binom{|L_{r'}|}{(\mathcal{P}_{i,0})_{p,r'}} \right].$$

Essentially, we represent each combination $[i_{r,j}]_j$ for each fixed r as an integer using the C functions, compress all the resulting integers using a mixed-radix system, and add the result by an offset to account for each profile in $\mathcal{P}_{i,0}$.

Other Optimizations

- Each matrix triplet (A, B, C) is encoded as the 27-bit integer $t(A, B, C) = \mathbb{B}(l(A) || l(B) || l(C))$, where $l(A)$ returns a list of the elements of matrix A in row-major order, and $\mathbb{B}(A) = \sum_{i=0}^{|A|-1} A_i 2^i$ for a list A .
- Any tensor of the form $T = \mathcal{G}(M)$ must have certain groups of its elements be equal, due to having a F-symmetric decomposition: we can define a list of lists of 6-tuples G_i , s.t. any such tensor T satisfies $\forall i, T_{G_i,0} = T_{G_i,1} = \dots = T_{G_i,|G_i|-1}$; we then define the *compression* of T to be $\tilde{T}_i = [T_{G_i,0}]_i$, which is stored as a bitset (since each of its elements is in $\{0, 1\}$) in a 32-bit integer array.

- Over mod 2, + and - between matrices and tensors are both equivalent to bitwise XOR, denoted as \oplus . This is also true for compressed tensors, since $\widetilde{(A+B)} = \widetilde{A} + \widetilde{B}$.
- We store a table $\widetilde{\mathbb{G}}_{i(A,B,C)} = \widetilde{[[\mathbb{G}((A,B,C))]]}$; thus, $\widetilde{[[\mathbb{G}(M)]]} = \bigoplus_{(A,B,C) \in M} \widetilde{\mathbb{G}}_{i(A,B,C)}$, which is faster to calculate than is the naive formula.
- Let $\mathbb{C}(A, f) := \mathbb{B}(A_{32f32(f+1)})$ be the evaluation of the f-th 32-bit chunk of A; for each $0 \leq f \leq \lfloor \frac{|G|-1}{32} \rfloor$ create a bitset \mathbb{F}_p of length 2^{32} with all bits initially 0; then for each $M \in D_{i,0}$ for some fixed i, set $\mathbb{F}_{f,\mathbb{C}(\widetilde{\mathbb{G}}(M),f)} \leftarrow 1$ for each f.
- For each $M \in D_{i,0}$ when we query whether $\exists M_0 \in D_{i,0}$ s.t. $\widetilde{[[\mathbb{G}(M_0)]]} = \widetilde{T} = \widetilde{T} \oplus \widetilde{[[\mathbb{G}(M_1)]]}$,

we first check if

$$\mathbb{F}_{f,\mathbb{C}(\widetilde{T},f)} = 1$$

if not, then we know for sure that no such M_0 exists. In practice, about 90 to about 99.9+ percent of all search queries terminate during this phase.

- We define an arbitrary function

$$\mathbb{H}(A) = \left(\bigoplus_f \mathbb{C}(A, f) \right) \bmod H,$$

where $H = 10000019$ is the smallest prime number above 10 million. We group each set of matrix triplets $M \in D_{i,0}$ by their hash $\mathbb{H}(\widetilde{[[\mathbb{G}(M)]]})$ and sort each group lexicographically by their compressed tensor $\widetilde{[[\mathbb{G}(M)]]}$. When searching for a given compressed tensor $\widetilde{\gamma}$ in $D_{i,0}$ if all filters are passed, we only have to binary search $\mathbb{H}(\widetilde{\gamma})$ rather than binary search over all elements of $D_{i,0}$.

- For each profile $P_i \in \mathcal{P}_{i,1}$, we depth-first search over all sets of matrix triplets in $D_{i,i}$ with profile P_i ; when we iterate over the last matrix triplet, we switch to a for loop instead of using recursion, which significantly improves iteration speed.

4. Results

Each test case run on our program was run on a 2019 MacBook Pro, 2.3 GHz 8-Core Intel Core i9, 16 GB 2667 MHz DDR4. In each test case, we ran our program with the -Xmx20g compiler flag. The results are shown in Table 1.

For the first test case, where we searched for a decomposition of $\langle 2, 2, 2 \rangle$ with only cyclic symmetry (i.e. forcing $\sigma = (0, 1)$), we found the following rank-7 decomposition:

$$M = \left[\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \right), \right. \\ \left. \left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right) \right].$$

Our results show that under cyclic+S₃ or cyclic+shift symmetry, there is no integer tensor decomposition of $\langle 3, 3, 3 \rangle$ with rank ≤ 23 . Under cyclic+flip symmetry, there is no such decomposition where each rank-1 tensor has at most 10 nonzero elements, and there is no integer decomposition with rank ≤ 22 where each rank-1 tensor has at most 14 nonzero elements.

However, for cyclic+S₃ symmetry, our program found a rank 25 decomposition; this can occur due to

$$\cup_i \{P_0 + P_1 | P_0 \in \mathcal{P}_{i,0}, P_1 \in \mathcal{P}_{i,1}\}$$

containing some extra profiles besides the ones in \mathcal{P} , potentially allowing some decompositions with rank $> R$ to be found:

$$M = \left[\left(\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \right. \\ \left. \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \right) \right].$$

Our source code can be found at <https://github.com/coolcomputery/Matrix-Multiplication-Tensor-Decomposition/blob/79500ae287090ac08c502425727eb56ccbad86fe/SymmetricMod2.java>.

Sidenote

It is known that $\langle 3, 3, 3 \rangle$ must have rank at least 19 (Blaser, 2003). However, in our test cases, removing all profiles with rank < 19 did not significantly reduce the total search space size (e.g. for $n = 3, R = 22, Z = 14, |D|$ goes from 27,058,964,881,796,966,735 to 27,042,530,413,760,623,197), so we did not use this result in our program.

5. Future Work

We hope to make our program faster by finding clever reductions in the search space and improving running time and memory constant factors. We also hope to try other symmetries that do not fit the form we used, as well as less restrictive symmetries, such as just cyclic symmetry.

6. Acknowledgements

We would like to thank MIT UROP for their support in making this research possible, especially Prof. Virginia Vassilevska Williams for her suggestion to work over mod 2 instead of the integers.

References

[Strassen, 1969] Strassen, Volker. (1969). Gaussian elimination is not optimal. *Numerische Mathematik*, 13, pp. 354–356. Retrieved on March 30, 2022 from <https://doi.org/10.1007/BF02165411>

[Blaser, 2013] Blaser, Markus. (2013). *Fast Matrix Multiplication*. Theory of Computing Library, Graduate Surveys 5, pp. 1-60. Retrieved on March 30, 2022 from <https://theoryofcomputing.org/articles/gs005/>

[Ballard, 2017] Ballard, Grey. (2017). *Discovering Fast Matrix Multiplication Algorithms via Tensor Decomposition*. SIAM Conference on Computational Science & Engineering. Retrieved on March 30, 2022 from <http://perso.ens-lyon.fr/bora.ucar/tensors-cse17/ballard-talk.pdf#10>

Symmetry	n	R	Z	$ D $	$[(D_{i,0} , D_{i,1})]_i$	Lowest rank found	Time (s)
cyclic	2	7	64	11 531 489	[(1 661, 18 726)]	7	0.659
cyclic+ S_3	3	23	729	37 719 901 683	[(231 908, 6 676 860)]	25	103.653
cyclic+shift	3	23	729	3 709 538 433 832 811 175	[(471 136 578, 13 027 173 078), (846 164 340, 4 261 041 877)]	\emptyset	9157.055
cyclic+flip	3	23	10	8 654 054 711 339 819 228	[(999 999 975, 14 315 526 502), (629 321 569, 14 314 120 510)]	\emptyset	12183.273
cyclic+flip	3	22	14	27 058 964 881 796 966 735	[(999 881 752, 88 810 777 763), (96 123 615, 873 415 421)]	\emptyset	18876.996

Table 1: List of results and running times of our search algorithm over various search spaces, as well as the sizes of the sets it uses for searching.

[Alman & Williams, 2021] Alman, Josh & Williams, Virginia Vassilevska. (2021). A Refined Laser Method and Faster Matrix Multiplication. 32nd Annual ACM-SIAM Symposium on Discrete Algorithms. Retrieved on March 30, 2022 from <https://arxiv.org/abs/2010.05846>

[Smirnov, 2013] Smirnov, Alexey V.. (2013). The Bilinear Complexity and Practical Algorithms for Matrix Multiplication. ISSN 09655425, Computational Mathematics and Mathematical Physics, 2013, Vol. 53, No. 12, pp. 1781–1795. Retrieved on March 30, 2022 from <https://cs.uwaterloo.ca/~eschost/Exam/Smirnov.pdf>

[Laderman, 1976] Laderman, Julian D.. (1976). A Non-Commutative Algorithm for Multiplying 3x3 Matrices Using 23 Multiplications. Bulletin of the American Math Society. Volume 82, Number 1, 126-128. Retrieved on March 30, 2022 from <https://www.ams.org/journals/bull/1976-82-01/S0002-9904-1976-13988-2/S0002-9904-1976-13988-2.pdf>

[Courtois, Bard, & Hulme, 2011] Courtois, Nicolas T., Bard, Gregory V., & Hulme, Daniel. (2011). A New General-Purpose Method to Multiply 3x3 Matrices Using Only 23 Multiplications. Retrieved on March 30, 2022 from <https://arxiv.org/abs/1108.2830>[Tichavsky, 2021] Tichavsky, Petr. (2021). Characterization of Decomposition of Matrix Multiplication Tensors. Retrieved on March 30, 2022 from <https://arxiv.org/abs/2104.05323>

[Blaser, 2003] Blaser, Markus. (2003). On the complexity of the multiplication of matrices of small formats. Journal of Complexity, 19, pp. 43-60. Retrieved on March 30, 2022 from [https://doi.org/10.1016/S0885-064X\(02q00007-9](https://doi.org/10.1016/S0885-064X(02q00007-9)

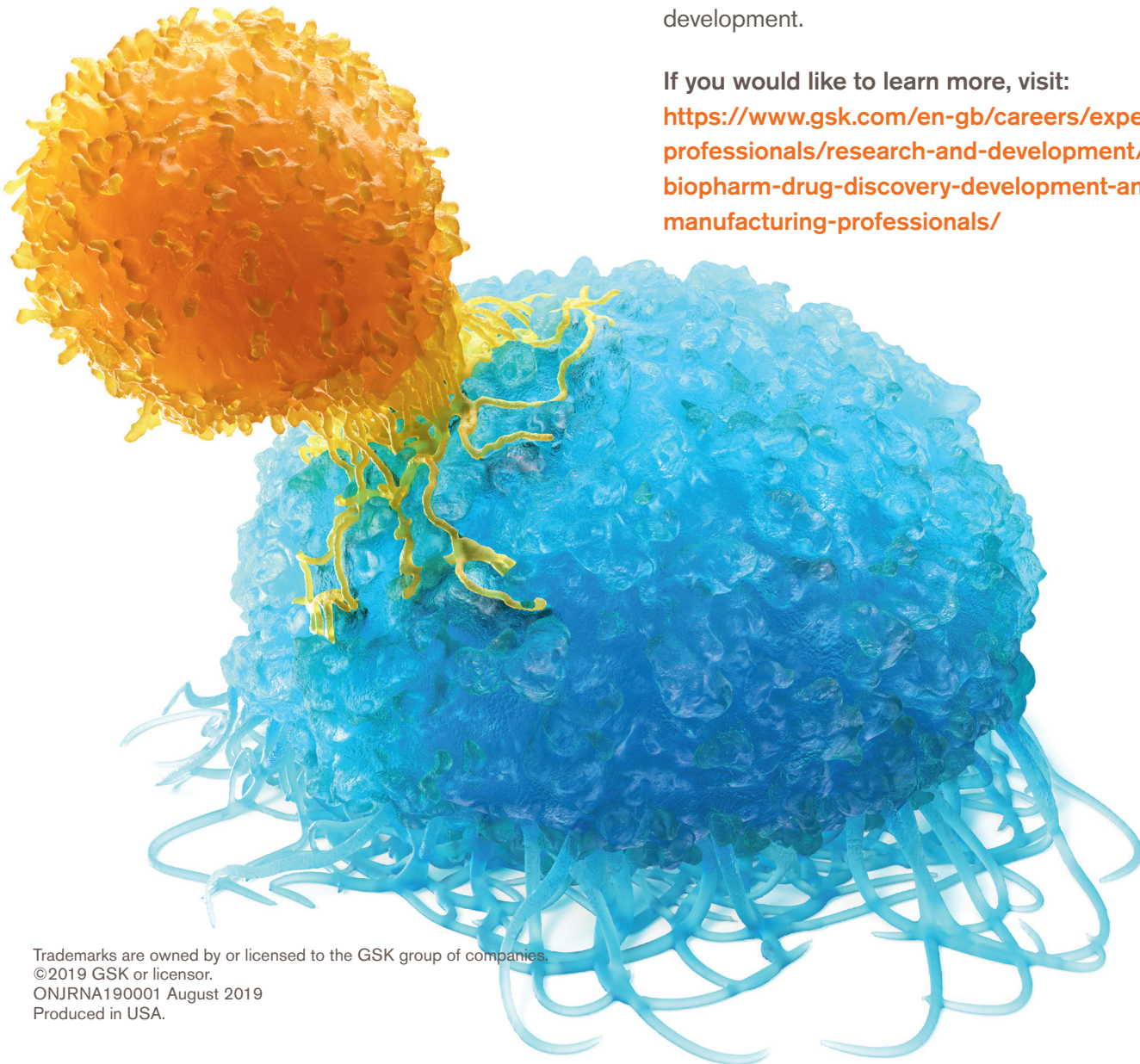
A healthcare company with a special purpose

Our goal is to be one of the world's most innovative, best performing and trusted healthcare companies. We believe that Biopharm medicines are key to achieving our ambitions in immuno-oncology and other immunologically driven diseases. As a global enterprise with activities in the US and the UK, we have a thriving team of Biopharmaceutical and Data Scientists across discovery, development and manufacturing.

We are growing our team to deliver our portfolio of exciting Biopharm medicines to patients around the world and have opportunities available across all disciplines for junior scientists, experienced professionals and leaders in biopharmaceutical development.

If you would like to learn more, visit:

<https://www.gsk.com/en-gb/careers/experienced-professionals/research-and-development/biopharm-drug-discovery-development-and-manufacturing-professionals/>



SCIENCE WILL WIN™

At Pfizer, we pursue groundbreaking science every single day, fueled by one purpose: to develop **breakthroughs that change patients' lives**. We deliver on this purpose by building courageous, diverse teams of experts who contribute to all aspects of what we do and providing them with state-of-the-art resources, technology, and facilities.

Discover how you can help ensure that Science Will Win – including opportunities in Cambridge – at careers.pfizer.com.





Looking for a job that makes a difference?

At Pall Corporation, we are unified by a singular drive: to take on our customers' biggest challenges and deliver complete solutions.

To resolve the critical problems that stand in the way of achieving their goals.

To help safeguard health.

Pall's technologies play key roles in the development and manufacture of life-saving drugs that range from Ebola and COVID-19 vaccines to cancer-curing monoclonal antibodies. Our portfolio helps customers bring vital drugs to market faster at lower costs.

Are you ready to join a team that is making an impact in lives around the world every day?

Our company offers career opportunities at every level, giving you the chance to join our globally diverse team and apply your talents to meaningful work.

R&D — Process Engineering — Operations
Sales — Marketing — and more!

No matter what, no matter where, we innovate and collaborate to deliver the one thing our customers need most:

The unsolvable, **solved.**



If you are looking for a job that makes a difference and gives you a true purpose each day, join our biotech team.

Discover the opportunities waiting for you here!

Visit <http://bit.ly/PALLJobs>



Scan me